Supplementary Information for
**Rural outdoor air pollution exceeds urban levels in the Delhi megacity region**

Mark Joseph Campmier[1], Sofiya Rao[2], Amarendra Singh[2], Saumya Singh[3], María Teresa Hernández[1], Nidhi Malik[4], Swagata Dey[4], Ajay Pillarisetti[1], Kalpana Balakrishnan[5], Sagnik Dey* [2,6], Joshua S Apte*[1,7]

[1] School of Public Health, University of California, Berkeley, CA 94720, USA
[2] Centre for Atmospheric Sciences, Indian Institute of Technology Delhi, New Delhi, Delhi 110016, India
[3] Ecosystem and Environment Management Area, Indian Institute of Forest Management, Bhopal 462003, India
[4] Center for Study of Science, Technology and Policy, Bengaluru 560094, India
[5] 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
[6] Department of Health, Policy and Management, Korea University, Seoul 02841, Republic of Korea
[7] Department of Civil and Environmental Engineering, University of California, Berkeley, Berkeley, CA 94720, USA

* Corresponding Authors: sagnik@cas.iitd.ac.in and apte@berkeley.edu

**This PDF file includes:**

**Tables S1-S5**
**Figures S1-S17**
**SI References**

## SI Tables

**Table S1.** Urban settlement type site-wise data summary.

| Site[#] | District | Install Date | Population (people/km$^2$)[*] | RWI[+] | Valid Data Points | PM2.5 (ug/m$^3$)[^] |
|---|---|---|---|---|---|---|
| DL-NCT-1 | East Delhi | 2024-Jul-01 | 33464 | 0.696 | 3078 | 55.3 (143.1) |
| DL-NCT-2 | North West Delhi | 2024-Jul-08 | 20453 | 0.119 | 2000 | 65.1 (114.2) |
| DL-NCT-3 | West Delhi | 2024-Jul-08 | 23166 | 0.706 | 3417 | 60.9 (131.1) |
| DL-NCT-4 | West Delhi | 2024-Jul-10 | 31060 | 0.119 | 2382 | 103.1 (168.5) |
| DL-NCT-5 | North Delhi | 2024-Jul-15 | 30509 | 0.119 | 3177 | 64.9 (152.5) |
| DL-NCT-6 | South Delhi | 2024-Jul-17 | 14410 | 1.064 | 2145 | 108.9 (153.9) |
| DL-VT-10A | North West Delhi | 2024-Jul-01 | 9697 | 0.853 | 2534 | 73.0 (179.4) |
| DL-VT-11B | West Delhi | 2024-Jul-01 | 32107 | 0.119 | 3371 | 59.1 (116.5) |
| DL-VT-13A | South Delhi | 2024-Jul-01 | 25958 | 0.151 | 3438 | 46.7 (73.4) |
| DL-VT-13B | South Delhi | 2024-Jul-01 | 25745 | 0.151 | 3663 | 44.3 (95.0) |
| HR-RT-2 | Rohtak | 2024-Jul-01 | 6376 | 0.199 | 2845 | 66.9 (153.1) |
| HR-RT-3 | Rohtak | 2024-Jul-03 | 4605 | 0.199 | 3586 | 61.5 (129.7) |
| HR-RT-6 | Rohtak | 2024-Jul-01 | 4242 | 0.199 | 3473 | 70.4 (151.7) |
| HR-RT-7 | Rohtak | 2024-Jul-01 | 6276 | 0.199 | 3346 | 63.7 (132.1) |
| HR-VT-14A | Faridabad | 2024-Jul-01 | 10872 | 1.153 | 3598 | 51.0 (98.9) |
| HR-VT-14B | Faridabad | 2024-Jul-01 | 10858 | 1.153 | 3602 | 48.8 (90.0) |
| HR-VT-15A | Faridabad | 2024-Jul-01 | 13442 | 0.206 | 2892 | 47.2 (85.6) |

[#] Site-id follows the formula: (STATE-CLUSTER-NUMBER)
* Population density data from the Global Human Settlement Layer project (1) aggregated within 5-km buffers (as in main text Figure 5) or 100-km$^2$
[+] Relative Wealth Index [RWI] from the AI Data for Good project (2) in native resolution of 5.76-km$^2$
[^] Campaign median $PM_{2.5}$ with interquartile range from our sensors

**Table S2.** Rural settlement type site-wise data summary.

| Site[#] | District | Install Date | Population (people/km$^2$)[*] | RWI[+] | Valid Data Points | PM2.5 (ug/m$^3$)[^] |
|---|---|---|---|---|---|---|
| DL-VT-9A | North West Delhi | 2024-Jul-01 | 1493 | 0.575 | 2775 | 72.7 (107.9) |
| DL-VT-9B | North West Delhi | 2024-Jul-01 | 1445 | 0.575 | 2630 | 60.2 (73.0) |
| HR-GH-1 | Sonipat | 2024-Jul-01 | 1875 | -0.435 | 3510 | 60.8 (105.4) |
| HR-GH-2 | Sonipat | 2024-Jul-02 | 1895 | -0.435 | 2998 | 58.3 (143.4) |
| HR-GH-3 | Sonipat | 2024-Jul-02 | 1914 | 0.645 | 3130 | 53.1 (135.1) |
| HR-GH-5 | Sonipat | 2024-Jul-01 | 1885 | -0.435 | 3168 | 61.3 (133.1) |
| HR-GH-6 | Sonipat | 2024-Jul-02 | 1958 | 0.645 | 2630 | 57.2 (94.3) |
| HR-GH-8 | Sonipat | 2024-Jul-02 | 1781 | 0.645 | 1978 | 77.0 (167.1) |
| HR-HT-1A | Panipat | 2024-Jul-30 | 369 | 0.632 | 2823 | 97.1 (201.8) |
| HR-HT-1B | Panipat | 2024-Jul-30 | 353 | 0.632 | 2481 | 63.6 (168.7) |
| HR-HT-2A | Panipat | 2024-Jul-30 | 379 | -0.327 | 2383 | 76.7 (186.8) |
| HR-HT-2B | Panipat | 2024-Jul-30 | 377 | -0.327 | 2743 | 79.0 (187.3) |
| HR-HT-3A | Sonipat | 2024-Jul-31 | 468 | 0.015 | 2180 | 146.7 (221.3) |
| HR-HT-3B | Sonipat | 2024-Jul-31 | 591 | 0.015 | 2846 | 100.2 (214.5) |
| HR-HT-4A | Sonipat | 2024-Jul-31 | 436 | -0.218 | 2187 | 101.8 (183.9) |
| HR-HT-5A | Rohtak | 2024-Aug-22 | 333 | -0.173 | 2385 | 163.8 (237.4) |
| HR-HT-6B | Rohtak | 2024-Jul-31 | 369 | 0.054 | 2156 | 139.1 (259.0) |
| HR-IN-VL-3 | Sonipat | 2024-Sep-10 | 357 | 0.15 | 1874 | 148.8 (192.1) |
| HR-IN-VL-8 | Sonipat | 2024-Sep-11 | 322 | 0.54 | 1907 | 176.9 (202.6) |
| HR-PN-3 | Panipat | 2024-Aug-06 | 971 | 0.632 | 2570 | 92.0 (183.8) |
| HR-VT-16A | Faridabad | 2024-Jul-01 | 331 | 0.486 | 3461 | 69.9 (124.8) |
| HR-VT-1A | Jind | 2024-Jul-11 | 357 | -0.481 | 3326 | 52.7 (151.1) |
| HR-VT-2A | Jind | 2024-Jul-01 | 480 | 0.143 | 3283 | 72.1 (130.7) |
| HR-VT-2B | Jind | 2024-Jul-01 | 479 | 0.143 | 3514 | 88.7 (148.3) |
| HR-VT-3A | Sonipat | 2024-Jul-01 | 582 | 0.586 | 3416 | 67.5 (116.3) |
| HR-VT-3B | Sonipat | 2024-Jul-01 | 557 | 0.586 | 3215 | 64.7 (129.4) |
| HR-VT-4B | Sonipat | 2024-Jul-02 | 1174 | 0.645 | 2118 | 82.6 (149.8) |
| HR-VT-5A | Sonipat | 2024-Jul-01 | 391 | 0.328 | 1927 | 46.8 (36.7) |
| HR-VT-5B | Sonipat | 2024-Jul-11 | 391 | 0.328 | 2725 | 84.7 (154.5) |
| HR-VT-6A | Sonipat | 2024-Jul-01 | 565 | -0.094 | 2204 | 56.0 (55.2) |
| HR-VT-7A | Sonipat | 2024-Jul-03 | 1087 | 0.777 | 2779 | 49.3 (69.9) |
| HR-VT-7B | Sonipat | 2024-Jul-01 | 1087 | 0.777 | 2552 | 50.0 (49.8) |

[#] Site-id follows the formula: (STATE-CLUSTER-NUMBER)
***** Population density data from the Global Human Settlement Layer project (1) aggregated within 5-km buffers (as in main text Figure 5) or 100-$km^2$
[+] Relative Wealth Index [RWI] from the AI Data for Good project (2) in native resolution of 5.76-$km^2$
[^] Campaign median $PM_{2.5}$ with interquartile range from our sensors

**Table S3.** Sensor network monthly median and interquartile ranges for each of the campaign months, July-November 2024, stratified by settlement-type, urban and rural.

| | Urban $PM_{2.5}$ ($\mu g/m^3$) | Rural $PM_{2.5}$ ($\mu g/m^3$) |
|---|---|---|
| July | 43.7 (22.5) | 51.7 (26.8) |
| August | 24.7 (14.7) | 31.7 (17.8) |
| September | 37.9 (30.7) | 45.7 (32.7) |
| October | 110.7 (104.1) | 124.8 (122.0) |
| November | 255.4 (123.7) | 297.9 (145.6) |

**Table S4.** Correlation of time series components between rural and urban sites

| | **Peak Rural vs. Urban†** | **Baseline Rural vs. Urban‡** |
|---|---|---|
| **July** | 0.39 | 0.71 |
| **August** | 0.26 | 0.73 |
| **September** | 0.33 | 0.81 |
| **October** | 0.46 | 0.90 |
| **November** | 0.38 | 0.85 |

† Pearson's correlation coefficient between the hourly time series of peak components: median of all rural sites vs. median of all urban sites.

‡ Pearson's correlation coefficient between the hourly time series of baseline components: median of all rural sites vs. median of all urban sites.

**Table S5.** Site-wise calibration coefficients derived performing an Ordinary-Least Squares regression fit. All calibrations below are annual models, which were determined to perform more similarly across seasons than seasonal models which carry stronger trans-seasonal bias. See Campmier et al (2023), and Barkjohn et al (2021, 2022) for more details (3–5).

| Model | PA | RH | Intercept |
|---|---|---|---|
| USA | 0.546 | -0.086 | 5.75 |
| Delhi | 0.555 | -0.878 | 47.1 |
| Hamirpur | 0.485 | -0.487 | 36.1 |
| Bengaluru | 0.515 | -0.139 | 14.1 |

## SI Figures

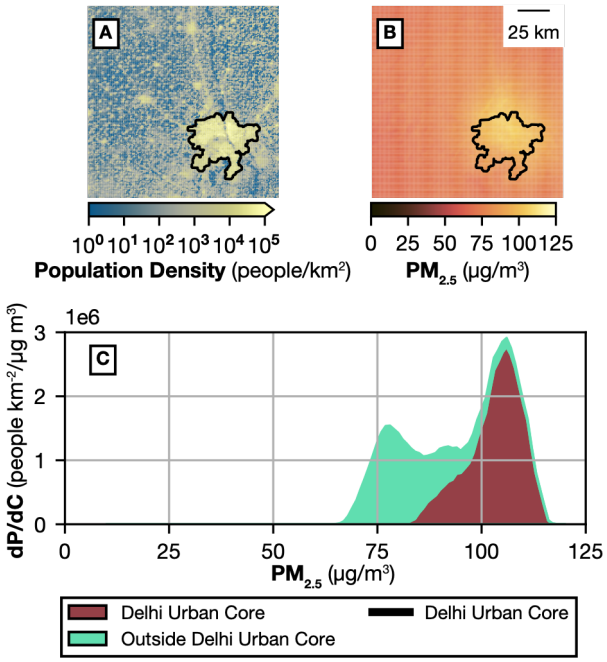



**Figure S1.** Satellite-derived $PM_{2.5}$ concentrations inside and outside of Delhi's urban core. **A.** The population density from Schiavina et al. (1) with boundaries showing Delhi's urban core, defined here as the area with pixels corresponding to population density higher than $10^4$ people/km$^2$, which encompasses the Delhi-Gurugram-Noida-Faridabad-Ghaziabad urban agglomeration. **B.** The satellite-derived annual $PM_{2.5}$ from Shen et al. (2024) (6) in the Delhi urban core for the year 2022. Extent of maps in A and B corresponds the the study area for our field measurement campaign. **C.** Population-concentration distributions (dP/dC) for the ~ 68 million people residing within our study area. Note that satellite remote sensing suggests that concentrations in the predominantly rural areas outside of Delhi's urban core are substantially lower than in the Delhi urban core.

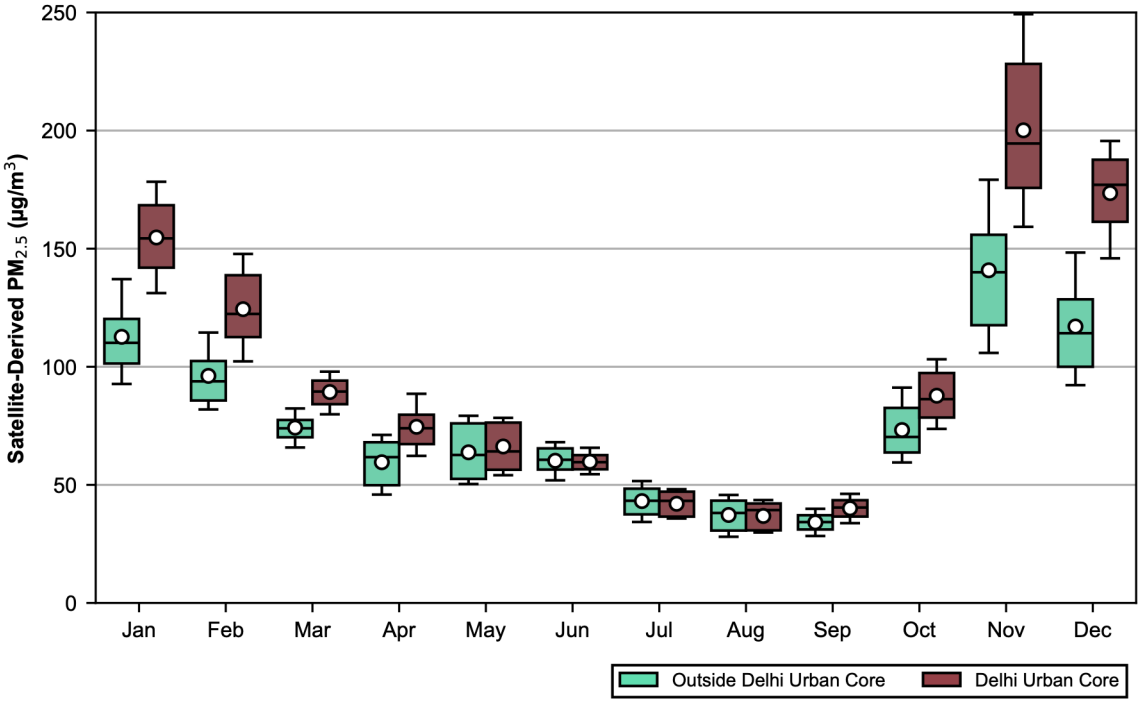



**Figure S2.** Satellite-derived monthly $PM_{2.5}$ (6) outside and within the Delhi urban core for the year 2022. While both urban and rural areas are extremely polluted and follow the same seasonal trends, these satellite-derived measurements suggest that concentrations in the Delhi urban core greatly exceed levels outside the urban core during the yearly maximum in November and December.

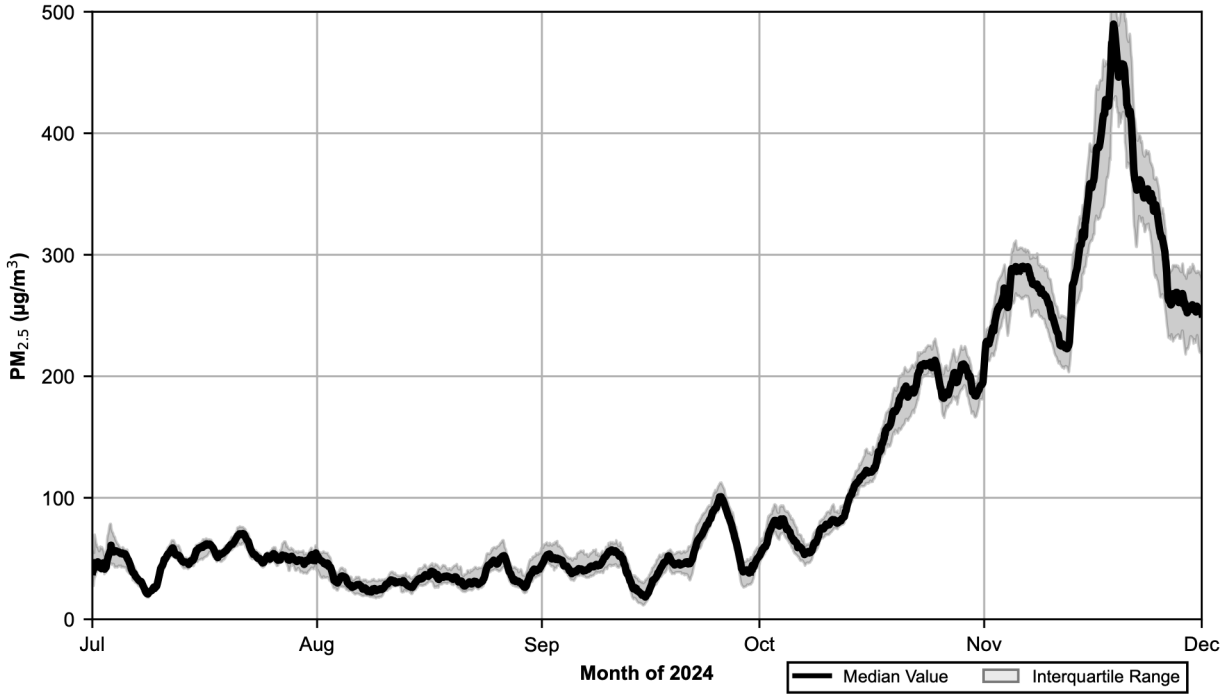



**Figure S3.** Time series of median and interquartile range $PM_{2.5}$ concentration from this campaign sensors (n = 49) from July through November 2024. Data are presented with a 72-hour rolling mean smoothing. Concentrations rapidly rise in October and peak in mid-November.

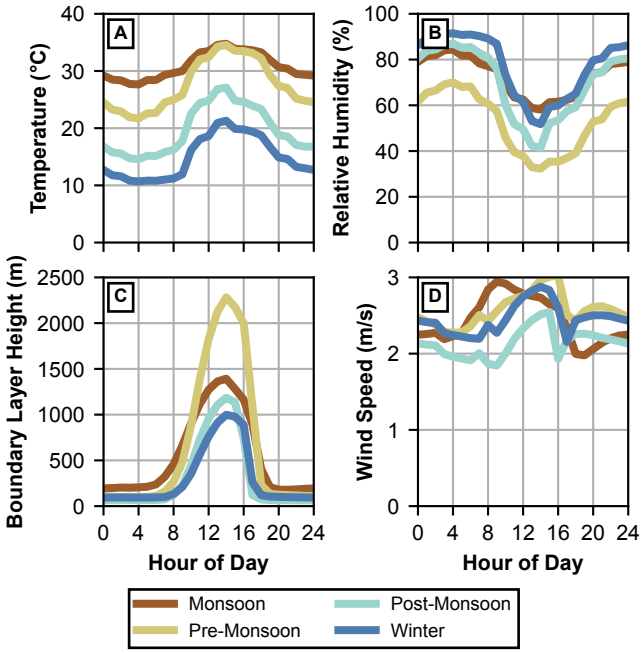



**Figure S4.** Seasonal diurnal mean meterological conditions for the campaign area shown in SI Appendix Figure S1 from ERA5 reanlysis products. **A.** temperature; **B.** relative humidity (RH); **C.** boundary layer height; and **D.** wind speed for the region of interest.

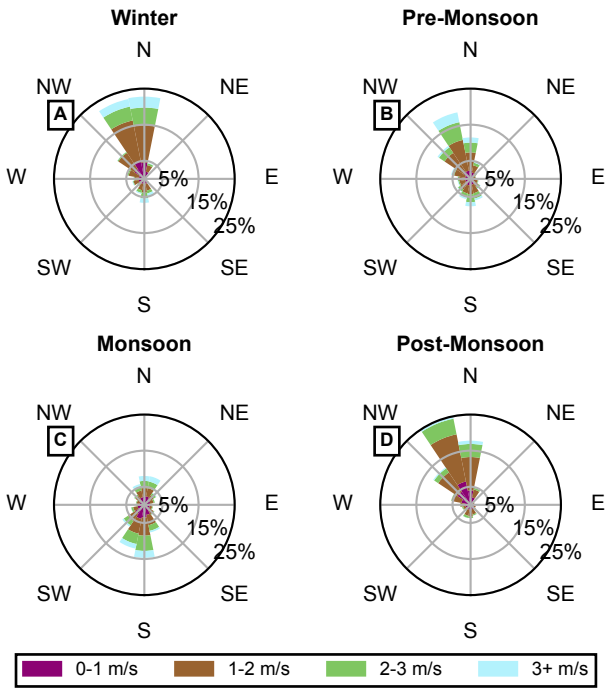



**Figure S5.** Mean seasonal wind roses for the campaign region based on ERA5 reanalysis data. **A.** winter; **B.** pre-monsoon; and **D.** post-monsoon all show the dominant wind direction from the NNW. **C.** The monsoon wind rose depicts opposite directionality with primarily southerly winds.

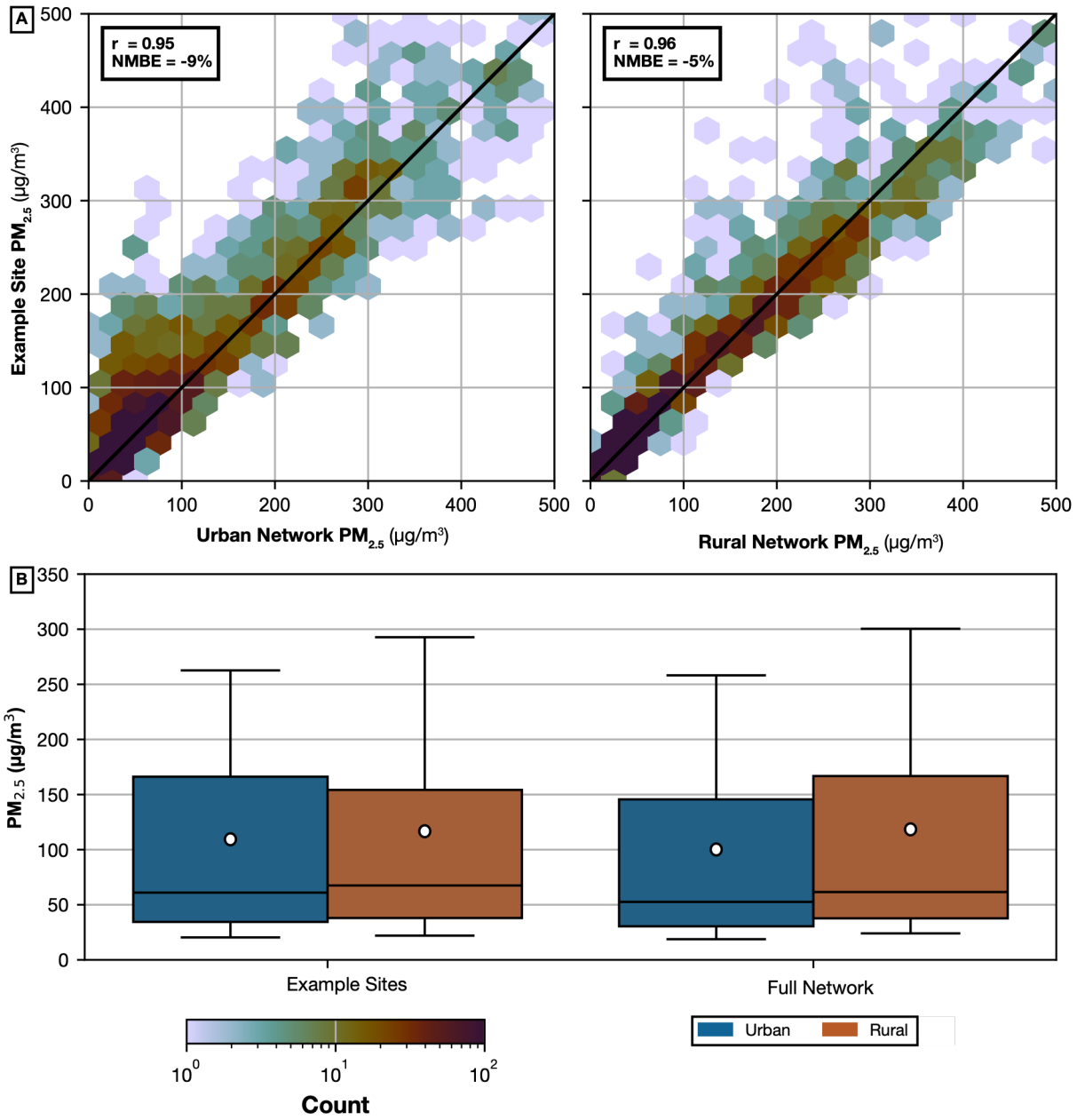



**Figure S6.** Comparison between $PM_{2.5}$ at two urban and rural example sites presented in Figure 1 with the distribtion of concentations from the the full network. **A.** Comparison of the marginal distributions of the example sites from the urban and rural sites respectively. Both example sites are representative samples of the larger overall site-type distributions. **B.** Comparison of the example site distributions (urban versus rural), as well as the full network, respectively. The example sites are well correlated with their sub-networks at large and exhibit similar differences in central tendency.

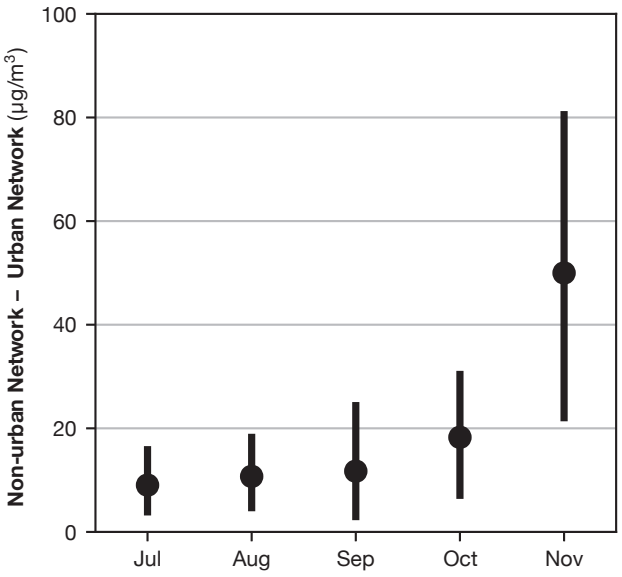



**Figure S7.** Evaluation of the sensitivity of the estimated concentration increment in rural vs. urban sites to sampling error arising from the finite number of measurement sites. We performed bootstrap resampling to estimate the 95% confidence interval in the difference between in the median monthly average concentration at rural vs. urban sites. This anlaysis indicates that our core conclusion that rural concentrations exceed urban concentrations is robust to concerns of sampling error. Larger rural vs. urban differences with broader confidence intervals in October and November reflect the generally higher concentrations moving into most the polluted months of the year.

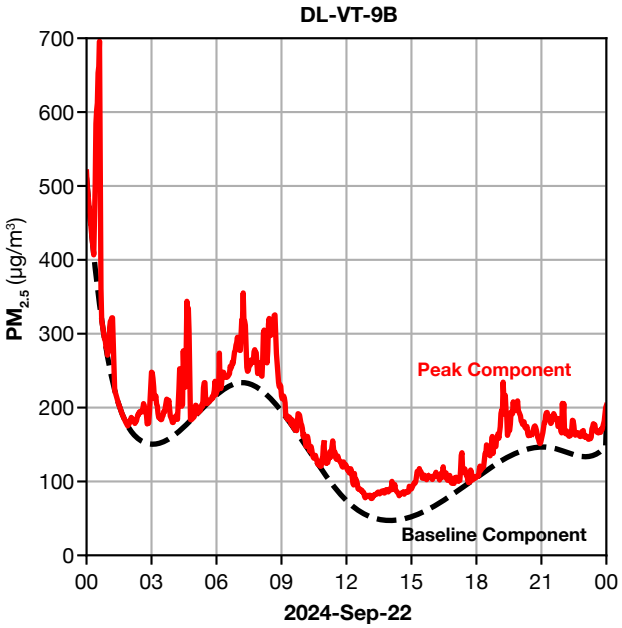



**Figure S8**. Illustration of spline decomposition technique used to estimate the baseline and peak components of the total $PM_{2.5}$ signal. This specific example shows data from the rural site DL-VT-9B on 2024 September 22. The peak enhancement demonstrates the isolated short-term variability, reflective of local combustion plumes near this community-based site.

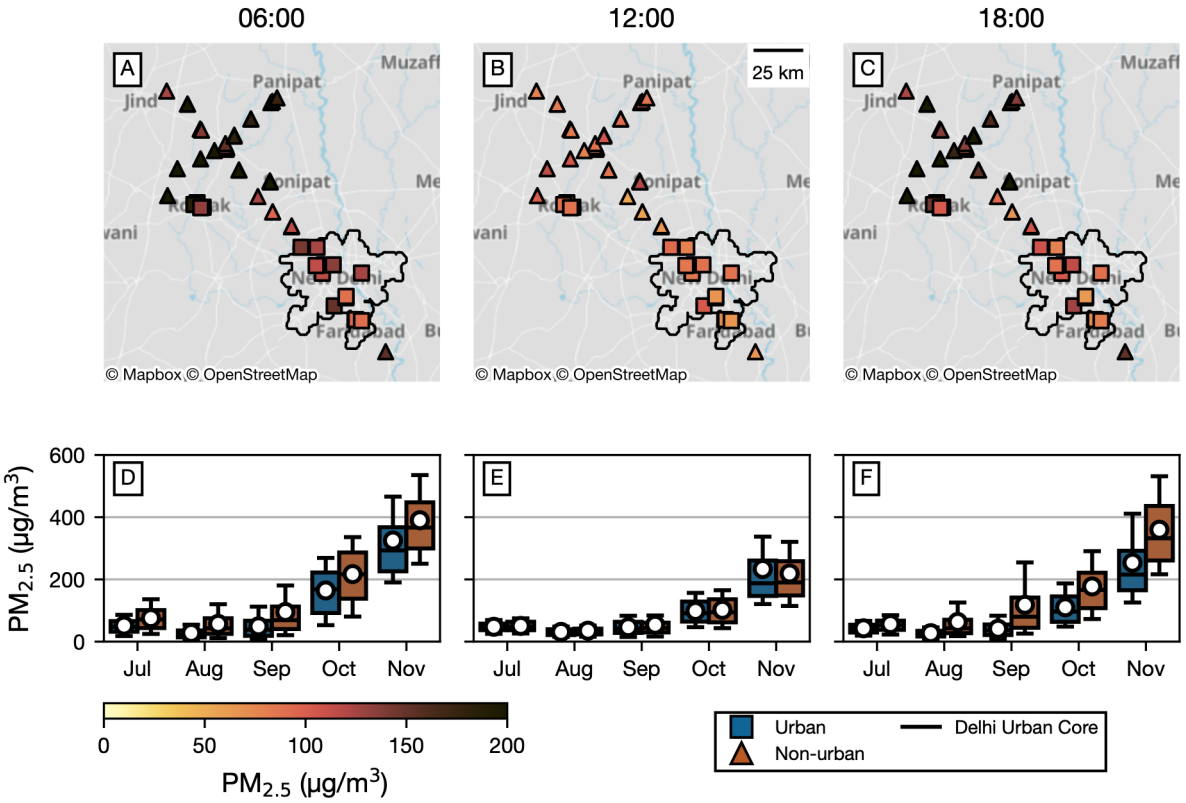



**Figure S9. A.** The campaign median in the morning, noon, and evening, respectively. **B.** The morning, noon, and evening monthly distributions stratified by urban and rural sites. During morning and evening, the rural sites are more polluted than the urban sites across the months. On the other hand, urban sites are more polluted than rural sites at noon each month.

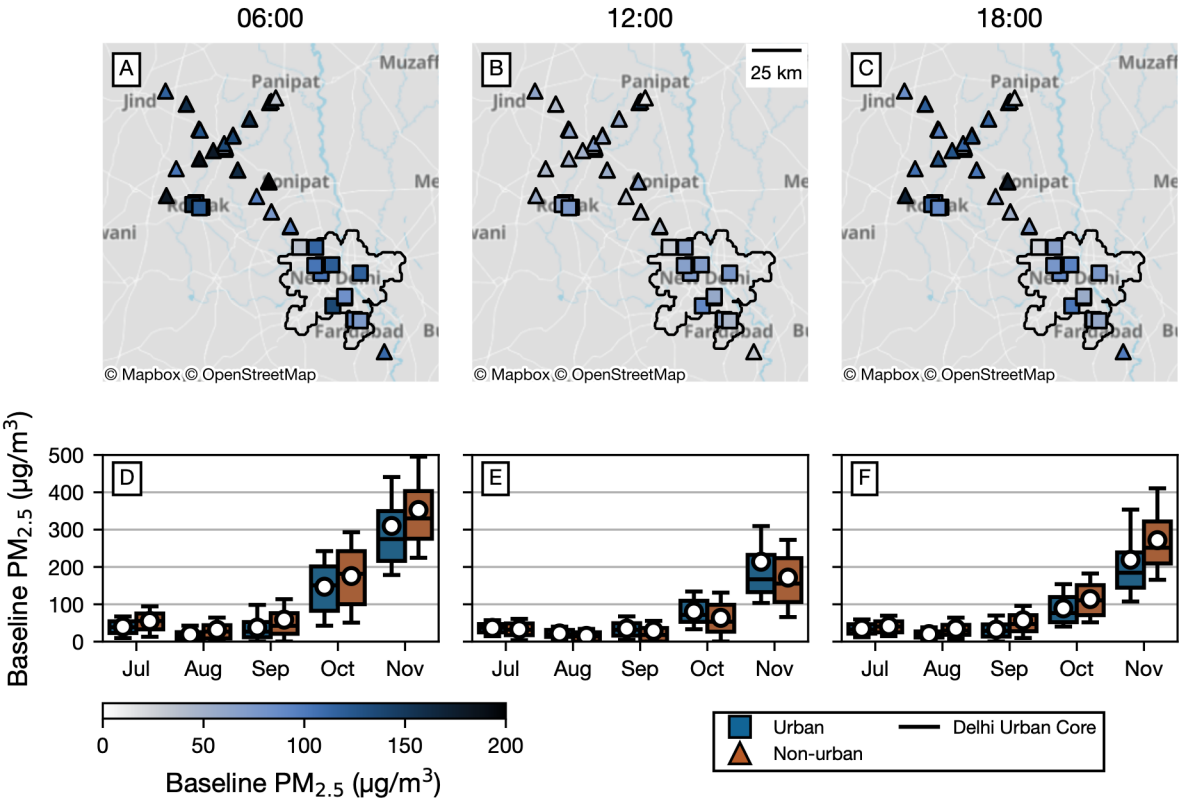



**Figure S10. A.** The campaign baseline component $PM_{2.5}$ median in the morning, noon, and evening, respectively. **B.** The morning, noon, and evening monthly baseline component distributions stratified by urban and rural sites. Baseline component concentrations are similar across urban and rural sites all all times of day, except evening in November, when crop stubble burning is most intense.

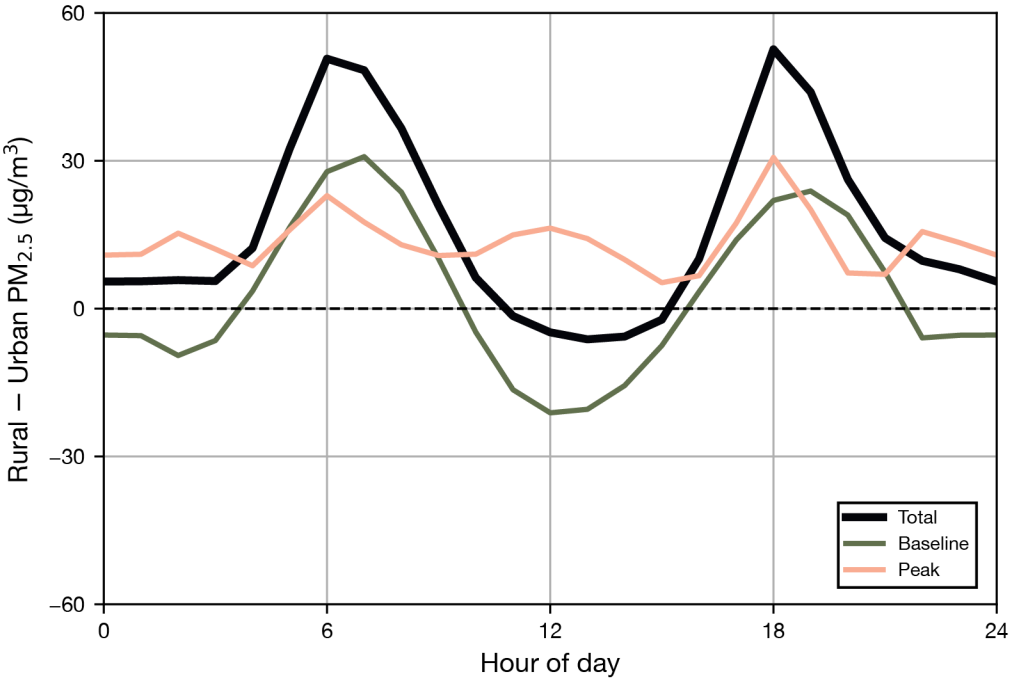



**Figure S11:** Campaign-length aggregated diurnal profiles of mean difference between the mean rural network signal and mean urban network signal by Baseline component $PM_{2.5}$, Peak component $PM_{2.5}$, and total $PM_{2.5}$. The directionality of the total $PM_{2.5}$ signal wavers, corresponding to shifts in the magnitude of Baseline component $PM_{2.5}$. In the morning (05:00 – 10:00 LT), evening (16:00 – 20:00 LT), and night (20:00 -05:00 LT) windows the rural sites feature more Baseline component $PM_{2.5}$, while during midday (10:00 – 16:00 LT), the magnitude of the Baseline component $PM_{2.5}$ is higher at the urban sites. The Peak component $PM_{2.5}$ at the rural sites is always elevated above the urban sites, reflecting a consistently higher household or hyperlocal signal.

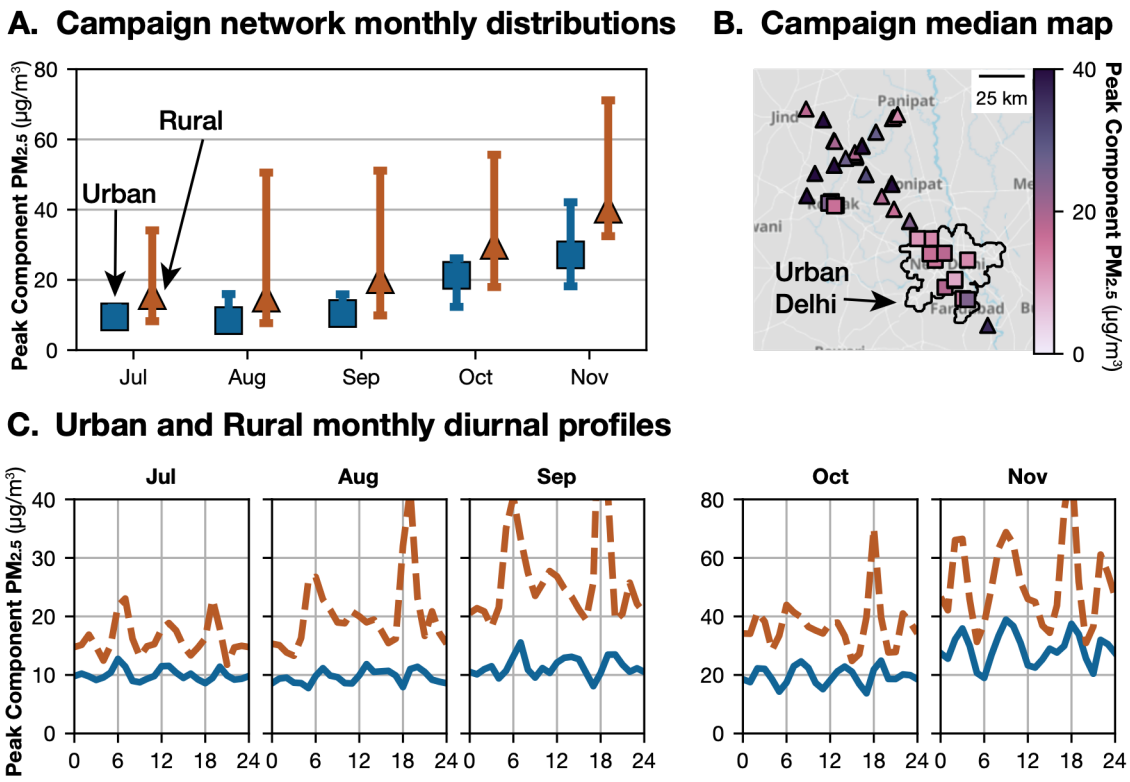



**Figure S12.** Summary of peak component $PM_{2.5}$ stratified by rural vs. urban sites. **A.** the monthly medians with interquartile range. **B.** Map of the entire network with markers indicating the settlement type and colored by campaign median mass concentration. **C.** The mean diurnal trend for July-November, respectively. Both in aggregate and monthly, the rural peak component $PM_{2.5}$ concentrations exceed the urban sub-network concentrations. This effect appears not to be limited exclusively to the Delhi urban core, as the "off-axis" city, Rohtak, also features substantially lower peak component $PM_{2.5}$ than nearby rural sites. Unlike total $PM_{2.5}$, rural peak component $PM_{2.5}$ is higher than urban peak component $PM_{2.5}$ across all hours of the day in each month, with a smaller but persistent difference at noon.

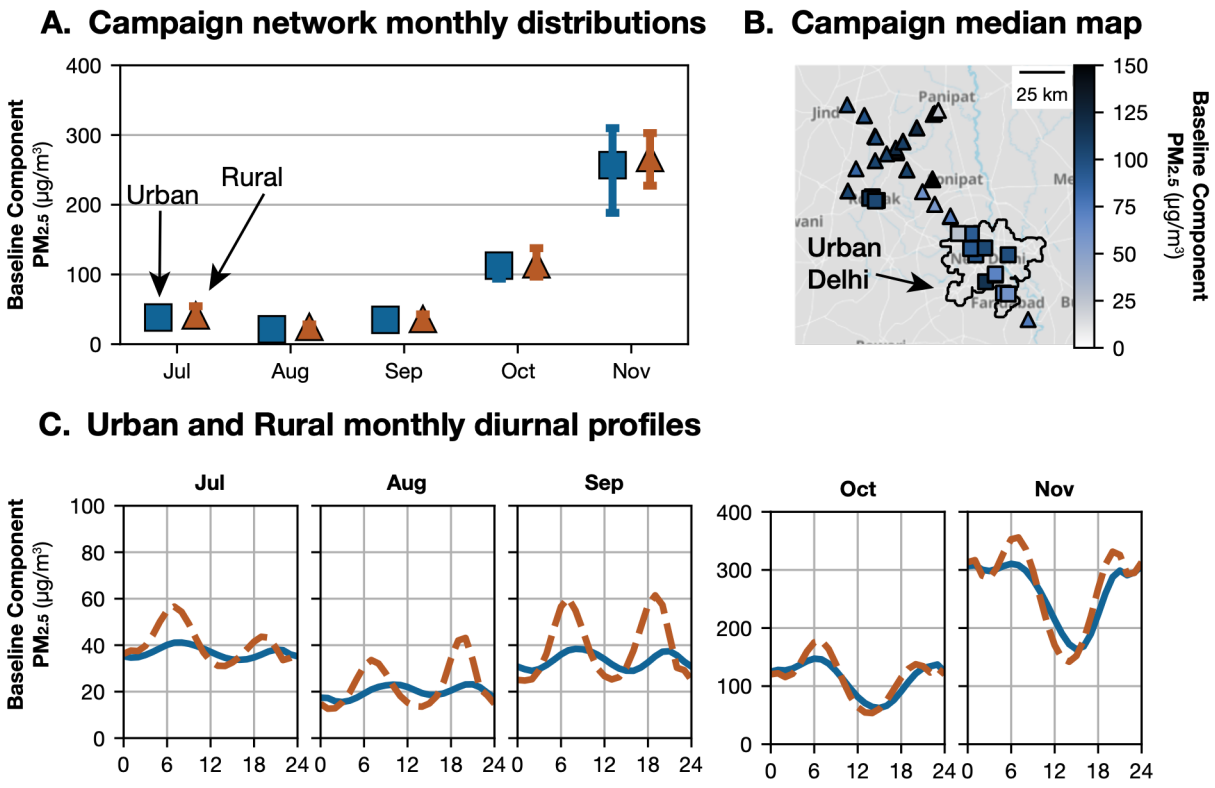



**Figure S13.** Summary of baseline component $PM_{2.5}$ stratified by urban sites versus rural sites. **A.** The monthly medians with interquartile range. **B.** Map of the entire network with markers indicating the settlement type and colored by campaign median mass concentration. **C.** The mean diurnal trend for July-November, respectively. Baseline component $PM_{2.5}$ trends do not demonstrate a consistent spatial difference between the rural and urban sites, with similar trends with aggregating by campaign length. During July-September, baseline component $PM_{2.5}$ is elevated at rural sites in the morning and evening hour blocks, but reports lower concentrations than at the urban sites, reflecting the stronger impacts of non-local sources in urban community-hosted sites.

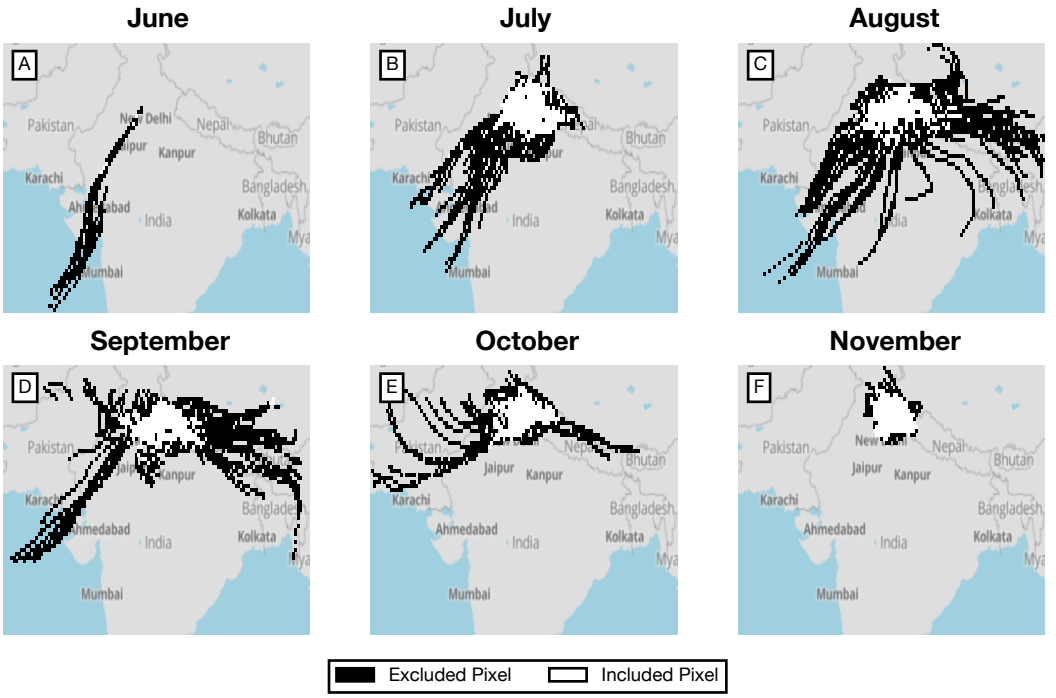



**Figure S14.** The monthly wind trajectories calculated using the HYSPLIT model. **A.** June; **B.** July; **C.** August; **D.** September; **E.** October; and **F.** November monthly simulations. Pixels are excluded from analysis that fall beneath the 75$^{th}$ percentile in trajectory intensity or total number of trajectories passing through a given pixel (0.5° x 0.5°).

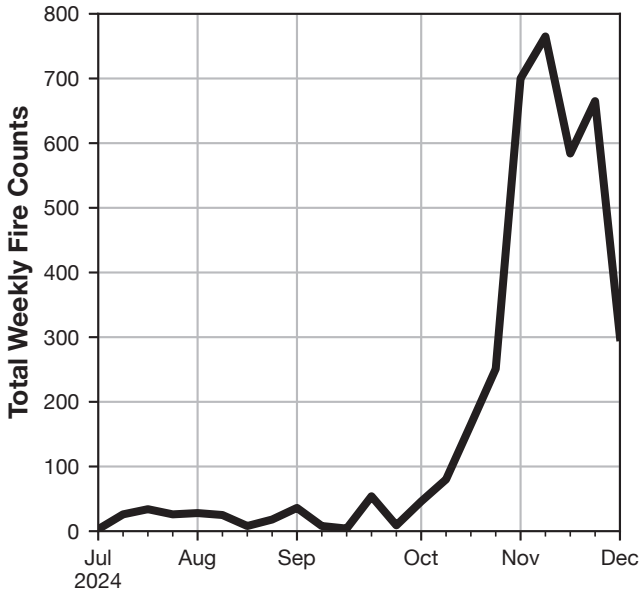



**Figure S15.** Time series of the weekly cumulative fire counts from the NASA FIRMS product during the campaign period (2024 July – November). Rather than simply use the campaign area shown (see SI Appendix Figure S1), we first calculated the monthly footprint and included fires only found in the footprint. In this way, we can better estimate fire events impacting our sites than simply using a buffer around our sites. Generally, fire events are dominated by a relatively short period occurring in early November when coinciding with both crop stubble burning and the wind trajectories (at 50 m) shifting to the northwest, where stubble burning activity is most intense in India.

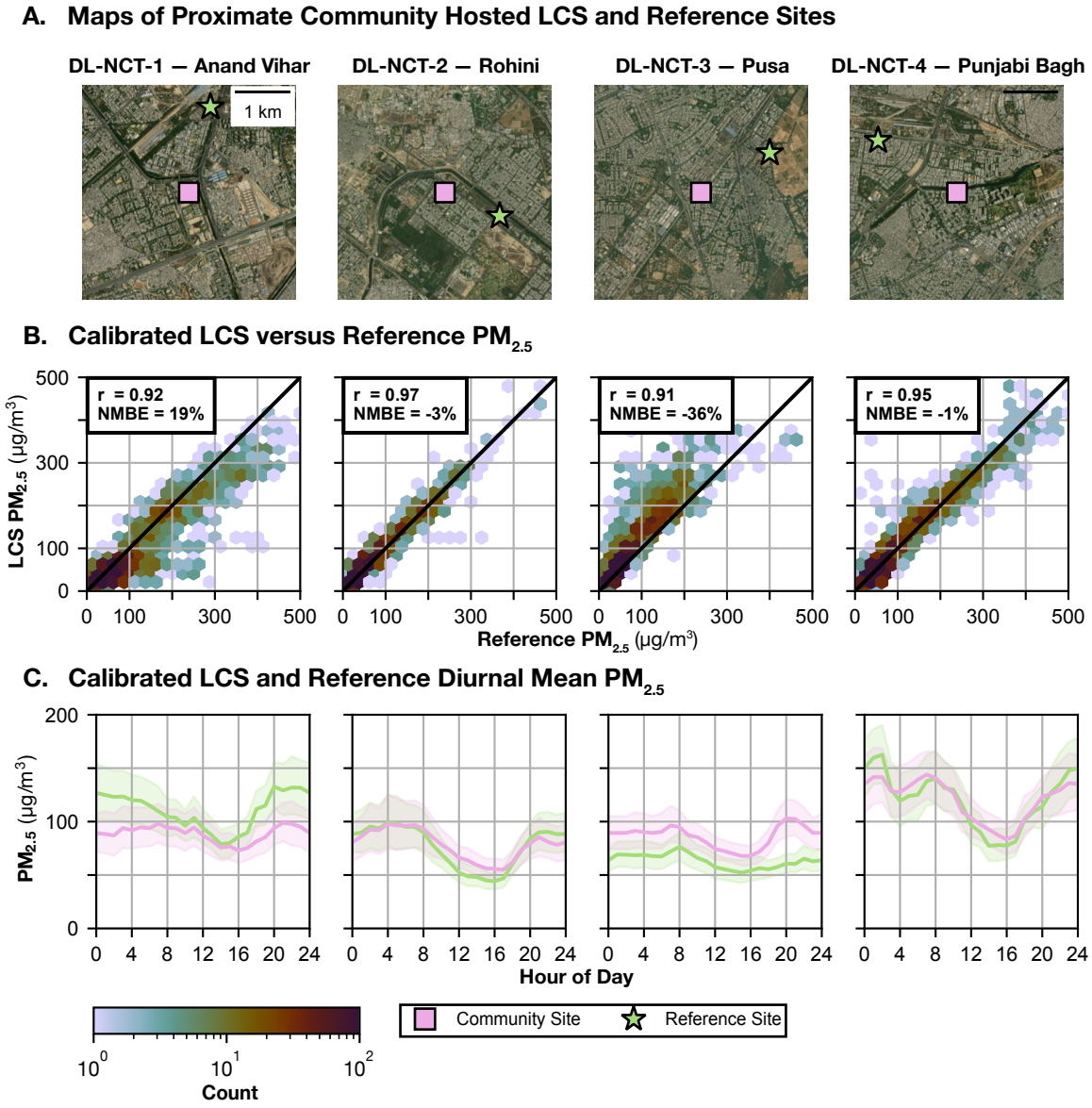



**Figure S16. A.** Maps of four urban LCS sites (DL-NCT-1, DL-NCT-2, DL-NCT-3, DL-NCT-4) in our network, and their nearest (within ~ 2 km) regulatory monitoring site (Anand Vihar, Rohini, Pusa, Punjabi Bagh) with satellite imagery. **B.** The marginal distribution with a 1:1 line between the LCS site and its regulatory counterpart. **C.** The campaign mean diurnal profiles with 95% confidence intervals for both the LCS and corresponding regulatory sites. Across all sites, the correlation is remarkably high, with low bias at two sites (DL-NCT-2, DL-NCT-4), positive bias at one site (DL-NCT-1), and negative bias at one bias (DL-NCT-3. This range of biases is expected given the differences in siting principles for regulatory versus community exposure monitoring.

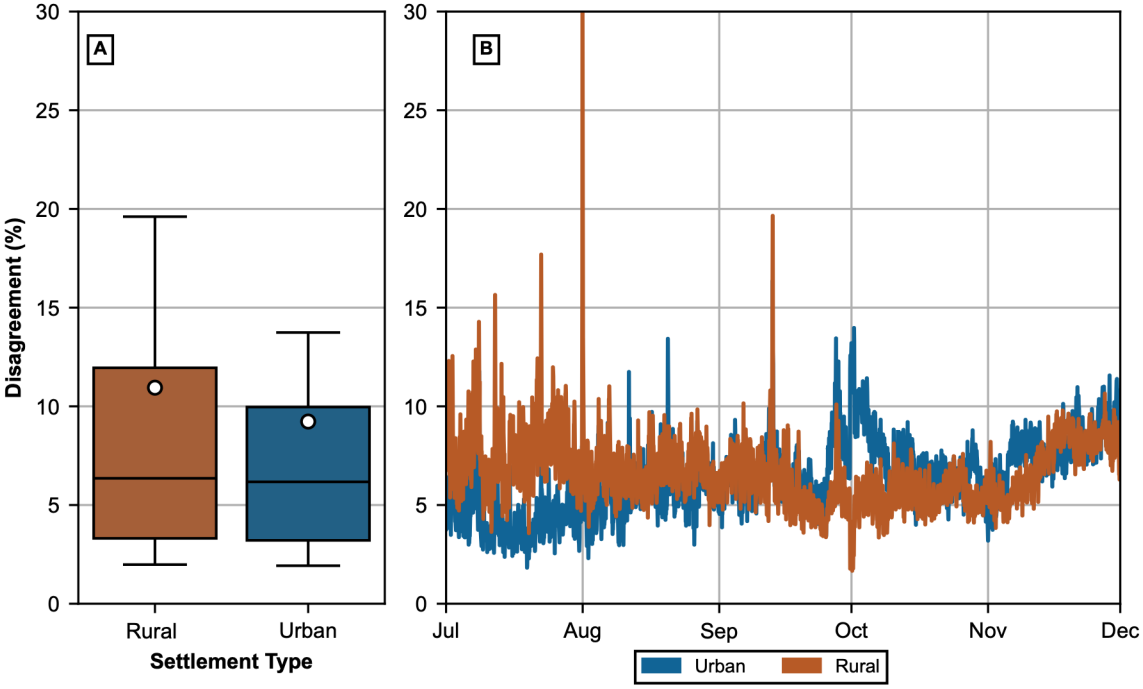



**Figure S17. A.** The channel disagreement distributions of the rural versus urban sites. **B.** Time series of median channel disagreement by urban versus rural sites. Channel disagreement is constrained to within 20% for the majority of the campaign for all sites, with little difference between the urban and rural sites.

## SI References


1. M. Schiavina, *et al.*, *GHSL Data Package 2022* (2022).

2. G. Chi, H. Fang, S. Chatterjee, J. E. Blumenstock, Microestimates of wealth for all low- and middle-income countries. *Proceedings of the National Academy of Sciences* **119**, e2113658119 (2022).

3. 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. *Atmospheric Measurement Techniques* **16**, 4357–4374 (2023).

4. 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 Measurement Techniques* **14**, 4617–4637 (2021).

5. 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).

6. S. Shen, *et al.*, Enhancing Global Estimation of Fine Particulate Matter Concentrations by Including Geophysical a Priori Information in Deep Learning. *ACS EST Air* **1**, 332–345 (2024).