Fog Risk Monitoring and Assessment for India Using Bayesian Networks and the ECMWF IFS Ensemble Prediction System

Sarath K. Guttikunda[1], **Nishadh Kalladath**[2], Robert R. Tucci[3], Jully Ouma[2], Ahmed Amdihun[2], Sai Krishna Dammalapati[1]

[1]Urban Emissions Information, Goa, India [2]IGAD Climate Prediction and Applications Centre (ICPAC), Nairobi, Kenya [3]www.ar-tiste.xyz

Introduction

Dense winter fog disrupts **aviation, transport and daily life** across northern India; the Indo-Gangetic Plain (IGP) regime is dominantly **radiation fog** under calm, near-saturated, aerosol-rich conditions, with the stable nocturnal boundary layer co-trapping $PM_{2.5}/NO_x$.

Objective. Building on Parde et al. (2022) and the BOFFIN-Melbourne Bayesian Decision Network (Boneh et al. 2015), we deliver a **continuous district-level (admin-2) risk monitoring system** that fuses the ECMWF IFS ensemble (51 members, 0.25°, 53 leads) with VIIRS / Sentinel-5P satellite indicators via a **Bayesian Network** emitting a four-level risk decision support advisories (Monitor / Evaluate / Assess / Actionable_Risk).

Materials and methods

Three design choices distinguish this pipeline:

1. **Open data, open infrastructure.** ECMWF IFS open-data archive (S3, anonymous) is streamed via GRIB-Index-Kerchunk (GIK) parquet manifests, decoded by byte-range reads on Coiled/Dask workers, and written to a publicly readable **Icechunk store on `source.coop`**. No proprietary feed; the materialised store is consumable by any Zarr-aware client.
2. **Bayesian Network with soft evidence.** A BN is a flow-chart of *what depends on what*: each node is a variable (aerosol, moisture, fog, stagnation), a per-node Conditional Probability Table holds expert-informed conditionals on fog_aq_risk, and Bayes' rule fuses new evidence with priors. Evidence enters as hard (observations), soft (Ensembles), or virtual (blended data).
3. **AI-assisted, version-controlled development.** The Python prep (`fog_data_prep.py`) and Julia BN engine (Bâgaev et al. 2023) (`fog_bn_ibf_v1.jl`) are co-developed using **Claude Code** (Anthropic, model `claude-opus-4-7`, accessed March-May 2026) as a coding assistant against the public repo `github.com/nishadhka/bn-airquality`..

Data Streaming ECMWF IFS

The IFS open-data 00Z 51-member ensemble lives on the public `ecmwf-forecasts` S3 bucket; E4DRR `gik-ecmwf-par` parquets give a byte-offset map per GRIB message. Each Coiled/Dask worker handles one *(date, member)* task end-to-end on the bbox $[20, 100°\mathrm{E}] \times [15°\mathrm{S}, 40°\mathrm{N}]$ at 0.25° (221×321), pulling 10 fog-relevant variables (`t2m, d2m, u10, v10, msl, strd, ssrd, tp, r_pl, t_pl`) via parallel byte-range S3 GETs and decoding with `gribberish`. Per-member streaming writes to Icechunk cap coordinator memory at ∼150 MiB. For 2025-01-01: **510 chunks, 2.74 GB, 21.7 min** on a 15-node Coiled **`e2-standard-4`** cluster; output at `source.coop/nishadhka/aq-icechunk-store-ifs/`.

Antecedent satellite AOD

A separate Icechunk store holds daily L3 AOD — **VIIRS NOAA-20** (0.1°, default) or **MODIS MAIAC** (0.05°). The prep script pulls a **7-day antecedent mean** over $[d{-}7, d{-}1]$ per district, motivated by the IGP anti-cyclonic aerosol-buildup-and-flush cycle (Parde 2022 §3.2(d); Dey 2010); replaces the v2 state-DJF climatology fallback.

Per-zone evidence extraction

Per ensemble member m and lead-time ℓ, a fuzzy fog-formation index $F_{m,\ell} \in [0, 1]$ at each pixel is the product of three sigmoid factors, one per physical pivot:

Near-saturation: factor → 1 as RH crosses **92%** (bbox p90 = 82%).

Per-zone evidence extraction (cont.)

Calm wind: factor → 1 as 10-m wind drops below **4 m s^{-1}** (bbox mean = 4 m s^{-1}).

No washout: factor decays once cumulative precipitation $\mathrm{tp} \geq$ **2 mm** — the rain-washout threshold (rainy pixels cannot hold fog).

Bayesian network

Four evidence-parent nodes (with an optional fifth tail-risk parent for ablation) feed a hidden risk node; a deterministic CRMA decision is computed from the posterior via a cost-loss trigger.

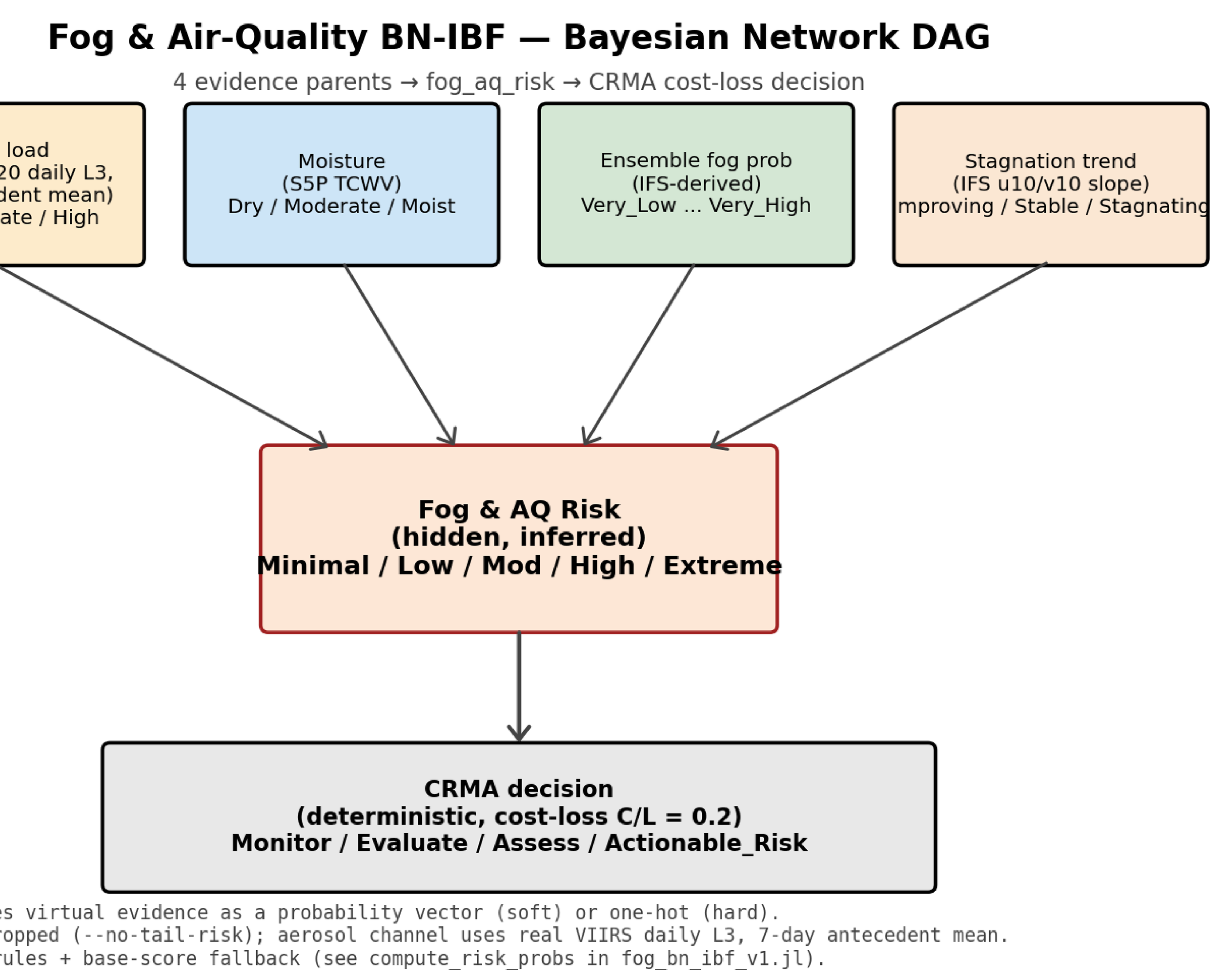

Figure 1. Fog & Air-Quality BN-IBF directed acyclic graph (4-parent topology; `tail` is an optional fifth parent enabled with `--include-tail-risk`).

Nodes, states, evidence

Node	States	Source
`aer`	Low / Mod / High (3)	VIIRS L3 7-day antecedent AOD
`mois`	Dry / Mod / Moist (3)	S5P TCWV daily L3
`fog`	VL/L/M/H/VH (5)	IFS ifs_fog_prob
`stag`	Imp / Stable / Stag (3)	IFS WS slope
`tail`†	Nil/L/Mod/H (4)	IFS ens-max F p95
`fog_aq_risk`	Min/L/Mod/H/Ext (5)	*hidden, inferred*
`crma_state`	Mon/Eval/Ass/Act (4)	cost-loss $C/L{=}0.2$

†optional ablation parent.

CRMA cost-loss decision

The 7-day VIIRS antecedent prioritizes: the **Rajasthan / MP dust + emission plateau** (AOD 1.0–1.5; Anoopgarh, Ganganagar, Tonk, Kota) and the **IGP industrial corridor** (Delhi 1.28, UP 1.42, Bihar 1.00) dominate the top-quartile, leading to Actionable Risk.

Results

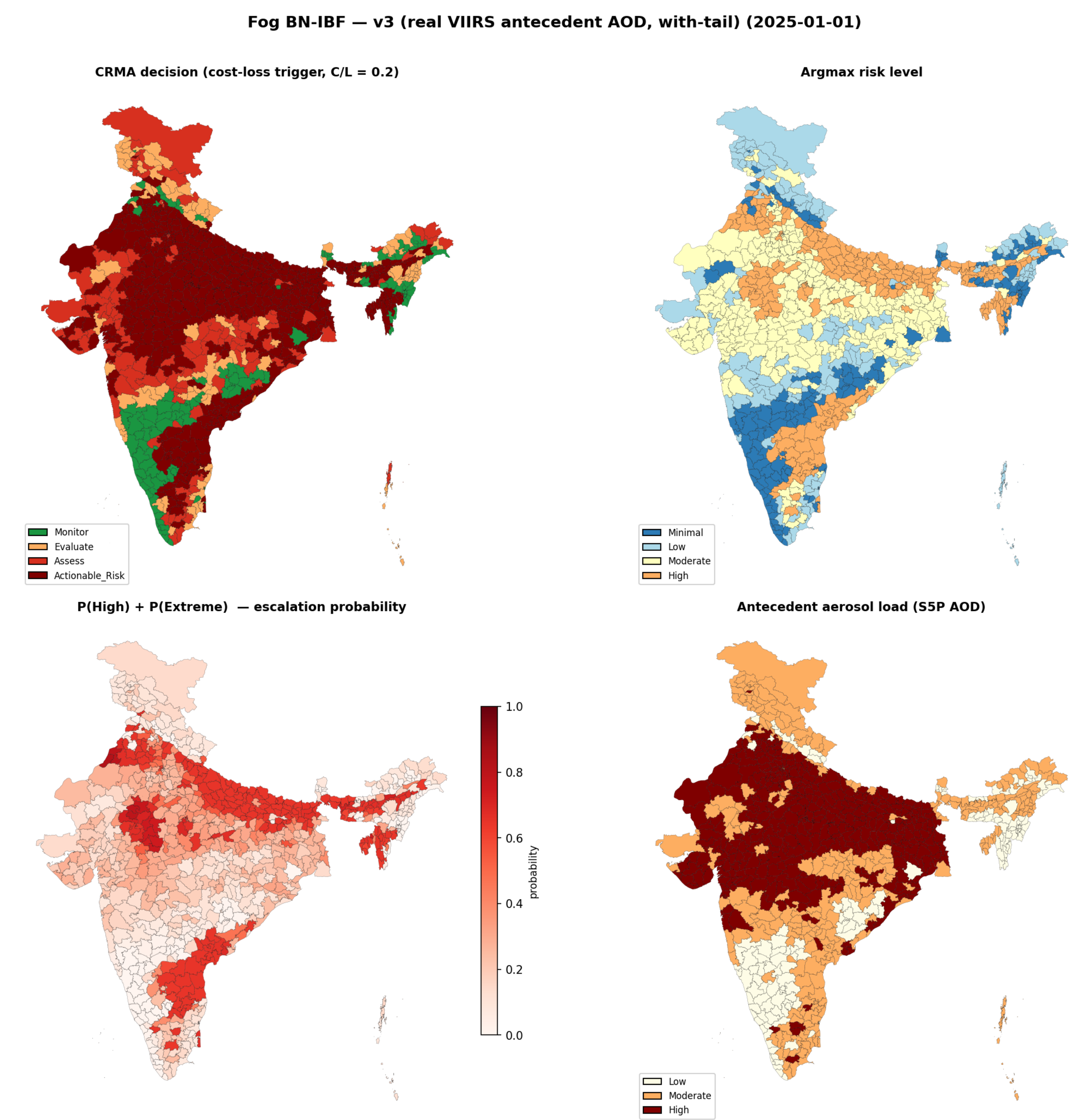

Figure 2. BN posteriors over 709 Indian districts, IFS init 2025–01–01, **7-day VIIRS antecedent AOD**. **Top-left:** CRMA decision ($C/L = 0.2$). **Top-right:** argmax fog_aq_risk. **Bottom-left:** P(High)+P(Ext). **Bottom-right:** antecedent aerosol load.

Discussion & outlook

Live 7-day VIIRS antecedent AOD elevates the high-aerosol + stagnation + fog co-occurrence directly to **Actionable_Risk**. **Next:** METAR / SYNOP validation; temporal chaining with multiple dates $\mathrm{risk}_{t-1} \to \mathrm{risk}_t$ via RxInfer.jl julia library.

Acknowledgements & references

Supported by **CRAF'd** (Complex Risk Analytics Fund) E4DRR project at ICPAC: `icpac-igad.github.io/e4drr`. We thank ECMWF for the IFS open-data archive, NASA Earthdata, github repos and the source.coop for data storage, E4DRR `gik-ecmwf-par` parquet manifests.

Parde et al. 2022, *Atmosphere* 13(10), 1608. Boneh et al. 2015, *Wea. Forecasting* 30(5), 1218–1233. Dey & Di Girolamo 2010, *JGR* 115(D15). Bâgaev et al. 2023, *JOSS* (RxInfer.jl).

nishadh.kalladath@igad.int