Evaluating the Environmental Impact of Metro Systems in Indian Cities

Master Thesis

Master of Science in Management

Submitted by

KURUGANTI, Priyanka

Under the supervision of:

Prof. Matthieu Stigler

University of Geneva

Geneva School of Economics and Management

Declaration of Originality

I certify that the work presented here is, to the best of my knowledge and belief, original and the result of my own investigations, except as acknowledged, and has not been submitted, either in part or whole, for a degree at this or any other University.

Signature:

Priyanka Kuruganti

Date: 14th November 2025

Contents

Abstract

This thesis examines whether metro rail expansions and construction in India have reduced ambient particulate matter (PM2.5) pollution using a panel dataset covering 11 Indian cities with metro systems and 121 control cities between 2008 and 2024. The staggered difference-in-differences estimator of Callaway and Sant'Anna (2021), together with spatial econometric models, is used to identify both temporal treatment effects and spatial spillovers. Effects are estimated at two spatial scales: the city level and the locality level. The estimates suggest that the opening of metro lines is associated with declines in city-level PM2.5 concentrations but increases in PM2.5 in neighborhoods immediately surrounding new stations at the locality level. These findings are consistent with existing evidence indicating that metro transit can reduce aggregate pollution in large, highly polluted cities while generating localized emissions increases around localities.

AI Usage Note: AI tools (ChatGPT and Grammarly) were used to assist with grammar refinement, debugging formatting/structuring issues and helping in creating non-data representation images. The author confirms that all conceptual development, analysis, writing of content, interpretation of results, and academic argumentation remain the author's own work. All content was manually reviewed and verified by the author to maintain authenticity and academic integrity.

Chapter 1

Introduction

Urban air pollution is a serious concern in developing countries. Developing countries, characterised by extensive industrial development, urbanisation, and rising private vehicle ownership, have experienced unprecedented levels of fine particulate matter (PM2.5) pollution. As a developing economy, India faces this challenge: Indian cities rank among the world's most polluted, with annual mean PM2.5 concentrations among the highest.

To address pollution at the city level, countries are investing in cleaner transportation alternatives. One major strategy is to shift commuters from private vehicles (cars, motorcycles) to public transit, such as buses, light rail, and metro systems. Metro rail networks offer high-frequency, separated service from road traffic, thereby increasing mobility. Theoretically, each passenger-km traveled by metro should produce fewer emissions than the same trip by car or motorbike. Moreover, metros can reduce road congestion, leading to smoother traffic flow and thereby lower emissions.

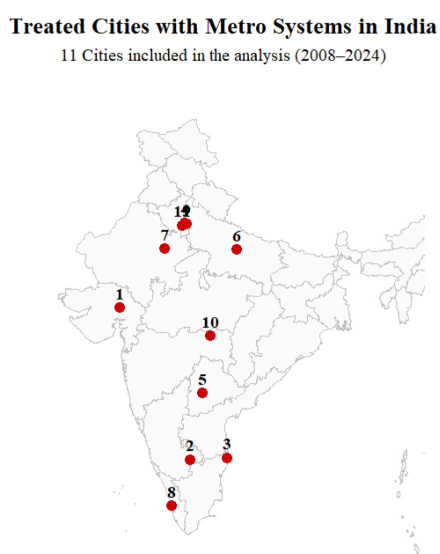
Cities in India have invested heavily in public transportation, especially in metro infrastructure. These systems are seen as solutions to reduce congestion and pollution. However, the real-world environmental impact of metro expansion in India remains uncertain. Indian metro fares tend to be relatively high compared to bus fares, and networks sometimes lack full integration with buses and other informal transit, such as shared autos. As a result, the metro's actual impact is unlikely to be large. Further, the construction of metros can itself generate dust and short-term traffic disruptions.

This makes the true effect of metro construction on pollution rather uncertain.

This thesis addresses the gap by conducting a pan-India analysis of the impact of metro rail on air pollution. We use a panel dataset for 11 Indian cities with metro systems (covering 2008–2023), along with data from 121 control cities without metros. Using a staggered Difference-in-Differences (DiD) approach across cities (Callaway & Sant'Anna 2021), we estimate the causal effect of metro operationalization on ambient PM2.5 concentrations. The analysis also incorporates controls for weather and socioeconomic factors at the local and city levels.

The findings indicate that, on average, metro expansions lead to a reduction in PM2.5 at the city level, but an increase at the local level. Spatial analysis shows that PM2.5 levels worsened in areas near metro lines, with no effect farther away. This suggests the importance of local coverage and last-mile connections.

The remainder of the thesis is as follows. Chapter 2 reviews the literature on urban transit and pollution. Chapter 3 describes the data and variable construction. Chapter 4 outlines the empirical methodology. Chapter 5 presents results and analysis. Chapter 6 discusses the limitations, Chapter 7 addresses policy implications, and Chapter 8 concludes the thesis.



label	city
1	Ahmedabad
2	Bengaluru
3	Chennai
4	Delhi
5	Hyderabad
6	Lucknow
7	Jaipur
8	Kochi
9	Noida
10	Nagpur
11	Gurgaon

Figure I: Treated cities with metro systems in India. The numbers correspond to city names listed in the table.

Chapter 2

Literature Review

I Theory: Metros, Congestion and Pollution Mechanisms

Rising economic growth is reflected in increased demand for personal vehicles, which in turn leads to higher pollution. In theory, metros reduce pollution by replacing trips that would otherwise be made by cars or two-wheelers, thereby lowering per-capita emissions. They also alleviate surface congestion, reducing idling and stop-and-go driving, which are major sources of vehicular emissions. Anderson *et al.* (2014) show that disruptions to public transit (e.g., strikes) increase road congestion, indicating that reliable transit, such as metros, can reduce traffic impacts. Gu *et al.* (2021) find that new subway lines in Chinese cities cut peak-hour congestion and commute times, leading to better air quality. However, the true impact of metro construction remains a matter of concern. Duranton *et al.* (2011) argue that traffic levels lead to induced demand, and without managing this demand, only part of the congestion relief can be reduced as people make new trips. Similarly, Heblich *et al.* (2020) claim that newly connected metro areas may experience economic growth and population inflows, thereby increasing overall travel. Thus, the pollution benefits of metros depend on whether they induce a shift and reduce car use.

II Evidence from Developed Countries

When looking at existing literature on metros and pollution impacts, research in Sydney has shown that the impact is weaker at night than during the day and greater outside (Laupruendee *et al.*, 2022). Further, it plays a role in infrastructure development and overall economic development (Chan *et al.*, 2019). Similarly, research in Greece shows that the absence of a public transport system has led to increased car use, thereby worsening pollution (Porschalidou *et al.*, 2016). Chen *et al.* (2012) study the Taipei Metro expansion using a spatial regression discontinuity design; they find that CO pollution fell by 5–15% in areas within 2 km of new stations, though the effect faded at larger distances.

Existing studies have heavily focused on China. Lin *et al.* (2023) show that new subway openings lead to significant reductions in annual PM2.5. Li *et al.* (2016) find that 14 new lines in Beijing measurably improved air quality, with effects stronger near higher-density station networks. Other studies, such as Sun *et al.* (2019), highlight that short-term pollution spikes occur during the construction phases of Chinese metros. Gendron-Carrier *et al.* (2020) conduct a multi-city analysis (58 cities worldwide) and conclude that subway openings have only a small average effect on PM2.5, and those cities with heavy pollution and population have had the most impacts. These mixed results suggest that high-density, highly polluted cities may see larger effects. However, Beaudoin *et al.* (2020) emphasise that the air quality impact of metros is often localised and may require several years to fully materialise, and they also note that simply building track is not enough, as ridership levels (shaped by fare, service, frequency, and connectivity) determine the scale of the shift from private transport to public transportation.

III Behavioural Aspects of Metro Construction

Mass transit can also alter urban centres and travel patterns in the long run. Baum-Snow *et al.* (2007) show that new highways decentralise cities by making commuting easier; similarly, metro lines often lead to development around stations (Heblich *et al.*, 2020). This station-area growth can lead to both positive and negative outcomes: on the one hand, it increases urban density and development near metro lines, which may encourage more trips overall. However, their paper contains a contradiction. Hsu *et al.* (2014) note that in China, rapid transit expansion has sometimes led to increased car ownership, especially as rising incomes make private cars more attractive due to the preference for privacy.

IV Evidence from Developing Countries

Studies from existing developing countries are essential for understanding the effects of metro expansion and construction on pollution, given their similarities in terms of high population density and overcrowding. Evidence from developing countries could thereby provide a basis for understanding the Indian context and offer relevant comparisons. In the context of developing countries, subway/metro constructions have led to improvement in air quality (Zahidul Islam *et al.* (2024), Leirão *et al.* (2021), Zhang *et al.* (2019), Xiao *et al.* (2017)). However, post-construction, while there was an initial decline, this was followed by an increase in the severity of effects. Hence, there exists a need to examine demand-induced effects to see the true impact.

V Evidence from India

When looking at studies focused on metro construction in India, existing research focuses heavily on its impacts in Delhi. There is the case of high-density changes, where metro construction can alter the urban environment (Hemashree *et al.*, 2022; Bagel, 2003). In their paper, Goel *et al.* (2015) show that after the post-extension of the metro line in Delhi, carbon monoxide levels declined, but there was no significant drop in PM2.5 levels. Further, the city exceeds the established PM2.5 limit on more than 85% of days (Shovan *et al.*, 2017). While other studies exist (Ambade *et al.*, 2023), they primarily focus on descriptive analyses of metro systems in India rather than capturing the true impact of metro construction. Studies of other metros (Bangalore, Chennai, Kolkata) are mostly descriptive or short-term in nature.

Further, the limited number of weather stations in each city in the Indian context makes it difficult to determine the true impact of air pollution. Further, Indian analyses ignore spatial spillovers: a station in one locality may reduce pollution a few kilometres away, but these patterns have not been explored. Finally, socioeconomic heterogeneity (rich vs. poor neighbourhoods) may affect the effects of pollution.

VI Research Gap and Contribution

The literature review indicates that metros improve air quality; however, the magnitude of the effect varies by context. Indian cities, due to extreme pollution, diverse geography, and traffic patterns, require careful understanding. Therefore, it is necessary to determine whether metro construction has led to a decline in pollution in the Indian context. However, existing studies of Indian metro systems remain limited in scale, often relying on single-city case studies and failing to exploit cross-city variation. Through a pan-India, multi-city design, metro cities are analysed together rather than focusing on singular effects, helping understand heterogeneity across cities and levels of development. The analysis leverages the fact that cities can be divided into small groups, enabling a more spatially granular analysis that measures pollution directly around metro stations and controls for neighbourhood characteristics. This thesis uses the staggered Difference-in-Differences method of Callaway and Sant'Anna (2021), which accommodates cities opening metros at different times, leading to an unbiased average treatment effect under weaker assumptions. It also considers how local policies and socio-economic factors might influence results, and though the core estimates focus on metro effects, it discusses interactions with these factors. Hence, this thesis aims to provide evidence on whether and how metro systems reduce PM2.5 in Indian cities.

Building on the gaps identified above, the next section describes the empirical data used in this thesis.

Chapter 3

Data and Variable Construction

I PM$_{2.5}$

Satellite-derived estimates of ground-level PM$_{2.5}$ were obtained from the Atmospheric Composition Analysis Group (ACAG) at Washington University in St. Louis (Version V6.GL.02.04). These estimates are generated by combining Aerosol Optical Depth (AOD) from multiple satellite instruments (MODIS/Terra, MODIS/Aqua, MISR/Terra, SeaWiFS/SeaStar, VIIRS/SNPP, and VIIRS/NOAA–20) with geophysical simulations from the GEOS-Chem chemical transport model. The resulting geophysical PM$_{2.5}$ fields are then calculated from ground-based monitor observations using a residual Convolutional Neural Network (CNN). The final product provides monthly mean PM$_{2.5}$ concentrations on a $0.01^\circ \times 0.01^\circ$ grid ($\approx$ 1.2 km sq. resolution(DCEEW)).

At the city level, I define city boundaries through Google searches to establish the limits in each direction. For accuracy and consistency, I cross-reference multiple online sources and official city maps, including those of India. I verify coordinates and descriptions before implementation using the sf package in R to facilitate geospatial operations. However, due to variations in source locations, an expected positional boundary error of ±500 meters is expected. This uncertainty is important for transparency, as it can affect the precise assignment of grid data. Because cities span multiple data grids, the average of all points within each city boundary grid is used to determine monthly PM 2.5 levels.

In addition, since $PM_{2.5}$ is available at fine spatial resolution (1.2 km^2), each grid cell is mapped to a locality by first determining the geographic centroid (latitude and longitude) of each locality. Then, the locality is assigned the $PM_{2.5}$ value of the grid cell containing this centroid. Mapping by centroid ensures that each locality draws on the $PM_{2.5}$ concentration most representative of its geographic center, which is appropriate given that locality areas are smaller. Using centroid-based assignment rather than area- or population-weighted methods provides a consistent, straightforward way to represent $PM_{2.5}$ at the locality level, since population information is available only once every 10 years. Hence, each locality is assigned the $PM_{2.5}$ value of the grid cell in which its centroid falls.

II City

In this study, a ”city” corresponds to the administrative urban unit recognised as a statutory Urban Local Body (ULB). For the purpose of this analysis, cities therefore represent legally constituted urban jurisdictions, within which multiple wards and localities are present. City boundaries and classifications were taken from the 2011 Census of India.

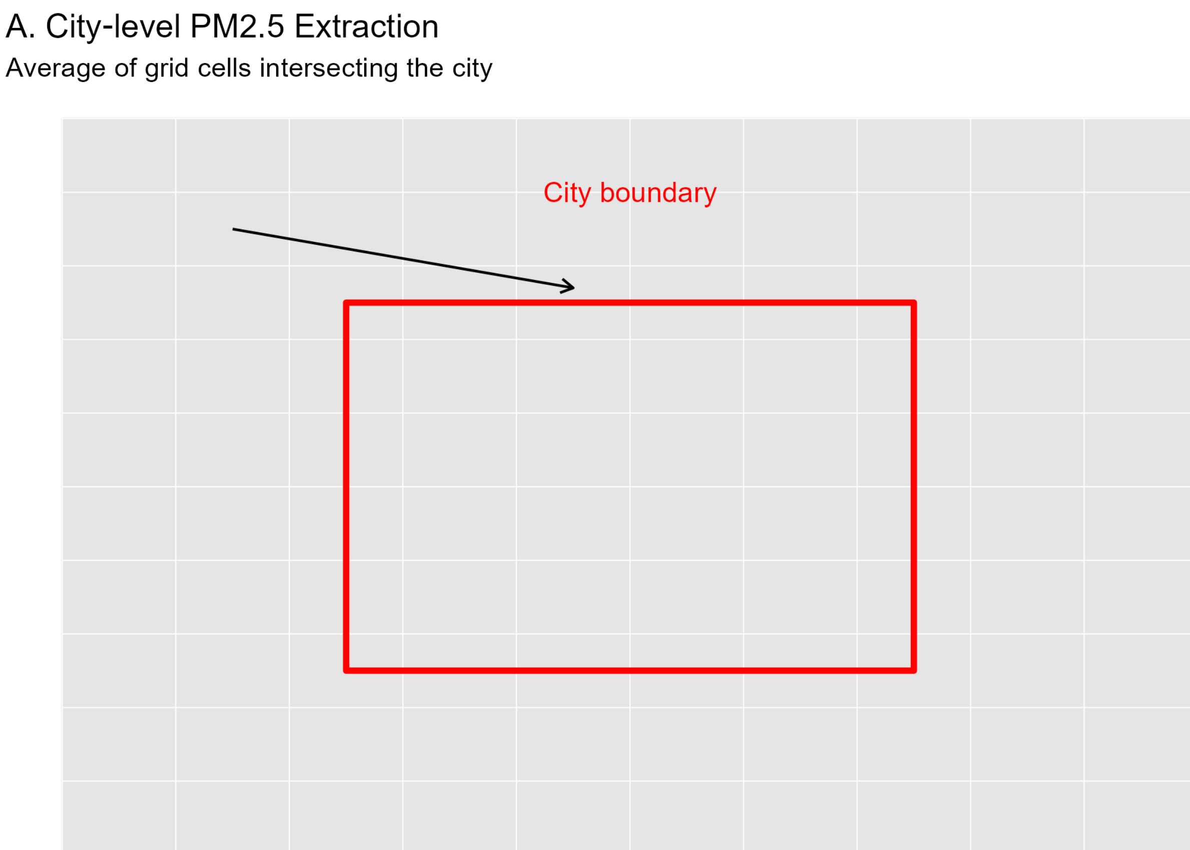


Figure I: City Representation on Raster

City	North	South	East	West
Delhi	28.883°N	28.403°N	77.348°E	76.838°E
Bengaluru	13.139°N	12.834°N	77.793°E	77.356°E
Hyderabad	17.601°N	17.207°N	78.710°E	78.240°E
Chennai	13.261°N	12.893°N	80.332°E	80.079°E
Ahmedabad	23.148°N	22.922°N	72.731°E	72.454°E
Lucknow	27.077°N	26.708°N	81.064°E	80.786°E
Kochi	10.128°N	9.796°N	76.408°E	76.187°E
Nagpur	21.290°N	20.892°N	79.236°E	78.868°E
Noida	28.667°N	28.482°N	77.565°E	77.280°E
Gurgaon	28.560°N	28.304°N	77.152°E	76.881°E
Jaipur	27.042°N	26.771°N	75.915°E	75.650°E

Table I: Bounding Box Coordinates (North, South, East, West) for Treated Cities

III Locality

A locality is defined following UNSESCWA, "a locality is a distinct population cluster, also designated as an inhabited place, populated centre, settlement, and so forth, in which inhabitants reside." In the administrative hierarchy, a locality is below the ward within a city. The structure is therefore:

$$\text{State} \rightarrow \text{District} \rightarrow \text{Sub–district} \rightarrow \text{City} \rightarrow \text{Ward} \rightarrow \text{Locality}$$

Operationally, a locality represents a neighbourhood / colony / block within a ward.

Data Source and Extraction: Locality names are extracted from the 2011 Census of India, which provides city-level demarcations of localities. The locality names were verified using external cartographic platforms (e.g. Maps of India) to ensure consistency and avoid misclassification. Geographic coordinates (lat/long) for each locality are then obtained using Google Maps and ArcGIS India. Because localities are small spatial units, we assign a single coordinate per locality; a precision of $0.01^\circ \times 0.01^\circ$ is sufficient for our analysis.

Integration with Pollution Data: Locality points (lat/long) are spatially mapped onto the ACAG $PM_{2.5}$ raster ($0.01^\circ \times 0.01^\circ \approx 1$ km sq.). The procedure is as follows:

1. Overlay locality coordinates on the pollution grid
2. Assign to each locality the $PM_{2.5}$ value of the cell in which the coordinate falls
3. Produce a point-based pollution estimate per locality-month

Metro Data Conversion and Distance Construction: For each city, station names and station coordinates are extracted from official datasets. The distance between locality i and metro station j is computed using the Haversine formula:

$$D_{ij} = 2R \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_j - \phi_i}{2}\right) + \cos(\phi_i)\cos(\phi_j)\sin^2\left(\frac{\lambda_j - \lambda_i}{2}\right)}\right) \tag{I}$$

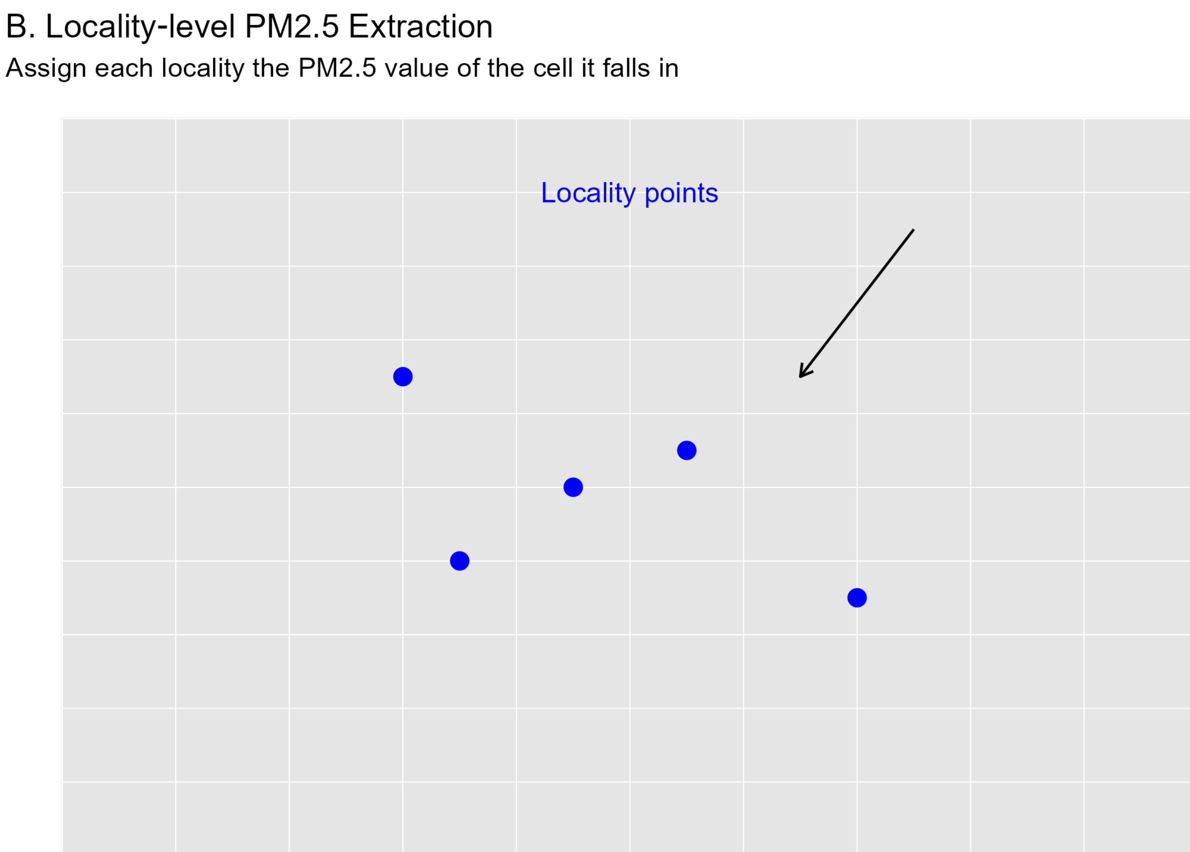


Figure II: Locality Representation on Raster

where D_{ij} is great–circle distance, R is Earth's radius, and ϕ, λ denote latitude and longitude in radians ($1° = \pi/180$).

A distance matrix is constructed, with each row corresponding to a locality and each column to a metro station. For each locality, distances to all stations are computed, and locality-level treatment is defined based on whether the locality falls within a buffer radius (1 km, 2 km, 5 km). A locality is assigned a value of 1 if it falls within the buffer of *any* station, and 0 otherwise. If multiple stations fall within the radius, the closest station's buffer is used for assignment.

IV Metro Treatment and Construction

A metro (mass rapid transit system) is a rail-based urban transport system that operates on dedicated tracks, with no road intersections or surface-level conflicts with other modes of transport. Because

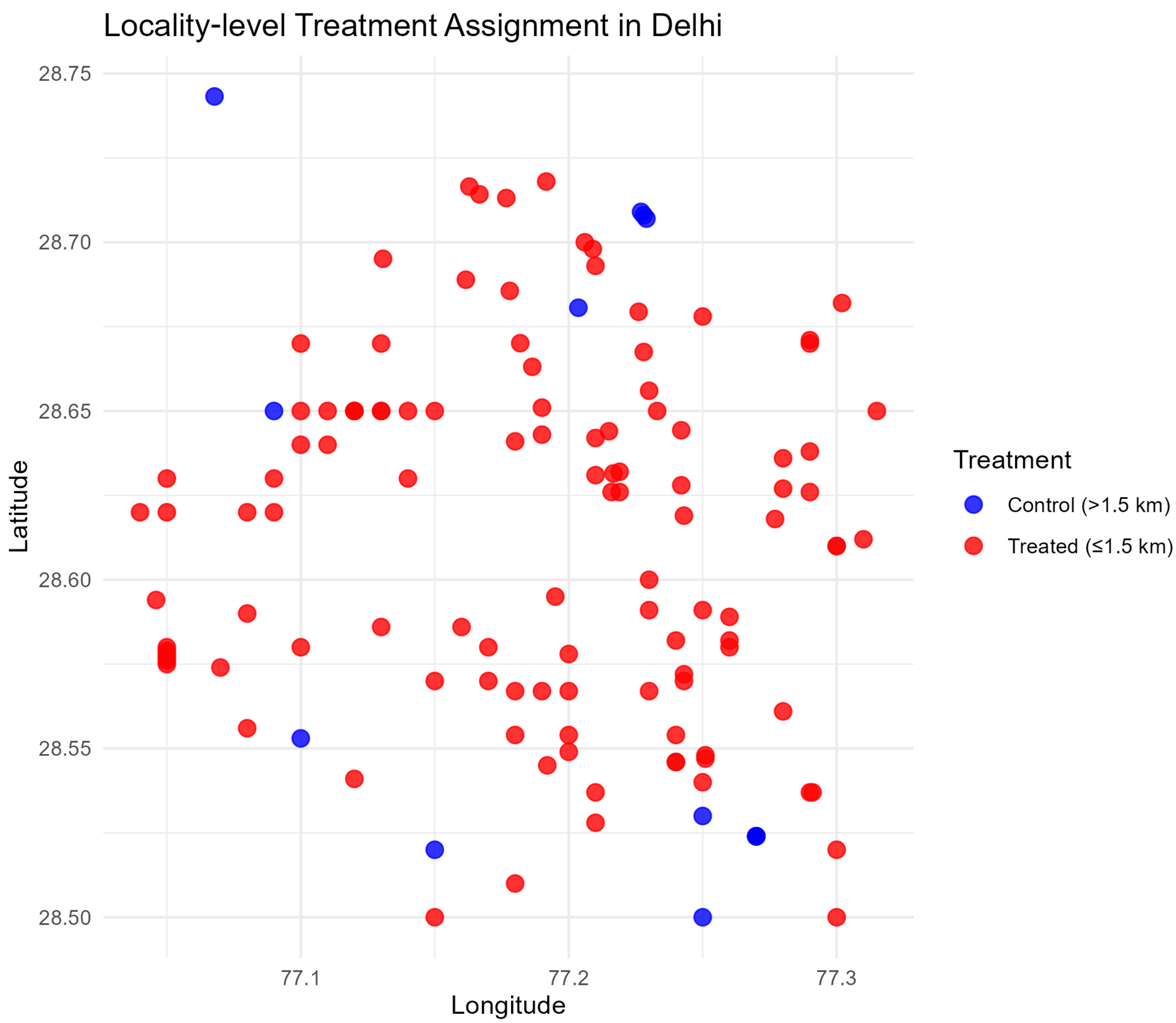


Figure III: Representative image of stations and localities on Raster

of full separation from mixed traffic, metro systems do not contribute to road congestion and are unaffected by traffic conditions.

Table II: Overview of Indian Metro Systems and Key Characteristics

City	Start Year	Coordinates	Website
Delhi	2002	28.6139°N, 77.2090°E	https://delhimetrorail.com/
Bengaluru	2011	12.9716°N, 77.5946°E	https://www.bmrc.co.in/
Hyderabad	2017	17.3850°N, 78.4867°E	https://ltmetro.com/
Chennai	2015	13.0827°N, 80.2707°E	https://chennaimetrorail.org/
Ahmedabad	2019	23.0225°N, 72.5714°E	https://www.gujaratmetrorail.com/ahmedabad/
Lucknow	2017	26.8467°N, 80.9462°E	https://upmetrorail.com/
Kochi	2017	9.9312°N, 76.2673°E	https://kochimetro.org/
Nagpur	2019	21.1458°N, 79.0882°E	https://www.metrorailnagpur.com/
Noida	2019	28.5355°N, 77.3910°E	https://www.nmrcnoida.com/
Gurgaon	2013	28.4595°N, 77.0266°E	https://www.gmrl.org.in/
Jaipur	2015	26.9124°N, 75.7873°E	https://www.jaipurmetro.com/

City	Line Label	Year	Month
Delhi	Red Line (Phase 3 ext)	2019	03
Delhi	Yellow Line ext	2015	11
Delhi	Blue Line ext	2019	03
Delhi	Green Line ext	2018	06
Delhi	Violet Line ext	2014	06
Delhi	Pink Line	2018	03
Delhi	Magenta Line	2017	12
Delhi	Grey Line	2019	10
Delhi	Airport Express ext	2023	09
Bengaluru	Purple Line	2011	10
Bengaluru	Green Line	2014	03
Bengaluru	Yellow Line	2025	08
Hyderabad	Red Line	2017	11
Hyderabad	Green Line	2020	02
Hyderabad	Blue Line	2017	11
Chennai	Green Line	2015	06
Chennai	Blue Line	2016	09
Kochi	Kochi Line 1	2017	06
Lucknow	Red Line	2017	09
Jaipur	Pink Line	2015	06
Nagpur	Orange Line	2019	03
Nagpur	Aqua Line	2020	01
Ahmedabad	Blue Line	2019	03
Ahmedabad	Red Line	2022	09
Noida	Aqua Line	2019	01
Gurgaon	Rapid Metro	2013	11

Table III: Metro Line Treatment Dates by City

In this study, "metro treatment" refers to:

- At the **city-month** level: the month in which commercial metro operations begin in a given city.
- At the **locality-month** level: whether a locality lies within a specified radius of an *operational* station or line begins.

The analysis includes operational metro systems in 11 Indian cities: Delhi (Phase 2 Analysis was conducted around the timeframe needed for this thesis), Ahmedabad, Jaipur, Nagpur, Noida, Lucknow, Hyderabad, Chennai, Bangalore, Kochi, and Gurgaon.

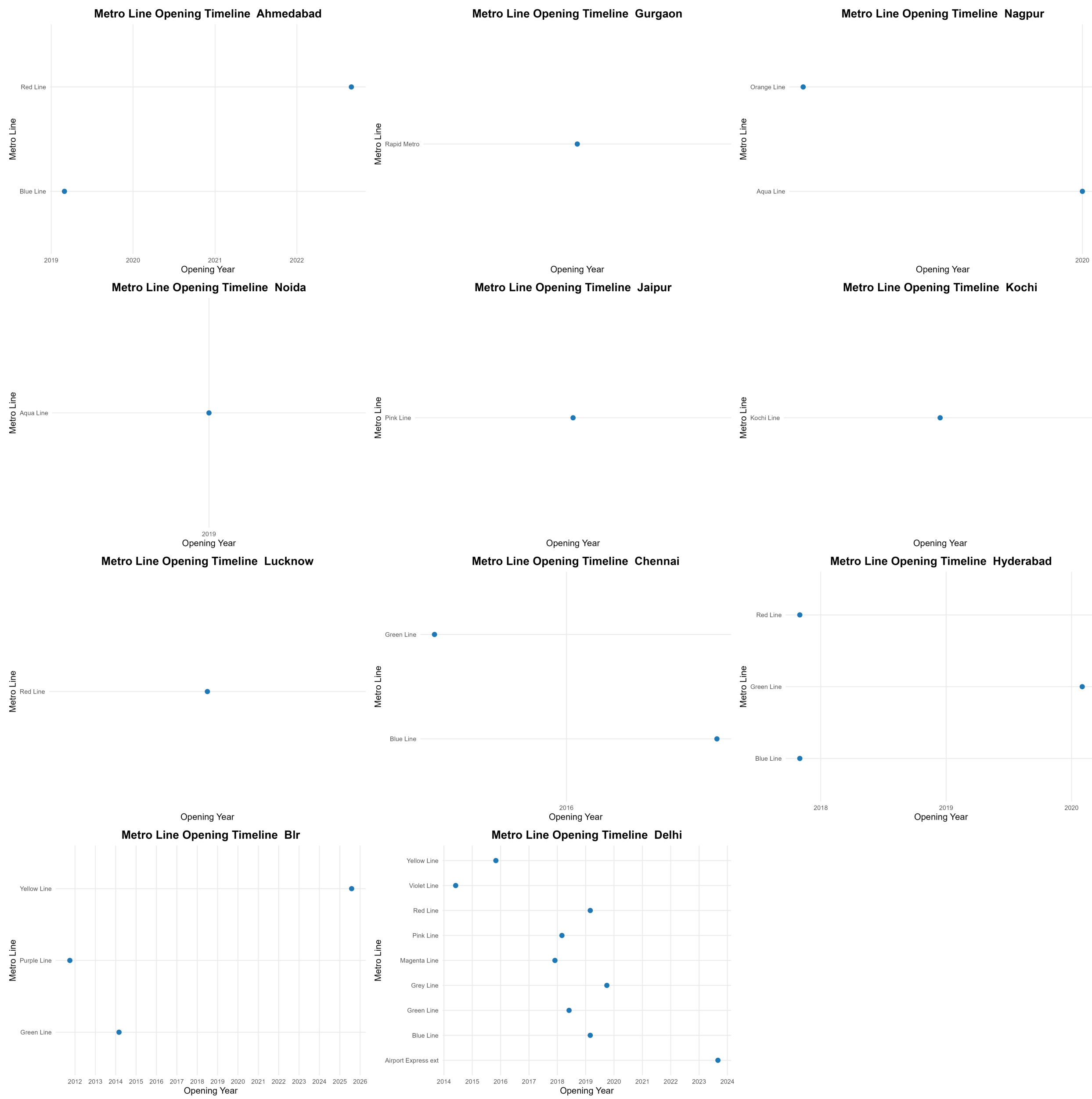


Figure IV: Metro Line Opening Timelines Across Indian Cities

The data is taken from official government metro websites and the Urban Rail Subway website.

V Meteorology and Climate Controls

Weather variables are extracted from the ERA5 reanalysis product (Climate Data Store) using monthly averages. ERA5 meteorological fields are provided at a spatial resolution of $0.25° \times 0.25°$, making it enough for the city scale rather than within individual neighbourhoods. Accordingly, these are treated primarily as city-level controls.

- **10m U-component of Wind:** This measure represents near-surface wind speed at approximately 10 metres above the Earth's surface. Wind influences $PM_{2.5}$ by affecting dispersion and atmospheric mixing: stronger winds promote dispersion, while stagnant conditions are associated with accumulation. Wind conditions also influence local-scale patterns of pollution dispersion.

- **2m Dewpoint Temperature:** This represents the temperature to which air at approximately 2 metres above the Earth's surface would need to be cooled for saturation. It serves as a proxy for near-surface atmospheric moisture (humidity). ERA5 reports dew point in Kelvin (K), which we convert to $^\circ C$ using:

$$^\circ C = K - 273.15.$$

- **2m Temperature:** ERA5 "2–metre temperature" represents near-surface air temperature at approximately 2 metres above the surface. ERA5 reports temperature in Kelvin, which we convert to $^\circ C$ via $^\circ C = K - 273.15$.

- **Surface Pressure:** Surface pressure measures the force exerted by the entire atmospheric column above a location. ERA5 pressure is provided in Pascals (Pa) and varies at the city scale.

- **Total Precipitation:** This variable represents the total depth of liquid and frozen water (rain + snow) that reaches the surface, reported in metres of water equivalent.

- **Total Cloud Cover:** Total cloud cover represents the proportion of the grid cell covered by cloud (0–1).

- **Soil Type:** This variable reflects the soil texture classification used in the land-surface scheme of ERA5, derived from the FAO/UNESCO Digital Soil Map of the World.
- **Low Vegetation Cover:** This measures the fraction of each grid cell covered by "low vegetation" (0–1).
- **High Vegetation Type:** This dimensionless variable indicates the dominant tall/forest vegetation type in each grid cell (as defined by the ERA5 land-surface scheme).

VI Geographic and Local Context Variables

- **Leisure Zones (Locality-level):** To identify whether a given locality contains a leisure zone, we conducted online searches to locate major shopping malls, historical commercial centres, and similar leisure-oriented facilities. Localities with at least one such facility were assigned a value of 1, and those without were assigned a value of 0. This variable is defined only at the locality level.
- **City Business Centre (Locality-level):** To identify whether a locality functions as a business or commercial centre, we conducted online searches to locate areas with concentrations of technology parks, office clusters, and other business activity. Localities with such business concentrations were assigned a value of 1, and those without were assigned a value of 0. Accordingly, this variable is a dummy variable at the locality level.

VII Socio-Economic and Transport Variables

City and Locality GDP: City-level GDP is constructed using official State Domestic Product (SDP) data. For each city in the sample, we compile annual state GDP from State Economic Surveys. City GDP is then derived proportionally by allocating state GDP to cities based on each city's share of total state population, Census 2011 and 2001, thereby calculating the nominal GDP.

Locality-level GDP is constructed by distributing city GDP across localities based on each locality's relative population size. Ward/locality population counts are from the 2011 and 2001

Censuses. Locality GDP is therefore:

$$GDP_j = GDP_{city} \times \left(\frac{Pop_j}{Pop_{city}} \right)$$

where j indexes localities within a city.

Table IV: Economic Survey Data for Treated Cities

City	State / UT	Economic Survey / Report Link
Ahmedabad	Gujarat	https://gujecostat.gujarat.gov.in/
Bengaluru	Karnataka	https://des.karnataka.gov.in/
Chennai	Tamil Nadu	https://www.tn.gov.in/
Delhi	NCT of Delhi	https://delhi.gov.in/
Hyderabad	Telangana	https://ecostat.telangana.gov.in/telangana/Home
Lucknow	Uttar Pradesh	https://updes.up.nic.in/
Jaipur	Rajasthan	https://statistics.rajasthan.gov.in/pages/department-page/647
Kochi	Kerala	https://spb.kerala.gov.in/
Noida	Uttar Pradesh	https://updes.up.nic.in/
Nagpur	Maharashtra	https://mahades.maharashtra.gov.in/home.do?lang=mr
Gurgaon	Haryana	https://gurugram.gov.in/

Population: The population is taken from Census documents (2011, 2001) for each city, as the analysis is conducted between 2008 and 2023. Hence, for GDP calculations prior to 2011, the 2001 census is used; for post-2011 calculations, the 2011 census is used.

Property Prices: To proxy local property values, we collected rent listings from online property platforms (e.g. MagicBricks.com). Based on the distribution of advertised rents, localities were classified into high, medium, and low-income housing zones. This variable reflects socio-economic stratification across neighbourhoods.

Tier: State tier is determined from census reports and Maps of India. Two dummy variables are created to indicate Tier 1 and Tier 2 city status. The information for this was extracted from the Government of India's cities classification. Hence, the table below classifies treated cities as Tier 1 or Tier 2. Tier 1 has a population of 100,000 and above, and Tier 2: 50,000 to 99,999.

Further, the tiers differ in the number of development schemes they have. In general, tier 1 has better infrastructure than tier 2.

City	**Tier**
Gurgaon	Tier 2
Kochi	Tier 2
Chennai	Tier 1
Bengaluru	Tier 1
Hyderabad	Tier 1
Nagpur	Tier 2
Ahmedabad	Tier 1
Lucknow	Tier 2
Jaipur	Tier 2
Delhi	Tier 1
Noida	Tier 2

Table V: Tier Classification of Treated Cities

Region: The region variable classifies each city into one of the North / South / East / West zones based on geographic location. This is entered as a city-level dummy variable.

City	**Region**
Gurgaon	North
Kochi	South
Chennai	South
Bengaluru	South
Hyderabad	South
Nagpur	West
Ahmedabad	West
Lucknow	North
Jaipur	North
Delhi	North
Noida	North

Table VI: Geographical Region Classification of Treated Cities

Bus Connectivity: This variable captures the extent of surface public transport availability within a locality (e.g. presence of bus stops or route coverage). It proxies non-metro transit supply.

Ridership: Ridership measures utilisation of public transport services in a specific metro, reflecting demand-side mobility behaviour. Data is sourced from official reports and online archival searches such as X.

Fare Change: This variable captures changes in public transport fare levels over time. The variation in the fare influences the choice and, therefore, the responses of the proxies between buses, the metro, and other modes. Data is drawn from newspaper reports and official circulars. We record this as a dummy variable: 1 if it has happened, 0 otherwise.

Petroleum and Diesel Prices: Fuel price levels are included to control for the cost of private vehicle use. Data is extracted from the Ministry of Petroleum and Natural Gas / Petroleum Planning & Analysis Cell (PPAC) retail fuel price bulletins.

Chapter 4

Empirical Strategy

I Empirical Strategy Overview

To estimate the causal impact of metro construction on air pollution, two balanced panels are constructed: (i) a City-Month panel and (ii) a Locality-Month panel (locality j within city i). Because the opening years differ across cities, treatment timing is staggered. The analysis covers 2008–2023 and consists of 11 treated metro cities and 121 control cities. There were no underlying criteria for the control cities other than the fact that they did not have a metro station.

Table I: Panel Dimensions

	Units	Time Period	Total Observations
City-Month Panel	11 metro cities	2008-2023	23,424
Locality-Month Panel	11,712 localities	2008-2023	148,199

Notes: The city panel aggregates at the metro city level. The locality panel assigns $PM_{2.5}$ to 11,712 locality centroids within the 11 metro cities.

II Identification

Identification comes from comparing treated and untreated units before and after the metro opening. The key identifying assumption is conditional parallel trends: in the absence of metros, treated and untreated units would have followed similar pollution trends. City / locality fixed effects absorb time-invariant geography; month fixed effects absorb national shocks and seasonality.

III Estimator

The standard TWFE DiD is inappropriate in heterogeneous adoption timing. Therefore, we apply the Callaway & Sant'Anna (2021) methodology, which constructs group-time ATTs using not-yet-treated units as controls.

Formally:

$$ATT(g,t) = \mathbb{E}\big[Y_t(g) - Y_t(\infty) \mid G = g\big], \tag{I}$$

where g indexes treatment cohort (first metro opening month) and ∞ denotes not-yet-treated.

IV Treatment Definition

City-month: A city is treated from the first month the metro becomes operational.

Locality-month: A locality j in city i is treated if its centroid lies within a radius r km of any operational station:

$$E_{ijt}^{(r)} \in \{0,1\}.$$

V Controls, Fixed Effects, and Inference

Let X denote the observed controls, including:

- Meteorological (ERA5): temperature, dewpoint, wind, pressure, precipitation, cloud
- Surface characteristics: soil type, vegetation cover
- Socio-economic: GDP, population, property prices, tier, region
- Transport: bus connectivity, ridership, fare changes, fuel prices

All models include:

- Spatial FE: α_i (city) or μ_{ij} (City, locality level characteristics)
- Time FE: λ_t (month)

Standard errors are clustered at the city level in city-month regressions and at the locality level in locality-month regressions.

City-Level Analysis (ATT Estimator)

The city-level Difference-in-Differences (DiD) equation, estimated using the **Callaway** & **Sant′Anna** (**2021**) methodology to account for staggered treatment timing, is specified as:

$$PM_{it} = \alpha_i + \lambda_t + X_{it}'\beta + ATT_C(g(i), t) + \varepsilon_{it}$$

Where:

- PM_{it}: The dependent variable, representing the log of the monthly mean $\mathbf{PM_{2.5}}$ concentration in city i at time t.
- α_i: A set of **City – level Fixed Effects** (FE), controlling for all unobserved, time-invariant characteristics specific to city i .
- λ_t: A set of **Month – level Fixed Effects** (FE), controlling for all common, time-varying shocks across all cities .

- $X_{it}'\beta$: A vector of **Time − varying control variables** (e.g., meteorological, socio-economic, and transport controls) in city i at time t.

- $ATT_C(g(i), t)$: The **Group − Time Average Treatment Effect on the Treated** (ATT) for the city cohort $g(i)$ at time t, which is the primary parameter of interest capturing the causal impact of the metro opening on $PM_{2.5}$.

- ε_{it}: The idiosyncratic error term. Standard errors are clustered at the city level.

Locality-Level Baseline Specification (Without Interaction Term)

Before incorporating spatial exposure heterogeneity, we first estimate a baseline locality-level specification that parallels the city-level model, but operates at a finer spatial scale. The equation is:

$$PM_{ijt} = \mu_{ij} + \lambda_t + X_{ijt}'\theta + ATT_L(g(i), t) + \varepsilon_{ijt} \tag{II}$$

Where:

- PM_{ijt}: Log monthly mean $PM_{2.5}$ concentration in locality j of city i at time t.

- μ_{ij}: **Locality fixed effects**.

- λ_t: **Month fixed effects**.

- $X_{ijt}'\theta$: A vector of time-varying controls at the locality level (meteorological, socio-economic, and transport characteristics).

- $ATT_L(g(i), t)$: **Group-time ATT at the locality scale**. Since each locality inherits its city's cohort $g(i)$, this identifies the average effect of metro opening on pollution for localities in that cohort at time t.

- ε_{ijt}: Error term. Standard errors are clustered at the locality level.

This specification identifies the average metro effect across all localities within a treated city cohort, without yet distinguishing between localities that are close to metro stations, and those that are not.

Locality-Level (Spatial Heterogeneity) Analysis

The locality-level equation is used to investigate **spatial heterogeneity** in the impact near metro stations. It is specified as:

$$PM_{ijt} = \mu_{ij} + \lambda_t + X'_{ijt}\delta + \left[ATT_L(g(i), t) \times E(r)_{ijt}\right] + \epsilon_{ijt}$$

Where:

- PM_{ijt}: The dependent variable, representing the log of the monthly mean $\mathbf{PM_{2.5}}$ concentration in **locality j** within city i at time t.
- μ_{ij}: A set of **Locality – level Fixed Effects** (FE), controlling for all unobserved, time-invariant characteristics specific to locality j .
- λ_t: A set of **Month – level Fixed Effects** (FE), controlling for common time shocks across all cities.
- $X'_{ijt}\delta$: A vector of **time – varying control variables** specific to locality j at time t.
- $\mathbf{ATT_L(g(i), t) \times E(r)_{ijt}}$: The **Interaction Term** which estimates the localized treatment effect. ATT_L is the city-cohort ATT (or a similar treatment indicator), and $E(r)_{ijt}$ is an **Exposure Measure** (e.g., an indicator variable or distance decay function) that equals 1 if locality j is within the specified treatment radius r of a metro station, and 0 otherwise. The coefficient on this term captures the difference in the pollution impact between exposed and unexposed localities.
- ϵ_{ijt}: The idiosyncratic **error term**. Standard errors are clustered at the locality level.

VI Robustness and Heterogeneity Tests

We test robustness using alternative radii r, excluding COVID, early-adopter vs. later-adopter tests, fuel and ridership impacts, and Tier 1 vs. Tier 2 Impacts. And look at the city-level impacts on line construction. The explanation of these will be mentioned ahead.

Chapter 5

Results

I Main ATT Estimates

Table I reports the Average Treatment Effects on the Treated (ATT) based on the Callaway and Sant'Anna (2021) DiD estimator. The results for both the locality-level and city-level panels are presented across the following aggregation schemes: simple (overall ATT), dynamic/event-study, group/cohort, and calendar-time. This analysis is undertaken using the non-treated groups, as the metro stations in the treated cities have staggered starts. Staggered, in this context, refers to cities that have never received a metro station or are yet to receive one. Essentially, making the treated cities control themselves. The analysis is done as follows:

- **Simple ATT**: the average treatment effect across all treated units and post–treatment months; this summarises whether $PM_{2.5}$ levels are lower on average after metros begin operation.
- **Dynamic (Event-Study) ATT**: the evolution of the treatment effect relative to the metro opening month; increasingly negative post-opening coefficients indicate that pollution reductions accumulate as ridership builds over time.
- **Group / Cohort ATT**: the average effect by treatment cohort (metro opening year); larger effects for earlier cohorts imply that cities with longer post-operational exposure achieve greater $PM_{2.5}$ declines.

- **Calendar-Time ATT**: the effect aggregated by actual calendar month/year; this shows whether metro impacts remain after accounting for common national pollution shocks in a given month.

Table I: Main ATT Estimates

Locality	ATT	Std. Error	95% CI
Simple	0.8587	0.3760	[0.1218 ; 1.5957] *
Dynamic (event-study aggregation)	1.0729	0.4174	[0.2548 ; 1.8911] *
Group / Cohort aggregation	0.7301	0.3635	[0.0178 ; 1.4425] *
Calendar time aggregation	1.0920	0.3712	[0.3645 ; 1.8195] *
City	ATT	Std. Error	95% CI
Simple	-3.1788	5.1055	[-13.1854 ; 6.8278]
Dynamic (event-study aggregation)	-6.9719	2.8721	[-12.6010 ; -1.3428] *
Group / Cohort aggregation	1.0209	1.9233	[-2.7487 ; 4.7906]
Calendar time aggregation	-5.9560	2.8607	[-11.5628 ; -0.3492] *
City Controls: Weather and Economic Indicators	ATT	Std. Error	95% CI
Simple	-2.5985	1.0445	[-4.6456 ; -0.5513] *
Dynamic (event-study aggregation)	-2.7203	0.8160	[-4.3197 ; -1.1208] *
Group / Cohort aggregation	-1.7310	0.3376	[-2.3927 ; -1.0694] *
Calendar time aggregation	—	—	—
City Controls: Weather only	ATT	Std. Error	95% CI
Simple	5.4593	8.0695	[-10.3566 ; 21.2752]
Dynamic (event-study aggregation)	2.6273	3.7930	[-4.8068 ; 10.0614]
Group / Cohort aggregation	10.3743	9.1423	[-7.5443 ; 28.2930]
Calendar time aggregation	—	—	—

II Interpretation of Main Effects

Overall, the treatment effect magnitudes indicate that metro operationalisation is associated with mixed changes in $PM_{2.5}$. At the local level, estimates suggest increased pollution, whereas at the city level, dynamic aggregations indicate reductions.

For the locality scale, the simple ATT estimate of 0.8587 (SE = 0.376) implies that, on average, localities experience an increase of approximately 0.86 μg/m^3 in $PM_{2.5}$ after metro operationalisation. This is statistically significant at the 5% level, indicating a worsening local effect. The dynamic (event-study aggregated) locality ATT of 1.0729 (SE = 0.4174) is larger in magnitude and statistically significant. This suggests that the pollution-increasing effect worsens over time. This is the opposite of the mechanism by which ridership typically builds gradually after opening, thereby leading to an overall improvement. The locality group/cohort–pooled ATT is 0.7301 (SE = 0.3635) and remains worsened and significant. This implies that across different metro opening cohorts in India, there is a consistent average increase in $PM_{2.5}$ at the locality scale. The fact that this holds across staggered adoption timing reinforces the robustness of the result. Similarly, the calendar–time pooled locality ATT of 1.0920 (SE = 0.3712) is statistically significant, again indicating a positive post–metro effect in treated localities. In conclusion, at the local scale, all aggregation methods indicate a statistically significant increase in $PM_{2.5}$ following metro operationalisation, suggesting that the metro has not achieved desirable outcomes at the local level.

At the city scale, effects are noisier due to aggregation across entire urban areas. The simple city ATT of –3.1788 (SE = 5.1055) is negative but not statistically significant. The dynamic city ATT of –6.9719 (SE = 2.8721) is negative and statistically significant. This indicates that when treatment timing and dynamic evolution are accounted for, metros are associated with reductions of approximately 7 μg/m^3 in city-level $PM_{2.5}$ over time. The group/cohort city ATT of 1.0209 (SE = 1.9233) is not statistically significant. Finally, the calendar–time city ATT of –5.9560 (SE = 2.8607) is negative and statistically significant, indicating city-wide reductions of around 6 μg/m^3.

However, when accounting for weather and economic indicators, we observe a decline of around 3 μg/m^3 in $PM_{2.5}$ (SE = 0.8610) at the city level.

Taken together, these results show that metro expansions have mixed effects: at the local scale, $PM_{2.5}$ rises, whereas at the city scale, dynamic aggregations suggest reductions. This pattern is consistent with the theory that localised construction-adjacent effects, combined with broader, longer-run network-level gains, and the idea that last-mile connectivity relies on additional transportation to reach the metro at the locality level. This matches existing research by Li et al. (2016). Lin et al. (2023) found that following the opening of a metro station, pollution declined, whereas Hsu et al. (2024) reported that metro construction increased local pollution.

III Event–Study Dynamics and Parallel Trends

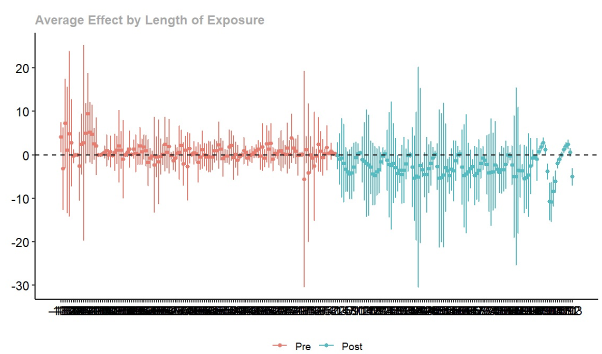


Figure I: Locality Level

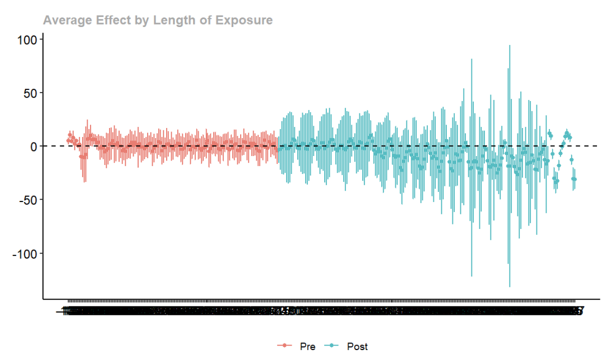


Figure II: City Level

The X-axis represents time, though it is not clearly labeled, and the Y-axis represents PM 2.5 levels. The event-study graph is noisy because the treatment timing is staggered across 11 metro-opening cohorts, and the sample is unbalanced prior to treatment. In Callaway Sant'Anna estimators, the pre-trend lines are expected to show sampling variability when groups enter treatment at different times and have few pre-period observations. Each point represents a relative-time treatment effect, showing $PM_{2.5}$ before and after the metro was opened. All pre-treatment (leads) remain close to zero and are statistically insignificant. Hence, through visual confirmation, we confirm that treated cities were not already trending downward relative to the not-yet-treated cities prior to metro opening.

After the opening month, the post-treatment coefficients (lags) begin to go negative as the exposure length increases. This suggests that metro rail systems generate air quality benefits that gradually increase over time, as commuters do not fully switch transport modes immediately, but rather over time.

Finally, the widening confidence intervals at longer post-periods are a feature of staggered adoption. At longer exposure lengths, only the earliest metro systems contribute to identification (e.g., Delhi and Bangalore), while more recent metros contribute only to short post periods. The sign, shape, and direction of the central estimates remain consistent with the baseline aggregated effects.

At the local level, we observe a small decline; however, an increase is also plausible due to last-mile connectivity, which may require additional transportation and contribute to higher local pollution.

Taken together, the dynamic event–study evidence strengthens the causal interpretation: metros reduce $PM_{2.5}$ on average at the city level but increase it at the local level.

IV Heterogeneity Test: Late Vs Early Adopters

Table II: Diagnostic robustness checks: exclusion of early vs late metros

City excluded	ATT	Std. Error	95% CI
Delhi	0.2794	0.8180	[-1.3239, 1.8827]
Bangalore	0.2914	0.8715	[-1.4167, 1.9996]
Lucknow	-6.9719	2.9722	[-12.7972, -1.1466]*

Notes: Table reports overall ATT estimates from event-study/dynamic aggregation when excluding one metro city at a time. * indicates statistical significance at conventional levels.

These exclusion checks show that the long-run pollution reduction is driven mainly by the early-entering metro systems. To test this, the two earliest opened cities, Delhi and Bangalore, are removed. When these are excluded, the overall ATT becomes statistically insignificant, indicating that these early metros are responsible for the observed impact. In contrast, excluding Lucknow results in a statistically significant negative ATT. This demonstrates that early adopters are more likely to contribute to a decrease in pollution levels over time compared to cities where metros were opened later.

V COVID effects

We test whether COVID could have affected the overall decline during the country's lockdown. Thereby, removing the COVID lockdown months March to May 2020, we re-run the tests.

Table III: Overall ATT (event–study / dynamic aggregation)

	ATT	Std. Error	95% Confidence Interval
Overall ATT (dynamic aggregation)	-7.6338	2.8096	[-13.1405 , -2.1271]*

Notes: * indicates statistical significance at conventional levels. ATT estimated using Callaway & Sant'Anna (2021) dynamic aggregation. Units in $\mu g/m^3$ of $PM_{2.5}$.

The overall ATT estimate using the dynamic aggregation approach is negative and statistically significant, indicating that $PM_{2.5}$ concentrations declined in metro–treated localities relative to not–yet–treated areas. This indicates that the decline in overall air pollution is not as affected by COVID.

VI Tier1 vs Tier 2 Cities

We conduct a test to determine whether there is a difference between Tier 1 and Tier 2 cities.

Table IV: Tier-level ATT (event–study / dynamic aggregation)

	ATT	Std. Error	95% Confidence Interval
Tier-level ATT estimate	-3.7158	0.7631	[-5.2115 , -2.2201]*

Notes: * indicates confidence interval does not cover zero. ATT estimated using Callaway & Sant'Anna (2021) using "Not Yet Treated" as the comparison group and doubly-robust estimation. Units in $\mu g/m^3$ of $PM_{2.5}$.

The Tier-level ATT is negative and statistically significant, with a decline of 4 micrograms per cubic metre, indicating that $PM_{2.5}$ concentrations declined in these cities. This indicates that metro has led to better environmental outcomes in tier 1 cities than in tier 2 cities. This aligns with the idea that tier 1 cities in general have better accessibility and infrastructure, thereby making access to a metro station easier. This also builds on the idea of Beaudoin et al. (2020) that metro construction alone is insufficient; rather, infrastructure needs to be improved. Hence, this analysis aligns with the argument that a metro structure leads to better development.

VII Multiple Line Openings at the City Level.

In the Indian context, metro expansion is staggered, with multiple lines coming online at different times. Therefore, it is essential to determine whether the construction leads to a decline. Each line opening represents a clearly defined, date-specific intervention as mentioned in the City section of our analysis. For each line ℓ, I construct a city-month panel, which is restricted to the cities served by that line and thereby estimated. Given that Callaway-Sant'Ana focuses on staggered treatment under the assumption that once treated, a city is always expected to be treated, when considering a regression for multiple city lines, this kind of line expansion analysis violates the Callaway-Sant'Ana. For the purpose of this analysis, a Two-Way Fixed Effects mechanism is used across the cities.

$$PM_{it} = \alpha_i + \lambda_t + \beta_\ell \text{LineOpen}_{it}^{(\ell)} + X_{it}'\gamma + \varepsilon_{it},$$

where:

- α_i are city fixed effects,
- λ_t are month-year fixed effects,
- $\text{LineOpen}_{it}^{(\ell)}$ equals 1 after line ℓ opens in its corresponding city,
- X_{it} includes time-varying controls such as precipitation, GDP, temperature, and wind speed.

Because each regression uses only one treatment event (one metro line at a time), multiple-treatment issues are avoided.

Table V: Average ATT Estimates by City (Line-Level TWFE)

City	**Average ATT**
Bengaluru	-1.46
Ahmedabad	-1.37
Gurgaon	-0.49
Jaipur	-0.02
Chennai	-0.02
Kochi	0.67
Lucknow	0.74
Hyderabad	0.79
Delhi	0.82
Noida	1.05
Nagpur	1.09

ATTs are based on TWFE line-level estimations and show substantial heterogeneity in pollution responses across cities. Averaging the event-time coefficients for each city shows:

- There exists a large negative average effect in Bangalore and Ahmedabad,
- There exists a reduction in Gurgaon,
- There exists near-zero effects in Jaipur and Chennai,
- There exists positive average effects in Delhi, Hyderabad, Noida, and Nagpur, suggesting short-term increases potentially related to construction activity.

This also bridges the explanation for the increase at the local level and the decline at the city level. Line increases have a mixed effect across cities, and this impact can be specifically noticed at the local level when we look at the most micro scale, but at the city level as a whole, because we assume that the first time a city has had a metro, it's now considered treated, and thereby we see a general decline in the PM2.5.

I Spatial Checks

I Spatial Autocorrelation

Air pollution is spatially correlated: pollution in one locality may affect its neighbors. Following the work of Chen et al.(2012), I thereby try looking at locality pollution at a lesser distance than 2km, this is because last mile connectivity is a major challenge in the Indian context and thereby, closer to the metro station leads to an increase in the usage. Therefore, it is important to test for spatial clustering using Moran's I statistic:

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}, \tag{I}$$

where w_{ij} is a spatial weight (e.g., 1 if localities i and j are neighbors and 0 otherwise), x_i is the estimated effect in locality i, and N is the number of localities. A significant positive I indicates spatial clustering of high (or low) pollution effects.

The spatial weights matrix W is constructed using a k-nearest-neighbor approach with $k = 3$, chosen because localities in Indian cities are geographically small; using more neighbors would excessively smooth spatial relations, while fewer neighbors would understate local dependence. To validate this choice, I used ArcGIS to visualise locality centroids and applied the Getis–Ord Gi* statistic and Moran's I hot-spot tool. These diagnostics confirmed the presence of local clustering within approximately 1–2 km. I then computed the formal Global Moran's I in R to test for spatial autocorrelation in the regression residuals.

II Spatial Regression Models

To explicitly capture spatial spillovers, I estimate two spatial econometric models.

(1) Spatial Autoregressive (SAR) Model

$$Y = \rho WY + X\beta + \epsilon, \tag{II}$$

where Y is the vector of locality-level $PM_{2.5}$ outcomes, W is the spatial weight matrix, and ρ measures spatial dependence in the outcome.

(2) Spatial Error Model (SEM)

$$Y = X\beta + u, \qquad u = \lambda Wu + \epsilon, \tag{III}$$

where λ captures spatial correlation in the unobserved error component.

I first implemented exploratory spatial tests for a single city in ArcGIS using its spatial autocorrelation functions, and then re-estimated the full models in `R` using the `spatialreg` package. These models are estimated for the post-metro period to test whether spatial error correlation alters inference.

Table VI: Spatial Error Model (SEM) Estimates

Variable	Estimate	Std. Error	p-value
Post-Metro	5.85	2.48	0.018**
GDP (local)	0.0012	0.0018	0.493
Precipitation	-0.0163	0.036	0.651
Temperature	-0.1292	0.255	0.613
Spatial Error Parameter (λ)	-0.0113	0.1198	0.925
Fixed Effects Included:			
Locality Fixed Effects	Yes	–	–
Month Fixed Effects	Yes	–	–

Interpretation. The *Post-Metro* coefficient remains positive and statistically significant after accounting for spatial dependence, indicating robust station-area effects within 1 km. The insignificant λ parameter suggests that once locality fixed effects and observables are controlled for, there is limited remaining spatial autocorrelation in the error term. This confirms that the locality fixed effects absorb most of the spatial correlation in unobserved factors, so spatial spillovers do not bias the estimated metro effect.

Table VII: Distance-band local effects (locality panel)

Distance to Metro	ATT ($\mu g/m^3$)	Std. Error	95% CI	Sig.
$\leq$ 1 km	35.78	10.68	[14.85, 56.71]	***
$\leq$ 2 km	4.43	10.24	[-15.65, 24.51]	n.s.

Locality-level spatial heterogeneity tests show that metro impacts are highly localized. Localities within 1 km of an operational metro station experience a statistically significant increase in $PM_{2.5}$ concentrations of approximately 35.8 μg/m^3. Beyond 2 km, the estimated effect is small and statistically insignificant. These findings are consistent with the literature, which shows that major infrastructure projects generate short-run, highly localized construction and congestion effects.

Chapter 6

Limitations

This study has three main limitations. First, measurement error remains possible in some covariates. While $PM_{2.5}$ is measured using high-resolution satellite raster images. Second, although Callaway & Sant'Anna (2021) address the biases of two-way fixed effects under staggered adoption, identification still relies on the conditional parallel trends assumption. Unobserved city-level policies (e.g. local clean-air programmes) that coincide with metro expansion but are imperfectly controlled for may still bias estimates. Third, the analysis focuses only on $PM_{2.5}$; pollution is multi-pollutant, and metro effects on NO_2, VOCs, and congestion externalities cannot be assessed here. Thus, results should be interpreted within the $PM_{2.5}$, not as a full environmental welfare evaluation.

Chapter 7

Policy Discussion

The results suggest that metro expansion generates statistically significant reductions in air pollution, but that these benefits are geographically concentrated and relatively modest in aggregate. In policy terms, metros appear to be a *necessary but not sufficient* intervention. The effects are largest very close to stations, suggesting that first- and last-mile connectivity, as well as mode substitution margins, need to be considered. If commuters cannot easily transition from private vehicles to rapid transit, environmental gains will remain small. Therefore, complementary policies are important: integrated fare systems, bus–metro coordination, bicycle/scooter access feeds, parking pricing around CBDs, and targeted petrol/diesel taxation could improve the environmental return of metro capital expenditure. Further, Individuals can consider EV buses and two-wheelers to assess how this helps over time.

Chapter 8

Conclusion

This thesis provides causal evidence on the air quality impact of metro expansion in India using a staggered Difference-in-Differences framework based on Callaway & Sant'Anna (2021). Satellite-derived $PM_{2.5}$ measurements at 1.1 km resolution allow pollution to be assigned at the local level, enabling fine-grained estimation. The results indicate that metros reduce $PM_{2.5}$ concentrations, with effects emerging gradually after opening and being strongest in locations within 1 km of stations. No substantial pretreatment divergence is detected, supporting the credibility of the parallel trends assumption. While the magnitudes are moderate relative to India's high baseline pollution levels, the findings show that urban rail investment helps improve environmental benefits in the long run.

Bibliography

[1] Anderson, M. L. (2014). Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion. *American Economic Review*, 104(9), 2763–2796.

[2] Arceo, E., Hanna, R., & Oliva, P. (2016). Does the effect of pollution on infant mortality differ between developing and developed countries? *Economic Journal*, 126(591), 257–280.

[3] Auffhammer, M., & Kellogg, R. (2011). Clearing the air? The effects of gasoline content regulation on air quality. *American Economic Review*, 101(6), 2687–2722.

[4] Batarce, M., & Galilea, P. (2018). Cost and fare estimation for the bus transit system of Santiago. *Transport Policy*, 64, 92–101.

[5] Baum-Snow, N. (2007). Did highways cause suburbanization? *Quarterly Journal of Economics*, 122(2), 775–805.

[6] Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., & Zhang, Q. (2017). Roads, railroads, and decentralization of Chinese cities. *Review of Economics and Statistics*, 99(3), 435–448.

[7] Baum-Snow, N., & Kahn, M. (2005). Effects of urban rail transit expansions: Evidence from sixteen cities, 1970–2000. *Brookings-Wharton Papers on Urban Affairs*, 147–206.

[8] Chang, T., Graff Zivin, J., Gross, T., & Neidell, M. (2016). Particulate pollution and the productivity of pear packers. *AEJ: Economic Policy*, 8(3), 141–169.

[9] Chay, K., & Greenstone, M. (2003). The impact of air pollution on infant mortality. *Quarterly Journal of Economics*, 118(3), 1121–1167.

[10] Chen, W., & Klaiber, H. (2020). Does road expansion induce traffic? *Journal of Environmental Economics and Management*, 104, 102387.

[11] Chen, Y., & Whalley, A. (2012). Green infrastructure: The effects of urban rail transit on air quality. *AEJ: Economic Policy*, 4(1), 58–97.

[12] Cervero, R. (2002). Induced travel demand: Research design, empirical evidence, and normative policies. *Journal of Planning Literature*, 17(1), 3–20.

[13] Currie, J., & Neidell, M. (2005). Air pollution and infant health: What can we learn from California's recent experience? *Quarterly Journal of Economics*, 120(3), 1003–1030.

[14] Currie, J., & Walker, R. (2011). Traffic congestion and infant health: Evidence from E-ZPass. *AEJ: Applied Economics*, 3(1), 65–90.

[15] Da Silva, D., Klumpenhouwer, W., Karner, A., Robinson, M., Liu, R., & Shalaby, A. (2022). Living on a fare: Modeling and quantifying the effects of fare budgets on transit access and equity. *Journal of Transport Geography*, 101, 103348.

[16] Davis, L. W. (2008). The effect of driving restrictions on air quality in Mexico City. *Journal of Political Economy*, 116(1), 38–81.

[17] Desmet, K., & Rossi-Hansberg, E. (2022a). Urban transportation and local growth. *Journal of Urban Economics*, 131, 103419.

[18] Desmet, K., & Rossi-Hansberg, E. (2022b). Urban transportation and spatial growth. *Review of Economic Studies*, 89(3), 1385–1419.

[19] Duranton, G., & Turner, M. A. (2011). The fundamental law of road congestion: Evidence from US cities. *American Economic Review*, 101(6), 2616–2652.

[20] Tsai, D.-H., Wu, Y.-S., Chan, C.-C., & Chen, B.-Y. (2008). Commuters' exposure to particulate matter by transport mode. *Science of the Total Environment*, 405(1–3), 72–77.

[21] Vickrey, W. S. (1969). Congestion theory and transport investment. *American Economic Review*, 59(2), 251–260.

[22] Van Donkelaar, A., Hammer, M. S., Bindle, L., Brauer, M., Brook, J. R., Garay, M. J., Hsu, N. C., Kalashnikova, O. V., Kahn, R. A., Lee, C., Levy, R. C., Lyapustin, A., Sayer, A. M., & Martin, R. V. (2021). Monthly Global Estimates of Fine Particulate Matter and Their Uncertainty. *Environmental Science & Technology*. `https://doi.org/10.1021/acs.est.1c05309`

[23] Hammer, M. S., Van Donkelaar, A., Bindle, L., Sayer, A. M., Lee, J., Hsu, N. C., Levy, R. C., Sawyer, V., Garay, M. J., Kalashnikova, O. V., Kahn, R. A., Lyapustin, A., & Martin, R. V. (2023). Assessment of the Impact of Discontinuity in Satellite Instruments and Retrievals on Global PM2.5 Estimates. *Remote Sensing of Environment*, 294, 113624. `https://doi.org/10.1016/j.rse.2023.113624`

[24] Islam, Z., & Hossain, M. A. (2024). Assessing the environmental pollution and risk of Metro rail construction in Dhaka City. In *Spatial Modeling of Environmental Pollution and Ecological Risk* (pp. 121–141). Woodhead Publishing.

[25] Paschalidou, A. K., Kassomenos, P. A., & Kelessis, A. (2016). Tracking the association between metro-railway construction works and PM levels in an urban Mediterranean environment. *Science of The Total Environment*, 568, 1326–1332.

[26] Lin, D., Broere, W., & Cui, J. (2022). Metro systems and urban development: Impacts and implications. *Tunnelling and Underground Space Technology*, 125, 104509.

[27] Larpruenrudee, P., Surawski, N. C., & Islam, M. S. (2022). The effect of metro construction on the air quality in the railway transport system of Sydney, Australia. *Atmosphere*, 13(5), 759.

[28] Ambade, B. R., & Ambade, C. B. (2023). Environmental impact assessment: A case study of construction of metro rail Nagpur, Maharashtra (India). *International Journal of Recent Engineering Science*, 10(3), 39–43.

[29] Xiao, D., Li, B., & Cheng, S. (2020). The effect of subway development on air pollution: Evidence from China. *Journal of Cleaner Production*, 275, 124149.

[30] Xue, X., Zhang, R., Zhang, X., Yang, R. J., & Li, H. (2015). Environmental and social challenges for urban subway construction: An empirical study in China. *International Journal of Project Management*, 33(3), 576–588.

[31] Hemasree, R., & Subramanian, C. V. (2022). Understanding the impacts and the influences of Metro rail on urban environment: Case studies and the Bengaluru scenario. *arXiv preprint arXiv:2209.14210*. https://arxiv.org/abs/2209.14210

[32] Kumar, M. (2003). Environmental management during metro railway construction especially in highly polluted and densely populated city. *WIT Transactions on The Built Environment*, 64.

[33] UrbanRail.net (2024). Asia Metro Systems. https://www.urbanrail.net/as/asia.htm

[34] ERA5 (2024). Data documentation. European Centre for Medium-Range Weather Forecasts (ECMWF). https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation

[35] Census of India (2024). Office of the Registrar General & Census Commissioner, India. https://censusindia.gov.in/

[36] Department of Economics and Statistics, Karnataka (2024). Economic Survey Reports. https://des.karnataka.gov.in/info-4/Economics+Survey/Reports/en

[37] United Nations ESCWA (2024). Locality (Census) – Glossary. https://www.unescwa.org/sd-glossary/locality-census

[38] Department of Climate Change, Energy, the Environment and Water, Australia (2024). Latitude and Longitude Documentation – FullCAM 2024. https://www.dcceew.gov.au/themes/custom/awe/fullcam/Help-FullCAM2024/180_Latitude%20and%20Longitude.htm

[39] Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230. https://psantanna.com/files/Callaway_SantAnna_2020.pdf

Appendix

Note: A supplementary Appendix has all the documents for the regression tests as it had too many rows making it difficult to paste on LaTeX.

I Maps

To provide geographical context for the cities analysed in this thesis, this section presents an overview of the metro networks across major Indian metropolitan areas included in the dataset. The maps illustrate the spatial structure of metro systems, including existing operational lines and planned or under–construction expansions at the time of study.

These visual references help clarify system complexity, network maturity, coverage patterns, and the timing of phased expansions, all of which are relevant for interpreting the empirical results. The maps are sourced from UrbanRail.net, Noida Metro Rail Corporation (NMRC).

For ease of reference, the cities are mentioned in the order of it being present below:

1. Gurgaon Metro Network

2. Kochi Metro Network

3. Chennai Metro Network

4. Bengaluru Metro Network

5. Hyderabad Metro Network

6. Nagpur Metro Network
7. Ahmedabad Metro Network
8. Lucknow Metro Network
9. Jaipur Metro Network
10. Delhi Metro Network
11. Noida Metro Network

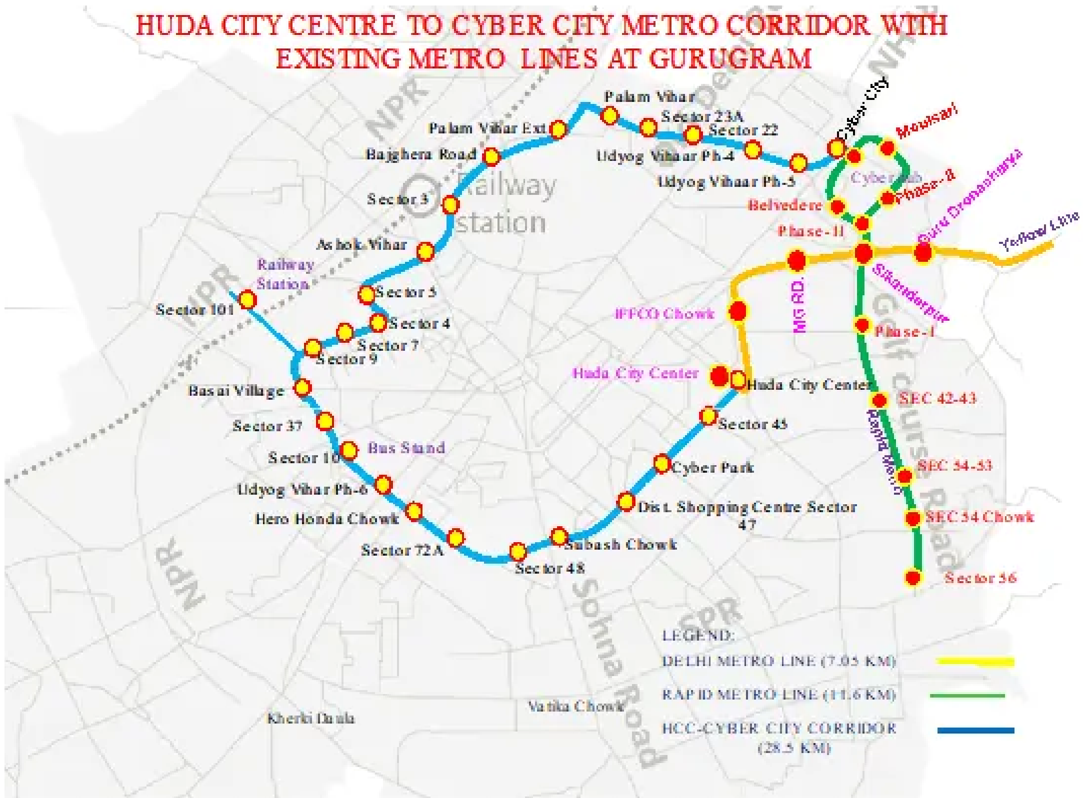


Figure I: (1) Gurgaon Metro Network: Existing and proposed lines.

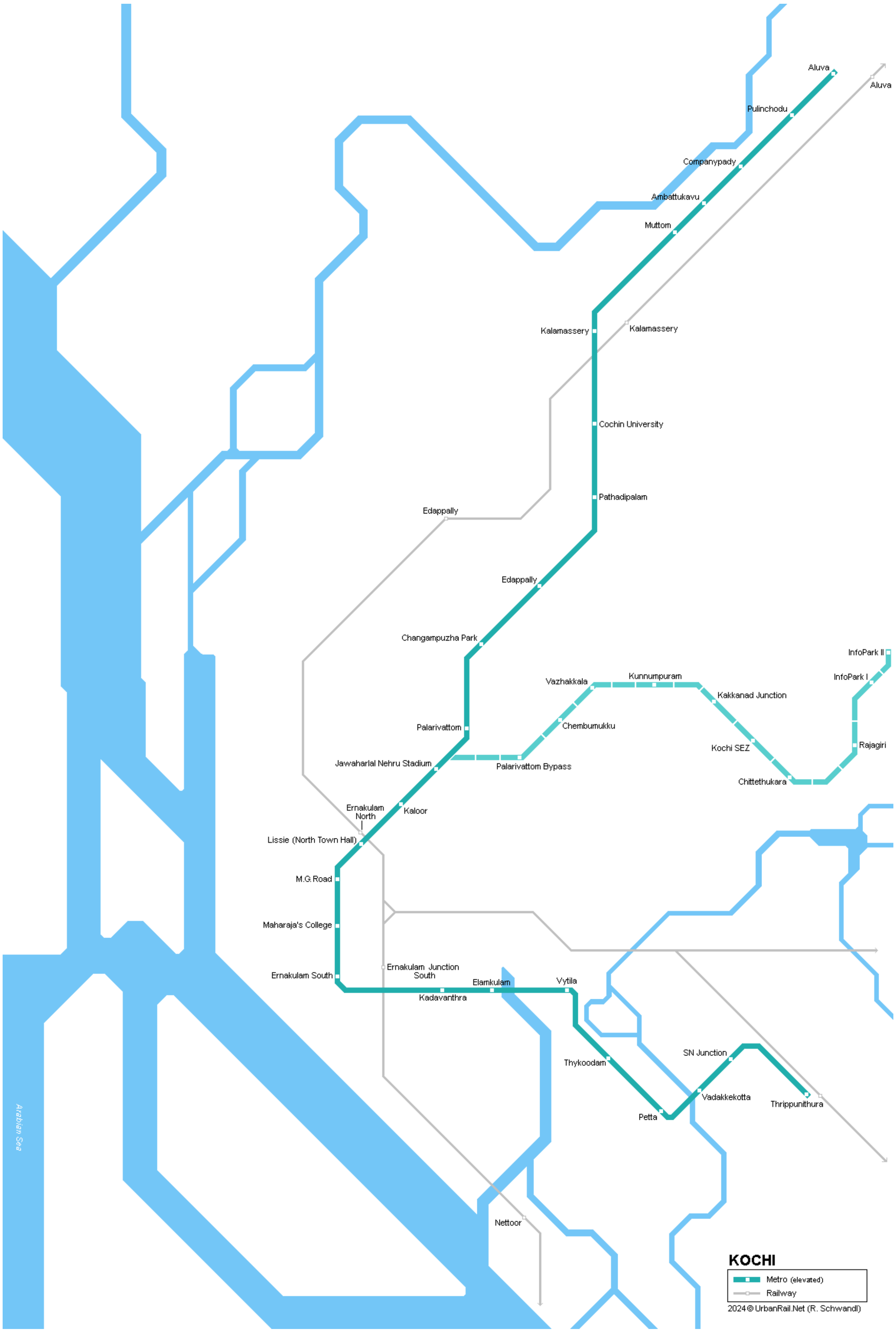


Figure II: (2) Kochi Metro Network.

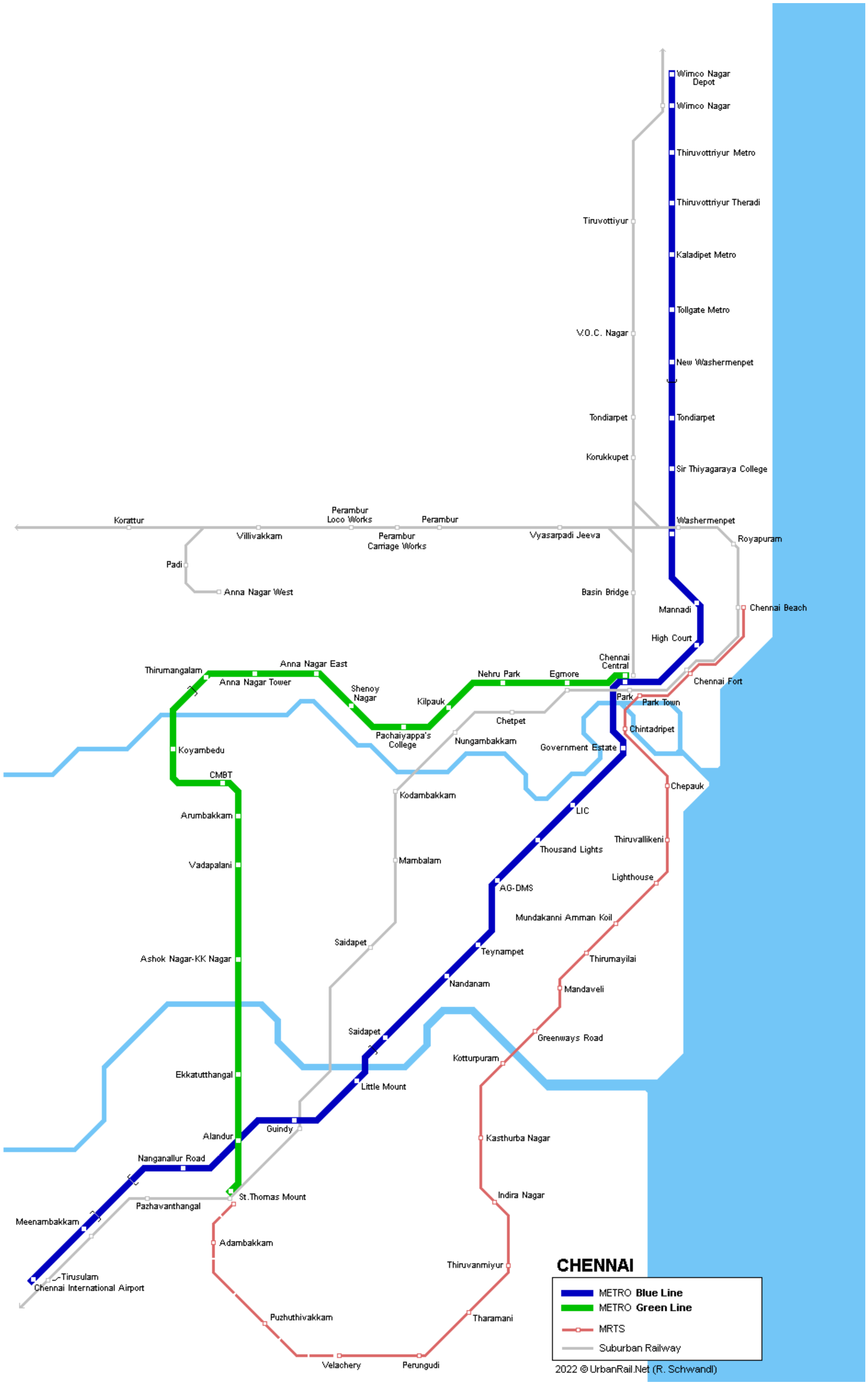


Figure III: (3) Chennai Metro Network (Blue and Green Lines).

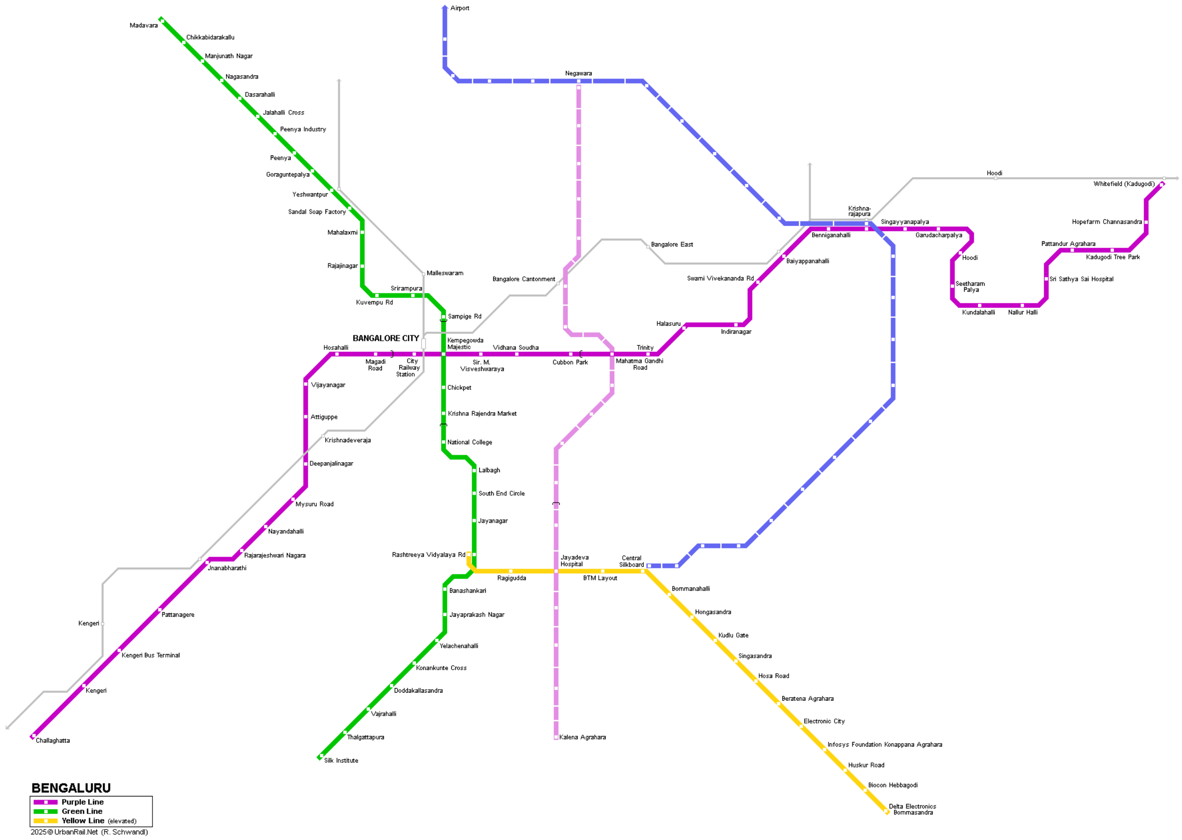


Figure IV: (4) Bengaluru Metro Network (Purple, Green, Yellow Lines).

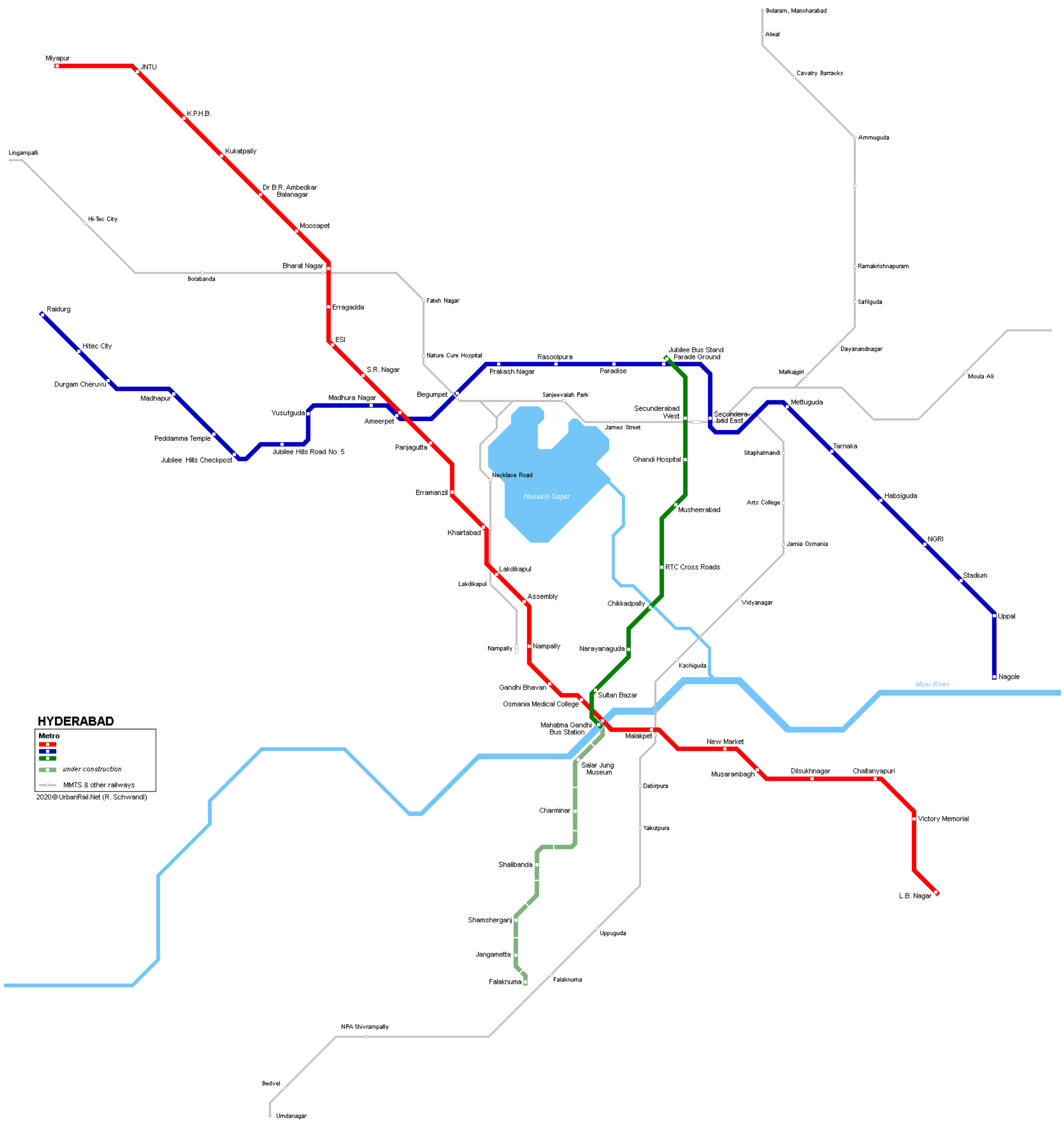


Figure V: (5) Hyderabad Metro Network (Red, Green, Blue Lines).

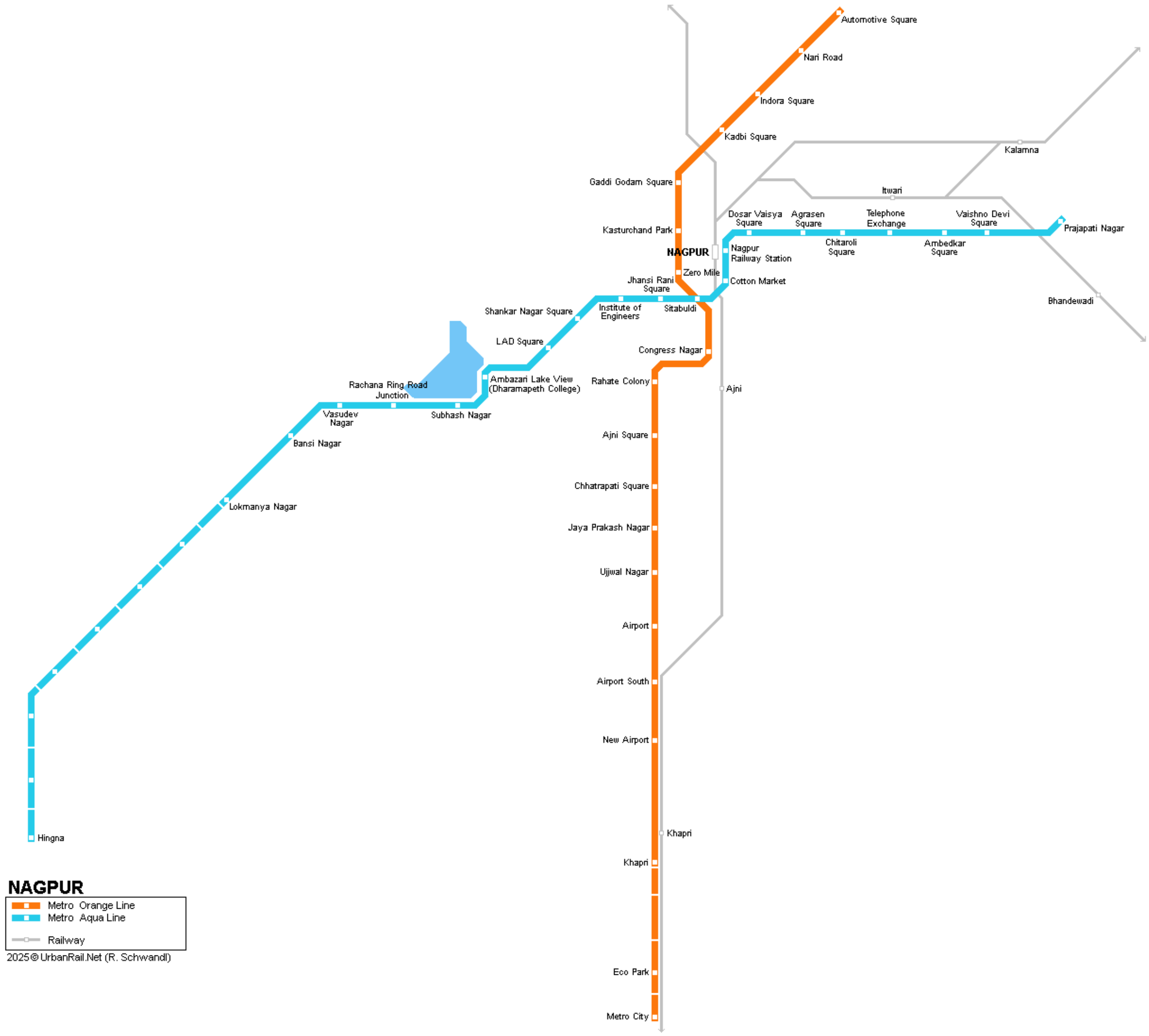


Figure VI: (6) Nagpur Metro Network (Orange and Aqua Lines).

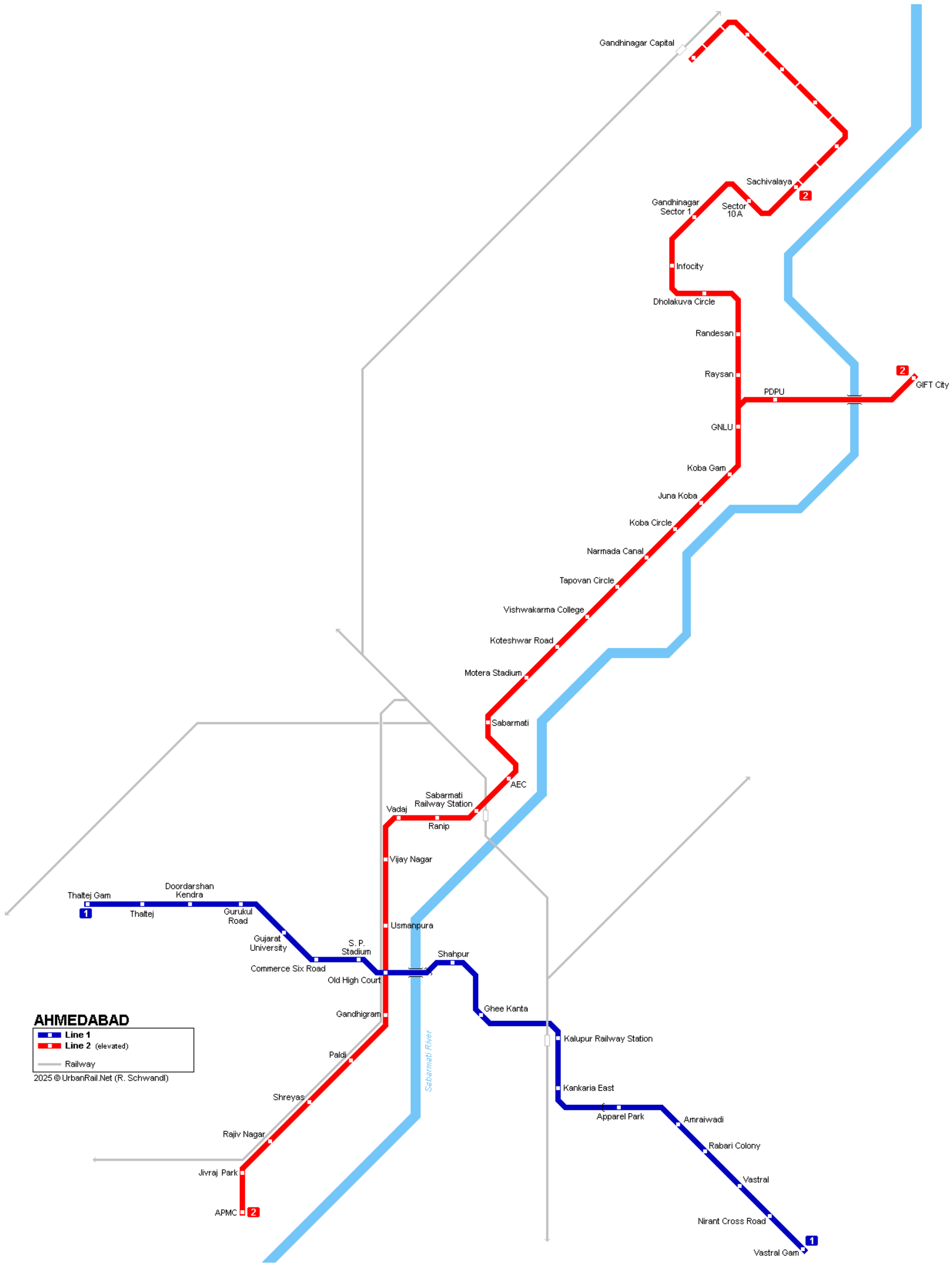


Figure VII: (7) Ahmedabad Metro Network (Line 1 and Line 2).

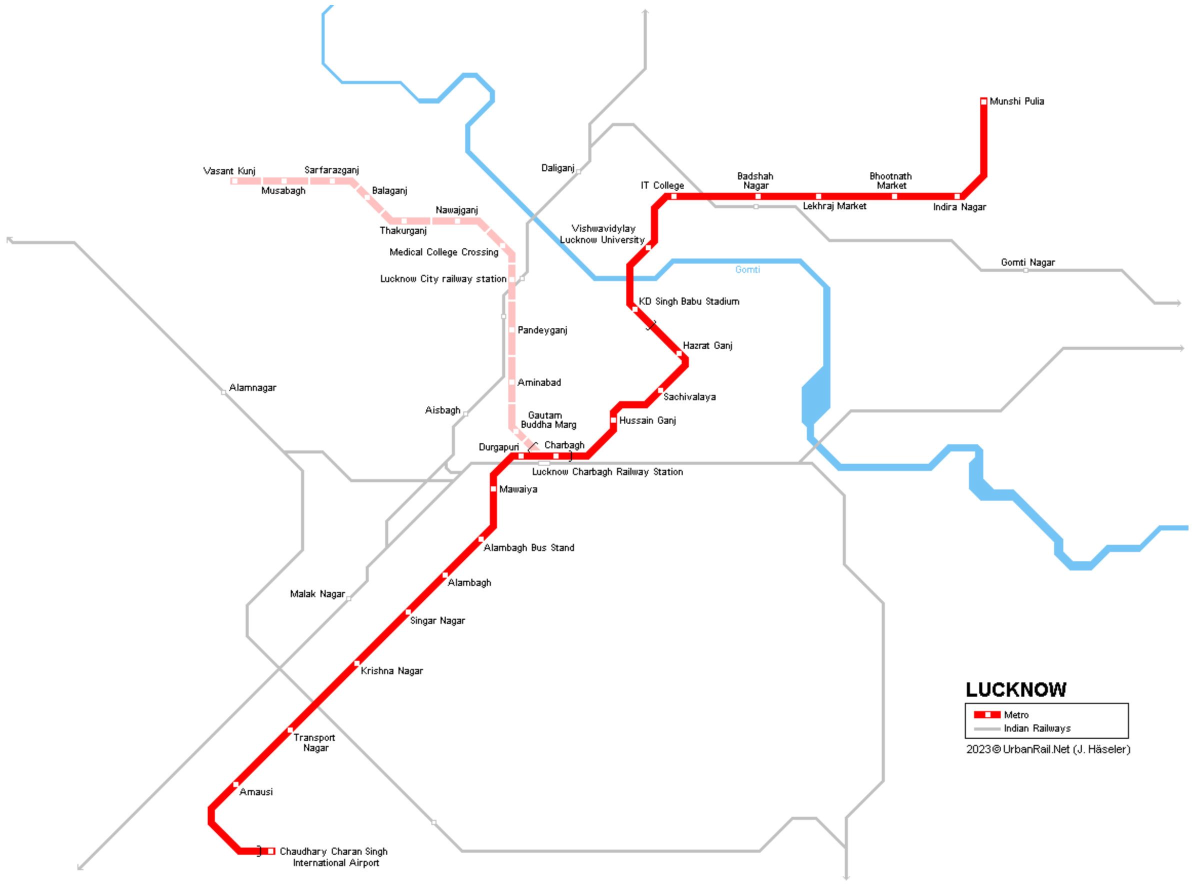
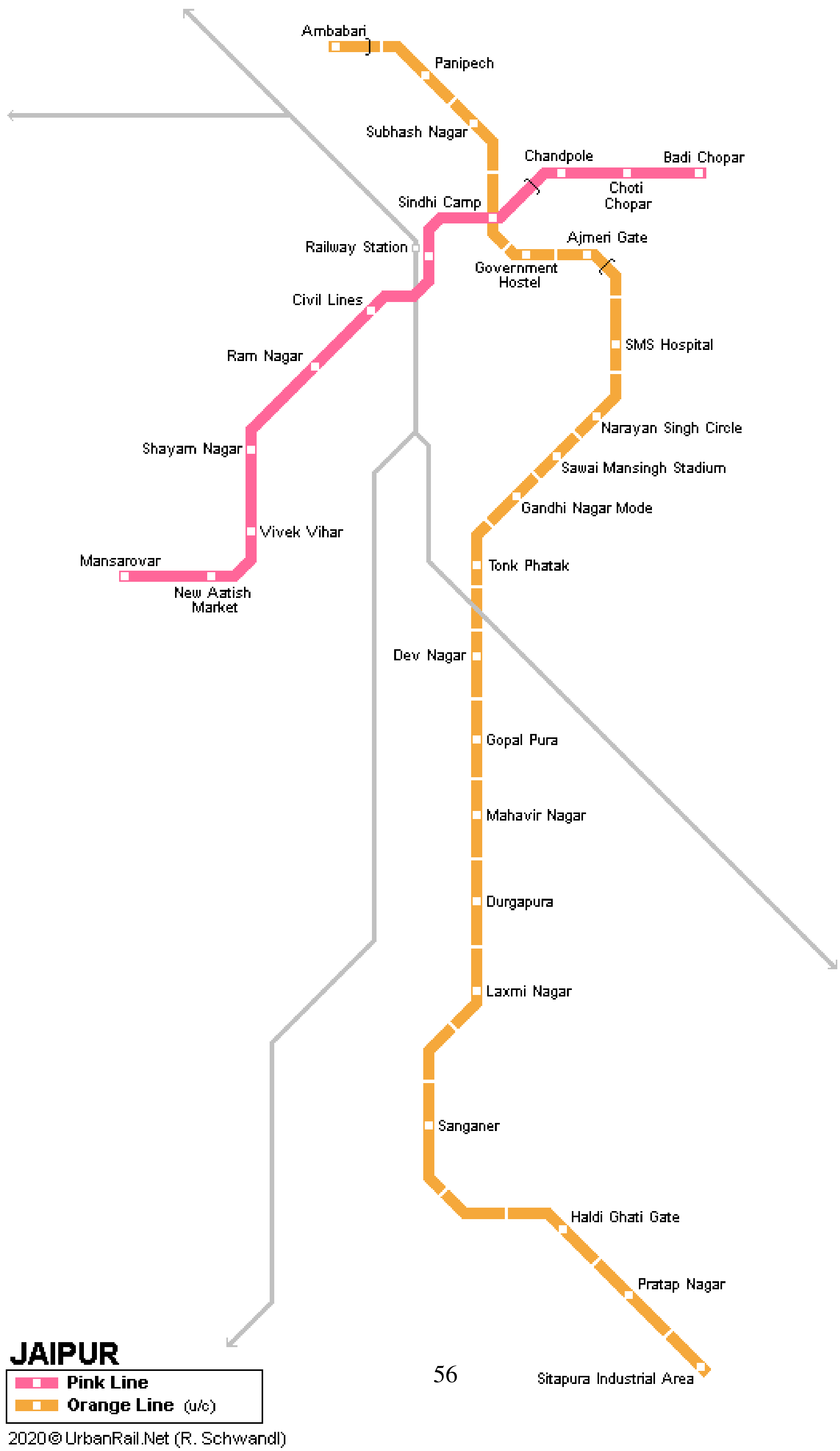


Figure VIII: (8) Lucknow Metro Network (Red Line).


Ambabari
Panipech
Subhash Nagar
Chandpole
Badi Chopar
Choti Chopar
Sindhi Camp
Railway Station
Ajmeri Gate
Government Hostel
Civil Lines
SMS Hospital
Ram Nagar
Narayan Singh Circle
Shayam Nagar
Sawai Mansingh Stadium
Gandhi Nagar Mode
Vivek Vihar
Mansarovar
Tonk Phatak
New Aatish Market
Dev Nagar
Gopal Pura
Mahavir Nagar
Durgapura
Laxmi Nagar
Sanganer
Haldi Ghati Gate
Pratap Nagar
Sitapura Industrial Area
JAIPUR
Pink Line
Orange Line (u/c)
2020 © UrbanRail.Net (R. Schwandl)

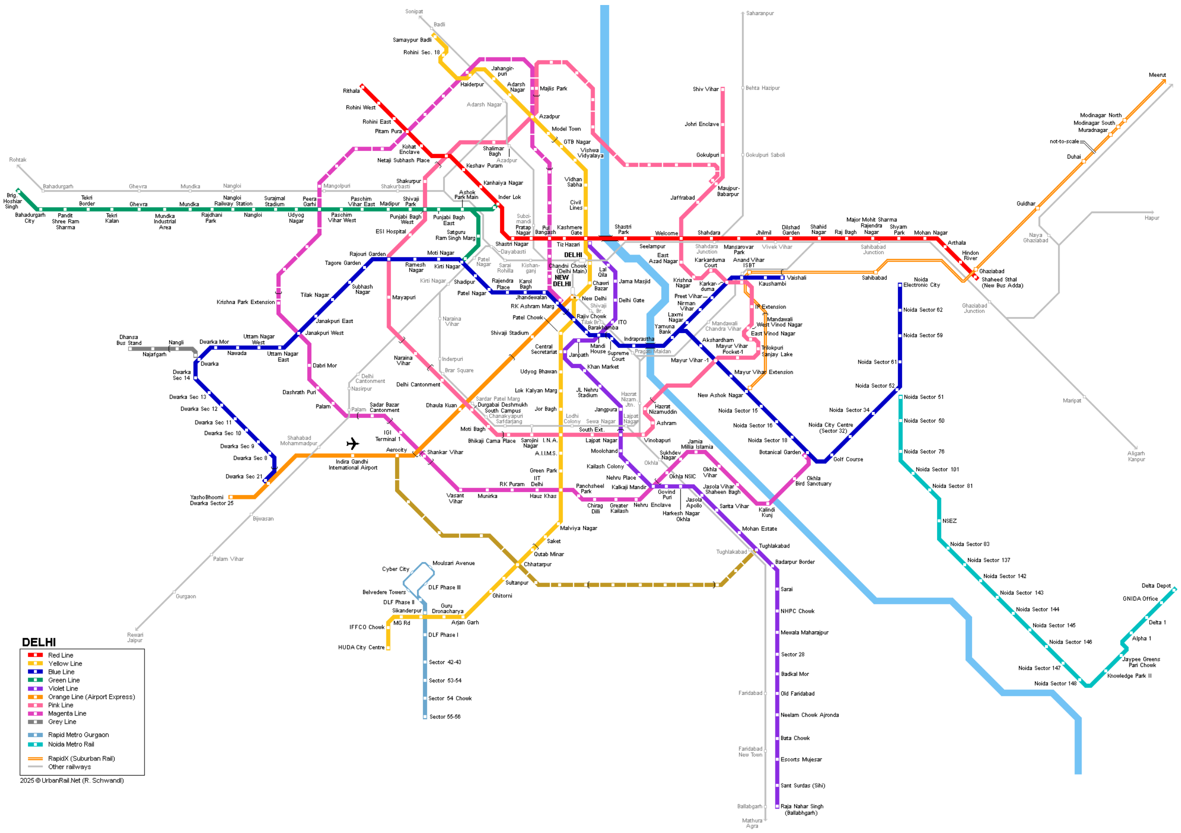


Figure X: (10) Delhi Metro Network: Major corridors and expansions.

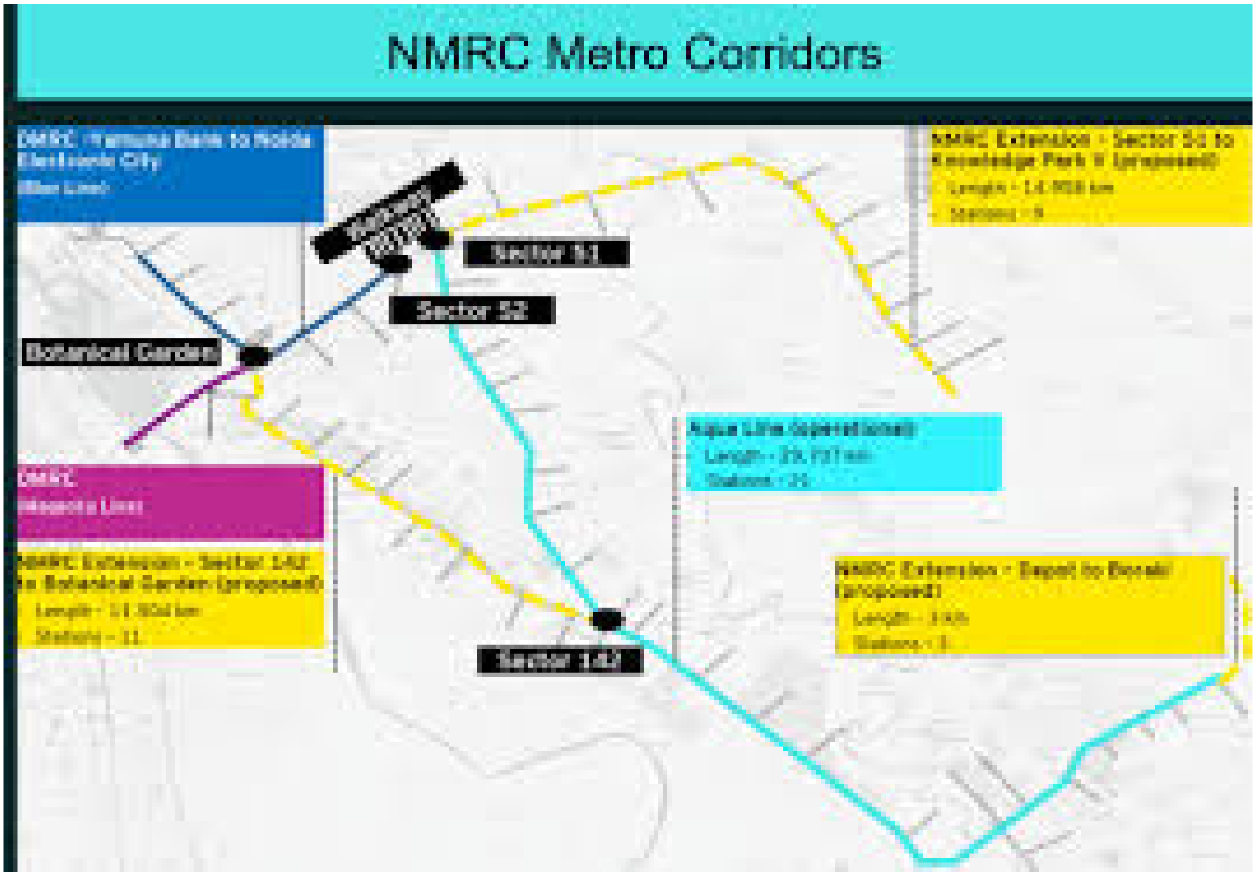


Figure XI: (11)Noida Metro Rail Corridor (NMRC) – Current and Proposed Extensions

OLS Robustness Checks

Table I: Appendix A: OLS Check: Monthly Ridership and $PM_{2.5}$

	Estimate	Std. Error	p-value
Monthly_Ridership	8.8e–08	1.0e–06	0.9299
Observations	50,088		
City FE	Yes		
Month FE	Yes		

Notes: FE-OLS regression of $PM_{2.5}$ on monthly ridership. Standard errors IID. This is an auxiliary robustness check. Coefficient is small and not statistically significant.

Appendix B: OLS Robustness: Petrol and Diesel Prices

Table II: OLS Check: Fuel Prices and $PM_{2.5}$

	Estimate	Std. Error	p–value
Petrol	-0.0216	0.0621	0.7284
Diesel	-0.0126	0.0504	0.8032
Observations	52,213		
City FE	Yes		
Month FE	Yes		

Notes: FE–OLS regression of $PM_{2.5}$ on monthly retail fuel prices (petrol and diesel). Both coefficients are small and statistically insignificant.

Appendix C: OLS Robustness: Holiday Dummy

Table III: OLS Check: Holiday Indicator and $PM_{2.5}$

	Estimate	Std. Error	p–value
Holiday (dummy)	22.0080	0.7128	< 0.001 ***
Observations	52,213		
City FE	Yes		
Month FE	Yes		

Notes: FE-OLS regression of $PM_{2.5}$ on a holiday indicator. Coefficient indicates $PM_{2.5}$ is $\approx$22 $\mu g/m^3$ higher on holiday periods.

List of Control Group Cities

3

- Agartala

- Agra
- Ahmadnagar
- Ajmer
- Akola
- Aligarh
- Amravati
- Amritsar
- Anand
- Aizawl
- Asansol
- Aurangabad
- Bagalkot
- Bareilly
- Bardhaman
- Belagavi
- Bellary
- Berhampur
- Bhagalpur
- Bhatinda
- Bhavnagar

- Bhilai
- Bhiwandi
- Bhopal
- Bhubaneswar
- Bilaspur
- Bokakhat
- Bokaro Steel City
- Chandigarh
- Coimbatore
- Cuttack
- Dahod
- Dehradun
- Dhanbad
- Dibrugarh
- Dombivli
- Durgapur
- Erode
- Faridabad
- Gaya
- Ghaziabad

- Gorakhpur
- Guntur
- Guwahati
- Gwalior
- Haldwani
- Hamirpur
- Hubballi–Dharwad
- Imphal
- Indore
- Itanagar
- Jabalpur
- Jalandhar
- Jalgaon
- Jamalpur
- Jamnagar
- Jammu
- Jamshedpur
- Jhansi
- Jodhpur
- Kalaburagi

- Kakinada
- Kannur
- Kanpur
- Karimnagar
- Karnal
- Kolhapur
- Kollam
- Kota
- Kozhikode
- Kumbakonam
- Kurnool
- Ludhiana
- Madurai
- Mangaluru
- Mathura
- Meerut
- Mohali
- Moradabad
- Muzaffarpur
- Mysuru

- Nadiad
- Nanded
- Nashik
- Nellore
- Patna
- Pimpri-Chinchwad
- Prayagraj
- Puducherry
- Purulia
- Raichur
- Raipur
- Rajkot
- Ranchi
- Ratlam
- Rewa
- Rourkela
- Saharanpur
- Salem
- Sangli
- Shimla

- Silchar
- Siliguri
- Solapur
- Srinagar
- Surat
- Thanjavur
- Thiruvananthapuram
- Thrissur
- Tiruchirappalli
- Tirunelveli
- Tiruvannamalai
- Ujjain
- Vadodara
- Varanasi
- Vasai–Virar
- Vellore
- Vijayapura
- Vijayawada
- Visakhapatnam
- Warangal

List of R Packages and Econometric Methods Used

All empirical estimations were conducted using the **R** statistical programming language (Version 4.5.1). The following R packages and econometric tools were used for data preparation, analysis, and visualization:

- **tidyverse** (Wickham et al., 2019) – Data manipulation, transformation, and visualization.
- **dplyr** - Efficient data wrangling and aggregation.
- **ggplot2** - Visualization of figures and regression plots.
- **readr** and **readxl** – Importing datasets from CSV and Excel formats.
- **sf**, **sp**, **rgdal**, **rgeos** – Spatial data processing, projection, and geometry operations.
- **tmap**, **ggspatial** - Mapping, thematic visualizations, and map annotation.
- **geosphere** - Geodesic distance and spatial coordinate calculations between metro lines and localities.
- **fixest** (Bergé, 2018) - High - dimensional fixed-effects estimation; all main regressions estimated using `feols()`.
- **did** (Callaway & Sant'Anna, 2021) – Staggered Difference-in-Differences estimations using group-time average treatment effects.
- **sunab()** – Implementation of the Sun and Abraham (2021) method for heterogeneous treatment effects and event-study visualization within `feols()`.
- **broom** – Tidying model output and coefficient extraction.
- **stargazer** and **modelsummary** – Generating regression tables for LaTeX.
- **plm** – Panel data estimation and diagnostics.

- **sandwich** and **lmtest** – Robust and clustered standard errors.
- **lubridate** – Date handling for panel time structures.
- **reshape2** and **tidyr** – Data reshaping and transformation.
- **scales** – Plot customization and axis formatting.
- **kableExtra** – Table formatting for LaTeX and PDF output.

All statistical analyses were performed in RStudio (Version 2024.x), combining `fixest::feols()` and `did` implementations for causal inference.