The weekly model record

Assessment: 18 October 2026. Training data extend through 17 October 2026 at 16:00 IST.

This record shows how two model families predict pollutant concentrations. It is a record of model performance, not a summary of the week's air quality.

RF models trained	BRT models trained	BRT screening results
7 of 7	7 of 7	4 pass; 3 fail

What changed since 12 October

The largest RF out-of-bag decline was SO2, at -0.0022 R^2. PM2.5 BRT scored 0.7431, which was 0.0151 below its screening floor.

What the scores establish

The models have been fitted and scored under the tests described here. A completed fit does not mean that the model passed its screening test. These scores do not by themselves establish a change in emissions, the cause of pollution, or a policy effect.

The RF and BRT scores use different test designs. A larger number in one column cannot establish that one family is better for future dates. This public record does not certify which models are currently used in the daily report.

Pages 2 and 3 give the complete current score tables. Page 4 records sample sizes and settings. Pages 5 and 6 retain the score history. Pages 7 and 8 explain the tests and the source record.

Random forest scores

R² is dimensionless. The change is against the 12 October assessment, not the previous Sunday.

Pollutant	OOB R²	OOB change	Forward CV R²	CV change
PM2.5	0.7901	+0.0000	0.3942	+0.0001
PM10	0.7411	-0.0004	0.3784	-0.0012
NO2	0.6178	-0.0010	0.2277	-0.0086
NOx	0.6342	+0.0005	0.3607	+0.0056
SO2	0.5769	-0.0022	0.0854	-0.0089
CO	0.5490	-0.0019	0.2787	-0.0009
O3	0.5834	-0.0009	0.3327	-0.0024

The more demanding time test

The RF code sorts observations by time into five blocks. It fits on earlier blocks and validates on the next block, giving up to four forward tests. The reported CV score averages the valid tests. It does not train on later blocks to predict earlier ones.

Out-of-bag predictions use trees whose training sample omitted that observation. Their test observations remain within the same overall time period. Nearby hours and shared conditions can make this easier than prediction into a later time block.

Every current forward CV score is lower than its OOB score. That gap is a reason to distinguish future-time performance from a fit within the historical record. It is not, by itself, evidence of a particular physical cause.

Boosted tree scores

The BRT test holds out a random 20% of rows. Its floor is the fixed 30 May RF anchor minus 0.0300 R².

Pollutant	Held-out R²	Change	Train R²	Floor	Screen
PM2.5	0.7431	-0.0004	0.7436	0.7582	Fail
PM10	0.6704	-0.0004	0.6730	0.7173	Fail
NO2	0.5918	+0.0002	0.5982	0.4918	Pass
NOx	0.6378	+0.0015	0.6413	0.5720	Pass
SO2	0.5382	-0.0019	0.5416	0.5650	Fail
CO	0.5527	-0.0017	0.5520	0.4209	Pass
O3	0.6263	+0.0035	0.6300	0.5533	Pass

NO2, NOx, CO and O3 pass this recorded screen. PM2.5, PM10 and SO2 fail it. These are statistical screening results, not records of whether a file was transferred or a model was adopted.

The floor compares a random held-out BRT score with a historical RF OOB anchor. These validation designs differ. Passing this screen is therefore not a matched comparison of the two model families, nor a guarantee of performance on later dates.

The training and held-out scores are close. That is useful to report, but it does not prove the absence of overfitting or show that the model has isolated weather from all other influences. The held-out rows were drawn from the same time range.

Samples and model settings

Counts are pollutant-specific rows with the required inputs. The panels are not a count of unique people or days.

Pollutant	Rows	Stations	BRT train	BRT test
PM2.5	1,889,465	66	1,511,572	377,893
PM10	1,871,050	60	1,496,840	374,210
NO2	1,632,698	62	1,306,159	326,539
NOx	1,427,568	59	1,142,055	285,513
SO2	1,241,527	52	993,222	248,305
CO	1,489,442	61	1,191,554	297,888
O3	1,487,492	60	1,189,994	297,498

Settings recorded for this assessment

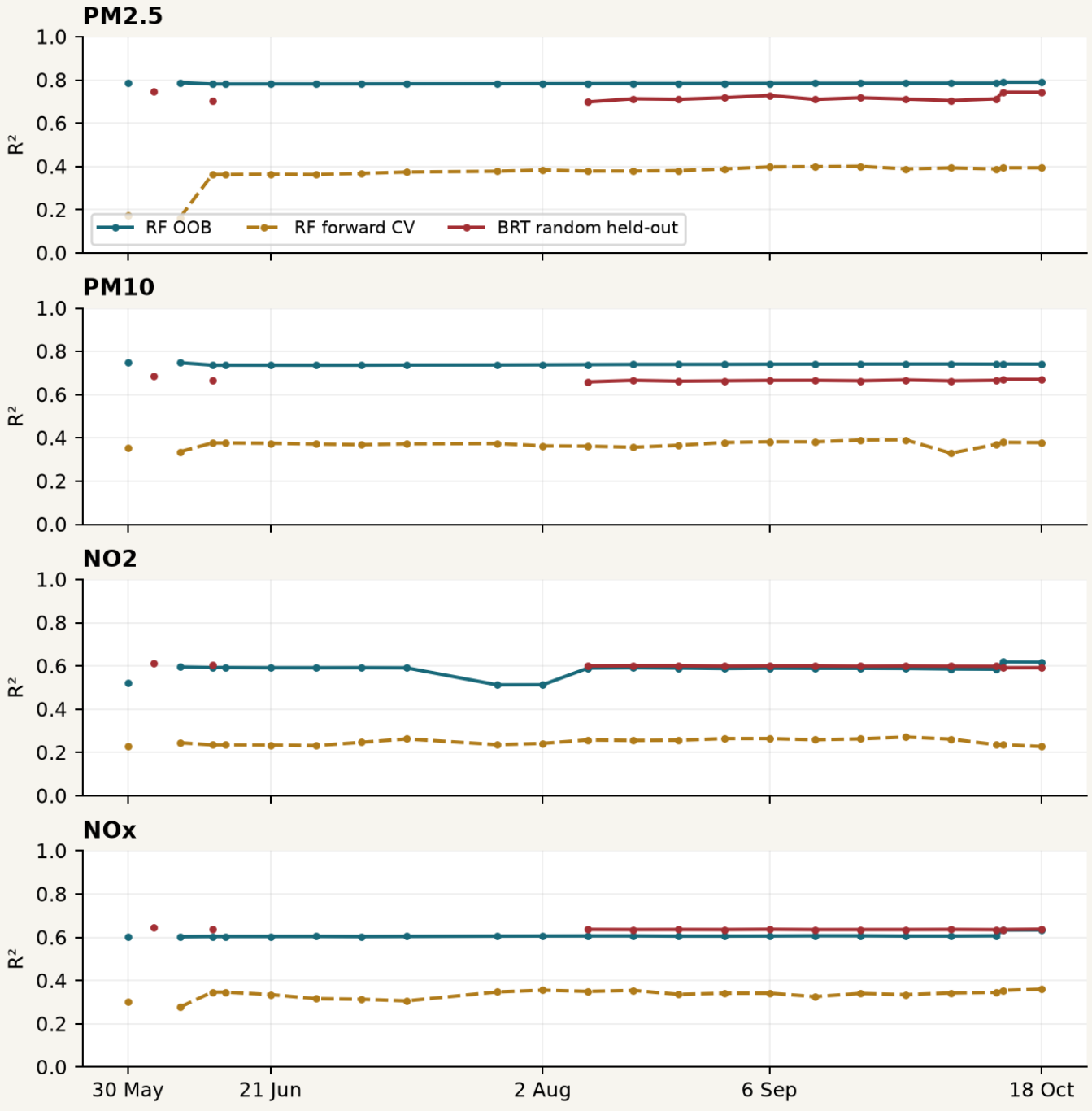
RF uses 800 trees for the final fit. Six tuning combinations test mtry of 3, 4 or 5 and minimum node size of 5 or 10, with 500 trees per tuning fit. Its validation uses five chronological blocks and up to four forward tests.

BRT uses 5,000 trees, interaction depth 6, shrinkage 0.01, bag fraction 0.5 and minimum node size 10. The split uses seed 42. Each test panel excludes stations absent from its training panel; the current reported train and test counts sum to the full panel.

Both model formulas include wind, temperature, boundary-layer height, rain and recent weather history, together with station, hour, weekday and a time trend. Including these variables does not make their separate causal effects identifiable.

The score history

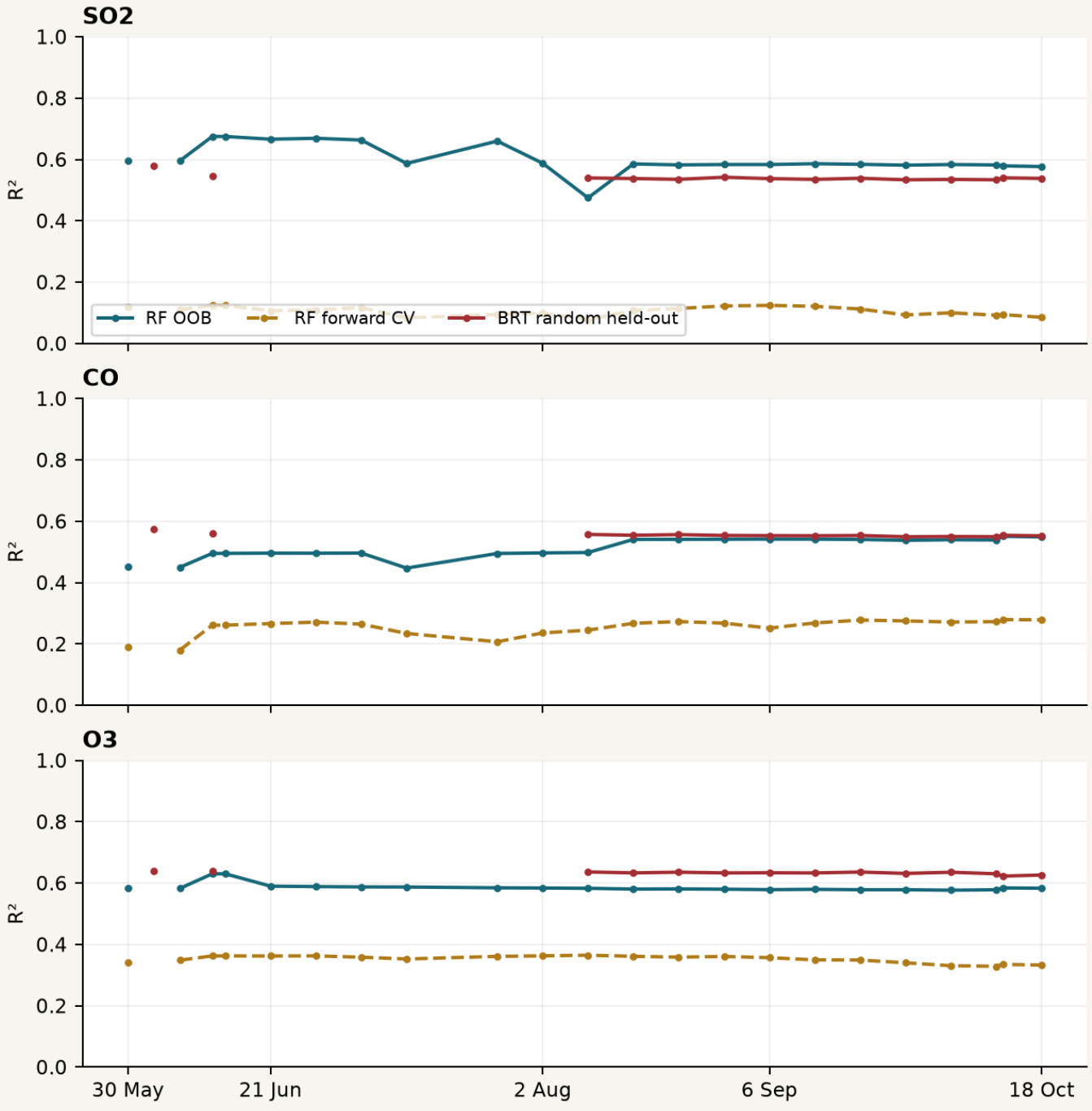
23 recorded assessment dates from 30 May to 18 October 2026. Some assessments were outside the weekly schedule.



Gaps mark unavailable scores, rather than zero. The three lines retain their different test designs. Changes in data, station coverage, settings or validation can also change scores, so these lines alone do not measure a physical change in Delhi's air.

The score history

23 recorded assessment dates from 30 May to 18 October 2026. Some assessments were outside the weekly schedule.



Gaps mark unavailable scores, rather than zero. The three lines retain their different test designs. Changes in data, station coverage, settings or validation can also change scores, so these lines alone do not measure a physical change in Delhi's air.

How to read the tests

These explanations describe the inspected training code. They are not a certification of daily model selection.

Two tree-based model families

A random forest combines many trees fitted to resampled observations. Boosted trees add trees sequentially to reduce the errors left by earlier trees. The recorded settings govern each fit; more trees do not guarantee convergence or better future predictions.

A score has a test population

For these recorded tests, R^2 is one minus squared prediction error divided by the variation in the test observations around their mean. A perfect score is 1. A negative score can be worse than predicting that test set's mean. It is not a percentage reduction in pollution.

Random rows are not future periods

The BRT split randomly separates rows from the same time range. Nearby observations can share weather and pollution conditions. The forward RF test instead predicts a later time block from earlier blocks. Their scores should be read alongside these differences.

Weather adjustment needs assumptions

A model can estimate concentrations under sampled weather conditions while holding other inputs fixed. Interpreting the result requires adequate data, model skill and a defensible sampling design. An adjusted estimate does not itself identify an emissions source or prove that a policy caused a change.

Sources and limits

The score tables and recorded history provide an auditable account of the assessment.

The assessment record

The dated RF and BRT score records were generated on 18 October 2026. The RF record gives a common training cutoff of 17 October at 16:00 IST for all seven pollutants. The history has 252 family-pollutant rows across 23 assessment dates. The immediate previous assessment is 12 October.

The CSV history and the current JSON score records were checked for agreement. The panel totals, threshold subtraction and all seven BRT verdicts were independently checked. The validation split and model settings were read in the training source. This publication did not rerun model fitting or independently rebuild the raw training panels.

What remains outside this record

This report does not certify current deployment, successful transfer, forecast accuracy under new conditions, or the scientific validity of every daily adjusted estimate. It does not replace the daily air-quality report. Comparisons of model families need a shared test design and matched data before supporting a selection decision.

Public downloads and method references

Model-performance archive and companion score downloads
Wright and Ziegler (2017), ranger, Journal of Statistical Software
Friedman (2001), Greedy Function Approximation, Annals of Statistics

The references explain model methods. The dated score records support the numbers here. Small changes are reported to four decimals, matching those score records. Rounding can conceal smaller differences.