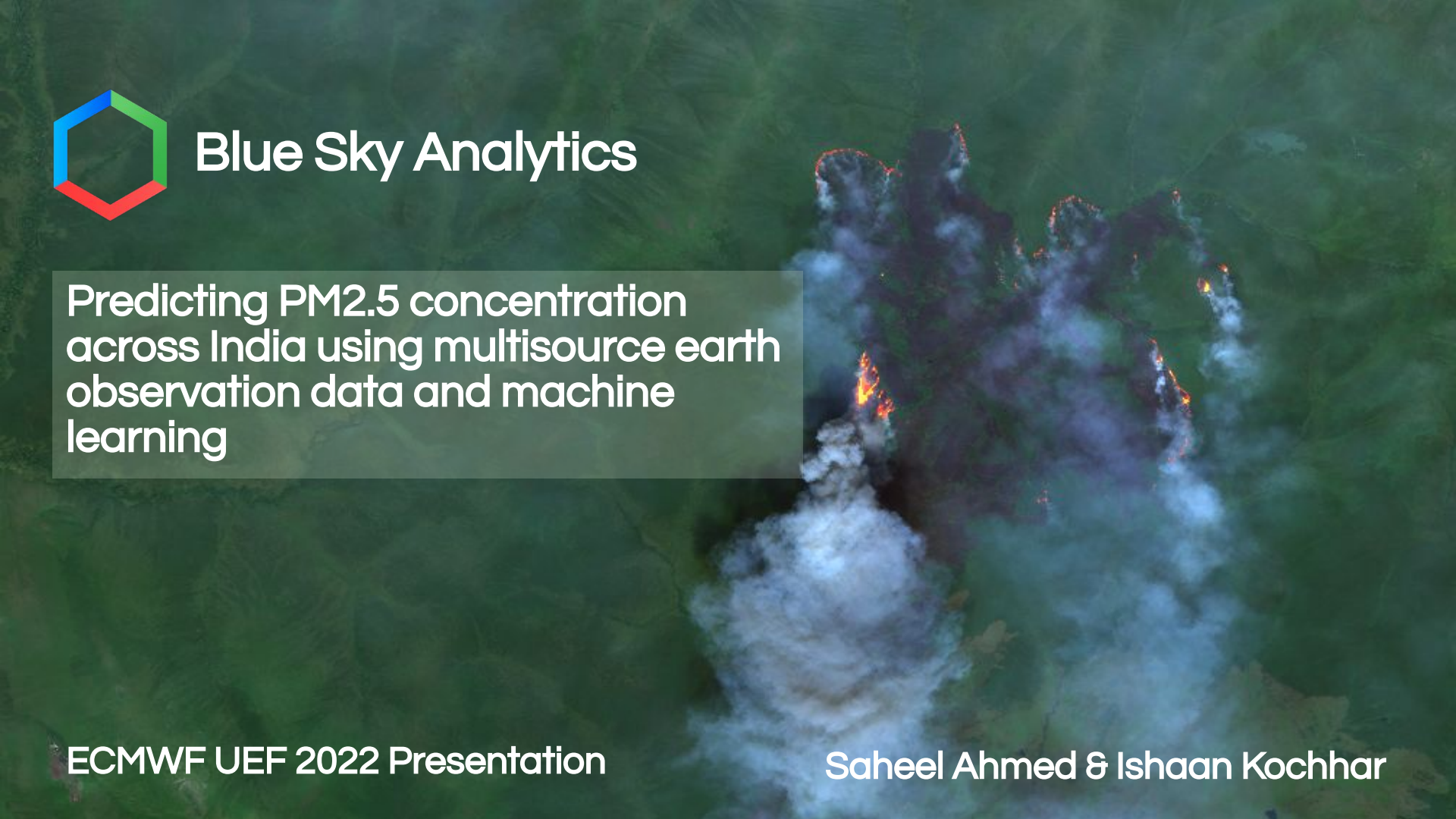
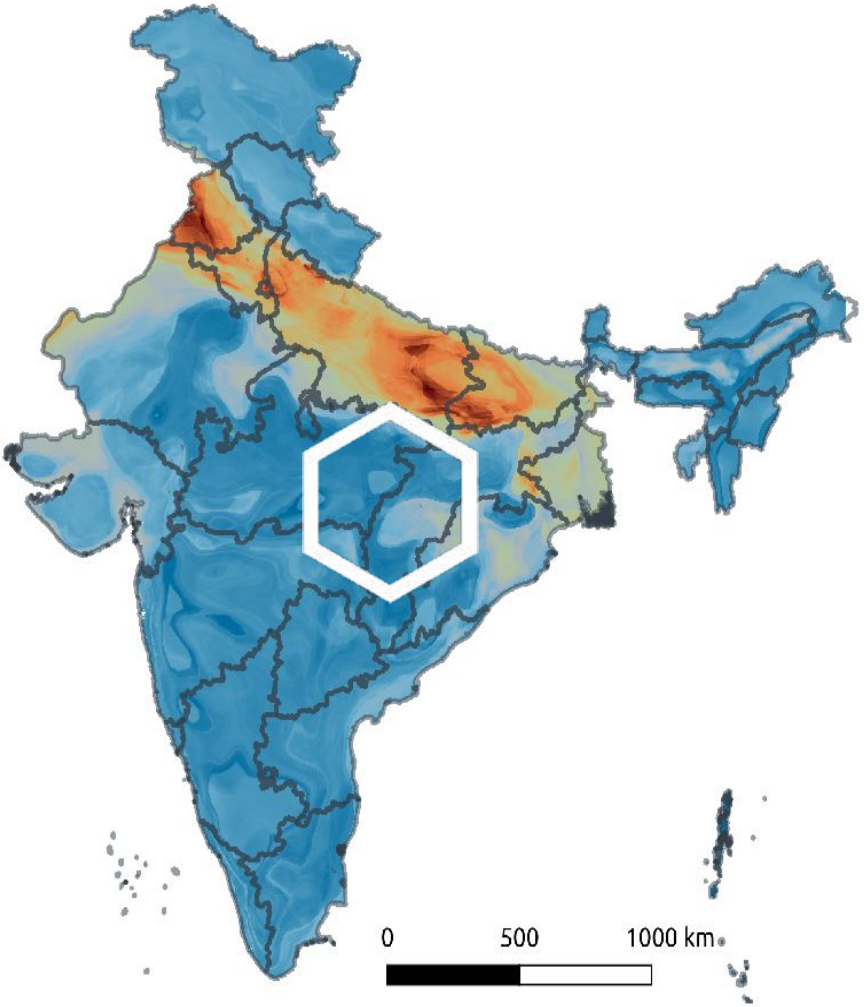


Blue Sky Analytics

Predicting PM2.5 concentration across India using multisource earth observation data and machine learning

ECMWF UEF 2022 Presentation

Saheel Ahmed & Ishaan Kochhar



AI + Satellites = Environmental Datasets

By combining the power of AI & Satellites, Blue Sky Analytics aims to build the equivalent of a 'Bloomberg for Environmental Data'

We build cutting-edge technology to monitor environmental parameters like **air pollution, fires, water quality, industrial emissions etc.** using satellite imagery.

Proprietary Artificial Intelligence analyzes this data to monitor & predict trends over space & time.

MIT SOLVE esa

© Shutterstock / 24Novembers / WWF

What is Blue Sky's Approach?

Ground sensors are expensive, resulting in **limited spatial coverage.**

India has 250 air quality monitors for an area of 3.2 million km²

Blue Sky combines data from **satellites, ground sensors using AI to build a spatially continuous dataset**

BreeZo creates a mosaic of satellite, meteorological and ground data

Meteorological Data
For our algorithm we use indicators like total precipitation, surface pressure, etc. We obtain all the meteorological data from the European Centre for Medium-Range Weather Forecasts.

Ground Measurements
We use data from almost 250+ monitoring stations and third-party ground monitors through partnerships. This allows us to provide data from different sources increasing its objectivity.

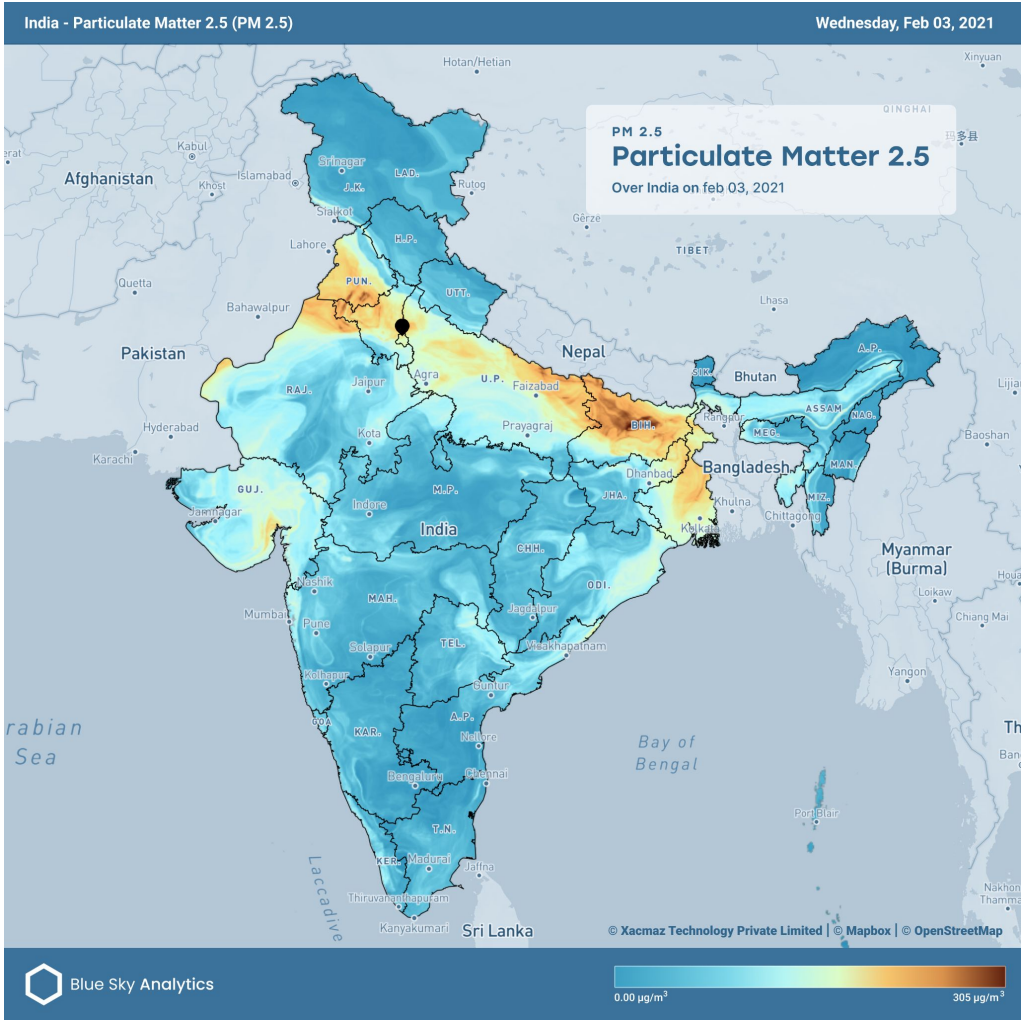
Breezo Smart AQ

Raw Data Sources

BreeZo source its data from from multiple sources from **MODIS, Sentinel 5P and ECMWF** services on a daily basis. We also ingest data from a network of AERONET sensors to calibrate **Aerosol Optical Depth** (AOD) data.

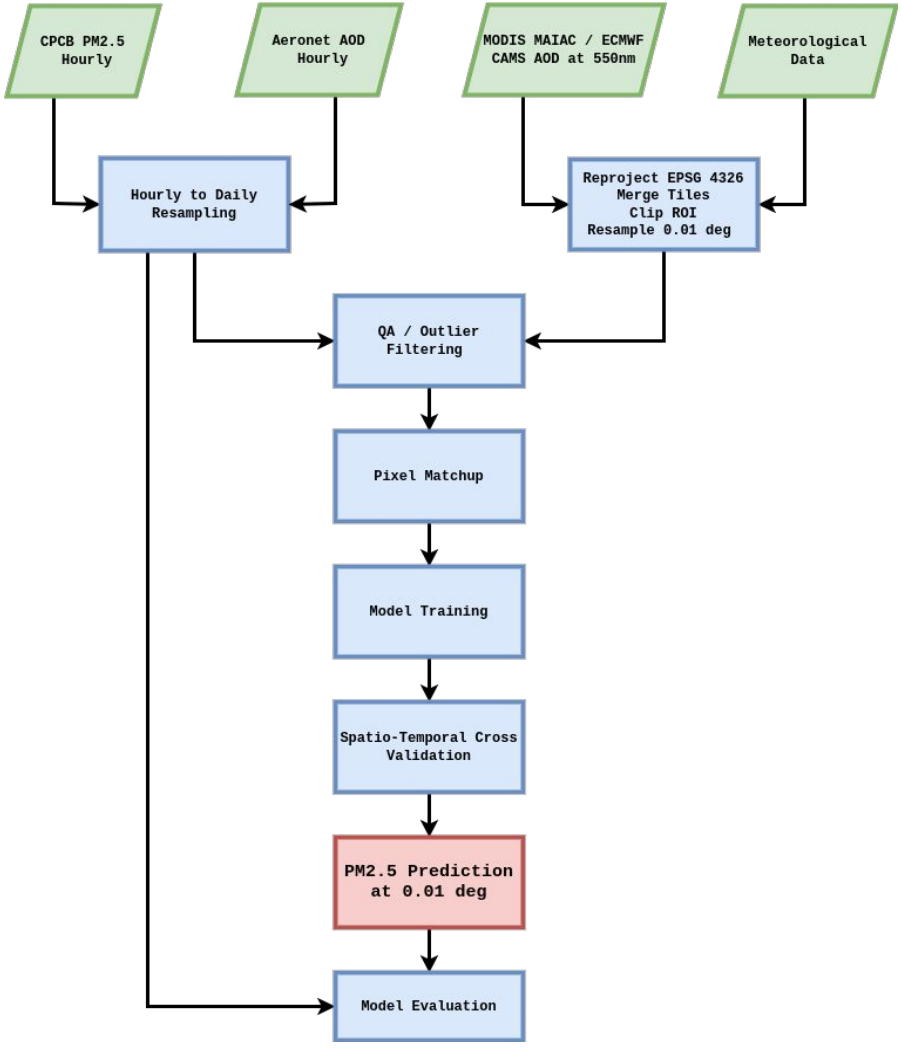
Results

The SmartAQ model is robust and can accurately estimate PM 2.5 concentrations by taking advantage of the **ensemble regression approach**, synergy of **space-time information**, and the **high-resolution and wide-spatial-coverage** of the AOD product.



Methodology Overview

1. Resampled hourly ground observations (PM2.5 & AOD) to daily.
2. Due to different spatial and temporal resolution all the satellite and meteorological data are uniformly resampled to the same spatial size 0.01 degree (~1km).
3. Satellite AOD validation against Aeronet measurements.
4. Outlier filtering, quality bands masking.
5. Pixel based sample extraction corresponding to each ground measurement.
6. Applying ensemble random forest regression.
7. Spatio-temporal cross validation and model evaluation.
8. Fine tuning and permutations feature importance.

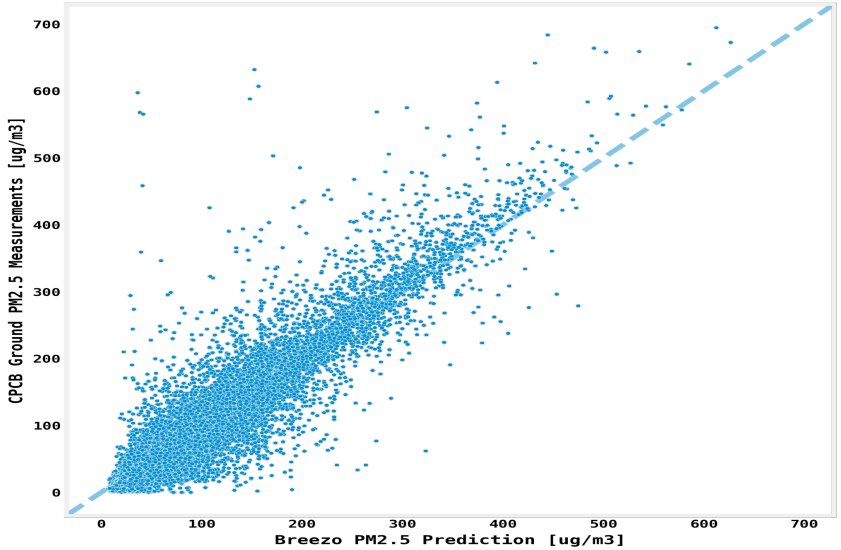


Results

Regression

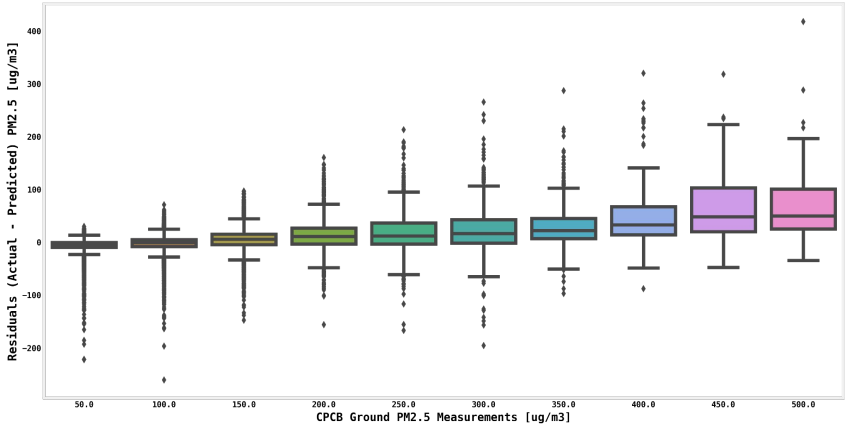
Regression statistics of daily satellite based Breezo predictions and CPCB ground measurements over India.

Achieved an R2 score: 0.88, MAE: 12.72 ug/m3, RMSE: 24 ug/m3



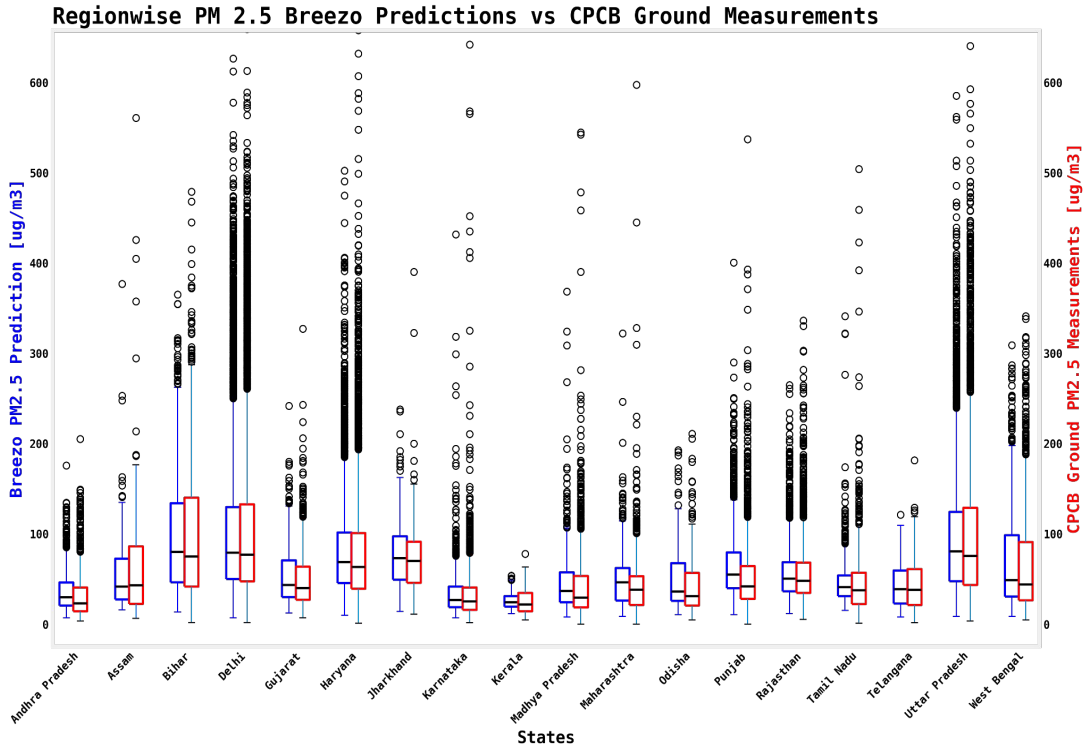
Residuals

Variation in the bias in daily satellite based Breezo predictions relative to CPCB ground measurements with an increase in the PM2.5 level. The boxplot represents 5-95 percentile levels.



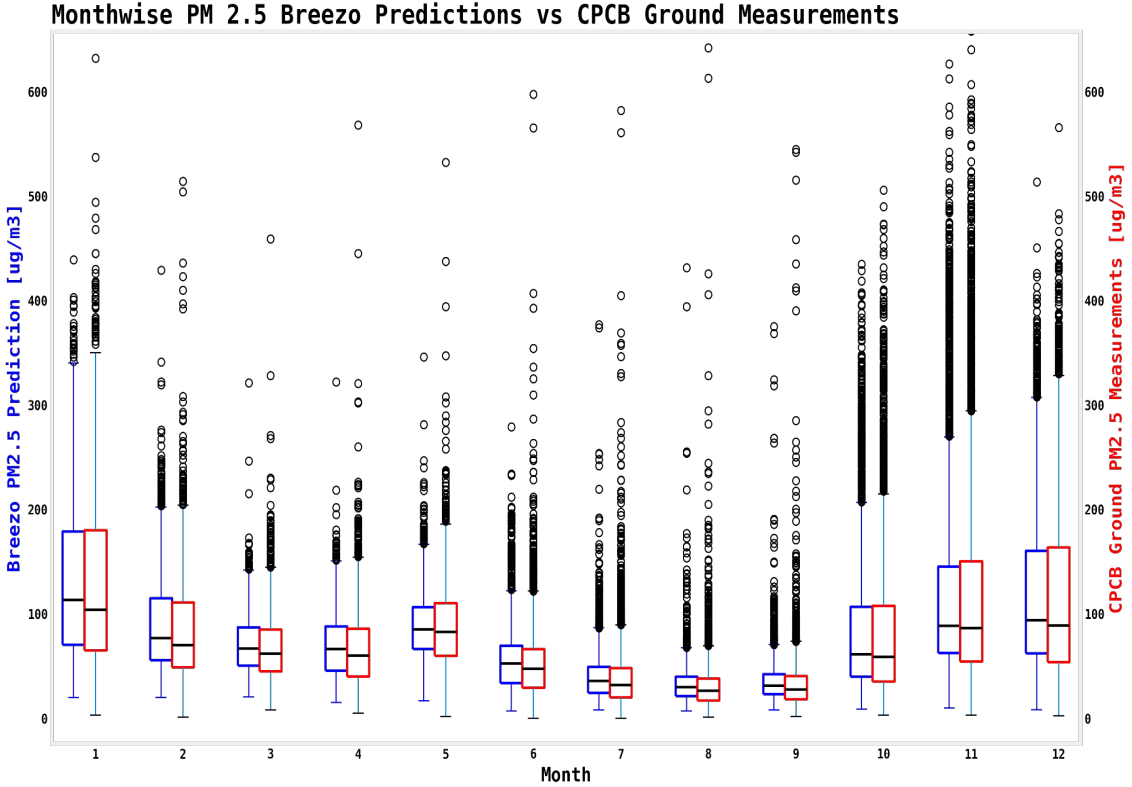
Spatial Trends

- The regional distribution of predictions is also in line with ground measurements.
- Majority falling within the 5 to 95 percentile range.
- Also, show variations in pollution range, with a higher level of pollution in regions of Indo-Gangetic Plains and lower in coastal and hilly states.



Temporal Trends

- Month-wise prediction comparison with ground measurements.
- Higher pollution range in Oct, Nov, Dec, and Jan primarily due to lower temperatures and annual biomass burning practice
- Lower in July and Aug due to the onset of monsoon.



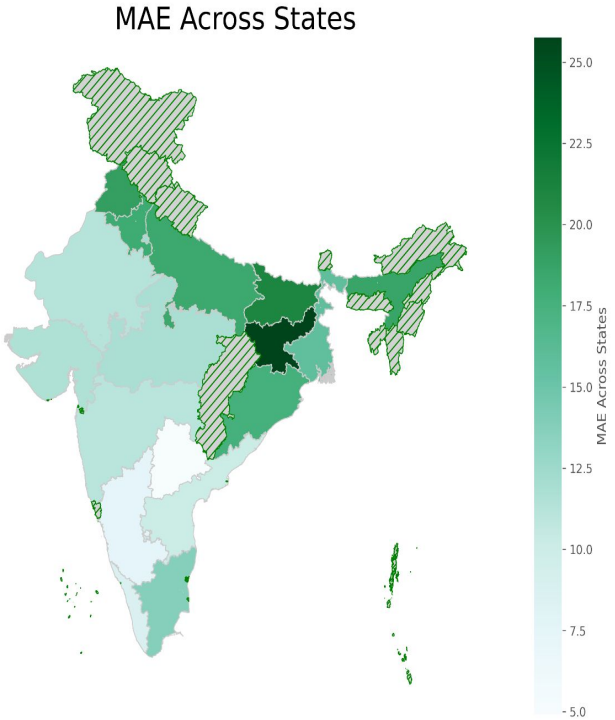
Limitations and Trade Offs

Lack of ground monitors in certain areas

Hilly states like Jammu & Kashmir, Himachal Pradesh, Uttarakhand and **northeastern states** like Tripura, Mizoram, Assam, Meghalaya etc. **Causing explicit bias** in these states which can't be validated due to lack of ground truth.

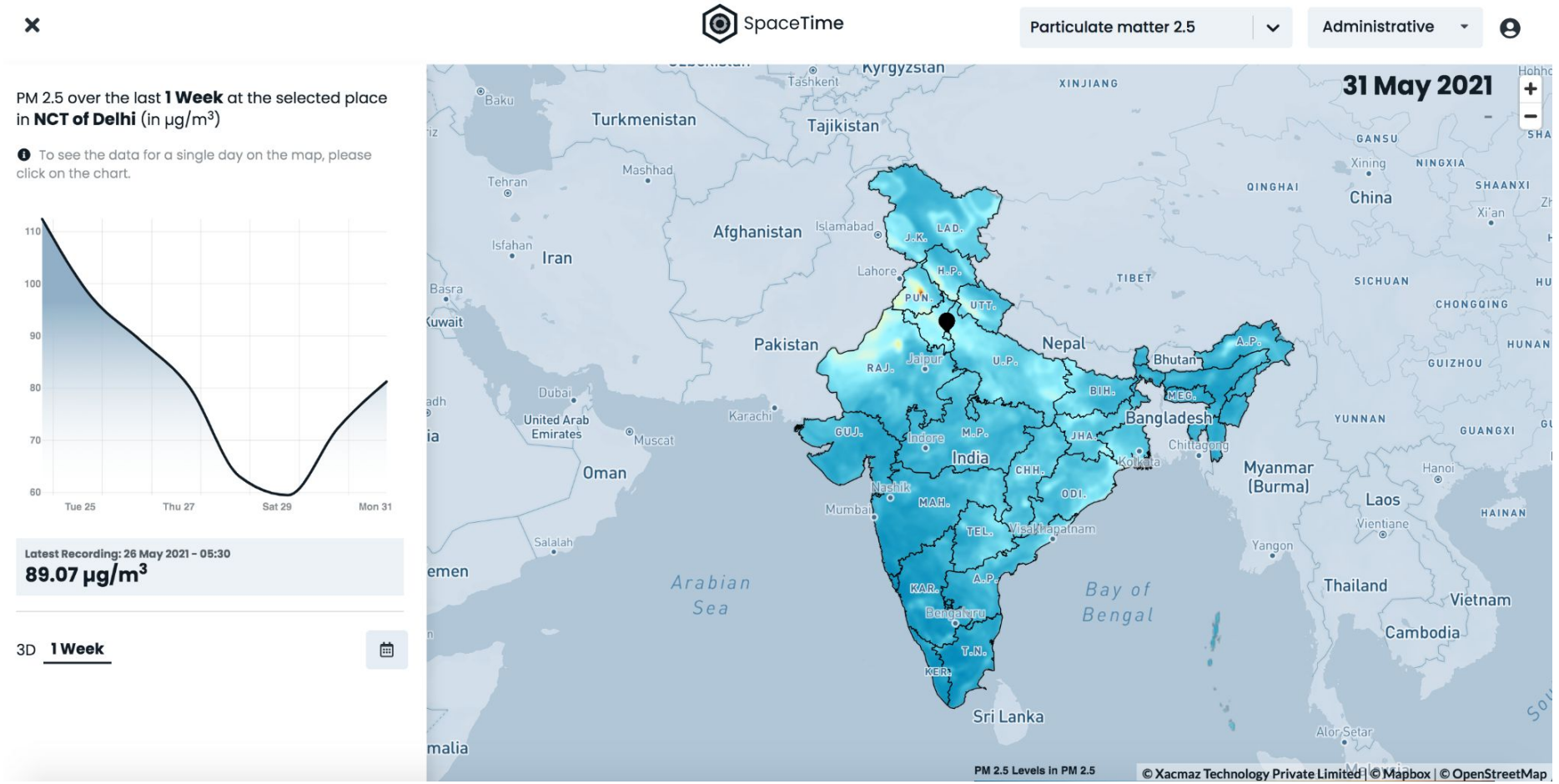
Rural vs Urban Air Quality

Most of the ground monitors are stationed in cities which makes it very hard to justify high pollution in any rural area. We can still analyse the results by comparing urban vs rural pollution trends but we won't be able to validate it.



Source: Blue Sky Analytics, 2019

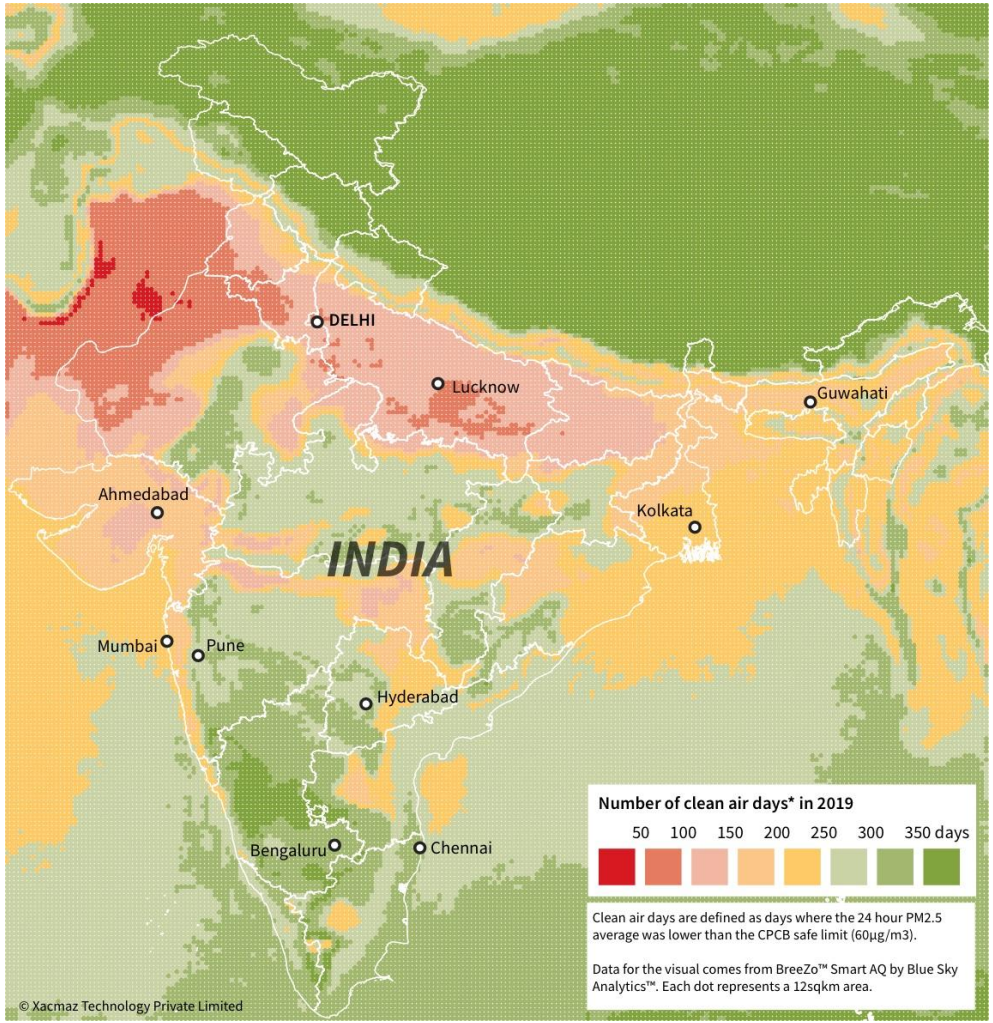
BreeZo: Satellite Data & AI- derived Air Quality Dataset at 1km2 resolution



BreeZo: Research Use-Case

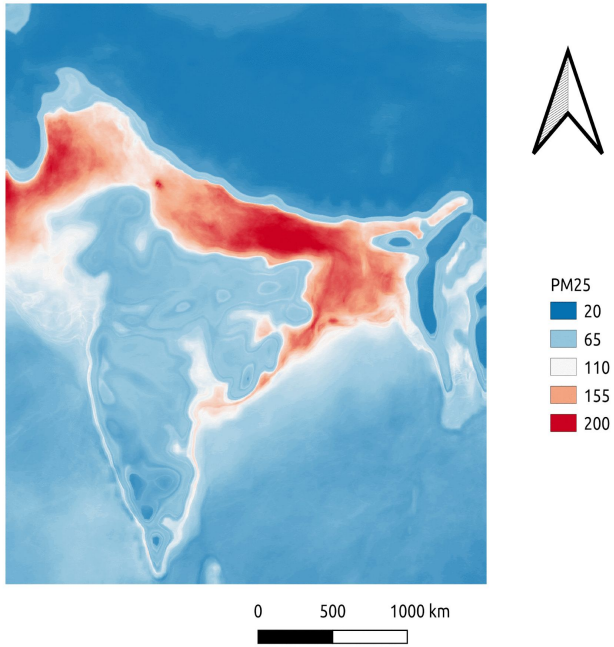
Our data was recently used by the **International Energy Agency**, in the India Energy Outlook, 2020 to demonstrate the **number of clean air days** in India.

You can see this map on the right.



Blue Sky Analytics

Average PM25 January 2019



Contact : tej@blueskyhq.io saheel@blueskyhq.io ishaan@blueskyhq.io