Citizen Monitoring and Corporate Environmental Behavior: Evidence from PurpleAir Sensors

Trang Nguyen*

May 2025

Please do not share – comments are welcome.

Abstract

This study examines how citizen-led environmental monitoring impacts corporate pollution behavior. Using the rapid proliferation of PurpleAir sensors — consumer-grade devices installed by local communities to track real-time air quality — I provide robust empirical evidence that citizen monitoring significantly reduces industrial fugitive air emissions reported to the Environmental Protection Agency (EPA). My stacked difference-in-differences analysis shows that facilities exposed to citizen monitoring reduce fugitive emissions by approximately 6.4% within three years. I strengthen the causal interpretation of these findings through instrumental variable analysis that exploits wildfire-induced sensor installations. Mechanism tests indicate that citizen monitoring prompts changes in regulatory oversight — regulators increase inspections and decrease the frequency of routine stack tests, suggesting an adaptation of enforcement strategies. The effectiveness of citizen monitoring is strongest where internal oversight is weaker and where pollution levels are higher, highlighting its potential to enhance existing regulatory frameworks, particularly where formal oversight may be less comprehensive. These findings demonstrate how grassroots technological advancements democratize monitoring, reshaping corporate accountability and environmental governance.

Keywords: citizen monitoring, corporate compliance, environmental disclosure, information asymmetry, regulatory enforcement

JEL classification: Q53, Q58, K32, M14

*Email: trnguyen@hbs.edu (Trang Nguyen). I am deeply grateful for the support and advice I received from my dissertation committee: Jonas Heese (chair), Shirley Lu, George Serafeim, and Charles Wang. I thank Peter Tufano, Aliya Korganbekova, Luis Ballesteros (discussant), Quan Le, conference participants at the 7th Greater Boston Corporate Governance Workshop, the 2025 Rutgers Accounting Doctoral Symposium, and seminar participants at Harvard Business School and the Climate and Sustainability Impact Lab at Harvard. I thank Andrew White from PurpleAir for his assistance with accessing the PurpleAir API and for providing helpful context on the institutional details of PurpleAir sensors. All errors are my own.

1 Introduction

Federal air monitoring networks remain critically sparse despite their regulatory importance. As of 2021, approximately 120 million Americans live in counties without any federal air quality monitors, creating significant "monitoring deserts" where communities lack basic information about their air quality (Coursen, 2021). Even in monitored areas, evidence suggests monitors are often strategically placed in cleaner locations (Grainger and Schreiber, 2019), potentially underestimating true pollution exposure, while many collect data only once every three or six days (Zou, 2021). These limitations — in spatial coverage, strategic placement, and temporal resolution — make it exceptionally difficult to detect localized pollution events or establish accurate pollution patterns. This insufficient infrastructure for monitoring industrial emissions represents a fundamental gap in environmental governance (Babich, 2018).

Citizen-deployed technologies have emerged to address these monitoring gaps. PurpleAir has become the most prominent community sensor network in the United States, expanding from just 138 devices in 2016 to over 10,700 active monitors by 2024 — substantially outpacing regulatory-grade equipment deployment. These affordable, internet-connected devices provide frequent measurements and automatically upload readings to a publicly accessible real-time map. The Environmental Protection Agency (EPA) has formally recognized these sensors' utility by incorporating their data into official air quality resources, allowing communities to generate high-resolution air quality data in previously unmonitored areas (EPA, 2022). While this expansion enhances environmental transparency, the extent to which citizen-based monitoring can influence corporate environmental behavior remains an open empirical question, reflecting the tension between growing transparency and the risk that such data lacks institutional credibility.

The effect of PurpleAir sensor deployment on facility-level air emissions is theoretically ambiguous. On one hand, citizen-deployed sensors could strengthen environmental monitoring by reducing information asymmetries between facilities and communities (Konar and Cohen, 1997). This enhanced transparency could enable detection of emission spikes and create pressure for improved compliance (Earnhart and Friesen, 2013). Additionally, facilities might anticipate heightened regulatory scrutiny following sensor deployment, as agencies could leverage citizen-generated data to prioritize inspection and enforcement actions (Colmer et al., 2023). Such monitoring might prove especially consequential where existing oversight is limited — for instance, for facilities with less sophisticated internal controls that might otherwise escape detection (Blackman et al., 2004). Furthermore, the impact may be particularly pronounced for high-emitting facilities, as these operations create more detectable pollution signatures that citizen monitors can identify and attribute, thereby focusing public and regulatory attention on the most significant contributors (Olken, 2007; Blackman et al., 2004).

Alternatively, PurpleAir sensors may prove ineffective as monitors of industrial air pollution for multiple reasons. First, many firms already face strong incentives to comply with environmental regulations due to potential legal penalties and significant reputational costs from violations (Karpoff et al., 2005; Konar and Cohen, 2001), potentially diminishing the marginal benefit of additional monitoring (Shimshack and Ward, 2008). Second, low-emitting facilities may remain largely unaffected by citizen monitoring, as their minimal pollution levels fall below detection thresholds or public concern thresholds that would trigger community action (Castell et al., 2017; Diviacco et al., 2022a,b). Third, these consumer-grade devices lack the precision, calibration, and institutional authority of regulatory monitors, potentially undermining their credibility when attributing emissions to specific sources (Duflo et al., 2013). Finally, proximity to highways or industrial clusters may contaminate readings, obscuring the true source of pollution and reducing individual facilities' accountability (Fowlie et al., 2019). Whether PurpleAir sensor deployment meaningfully impacts facility-level emissions thus remains to be empirically tested.

I answer this question by leveraging data from two primary sources to examine how PurpleAir sensor presence affects facility-level emissions. First, I construct a dataset combining PurpleAir sensor locations with facility coordinates to determine sensors within specified proximity of each facility. Second, I use the EPA's Toxic Release Inventory (TRI) data, which distinguishes between fugitive emissions (unintended releases from equipment leaks and other non-point sources) and stack emissions (releases through controlled vents). Since PurpleAir sensors primarily detect fugitive emissions, I select fugitive air as my main dependent variable and use stack air as a falsification test.

I employ a stacked difference-in-differences framework exploiting the staggered installation of PurpleAir sensors from 2016 to 2020. The treatment group consists of facilities with at least one PurpleAir sensor within a 5-mile radius, whereas the control group consists of facilities with no sensors within a 10-mile radius. All specifications include facility$\times$cohort fixed effects to control for time-invariant facility characteristics and industry-year$\times$cohort fixed effects to account for industry-specific temporal trends. I also add time-varying county-level controls to address potential confounding factors and adjust for spatial dependence in emissions among nearby facilities. Additionally, I apply entropy balancing on pre-period total releases to ensure comparability between treatment and control facilities prior to sensor installation.

My main results show that facilities with at least one PurpleAir sensor within a 5-mile radius see a 6.4% reduction in fugitive air emissions over the subsequent three years. For the average facility, this represents a decrease of approximately 365 pounds per year — more than 18 times the median facility's total emissions of just 20 pounds. This substantial reduction appears concentrated among higher-emitting facilities where environmental impact is greatest, with stronger effects when sensors are positioned closer to facilities or when multiple sensors are present. To ensure that pollution attribution challenges do not drive the results, I exclude facilities near highways or in industrial clusters. The findings remain consistent, indicating that the main effects are not artifacts of measurement contamination.

To mitigate endogeneity concerns that citizens may install sensors in already-polluted areas, I employ an instrumental variable approach following Borgschulte et al. (2024b). I use the number of smoke days as an instrument for PurpleAir sensor installation, as wildfire smoke constitutes a plausibly exogenous shock that increases the probability of sensor deployment without directly affecting facility-level emission behavior. The first-stage regression confirms that wildfire smoke exposure significantly increases the likelihood of treatment, satisfying the instrument relevance condition. The second-stage estimates corroborate my main findings: Wildfire smoke-induced PurpleAir sensors lead to decreased next year's fugitive air emissions at the facility level, strengthening the causal interpretation of my results.

My analysis of the mechanisms underlying citizen monitoring's effect on emissions reveals notable shifts in regulatory behavior. Specifically, air inspections by state agencies and the EPA increase significantly when PurpleAir sensors are installed nearby. This suggests that real-time citizen data enables regulators to identify and respond to potential emission irregularities more efficiently. At the same time, state agencies conduct fewer scheduled air stack tests, potentially substituting targeted, data-driven inspections for routine point-source evaluations. This shift in regulatory approach — combining heightened inspection intensity with reduced reliance on standardized tests — reflects how citizen-generated environmental data reshapes oversight strategies.

Further cross-sectional analyses reveal that the impact of PurpleAir sensors is heterogeneous. Emission reductions are significantly larger at facilities relying on less sophisticated internal emissions estimation methods, suggesting that citizen monitoring is particularly effective where internal oversight may be weaker. Similarly, sensors induce greater emission reductions at facilities with higher baseline pollution levels. These patterns indicate that citizen monitoring most effectively disciplines corporate behavior where existing oversight is less stringent and pollution is more severe, thereby targeting facilities that likely pose greater environmental and public health risks.

This study makes several contributions to the literature on environmental compliance and the intersection of public and private enforcement mechanisms. First, I provide systematic evidence that citizen-driven environmental monitoring, enabled by consumer-grade PurpleAir sensors, effectively reduces corporate fugitive air emissions. This finding reveals a nuanced relationship — PurpleAir sensors act as substitutes for formal air pollution data collection but function as complements to regulatory enforcement. This dual role extends existing research in an important direction: prior accounting and economic studies predominantly examine formal regulatory oversight and its limitations (Shimshack and Ward, 2005; Earnhart and Friesen, 2013), while complementary work shows that private governance mechanisms often reinforce state oversight (Short and Toffel, 2008; Potoski and Prakash, 2005; Fowlie et al., 2012). My study demonstrates how citizen-driven monitoring fills critical information gaps that constrain formal enforcement while enhancing regulatory effectiveness, contributing directly to our understanding of public-private enforcement interactions in environmental contexts.

Second, this research examines how ordinary citizens can function as effective monitors in corporate governance. The corporate governance literature has established that informal monitors — entities beyond regulators, courts, and auditors — significantly influence firm behavior. Media watchdogs (Miller, 2006; Dyck et al., 2008; Jiang and Kong, 2024), NGOs (Daubanes and Rochet, 2019), and whistleblowers (Dyck et al., 2010; Bowen et al., 2010) all raise the expected cost of misconduct. Yet ordinary citizens acting independently remain largely unexplored as potential monitors. Existing studies examine how citizens influence firm behavior through complaints (Colmer et al., 2023), social-media or other online appeals (Heese and Pacelli, 2024; Buntaine et al., 2024), but none investigate settings where citizens themselves generate the underlying quantitative data used for monitoring.

Third, this research demonstrates how technological innovation democratizes environmental oversight. Prior literature has largely focused on how regulators or corporations leverage costly satellite imagery for improved disclosure or internal controls (Gu et al., 2023). In contrast, I show how citizens independently deploy affordable sensor technologies to influence corporate environmental behavior, transforming accountability by enabling citizens to become data producers rather than merely data consumers.

Finally, to my knowledge, this is the first research to separately analyze fugitive air emissions rather than relying on aggregate toxic release metrics. While EPA monitoring primarily focuses on stack emissions (Babich, 2018), fugitive emissions contribute similarly to health risks despite being only one-third the volume (EPA, 2018). This study demonstrates how citizen monitoring can effectively reduce these under-regulated but equally harmful pollution sources, suggesting policymakers should promote access to monitoring technologies, integrate citizen-generated data into regulatory frameworks, and focus agency oversight where citizen monitoring is absent.

2 Institutional Background and Literature Review

2.1 Traditional Air Quality Monitoring and Its Limitations

Air quality monitoring in the United States has long been a critical function of state and federal agencies, most notably the Environmental Protection Agency (EPA). These governmental bodies have developed extensive networks of monitoring stations aimed at measuring air pollutants and ensuring compliance with established air quality standards. However, despite their importance, these traditional air quality monitoring systems are subject to several notable limitations that affect their overall efficacy.

One major limitation is sparse spatial coverage. Of the 3,100 counties in the U.S., only about 21% have monitors reporting fine particulate matter $PM_{2.5}$, and nearly half of these counties have just a single monitor (Sullivan et al., 2018). In such cases, it is standard practice to assume that readings from one monitor are representative of the entire county — an assumption that overlooks pollution heterogeneity. Nationally, roughly 120 million Americans live in counties without any EPA $PM_{2.5}$ monitoring (Coursen, 2021). Even where monitors exist, the typical density — about one monitor per 750 square miles — offers insufficient granularity to detect neighborhood-level pollution hot spots (Zivin et al., 2024; Desouza and Kinney, 2021).

Second, the placement of monitors is not random and maybe influenced by political or economic considerations. State regulators exercise discretion in siting monitors and may strategically locate them in cleaner areas to avoid triggering regulatory thresholds that could constrain industrial development. A growing body of economic research documents evidence of such "gaming" behavior, both in monitor siting (Grainger and Schreiber, 2019) and in timing, where polluters may adjust emissions in anticipation of monitoring schedule (Zou, 2021).

Finally, existing monitors often lack the temporal resolution necessary to respond to real-time public health risks. Many $PM_{2.5}$ monitors sample only once every three or six days, limiting their ability to capture short-term pollution spikes (Zou, 2021).

Taken together, these limitations highlight the structural weaknesses of traditional monitoring systems and motivate the need for complementary approaches — such as citizen-deployed sensor networks — that offer both greater spatial and temporal responsiveness.

2.2 Institutional Background on PurpleAir Sensors

2.2.1 Technology, Performance, and Adoption

In response to the spatial and temporal limitations of traditional regulatory air quality monitoring, citizen-led technologies have emerged as important alternatives. The PurpleAir sensor network, a widely adopted, low-cost system for measuring $PM_{2.5}$, powerfully exemplifies this trend. Its origins are deeply rooted in citizen-driven environmental concern: founded in 2015 by engineer Adrian Dybwad, who was initially motivated by heavy dust pollution from a gravel mine near his Utah home. Seeking an objective measure of the air quality he and his neighbors were inhaling, Dybwad began developing and distributing sensors. After he initially gave away roughly 80 devices, the ensuing community interest and demand spurred the formal creation of PurpleAir. As Dybwad himself noted, while technology like sensors can "get the ball rolling" by raising awareness, effective change often requires further civic engagement, such as community participation in local governance — a path he successfully pursued regarding the mine near his home (Newitz, 2020).

From these community-centered beginnings, PurpleAir has grown rapidly into one of the world's largest decentralized environmental sensor platforms, with widespread adoption by residents, schools, researchers, and community organizations. Between 2016 and 2024, approximately 21,000 PurpleAir sensors were deployed across the contiguous United States, far exceeding the roughly 1,000 $PM_{2.5}$ monitors in the EPA's regulatory network. These compact devices, which are typically installed by users as shown in Appendix 1, retail for between \$140- \$300, or roughly one-hundredth the cost of a regulatory-grade instrument (Desouza and Kinney, 2021). Research suggests these sensors explicitly substitute for formal monitoring infrastructure — Zivin et al. (2024) find that a census tract located 1 mile closer to an EPA monitor has 0.8% fewer outdoor PurpleAir monitors. Once installed and connected to the internet, the sensors provide frequent, real-time data uploads to PurpleAir real-time map, unless the owner actively opts out.

PurpleAir sensors are equipped with laser particle counters that continuously measure $PM_{2.5}$, offering both high spatial and temporal resolution. They are especially suitable for identifying short-term pollution events and local hotspots (Sun et al., 2022). However, their low-cost design introduces challenges. Sensor degradation, high humidity, and temperature variation can all bias results over time (Desouza and Kinney, 2021). These limitations are well documented, but evaluation studies have also shown promising performance. The Air Quality Sensor Performance Evaluation Center, operated by California's South Coast Air Quality Management District, found high correlations between PurpleAir and federal reference monitors in controlled lab settings (Krebs et al., 2021; AQMD, 2019).

To improve data comparability, EPA researchers developed a public correction algorithm for

PurpleAir sensors, adjusting for environmental biases — particularly humidity — that affect optical readings (EPA, 2022). This correction enables the integration of PurpleAir data into the EPA's AirNow Fire and Smoke Map, which provides real-time pollution information during wildfire events (see Appendix 2 for an illustration of this integration).[1] The corrected data are now also used by Google Maps, South Coast AQMD, NASA, Windy.com, and Weather Underground to support public mapping platforms, satellite calibration, and hyper-local pollution tracking.[2] These diverse integrations across government, scientific, and commercial users reflect growing recognition of PurpleAir's value in expanding access to real-time air quality information and supporting more responsive, community-level environmental decision-making.

2.2.2 Policy Use and Legal Standing

While PurpleAir has been embraced as a public engagement tool, its role in formal regulatory processes remains debated. The EPA promotes PurpleAir through sensor loan programs and community grant initiatives, including a $50 million national program launched in 2021 to expand air-monitoring access in underserved communities most affected by pollution and COVID-19 (EPA, 2021). These investments reflect a federal acknowledgment of the promise of citizen-generated data in addressing gaps in environmental monitoring.

Even though PurpleAir sensors are not certified for regulatory enforcement, multiple states have taken steps to further restrict their legal standing. West Virginia, Louisiana, and Kentucky have enacted laws limiting the admissibility of community-collected air quality data, effectively preventing PurpleAir sensor readings from being used in courts or enforcement actions. Appendix 3 summarizes these recent laws and their implications for citizen-led environmental monitoring.

[1] The Fire and Smoke Map is available at `https://fire.airnow.gov/`, accessed May 8, 2025.

[2] PurpleAir Data Use Cases is available at `https://www2.purpleair.com/pages/purpleair-data-use-cases`, accessed May 8, 2025.

In short, PurpleAir sensors represent a promising and rapidly evolving, yet institutionally unsettled, model of decentralized environmental governance. Their real-time, publicly accessible data streams expand transparency, but their influence on policy and enforcement depends on legal and institutional frameworks that are willing to acknowledge and integrate citizen science.

2.3 Literature Review

2.3.1 Public and Private Enforcement in Environmental Compliance

A growing body of work studies how public oversight and private-sector initiatives jointly shape environmental compliance. Classical deterrence models stress inspections, monitoring, and monetary penalties administered by regulators (Shimshack and Ward, 2005; Earnhart and Friesen, 2013), yet theoretical work highlights the resource constraints and information gaps that can limit agencies' reach (Schantl and Wagenhofer, 2020). In response, researchers have examined non-state mechanisms that can reinforce government efforts. Voluntary management standards such as ISO 14001 are associated with fewer months out of Clean Air Act non-compliance, indicating that certification helps firms meet regulatory expectations rather than replacing them (Potoski and Prakash, 2005). Under the U.S. EPA's Audit Policy, facilities are most likely to self-report violations shortly after they have been inspected or cited, showing that regulatory scrutiny stimulates private disclosure and remediation (Short and Toffel, 2008). Evaluations of the NO_x Budget Program likewise reveal that rigorous auditing remains essential for cap-and-trade markets to deliver large pollution reductions (Fowlie et al., 2012).

Despite these insights, one important frontier is largely unexplored: informal, citizen-driven monitoring made possible by cheap, networked sensors. Because communities can now collect high-frequency pollution data independent of agency infrastructure, these technologies have the potential both to substitute for official data collection and to complement enforcement by spotlighting violations regulators might otherwise miss.

2.3.2 Informal Monitoring and Decentralized Enforcement

Informal monitors — entities beyond traditional regulatory authorities — significantly influence corporate behavior by raising the expected costs of misconduct. Media coverage plays a critical role in detecting and deterring corporate malfeasance (Miller, 2006; Dyck et al., 2008; Core et al., 2008), with multiple studies documenting the environmental consequences of declining local news coverage. Both Jiang and Kong (2024) and Heese et al. (2022) show that newspaper closures and the resulting drop in local media scrutiny lead to significant increases in toxic releases and environmental violations, underscoring the disciplining role of press oversight. NGO activities (Daubanes and Rochet, 2019) and whistleblower actions (Dyck et al., 2010; Bowen et al., 2010) have also been shown to effectively detect and deter corporate misconduct.

Community-based monitoring represents another form of informal oversight that complements formal enforcement, particularly in settings where regulatory capacity is constrained. Pargal and Wheeler (1996) demonstrate that local community pressure significantly influenced industrial pollution in Indonesia, while Olken (2007) show that grassroots monitoring reduced corruption in Indonesian village road projects. However, existing research has primarily examined how citizens leverage information either publicly disclosed by regulators or observed directly through visible pollution events. Recent studies have explored citizen influence through formal complaints (Colmer et al., 2023), social media or other online appeals (Heese and Pacelli, 2024; Buntaine et al., 2024), but have not investigated settings where citizens themselves generate quantitative environmental data.

2.3.3 Technological Advances and Citizen Monitoring

Recent technological innovations have significantly expanded decentralized environmental monitoring capabilities. Affordable, internet-connected sensors now allow citizens to independently measure and share local environmental conditions in real-time, promising increased data density and enhanced transparency in previously unmonitored areas. The accounting and finance literature has primarily focused on how regulators and corporations leverage sophisticated technologies like satellite imagery (Kang et al., 2021; Gu et al., 2023), with far less less attention to how accessible consumer-grade technologies might reshape environmental governance.

A growing body of work explores citizen science initiatives and sensor networks, largely focusing on technological aspects, data validation, and adoption patterns (Morawska et al., 2018; Snyder et al., 2013). Existing research on PurpleAir sensors has examined demographic factors influencing deployment and pollution concentration correlations (Zivin et al., 2024; Desouza and Kinney, 2021; Sun et al., 2022). However, despite this attention to sensor distribution, there remains a critical gap in empirical evidence on whether citizen monitoring fundamentally influences regulated entities' behavior.

This study addresses this gap by analyzing how PurpleAir sensor deployment affects corporate fugitive emissions—pollution that often escapes formal regulatory attention but is detectable by ground-level sensors. This provides the first systematic evidence on this behavioral link, extending the literature on informal monitoring and technological transparency by demonstrating how community-led data collection can influence corporate environmental performance.

2.4 Hypothesis Development

The preceding review highlights the established role of informal monitors in influencing corporate behavior and the transformative potential of new technologies in expanding citizen oversight. Building on this, I hypothesize that citizen-deployed PurpleAir sensors influence corporate behavior primarily through greater environmental transparency that can prompt regulatory scrutiny and mobilize community activism. These pressures increase the likelihood that real-time pollution spikes will be detected and publicly attributed to specific facilities, in turn encouraging proactive reductions in fugitive emissions. In essence, firms aware of nearby sensor installations may act preemptively to avoid negative publicity or penalties associated with elevated pollution levels, thus strengthening environmental compliance. Therefore, I posit the following hypothesis regarding the primary outcome:

H1: Facilities exposed to nearby PurpleAir sensors will significantly reduce fugitive air emissions reported to the EPA, relative to facilities without nearby sensors.

If H1 is supported, it suggests that citizen monitoring through affordable sensor technology complements traditional regulatory oversight and expands accountability mechanisms for corporate environmental behavior. To understand the pathways through which this effect operates, I further hypothesize about specific mechanisms.

Increased transparency from citizen sensors may alter the behavior of formal regulators. If regulators use citizen-generated data to inform their oversight activities, we should observe changes in their enforcement patterns. Specifically, regulators might increase inspections in areas with new sensor coverage, as the data provides insights into potential emission events or areas of concern.

They might also shift resources, potentially adjusting routine activities like pre-scheduled stack tests if sensor data allows for more targeted and responsive oversight. This leads to my second hypothesis:

H2: The installation of PurpleAir sensors near a facility will lead to (a) an increase in regulatory inspections by state agencies, and (b) a decrease in routine stack tests, as regulators adapt their oversight strategies.

The impact of citizen monitoring is unlikely to be uniform across all facilities. Facilities with less sophisticated internal emissions monitoring might be more uncertain about their own compliance and thus more responsive to external visibility. Similarly, high-emitting facilities are more likely to be scrutinized once their pollution becomes more apparent. This leads to my final hypothesis regarding heterogeneous effects:

H3: The emission-reducing effect of PurpleAir sensors will be stronger for: (a) facilities with less sophisticated internal emissions estimation methods. (b) facilities with higher baseline level of total releases.

If supported, these hypotheses would offer a nuanced examination of how citizen monitoring operates. By generating localized, real-time environmental data, citizen sensors enhance the information available to both regulators and the public. The subsequent tests on regulatory responses and heterogeneous impacts aim to demonstrate how this enhanced transparency ultimately complements formal regulatory frameworks, improving environmental outcomes by targeting oversight where it is most needed or where firms are most susceptible to public pressure.

3 Data and Sample

3.1 Facility-level Emissions Data

I obtain facility-level emissions data from the EPA's Toxic Release Inventory (TRI) program for 2013–2023. The TRI is a comprehensive database that tracks toxic chemical management by industrial and federal facilities, requiring eligible sites to report their annual releases and waste practices, including on-site air emissions, water discharges, land disposals, as well as off-site transfers. Importantly, the data distinguishes between two types of on-site air emissions: *fugitive* releases (uncontrolled emissions from leaks, evaporative losses, and diffuse sources) and *stack* releases (emissions from controlled vents, smokestacks, and other point sources).

While most prior studies focus on *Total Toxic Releases* as their primary dependent variable (Akey and Appel, 2021; Xu and Kim, 2022; Thomas et al., 2022; Duchin et al., 2025), I examine fugitive air emissions. PurpleAir sensors operate by detecting pollutants that physically pass through the device.[3] As these sensors are typically installed at residential heights (5–7 feet above ground, see Appendix 1), they are consequently most effective at capturing pollutants present at these ground-level altitudes. Fugitive emissions tend to disperse at ground level and remain concentrated near the source, whereas stack emissions, released from heights often exceeding 30 feet, disperse vertically and over wider areas.[4] Furthermore, the different regulatory landscapes for these emission types make fugitive emissions a particularly salient focus for studying the impact of citizen monitoring. Stack emissions are subject to formal reporting and abatement measures, while

[3] Interview with PurpleAir representative, September 2024.

[4] Consequently, surrounding populations are less likely to be directly exposed to pollutants from stack emissions than from fugitive emissions, making the latter more detectable by community-based sensors like PurpleAir and more relevant for evaluating localized monitoring impacts (EPA, 2018).

fugitive emissions are often under-monitored and under-regulated (Babich, 2018), despite posing significant risks to nearby communities (Scheer, 2025). This disparity suggests that citizen-led monitoring has greater potential to provide novel information and influence corporate behavior concerning fugitive emissions, where formal oversight may be less comprehensive and the 'information gap' filled by sensors is larger. Thus, fugitive emissions serve as my primary dependent variable, with stack emissions used as a falsification test.[5]

Because PurpleAir sensors do not offer a complete pre-treatment time series, and satellite data lack the necessary facility-level attribution (Holloway et al., 2021), I rely on TRI data, which span the full study period. Although self-reported, TRI data are subject to annual EPA quality checks that flag inconsistencies and prompt corrections, enhancing their credibility (EPA, 2013).

For the sample selection, I exclude facilities that report zero air releases under TRI, thereby focusing on sites where PurpleAir monitors can plausibly influence regulatory scrutiny or community attention. Panel A of Table 1 summarizes the fugitive air and stack air emissions in this study. For each facility-year observation, I aggregate the total pounds released across all chemical substances. The final sample includes 226,508 facility-year observations. On average, pounds released from fugitive air emissions represent approximately 24% of those from stack emissions, aligning with national trends over the past decade (EPA, 2018).[6] The large standard deviation for both reflect a highly right-skewed distribution.

[5]Dependent variable selection is motivated by sensor detection capabilities and regulatory oversight differences between fugitive and stack emissions.

[6]The ratio of fugitive to stack air emissions has gradually increased - from about one-fourth in 2013 to about one-third in 2022 - indicating a slow rise in the relative contribution of fugitive emissions over time.

3.2 Citizen-Installed PurpleAir Sensor Data

To construct a complete dataset of all publicly accessible outdoor PurpleAir sensors installed across the United States, I collect sensor data using the PurpleAir API. The dataset covers the years 2016 through 2023 and includes all sensors that were publicly registered during this period.

Each sensor record includes a unique identifier, geographic coordinates, installation date, and removal date (if applicable). PurpleAir assigns a "position rating" to indicate location accuracy; to ensure accurate distance measurements between sensors and facilities, I retain only those sensors with the highest accuracy rating (5). To ensure stability in the sensor data, I exclude sensors that remain active for less than one month.

Using these geographic coordinates, I geocode each sensor to identify its state, county, and ZIP code, allowing me to restrict the sample to the contiguous United States. For each sensor and year in the sample period, I then determine whether the sensor falls within specific radii (ranging from one to five miles) of industrial facilities. Finally, I aggregate these sensor-facility-year observations to the facility-year level, calculating the total number of active PurpleAir sensors within each distance threshold for every facility in each year.

3.3 Wildfire Smoke Data

To construct an instrument for my instrumental variable analysis, I develop a facility-level measure of wildfire smoke exposure. I adapt the methodology of Borgschulte et al. (2024b) to construct a facility-level measure of wildfire smoke exposure. Their approach draws on daily smoke plume polygons identified by the National Oceanic and Atmospheric Administration's Hazard Mapping

System (HMS).[7] The HMS uses geostationary satellites to observe fires and smoke emissions across the contiguous United States, producing high-resolution imagery multiple times per day (Ruminski et al., 2006). Based on these observations, smoke analysts create georeferenced polygons that represent the spatial extent of visible wildfire smoke plumes.

While Borgschulte et al. (2024b) define exposure at the county level, I refine the measurement to the facility-level to suit the level of analysis in my paper. Specifically, a facility is considered exposed on a day if a one-mile radius around it is fully covered by a detected smoke plume. I then aggregate daily exposures to the annual level and compute the total number of wildfire smoke days per facility per year.[8]. On average, a facility is exposed to wildfire smoke for approximately 42 days per year, with a median of 36 days. This corresponds to roughly one to one and a half months of annual exposure.

3.4 Regulator Monitoring Data

To analyze regulatory responses following the installation of PurpleAir sensors, I examine which level of government — local, state, or federal — is responsible for enforcement actions. I obtain inspection histories of facilities from the EPA's Enforcement and Compliance History Online (ECHO) database. This dataset provides comprehensive compliance and enforcement information for regulated facilities, including the identities of inspected facilities, the type and timing of inspections, and the agency responsible. While the EPA oversees the national environmental enforcement framework, it delegates much of the routine permitting, monitoring, and enforcement activity to state air agencies, and in some cases, to local regulators (EPA, 2015). State agencies are

[7] Miller et al. (2021) originally developed the daily smoke exposure data, while Borgschulte et al. (2024b) linked these data to economic outcomes.

[8] To validate this approach, I also construct a county-level measure and confirm it aligns with county smoke-day counts reported in Borgschulte et al. (2024a) and Borgschulte et al. (2024b)

primarily responsible for implementing and enforcing air quality standards for common pollutants, including issuing permits, conducting inspections, and enforcing compliance with facility-level emission limits. Consistent with this institutional structure, the data show that state agencies carry out the overwhelming majority of regulatory activity, including 75% of all inspections and 61% of stack tests.

3.5 Other Data

To further contextualize the relationship between PurpleAir sensor deployment and local economic or infrastructural conditions, I supplement facility data with county-level demographic and geographic information. I obtain county-level data on population and unemployment rates from the U.S. Census Bureau. In addition, I use TIGER/Line shapefiles to calculate the distance from each facility to the nearest major highway. On average, facilities are 11.3 miles away from the nearest highway. In alternative specifications, I exclude facilities within 0.5, 1, or 2 miles of a major highway to ensure that the measured emissions are not driven by nearby traffic (Karner et al., 2010; Shapiro and Walker, 2018).

Furthermore, to account for potential spatial interdependencies between facilities, I incorporate control variables that capture local spatial spillovers in air emissions, following Delgado and Florax (2015). I assume that interference is negligible beyond 10 miles (i.e., SUTVA holds for facilities more than 10 miles apart). I compute continuous spatial weights using a piecewise linear decay function — weights decay linearly for facilities within 10 miles and are set to zero otherwise. For each facility in a given year, the spatial lag (a weighted average of log emissions across neighboring facilities) is computed using these weights.[9]

Next, I exclude facilities located in industrial clusters, where disentangling pollution attributable to a single facility is challenging. Clusters are defined using alternative spatial thresholds (0.5 and 1 mile) and density criteria (at least 3 or 5 facilities within the specified radius). On average, between 64% and 96% of facility-year observations are not located in a cluster, depending on the specific definition used.

Finally, to investigate potential heterogeneity in the impact of citizen monitoring, I construct two facility-level variables that capture differences in emissions reporting practices and baseline pollution levels. First, I use facilities' TRI Form R disclosures to measure the quality of emissions estimation. Facilities report their method for estimating emissions (e.g., direct measurement, site-specific factors, or generic emission factors), and I create an indicator, *Low Quality Emissions Estimation*, which equals 1 for facilities relying on generic emission factors (method code E1) and 0 otherwise. Approximately 70% of facility-year observations fall in this category, reflecting a widespread reliance on standardized, less precise methods. Second, I define an indicator, *High Emitter*, which equals 1 for facilities with above-median fugitive air emissions during the pre-treatment period, thereby capturing heterogeneity in baseline pollution that may drive differential responses to monitoring.

3.6 Sample Construction and Summary Statistics

For the stacked difference-in-differences (DiD) framework, I construct a panel dataset comprising 226,508 facility-year observations from 2013 to 2023. Table 1, Panel A, presents summary

[9]For example, if Facility A has three neighbors within 10 miles—with log emissions of 10, 12, and 14 and weights of 0.8, 0.5, and 0.2, respectively, then the spatial lag is computed as follows: $(10\times0.8+12\times0.5+14\times0.2)/(0.8+0.5+0.2) = 11.27$.

statistics for the key variables used in the baseline analysis. Panel B reports descriptive statistics for the instrumental variable analysis, showing that the average facility experiences full wildfire smoke coverage for 42 days per year.

4 Empirical Strategy and Results

4.1 Empirical Strategy

I employ a stacked difference-in-differences (DiD) framework that exploits the staggered installation of PurpleAir sensors between 2016 and 2020. On average, each cohort includes approximately 1,588 treated facilities and 11,638 control facilities. Figure 1 shows the geographical distribution of treated versus control facilities in the contiguous United States. I stack these cohorts across all years and estimate the following facility-cohort-year panel regression:

$$\textit{Fugitive Air Emissions}_{f,c,t} = \beta_0 + \beta_1 \textit{PA Sensor within 5 mi} \times \textit{Post}_{f,c,t} + \sum \textit{Controls} + \varphi + \tau + \varepsilon_{f,c,t} \tag{1}$$

The main dependent variable is *Fugitive Air Emissions*$_{f,c,t}$, measured as the natural logarithm of one plus total on-site fugitive air emissions reported by facility f in cohort c during year t. *PA Sensor within 5 mi* $\times$ *Post* is an indicator variable that equals one for treated facility-year observations in the 3 years after at least one PurpleAir sensor is installed within a 5-mile radius of the facility and zero in the three years before PurpleAir sensor installation. The control group consists of facilities with no PurpleAir sensors within a 10-mile radius. The model includes facility-by-cohort fixed effects (φ) and industry-year-by-cohort fixed effects (τ) to account for both time-invariant facility characteristics and broader industry-specific shocks over time. The null hypothesis is that

$\beta_1 = 0$ — that is, the installation of a PurpleAir sensor within close proximity of a facility has no impact on that facility's fugitive air emissions. The first difference is the change within the facility before and after the presence of PurpleAir sensor, and the second difference is the difference between the treated facilities and the control facilities. In other words, this design allows me to identify the effect on facilities with at least one PurpleAir sensor within a 5-mile radius using the time trend of the control facilities as a counterfactual.

I incorporate several time-varying *Controls* at the facility and county levels. At the facility level, I control for total emissions in the prior year to account for historical pollution patterns that may influence both current emissions and the likelihood of nearby sensor installation. At the county level, I include population and unemployment rates to adjust for macroeconomic conditions that could influence industrial activity. I also include a variable capturing spatial dependence in emissions among nearby facilities to account for regional pollution spillovers. Finally, I apply entropy balancing of pre-treatment total emissions to ensure covariate balance between treated and control facilities, enhancing comparability prior to sensor deployment. All control variables are winsorized at the 1st and 99th percentiles.

4.2 Main Result

I present stacked DiD estimates examining the effect of Purple Air sensor installation on facility-level fugitive air emissions in Table 2. The key coefficient of interest is on the interaction term *PA Sensor within 5 mi* $\times$ *Post*, which captures the average treatment effect on the treated (ATT). The columns reflect a progression from baseline models with minimal controls to fully saturated specifications addressing multiple sources of confounding.

In columns 1-3, I begin with less saturated models. I observe that sensor presence is associated with significant reductions in fugitive emissions. The baseline model without controls shows substantial post-treatment reductions, though it also reveals pre-existing differences between treated and control facilities, underscoring the need to account for facility-specific heterogeneity. Adding facility-fixed effects addresses this concern by controlling for time-invariant facility characteristics. Further incorporating year-fixed effects in column 3 yields a coefficient that is approximately three times smaller, suggesting that part of the initially observed effect was driven by broad temporal trends affecting all facilities.

In columns 4-6, I introduce industry-year-fixed effects and additional controls. The model with industry-year-fixed effects in column 4 yields similar result to the model with industry-fixed effects. The results remain stable when including lagged emissions, county-level controls, and adjustments for spatial dependence.

Column 6 shows the most rigorous specification, which implements entropy balancing to reweight treated and control facilities on pre-treatment total releases. This final estimate indicates that facilities within a 5-mile radius of at least one PurpleAir sensor experience a 6.4% reduction in fugitive air emissions ($e^{-0.066} - 1 \approx -0.064$). For the average facility emitting 5,709 pounds annually, this represents a decrease of approximately 365 pounds per year. This effect is economically meaningful, amounting to more than 18 times the total annual emissions of the median facility, which emits just 20 pounds.

Figure 2 displays the dynamic treatment effects using the specification from column 6 of Table 2. The graph plots coefficient estimates across a seven-year window surrounding sensor deployment, spanning from three years pre-installation to three years post-installation. In the pre-event period, coefficient estimates are stable and statistically indistinguishable from zero, providing strong support for the parallel trends assumption underlying my DiD approach.

Finally, I do not use alternative fixed-effects structures such as county-year or state-year due to data limitations. Many counties contain only a single facility — often the only treated facility in the area — resulting in limited within-county variation in treatment timing. Since treatment is defined geographically based on proximity to sensors rather than administrative boundaries, I am not primarily concerned with capturing within-county differences. Moreover, including county-year fixed effects absorbs substantial identifying variation and leads to unstable estimates; in untabulated results, the estimated coefficient reverses sign and becomes statistically insignificant. While state-year fixed effects are less restrictive, they aggregate across heterogeneous local conditions and risk masking meaningful treatment effects. For these reasons, geographic-time fixed effects are not employed in this paper, as they risk over-controlling and attenuating the estimated impact. To address concerns about sparse data structure, I restrict the sample to observations that would not be dropped in a county-year fixed effects model, ensuring results are not driven by sample composition or singleton groups.

4.3 Enhancing Causal Identification

4.3.1 Sensitivity to Different Treatment Assignment

To assess the robustness of the estimated treatment effects, table 3 examines the sensitivity of the results to alternative definitions of treatment based on spatial proximity and sensor density. In Panel A, I systematically vary the spatial radius used to define treatment status, moving from the baseline 5-mile threshold to 3 miles and 1 mile. The effect magnitude increases monotonically as the treatment radius contracts, with fugitive air emission reductions growing from 6.4% at 5 miles to 6.8% at 3 miles to approximately 8.2% at 1 mile. This pattern aligns with the intuition that monitoring effectiveness intensifies when sensors are positioned closer to potential emission sources, likely because of improved attribution and signal strength.

Panel B explores variation in sensor density by modifying the minimum number of sensors required for treatment classification while maintaining the 5-mile radius. When raising the threshold from one sensor to two or three sensors, the estimated emission reductions increase progressively from 6.4% for one sensor to 7.1% for two sensors to 7.6% for three sensors. This pattern suggests that greater monitoring coverage enhances detection capability and amplifies pressure on facilities to reduce emissions.

The consistent gradients in treatment effects across both dimensions – proximity and density – provide compelling evidence that the emissions reductions are specifically attributable to Purple Air sensor presence rather than confounding factors. If the observed effects were driven by omitted variables or coincidental trends, we would not expect such a systematic variation in impact magnitude as the sensor placement becomes more proximate or more concentrated. These patterns are most consistent with the interpretation that citizen monitoring, rather than other concurrent developments, is shaping facility-level environmental outcomes.

4.3.2 Falsification Test

To test whether the observed reductions in fugitive emissions are credibly driven by PurpleAir sensor detection, I conduct a falsification test using stack air emissions — pollutants that the sensors are unlikely to capture reliably. Table 4 mirrors the structure of Table 3, but substitutes stack emissions for fugitive emissions as the dependent variable. This test relies on the technical limitation that PurpleAir sensors are designed to detect particulate matter near ground level and are therefore not well-positioned to capture emissions released through tall stacks, which are typically elevated 30 to 100 feet above ground and disperse widely before reaching sensor height.

As expected, I find no consistent or statistically significant relationship between PurpleAir sensor presence and stack air emissions. Across all specifications, the coefficients are small in magnitude and mostly statistically indistinguishable from zero. The lone exception is a positive and weakly significant effect in the most restrictive specification (*PA Sensor within 1 mi* $\times$ *Post*, significant at the 10% level), which is likely spurious given the absence of any systematic pattern across other distance or density thresholds.

This falsification test complements the main findings in two ways. First, it demonstrates outcome specificity: facilities appear to reduce emissions that PurpleAir sensors are capable of detecting (i.e., fugitive emissions), while stack emissions — less visible to ground-level sensors — remain unaffected. Second, it helps rule out alternative explanations such as time-varying omitted variables or general environmental trends, which would presumably affect both fugitive and stack emissions. The contrast between significant reductions in fugitive emissions and null effects on stack emissions reinforces the interpretation that the observed behavioral responses are driven by increased visibility through citizen-based monitoring.

4.3.3 Addressing Measurement Attribution Concerns

A key concern with citizen-deployed air sensors is the potential difficulty of attributing pollution readings to specific facilities, especially in areas with non-industrial pollution sources or multiple nearby emitters. Table 5 addresses this challenge through targeted sample restrictions.

Panel A examines whether proximity to major roadways affects the results by excluding facilities located within various distances (0.5, 1, and 2 miles) from highways. Across these increasingly restrictive samples, the estimated treatment effects remain robust - and in fact become larger in magnitude. When facilities within 2 miles of a highway are excluded, the estimated effect corresponds to an 8.0% reduction in fugitive emissions, compared to the baseline estimate of 6.4%.

Panel B addresses the challenge of industrial co-location by removing facilities situated in dense industrial clusters. I define clusters using two distance thresholds (0.5 and 1 mile) and two size cutoffs (minimum of 3 or 5 facilities). Across all definitions, the estimated treatment effects remain statistically significant and comparable in magnitude. The most restrictive definition - a cluster defined as 3 or more facilities within a 1-mile radius - yields a 7.4% reduction in emissions, again larger than the baseline.

The consistency and strengthening of results across these attribution-restricted samples reinforces the validity of the main findings. The fact that larger effects are observed when attribution is more credible suggests that PurpleAir sensors directly influence facility behavior rather than capturing coincidental environmental variation.

4.3.4 IV Design - Strengthening Causal Inference with Wildfire Smoke

To assess whether the observed relationship between PurpleAir sensor installation and emission reductions is causal — rather than driven by endogenous sensor placement (Zivin et al., 2024; Desouza and Kinney, 2021) — I implement an instrumental variable (IV) strategy that exploits exogenous variation in wildfire smoke exposure. Following Borgschulte et al. (2024b), I use the number of wildfire smoke days to instrument for PurpleAir sensor installation and analyze its effect on next-year fugitive emissions.

This identification strategy rests on the premise that wildfire smoke events raises public concern about air quality, prompting greater sensor adoption, while remaining plausibly unrelated to facility-level emissions behavior, except through this monitoring channel. To support the exclusion restriction, I focus on next-year emissions rather than contemporaneous outcomes to avoid capturing any direct short-term effects of smoke exposure on industrial operations, such as evacuation protocols or staffing adjustments during fire events.

The first-stage results in Table 6 Panel A confirm that exposure to wildfire smoke significantly predicts sensor installation. This finding is consistent with Coury et al. (2024) and Zivin et al. (2024), who show that pollution shocks from wildfire events lead to substantial uptake of low-cost air quality monitors. The coefficient of 0.004 indicates that each additional wildfire smoke day increases the likelihood that a facility is in the treated group and post-treatment period by 0.4 percentage points, consistent with the idea that wildfire smoke influences the likelihood of sensor deployment. In the second stage (Panel B), instrumented sensor presence is associated with reduced fugitive emissions in the following year, reinforcing the core findings of the main analysis. The estimated second-stage coefficient of –0.211 implies that wildfire-induced sensor installation leads to a 19% reduction in fugitive air emissions in the following year, which is roughly three times the magnitude of the 6.4% reduction estimated in the DiD baseline specification. This larger magnitude may reflect that the IV strategy captures longer-run or more persistent behavioral responses to sensor installation, whereas the DiD estimate may primarily reflect shorter-term impacts. Alternatively, the difference may arise from the IV identifying a local average treatment effect (LATE) for facilities whose monitoring behavior is more responsive to wildfire-driven salience, which could differ from the average effect captured in the DiD.

5 Additional Tests

5.1 Understanding the Mechanism

5.1.1 Mechanism Test: Regulatory Response

To understand how citizen monitoring influences corporate environmental behavior, I examine changes in regulatory oversight following PurpleAir sensor deployment. Table 7 presents results on how sensor presence affects air compliance activities — inspections and stack tests — conducted by different enforcement authorities. Because they are count data, I estimate a Poisson fixed effects model, which is more appropriate for count data than using log-transformations and helps address concerns raised by Cohn et al. (2022) regarding bias and interpretability in logged count regressions.

The findings in Panel A of Table 7 indicate that sensor installation is associated with a significant increase in air inspections by state agencies and the EPA. The estimated coefficient for state inspections corresponds to a 5.8% increase in inspection frequency, or about 0.08 additional inspections per facility-year, relative to a mean of 1.45. The effect is larger for EPA inspections, which rise by 40.4%, amounting to 0.12 additional inspections per facility-year from a mean of 0.31.[10]

Meanwhile, Panel B of Table 7 shows a statistically significant reduction in stack testing by state agencies in the Poisson specification, with an estimated 13.2% decline (approximately 0.35 fewer tests per facility-year, relative to a mean of 2.65).

[10]When estimating the Poisson fixed effects model with facility and industry-year fixed effects, a substantial number of observations are dropped because many facilities exhibit no variation in inspection activity over time — either receiving no inspections or the same number in every year. Since Poisson fixed effects rely on within-facility variation to identify coefficients, any unit with no temporal variation in the dependent variable provides no identifying information and is excluded from the estimation.

This pattern — an increase in state-level inspections alongside a reduction in stack tests — suggests a shift in regulatory strategy following the deployment of citizen-operated sensors. Regulators appear to move away from potentially more resource-intensive or pre-scheduled compliance checks like stack tests toward more nimble and targeted inspections, likely informed by real-time data from PurpleAir sensors. The interaction between citizen monitoring and formal enforcement may enable greater regulatory precision, as continuous data streams allow agencies to prioritize oversight more effectively. This shift toward more dynamic, data-guided enforcement likely plays an important role in driving the observed reductions in fugitive emissions, as facilities adjust to more focused scrutiny and an information-rich compliance environment.

5.2 Heterogeneous Impacts of Installation of PurpleAir sensors

Table 8 examines how the effectiveness of citizen monitoring varies across facility characteristics, revealing systematic patterns in which facilities respond most strongly to PurpleAir sensor deployment.

Panel A shows that PurpleAir sensors have substantially stronger effects on facilities that use less sophisticated emissions estimation methods. Facilities relying on published emission factors — rather than site-specific measurements or customized calculations — reduce their fugitive emissions by approximately 10.5% more than those using higher-quality estimation approaches. In contrast, facilities with more rigorous measurement methods exhibit no meaningful response. This pattern suggests that external scrutiny is more influential where internal oversight is weaker. Facilities with low-quality emissions estimation may be less confident in their internal monitoring and thus more responsive to outside visibility, reducing emissions as a precaution to avoid regulatory or legal risks prompted by sensor-based public exposure.

Panel B shows that citizen monitoring has greater impact at facilities with higher baseline pollution levels. Among above-median emitters, the presence of PurpleAir sensors is associated with an additional 12.7% reduction in fugitive emissions. Facilities with lower pre-treatment emissions show little or no change, implying that citizen monitoring may have limited influence where pollution levels are already low.

These cross-sectional tests indicate that citizen monitoring through PurpleAir sensors has its strongest impact where internal oversight mechanisms are weaker and where pollution levels are higher. These findings suggest that citizen-deployed sensors may be particularly valuable for addressing regulatory gaps at facilities with less stringent internal controls and greater environmental footprints - precisely the settings where additional oversight might yield the largest public health benefits.

6 Conclusion

This paper examines how citizen-led environmental monitoring impacts corporate pollution behavior through the deployment of PurpleAir sensors. I find that facilities exposed to citizen monitoring reduce their fugitive air emissions by 6.4% within three years of sensor installation. These results, robust to various identification strategies including instrumental variable approach using wildfire smoke exposure, demonstrate that citizen monitoring serves as an effective complement to traditional regulatory oversight.

This study makes four key contributions to the literature on environmental regulation, corporate oversight, and citizen science. First, it provides systematic evidence that citizen-deployed sensors reduce industrial pollution and reveals a nuanced public-private interaction: citizen monitoring substitutes for formal data collection in resource-constrained areas while complementing regulatory enforcement through enhanced targeting, contributing to our understanding of how public and private enforcement mechanisms intersect.

Second, it expands the corporate governance literature on informal monitoring by demonstrating that ordinary citizens, by generating quantitative data themselves, can effectively influence corporate behavior, moving beyond traditional roles as complainants or social media advocates.

Third, it highlights how affordable technological innovation democratizes environmental oversight, empowering citizens as data producers rather than mere consumers, distinct from regulator or corporate use of costly technologies like satellite imagery.

Fourth, by being the first to separately analyze the impact on fugitive air emissions — often under-regulated yet posing significant health risks comparable to stack emissions — this research underscores the unique capability of citizen monitoring to address specific, harmful pollution types often overlooked by aggregate metrics and traditional regulatory focus.

As the scope and complexity of environmental challenges persist, and recognizing that traditional regulatory monitoring may not always achieve comprehensive coverage, these findings underscore the growing importance of citizen participation in environmental governance. This is particularly true in areas where existing oversight mechanisms have limitations or for pollution types, such as fugitive emissions, that are inherently difficult to monitor comprehensively. The emergence of affordable, networked monitoring technologies appears to be reshaping the landscape of corporate environmental accountability in ways that policymakers, regulators, and communities should recognize and leverage to foster better environmental outcomes.

References

Akey, P. and Appel, I. (2021). The limits of limited liability: Evidence from industrial pollution. *The Journal of Finance*, 76(1):5–55.

AQMD (2019). Field evaluation: Purpleair (pa-ii) pm sensor. Available at: https://www.aqmd.gov/docs/default-source/aq-spec/field-evaluations/purple-air-pa-ii—field-evaluation.pdf, accessed May 8, 2025.

Babich, A. (2018). The unfulfilled promise of effective air quality and emissions monitoring. *Geo. Envtl. L. Rev.*, 30:569.

Blackman, A., Afsah, S., and Ratunanda, D. (2004). How do public disclosure pollution control programs work? evidence from indonesia. *Human Ecology Review*, pages 235–246.

Borgschulte, M., Deal, C., and Mazumder, B. (2024a). The broadening impact of rising wildfire smoke in the united states. *Chicago Fed Letter*, 500:1–6.

Borgschulte, M., Molitor, D., and Zou, E. Y. (2024b). Air pollution and the labor market: Evidence from wildfire smoke. *Review of Economics and Statistics*, 106(6):1558–1575.

Bowen, R. M., Call, A. C., and Rajgopal, S. (2010). Whistle-blowing: Target firm characteristics and economic consequences. *The Accounting Review*, 85(4):1239–1271.

Buntaine, M. T., Greenstone, M., He, G., Liu, M., Wang, S., and Zhang, B. (2024). Does the squeaky wheel get more grease? the direct and indirect effects of citizen participation on environmental governance in china. *American Economic Review*, 114(3):815–850.

Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D., and Bartonova, A. (2017). Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environment international*, 99:293–302.

Cohn, J. B., Liu, Z., and Wardlaw, M. I. (2022). Count (and count-like) data in finance. *Journal of Financial Economics*, 146(2):529–551.

Colmer, J. M., Evans, M. F., and Shimshack, J. (2023). Environmental citizen complaints.

Core, J. E., Guay, W., and Larcker, D. F. (2008). The power of the pen and executive compensation. *Journal of financial economics*, 88(1):1–25.

Coursen, D. (2021). Environmental justice requires adequate air quality monitoring system. *Bloomberg Law*.

Coury, M., Falconer, L., and La Nauze, A. (2024). Wildfire smoke and private provision of public air-quality monitoring. *Journal of Environmental Economics and Management*, 127:103036.

Daubanes, J. and Rochet, J.-C. (2019). The rise of ngo activism. *American Economic Journal: Economic Policy*, 11(4):183–212.

Delgado, M. S. and Florax, R. J. (2015). Difference-in-differences techniques for spatial data: Local autocorrelation and spatial interaction. *Economics Letters*, 137:123–126.

Desouza, P. and Kinney, P. L. (2021). On the distribution of low-cost pm2. 5 sensors in the us: demographic and air quality associations. *Journal of exposure science & environmental epidemiology*, 31(3):514–524.

Diviacco, P., Iurcev, M., Carbajales, R. J., and Potleca, N. (2022a). First results of the application of a citizen science-based mobile monitoring system to the study of household heating emissions. *Atmosphere*, 13(10):1689.

Diviacco, P., Iurcev, M., Carbajales, R. J., Potleca, N., Viola, A., Burca, M., and Busato, A. (2022b). Monitoring air quality in urban areas using a vehicle sensor network (vsn) crowdsensing paradigm. *Remote Sensing*, 14(21):5576.

Duchin, R., Gao, J., and Xu, Q. (2025). Sustainability or greenwashing: Evidence from the asset market for industrial pollution. *The Journal of Finance*, 80(2):699–754.

Duflo, E., Greenstone, M., Pande, R., and Ryan, N. (2013). Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from india. *The Quarterly Journal of Economics*, 128(4):1499–1545.

Dyck, A., Morse, A., and Zingales, L. (2010). Who blows the whistle on corporate fraud? *The Journal of Finance*, 65(6):2213–2253.

Dyck, A., Volchkova, N., and Zingales, L. (2008). The corporate governance role of the media: Evidence from russia. *The Journal of finance*, 63(3):1093–1135.

Earnhart, D. and Friesen, L. (2013). Can punishment generate specific deterrence without updating? analysis of a stated choice scenario. *Environmental and Resource Economics*, 56:379–397.

EPA (2013). Tri data quality process. Available at: https://www.epa.gov/toxics-release-inventory-tri-program/tri-data-quality-process, accessed May 8, 2025.

EPA (2015). Government partnerships to reduce air pollution. Available at: https://www.epa.gov/clean-air-act-overview/government-partnerships-reduce-air-pollution, accessed May 6, 2025.

EPA (2018). Air releases. Available at: https://www.epa.gov/trinationalanalysis/air-releases, accessed May 6, 2025.

EPA (2021). Enhanced air quality monitoring funding under the arp. Available at: https://www.epa.gov/arp/enhanced-air-quality-monitoring-funding-under-arp, accessed May 6, 2025.

EPA (2022). Epa research improves air quality information for the public on the airnow fire and smoke map. Available at: https://www.epa.gov/sciencematters/epa-research-improves-air-quality-information-public-airnow-fire-and-smoke-map, accessed May 8, 2025.

Fowlie, M., Holland, S. P., and Mansur, E. T. (2012). What do emissions markets deliver and to whom? evidence from southern california's nox trading program. *American Economic Review*, 102(2):965–993.

Fowlie, M., Rubin, E., and Walker, R. (2019). Bringing satellite-based air quality estimates down to earth. In *AEA Papers and Proceedings*, volume 109, pages 283–288. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Grainger, C. and Schreiber, A. (2019). Discrimination in ambient air pollution monitoring? In *AEA Papers and Proceedings*, volume 109, pages 277–282. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Gu, Y., Dai, J., and Vasarhelyi, M. A. (2023). Audit 4.0-based esg assurance: An example of using satellite images on ghg emissions. *International Journal of Accounting Information Systems*, 50:100625.

Heese, J. and Pacelli, J. (2024). The monitoring role of social media. *Review of Accounting Studies*, 29(2):1666–1706.

Heese, J., Pérez-Cavazos, G., and Peter, C. D. (2022). When the local newspaper leaves town: The effects of local newspaper closures on corporate misconduct. *Journal of Financial Economics*, 145(2):445–463.

Holloway, T., Miller, D., Anenberg, S., Diao, M., Duncan, B., Fiore, A. M., Henze, D. K., Hess, J., Kinney, P. L., Liu, Y., et al. (2021). Satellite monitoring for air quality and health. *Annual review of biomedical data science*, 4(1):417–447.

Jiang, J. X. and Kong, J. (2024). Green dies in darkness? environmental externalities of newspaper closures. *Review of Accounting Studies*, 29(4):3564–3599.

Kang, J. K., Stice-Lawrence, L., and Wong, Y. T. F. (2021). The firm next door: Using satellite images to study local information advantage. *Journal of Accounting Research*, 59(2):713–750.

Karner, A. A., Eisinger, D. S., and Niemeier, D. A. (2010). Near-roadway air quality: synthesizing the findings from real-world data. *Environmental science & technology*, 44(14):5334–5344.

Karpoff, J. M., Lott, Jr, J. R., and Wehrly, E. W. (2005). The reputational penalties for environmental violations: Empirical evidence. *The Journal of Law and Economics*, 48(2):653–675.

Konar, S. and Cohen, M. A. (1997). Information as regulation: The effect of community right to know laws on toxic emissions. *Journal of environmental Economics and Management*, 32(1):109–124.

Konar, S. and Cohen, M. A. (2001). Does the market value environmental performance? *Review of economics and statistics*, 83(2):281–289.

Krebs, B., Burney, J., Zivin, J. G., and Neidell, M. (2021). Using crowd-sourced data to assess the temporal and spatial relationship between indoor and outdoor particulate matter. *Environmental science & technology*, 55(9):6107–6115.

Miller, G. S. (2006). The press as a watchdog for accounting fraud. *Journal of Accounting Research*, 44(5):1001–1033.

Miller, N., Molitor, D., and Zou, E. (2021). A causal concentration-response function for air pollution: Evidence from wildfire smoke. *Work. Pap., Coll. Bus., Univ. Ill., Urbana-Champaign*.

Morawska, L., Thai, P. K., Liu, X., Asumadu-Sakyi, A., Ayoko, G., Bartonova, A., Bedini, A., Chai, F., Christensen, B., Dunbabin, M., et al. (2018). Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? *Environment international*, 116:286–299.

Newitz, A. (2020). I live in california. how do i know it's safe to go outside? Available at: https://www.nytimes.com/2020/09/04/opinion/california-wildfires-air-quality.html, accessed May 13, 2025.

Olken, B. A. (2007). Monitoring corruption: evidence from a field experiment in indonesia. *Journal of Political Economy*, 115(2):200–249.

Pargal, S. and Wheeler, D. (1996). Informal regulation of industrial pollution in developing countries: evidence from indonesia. *Journal of political economy*, 104(6):1314–1327.

Potoski, M. and Prakash, A. (2005). Green clubs and voluntary governance: Iso 14001 and firms' regulatory compliance. *American journal of political science*, 49(2):235–248.

Ruminski, M., Kondragunta, S., Draxler, R., and Zeng, J. (2006). Recent changes to the hazard mapping system. In *Proceedings of the 15th International Emission Inventory Conference*, volume 15, page 18. National Oceanic and Atmospheric Administration.

Schantl, S. F. and Wagenhofer, A. (2020). Deterrence of financial misreporting when public and private enforcement strategically interact. *Journal of Accounting and Economics*, 70(1):101311.

Scheer, M. (2025). Attorneys: Analysis of 'fugitive emissions' heighten goodyear cancer risk. Available at: https://www.yahoo.com/news/attorneys-analysis-fugitive-emissions-heighten-214600942.html, accessed May 6, 2025.

Shapiro, J. S. and Walker, R. (2018). Why is pollution from us manufacturing declining? the roles of environmental regulation, productivity, and trade. *American economic review*, 108(12):3814–3854.

Shimshack, J. P. and Ward, M. B. (2005). Regulator reputation, enforcement, and environmental compliance. *Journal of Environmental Economics and Management*, 50(3):519–540.

Shimshack, J. P. and Ward, M. B. (2008). Enforcement and over-compliance. *Journal of Environmental Economics and Management*, 55(1):90–105.

Short, J. L. and Toffel, M. W. (2008). Coerced confessions: Self-policing in the shadow of the regulator. *The Journal of Law, Economics, & Organization*, 24(1):45–71.

Snyder, E. G., Watkins, T. H., Solomon, P. A., Thoma, E. D., Williams, R. W., Hagler, G. S., Shelow, D., Hindin, D. A., Kilaru, V. J., and Preuss, P. W. (2013). The changing paradigm of air pollution monitoring. *Environmental science & technology*, 47(20):11369–11377.

Sullivan, D. M., Krupnick, A., et al. (2018). Using satellite data to fill the gaps in the us air pollution monitoring network. *Resources for the future working paper*, 18.

Sun, Y., Mousavi, A., Masri, S., and Wu, J. (2022). Socioeconomic disparities of low-cost air quality sensors in california, 2017–2020. *American Journal of Public Health*, 112(3):434–442.

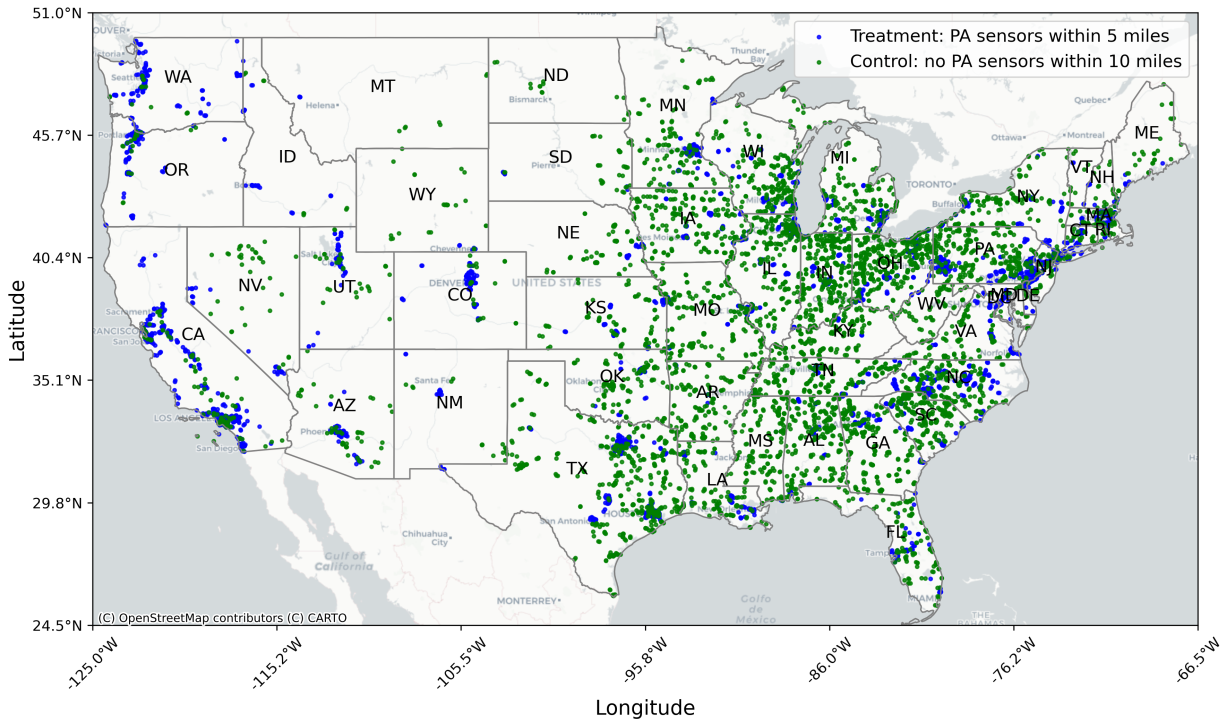
Thomas, J., Yao, W., Zhang, F., and Zhu, W. (2022). Meet, beat, and pollute. *Review of Accounting Studies*, 27(3):1038–1078.

Xu, Q. and Kim, T. (2022). Financial constraints and corporate environmental policies. *The Review of Financial Studies*, 35(2):576–635.

Zivin, J. S. G., Krebs, B., and Neidell, M. J. (2024). Private adoption of public good technologies: The case of purpleair. Technical report, National Bureau of Economic Research.

Zou, E. Y. (2021). Unwatched pollution: The effect of intermittent monitoring on air quality. *American Economic Review*, 111(7):2101–2126.

Figure 1: Treatment versus Control Facilities



Treatment status is assigned to facilities with at least one PurpleAir sensor within a 5-mile radius, while control facilities have no sensors within 10 miles. The treatment year corresponds to the initial year a facility is exposed to a PurpleAir sensor within its 5-mile vicinity.

Figure 2: Event Study for Main Result

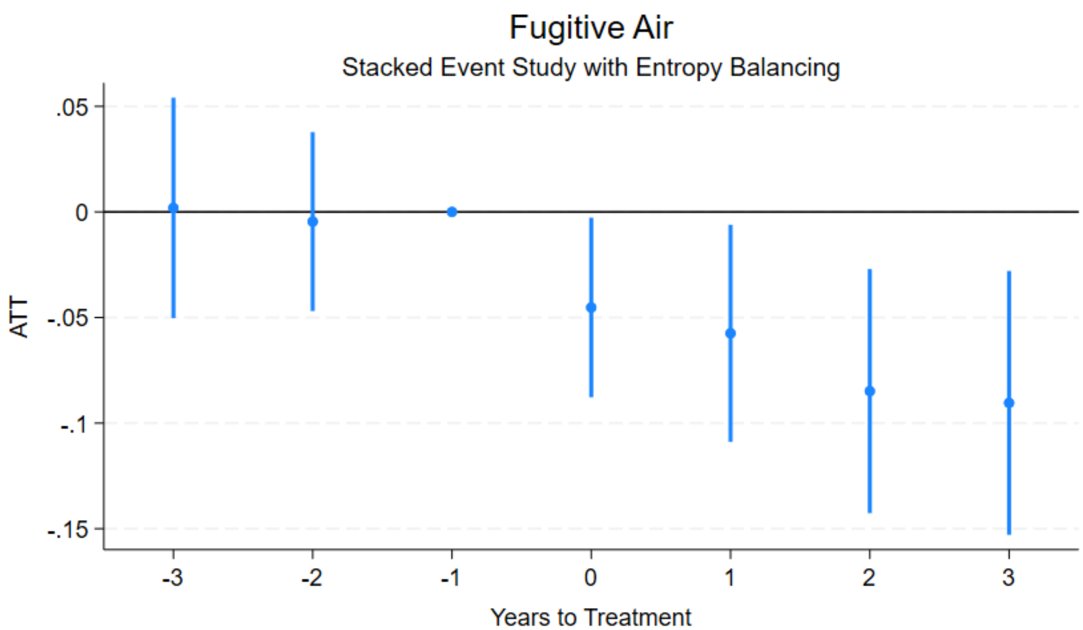


Table 1: Descriptive Statistics

Panel A: Summary Statistics for the DiD Design

	count	mean	sd	min	p50	max
Fugitive Air	226,508	5,708.95	20,172.13	0	20	128,257
Stack Air	226,508	23,833.50	77,739.38	0	75	483,699
Ln(Total Releases)$_{t-1}$	226,508	7.20	4.10	0	8	15
Ln(Population)	226,508	11.51	1.44	9	11	16
Unemployment Rate	226,508	4.94	1.79	2	5	12
Spatial Dependence in Fugitive Emissions	226,508	2.88	2.11	0	3	9
Spatial Dependence in Stack Emissions	226,508	3.53	2.57	0	3	10
Distance from the Nearest Highway	226,508	11.26	18.40	0	4	190
Not in a Cluster (5 Facilities within 0.5 mi radius)	226,508	0.96	0.20	0	1	1
Not in a Cluster (5 Facilities within 1 mi radius)	226,508	0.84	0.37	0	1	1
Not in a Cluster (3 Facilities within 0.5 mi radius)	226,508	0.83	0.38	0	1	1
Not in a Cluster (3 Facilities within 1 mi radius)	226,508	0.64	0.48	0	1	1
Local Inspections	226,508	0.04	0.73	0	0	87
State Inspections	226,508	0.75	2.51	0	0	284
EPA Inspections	226,508	0.03	0.30	0	0	34
Local Stack Tests	226,508	0.05	1.95	0	0	370
State Stack Tests	226,508	0.61	3.26	0	0	195
EPA Stack Tests	226,508	0.00	0.07	0	0	12
Low Quality Estimation	226,508	0.70	0.46	0	1	1
High Emitter	226,508	0.50	0.50	0	0	1

Panel B: Summary Statistics for the IV Design

	count	mean	sd	min	p50	max
Wildfire Smoke Days	208,167	41.85	32.92	0	36	200
Fugitive Air$_t+1$	208,167	5,855.88	20,452.95	0	20	128,257
Ln(Total Releases)$_{t-1}$	208,167	7.30	4.08	0	8	15
Ln(Population)	208,167	11.48	1.44	9	11	16
Unemployment Rate	208,167	4.97	1.79	2	5	12
Spatial Dependence in Fugitive Emissions	208,167	2.89	2.12	0	3	9

Table 2: PurpleAir Sensors and Air Releases

	(1)	(2)	(3)	(4)	(5)	(6)
			Ln(Fugitive Air)			
PA Sensor within 5 mi × Post	-0.151***	-0.110***	-0.053**	-0.055**	-0.051**	-0.066***
	(0.023)	(0.019)	(0.021)	(0.022)	(0.021)	(0.022)
PA Sensor within 5 mi	-0.414***					
	(0.053)					
Ln(Total Releases)$_{t-1}$					0.149***	0.137***
					(0.004)	(0.007)
Ln(Population)					-0.075	0.176
					(0.199)	(0.312)
Unemployment Rate					-0.030***	-0.013*
					(0.005)	(0.007)
Spatial Dependence in Fugitive Emissions					-0.004	-0.004
					(0.005)	(0.008)
N	226508	226508	226508	226508	226508	226508
Adj. R-squared	0.002	0.893	0.893	0.894	0.897	0.897
Facility × Cohort FE		Yes	Yes	Yes	Yes	Yes
Year × Cohort FE			Yes			
Industry-Year × Cohort FE				Yes	Yes	Yes
Entropy Balancing						Yes

This table presents stacked DiD estimates examining the impact of PurpleAir sensor presence on facility-level fugitive air emissions. The dependent variable is *Ln(Fugitive Air)*, winsorized at the 1% level. Treatment status is assigned to facilities with at least one PurpleAir sensor within a 5-mile radius, while control facilities have no sensors within 10 miles. The treatment year corresponds to the initial year a facility is exposed to a PurpleAir sensor within its 5-mile vicinity. *PA Sensor within 5 mi × Post* is an indicator variable that equals one for treated facility-years following sensor deployment, and zero otherwise. Standard errors are clustered at the facility×cohort level and reported below the coefficients. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 3: Sensitivity to Different Treatment Definitions

Panel A: Changing Treatment Radii			
	(1)	(2)	(3)
	Treatment Radii		
	5 miles	3 miles	1 mile
	Ln(Fugitive Air)		
PA Sensor within X mi × Post	-0.066***	-0.070***	-0.086*
	(0.022)	(0.027)	(0.050)
N	226508	214293	196995
Adj. R-squared	0.897	0.897	0.901
Controls	Yes	Yes	Yes
Facility × Cohort FE	Yes	Yes	Yes
Industry-Year × Cohort FE	Yes	Yes	Yes
Entropy Balancing	Yes	Yes	Yes

Panel B: Changing Number of Sensors Threshold			
	(1)	(2)	(3)
	Min. No. of Sensors		
	1 sensor	2 sensors	3 sensors
	Ln(Fugitive Air)		
PA Sensor within 5 mi × Post	-0.066***	-0.074***	-0.079***
	(0.022)	(0.026)	(0.030)
N	226508	213537	207892
Adj. R-squared	0.897	0.897	0.897
Controls	Yes	Yes	Yes
Facility × Cohort FE	Yes	Yes	Yes
Industry-Year × Cohort FE	Yes	Yes	Yes
Entropy Balancing	Yes	Yes	Yes

This table examines the robustness of the primary results tabulated in Table 2 by presenting stacked DiD estimates of the effect of PurpleAir sensor presence on facility-level fugitive air emissions across varying treatment definition specifications. The dependent variable is *Ln(Fugitive Air)*, winsorized at the 1% level. Panel A reports sensitivity analyses for different treatment radii: the main specification (5 miles) alongside alternative thresholds (3 miles and 1 mile). Panel B examines sensitivity to the number of sensors: the main specification (at least one sensor) alongside more stringent requirements (at least 2 sensors or at least 3 sensors). In all specifications, control facilities have no PurpleAir sensors within 10 miles. The treatment year corresponds to the initial year a facility meets the respective treatment criterion. All specifications include facility×cohort and industry-year×cohort fixed effects, with standard errors clustered at the facility×cohort level. Standard errors are reported below the coefficients, ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4: Falsification Test

Panel A: Changing Treatment Radii

	(1)	(2)	(3)
	Treatment Radii		
	5 miles	3 miles	1 mile
	Ln(Stack Air)		
PA Sensor within X mi × Post	0.032	0.037	0.087*
	(0.021)	(0.025)	(0.048)
N	226508	214293	196995
Adj. R-squared	0.928	0.928	0.930
Controls	Yes	Yes	Yes
Facility × Cohort FE	Yes	Yes	Yes
Industry-Year × Cohort FE	Yes	Yes	Yes
Entropy Balancing	Yes	Yes	Yes

Panel B: Changing Number of Sensors Threshold

	(1)	(2)	(3)
	Min. No. of Sensors		
	1 sensor	2 sensors	3 sensors
	Ln(Stack Air)		
PA Sensor within 5 mi × Post	0.032	0.011	0.016
	(0.021)	(0.025)	(0.027)
N	226508	213537	207892
Adj. R-squared	0.928	0.927	0.927
Controls	Yes	Yes	Yes
Facility × Cohort FE	Yes	Yes	Yes
Industry-Year × Cohort FE	Yes	Yes	Yes
Entropy Balancing	Yes	Yes	Yes

This table presents falsification tests for the results of Table 3 by analyzing the effect of PurpleAir sensor presence on stack air emissions rather than fugitive emissions. The dependent variable is *Ln(Stack Air)*, winsorized at the 1% level. Panel A reports sensitivity analyses for different treatment radii: the main specification (5 miles) alongside alternative thresholds (3 miles and 1 mile). Panel B examines sensitivity to the number of sensors: the main specification (at least one sensor) alongside more stringent requirements (at least 2 sensors or at least 3 sensors). In all specifications, control facilities have no PurpleAir sensors within 10 miles. The treatment year corresponds to the initial year a facility meets the respective treatment criterion. All specifications include facility×cohort and industry-year×cohort fixed effects, with standard errors clustered at the facility×cohort level. Standard errors are reported below the coefficients, ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 5: Addressing Measurement Attribution Concerns

Panel A: Excluding Facilities Near Highways

	(1)	(2)	(3)
	Excluding Facilities Within		
Distance to the nearest highway	0.5 mile	1 mile	2 miles
	Ln(Fugitive Air)		
PA Sensor within 5 mi × Post	-0.063**	-0.079***	-0.083**
	(0.025)	(0.030)	(0.038)
N	197977	173075	141241
Adj. R-squared	0.899	0.900	0.902
Controls	Yes	Yes	Yes
Facility × Cohort FE	Yes	Yes	Yes
Industry-Year × Cohort FE	Yes	Yes	Yes
Entropy Balancing	Yes	Yes	Yes

Panel B: Excluding Facilities in a Cluster

	(1)	(2)	(3)	(4)
	Excluding Facilities in a Cluster			
Radius for cluster consideration	0.5 mile	1 mile	0.5 mile	1 mile
Min. points to form a cluster	5 facilities		3 facilities	
	Ln(Fugitive Air)			
PA Sensor within 5 mi × Post	-0.060***	-0.065**	-0.052**	-0.077**
	(0.022)	(0.025)	(0.025)	(0.032)
N	216716	189379	187810	145032
Adj. R-squared	0.898	0.899	0.897	0.898
Controls	Yes	Yes	Yes	Yes
Facility × Cohort FE	Yes	Yes	Yes	Yes
Industry-Year × Cohort FE	Yes	Yes	Yes	Yes
Entropy Balancing	Yes	Yes	Yes	Yes

This table presents stacked DiD estimates examining the effect of PurpleAir sensor presence on facility-level fugitive air emissions on subsamples that address potential measurement attribution concerns. The dependent variable is *Ln(Fugitive Air)*, winsorized at the 1% level. Panel A reports results after excluding facilities located within specified distances (0.5, 1, or 2 miles) of the nearest highway, addressing concerns that roadway pollution might contaminate sensor readings. Panel B reports results after excluding facilities located in industrial clusters, where pollution attribution to specific sources may be difficult. Clusters are defined using two alternative distance thresholds (0.5 and 1 mile) and two alternative minimum facility requirements (3 or 5 facilities). Treatment status is assigned to facilities with at least one PurpleAir sensor within a 5-mile radius, while control facilities have no sensors within 10 miles. The treatment year corresponds to the initial year a facility is exposed to a PurpleAir sensor within its 5-mile vicinity. All specifications include facility×cohort and industry-year×cohort fixed effects, with standard errors clustered at the facility×cohort level. Standard errors are reported below the coefficients, ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 6: Wildfire Smoke Day as Instrumental Variable

Panel A: First-stage and reduced-form estimates		
	(1)	(2)
	PA Sensor Entry	Ln(Fugitive Air)$_{t+1}$
Wildfire Smoke Days	0.004***	-0.001***
	(65.77)	(-3.85)
N	208167	208167
Adj. R-squared	0.394	0.895
Facility$\times$ Cohort FE	Yes	Yes
Industry$\times$ Cohort FE	Yes	Yes
Controls	Yes	Yes
Entropy Balancing	Yes	Yes

Panel B: IV estimates		
PA $\widehat{\text{Sensors}}$ Entry	-	-0.211***
	-	(-3.85)
N	-	208167
Adj. R-squared	-	0.003
Facility$\times$ Cohort FE	-	Yes
Industry$\times$ Cohort FE	-	Yes
Controls	-	Yes
Entropy Balancing	-	Yes
Kleibergen-Paap F	-	4326.4

This table presents an instrumental variables analysis examining the impact of PurpleAir sensor presence on next year's fugitive air emissions at the facility level. The main dependent variable is *Ln(Fugitive Air)*$_{t+1}$, winsorized at the 1% level. Panel A reports the first-stage and reduced-form estimates. *Wildfire Smoke Days* is the annual count of days during which a facility is fully enveloped by a wildfire smoke plume. *PA Sensor Entry* is an indicator set to 1 in the three years after a sensor is deployed within a 5-mile radius of a facility. In Panel B, *Wildfire Smoke Days* serves as an instrument for PurpleAir sensor deployment within this 5-mile radius. All specifications include facility$\times$cohort and industry$\times$cohort fixed effects, with standard errors clustered at the facility$\times$cohort level. T-statistics are reported below the coefficients, ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 7: Mechanism Test: Regulatory Response via Air Compliance Activities

Panel A: Air Inspections

	(1)	(2)	(3)
	Regulatory Authority		
	Local	State	EPA
	Count of Air Inspections		
PA Sensor within 5 mi × Post	0.091	0.056^{*}	0.339^{***}
	(0.114)	(0.032)	(0.114)
Outcome mean	2.175	1.449	0.305
N	4573	116913	18678
Pseudo R-squared	0.542	0.481	0.266
Controls	Yes	Yes	Yes
Facility × Cohort FE	Yes	Yes	Yes
Industry-Year × Cohort FE	Yes	Yes	Yes

Panel B: Air Stack Tests

	(1)	(2)	(3)
	Regulatory Authority		
	Local	State	EPA
	Count of Air Stack Tests		
PA Sensor within 5 mi × Post	0.065	-0.141^{**}	-0.334
	(0.196)	(0.068)	(0.986)
Outcome mean	4.847	2.645	0.748
N	2132	52652	416
Pseudo R-squared	0.797	0.566	0.317
Controls	Yes	Yes	Yes
Facility × Cohort FE	Yes	Yes	Yes
Industry-Year × Cohort FE	Yes	Yes	Yes

This table presents stacked DiD estimates examining the impact of PurpleAir sensor presence on regulatory air compliance activities. Panel A reports results for *Count of Air Inspections* while Panel B reports results for *Count of Air Stack Tests*, both disaggregated by the responsible regulatory authority (local agencies, state agencies, or EPA). Treatment status is assigned to facilities with at least one PurpleAir sensor within a 5-mile radius, while control facilities have no sensors within 10 miles. The treatment year corresponds to the initial year a facility is exposed to a PurpleAir sensor within its 5-mile vicinity. *PA Sensor within 5 mi × Post* is an indicator variable that equals one for treated facility-years following sensor deployment, and zero otherwise. All specifications include facility×cohort and industry-year×cohort fixed effects, with standard errors clustered at the facility×cohort level. Standard errors are reported below the coefficients, ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 8: Cross-Sectional Tests

Panel A: Emissions Estimation Method

	(1) Ln(Fugitive Air)
PA Sensor within 5 mi × Post × Low Quality Emissions Estimation	-0.111***
	(0.037)
PA Sensor within 5 mi × Post	-0.002
	(0.029)
N	226508
Pseudo R-squared	0.897
Controls	Yes
Facility × Cohort FE	Yes
Industry-Year × Cohort FE	Yes
Entropy Balancing	Yes

Panel B: Pre-Treatment Total Releases

	(1) Ln(Fugitive Air)
PA Sensor within 5 mi × Post × High Emitter	-0.136***
	(0.040)
PA Sensor within 5 mi × Post	-0.006
	(0.026)
N	226508
Pseudo R-squared	0.897
Controls	Yes
Facility × Cohort FE	Yes
Industry-Year × Cohort FE	Yes
Entropy Balancing	Yes

This table presents stacked DiD estimates examining the cross-sectional variation in the results of Table 2. Panel A shows the result from cross-sectional test based on emissions estimation method. *Low Quality Emissions Estimation* equals 1 if the facility uses published emission factors to estimate its air emission releases, and 0 if the facility relies on more site-specific methods such as monitoring data, mass balance, site-adjusted emission factors, or engineering calculations. Panel B shows the result from cross-sectional test based on pre-treatment total releases. *High Emitter* equals 1 if the pre-treatment total releases are above the median, and 0 otherwise. The dependent variable is *Ln(Fugitive Air)*, winsorized at the 1% level. Treatment status is assigned to facilities with at least one PurpleAir sensor within a 5-mile radius, while control facilities have no sensors within 10 miles. The treatment year corresponds to the initial year a facility is exposed to a PurpleAir sensor within its 5-mile vicinity. *PA Sensor within 5 mi × Post* is an indicator variable that equals one for treated facility-years following sensor deployment, and zero otherwise. All specifications include facility×cohort and industry-year×cohort fixed effects, with standard errors clustered at the facility×cohort level. Standard errors are reported below the coefficients, ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Appendix

Appendix 1: Example of a PurpleAir Sensor being Installed for Air Quality Monitoring

ENERGY & ENVIRONMENT

Fort Bend environmental organization to install air monitors to share data with EPA

The Fort Bend County Environmental Organization hopes that the data sharing will prompt officials to install regulatory air monitors that can be used to enforce air quality standards.

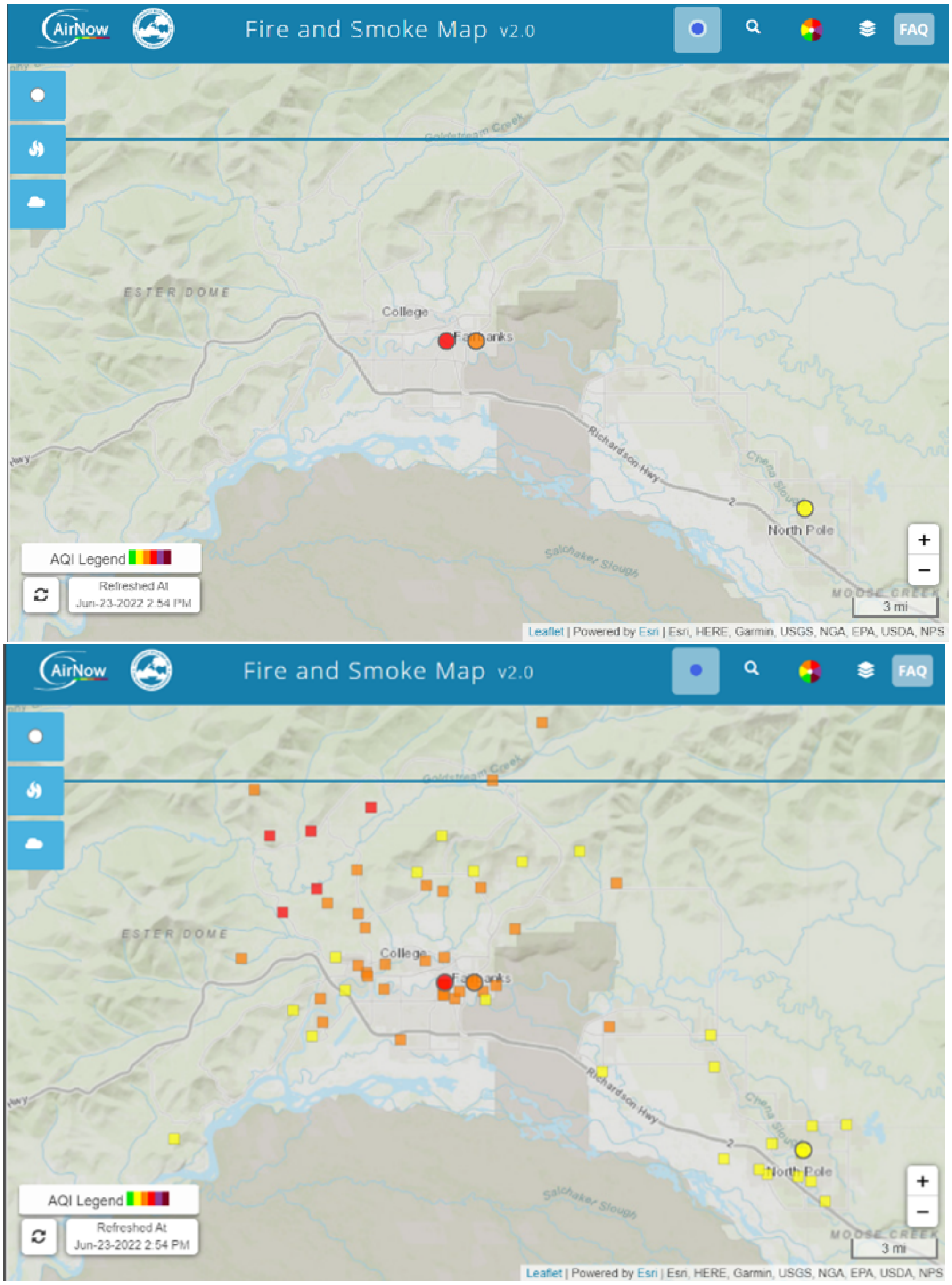
Natalie Weber, Fort Bend County Bureau | June 25, 2024, 4:18 PM

Katie Watkins/Houston Public Media

FILE: Juan Flores with Air Alliance Houston installs a PurpleAir monitor at a home in Jacinto City.

Source: Houston Public Media

Appendix 2: PurpleAir Sensor Integration into AirNow.gov



Two images from the AirNow Fire and Smoke Map show conditions near Fairbanks, AK on June 23, 2022. The image on the top shows the location of regulatory monitors only, while the image on the bottom shows a larger area covered by both regulatory monitors (circles) and PurpleAir sensors (squares). Source: U.S. Environmental Protection Agency.

Appendix 3: Recent State Legislation Affecting Community Air Monitoring Data

State	Bill (Session/Year)	Key Language	Current Status	Why It Matters
Louisiana	Senate Bill 503 (Act 181, Regular Session 2024) – "Community Air Monitoring Reliability Act"	Requires use of federal reference methods for alleged non-compliance; community-sensor data not admissible for enforcement or legal purposes.	Enacted. Signed by Gov. Landry on May 23, 2024.	Sets technical and QA standards most PurpleAir devices don't meet, limiting their use in lawsuits or regulatory challenges.
Kentucky	House Bill 137 (Regular Session 2025) – "An Act relating to air quality monitoring"	Limits air-pollution cases to "EPA-approved" methods with "scientifically defensible, quality-assured data," effectively excluding consumer sensors.	Enacted. Became law without governor's signature on Mar 25, 2025 (Acts Chapter 79)	Narrows admissible evidence in court or regulatory actions, sidelining PurpleAir data.
West Virginia	House Bill 5018 (Regular Session 2024)	Bars "data collected by community air-monitoring programs" from use by DEP or in Clean Air Act enforcement.	Pending. Passed House; stalled in Senate Judiciary (no final vote in 2024)	Would exclude PurpleAir readings from violation notices or court evidence.