Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Uncertainty and proportionality when artificial intelligence forecasts trigger environmental restrictions

Tamil Iniyan A

tamiliniyana24@students.cutn.ac.in

Perspective

Keywords: Artificial intelligence, Decision support systems, Environmental policy, Air quality forecasting, Uncertainty quantification, Proportionality

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11241325/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Uncertainty and proportionality when artificial intelligence forecasts trigger environmental restrictions

Tamil Iniyan A.

Independent Legal Researcher; LL.M. (Corporate Law and Industrial Jurisprudence),

Central University of Tamil Nadu | ORCID: 0009-0005-5562-6291

tamiliniyana24@students.cutn.ac.in — corresponding author

Abstract

Some environmental regulators now trigger restrictions on forecasts before exceedances occur, while machine-learning methods are proposed for the forecasting pipelines behind them. Delhi-NCR's Graded Response Action Plan (GRAP) directs advance invocation on numerical air-quality forecasts, yet reported practice often followed observed readings. Judicial and regulatory directions demand anticipatory action, yet none reviewed here specifies the confidence a forecast must carry before it justifies coercion. This article asks what validation, reproducibility and uncertainty-quantification thresholds an AI-enabled forecast must meet before triggering binding restrictions. Using a functional comparative method, it separates physics-based forecasts and digital dissemination from the machine-learning layer that makes a pipeline AI-enabled. GRAP, with the Supreme Court's timing orders, is the single test case. Instruments from India, China, France and California's air districts are coded for observed, forecast-based and discretionary triggers and read alongside forecast-verification studies. This article argues that proportionality and reasoned decision-making require restrictions to reflect forecast confidence as well as predicted severity, and that precaution licenses early action only if uncertainty is disclosed. One implication is that legal reliance on an AI-enabled forecast should depend on a disclosed validation dossier, because machine-learning layers add opacity and drift. Disclosure supports human oversight and lets affected people contest restrictions built on overconfident models. The contribution is a test mapping calibrated confidence onto the intensity of proportionality review. The assessment is prospective and doctrinal, evaluating design adequacy rather than operational performance; conclusions remain contingent.

Keywords: Artificial intelligence; Decision support systems; Environmental policy; Air quality forecasting; Uncertainty quantification; Proportionality

1 Introduction

Delhi's Graded Response Action Plan (GRAP) is written to act before the air deteriorates. The schedule in force was issued on 21 November 2025. It directs the Commission for Air Quality Management in the National Capital Region and Adjoining Areas (CAQM) to invoke each stage in advance of the projected Air Quality Index (AQI) level [1]. It relies on a dynamic model and day-to-day forecasts from the India Meteorological Department (IMD) and the Indian Institute of Tropical Meteorology (IITM), and higher projected levels must be likely to persist [1]. Practice has often followed the observed reading instead. An independent assessment found that Stages III and IV were imposed on observed AQI, not forecasts, in winters 2023–24 and 2024–25 [2]. A press analysis of GRAP orders from January 2025 to January 2026 reached a similar conclusion for most impositions [3].[1]

The Supreme Court has pressed in the same direction. In its order of 18 November 2024, the Court stated that the CAQM's duty is to begin Stage III and IV actions in anticipation of the AQI crossing a threshold [4]. The CAQM, it added, cannot wait for the AQI to improve [4]. The order also kept Stage IV in force even if the AQI fell below 450 until further orders [4]. In March 2026 the Court closed the parent petition and directed the registration of five *suo motu* proceedings, and it stated that its earlier directions remain operative [5].

What these directions leave open is how reliable a forecast must be before it may set coercive measures in motion. The 2024 order speaks of anticipation. The schedule speaks of projections that are likely to persist. Neither states, in the text reviewed here, the error rate, calibration or validation record that a forecast must show [1,4]. The 2024 order did notice a related question about inputs, namely which agency's data should count as the official AQI, and left it for hearing [4]. The gap matters because the measures are not trivial. The higher stages restrict construction activity and vehicle entry [2], and the Court has listed compensation for workers affected by suspension of work among the issues for the new proceedings [5]. The forecasts are also imperfect where it counts. Day-1 forecasts detected one of 15 episodes above AQI 400 in winter 2023–24 and five of 14 in 2024–25 [2].[2]

[1] Both assessments are secondary. The independent assessment compares day-1 forecasts scraped from the forecast portal with observations published by the Central Pollution Control Board (CPCB) [2]. The press analysis reviews CAQM orders and counts 13 of 17 impositions as reactive [3]. Section 3.3 and Table 2 test both against the CAQM orders themselves.

[2] Probability of detection is the share of observed episodes that were forecast. The independent assessment averages the day-1 forecast to the CPCB convention of 4 pm to 4 pm. AQI data were missing for 13 days from the forecast portal and 2 days from CPCB in 2024–25, so the counts carry that limitation [2].

The gap will widen if machine learning enters the pipeline. Today the forecasts that feed GRAP come from a physics-based chemical transport model with data assimilation [2,6,7], and the independent assessment recommends adding machine-learning models to correct forecast errors [2]. Research on machine-learning air-quality forecasting is expanding, including in this journal [8,9]. Once a learned layer helps decide when a restriction begins, the legal question changes. It is no longer only whether the model is accurate enough. It is what a decision-maker must be able to show about a layer whose behaviour is learned from data and may drift.

This article asks what validation, reproducibility and uncertainty-quantification thresholds a forecast enabled by artificial intelligence (AI) must meet before it may trigger binding restrictions. GRAP, with the Supreme Court's timing orders, is the single test case. Using a functional comparative method, the article also examines France, California's air districts and China. It argues that proportionality and reasoned decision-making require restrictions to reflect forecast confidence as well as predicted severity, and that precaution licenses early action under uncertainty provided that uncertainty is disclosed. It proposes that legal reliance on an AI-enabled forecast should depend on a disclosed validation dossier.

The article adds to two literatures. Work on forecast verification treats reliability as a scientific property [10-13]. Work on responsible AI for environmental science treats trust as an ethical aim [14]. The reviewed literature does not offer a legal standard that turns measured reliability into the intensity of scrutiny a restriction must survive. This article addresses that gap directly.

Section 2 sets out the concepts and method. Section 3 describes GRAP as design and practice, and Section 4 turns to the courts. Section 5 compares France, California and China. Section 6 develops the confidence-scaled test, and Section 7 proposes the validation dossier. Section 8 discusses limits, and Section 9 concludes.

2 Concepts and method

2.1 A three-layer pipeline and a functional test of AI-enablement

A forecast that triggers a restriction travels through a pipeline of three layers. The first is a numerical forecast: a physics-based chemical transport model, usually with data assimilation, that simulates emissions, weather and chemistry. The second is a machine-learning layer that post-processes, corrects, emulates or replaces part of that forecast. The third is digital dissemination: the portals, bulletins and alerts that carry the result to decision-makers and the

public. The call for this collection asks authors to separate AI-enabled solutions from broader digital interventions. The three layers make that separation concrete.

This article treats a component as AI-enabled, for legal purposes, when two conditions hold. Its mapping from inputs to outputs is learned from data instead of being fixed by physical law or engineering rule. Its output also changes the trigger decision instead of only its presentation. Dissemination tools fail the first condition. A conventional numerical model fails it too, however complex. A machine-learning bias-correction layer that shifts a forecast across a stage threshold meets both. The test is narrower than the definition in the EU Artificial Intelligence Act, which turns on inference from inputs and varying levels of autonomy [15]. It isolates the feature that matters here, namely behaviour that is learned and may change with the data.

The test also fixes the present position. The forecasts that feed GRAP come from the Air Quality Early Warning System (AQEWS), which runs a chemical transport model with data assimilation [2,6,7]. The reviewed sources document no operational machine-learning layer in that chain.[3] AI-enablement is therefore a prospective question. The article evaluates the design adequacy of legal standards, not the operational performance of an AI-enabled system. Hybrid designs that combine physical models with machine learning are a direction of travel in Earth-system science [16], so the prospect is not remote.

2.2 Forecast verification in legal terms

The legal argument needs a small vocabulary from forecast verification [10-13]. Bias is the average difference between forecast and observation. Skill is accuracy relative to a reference forecast. For a categorical forecast, such as whether the AQI will exceed a threshold, the probability of detection is the share of observed episodes that were forecast. The false alarm ratio is the share of forecast episodes that did not occur. Calibration is the agreement between stated probabilities and observed frequencies. Sharpness is how concentrated the forecast distribution is. A prediction interval has good coverage when the stated share of observations falls inside it. Lead time is how far ahead the forecast is made. In Delhi the forecasts perform better on the first two days than on the third [2].

Two points follow for law. These measures apply to any forecast, not only to AI-enabled ones. What differs for a learned layer is that its errors can change with the data on which it

[3] The AIRWISE paper describes the core of the operational system as a chemical transport model with three-dimensional variational data assimilation [6]. The independent assessment lists machine-learning bias correction as a recommendation, not as a feature of the system [2].

was trained, so verification must be repeated and disclosed. Proper scoring rules, which reward honest probabilities [12,13], also matter. A restriction scaled to confidence needs a forecast that reports its confidence honestly.

2.3 Comparative design

The comparison is functional [17]. The *tertium comparationis* is one system-neutral sentence: how a legal order authorises coercive environmental measures on the strength of uncertain forecasts. India supplies the test case. France, California's air districts and China are chosen because each lets forecasts, or warnings built on forecasts, set restrictions in motion, at different levels of government and through different decision-makers. The comparison is constitutive, not ornamental, because the comparators supply the design features tested in Section 6.

Each comparator is answered on four questions: the regulatory framework, its similarity to India, its difference from India and the lesson that could be implemented in India. Instruments from all four jurisdictions are coded for trigger basis (observed, forecast-based or discretionary), decision-maker, lead time, confidence requirement and rollback. Table 1 sets out the coding with a methodological note.

The method has limits. Functional comparison presumes that different systems solve the same problem, and critics argue that it neglects legal culture [18] and the politics of comparison [19]. Comparing a civilian system such as France with a common-law system such as India calls for care about which rules actually operate, a point that Sacco's work on legal formants makes [20]. The regimes sit at national, prefectoral and regional levels, so the comparison is between functions, not institutions. Every claim that one design is better is tested against Kamba's framework [21].

2.4 Search protocol and status of the assessment

The claim that no reviewed instrument states the confidence a forecast must carry is a negative claim, and rests on the search conducted across CAQM's own notifications, the Supreme Court's orders, government and legislative databases, and the comparative, verification and proportionality literature cited throughout this article. Press reports and think-tank studies are used for practice claims only and never as sole support for a doctrinal claim. The assessment is dated 6 October 2026 and is prospective and doctrinal.

3 GRAP: design and practice

3.1 Legal architecture

The Commission for Air Quality Management in National Capital Region and Adjoining Areas Act, 2021 (Act 29 of 2021) gives the CAQM statutory power to issue directions for air quality [1,22]. Section 12(2)(xi) lets it direct action to control pollution from any source it identifies, and section 14 makes non-compliance an offence [22].[4] The Supreme Court's judgment of 12 March 2026 records that the Court itself facilitated the CAQM's creation, and that GRAP is the Commission's central instrument [5]. The GRAP schedule now in force is dated 21 November 2025 [1]. It sets four stages by AQI band, and directs the Sub-Committee to invoke each stage in advance of the corresponding band, provided the projection is likely to persist [1].

Clause 3 supplies a second trigger. Even where forecasts show no breach, an extreme meteorological condition or an episodic event that breaches the threshold invokes that stage immediately [1]. Two logics sit side by side, then. One is anticipatory and forecast-led. The other is reactive and event-led, and it fires exactly when the first was supposed to have pre-empted it. The schedule does not rank them, or say which controls when a forecast and an observed breach point in different directions.

3.2 Forecast pipeline and lead-time skill

Delhi's forecasts come from the Air Quality Early Warning System, which the Ministry of Earth Sciences launched in 2018 and paired with a Decision Support System in 2021 [2,6]. The core model is WRF-Chem at 400-metre resolution, run by the Indian Institute of Tropical Meteorology, with data assimilation from monitoring stations, satellite aerosol depth and fire counts [2,7]. A second model, IMD-SILAM, feeds the Meteorological Department's own bulletins [2]. Neither is a machine-learning system in current operation. The reviewed literature recommends adding one, specifically to correct bias in the physics-based output [2].

An independent verification study gives the first public account of how well this pipeline performs [2]. For episodes above AQI 400, the day-1 forecast correctly flagged 1 of 15 in winter 2023-24 and 5 of 14 in 2024-25 [2]. The false alarm ratio ran near 90 per cent in the first year and 38 per cent in the second [2]. For episodes above AQI 300, detection exceeded 90 per cent in both years [2]. Pollutant concentrations were consistently under-predicted, by roughly 18 to 25 micrograms per cubic metre for PM2.5 [2]. A separate account by the

[4] Section 12(2)(xi) empowers the Commission to issue binding written directions, and section 14 makes non-compliance an offence punishable with up to five years' imprisonment or a fine of up to one crore rupees, or both.

system's own developers reports that skill is strongest on the first two days and weaker on the third.[5] Two things follow together, and they cut against each other. Forecasts of the milder categories are reasonably reliable. Forecasts of exactly the severe episodes GRAP is built to intercept are not, and grow less reliable the further out they reach.

Table 1. AQEWS forecast performance for 'severe and above' (AQI>400), day-1 forecast

Winter	Accuracy (%)	False alarm ratio (%)	Probability of detection (%)	False negative rate
2023–24	82.73	90.9	6.67	0.93
2024–25	91.24	37.5	35.71	0.64

Methodological note. (a) Figures are reproduced from an independent verification study comparing WRF-Chem (400m) day-1 forecasts against CPCB observations [2]; thresholds follow the study's own categories. (b) Computed by the study's authors, not by the present author. (c) Alternative readings: AQI data were unavailable for 13 days (AQEWS) and 2 days (CPCB) in 2024–25, which may bias the 2024–25 figures in an unstated direction [2]. (d) Assessment date of the underlying study: October 2025 [2]. (e) These are retrospective, operational performance figures for the existing physics-based pipeline; they are not a claim about any AI-enabled layer, which remains prospective in this article (Section 2.1).

3.3 Practice, coded by trigger basis

Design and practice diverge in the direction the doctrine most needs to worry about. The same verification study concludes that the CAQM imposed GRAP Stages III and IV on observed AQI, not on forecasts, in both winters it reviewed [2]. A separate review of orders across the following winter reaches a similar count [3]. Table 2 sets out a sample of individually sourced impositions to illustrate the coding. It is not the full run the press review claims to have counted, and that fuller log is a task for the next verification pass.

The pattern is not universal. A Stage III order of 16 January 2026 invokes the stage explicitly on a forecast that AQI would breach 400 within days, with the observed level still at 354. Officials described it as a proactive measure [23,24]. That order shows the schedule's anticipatory logic operating as written. The more common pattern in the sample is the reactive one. An order dated 13 December 2025 records that the AQI had already breached 450 before Stage IV was invoked with immediate effect [25,26]. A separate order the same day recorded a rising trend already past the Stage III threshold [27]. Revocation follows the same logic in reverse. The Stage IV revocation of 24 December 2025 cites an observed improvement to AQI 271, not a forecast that the improvement would hold [26]. Both

[5] This claim is drawn from Ghude et al., as summarised at second hand in the verification study [2]. The primary AIRWISE paper [6] should be checked directly for the same claim before submission.

reviewed winters likewise lifted a stage once the observed AQI fell back under the lower band, and the schedule states no separate test for revocation [2].

Table 2. Sample of individually sourced GRAP impositions (illustrative, not exhaustive)

Date	Stage	Observed AQI at order	Forecast basis stated	Trigger coded	Source
12–14 Nov 2024	III	Crossed 401 on 12/13 Nov; invoked from 14 Nov	Not stated as the basis	Observed (delayed)	[4]
17–18 Nov 2024	IV	Crossed 450 on 17 Nov, 6pm; implemented from 18 Nov	Not stated as the basis	Observed	[4]
13 Dec 2025	III then IV	Rising trend crossed 401; later breached 450 same day	Not stated as the basis	Observed	[25,27]
24 Dec 2025	Revocation of IV	Improved to 271	Not the stated basis for revocation	Observed (revocation)	[26]
16 Jan 2026	III	354 (below the 401 threshold)	Forecast of breach within days, described as a proactive measure	Forecast-based	[23,24]

Methodological note. (a) "Observed" = the order's own reported text cites the AQI level at or after the threshold was crossed as its stated basis; "Forecast-based" = the order cites a projected future level as its stated basis. (b) Coded by the author from the CAQM orders and news reports cited in each row; not independently cross-checked by a second coder. (c) Alternative readings: several orders cite both a rising observed trend and a forecast in the same sentence; where a discretionary or precautionary phrase appears alongside a crossed threshold, the row is coded by the reading given greatest emphasis in the reported text, which is a judgment call. (d) Assessment date: 22 September 2026. (e) This is a small, individually verified sample, not the full population of impositions across the period; a complete order-by-order log remains outstanding, and the aggregate 13-of-17 figure reported elsewhere should be treated as a secondary count pending that log [3].

4 Judicial overlay and doctrine

4.1 The duty to act early

The Supreme Court's order of 18 November 2024 states the doctrine most directly. Stage III requires action once the AQI crosses 401. The record shows the Sub-Committee waited to invoke it from 14 November, although the level had already crossed on 12 or 13 November [4]. The Court held that approach contrary to its own order of 29 October 2018, which had

directed the Commission's predecessor to take pre-emptive steps without strict adherence to the stages [4].[6] It called the Commission's approach entirely wrong, and said in terms that the Commission cannot wait for the AQI to improve [4]. It also kept Stage IV in force even after the AQI fell below 450, until further orders [4].

A further order, reported at 5 December 2024, brought the duty down to specific numbers: Stage III at AQI 350, Stage IV at 400 [2]. Both sit ahead of the schedule's own 401 and 451 bands [1].[7] The CAQM's schedule now carries a matching clause: even where the forecast shows no breach, an extreme condition that produces one triggers the stage at once [1]. Read together, the orders and the amended clause push toward earlier and more frequent triggering. Whether the underlying forecast is reliable enough to justify that earlier point is not a question either order asks.

4.2 Precaution, proportionality and reasons

Indian environmental doctrine reads precaution as a licence to act before scientific certainty, not as a demand for more of it before acting. Vellore Citizens' Welfare Forum v. Union of India [28] read the precautionary principle into Article 21 and shifted the burden onto the party proposing an activity to show it is environmentally benign. A.P. Pollution Control Board v. Prof. M.V. Nayudu (Retd.) [29] went further and cautioned against demanding the same rigour of proof in environmental cases that ordinary adjudication expects, precisely because the underlying science is often unsettled. Applied to GRAP, that doctrine supports the Supreme Court's 2024 orders. It sits awkwardly with a rule that would raise the evidentiary bar before a restriction may bind, which is what a confidence threshold does.

Proportionality supplies the other half of the test. Modern Dental College and Research Centre v. State of Madhya Pradesh [30] adopted a four-stage structure: legitimate purpose, a rational connection between measure and purpose, necessity, and balance. Necessity asks whether a less restrictive measure would serve as well; balance weighs the measure's benefit against the burden it imposes. None of that structure turns on how confident the underlying prediction is. It asks whether the restriction fits the harm, not whether the harm was reliably forecast. Kranti Associates Pvt. Ltd. v. Masood Ahmed Khan [31] supplies the missing piece from a different angle: a body exercising power that affects rights must record reasons, and

[6] The order of 29 October 2018 is quoted from the Supreme Court's own later recitation of it [4]; the 2018 order itself has not yet been retrieved and should be checked directly.

[7] The 5 December 2024 order and its 350/400 figures are known only through the verification study's account [2]; the order itself should be retrieved and cited directly before submission.

an unreasoned order is itself arbitrary. A GRAP order that states an AQI figure and a stage, without saying anything about the forecast's own track record, meets the letter of that duty and very little of its spirit.

4.3 The asymmetry

A false negative and a false alarm sit on opposite sides of the same forecast, and the record examined here has attended to only one of them. Every hearing traced in this article concerns delay: a stage invoked too late, a restriction eased too soon [2,4]. None asks what a restriction costs when the severe episode it was invoked to intercept does not, in fact, arrive. Construction bans and vehicle restrictions bear on daily-wage labour and haulage, and the reviewed literature on GRAP names that cost without quantifying it [2]. This is not a claim that the Court was wrong to press for earlier action. It is a claim that the asymmetry between the two error types has gone unexamined. A system built entirely to guard against the first kind of error carries no correspondingly explicit check on the second.

5 Comparative analysis

The comparators are chosen for a shared feature, not a shared architecture. Each authorises coercive restriction on a forecast of a pollution episode, or on that forecast's persistence, before the episode is observed to have occurred. What differs is who decides, at what level of government, and what the trigger actually requires. That variation is the material of the comparison. Each system is tested here on four questions: its regulatory framework, its similarity to the Indian design, its difference from it, and the lesson it offers.

5.1 France

France's Code de l'environnement, art. L223-1, requires the *préfet* to inform the public and take measures once an alert threshold is reached or is at risk of being reached [32].[8] That second limb is explicitly anticipatory, in the same register as GRAP's advance-invocation clause. Articles R223-1 to R223-5 delegate the operative detail to a joint ministerial order, which must fix trigger conditions by reference to measured levels, foreseeable developments

[8] Full official title and the consolidated version date of the provision to be confirmed against Légifrance at the audit stage.

and population exposure [33]. The governing instrument is the *arrêté* of 7 April 2016, made jointly by the ministers for environment, health, agriculture, economy and transport [34].[9]

Similarity. Like GRAP, the French design ties a coercive measure to a forecast component, not only to an observed reading, and it names foreseeable evolution as a factor the triggering order must weigh [33]. *Difference.* The French decision-maker is a single named official, the departmental prefect, answerable through ordinary administrative-law channels, where the CAQM is a specialised multi-member commission whose sub-committee decides collectively [1,32]. The trigger is also structured around two named thresholds, information-and-recommendation and alert, rather than four graded stages [32]. *Lesson.* Locating the decision in one accountable office, rather than a sub-committee acting under a schedule, is one way to keep responsibility traceable when a forecast turns out wrong. That traceability does not by itself supply a confidence standard. It does identify who would answer for its absence.

5.2 California's air districts

California regulates by air basin, not by state-wide rule, and the two largest basins each run a forecast-triggered burn ban. The Bay Area Air Quality Management District's Regulation 6, Rule 3, makes it unlawful to burn wood on a day the District forecasts will exceed a PM2.5 threshold [35]. Amendments the District's Board adopted on 1 October 2025 lowered that threshold and made the alert year-round rather than seasonal [35].[10] The South Coast Air Quality Management District's Rule 445 runs the same design under the name Check Before You Burn [36]. Its threshold was most recently proposed for reduction, from 29 to 25 micrograms per cubic metre, in a rule-making noticed in September 2025 [36].[11]

Similarity. Both districts trigger entirely on a next-day forecast; there is no observed-breach fallback comparable to GRAP's clause 3, and no discretionary override on the record reviewed [35,36]. *Difference.* The restriction reaches one narrow behaviour, residential wood burning, not the broad construction and vehicle measures GRAP's higher stages impose [35]. Enforcement runs through a graduated ticket, too, rather than the closure and entry-ban apparatus used in Delhi [1]. *Lesson.* A single-purpose, forecast-only trigger is administrable

[9] The arrêté's full Journal officiel citation to be added; the version consulted here is a secondary compilation, not the Journal officiel text itself.

[10] The exact revised threshold figure adopted on 1 October 2025 was not confirmed in the sources reviewed and should be checked against the Air District's adopted rule text before submission.

[11] Whether the proposed 25 microgram threshold was adopted, and on what date, was not confirmed in the sources reviewed.

partly because the restricted conduct is narrow and the harm of a false alarm is a lost evening's fire, not a halted construction site. Whether the design scales to a stage that idles daily-wage labour is doubtful, precisely because that asymmetry of stakes is what makes a bare forecast trigger tolerable here.

5.3 China

China's Law on the Prevention and Control of Atmospheric Pollution, as revised in 2018, requires local governments to grade pollution warnings [37]. Once a warning issues, a contingency plan may command production limits, vehicle restrictions and school closures [37].[12] The framework law leaves the warning grades themselves to be set beneath it. Beijing's own contingency plan supplies the operative detail at that level. A red alert issues on a *forecast* daily AQI above 200 sustained for four days, above 300 for two days, or a forecast reaching 500 outright [38].[13]

Similarity. The persistence requirement is functionally the same device as GRAP's own clause that a projected level must be likely to sustain before it counts [1,38]. Both systems also grade the response to forecast severity. *Difference.* China states the persistence condition in hours, not in a general phrase [38]. It also states that condition locally, in the contingency plan a city government adopts, rather than in the single national schedule GRAP uses for the whole of the NCR [1]. *Lesson.* Converting persistence into a stated number of hours removes one axis of discretion the CAQM's schedule currently leaves open. It does so, though, without touching the harder question: how confident must the underlying forecast be over that stated horizon?

Table 1 draws the four regimes together on a common schema, coded for legal basis, decision-maker, trigger basis, persistence condition and discretionary override.

Table 3. Trigger schema across four regimes

Regime	Legal basis	Decision-maker	Trigger basis	Persistence/lead-time condition	Discretionary override

[12] Article numbers cited from a secondary translation; the current Chinese text and any amendment since 2018 should be checked before submission.

[13] The Beijing figures are drawn from the municipality's published contingency plan; whether the plan has since been superseded should be checked before submission.

Regime	Legal basis	Decision-maker	Trigger basis	Persistence/lead-time condition	Discretionary override
India (GRAP)	CAQM Act 2021, s.12; GRAP Schedule (rev. 21.11.2025), cll. 1-3 [1,22]	CAQM Sub-Committee	Forecast-based, with observed fallback	"Likely to sustain" (unquantified) [1]	Yes — extreme/episodic event overrides forecast, cl. 3 [1]
France	Code de l'environnement, art. L223-1; arts. R223-1–R223-5; arrêté of 7.4.2016 [32,33,34]	Départmental préfet	Observed or "at risk of being reached"	Not separately stated in the texts reviewed [32]	Not identified in the texts reviewed
California (BAAQMD/SCAQMD)	Reg. 6, Rule 3 (BAAQMD); Rule 445 (SCAQMD) [35,36]	District Executive Officer	Forecast-only (next-day)	Single-day forecast; no persistence condition [35,36]	None identified
China	Air Pollution Prevention and Control Law 2018, arts. 96–97; Beijing Contingency Plan [37,38]	Local government (graded by warning level)	Forecast-based, graded by level	Stated in hours (e.g. AQI>200 for 96h, or >300 for 48h) [38]	Not identified in the texts reviewed

Methodological note. (a) Classification criteria: "forecast-based" denotes a trigger whose text names a projected or forecast value as sufficient; "observed fallback" or "observed" denotes reliance on a measured reading. (b) Coded by the author from the primary and secondary sources cited in each cell; a single coder, not cross-checked by a second reader at this draft stage. (c) Alternative readings: the French "at risk of being reached" language could be read as closer to a discretionary standard than a forecast trigger; that reading is not adopted here but is not excluded. (d) Assessment date: 22 September 2026. (e) This table records the design stated in each instrument's text, not its operational record; Table 2 addresses the Indian design against its own practice.

6 Confidence-scaled proportionality

6.1 Where confidence enters the test

Modern Dental College's four stages ask about purpose, fit, necessity and balance [30]. None of the four, on its face, asks how sure the decision-maker was that the harm would occur. That omission is not a flaw in the test. It reflects an assumption, reasonable for most regulatory contexts, that the factual premise for a measure is either established or not yet ripe for one. GRAP does not fit that assumption. It is built to act on a premise, a projected AQI, that is stated with a number but never with an error bar. Confidence belongs inside the test's own structure, not as an external check on it. It bears most directly on necessity. A measure

that would be proportionate against a highly reliable forecast may fail necessity against one barely better than a coin toss. A less restrictive advisory would serve just as well until the picture clarifies. It bears on balance too. The weight on one side of that scale is not the harm itself but the expected harm, and expectation is exactly what a forecast's confidence interval is for.

6.2 A cost-loss bridge

Meteorology already has a name for this problem, and a simple answer to it. A decision-maker facing a forecast should act when the cost of acting is less than the probability of the event times the loss avoided by acting [10,13].[14] That cost-loss ratio is not a legal test, but it translates directly into one. Where the cost of the higher-stage measures is high, construction halted, vehicles turned back, the probability the forecast must clear before that measure is proportionate rises with it. Where the measure is cheap, an advisory, a public notice, a lower probability justifies it. This is not a new value inserted into Indian doctrine. It is the ordinary balancing stage of Modern Dental College, given a number to balance.

6.3 Reconciling precaution with proportionality

The tension flagged in Section 4.2 can now be stated precisely, and resolved without asking Vellore Citizens or Nayudu to do work they were never meant to do. Precaution answers a different question from the one asked here. It says action need not wait for certainty that a harm will occur. It does not say a decision-maker may treat an unquantified guess as equivalent to a well-verified forecast, or that disclosure of the guess's reliability would somehow undercut the licence to act early.[15] Acting under uncertainty and disclosing that uncertainty are not opposed to each other. A rule that suppresses disclosure, to preserve the appearance of confident action, is itself a choice. It forecloses the affected party's ability to contest the measure on its own terms. Precaution licenses Stage III at AQI 350 instead of 401 [2]. It does not license silence about how often a forecast at that point has been wrong before.

Table 4. Confidence-severity matrix (normative proposal)

[14] The cost-loss framework is standard in forecast verification texts [10,11]; a direct citation to its originating source, rather than the general verification texts, should be added at the audit stage.

[15] This paragraph should be anchored to named sources on regulatory science and uncertainty at the audit stage; it is left here as a reasoned claim pending that citation, not attributed to any specific author.

Predicted severity	Low confidence (wide interval / poor recent calibration)	High confidence (narrow interval / good recent calibration)
Advisory range (below Stage I)	Public advisory permissible; no coercive measure	Public advisory permissible; no coercive measure
Stages I–II	Advisory plus voluntary measures; a coercive measure requires a disclosed dossier and a stated fallback if the forecast proves wrong	Coercive measure permissible with a disclosed dossier
Stages III–IV	A coercive measure requires either corroborating observed data or a dossier showing the specific forecast's own recent skill at this lead time; bare model output does not suffice standing alone	Coercive measure permissible; dossier disclosure remains mandatory under Section 7

This table states a proposal developed in Section 6, not a coded description of any existing instrument. It applies the cost-loss logic of Section 6.2 to GRAP's own four stages and is offered as a starting point for the reform discussion in Section 8, not as a finished rule.

7 The validation dossier

A workable rule needs a concrete deliverable, not only a principle, and design-support literature outside law supplies a template. A recent decision-support-system study builds its tool through a staged validation process combining quantitative testing against past cases with structured expert review before the tool is treated as fit for use [39].[16] That structure travels well into this setting. A validation dossier, on that model, would disclose the forecast's skill and calibration over a stated period, and its lead-time degradation. It would also name the bias-correction method applied, if the layer is AI-enabled, and the date the assessment was last run. It would state plainly whether the layer's classification is prospective, a design-adequacy claim, or operational, a claim about how the system has actually performed, and would not conflate the two.

Confidence in an automated tool is itself measurable. A recent model for scoring risk in AI-assisted judgment, under a structured, criteria-based method, offers one way to make that measurement auditable rather than impressionistic [40].[17] None of this requires new primary legislation. The Information Technology Act, 2000, sections 70 and 70A, already lets

[16] The PRABS study concerns brownfield risk assessment, not air quality; it is used here only for its five-stage validation method, not for any substantive claim about forecasting.

[17] This reference was identified by its abstract and publication record; its applicability to forecast-confidence scoring specifically, rather than construction-risk scoring generally, should be checked before the claim built on it is relied on.

government designate a system as protected critical information infrastructure [41]. Security obligations follow that designation, a route available if a forecasting pipeline is ever brought within it. The CERT-In Directions of 28 April 2022 already require certain entities to report cyber incidents and preserve logs [42]. That is most of the audit trail a dossier would need for the integrity of its inputs. Where the pipeline touches personal location or health data, for instance in a last-mile alert system, the Digital Personal Data Protection Act, 2023, governs that layer separately from anything proposed here [43]. The dossier does not need a new statute. It needs a rule that legal reliance on the forecast is conditioned on the dossier's existence and its disclosure to whoever the restriction binds.

8 Discussion and limits

Three limits qualify what has been argued. First, the gap between GRAP's design and its practice is not itself a confidence problem. It is a compliance problem, and no dossier fixes a Sub-Committee's decision to wait for an observed reading it was already required to anticipate. The dossier answers a different, narrower question: once a decision-maker does rely on a forecast, what must that reliance be able to show. Second, the comparators are not equally weighted. France and China regulate through instruments built for exactly this kind of graded response. California's districts regulate one behaviour with one blunt tool, and the lesson drawn from them in Section 5.2 is correspondingly modest. Third, this assessment is prospective and doctrinal. It tests whether the legal standard, as it stands, is adequate to a forecast pipeline that has not yet gone AI-enabled in documented operation [2]. It does not test whether any existing tool performs well or badly. That is a deliberate limit, not an oversight: a standard is worth stating before the harder empirical question of a specific system's performance can even be asked.

9 Conclusion

GRAP is written to act on forecasts and is shown, on the best available evidence, to act mostly on observations instead [2,3]. The Supreme Court has pressed hard against that gap, and rightly so, without once asking how reliable the forecast behind an earlier trigger needs to be [4]. France, California and China each show a different piece of an answer. One offers a named accountable decision-maker, one a single-purpose trigger scaled to the stakes it carries, and one a persistence condition stated in hours rather than a general phrase. None of the three states a confidence threshold outright, which suggests the gap identified here is not a peculiarly Indian failure. A validation dossier, disclosed wherever a forecast is relied on to

bind, would let proportionality do the work precaution was never meant to do alone. It would scale the restriction not only to how bad the forecast harm would be, but to how much the forecast itself can be trusted.

Declarations

Funding

This research received no external funding.

Competing interests

The author declares no competing interests.

Data availability

Not applicable.

Ethics declarations

Not applicable. This study did not involve human participants, animal subjects, or the collection of primary data.

Clinical trial number

Not applicable.

Consent to participate

Not applicable.

Consent to publish

Not applicable.

References

1. Commission for Air Quality Management in National Capital Region and Adjoining Areas. Graded Response Action Plan (GRAP) for the National Capital Region (NCR) (Revision: 21.11.2025) [Internet]. New Delhi: CAQM; 2025 [cited 2026 Sep 21]. Available from: https://caqm.nic.in/WriteReadData/LINKS/GRAP%20Schedule7e3bd726-7308-484f-995f-7cb3fa705b8d.pdf
2. Ignatious S, Rafiuddin M. How well can Delhi predict air quality? Insights for India's decision support systems [Internet]. New Delhi: Council on Energy, Environment and

Water; 2025 Oct 1 [cited 2026 Sep 21]. Available from: https://www.ceew.in/publications/designing-effective-air-quality-decision-support-and-pollution-warning-alert-systems

3. ThePrint. Delhi brought GRAP reactively after AQI crossed limit—13 out of 17 times this winter [Internet]. ThePrint; 2026 Feb [cited 2026 Sep 21]. Available from: https://theprint.in/environment/delhi-brought-grap-reactively-after-aqi-crossed-limit-13-out-of-17-times-this-winter/2855807/
4. M.C. Mehta v. Union of India. Writ Petition (Civil) No. 13029/1985 (In re: Commission for Air Quality Management (Air Pollution)), record of proceedings of 18 November 2024, Supreme Court of India (Oka and Masih JJ.) [Internet] [cited 2026 Sep 21]. Available from: https://indiankanoon.org/doc/122095945/
5. M.C. Mehta v. Union of India and Others. Writ Petition (Civil) No. 13029/1985, judgment of 12 March 2026, 2026 INSC 383, Supreme Court of India (Surya Kant CJ, Bagchi and Pancholi JJ.) [Internet] [cited 2026 Sep 21]. Available from: https://www.supremecourtcases.com/m-c-mehta-v-union-of-india-and-others-46/
6. Ghude SD, Govardhan G, Kumar R, Yadav PP, Jat R, Debnath S, et al. Air quality warning and integrated decision support system for emissions (AIRWISE): enhancing air quality management in megacities. Bull Am Meteorol Soc. 2024;105(12). https://doi.org/10.1175/BAMS-D-23-0181.1
7. Govardhan G, Ghude SD, Kumar R, Sharma S, Gunwani P, Jena C, et al. Decision support system version 1.0 (DSS v1.0) for air quality management in Delhi, India. Geosci Model Dev. 2024;17:2617-2640. https://doi.org/10.5194/gmd-17-2617-2024
8. Rana R, Kumar N. Air pollution forecasting trends and regional patterns in the Asia Pacific through bibliometric analysis. Discov Environ. 2025;3:203. https://doi.org/10.1007/s44274-025-00392-7
9. Rybarczyk Y, Zalakeviciute R. Machine learning approaches for outdoor air quality modelling: a systematic review. Appl Sci. 2018;8(12):2570. https://doi.org/10.3390/app8122570
10. Murphy AH, Winkler RL. A general framework for forecast verification. Mon Weather Rev. 1987;115(7):1330-1338. https://doi.org/10.1175/1520-0493(1987)115<1330:AGFFFV>2.0.CO;2
11. Jolliffe IT, Stephenson DB, editors. Forecast verification: a practitioner's guide in atmospheric science. 2nd ed. Chichester: Wiley; 2012.
12. Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc. 2007;102(477):359-378. https://doi.org/10.1198/016214506000001437
13. Gneiting T, Balabdaoui F, Raftery AE. Probabilistic forecasts, calibration and sharpness. J R Stat Soc Series B Stat Methodol. 2007;69(2):243-268. https://doi.org/10.1111/j.1467-9868.2007.00587.x
14. McGovern A, Ebert-Uphoff I, Gagne DJ 2nd, Bostrom A. Why we need to focus on developing ethical, responsible, and trustworthy artificial intelligence approaches for environmental science. Environ Data Sci. 2022;1:e6. https://doi.org/10.1017/eds.2022.5

15. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Off J Eur Union. 2024;L 2024/1689.

16. Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, et al. Deep learning and process understanding for data-driven Earth system science. Nature. 2019;566(7743):195-204. https://doi.org/10.1038/s41586-019-0912-1

17. Zweigert K, Kötz H. An introduction to comparative law. 3rd ed. Weir T, translator. Oxford: Clarendon Press; 1998.

18. Legrand P. The impossibility of 'legal transplants'. Maastricht J Eur Comp Law. 1997;4(2):111-124.

19. Frankenberg G. Critical comparisons: re-thinking comparative law. Harv Int Law J. 1985;26(2):411-455.

20. Sacco R. Legal formants: a dynamic approach to comparative law (installment I of II). Am J Comp Law. 1991;39(1):1-34. Installment II appears at 39(2):343-401.

21. Kamba WJ. Comparative law: a theoretical framework. Int Comp Law Q. 1974;23(3):485-519.

22. The Commission for Air Quality Management in National Capital Region and Adjoining Areas Act, 2021 (Act 29 of 2021), ss. 12(2)(xi), 14 [Internet]. New Delhi: Government of India [cited 2026 Oct 6]. Available from: https://www.indiacode.nic.in/bitstream/123456789/16898/1/A2021-29.pdf

23. Business Today Desk. GRAP-III curbs imposed across Delhi-NCR as AQI threatens to turn severe [Internet]. Business Today; 2026 Jan 16 [cited 2026 Sep 22]. Available from: https://www.businesstoday.in/amp/india/story/grap-iii-curbs-imposed-across-delhi-ncr-as-aqi-threatens-to-turn-severe-check-what-is-banned-511273-2026-01-16

24. Deccan Herald (PTI). Grap 3 curbs enforced in Delhi-NCR as air quality likely to turn 'severe' in coming days [Internet]. Deccan Herald; 2026 Jan 16 [cited 2026 Sep 22]. Available from: https://www.deccanherald.com/amp/story/india%2Fdelhi%2Fgrap-3-curbs-enforced-in-delhi-ncr-as-air-quality-likely-to-turn-severe-in-coming-days-3864658

25. Commission for Air Quality Management in National Capital Region and Adjoining Areas. Order No. 120017/27/GRAP/2021/CAQM, 13 December 2025 (GRAP Stage-IV invocation) [Internet]. New Delhi: CAQM [cited 2026 Sep 22]. Available from: https://caqm.nic.in/WriteReadData/LINKS/GRAP%20Stage-IV%20invocation%20order%2013122025ed5d1f9d-5ea1-4238-896c-17979da39e91.pdf

26. Press Information Bureau. CAQM revokes GRAP IV as AQI improves to 271 [Internet]. New Delhi: PIB; 2025 Dec 24 [cited 2026 Sep 22]. Available from: https://www.pib.gov.in/PressReleasePage.aspx?PRID=2208266&reg=3&lang=1

27. IANS. Delhi slips into 'severe' air category, stage-III GRAP curbs enforced across NCR [Internet]. The Hawk; 2025 Dec 13 [cited 2026 Sep 22]. Available from:

https://www.thehawk.in/news/science/delhi-slips-into-severe-air-category-stage-iii-grap-curbs-enforced-across-ncr

28. Vellore Citizens' Welfare Forum v. Union of India, (1996) 5 SCC 647 (Supreme Court of India)
29. A.P. Pollution Control Board v. Prof. M.V. Nayudu (Retd.), (1999) 2 SCC 718 (Supreme Court of India)
30. Modern Dental College and Research Centre v. State of Madhya Pradesh, (2016) 7 SCC 353 (Supreme Court of India)
31. Kranti Associates Pvt. Ltd. v. Masood Ahmed Khan, (2010) 9 SCC 496 (Supreme Court of India)
32. Code de l'environnement, art. L223-1 [Internet]. Paris: Légifrance [cited 2026 Sep 22]. Available from: https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000039785271/
33. Code de l'environnement, arts. R223-1 to R223-5 [Internet]. Paris: Légifrance [cited 2026 Sep 22]. Available from: https://www.legifrance.gouv.fr/codes/id/LEGISCTA000006159472/
34. Arrêté du 7 avril 2016 relatif au déclenchement des procédures préfectorales en cas d'épisodes de pollution de l'air ambiant [Internet]. Paris: Journal officiel de la République française [cited 2026 Sep 22]. Available from: https://www.legifrance.gouv.fr/loda/id/JORFTEXT000032376671/
35. Bay Area Air Quality Management District. Regulation 6, Rule 3: Wood-Burning Devices, as amended 1 October 2025 [Internet]. San Francisco: BAAQMD [cited 2026 Sep 22]. Available from: https://www.baaqmd.gov/rules-and-compliance/rules/reg-6-rule-3-woodburning-devices
36. South Coast Air Quality Management District. Proposed Amended Rule 445 - Wood-Burning Devices, notice posted 5 September 2025 [Internet]. Diamond Bar (CA): South Coast AQMD [cited 2026 Sep 22]. Available from: https://ceqanet.lci.ca.gov/2025090213
37. Law of the People's Republic of China on the Prevention and Control of Atmospheric Pollution (as amended 2018), arts. 96-97 [Internet] [cited 2026 Sep 22]. Available from: http://en.cciced.net/NEWSCENTER/LatestEnvironmentalandDevelopmentNews/20150 9/P020160908584777560703.pdf
38. Beijing Municipal People's Government. Heavy Air Pollution Contingency Plan of Beijing Municipality [Internet]. Beijing [cited 2026 Sep 22]. Available from: https://english.beijing.gov.cn/latest/lawsandpolicies/202007/t20200723_1957677.html
39. Mahammedi C, Mahdjoubi L, Booth C, Butt TE, Al-mhdawi MKS. The design and validation of a decision support system (DSS) for the preliminary risk assessment of brownfield sites (PRABS). Smart Sustain Built Environ. 2024. https://doi.org/10.1108/SASBE-11-2023-0364

40. Mohamed AH, Al-Mhdawi MKS, Shadlooye G, Ojiako U. Development of a hierarchical fuzzy assessment model for quantifying GenAI risks in sustainable construction projects. Front Built Environ. 2026;12:1842237. https://doi.org/10.3389/fbuil.2026.1842237

41. Information Technology Act, 2000 (Act 21 of 2000), ss. 70, 70A [Internet]. New Delhi: Government of India [cited 2026 Sep 22]. Available from: https://aicte.gov.in/sites/default/files/itact2000.pdf

42. Indian Computer Emergency Response Team. Directions under sub-section (6) of section 70B of the Information Technology Act, 2000 relating to information security practices, procedure, prevention, response and reporting of cyber incidents, 28 April 2022 [Internet]. New Delhi: CERT-In [cited 2026 Sep 22]. Available from: https://www.cert-in.org.in

43. Digital Personal Data Protection Act, 2023 (Act 22 of 2023) [Internet]. New Delhi: Government of India [cited 2026 Sep 22]. Available from: https://www.indiacode.nic.in