Research Square

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

Gasping for Data: Evaluating the Adequacy of Continuous Ambient Air Quality Monitoring Network of India

Abhinav Kapoor

kapoor.abhinav123@gmail.com

Indian Institute of Management Bangalore

Somnath Bera

Central University of South Bihar

Research Article

Keywords: Continuous Ambient Air Quality Monitoring, Spatial Representativeness, Population Exposure, Data Availability, Environmental Governance, India

Posted Date: June 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9242405/v1

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

Additional Declarations: No competing interests reported.

Abstract

Air pollution in India represents a significant health and economic burden, requiring the development of a national ambient air quality monitoring framework which is representative, continuous, reliable and reports real-time data. This paper evaluates the adequacy of the Continuous Ambient Air Quality Monitoring System (CAAQMS) in India by cross-referencing their spatial distribution with the country's population density. The analysis employs a tiered proximity framework comprising a 2 km precision zone, a 10 km representative zone, and a 50 km proxy radius to estimate population exposure across varying levels of data fidelity. Results indicate that 40 percent of the national population, approximately 588 million people, reside beyond 50 kilometers of any functional station, with coverage absent entirely in Ladakh and Lakshadweep. In states such as Himachal Pradesh, over 85 percent of residents fall outside the 50 kilometer threshold, while even in Delhi, 65 percent of the population lives beyond the high-precision 2 kilometer zone. Analysis of data reported between January 2021 and December 2025 reveals that infrastructure presence does not guarantee data continuity. While Punjab, Puducherry, and Nagaland recorded availability above 88 percent, several states reported data on fewer than 30 percent of days in 2023 and 2024. These gaps in spatial reach and reporting consistency undermine the reliability of time-series data essential for an analysis of air pollution trends and their subsequent policy evaluation.

1 Introduction

The adverse impacts of pollution are rising across India, severely affecting its current and future economic output and human resources. Air pollution significantly contributes to increasing mortality rates, placing a disproportionate burden on healthcare infrastruc- ture(Abdurrahman et al., 2020; Bharath et al., 2020; Markandeya et al., 2021). The issue has not only been persistent in Delhi NCR, but also in various regions of India including Bangalore(Bhagat & Saha, 2021; Govindaraj et al., 2021; Nair et al., 2023), Mumbai(Ghosh, 2025; Singh et al., 2023), Jaipur(Dadhich et al., 2018; Soni et al., 2018), Chennai (Chelani, 2016; Rana & Kumar, 2024) and Hyderabad(Gautam et al., 2021; Yeasin et al., 2023). It will be fair to claim that mitigating air pollution has become one of the primary objectives for both central and state governments.

There are two methodologies using which air quality data is collected in India. The National Air Monitoring Programme (NAMP) tracks long-term trends for four pri- mary pollutants (SO_2, NO_2, PM_{10}, and $PM_{2.5}$) through manual sampling conducted twice weekly. This process involves monitoring gaseous pollutants every 4 hours, PM_{10} every 8 hours, and $PM_{2.5}$ over a full 24-hour period, culminating in 104 measurements per year (Central Pollution Control Board, 2024). While NAMP is vital for calculating annual averages for policy compliance, its reliance on manual collection creates significant temporal gaps; for instance, the most recent available data under this stream corresponds only to 2023 (Central Pollution Control Board, 2024). The second methodology, the Con- tinuous Ambient Air Quality Monitoring System (CAAQMS) provides real-time, hourly data for an expanded suite of pollutants,

including SO_2, NO_2, PM_{10}, $PM_{2.5}$, NH_3, CO, and O_3. Apart from this, lead, arsenic and nickel are monitored as a fraction of PM_{10} and are only undertaken as special monitoring done for Diwali by Central Pollution Con- trol Board (CPCB). While all these data are important and provide valuable insight to long term policy making, CAAQMS data is especially essential for identifying localized hotspots and analyzing the rapid temporal fluctuations characteristic of urban pollution cycles. CAAQMS also serves as the functional basis for the daily Air Quality Index (AQI) and immediate public health alerts, offering a more accessible and up-to-date platform for environmental management.

This paper evaluates the adequacy of India's national CAAQMS by cross referenc- ing its spatial distribution with high resolution WorldPop demographic data. Adopting a tiered proximity framework comprising a 2 km accurate radius, a 10 km representative zone, and a 50 km proxy area, the study estimates population exposure across varying levels of data fidelity.

2 Literature Review

Limited air quality monitoring hinders accurate AQI assessment and subsequent miti- gation across many regions of India (Roychowdhury et al., 2023). This phenomenon is expected to obscure unrecognised sources of pollution thus making it hard to make im- provements. Therefore, a more equitable air quality monitoring distribution is needed. Placement of AQI monitoring sations has been a subject of research for a long time. Key frameworks include Population Weighted Pollution Concentration and Weighted Spatial Variability(Munir et al., 2019).Infrastructural deficits and spatial mismatches are symp- tomatic of broader challenges observed across low- and middle-income countries, where the deployment of monitoring networks is frequently dictated by economic capacity and institutional responsiveness rather than environmental problem pressure alone (Schoch et al., 2026). In the Indian context, recent spatial assessments re-emphasise the severity of this deficit, revealing that 62 percent of the national population resides entirely out- side the 50 km coverage radius of the real-time CAAQMS grid, leaving millions without reliable daily air quality alerts (Roychowdhury et al., 2023). Furthermore, the existing monitoring sites are predominantly situated to measure short-term pedestrian or local- ized traffic exposure, failing to adequately characterize long-term population exposure across diverse urban and peri-urban micro-environments (Ecole polytechnique fédérale de Lausanne (EPFL), 2021). Establishing the true geographical representativeness of these individual monitoring point locations remains a complex scientific challenge, as it requires stringent classifications based on highly localized emission categories, dispersion conditions, and demographic distributions (Spangl et al., 2007).

To address this evidentiary gap, recent scholarship has turned toward proxy-based estimations to simulate air quality metrics in regions currently lacking ground-level in- strumentation. Consequently, there has been an intensification of academic interest in high-resolution, predictive air quality modeling to identify localized hotspots and tem- poral fluctuations. These predictive frameworks generally follow three methodological categories: numerical simulation, statistical methods, and machine learning.

Earlier research on air quality prediction primarily relied on numerical simulation, which uses mathematical frameworks to model air quality changes based on the chemical and physical processes in the atmosphere. These models, such as the Nested Air Quality Prediction Modeling System(Wang et al., 2001), the Weather Research and Forecasting (WRF) model (Liang, 2012; Powers, 2017), and the Community Multiscale Air Quality (CMAQ) model (Tang, 2017; Yamaji, 2006), typically combine air quality data with meteorological data. While these models reach a high level of significance, they place heavy demands on datasets and operate under the assumption that pollutant discharge is constant and predictable. In reality, emissions are stochastic and occur randomly(Stern, 2008), making these complex calculations less applicable for continuous assessment.
In response to these inadequacies, Statistical Methods have gained popularity by exploring patterns within observed data rather than relying on complex meteorological data (U. Kumar & Jain, 2010; J. Zhang & Li, 2022; W. Zhang et al., 2022; Zhao et al., 2022). These models are computationally efficient but often assume that dependent and independent variables are linearly correlated(Hao et al., 2021).This is a relationship that is seldom applicable in complex air quality dynamics. Despite this limitation, statistical tools like the Auto Regressive Moving Average (ARMA) model have been widely imple- mented to predict pollutant concentrations at urban traffic sites, such as in Delhi, and multiple linear regression models have assisted governments in making traffic control de- cisions(U. Kumar & Jain, 2010). Therefore, statistical methods are not ideal if one wants to deal with missing data in the form of absent stations and incomplete observations when stations get dysfunctional.

Machine Learning (ML) represents a significant leap forward because these meth- ods perform well even when the relationship between variables is non-linear (Dai et al., 2022). Hybrid models utilizing multilayer perception, support vector machines, and land use regression have demonstrated improved performance across diverse regions. However, standard ML algorithms still have limited capacity for long-term, multi-step predictions and capturing deep temporal data properties(Zhao et al., 2023). Deep Learning, a spe- cialized branch of ML, addresses these gaps. For instance, Long Short-Term Memory (LSTM) networks are highly effective at solving long-distance dependencies, often out- performing traditional ARMA models in high-density areas(X. Li, 2017). Although these advancements significantly improve temporal accuracy, a critical challenge remains in analyzing the spatial characteristics of air quality data. While we can now predict air quality over time with greater precision, analyzing its spatial nature, how pollution at one site relates to neighboring stations, is still an emerging field. Recent scholarship indicates that incorporating spatial correlations can markedly reduce prediction errors.

Since ground stations provide only point-based measurements, achieving contin- uous spatial coverage is difficult regardless of their location (Brauer et al., 2019; Gut- tikunda et al., 2023; Kawano et al., 2025; P. Kumar, 2025; Mathew et al., 2024). As a result, satellite remote sensing methods have gained significant popularity in recent years for monitoring PM2.5 concentrations, providing extensive spatial and temporal coverage, particularly in regions lacking sufficient ground-based monitoring networks (T. Li et al., 2017; Van Donkelaar et al., 2015; Zhu et al., 2023). Despite the logistical difficulties, ground-based networks are considered the gold standard for air quality measurement, as they offer direct, real-time, and location-specific data that are critical for validating satellite-based estimates and for understanding local

pollution dynamics (Kawano et al., 2025; Verghese & Nema, 2023). However, despite most monitoring stations being located in urban areas, the overall density of sites in India remains modest, which limits comprehensive spatial coverage for large-scale air quality monitoring.

3 Data and Methodology

3.1 Data Source

Air quality assessment in India utilizes both manual and continuous methodologies. Uti- lizing these methodologies, the Central Pollution Control Board (CPCB) publishes air quality data in three forms. The first form is by releasing live AQI coverage data from Continuous Ambient Air Quality Monitoring Systems (CAAQMS). The second format is through the daily AQI bulletin, which reports air quality in different cities and towns each day. These readings are also gathered from the CAAQMS network. The CAAQMS has expanded significantly, growing from fewer than 50 stations in 2015 to over 220 by 2019 (Sharma & Mauzerall, 2022). As of January 2026, this study identifies 564 functional continuous monitoring stations across all states and union territories, with the exceptions of Goa, Ladakh, and Lakshadweep, which lacked operational continuous infrastructure at the time of analysis. The third format is annual AQI reports, which contain data from manual stations. Manual monitoring infrastructure remains the primary source of geographic breadth for air quality surveillance. By September 2022, the Central Pol- lution Control Board (CPCB) reported 883 operating manual stations across 379 cities and towns, covering 28 states and 7 Union Territories (Roychowdhury et al., 2023). This network expanded by November 2024, with the National Air Quality Monitoring Pro- gramme (NAMP) operating 966 manual stations across 419 districts (Central Pollution Control Board, 2024). The monitoring of pollutants is carried out for 24 hours (a 4- hourly sampling for gaseous pollutants and an 8-hourly sampling for particulate matter) twice a week, to have 104 observations in a year. The monitoring is carried out with the help of the CPCB, State Pollution Control Boards (SPCB), Pollution Control Com- mittees (PCC) and National Environmental Engineering Research Institute (NEERI). The CPCB co-ordinates with these agencies to ensure uniformity and consistency of air quality data and provides technical and financial support for operating the monitoring stations(Ministry of Environment, Forest & Climate Change, 2019). However, data from these stations is not real time nor is it published in daily reports. The most recent report containing manual station data is from 2023.

The CPCB collects data from its 564 CAAQMS at a high frequency to facilitate real-time air quality assessment. According to the official protocol, these automated stations are designed to generate and transmit data every 15 minutes to the CPCB online portal. This raw 15-minute data is then used as the foundation for all subsequent reporting. On an hourly basis, this data is aggregated to calculate the Air Quality Index (AQI), which is the primary figure communicated to the public. For a daily value to be considered valid and reliable for regulatory purposes, CPCB mandates that a station must provide at least 16 hours of verified data for key pollutants like PM2.5 or PM10. This systematic process, from 15-minute collection to daily validation, ensures a continuous and automated data stream that, according to the CPCB, precludes any possibility of human intervention or manipulation in the reported figures .

3.2 Data Collection

CAAQMS data is publicly accessible through the CPCB portal, where it can be retrieved at either the city or individual station level. The CPCB is a statutory organization which is part of the Ministry of Environment, Forest and Climate Change. As of December 2025, it monitored a national network of 564 operational stations. While station-level data was obtained for all identified sites, the dataset contains gaps; AQI readings are naturally unavailable for periods preceding its commissioning or during intervals of tech- nical downtime and maintenance of these stations. While the Central Pollution Control Board Dashboard (2026) mentions a total of Stations as 589, data for only 564 stations is publicly available from its database at the time of analysis.

Air quality data were collected at six-day intervals corresponding to every 5th, 10th, 15th, 20th, 25th, and 30th day of each month, covering the period from January 2021 to December 2025. This systematic five-day sampling interval was designed to enable consistent temporal coverage while facilitating the detection of gaps in data availability from individual stations. The collected data were subsequently compiled and analyzed to determine the number of valid readings obtained from each station over the study period. All data were sourced from the Central Pollution Control Board web portal, accessible at the CPCB website.

To determine the spatial footprint of the CAAQMS monitoring network, station coordinates were estimated via the Google Maps API based on CPCB locality data, as exact GPS coordinates are not publicly disclosed. While these represent approximate locations, they enable a systematic assessment of the reach in the network. The anal- ysis employs administrative boundaries as of January 2026, comprising 28 states and 8 union territories. Among these 36 entities, Lakshadweep, Ladakh, Daman and Diu, and Goa lack functional CAAQM stations within their borders. However, geographic prox- imity provides proxy coverage; Goa is situated near the KHB Colony station in Karwar (managed by Karnataka State Pollution Control Board), and Dadra and Nagar Haveli is serviced by the GIDC station in Vapi (managed by Gujarat State Pollution Control Board).

3.3 Methodology adopted for estimating population coverage

We evaluate the design of air quality monitoring network based on population in different regions of India. Population estimates are based on the 100mx100m spatial distribution of population in 2020 developed by the WorldPop research programme, based in the School of Geography and Environmental Sciences at the University of Southampton(Bondarenko et al., 2025). Air quality reported by a station is understood to be accurate represen- tation of ambient air in its 2km radius, while it can be considered fairly representative for ambient air 2-10km away from the station. If no major topographical or human-made features exist then the air quality reported at a station can be considered a good proxy for ambient air 10-50km away from the station(Roychowdhury et al., 2023). The 2019 UNICEF report “Silent Suffocation in Africa: Air Pollution is a Growing Menace, Affecting the Poorest Children the Most” uses

same principle to estimate percentage of children living within a 50km radius of air monitoring stations (UNICEF, 2019, 2021). To characterize the demographic reach of the air quality monitoring network, this study employs a spatial buffer analysis based on the above proximity thresholds.

Population coverage is assessed for residents situated within 2km, 10km, and 50km Euclidean (straight-line) distances from each monitoring station. Technically, these zones are modeled as geospatial buffers (or circular catchments) centered on coordinates for each station. The aggregate population residing within these defined geometries is then quantified. In densely monitored metropolitan areas such as Delhi and Mumbai, where inter-station distances are frequently less than 50km, significant spatial overlap between buffers occurs. To ensure the statistical integrity of the effective population coverage, a geometric union is performed on the overlapping buffers. Such a computational ad- justment eliminates any instances of double-counting, ensuring that individuals situated within the reach of multiple sensors are accounted for as a single unit in the total coverage estimate.

4 Results

4.1 Spatial Disparities in Monitoring Infrastructure

The geographical distribution of ambient air quality monitoring stations in India reveals significant inter-state and intra-state disparities in air pollution surveillance infrastruc- ture. The analysis encompasses monitoring data from 28 states and 8 union territories. These stations are located in 372 districts nationwide. Maharashtra leads with the most extensive network, closely followed by Delhi (NCT). The spatial pattern, as depicted in Fig. 1, indicates that CAAQMS stations are densely located in states/UTs such as

Maharashtra, Karnataka, the NCT of Delhi, and other major urban centers, whereas the northeastern states, smaller union territories, and remote areas have very limited moni- toring coverage. This is a heterogeneous distribution of monitoring stations and overlooks the potential of pollution dynamics in peri-urban and rural landscapes where different emission sources and exposure patterns may prevail. The analysis reveal several gaps in data: 40% of the national population, approximately 588 million people, reside more than 50 km from any monitor. While urban centers like Delhi show high regional pres- ence, over 65% of their residents still live outside the high accuracy 2 km zone. In stark contrast, states like Himachal Pradesh and Sikkim leave over 85% of their populations unmonitored, while Ladakh and Lakshadweep lack coverage entirely.

A distinct urban-rural divide characterizes the monitoring network, with approx- imately 73% of stations located in just 47 major urban districts. Mumbai Suburban district alone hosts 21 stations, while Bengaluru Urban district contains 14 stations, this indicates a strong correlation between monitoring density and urban population concen- tration. The Northern Indian states, which are recognised to have very high pollution levels have multiple stations in Delhi and parts of Uttar Pradesh but sparse infrastruc-ture in Bihar and West Bengal. Northern and western regions generally demonstrate better monitoring

coverage compared to eastern and northeastern regions. This points to significant regional disparities in environmental monitoring investment. At the district level, 214 districts have only one monitoring station, limiting their ability to capture intra-district spatial variations in air quality.

The current distribution of CAAQMS stations highlights the information gaps that have a negative impact on air quality management and the development of public health policies. Several states and union territories lack any monitoring infrastructure, creating blind spots in the national air quality assessment framework. The current urban-centric approach may underestimate population exposures in rapidly industrializing peri-urban areas and agricultural regions experiencing crop residue burning.

4.2 Methodological and Data Availability Gaps

Beyond the mere presence of monitoring infrastructure, the completeness and continuity of data reporting represent equally critical dimensions of air quality surveillance. The heatmap in Fig. 2 illustrates the percentage of days for which AQI data was actually reported by each state since 2021, revealing stark disparities in data availability. Punjab, Puducherry, and Nagaland emerge as the top performers, with data availability rates of 90.9%, 90.6%, and 88.2% respectively, indicating relatively consistent reporting through- out the year. The NCT of Delhi, despite its dense monitoring network, achieved a data availability rate of 81.3%, while Chandigarh and Mizoram reported 45.9% and 83.5% respectively. It can be seen that station presence alone does not guarantee continuous data streams for public health monitoring and policy formulation.

The extent of data completeness reveals a concerning pattern where several states with moderate monitoring infrastructure exhibit severely limited data availability. States such as Nagaland and Mizoram, despite having fewer monitoring stations, demonstrate relatively high data completeness, suggesting that operational efficiency and maintenance protocols significantly influence data continuity. Conversely, some states with substantial monitoring infrastructure including Karnataka, Haryana, Tamil Nadu show low data availability rates in atleast four of the five years for which data was collected. This points to potential challenges in equipment maintenance, data transmission, or reporting protocols. This disconnect between infrastructure presence and data output points at the need for comprehensive quality control measures and standardized reporting mechanisms across all states. Without consistent data availability, even well-distributed monitoring networks fail to provide the continuous time-series data essential for trend analysis, health impact assessments, and policy evaluation.

4.3 Population Coverage and Accessibility Analysis

The spatial distribution of air quality monitoring infrastructure in India reveals a signif- icant disparity when viewed through the lens of population coverage. While the national aggregate suggests a broad reach at the regional level, the density of monitoring stations within immediate proximity to the population remains relatively low. As shown in Fig. 3, the all India average for population coverage within a 2 km radius stands at only 3.8% which indicates that for the vast majority of the Indian population, hyper-local air quality data which is most representative of immediate human exposure is largely unavailable.

At the regional scale, the coverage improves significantly, with 60.0% of the na- tional population living within 50 km of a monitoring station. However, this leaves a substantial 40% of the population in a monitoring blind spot, where the nearest sensor is over 50 km away. As illustrated in Fig. 3, the transition from 10 km coverage (18.6%) to 50 km coverage (60.0%) highlights that while regional networks are expanding, there is a critical need for infrastructure to bridge the gap between regional presence and local utility.

The performance of individual states varies drastically, as evidenced by the tiered data in Fig. 3. Union Territories which are urbanized and relatively smaller in area like Delhi and Chandigarh lead the nation, achieving 100% coverage at the 50 km tier and maintaining the highest 2 km local density at 33.0% and 31.0% respectively. In stark con- trast, large states such as Jharkhand and Himachal Pradesh exhibit 2 km coverage rates well below 1%, reflecting a severe infrastructure deficit in protecting these populations.

5 Discussion

The spatial distribution of the CAAQMS network reflects a monitoring framework shaped by population exposure. The evidence behind this is the overwhelming and dispropor- tionate presence of CAAQMS stations in cities which have large populations. Approxi- mately 73 percent of all stations are concentrated within 47 urban districts, and while states such as Maharashtra and the National Capital Territory of Delhi possess dense monitoring clusters, an estimated 588 million individuals comprising 40 percent of the national population reside more than 50 kilometers from any functional station. This disparity is particularly acute in Ladakh, Lakshadweep, Himachal Pradesh, and Sikkim, where infrastructure remains insufficient to capture ambient air quality, effectively ex- cluding rural and peri-urban populations from the national assessment framework. The concentration of stations in metropolitan districts such as Mumbai Suburban and Ben- galuru Urban, while administratively logical, leaves these populations without recourse to real-time data, as emission profiles in peri-urban zones differ substantially from ur- ban cores. Agricultural burning, brick kilns, and unregulated industrial activity generate pollution dynamics that urban-centric placement is ill-equipped to capture. Although aggregate statistics indicate that 60 percent of the population now falls within the 50 kilometer threshold, reflecting incremental improvement since the 2023 baseline estab- lished by Roychowdhury et al. (2023), such figures obscure localized deficiencies. Spatial reach at the 50 kilometer tier functions as a proxy measure at best, since station rep- resentativeness degrades considerably with distance, particularly in regions with varied topography.

The data availability analysis between 2021 and 2025 exposes a structural discon- nect between monitoring infrastructure and functional surveillance. States with moder- ate station counts, including Nagaland and Mizoram, sustained higher data completeness rates than several states with far larger networks, which points to the primacy of oper- ational governance over hardware investment. The improvement observed in Chhattis- garh and Rajasthan, where availability rose from near zero in 2021 to above 80 percent by 2025, indicates that institutional attention and maintenance protocols can produce measurable gains within short timeframes. By contrast, the persistent underperformance of Jharkhand

and Jammu and Kashmir across all five years of the study period reflects entrenched systemic deficits rather than transitional technical failures. The implications for time-series analysis are consequential: irregular and incomplete data records preclude the identification of pollution trends, seasonal patterns, and the evaluation of policy interventions introduced under the National Clean Air Programme.

The adequacy of an air quality monitoring network cannot be assessed on the basis of station count or geographic spread alone. Schoch et al. (2026) demonstrate that monitoring density across low- and middle-income countries correlates with income levels and the character of political institutions, suggesting that the deficits identified in this analysis are structurally conditioned rather than incidental. The Indian case illustrates this pattern, where democratic accountability has driven network expansion in politically visible urban centers while more remote and less electorally salient populations remain outside the coverage radius. Furthermore, the current monitoring network faces inade- quate data generation, limited completeness, and poor quality control, and instances of strategic misreporting have been documented in various global contexts where monitors are placed in relatively cleaner areas to avoid scrutiny (Schoch et al., 2026). Hybrid mon- itoring approaches that integrate reference-grade stations with low-cost sensor networks and satellite-derived aerosol optical depth data offer a methodologically viable path to- ward filling spatial gaps that ground-based infrastructure alone cannot address within current fiscal and logistical constraints. Strategic deployment guided by emission inventories, population exposure modeling, and land-use classification would allow incremental network expansion to yield disproportionate gains in coverage relative to cost.

6 Conclusion

The findings of this study confirm that India's CAAQMS faces structural deficits in both spatial reach and operational continuity. With 40 percent of the national population, approximately 588 million residents, located beyond 50 kilometers of any functional sta- tion, the network fails to provide reliable ambient air quality data to a substantial share of those most in need of health protection. The concentration of 73 percent of stations within 47 urban districts, alongside the complete absence of monitoring infrastructure in Ladakh and Lakshadweep, reflects a distribution pattern driven by institutional capacity and population density rather than pollution burden or public health need. At the hyperlocal scale, national coverage within 2 kilometers stands at 3.8 percent, which means that precise, location-specific exposure data remains unavailable to the vast majority of the population.

Data availability analysis between 2021 and 2025 further reveals that infrastructure deployment alone does not translate into consistent surveillance. States such as Jharkhand and Jammu and Kashmir recorded data availability below 30 percent across the study period, while Maharashtra, despite operating 97 stations, reported below 35 percent availability in 2021. These patterns indicate systemic failures in equipment maintenance, data transmission protocols, and quality control. Conversely, states such as Punjab, Puducherry, and Nagaland demonstrated that high data completeness is achievable even within networks of moderate scale, suggesting that operational management and institutional commitment are as consequential as the number of stations deployed.

Addressing these deficiencies requires a dual policy response. Expansion of the monitoring network must prioritize underserved states, rural districts, and regions with documented pollution sources that currently lack coverage. In parallel, regulatory authorities must establish standardized protocols for data validation and reporting to ensure that existing stations generate continuous, usable time-series data. The integration of low-cost sensor networks and satellite-derived estimates offers a supplementary path for bridging spatial gaps in the near term. Without simultaneous investment in both network reach and data reliability, the CAAQMS will continue to underperform as a foundation for public health communication and evidence-based environmental governance.

7 Study Limitations

A primary limitation of the present study concerns spatial uncertainty. Because exact GPS coordinates for the CAAQMS stations are not publicly disclosed, station locations were approximated via the Google Maps API using available locality data. Although these estimated coordinates provide a reliable baseline for systematically assessing the broader geographic reach of the network, generating spatial buffers from approximated centroids introduces a potential margin of error in the population exposure calculations. Specifically, if a physical sensor is actually situated in a localized area with a significantly different population density than its corresponding API estimated coordinate, the demo- graphic counts aggregated within the high precision boundaries (such as the 2 km radius) may inadvertently overestimate or underestimate the actual population served by that station.

Declarations

Acknowledgments

The authors acknowledge the Central Pollution Control Board, from where the data for this study were downloaded. They also thank the reviewers for their time and valuable comments.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

All authors contributed equally to the study conception, design, material preparation, data collection, analysis, and manuscript preparation. The first draft of the manuscript was written by Abhinav Kapoor and subsequently, both authors commented on previous versions of the manuscript. Both authors read and approved the final manuscript.

Data Availability Statement

The datasets generated during and/or analysed during the current study are available on the Central Pollution COntrol Board Portal (https://airquality.cpcb.gov.in/AQI_India/). The data will also be made available from the corresponding author on reasonable request.

Ethics Approval

This study did not involve human participants, their data, or biological material, nor did it involve animals, their data, or biological material. Therefore, no ethical approval was required. The analysis was conducted using publicly available secondary data, and no ethical clearance was necessary.

Consent to Participate

Not applicable. This study did not involve human participants.

Consent to Publish

Not applicable. This manuscript does not contain any individual person's data in any form (including any individual details, images, or videos).

References

1. Abdurrahman, M., Chaki, S., & Saini, G. (2020). Stubble burning: Effects on health & environment, regulations and management practices. *Environmental Advances*, *2*. https://doi.org/10.1016/j.envadv.2020.100011 Export Date: 06 February 2026; Cited By: 241.
2. Bhagat, J., & Saha, G. (2021). Pollutant PM2.5 Multi Step Prediction under Seasonal Influences across 13 Indian Cities. https://doi.org/10.1109/i-PACT52855.2021. 9696866 Export Date: 06 February 2026; Cited By: 4.
3. Bharath, M., Srinivasalu, S., Natesan, U., Ayyamperumal, R., Karuppannan, S., Gopalakr- ishnan, G., & Nazir, N. (2020). Air pollution improvement and mortality rate during COVID-19 pandemic in India: Global intersectional study. *Air Quality, Atmosphere and Health*, *13*(11), 1375–1384. https://doi.org/10.1007/s11869-020- 00892-w Export Date: 06 February 2026; Cited By: 65.
4. Bondarenko, M., Priyatikanto, R., Tejedor Garavito, N., Zhang, W., McKeen, T., Cun- ningham, A., Woods, T., Hilton, J., Cihan, D., Nosatiuk, B., et al. (2025). Con- strained estimates of 2015-2030 total number of people per grid square at a reso- lution of 3 arc (approximately 100m at the equator) r2025a version v1.
5. Brauer, M., Guttikunda, S. K., K A, N., Dey, S., Tripathi, S. N., Weagle, C., & Martin, R. V. (2019). Examination of monitoring approaches for ambient air pollution: A case study for india [Type: Article]. *Atmospheric Environment*, *216*. https ://doi.org/10.1016/j.atmosenv.2019.116940
6. Central Pollution Control Board. (2024). *Ambient air quality monitoring network (man- ual) under NAMP as on 19.11.2024* (tech. rep.). Central Pollution Control Board, Ministry of Environment, Forest

and Climate Change. Retrieved February 12, 2026, from https://cpcb.nic.in/monitoring-network-3/

7. Central Pollution Control Board Dashboard. (2026). *Continuous ambient air quality mon- itoring (caaqm) dashboard* [Accessed on 20 March 2026]. Retrieved March 20, 2026, from https://airquality.cpcb.gov.in/ccr/#/caaqm-dashboard-all/caaqm-landing
8. Chelani, A. (2016). Long-range correlations in air quality time series: Effect of differencing and shuffling. *Aerosol and Air Quality Research*, *16*(9), 2302–2313. https://doi.org/10.4209/aaqr.2016.04.0139 Export Date: 06 February 2026; Cited By: 4.
9. Dadhich, A., Goyal, R., & Dadhich, P. (2018). Assessment of spatio-temporal variations in air quality of Jaipur city, Rajasthan, India. *Egyptian Journal of Remote Sensing and Space Science*, *21*(2), 173–181. https://doi.org/10.1016/j.ejrs.2017.04.002 Export Date: 06 February 2026; Cited By: 58.
10. Dai, H., Huang, G., Zeng, H., & Zhou, F. (2022). Pm2.5 volatility prediction by xgboost- mlp based on garch models. *J. Clean. Prod.*, *356*. https://doi.org/10.1016/j. jclepro.2022.131898
11. Ecole polytechnique fédérale de Lausanne (EPFL). (2021, September). *Clean air project in india detailed report on strategies for improving air quality monitoring* (tech. rep.) (Supported by: Swiss Agency for Development and Cooperation (SDC)). Lausanne, Switzerland.
12. Gautam, A., Kumar, S., Gautam, S., Anand, A., Kumar, R., Joshi, A., Bauddh, K., & Singh, K. (2021). Pandemic induced lockdown as a boon to the Environment: Trends in air pollution concentration across India. *Asia-Pacific Journal of Atmo- spheric Sciences*, *57* (4), 741–756. https://doi.org/10.1007/s13143-021-00232-7 Export Date: 06 February 2026; Cited By: 22.
13. Ghosh, D. (2025). Assessing air quality extremes: A comparative extreme value analysis of metropolitan cities across India and the world. *Environmental Monitoring and Assessment*, *197* (3). https://doi.org/10.1007/s10661-025-13754-8 Export Date: 06 February 2026; Cited By: 2.
14. Govindaraj, R., Sarkar, A., Jyotsna Rao, J., & Pal, U. (2021). Comparison of Indoor Air Pollution in a Residence in Bangalore, India: A Weekday-Weekend Analysis. https://doi.org/10.1109/IBSSC53889.2021.9673459 Export Date: 06 February 2026; Cited By: 3.
15. Guttikunda, S., Ka, N., Ganguly, T., & Jawahar, P. (2023). Plugging the ambient air monitoring gaps in india's national clean air programme (NCAP) airsheds [Type: Article]. *Atmospheric Environment*, *301*. https://doi.org/10.1016/j.atmosenv. 2023.119712
16. Hao, Y., Gai, Z., Yan, G., Wu, H., & Irfan, M. (2021). The spatial spillover effect and nonlinear relationship analysis between environmental decentralization, govern- ment corruption and air pollution: Evidence from china. *Sci. Total Environ.*, *763*. https://doi.org/10.1016/j.scitotenv.2020.144183
17. Kawano, A., Kelp, M., Qiu, M., Singh, K., Chaturvedi, E., Dahiya, S., Azevedo, I., & Burke, M. (2025). Improved daily PM2.5estimates in india reveal inequalities in recent enhancement of air quality [Type: Article]. *Science Advances*, *11*(4). https://doi.org/10.1126/sciadv.adq1071
18. Kumar, P. (2025). Exploring the influence of human activities and the COVID-19 lock- down on urban air quality in Rajasthan, India. *Theoretical and Applied Climatol- ogy*, *156*(4). https://doi.org/10.1007/s00704-025-05445-8 Export Date: 06 February 2026; Cited By: 2.

19. Kumar, U., & Jain, V. (2010). Arima forecasting of ambient air pollutants (o3, no, no2 and co). *Stoch. Environ. Res. Risk Assess.*, *24*. https://doi.org/10.1007/s00477- 009-0361-8
20. Li, T., Shen, H., Yuan, Q., Zhang, X., & Zhang, L. (2017). Estimating ground-level PM2.5 by fusing satellite and station observations: A geo-intelligent deep learning approach [Type: Article]. *Geophysical Research Letters*, *44*(23), 11, 985–11, 993. https://doi.org/10.1002/2017GL075710
21. Li, X. (2017). Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environ. Pollut.*, *231*. https ://doi.org/10.1016/j.envpol.2017.08.114
22. Liang, X. .-. Z. (2012). Regional climate-weather research and forecasting model. *Bull. Am. Meteorol. Soc.*, *93*. https://doi.org/10.1175/BAMS-D-11-00180.1
23. Markandeya, n., Verma, P., Mishra, V., Singh, N., Shukla, S., & Mohan, D. (2021). Spatio-temporal assessment of ambient air quality, their health effects and im- provement during COVID-19 lockdown in one of the most polluted cities of In- dia. *Environmental Science and Pollution Research*, *28*(9), 10536–10551. https://doi.org/10.1007/s11356-020-11248-3 Export Date: 06 February 2026; Cited By: 29.
24. Mathew, A., Shekar, P. R., Nair, A. T., Mallick, J., Rathod, C., Bindajam, A. A., Al- harbi, M. M., & Abdo, H. G. (2024). Unveiling urban air quality dynamics during COVID-19: A sentinel-5p TROPOMI hotspot analysis [Type: Article]. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-72276-4
25. Ministry of Environment, Forest & Climate Change. (2019). *National clean air programme (ncap)* (N. K. Sundaray & S. R. Bhardwaj, Eds.; Programme Document). Gov- ernment of India. New Delhi, India. https://prana.cpcb.gov.in/#/resources
26. Munir, S., Mayfield, M., Coca, D., & Jubb, S. A. (2019). Structuring an integrated air quality monitoring network in large urban areas – discussing the purpose, criteria and deployment strategy. *Atmospheric Environment: X*, *2*. Retrieved February 22, 2026, from https://eprints.whiterose.ac.uk/id/eprint/145671/
27. Nair, A., Khare, S., & Thakur, A. (2023). Air Quality Index Prediction of Bangalore City Using Various Machine Learning Methods. *615 LNNS*, 391–406. https://doi.org/ 10.1007/978-981-19-9304-6_37 Export Date: 06 February 2026; Cited By: 5.
28. Powers, J. G. (2017). The weather research and forecasting model: Overview, system efforts, and future directions. *Bull. Am. Meteorol. Soc.*, *98*. https://doi.org/10. 1175/BAMS-D-15-00308.1
29. Rana, R., & Kumar, N. (2024). Smart Air: A Spatiotemporal Attention Based Deep Learning Approach for Accurate PM2.5 and PM10 Forecasting. *Earth Systems and Environment*. https://doi.org/10.1007/s41748-024-00551-5 Export Date: 06 February 2026; Cited By: 1.
30. Roychowdhury, A., Somvanshi, A., & Kaur, S. (2023, July). *Status of air quality moni- toring in india: Spatial spread, population coverage and data completeness* (tech. rep.) (Lead research: Avikal Somvanshi). Centre for Science and Environment. New Delhi. Retrieved February 12, 2026, from https://www.cseindia.org/Note- AQM-Network-analysis.pdf

31. Schoch, M., Fournier De Lauriere, C., & Bernauer, T. (2026). Large gaps in monitoring urban air pollution in low- and middle- income countries associated with economic conditions and political institutions. *PLoS One*, *21*(1), e0321358. https://doi.org/ 10.1371/journal.pone.0321358
32. Sharma, D., & Mauzerall, D. (2022). Analysis of air pollution data in india between 2015 and 2019. *Aerosol and Air Quality Research*, *22*(2), 210204. https://doi.org/10. 4209/aaqr.210204
33. Singh, A., Soni, P., Krishnan, D., & Potle, I. (2023). Ensemble Learning with Hybrid Modelling for Multivariate AQI, PM2.5, and PM10 Forecasting in Mumbai. https://doi.org/10.1109/CSET58993.2023.10346745 Export Date: 06 February 2026; Cited By: 1.
34. Soni, M., Payra, S., & Verma, S. (2018). Particulate matter estimation over a semi arid region Jaipur, India using satellite AOD and meteorological parameters. *Atmo- spheric Pollution Research*, *9*(5), 949–958. https://doi.org/10.1016/j.apr.2018.03. 001 Export Date: 06 February 2026; Cited By: 55.
35. Spangl, W., Schneider, J., Moosmann, L., & Nagl, C. (2007). *Final report: Representative- ness and classification of air quality monitoring stations* (tech. rep. No. REPORT REP-0121). Umweltbundesamt GmbH. Vienna, Austria.
36. Stern, R. (2008). A model inter-comparison study focussing on episodes with elevated pm10 concentrations. *Atmos. Environ.*, *42*. https://doi.org/10.1016/j.atmosenv. 2008.01.068
37. Tang, Y. (2017). A case study of aerosol data assimilation with the community multi- scale air quality model over the contiguous united states using 3d-var and optimal interpolation methods. *Geosci. Model Dev.*, *10*. https://doi.org/10.5194/gmd-10- 4743-2017
38. UNICEF. (2019, June). *Silent suffocation in africa air pollution is a growing menace, affecting the poorest children the most* (Report). United Nations Children's Fund (UNICEF). New York, NY. https://www.unicef.org/reports/silent-suffocation- in-africa-air-pollution-2019
39. UNICEF. (2021). *The climate crisis is a child rights crisis: Introducing the children's climate risk index* (tech. rep.). United Nations Children's Fund (UNICEF). New York, NY. https://www.unicef.org/media/105376/file/UNICEF-climate-crisis- child-rights-crisis.pdf
40. Van Donkelaar, A., Martin, R. V., Brauer, M., & Boys, B. L. (2015). Use of satellite ob- servations for long-term exposure assessment of global concentrations of fine particulate matter [Type: Article]. *Environmental Health Perspectives*, *123*(2), 135–143. https://doi.org/10.1289/ehp.1408646
41. Verghese, S., & Nema, A. K. (2023). Assessment of air quality monitoring network for adequacy in representation of urban PM2.5 and coverage efficiency: A case study of delhi, india [Type: Article]. *Environmental Engineering Science*, *40*(9), 373–380. https://doi.org/10.1089/ees.2023.0029
42. Wang, Z., Maeda, T., Hayashi, M., Hsiao, L. .-. F., & Liu, K. .-. Y. (2001). A nested air quality prediction modeling system for urban and regional scales: Application for high-ozone episode in Taiwan. *Water Air Soil Pollut.*, *130*. https://doi.org/10. 1023/A:1013833217916
43. Yamaji, K. (2006). Analysis of the seasonal variation of ozone in the boundary layer in east Asia using the community multi-scale air quality model: What controls surface ozone levels over Japan? *Atmos. Environ.*, *40*. https://doi.org/10.1016/j. atmosenv.2005.10.067

44. Yeasin, M., Paul, R., Das, S., Deka, D., & Karak, T. (2023). Change in the air due to the coronavirus outbreak in four major cities of India: What do the statistics say? *Journal of Hazardous Materials Advances*, *10*. https://doi.org/10.1016/j.hazadv. 2023.100325 Export Date: 06 February 2026; Cited By: 2.
45. Zhang, J., & Li, S. (2022). Air quality index forecast in Beijing based on CNN-LSTM multi-model. *Chemosphere*, *308*. https://doi.org/10.1016/j.chemosphere.2022. 136180 Export Date: 18 February 2026; Cited By: 166; Correspondence Address: S. Li; School of Mathematical Sciences, Shanxi University, Taiyuan, China; email: lisy75@sxu.edu.cn; CODEN: CMSHA.
46. Zhang, W., Zhang, R., Wu, C., Goh, A., & Wang, L. (2022). Assessment of basal heave stability for braced excavations in anisotropic clay using extreme gradient boosting and random forest regression. *Underground Space (China)*, *7* (2), 233–241. https://doi.org/10.1016/j.undsp.2020.03.001 Export Date: 18 February 2026; Cited By: 99; Correspondence Address: R. Zhang; School of Civil Engineering, Chongqing University, Chongqing, 400045, China; email: zhangrh@cqu.edu.cn.
47. Zhao, Z., Wu, J., Cai, F., Zhang, S., & Wang, Y. .-. G. (2022). A statistical learning framework for spatial-temporal feature selection and application to air quality index forecasting. *Ecol. Ind.*, *144*. https://doi.org/10.1016/j.ecolind.2022.109416
48. Zhao, Z., Wu, J., Cai, F., Zhang, S., & Wang, Y.-G. (2023). A hybrid deep learning framework for air quality prediction with spatial autocorrelation during the covid- 19 pandemic. *Scientific Reports*, *13*(1), 1015.
49. Zhu, S., Tang, J., Zhou, X., Li, P., Liu, Z., Zhang, C., Zou, Z., Li, T., & Peng, C. (2023). Research progress, challenges, and prospects of PM2.5 concentration estimation using satellite data [Type: Article]. *Environmental Reviews*, *31*(4), 605–631. https://doi.org/10.1139/er-2022-0125

Figures

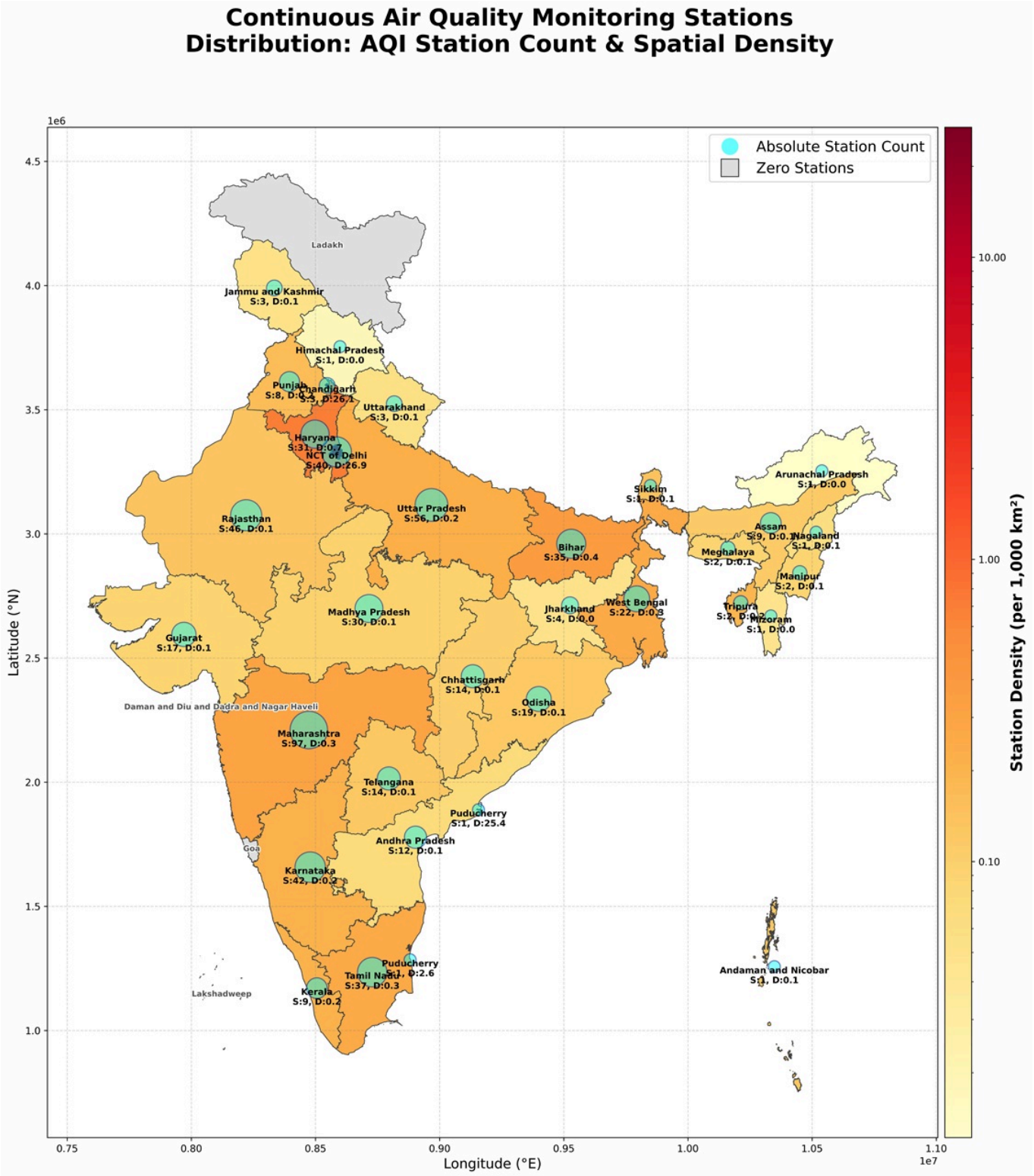


Figure 1

State-wise distribution of AQI monitoring stations across India.

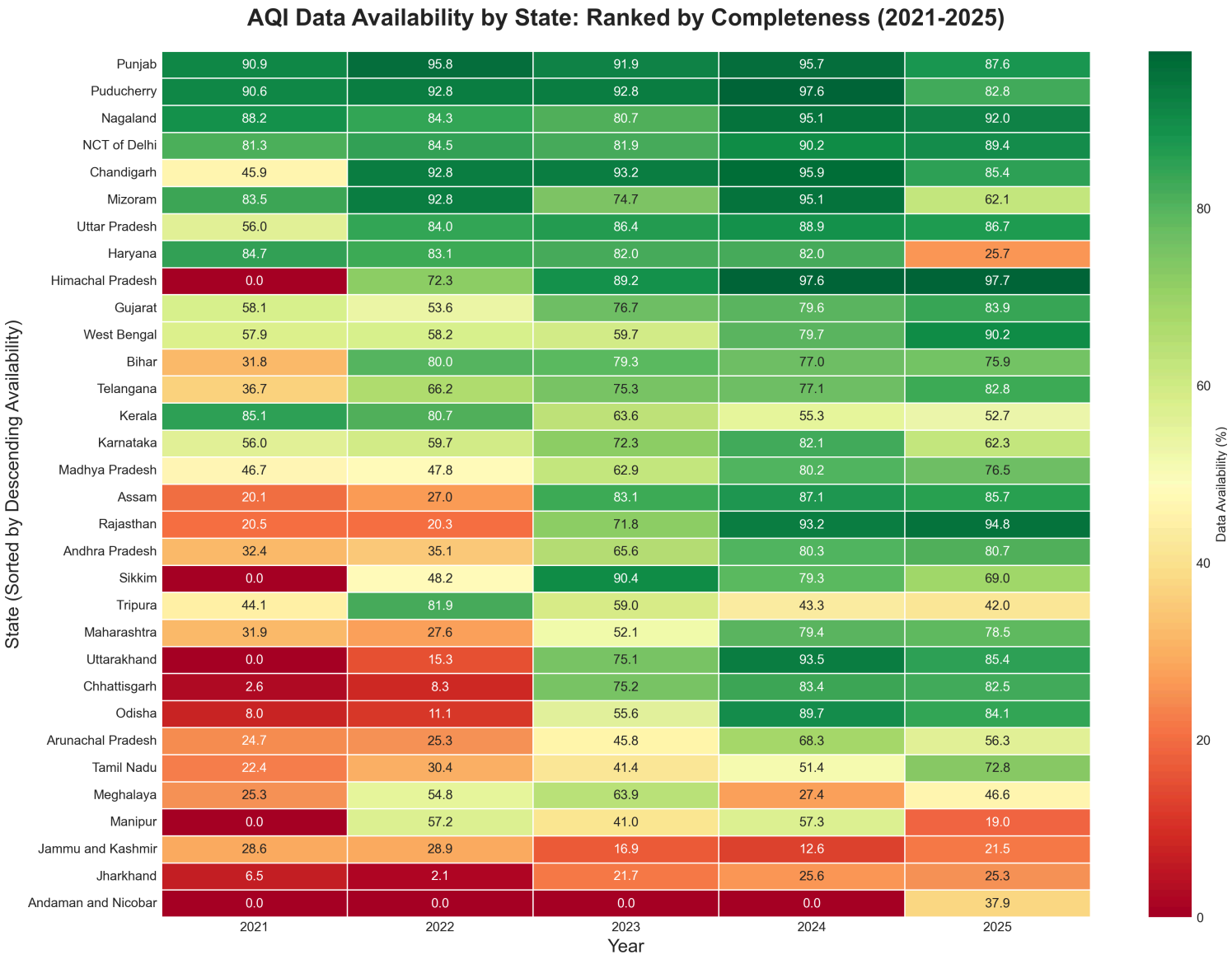


Figure 2

State-wise comparison of AQI data availability for the years 2021 to 2025.

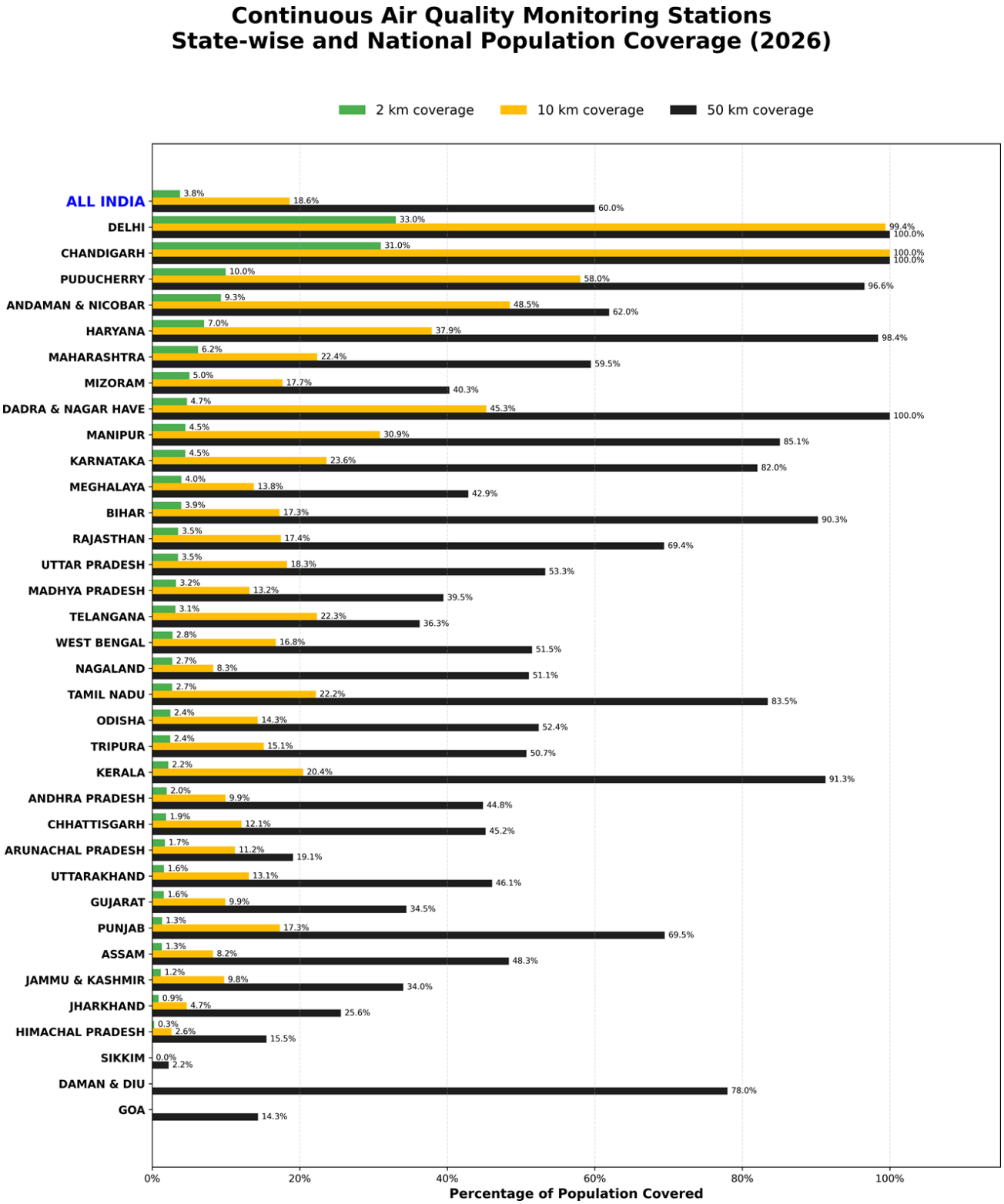


Figure 3

State-wise and National Population Coverage of Continuous Ambient Air Qual- ity Monitoring Networks across 2 km, 10 km, and 50 km Tiers.