Supplement to: Solar-tidal phase and amplitude in middle-atmosphere winds investigated from an adaptive rational neural operator, two reanalyses and meteor-radar observations

Yang Wu[1,2], Qixiang Liao[1,*], Zhihui Li[1], Zheng Sheng[1,*]

[1]College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China
[2]Key Laboratory of Digital Economy and Social Computing Science of Gansu, Lanzhou 730020, China

Correspondence to: Qixiang Liao (liaoqixiang2013@126.com) and Zheng Sheng (19994035@sina.com)

S1 Multi-horizon forecast skill and seasonal robustness

Separate direct models were trained for the 6, 12 and 24 h horizons using the purged, lead-aligned splits of Sect. 3.1 of the main text. Direct forecasts, a three-hour autoregressive rollout and ERA5 persistence were evaluated on the same 4304 test starts, and paired sample-level differences were assessed using 5000 seven-day block-bootstrap resamples.

Table S1 gives the level-mean results. The advantage over persistence is large at short lead and erodes steadily, from 46.3% at 3 h to 40.3% at 6 h, 26.9% at 12 h and 6.5% at 24 h. The direct models and the autoregressive rollout perform comparably at 6 and 12 h, at 3.597 against 3.427 m s^{-1} and 5.388 against 5.209 m s^{-1}. The ordering separates decisively at 24 h, where the rollout has accumulated error to 8.169 m s^{-1} and falls behind both the direct model at 7.560 m s^{-1} and persistence at 8.083 m s^{-1}.

Table S1. Level-mean wind-speed RMSE by forecast horizon, in m s^{-1}.

Lead (h)	ATMNO direct	Persistence	3 h rollout	Improvement over persistence (%)
3	2.215	4.121	2.215	46.30
6	3.597	6.022	3.427	40.27
12	5.388	7.375	5.209	26.94
24	7.560	8.083	8.169	6.47

The altitude dependence in Table S2 shows that the narrowing at long lead is not uniform. At 20 and 30 km the 24 h direct model is barely better than persistence, at 3.495 against 3.719 m s^{-1} and 4.826 against 4.848 m s^{-1}. At 80 km a substantial margin survives, at 7.501 against 8.841 m s^{-1}. The levels where the forecast retains most value at 24 h are therefore the upper ones, which is the opposite of what the absolute error magnitudes alone would suggest. Persistence benefits most where the diurnal cycle is most nearly repeatable, and the ERA5 diurnal signal is better constrained in the stratosphere than in the mesosphere.

Table S2. Per-altitude direct-model (D) and persistence (P) wind-speed RMSE, in m s^{-1}.

Altitude (km)	D 3 h	P 3 h	D 6 h	P 6 h	D 12 h	P 12 h	D 24 h	P 24 h
20	1.231	1.874	1.850	2.718	2.586	3.266	3.495	3.719
30	1.553	2.445	2.398	3.434	3.486	4.051	4.826	4.848
40	2.283	3.833	3.680	5.452	5.289	6.689	7.064	7.434
50	2.979	5.155	4.826	7.691	7.061	9.792	9.379	10.427
60	2.912	5.380	4.864	8.217	7.511	10.496	10.367	10.744
70	2.558	5.225	4.304	7.759	6.688	9.598	10.291	10.571
80	1.990	4.937	3.254	6.882	5.095	7.729	7.501	8.841

The behaviour of persistence itself explains most of the 24 h result. Sampling persistence error every 3 h out to 36 h, as Table S3 shows, reveals that it is not monotonic. Vector RMSE rises to a maximum of 12.058 m s^{-1} at 18 h, falls to a local minimum of 11.508 m s^{-1} at 24 h, then resumes rising. Speed RMSE shows the same pattern more weakly, with 8.044 m s^{-1} at 21 h and 8.083 m s^{-1} at 24 h bracketing a shallow trough. A 24 h lead corresponds to one full diurnal and two full semidiurnal cycles, so the tidal component of the wind returns approximately to its initial phase and persistence recovers part of the signal it lost at intermediate leads. The 6.5% margin at 24 h is therefore measured against an unusually favourable configuration for persistence, and the direct model still improves on it after purging. Establishing where the genuine predictability ceiling lies would require leads chosen to avoid tidal commensurability, such as 9, 15 or 21 h.

Table S3. Persistence error as a function of forecast lead time, in m s^{-1}.

Lead (h)	Speed RMSE	Vector RMSE
3	4.125	6.228
6	6.027	9.250
9	6.895	10.625
12	7.381	11.313
15	7.839	11.894
18	8.095	12.058
21	8.044	11.661
24	8.083	11.508
27	8.757	12.618
30	9.434	13.823
33	9.821	14.469
36	10.060	14.767

Seasonal behaviour is dominated by the winter stratosphere, as Table S4 and Fig. S1c and d show. DJF speed RMSE reaches 4.060 m s^{-1} at 50 km and 4.020 m s^{-1} at 60 km, against 2.070 to 2.570 m s^{-1} at the same levels in the other three seasons. The level-mean DJF error of 2.820 m s^{-1} is roughly 50% above the MAM minimum of 1.860 m s^{-1}. The excess concentrates at 50–60 km, where the winter polar vortex edge and the associated planetary-wave activity produce the largest wind variance in this sector. Direction MAE does not follow the same pattern. It is largest in MAM at 5.810° and smallest in JJA at 4.340°, and the DJF value of 5.560° is unremarkable. Speed error tracks variance amplitude while direction error tracks the frequency of weak-wind and transitional conditions.

Table S4. Seasonal ATMNO skill. Direction MAE uses grid points with observed wind speed above 3 m s^{-1}.

Season	Metric	20 km	30 km	40 km	50 km	60 km	70 km	80 km	Mean
DJF	Speed RMSE	1.390	1.780	2.750	4.060	4.020	3.200	2.530	2.820
MAM	Speed RMSE	1.190	1.400	2.100	2.140	2.070	2.260	1.850	1.860
JJA	Speed RMSE	1.160	1.460	2.110	2.570	2.370	2.440	1.760	1.980
SON	Speed RMSE	1.200	1.540	2.150	2.790	2.790	2.300	1.820	2.080
DJF	Direction MAE (°)	4.560	5.210	6.700	6.880	4.410	5.380	5.750	5.560
MAM	Direction MAE (°)	6.830	5.610	8.280	6.810	4.630	4.380	4.120	5.810
JJA	Direction MAE (°)	5.910	3.600	3.790	4.340	4.230	3.530	5.010	4.340
SON	Direction MAE (°)	5.220	6.010	7.600	5.460	3.960	4.480	4.720	5.350

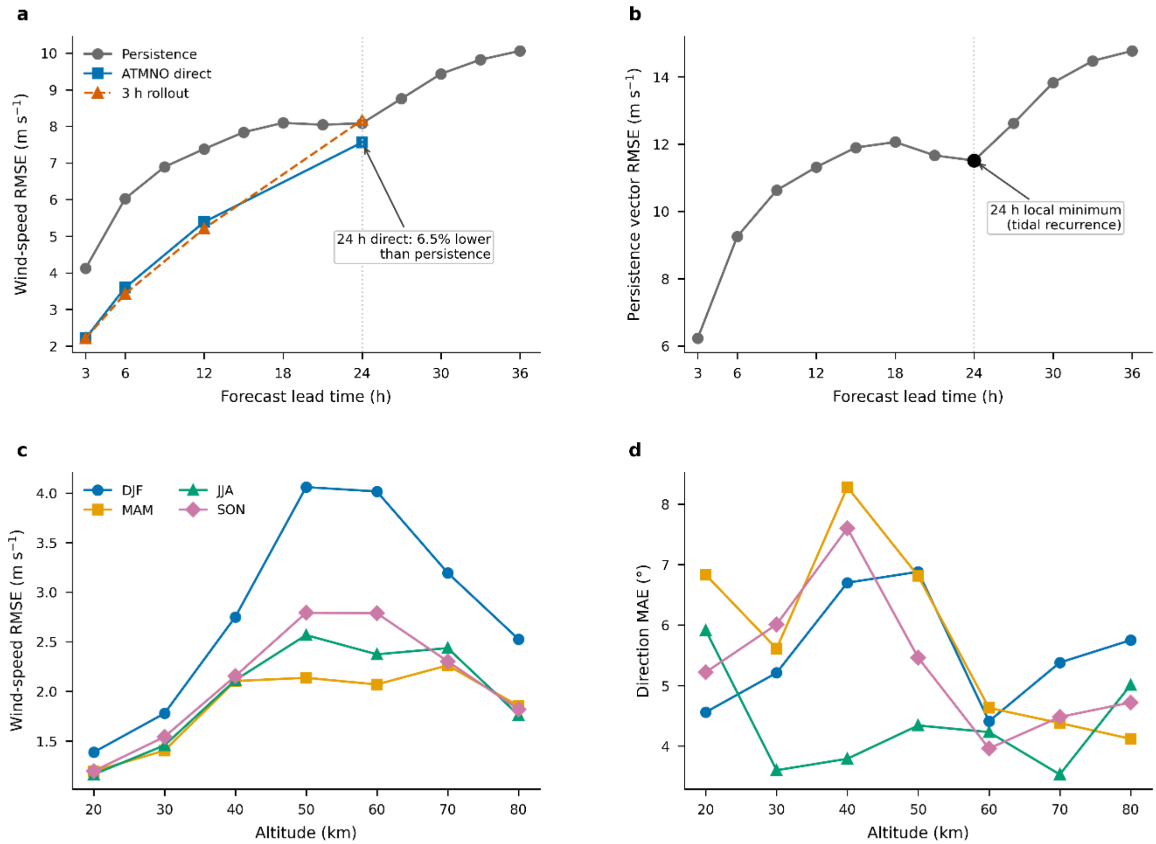


Figure S1. Forecast skill across lead time and season. (a) Test-set wind-speed RMSE against lead for persistence, the direct ATMNO models and the three-hour autoregressive rollout. (b) Persistence speed and vector RMSE sampled every 3 h, showing the local minimum at 24 h. (c, d) Seasonal wind-speed RMSE and direction MAE by altitude.

S2 Full harmonic-fidelity comparison

This section reports the altitude-resolved and month-resolved comparison summarised in Sect. 4.6 of the main text, together with the harmonic fidelity of the reduced configurations of Sect. 4.2. Complete calendar months serve as the paired unit, which sidesteps the autocorrelation problem affecting sample-level comparison, since month-to-month dependence in these metrics is weak. Seventeen months from June 2024 to October 2025 were available.

Table S5 gives the level-mean metrics with 95% month-bootstrap intervals. Reading down the phase column, ATMNO is lowest for all three harmonics against both comparators, but the separation relative to the interval widths differs greatly by harmonic. For the diurnal component the three models span 4.33 to 4.56°, against individual intervals roughly 0.9° wide, so the ordering is consistent but the models are close. For the semidiurnal component the span is 10.12 to 10.79° against intervals of about 1.5°. For the terdiurnal component the span is 37.43 to 43.60° against intervals of about 5°, and the separation is unambiguous. The same ordering does not hold for the other three metrics, where Factformer has marginally lower complex NRMSE and amplitude NMAE and marginally higher pattern coherence for the diurnal and semidiurnal components.

Table S5. Level-mean harmonic-fidelity metrics over 17 complete test months, with 95% month-bootstrap intervals.

Harmonic	Model	Complex NRMSE	Phase MAE (°)	Pattern coherence	Amplitude NMAE
S1	ATMNO	0.1461, 0.1336 to 0.1599	4.33, 3.88 to 4.82	0.9768, 0.9731 to 0.9802	0.0839, 0.0778 to 0.0908
S1	Factformer	0.1426, 0.1301 to 0.1560	4.50, 4.03 to 5.02	0.9786, 0.9753 to 0.9817	0.0800, 0.0735 to 0.0872
S1	MWT	0.1510, 0.1388 to 0.1644	4.56, 4.12 to 5.02	0.9757, 0.9719 to 0.9792	0.0881, 0.0815 to 0.0955
S2	ATMNO	0.3131, 0.2944 to 0.3313	10.12, 9.38 to 10.91	0.9042, 0.8952 to 0.9133	0.1753, 0.1650 to 0.1857
S2	Factformer	0.3106, 0.2928 to 0.3295	10.79, 10.00 to 11.65	0.9091, 0.9011 to 0.9176	0.1690, 0.1596 to 0.1785
S2	MWT	0.3147, 0.2986 to 0.3311	10.61, 9.92 to 11.34	0.9045, 0.8969 to 0.9126	0.1766, 0.1679 to 0.1855
S3	ATMNO	0.7996, 0.7615 to 0.8403	37.43, 35.03 to 40.00	0.5189, 0.4886 to 0.5473	0.3847, 0.3714 to 0.3981
S3	Factformer	0.8593, 0.8155 to 0.9042	43.60, 40.66 to 46.68	0.4980, 0.4726 to 0.5223	0.3915, 0.3787 to 0.4049
S3	MWT	0.8298, 0.7907 to 0.8699	39.70, 37.28 to 42.25	0.4991, 0.4698 to 0.5285	0.3949, 0.3813 to 0.4085

The paired comparison in Table S6 sharpens the pattern. Of the 24 paired tests, the six phase comparisons all favour ATMNO with Holm-adjusted $P \leq 0.007$ and win fractions of 0.824 to 1.000. Every one of the six diurnal and semidiurnal comparisons on complex NRMSE, amplitude NMAE and pattern coherence against Factformer favours Factformer, with win fractions of 0 to 0.294. Against MWT the picture is more favourable to ATMNO: all four diurnal metrics favour ATMNO significantly, the semidiurnal complex NRMSE, amplitude NMAE and pattern coherence are statistically indistinguishable at $P = 0.808$, 0.808 and 0.927, and all four terdiurnal metrics favour ATMNO. The consistency of the phase result across two comparators and three harmonics, with win fractions at or near unity for five of the six, is the strongest single indication that the adaptive basis affects the represented timing specifically. A metric improving by chance would not do so in the same direction for every harmonic against both comparators while the neighbouring metrics moved the other way. Multiplicity was controlled within each model-by-harmonic group of four metrics rather than across all 24 tests.

Table S6. Paired monthly comparisons for all three solar-tidal harmonics. Positive improvement favours ATMNO. *n* = 17 months.

Comparator, harmonic	Metric	ATMNO improvement, 95% CI	Win fraction	Holm-adj. *P*
Factformer, S1	Complex NRMSE	−0.0035, −0.0051 to −0.0021	0.118	1
Factformer, S1	Amplitude NMAE	−0.0039, −0.0050 to −0.0028	0	1
Factformer, S1	Phase MAE (°)	0.173, 0.093 to 0.256	0.824	0.007
Factformer, S1	Pattern coherence	−0.0018, −0.0023 to −0.0014	0	1
Factformer, S2	Complex NRMSE	−0.0025, −0.0056 to 0.0002	0.294	1
Factformer, S2	Amplitude NMAE	−0.0063, −0.0088 to −0.0038	0.176	1
Factformer, S2	Phase MAE (°)	0.672, 0.527 to 0.824	1	< 0.001

Factformer, S2	Pattern coherence	−0.0049, −0.0072 to −0.0027	0.118	1
Factformer, S3	Complex NRMSE	0.0597, 0.0517 to 0.0677	1	< 0.001
Factformer, S3	Amplitude NMAE	0.0068, 0.0024 to 0.0108	0.765	0.055
Factformer, S3	Phase MAE (°)	6.168, 5.527 to 6.820	1	< 0.001
Factformer, S3	Pattern coherence	0.0209, 0.0130 to 0.0286	0.824	0.002
MWT, S1	Complex NRMSE	0.0049, 0.0037 to 0.0061	1	< 0.001
MWT, S1	Amplitude NMAE	0.0041, 0.0030 to 0.0052	0.941	< 0.001
MWT, S1	Phase MAE (°)	0.228, 0.176 to 0.281	1	< 0.001
MWT, S1	Pattern coherence	0.0011, 0.0007 to 0.0016	0.941	< 0.001
MWT, S2	Complex NRMSE	0.0017, −0.0025 to 0.0057	0.588	0.808
MWT, S2	Amplitude NMAE	0.0013, −0.0021 to 0.0045	0.706	0.808
MWT, S2	Phase MAE (°)	0.489, 0.351 to 0.638	0.941	< 0.001
MWT, S2	Pattern coherence	−0.0003, −0.0029 to 0.0022	0.529	0.927
MWT, S3	Complex NRMSE	0.0302, 0.0209 to 0.0388	0.941	< 0.001
MWT, S3	Amplitude NMAE	0.0102, 0.0070 to 0.0137	0.941	< 0.001
MWT, S3	Phase MAE (°)	2.264, 1.666 to 2.868	0.941	< 0.001
MWT, S3	Pattern coherence	0.0198, 0.0137 to 0.0256	0.941	< 0.001

The altitude dependence in Figs. S2 to S4 shows that the terdiurnal advantage is not uniform with height. Complex NRMSE for all three models decreases monotonically from about 1.25 to 1.35 at 20 km to about 0.35 at 80 km, and the separation between ATMNO and Factformer is widest between 20 and 50 km, closing to near zero above 70 km. Phase MAE behaves similarly, with the largest separation at 20–40 km. Absolute fidelity runs the other way. Complex NRMSE exceeds unity below 40 km, so the terdiurnal error is comparable to the terdiurnal signal itself, and falls below 0.5 only above 60 km. The relative advantage is concentrated where absolute skill is worst. Month-to-month consistency in Fig. S5 shows a pronounced annual cycle in all three models, with complex NRMSE peaking in boreal autumn and winter and reaching a minimum in spring. ATMNO tracks below Factformer in 16 of the 17 months for the terdiurnal component, so the ordering is stable rather than driven by a few outlying months.

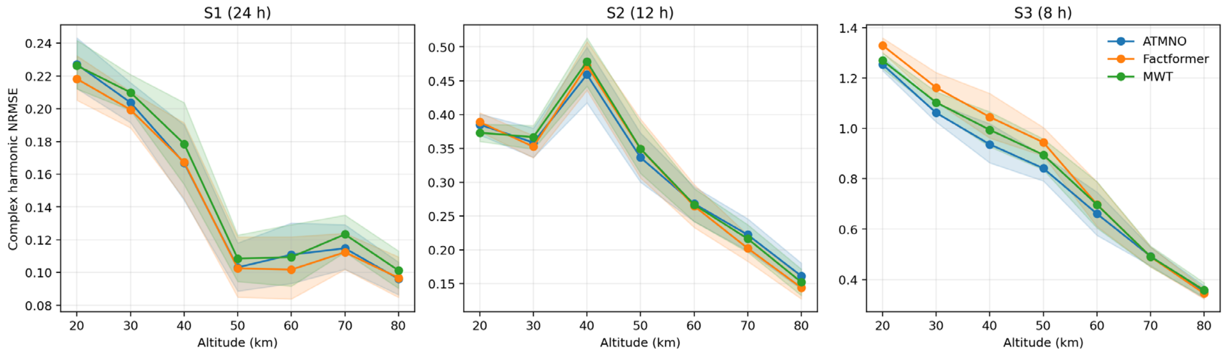


Figure S2. Complex harmonic NRMSE by altitude for the diurnal, semidiurnal and terdiurnal components. Lines show the 17-month mean and shading shows month-bootstrap intervals.

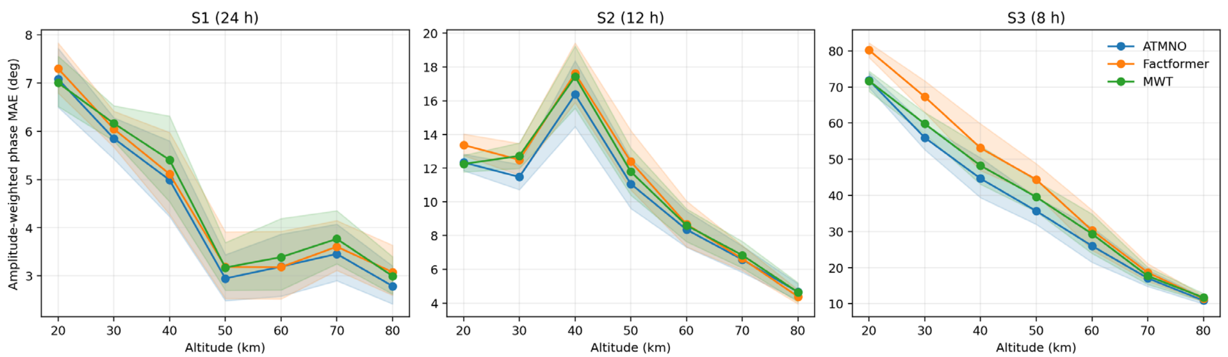


Figure S3. Amplitude-weighted harmonic phase MAE by altitude.

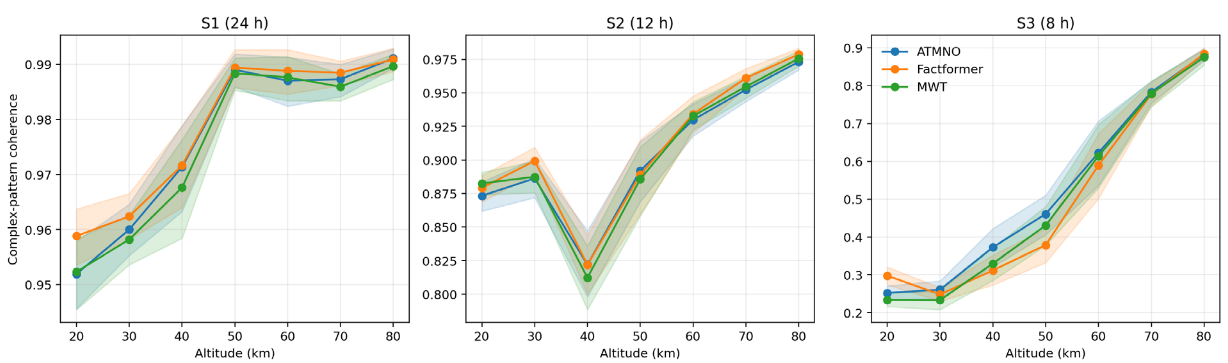


Figure S4. Complex-pattern coherence by altitude.

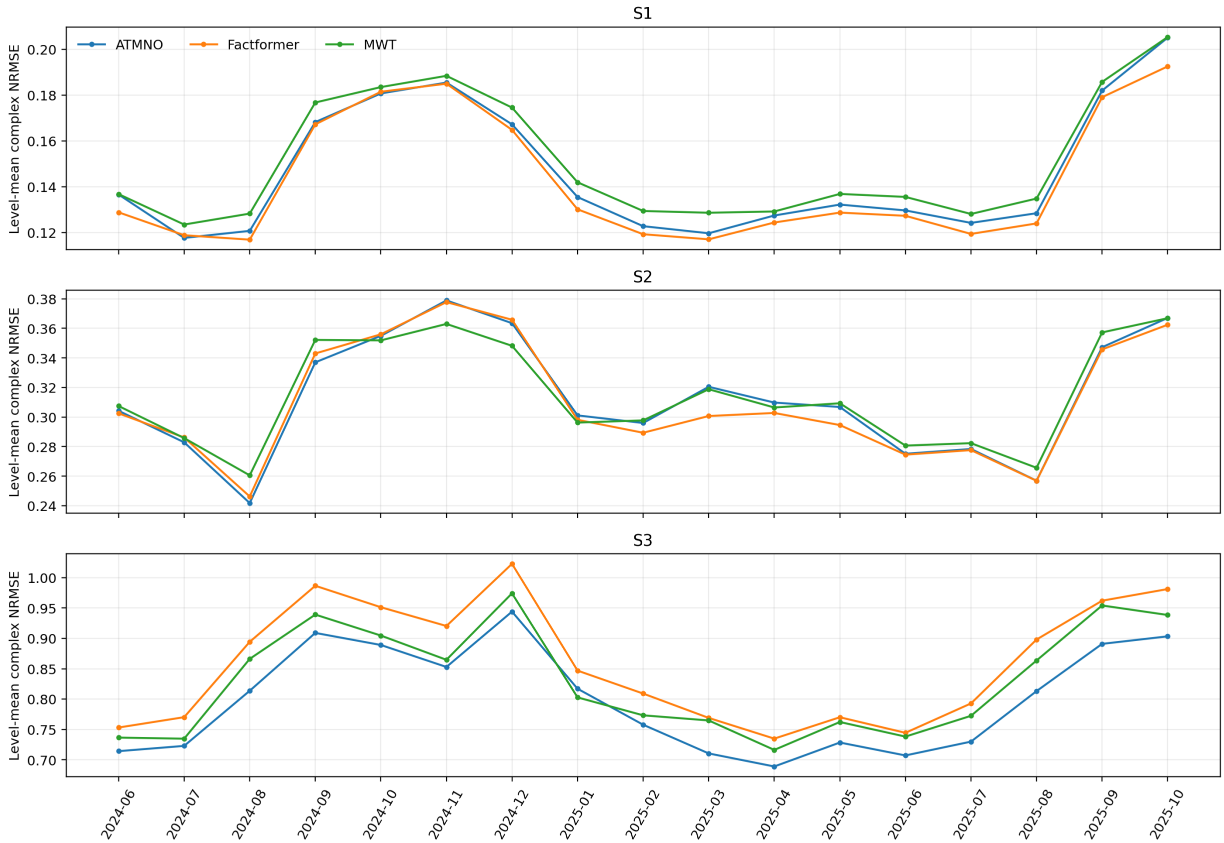


Figure S5. Monthly level-mean complex NRMSE. Each point is one complete calendar month.

Table S7 applies the same decomposition to the four reduced configurations. All 24 paired monthly comparisons against the full model favour the full model, with Holm-adjusted $P \leq 0.001$. The pattern across configurations tracks the aggregate ordering of Sect. 4.2 of the main text, with one difference that bears on the main argument. The Fourier-basis variant, which is worst on aggregate RMSE, raises the level-mean terdiurnal phase error by 2.12° relative to the full model. That figure sits next to the terdiurnal phase advantage of ATMNO over MWT, which is 2.264°. Replacing the adaptive rational basis with a uniform Fourier basis therefore costs approximately as much terdiurnal phase fidelity as the full model gains over its strongest comparator on that metric, which is what would be expected if the adaptive placement of spectral resolution is the mechanism behind the phase result. The diurnal and semidiurnal phase degradations are smaller in absolute terms, at 0.25 to 0.37° and 0.38 to 0.98°, but proportionally comparable at 6 to 9% of the full-model values. Complex NRMSE degrades by 5 to 7% for the diurnal, 4 to 9% for the semidiurnal and 3 to 5% for the terdiurnal component.

Table S7. Level-mean harmonic fidelity of reduced temporal configurations over 17 complete test months.

Configuration	Speed RMSE (m s^{-1})	Phase MAE (°) S1	Phase MAE (°) S2	Phase MAE (°) S3	Complex NRMSE S1	Complex NRMSE S2	Complex NRMSE S3

ATMNO, full	2.215	4.33	10.12	37.43	0.1461	0.3131	0.7996
Mamba only	2.295	4.70	11.10	39.43	0.1554	0.3360	0.8265
Fixed poles	2.306	4.59	10.50	38.58	0.1537	0.3277	0.8255
TM only	2.312	4.58	10.63	39.85	0.1540	0.3249	0.8331
Fourier basis	2.325	4.70	10.94	39.55	0.1567	0.3398	0.8366

S3 Station-level meteor-radar forecast validation

The collocation protocol was fixed before the comparison was run. Three-hour ATMNO forecasts were collocated with quality-controlled hourly meteor-radar winds at the nominal 80 km level at Beijing-Shisanling, Mohe and Wuhan. Radar zonal wind is positive eastward and meridional wind positive northward as supplied, and no data-dependent sign reversal and no temporal-shift optimisation was applied at any point. The primary comparison uses the vector mean of observations at $t-1$, t and $t+1$ h, labelled at the model valid time t. ERA5 persistence uses the collocated analysis at $t-3$ h. The valid-time ERA5 analysis is retained as a reference rather than as a forecast baseline, since it has access to information the forecast does not, and its role is to show how much of the achievable agreement with the radars the forecast retains.

Table S8 gives the scalar verification. Three features stand out. The biases are large and negative for wind speed at every station, at −11.33, −10.40 and −10.39 m s^{-1} at Beijing, Mohe and Wuhan, and they are essentially identical for ATMNO, persistence and the valid-time analysis. A bias shared by all three is a property of ERA5 rather than of the forecast. The component biases are much smaller than the speed bias, ranging from −6.91 to +1.50 m s^{-1}, which is the signature of a systematic underestimate of wind magnitude rather than a directional offset. ATMNO improves on persistence in R^2 for every variable at every station, and the improvement is largest for the meridional component: at Mohe R^2 rises from 0.095 to 0.335, and at Wuhan from 0.091 to 0.285. Persistence has negative R^2 for wind speed at Mohe, at −0.028, so it performs worse there than the observed mean.

Table S8. Direct scalar validation against meteor-radar winds at nominal 80 km.

Station	Source	Variable	Bias	RMSE	MAE	R^2
Beijing (N=782)	ATMNO	U	-6.91	23.79	19.17	0.493
		V	-0.08	20.54	16.23	0.432
		Speed	-11.33	21.78	17.23	0.349
	Persistence	U	-6.82	24.97	20.06	0.434
		V	-0.05	22.67	17.85	0.215
		Speed	-11.39	22.44	17.62	0.302
	Valid-time ERA5	U	-6.73	23.62	19.01	0.497
		V	-0.07	20.57	16.26	0.429

		Speed	-11.35	21.82	17.32	0.353
Mohe (N=1335)	ATMNO	U	-2.42	21.02	17.03	0.564
		V	-1.94	18.94	15.55	0.335
		Speed	-10.4	19.62	15.67	0.138
	Persistence	U	-2.42	22.74	18.48	0.477
		V	-1.86	21.19	17.25	0.095
		Speed	-10.41	20.78	16.37	-0.028
	Valid-time ERA5	U	-2.55	21.09	17.07	0.561
		V	-1.92	19.04	15.62	0.326
		Speed	-10.49	19.77	15.77	0.125
Wuhan (N=1327)	ATMNO	U	-6.22	27.89	22.81	0.473
		V	1.5	21.64	17.44	0.285
		Speed	-10.39	23.24	18.77	0.349
	Persistence	U	-6.09	28.65	23.43	0.44
		V	1.57	23.31	18.68	0.091
		Speed	-10.42	23.5	18.93	0.334
	Valid-time ERA5	U	-6.06	27.99	22.89	0.467
		V	1.51	21.7	17.5	0.278
		Speed	-10.36	23.34	18.86	0.344

The vector and directional verification in Table S9 places ATMNO between persistence and the valid-time analysis at every station, and much closer to the latter. At Beijing the vector RMSE is 31.42 m s^{-1} for ATMNO against 31.32 for the valid-time analysis and 33.72 for persistence. At Mohe, ATMNO at 28.30 is marginally better than the valid-time analysis at 28.42. Direction MAE follows the same ordering. That a three-hour forecast should match the valid-time analysis this closely follows from the amplitude deficit. When the analysis itself agrees with the radars only moderately, a forecast reproducing the analysis faithfully inherits almost all of that agreement, and the marginal loss from the three-hour lead is small by comparison. The strictly causal harmonic baseline is included for scale, with within-20° hit rates of 5.2 to 9.6% against 24.5 to 26.8% for ATMNO.

Table S9. Vector-wind and direction validation against meteor-radar winds.

Station	Source	Vector RMSE	Vector MAE	Direction MAE (deg)	Within 20 deg(%)
Beijing	ATMNO	31.42	27.66	56.72	26.5
	Persistence	33.72	29.68	65.6	20.5
	Valid-time ERA5	31.32	27.56	56.04	27

	Harmonic baseline	68.70	61.22	110.17	5.2
Mohe	ATMNO	28.3	25.43	54.76	26.8
	Persistence	31.08	27.96	62.88	20.8
	Valid-time ERA5	28.42	25.51	54.88	27.5
	Harmonic baseline	98.58	91.22	105.91	9.4
Wuhan	ATMNO	35.3	31.84	63.46	24.5
	Persistence	36.93	33.34	68.25	23.2
	Valid-time ERA5	35.42	31.92	63.55	24.9
	Harmonic baseline	55.39	48.63	102.55	9.6

Radar and model samples are serially correlated, so differences were assessed with a seven-day paired calendar-block bootstrap. Table S10 shows that all nine intervals, covering three variables at three stations, exclude zero and favour ATMNO. The largest differences are in the meridional component, from −1.666 to −2.250 m s^{-1}, and the smallest in wind speed at Wuhan, at −0.257 m s^{-1} with an interval of −0.406 to −0.109. Block counts are 17 at Beijing and 25 at Mohe and Wuhan.

Table S10. Seven-day paired block-bootstrap RMSE differences between ATMNO and persistence, in m s^{-1}. Negative values favour ATMNO.

Station	Variable	ATMNO−persistence RMSE	95% CI lower	95% CI upper	Blocks
Beijing	U	-1.181	-1.696	-0.718	17
	V	-2.133	-2.888	-1.367	17
	Speed	-0.662	-1.026	-0.338	17
Mohe	U	-1.714	-2.272	-1.151	25
	V	-2.25	-2.813	-1.709	25
	Speed	-1.165	-1.592	-0.758	25
Wuhan	U	-0.757	-1.01	-0.538	25
	V	-1.666	-2.026	-1.296	25
	Speed	-0.257	-0.406	-0.109	25

Radar and model are sampled on different clocks, so the comparison was repeated under two alternative zero-shift alignments: a left-aligned three-hour window covering $t-2$, $t-1$ and t, and an exact-hour match with no averaging. Table S11 shows that ATMNO is ahead of persistence for every station, variable and alignment, so the ranking does not depend on window placement. Absolute errors are lowest under the pre-declared centred product, which remains the primary analysis, and the exact-hour match is worst for both models, as expected from the reduced averaging.

Table S11. Zero-shift temporal-alignment sensitivity, RMSE in m s^{-1}. The centred three-hour product is the pre-declared primary analysis.

Station	Variable	Centred 3h: ATMNO	Centred 3h: Persist.	Left 3h: ATMNO	Left 3h: Persist.	Exact hour: ATMNO	Exact hour: Persist.
Beijing	U	23.79	24.97	24.28	25.59	25.94	27.11
Beijing	V	20.54	22.67	20.80	23.43	23.10	25.23
Beijing	Speed	21.78	22.44	22.14	22.83	24.42	24.90
Mohe	U	21.02	22.74	21.58	23.50	23.42	25.15
Mohe	V	18.94	21.19	19.39	21.90	21.04	23.14
Mohe	Speed	19.62	20.78	20.11	21.18	22.16	23.29
Wuhan	U	27.89	28.65	27.88	28.76	30.07	30.91
Wuhan	V	21.64	23.31	21.91	23.84	24.03	25.58
Wuhan	Speed	23.24	23.50	22.93	23.17	25.49	25.77

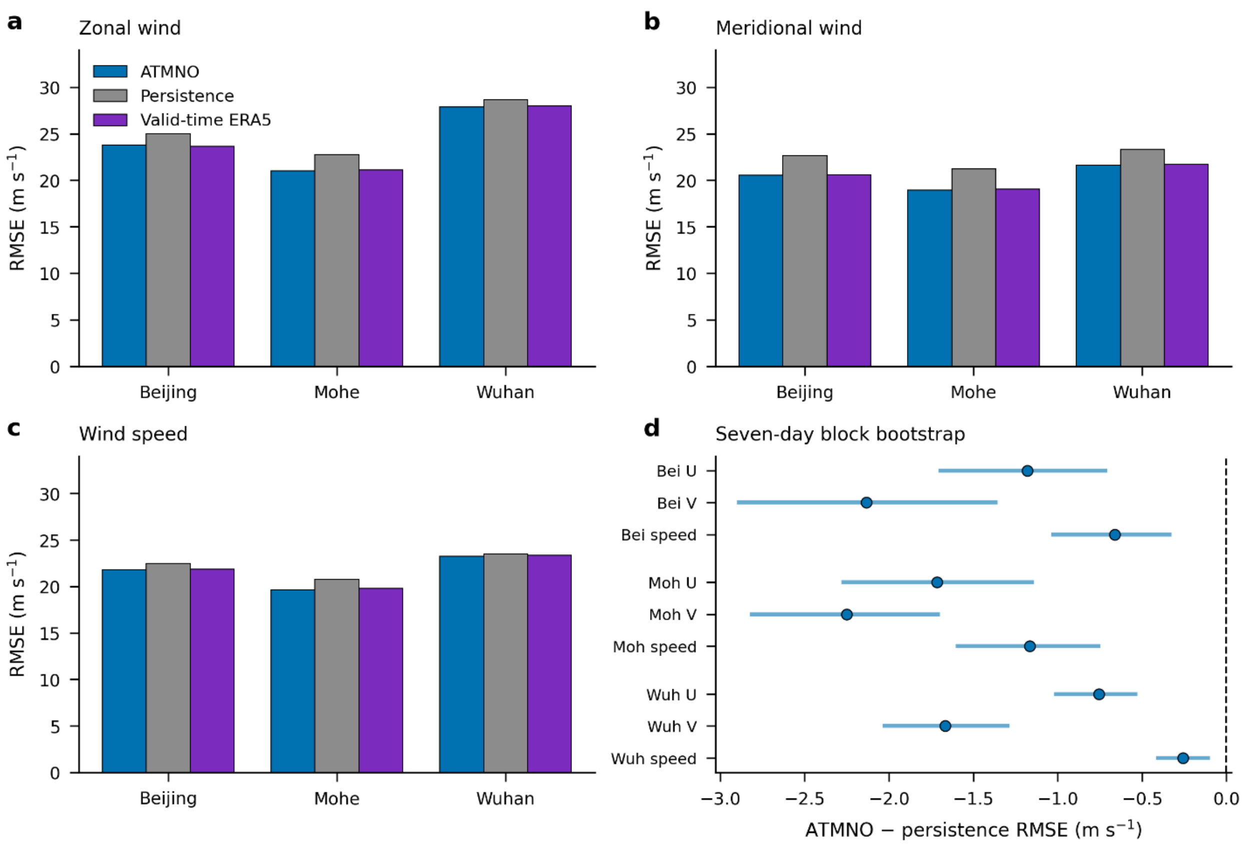


Figure S6. Validation against meteor-radar winds at the nominal 80 km level. (a) to (c) Component and wind-speed RMSE at the three stations for ATMNO, ERA5 persistence and the valid-time ERA5 analysis. (d) Paired seven-day block-bootstrap differences between ATMNO and persistence, with 95% intervals.

S4 Functional intervention experiments

The pole analysis of Sect. 4.3 of the main text establishes that certain rational components sit at tidal frequencies. Whether those components carry the tidal information the model uses is a separate question, and the experiments reported here separate the two.

The design is an inference-time intervention. Rather than suppressing fixed pole indices, which would confound the frequency of interest with whatever else a particular index has learned, poles were selected independently for every sample and every spatial token whenever their instantaneous equivalent periods fell within pre-defined ±10% bands around 24, 12 or 8 h. The selection is therefore dynamic and frequency-defined, and it adapts to the input dependence of the pole positions.

Suppressing the diurnal, semidiurnal and terdiurnal bands individually increased the level-mean wind-speed RMSE by 0.18, 0.20 and 0.80%, and joint suppression increased it by 1.46%. Seven-day block-bootstrap intervals excluded zero for all four aggregate changes, so the components do participate in prediction. Three features of the result argue against a localised reading.

The effects are not frequency-selective in the way a one-pole-to-one-tide correspondence would require. Suppressing the diurnal band degraded several harmonics rather than principally the diurnal one, and the same held for the semidiurnal band, with a change in the matching harmonic complex error of 0.001 in both cases, as Table S12 shows. Only the terdiurnal band produced a clearly diagonal response, with a matching complex-error change of 0.025 and a phase MAE change of 2.414°, and that response did not remain significant after correction across all evaluated metrics.

More decisively, the joint tidal-band suppression produced a smaller RMSE increase than energy-matched non-tidal interventions, at 0.0324 against 0.0516 m s^{-1}. Removing an equivalent amount of energy from components at non-tidal frequencies degraded the forecast more than removing it from the tidal bands. The magnitudes are also small in absolute terms, since a 1.46% degradation from removing every tide-frequency rational component is not what a localised mechanism would produce.

Predictive information is therefore distributed redundantly across the rational basis and the parallel state-space pathway, and the poles constitute an auditable frequency coordinate system rather than a set of isolated causal oscillators. Suppressing a component at inference time on a network trained with that component present probes a counterfactual the network never encountered, and a redundant representation will compensate. Retraining with the tidal bands masked throughout would answer a different question.

Table S12. Dynamic tide-frequency interventions and matched controls.

Intervention	Change in speed RMSE (m s^{-1})	Change in matching complex error	Change in phase MAE (°)	Reading
Dynamic S1	0.004	0.001	0.034	No diagonal selectivity
Dynamic S2	0.005	0.001	0.153	No diagonal selectivity
Dynamic S3	0.018	0.025	2.414	Exploratory phase-selective

					response
Dynamic combined	S1, S2 and S3	0.032	—	—	Smaller than matched controls
Energy-matched control	non-tidal	0.052	—	—	Tidal set not uniquely important

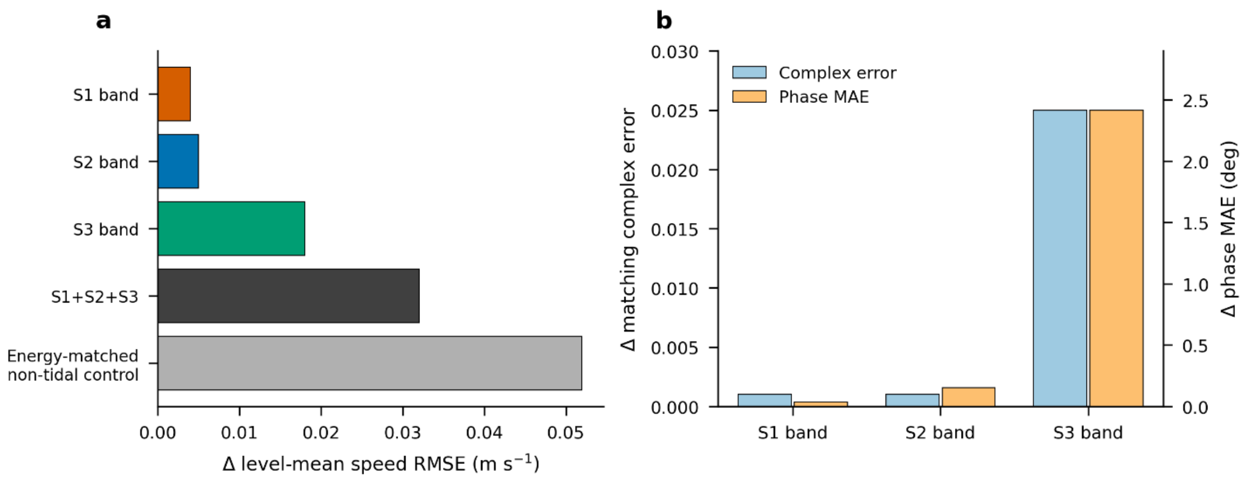


Figure S7. Dynamic frequency-band interventions. (a) Change in level-mean wind-speed RMSE for each band, for the combined tidal intervention and for the energy-matched non-tidal control. (b) Change in the matching harmonic complex error and in phase MAE for the three single-band interventions.

S5 Spatial and temporal spectral fidelity

Aggregate RMSE says nothing about which spatial scales a model reproduces. A model can score well by rendering the large scales accurately while smoothing away grid-scale variance, and in a field dominated by low-wavenumber structure this trade is barely visible in the mean error. Spatial fidelity was examined along two complementary lines.

The first line asks whether any smoothing is specific to ATMNO, and for that a cross-model analysis was run on the existing test predictions from ATMNO, Factformer and MWT. Zonal and meridional one-dimensional spectra were treated separately over the 4295 common test starts remaining after the boundary purge of Sect. 3.1 of the main text. Wavenumber was normalised by the Nyquist wavenumber along each axis, with low, mid and high bands defined as 0 to 0.25, 0.25 to 0.5, and above 0.5. Magnitude-squared coherence was computed per wavenumber and weighted by target energy within each band, and uncertainty came from 5000 paired seven-day block-bootstrap resamples.

No model dominates. ATMNO has the highest level-mean low- and mid-wavenumber coherence in both directions, at 0.925 and 0.199 zonally against 0.910 and 0.130 for Factformer and 0.911 and 0.184 for MWT, and at 0.969 and 0.415 meridionally against 0.963 and 0.310 for Factformer and 0.962 and 0.372 for MWT. That advantage stops at high wavenumbers. ATMNO high-band coherence is 0.013 zonally and 0.060 meridionally, below Factformer at 0.055 and 0.076.

Against MWT the zonal difference is also negative, while the meridional difference is small with a 95% interval spanning zero. High-band energy ratios tell the same story from the other side. ATMNO retains 0.307 of the ERA5 high-band energy zonally and 0.265 meridionally, against 0.467 and 0.466 for Factformer. MWT goes the other way and over-amplifies high-band energy in the level mean, by factors of 4.81 zonally and 1.33 meridionally, driven mainly by the upper levels. ATMNO under-represents grid-scale variance, Factformer under-represents it less, and MWT over-represents it.

The second line offers a complementary scalar view of ATMNO on its own, using a radial rather than directional decomposition. Table S13 shows that total spatial spectral energy tracks the ERA5 target closely at every level, with ratios between 0.991 and 1.018. The designated radial high-wavenumber fraction behaves very differently. It sits between 5.3 and 17.7% from 20 to 50 km, climbs to 65.0% at 60 km, exceeds unity at 70 km at 1.952, and returns to 68.8% at 80 km. The directional and radial diagnostics use different spectral decompositions and different band definitions, so their ratios are not expected to agree numerically. Both identify high-wavenumber fidelity as the principal spatial limitation, and both show that the character of that limitation changes above 50 km, from suppression at the lower levels to over-amplification at 70 km. The ERA5 high-wavenumber variance is itself small and weakly constrained at those levels, which makes the ratio unstable.

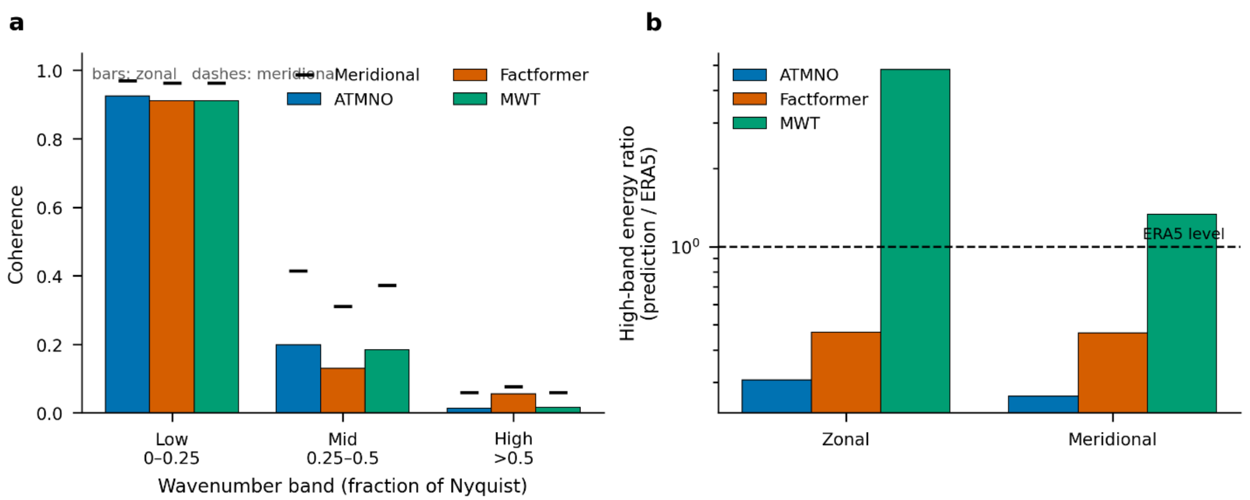
Table S13. Spatial spectral-energy ratios by altitude.

Altitude (km)	Total energy ratio, prediction to ERA5	High-wavenumber energy ratio
20	1.018	0.053
30	1.017	0.074
40	1.006	0.074
50	1.004	0.177
60	1.003	0.650
70	0.991	1 .952
80	1.006	0.688

Temporal spectral fidelity is markedly better than spatial. Table S14 compares the predicted and ERA5 harmonic amplitudes level by level. Agreement is close throughout. The largest relative discrepancy for the diurnal component is 4.1%, at 2.516 against 2.619 m s^{-1} at 30 km, and at 80 km the predicted amplitude of 10.086 m s^{-1} slightly exceeds the ERA5 value of 9.736 m s^{-1}. Semidiurnal amplitudes agree to within 10% at every level, and terdiurnal amplitudes are systematically low by 3 to 27%, with the largest shortfall at 30 km. Figure S9 shows the same result spectrally, with the predicted power spectral density reproducing the diurnal, semidiurnal and terdiurnal peaks and their altitude dependence at 30, 50 and 80 km, and a slight deficit in the inter-peak continuum at 30 and 50 km. The contrast between the two lines of evidence is systematic. The model reproduces the temporal spectrum of its target well and the spatial spectrum only partially, which is the pattern the architecture would predict, since its distinguishing feature is an adaptive temporal basis while its spatial module is a conventional factorized Fourier operator.

Table S14. Predicted and ERA5 temporal harmonic amplitudes by altitude, in m s^{-1}.

Altitude (km)	S1 predicted	S1 ERA5	S2 predicted	S2 ERA5	S3 predicted	S3 ERA5
20	3.437	3.618	1.832	1.973	0.449	0.528
30	2.516	2.619	1.393	1.532	0.331	0.455
40	4.190	4.199	1.142	1.288	0.372	0.498
50	7.832	7.744	1.234	1.381	0.320	0.381
60	9.258	9.058	2.269	2.303	0.463	0.486
70	7.431	7.193	3.403	3.343	0.598	0.604
80	10.086	9.736	6.419	6.357	1.042	1.103



230 **Figure S8. Directional spatial spectral fidelity across model architectures. (a) Level-mean target-energy-weighted magnitude-squared coherence in the low, mid and high wavenumber bands. Bars give zonal values and dashes meridional values. (b) High-band spectral-energy ratio relative to ERA5 on a logarithmic scale, where the dashed line marks exact agreement.**

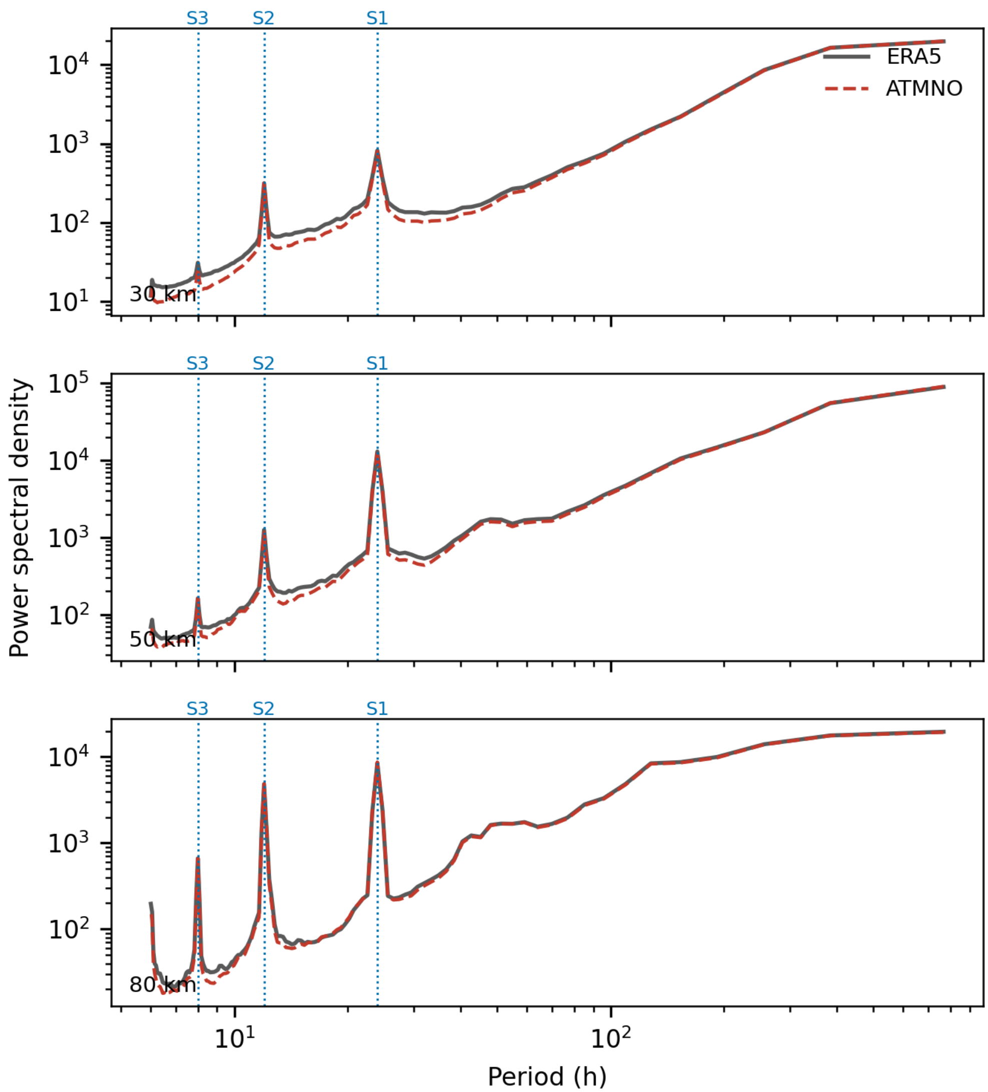


Figure S9. Temporal spectral fidelity at 30, 50 and 80 km. Predicted spectra reproduce the dominant diurnal, semidiurnal and terdiurnal peaks and their altitude dependence.

S6 Stratospheric circulation transitions

This section reports an exploratory diagnostic. The question is whether the learned pole periods respond to changes in the background circulation, which would be a further indication that the representation tracks something physical. What follows is a sign pattern across four winters rather than a validated event classifier, and Sect. 5.4 of the main text states the associated caveat.

Event dates are based on a regional 65–143° E mean zonal-wind reversal at 60° N and 30 km. Table S16 lists all reversals detected by this diagnostic over 2016 to 2025. Many are short-lived, at three to twelve days, and a smaller number persist for more than 90 days and correspond to final warmings. Four events were selected for detailed tracking: the 2024/25 final warming, which falls in the test period, and three comparison winters, the 2019/20 final warming and the 2023/24 and 2020/21 midwinter events, which fall in training or validation periods and are used only to interpret the learned representation.

Table S15 gives the phase-resolved pole statistics. Across the four winters, the planetary-scale pole period lengthened from pre-event to post-event in the two final warmings, by 32.55 h in 2024/25 and 52.80 h in 2019/20, and shortened in the two midwinter major warmings, by 15.39 h in 2023/24 and 56.80 h in 2020/21. The tidal poles moved far less. The terdiurnal pole changed by at most 0.15 h across all four events and the semidiurnal pole by at most 0.76 h. The diurnal pole shifted appreciably during the two final warmings, by 2.95 and 2.78 h, and little during the midwinter events, by 0.43 and 0.80 h.

With two events of each type the sign pattern would arise by chance one time in eight under a null of random signs. The differential response, with the planetary pole large and the tidal poles nearly static, is internally consistent with the expectation that a circulation transition alters planetary-wave conditions while leaving solar forcing unchanged. A larger event set drawn wholly from held-out data would be needed before this could be treated as more than hypothesis-generating.

Table S15. Phase-resolved statistics of tracked poles in four winters, equivalent period in h.

Winter(split)	Pole	Background	Pre	Onset	Post	Post−pre
2024/25(Test) Final warming	Diurnal S1 (P8)	23.38	24.5	26.09	27.45	2.95
	Semidiurnal S2(P15)	11.78	12.37	12.78	12.85	0.48
	Terdiurnal S3 (P16)	7.41	7.36	7.48	7.51	0.15
	Planetary (P29)	119.54	126.04	151.04	158.59	32.55
2019/20(Training) Final warming	Diurnal S1 (P8)	23.59	24.86	25.19	27.64	2.78
	Semidiurnal S2(P15)	11.83	12.58	12.51	12.62	0.04
	Terdiurnal S3 (P16)	7.47	7.54	7.56	7.54	0.00
	Planetary (P29)	139.15	164.07	213.35	216.87	52.8
2023/24(Validation) Midwinter SSW	Diurnal S1 (P8)	22.84	23.12	22.22	23.55	0.43
	Semidiurnal S2(P15)	11.82	11.8	11.93	12.56	0.76
	Terdiurnal S3 (P16)	7.50	7.36	7.49	7.49	0.13
	Planetary (P29)	154.33	148.28	151.82	132.89	−15.39
2020/21(Training) Midwinter SSW	Diurnal S1 (P8)	22.41	22.66	24.63	23.46	0.80
	Semidiurnal S2(P15)	11.65	11.8	11.86	12.11	0.31
	Terdiurnal S3 (P16)	7.52	7.47	7.57	7.55	0.08
	Planetary (P29)	146.5	202.46	203.78	145.66	−56.80

Table S16. Winter and spring westerly reversals detected by the regional diagnostic, 2016 to 2025.

Reversal date	Duration (days)	Minimum U (m s^{-1})
2016-03-05	13	−26.99
2016-03-29	144	−18.99
2017-04-08	15	−6.78
2018-02-10	19	−35.92
2018-04-18	124	−12.25
2018-12-28	4	−8.44
2019-03-01	121	−21.76
2020-01-25	3	−14.31
2020-03-03	116	−19.86
2020-11-04	12	−23.99
2020-12-04	6	−5.66
2021-02-24	122	−12.20
2022-01-17	6	−7.09
2023-01-04	3	−6.36
2023-02-19	12	−8.82
2023-11-18	3	−6.17
2024-01-04	30	−11.12
2024-02-25	3	−0.72
2024-03-21	96	−12.90
2025-01-09	3	−5.87
2025-02-20	125	−10.47

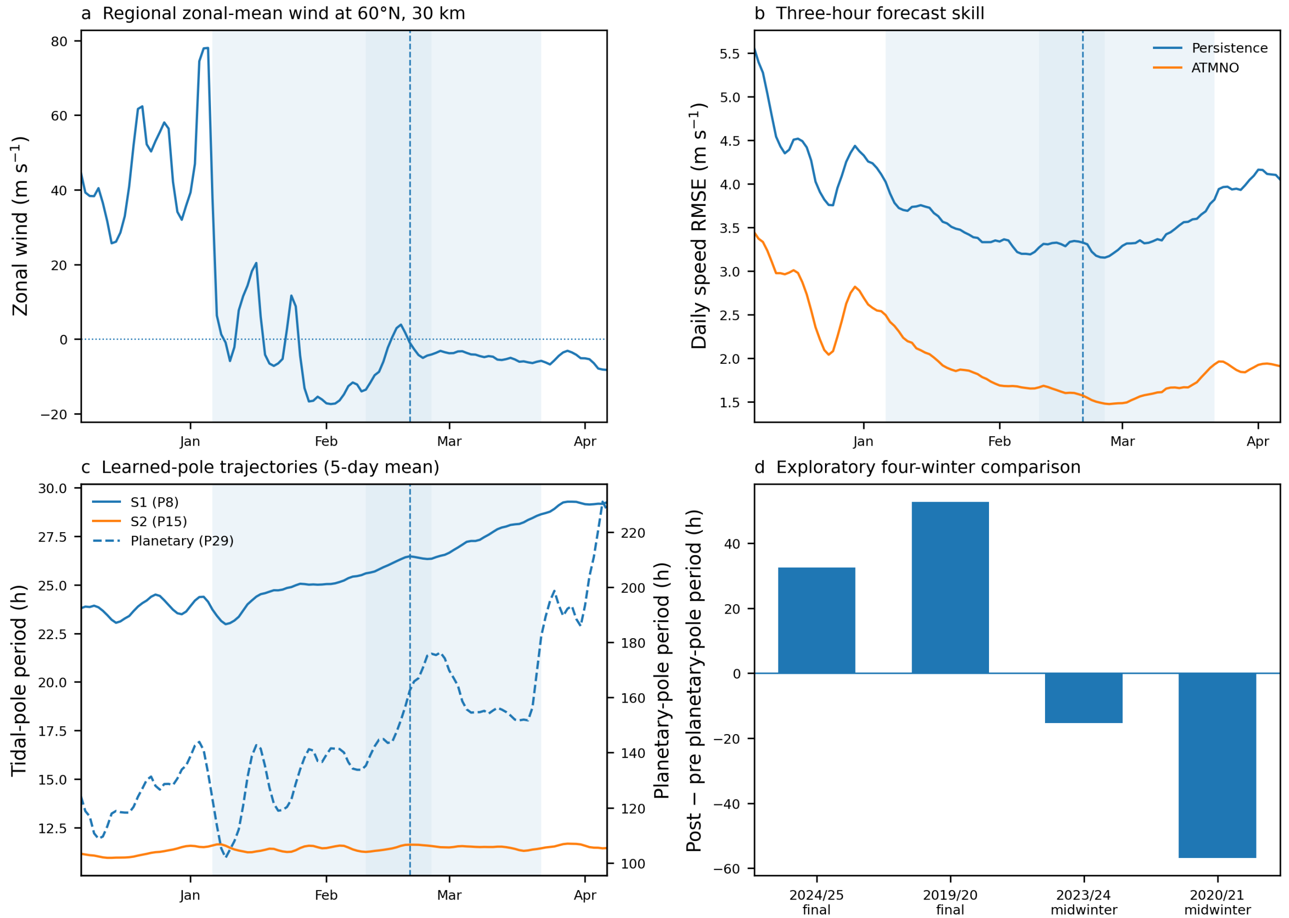


Figure S10. Exploratory circulation-transition diagnostics. (a) Regional zonal-mean wind at 60° N and 30 km around the 2024/25 final warming. (b) Five-day-smoothed daily three-hour wind-speed RMSE for ATMNO and persistence. (c) Five-day-smoothed diurnal, semidiurnal and planetary learned-pole periods. (d) Pre-to-post change in the planetary-pole period for the four illustrated winters.

References

He, M. and Chau, J. L.: Mesospheric semidiurnal tides and near-12 h waves through jointly analyzing observations of five specular meteor radars from three longitudinal sectors at boreal midlatitudes, Atmos. Chem. Phys., 19, 5993–6006, https://doi.org/10.5194/acp-19-5993-2019, 2019.

Holm, S.: A simple sequentially rejective multiple test procedure, Scand. J. Stat., 6, 65–70, https://www.jstor.org/stable/4615733, 1979.

Stober, G., Kuchar, A., Pokhotelov, D., Liu, H., Liu, H.-L., Schmidt, H., Jacobi, C., Baumgarten, K., Brown, P., Janches, D., Murphy, D., Kozlovsky, A., Lester, M., Belova, E., Kero, J., and Mitchell, N.: Interhemispheric differences of mesosphere–

lower thermosphere winds and tides investigated from three whole-atmosphere models and meteor radar observations, Atmos. Chem. Phys., 21, 13855–13902, https://doi.org/10.5194/acp-21-13855-2021, 2021.