https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

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)

Abstract. Winds between 20 and 80 km couple the lower and upper atmosphere through solar tides, planetary waves and gravity-wave forcing, and they set the operating envelope for near-space flight. Observations thin rapidly above the lower stratosphere, so reanalysis is in practice the only systematic source for studying and predicting this layer. Reanalyses reproduce the canonical solar-tidal harmonics here, and satellite comparisons of temperature indicate that their tidal amplitudes weaken too rapidly above the upper stratosphere. How much tidal wind amplitude survives near the mesopause has not been measured against ground-based observations, and the consequences for reanalysis-trained prediction have not been drawn. Winds from three meteor radars spanning 23° of latitude are compared with collocated ERA5, and an adaptive Takenaka–Malmquist neural operator is trained on seven-level ERA5 vector winds over eastern Asia and evaluated against eleven capacity-matched architectures, MERRA-2 and the radar observations. ERA5 retains 15 to 36%of the observed semidiurnal and 30 to 82%of the observed diurnal wind amplitude at 80 km. The operator attains the lowest wind-speed and direction errors evaluated, 2.215 $\mathrm{ms^{-1}}$ and 5.2°, and improves on three-hour ERA5 persistence at all three radar sites. Without frequency supervision its learned poles converge on the semidiurnal and terdiurnal harmonics, vary by only 4 to 11%across the domain, and reappear after retraining on MERRA-2. The amplitude deficit constrains what any reanalysis-trained model of this layer can reproduce, and the recovered frequencies give a representation that can be audited directly against observations.

1 Introduction

The middle atmosphere between roughly 20 and 80 km spans the upper stratosphere, the stratopause and the lower mesosphere. Winds in this layer organise vertical coupling through gravity-wave breaking, solar atmospheric tides, planetary-wave propagation, the quasi-biennial oscillation and sudden stratospheric warmings (Baldwin et al., 2001; Fritts and Alexander, 2003; Oberheide et al., 2009; He et al., 2024). They also set the operating envelope for stratospheric balloons, solar-powered unmanned aerial vehicles and near-space airships (Gonzalo et al., 2018; Zheng et al., 2024; Qi et al., 2026; Sheng et al., 2025). Dedicated forecasts nonetheless remain scarce. Observations become sparse above the lower

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

stratosphere, and operational systems treat much of the region as a sponge-damped model top rather than as a forecast target. Reanalysis is therefore the practical source for systematic model development, while independent radar observations establish how much of the upper-level variability such a source retains.

Data-driven weather prediction has advanced quickly. FourCastNet (Pathak et al., 2022), Pangu-Weather (Bi et al., 2023), GraphCast (Lam et al., 2023) and GenCast (Price et al., 2025) now match or surpass operational numerical weather prediction at medium range. NeuralGCM (Kochkov et al., 2024), together with the foundation models Aurora (Bodnar et al., 2025) and Aardvark Weather (Allen et al., 2025), has broadened the scope of machine-learning forecasting further. These systems target tropospheric variables almost exclusively, and their applicability above the tropopause has not been examined systematically. Three features of the middle atmosphere make the extension non-trivial. Observational constraints weaken rapidly with height, so reanalysis carries more model and less measurement as altitude increases (Gelaro et al., 2017; Hersbach et al., 2020). Temporal variance concentrates in altitude-dependent tidal and planetary-wave frequencies rather than spreading across the spectrum. The horizontal vector, not the speed alone, governs momentum transport, wave propagation and directional forecast utility.

These features expose a limitation shared by current architectures. The Fourier neural operator (Li et al., 2021) and its relatives capture global spatial structure efficiently through spectral convolution, but their treatment of time is restricted. Channel concatenation discards explicit temporal structure, and recurrent coupling introduces sequential bottlenecks over long input windows. State-space models such as Mamba (Gu and Dao, 2024) handle long sequences at linear cost yet learn spectral structure only implicitly. A standard Fourier temporal basis distributes resolution uniformly, which is inefficient when the energy sits at a small number of unevenly spaced frequencies. In the ERA5 winds analysed, 1.3 to 3.6%of frequency modes carry 80%of the variance, and the solar-tidal harmonics appear as narrow peaks near 24, 12 and 8h that are not uniformly spaced on the frequency axis. A uniform basis therefore spends most of its resolution where there is little to resolve.

Several threads inform the design adopted here. Neural operators learn maps between function spaces and provide a grounded framework for data-driven physical modelling (Kovachki et al., 2023). Beyond the Fourier neural operator (Li et al., 2021) and DeepONet (Lu et al., 2021), the adaptive Fourier neural operator (Guibas et al., 2022) and the spherical Fourier neural operator (Bonev et al., 2023) extend the family to token mixing and stable rollouts on the sphere. The factorized Fourier operator (Tran et al., 2023), multiwavelet operators (Gupta et al., 2021), latent-spectral models (Wu et al., 2023) and transformer-based solvers (Li et al., 2023; Wu et al., 2024a) constitute the operator baselines adopted here. Structured state-space models offer an efficient route to long-range sequence modelling, beginning with S4 (Gu et al., 2022) and continuing through Mamba (Gu and Dao, 2024) and its structured state-space duality formulation (Dao and Gu, 2024). Recent time-series variants (Cai et al., 2024; Ahamed and Cheng, 2024; Wang et al., 2025a) and frequency–state-space hybrids (Wang et al., 2024a; Wu et al., 2024b, 2025, 2026) nonetheless operate in the time domain without an explicit, adaptive frequency decomposition. The Takenaka–Malmquist system (Takenaka, 1925; Malmquist, 1926) builds orthonormal bases in the Hardy space from complex poles inside the unit disk, achieving non-uniform spectral concentration

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

controlled by pole position rather than the uniform resolution of the Fourier basis. It is widely used for the identification of resonant systems (Wahlberg, 1991; Ninness and Gustafsson, 1997; Heuberger et al., 2005) and for efficient signal representation (Szabó, 2003), but not, to our knowledge, as a learnable temporal module for atmospheric prediction.

The dynamics this layer hosts are themselves demanding. The quasi-biennial oscillation (Baldwin et al., 2001), gravity-wave driving (Fritts and Alexander, 2003), atmospheric tides (Oberheide et al., 2009; Vincent, 2015) and sudden stratospheric warmings (Limpasuvan et al., 2004; Charlton and Polvani, 2007; Gasperini et al., 2023) all leave signatures in the wind field. Against this, data-driven studies of middle-atmosphere wind remain few. One distilled an analogue ensemble into a network below 30 km (Candido et al., 2020). Another used single-station meteor-radar data in the mesosphere and lower thermosphere (Mauricio et al., 2024). Earlier work established the general feasibility of learned forecasting (Scher, 2018; Weyn et al., 2019). Separately, the representation of solar tides in modern reanalyses has been assessed against satellite and ground-based measurements (Sakazaki et al., 2012, 2018; Keckhut et al., 2023), and meteor-radar chains have characterised mesospheric tidal variability directly, including its modulation by the quasi-biennial oscillation over China (He and Chau, 2019; Stober et al., 2020, 2021; Wang et al., 2024b). These two literatures have largely not met. Reanalysis tidal fidelity has been quantified mostly in temperature and mostly from satellite comparison, and the implications for reanalysis-trained prediction have not been drawn.

This study addresses that gap. Three meteor radars establish how much tidal amplitude ERA5 retains at 80 km, which constrains what any reanalysis-trained model of this region can reproduce. An adaptive temporal representation then tests whether the residual signal can be exploited. The model introduced here, the adaptive Takenaka–Malmquist neural operator (ATMNO), runs two complementary temporal paths in parallel. An adaptive Takenaka–Malmquist decomposition predicts complex poles from each input and places spectral resolution where the signal concentrates, a bidirectional Mamba path captures transient sequence structure, and a learned gate combines them. Spatial structure is handled by a factorized Fourier operator with a complementary large-kernel local branch, and the wind vector is predicted jointly.

The remainder of the paper is organised as follows. Section 2 describes the ERA5 and MERRA-2 fields and the meteor-radar winds. Section 3 sets out the forecast problem, the architecture, the harmonic decomposition and the statistical treatment. Section 4 presents the results in seven parts: aggregate forecast skill, the contribution of each temporal component, the frequencies the adaptive basis recovers, their corroboration in the ERA5 spectra, their reproducibility across reanalyses, harmonic phase fidelity relative to the strongest baselines, and the meteor-radar constraints. Section 5 interprets these results and states the limitations, and Sect. 6 concludes. The Supplement contains the multi-horizon and seasonal diagnostics, the full harmonic comparison, the station-level radar verification, the intervention experiments, the spectral-fidelity analysis and the circulation-transition diagnostics.

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

2 Data

2.1 ERA5

We use ERA5 (Hersbach et al., 2020) over 0–60°N and 65–143°E at 3-hourly resolution from January 2016 to December 2025, giving 28736 time steps on a 61×79 grid at 1° spacing. Seven altitude levels are mapped from the ERA5 model levels to geometric altitude by nearest-level assignment, with a deviation of at most 0.6 km. The 20 km level maps to L49 at 20.3 km, 30 km to L31 at 29.9 km, 40 km to L20 at 39.9 km, 50 km to L13 at 50.2 km, 60 km to L8 at 60.1 km, 70 km to L4 at 69.4 km, and 80 km to L1 at 80.3 km. This mapping reproduces the expected mesospheric jet maximum near 60 km with a reversal aloft, and the 80 km level approximates the mesopause, extending the predictive range above the 70 km ceiling of prior data-driven studies.

ERA5 assimilates few direct wind observations above the lower stratosphere, so the upper levels are constrained largely by the forecast model. Above roughly 50 km the fields are additionally affected by the sponge layer applied near the model top. Section 4.7 quantifies what this means for tidal amplitude at 80 km.

2.2 MERRA-2

MERRA-2 (Gelaro et al., 2017) is used over the same domain, the same seven altitude levels, the same ten-year period and the same 3-hourly sampling. It serves as an independent spectral reference and as a separate training source for a model retrained from scratch, which tests whether the frequencies recovered from ERA5 reflect the method or a property of one reanalysis.

2.3 Meteor-radar winds

Independent 2024 middle- and upper-atmospheric wind data were obtained for the Mohe, Beijing-Shisanling and Wuhan stations of the Beijing National Observatory of Space Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, through the Geophysics Center, National Earth System Science Data Center. The three sites lie at 53.5, 40.3 and 30.5° N and operate MDR6 meteor radars at 38.9 MHz. Background winds are derived from Doppler shifts of underdense meteor-trail echoes using standard meteor-radar processing and poly-pulse-pair velocity estimation (Hocking et al., 2001; Reid, 2024). Records carrying platform error code 1 were discarded. The supplied products define zonal wind as positive eastward and meridional wind as positive northward, and no data-dependent sign reversal was applied at any stage.

For the harmonic-amplitude comparison of Sect. 4.7, hourly winds at the nominal 80 km level were aggregated to the 3-hourly ERA5 grid over 11 July to 31 December 2024. For direct forecast validation, observations at $t-1$, t and t+1h were vector-averaged and compared with the forecast valid at t, with zero temporal shift. Three-hour persistence is the collocated ERA5 wind at $t-3$ h, and the valid-time ERA5 analysis is retained as a reference rather than as a forecast baseline. Sample sizes after quality control are 782 at Beijing, 1335 at Mohe and 1327 at Wuhan.

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

3 Methods

3.1 Forecast, data splitting, and evaluation

Near-space wind-field prediction is formulated as a spatiotemporal operator-learning problem. Let u_t denote the stacked zonal and meridional wind field at time step t, with $2C$ channels over a spatial grid of $H \times W$, where C is the number of altitude levels. Given T consecutive fields, the operator G_θ predicts the field at the next step,

$$\hat{u}(t+1) \;=\; G\_\theta\big(\, u(t-T+1),\; \dots,\; u(t)\,\big) \qquad (1)$$

In our configuration C =7 levels at 20, 30, 40, 50, 60, 70 and 80 km, the grid is 61×79, and T =16, a 48h window. The two output channels per level give the wind speed s and the wind direction $\theta = atan2(v,u)$, so the operator delivers the full horizontal wind vector.

The data are split chronologically to prevent temporal leakage. The first 70%, or 20115 steps, is used for training, the next 15%, or 4310 steps, for validation, and the final 15%, or 4311 steps, for testing. Each level's zonal and meridional components are normalised independently with a robust scaler centred on the training-set median and scaled by the interquartile range. The fitted training statistics are then held fixed for validation, test, rollout and seasonal evaluation. Wind components beyond prescribed physical bounds are clipped before normalisation, affecting fewer than 0.1% of values.

Adjacent 3-hourly samples share 15 of their 16 input steps, so successive test cases are strongly autocorrelated. This does not bias the point estimates, but a naive count of test samples overstates the effective sample size, and per-sample significance tests must therefore be read as descriptive. Two measures were applied to the splits themselves, and a third to the inference (Sect. 3.9). For the 6, 12 and 24 h experiments the fixed chronological boundaries were purged using the maximum 24 h lead, so that no raw timestamp is shared across training, validation and test subsets. All three direct models therefore use identical sample starts, at 20 092 training, 4287 validation and 4304 test samples, and direct forecasts and persistence are evaluated on the same test starts.

3.2 The ATMNO operator

ATMNO stacks L = 3blocks at a hidden width of D=56, for a total of 1.16 million parameters. Figure 1 summarises the design. Each time step is embedded independently, the temporal core operates on a pooled token grid, and the final step is decoded back to full resolution.

The wind field, concatenated with normalised latitude and longitude, is lifted to the model dimension by a convolutional encoder whose second convolution uses stride two to reduce the grid to about 31 × 40. A learnable temporal positional encoding and a coordinate-derived spatial encoding are then added. Within each block a factorized Fourier operator retaining the lowest 16 modes along the height and width axes is summed with a large-kernel-attention branch (Guo et al., 2023),

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

followed by a residual connection and group normalisation. A depthwise three-dimensional convolution along the time axis at full resolution supplies local temporal coupling. The spatial features are pooled to a 16 × 20 token grid, where the adaptive Takenaka–Malmquist path and the bidirectional Mamba path process each token's length-T sequence in parallel and are merged by a learned gate. The result is interpolated back to full resolution and passed through a feed-forward network, with pre-layer-normalisation and residual connections wrapping every sub-operation. The last-step representation is upsampled to the full 61 × 79 grid, refined by a large-kernel residual block, and read out to the $2C = 14$ output channels.

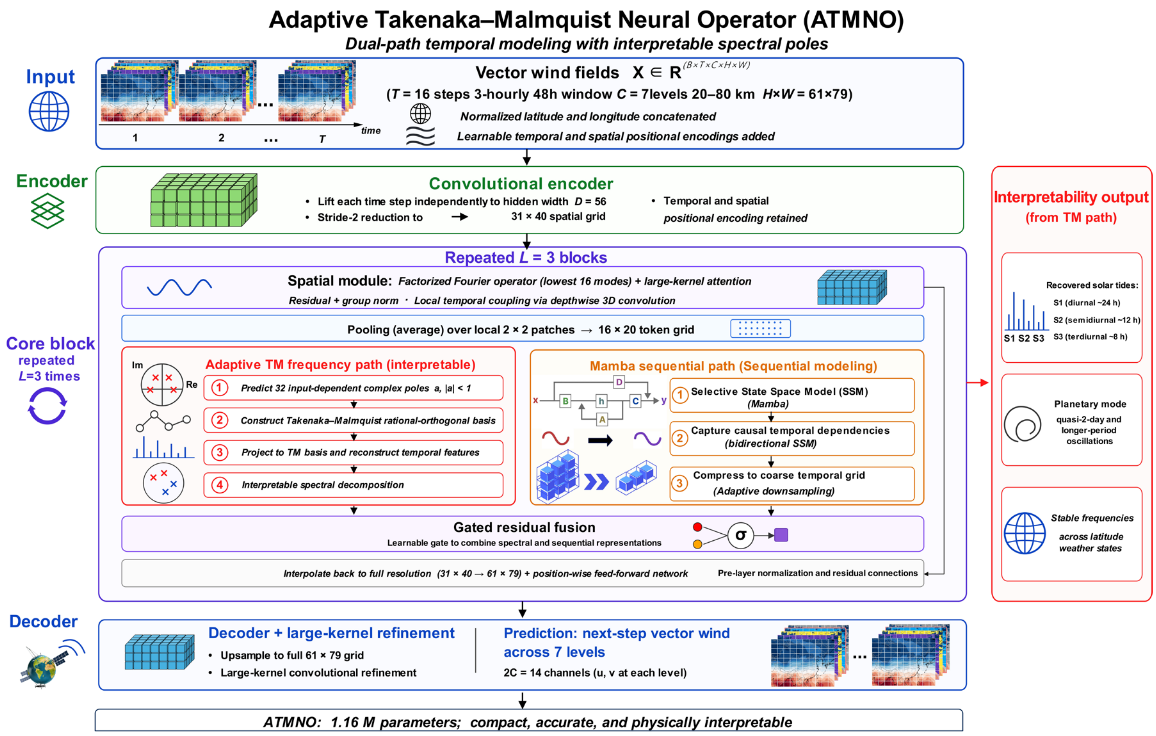


Figure 1. Architecture of ATMNO. Vector-wind input is processed by three blocks combining a factorized Fourier and large-kernel spatial module with parallel adaptive Takenaka–Malmquist and bidirectional Mamba temporal paths, then decoded to seven-level zonal and meridional wind forecasts.

3.3 Adaptive Takenaka–Malmquist temporal representation

Given complex poles α_k with modulus below unity, the Takenaka–Malmquist basis functions are

$$\varphi_k(z)=\frac{\sqrt{1-|\alpha_k|^2}}{1-\overline{\alpha_k}z}\prod_{j<k}\frac{z-\alpha_j}{1-\overline{\alpha_j}z} \qquad (2)$$

The leading Cauchy kernel concentrates spectral energy near the pole angle, and the Blaschke product enforces orthonormality in the Hardy space. The pole angle sets the centre frequency and the modulus controls bandwidth. Poles near the unit circle give narrow, frequency-selective filters, whereas poles near the origin approach broadband Fourier modes.

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

When all poles sit at the origin, the system reduces to the standard Fourier basis with uniform resolution. The basis can therefore place resolution at a small number of arbitrary, unevenly spaced frequencies, which is what the middle-atmosphere spectrum requires.

Rather than fixing the poles in advance, a two-layer perceptron acting on the temporal-mean feature predicts each complex pole and maps it to the open unit disk,

$$\hat{\alpha}_k = \tanh\left(p_k^{\mathrm{re}}\right) + i\tanh\left(p_k^{\mathrm{im}}\right), \qquad |\hat{\alpha}_k| \le 0.95 \tag{3}$$

The modulus is clipped at 0.95 with the angle preserved, for numerical stability. The poles are initialised on a radius-0.5 circle to encourage diverse frequency coverage, and 32 poles are used. The temporal sequence is decomposed, transformed in coefficient space and reconstructed,

$$c_k = \sum_t z_t \varphi_k\left(e^{i\theta_t}\right) \tag{4}$$

$$\hat{c}_k = f_\phi(c_k) \tag{5}$$

$$z_t^{TM} = \sum_k \hat{c}_k \varphi_k\left(e^{i\theta_t}\right) \tag{6}$$

The learned coefficient transform can amplify, attenuate or phase-shift individual frequency components according to their predictive value. Each pole angle is a frequency and each modulus a bandwidth, so the temporal representation is directly readable as a spectrum. This property makes the analysis of Sect. 4.3 possible.

3.4 Sequential path and gated fusion

The second temporal path is a bidirectional Mamba module (Gu and Dao, 2024) with state dimension 16, convolutional kernel 4 and expansion factor 2. Bidirectional processing is admissible because the input window is fully observed at inference time, and no information from after the forecast valid time enters the model.

The two temporal outputs are combined by concatenation, projection and a gated residual connection:

$$z_{fused} = W_f\left[z^{TM}; z^{Mamba}\right] \tag{7}$$

$$g = \sigma\left(W_g\left[z_{fused}; z_{res}\right]\right) \tag{8}$$

$$z_{out} = g \odot z_{fused} + (1-g) \odot z_{res} \tag{9}$$

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

Here z_{res} is the representation before temporal processing, so the sigmoid gate controls how much temporal modification is applied at each location and time step.

3.5 Pole extraction and the frequency-locking criterion

Pole statistics are pooled over 153600 spatiotemporal token locations drawn from 480 test windows, giving the distribution of each of the 32 poles in the first temporal core. A pole is called frequency-locked when its pooled angular standard deviation satisfies $\sigma_\theta < 25°$ and its mean modulus exceeds 0.35. In the tidal bands the angular threshold corresponds to a period uncertainty narrower than the separation between adjacent harmonics. The modulus threshold excludes near-origin poles, which approach broadband Fourier modes and carry no meaningful centre frequency. Both thresholds were fixed before the pole analysis was run.

Applying the criterion leaves the six entries of Table 4 together with a small number of further locked poles outside the tidal bands, and the remainder are broadband or angularly diffuse. The selectivity of the criterion is itself informative, since an architecture producing narrow angular distributions indiscriminately would offer no discrimination between frequencies. Figure 5c shows the locking plane with the criterion marked.

3.6 Baseline architectures and training

Eleven baselines were evaluated: Factformer (Li et al., 2023), MWT (Gupta et al., 2021), AFNO (Guibas et al., 2022), F-FNO (Tran et al., 2023), Swin Transformer (Liu et al., 2021), ConvLSTM (Shi et al., 2015), FNO (Li et al., 2021), UNet (Ronneberger et al., 2015), Transolver (Wu et al., 2024a), DeepONet (Lu et al., 2021) and LSM (Wu et al., 2023). All use the same ERA5 fields, the same 16-step input window, the same chronological splits, the same robust normalisation, the same composite vector loss, the same optimiser settings and the same validation-based checkpoint-selection criterion, within a parameter budget of approximately 0.9 to 1.2 million.

Two non-learned references accompany the architecture comparison and are reported alongside it. Three-hour ERA5 persistence gives the skill obtainable by assuming the field unchanged, and a station-wise harmonic model comprising a linear trend plus diurnal, semidiurnal and terdiurnal terms gives the skill obtainable from tidal climatology alone. Both are evaluated on identical samples and appear in Sect. 4.7 and in Supplement Sects. S1 and S3.

The loss is computed on the normalised vector prediction together with derived quantities, so that direction error cannot be hidden by suppressing speed:

$$\mathcal{L} = \mathrm{MSE}(u,v) + 5\,\mathrm{MSE}(s) + 0.5\left(1 - \cos\Delta\theta\right) \qquad (10)$$

Optimisation uses AdamW (Loshchilov and Hutter, 2019) at a learning rate of 3×10^{-4} and weight decay of 10^{-5}, with a cosine-annealing schedule to a minimum of 10^{-6}, gradient clipping at norm 1.0, a batch size of 8, and early stopping with patience 12 on validation loss.

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

3.7 Evaluation metrics

Predicted zonal and meridional winds are inverse-transformed with the fixed per-level robust scalers before evaluation. Speed RMSE and MAE are computed from the scalar speed difference over all samples and grid points. Vector RMSE pools the two component errors. The angular error $\Delta\theta$ is obtained from the normalised dot product of the predicted and observed vectors. Direction is ill-defined under weak winds, so direction MAE and the fractions within 20 and 30° are evaluated only where the observed wind speed exceeds 3 ms^{-1}. The same threshold applies to Table 2, the spatial direction-error maps and the seasonal diagnostics. Anomaly correlation is the Pearson correlation between predicted and observed wind-speed anomalies after mean removal.

3.8 Harmonic decomposition and tidal-fidelity metrics

ATMNO, Factformer and MWT predictions were analysed over 17 complete calendar months from June 2024 to October 2025. For each model, month, altitude, grid point and wind component, an intercept, a linear trend and joint cosine–sine terms at 24, 12 and 8h were fitted. The identical decomposition was applied to the ERA5 target and, in Sect. 4.7, to the radar winds, so that all three sources pass through the same estimator. This follows the practice adopted in meteor-radar and model intercomparisons, where processing asymmetry can otherwise masquerade as physical difference (Stober et al., 2021).
Harmonic fidelity is summarised by normalised complex-coefficient error, normalised amplitude error, amplitude-weighted circular phase MAE and complex spatial-pattern coherence. Model differences are paired by month. Confidence intervals come from 10000 paired month-bootstrap draws, and one-sided Wilcoxon tests are adjusted across the evaluated harmonic metrics within each model comparison by the Holm procedure (Holm, 1979). The same predefined decomposition and endpoints were used for all three models and were fixed before the comparison was run.

3.9 Statistical treatment

For the architecture comparison, paired per-sample wind-speed errors are summarised by the fraction of test cases on which ATMNO has lower error and by two-sided Wilcoxon signed-rank tests, with the corresponding block-bootstrap intervals given alongside. The multi-horizon comparison between direct forecasts and persistence uses 5000 paired seven-day block-bootstrap resamples of the sequential sample-level speed-RMSE difference. The tidal-fidelity comparison uses complete calendar months as paired units. Meteor-radar forecast differences use seven-day calendar-block bootstrap intervals with 17 to 25 blocks per station; Dynamic frequency-band interventions use seven-day block bootstrap for aggregate RMSE together with empirical energy-matched non-tidal controls, and the full intervention results are reported in Supplement Sect. S4.

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

4 Results

4.1 Vector-wind forecast skill

Table 1 reports the validation-selected checkpoint for each architecture. Across the twelve configurations the domain-mean wind-speed RMSE spans 2.215 to 8.570 ms^{-1}. The operator family occupies the lower half of that range, and the two general-purpose sequence models, ConvLSTM and Swin Transformer, fall among the weaker operators. LSM did not converge under the vector objective because its instance normalisation is unstable with the composite speed-weighted loss, and it is retained only as a negative reference.

ATMNO attains the lowest wind-speed RMSE, 2.215 ms^{-1}, and the lowest direction MAE, 5.2°. Its zonal and meridional errors, 2.350 and 1.765 ms^{-1}, are likewise the lowest evaluated. The margin over the two strongest baselines is 4.7%against Factformer and 5.1%against MWT. Paired at the sample level, ATMNO has the lower speed error in 88.6%of test cases against Factformer and 92.8%against MWT. The corresponding seven-day block-bootstrap mean differences are 0.110 ms^{-1}, with a 95%confidence interval of 0.097 to 0.124, and 0.125 ms^{-1}, with an interval of 0.115 to 0.136. Both intervals exclude zero.

Both non-learned references are substantially weaker. Three-hour ERA5 persistence gives a level-mean speed RMSE of 4.121 ms^{-1}, so the learned models improve on it by 44 to 46 %. The station-wise harmonic model, evaluated against the radar winds in Sect. 4.7, gives vector-wind errors two to three times those of ATMNO. Neither reference is competitive at this lead, and the ordering among architectures in Table 1 is therefore not an artefact of a strongly periodic target that any method would reproduce.

Table 1. Vector-wind forecast skill on ERA5. Direction MAE uses grid points with observed wind speed above 3 ms^{-1}.

Model	Family	Params	u RMSE (ms^{-1})	v RMSE (ms^{-1})	Speed RMSE (ms^{-1})	Dir. MAE (°)
ATMNO	operator	1.16 M	**2.350**	**1.765**	**2.215**	**5.2**
Factformer	operator	1.01 M	2.464	1.879	2.323	5.5
MWT	operator	1.04 M	2.486	1.901	2.333	5.7
AFNO	operator	0.99 M	2.624	1.983	2.473	6.0
F-FNO	operator	1.09 M	2.769	2.048	2.590	6.1
Swin Transformer	general	0.92 M	2.750	2.236	2.613	6.6
ConvLSTM	general	1.05 M	2.883	2.313	2.709	6.7
FNO	operator	1.20 M	2.924	2.160	2.739	6.6
UNet	general	0.99 M	2.874	2.020	2.787	6.0
Transolver	operator	1.14 M	3.063	2.108	2.852	6.5
DeepONet	operator	0.99 M	4.849	2.867	4.496	10.5
LSM	operator	1.22 M	10.269	3.348	8.570	19.1

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

Per-altitude skill is summarised in Table 2. ATMNO records the lowest speed RMSE at each of the seven levels as well as the lowest mean. The zonal error exceeds the meridional error at every level, 1.288 against 1.107 ms^{-1} at 20 km and 3.157 against 2.387 ms^{-1} at the 50 km maximum. This gap tracks the larger climatological variance of the zonal wind at these altitudes rather than any anisotropy of the model. Speed RMSE peaks near 50–60 km, where wind variability is greatest, and decreases toward 80 km. Direction MAE ranges from 4.3 to 6.5° across levels and averages 5.2°, and more than 93%of valid-direction predictions fall within 20° of the truth at every altitude. Predicted and observed wind speeds correlate at anomaly correlation coefficients of 0.990 to 0.994 at every level, so the dominant errors are in amplitude rather than in large-scale pattern.

Table 2. Per-altitude skill of the final ATMNO checkpoint. Direction metrics use grid points with observed wind speed above 3ms^{-1}.

Altitude	*u* RMSE	*v* RMSE	Speed RMSE	Speed MAE	Dir. MAE (°)	< 20°(%)	ACC
20 km	1.288	1.107	1.231	0.939	5.5	96.5	0.991
30 km	1.648	1.321	1.553	1.186	5.1	96.3	0.994
40 km	2.439	1.965	2.283	1.703	6.5	93.8	0.993
50 km	3.157	2.387	2.979	2.090	5.7	95.3	0.990
60 km	3.067	2.115	2.912	2.015	4.3	97.1	0.994
70 km	2.720	1.768	2.558	1.834	4.4	97.2	0.993
80 km	2.133	1.692	1.990	1.496	4.9	96.6	0.990
Mean	2.350	1.765	2.215	1.610	5.2	96.1	0.992

Pooling the components gives a mean vector RMSE of 2.94 ms^{-1}. Figure 2 places the architecture comparison in context, Fig. 3 maps the spatial distribution of the test-set errors at all seven levels, and Fig. 4 gives the pointwise predicted-versus-observed densities with level-specific calibration statistics.

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

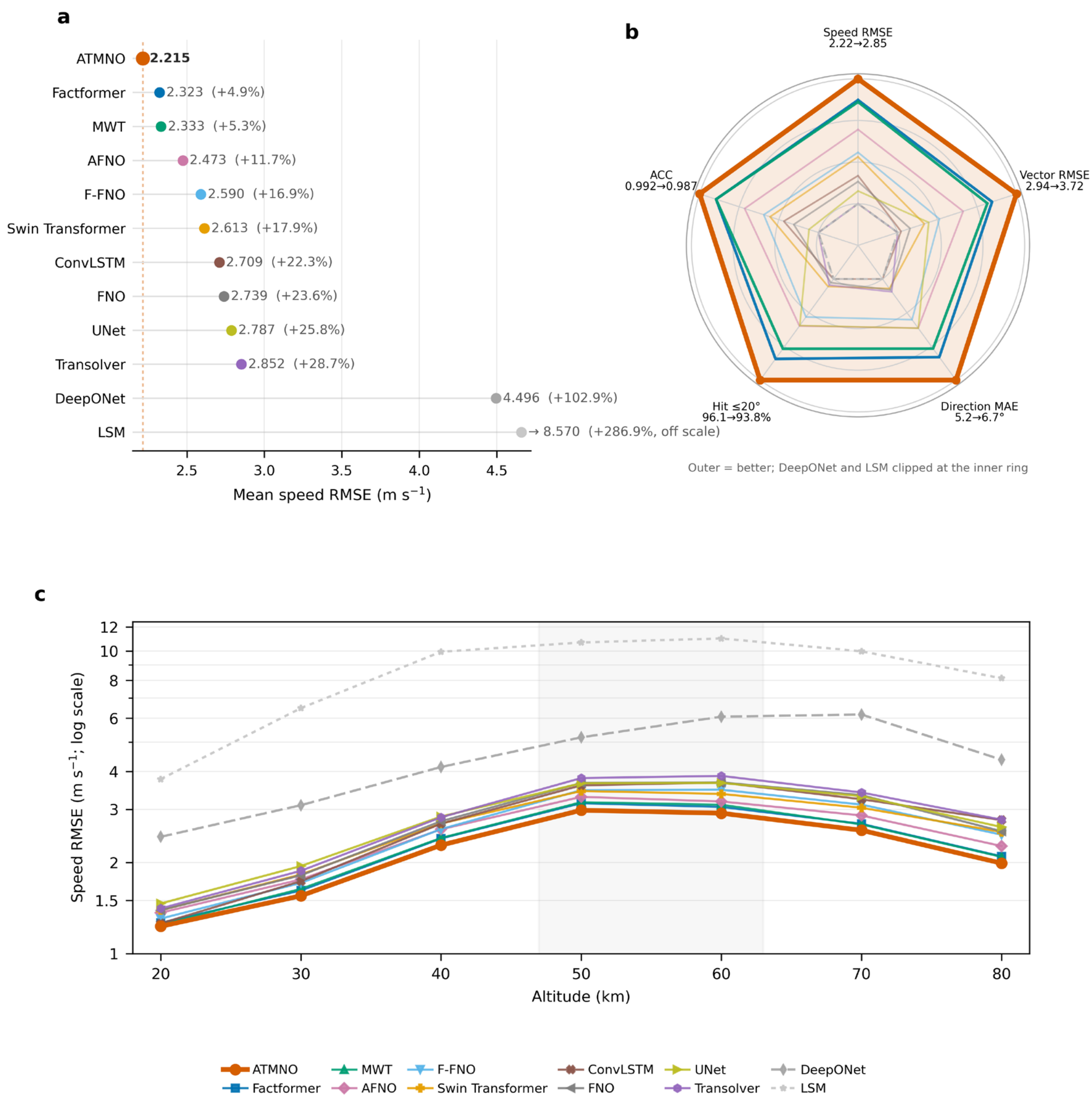


Figure 2. Comparative vector-wind forecast skill. (a) Domain-mean wind-speed RMSE across the evaluated architectures, with seven-day block-bootstrap 95%intervals. (b) Wind-speed RMSE by altitude on a logarithmic scale.

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

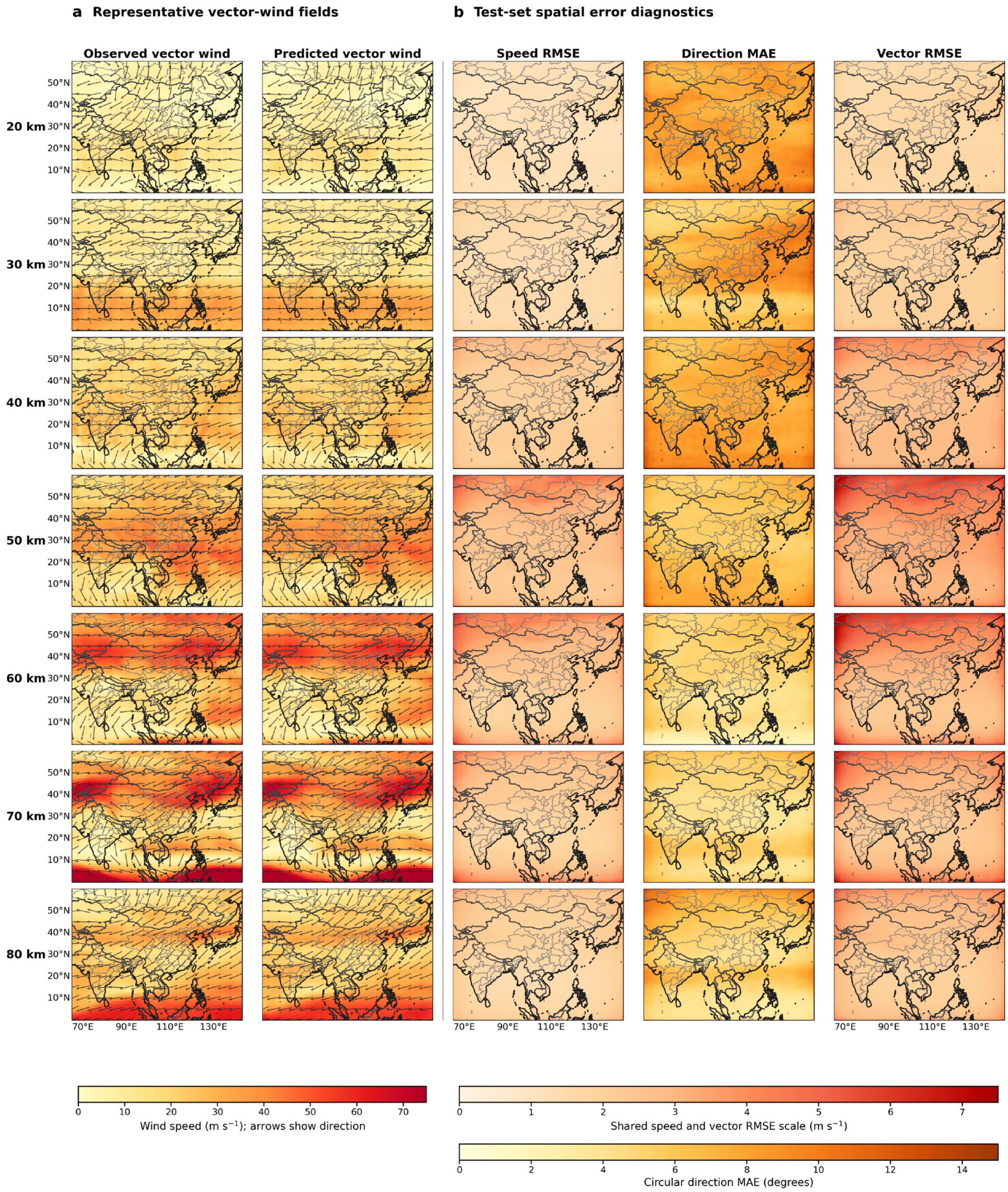


Figure 3. All-level spatial reconstruction and error diagnostics, 20–80 km. (a) Representative observed and predicted vector winds. (b) Test-set speed RMSE, direction MAE and vector RMSE. Coastlines are from Natural Earth and no national boundaries are drawn.

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

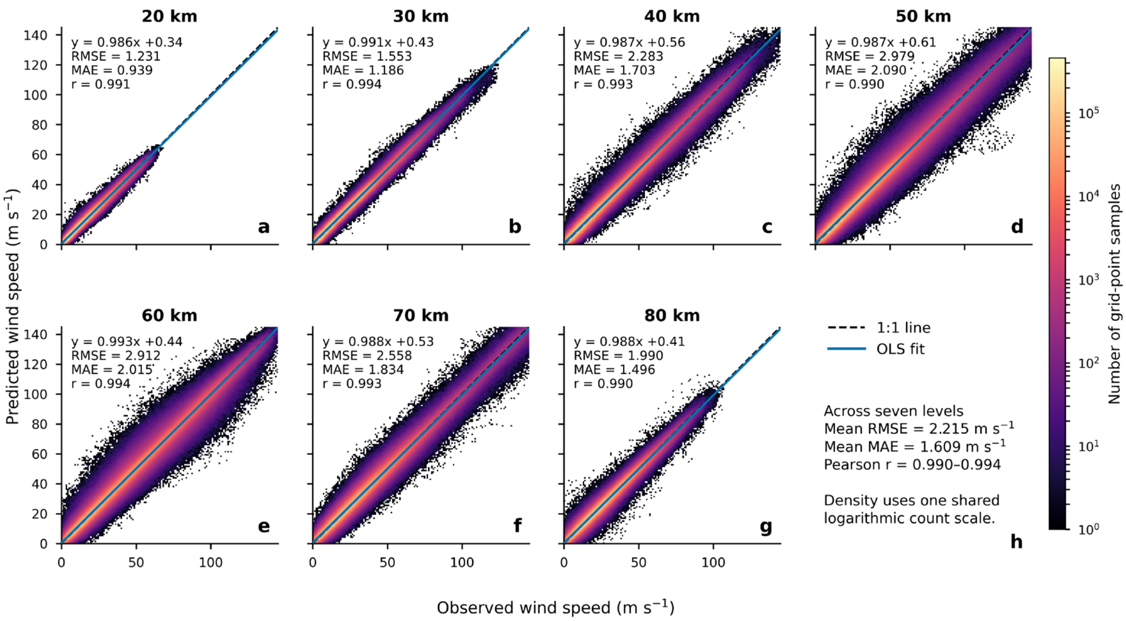


Figure 4. Predicted versus observed wind speed by altitude. Panels (a) to (g) show density histograms with one-to-one and ordinary-least-squares lines for 20–80 km. Across-level statistics and the shared logarithmic density scale are given in the legend.

4.2 Contribution of the temporal components

Four reduced configurations isolate the role of each element of the temporal core. All were trained under identical data, splits, normalisation, loss, optimiser settings and checkpoint-selection criteria, and Table 3 reports the outcome.

Every reduction degrades performance. Retaining only the Mamba path costs 3.61 %, and retaining only the rational path costs 4.38 %. Neither path is sufficient on its own, and the two single-path variants lie within 0.8%of each other, which indicates that they carry substantially overlapping information.

The two variants that isolate the adaptivity of the basis rather than its presence are more informative. Fixing the poles at their initialisation, so that the rational basis is present but non-adaptive, costs 4.11 %, more than removing the rational path altogether. Replacing the rational basis with a uniform Fourier basis of matched parameter count costs 4.97 %, the largest degradation of the four. A fixed set of 32 poles at radius 0.5 imposes a spectral prior that is unhelpful when it does not match the data, whereas the Mamba path alone imposes none. The benefit therefore lies in the adaptivity of the pole positions rather than in the rational form itself. Section 4.6 shows that the same ordering holds for harmonic fidelity, and that the Fourier-basis substitution is costliest for the terdiurnal phase.

Table 3. Temporal-component configurations of ATMNO.

Variant	Params	Speed RMSE (ms^{-1})	Change relative to full model
ATMNO, full: TM and Mamba	1.164 M	2.215	—

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

Mamba only, no TM	1.115 M	2.295	+3.61 %
Fixed poles, non-adaptive	1.143 M	2.306	+4.11 %
TM only, no Mamba	0.929 M	2.312	+4.38 %
Fourier basis replacing TM	1.143 M	2.325	+4.97 %

4.3 Frequencies recovered by the adaptive basis

Each pole angle is a frequency and each modulus a bandwidth, so the temporal core admits a direct spectral reading. Pooling 153600 spatiotemporal token locations from 480 test windows recovers the distribution of each of the 32 poles in the first temporal core. Table 4 lists those satisfying the frequency-locking criterion, together with their locations relative to the solar-tidal bands.

A subset of the poles concentrates tightly. Two sit in the terdiurnal band with median equivalent periods of 7.5 and 8.2h and angular spreads of 7.5 and 10.4°. A third has a mean angle of −88.2°, only 1.8° from the exact 12h angular frequency, with a median equivalent period of 12.9h and an angular spread of 11.1°. The diurnal band is less tightly defined. P17 lies near 24.3h but varies strongly with latitude, whereas the more angle-stable P8 is centred near 28.0h. A further pole, P29, sits near 222.9h and corresponds to a low-frequency, planetary-scale component. Its period lies beyond the 48h input window, so Table 4 reports it without a quantitative period interpretation.

Coincidence with a tidal period is necessary but not sufficient, so the leading alternative was tested directly. Near-inertial oscillations have a period of 12h divided by the sine of latitude, which more than doubles across this domain, and a pole tracking that mechanism should inherit the same latitude dependence. Figure 5d bins each learned pole by latitude. The two terdiurnal poles vary by only 4.4 and 8.6% across the domain and the semidiurnal pole by 10.7%, against a theoretical inertial period that changes by more than a factor of two over the same range. The semidiurnal pole anti-correlates with the inertial curve, at r=−0.92. The diurnal-band candidates behave differently: P8 varies by 15.5% with weak inertial correlation at r=−0.22, and P17 by 39.5% at r=+0.05.

These results support a hierarchy rather than a uniform conclusion. The semidiurnal and terdiurnal poles are frequency-locked, and their latitude behaviour matches that expected of migrating, sun-synchronous components whose frequencies are set by solar forcing rather than by local rotation. The diurnal band is broader and more run-dependent, and Sect. 4.5 shows that it does not survive retraining on a second reanalysis.

Table 4. Selected stable and tide-band poles of the main ATMNO checkpoint, block 0.

Pole	θ (°)	θ std (°)	Median period (h)	Modulus	Latitude variation	Interpretation
P16	−145.8	7.5	7.5	0.512	4.4 %	Terdiurnal S3
P26	−128.0	10.4	8.2	0.695	8.6 %	Terdiurnal S3
P15	−88.2	11.1	12.9	0.588	10.7 %	Semidiurnal S2

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

P17	+49.7	21.2	24.3	0.363	39.5 %	Diurnal-band candidate S1
P8	−40.2	8.5	28.0	0.494	15.5 %	Diurnal-band candidate S1
P29	−2.9	7.1	222.9	0.504	—	Low-frequency, planetary scale

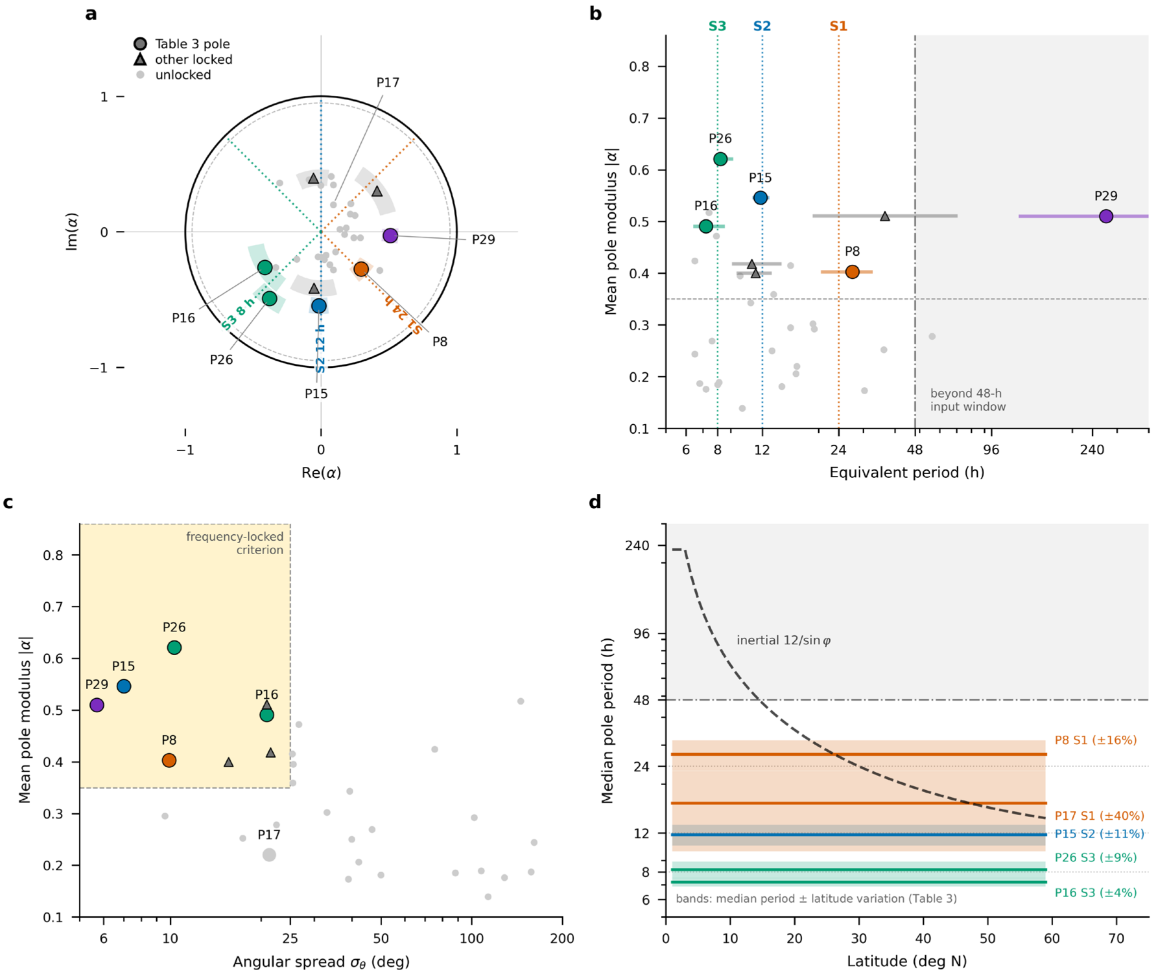


Figure 5. Learned Takenaka–Malmquist poles and solar-tidal frequency bands. (a) Mean pole positions in the unit disk, with rays at the exact diurnal, semidiurnal and terdiurnal angular frequencies and wedges showing each locked pole's angular spread. (b) Mean pole modulus against median equivalent period. Bar length spans the period range implied by the angular spread and is inversely related to how tightly a frequency is locked. (c) Locking plane: angular spread against modulus, with the frequency-locked marked. (d) Latitude dependence of the tide-band poles compared with the near-inertial period. Shading in (b) to (d) marks periods beyond the 48h input window.

4.4 Spectral corroboration in the ERA5 winds

The pole analysis is internal to the model, and the same frequencies can be tested against the data the model was trained on. Figure 6 shows the per-grid-point Welch (1967) power spectral density of the ERA5 zonal and meridional winds at each

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

altitude, power-averaged over the domain, with the learned tide-band periods overlaid. Table5 quantifies spectral concentration and tidal-band energy.

The spectrum is strongly non-uniform at every level. Of 1025 frequency bins, 13 to 37, or 1.3 to 3.6%, contain 80% of the variance, and 36 to 71 contain 90%. This concentration is a property of the data, established independently of any model, and it is what motivates adaptive rather than uniformly spaced frequency resolution.

Distinct diurnal, semidiurnal and terdiurnal peaks occur at every level. Their energy varies non-monotonically between adjacent levels but strengthens overall toward the upper mesosphere. Between 20 and 80km the diurnal fraction rises from 1.70 to 4.15%, the semidiurnal from 0.85 to 2.68%, and the terdiurnal from 0.24 to 0.48%. The learned terdiurnal poles at 7.5 and 8.2h and the semidiurnal pole centred angularly near the 12h frequency align with the corresponding spectral bands, whereas the diurnal representation is broader around the 24h band. This reproduces, from an entirely separate diagnostic, the hierarchy found in Sect. 4.3.

These periods and their vertical structure match established tidal phenomenology. Periods near 24, 12 and 8h correspond to the canonical migrating harmonics of the solar day, and previous evaluations of modern reanalyses and satellite products identify the same structures (Sakazaki et al., 2018; Keckhut et al., 2023). Strengthening toward the upper levels is consistent with the upward amplification of thermally forced tides as density decreases (Fritts and Alexander, 2003; Oberheide et al., 2009), although the level-to-level changes are not strictly monotonic. The increasing relative prominence of the semidiurnal component at 80km matches the known transition toward semidiurnal dominance in the upper mesosphere (Smith, 2012; Garcia, 2023). Latitude stability is what migrating, sun-synchronous tides would produce. Amplitude comparison requires more care. ERA5 damps tidal amplitudes above about 50km through its upper-level sponge layer, and reanalysis amplitudes in the upper stratosphere and mesosphere are known to be weaker than satellite observations (Sakazaki et al., 2012). Classical migrating-tide calculations driven by tropospheric latent heating likewise place strong diurnal and semidiurnal components at these periods (Hagan and Forbes, 2002, 2003). Section 4.7 provides the independent frequency check and quantifies the amplitude limitation directly.

Table 5. ERA5 spectral concentration and tidal-band energy fractions. Welch spectra are computed per grid point and power-averaged over the domain.

Altitude	Modes for 80 %	Modes for 90 %	S1, 24 h	S2, 12 h	S3, 8 h
20 km	22, or 2.1 %	65	1.70 %	0.85 %	0.24 %
30 km	16, or 1.6 %	37	1.28 %	0.68 %	0.23 %
40 km	15, or 1.5 %	36	1.76 %	0.58 %	0.24 %
50 km	15, or 1.5 %	38	3.07 %	0.58 %	0.18 %
60 km	13, or 1.3 %	39	3.04 %	0.72 %	0.17 %
70 km	14, or 1.4 %	43	2.00 %	0.89 %	0.19 %
80 km	37, or 3.6 %	71	4.15 %	2.68 %	0.48 %

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

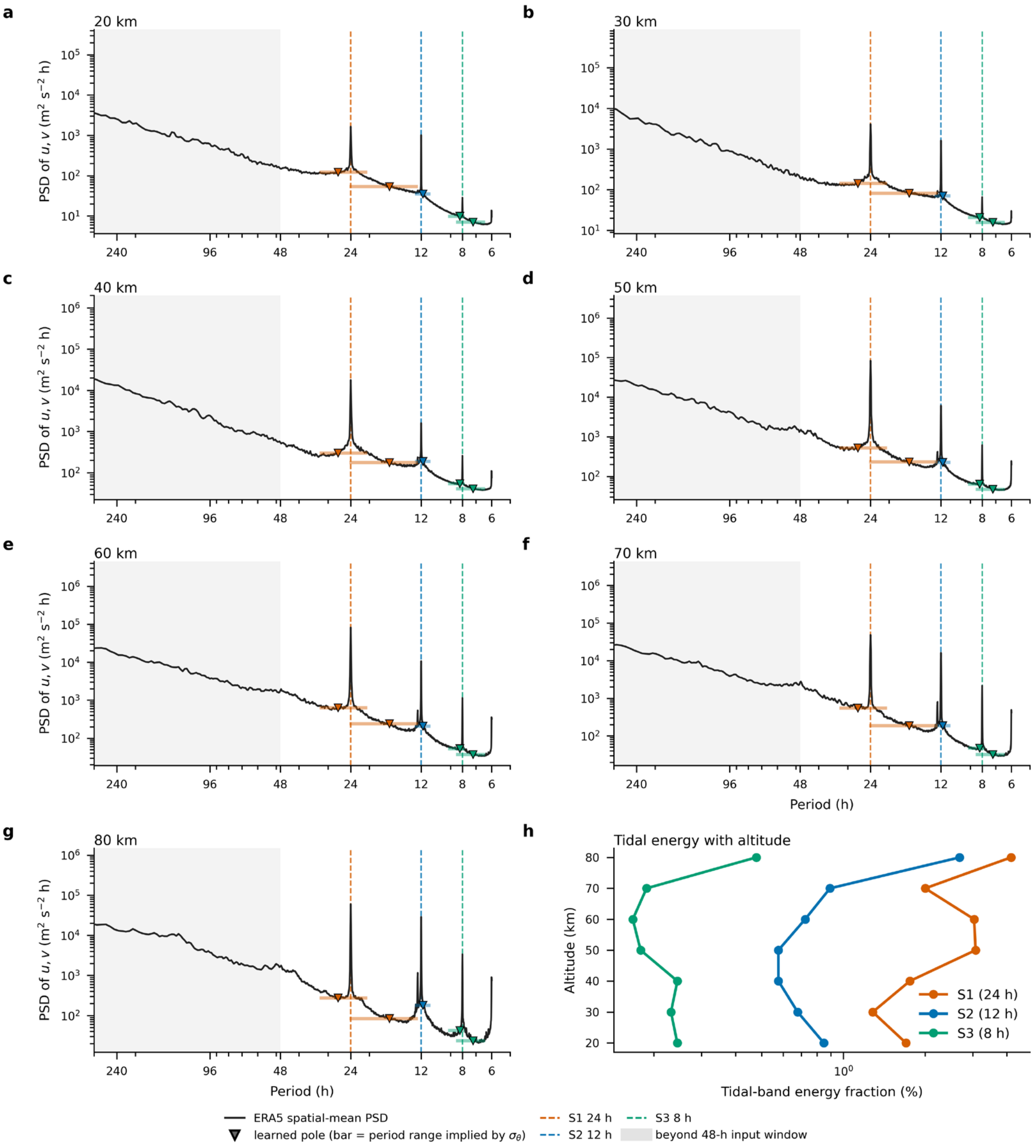


Figure 6. ERA5 spectra and tidal-band energy. Panels (a) to (g) show spatial-mean Welch spectra at 20–80 km with tidal and learned-pole periods. (h) Diurnal, semidiurnal and terdiurnal energy fractions versus altitude.

4.5 Cross-reanalysis reproducibility

The recovery established so far rests on ERA5 alone. The spectral and pole analyses were therefore repeated on MERRA-2 over the same domain, the same seven levels, the same ten-year period and the same 3-hourly sampling, combining direct spectral analysis, zero-shot application of the ERA5-trained checkpoint, and a model retrained from scratch. Figure 7a and b summarise the outcome.

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

The MERRA-2 spectra reproduce diurnal, semidiurnal and terdiurnal peaks at every altitude, as Table 6 shows. Their energy increases strongly with height, particularly above 50 km, though not strictly monotonically at the lower levels. The diurnal fraction changes from 1.26%at 20 km to 1.06%at 30 km before reaching 22.27%at 80 km, while the semidiurnal rises from 0.45 to 8.22%and the terdiurnal from 0.02 to 0.85 %. The absolute fractions at the upper levels are several times those in ERA5, which reflects how differently the two reanalyses treat the upper mesosphere and is taken up in Sect. 5.2.

Applied without retraining, the ERA5 model retains a 12.3h semidiurnal lock and terdiurnal poles at 7.8 and 8.3 h. A separately retrained MERRA-2 model recovers semidiurnal poles at 11.7 and 12.1h and a terdiurnal pole at 7.7 h, with a mean speed RMSE of 1.98 ms^{-1} and a direction MAE of 4.2° on its own test segment. The diurnal component is again the weak case. A candidate occurs near 27h in zero-shot transfer but does not meet the locking criterion after MERRA-2 retraining.

The pattern across all three settings matches that of Sect. 4.3 and 4.4. Semidiurnal and terdiurnal recovery is reproducible across reanalyses and across independent training runs, and diurnal recovery is not. The recovered periods are stable even though the two reanalyses disagree substantially on tidal amplitude, which separates the frequency result from the amplitude result.

Table 6. MERRA-2 tidal-band energy fractions across seven altitude levels.

Altitude	S1, 24 h	S2, 12 h	S3, 8 h
20 km	1.26 %	0.45 %	0.02 %
30 km	1.06 %	0.27 %	0.02 %
40 km	1.15 %	0.23 %	0.02 %
50 km	3.52 %	0.45 %	0.04 %
60 km	8.92 %	1.08 %	0.09 %
70 km	9.21 %	3.29 %	0.27 %
80 km	22.27 %	8.22 %	0.85 %

4.6 Tidal phase fidelity across model architectures

Recovering a frequency internally is not the same as predicting the corresponding field well. Harmonic fidelity was therefore compared across ATMNO, Factformer and MWT over 17 complete test months, using the decomposition of Sect. 3.8. Figure 8a to c shows the altitude dependence, and Supplement Sect. S2 gives the full tables.

The advantage is specific to phase. Amplitude-weighted phase MAE is lowest for ATMNO at every harmonic and against both comparators, and all six paired monthly comparisons remain significant after Holm correction within each model pair. The magnitudes differ sharply between harmonics. Against Factformer the diurnal and semidiurnal phase improvements are 0.173 and 0.672°, with 95%confidence intervals of 0.093 to 0.256 and 0.527 to 0.824, against level-mean phase errors of

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

4.33 and 10.12°. The terdiurnal margin is far larger, at 6.168° against Factformer, with an interval of 5.527 to 6.820, and 2.264° against MWT, with an interval of 1.666 to 2.868.

No other harmonic metric follows this pattern. Against Factformer, ATMNO has slightly higher diurnal and semidiurnal complex NRMSE, higher amplitude NMAE and marginally lower pattern coherence. Against MWT, ATMNO is better across all diurnal metrics and equivalent for the semidiurnal. For the terdiurnal component ATMNO additionally reduces the complex harmonic error by 6.9% relative to Factformer and 3.6% relative to MWT, and raises complex-pattern coherence by 0.021 and 0.020. The improvement is therefore concentrated in the phase of the tidal field rather than in its amplitude or its spatial pattern, an asymmetry taken up in Sect. 5.1.

Absolute terdiurnal fidelity remains low. The level-mean terdiurnal complex NRMSE is 0.80 for ATMNO, approaches or exceeds unity below 40 km, and falls below 0.5 only above 60 km. The terdiurnal result is accordingly a relative improvement in phase and complex-field fidelity within a weakly resolved component.

The reduced configurations of Sect. 4.2 degrade harmonic fidelity in the same order as they degrade aggregate skill. Replacing the adaptive rational basis with a uniform Fourier basis raises the level-mean terdiurnal phase error by 2.12°, which is comparable to the entire terdiurnal phase advantage of ATMNO over MWT. All 24 paired monthly comparisons against the full model favour the full model, with Holm-adjusted $P \leq 0.001$. That the adaptive basis matters most for phase, and most for the harmonic that is worst resolved, is consistent with the distributed representation indicated by the interventions of Supplement Sect. S4.

4.7 Meteor-radar constraints on tidal amplitude and forecast skill

The analyses above rest on reanalysis, which in the upper mesosphere is a model product. Direct observations close the validation chain.

Diurnal, semidiurnal and terdiurnal harmonic amplitudes were fitted separately for each station and wind component, with the identical least-squares fit applied to ERA5 at the nearest grid point and level. Table 7 lists the results, and Fig. 7c and d display them. All three harmonics are detected at every station in both wind components. The frequencies the model selected are therefore present in the observed mesosphere across more than 23° of latitude, an observational check independent of the reanalysis spectra of Sect. 4.4.

The amplitudes tell a different story. ERA5 retains 30 to 82%of the observed diurnal amplitude and 15 to 36%of the observed semidiurnal amplitude at 80 km, while terdiurnal amplitudes are approximately 1 ms^{-1} or less in both sources. The deficit is largest at Wuhan, the lowest-latitude station, where ERA5 retains 30%of the observed diurnal zonal amplitude and 29%of the semidiurnal. It is smallest at Mohe, where diurnal retention reaches 81 to 82%but semidiurnal retention still falls to 15 to 20 %. The semidiurnal deficit is thus both large and consistent across three stations spanning 23° of latitude, whereas the diurnal deficit varies strongly with latitude.

Table 7. Tidal amplitudes at 80 km from meteor-radar winds over 11 July to 31 December 2024 and from collocated ERA5, in ms^{-1}.

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

Station	Comp.	S1 obs.	S1 ERA5	S2 obs.	S2 ERA5	S3 obs.	S3 ERA5
Mohe, 53.5° N	*u*	5.33	4.33	5.08	1.00	1.07	0.38
Mohe, 53.5° N	*v*	6.92	5.68	5.55	0.81	1.45	0.34
Beijing, 40.3° N	*u*	5.51	2.83	6.62	1.71	0.44	0.33
Beijing, 40.3° N	*v*	7.26	4.72	6.33	2.01	0.70	0.68
Wuhan, 30.5° N	*u*	8.03	2.38	7.95	2.30	1.18	0.30
Wuhan, 30.5° N	*v*	12.49	4.34	8.75	3.17	1.29	0.61

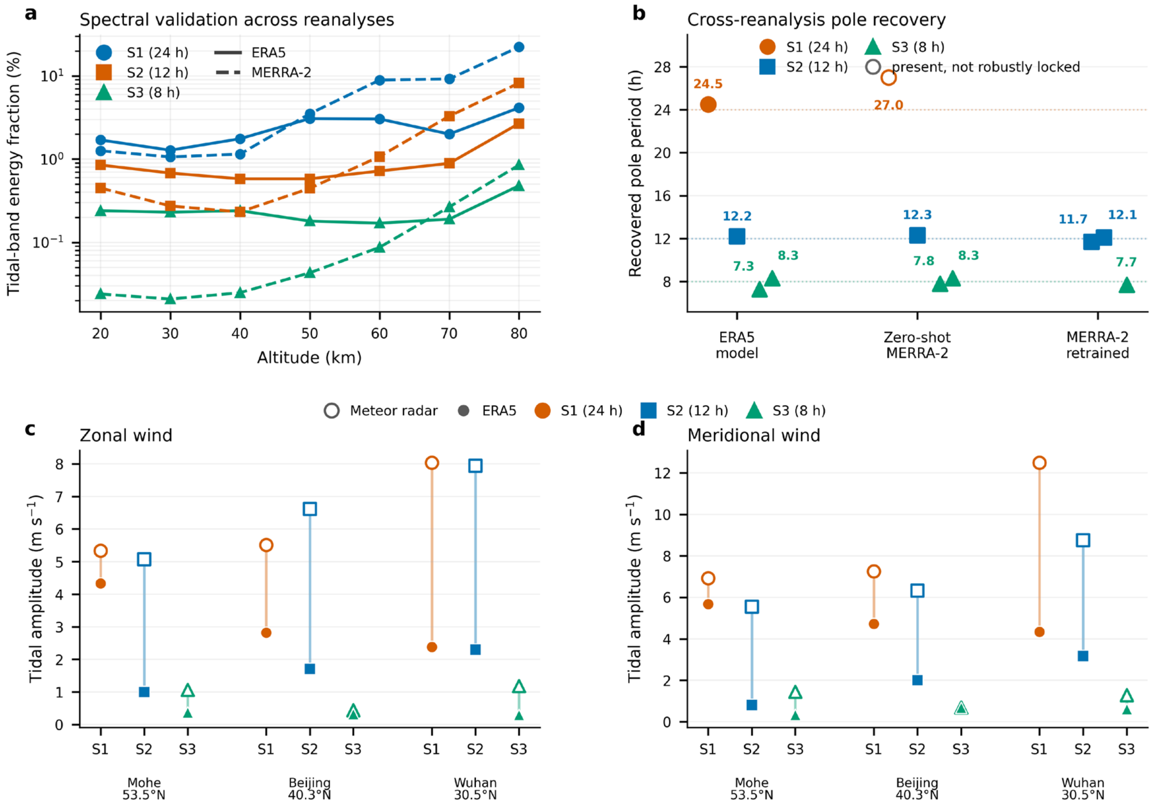


Figure 7. Cross-reanalysis and meteor-radar validation. (a) ERA5 and MERRA-2 tidal-band energy. (b) Recovered pole periods across three reanalysis settings. (c, d) Meteor-radar and collocated ERA5 tidal amplitudes at 80 km for the zonal and meridional components.

Under the pre-specified centred three-hour collocation, ATMNO reduced vector-wind RMSE relative to ERA5 persistence by 6.8, 9.0 and 4.4% at Beijing, Mohe and Wuhan. Improvements occurred in both components at all three stations, with zonal-wind RMSE reductions of 2.6 to 7.5% and meridional-wind reductions of 7.1 to 10.6%. Direction MAE fell from 65.6 to 56.7° at Beijing, from 62.9 to 54.8° at Mohe, and from 68.3 to 63.5° at Wuhan. ATMNO remained close to the valid-time ERA5 analysis, so the forecast retained most of the observational agreement of its reanalysis target. Figure 8d and e display

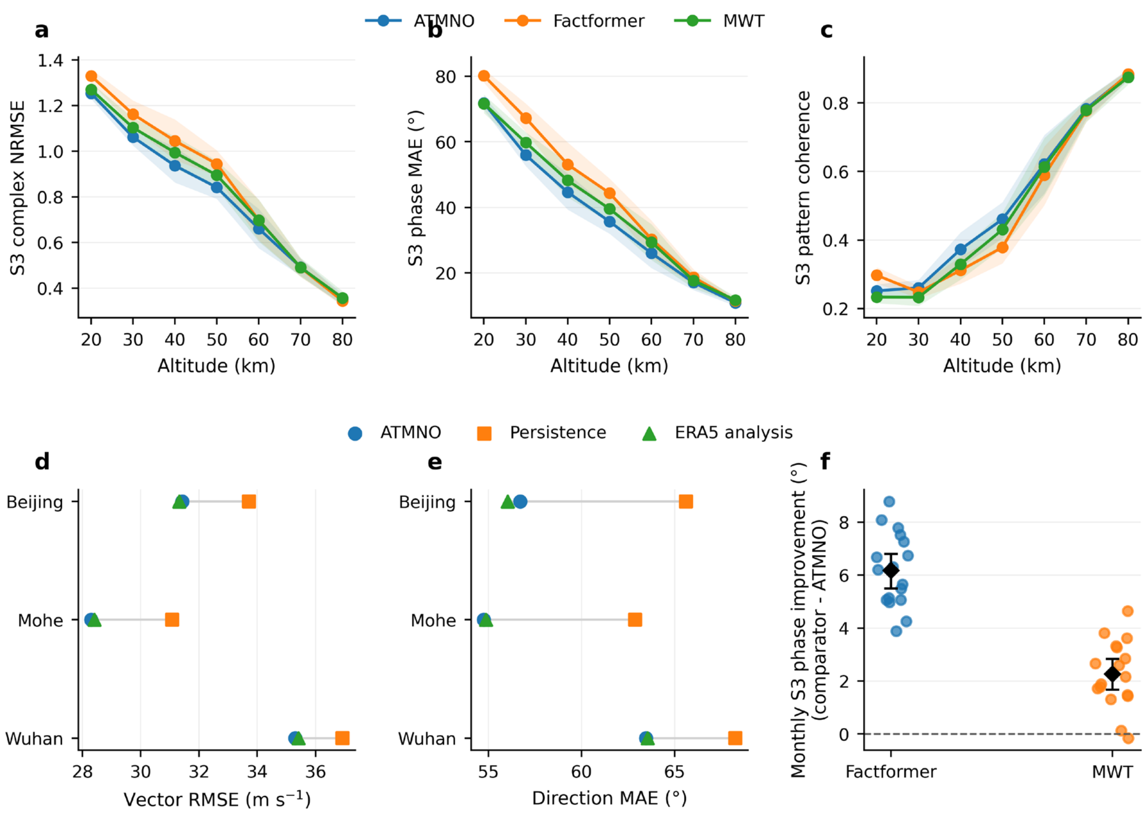
https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

these comparisons, and Supplement Sect. S3 gives component-level scores, paired block-bootstrap intervals and temporal-alignment sensitivities. The ranking of ATMNO over persistence holds for every station, variable and alignment tested.

Absolute errors remain substantial, including a wind-speed underestimate of approximately 10 to 11 ms^{-1} at every station. The same bias appears in persistence and in the valid-time analysis, so it is a property of ERA5 at this altitude rather than of the forecast.

The strictly causal harmonic baseline provides a further reference. Fitting a station-wise trend plus diurnal, semidiurnal and terdiurnal terms gives vector-wind RMSE of 68.7, 98.6 and 55.4 ms^{-1} at Beijing, Mohe and Wuhan, two to three times the error of ATMNO, with direction MAE above 100° at every station. Its calibration and evaluation windows fall in different seasons and it carries no day-to-day variability, so it represents a lower bound on harmonic-extrapolation skill. It nonetheless establishes that tidal climatology alone accounts for only a small part of the predictable variance at 80km, and that the phase fidelity documented in Sect. 4.6 is not climatology recovered by another route.

ATMNO models retained lower level-mean wind-speed RMSE than persistence at 6, 12 and 24h, with relative reductions of 40.3, 26.9 and 6.5%. The narrowing at 24h coincides with a local minimum in persistence error associated with tidal recurrence. Seasonal errors were largest in DJF and concentrated near 50–60 km. Supplement Sect. S1 reports the full lead-aligned results, block-bootstrap intervals, altitude-resolved errors and seasonal diagnostics.



https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

Figure 8. Tidal forecast fidelity and independent meteor-radar validation. (a) to (c) Terdiurnal complex NRMSE, amplitude-weighted phase MAE and complex-pattern coherence by altitude for ATMNO, Factformer and MWT over 17 complete test months. Shading denotes 95%month-bootstrap intervals. (d, e) Vector-wind RMSE and direction MAE at the nominal 80 km level against meteor-radar winds at Beijing, Mohe and Wuhan, comparing ATMNO three-hour forecasts, ERA5 persistence and the valid-time ERA5 analysis. (f) Paired monthly terdiurnal phase-error improvement, defined as comparator minus ATMNO. Diamonds show mean differences with 95%month-bootstrap intervals.

5 Discussion

5.1 What the learned poles represent

Four diagnostics converge on the semidiurnal and terdiurnal poles. They recur across ERA5 checkpoints and MERRA-2 experiments, remain close to latitude-independent, align with peaks in the underlying wind spectra, and correspond to harmonics detected in independent meteor-radar observations. Three of the four are external to the model. The diurnal pole does not survive the same tests and is treated throughout as a supporting rather than a decisive result.

The poles are not a set of isolated causal carriers. The interventions of Supplement Sect. S4 suppress dynamically identified tide-frequency rational components at inference time and produce only modest degradation: increases of 0.18, 0.20 and 0.80% in level-mean speed RMSE for the three bands individually, and 1.46% jointly. Joint tidal suppression produces a smaller increase than energy-matched non-tidal interventions, at 0.0324 against 0.0516 ms^{-1}. A representation in which individual poles carried individual tides would not behave this way.

The poles are better described as a frequency-auditable coordinate system embedded in a distributed predictor. The architecture supports that reading. The rational basis, the coefficient transform, the gating and the parallel Mamba pathway can encode the same predictive information redundantly, and the ablations of Sect. 4.2 show that removing either temporal path costs skill without either path being individually sufficient. The same reading explains the asymmetry of Sect. 4.6. If tidal information is distributed and partially redundant, an adaptive frequency basis should sharpen the timing of the represented oscillation, where placing resolution precisely at the right frequency matters most, while contributing comparatively little to amplitude or spatial pattern, which depend on the whole predictor. That is the observed pattern.

One alternative explanation deserves attention. The recovered frequencies might reflect the spectral structure of the training data rather than a physical mechanism the model has captured, and a 48h input window contains two diurnal, four semidiurnal and six terdiurnal cycles, which is more than enough for periodic extrapolation. The distinction is less sharp than it first appears, since the spectral structure of the training data is itself imposed by the tides. Two observations bear on it. The latitude discrimination of Sect. 4.3 is not a property the training loss rewards, because nothing in the objective asks that pole periods be latitude-invariant. The phase-fidelity improvement of Sect. 4.6 concerns predicted fields rather than internal representations, and the harmonic baseline of Sect. 4.7 shows that periodic extrapolation alone reproduces only a small part of the predictable variance. Both are more readily explained if the basis has organised itself around a genuine periodicity than if it were fitting spectral bulk.

5.2 The reanalysis amplitude deficit

ERA5 retains 15 to 36%of the observed semidiurnal wind amplitude at 80 km. This result stands apart from the model and bears on a wide class of work: models trained on these fields, evaluations that treat them as truth, and studies that infer mesospheric tidal amplitude from them. ERA5 and MERRA-2 are the default sources wherever middle-atmosphere observations are sparse.

The finding extends what is known from temperature. Sakazaki et al. (2012) reported that reanalysis diurnal tidal amplitudes fall below satellite observations in the upper stratosphere and mesosphere, and Sakazaki et al. (2018) found that agreement among reanalyses is good in the lower stratosphere but degrades markedly above the upper stratosphere, where fewer observations are assimilated and the forecast model carries more weight. The present comparison adds a wind-based, ground-based estimate at a single well-defined altitude together with a magnitude. For the semidiurnal component the shortfall is roughly a factor of three to seven, and it is consistent across 23° of latitude.

The physical mechanism is not obscure. The ERA5 sponge layer damps wave amplitudes near the model top, and the 70 and 80km levels used here sit within the region of strongest damping. The MERRA-2 comparison of Sect. 4.5 supports this reading from another direction, since the two reanalyses differ by several times in upper-level tidal energy fraction while agreeing on the periods themselves. Whichever product is closer to the observations, the disagreement between them is itself an indication of how weakly constrained mesospheric tidal amplitude is in reanalysis.

Less obvious is the asymmetry between components. The diurnal deficit varies strongly with latitude, from 30 to 82 %, while the semidiurnal deficit does not, at 15 to 36%. Two mechanisms could produce this. The diurnal tide, with its longer vertical wavelength at low latitudes and its trapped modes at higher latitudes, may be differentially affected by vertical damping. Alternatively the mix of migrating and non-migrating components may differ between harmonics at these three sites in a way the collocated comparison cannot resolve. Distinguishing the two would require either a longitudinally denser radar chain of the kind used by He and Chau (2019) and Wang et al. (2024b), or a specified-dynamics model run. This is the most concrete follow-up the present dataset suggests.

5.3 Scale-selective spatial fidelity

Spatial fidelity is not uniform across wavenumber, and the cross-model comparison of Supplement Sect. S5 shows that the pattern is not peculiar to one architecture. ATMNO retains the strongest low- and mid-wavenumber coherence with ERA5 among the three models compared, at 0.925 and 0.199 zonally against 0.910 and 0.130 for Factformer and 0.911 and 0.184 for MWT. Its high-wavenumber variance and zonal high-wavenumber coherence are more strongly suppressed than those of Factformer. MWT instead over-amplifies high-wavenumber energy at several upper levels, by a level-mean factor of 4.81 zonally. Each model trades scale fidelity differently, and none of the three reproduces the full spectrum. The grid-scale suppression in ATMNO is a property of the model rather than of the target field.

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

5.4 Limitations

Several constraints bound the scope of these results.

The evaluation spans 0–60°N and 65–143°E. This domain was set by the three meteor radars that provide the observational anchor, without which the amplitude result could not be obtained. The harmonics examined are migrating, sun-synchronous components whose periods are fixed by solar forcing rather than by local dynamics, and Sect. 4.3 shows that they vary by only 4 to 11%across 60° of latitude within the domain. The reanalysis amplitude deficit is therefore expected to reflect the upper-level treatment of the reanalyses rather than a property of East Asia. That expectation is an inference, and testing it requires radar coverage in other longitude sectors.

Above roughly 50 km ERA5 is a sponge-damped model product. A model trained to reproduce it learns that product, including its suppressed amplitudes, which is why the radar comparison rather than reanalysis verification terminates the validation chain. Absolute forecast errors against the radars remain large, including the 10 to 11 ms^{-1} speed underestimate reported in Sect. 4.7.

Three stations in a single longitude sector cannot separate migrating from non-migrating tides. The radar amplitudes are local superpositions of both, together with local gravity-wave variance, so the retention ratios of Sect. 4.7 indicate the magnitude of the deficit rather than exact retention coefficients. Whether non-migrating components inflate or deflate the ratios depends on their contribution at these sites, which the present sampling cannot determine. A factor-of-three-to-seven semidiurnal shortfall is nonetheless large enough that plausible aliasing corrections would not remove it. The radar altitude assignment is also nominal, so the comparison with the ERA5 80km level carries a vertical-registration uncertainty. The radar record covers roughly half of one year at each site, which limits what can be said about seasonal or interannual variation in the deficit.

The terdiurnal band in the mesosphere can receive contributions from non-linear wave–wave interaction and from gravity-wave forcing in addition to direct solar forcing. The poles localise a frequency without identifying its source, so the terdiurnal result concerns an 8h spectral band, with the tidal attribution inherited from the literature.

The eleven baselines are general-purpose operators reimplemented at matched capacity rather than tuned middle-atmosphere forecast systems, and no dynamical forecast is included. The comparison of Sect. 4.1 therefore establishes relative architectural merit within a controlled setting. A comparison against a specified-dynamics whole-atmosphere model would situate the result differently and requires simulation infrastructure outside the present scope.

The 48h input window spans several cycles of each harmonic examined, so part of the three-hour skill reflects periodic extrapolation rather than learned dynamics; the harmonic reference of Sect. 4.7 bounds how large that part can be, and the multi-horizon results of Supplement Sect. S1 show how quickly the advantage over persistence erodes with lead.

The terdiurnal advantage is relative. A level-mean complex NRMSE of 0.80 means the field is not skilfully reconstructed at most altitudes.

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

The learned representation shows regime-sensitive pole shifts across four examined winters, as Supplement Sect. S6 documents. The event set is small and spans training, validation and test periods, so it cannot establish a discriminator and is treated as hypothesis-generating.

6 Conclusions

Two questions motivated this study: how much solar-tidal information reanalysis retains in the upper mesosphere, and whether an adaptive temporal representation can make use of what remains.

Against meteor-radar winds at three stations spanning 23° of latitude, ERA5 retains 15 to 36%of the observed semidiurnal and 30 to 82%of the observed diurnal wind amplitude at 80 km. An adaptive Takenaka–Malmquist neural operator of 1.16 million parameters attains a wind-speed RMSE of 2.215 ms^{-1} and a direction MAE of 5.2°, the lowest values among eleven capacity-matched architectures, and improves on three-hour ERA5 persistence by 4.4 to 9.0%in vector-wind error at the three radar sites. Without frequency supervision the learned poles converge on the semidiurnal and terdiurnal solar harmonics, with recovered periods varying by only 4 to 11%across the domain and reappearing after retraining on MERRA-2. Amplitude-weighted phase error is lower than for the two strongest operator baselines at all three harmonics across 17 complete test months, with the largest margin at the terdiurnal period.

These results fit together in a specific way. The latitude stability of the recovered periods is inconsistent with a near-inertial origin and matches what migrating, sun-synchronous tides produce, while ablation shows that the advantage rests on the adaptivity of the pole positions rather than on the rational form alone: a non-adaptive rational basis performs worse than no rational basis, and a uniform Fourier substitution is costliest of the four reduced configurations. Suppressing tide-frequency components at inference time nonetheless degrades skill less than energy-matched non-tidal interventions do, so the poles are best understood as a frequency-auditable coordinate system embedded in a distributed predictor rather than as carriers of individual tides. The improvement is specific to phase, which is what an adaptive frequency basis would be expected to sharpen.

The amplitude result extends to wind, and to a well-defined altitude, what has been reported for reanalysis temperature tides against satellite measurements (Sakazaki et al., 2012, 2018). The frequencies themselves agree with meteor-radar climatologies of the mesospheric tidal field (He and Chau, 2019; Stober et al., 2021; Wang et al., 2024b). The disagreement between ERA5 and MERRA-2 on upper-level tidal energy, which reaches several times at 80 km while the periods agree, indicates how weakly constrained mesospheric tidal amplitude is in reanalysis.

Several caveats bound these conclusions. The domain is regional and set by the radar network, the radar record covers roughly half of one year, and three stations in one longitude sector cannot separate migrating from non-migrating tides, so the retention ratios indicate the magnitude of the deficit rather than exact coefficients. The 8h band may receive non-tidal contributions that the poles cannot distinguish. The 48h input window spans several cycles of each harmonic, so part of the short-lead skill reflects periodic extrapolation. Absolute terdiurnal fidelity remains low at most altitudes. Confirming that the

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

amplitude deficit is a property of the reanalyses rather than of this region requires radar coverage in other longitude sectors, and separating migrating from non-migrating contributions requires a longitudinally denser chain.

The implications reach beyond the architecture. Any assessment of mesospheric tidal amplitude that relies on reanalysis inherits a deficit of this magnitude, and any model trained on these fields inherits it as a ceiling that no architectural improvement can lift. Progress in data-driven prediction above the tropopause therefore depends less on reducing forecast error against reanalysis than on establishing how much of the atmosphere's variability the reanalysis contains, which requires the kind of ground-based comparison presented here. The frequency-auditable representation offers one route to that assessment, since the frequencies a model has organised itself around can be tested directly against observations rather than inferred from its error statistics. The same adaptive rational framework may prove useful in other geophysical settings whose dominant frequencies shift with location, altitude or circulation regime.

Code availability

The ATMNO implementation is not publicly released. The source code, trained checkpoints and baseline implementations are available to the editor and to referees during the review process on request to the corresponding authors, and to other readers from the corresponding authors on reasonable request.

Data availability

ERA5 is available from the Copernicus Climate Data Store at https://cds.climate.copernicus.eu/, and MERRA-2 from the NASA Goddard Earth Sciences Data and Information Services Center at https://disc.gsfc.nasa.gov/. The 2024 Mohe, Beijing-Shisanling and Wuhan middle- and upper-atmospheric wind datasets are available by online order from the Geophysics Center, National Earth System Science Data Center, at http://www.geodata.cn. The processed evaluation summaries underlying every figure and table are available from the corresponding authors on reasonable request.

Author contributions

YW conceived the study, developed the model and software, performed the experiments and data curation, and prepared the original draft. QL reviewed the manuscript. ZL contributed data curation. ZS reviewed and edited the manuscript, supervised the project and acquired funding.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

The 2024 meteor-radar wind data were provided by the Beijing National Observatory of Space Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, through the Geophysics Center, National Earth System Science Data Center. We thank Guozhu Li for contributing the datasets and acknowledge data support from the National Earth System Science Data Center, National Science and Technology Infrastructure of China, at http://www.geodata.cn.

Financial support

This work was supported by the National Natural Science Foundation of China (grant nos.42275060, 42405065, 42474225, and 42305048) and the Independent Innovation Science Fund of the National University of Defense Technology (grant nos. 24-ZZCX-JDZ-45 and 25-ZZCX-BC-10). We are grateful for the strong support from the "Western Light" Cross-Team Project of the Chinese Academy of Sciences, Key Laboratory Cooperative Research Project..

References

Ahamed, M. A. and Cheng, Q.: TimeMachine: a time series is worth 4 Mambas for long-term forecasting, Front. Artif. Intell. Appl., 392, 1688–1695, https://doi.org/10.3233/FAIA240677, 2024.

Allen, A., Markou, S., Tebbutt, W., Requeima, J., Bruinsma, W. P., Andersson, T. R., Herzog, M., Lane, N. D., Chantry, M., Hosking, J. S., and Turner, R. E.: End-to-end data-driven weather prediction, Nature, 641, 1172–1179, https://doi.org/10.1038/s41586-025-08897-0, 2025.

Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander, M. J., Hirota, I., Horinouchi, T., Jones, D. B. A., Kinnersley, J. S., Marquardt, C., Sato, K., and Takahashi, M.: The quasi-biennial oscillation, Rev. Geophys., 39, 179–229, https://doi.org/10.1029/1999RG000073, 2001.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q.: Accurate medium-range global weather forecasting with 3D neural networks, Nature, 619, 533–538, https://doi.org/10.1038/s41586-023-06185-3, 2023.

Bodnar, C., Bruinsma, W. P., Lucic, A., Stanley, M., Allen, A., Brandstetter, J., Garvan, P., Riechert, M., Weyn, J. A., Dong, H., Gupta, J. K., Thambiratnam, K., Archibald, A. T., Wu, C.-C., Heider, E., Welling, M., Turner, R. E., and Perdikaris, P.: A foundation model for the Earth system, Nature, 641, 1180–1187, https://doi.org/10.1038/s41586-025-09005-y, 2025.

Bonev, B., Kurth, T., Hundt, C., Pathak, J., Baust, M., Kashinath, K., and Anandkumar, A.: Spherical Fourier neural operators: learning stable dynamics on the sphere, Proc. Mach. Learn. Res., 202, 2806–2823, https://proceedings.mlr.press/v202/bonev23a.html, 2023.

Cai, X., Zhu, Y., Wang, X., and Yao, Y.: MambaTS: improved selective state-space models for long-term time-series forecasting, arXiv [preprint], https://doi.org/10.48550/arXiv.2405.16440, 26 May 2024.

Candido, S., Singh, A., and Delle Monache, L.: Improving wind forecasts in the lower stratosphere by distilling an analog ensemble into a deep neural network, Geophys. Res. Lett., 47, e2020GL089098, https://doi.org/10.1029/2020GL089098, 2020.

Charlton, A. J. and Polvani, L. M.: A new look at stratospheric sudden warmings. Part I: climatology and modeling benchmarks, J. Climate, 20, 449–469, https://doi.org/10.1175/JCLI3996.1, 2007.

Crameri, F., Shephard, G. E., and Heron, P. J.: The misuse of colour in science communication, Nat. Commun., 11, 5444, https://doi.org/10.1038/s41467-020-19160-7, 2020.

Dao, T. and Gu, A.: Transformers are SSMs: generalized models and efficient algorithms through structured state space duality, Proc. Mach. Learn. Res., 235, 10041–10071, https://proceedings.mlr.press/v235/dao24a.html, 2024.

Fritts, D. C. and Alexander, M. J.: Gravity-wave dynamics and effects in the middle atmosphere, Rev. Geophys., 41, 1003, https://doi.org/10.1029/2001RG000106, 2003.

Garcia, R. R.: On the structure and variability of the migrating diurnal temperature tide observed by SABER, J. Atmos. Sci., 80, 687–704, https://doi.org/10.1175/JAS-D-22-0167.1, 2023.

Gasperini, F., Jones, M., Jr., Harding, B. J., and Immel, T. J.: Direct observational evidence of altered mesosphere–lower-thermosphere mean circulation from a major sudden stratospheric warming, Geophys. Res. Lett., 50, e2022GL102579, https://doi.org/10.1029/2022GL102579, 2023.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J. Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Gonzalo, J., López, D., Domínguez, D., García, A., and Escapa, A.: On the capabilities and limitations of high-altitude pseudo-satellites, Prog. Aerosp. Sci., 98, 37–56, https://doi.org/10.1016/j.paerosci.2018.03.006, 2018.

Gu, A. and Dao, T.: Mamba: linear-time sequence modeling with selective state spaces, Proc. COLM, https://openreview.net/forum?id=tEYskw1VY2, 2024.

Gu, A., Goel, K., and Ré, C.: Efficiently modeling long sequences with structured state spaces, Proc. ICLR, https://openreview.net/forum?id=uYLFoz1vlAC, 2022.

Guibas, J., Mardani, M., Li, Z., Tao, A., Anandkumar, A., and Catanzaro, B.: Adaptive Fourier neural operators: efficient token mixers for transformers, Proc. ICLR, https://doi.org/10.48550/arXiv.2111.13587, 2022.

Guo, M.-H., Lu, C.-Z., Liu, Z.-N., Cheng, M.-M., and Hu, S.-M.: Visual attention network, Comput. Vis. Media, 9, 733–752, https://doi.org/10.1007/s41095-023-0364-2, 2023.

Gupta, G., Xiao, X., and Bogdan, P.: Multiwavelet-based operator learning for differential equations, Adv. Neural Inf. Process. Syst., 34, 24048–24062, https://proceedings.neurips.cc/paper_files/paper/2021/hash/c9e5c2b59d98488fe1070e744041ea0e-Abstract.html, 2021.

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

Hagan, M. E. and Forbes, J. M.: Migrating and nonmigrating diurnal tides in the middle and upper atmosphere excited by tropospheric latent heat release, J. Geophys. Res.-Atmos., 107, 4754, https://doi.org/10.1029/2001JD001236, 2002.

Hagan, M. E. and Forbes, J. M.: Migrating and nonmigrating semidiurnal tides in the upper atmosphere excited by tropospheric latent heat release, J. Geophys. Res.-Space, 108, 1062, https://doi.org/10.1029/2002JA009466, 2003.

He, M. and Chau, J. L.: Mesospheric semidiurnal tides and near-12h 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.

He, Y., Zhu, X., Sheng, Z., and He, M.: Identification of stratospheric disturbance information in China based on the round-trip intelligent sounding system, Atmos. Chem. Phys., 24, 3839–3856, https://doi.org/10.5194/acp-24-3839-2024, 2024.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. Roy. Meteor. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Heuberger, P. S. C., Van den Hof, P. M. J., and Wahlberg, B. (Eds.): Modelling and Identification with Rational Orthogonal Basis Functions, Springer, London, UK, https://doi.org/10.1007/1-84628-178-4, 2005.

Hocking, W. K., Fuller, B., and Vandepeer, B.: Real-time determination of meteor-related parameters utilizing modern digital technology, J. Atmos. Sol.-Terr. Phy., 63, 155–169, https://doi.org/10.1016/S1364-6826(00)00138-3, 2001.

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

Keckhut, P., Lefebvre, T., Hauchecorne, A., Meftah, M., and Khaykin, S.: Detection of migrating and non-migrating atmospheric tides derived from ERA5 temperature meteorological analyses, Atmosphere, 14, 895, https://doi.org/10.3390/atmos14050895, 2023.

Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., Klöwer, M., Lottes, J., Rasp, S., Düben, P., Hatfield, S., Battaglia, P., Sanchez-Gonzalez, A., Willson, M., Brenner, M. P., and Hoyer, S.: Neural general circulation models for weather and climate, Nature, 632, 1060–1066, https://doi.org/10.1038/s41586-024-07744-y, 2024.

Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., and Anandkumar, A.: Neural operator: learning maps between function spaces with applications to PDEs, J. Mach. Learn. Res., 24, 1–97, https://jmlr.org/papers/v24/21-1524.html, 2023.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., and Battaglia, P.: Learning skillful medium-range global weather forecasting, Science, 382, 1416–1421, https://doi.org/10.1126/science.adi2336, 2023.

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A.: Fourier neural operator for parametric partial differential equations, Proc. ICLR, https://openreview.net/forum?id=c8P9NQVtmnO, 2021.

Li, Z., Shu, D., and Barati Farimani, A.: Scalable transformer for PDE surrogate modeling, Adv. Neural Inf. Process. Syst., 36, 28010–28039, https://doi.org/10.52202/075280-1216, 2023.

Limpasuvan, V., Thompson, D. W. J., and Hartmann, D. L.: The life cycle of the Northern-Hemisphere sudden stratospheric warmings, J. Climate, 17, 2584–2596, https://doi.org/10.1175/1520-0442(2004)017<2584:TLCOTN>2.0.CO;2, 2004.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B.: Swin Transformer: hierarchical vision transformer using shifted windows, Proc. IEEE/CVF Int. Conf. Comput. Vis., 9992–10002, https://doi.org/10.1109/ICCV48922.2021.00986, 2021.

Loshchilov, I. and Hutter, F.: Decoupled weight decay regularization, Proc. ICLR, https://openreview.net/forum?id=Bkg6RiCqY7, 2019.

Lu, L., Jin, P., Pang, G., Zhang, Z., and Karniadakis, G. E.: Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators, Nat. Mach. Intell., 3, 218–229, https://doi.org/10.1038/s42256-021-00302-5, 2021.

Malmquist, F.: Sur la détermination d'une classe de fonctions analytiques par leurs valeurs dans un ensemble donné de points, C. R. 6e Congr. Math. Scand., 253–259, 1926.

Mauricio, C., Suclupe, J., Milla, M., López de Castilla, C., Kuyeng, K., Scipion, D., and Rodriguez, R.: Short-term prediction of horizontal winds in the mesosphere and lower thermosphere over coastal Peru using a hybrid model, Front. Astron. Space Sci., 11, 1442315, https://doi.org/10.3389/fspas.2024.1442315, 2024.

Ninness, B. and Gustafsson, F.: A unifying construction of orthonormal bases for system identification, IEEE T. Automat. Contr., 42, 515–521, https://doi.org/10.1109/9.566661, 1997.

Oberheide, J., Forbes, J. M., Häusler, K., Wu, Q., and Bruinsma, S. L.: Tropospheric tides from 80 to 400 km: propagation, interannual variability, and solar-cycle effects, J. Geophys. Res., 114, D00I05, https://doi.org/10.1029/2009JD012388, 2009.

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., Kurth, T., Hall, D., Li, Z., Azizzadenesheli, K., Hassanzadeh, P., Kashinath, K., and Anandkumar, A.: FourCastNet: a global data-driven high-resolution weather model using adaptive Fourier neural operators, arXiv [preprint], https://doi.org/10.48550/arXiv.2202.11214, 22 February 2022.

Price, I., Sanchez-Gonzalez, A., Alet, F., Andersson, T. R., El-Kadi, A., Masters, D., Ewalds, T., Stott, J., Mohamed, S., Battaglia, P., Lam, R., and Willson, M.: Probabilistic weather forecasting with machine learning, Nature, 637, 84–90, https://doi.org/10.1038/s41586-024-08252-9, 2025.

Qi, L., Yang, X., Deng, X., Bai, F., and Ma, Z.: Stratospheric airship flight trajectory envelope prediction based on wind field uncertainty model, Chin. J. Aeronaut., 39, 103956, https://doi.org/10.1016/j.cja.2025.103956, 2026.

Reid, I. M.: Meteor radar for investigation of the MLT region: a review, Atmosphere, 15, 505, https://doi.org/10.3390/atmos15040505, 2024.

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: convolutional networks for biomedical image segmentation, Proc. MICCAI, 234–241, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

Sakazaki, T., Fujiwara, M., Zhang, X., Hagan, M. E., and Forbes, J. M.: Diurnal tides from the troposphere to the lower mesosphere as deduced from TIMED/SABER satellite data and six global reanalysis data sets, J. Geophys. Res.-Atmos., 117, D13108, https://doi.org/10.1029/2011JD017117, 2012.

Sakazaki, T., Fujiwara, M., and Shiotani, M.: Representation of solar tides in the stratosphere and lower mesosphere in state-of-the-art reanalyses and in satellite observations, Atmos. Chem. Phys., 18, 1437–1456, https://doi.org/10.5194/acp-18-1437-2018, 2018.

Scher, S.: Toward data-driven weather and climate forecasting: approximating a simple GCM with deep learning, Geophys. Res. Lett., 45, 12616–12622, https://doi.org/10.1029/2018GL080704, 2018.

Sheng, Z., He, Y., Wang, S., Chang, S., Leng, H., Wang, J., Zhang, J., Wang, Y., Zhang, H., Sui, H., Song, Y., Wu, G., Guo, S., Chai, J., Feng, W., and Song, J.: Dynamics, chemistry, and modeling studies in the aviation and aerospace transition zone, Innovation, 6, 101012, https://doi.org/10.1016/j.xinn.2025.101012, 2025.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-C.: Convolutional LSTM network: a machine learning approach for precipitation nowcasting, Adv. Neural Inf. Process. Syst., 28, 802–810, https://proceedings.neurips.cc/paper/2015/hash/07563a3fe3bbe7e3ba84431ad9d055af-Abstract.html, 2015.

Smith, A. K.: Global dynamics of the mesosphere and lower thermosphere: the roles of variability and forcings, Surv. Geophys., 33, 1177–1230, https://doi.org/10.1007/s10712-012-9196-9, 2012.

Stober, G., Baumgarten, K., McCormack, J. P., Brown, P., and Czarnecki, J.: Comparative study between ground-based observations and NAVGEM-HA analysis data in the mesosphere and lower thermosphere region, Atmos. Chem. Phys., 20, 11979–12010, https://doi.org/10.5194/acp-20-11979-2020, 2020.

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.

Szabó, Z.: Rational Orthonormal Functions and Applications, Ph.D. thesis, Eötvös Loránd University, Budapest, Hungary, 2003.

Takenaka, S.: On the orthogonal functions and a new formula of interpolation, Jpn. J. Math., 2, 129–145, https://doi.org/10.4099/jjm1924.2.0_129, 1925.

Tran, A., Mathews, A., Xie, L., and Ong, C. S.: Factorized Fourier neural operators, Proc. ICLR, https://openreview.net/forum?id=tmIiMPl4IPa, 2023.

Vincent, R. A.: The dynamics of the mesosphere and lower thermosphere: a brief review, Prog. Earth Planet. Sci., 2, 4, https://doi.org/10.1186/s40645-015-0035-8, 2015.

https://doi.org/10.5194/egusphere-2026-4972
Preprint. Discussion started: 25 August 2026

Wahlberg, B.: System identification using Laguerre models, IEEE T. Automat. Contr., 36, 551–562, https://doi.org/10.1109/9.76361, 1991.

Wang, J., Li, N., Yi, W., Xue, X., Reid, I. M., Wu, J., Ye, H., Li, J., Ding, Z., Chen, J., Li, G., Tian, Y., Chang, B., Wu, J., and Zhao, L.: The impact of quasi-biennial oscillation (QBO) disruptions on diurnal tides over the low- and mid-latitude mesosphere and lower thermosphere (MLT) region observed by a meteor radar chain, Atmos. Chem. Phys., 24, 13299–13315, https://doi.org/10.5194/acp-24-13299-2024, 2024b.

Wang, S., Wu, H., Shi, X., Hu, T., Luo, H., Ma, L., Zhang, J. Y., and Zhou, J.: TimeMixer: decomposable multiscale mixing for time-series forecasting, Proc. ICLR, https://openreview.net/forum?id=7oLshfEIC2, 2024a.

Wang, Z., Kong, F., Feng, S., Wang, M., Yang, X., Zhao, H., Wang, D., and Zhang, Y.: Is Mamba effective for time-series forecasting?, Neurocomputing, 619, 129178, https://doi.org/10.1016/j.neucom.2024.129178, 2025a.

Welch, P.: The use of the fast Fourier transform for the estimation of power spectra, IEEE T. Acoust. Speech, 15, 70–73, https://doi.org/10.1109/TAU.1967.1161901, 1967.

Weyn, J. A., Durran, D. R., and Caruana, R.: Can machines learn to predict weather?, J. Adv. Model. Earth Sy., 11, 2680–2693, https://doi.org/10.1029/2019MS001705, 2019.

Wu, H., Hu, T., Luo, H., Wang, J., and Long, M.: Solving high-dimensional PDEs with latent spectral models, Proc. Mach. Learn. Res., 202, 37417–37438, https://proceedings.mlr.press/v202/wu23f.html, 2023.

Wu, H., Luo, H., Wang, H., Wang, J., and Long, M.: Transolver: a fast transformer solver for PDEs on general geometries, Proc. Mach. Learn. Res., 235, 53681–53705, https://proceedings.mlr.press/v235/wu24r.html, 2024a.

Wu, Y., Qian, C., and Huang, H.: Enhanced air quality prediction using a coupled DVMD Informer-CNN-LSTM model optimized with dung beetle algorithm, Entropy, 26, 534, https://doi.org/10.3390/e26070534, 2024b.

Wu, Y., Meng, X., Hu, H., Zhang, J., Dong, Y., and Lu, D.: Affirm: interactive Mamba with adaptive Fourier filters for long-term time-series forecasting, Proc. AAAI Conf. Artif. Intell., 39, 21599–21607, https://doi.org/10.1609/aaai.v39i20.35463, 2025.

Wu, Y., Jin, G., Li, X., Huang, H., Sheng, Z., and Liao, Q.: A novel IDBO-VMD-iTransformer framework for air quality index prediction: multi-strategy optimization and environmental sustainability assessment, Process Saf. Environ., 207, 108379, https://doi.org/10.1016/j.psep.2025.108379, 2026.

Zheng, B., Zhu, M., Guo, X., Ou, J., and Yuan, J.: Path planning of stratospheric airship in dynamic wind field based on deep reinforcement learning, Aerosp. Sci. Technol., 150, 109173, https://doi.org/10.1016/j.ast.2024.109173, 2024.