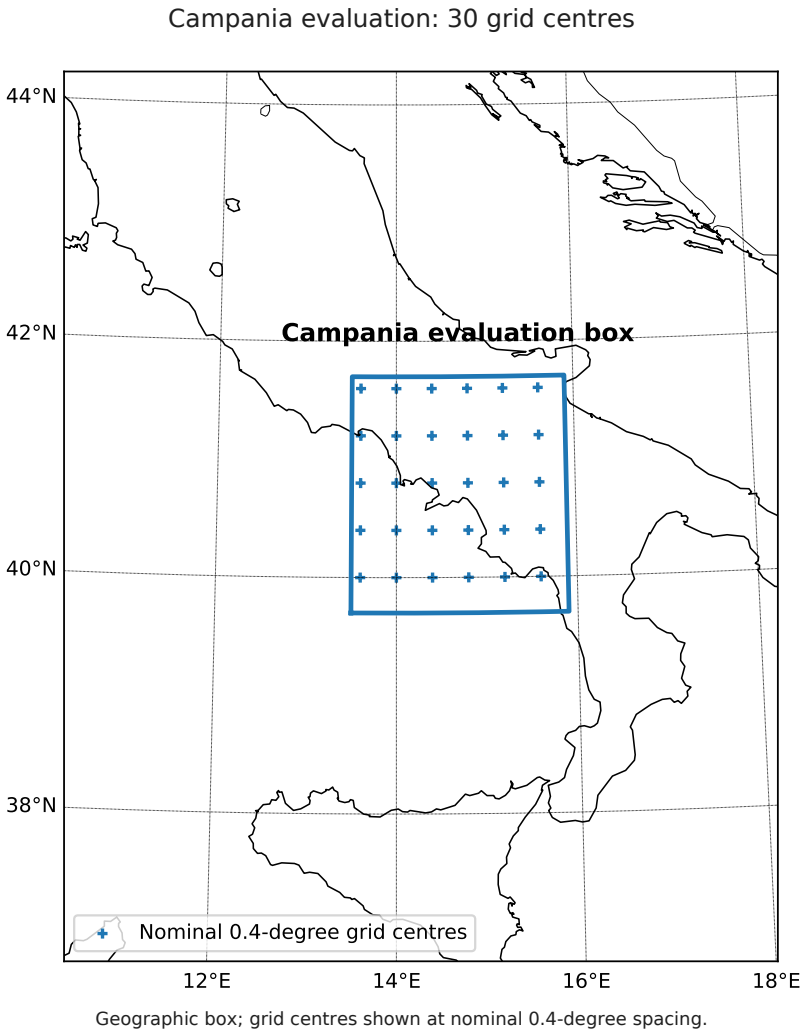
Supplementary Information

Independent evaluation of Aurora air pollution forecasts beyond benchmark skill

Angelo Riccio

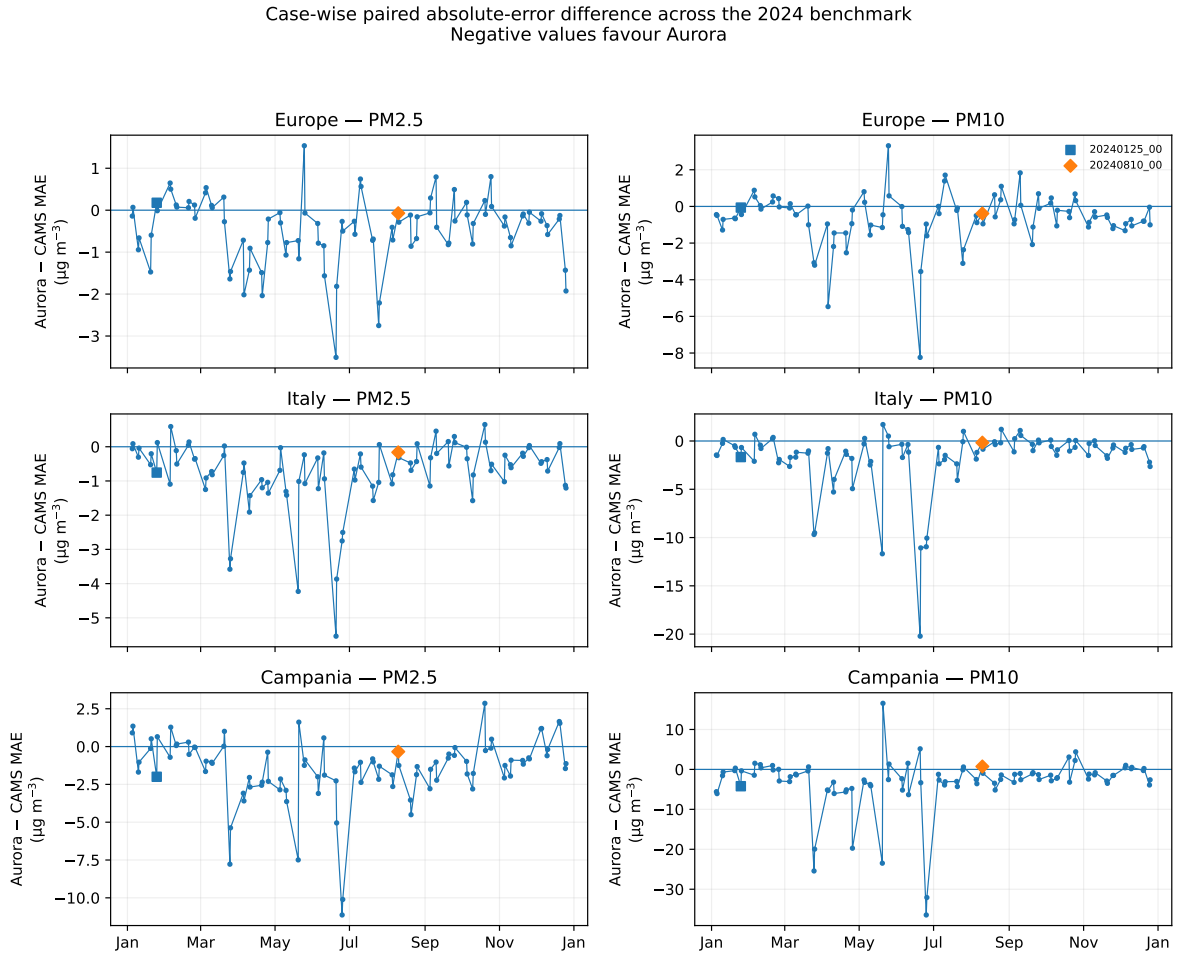
Supplementary spatial support

The Campania scoring box includes 30 grid centres; its outline is not an administrative mask. The three nested evaluation boxes are shown in Fig. 5 of the main manuscript.



Supplementary Fig. S1. Campania evaluation box and the 30 Aurora grid centres contributing to its regional scores.

Meteorological stress tests. The highlighted initializations are 25 January 2024 00 UTC (cold and stagnant) and 10 August 2024 00 UTC (warm and well ventilated). In Campania, the cold case lies at the 19.8th and 31.3rd percentiles of the 96-case ΔMAE distribution for PM_{10} and $PM_{2.5}$; the warm case lies at the 88.5th and 74.0th percentiles. Lower percentiles favour Aurora.



Supplementary Fig. S2. Case-wise mean paired absolute-error difference, $\Delta|e| = |A - O| - |C - O|$, for the 96 sampled 2024 initializations. Panels a–f show $PM_{2.5}$ and PM_{10} in Europe, Italy and Campania, respectively, from top to bottom. Negative values favour Aurora. The blue square marks the 25 January cold case; the orange diamond marks the 10 August warm case. Blue circles show the remaining initializations.

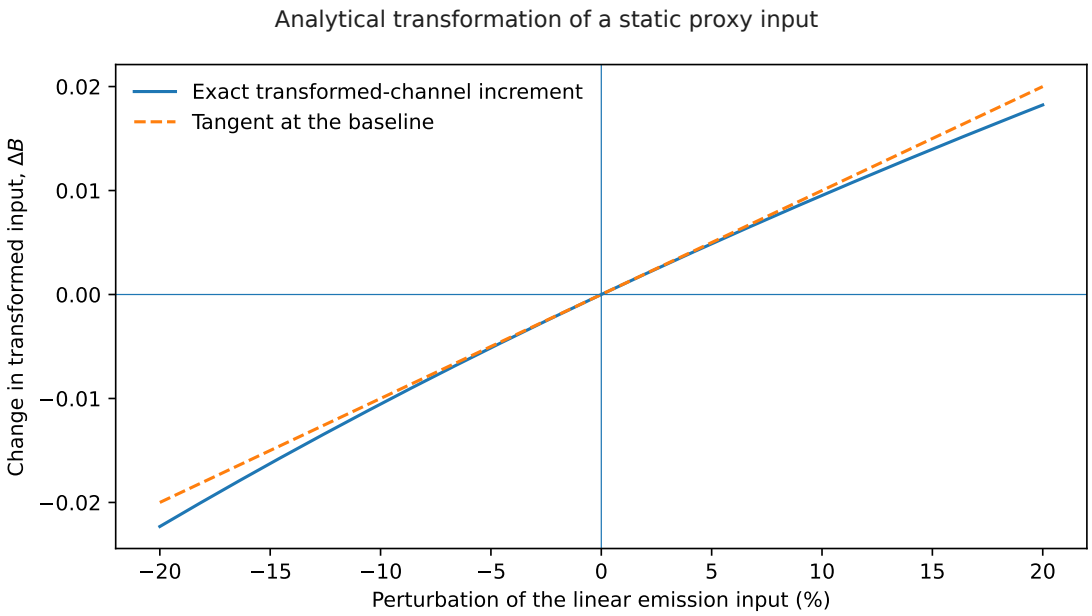
Supplementary Table S1. Selected meteorological stress tests. Bias and RMSE are in $\mu g\,m^{-3}$. Aurora (A) and CAMS (C) use identical station–time records.

Box	Episode	Pollutant	Bias A	Bias C	RMSE A	RMSE C
Europe	Cold/stagnant	PM_{10}	-5.77	-6.75	15.77	15.55
Europe	Cold/stagnant	$PM_{2.5}$	-4.49	-4.46	13.00	12.42
Europe	Warm/high-radiation	PM_{10}	1.44	-0.55	8.19	9.22
Europe	Warm/high-radiation	$PM_{2.5}$	1.21	0.95	5.35	5.96
Campania	Cold/stagnant	PM_{10}	-28.88	-32.69	58.73	58.49
Campania	Cold/stagnant	$PM_{2.5}$	-28.18	-29.82	57.58	56.54
Campania	Warm/high-radiation	PM_{10}	-0.66	-0.48	8.07	7.43
Campania	Warm/high-radiation	$PM_{2.5}$	-0.43	1.07	4.89	5.41

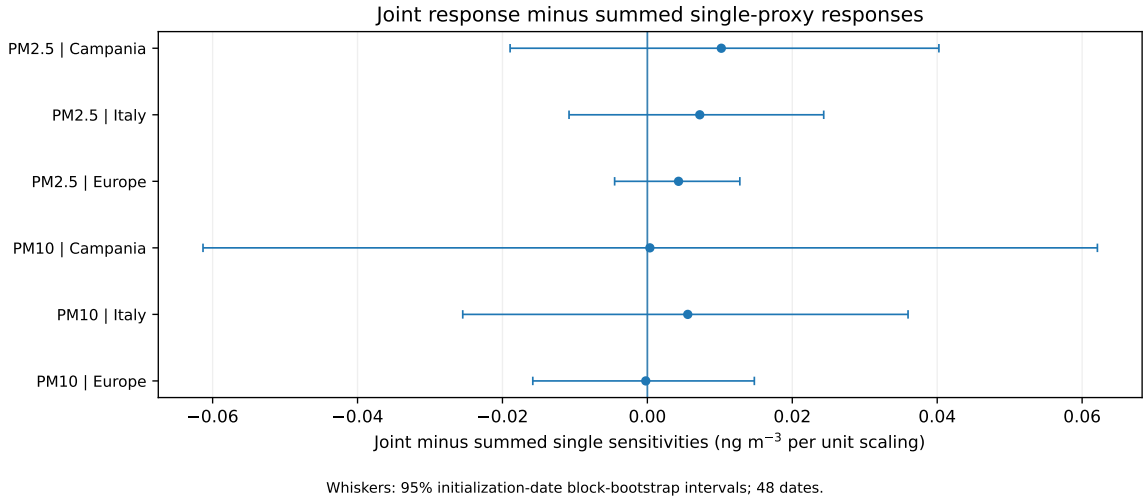
Supplementary Table S2. Seasonal ΔMAE (Aurora minus CAMS; $\mu g\,m^{-3}$) and fraction of the 96 initializations with lower Aurora mean absolute error. Seasonal values are descriptive point estimates; no seasonal confidence intervals are assigned.

Box	Pollutant	DJF	MAM	JJA	SON	Fraction cases A better
Europe	PM_{10}	-0.37	-0.94	-0.94	-0.38	0.740
Europe	$PM_{2.5}$	-0.29	-0.59	-0.79	-0.14	0.750
Italy	PM_{10}	-0.91	-2.68	-2.93	-0.49	0.823
Italy	$PM_{2.5}$	-0.32	-1.24	-1.15	-0.30	0.833
Campania	PM_{10}	-1.00	-5.25	-4.85	-1.27	0.781
Campania	$PM_{2.5}$	+0.05	-2.24	-2.67	-0.99	0.812

Supplementary proxy-response diagnostics



Supplementary Fig. S3. Analytical change in the transformed static input above its logarithmic floor (solid blue line) and its local tangent at the unperturbed input (dashed orange line).



Supplementary Fig. S4. Regional-mean difference between the joint D_{10} response and the sum of four single-input D_{10} responses. Points show mean residuals in ng m^{-3} per unit scaling; whiskers show 95% intervals from 20,000 initialization-date block-bootstrap resamples of 48 dates, retaining both cycles and all four leads.