Big Data Resources to Support Research Opportunities on Air Pollution Analysis in India

Sarath K. Guttikunda[1,2](✉)

[1] TRIP-C, Indian Institute of Technology, New Delhi, India
sguttikunda@urbanemissions.info
[2] Urban Emissions, New Delhi, India

Abstract. Most debates on air quality in India are (often) limited to big cities like Delhi, Mumbai, Kanpur, Pune, Hyderabad, and Kolkata, even though most of India's population lives in Tier-2, Tier-3, and smaller towns. There is little by way of local measurements for ground truthing or an assessment of sources contributing to air pollution problems in urban and rural areas or the growing health impacts associated with these pollution levels. The Air Pollution kNowledge Assessment (APnA) city program, launched in 2017, is an attempt to fill this lacuna of information, with an objective to create a baseline database for air pollution in Indian cities using open-access reanalysis data, satellite imagery, and satellite retrievals to inform policymakers as they evaluate the evolution of pollution and chart out strategies to improve air quality. This paper is based on the presentations delivered at two workshops - MAQTDS 2022, held as part of DASFAA 2022, and DCAAQ at BDA2021 - outlining an overview of air quality in India and opportunities for research to support air quality analysis using bigdata resources.

Keywords: India · Air quality · Bigdata · Satellite retrievals · NCAP · Air quality management · Emissions

1 Introduction – Air Quality in India

More than 50 Indian cities are ranked among the top 100 with the worst annual $PM_{2.5}$ averages, with Delhi taking the top spot among the capital cities worldwide in 2020 (https://www.iqair.com). Between 1998 and 2020 India's annual average $PM_{2.5}$ values have at least doubled [1]. On India's air quality index (AQI) scales, pollution levels over the Indo-Gangetic plain (IGP) moved from poor to very poor and severe conditions and the Central India region moved from moderate to poor conditions. At the administrative level, number of districts complying with India's annual ambient standard of 40 $\mu g/m^3$ dropped from 440 to 255 (out of 640 districts as per Census 2011) and number states dropped from 29 to 21 (out of 36, including union territories). Traditionally, these increases are observed over the cities. However, in the recent reanalysis databases which combine satellite retrievals, similar trends were observed over the rural areas. In these 23 years, total population complying with the annual ambient standard dropped from 60.5% to 28.4%, with most of this change coming from non-urban areas in IGP. In the

U. K. Rage et al. (Eds.): DASFAA 2022 Workshops, LNCS 13248, pp. 389–401, 2022.
https://doi.org/10.1007/978-3-031-11217-1_28

same period, the population exposed to poor, very poor, and severe AQI levels increased from 0.0% to 17.8%. In 2020, only a small portion of India's population lived in areas complying with World Health Organization (WHO)'s new guideline of 5 g/m^3.

According to the Global Burden of Disease (GBD) analysis, an estimated 1.2 million premature deaths in India can be traced back to exposure due to outdoor PM2.5 pollution levels [2]. According to GBD-Mapping Air Pollution Sources (MAPS) program, approximately 80% of the pollution originated from fossil fuel combustion and resuspended dust and the remainder coming from natural activities like sea salt, dust storms, and some agricultural activities [2, 3].

In 2019, India's Ministry of Environment Forests and Climate Change (MoEFCC) announced the National Clean Air Programme (NCAP) [4]. Under the programme, 132 non-attainment cities (i.e., cities that did not meet the annual ambient standard in 2017) were asked to prepare action plans to reduce their ambient PM2.5 pollution levels by 20–30% by 2024, compared to the pollution levels recorded in 2017. Individual cities have started to assimilate information on emissions and pollution loads to support the action plans.

In air quality management, general practice is to rely on monitoring data, which is basically snippets of information, both spatially and temporally. In India, there is an acute lack of ambient monitoring efforts in most cities to build a story just on that database. Using CPCB's own thumb rules, India requires 4000 stations across India and as of February 2022 there are only 340 in operation. Even the surveys and tests conducted to understand emissions are spread across temporally. For example, a pool of emission factors tests for vehicles was last conducted in 2010 as part of CPCB's 6-city study and there was one more round for a sample of vehicles in Pune in 2018. While these snippets of information are useful for ground truthing and expanding our understanding of the pollution loads and source strengths, the monitoring database needs to expand beyond the current capacity.

On the other side, we have the atmospheric modelling community, combining a larger pool of data from multiple resources including satellite retrievals and chemical transport models, helping us build patterns in emissions, pollution, and activity data, all in the hope of plugging the gaps in the monitoring data. Figure 1 presents a schematic of major components of air quality modelling. All of them are data intensive, computationally challenging, and require substantial personnel training to move forward from planning to execution. (a) Emissions modelling for both aggregate emissions and spatial/temporal allocations need a lot of data on source strengths, source locations, source emission control performance, and proxies for allocation of emissions at various scales (smaller the grid size, larger the need for proxies for finer distribution). (b) Meteorological modelling is streamlined with the existence of multiple global forecasting systems and agencies distributing the 3-dimensional fields to support scientific research and communications. For example, NASA's GFS and ESA's ECMWF. In India, one such system is maintained by the Indian Meteorological Department (IMD), which issues 10-day sub-regional forecasts and some feeds customized for fishery and farming communities. However, if the need is for meteorological data at a finer resolution, say 1-km over a city airshed, then downscaling models like WRF must be adapted, which require large computational capacity and personnel training. (c) Chemical transport modelling can vary in size and application depending on the requirements. Multiple models exist to accommodate these needs – models like CMAQ, CAMx, WRF-chem, GEOS-chem, and CHIMERE can help simulate multiple pollutants with full chemistry and evaluate the impacts of advection and chemistry at urban, regional, and global scales, including source apportionment; and models like inMAP and GAINS can help integrate chemical transport model results to evaluate scenarios and health impacts. (d) Validation is the central pillar of the whole modelling exercise, which is dependent on the monitoring data. There is no limit on data that can be used for validating and calibrating the modelling results, as long we have sample large enough to represent reality and represent the modelling domains spatially and temporally. (e) And finally, dissemination for public awareness and policy dialogue, which requires a completely different set of teams to take the message forward.

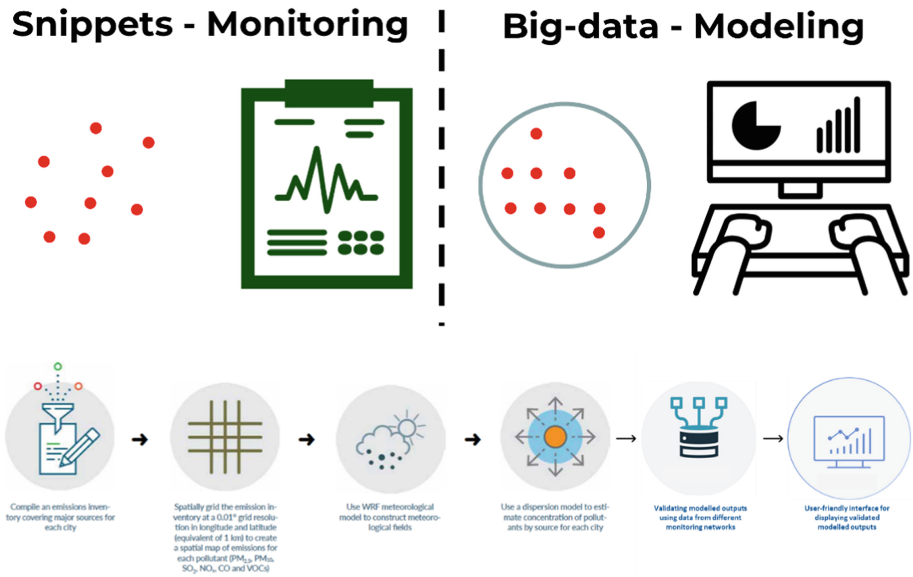


Fig. 1. Schematic of data management required for air quality analysis

While the kind of information gathered from monitoring and modelling exercises is different in shapes and sizes, both are integral pillars of an air quality management campaign, both needing snippets of information like surveys and pattern building from large (to very large) information databases. This paper presents a summary of research opportunities of using bigdata to study India's air quality.

2 Big Data Research

2.1 Support to Ambient Monitoring Efforts

As of February 2022, there are 340 continuous monitoring stations operating across India covering 174 cities with at least one station. Delhi (40), Mumbai (21), Bengaluru (10), Ahmedabad (8) and Pune (8) are few cities with multiple stations. The total monitors count translates to 0.25 per million population and in most cases is not a representative sample for regulatory and research grade pollution analysis [5, 6]. This density factor is the lowest among the big countries - China (1.2), the USA (3.4), Japan (0.5), Brazil (1.8) and most European countries (2–3). In addition to the continuous stations, CPCB also operates 800 manual stations to collect 24-h average pollution levels for up to 104 days in a year.

Meteorology, population, and human settlements databases were accessed to support the monitoring network design under NCAP, starting with determining a city's representative airshed. A city's airshed is determined using urban-rural classifications, landuse information, and an understanding of the known emission sources in the immediate vicinity of the city's administrative boundary. Human settlements layer is used to estimate the urban and rural shares of area and population in the city's airshed. The minimum number of sampling sites for each airshed is determined using the population information and protocols established by CPCB [7] and the sampling frequency is determined using the meteorological information (Table 1).

Table 1. Source and use case of open GIS databases

Field	Database	Design component
Meteorology	Weather Research and Forecasting (WRF) model with global inputs from NOAA's National Centres for Environmental Prediction (NCEP) [8] was used to build 3-dimensional meteorological fields, such as wind speeds, wind directions, temperature, relative humidity, pressure, precipitation, mixing layer heights, and surface threshold velocities at 1-h temporal resolution for base year 2018	# Sampling seasons
Population	Census-India database at the district level [9] and Landscan of Oakridge National Laboratory [10] were used to create 0.01° resolution population database for the city airsheds. The raw databases are available at 30 s spatial resolution	# Sampling sites
Global Human settlement (GHS)	GHS layer of Landsat satellite imagery was used to designate the city airshed grids and the gridded population as urban and rural [11]	# Sampling sites

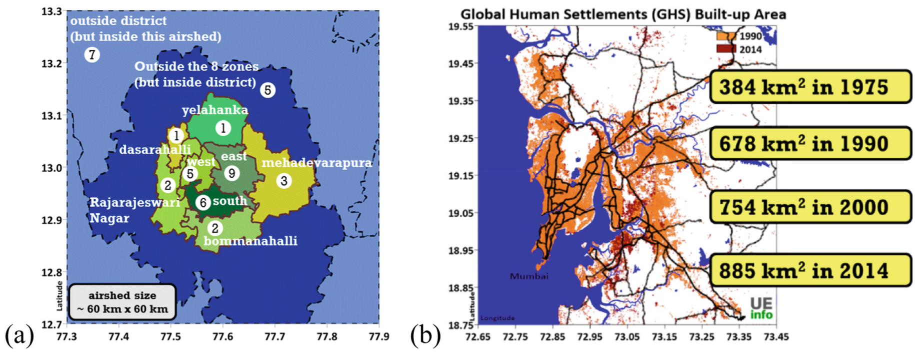


Fig. 2. (a) Allocation of monitoring sites by zones in Bengaluru (b) Expansion of built-up area in the Greater Mumbai airshed.

Overlapped with commercial activity density information in the form of number of hotels, hospitals, schools, parks, malls, markets, apartment complexes, industrial estates, worship sites, banks, eateries, fuel stations, and traffic spots like parking and stops, the recommended number of stations can be further assigned to a zone or a sub-district for better spatial representation – Fig. 2a presents an example for allocating 41 recommended stations in Bengaluru's 8 zones, peri-urban area surrounding the city administrative boundary, and the background rural areas in the airshed. The Human Settlement Built-Up maps can also be utilized to systematically shift the location of the monitoring sites as the city expands – Fig. 2b presents an example of expansion of the Greater Mumbai region, with the built-up area increasing from 384 km^2 in 1975 to 885 km^2 in 2014, which is a proxy for increasing demand for commercial and transport amenities, construction material, and together also increasing the demand for additional ambient monitoring for better representation of the activities.

2.2 Use of Satellite Retrievals

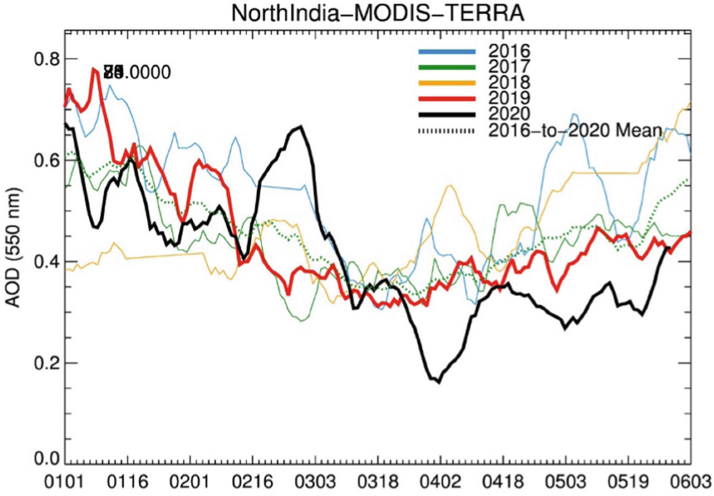
In 2020, COVID-19 lockdowns in March and April provided a glimpse into what is possible when the emissions are eliminated or reduced at all the known sources. In India, starting on March 24th, 2020, four lockdowns were announced (for 21, 21, 19, and 14 days), with the strictest regulations during the first lockdown period and slowly easing the restrictions by the end of the fourth. Thus resulting in better air quality across India with most improvements observed during the first period [12]. Lockdown periods featured the following regulations - (a) all the offices implemented work-from-home and all the schools, colleges, universities, training institutions, markets, malls, religious centers, and other public spaces were shut - this reduced most of the demand for passenger movement on the roads (b) all the shops and small-scale industries in the urban and rural areas were shut - with exceptions introduced after for essential food and medicine supply chains (c) all the construction activities were banned including brick manufacturing - this reduced the dust loading in the hotspots, debris movement, and construction freight movement (d) all the open waste burning activities were banned - this was possible since movement on the roads and inside/outside the residential communities was restricted

(e) all the passenger and public transport movement was stopped - with the exception of police, press, and medical practitioners and some with special permissions on a need for basis (f) all the freight movement was stopped on the highways and at the interstate border crossing - this was eased after the first week in response to supply shortages for essential goods in the cities (g) heavy industries (like power plants, refineries, fertilizers, cement, iron and steel, and other ore processing units) limited their operational times and fuel consumption loads, in response to a lower demand (h) road dust resuspension was at the minimum with reduced traffic on the roads and no construction activity.

While no primary surveys were conducted to ascertain these changes in the sectoral activity, the satellite observations provided before and after measurements to study this natural experiment in detail. The data from the ground monitors and the satellite retrievals also provided necessary data to study new research questions, which in the past would have been possible only in theory or lab experiments. Such as (a) impact of low emission densities on ozone photochemistry [13] (b) evaluation of NOx-VOC control regimes [14] (c) estimation of true background concentrations for cities, in the absence of all or most of the major emission sources [15].

The air quality during the lockdown periods is one example where the use of bigdata was demonstrated to explain the extreme lows and evaluation of daily trends. PM2.5 concentrations dropped across the country at the start of the lockdown periods and with every phase there is a marked increase in the average numbers was observed. On average, every lockdown period witnessed at least 25% drop across India, most (as high as 70%) coming from the cities. PM pollution is affected by all the known sources and all the regulations discussed above led to these drops [16, 17]. A climatological analysis of the satellite retrieval based AOD estimated a drop of 50% in the PM2.5 concentrations at the start of the lockdowns and slowing catching up to the decadal averages at the end of the 4th lockdown (see Fig. 3) [18]. The 4 lockdown periods ended on May 31st, 2020. Starting on June 1st, 2020, restrictions started to ease in phases, with individual states either continuing or easing them at their discretion. Similar databases are available for other pollutants – SO2, NO2, CO, HCHO, and Ozone, all of which can used to not only study the impact of chemistry and but also can be used to plug the gap in the monitoring efforts.

It is important to understand that the global models and satellite retrievals come with a lot of assumptions. While these broad insights are very helpful, limitations must be understood before applying these databases for public and policy use. Some of the limitations include (a) global models run at a coarse spatial resolution. For example, the reanalysis results presented in the introduction are from a model with 0.5-degree resolution, which cannot capture the core urban activities (b) the satellite retrievals used are just passes over India (a snapshot) and not geostationary with longer time stamps for a given day and (c) India's on-ground monitoring network is not wide enough to feed these models for representative calibration. Despite the limitations, satellite data retrievals during the COVID-19 pandemic and the associated lockdowns provided a new normal for Indian cities – a realization that "clean air" and "blue skies" is possible also during the times when it is not raining or windy. Some hard decisions are required to achieve such a reduction in the emissions at all the sources and a change in how the cities and regions manage air quality to sustain the benefits, and in this process bigdata can



(a)

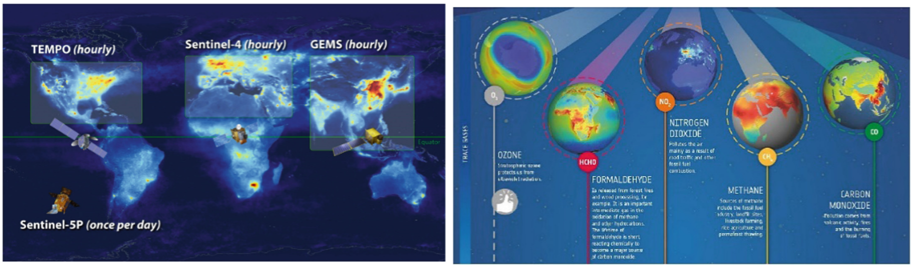


Fig. 3. (a) Satellite retrieval based AOD, averaged over North India for the lockdown period days between 2016 and 2020 [18] and [personal communication with co-author Dr Pawan Gupta, NASA] (b) Satellite clusters from NASA, ESA, and South Korea to support daily air quality analysis for multiple pollutants.

help nudge the change. Over Asia, GEMS system is expected to provide geostationary data for India in the coming years.

2.3 Use of Meteorological Data

Meteorology plays an integral part of pollution's ups and downs. It is responsible for the movement of emissions from source to receptor regions depending on the wind speeds and direction, for chemical evolution of the various pollutants within the gaseous phase and from gaseous to aerosol phase depending on the temperature, relative humidity, and pressure components which drive the chemical kinetics, and scavenging of pollution in the form of dry and wet deposition depending on wind speeds, landcover, and precipitation rates. While all these components are standardized in multiple chemical transport models, meteorology also plays a critical role in (a) emission modulation and (b) early pollution alert system.

The meteorological models like WRF can build high resolution lightning and dust storm emissions in forecasting and hindcasting mode. Both these sources are uncertain and depend not only on the model formulation to initiate these natural emissions, but also depend on multiple bigdata resources like landuse-landcover, seasonality in the soil moisture content, and cloud-cover information, all of which can be assimilated using multiple satellite products [19]. For example, NASA's MODIS satellite products include an 8-day ensemble landcover information and leaf area index, which is a direct input for dust-generation modules in WRF and biogenic-emissions generation module MEGAN [20].

Meteorological information is also useful in adjusting an emissions inventory in a dynamic mode. For example, precipitation rates at grid and hourly scale can be used to turn on or off resuspension emissions depending on a threshold. At the same time, soil moisture content after the rains can be used to decide when to turn on dust resuspension. For large scale dust storms over arid regions, formulation includes this if-then clause as a default. However, for urban scale assessments, where dust resuspension from on-road and construction activities is high, this correction must be linked to fine resolution meteorological data to adjust the emissions automatically, before the number is entered into the chemical transport model for further processing.

A similar correction can be employed in the space heating sector when the surface temperature and air temperature at 2 m drops below a threshold. In India, the during the winter months, surface temperature can drop to under 15 ˚C which triggers the need for heating. In most non-urban places, this demand is met by burning biomass, coal, and in some cases waste [21].

Summary of meteorological statistics for 2 cities is presented in Fig. 4 – Lucknow from North India and Hyderabad from South India (and summaries for 640 Indian districts is available at https://urbanemissions.info). For Lucknow, surface temperature is low for substantial number of hours during the winter nighttime, informing that the space heating emissions are an important part of Lucknow's inventory. For Hyderabad, while space heating is not a major component of its inventory, the wind directions alter significantly between the summer and the winter months, informing that the regional sources outside the city in the respective directions are important to track, since they have the right conditions to effect Hyderabad's air quality, as part of outside ("boundary") contributions at the chemical transport modelling stage [22, 23]. These deductions and dynamic adjustments are not possible from the use of just measurements at 1 or 10 locations in a city, but only possible when 3-dimensional high resolution meteorological modelling is conducted.

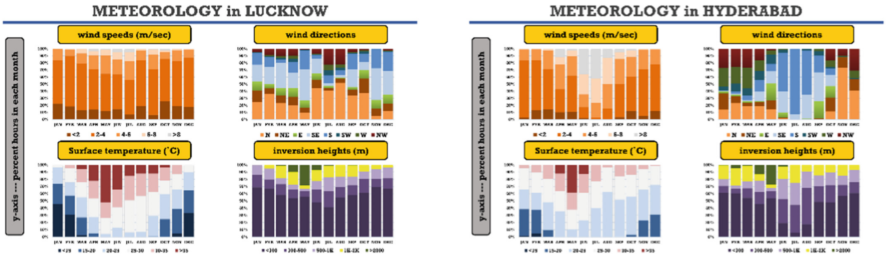


Fig. 4. Summary of wind speed, wind direction, temperature, and mixing heights as % hours in various bins in each month for 2 cities – Lucknow from North India and Hyderabad South India

2.4 Use of Google Earth Services

High resolution image processing is fast becoming an integral resource with easy access to multiple satellite feeds and algorithms to retrieve information for immediate use, not only among the air pollution modelling community, but also other areas like flood management and water resource management [24–26]. One free resource is Google Earth imagery, which has good spatial resolution to spot roads, landuse types, large industries, and large landfills. Below are examples of brick kilns spotted in Punjab and outside Mumbai, landfill in Mumbai, and rock quarries outside Pune (see Fig. 5).

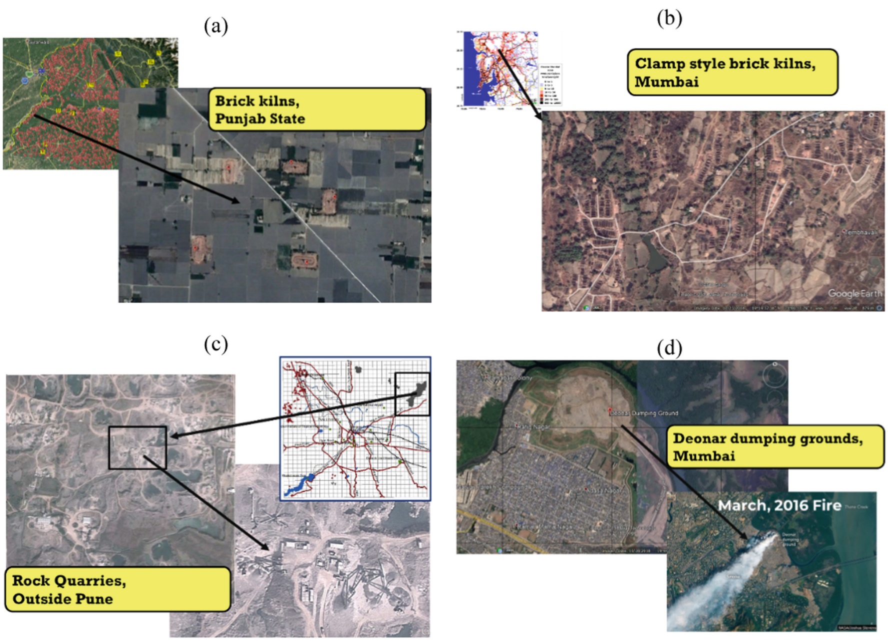


Fig. 5. Examples of activities spotted using Google Earth imagery (a) ~3000 fixed stack brick kilns in the state of Punjab (b) clamp style brick kilns outside Mumbai (c) ~11 km^2 of rock quarries outside Pune (d) ~2 km^2 landfill in Mumbai processing ~8000 tons per day waste from the city

While just the location and size of an entity is not enough to adjust the emissions, this information is useful when overlapped with other satellite products like aerosol optical depth, columnar NO_2 concentrations and columnar SO_2 concentrations, which can help deduce the source strengths. Knowing the sources around a city also helps in better storytelling of the air pollution problem, along with the confidence to say whether the source is officially part of the calculations or not.

For example, most of the rock quarries outside Pune are unofficial, running off-grid, and using engines likely banned in the city for crushing rocks and transporting locally. Most of the diesel used for this activity is not accounted in the official records of the city, but the emissions from the quarry activities will affect the air quality in the middle of the city.

For example, the clamp style brick kilns (showing outside Mumbai) are haphazardly placed around an area, with bricks piled along with biomass and coal mix to burn and bake. While the fixed stack kilns (showing in Punjab) are easy to spot, the clamp style kilns can only be mapped as an area and use it for back of the envelope calculations. This is the most inefficient way of manufacturing bricks and knowing where they are is the most useful information for air quality analysis and management.

2.5 Use of Google Maps Services

This is the only example which is a paid service. From the Google Maps distance API service, with each call, between 2 points, data can be extracted on total distance, total current time taken to travel (includes congestion times), total typical time taken to travel (with no congestion times), information on each of the segments (turns) along the way including segment distances. The base information from each of these calls is enough to estimate current average traffic speeds by grid (defined as ~1-km^2). For 30 Indian cities and non-Indian cities, such data was extracted and averaged at grid level for further processing (see Fig. 6).

Some applications include (a) development of speed profiles and congestion zones in the city to support urban transport planning (b) modulation of the vehicle emissions profile with average vehicle speed at grid level, such as higher CO and VOC emissions at speeds under 10 kmph to indicate incomplete combustion in the engines (c) dynamic adjustment of road dust resuspension, such as turning of resuspension when the grid speeds are under 10 kmph. The later 2 options can play a key role in altering photochemistry and ozone sensitivity to change in NO_x to VOC emission ratios and absence of dust particles for reactions.

2.6 Use of Open Street Maps (OSM) Database

An open and widely used resource is OSM database, which can provide several useful GIS layers information for most cities worldwide (https://download.geofabrik.de). For examples, roads (differentiating primary, secondary, motorable, and unclassified), railway lines, and commercial activity information in the form of number of hotels, hospitals, schools, parks, malls, markets, apartment complexes, industrial estates, worship sites, banks, eateries, fuel stations, and traffic spots like parking and stops. This is a crowd sourced database, so some level of ground truthing is advised before full use of the layers.

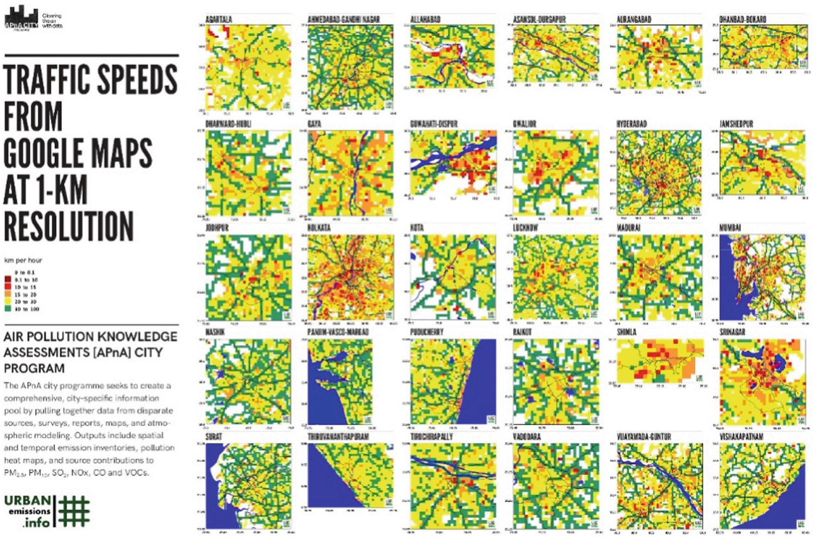


Fig. 6. Average traffic speeds at 1-km grid resolution for 30 Indian cities, extracted using the Google maps distance API service (Red colour indicates speeds under 15 kmph)

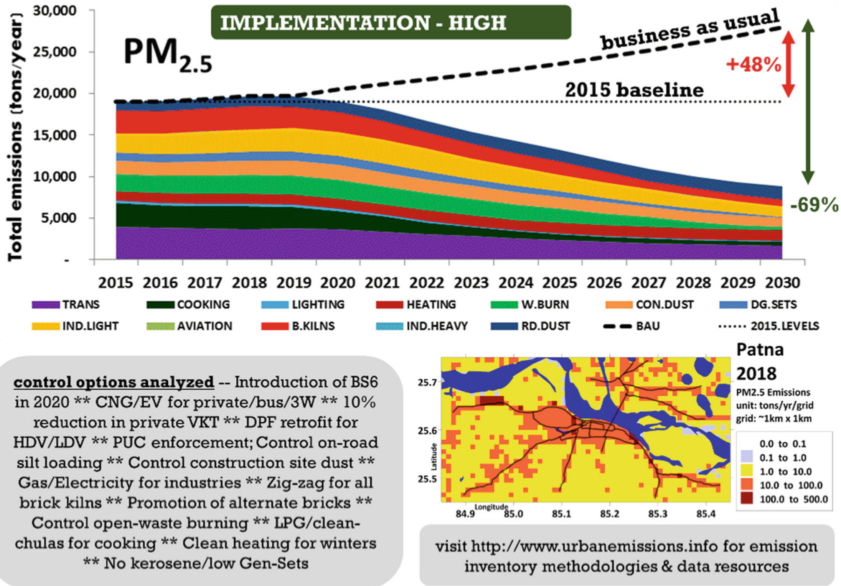


Fig. 7. Data integration to understand emission and pollution patterns and to support clean air action plans by evaluating their potential to reduce pollution (example: Patna under NCAP)

3 Conclusions

Air quality modelling through the stages of emissions, meteorology, and pollution, followed by public dissemination of the information generated is nothing short of art (see Fig. 1). At every stage, there is a lot of information (old and new) available in the public domain, which can be integrated to build defendable emissions and pollution maps to study "what-if" scenarios in support of clean air action plans. Figure 7 presents a summary of such an analysis conducted for the city of Patna – what will be impact of full implementation of actions listed under NCAP.

There is no second guessing that the ambient monitoring network must be expanded – not only in the cities, but also rural areas where similar growing trends are vividly visible. The monitoring data forms the basis for validating the bigdata. We also need more local level efforts to strengthen our understanding – this includes both bottom-up emissions and top-down source apportionment studies.

Acknowledgements. We would like to acknowledge and thank Dr Girish Agrawal for the invitation to present and submit this manuscript. This research received no external funding. The author declares no conflict of interest with the conference, conference organizers, and special issue editors.

References

1. van Donkelaar, A., et al.: Monthly global estimates of fine particulate matter and their uncertainty. Environ. Sci. Technol. **55**, 15287–15300 (2021). https://doi.org/10.1021/acs.est.1c05309
2. McDuffie, E.E., et al.: Source sector and fuel contributions to ambient PM2.5 and attributable mortality across multiple spatial scales. Nat. Commun. **12**, 3594 (2021). https://doi.org/10.1038/s41467-021-23853-y
3. Balakrishnan, K., et al.: The impact of air pollution on deaths, disease burden, and life expectancy across the states of India: the Global Burden of disease study 2017. Lancet Planet. Health **3**, e26–e39 (2019). https://doi.org/10.1016/s2542-5196(18)30261-4
4. Ganguly, T., Selvaraj, K.L., Guttikunda, S.K.: National Clean Air Programme (NCAP) for Indian cities: review and outlook of clean air action plans. Atmos. Environ. X **8**, 100096 (2020). https://doi.org/10.1016/j.aeaoa.2020.100096
5. Brauer, M., et al.: Examination of monitoring approaches for ambient air pollution: a case study for India. Atmos. Environ. **216**, 116940 (2019). https://doi.org/10.1016/j.atmosenv.2019.116940
6. Pant, P., et al.: Monitoring particulate matter in India: recent trends and future outlook. Air Qual. Atmos. Health **12**(1), 45–58 (2018). https://doi.org/10.1007/s11869-018-0629-6
7. CPCB. Guidelines for Ambient Air Quality Monitoring; Central Pollution Control Board, Ministry of Environment Forests and Climate Change, Government of India: New Delhi, India (2003)
8. NCEP. National Centers for Environmental Prediction. http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html. Accessed 15 Aug 2020
9. Census-India. Census of India 2011, The Governement of India, New Delhi, India (2011)
10. Rose, A.N.; McKee, J.J.; Urban, M.L.; Bright, E.A.; Sims, K.M.: LandScan 2018 (2019)

11. Pesaresi, M., et al.: GHS built-up grid, derived from Landsat, multitemporal (1975, 1990, 2000, 2014). European Commission, Joint Research Centre, JRC Data Catalogue (2015)
12. CPCB: Impact of lockdowns 25th March to 15th April on air quality (2020)
13. Kumar, A.H., Ratnam, M.V., Jain, C.D.: Influence of background dynamics on the vertical distribution of trace gases (CO/WV/O3) in the UTLS region during COVID-19 lockdown over India. Atmos. Res. **265**, 105876 (2022). https://doi.org/10.1016/j.atmosres.2021.105876
14. Rathod, A., Sahu, S.K., Singh, S., Beig, G.: Anomalous behaviour of ozone under COVID-19 and explicit diagnosis of O3-NOx-VOCs mechanism. Heliyon **7**, e06142 (2021). https://doi.org/10.1016/j.heliyon.2021.e06142
15. Beig, G., et al.: Towards baseline air pollution under COVID-19: implication for chronic health and policy research for Delhi, India. Current Sci. **119**, 00113891 (2020)
16. Gkatzelis, G.I., et al.: The global impacts of COVID-19 lockdowns on urban air pollution: a critical review and recommendations. Element. Sci. Anthrop. 9, 1–46 (2021). https://doi.org/10.1525/elementa.2021.00176
17. Ravindra, K., Singh, T., Biswal, A., Singh, V., Mor, S.: Impact of COVID-19 lockdown on ambient air quality in megacities of India and implication for air pollution control strategies. Environ. Sci. Pollut. Res. **28**(17), 21621–21632 (2021). https://doi.org/10.1007/s11356-020-11808-7
18. Sathe, Y., Gupta, P., Bawase, M., Lamsal, L., Patadia, F., Thipse, S.: Surface and satellite observations of air pollution in India during COVID-19 lockdown: implication to air quality. Sustain. Cities Soc. **66**, 102688 (2021). https://doi.org/10.1016/j.scs.2020.102688
19. Tinmaker, M.I.R., et al.: Relationships among lightning, rainfall, and meteorological parameters over oceanic and land regions of India. Meteorol. Atmos. Phys. **134**(1), 1–11 (2021). https://doi.org/10.1007/s00703-021-00841-x
20. Sindelarova, K., et al.: High-resolution biogenic global emission inventory for the time period 2000–2019 for air quality modelling. Earth Syst. Sci. Data **14**, 251–270 (2022). https://doi.org/10.5194/essd-14-251-2022
21. Chowdhury, S., Dey, S., Guttikunda, S., Pillarisetti, A., Smith, K.R., Di Girolamo, L.: Indian annual ambient air quality standard is achievable by completely mitigating emissions from household sources. Proc. Natl. Acad. Sci. USA **116**, 10711–10716 (2019). https://doi.org/10.1073/pnas.1900888116
22. Guttikunda, S.K., Nishadh, K.A., Jawahar, P.: Air pollution knowledge assessments (APnA) for 20 Indian cities. Urban Climate **27**, 124–141 (2019). https://doi.org/10.1016/j.uclim.2018.11.005
23. UEinfo: Air Pollution knowledge Assessments (APnA) city program covering 50 airsheds and 60 cities in India (2019). https://www.urbanemissions.info
24. Chithra, K., Binoy, B.V., Bimal, P.: Spatial mapping of the flood-affected regions of Northern Kerala: a case study of 2018 Kerala floods. J. Indian Soc. Rem. Sens. **50**, 677–688 (2021). https://doi.org/10.1007/s12524-021-01485-5
25. Goel, R., Miranda, J.J., Gouveia, N., Woodcock, J.: Using satellite imagery to estimate heavy vehicle volume for ecological injury analysis in India. Int. J. Inj. Contr. Saf. Promot. **28**, 68–77 (2021). https://doi.org/10.1080/17457300.2020.1837886
26. Lee, J., et al.: Scalable deep learning to identify brick kilns and aid regulatory capacity. Proc. Natl. Acad. Sci. **118**, e2018863118 (2021). https://doi.org/10.1073/pnas.2018863118