ISSN 0252-1075
Contribution from IITM
Technical Report No.TR -08
ESSO/IITM/MAQWS/TR/01(2025)/203

IITM

Air pollution forecasts over New Delhi: Multi-model inter-comparison

Indian Institute of Tropical Meteorology,
Ministry of Earth Sciences (MoES),
Dr. Homi Bhabha Road, Pashan, Pune – 411 008
https://www.tropmet.res.in

ISSN 0252-1075
Contribution from IITM
Technical Report No.TR -08
ESSO/IITM/MAQWS/TR/01(2025)/203

Air Pollution Forecasts over Delhi: Multi-model Intercomparison

Lead Authors

Gaurav Govardhan, Prafull P. Yadav, Chandrakala Bharali, Pritanjali Shende, Aditi Rathod, Rajmal Jat, and Sachin D. Ghude.

Contributory Authors

(in alphabetical order by last name)

T. Anurose, S. D. Attri, Shweta Bhati, Partha Bhattacharjee, Sreyashi Debnath, Johannes Fleming, Pawan Gupta, A. Jayakumar, Chinmay Jena, Gayatri Kalita, Rajesh Kumar, Alqamah Sayeed, Junhyeon Seo, Rajanikant Shinde, and Vijay Soni.

Indian Institute of Tropical Meteorology, Pashan, Pune – 411008

Corresponding Author:
Dr. Sachin D. Ghude
Scientist - F and Project Head
Metropolitan Air Quality and Weather Forecasting Services
Indian Institute of Tropical Meteorology,
Dr. Homi Bhabha Road, Pashan, Pune - 411 008, India.
E-mail: sachinghude@tropmet.res.in
Phone: +91-(0)20-25904494

Indian Institute of Tropical Meteorology (IITM)
Ministry of Earth Sciences (MoES)
PUNE, INDIA
https://www.tropmet.res.in

DOCUMENT CONTROL SHEET

Ministry of Earth Sciences (MoES)
Indian Institute of Tropical Meteorology (IITM)

ESSO Document Number
ESSO/IITM/MAQWS/TR/01(2025)/203

Title of the Report
Air Pollution Forecasts over Delhi: Multi-model Intercomparison

Lead Authors
Gaurav Govardhan , Prafull P. Yadav, Chandrakala Bharali, Pritanjali Shende, Aditi Rathod, Rajmal Jat and Sachin D. Ghude

Contributory Authors: (in alphabetical order by last name)
T. Anurose, S. D. Attri, Shweta Bhati, Partha Bhattacharjee, Sreyashi Debnath, Johannes Fleming, Pawan Gupta, A. Jayakumar, Chinmay Jena, Gayatri Kalita, Rajesh Kumar, Alqamah Sayeed, Junhyeon Seo, Rajanikant Shinde, and Vijay Soni

Reviewer
Dr. K.J Ramesh - Ex-DG IMD, Ex- Member (Technical) CAQM

Type of Document
Technical Report

Number of pages and figures
36, 5

Number of references
35

Keywords
Air quality Index, GRAP, air quality forecasts, Delhi, WRF-Chem, GEOS-ML, GEOS-FP, GEOS-ML, SILAM, GEFS-Aerosols, DM-Chem, model-intercomparison

Security classification
Open

Distribution
Unrestricted

Date of Publication
July 2025

Table of Contents

List of Tables

List of Figures

Abstract

Air pollution remains one of the most serious environmental and public health challenges in urban regions like Delhi, where fine particulate matter ($PM_{2.5}$) often reaches hazardous levels. Effective forecasting of $PM_{2.5}$ concentrations is essential for timely air quality management, public health advisories, and policy decisions. This study provides a comprehensive evaluation of multiple regional and global air quality forecasting models to assess their performance in predicting $PM_{2.5}$ levels and the associated Air Quality Index (AQI) in Delhi. The models assessed include three regional systems—WRF-Chem, SILAM, and DM-Chem—and four global systems—IFS, GEOS-FP, GEFS-Aerosols, and a data-driven machine learning-based model known as GEOS-ML. These models differ in terms of their resolution, underlying physics, data assimilation strategies, and forecast techniques. To evaluate their accuracy, model outputs were compared against hourly ground-based air quality measurements from 39 monitoring stations located across Delhi.

Among all the models analyzed, the WRF-Chem-based Air Quality Early Warning System (AQEWS) consistently demonstrated the highest accuracy. It closely matched observed $PM_{2.5}$ levels across both normal and high-pollution episodes, making it a highly reliable tool for operational air quality forecasting in Delhi. The machine learning-based GEOS-ML model also performed well, successfully capturing daily variations and pollution patterns through advanced data analysis methods. The DM-Chem model showed reasonable performance at both high-resolution and coarser configurations. On the other hand, some global models such as IFS and GEOS-FP were able to capture general pollution trends but showed limitations in precision. The GEFS-Aerosols model exhibited relatively lower consistency, particularly during severe pollution events. SILAM displayed noticeable discrepancies during high-concentration periods, highlighting the need for further refinement.

On the basis of an average performance index in simulating the $PM_{2.5}$ over Delhi, the WRF-Chem based AQEWS ranks first at a score of 85, followed by GEOS-ML and DM-Chem (330 m) at 69, which are further followed by DM-Chem (1.5 km) and IFS with scores of 61 and 59 respectively. The other partners including SILAM, GEOS-FP, and GEFS-aerosols reach up to the performance index values of 58, 46, and 46 respectively. This evaluation underscores the critical need for rigorous, data-driven assessment of forecasting systems before operational deployment. By identifying the most dependable models, policymakers and environmental agencies can enhance the effectiveness of air pollution warnings, take proactive mitigation measures, and better safeguard public health in pollution-prone regions like Delhi.

Keywords: Air quality Index, GRAP, air quality forecasts, Delhi, WRF-Chem, GEOS-ML, GEOS-FP, GEOS-ML, SILAM, GEFS-Aerosols, DM-Chem, model-intercomparison

Executive Summary

Air pollution remains one of the most critical environmental and public health challenges in Delhi, with fine particulate matter ($PM_{2.5}$) levels frequently exceeding safe limits, especially during the post-monsoon and winter seasons. To address this issue and support timely interventions and public health actions, various air quality forecasting models have been developed and deployed by national and international institutions. This technical report provides a comprehensive evaluation of seven key forecasting models used between 2022 and 2025 to predict $PM_{2.5}$ concentrations and Air Quality Index (AQI) levels across Delhi. Seven models were assessed, including regional systems - WRF-Chem, DM-Chem (at 1.5 km and 330 m resolutions), and SILAM as well as global models such as GEOS-ML (a machine learning-based system), GEOS-FP, GEFS-Aerosols, and IFS. The study assessed model performance by comparing forecast outputs for the post-monsoon and winter seasons (October to January) of the years 2022-23, 2023-24, and 2024-25, with ground-based air quality measurements from more than 40 monitoring stations. Multiple statistical metrics and AQI-based categorical evaluations were used to determine their accuracy in predicting $PM_{2.5}$ concentrations and pollution severity levels. In short, how well each model captured pollution trends and severity was evaluated and is addressed in this report.

Among all the systems, the WRF-Chem-based Air Quality Early Warning System (AQEWS) demonstrated the most consistent and accurate performance across seasons and forecast horizons. It reliably captured both typical and extreme pollution events, making it the most dependable model for operational use in Delhi. The DM-Chem model, particularly at 330 m resolution, also performed well, especially in short-term forecasts. The machine learning-based GEOS-ML model showed strong results for moderate pollution levels and demonstrated potential for rapid adaptation to local conditions. However, it showed reduced accuracy during extreme pollution episodes. In contrast, the global models-GEOS-FP, GEFS-Aerosols, and IFS—struggled to predict high pollution events effectively due to their coarser spatial resolution and lack of localized calibration. These models often failed to detect critical AQI categories, which are essential for public health alerts and emergency responses. Based on their average performance index in simulating $PM_{2.5}$ levels over Delhi, the WRF-Chem-based AQEWS outperforms others with a leading score of 85. This is followed by GEOS-ML and the high-resolution DM-Chem (330 m), both scoring 69. The DM-Chem (1.5 km) and IFS models follow with scores of 61 and 59, respectively. Other participating models, including SILAM, GEOS-FP, and GEFS-Aerosols, demonstrate relatively lower performance, with scores of 58, 46, and 46, respectively.

Overall, the findings highlight the superiority of high-resolution, regionally customized forecasting systems for complex urban environments like Delhi. While global models offer broader coverage, they fall short in precision for localized scenarios like Delhi. The results support continued investment in localized emissions inventories, advanced data assimilation techniques, and model refinement to enhance prediction capabilities, forecast reliability and support timely public health interventions. Strengthening such forecasting infrastructure will play a crucial role in guiding policy actions, optimizing warning systems like GRAP, and better protect public health during severe pollution episodes in India's most polluted urban areas.

1. Introduction

1.1 Background and Context

Air pollution is one of the most pressing environmental and public health challenges confronting India today. Among the country's urban centers, Delhi, and its surrounding National Capital Region (NCR) consistently rank among the most polluted areas globally. The region frequently records annual average $PM_{2.5}$ concentrations several times higher than World Health Organization (WHO) guidelines, with daily concentrations often exceeding safe limits for extended periods. Episodes of severe air pollution are particularly prevalent during the post-monsoon (October–November) and winter (December–February) seasons, during which air quality index (AQI) values often reach hazardous levels (Govardhan et al., 2024; Jena et al., 2021; Ghude et al., 2020). Multiple factors contribute to the recurring pollution crisis in Delhi. The region is characterized by dense urbanization, high population density, rapid industrial expansion, and a growing vehicle fleet, all of which generate substantial anthropogenic emissions. Key contributors to the air pollution in the city include vehicular exhaust, industrial emissions, residential fuel combustion, construction dust, power generation, road dust, and open waste burning (ARAI & TERI, 2018; TERI, 2021). In addition to these local sources, transboundary pollution from neighbouring states, especially during agricultural burning seasons, significantly exacerbates the situation. The practice of crop residue burning, particularly in Punjab and Haryana during October and November, introduces massive quantities of particulate matter and trace gases into the atmosphere, which are subsequently transported towards Delhi by prevailing westerly winds (Jethva et al., 2019; Kulkarni et al., 2020; Lan et al., 2022). Seasonal meteorology plays a pivotal role in modulating pollutant concentrations. During winter, temperature inversions, low wind speeds, and stable atmospheric layers restrict vertical mixing and trap pollutants near the surface, compounding the already high emission load (Yadav et al., 2022). These conditions make the region particularly vulnerable to episodic pollution spikes, or "air pollution emergencies," requiring timely intervention and public advisories.

1.2 Public Health Implications

The health effects of prolonged exposure to poor air quality are severe and far-reaching. Particulate matter (especially those with their effective diameter less than 2.5 m, i.e. $PM_{2.5}$) can penetrate deep into the respiratory tract, leading to both short-term and chronic health conditions. The rise in respiratory illnesses, including asthma, bronchitis, and chronic obstructive pulmonary disease (COPD), has been consistently linked to high pollution levels in Delhi (Jat et al., 2024; Lelieveld et al., 2020). Cardiovascular stress, developmental impacts in children, and increased mortality risks further underscore the scale of the crisis. In addition to direct health outcomes, pollution also imposes a significant socio-economic burden. Increased healthcare costs, reduced labour productivity, absenteeism in schools and workplaces, and damage to infrastructure all contribute to a broader disruption of societal well-being. Consequently, early warning systems and accurate forecasting tools are essential not only for scientific understanding but also for protecting public health and informing policy interventions.

1.3 Government Initiatives and Policy Framework

Recognizing the gravity of the issue, the Government of India has taken multiple steps to address air pollution in Delhi-NCR. The Commission for Air Quality Management (CAQM) was established as a centralized authority to coordinate pollution control measures across the region. One of CAQM's key instruments is the Graded Response Action Plan (GRAP), a tiered policy mechanism that activates targeted interventions-such as halting construction activities, restricting vehicular movement, or shutting down industries-based on real-time and forecast AQI levels. The different GRAP stages are implemented or revoked based on the forecast of AQI. Thus governing authorities need a reliable, trustworthy and accurate forecast of AQI to take policy-level decisions which are likely to affect the lives of more than 2 million people

living in the city of Delhi. Realizing the need of accurate air quality forecasts, in 2018, the Air Quality Early Warning System (AQEWS) was launched as a collaborative effort between the Indian Institute of Tropical Meteorology (IITM), the India Meteorological Department (IMD), and the National Center for Atmospheric Research (NCAR), USA. AQEWS provides 72-hour air quality forecasts at a spatial resolution of 10 km, 2 km and 400 meters for Delhi and its surrounding areas. The forecasts are updated daily and made available to the public and policymakers through the AQEWS website (https://ews.tropmet.res.in/), forming a critical tool in preemptive air quality management (Ghude et al., 2020; Kumar et al., 2020).

Apart from AQEWS, the following air quality forecasting systems are in operational or experimental use for Delhi:

- System for Integrated modeling of Atmospheric coMposition (SILAM) by India Meteorological Department (IMD)
- Delhi Model with Chemistry (DM-Chem) by the National Centre for Medium Range Weather Forecasting (NCMRWF), MoES, India (Jayakumar et al., 2021, 2025)
- Global Ensemble Forecast System with Aerosols (GEFS-Aerosols) by the National Oceanic and Atmospheric Administration, USA (NOAA) and National Centre for Environmental Prediction, USA (NCEP) (Zhang et al., 2022)
- Goddard Earth Observing System- Forward Processing (GEOS-FP) model and Goddard Earth Observing System-Machine Learning (GEOS-ML) model by NASA (Lucchesi, 2018; Knowland et al., 2022)
- Copernicus Atmosphere Monitoring Service (CAMS) by the European Centre for Medium Range Weather Forecasting (ECMWF) (Peuch et al., 2022)

Each model varies in terms of resolution, emissions input, data assimilation methods, and chemical mechanisms, offering diverse strengths and limitations. For example, GEFS-Aerosols operates globally at a coarser resolution (0.25°), while GEOS-FP integrates machine learning in its modelling pipeline.

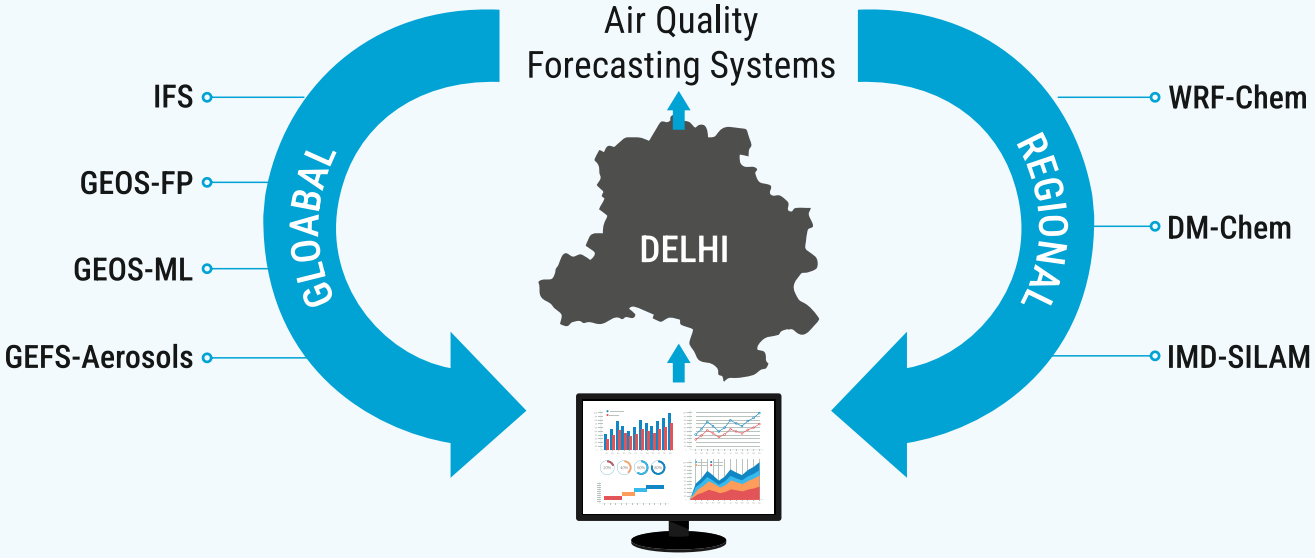


Figure 1: Schematic showing air quality forecasting systems for Delhi, India.

No single model can perfectly capture the complex chemical and physical processes driving air quality in Delhi. Factors such as uncertainties in emissions, meteorological variability, and model parameterizations can lead to divergent forecast outcomes. Multi-model comparisons allow decision-makers to quantify forecast confidence. If multiple systems converge on a prediction of high pollution levels, policy responses such as GRAP stages can be initiated with greater assurance. Conversely, discrepancies among models may prompt closer scrutiny or additional observational support. Several recent studies have assessed the relative performance of different models in forecasting particulate matter concentrations and AQI levels in Delhi, revealing that model performance varies by season, pollution type, and pollution event intensity. These insights are valuable for both scientific refinement and operational use.

This report presents the first comprehensive multi-model assessment of air quality forecasts for Delhi. It outlines the modeling systems utilized in this study and evaluates their performance by comparing model forecasts with ground-based observational data collected during the 2022–2025 post-monsoon (October–December) and winter (January) seasons.

2. Model details

2.1 Air Quality Early Warning System (AQEWS) based on WRF-Chem

The Air Quality Early Warning System (AQEWS) is a comprehensive operational framework designed to deliver high-resolution air quality based on the Weather Research and Forecasting model coupled with Chemistry (WRF-Chem, Version 3.9.1) (Grell et al., 2005). It leverages the WRF-Chem model in a three-level nested domain configuration to provide high-resolution air quality forecasts over Delhi and surrounding regions. The outermost domain (d01) covers the Indian subcontinent with a grid spacing of 10 km × 10 km, which is dynamically downscaled to a second domain (d02) focussing on the National Capital Region (NCR) at 2 km × 2 km resolution. The innermost domain (d03) is centered over Delhi with very high spatial resolution of 400 m × 400 m, enabling neighbourhood - scale pollutant simulations.

Meteorological initial and boundary conditions are provided by the IITM-Global Forecasting System (IITM-GFS, T1534) with 12.5 km resolution, updated every three hours using Ensemble Kalman Filtering (EnKF) (Mukhopadhyay et al., 2019). Chemical boundary conditions for the outermost domain are supplied by six-hourly, 10-year climatological data from the MOZART-4 (Model for Ozone and Related Chemical Tracers, Version 4) chemical transport model (Emmons et al., 2010). These outputs are dynamically passed to inner domains every three hours to ensure consistent chemical forcing. The WRF-Chem model integrates a chemistry scheme known as MOZCART, combining gas-phase chemistry from MOZART-4 with aerosol modules from GOCART (Goddard Chemistry Aerosol Radiation and Transport model). Forecasts are produced 10 days in advance for d01 and up to 96 hours for domains d02 and d03.

A hierarchical and region-specific emissions strategy is adopted across the nested domains to improve model accuracy. For domain d01 (Indian subcontinent), anthropogenic emissions are based on the EDGAR-HTAP (Emissions Database for Global Atmospheric Research- Hemispheric Transport of Air Pollution) v2.2 global emissions inventory at 0.1° × 0.1° resolution, adjusted using regional scaling factors from Venkataraman et al. (2018) to reflect recent Indian emission trends. Emissions for domain d02 (NCR) are taken from the 2016 inventory by provided by The Energy Research and Resources Institute (TERI) and the Automotive Research Association of India (ARAI) (TERI & ARAI, 2018), originally at 4 km × 4 km and regirded to 2 km × 2 km while conserving mass. Whereas, the innermost domain (domain d03; Delhi city) uses a high-resolution 400 m × 400 m inventory developed under the System of Air Quality and Weather Forecasting And Research (SAFAR) project (Beig et al., 2018) for the year 2018. To ensure consistency and avoid overlap, the high-resolution SAFAR emissions from d03 are upscaled into d02, replacing TERI values over Delhi. Similarly, the adjusted d02 emissions are upscaled into d01 to overwrite EDGAR-HTAP emissions over the NCR region. At each stage, mass conservation is strictly maintained. Diurnal emission cycles are applied based on the temporal profiles derived from Govardhan et al. (2019), improving the model's ability to capture hourly pollution variations.

Biogenic emissions are dynamically estimated using MEGAN (Model of Emissions of Gases and Aerosols from Nature ; Guenther, 2007). Fire emissions use the Fire Inventory from NCAR (FINN) (Wiedinmyer et al., 2011) inventory, supplemented with a historical database (2002–2018) using MODIS and VIIRS active fire count information retrieved by the Moderate Resolution Imaging Spectroradiometer (MODIS) and Visible Infrared Imaging Radiometer Suite (VIIRS) instruments on-board the Aqua and Terra satellites of the A-train chain of satellites. Only the high-confidence fire pixels (>50%) are included in the methodology, enabling grid-specific, near-real-time emission estimates. To improve initial conditions, AQEWS integrates satellite and surface observations of aerosol optical depth (AOD), $PM_{2.5}$, and PM_{10}, using a three-dimensional variational (3D-VAR) data assimilation module utilizing the Grid point Statistical Interpolation (GSI) framework (Wu et al., 2002). This assimilation reduces forecast uncertainty and enhances the system's ability to simulate evolving pollution episodes across spatial and temporal scales.

High-resolution surface-level $PM_{2.5}$ observations are obtained in near real-time from a network of approximately 370 air quality monitoring stations distributed across India. These stations are operated under the coordinated efforts of the Central Pollution Control Board (CPCB), the Delhi Pollution Control Committee (DPCC), and the Indian Institute of Tropical Meteorology (IITM). Of these, around 40 monitoring stations are situated within the Delhi-NCR region, providing dense spatial coverage critical for regional air quality assessment.

To ensure the reliability of the data used in both chemical data assimilation and forecast verification, a rigorous quality control process is applied. Observations with $PM_{2.5}$ concentrations exceeding 1500 µg/m³, which are typically indicative of sensor anomalies or extreme outliers, are excluded. Additional filters are applied to eliminate values affected by instrumental malfunctions or operational inconsistencies. The resulting data set supports the generation of robust air quality forecasts and model evaluation over Delhi and surrounding areas. Air quality forecasts and related products are made accessible to decision-makers, researchers, and the public through a public dissemination portal, available at https://ews.tropmet.res.in/.

2.2 Global Ensemble Forecast System – Aerosols (GEFS-Aerosols)

The Global Ensemble Forecast System Aerosols (GEFS-Aerosols) is an advanced global atmospheric modelling system developed by NOAA in the United States. Officially implemented in 2020, it is part of NOAA's broader effort to improve environmental forecasting through the Unified Forecast System (UFS)—a state-of-the-art, community-driven Earth system modeling framework. GEFS-Aerosols is a specialized version (v1) of the GEFS model (Zhang et al., 2022) predicts aerosols originating from both anthropogenic and natural sources, including emissions from dust storms, biomass burning, and sea spray. The model accounts for key aerosol species such as sulfate, black carbon, organic carbon, dust, and sea salt. This model was developed through collaboration among several NOAA research divisions, including the Global Systems Laboratory (GSL), Chemical Sciences Laboratory (CSL), Air Resources Laboratory (ARL), and Environmental Modelling Center (EMC). GEFS-Aerosols integrates an aerosol module directly within the Finite Volume third version Global Forecast System (FV3GFS), enabling it to simulate how aerosols interact with the atmosphere in real time. Its chemical component is based on WRF-Chem, and it uses aerosol behaviour modules from the GOCART system (Chin et al., 2000).

GEFS-Aerosols includes enhanced dust prediction using the NOAA's FENGSHA dust scheme (Tong et al., 2017) and more accurate representation of smoke from fires through a plume-rise module adapted from the HRRR-Smoke system (Freitas et al., 2007). It incorporates emissions data from a wide range of sources: biomass burning emissions are based on satellite-derived fire radiative power from the Global Biomass Burning Emissions Product (GBBEPx v3), while anthropogenic (human-caused) emissions are drawn from the internationally recognized Community Emissions Data System (CEDS; Zhang el al., 2022).
GEFS-Aerosols provides global, five-day forecasts at a high horizontal resolution of 0.25 degrees. These forecasts are essential for predicting air pollution events, monitoring long-range transport of dust and smoke, and supporting air quality management, public health advisories, and climate research efforts. While originally designed for scientific and operational use, tools like GEFS-Aerosols are increasingly important for informing the public and decision-makers about evolving air quality conditions at the global scale.

2.3 System for Integrated modeLling of Atmospheric coMposition (SILAM)

The India Meteorological Department (IMD) operates the System for Integrated modeLling of Atmospheric coMposition (SILAM), a global-to-meso-scale hybrid Eulerian–Lagrangian dispersion model designed to simulate atmospheric composition and air quality. SILAM captures key pollutants such as SO_2, NOx, O_3, $PM_{2.5}$, and PM_{10}, and integrates both Eulerian and Lagrangian transport schemes along with eight chemico-physical transformation modules. These modules encompass processes such as basic acid chemistry, secondary aerosol formation, tropospheric and stratospheric ozone chemistry, radioactive decay, aerosol dynamics, and pollen transformations. The model also includes advanced data assimilation capabilities, utilizing 3D- and 4D-variational techniques, as well as Ensemble Kalman Filter and Smoother

(EnKF/EnKS) approaches. For its regional application over India, SILAM generates four-day forecasts at a 3 km horizontal resolution, covering the entire Indian subcontinent. It is driven by meteorological fields from the operational 3 km Weather Research and Forecasting (WRF) model. The forecast outputs include all criteria pollutants—namely Pm10, $PM_{2.5}$, CO, NO_2, SO_2, and O_3—alongside additional species such as dust, H_2O_2, HCHO, HNO_3, H_2O, HONO, and NH_3. SILAM's initialization strategy integrates outputs from its global run, the previous day's regional run, and the latest meteorological boundary conditions to ensure consistent and accurate forecasts.

The emission inputs to SILAM leverage the CAMS-GLOB v2.1 inventory at 0.1° resolution, complemented by EDGAR v4.3.2 data to represent anthropogenic coarse and fine mineral particulate emissions. Moreover, sector-specific emissions from sources such as public power, industry, road transport, shipping, off-road dust, agricultural waste burning, and other stationary combustion activities are comprehensively detailed in a source term file, which defines attributes like source name, geographic coordinates, plume rise parameters, emission rates, and temporal profiles. SILAM's robust chemical and aerosol modules include a simplified equilibrium scheme for secondary inorganic aerosols, the Volatility Basis-Set (VBS) framework for secondary organic aerosols, the Carbon Bond Mechanism 5 (CBM5) for gas-phase chemistry enhanced with secondary organic compounds, and the DMAT_SULPHUR scheme specifically adapted for sulfur oxidation under Indian conditions. This sophisticated configuration allows SILAM to deliver high-resolution, chemically detailed forecasts of atmospheric composition across India, effectively accounting for a wide range of chemical transformations and physical dispersion processes (Tiwari 2022).

2.4 Integrated Forecasting System (IFS)

The Copernicus Atmosphere Monitoring Service (CAMS), operated by the European Centre for Medium-Range Weather Forecasts (ECMWF), provides advanced global analyses and forecasts of atmospheric composition. At the heart of this capability is the Integrated Forecasting System (IFS), which ECMWF originally developed for numerical weather prediction and has since been extended to include atmospheric chemistry and aerosols. The IFS employed by CAMS integrates an enhanced version of the Carbon Bond 2005 (CB05) chemical mechanism (as detailed by Flemming et al., 2017), which simulates gas-phase chemistry, alongside a dedicated aerosol module. This synergy enables the system to capture complex interactions between gases and particles in the atmosphere. CAMS routinely delivers five-day global forecasts, encompassing 56 reactive trace gases in the troposphere, detailed representations of stratospheric ozone, and seven distinct aerosol species, including desert dust, sea salt, organic matter, black carbon, sulphate, nitrate, and ammonium. To initialize these forecasts, CAMS employs a sophisticated Four-Dimensional Variational data assimilation (4D-VAR) framework, which blends previous forecasts with a wealth of satellite observations. This assimilation incorporates data on AOD, ozone (O_3), carbon monoxide (CO), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2), thereby ensuring that the model starts from a state closely aligned with real-world atmospheric conditions. The forecasting system further integrates anthropogenic emissions sourced from the CAMS-GLOB-ANT inventory (Soulie et al., 2024) and biomass burning emissions from the Global Fire Assimilation System (GFAS) (Kaiser et al., 2012). Meanwhile, biogenic emissions, such as those from vegetation, are estimated offline using the MEGAN v2.1 model. A significant enhancement occurred on 23 June 2023, when CAMS upgraded its operational system from IFS Cycle 47r3 to 48r1. This upgrade introduced several key advancements: anthropogenic emissions were updated from CAMS-GLOB-ANT v2.1 to v5.3, improving the spatial and temporal resolution and the representation of emission sectors. The system also began explicitly simulating secondary organic aerosols, which are crucial for capturing fine particulate matter dynamics. Additionally, a new stratospheric chemistry scheme (BASCOE) was incorporated to better represent ozone and related processes in the upper atmosphere. This scheme operates alongside the AER aerosol module, which handles the formation of secondary inorganic aerosols such as sulphate and nitrates. Through this highly integrated modelling framework-combining advanced gas-phase chemistry, detailed aerosol processes, dynamic emissions, and continuous assimilation of observational data-CAMS delivers comprehensive, high-resolution forecasts and analyses of global atmospheric composition, serving critical applications ranging from air quality management to climate monitoring and public health advisories.

2.5 Delhi Model for Chemistry

The DM-Chem model is a sophisticated, regionally nested configuration of the Met Office Unified Model (MetUM), purpose-built to simulate atmospheric composition and provide high-resolution air quality forecasts. It is designed with a dual-domain approach: an outer domain at 1.5 km horizontal grid spacing (DM-Chem 1.5 km), which captures meteorological and chemical transport processes over the neighbouring states surrounding Delhi, and an inner nested domain at an exceptionally fine 330 m resolution (DM-Chem 330 m), focused specifically on Delhi to resolve urban-scale processes. This nesting enables seamless interaction between broader regional dynamics and detailed urban features. For representing anthropogenic emissions, DM-Chem employs the widely recognized EDGAR inventory for the outer 1.5 km domain, ensuring consistency with global emission datasets, while utilizing the SAFAR high-resolution emission inventory for the 330 m domain, tailored to capture the complex and localized emission patterns within Delhi. Additionally, the model integrates near-real-time fire emissions data from GFAS, which allows it to account for episodic biomass burning events that significantly influence regional air quality. At its core, DM-Chem incorporates the United Kingdom Chemistry and Aerosol (UKCA) module, enabling comprehensive simulation of atmospheric chemical and aerosol processes. It captures phenomena such as new particle formation, condensation, coagulation, and deposition by employing the advanced Global Model of Aerosol Processes (GLOMAP)-mode aerosol scheme (Mann et al., 2010), which is well-established for resolving aerosol size distributions and chemical evolution.

A standout feature of the DM-Chem system is its urban canopy scheme, which is enhanced by explicit incorporation of local urban morphology data (Theethai-Jacob et al., 2023), thereby allowing it to accurately represent heat fluxes, turbulence, and pollutant dispersion within complex cityscapes. Moreover, DM-Chem is equipped with an advanced double-moment cloud microphysics scheme that couples aerosols with cloud processes, facilitating realistic simulations of aerosol-cloud interactions which are critical for both weather and air quality. The inclusion of a detailed irrigation representation further improves its ability to simulate land-atmosphere exchanges, particularly relevant for agricultural regions of the Indo-Gangetic plain. Collectively, these features make DM-Chem exceptionally well-suited for capturing multi-scale interactions between meteorology, chemistry, and urban processes, thus providing robust city-scale air quality forecasts. For a comprehensive technical description and the latest enhancements to the DM-Chem system, readers are referred to Jayakumar et al. (2025).

2.6 Goddard Earth Observing System Forward Processing

The Goddard Earth Observing System Forward Processing (GEOS-FP) model is National Aeronautics and Space Administration's (NASA's) cutting-edge atmospheric forecasting system that provides near-real-time global data on weather, air quality, and atmospheric composition. NASA advanced weather and air quality forecasting system that helps scientists predict everything from daily weather patterns to pollution levels around the world (Luccesi, 2018). It is developed by NASA's Global Modeling and Assimilation Office (GMAO), is a high-resolution atmospheric data assimilation and forecasting system that provides near-real-time (NRT) analyses of global meteorology and atmospheric composition (Luccesi, 2018). The system operates at an exceptionally high resolution (0.25° latitude × 0.3125° longitude grid, about 25 km) with 72 vertical layers from the surface to the edge of space (0.01 hPa), allowing it to capture fine-scale atmospheric processes. GEOS-FP combines satellite observations, ground measurements, and advanced computer modeling to create detailed 3D maps of our atmosphere updated every 6 hours. The model's data assimilation framework employs the GSI 3D-Var system (Luccesi, 2018; Rienecker et al., 2008). Each assimilation cycle processes approximately 2 million observations, including satellite radiances (e.g., AIRS, AMSU-A), radiosondes, aircraft reports (AMDAR), and surface measurements, with rigorous quality control to exclude cloud-contaminated data (Keller et al., 2021). To minimize perturbations, GEOS-FP implements an Incremental Analysis Update (IAU) technique, which incrementally adjusts the model state over a 6-hour window (Molod et al., 2012).

GEOS-FP delivers outputs at high temporal and spatial resolution, including hourly 2D fields (e.g., surface fluxes) and 3-hourly 3D fields (e.g., winds, temperature) on both native model layers and pressure levels (42 levels from 1000 hPa to 0.1 hPa) (Keller et al., 2021). These outputs drive applications such as chemical transport modeling (e.g., the GEOS-Chem framework for aerosol and trace gas simulations at nested resolutions) and field campaign support (e.g., providing interpolated meteorological data for the ATom mission) (Luccesi, 2018; Newman and Pawson, 2021). The model's Unified Chemistry Extension (UCX), integrated into the GEOS Composition Forecast (GEOS-CF) system, enables 5-day forecasts of stratospheric ozone and pollutants, demonstrating superior performance to climatological parameterizations during polar vortex events (Keller et al., 2021). The system updates itself every few hours with new data, allowing it to make predictions that are both timely and accurate (Molod et al., 2012). NASA regularly updates it with new capabilities, such as improved ozone tracking and better integration of satellite data. Its predictions inform everything from daily air quality alerts to long-term climate research, demonstrating how advanced technology can help us navigate environmental challenges.

2.7 Global Earth Observing System - Machine Learning

NASA is supercharging its Earth observation systems with artificial intelligence through Global Earth Observing System - Machine Learning (GEOS-ML). GEOS-ML framework represents an innovative integration of artificial intelligence (AI) and traditional physics-based modelling within NASA's GEOS infrastructure. Designed to augment the capabilities of established systems like GEOS-FP and GEOS-CF (Composition Forecasting), GOES-ML leverages machine learning (ML) to address persistent challenges in data assimilation, forecast accuracy, and atmospheric composition analysis. It uses convolutional neural networks (similar to what powers facial recognition) to analyze atmospheric data across NASA's high-resolution global grid. This innovative approach combines NASA's decades of satellite data with cutting-edge AI to tackle some of meteorology's toughest challenges. It enhances the GSI system—used in GEOS-FP for data assimilation (Rienecker et al., 2008)—by employing ML algorithms to optimize the integration of heterogeneous observational data (Gupta et al., 2021). This approach can reduce biases in satellite radiance measurements (e.g., from AIRS or OMPS-LP instruments) and improve the quality control of inputs from ground stations, aircraft (AMDAR), and radiosondes, leading to more accurate initial conditions for forecasts.

A critical application of GEOS-ML lies in refining the GEOS-CF system (Keller et al., 2021), which provides 5-day global forecasts of atmospheric composition (e.g., ozone, $PM_{2.5}$, and wildfire smoke). By training ML models on high-resolution GEOS-FP outputs (0.25°×0.3125° grid, 72 vertical layers), GEOS-ML can identify and correct systematic errors in chemical transport simulations, particularly for reactive species like SO2 or aerosols that exhibit complex nonlinear behaviours. Hybrid physics-ML approaches, such as embedding convolutional neural networks (CNNs) within GEOS-CF's Unified Chemistry Extension (UCX), have shown promise in stratospheric ozone prediction during polar vortex events, outperforming traditional parameterizations. Additionally, GOES-ML supports field campaigns like the ATom mission by enabling dynamic interpolation of GEOS-FP meteorological data to flight paths, where ML algorithms enhance tracer tracking and reduce uncertainties in atmospheric sampling.

Beyond forecasting, GOES-ML facilitates real-time applications such as air quality alerts and wildfire smoke dispersion modelling. The framework also addresses computational bottlenecks: ML-based emulators can approximate costly model components (e.g., radiative transfer calculations), reducing runtime while maintaining fidelity. However, challenges remain, including the need for robust uncertainty quantification in ML outputs and the integration of explainable AI (XAI) techniques to ensure transparency for operational users. By bridging the gap between cutting-edge ML and proven Earth system models, GEOS-ML exemplifies the transformative potential of AI in advancing atmospheric science and operational forecasting. At its heart, GEOS-ML works like a constantly learning weather assistant - it takes in billions of data points from satellites like GEOS-R, ground sensors, and weather balloons, then uses machine learning algorithms to spot patterns human scientists might miss.

3. Results

3.1 Multi-Season Comparative Analysis of $PM_{2.5}$ Forecast Performance (2022–2025)

This section presents a comprehensive comparative analysis of $PM_{2.5}$ forecast performance for three successive high-pollution seasons in Delhi: October to January of 2022–23, 2023–24, and 2024–25. Using a consistent set of statistical metrics—including Root Mean Square Error (RMSE), Index of Agreement (IOA), Normalized Mean Bias (NMB), Mean Fractional Bias (MFB), Fraction within a Factor of Two (FAC2), correlation coefficient (R), and Performance Index (PI)—as well as categorical evaluation indicators like Accuracy, False Alarm Ratio (FAR), Critical Success Index (CSI), we examine how eight air quality models performed across forecast lead times and pollution severity levels. The analysis for the model performance is represented in Figure 2, Figure 3 and Figure 4 respectively. A heat map of R presenting the differences in the modelled $PM_{2.5}$ concentrations and the ground observations can be seen in Figure 2. Figure 3 shows the estimated MBF and Figure 4 presents the calculated statistical analysis to understand the model performance with respect to the ground-observations. In the following section we highlight the key points with regards to the performance of each of the models in capturing surface-level $PM_{2.5}$ in comparison with the observations (which are carried out over more than 40 CPCB, DPCC and IITM observatories across the city of Delhi).

3.1.1: WRF-Chem Performance Across Seasons

WRF-Chem consistently led across all forecast days and years.

2022-23: RMSE increased modestly from 71.84 µg/m³ on Day 1 to 82.2 µg/m³ by Day 3, while IOA declined slightly from 0.80 to 0.78. The PI, though strong overall, decreased from 85 on Day 1 to 79 by Day 3.

2023-24: Forecast skill improved, with RMSE ranging from 66.13 to 78.56 µg/m³ and IOA from 0.85 (Day 1) to 0.79 (Day 3). PI remained high across all days, starting at 86 and dipping to 79. Correlation (R) dropped from 0.74 to 0.65, while FAC2 stayed robust (0.91–0.84), indicating reliable amplitude prediction.

2024-25: Only Day 1 data was available, but performance remained solid. RMSE was 90.01 µg/m³, IOA was 0.80, and R was 0.70. The PI held strong at 84.

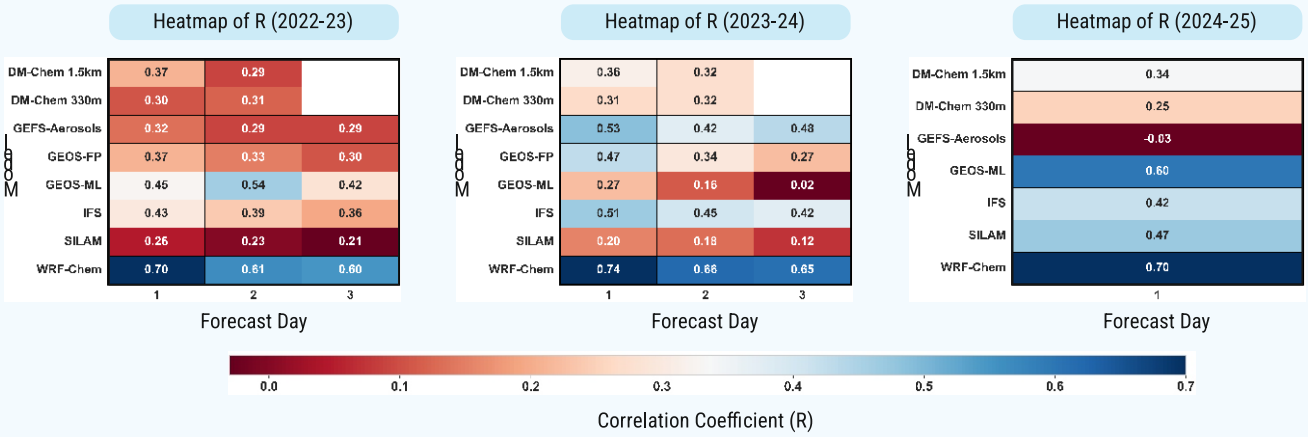


Figure 2: Heat map plot of Pearson's coefficient (R) showing co-relation between various model simulated and observed $PM_{2.5}$ concentrations.

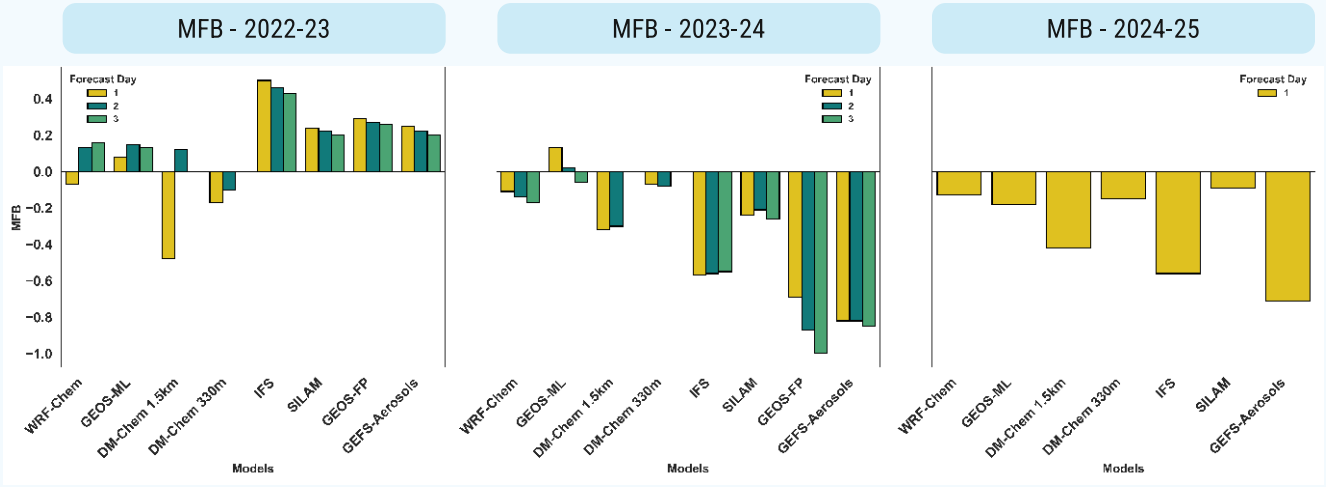


Figure 3: Mean fractional Bias (MFB) analysis performed for the various models and observed $PM_{2.5}$ concentrations.

3.1.2: GEOS-ML Performance

GEOS-ML exhibited moderate and variable performance across seasons, with a notable recovery in 2024–25.

2022-23: IOA held steady around 0.67–0.68, while RMSE increased from 86.71 µg/m³ on Day 1 to 106.9 µg/m³ by Day 3. The PI declined consistently, from 75 to 59.

2023-24: Forecast skill deteriorated with lead time. RMSE rose from 93.3 to 98.73 µg/m³ across Days 1-3, IOA dropped significantly from 0.57 to 0.37, and R fell dramatically to 0.02 by Day 3. The PI followed suit, declining from 68 to 53. Despite this, the model retained modest capability in capturing severe pollution.

2024-25: A substantial rebound occurred. For Day 1, GEOS-ML achieved RMSE = 85.39 µg/m³, IOA = 0.70, and PI = 80—its best in three years.

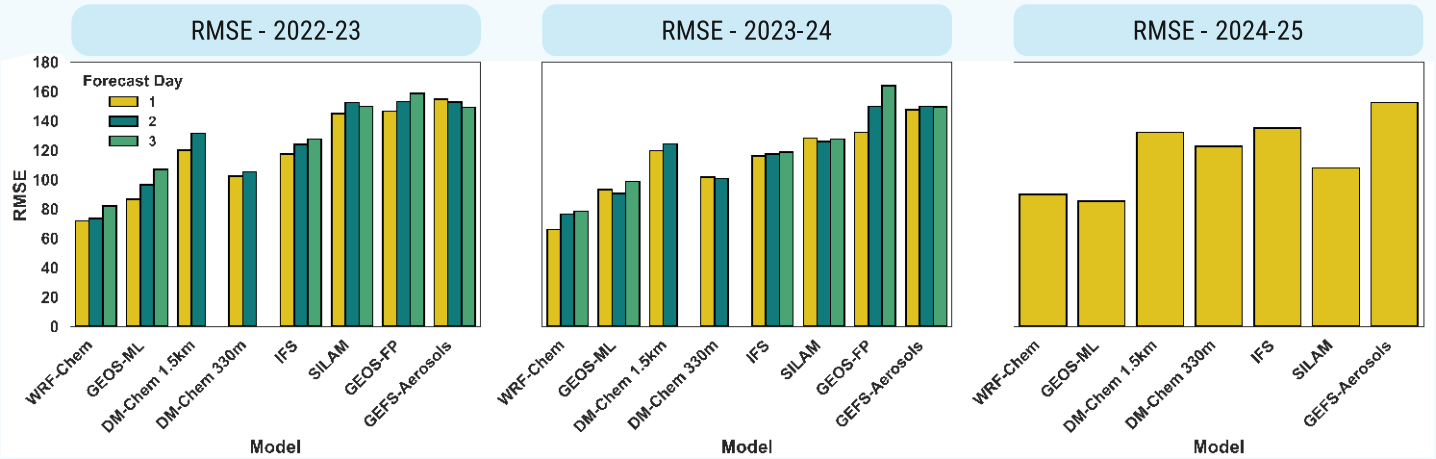


Figure 4: Statistical analysis (RSME, IOA, FAC2, PI) performed for 3 different study period (2022-23, 2023-24, 2024-25) to understand the performance of 7 models (WRF-Chem, GEOS-ML, DM-CHem, IFS, SILAM and GEFS-Aerosols).

3.1.3: DM-Chem Performance (1.5 km and 330 m Resolutions)

The high-resolution DM-Chem system, especially the 330 m version, showed steady gains over time, though with day-wise and version-based differences.

2022-23: The 1.5 km model struggled with RMSE exceeding 120 µg/m³ and PI near 55. The 330 m version, however, outperformed it with IOA up to 0.65 and PI around 67.

2023-24: Both resolutions improved. The 330 m model had RMSE near 101 µg/m³ and maintained a stable PI of 69, with IOA improving modestly. The 1.5 km version still lagged in PI (around 60).

2024-25: The 330 m version remained competitive despite a higher RMSE of 122.71 µg/m³ and a drop in IOA to 0.52. However, its PI held at 69. The 1.5 km model's PI declined to 58, confirming the advantage of ultra-high resolution.

3.1.4: IFS Forecasts

IFS maintained consistently weak performance in both continuous and categorical statistics, with minimal capacity to detect peak events.

2022-23: IOA fell from ~0.65 to 0.62 across forecast days. RMSE increased to 127.5 µg/m³, and PI dropped to 55 by Day 3.

2023-24: Under-performance continued. IOA averaged 0.56, RMSE reached 118.7 µg/m³, and R dropped to 0.42. PI remained around 55 but lacked robustness across lead times.

2024-25: Only Day 1 results are available, with IOA = 0.52 and no clear PI reported.

3.1.5: SILAM's performance

SILAM's performance improved steadily, particularly in 2024–25, but still lagged regional models.

2022-23: RMSE remained high (144.8 to 149.7 µg/m³), IOA declined to 0.52 by Day 3, and PI dropped to 48.

2023-24: IOA improved slightly (0.53 to 0.48), while RMSE stabilized around 128 µg/m³. PI increased modestly to 55-58.

2024-25: A marked performance leap was observed. RMSE dropped to 107.88 µg/m³, IOA rose to 0.66, and PI reached 73-its highest to date.

3.1.6: GEOS-FP and GEFS-Aerosols evaluations

Both global models performed poorly in every season, with low resolution and lack of regional tuning contributing to widespread underprediction.

GEOS-FP: RMSE worsened each year, climbing from 132.17 µg/m³ in 2022–23 to 163.94 µg/m³ by 2023–24. IOA remained below 0.51 throughout, with no improvement in PI or categorical scores.

GEFS-Aerosols: Slightly better in 2022–23 (not fully reported), but by 2023–24, PI was only 43 and CSI stayed below 1%. In 2024–25, PI dropped further to 33. These models remain the least reliable for urban-scale forecasting in India.

3.1.7: Summary and Implications

Across three winter seasons, WRF-Chem consistently outperformed all other models in both statistical accuracy and categorical event detection, reaffirming the effectiveness of high-resolution, regionally tuned forecasting systems. DM-Chem (330 m) demonstrated notable gains, particularly in short-term forecasts. Global models such as GEOS-FP and GEFS-Aerosols underperformed, underscoring the need for regional configurations in complex urban environments like Delhi. The seasonal progression indicates incremental improvements for some models (e.g., GEOS-ML, SILAM), while others plateaued or declined. These results emphasize the importance of sustained model development, high-resolution emission inventories, and data assimilation to enhance urban air quality forecasting.

3.2 Evaluation of Model Performance Across AQI Categories

Assessing forecasting model performance using AQI categories provides a clear and health-relevant perspective, aligning technical accuracy with public understanding. By using the Air Quality Index as a benchmark, model outputs can be meaningfully evaluated for their ability to predict air pollution levels with direct health implications. Figure 5 summarizes the categorical performance of various models for the polluted season during 2022–25 across three AQI severity levels: Unhealthy (AQI >= 200), Very Unhealthy (AQI >= 300), and Critical (AQI >= 400). Key evaluation metrics include Accuracy (higher is better), False Alarm Rate (FAR; lower is better), and Critical Success Index (CSI; higher is better), offering a multi-dimensional framework for comparing predictive skill. In the following section we highlight the key points with regards to the performance of each of the models in capturing AQI in comparison with the observations.

3.2.1: WRF-Chem performance for AQI

2022-23: Accuracy for Unhealthy category remained above 93% across days, with CSI ≥88.1%. FAR was consistently low (≤2.3%), indicating few false alarms. Very Unhealthy category showed CSI decline from 89.5% (Day 1) to 81.3% (Day 3), but maintained <5% FAR. For Critical AQI, however, CSI stayed low (~28–34%), and FAR rose to 26% by Day 3.

2023-24: Highest categorical forecast skill across all AQI categories. Unhealthy and Very Unhealthy events had accuracy >94%, CSI >90%, and FAR <2.5% through Day 3. Critical AQI showed some gains in Day 1 CSI (37.35%), though it declined by Day 3. FAR remained high (~26%).

2024-25: Notable decline in performance. Unhealthy accuracy fell to 85.56%, and FAR rose to 8.36%. Very Unhealthy saw CSI drop to 73.69% and FAR spike to 9.8%. Critical AQI remained challenging with CSI at 38.4% and FAR >31%.

3.2.2: GEOS-ML evaluation for AQI

2022-23: Maintained 100% accuracy on Day 1 for Unhealthy AQI, dropping to 95% by Day 3. CSI stayed >90%, while FAR stayed under 3%. Critical AQI showed limited skill (CSI ~13.2% Day 3), with high FAR (~49%).

2023-24: Peak performance. Unhealthy and Very Unhealthy categories achieved CSI >96% across all days. FAR was minimal (~1.9–3.3%). Critical AQI saw CSI of 40.1% on Day 1 but dropped significantly by Day 3. FAR remained high (~48.7%).

2024-25: Large drop. Unhealthy category accuracy fell to 82.4%, and CSI to 81.88%. Very Unhealthy CSI declined to 67.7%, and FAR rose to 15.76%. Critical AQI showed poor CSI (23.3%) with improved but still high FAR (19.2%).

3.2.3: DM-Chem (1.5 km and 330 m) AQI evaluation

2022-23: Unhealthy category forecasts achieved 85.1–94.6% accuracy, with higher CSI and lower FAR at finer resolution. Very Unhealthy category had moderate CSI (~75–84%). Critical AQI performance was poor (CSI ~6–12%) with FAR exceeding 70%.

2023-24: Improved results. 330 m model reached >96% accuracy and >88% CSI for Unhealthy. FAR stayed ~6–8%. Critical events remained difficult—CSI peaked at 11.1%, with FAR ~75%.

2024-25: Major performance drop. Unhealthy accuracy fell below 75%, and Critical AQI forecasts were largely ineffective. CSI was as low as 12% at 330 m, with FAR >74%.

3.2.4: SILAM's performance

2022-23: Unhealthy category accuracy was ~84%, with FAR <2%. Very Unhealthy category had moderate CSI (~60%), but Critical events had negligible CSI (~3%) and FAR ~92%.

2023-24: Improved Unhealthy category CSI (~87%) with slightly higher FAR. Critical AQI performance remained weak (CSI ~5.8%, FAR ~88.5%).

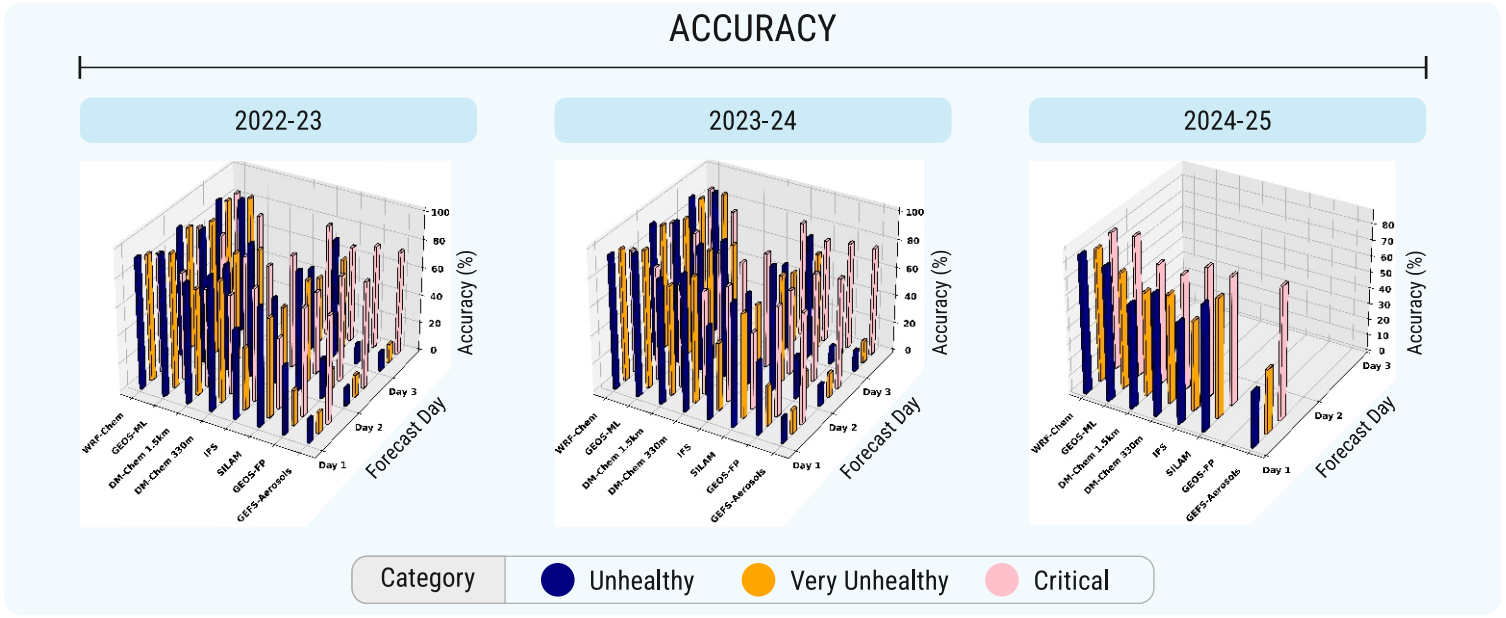
2024-25: Accuracy fell slightly to 77.1% for Unhealthy category, and FAR increased to 15.3%. For Critical category AQI, CSI improved to 20.4%, while FAR dropped slightly to 55.1% — still high but showing some gains.

3.2.5: IFS forecasts

All years: Unhealthy and Very Unhealthy categories showed flat performance: accuracy ~60–65%, CSI ~30–50%, FAR = 0% due to no detection. Critical AQI showed accuracy >80%, but CSI never rose above 0.3%, indicating a total miss on high pollution events.

3.2.6: GEOS-FP and GEFS-Aerosols evaluation

All years: Accuracy remained low (often <20%), and CSI values were near zero across all AQI categories. Critical AQI forecasts had high accuracy (~75–80%) but zero detection skill, with CSI = 0 and FAR = 0, suggesting complete under prediction.



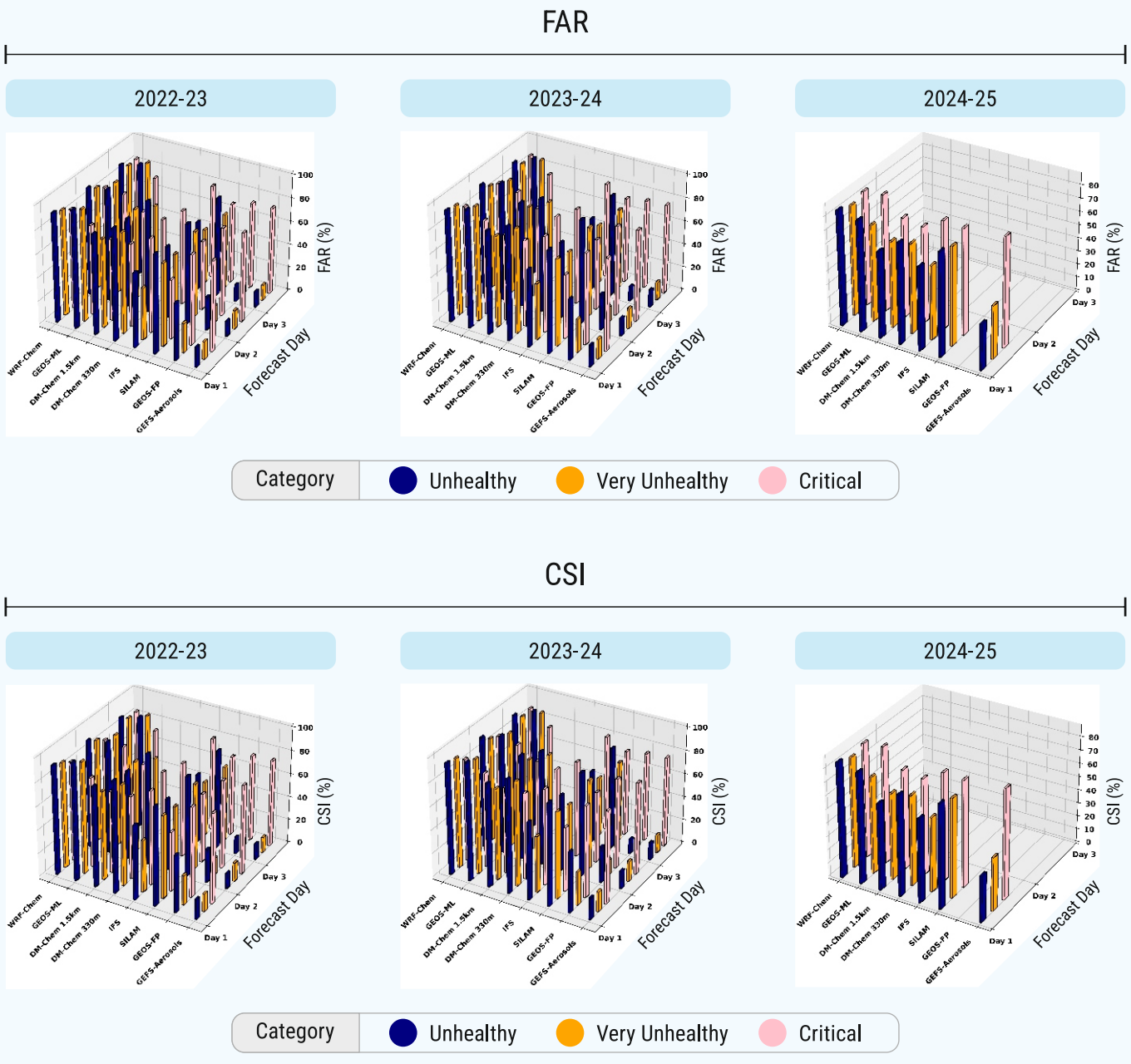


Figure 5: Performance Metrics of Air Quality Forecasting Models across AQI Categories for the polluted seasons (October--January) 2022-23, 2023-24 and 2024-25

3.2.7: Summary and Implications

High-resolution regional models like WRF-Chem and DM-Chem (330 m) consistently outperformed global systems in forecasting AQI categories of public health significance. Machine-learning-enabled models like GEOS-ML also showed promise for moderate pollution levels but struggled with extremes. In contrast, global models demonstrated poor performance, particularly for Critical episodes, due to coarse resolution and limited emission sensitivity. The stark contrast in predictive accuracy reinforces the need for localized modelling efforts, data assimilation, and urban-specific parameterizations to support operational forecasting and public health interventions in megacities like Delhi.

3.3 Categorical Forecast Performance

To present a consolidated evaluation of categorical AQI forecasts, a comparative performance summary is shown in Table 4. The table facilitates side-by-side assessment of air quality forecast models across Day 1, Day 2, and Day 3 lead times for three critical AQI categories: Unhealthy, Very Unhealthy, and Critical. It includes key verification metrics—Probability of Detection (POD), Success Ratio (SR), and Critical Success Index (CSI)—to offer an integrated view of model performance across different pollution severity thresholds. Each metric provides insight into a distinct aspect of forecast quality. The POD reflects how often observed high-pollution events were correctly predicted, while the SR (equivalent to 1 minus the False Alarm Ratio) indicates the proportion of predicted events that occurred. The CSI combines both hit and false alarm rates, offering a balanced measure of overall accuracy.

3.3.1: Unhealthy Category AQI Forecasts

For this category, most models demonstrated strong performance on Day 1, with high POD and SR values indicating good detection capability and few false alarms. WRF-Chem achieved a POD of 95.17% and SR of 99.03%, reflecting a strong balance between sensitivity and precision. Both GEOS-ML and DM-Chem (330 m) outperformed slightly, achieving 100% POD and SR, indicating perfect detection and no false alarms on the first day. In contrast, IFS yielded a lower POD of 61.55%, though with an SR of 100%, suggesting that while it rarely issued false alarms, it failed to identify a substantial number of actual events.

On Day 2, performance began to diverge. GEOS-ML retained perfect POD and SR, while DM-Chem (330 m) remained robust with POD at 96.11% and SR at 92.96%. SILAM showed a moderate decline (POD: 87.34%, SR: 89.66%), and other models exhibited more significant degradation. By Day 3, most models showed further reductions. SILAM's POD dropped to 85.47%, while DM-Chem (330 m) sustained high performance (POD: 96.26%, SR: 91.7%).

3.3.2: Very Unhealthy Category AQI Forecasts

Greater variability was observed across models in this category. On Day 1, GEOS-ML continued to lead with 100% POD and SR, and DM-Chem (330 m) followed closely (POD: 95.31%, SR: 91.92%). IFS showed limited skill with a POD of 62.58% despite perfect SR, again highlighting under-detection. On Day 2, most models experienced modest declines. DM-Chem (330 m) recorded a POD of 94.13% and SR of 92.2%, indicating continued reliability. By Day 3, models like WRF-Chem exhibited a drop in detection ability (POD: 88.5%, SR: 96%), suggesting reduced effectiveness for predicting high-severity pollution events. Lower-resolution global models (e.g., IFS, GEOS-FP) maintained minimal detection, while GEFS-Aerosols failed to register any events in this category.

Table 1: Intercomparison of Air Quality Prediction Models for Different Health Risk Categories

- AQEWS (WRF-Chem)

Healthwise AQI Category	Unhealthy		Very Unhealthy		Critical	
Forecast Day	POD (%)	SR (%)	POD (%)	SR (%)	POD (%)	SR (%)
Day 1	95.17	99.3	92.6	98.62	41.1	77.78
Day 2	94.56	98.87	88.78	95.76	31.72	74.87
Day 3	93.37	97.84	86.52	95.06	31	73.9

- GEOS-ML

Healthwise AQI Category	Unhealthy		Very Unhealthy		Critical	
Forecast Day	POD (%)	SR (%)	POD (%)	SR (%)	POD (%)	SR (%)
Day 1	100	100	100	97.19	63.1	52.38
Day 2	98.2	100	99.64	97.36	46.82	63.4
Day 3	98.12	100	96.25	96.77	14	51.27

- DM-Chem 1.5km

Healthwise AQI Category	Unhealthy		Very Unhealthy		Critical	
Forecast Day	POD (%)	SR (%)	POD (%)	SR (%)	POD (%)	SR (%)
Day 1	89.69	97	79.59	94.01	16.33	25.87
Day 2	89.39	98.48	80.43	96.1	16.14	26.11

- DM-Chem 330km

Healthwise AQI Category	Unhealthy		Very Unhealthy		Critical	
Forecast Day	POD (%)	SR (%)	POD (%)	SR (%)	POD (%)	SR (%)
Day 1	99.72	96.11	95.31	91.91	15.6	25.5
Day 2	99.11	96.96	15.6	25.44	10.82	17.3

- IFS

Healthwise AQI Category	Unhealthy		Very Unhealthy		Critical	
Forecast Day	POD (%)	SR (%)	POD (%)	SR (%)	POD (%)	SR (%)
Day 1	61.54	100	34.89	100	0	NAN
Day 2	60.55	100	39.29	100	0	NAN
Day 3	58.25	100	39.93	99.04	0	NAN

- SILAM

Healthwise AQI Category	Unhealthy		Very Unhealthy		Critical	
Forecast Day	POD (%)	SR (%)	POD (%)	SR (%)	POD (%)	SR (%)
Day 1	85.45	100	69.38	98.66	8.17	9.5
Day 2	87.34	99.78	72.68	100	6.29	7.3
Day 3	85.46	100	64.36	98.63	8.4	14.28

- GEOS-FP

Healthwise AQI Category	Unhealthy		Very Unhealthy		Critical	
Forecast Day	POD (%)	SR (%)	POD (%)	SR (%)	POD (%)	SR (%)
Day 1	50.73	100	23.06	96.05	0	NAN
Day 2	30.277	100	6.4	93.18	0	NAN
Day 3	12.76	100	1.53	100	0	NAN

- GEFS-Aerosols

Healthwise AQI Category	Unhealthy		Very Unhealthy		Critical	
Forecast Day	POD (%)	SR (%)	POD (%)	SR (%)	POD (%)	SR (%)
Day 1	12.18	100	0	NAN	0	NAN
Day 2	8.96	100	0	NAN	0	NAN
Day 3	9.28	100	0.6	55	0	NAN

3.3.3: Critical Category AQI Forecasts

Forecasting performance for the Critical AQI category remained challenging across all models and lead times. WRF-Chem was the most consistent, though its performance declined with forecast horizon. On Day 1, it achieved a POD of 51% and SR of 77.8%; by Day 3, those values declined to 31% and 73.9%, respectively. These results suggest fewer false alarms but increasing event misses over time.

GEOS-ML, which performed strongly in less severe categories, showed a marked performance drop: POD fell from 58% on Day 1 to 14% on Day 3, with a mean SR of approximately 51%, highlighting difficulty in capturing the most extreme pollution episodes.

The DM-Chem models showed some limited detection capacity. At 330 m resolution, the POD was 15.6% and SR 25.45% on Day 1, declining to 10.8% POD and 17.31% SR by Day 2. High FAR values exceeding 70% further underscored the unreliability of predictions for this category.
SILAM recorded POD values between 6.29% and 8.4%, with SR on Day 1 at 9.51%, suggesting frequent false alarms and limited detection ability.

Other global-scale models (GEFS-Aerosols, IFS, and GEOS-FP) consistently recorded zero POD for Critical AQI across all days, indicating a complete failure to detect any extreme pollution events.

3.4 Summary and Implications

In summary, while models like WRF-Chem, GEOS-ML, and DM-Chem exhibited high accuracy for Unhealthy and Very Unhealthy AQI categories, all models showed significant difficulty forecasting Critical events. Only a few models captured these severe episodes with modest skill, as reflected by low CSI values across the board. The results highlight the need for improved resolution, model physics, and data assimilation techniques to enhance prediction capabilities for extreme pollution conditions.

4. Conclusions

This report presents a detailed inter-comparison of several air quality forecasting models in predicting surface-level $PM_{2.5}$ concentrations and corresponding AQI categories over Delhi during the high-pollution months of October to January. Emphasis was placed on evaluating the models' statistical performance across three forecast days using metrics such as Root Mean Square Error (RMSE), Index of Agreement (IOA), Normalized Mean Bias (NMB), and Performance Index (PI).

Among the evaluated models, WRF-Chem consistently delivered the most accurate predictions, particularly on the first forecast day. It recorded the lowest RMSE of 66.13 µg/m³ and the highest IOA of 0.85, demonstrating a strong capacity to capture temporal variations in $PM_{2.5}$ levels. GEOS-ML and DM-Chem (330 m resolution) followed with moderate accuracy—recording RMSEs of 93.3 µg/m³ and 101.71 µg/m³, and IOAs of 0.54 and 0.59, respectively. While these models reflected reasonable agreement with observations, they trailed behind WRF-Chem in both error minimization and consistency. In contrast, global-scale models such as IFS and GEFS-Aerosols showed substantial underestimation of $PM_{2.5}$ concentrations. IFS reported an RMSE of 115.96µg/m³ with a large negative bias (NMB = −46%), while GEFS-Aerosols fared worse with an RMSE of 147.46 µg/m³ and NMB of −63%. These models struggled to resolve localized pollution events, a limitation attributed to their coarse spatial resolution and less refined treatment of urban emissions.

The Performance Index (PI) ranking further reinforced the superiority of WRF-Chem. On Day 1, it achieved a PI of 87, qualifying it as "Very Good" in forecast quality. GEOS-ML and DM-Chem (330 m) earned PI values of 70 and 69, respectively-placing them in the "Good" category. In contrast, global models such as GEOS-FP and GEFS-Aerosols registered PI values below 60, indicating poor overall performance. In terms of AQI category forecasting, WRF-Chem again led with high accuracy levels: 94.87% for the "Unhealthy" category and 93.05% for "Very Unhealthy" on Day 1. Accuracy dropped slightly to 86.75% in the "Critical" category, reflecting the broader challenge shared by all models in identifying extreme pollution levels. GEOS-ML displayed high accuracy for "Unhealthy" air quality but was less effective for "Critical" conditions, with reduced accuracy (79.82%) and an elevated false alarm rate (47.61%).

The overall findings highlight the value of high-resolution regional model like WRF-Chem for short-term urban air quality forecasting. The model's superior performance in both statistical and categorical metrics underscores the importance of spatial resolution and local emission representation in capturing the complex pollution dynamics of mega cities like Delhi. While significant progress has been made, persistent challenges—particularly in forecasting extreme pollution events—signal the need for further refinement. Future model development should prioritize enhancements in physical parameterizations, urban emissions data, and data assimilation techniques. High-resolution dynamic emissions inventory using real-time data from mobile platforms like drones and unmanned air vehicles may also be prioritized to further improve the forecasts of the extreme events. These enhancements will be essential in supporting robust, reliable, and actionable air quality forecasts for urban environments facing severe pollution episodes.

5. References

- Description and evaluation of GLOMAP-mode: a modal global aerosol microphysics model for the UKCA composition-climate model G. W. Mann, K. S. Carslaw, D. V. Spracklen, D. A. Ridley, P. T. Manktelow, M. P. Chipperfield, S. J. Pickering, and C. E. Johnson

- Emmons, L. K., and Coauthors, 2010: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4). Geosci. Model Dev., 3, 43–67, https://doi.org/10.5194/gmd-3-43-2010.

- Flemming, J., Benedetti, A., Inness, A., Engelen, R.J., Jones, L., Huijnen, V., Remy, S., Parrington, M., Suttie, M., Bozzo, A., & Peuch, V.H. (2017), The CAMS interim reanalysis of carbon monoxide, ozone, and aerosol for 2003–2015, Atmos. Chem. Phys., 17(3), 1945-1983

- Ghude, S. D., et al. (2020), Evaluation of $PM_{2.5}$ forecast using chemical data assimilation in the WRF-Chem model: A novel initiative under the Ministry of Earth Sciences Air Quality Early Warning System for Delhi, India, Curr. Sci., 118, 1803–1815.

- Govardhan, G., Ghude, S. D., Kumar, R., Sharma, S., Gunwani, P., Jena, C., Yadav, P., Ingle, S., Debnath, S., Pawar, P., Acharja, P., Jat, R., Kalita, G., Ambulkar, R., Kulkarni, S., Kaginalkar, A., Soni, V. K., Nanjundiah, R. S., and Rajeevan, M. (2024), Decision Support System version 1.0 (DSS v1.0) for air quality management in Delhi, India, Geosci. Model Dev., 17, 2617–2640, https://doi.org/10.5194/gmd-17-2617-2024.

- Govardhan, G., S. K. Satheesh, K. K. Moorthy, and R. Nanjundiah, 2019: Simulations of Black Carbon Over Indian Region: Improvements and Implications of Diurnality in Emissions. Atmos. Chem. Phys. Discuss., 1–25, https://doi.org/10.5194/acp-2019-152.

- Grell, G. A., S. E. Peckham, R. Schmitz, S. A. McKeen, G. Frost, W. C. Skamarock, and B. Eder, 2005: Fully coupled "online" chemistry within the WRF model. Atmos. Environ., 39, 6957–6975, https://doi.org/10.1016/j.atmosenv.2005.04.027.

- Guenther, A., 2007: Erratum: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature) (Atmospheric Chemistry and Physics (2006) 6 (3181-3210)). Atmos. Chem. Phys., 7, 4327, https://doi.org/10.5194/acp-7-4327-2007

- Gupta, P., Zhan, S., Mishra, V., Aekakkararungroj, A., Markert, A., Paibong, S., & Chishtie, F. (2021), Machine learning algorithm for estimating surface $PM_{2.5}$ in Thailand, Aerosol and Air Quality Research, 21(11), 210105. Aerosol and Air Quality Research, 11. https://doi.org/10.4209/aaqr.210105

- Guttikunda, S. K., et al. (2023), What Is Polluting Delhi's Air? A Review from 1990 to 2022, Sustainability, 15, 4209.

- Jayakumar, A., Gordon, H., Francis, T., Hill, A. A., Mohandas, S., Sandeepan, B. S., Mitra, A. K., and Beig, G. (2021), Delhi Model with Chemistry and aerosol framework (DM-Chem) for high-resolution fog forecasting, Q. J. R. Meteorol. Soc., 147(741), 3957-3978.

- Jayakumar, et al. (2025), Development of an integrated modeling framework for visibility and air quality forecasting in Delhi, Bull. Am. Meteorol. Soc., https://doi.org/10.1175/BAMS-D-24-0194.1.

- Jena, C., et al. (2021), Performance of high resolution (400 m) $PM_{2.5}$ forecast over Delhi, Sci. Rep., 11, 1–9.

- Jethva, H., Torres, O., Field, R. D., et al. (2019), Connecting Crop Productivity, Residue Fires, and Air Quality over Northern India, Sci Rep, 9, 16594.

- Kaiser, J. W., Heil, A., Andreae, M. O., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J.-J., Razinger, M., Schultz, M. G., Suttie, M., & van der Werf, G. R. (2012), Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, Biogeosciences, 9, 527–554, https://doi.org/10.5194/bg-9-527-2012.

- Keller, C.A., Knowland, K.E., Duncan, B.N., Liu, J., Anderson, D.C., Das, S., Lucchesi, R.A., Lundgren, E.W., Nicely, J.M., Nielsen, E., Ott, L.E., Saunders, E., Strode, S.A., Wales, P.A., Jacob, D.J., Pawson, S., 2021. Description of the NASA GEOS Composition Forecast Modeling System GEOS-CF v1.0. Journal of Advances in Modeling Earth Systems 13. https://doi.org/10.1029/2020MS002413

- Knowland, K. E., Keller, C. A., Wales, P. A., Wargan, K., Coy, L., Johnson, M. S., Liu, J., Lucchesi, R. A., Eastham, S. D., Fleming, E., and Liang, Q. (2022), NASA GEOS composition forecast modeling system GEOS-CF v1. 0: Stratospheric composition, J. Adv. Model. Earth Syst., 14(6), e2021MS002852, https://doi.org/10.1029/2021MS002852.

- Kulkarni, S. H., et al. (2020), How Much Does Large-Scale Crop Residue Burning Affect the Air Quality in Delhi?, Environ. Sci. Technol., 54, 4790–4799.

- Kumar, R., Ghude, S. D., Biswas, M., Jena, C., Alessandrini, S., Debnath, S., Kulkarni, S., Sperati, S., Soni, V. K., Nanjundiah, R. S., & Rajeevan, M. (2020), Enhancing accuracy of air quality and temperature forecasts during paddy crop residue burning season in Delhi via chemical data assimilation, J. Geophys. Res.-Atmos., 125, 1–16.

- Lan, R., Eastham, S. D., Liu, T., et al. (2022), Air quality impacts of crop residue burning in India and mitigation alternatives, Nat Commun, 13, 6537.

- Lelieveld, J., Klingmüller, K., Pozzer, A., Burnett, R. T., Haines, A., & Ramanathan, V. (2020), Effects of fossil fuel and total anthropogenic emission removal on public health and climate, Proc. Natl. Acad. Sci. U.S.A., 117(49), 30087-30094.

- Lucchesi, R., 2018. File Specification for GEOS FP (Forward Processing), NASA GMAO, File Specification for GEOS-5 FP (Forward Processing). National Technical Reports Library, NTIS, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland 20771.

- Molod, A., Takacs, L., Suarez, M., Bacmeister, J., Song, I.-S., Eichmann, A., 2012. Technical Report Series on Global Modeling and Data Assimilation (Technical Memorandum (TM) No. NASA/TM–2012-104606). National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland 20771.

- Mukhopadhyay, P., and Coauthors, 2019: Performance of a very high-resolution global forecast system model (GFS T1534) at 12.5 km over the Indian region during the 2016– 2017 monsoon seasons. J. Earth Syst. Sci., 128, 155, https://doi.org/10.1007/s12040-019-1186-6.

- Newman, P.A., Pawson, S., 2021. ATom: GEOS-5 Derived Meteorological Conditions and Tagged Tracers Along Flight Tracks. ORNL Distributed Active Archive Center, Oak Ridge, Tennessee, USA.

- Peuch, V. H., Engelen, R., Rixen, M., Dee, D., Flemming, J., Suttie, M., et al. (2022), The Copernicus atmosphere monitoring service from research to operations, Bull. Am. Meteorol. Soc., 103(12), E2650–E2668.

- Rienecker, M., Suarez, M.J., Todling, R., Bacmeister, J., Takacs, L., Liu, H.C., Gu, W., Sienkiewicz, M., Koster, R.D., Gelaro, R., Stajner, I., Nielsen, J.E., 2008. The GEOS-5 Data Assimilation System— Documentation of Versions 5.0.1, 5.1.0, and 5.2.0 (Technical Memorandum (TM) No. NASA/TM–2008–104606, Vol. 27). National Aeronautics and Space Administration, Goddard Space Flight Center Greenbelt, Maryland 20771.

- Soulie, A., Granier, C., Darras, S., Zilbermann, N., Doumbia, T., Guevara, M., Jalkanen, J.-P., Keita, S., Liousse, C., Crippa, M., Guizzardi, D., Hoesly, R., & Smith, S. J. (2024), Global anthropogenic emissions (CAMS-GLOB-ANT) for the Copernicus Atmosphere Monitoring Service simulations of air quality forecasts and reanalyses, Earth Syst. Sci. Data, 16, 2261–2279, https://doi.org/10.5194/essd-16-2261-2024.

- TERI & ARAI: Source apportionment of PM2. 5 and PM10 concentrations of Delhi NCR for identification of major sources, TERI and ARAI, https://www.teriin.org/sites/default/files/2018-08/Report_SA_AQMDelhi-NCR_0.pdf, 2018.

- Theethai Jacob, A., Jayakumar, A., Gupta, K., Mohandas, S., Hendry, M. A., Smith, D. K. E., et al. (2023), Implementation of the urban parameterization scheme in the Delhi model with an improved urban morphology, Q. J. R. Meteorol. Soc., 149, 40–60.

- Tiwari A, Soni VK, Jena C, Kumar A, Bist S, Kouznetsov R (2022) Pre-Operational Validation of Air Quality Forecasting Model SILAM for India. J Pollut Eff Cont. 10: 343.

- Venkataraman, C., and Coauthors, 2018: Source influence on emission pathways and ambient PM 2.5 pollution over India (2015-2050). Atmos. Chem. Phys., 18, 8017–8039, DOI 10.1175/BAMS-D-23-0181.1. https://doi.org/10.5194/acp-18-8017-2018.

- Wiedinmyer, C., S. K. Akagi, R. J. Yokelson, L. K. Emmons, J. A. Al-Saadi, J. J. Orlando, and A. J. Soja, 2011: The Fire INventory from NCAR (FINN): A high resolution global model to estimate the emissions from open burning. Geosci. Model Dev., 4, 625–641, https://doi.org/10.5194/gmd-4-625-2011.

- Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-Dimensional Variational Analysis with Spatially Inhomogeneous Covariances. Mon. Weather Rev., 130, 2905–2916, https://doi.org/10.1175/1520-0493(2002)1302.0.CO;2.

- Zhang, L., Montuoro, R., McKeen, S. A., Baker, B., Bhattacharjee, P. S., Grell, G. A., Henderson, J., Pan, L., Frost, G. J., McQueen, J., Saylor, R., Li, H., Ahmadov, R., Wang, J., Stajner, I., Kondragunta, S., Zhang, X., and Li, F.: Development and evaluation of the Aerosol Forecast Member in the National Center for Environment Prediction (NCEP)'s Global Ensemble Forecast System (GEFS-Aerosols v1), Geosci. Model Dev., 15, 5337–5369, https://doi.org/10.5194/gmd-15-5337-2022, 2022. https://doi.org/10.5194/gmd-15-5337-2022

6. Acknowledgements

This work would not have been possible without the support and collaboration of numerous institutions, data providers, funding agencies, and individuals who have contributed directly or indirectly to this research. We express our deep gratitude to the Indian Institute of Tropical Meteorology (IITM), Pune, for providing the necessary institutional support. In particular, we acknowledge the Pratyush High-Performance Computing System at IITM, which enabled the execution of computationally intensive simulations and analyses required for this study. We also extend sincere thanks to the Director of IITM for their continued encouragement and insightful guidance throughout the course of this work. The contribution of the Library and Information Processing (LIP) Division at IITM, Pune, is also gratefully acknowledged for facilitating access to critical resources and literature that supported the research activities.

We would like to thank the Commission for Air Quality Management (CAQM) in the National Capital Region and adjoining areas for its overarching efforts toward coordinated air quality management. The policies and frameworks laid down by CAQM provided a relevant context for this study and underscore the need for robust forecasting tools to support evidence-based decision-making.We are immensely thankful to the Central Pollution Control Board (CPCB) for making observational air quality data publicly available, which served as the benchmark for model validation. These data were accessed through the CPCB website: https://airquality.cpcb.gov.in.

We sincerely acknowledge the contributions of international organizations whose global models and data were integral to this study. We thank the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Copernicus Atmosphere Monitoring Service (CAMS) for providing access to global atmospheric composition forecast data, available through the CAMS data portal: https://ads.atmosphere.copernicus.eu. We express our gratitude to NASA for supplying GEOS-FP and GEOS-ML datasets, which were used extensively in the intercomparison analysis. These products are supported by NASA's Health and Air Quality Applied Sciences Team and the Satellite Needs Working Group. The GEOS-FP data can be accessed at https://gmao.gsfc.nasa.gov/GEOS/ and the GEOS-ML data are available at https://aeronet.gsfc.nasa.gov.

We also thank the National Centers for Environmental Prediction (NCEP) and the National Oceanic and Atmospheric Administration (NOAA) for providing the GEFS-Aerosols forecast data used in this study. Special appreciation is extended to the India Meteorological Department (IMD) for enabling access to the SILAM model data and for their ongoing contributions to national air quality forecasting systems. We are deeply grateful to the reviewer Dr. K.J Ramesh (Ex- DG IMD, Ex Member (Technical) at CAQM) for his meticulous evaluation, thoughtful comments, and insightful suggestions. His valuable input has played a crucial role in strengthening the clarity, depth, and overall quality of this work. Finally, we acknowledge the efforts of all researchers, model developers, and data managers whose dedication to atmospheric science and open data sharing has made this multi-model evaluation study possible. Their collective work supports the broader mission of improving air quality forecasting and safeguarding public health in pollution-affected regions such as Delhi.

7. Data Availability

The data employed in this multi-model inter-comparison report would be made available upon request to the lead authors.

List of Abbreviations:

3D- and 4D-VAR	Three-Dimensional and Four-Dimensional Variational Data Assimilation
3D-VAR	Three-Dimensional Variational Data Assimilation
4D-VAR	Four-Dimensional Variational Data Assimilation
AER	Aerosol Module (within IFS for simulating secondary aerosols)
AOD	Aerosol Optical Depth
AQEWS	Air Quality Early Warning System
AQI	Air Quality Index
ARAI	Automotive Research Association of India
ARL	Air Resources Laboratory (NOAA)
BASCOE	Belgian Assimilation System for Chemical Observations from the Environment
CAMS	Copernicus Atmosphere Monitoring Service
CAMS-GLOB	Copernicus Atmosphere Monitoring Service Global Emission Inventory
CAMS-GLOB-ANT	CAMS Global Anthropogenic Emission Inventory
CAQM	Commission for Air Quality Management
Cb05	Carbon Bond 2005 (chemical mechanism)
CBM5	Carbon Bond Mechanism 5 (gas-phase chemistry mechanism)
CEDS	Community Emissions Data System
CNNs	Convolutional neural networks
CO	Carbon Monoxide
COPD	Chronic Obstructive Pulmonary Disease
CPCB	Central Pollution Control Board
CSL	Chemical Sciences Laboratory (NOAA)
DMAT_SULPHUR	Sulfur oxidation scheme (specifically adapted for India)
DM-Chem	Delhi Model with Chemistry
DPCC	Delhi Pollution Control Committee
ECMWF	European Centre for Medium-Range Weather Forecasts
EDGAR	Emissions Database for Global Atmospheric Research
EMC	Environmental Modelling Center (NOAA)
EnKF / EnKS	Ensemble Kalman Filter / Ensemble Kalman Smoother
FINN	Fire INventory from NCAR
FIRMS	Fire Information for Resource Management System
FRP	Fire Radiative Power
FV3GFS	Finite-Volume Cubed-Sphere Global Forecast System
GBBEPx	Global Biomass Burning Emissions Product – Experimental

GEFS	Three-Dimensional and Four-Dimensional Variational Data Assimilation
GEFS-Aerosols	Three-Dimensional Variational Data Assimilation
GEOS-CF	Four-Dimensional Variational Data Assimilation
GEOS-FP	Aerosol Module (within IFS for simulating secondary aerosols)
GEOS-FP	Aerosol Optical Depth
GEOS-ML	Air Quality Early Warning System
GFAS	Air Quality Index
GLOMAP	Automotive Research Association of India
GMAO	Air Resources Laboratory (NOAA)
GOCART	Belgian Assimilation System for Chemical Observations from the Environment
GOES-ML	Copernicus Atmosphere Monitoring Service
GRAP	Copernicus Atmosphere Monitoring Service Global Emission Inventory
GSI	CAMS Global Anthropogenic Emission Inventory
GSI	Commission for Air Quality Management
GSL	Carbon Bond 2005 (chemical mechanism)
H_2O	Carbon Bond Mechanism 5 (gas-phase chemistry mechanism)
H_2O_2	Community Emissions Data System
HCHO	Convolutional neural networks
HNO_3	Carbon Monoxide
HONO	Chronic Obstructive Pulmonary Disease
HRRR-Smoke	Central Pollution Control Board
HTAP	Chemical Sciences Laboratory (NOAA)
IAU	Sulfur oxidation scheme (specifically adapted for India)
IFS	Delhi Model with Chemistry
IGP	Delhi Pollution Control Committee
IITM	European Centre for Medium-Range Weather Forecasts
IMD	Emissions Database for Global Atmospheric Research
MEGAN	Environmental Modelling Center (NOAA)
MetUM	Ensemble Kalman Filter / Ensemble Kalman Smoother
MODIS	Fire INventory from NCAR
MOZART-4	Fire Information for Resource Management System
NAAQS	Fire Radiative Power
NASA	Finite-Volume Cubed-Sphere Global Forecast System
NASA	Global Biomass Burning Emissions Product – Experimental
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NCMRWF	National Centre for Medium Range Weather Forecasting
NCR	National Capital Region

NEMS	NOAA Environmental Modeling System
Nh_3	Ammonia
No_2	Nitrogen Dioxide
NOAA	National Oceanic and Atmospheric Administration
No	Nitrogen Oxides (NO + No_2)
NRT	Near-Real-Time
NSF	National Science Foundation
NUOPC	National Unified Operational Prediction Capability
O_3	Ozone
Pm_{10}	Particulate Matter ≤10 µm diameter
$PM_{2.5}$	Particulate Matter with diameter ≤ 2.5 micrometers
SAFAR	System of Air Quality and Weather Forecasting And Research
SILAM	System for Integrated modelling of Atmospheric coMposition
So_2	Sulfur Dioxide
TERI	The Energy and Resources Institute
UCX	Unified Chemistry Extension
UFS	Unified Forecast System
UKCA	United Kingdom Chemistry and Aerosol (module)
VBS	Volatility Basis Set (for Secondary Organic Aerosols)
VIIRS	Visible Infrared Imaging Radiometer Suite
WHO	World Health Organization
WRF	Weather Research and Forecasting model
WRF-Chem	Weather Research and Forecasting model coupled with Chemistry
XAI	explainable AI

Annexure

Formulae for calculation of statistics

- **Mean Fractional Bias (MFB)**

$$MFB = \frac{2}{N}\Sigma\frac{P_i + O_i}{P_i - O_i}$$

Pi are modelled and Oi are observed concentrations
Mean Fractional Bias (MFB) is a measure of the systematic bias in predictions. Negative values indicate underestimation, while positive values indicate overestimation.

- **Normalized Mean Bias (NMB)**

$$NMB = \frac{\Sigma(P_i - O_i)}{\Sigma O_i}$$

Normalized Mean Bias (NMB) represents the average relative bias between the predicted and observed values. NMB close to 0% indicates minimal bias. Positive NMB shows overprediction, and negative NMB shows underprediction.

- **Root Mean Square Error (RMSE)**

$$RMSE = \frac{1}{N}\sqrt{\Sigma(P_i - O_i)^2}$$

Root Mean Square Error (RMSE) measures the standard deviation of prediction errors. Lower RMSE values indicate higher accuracy.

- **Index of Agreement**

$$IOA = 1 - \frac{\Sigma(P_i - O_i)^2}{\Sigma(|P_i - \bar{O}| + |(O_i - \bar{O}|)^2}$$

Index of Agreement (IOA) measures how well values match observed values predicted. It ranges from 0 (poor agreement) to 1 (perfect agreement). Higher IOA values indicate better model performance, where ō is mean of the observations.

Factor-of-2 Fraction (FAC2) Equation

$$FAC2 = (Number\ of\ data\ points\ satisfying\ 0.5 \leq (P_i/O_i) \leq 2) / (Total\ number\ of\ data\ points)$$

Provides fraction (0-1) of modelled & observed pairs meeting this criterion (Pi are modelled and Oi are observed concentrations); dimensionless statistic, not sensitive to outliers.

Pollutant-Specific Performance Index (PI) Equation-

$$PI\ [PM_{2.5}] = 100 * AVG\ (FAC2 + \mathrm{R} + (1 - |MFB/2|))$$

Where

- FAC2 = Factor-of-2 Fraction (measure of error or scatter)
- R = Correlation Coefficient (measure of linearity of relationship)
- MFB = Mean Fractional Bias (measure of bias or offset)

Legend	PI (%)
Excellent	[90,100]
Very good	[80,90]
Good	[70,800]
Acceptable	[60,70]
Poor	[50,60]
Very poor	<50

Table A 1: A contingency table along with equations used to compute various skill scores for different AQI forecast categories

Observations

Forecasted	YES	NO
YES	a	b
NO	c	d

Statistic name	What it measures	Education	Unit	How to interpret
Accuracy (A)	Forecasts that correctly predicted the event or non-event.	A=(a+d)/(a+b+c+d) *100	%	Higher numbers are better
False Alarm Rate (FAR)	The percent of times a forecast of high pollution did not actually occur.	FAR = (b/(a+b)) *100	%	Smaller values are best
Probability of Detection (POD) or Hit rate	Ability to predict high pollution events (i.e., the percentage of forecasted high pollution events that actually occurred).	POD = (a/(a+c)) * 100	%	Higher numbers are best
Critical Success Index (CSI), also called Threat Score	How well the high-pollution events were predicted. Useful for evaluating rarer events like high-pollution days. It is not affected by a large number of correctly forecasted, low pollution events.	CSI = (a/(a+b+c)) * 100	%	Higher numbers are best
Success Ratio (SR)	The proportion of predicted high-pollution events that actually occurred, complementing CSI.	FAR = (a/(a+b)) *100	%	Higher values indicate fewer false alarms.

भारतीय उष्णदेशीय मौसम विज्ञान संस्थान

(पृथ्वी विज्ञान मंत्रालय, भारत सरकार का एक स्वायत्त संस्थान)

डॉ. होमी भाभा मार्ग, पाषाण, पुणे–४११००८, भारत.

INDIAN INSTITUTE OF TROPICAL METEOROLOGY

(An Autonomous Institute of the Ministry of Earth Sciences, Government of India)

Dr. Homi Bhabha Road, Pashan, Pune - 411 008, India

https://www.tropmet.res.in/