

SENAI, IIT MADRAS

# Data Science and IoT for Addressing Environmental Sanctity

Project *Kaatru*

# PROBLEM STATEMENT



## AWARENESS

We know surprisingly little about our surroundings.



## SPATIAL VARIATION

Ambient air quality can vary drastically between locations that are barely a few hundred meters apart.



## LIFE EXPECTANCY

3 million people across the globe die due to outdoor air pollution, of which India's burden is 1 million.



## QUALITY OF LIFE

Respiratory issues can be acquired and impact the wellbeing of individuals.



## EXPENSIVE

Existing air quality monitoring networks are sporadic and prohibitively expensive to expand.

1. I. Kheirbek, K. Ito, R. Neitzel, J. Kim, S. Johnson, Z. Ross, H. Eisl, T. Matte. Spatial variation in environmental noise and air pollution in New York City., J. Urban Health. 91 (2014) 415–431. doi:10.1007/s11524-013-9857-0.

2. World Health Organization. Ambient Air Pollution: A global assessment of exposure and burden of disease. World Health Organization. 1–131 (2016). Available at: doi:9789241511353



# ECONOMIC IMPACT OF AIR POLLUTION

## ILLNESS

Air Pollution contributes to 49% of asthma triggers; 30 million Indians suffer from asthma<sup>1</sup>.

## ANNUAL SPENDING

Indians spend INR 139.45 billion every year on asthma treatment(as of 2015)<sup>3</sup>

## COST OF TREATMENT

Asthma costs INR 44,095 per year per patient in India<sup>2</sup>.

## CONSEQUENCES

Poor air quality has an impact on child birth and mortality.

Patient reported asthma triggering factors (1)



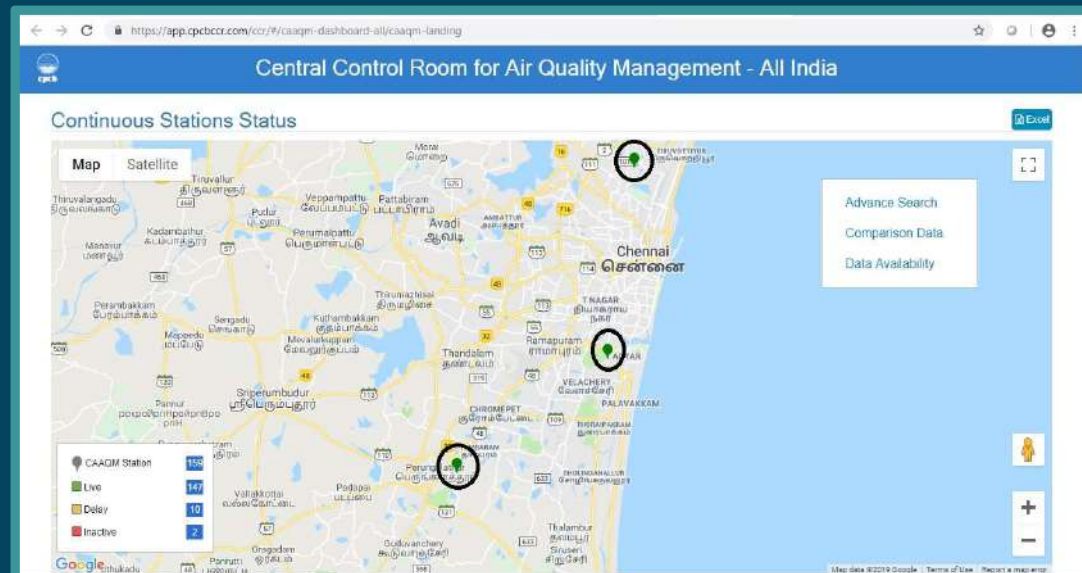
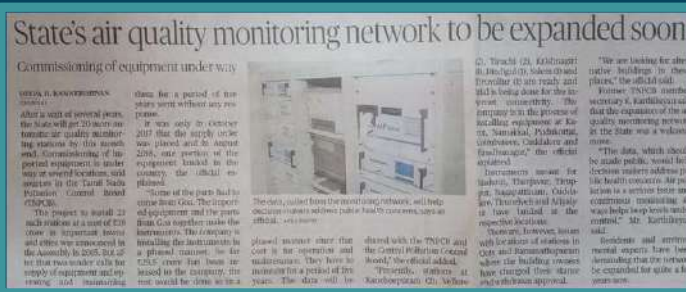
1. Asthma Insights and Management in India: Lessons Learnt from the Asia Pacific - Asthma Insights and Management (AP-AIM) Study; Sundeep S Salvi, Komalkirti K Apte, Raja Dhar, Pradeep Shetty, Rab A Faruqi, Philip J Thompson, Randeep Guleria, J. Accoc. Physicians of India, Sept. 2015
2. Economic burden of asthma among patients visiting a private hospital in South India; Surendran Aneeshkumar, Raj B Singh; Lung India, 2018
3. . Murthy KJ, Sastry JG. Economic burden of asthma. In: Burden of Disease in India. Background Papers. Available from: [http://www.who.int/macrohealth/action/NCMH\\_Burdenofdisease\\_%2829Sep%202005%29.pdf](http://www.who.int/macrohealth/action/NCMH_Burdenofdisease_%2829Sep%202005%29.pdf).



# EXISTING INFRASTRUCTURE



Conventional air quality monitoring station



Current distribution of CPCB owned monitoring stations across Chennai

Tamil Nadu's air quality monitoring network expansion plan:  
(The Hindu, June 19, 2019)  
**INR 39** crores to install **23** monitoring stations  
across **13** towns.



- <https://economictimes.indiatimes.com/news/politics-and-nation/noida-gets-its-first-automatic-air-ambient-monitoring-system/articleshow/58685064.cms?from=mdr>
- <https://app.cpcbcr.com/ccr/#/caaqm-dashboard-all/caaqm-landing>



# OBJECTIVES

- **Provide hyper-local environmental insights by combining information from multiple sources.**
- **Leverage Big Data Analytics, Data Science, Artificial Intelligence and IoT to develop:**
  - AI driven self calibrating high accuracy distributed sensing network
  - Intelligence for spatio-temporal hotspot identification
  - Country-wide pollution map
  - Personal exposure assessment (based on mode of transport)
  - Strategies for mitigation and abatement
- **Develop customized solutions for social and commercial impact.**
- **Develop high accuracy low cost sensing principles.**



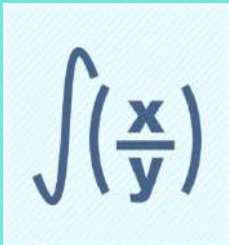
# COMPONENTS OF PROPOSED SOLUTION



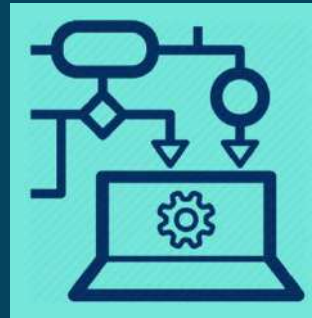
Satellite  
Remote  
Sensing



IoT | Sensor  
Network



Dispersion  
Models



Big Data and  
Data Science



# IoT | DISTRIBUTED MOBILE SENSOR NETWORK

A mobile, low cost sensor network is built by deploying IoT(Internet of Things) devices on vehicles. These devices, containing an array of sensors, act as agents for measuring multiple environmental parameters with high spatio-temporal resolution.

## Merits

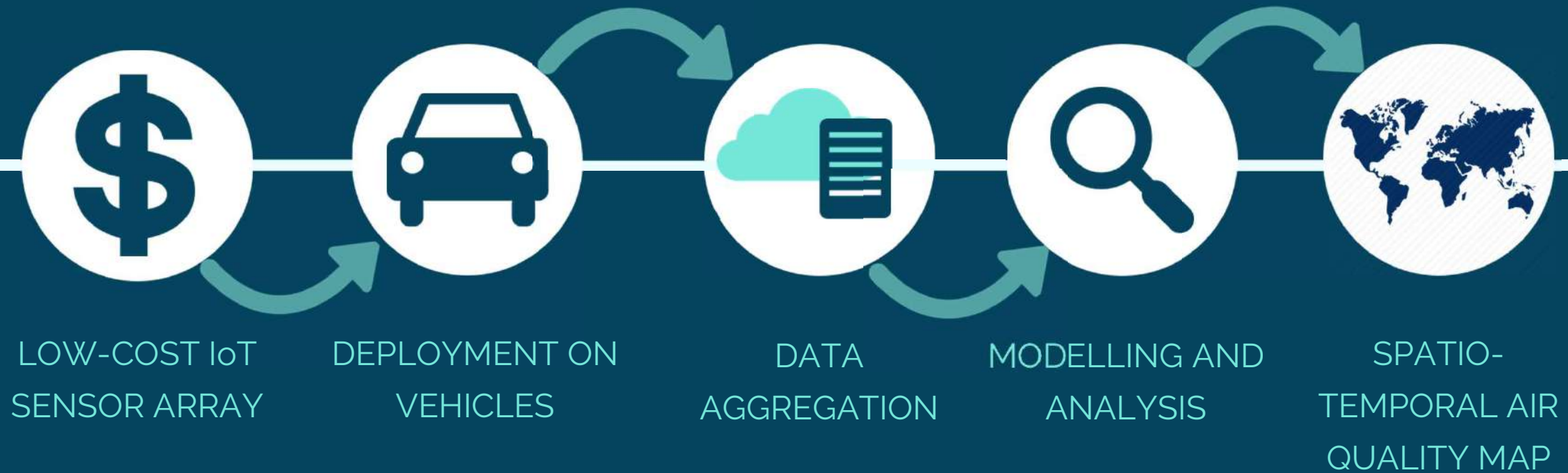
- Ground level information with high spatio-temporal resolution
- Real data from physical sensors

## Demerits

- Fidelity of low cost sensors is not established



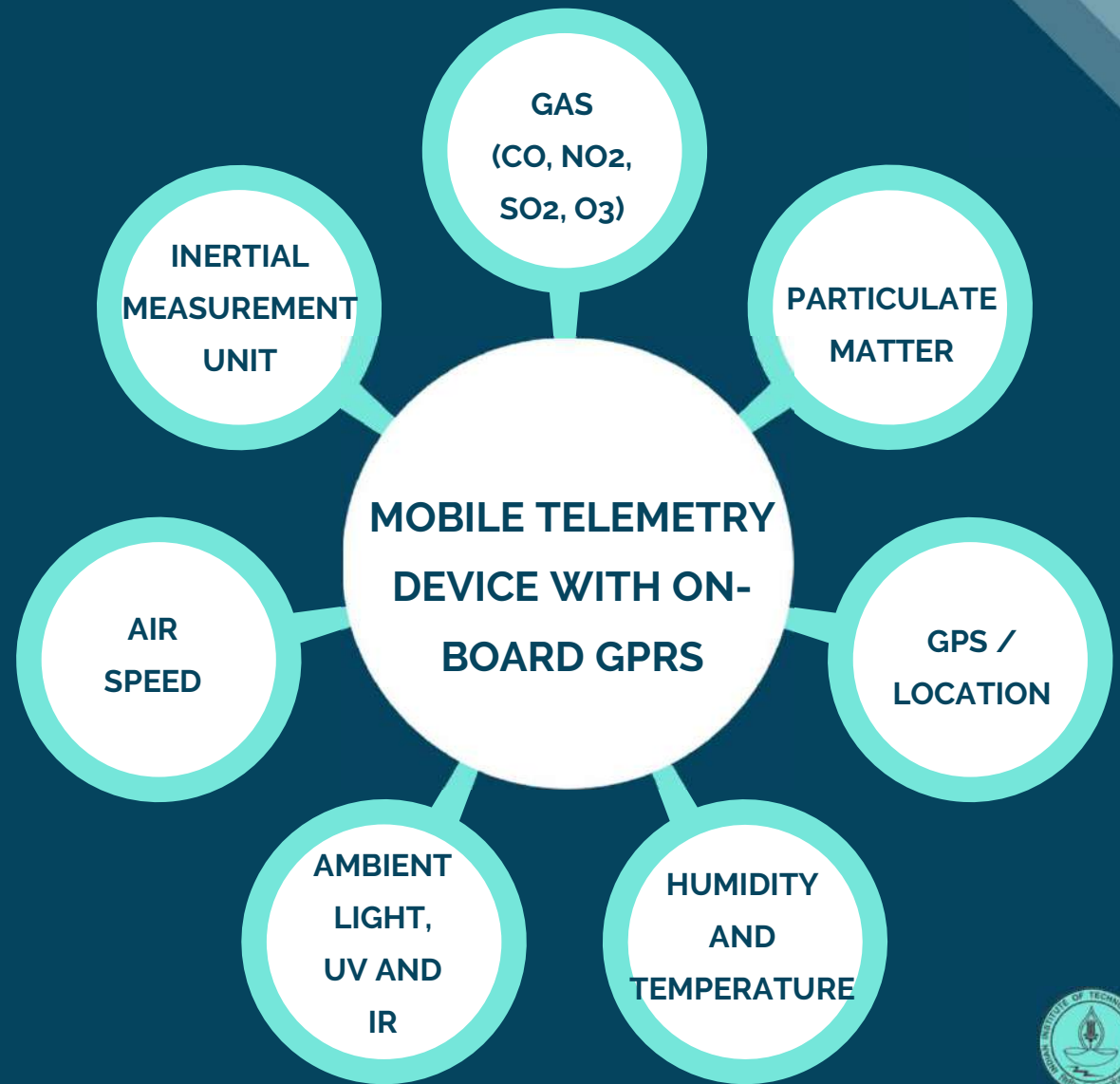
# SCHEMATIC OF SOLUTION





## SENSING DEVICE HARDWARE SPECIFICATIONS

- The hardware setup is a standalone, location aware telemetry device with bidirectional communication with a central server.
- 25 environmental parameters measured along with location and time
- Modular design of the hardware allows for integration of additional sensors based on demand



# SENSING DEVICES



Mobile device with active air sampling



Static device with passive air sampling

# DEVICE DEPLOYMENT



Test vehicle in Chennai carrying mobile sensing device on top



Mobile device fixed on top of E-Rickshaws in Gurugram



Mobile device on Motorbike in Gurgaon

# SATELLITE REMOTE SENSING

Data from satellites and satellite imagery products are used to estimate environmental parameters and determine land use patterns.

## Merits

- Extensive coverage

## Demerits

- Need ground measurements for calibration
- Low spatio-temporal resolution



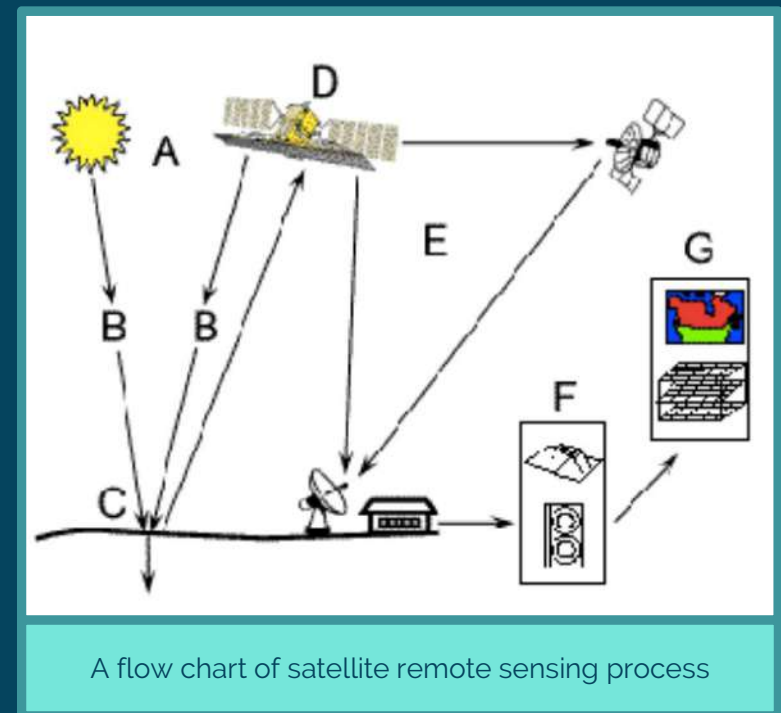
# FUNDAMENTALS OF SATELLITE REMOTE SENSING

- **OBJECTIVE:** To estimate spatial concentrations in vertical directions of atmosphere over a time period.

- **In the figure:**

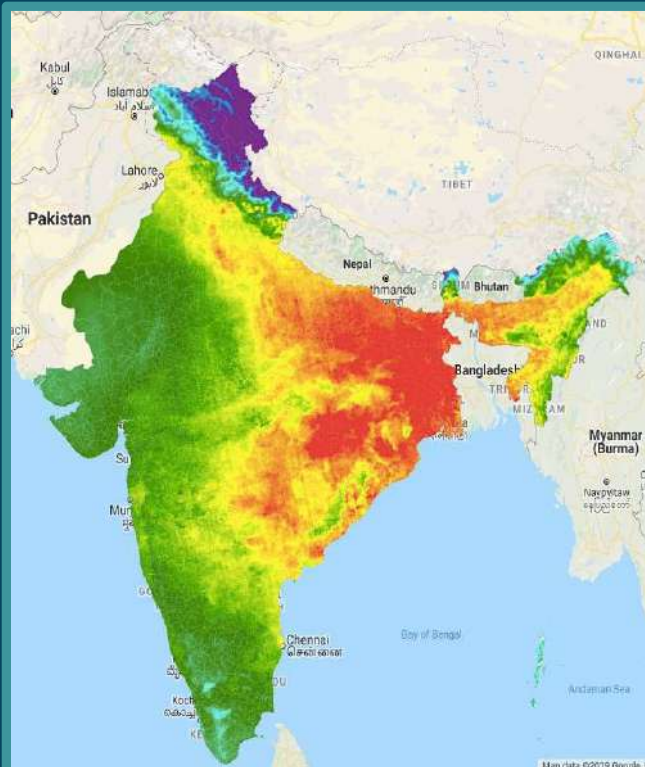
- A – Energy source / Illumination
- B – Radiation and Atmosphere
- C – Target interaction
- D – Sensor recording
- E – Transmission, Reception and Processing
- F – Interpretation and Analysis
- G – Application

- Our work comprises of F and G parts.

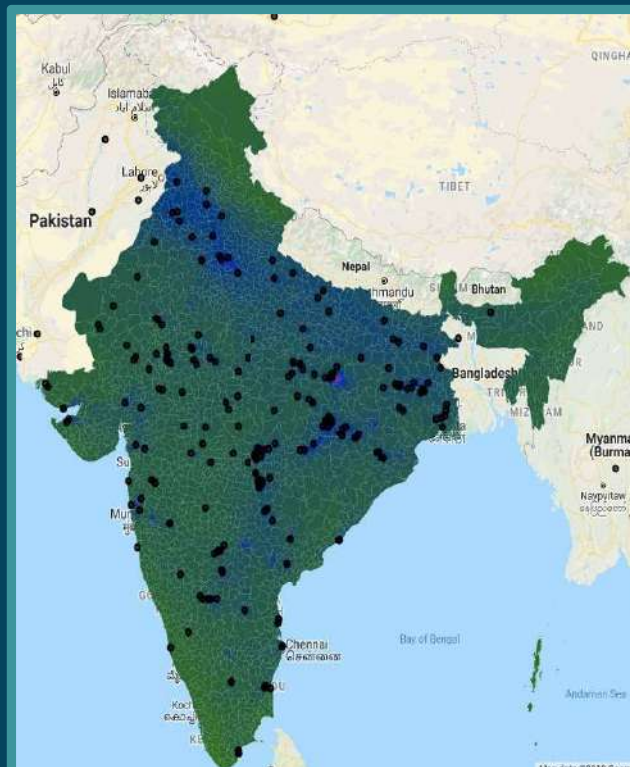




# GOOGLE EARTH ENGINE (GEE)



Carbon Mono Oxide (CO) Concentration,  
India, June 2019  
Red – highest concentrations, Blue – lowest



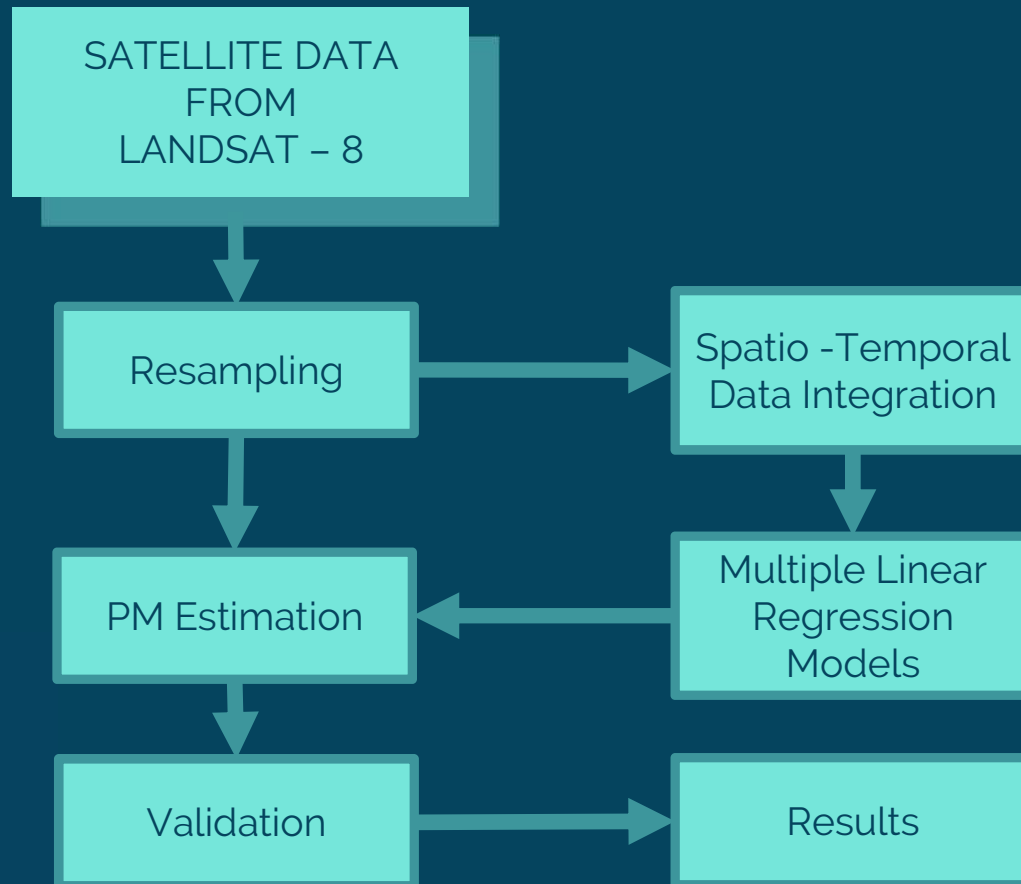
Nitrogen Di Oxide Concentration, India,  
June 2019  
Black dots – Thermal Power Plants

- **GEE** – An open source platform with catalog of satellite imagery and geospatial datasets as a platform for Earth science data and analysis.

- **Data source:**  
Sentinel-5P satellite



# PROPOSED APPROACH FOR ESTIMATION OF Particulate Matter (PM)



## Activities

- Creating pollution maps for various cities.
- Estimating average values for time period.
- Comparing results with that of sensor data and dispersion models.



# DISPERSION MODELLING

CFD/Transport principles would be used to model the dispersion of pollutions in macro environments. Multiscale agent based models will be developed as part of this work to estimate air pollutant concentration at different spatial and temporal scales.

## Merits

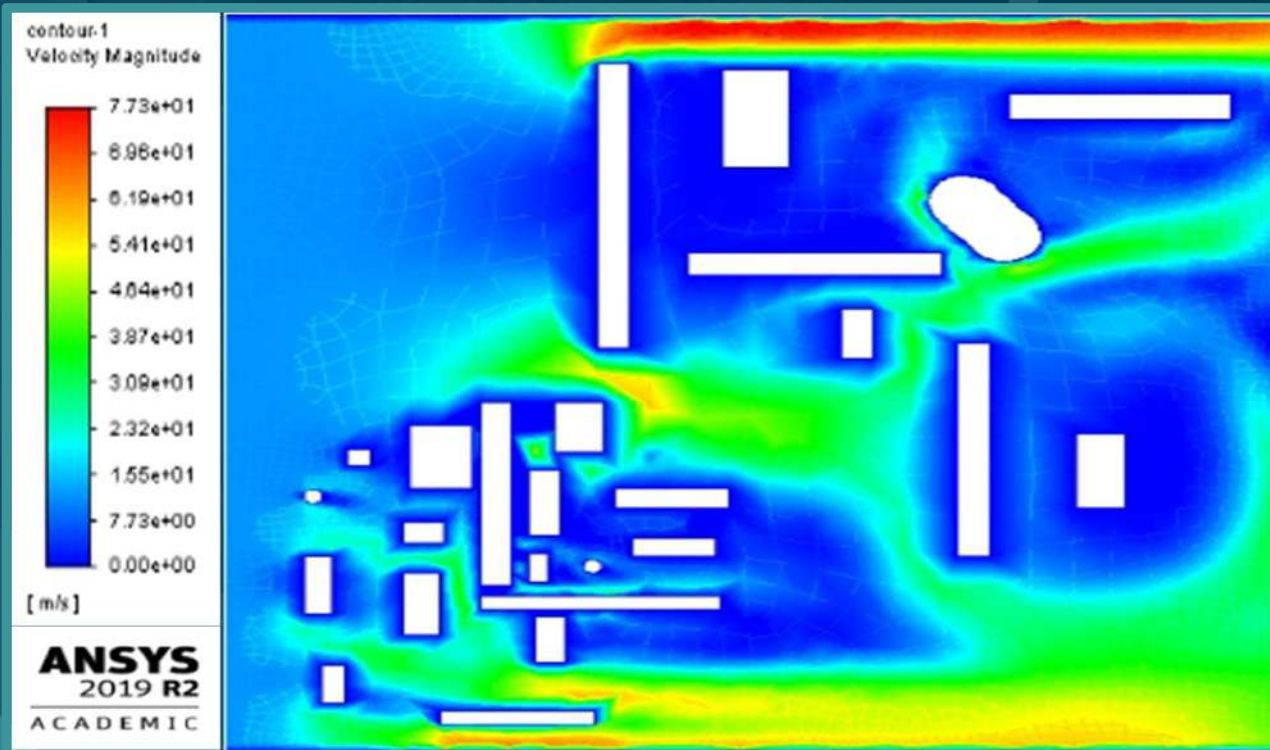
- Established method for air quality assessment.
- Models are scaleable in both space and time.

## Demerits

- Accuracy of the model is dependent on modelling approach and data source.
- Needs to be validated against ground measurements.



# DISPERSION MODELS



2-D analysis of velocity in a model city

- An urban environment presents many obstacles to dispersion of pollutants.
- Buildings, trees, walls, etc. can throttle or stagnate the flow of air.
- To understand transport behavior of pollutants, Air Dispersion Models are used





# CONVENTIONAL AIR DISPERSION MODELS

## GAUSSIAN DISPERSION MODEL

### MERITS

- Continuous emission model.
- Can capture overall, large-scale dispersion trend.

### DEMERITS

- Steady-state model.
- Suitable only for flat and open terrain.
- Dispersion parameters are assumed to be function only of velocities.

## LAGRANGIAN PUFF MODEL

### MERITS

- Simulates continuous puffs of pollutants.
- Unsteady state analysis possible.

### DEMERITS

- Puff diffusion is Gaussian.
- Suitable only for long distance transport of pollutants.

## CFD BASED MODEL

### MERITS

- Fully resolves transport equation.
- Complex geometries and obstacles can be taken into account.

### DEMERITS

- Introduces first and second order errors.
- Computationally expensive for complex geometries.

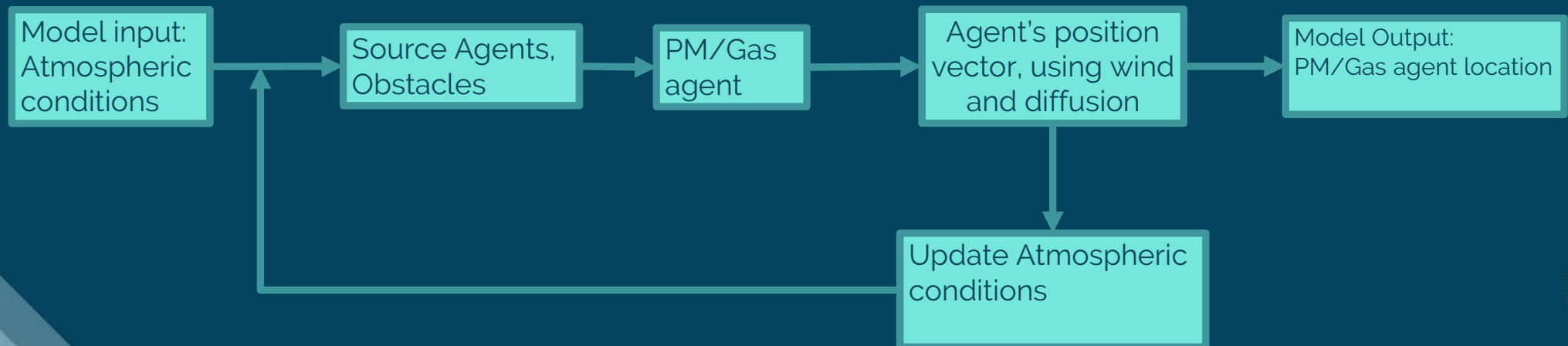


# PROPOSED AIR DISPERSION MODEL

## Agent Based Modeling (ABM):

- Models pollutants, surroundings and environment to be distinct agents.
- Agents follow set of predefined rules.
- Set of predefined rules define interactions between different agents .
- Captures Lagrangian perspective as description of agents and the Eulerian perspective as the description of environment.
- Low computational costs

## Use of ABM in air Dispersion:

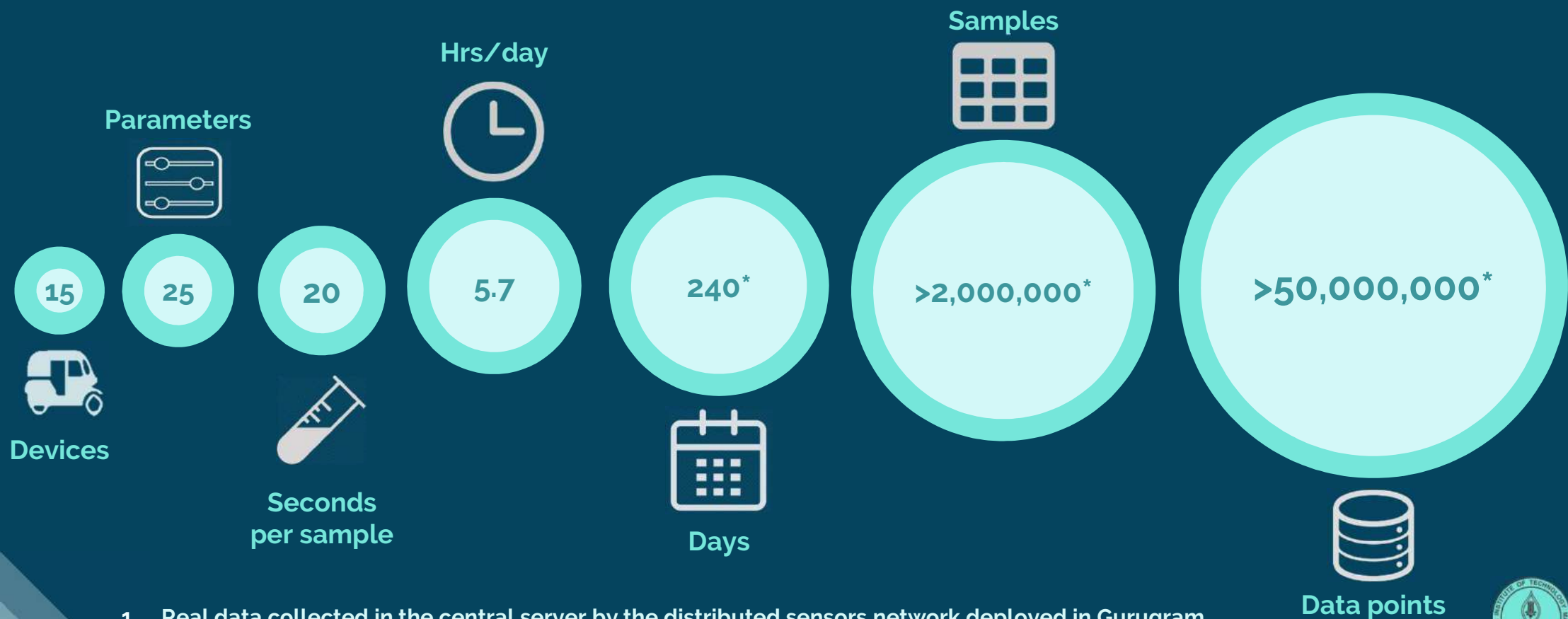


# BIG DATA AND DATA SCIENCE

- Big data and data science are the two cornerstones of this project.
- Raw environmental data collected from multiple sources are collated, processed and analyzed to obtain meaningful information.
- The enormous quantities, variety and velocity of data collected necessitates the need to incorporate a Big Data framework to manage and process data effectively.
- A variety of data science tools and techniques along with custom Machine Learning algorithms developed in-house are used to generate insights and solutions.
- Custom packages for curation and analysis of spatio-temporal data will be developed under the scope of this project.



# OVERVIEW OF DATA COLLECTED<sup>1</sup>



1. Real data collected in the central server by the distributed sensors network deployed in Gurugram and Chennai, India, from February 2019 through November 2019.

\* The sensor network continues to collect data



# CASE STUDY 1 – AMBATTUR, CHENNAI

- A pilot study was run in Ambattur – Moggapair – Korratur – Padi, areas in Chennai with one mobile device fitted on a van.
- The route was chosen such that the vehicle would travel 25 kilometres in each trip covering an effective area of roughly 13 sq. km (~5 x 2.5 km).
- Data collected over 30 days of study was condensed into a 24 hour frame for analysis.



Route covered by the vehicle during pilot study along with key locations.

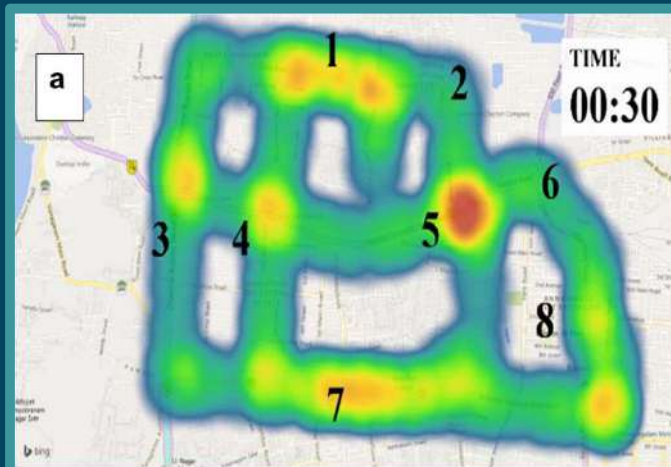


# CASE STUDY 1 – AMBATTUR, CHENNAI

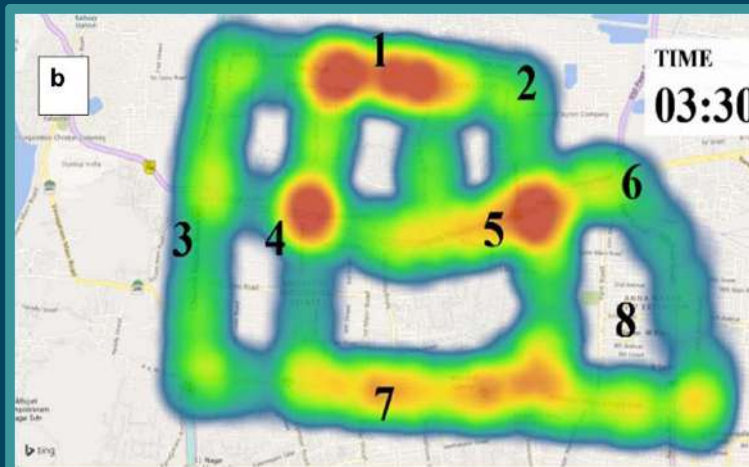
PM2.5(ug/m3) (No Aggregation)

19.00  
or Less

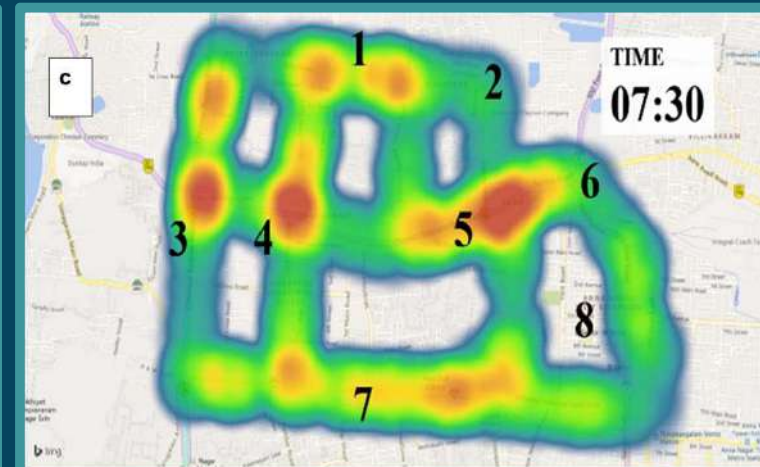
151.03  
or More



At 0030 hours there is very little vehicular pollution. There is however some traces of residual PM 2.5 at location 5. Location 5 is a narrow traffic junction with tall buildings around. This prevents particulate matter from dispersing and thus causing them to accumulate in the region.



At 0330 hours, there is significant spike in PM2.5 concentration at location 1; entrance to the state owned milk factory. This is attributed to the outflow of trucks that leave the factory to deliver the morning batch of milk.



At 0730 hours, PM2.5 concentration at location 1 has reduced, coinciding with the end of milk dispatch time window and in turn a decline in the number of trucks leaving the factory. However, the aftereffects of the trucks' movement can be seen on adjacent roads.



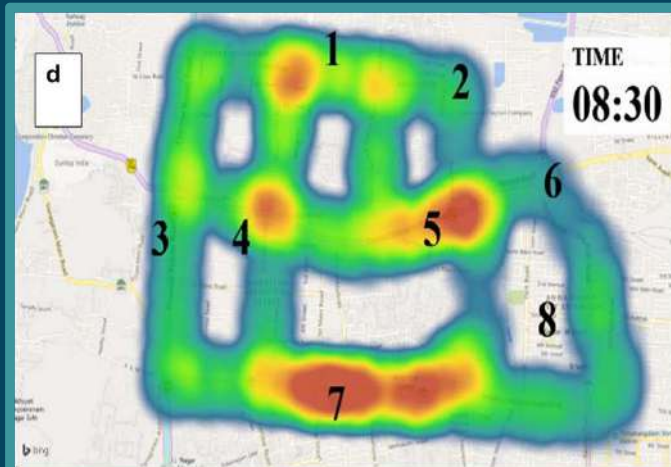


# CASE STUDY 1 – AMBATTUR, CHENNAI

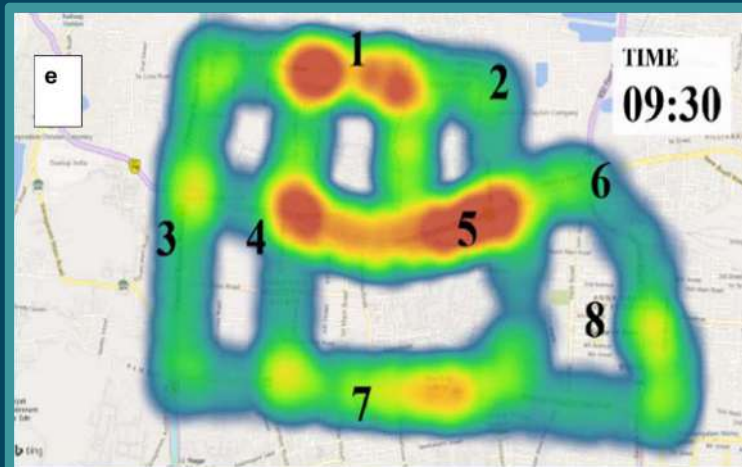
PM2.5( $\mu\text{g}/\text{m}^3$ ) (No Aggregation)

19.00  
or Less

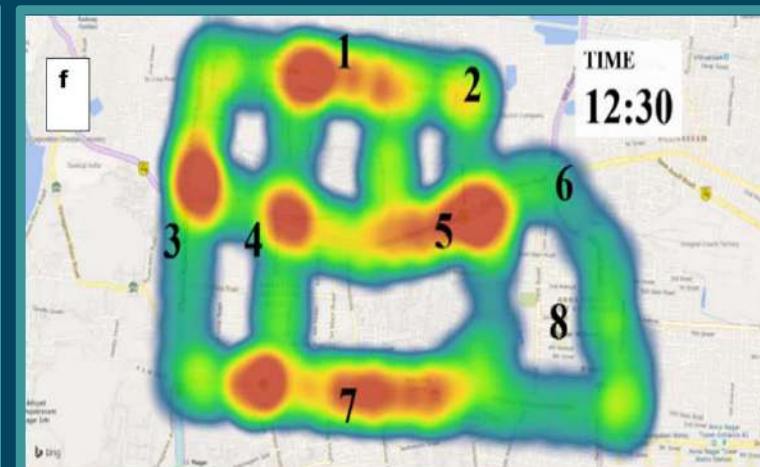
151.03  
or More



At 0830 hours, one can notice the sudden spike in PM2.5 concentration at location 7, i.e. the school area. It coincides with the school start time when students are dropped at the school in the morning. There are about 15 schools comprising both junior and senior sections within a 500 meter radius around the school area.



At 0930 hours, there is a drastic shift in the PM2.5 concentration away from the school zone and towards the road leading to the Ambattur Industrial Estate (locations 4 and 5). This shift is attributed to the reduction in traffic at the school zone while there's an increased traffic of employees working in the Ambattur industrial estate and adjacent areas. It should be noted that the PM concentration in school area changes within an hour.



At 1230 hours, another significant change is observed when there is a higher PM2.5 concentration at multiple locations. The spike at location 7 is attributed to the dismissal time for the junior sections of the school. By this time, regular traffic has picked up in locations 3, 4, and 5.

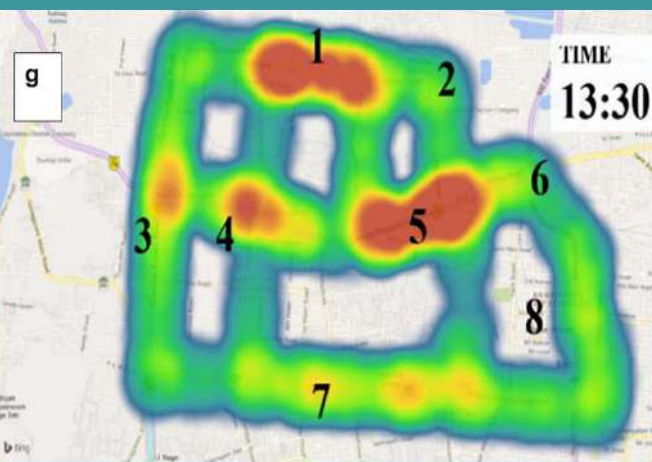


# CASE STUDY 1 – AMBATTUR, CHENNAI

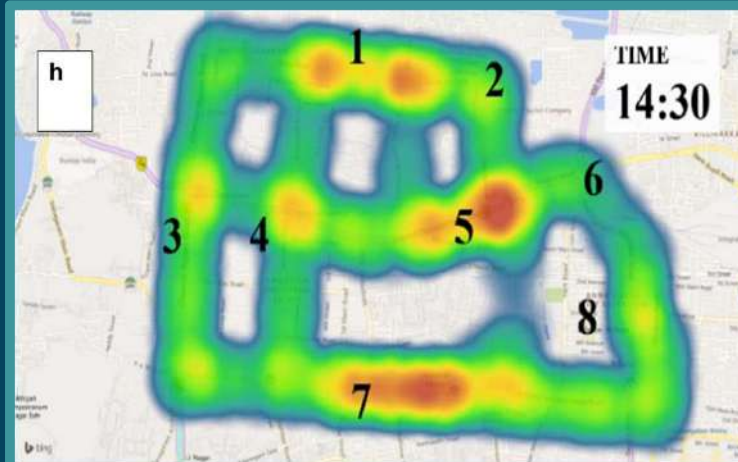
PM2.5(ug/m3) (No Aggregation)

19.00  
or Less

151.03  
or More



At 1330 hours, PM concentration near the school area (location 7) drops significantly within a short duration again. However, there is slight redistribution in the PM2.5 concentration in locations 3, 4, and 5.



At 1430 hours, the PM concentration spike once again at the school zone. This coincides with the dismissal time for the senior sections in school. It is worth noting that the PM2.5 concentration around location is really dynamic. This could be attributed to the relatively wider road at location 7, lack of tall buildings, and lesser traffic in general other than during school times. Another key point to be noted is that there is drop in the traffic in all the other segments which is captured by the lower PM concentration in those areas.

It is worth noting that the residential areas, 2 and 8 are generally cleaner throughout the day than the rest of the area. This could be attributed to the tree cover and lack of traffic.

There is very little to no PM2.5 accumulation in locations 3 and 6. This could probably be attributed to the availability of ample open area for the PM2.5 particles to disperse unlike locations 4 and particularly 6 which is one of the most congested parts in the study area.

Location 3 is the Chennai Bypass road which is a 6-lane expressway with very few buildings around and mostly open land on one side of the expressway. Location 6 is a flyover and the height of this location coupled with lack of tall buildings or trees, makes more room available for dispersion of PM2.5 particles.



# COUNTRY-WIDE POLLUTION MAP

- Machine learning algorithms are used to project air quality information across the entire country
- Low cost sensor network is deployed in reference cities.
- A list of features containing land use information and other city specific information is collected.
- Data from sensor network is mapped to the cities' features.
- This information is used in a machine learning framework to estimate air quality in cities across the country.



# PERSONAL EXPOSURE ASSESSMENT

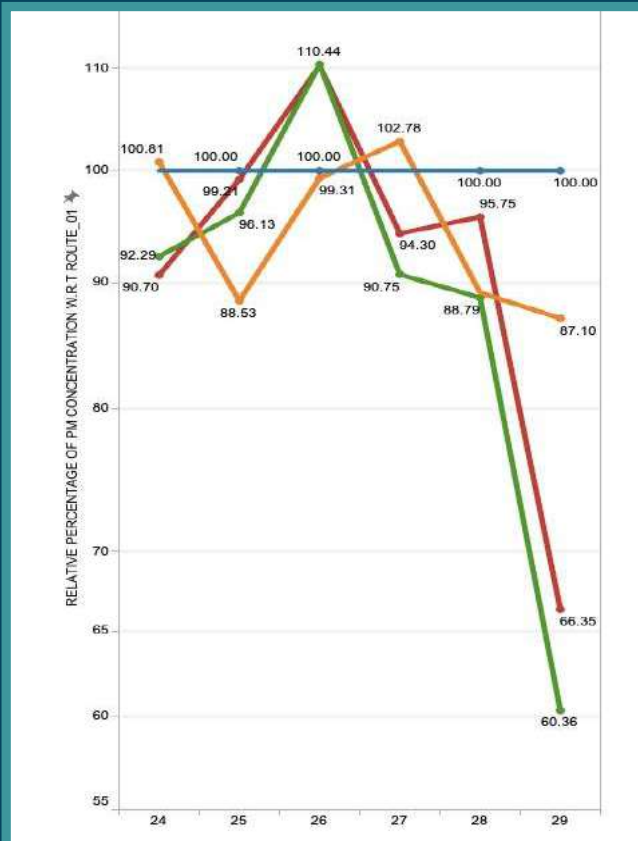
- Given hyperlocal air quality information is available, the next step is estimate exposure to individuals.
- Formulations for personal exposure would developed based on the following information:
  - Time activity of individuals
  - Exposure as a factor of mode of transport
  - Air quality along the route of travel as a function of space and time
  - Indoor air quality
- Ability to provide dynamic predictions of cleanest routes in real time.





# SAFEST ROUTE

- There exists variation between multiple routes between the same two locations.
- Based on data collected during Diwali of 2019, we compare PM<sub>2.5</sub> concentrations across the routes to identify the safest route
- Velachery Routes during Diwali 2019 (R):
  - 1 - Bypass road
  - 2 - A small detour through residential
  - 3 - Main road
  - 4 - Commercial + Residential route
- 27 Oct 2019- Diwali, more pollution estimated in residential areas.
- R2 is fairly better than R1.
- Pre-Diwali (26 Oct) estimate is significantly higher due to commercial activity and bursting crackers.



Relative percentage of PM<sub>2.5</sub> concentrations across the 4 routes in Velachery over 6 days in October 2019



4 routes in Velachery area chose for study





# STRATEGIES FOR MITIGATION AND ABATEMENT

- Develop actionable insights from the data
- Identification and classification of sources of pollution
- Suggest strategies for mitigation and abatement
- Study the effect of city planning initiatives on air quality.



## SUMMARY

- The project proposes to provide hyper-local environmental information.
- Low cost mobile distributed sensor network is used to obtain high resolution ground level information.
- Satellite remote sensing data is used to augment sensor network.
- Theoretical models are used to obtain high resolution data over larger areas.
- Big data, machine learning and artificial intelligence is used to combine data from all sources to obtain high resolution, accurate and reliable air quality information.
- Preliminary results from multiple pilot projects validate the efficacy of the technology.
- Customized solutions for impact are developed



# FUNDING SUPPORT



# CONTACT

**Prof. Raghunathan Rengaswamy**

MSB 150, Mechanical Sciences block  
Department of Chemical Engineering  
IIT Madras, India.

E-mail: [raghur@iitm.ac.in](mailto:raghur@iitm.ac.in)

**SENAI Lab:**

MSB 147, Mechanical Sciences block  
Department of Chemical Engineering  
IIT Madras, India.

Website:

<https://che.iitm.ac.in/~senai/>

