PNAS RESEARCH ARTICLE | ENVIRONMENTAL SCIENCES OPEN ACCESS

Rural outdoor air pollution exceeds urban levels in the Delhi megacity region

Mark Joseph Campmier[a], Sofiya Rao[b], Amarendra Singh[b], Saumya Singh[c], María Teresa Hernández[a], Nidhi Malik[d], Swagata Dey[d], Ajay Pillarisetti[a], Kalpana Balakrishnan[e], Sagnik Dey[b,f,1], and Joshua S. Apte[a,g,1]

Affiliations are included on p. 9.

Edited by Akkihebbal R. Ravishankara, Colorado State University, Fort Collins, CO; received September 24, 2025; accepted October 13, 2026

India routinely experiences some of the highest concentrations of fine particulate matter ($PM_{2.5}$) air pollution in the world. Although India's population predominantly lives in rural areas and smaller cities, policy efforts to date have emphasized air pollution in India's large cities. In contrast, there is limited understanding of air pollutant dynamics in India's rural and less urbanized areas. Here, we established a well-calibrated lower-cost $PM_{2.5}$ sensor network spanning a wide range of urbanicity (population densities: 300 to 30,000 people/km^2) outside and inside of Delhi. The network of 49 sites incorporated a 160 km measurement transect to capture ambient $PM_{2.5}$ trends of the Delhi megacity, as well as stratified measurements in villages, small cities, and Delhi itself. We report on data from 5 mo of operation (July–November 2024) spanning the cleanest and most polluted times of the year. $PM_{2.5}$ concentrations at rural sites consistently exceeded those at urban sites by about 20% during every month of our study. Analysis of diurnal $PM_{2.5}$ cycles showed sharp morning (approximately 06:00–10:00 LT) and evening (approximately 16:00–20:00 LT) concentration peaks at rural sites that were absent at urban sites. We attribute these events to widespread use of solid fuels for cooking, heating, and small-scale industry in suburban and rural settings. These striking enhancements in rural air pollution are missed by satellite remote sensing data, and our findings run contrary to the general societal perception that Indian cities are pollution hotspots relative to rural areas.

air quality | $PM_{2.5}$ | low-cost sensors | India | rural air pollution

Significance

Fine particle air pollution ($PM_{2.5}$) exposure is the largest environmental risk factor for death in India. Delhi is often considered the world's most polluted megacity, yet air quality measurements in surrounding rural and peri-urban areas is sparse. Using a 160 km transect of sensors across the urban–rural divide, we found that $PM_{2.5}$ is substantially higher in villages outside Delhi's urban core, which diverges from available evidence from satellite remote sensing observations. High morning and evening pollution from diverse combustion activities in/near villages, likely cooking, heating, and rural industry, drives elevated concentrations. Clean air strategies should prioritize pollution sources specific to rural communities to promote environmental health across the region.

Exposure to ambient fine particulate matter ($PM_{2.5}$)—particles smaller than 2.5 μm in aerodynamic diameter (d_p)—is the top environmental risk factor and fourth highest risk factor overall for premature mortality globally (1–3). India, the most populous nation in the world, faces the highest global particulate air pollution-related disease burden, with over 1 million premature deaths annually attributable to ambient and household $PM_{2.5}$ exposure (2, 4). Estimates of ambient $PM_{2.5}$ concentrations consistently demonstrate that Delhi, the capital of India, and its surrounding peri-urban and rural districts, are among the most polluted regions in the world (5–8). Satellite remote sensing products suggest annual mean $PM_{2.5}$ concentrations exceed at least 80 μg/m^3, more than double the World Health Organization Interim Target 1 guideline of 35 μg/m^3 [*SI Appendix*, Fig. S1; (9)]. Seasonally, during the postmonsoon period (October–November) and winter (December–February), short-term exposure can be much higher, with hourly means exceeding 500 to 1,000 μg/m^3 [*SI Appendix*, Figs. S2 and S3; (9–12)]. Most evidence suggests this is due to a combination of persistent synoptic-scale boundary layer inversions at the base of the Himalayas, and secondary $PM_{2.5}$ formation chemical mechanisms resulting in regional haze [*SI Appendix*, Figs. S3 and S4; (5, 8, 13–15)].

The Indian government has implemented policies to intervene across multiple sectors nationwide. The "National Clean Air Programme" (NCAP) designated nonattainment cities for improvement, funded targeted interventions at urban-specific sources (e.g., water sprinkling to reduce fugitive road dust), and built a network of ground monitoring stations to monitor compliance, approximately 40 of which are in the National Capital Region (NCR) alone (16–18). However, regulatory monitoring is largely absent beyond these urban pockets (19). Previous work utilizing satellite-based data has emphasized urban cores as polluted hotspots (20, 21). Our findings here suggest that satellite-based exposure estimates may struggle to resolve temporally variable urban–rural gradients. Nonetheless, the importance of reducing dependence on biomass fuel has been recognized as key to multisectoral development for smaller rural settlements, including reduced exposure to air pollution. The "Pradhan Mantri Ujjwala Yojana" (PMUY) and "Saubhagya" initiatives encouraged uptake of liquified petroleum gas (LPG) and improved electricity access,

Author contributions: M.J.C., S.R., S.S., M.T.H., N.M., Swagata Dey, A.P., K.B., Sagnik Dey, and J.S.A. designed research; M.J.C., S.R., A.S., M.T.H., N.M., Swagata Dey, Sagnik Dey, and J.S.A. performed research; M.J.C. contributed new reagents/analytic tools; M.J.C., A.S., Sagnik Dey, and J.A. analyzed data; and M.J.C., S.R., A.S., S.S., M.T.H., N.M., Swagata Dey, A.P., K.B., Sagnik Dey, and J.S.A. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2026 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

PNAS policy is to publish maps as provided by the authors.

[1]To whom correspondence may be addressed. Email: sagnik@cas.iitd.ac.in or apte@berkeley.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2526851123/-/DCSupplemental.

Published November 23, 2026.

Downloaded from https://www.pnas.org by 72.59.229.104 on November 23, 2026 from IP address 72.59.229.104.

respectively, in part to combat household biomass burning and kerosene emissions (22, 23). However, numerous factors, including access and affordability, result in the nearly universal practice of "stacking," or utilizing multiple appliances and fuels simultaneously (e.g., a household may use both LPG and biomass for cooking), thus complicating the potential impact of the policies on clean air and energy use (24).

While considered the "gold standard" for air pollution observation, reference-grade ground monitoring stations are expensive, both in terms of capital cost and maintenance. Consequently, most deployments are in relatively well-resourced urban cores (25). Satellite remote sensing has offered unprecedented improvements in the spatial resolution of air pollution assessment (26). However, temporal coverage depends on clear skies (i.e., no or few clouds) and is limited to daytime retrievals (27). Furthermore, satellites do not directly measure $PM_{2.5}$. They report aerosol optical depth (AOD), a useful proxy for accumulation mode particles (0.1 μm < particle aerodynamic diameter [d_p] < 1 μm), which prior work has shown to be useful for inferring ground-level $PM_{2.5}$ (27). Both remote sensing products and reference-grade monitoring networks are typically too coarse to resolve the spatial and temporal differences in key sources, such as household biomass burning, municipal waste fires, diesel generators, and other small-scale sources of combustion emissions. Consequently, estimates of impact from common sources of unevenly distributed emissions are likely biased low (28).

Here, we used particulate matter lower-cost sensors (PM LCS) to develop a dense, large-scale community-based network. PM LCS infer mass concentration from aerosol light scattering, thereby offering a cost-effective (less than $500 USD) alternative to reference grade instruments (typically more than $10,000 USD) (29). Diverse applications of PM LCS networks include wildfire smoke transport and building infiltration, household air pollution exposure and epidemiology, and environmental justice analyses (30–32). LCS require intensive quality assurance and calibration techniques, especially at extreme mass concentrations (i.e., >100 μg/m^3) (33–36). Furthermore, many such monitors are sensitive to only a narrow subset of accumulation mode particles (0.3 μm < d_p < 1 μm) (37, 38). Despite these constraints, PM LCS networks have demonstrated the capacity to leverage shared information in the network with other data sources (e.g., reference monitors and satellite data) to produce high-fidelity exposure datasets (39–41).

In this study, we deployed an extensive network of PurpleAir PA-II monitors to characterize $PM_{2.5}$ dynamics along the rural-to-urban gradient in the surroundings of Delhi. Here, we adapted Dijkstra et al.'s (42) "degrees of urbanicity" framework using North India-specific population thresholds to classify sensors as "urban" (Delhi and its major satellite cities), and "rural" (*Materials and Methods, Urban–Rural Classification*).

We define our domain for the study as the square area derived by delineating the Northwest-Southeast transect as the diagonal (Fig. 1*A*) resulting in an area of approximately 12,800 km^2. Throughout, we refer to this delineated area as the "region." Since our transect is organized in alignment with the dominant upwind-downwind pattern, we treat "regional processes" or "regional scale" as synonymous with synoptic or those occurring on a scale of 10^2 to 10^3 km.

Our network design consisted of 49 sites (17 urban, 32 rural) which were principally organized along a 160 km transect following the dominant upwind-downwind pattern of synoptic meteorology in the region. Measurements from the network adhered to rigorous collocation calibration protocols [*Materials and Methods* and *SI Appendix*, Tables S1 and S2 (34)]. We also deployed additional monitors along a rural transect orthogonal to the main transect to quantify intrarural concentration gradients (*SI Appendix*, Tables S1 and S2). Our network operated at full capacity from July–December 2024, spanning the moderately polluted monsoon season (median monthly $PM_{2.5}$ = 51 ± 12 μg/m^3) through the sharp ramp-up to extreme pollution in the winter (median monthly $PM_{2.5}$ = 340 ± 53 μg/m^3) (*SI Appendix*, Fig. S3).

We focus on resolving differences between the rural and urban sites, which helps contextualize the impact of overlapping clean air and energy policy objectives and the fidelity of existing environmental exposure monitoring infrastructure. The temporal resolving power of our community-hosted network allows us to explore how these differences evolve along multiple time scales.

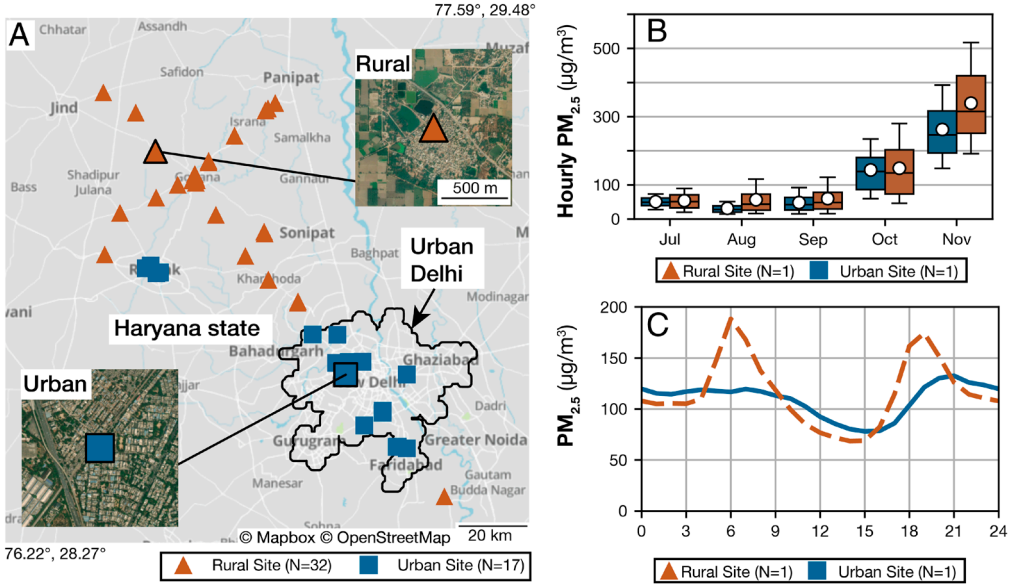


Fig. 1. (*A*) Map of the monitoring network with 32 rural (orange triangles) and 17 urban (blue squares) sites. Urban Delhi extent (as determined by population density) outlined in black. *Insets* show the land use around the two example sites (each had >93% uptime or 3,417 h of 3,672 total campaign hours). (*B*) Distribution of hourly $PM_{2.5}$ concentrations during each month of the campaign at the two example sites. Monthly boxplots show the time-averaged mean (circle), 10th (lower whisker), 25th (lower box line), 50th (line in box), 75th (upper box line), and 90th (upper whisker) percentile hourly concentrations for the two sites. (*C*) Average diurnal profile of $PM_{2.5}$ concentrations for the duration of campaign (152 d) for the example rural (orange) and urban (blue) sites.

Downloaded from https://www.pnas.org by 72.59.229.104 on November 23, 2026 from IP address 72.59.229.104.

We use a baseline-subtraction regression technique (43) to further decompose the total signal into a finely-resolved hyperlocal (i.e., <hourly timescale, ~10^{-1} km spatial scale) peak components from the dominant coarser (i.e., >hourly timescale, 10^0 to 10^2 km spatial scale) baseline components encompassing both regional pollution and dilute local plumes. Here, we define plumes as a localized dispersion of pollutants from a nearby source characterized by high temporal variability.

Results and Discussion

We report here on data from 49 sites which had a high data recovery rate (>50%) after being subjected to rigorous QA/QC (*Materials and Methods*). In general, the sensor network reproduced the expected seasonal trend of $PM_{2.5}$ concentrations increasing sharply from the relatively cleaner conditions during the monsoon season (July to early September; daily means 40 to 75 μg/m^3) to much more polluted levels starting in mid-October and November (daily means 150 to 500 μg/m^3).

Ambient $PM_{2.5}$ at Example Rural and Urban Sites. Urban and rural monitoring sites had systematically different air pollutant patterns (*SI Appendix*, Table S3). Fig. 1*A* shows the layout of the full study transect. In inset satellite imagery, we highlight two illustrative, contrasting network sites about 80 km apart: an urban site in Delhi (DL-NCT-3), and a rural village site (HR-VT-3A). Each site had nearly complete data recovery over the study period (i.e., 3417 hourly samples of 3,672 total campaign hours or ~93%). The urban site represents the complexity of pollution sources and land use in the Delhi urban core. Heavy industry, warehouses, traffic from both primary and secondary roadways, dense middle income housing, construction, and scattered greenery are within a square kilometer with a population of about 23,000 people (44). The rural site land use is more homogenous, with a cluster of small residential buildings within the approximately 250 m × 250 m Nuran Khera village, population ~3,000, surrounded primarily by farmland and tertiary roads (44).

While the rural site was "upwind" of Delhi by 80 km based on prevailing winds (*SI Appendix*, Figs. S4 and S5) and thus urban primary emissions are unlikely to have a major impact on the example rural site, we expect correlated $PM_{2.5}$ signals at sites across the region. This expectation holds because of similar primary sources (e.g., traffic congestion, heating), and synoptic scale meteorology (i.e., diurnal/seasonal patterns in atmospheric mixing/transport) and secondary formation processes. Local air pollution is inclusive of both "hyperlocal" household sources (<100 m from the sensor) as well as neighborhood sources (<1 km from the sensor).

Fig. 1*B* shows the hourly concentration distributions segregated by month, and stratified by urban vs. rural. The monthly means and medians of the hourly concentration distributions at rural sites are higher than or equivalent to urban concentrations. While both sites are well correlated throughout the campaign (Pearson's r = 0.87; n = 3,185), the median monthly difference in hourly concentrations between the two sites increases from about 10 μg/m^3 in July–October to 38 μg/m^3 in November. Additionally, the rural site features greater spread in hourly concentrations, with a higher range between the monthly 10th and 90th percentiles than the urban site. The mean diurnal trend in Fig. 1*C* reveals that the higher campaign median and interquartile range at the rural site is driven by sharper local maxima in the morning (05:00–08:00 local time [LT]) and evening (17:00–20:00 LT), but with a slightly lower midday trough (10:00–15:00 LT) at the rural site. As shown in the following sections, this contrasting pattern of sharp morning and evening maxima driving higher time-averaged concentrations was common throughout the rural sites in our domain (*SI Appendix*, Fig. S6).

Monthly $PM_{2.5}$ Trends. The full dataset from the entire sensor network (an average of 148 d/sensor) illustrates broadly similar diurnal and spatial patterns. Fig. 2 illustrates striking differences between urban and rural diurnal signals across all deployment months. In Fig. 2*A*, although the rural and urban sites are well correlated (median hourly Pearson's r = 0.96), the monthly rural median concentrations are persistently elevated above the urban median by 10 to 20 μg/m^3 (10 to 20%) in July–October and by over 40 μg/m^3 (11%) in November.

Fig. 2*B* further illustrates the rural vs. urban differences. Broadly, $PM_{2.5}$ concentrations in urban Delhi and some nearby rural settlements along the main transect are lower than the concentration further from the Delhi urban core. This suggests a "reverse" gradient in which the densely populated urban neighborhoods are exposed to lower outdoor $PM_{2.5}$ concentrations than rural sites. Bootstrapped estimates of 95% CI suggest that this pattern of higher concentrations at rural sites is highly robust to site selection and sampling error (*SI Appendix*, Fig. S7).

Differences in Diurnal Characteristics. We observed sharper divergence between urban and rural $PM_{2.5}$ when considering diurnal patterns (Fig. 3). We first consider the monsoon months

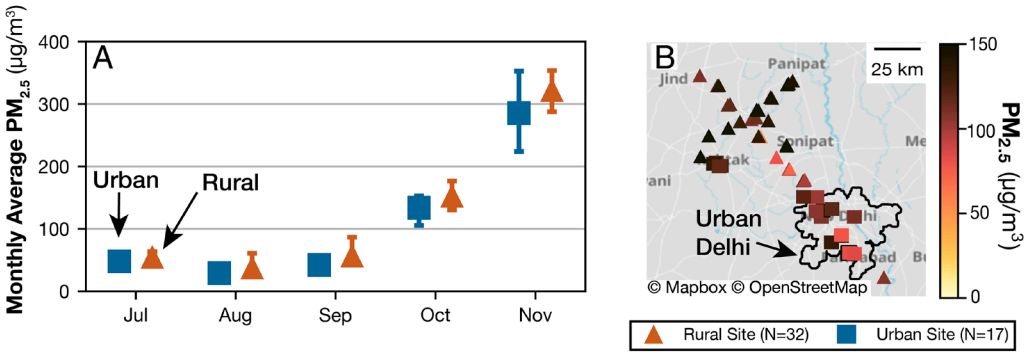


Fig. 2. Summary of campaign observations stratified by urban and rural sites. All summary statistics for the 49 sites (32 rural, 17 urban) are derived from distributions of hourly $PM_{2.5}$ data. Both rural and urban sites reported data for every hour of the 5-mo campaign (total of 3,672 h), with all sites reporting data more than 50% of the campaign. See *SI Appendix*, Tables S1 and S2 for details on exact sample size on a per site basis. (*A*) Distribution of the monthly mean concentrations among all sites, with the plotted range indicating the 10th and 90th percentile sites. (*B*) Map of the entire network, with markers indicating the settlement type and colored by campaign mean mass concentration. Both in aggregate and monthly, the rural subnetwork $PM_{2.5}$ concentrations exceed the urban subnetwork concentrations.

Downloaded from https://www.pnas.org by 72.59.229.104 on November 23, 2026 from IP address 72.59.229.104.

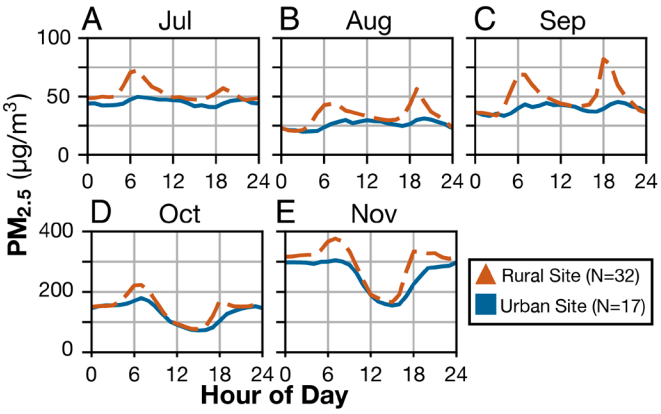


Fig. 3. Mean diurnal trend for July (*A*), August (*B*), September (*C*), October (*D*), and November (*E*), stratified by urban (17 sites) and rural (32 sites). All sites reporting data more than 50% of the campaign (total of 3,672 h), see *SI Appendix*, Tables S1 and S2 for details on exact sample size on a per site basis. On a diurnal basis, the daily maximum values occurring during the morning and evening hours are much higher at the rural sites than urban sites. However, at midday during the daily trough, values are similar at rural and urban sites.

(e.g., July–September). During this wet season in North India, biomass burning may be less common (e.g., less demand for heating, and less large-scale crop residue burning), and concentrations approach their annual minima. Diurnal profiles for the urban sites are relatively flat, with hourly variation of 15 to 30% (~ 10 μg/m^3) around the daily mean. At rural sites, in contrast, there is more pronounced diurnal variation. The rural sites consistently show sharp morning and evening maxima (respectively 05:00–08:00 and 17:00–20:00 LT). In the monsoon months, these two diurnal maxima at rural sites reach or exceed 1.5× the daily mean (i.e., +30 μg/m^3). Given that concentrations at mid-day (12:00–14:00 LT) and late night (23:00–01:00 LT) are roughly similar at rural and urban sites, we conclude that it is the morning and evening maxima that drive much of the difference in time-averaged concentrations between urban and rural sites.

In the postmonsoon months (October, November), the difference between the rural and urban diurnal patterns was larger in absolute magnitude compared to the monsoon months but was smaller in relative magnitude, given the generally higher concentrations during these heavily polluted months. Both rural and urban sites feature high morning maxima ([median ± interquartile range]; rural: 301 ± 148 μg/m^3; urban: 252 ± 168 μg/m^3), with somewhat less pronounced evening maxima (rural: 263 ± 159 μg/m^3; urban 210 ± 160 μg/m^3). Thus, the diurnal maxima in morning and evening were typically ~40 to 50 μg/m^3 higher at rural sites compared to urban sites in October and November. Nonetheless, during the mid-day concentration trough, when atmospheric dilution is at its strongest, concentrations were very similar at urban and rural sites.

Signal Decomposition: Peak and Baseline Components. We hypothesized that the sharply elevated morning and evening concentration maxima at rural sites are suggestive of a hyperlocal household air pollution signal that may be more prevalent in village settings than in urban Delhi and its surrounding cities. Our sensor network was designed to capture residential conditions that are relevant for human exposure. We intentionally deviated from regulatory monitor siting guidelines, which often dictate that measurements only be conducted in sheltered locations (i.e., parks) where hyperlocal plumes are unlikely to impact a monitor. Source-free sites are impractical given the abundance of biomass burning in rural areas. Furthermore, sheltered locations are also not representative of human exposures in much of Delhi. Unsurprisingly, the high-resolution (2 min) time series from across the network suggest that hyperlocal-scale combustion emissions (e.g., transportation, cooking, waste burning, heating, and other solid fuel use) substantially impacted exposures at many measurement locations.

We undertook the following analysis to quantify the roles that fresh plumes of primary $PM_{2.5}$ play in explaining the differing temporal patterns of $PM_{2.5}$ at urban and rural sites in our network. First, we applied a baseline subtraction regression technique previously documented (43) and validated in atmospheric observation studies to separate the high-frequency signal correlated with hyperlocal plumes from the more slowly varying signal that is representative of larger spatial scales inclusive of both diluted local plumes, secondary chemical formation, and regionally transported primary $PM_{2.5}$. For each site, we additively decomposed the PurpleAir 2-min $PM_{2.5}$ signal into subhourly short-term variations, which we refer to as "peak component $PM_{2.5}$" and longer daily variability, which we refer to as "baseline component $PM_{2.5}$" (*Materials and Methods* and *SI Appendix*, Fig. S8).

Separating $PM_{2.5}$ into baseline and peak components reveals how the peak component $PM_{2.5}$ contributes to driving the divergent patterns between rural and urban sites (Fig. 4 and *SI Appendix*, Figs. S9 and S10). Over all months of the study, the average baseline component $PM_{2.5}$ is only moderately elevated (+3 μg m^{-3}, +3%) at rural sites compared to urban sites, while the peak component is nearly twice as large (+13 μg m^{-3}, +81%) at rural sites compared to urban sites (*SI Appendix*, Fig. S11). Averaged over the campaign, we find that peak component $PM_{2.5}$ drives over 80% of the urban–rural concentration difference.

The diurnal cycles of baseline and peak components at rural and urban sites diverge in ways that lend further insight into the processes that are associated with the higher overall average $PM_{2.5}$ concentrations at rural sites (Fig. 4). Concentrations of peak component $PM_{2.5}$ are higher at rural than urban sites throughout the diurnal cycle (*SI Appendix*, Figs. S11 and S12). Notably, there is not a strong, consistent diurnal cycle in peak component $PM_{2.5}$ at our urban sites. While we observe some monthly variation in diurnal cycles of $PM_{2.5}$ among the rural sites (*SI Appendix*, Fig. S12*C*), the sharpest contrasts between rural and urban sites occur during the morning (06:00 LT) and evening (18:00 LT) windows across all seasons, in which peak component $PM_{2.5}$ rural sites exceeds urban site exposure by 2.5 to 3×. The peak component $PM_{2.5}$ reaches its highest values in the evening, especially in the postmonsoon months, where the most extreme rural sites show average monthly peak component contributions well in excess of 75 to 100 μg/m^3 in the evening (Fig. 4*H*).

While the peak $PM_{2.5}$ component drives the overall rural–urban enhancement, it is the baseline $PM_{2.5}$ component that more strongly modulates the diurnal and seasonal cycle of $PM_{2.5}$ (Fig. 4 and *SI Appendix*, Fig. S13). At rural and urban sites alike, the most marked feature of the baseline component $PM_{2.5}$ signal is its sharp (>5×) seasonal increase from July through November, reflecting the transition into the more stagnant meteorological conditions of the North Indian postmonsoon. Interestingly, the diurnal variation in the baseline $PM_{2.5}$ component is considerably stronger at the rural sites, with sharper peaks and deeper mid-day troughs as compared to the urban sites (Fig. 4 and *SI Appendix*, Fig. S13). Baseline component $PM_{2.5}$ is higher at rural sites during the morning and evening hours, but at midday is slightly lower at rural sites compared to urban sites (Fig. 4*A* and *SI Appendix*, Figs. S11 and S13). We expect that diluted plumes from morning and evening

Downloaded from https://www.pnas.org by 72.59.229.104 on November 23, 2026 from IP address 72.59.229.104.

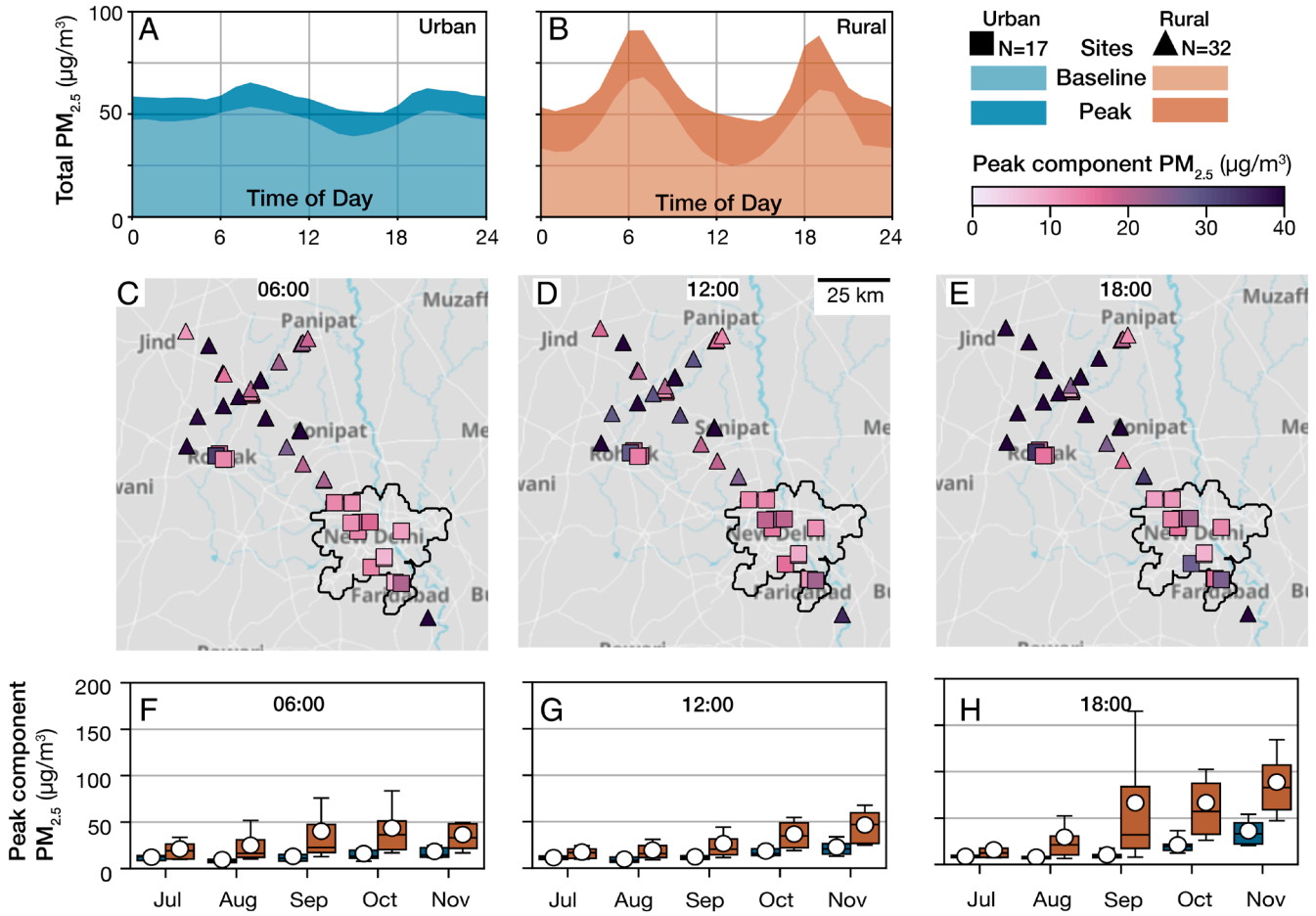


Fig. 4. Comparison of peak component $PM_{2.5}$ urban and rural site distributions at morning (06:00 LT), noon (12:00 LT), and evening (18:00 LT). (*A* and *B*) Decomposition of the diurnal profile into peak and baseline components for urban vs. rural sites. Data is the diurnal average for the duration of the campaign, 5-mo (3,672 h), see *SI Appendix*, Tables S1 and S2 for a site-wise breakdown of sample size. Note that larger peak component and greater diurnal variability for the rural sites (*B*) relative to the urban sites (*A*). (*C*–*E*) Comparison of campaign peak component $PM_{2.5}$ median in the morning (*C*), noon (*D*), and evening (*E*), respectively. (*F*–*H*) Morning (*F*), noon (*G*), and evening (*H*) monthly distributions of the monthly mean of hourly peak component $PM_{2.5}$, stratified by urban and rural sites. Note the considerably higher peak component $PM_{2.5}$ concentrations for the rural sites especially in evening hours.

cooking emissions may explain the observed diurnal differences in baseline $PM_{2.5}$. Additional possible explanations that would be consistent with our data could include rural–urban differences in underlying meteorology (e.g., stronger rural nighttime inversions caused by both the absence of a strong urban heat island effect and presence of agricultural irrigation), or the stronger effect of atmospheric chemistry in driving mid-day secondary PM formation at urban sites.

The overall pattern that emerges from this decomposition is that the peak component $PM_{2.5}$ is consistently and substantially elevated for rural sites—and especially so at some hours of the day—whereas the baseline component shows greater diurnal dynamic range that swings between urban and rural enhancements (*SI Appendix*, Fig. S11). Peak component $PM_{2.5}$ shows weak hourly correlation between rural and urban sites (monthly Pearson $R \sim 0.26$ to 0.46), whereas there is a stronger correlation of the baseline component $PM_{2.5}$ (monthly Pearson $R \sim 0.71$ to 0.90, *SI Appendix*, Table S4). To further illustrate the differences in the peak component $PM_{2.5}$ between urban and rural sites, Fig. 5 contrasts the population density profile across the NW/SE study transect against these localized-short term $PM_{2.5}$ enhancements. Fig. 5 notably emphasizes how the peak component $PM_{2.5}$ is consistently low at urban sites, and is elevated—but more heterogeneous—at rural sites.

These quantitative findings align with our understanding of on-the-ground realities in the villages and urban neighborhoods that formed our study sites. Many of the villages along our upwind transect, while lower than urban Delhi in total population, comprise tightly packed housing units. Nearly all the buildings that served as study sites were impacted by extensive combustion emissions within the surrounding 50 to 200 m. Biomass burning for cooking, water/space heating, and animal husbandry (e.g., fodder preparation) is ubiquitous. These activities take place disproportionately during the morning and evening, matching the observed temporal profiles of the peak component $PM_{2.5}$. There are also other localized combustion sources prevalent in rural areas, including small-scale industries, brick kilns, burning of municipal waste, and older vehicles that are no longer permitted to operate within the urban areas of Delhi. Our sensor siting strategy thus reflects the reality that localized sources of particulate matter strongly affect community-scale exposures in villages. In short, smoke plumes are a ubiquitous fact of daily life and breath.

In contrast, the situation in urban Delhi may be somewhat more complex. While urban and rural sites are impacted by high baseline component $PM_{2.5}$, the differences in peak component $PM_{2.5}$ indicate key differences in the relative importance of household sources. In more affluent parts of Delhi, domestic biomass burning is less common than at rural sites. Traffic, construction,

Downloaded from https://www.pnas.org by 72.59.229.104 on November 23, 2026 from IP address 72.59.229.104.

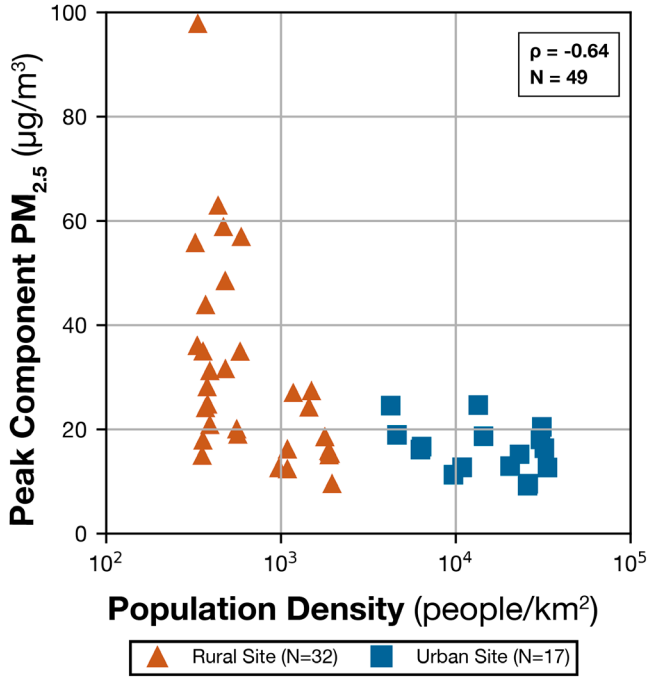


Fig. 5. Scatterplot of site-wise campaign (July–November; 3,672 campaign hours, see *SI Appendix*, Tables S1 and S2 for site-wise sample size breakdown) median peak component $PM_{2.5}$ vs. 5-km buffer of 1-km^2 population density cells. Rural sites feature higher heterogeneity in median peak component $PM_{2.5,}$ than the more tightly clustered urban sites. Generally, rural sites experience higher peak component $PM_{2.5}$. Consequently, Peak Component PM2.5 is moderately anticorrelated with population-density, meaning local activity is stronger at our community sites in rural (i.e., relatively population sparse) sites. Two very low population-density rural sites are omitted by the axis limits but contribute to the correlation metric calculation, see *SI Appendix*, Tables S1 and S2.

and commercial sources are the largest local primary $PM_{2.5}$ emissions source. In other lower-income neighborhoods of Delhi, domestic biomass burning remains somewhat prevalent, and unregulated small-scale industrial activity is intermingled within residential areas (8). These heterogeneous neighborhood conditions are evident in the spatial distribution of peak component $PM_{2.5}$ across Delhi (Fig. 4 *A–C*). For example, there were higher peak component contributions in the lower-income neighborhood sites in northwest Delhi, with lower levels in the affluent parts of South Delhi.

Role of Crop Stubble Burning. Much academic and policy attention has focused on crop stubble burning in Northwest India, which can produce intense pollution episodes approaching or exceeding 1,000 μg/m^3. Large-scale biomass burning is triggered by the transition from kharif (e.g., rice, maize, sugarcane) to rabi (e.g., wheat, barley, peas) crops in October and November (8, 45). During our study year, there was a roughly 5-wk period from October through late November with strong stubble-burning impacts, including a more extreme 2-wk interval [*SI Appendix*, Figs. S14 and S15; (46)]. Although daily average concentrations were elevated regionally during these episodes, the morning and evening gradient from more polluted rural upwind sites to less polluted urban downwind sites persisted (*SI Appendix*, Figs. S14 and S15). Midday trough data still suggested the opposite pattern, with urban sites narrowly more polluted than rural sites by about 3 to 10%, even after including postmonsoon burn data (*SI Appendix*, Fig. S10). Thus, sources of elevated rural $PM_{2.5}$ are unlikely to be only driven by agricultural stubble burning. This is also consistent with prior source apportionment work in Delhi, which shows that aged (i.e., multiple days to weeks) regional extreme pollution persists beyond the agricultural burning period (5, 8, 47).

Satellite Remote Sensing Comparison. Our core findings of higher $PM_{2.5}$ in rural settlements runs contrary to the perception, often inferred from satellite data, that urban areas are generally air pollution hotspots. Complimentary evidence from other work in South Asia, supports our results and has also stressed the urgency of source abatement in rural settlements (48, 49). It is therefore worthwhile to compare our results against other available observational datasets. Satellite remote sensing data (i.e., using aerosol optical depth; AOD) are arguably the most employed observational product for characterizing air pollution in regions of the world where in situ measurements are lacking. Fig. 6 presents a comparison of our results against two commonly utilized satellite remote sensing products: MAIAC AOD and the Atmospheric Composition and Analysis Group (ACAG, Washington University, St. Louis, MO) satellite $PM_{2.5}$ product (9, 26, 50, 51). Intriguingly, our findings diverge from these remote sensing products and point to possible temporal biases in the use of remote sensing products for understanding rural–urban gradients.

Fig. 6*A* shows the seasonal spatial distribution of $PM_{2.5}$ for 2022, the most recently available ACAG satellite-derived product as of writing, and representative of reported spatial-temporal trends in the 2020s. In contrast to our core result, the product reports the highest $PM_{2.5}$ concentrations are in the urban core. Remote sensing products are built on AOD measurements aggregated from repeated satellite overpasses during midday hours, often close to the local solar noon (10, 27, 52, 53). We therefore assert the remotely sensed $PM_{2.5}$ product best represents midday conditions. At midday our networkreported muted urban–rural differences (Fig. 3), with the background component of $PM_{2.5}$ higher at urban sites (*SI Appendix*, Fig. S10).

Next, we compared each site's time series of $PM_{2.5}$ measurements against the corresponding 1 km pixel value of daily overpass MAIAC AOD (Fig. 6). For this analysis, we focus only on the postmonsoon months (October and November) due to poor data availability in the cloudy monsoon months (July, August, September). We note that the $PM_{2.5}$-AOD relationships derived here may not transfer well to other seasons due to the unique $PM_{2.5}$ vertical atmospheric distribution during the onset of the agricultural burning season (see *Materials and Methods* subsection *Remote Sensing Data Processing* for more information on data quality assurance). Rather than comparing daily mean concentrations with the overpass AOD, it is more instructive to compare the in situ morning, noon, and evening concentrations with it. Stratifying by urban vs. rural sites reveals three key points. First, the in situ $PM_{2.5}$- to-AOD relationship is steeper in the mornings and in the evenings, reflecting the expected diurnal signal. Second, we find higher correlation (Pearson's $r \sim 0.79$ [rural], $r = 0.78$ [urban]) at noontime than in the mornings (0.54, 0.65) and evenings (0.37, 0.63). Third, and crucially, we note the overpass (i.e., approximately noontime) AOD has notably weaker correlation with the morning and evening concentrations at the rural sites, largely driven by the datapoints with a large peak component $PM_{2.5}$.

Several factors may account for the differences between ground-based sensors and satellite products. Aerosol optical depth observations acquired near solar noon may not resolve urban–rural differences in $PM_{2.5}$ dynamics that emerge during the morning and evening. In addition, the spatial resolution of currently available satellite products, approximately 1 km, may be too coarse to capture the contributions of small, ground-level combustion

Downloaded from https://www.pnas.org by 72.59.229.104 on November 23, 2026 from IP address 72.59.229.104.

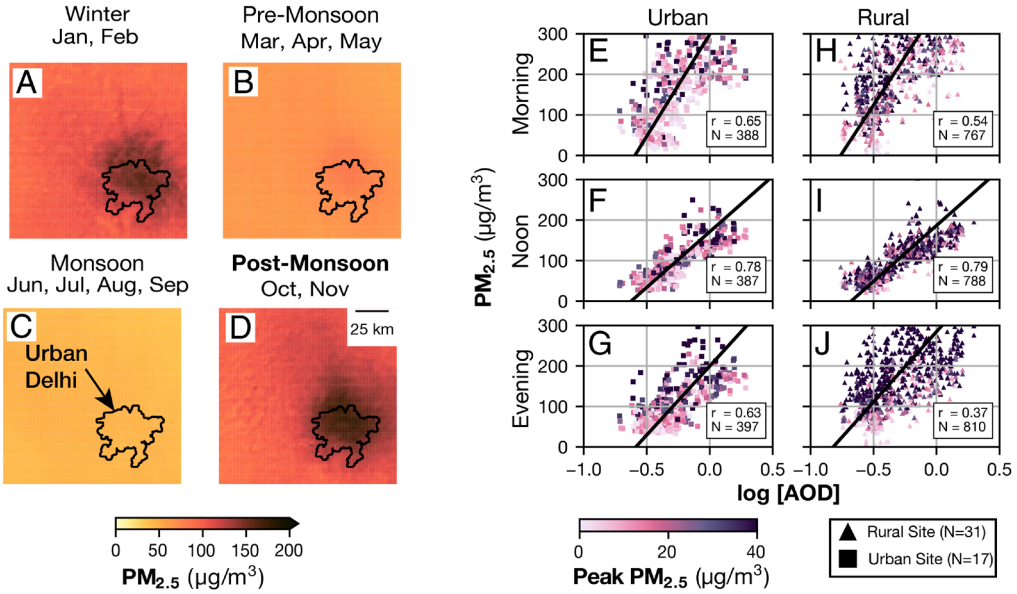


Fig. 6. Historical remote sensing data assimilation $PM_{2.5}$ product for year 2022, and comparison of satellite-derived AOD and the community-based LCS network. (*A–D*) One rural sensor did not report any valid paired data, resulting in 31 rural sites, rather than 32. Seasonal mean $PM_{2.5}$ for the Delhi region for the year 2022. In every season, urban Delhi is a clear regional hotspot especially in the postmonsoon (*D*). (*E–J*) Comparison for the postmonsoon of in situ $PM_{2.5}$ measurements against coincident daily satellite AOD. Plots are presented separately for urban (*E–G*) and rural sites (*H–J*) at three times of day (hourly means at: 06:00 [*E* and *H*], 12:00 [*F* and *I*] and 18:00 [*G* and *J*]). Each data point represents one day's spatially coincident measurement of the daily satellite AOD (collected at the midday satellite overpass), as compared against the in situ measurement at a specific time of day (06:00, 12:00, or 18:00). Markers are colored according to each site's estimate of the peak component $PM_{2.5}$. Best fit lines are determined by total-least-squares regression. Both rural and urban sites demonstrate strong linearity with AOD in the noon, with notable declines in correlation in the other windows. Furthermore, the rural sites with their higher peak component $PM_{2.5}$ demonstrate both poorer correlation as well as steeper slopes, indicating a fundamentally different AOD-$PM_{2.5}$ relationship that results from the sharp morning and evening maxima that are not detected by the midday satellite observations.

sources that substantially influence $PM_{2.5}$ exposure in many villages. In contrast, the in situ sensors in our network are often placed within tens or hundreds of meters of these sources and are thus more sensitive to their impact. Whereas current satellite remote sensing excels at capturing the mid-day spatial patterns at coarse spatial resolution across a large domain, this study's in situ measurements with small sensors can offer a complementary perspective about the hyperlocal exposure impacts of ubiquitous combustion emissions that are prevalent in the study region. Future work could build on the complementary strengths of small sensors and novel satellite products to build exposure products exploiting the strengths of both observational methods.

Implications. Aggregating across the campaign duration, rural sites experienced higher peak component $PM_{2.5}$ than urban sites by around 12.5 ± 3.3 μg/m^3 or a factor of 2×. Both rural and urban sites show monthly aggregated peak component $PM_{2.5}$ exposure rapidly increasing (Fig. 4 and *SI Appendix*, Fig. S12) during the transition from monsoon (July, August, September) to postmonsoon/winter (October, November). However, rural sites are exposed to higher peak component $PM_{2.5}$ across all months, and the absolute difference between the rural and urban site peak component exposure increases (July: 6.06 ± 1.21 μg/m^3, November: 13.0 ± 3.62 μg/m^3). The sharpest contrasts between rural and urban sites occurs during the morning and evening windows across all seasons, in which peak component exposure at rural sites exceeds urban site exposure by 2.5 to 3×. The higher frequency of short-term, localized plumes at rural sites likely indicates a higher prevalence of household pollution, such as biomass burning given the diurnal pattern.

While LCS networks enable dense spatial coverage that would be difficult to achieve with regulatory-grade monitors, they have key limitations. The Plantower sensor used in the PurpleAir node estimates concentration from particle counts but does not fully capture all particles across the size distribution, particularly ultrafine (d_p < 0.1 μm) and supramicron (d_p > 1 μm) particles (29, 37, 38). As a result, concentration estimates from Plantower-based monitors may be biased low for some sources, especially when local emissions are dominated by particle sizes that fall outside the instrument's most sensitive range. Prior work in the Delhi region suggests that the ultrafine fraction from primary emissions is substantial during morning and evening peaks (5), but these limitations are unlikely to introduce a directional bias that would alter our main conclusions regarding rural–urban gradients.

Another significant source of uncertainty in LCS campaigns arises from siting principles. Often, analyses of PurpleAir crowdsourced data rely on the high density of the network to identify poorly performing or improperly placed nodes (e.g., marked "outdoors" but placed indoors) (30, 54). Our network design relied on intentionally placed sensors following a rigorous protocol rather than crowdsourced nodes, alleviating the risk of sensor placement in nonideal locations, such as a kitchen or driveway. However, given the diversity of land use and sources in Delhi city proper, it is probable that our network misses urban "hotspots" such as traffic intersections, garbage dumps, and chemical manufacturing plants (55). Similarly, rural sites are impacted by local sources (e.g., unregulated older brick kilns, local traffic, and small construction sites). Therefore, we limit our interpretation of findings to community-scale exposure rather than source-specific findings.

This study demonstrates that concentrations of $PM_{2.5}$ in the rural areas surrounding Delhi may be even higher than in the urban core. Our community-based residential low-cost sensor network measured short-term variability driven by local sources, particularly household emissions, that remote sensing products and existing regulatory monitors systematically underestimate.

Downloaded from https://www.pnas.org by 72.59.229.104 on November 23, 2026 from IP address 72.59.229.104.

These ground-level measurements revealed characteristic diurnal and subhourly pollution patterns in rural communities that differ strongly from urban trends. Our work further demonstrates how LCS can provide critical data that complements existing regulatory networks and satellite monitoring capabilities. Given the ubiquity of domestic biomass burning in rural areas throughout the heavily polluted Indo-Gangetic Plains, future work could use a similar design to validate the generalizability of our findings beyond Delhi.

Clean air policy should focus on enforcement beyond urban Delhi. Under the existing National Clean Air Programme, enforcement of bans on biomass burning for small scale manufacturing has been beneficial for reducing the burden of local primary $PM_{2.5}$. Rural areas would benefit from regulating this practice, and in turn transport less $PM_{2.5}$ and precursors to downwind urban Delhi. In addition to focusing on sources common to both urban and rural environments, air quality across the region would benefit from focusing on sources disproportionately represented in rural environments. Our analysis shows that household air pollution continues to represent a key threat to environmental health at rural residences and requires tailored solutions not possible by merely extending existing urban Delhi regulations to rural settlements. Finally, our network corroborates extensive previous work demonstrating the seasonal impact of crop residue burning and its deleterious impact across rural and urban settlements. We also note that household air pollution continues to feature as a key contributor to higher $PM_{2.5}$ exposure at rural sites during the November crop burning period. Our findings may be relevant to the many other relatively understudied nonattainment regions of India with similar urban–rural dynamics.

Materials and Methods

Urban–Rural Classification. Here we distinguish between rural and urban by building on the "degrees of urbanicity" methodology from Dijkstra et al. (42). Although we employ different thresholds in defining high-density cells, we do so because of the uniquely high total population of North India (~500 million people). We first identify all cells with population greater than 10^4 people/km^2 (approximately the 95th percentile of the region including 0 population cells). Next, we concatenate all adjacent cells to identify the extent of "urban centers" or agglomerates. Finally, we close all concavities (e.g., parks, highways, landfills) within the agglomerate polygons. In this way, we identified the extent of the "Delhi Urban Core" (shown in regional maps in main text Figs. 1, 2, 4, and 6), a combination of the contiguous population dense areas of Delhi and its satellite cities, most notably Gurugram and Faridabad in Haryana state, and Noida in Uttar Pradesh state.

Network Design and Sensor Siting Strategy. We established a purposefully sited sensor network consisting of PurpleAir PA-II sensors distributed across the regions upwind and downwind of the urban core of Delhi. We prioritized a systematic, researcher-led study design to avoid possible pitfalls in inference that come from analyses of crowdsourced data. As of 2025, there are only 5 to 10 active PurpleAir sensors on the public-facing map in the Delhi region, most of which show poor quality assurance metrics as reported by PurpleAir. The core of our network design consisted of sites along two approximately perpendicular transects. The main "upwind-downwind" transect (total length 160 km) follows the typical prevailing winds from the rural regions to the northwest of the city (90 km), crosses through the increasingly urbanized regions of Delhi (50 km), and exits into downwind rural regions (30 km). A second crosswind transect (total length: 60 km) spanned the rural areas upwind of Delhi between two smaller cities, Rohtak (population ~400 k) and Panipat (~300 k). Along each transect, sensor sites were evenly spaced in approximately every 10 km, with sensors generally deployed in closely spaced pairs (within 100 to 300 m) to reduce the risk of influential but nonrepresentative sites. To facilitate site access and continuous electrical supply, measurement sites were generally located in residential areas within villages, towns, and cities. We supplemented measurements from these two transects with several sites in urban and peri-urban areas of Delhi, Rohtak, Panipat, and the town of Jind (~65,000 people). Finally, we placed two monitors in village (Bhatgaon, ~8,000 people) "off-axis" from the main transects to ensure the spatial-temporal patterns we observe are not unique to our transect.

We placed all sensors between the top of the ground floor (~3 m) and the mid-level of the second floor (~7.5 m). We ensured all sensors were placed outdoors, from a balcony or attached to a pole outside the house or apartment. The outdoor location was also specified to be away from adjacent local sources, such as kitchens, parking spaces, and tobacco smoking areas. Sensors provided real-time connectivity for online QA/QC via a cellular internet hotspot and were backed up using USB-powered battery banks to reduce the influence of sporadic power outages. We would offer a replacement sensor when persistent downtime or internal sensor disagreement were detected at the site. In a small number of cases (n = 2), the hosts could not support the campaign for the entirety of the study period, so we started new sites within about 1 km of the original site.

Our location selection criteria and siting guidelines helped reduce the bias that typically comes from using community-based sampling for ambient $PM_{2.5}$ measurements. We still expect the noncollocated LCS sites to measure higher concentrations than the nearest reference site since they are purposefully designed to avoid exaggerating the impact of any individual source. As a result, regulatory bodies in India typically choose parks or large university campuses to house their sites, resulting in lower measurements of $PM_{2.5}$ than a factory district or dense apartment complex (*SI Appendix*, Fig. S16). However, since much of the total variability of the $PM_{2.5}$ signal driven by processes (e.g., photochemical secondary formation) on larger spatial scales (>1 km) than relative to other pollutants, such as NO, we found our community LCS sites to be strongly correlated with reference sites (*SI Appendix*, Fig. S16).

Data Quality Assurance. Quality assurance (QA) procedures for LCS differ by pollutant, application, and measurement technology. Since the PurpleAir PA-II includes two replicate Plantower PMS5003 sensors to provide an internal quality assurance standard, it is common practice to compute the "disagreement" between the "A" and "B" Plantower sensors as shown in Equation A as a proxy for sensor degradation (56).

$$Disagreement = \frac{2 \times |A-B|}{A+B} \times 100\,\%. \quad [1]$$

Here, two issues are at play. On one hand, a permissive (high) data disagreement tolerance brings the risk of introducing biased data from malfunctioning sensors. On the other hand, a stringent (low) data disagreement tolerance produces a spatiotemporally unrepresentative dataset, because many sites and time periods could be censored. We found that a choice of a maximum of 30% hourly disagreement provided a robust tradeoff between maintaining robust internal node consistency and maximizing the number of valid data points (*SI Appendix*, Fig. S17). With this criterion, the median disagreement of sensors across the urban and rural networks never exceeded 20%, with less than 10% of all data points removed through the quality assurance process (*SI Appendix*, Fig. S17).

For a given site to reliably report monthly medians, we required a minimum of 50% uptime in each month. While our criteria are lower than the recommended level by US-EPA of 75%, we found through bootstrap analysis that we were able to recover qualitatively identical findings accounting for empirical CI (*SI Appendix*, Fig. S7). This result is largely due to the redundancy in our network design, meaning a single node's behavior does not strongly impact the subnetwork or transect distribution.

Calibration Model Selection. Previous work across diverse aerosol regimes has shown that the uncalibrated Plantower signal is highly precise, and strongly correlated with reference measurements of PM_1 and $PM_{2.5}$ ($r \geq 0.9$), although it is biased (33, 34). Our previous work in North India has demonstrated that Plantower instruments can be effectively calibrated with a relatively simple two-parameter linear regression model of the form (34):

$$C = \beta_P \times P + \beta_{RH} \times RH + \beta_0, \quad [2]$$

where C represents the corrected signal, P represents the $PM_{2.5}$ data from the PMS5003, RH means the relative humidity from a collocated sensor (as is the case with the PA-II), and the terms represent the regression coefficients fit by collocating the LCS with a reference-grade $PM_{2.5}$ monitor (*SI Appendix*, Table S5).

Downloaded from https://www.pnas.org by 72.59.229.104 on November 23, 2026 from IP address 72.59.229.104.

In Campmier et al. (34), an exhaustive approach was employed to iterate on many model forms and covariates to converge on a parsimonious and accurate model for PurpleAir units deployed in Delhi city center and rural North India, remarkably similar to the continental US models developed and deployed by the US-EPA (57, 58). The Delhi city-specific model reduced normalized bias to near 0% and constrained normalized RMSE on test data to under 35%. Evaluating on rural data resulted in more robust performance, with NRMSE below 20%. All linear models effectively reduced monthly bias to nearly 0% compared to bias as high as 70% in the uncalibrated data. In the supplement, we demonstrate that LCS calibrated with the Delhi model from Campmier et al. (34) report strongly collinear and low bias data relative to proximate reference monitors in urban Delhi (*SI Appendix*, Fig. S16).

Decomposition Model Selection. To decompose the $PM_{2.5}$ LCS signal into "peak" and "baseline" components, we adapt the methodology of Brantley et al. (43, 59). First, we estimate the baseline function by fitting a 10th percentile quantile regression with cubic spline smoothing on each day of sampling for each monitor separately. After subtracting the baseline function from each sampling day, the remainder distributions represent the "peak component" relative to the diurnal-daily baseline component (*SI Appendix*, Fig. S8). While this statistical technique may not perfectly fit the baseline component function on all days for all pollutants, validation exercises have found its performance comparable to more complex techniques and less biased than a moving percentile filter (59). The workflow follows the well-established mathematical framework previously published (43, 60, 61).

While the exact choice of quantile is subjective, sensitivity analysis has shown p_{10} is a consistent choice across pollutants (43). Given the 2-min input, the output of the quantile regressor can easily overfit, hence motivating the natural cubic spline smoothing. Comparison with meteorological data shows that peak component $PM_{2.5}$ is likely strongly correlated with local emissions activity, while baseline component $PM_{2.5}$ best captures the regional and synoptic scale processes, such as the influence of the boundary layer height (Spearman correlation [ρ] = 0.54 vs ρ = 0.28 for baseline component-PBLH vs peak component-PBLH).

Remote Sensing Data Processing. From the NASA Earthdata archive, we downloaded the Level-2 550 nm AOD acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on the Terra and Aqua satellites and processed by the Multiangle Implementation of Atmospheric Correction (MAIAC) algorithm. Specifically, we used the MCD19A2 V6.1 product, which reports one data point per day on a 1-km^2 grid (51). Terra and Aqua daytime flyover times correspond to approximately 10:30 LT and 13:30 LT, respectively. Since the MODIS-MAIAC dataset features many gaps (e.g., cloud cover), we averaged individual single days with multiple measurements into a single representative value for each day. This approach is commonly deployed given the relatively low variance in hourly AOD in the 10:00–14:00 LT window from ground-based AOD measurements.

MODIS-MAIAC data availability is poor during the rainy monsoon months (July, August, September) (51). Hence, we limit our investigation using the MODIS-MAIAC dataset to October and November. In addition to the intensive calibration and quality control procedures in the MAIAC algorithm, we excluded AOD measurements exceeding 2, coinciding with extreme episodes when the vertical aerosol distribution leads to unusually low accuracy ground-based comparisons, removing approximately 4% of the AOD data. When comparing MAIAC–MODIS AOD to our LCS network, we matched satellite granules to sensors within the same 1-km^2 grid cell. In the case of multiple sensors within the same pixel, we considered the pairings as independent comparisons since we are interested in not only the total $PM_{2.5}$-AOD relationship, but also the impact of peak component $PM_{2.5}$-AOD on the linearity of the $PM_{2.5}$-AOD correlation.

Data, Materials, and Software Availability. Data used in this study are freely available from PurpleAir, the Central Pollution Control Board, the Global Human Settlement Project, the Atmospheric Composition Analysis Group at Washington University, the National Aeronautics and Space Administration, the National Oceanic and Atmospheric Administration, AI for Good, and the European Centre for Medium-Range Weather Forecasts. More details are provided in the *SI Appendix* of the Manuscript. We have deposited PurpleAir sensor data, and code on Dryad at doi.org/10.5061/dryad.2rbnzs86h (9, 44, 51, 62–67).

ACKNOWLEDGMENTS. We are grateful to Open Philanthropy, Global Health Emerging Scholars (NIH FIC TW010540), and The Center on Contemporary India at UC Berkeley for their support. We are grateful to our community hosts for their participation in the study, as well as their patience and flexibility with project maintenance. We would like to thank our community liaisons: Kapil Dev, Rajkumar Dhiryan, Ashish Gupta, Ankush Kumar, and Rishabh Kumar. We are thankful to IIT Delhi, and CSTEP for institutional support. We are thankful to PurpleAir, CPCB, DPCC, and HSPCB for data access.

Author affiliations: [a]Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA 94720; [b]Centre for Atmospheric Sciences, Indian Institute of Technology Delhi, New Delhi, Delhi 110016, India; [c]Ecosystem and Environment Management Area, Indian Institute of Forest Management, Bhopal 462003, India; [d]Center for Study of Science, Technology and Policy, Bengaluru 560094, India; [e]Indian Council of Medical Research Center for Advanced Research on Air Quality, Climate and Health, Department of Environmental Health Engineering, Sri Ramachandra Institute for Higher Education and Research, Chennai 600116, India; [f]Department of Health, Policy and Management, Korea University, Seoul 02841, South Korea; and [g]Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720

1. C. J. L. Murray *et al.*, Global burden of 87 risk factors in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *Lancet (London, England)* **396**, 1223–1249 (2020).
2. A. Pandey *et al.*, Health and economic impact of air pollution in the states of India: The Global Burden of Disease Study 2019. *Lancet Planet. Health* **5**, e25–e38 (2021).
3. T. Vos *et al.*, Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1204–1222 (2020).
4. J. S. Apte, M. Brauer, A. J. Cohen, M. Ezzati, C. A. Pope, Ambient PM2.5 reduces global and regional life expectancy. *Environ. Sci. Technol. Lett.* **5**, 546–551 (2018).
5. S. Gani *et al.*, Submicron aerosol composition in the world's most polluted megacity: The Delhi Aerosol Supersite study. *Atmos. Chem. Phys. Katlenburg-Lindau* **19**, 6843–6859 (2019).
6. U. C. Dumka *et al.*, Year-long variability of the fossil fuel and wood burning black carbon components at a rural site in southern Delhi outskirts. *Atmos. Res.* **216**, 11–25 (2019).
7. P. Joshi, S. Dey, S. Ghosh, S. Jain, S. K. Sharma, Association between acute exposure to PM2.5 chemical species and mortality in megacity Delhi, India. *Environ. Sci. Technol.* **56**, 7275–7287 (2022).
8. S. K. Guttikunda *et al.*, What is polluting Delhi's air? A review from 1990 to 2022. *Sustainability* **15**, 4209 (2023).
9. S. Shen *et al.*, Enhancing global estimation of fine particulate matter concentrations by including geophysical a priori information in deep learning. *ACS ES&T Air* **1**, 332–345 (2024).
10. S. Dey *et al.*, A satellite-based high-resolution (1-km) ambient PM2.5 database for India over two decades (2000–2019): Applications for air quality management. *Remote Sens.* **12**, 3872 (2020).
11. M. Smaran, V. Vinoj, Evaluation of background black carbon concentration in India. *Aerosol Air Qual. Res.* **24**, 230300 (2024).
12. B. Rooney *et al.*, Impacts of household sources on air pollution at village and regional scales in India. *Atmos. Chem. Phys.* **19**, 7719–7742 (2019).
13. S. K. Guttikunda, B. R. Gurjar, Role of meteorology in seasonality of air pollution in megacity Delhi, India. *Environ. Monit. Assess.* **184**, 3199–3211 (2012).
14. Y. Chen *et al.*, Ammonium chloride associated aerosol liquid water enhances haze in Delhi, India. *Environ. Sci. Technol.* **56**, 7163–7173 (2022).
15. K. Patel *et al.*, Persistence of primary and secondary pollutants in Delhi: Concentrations and composition from 2017 through the COVID pandemic. *Environ. Sci. Technol. Lett.* **8**, 492–497 (2021).
16. T. Ganguly, K. L. Selvaraj, S. K. Guttikunda, National Clean Air Programme (NCAP) for Indian cities: Review and outlook of clean air action plans. *Atmos. Environ.* **8**, 100096 (2020).
17. S. K. Guttikunda, S. K. Dammalapati, G. Pradhan, Assessing air quality during India's National Clean Air Programme (NCAP): 2019–2023. *Atmos. Environ.* **343**, 120974 (2025).
18. D. Sarkar *et al.*, Pathways for India to reduce ambient air pollution health burden and achieve the Sustainable Development Goal (SDG-3.4). *Environ. Sci. Technol.* **59**, 4765–4777 (2025).
19. S. Dey, Air pollution in rural India: Ignored but not absent. Down To Earth (2021). https://www.downtoearth.org.in/pollution/air-pollution-in-rural-india-ignored-but-not-absent-75341 (Accessed 19 August 2025).
20. J. Singh, S. Payra, M. K. Mishra, S. Verma, An analysis of particulate pollution using urban aerosol pollution island intensity over Delhi, India. *Environ. Monit. Assess.* **194**, 874 (2022).
21. S. Jaganathan *et al.*, Nationwide analysis of air pollution hotspots across India: A spatiotemporal PM2.5 trend analysis (2008–2019). *Environ. Res.* **264**, 120276 (2025).
22. V. Sahu *et al.*, Assessment of a clean cooking fuel distribution scheme in rural households of India–"Pradhan Mantri Ujjwala Yojana (PMUY)". *Energy Sustainable Dev.* **81**, 101492 (2024).
23. J. Bedi, D. Toshniwal, Can electricity demand lead to air pollution? A spatio-temporal analysis of electricity demand with climatic conditions. *Renew. Sustainable Energy Rev.* **136**, 110413 (2021).
24. A. V. Shankar *et al.*, Everybody stacks: Lessons from household energy case studies to inform design principles for clean energy transitions. *Energy Policy* **141**, 111468 (2020).
25. R. V. Martin *et al.*, No one knows which city has the highest concentration of fine particulate matter. *Atmos. Environ.* **3**, 100040 (2019).
26. M. S. Hammer *et al.*, Global estimates and long-term trends of fine particulate matter concentrations (1998–2018). *Environ. Sci. Technol.* **54**, 7879–7890 (2020).

Downloaded from https://www.pnas.org by 72.59.229.104 on November 23, 2026 from IP address 72.59.229.104.

27. Z. Ma *et al.*, A review of statistical methods used for developing large-scale and long-term PM2.5 models from satellite data. *Remote Sens. Environ.* **269**, 112827 (2022).
28. D. Chatterjee *et al.*, Source contributions to fine particulate matter and attributable mortality in India and the surrounding region. *Environ. Sci. Technol.* **57**, 10263–10275 (2023).
29. D. H. Hagan, J. H. Kroll, Assessing the accuracy of low-cost optical particle sensors using a physics-based approach. *Atmos. Meas. Tech.* **13**, 6343–6355 (2020).
30. Y. Liang *et al.*, Wildfire smoke impacts on indoor air quality assessed using crowdsourced data in California. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2106478118 (2021).
31. S. Jain *et al.*, Identification of neighborhood hotspots via the cumulative hazard index: Results from a community-partnered low-cost sensor deployment. *GeoHealth* **8**, e2023GH000935 (2024).
32. M. A. Johnson *et al.*, Air pollutant exposure and stove use assessment methods for the Household Air Pollution Intervention Network (HAPIN) Trial. *Environ. Health Perspect.* **128**, 047009 (2020).
33. V. Sreekanth *et al.*, Inter-versus intracity variations in the performance and calibration of low-cost PM2.5 sensors: A multicity assessment in India. *ACS Earth Space Chem.* **6**, 3007–3016 (2022).
34. M. J. Campmier *et al.*, Seasonally optimized calibrations improve low-cost sensor performance: Long-term field evaluation of PurpleAir sensors in urban and rural India. *Atmos. Meas. Tech.* **16**, 4357–4374 (2023).
35. S. Jain, N. Zimmerman, Exploration of intra-city and inter-city PM2.5 regional calibration models to improve low-cost sensor performance. *J. Aerosol Sci.* **177**, 106335 (2024).
36. R. Sahu *et al.*, Validation of low-cost sensors in measuring real-time PM10 concentrations at two sites in Delhi National Capital Region. *Sensors* **20**, 1347 (2020).
37. E. Molina Rueda, E. Carter, C. L'Orange, C. Quinn, J. Volckens, Size-resolved field performance of low-cost sensors for particulate matter air pollution. *Environ. Sci. Technol. Lett.* **10**, 247–253 (2023).
38. J. Ouimette *et al.*, Fundamentals of low-cost aerosol sensor design and operation. *Aerosol Sci. Technol.* **58**, 1–15 (2024).
39. T. Zheng *et al.*, Gaussian process regression model for dynamically calibrating and surveilling a wireless low-cost particulate matter sensor network in Delhi. *Atmos. Meas. Tech.* **12**, 5161–5181 (2019).
40. J. Bi, A. Wildani, H. H. Chang, Y. Liu, Incorporating low-cost sensor measurements into high-resolution PM2.5 modeling at a large spatial scale. *Environ. Sci. Technol.* **54**, 2152–2162 (2020).
41. S. Madhwal *et al.*, Evaluation of PM2.5 spatio-temporal variability and hotspot formation using low-cost sensors across urban-rural landscape in lucknow, India. *Atmos. Environ.* **319**, 120302 (2024).
42. L. Dijkstra *et al.*, Applying the Degree of Urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. *J. Urban Econ.* **125**, 103312 (2021).
43. H. L. Brantley *et al.*, Characterization of spatial air pollution patterns near a large railyard area in Atlanta, Georgia. *Int. J. Environ. Res. Public Health* **16**, 535 (2019).
44. M. Schiavina *et al.*, GHSL data package 2022 (2022).
45. T. Liu *et al.*, Crop residue burning practices across north India inferred from household survey data: Bridging gaps in satellite observations. *Atmos. Environ.* **8**, 100091 (2020).
46. K. Patel *et al.*, Sources and dynamics of submicron aerosol during the autumn onset of the air pollution season in Delhi, India. *ACS Earth Space Chem.* **5**, 118–128 (2021).
47. P. Mangaraj *et al.*, Weak coupling of observed surface PM2.5 in Delhi-NCR with rice crop residue burning in Punjab and Haryana. *NPJ Clim. Atmos. Sci.* **8**, 1–9 (2025).
48. P. Lenschow *et al.*, Some ideas about the sources of PM10. *Atmos. Environ.* **35**, S23–S33 (2001).
49. S. S. Sethi, V. Vinoj, Transported dust modulates aerosol pollution domes over rapidly urbanizing Indian cities. *Commun. Earth Environ.* **6**, 554 (2025).
50. A. van Donkelaar *et al.*, Monthly global estimates of fine particulate matter and their uncertainty. *Environ. Sci. Technol.* **55**, 15287–15300 (2021).
51. N. Earth Science Data Systems, MODIS/Terra+aqua land aerosol optical depth daily L2G global 1 km SIN grid V061|NASA Earthdata. Earth Science Data Systems, NASA. Deposited 16 June 2025.
52. V. Katoch *et al.*, Addressing biases in ambient PM2.5 exposure and associated health burden estimates by filling satellite AOD retrieval gaps over India. *Environ. Sci. Technol.* **57**, 19190–19201 (2023).
53. C. J. Paciorek, Y. Liu, Limitations of remotely sensed aerosol as a spatial proxy for fine particulate matter. *Environ. Health Perspect.* **117**, 904–909 (2009).
54. P. deSouza, P. L. Kinney, On the distribution of low-cost PM 2.5 sensors in the US: Demographic and air quality associations. *J. Expo. Sci. Environ. Epidemiol.* **31**, 514–524 (2021).
55. T. Zheng, M. Bergin, G. Wang, D. Carlson, Local PM2.5 hotspot detector at 300 m resolution: A random forest-convolutional neural network joint model jointly trained on satellite images and meteorology. *Remote Sens.* **13**, 1356 (2021).
56. P. deSouza *et al.*, An analysis of degradation in low-cost particulate matter sensors. *Environ. Sci. Atmos.* **3**, 521–536 (2023).
57. K. K. Barkjohn, B. Gantt, A. L. Clements, Development and application of a United States-wide correction for $PM_{2.5}$ data collected with the PurpleAir sensor. Atmospheric. *Measure. Techniques* **14**, 4617–4637 (2021).
58. K. K. Barkjohn, A. L. Holder, S. G. Frederick, A. L. Clements, Correction and accuracy of PurpleAir PM2.5 measurements for extreme wildfire smoke. *Sensors* **22**, 9669 (2022).
59. B. Actkinson, K. Ensor, R. J. Griffin, SIBaR: A new method for background quantification and removal from mobile air pollution measurements. *Atmos. Meas. Tech.* **14**, 5809–5821 (2021).
60. R. Koenker, G. Bassett, Regression quantiles. *Econometrica* **46**, 33–50 (1978).
61. B. Lê Cook, W. G. Manning, Thinking beyond the mean: A practical guide for using quantile regression methods for health services research. *Shanghai Arch. Psychiatry* **25**, 55–59 (2013).
62. G. Chi, H. Fang, S. Chatterjee, J. E. Blumenstock, Microestimates of wealth for all low- and middle-income countries. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2113658119 (2022).
63. Central Pollution Control Board, *Central control room for air quality Management: All India CAAQMS data repository*. https://airquality.cpcb.gov.in/ccr/. Accessed 1 March 2025.
64. A. F. Stein *et al.*, NOAA's HYSPLIT atmospheric transport and dispersion modeling system. *Bull. Am. Meteorol. Soc.* **96**, 2059–2077 (2015).
65. H. Hersbach *et al.*, ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store. https://doi.org/10.24381/cds.adbb2d47. Accessed 1 March 2025.
66. PurpleAir, PurpleAir API. https://api.purpleair.com/. Accessed 1 March 2025.
67. M. J. Campmier *et al.*, Rural outdoor air pollution exceeds urban levels in the Delhi megacity region [Data]. Dryad. https://datadryad.org/dataset/doi:10.5061/dryad.2rbnzs86h. Deposited 30 October 2026.

Downloaded from https://www.pnas.org by 72.59.229.104 on November 23, 2026 from IP address 72.59.229.104.