

A PROPOSAL FOR INDIA · 2026

AURA. Seeing the air every Indian breathes — from space.

A satellite-grounded intelligence platform for Surface AQI prediction and Formaldehyde hotspot detection across all 28 states and 8 UTs.

• SURFACE AQI

• HCHO HOTSPOTS

• REAL-TIME ALERTS

• 22 LANGUAGES



The mandate.

Develop a surface Air Quality Index derived from satellite data, and identify **spatio-temporal HCHO hotspots** over India using TROPOMI, INSAT-3D, CPCB and fire datasets — with a focus on biomass burning seasons.

OBJECTIVE 1

Spatial maps of surface AQI across India

OBJECTIVE 2

High-resolution HCHO hotspot maps

OBJECTIVE 3

Major source regions identified

OBJECTIVE 4

Fire–HCHO transport analysis

ALIGNED WITH

- 01 National Clean Air Programme
- 02 SDG Target 11.6 — Urban environment
- 03 Chintan Shivir 2.0 priorities
- 04 ISRO Earth Observation mission

AIR POLLUTION IN INDIA

A silent emergency measured in **lives.**

1.67M

EXCESS DEATHS
ATTRIBUTED TO PM_{2.5} IN
INDIA / YEAR

99%

OF INDIANS BREATHE AIR
DIRTIER THAN WHO
GUIDELINE

₹2.6L Cr

ANNUAL ECONOMIC LOSS
FROM AIR POLLUTION

9 / 10

MOST POLLUTED CITIES
GLOBALLY ARE IN INDIA

Sources: Lancet Planetary Health 2021 · WHO Global Air Quality Guidelines · World Bank.

The 100 km blindspot.

Most of the global population lives more than **100 km** from an air quality monitor. In India, **~1,300 million people** have no local AQI signal at all.

CPCB ground stations	~1,500
Cities under NCAP	131
Villages with zero coverage	~640,000
Forest land unmonitored	21.7%



AIR QUALITY INDEX

One number. Six pollutants. One decision.

AQI compresses concentrations of CO, NO₂, O₃, SO₂, PM₁₀ and PM_{2.5} into a single rating used to warn citizens and trigger regulatory action.



PURPOSE A
Inform & caution the public about daily exposure risk.

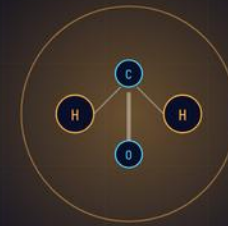
PURPOSE B
Trigger regulatory response by competent authority.

TODAY
Reported only where monitors exist — almost nowhere.

FORMALDEHYDE · HCHO

The fingerprint of **burning** India.

HCHO is a short-lived intermediate of VOC oxidation. It is the most reliable space-borne proxy for biogenic and pyrogenic VOC emissions, and a direct precursor to tropospheric ozone.



LIFETIME

~3 h

Snapshot of fresh emissions

RETRIEVED BY

GOME · SCIAMACHY · OMI
· TROPOMI

Two seasons. **One nation on fire.**

Each year India experiences two distinct biomass burning regimes that flood the troposphere with VOCs — visible from orbit as enormous HCHO columns.

APR — MAY

Forest fires

Western Ghats, Central India, Northeast. MODIS hotspot density peaks in late April.

BANDIPUR

SIMILIPAL

BANDHAVGARH

DZUKOU

OCT — NOV

Agricultural residue

Punjab, Haryana, western UP. Post-paddy stubble burning drives the Indo-Gangetic Plain into severe AQI.

SANGRUR

LUDHIANA

KARNAL

BAREILLY

A national priority. A funded mandate.

NCAP

20–40%

Reduction in PM_{2.5} & PM₁₀
across 131 non-attainment cities
by 2026.

SDG 11.6

Cities

Reduce adverse per-capita
environmental impact, including
air quality.

CHINTAN SHIVIR 2.0

Climate

Urban climate, air quality,
mitigation and sustainable
development.

“AURA directly serves ISRO's Earth Observation mandate by translating raw satellite signals into policy-grade public-health intelligence.”

THE PROJECT GOAL

Monitor surface AQI and identify HCHO hotspots over India using **satellites** + **sensors** + **AI**.

Bridging the gap between highly monitored urban centres and unmonitored rural India — at national scale, in real time.

SURFACE AQI DEVELOPMENT

Wall-to-wall AQI from space.

Predict ground-level concentrations of six pollutants from columnar satellite retrievals and meteorology, then render daily AQI as a continuous 1 km grid across India.

CO

NO₂

O₃

SO₂

PM_{2.5}

PM₁₀

TARGET RESOLUTION

1 km

spatial · daily temporal · 6 pollutants

COVERAGE

28 states · 8 UTs

~3.28 million km² · ~1.4 B people

HCHO RETRIEVAL

5.5 km

Sentinel-5P TROPOMI native · gridded to 0.05°

HOTSPOT THRESHOLD

$> \mu + 2\sigma$

Per pixel, seasonal baseline (2019–2025)

HCHO HOTSPOT MAPPING

**Where the air
is burning.**

High-resolution maps of formaldehyde columns over the Indian subcontinent during biomass burning seasons, with statistical hotspot detection and DBSCAN clustering.

SOURCE IDENTIFICATION

Pinpoint the **where** behind the haze.

REGION A

Indo-Gangetic Plain

Stubble burning, brick kilns, residential biomass, vehicular NO_x.

REGION B

Central forest belt

Dry-season fires across MP, Chhattisgarh, Odisha, Maharashtra.

REGION C

NE India & Himalayas

Jhum cultivation fires, transboundary smoke from Myanmar.

6+

DISTINCT HOTSPOT REGIONS EXPECTED

365 d

ANNUAL CONTINUOUS TIME SERIES

7 yr

HISTORICAL BASELINE 2019–2025

HYSPLIT

BACK-TRAJECTORY TRANSPORT MODEL

Four deliverables. One operational layer.

DELIVERABLE 1

Spatial maps of surface AQI

Daily 1 km grids, 6 pollutants + composite AQI.

DELIVERABLE 2

HCHO concentration & hotspot maps

Daily and seasonal, with cluster polygons.

DELIVERABLE 3

Major source region atlas

Annotated polygons with attributed source types.

DELIVERABLE 4

Fire → HCHO transport analysis

Correlation surfaces and trajectory bundles.

DATA SOURCES

Five orbits. One ground truth.

SATELLITE ·
AEROSOLS

INSAT-3D

AOD · half-hourly ·
MOSDAC

SATELLITE ·
GASES

Sentinel- 5P

TROPOMI · NO₂ SO₂
CO O₃ HCHO

GROUND ·
REFERENCE

CPCB CAAQM

~1,500 stations · hourly

SATELLITE ·
FIRE

MODIS / VIIRS

Active fire · 375 m / 1
km

REANALYSIS ·
MET

ERA5 / IMDAA / MERRA-2

Wind, RH, BL height, T

Combined daily volume

~ 240 GB

INSAT-3D · MOSDAC

India's geostationary eye.

Half-hourly Aerosol Optical Depth at 550 nm over the Indian region, retrieved by the INSAT-3D Imager — the temporal backbone for PM_{2.5} estimation.

CADENCE

~30 min

FOOTPRINT

~ 4 km

COVERAGE

Full disk

PRODUCT CHAIN

L1B → L2 AOD

→ PM_{2.5} derivation (DL)

→ AQI integration

[mosdac.gov.in / insat-3d-data-products](https://mosdac.gov.in/insat-3d-data-products)

SENTINEL-5P · TROPOMI

Five molecules. One pass per day.

NO₂

tropospheric column

SO₂

tropospheric column

CO

tropospheric column

O₃

tropospheric column

HCHO

tropospheric column

NATIVE PIXEL

5.5 × 3.5 km

OVERPASS

~13:30 local

ACCESS

GEE · DLR L3

CPCB · CAAQM

The ground truth.

Continuous Ambient Air Quality Monitoring network — ~1,500 stations across India delivering hourly reference-grade measurements. The supervised signal that anchors every DL model.

Total stations (operational)	1,500+
Pollutants reported	12
Cadence	hourly
Historical depth	~10 years



MODIS · VIIRS · FIRMS

Every fire, **timestamped.**

MODIS AQUA/TERRA

1 km

Twice-daily global fire detection since 2000.

VIIRS S-NPP/NOAA-20

375 m

Higher resolution, small fires & nighttime detection.

NRT DELIVERY

~ 3 hours

FIRMS near-real-time JSON / shapefile feed.

Use: extract biomass-burning windows · weight HCHO covariate · constrain BC-OC emissions.

REANALYSIS · METEOROLOGY

The atmosphere as **context.**

Surface concentration is driven as much by transport and mixing as by emissions. We ingest hourly meteorology from three independent reanalysis systems.

ECMWF

ERA5 · 0.25°

NCMRWF

IMDAA · 12 km Indian
region

NASA

MERRA-2 · 0.5° × 0.625°

T_{2m}

U,V wind

RH

BL height

Precip

SH flux

The technology stack.

INGEST

Earth Engine · CDS API
· FIRMS API · MOSDAC

STORAGE

Zarr on object store ·
Parquet metadata

COMPUTE

Python · xarray · Dask ·
PyTorch · CUDA

SERVE

FastAPI · MapLibre
vector tiles ·
WebSocket alerts

GEE

xarray

Dask

PyTorch

Lightning

WandB

Zarr

FastAPI

MapLibre

PostGIS

React Native

HYSPLIT

GOOGLE EARTH ENGINE

Planet-scale preprocessing.

GEE handles the heavy lifting — cloud-masking, reprojection, daily aggregation and export of Sentinel-5P and fire products directly to our training pipeline.

```
# aura/ingest/s5p_hcho.py
import ee
ee.Initialize()

col = ee.ImageCollection("COPERNICUS/S5P/OFFL/L3_HCHO")
    .filterDate(start, end)
    .filterBounds(india)
    .select("tropospheric_HCHO_column_number_density")

daily = col.map(qa_mask).reduce(ee.Reducer.mean())
return ee.batch.Export.image.toCloudStorage(daily, ...)
```

Three networks. One **surface estimate**.

SPATIAL

CNN

2D conv encoder over column + meteorology stacks. Captures gradients across grid cells.

TEMPORAL

LSTM

Sequence model over 30-day windows. Captures accumulation, washout, seasonality.

HYBRID

CNN-LSTM

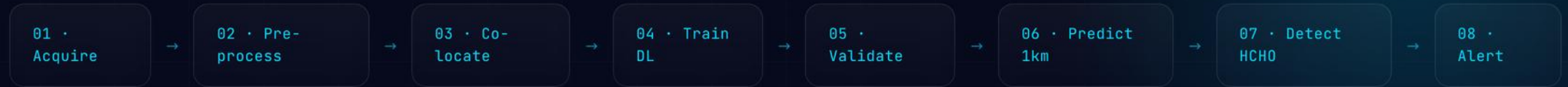
Spatio-temporal joint model. Production candidate per ablation study.

Loss

$$L = \text{MSE} + \lambda \cdot \text{Huber}(\delta=2)$$

END-TO-END PIPELINE

From photon to push notification.



OBJECTIVE 1 PATH
Columnar concentrations + meteorology → DL model
→ Surface AQI 1 km grid.

OBJECTIVE 2 PATH
HCHO L3 + fire mask → statistical baseline →
hotspot clusters + transport.

DATABASE · COLUMNAR

A 7-year cube of India from above.

Build a daily, gap-filled, India-wide cube of columnar multipollutant concentrations (NO₂, SO₂, CO, O₃, HCHO) plus AOD from INSAT-3D.

Temporal range	2019-01 → present
Spatial grid	0.05° × 0.05°
Variables	6 columns + 8 met
Estimated size	~ 4.2 TB Zarr

QA PIPELINE

- QA value ≥ 0.5 mask
- Cloud fraction < 0.3
- Solar zenith $< 70^\circ$
- Sea-pixel exclusion
- Daily mosaic + 3-day gap fill

DATABASE · SURFACE

Co-locating ground truth with the sky.

CPCB STATIONS

Hourly observations at ~1,500 sites resampled to the satellite overpass window.

```
df.set_index(['station', 'time'])  
.resample('1H').mean()
```

METEOROLOGY

ERA5 / IMDAA / MERRA-2 interpolated to station coordinates and overpass time.

```
xarray.sel(lat, lon,  
method='nearest', time=overpass)
```

~ 8.4 M

CO-LOCATED SAMPLES (2019–2025)

13

PREDICTOR VARIABLES PER SAMPLE

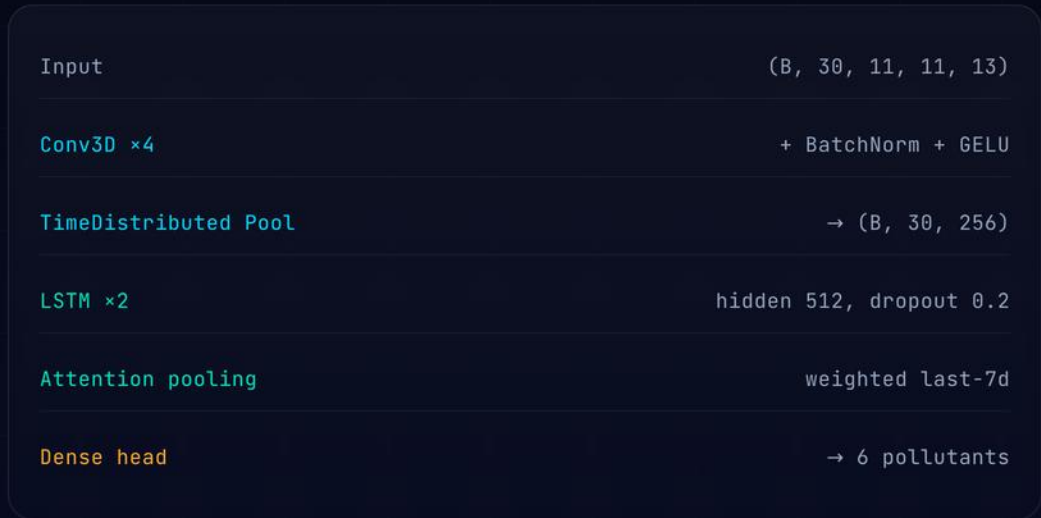
80/10/10

TRAIN / VAL / TEST SPLIT
(TEMPORAL BLOCK)

MODEL · ARCHITECTURE

CNN-LSTM spatio-temporal core.

A 4-layer CNN encodes the 11×11 pixel neighbourhood of each station; an LSTM rolls a 30-day sequence forward; a dense head emits the six pollutants jointly.



2.4 M PARAMS

A100 · 12 H

PYTORCH LIGHTNING

VALIDATION & INFERENCE

From station to subcontinent.

- 1 Hold-out validation**
Spatial K-fold (k=5) + temporal block hold-out on 2024–2025.
- 2 Compute statistics**
RMSE, R, MAE per pollutant per station — and per agro-climatic zone.
- 3 Predict 1 km grid**
Inference over India daily; composite AQI computed per CPCB sub-index method.
- 4 Render & publish**
Vector tiles → public app + open data portal.

TARGET ACCURACY

PM_{2.5} R²

≥

0.82

NO₂ R²

≥

0.78

O₃ R²

≥

0.74

SO₂ R²

≥

0.70

HCHO · PRE-PROCESSING

Cleaning the HCHO signal.

FILTER

QA value ≥ 0.5 · cloud fraction < 0.3 · solar zenith $< 70^\circ$.

REGRID

Bilinear resample to a 0.05° India lattice (700×800 cells).

BIAS CORRECT

Apply latitudinal scaling per Müller et al. 2024; offset against MAX-DOAS.

```
def qa_mask(img):  
    qa = img.select('qa_value')  
    return img.updateMask(qa.gte(0.5))
```

FIRE WINDOW EXTRACTION

When the country burns.

From MODIS + VIIRS active-fire detections we build daily fire-count rasters and isolate two operational windows per year — using a Gaussian kernel density estimate over the 7-day moving sum.

FOREST FIRES

Late March → mid-May;
peak in week 18.

STUBBLE BURNING

Late September → late
November; peak in week 44.

ANNUAL FIRE-COUNT SIGNATURE



SPATIO-TEMPORAL MAPPING

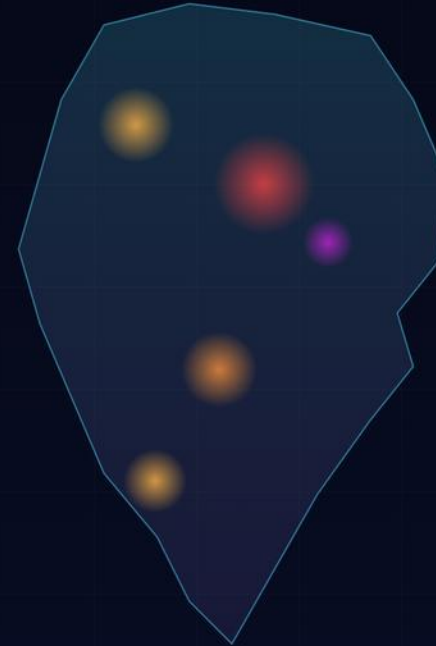
Rendering the HCHO field.

Daily mean tropospheric HCHO columns are rendered on the 0.05° India grid and animated across the burning windows to surface seasonal anomalies.

low

background

enhanced



HOTSPOT DETECTION

Statistics + clustering.

STATISTICAL THRESHOLD

$$\mu + 2\sigma$$

Per-pixel seasonal baseline 2019–2024. Flag any 2025 cell exceeding two standard deviations.

SPATIAL CLUSTERING

DBSCAN

$\epsilon = 25$ km, minPts = 8. Emits hotspot polygons with mean intensity and persistence days.

Validation

Hit-rate vs. CPCB benzene proxy stations · ≥ 0.75 target

TRANSPORT ANALYSIS

Following the wind.

Back-trajectories from HYSPLIT (driven by ERA5) traced 72 hours upstream from every hotspot centroid — quantifying the share of HCHO arriving from upwind fire pixels.

$r \geq 0.7$

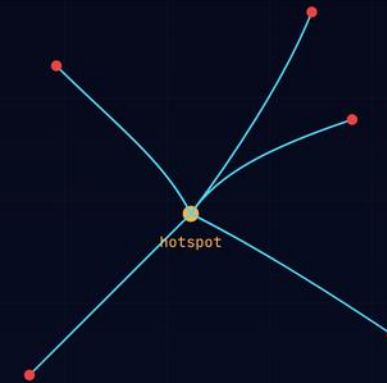
EXPECTED FIRE-
HCHO
CORRELATION IN
BURNING WEEKS

72 h

BACK-
TRAJECTORY
WINDOW

500 m

HYSPLIT
VERTICAL LEVEL



PROPOSED SOLUTION

AURA. A cloud-native AQI operating system for India.

SENSE

Aggregate municipal sensors, AURA microsensors and five satellite missions.

INFER

DL fusion → live 1 km AQI grid + HCHO hotspot polygons.

ACT

Multilingual push alerts to citizens + automated dispatch to authorities.

System architecture.

INGEST

- > INSAT-3D L2
- > Sentinel-5P L3
- > CPCB API
- > FIRMS NRT
- > Microsensor MQTT

PROCESS

- > GEE preprocessing
- > Zarr writer
- > Dask cluster
- > QA gates

INTELLIGENCE

- > CNN-LSTM AQI
- > HCHO baseline
- > DBSCAN clusters
- > HYSPLIT transport
- > LLM Assistant

DELIVER

- > Vector tiles
- > Push notifications
- > Authority webhooks
- > Multilingual TTS

Event bus · Kafka

Object store · Zarr / Parquet

Tile cache · CloudFront

Auth · OTP + ABHA

TIER 01 · METRO & URBAN

Integrate. Densify. Now.

Pipe existing municipal sensor feeds into AURA cloud; deploy AURA microsensors to fill micro-level gaps inside wards, industrial zones and traffic corridors.

42

NCAP TIER-1
CITIES LIVE IN
PHASE 1

500 m

MICRO-GRID
RESOLUTION

< 30 s

SPIKE-TO-
ALERT
LATENCY

DEPLOYMENT MANIFEST

Municipal feed integrations	REST + MQTT
Microsensor housing	IP66 · solar
Mounts	lamppost · panel
Telemetry	NB-IoT / LoRa
Calibration	co-location to CPCB

TIER 02 · SEMI-RURAL & DISTRICT TOWNS

The microsensor mesh.

HARDWARE

Compact mount

Mounts on lampposts and solar panels. 500–600 m sensing radius.

TELEMETRY

Continuous

Streams to AURA cloud; flags rapid fluctuation events for developers.

ACTION

Authority dispatch

Spike coordinates auto-shared with Pradhan, Tehsildar, BDO, DM.

Every Panchayat sees its own air — for the first time, in its own dialect.

TIER 03 · FOREST & EXTREME RURAL

Where ground hardware cannot go, orbit will.

In dense forests and remote regions we deploy no hardware. AURA relies exclusively on satellite intelligence — INSAT-3D, Sentinel-5P, MODIS — augmented by CPCB regional anchors.

EMERGENCY RESPONSE

- 01 Detect forest fire from MODIS NRT
- 02 Geocode to nearest village + Tehsil
- 03 Translate to local language
- 04 Dispatch coordinates + course of action
- 05 Activate ground manpower

REAL-TIME ALERTS

When the air turns, we tell **everyone**.

A WebSocket pipeline scans every fresh inference; AQI crossings above hazardous thresholds fan out as push notifications to registered users within their 5 km radius — and as structured webhooks to local authorities.

< 30 s

DETECT →
NOTIFY

5 km

DEFAULT
CITIZEN
RADIUS

22 + 11

SCHEDULED
LANGS +
DIALECTS



ACCESSIBILITY & LOCALIZATION

India, in India's words.

हिन्दी	বাংলা	ತುಳು	मराठी	தமிழ்	ગુજરાતી
ಕನ್ನಡ	മലയാളം	ਪੰਜਾਬੀ	ଓଡ଼ିଆ	অসমীয়া	اردو
संस्कृतम्	سنڌي	नेपाली	कोंकणी	मैथिली	डोगरी
বোডো	मणिपुरी	کشمیری	संथाली		

VOICE ASSISTANCE

Native-language TTS via AI4Bharat IndicTTS — for the visually impaired and low-literacy users.

DIALECTS

Bhojpuri, Awadhi, Maithili, Marwari, Chhattisgarhi, Tulu, Konkani and more.

ICONIC UI

Glanceable color + emoji status for sub-literate readers.

SPATIAL-AWARE AI

An assistant that knows your air.

Indic LLM grounded on the live AURA tensor — answers in the user's dialect with location- and source-specific mitigation suggestions.

USER · SANGRUR

मेरे पिंड में आज दम क्यों घुट रहा है?

AURA

आज पंजाब में 1,842 पराली जलाने की घटनाएँ हैं। हवा की दिशा NW है — आपके पिंड में धुआँ शाम 5 बजे तक रहेगा। मास्क पहनें, खिड़कियाँ बंद रखें।

USER

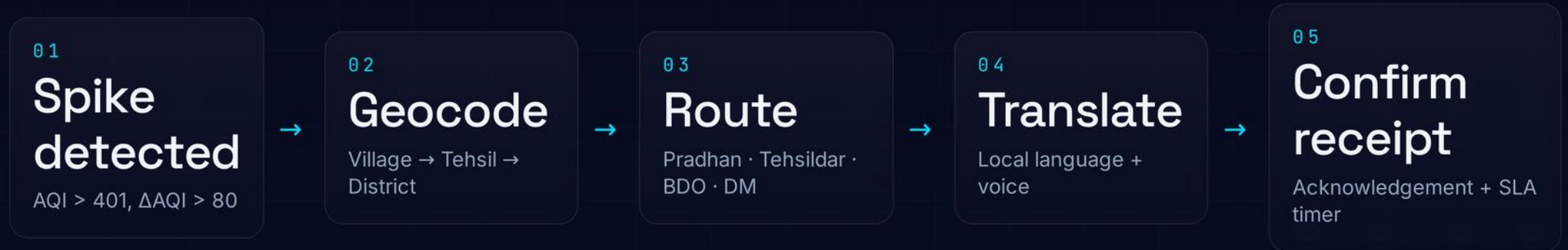
क्या करूँ कि अगले साल यह न हो?

AURA

पास का Happy Seeder केंद्र 4 km दूर है। PUSA Decomposer सब्सिडी 25 अक्टूबर तक उपलब्ध है। मैं फॉर्म खोल दूँ?

AUTHORITY DISPATCH

From signal to **officer** in under a minute.



4-tier
ESCALATION CHAIN TO DISTRICT
MAGISTRATE

SLA
HARDCODED 30-MIN OFFICER
RESPONSE WINDOW

Audit
APPEND-ONLY LEDGER OF EVERY
ALERT

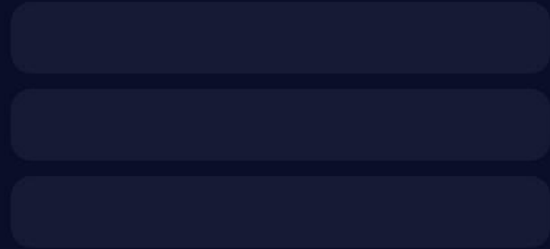
AURA MOBILE

One app. Three audiences.

CITIZEN

284

AQI · Poor

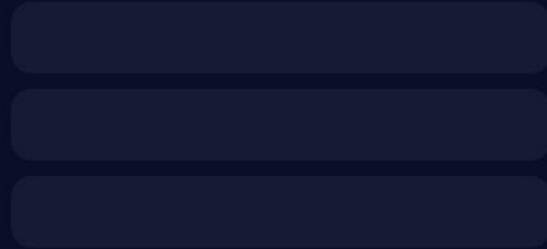


AQI now · 7-day forecast ·
health tips · voice assistant

OFFICER

12

Open spikes

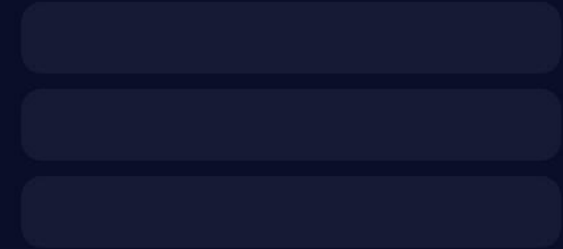


Live ward heatmap · spike
inbox · SLA tracker · weekly
digest

RESEARCHER

0.84

Model R²



Tile API · Zarr download ·
notebook templates · DOI
export

OBJECTIVE 01 METRICS

How we know it works.

ROOT MEAN SQUARE ERROR

RMSE

Per pollutant, per station, per season — and aggregated across India.

PEARSON CORRELATION

R

Linear agreement between predicted and observed surface concentration.

MEAN ABSOLUTE ERROR

MAE

Robust to outliers; primary metric for policy reporting.

PM_{2.5}

R² 0.82

MAE 8.4

NO₂

R² 0.78

MAE 5.1

O₃

R² 0.74

MAE 9.2

SO₂

R² 0.7

MAE 3.8

CO

R² 0.76

MAE 0.21

AOD

R² 0.85

MAE 0.05

OBJECTIVE 02 METRICS

Hotspots, judged honestly.

ACCURACY & CLARITY

Precision · recall · F1 against MAX-DOAS reference and CPCB benzene proxy.

MULTI-SOURCE INTEGRATION

Concordance across TROPOMI, MODIS, ERA5, CPCB measured by mutual information.

SCIENTIFIC INTERPRETATION

Peer review of source attribution by ISRO-NRSC and IIT-D atmospheric groups.

VISUALIZATION QUALITY

Cartographic clarity, color-blind safe palettes, temporal animation legibility.

METHODOLOGICAL INNOVATION

Novelty vs. baselines (geometric mean, percentile threshold, Getis-Ord G_i^*).

OPERATIONAL LATENCY

End-to-end ingestion → published hotspot polygon
SLA \leq 6 hours.

EXPECTED RESULTS

What India gets.

NATIONAL AQI CANVAS

Daily, 1 km, 7-year archive

Open data. CC-BY 4.0. DOI per release.

HCHO ATLAS

Seasonal hotspot polygons

Source attribution + transport bundles.

PUBLIC APP

22 scheduled languages

iOS · Android · Bharat OS.

AUTHORITY CONSOLE

SLA-tracked dispatch

For Pradhan → District Magistrate.

COST EFFICIENCY

Coverage at 1/40th the cost.

A traditional CPCB CAAQM station costs ~₹1.2 Cr to deploy and ~₹15 L / year to operate. AURA reaches one village for under ₹30,000 — and one square kilometre for free, from orbit.

CAAQM STATION
Capex per location

₹ 1.2 Cr

AURA MICROSENSOR
Capex per location

₹ 28 K

AURA SATELLITE-ONLY
Capex per km²

₹ 0

HUMAN IMPACT

If even 5% of severe days are avoided...

83,000

PREMATURE DEATHS
PREVENTED ANNUALLY

₹13K Cr

HEALTHCARE COSTS
AVERTED

11 M

ASTHMA EXACERBATIONS
AVOIDED

6 d

EARLIER SEASONAL
ACTION BY AUTHORITIES

Modelled on Health Effects Institute SoGA 2023 PM_{2.5} exposure-response coefficients, applied to a 5% avoidance of days exceeding AQI 301.

ROADMAP · 24 MONTHS

Phased to **delivery.**

Q1-Q2 Foundation

DATA INGESTION · 5 MISSIONS

ZARR CUBE · 2019-PRESENT

CPCB CO-LOCATION DATASET

Q3-Q4 Intelligence

CNN-LSTM V1 TRAINED

HCHO BASELINE + HOTSPOT DETECTOR

VALIDATION AGAINST HELD-OUT 2024

Q5-Q6 Product

MOBILE APP V1 · 4 LANGUAGES

AUTHORITY DISPATCH PILOT · PUNJAB + KARNATAKA

MICROSENSOR PILOT · 200 NODES

Q7-Q8 Scale

ALL 22 LANGUAGES + DIALECTS

NATIONAL ROLLOUT · 131 NCAP CITIES

OPEN DATA PORTAL + DOI RELEASES

THE ASK

A partnership with ISRO.

DATA ACCESS

INSAT-3D L1B & L2 via MOSDAC research licence. Bhuvan tile hosting for the public app.

COMPUTE & FUNDING

ISRO HPC quota for backfill training. Seed grant ₹3.2 Cr · 24 months.

VALIDATION NETWORK

Joint deployment with NRSC, IIT-D CARE & SAFAR sites for ground reference.

ATMOSPHERIC SCIENCE

3 PhD scientists · Remote sensing + atmospheric chemistry

ML & ENGINEERING

4 engineers · Geo-ML, MLOps, mobile

FIELD & POLICY

2 leads · CPCB liaison, panchayat outreach

AURA · 2026

Every Indian deserves to know the air they breathe.

Built for ISRO. Powered by satellites. Spoken in 22 languages.

CONTACT

team@aura.intelligence

REPOSITORY

github.com/aura-india

ISSUE

ISRO · No. 3 · Surface AQI & HCHO