Supplementary information

Physical fidelity and forecasting in compressed active matter representations

Leonel Guaman[1], Franklin Limaico[1], Hernán Morales-Navarrete[2]*

1 Escuela de Ciencias Físicas y Nanotecnología, Yachay Tech, Urcuquí, 100119, Ecuador

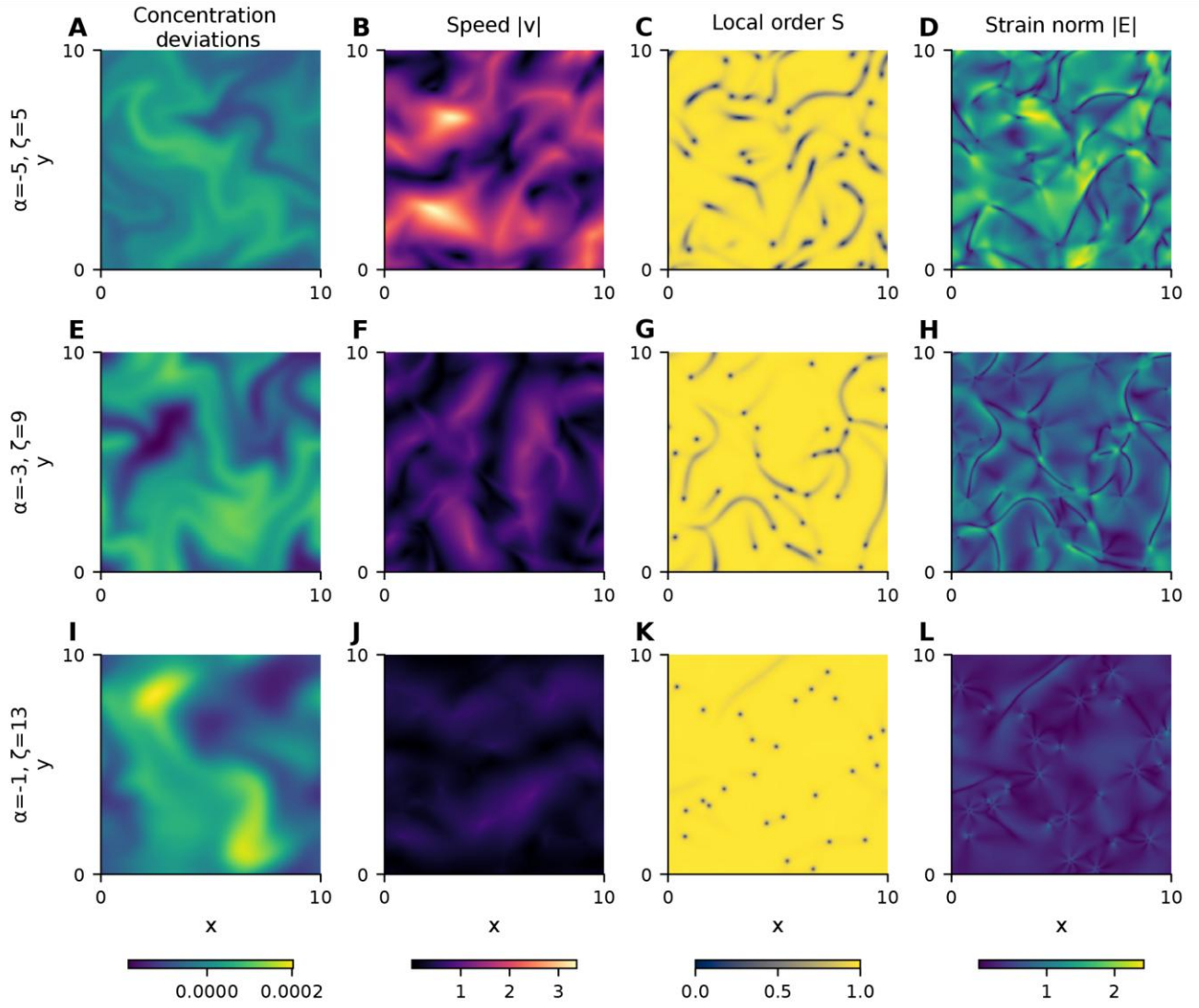
2 Bio-Cheminformatics Research Group and Facultad de Ingeniería y Ciencias Aplicadas, Universidad de Las Américas, Quito, 170504, Ecuador

* Correspondence: hernan.morales@udla.edu.ec

Contents

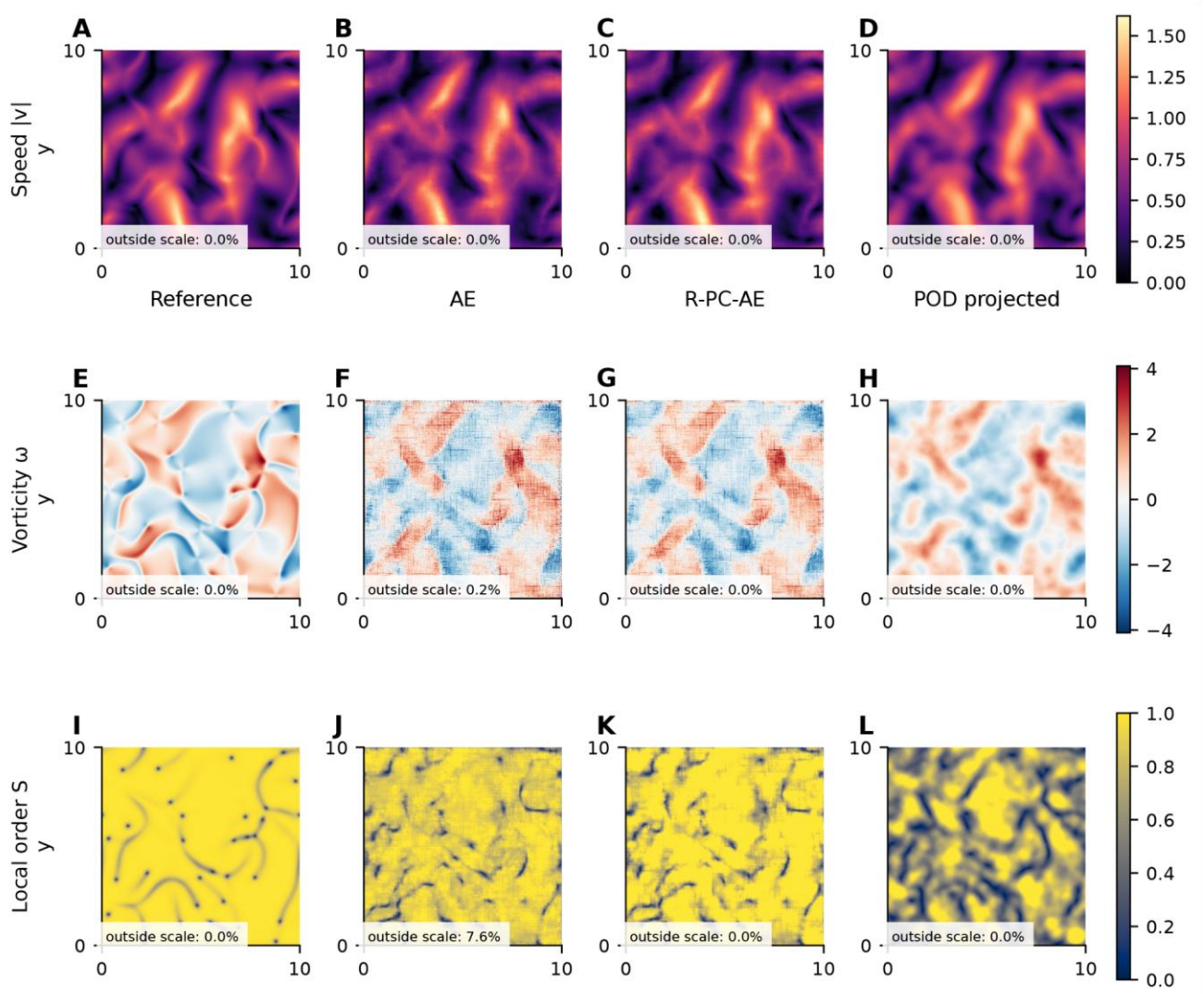
Supplementary Figures 1–10; Supplementary Data 1; Supplementary Methods; Supplementary Videos 1–3.

Supplementary Figure 1



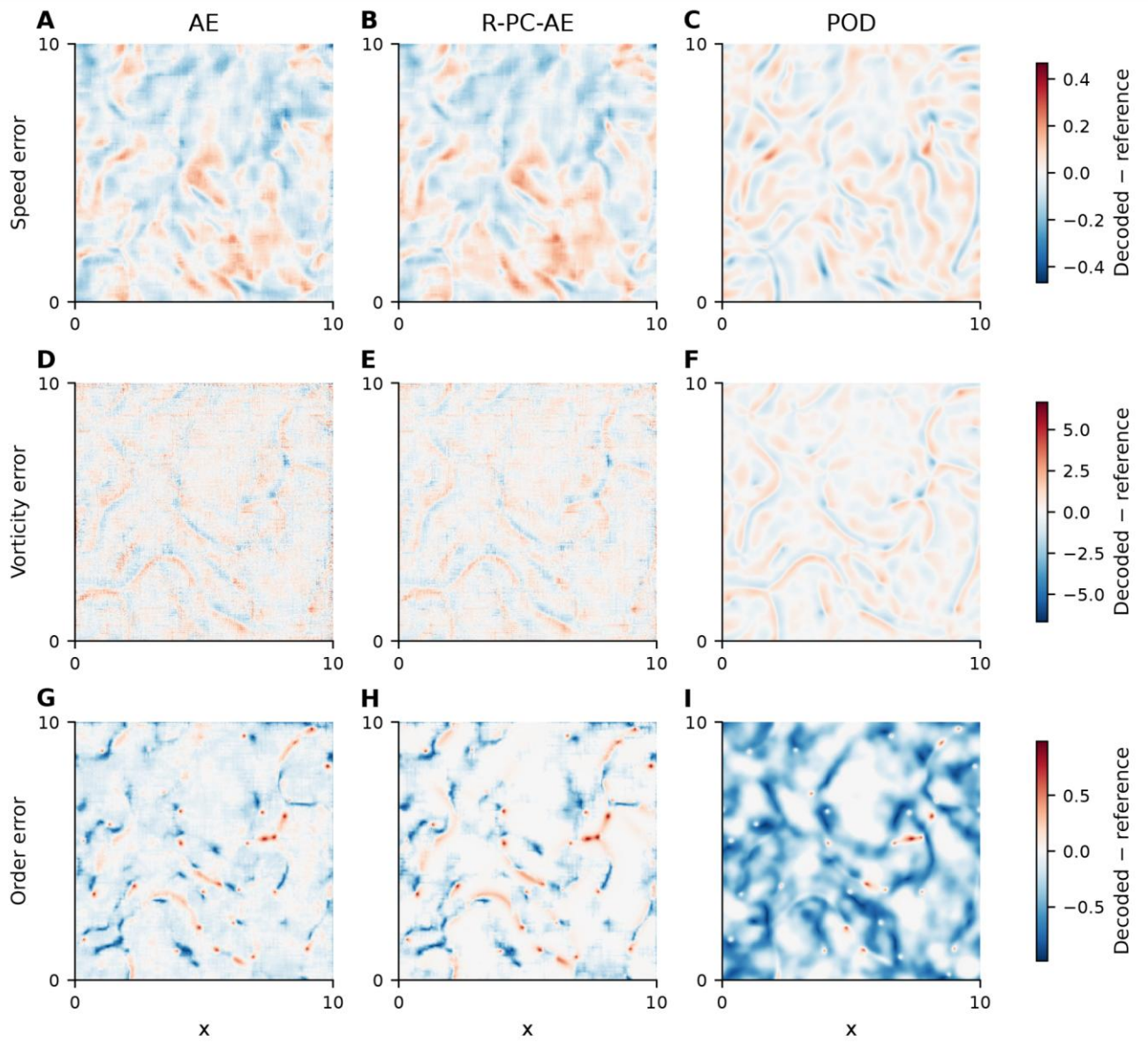
Supplementary Figure 1. Design-selected active-fluid examples. Rows show the minimum, median and maximum sorted parameter tuples at frame 40 ($t = 10$), using the first trajectory ID in each group. Columns show concentration deviations from $c = 1$, speed, local nematic order and the Frobenius norm of the stored strain-rate tensor. Each field shares its color scale across the three examples. Concentration color-bar values indicate $c - 1$, rather than absolute concentration. The examples were not chosen by reconstruction or prediction quality.

Supplementary Figure 2



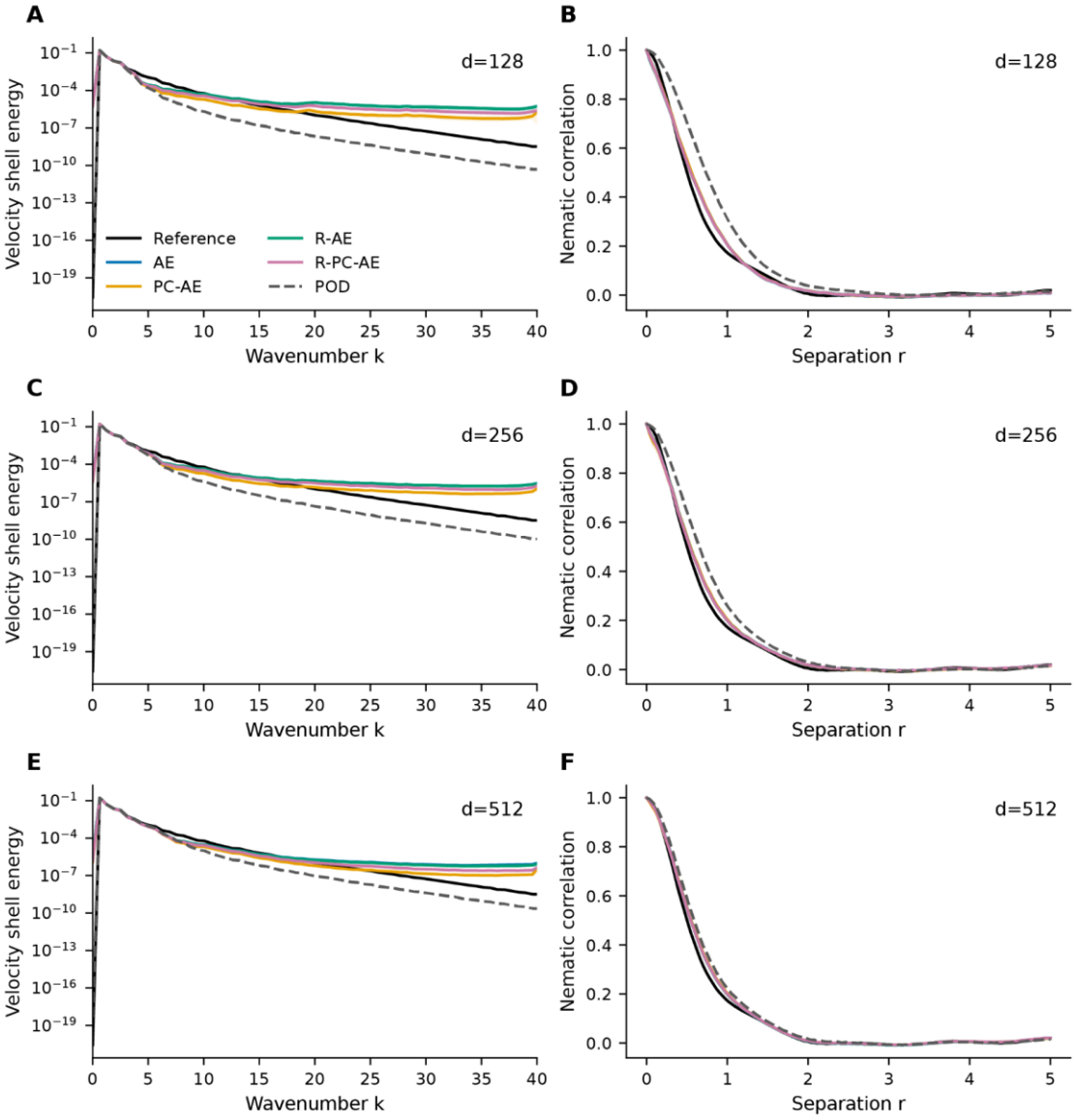
Supplementary Figure 2. Fixed reconstruction examples. Reference fields and AE, R-PC-AE direct corrected and projected POD reconstructions are shown at $d = 256$, using training seed 1729 for the neural models. Rows show speed, vorticity and local nematic order. The median parameter group, first trajectory ID and midpoint frame are chosen by a fixed rule. Color limits are shared within each row. Speed is displayed from zero to the reference maximum, vorticity uses symmetric limits determined by the maximum absolute reference value, and local order is displayed over [0,1]. Predictions outside these limits saturate, and the fraction of pixels outside the displayed range is annotated. Coordinates are physical domain coordinates. Signed errors with full-range common scales are shown in Supplementary Fig. 3.

Supplementary Figure 3



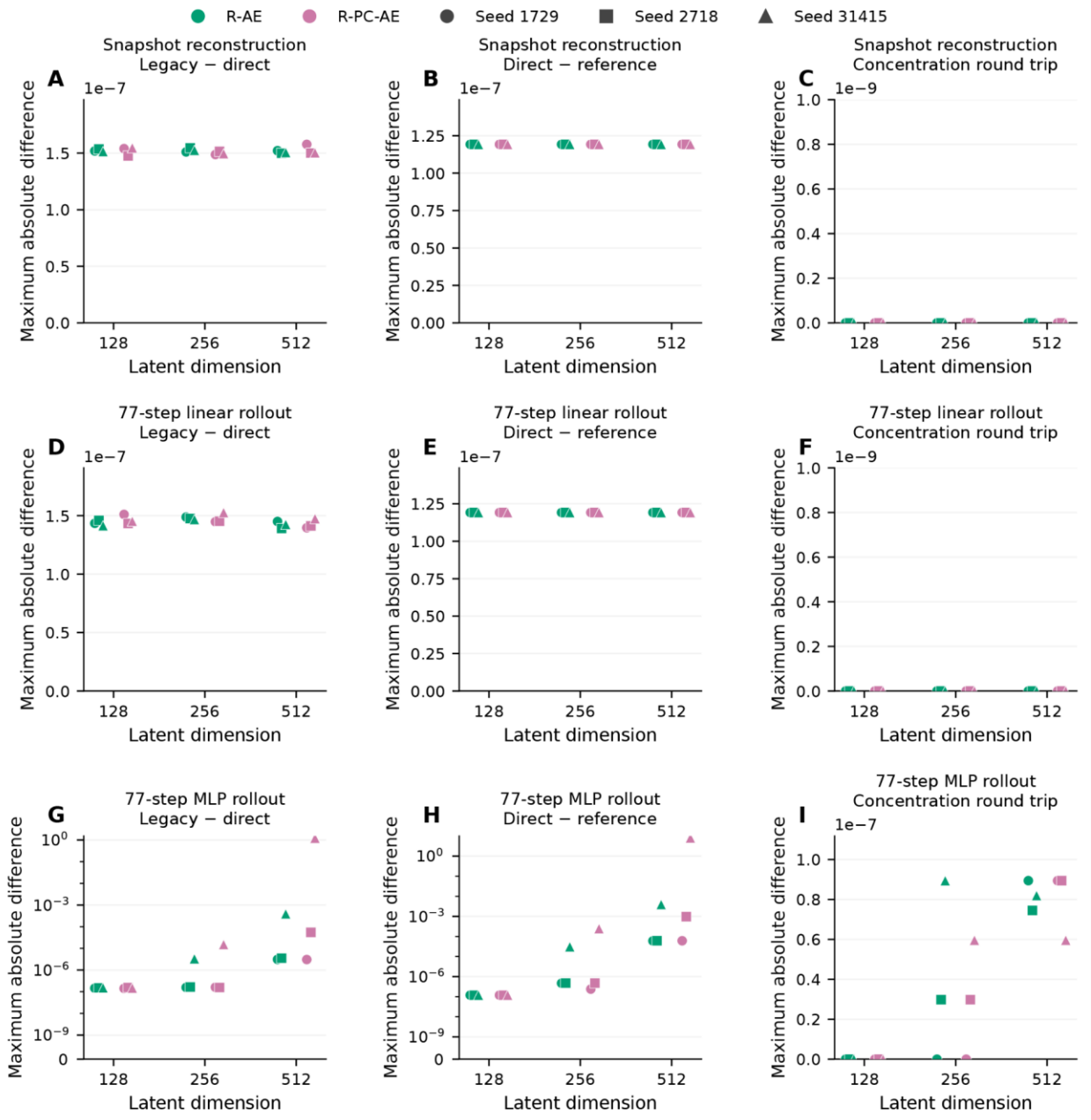
Supplementary Figure 3. Signed reconstruction error maps. Decoded-minus-reference speed, vorticity and local order for the same fixed frame and models as Supplementary Figure 2. Each row shares a symmetric color scale spanning all finite errors across the three reconstructions. Errors are in the corresponding physical field units, rather than the normalized aggregate F or Q scores. These fixed-frame maps show spatial discrepancies and cannot establish statistical superiority.

Supplementary Figure 4



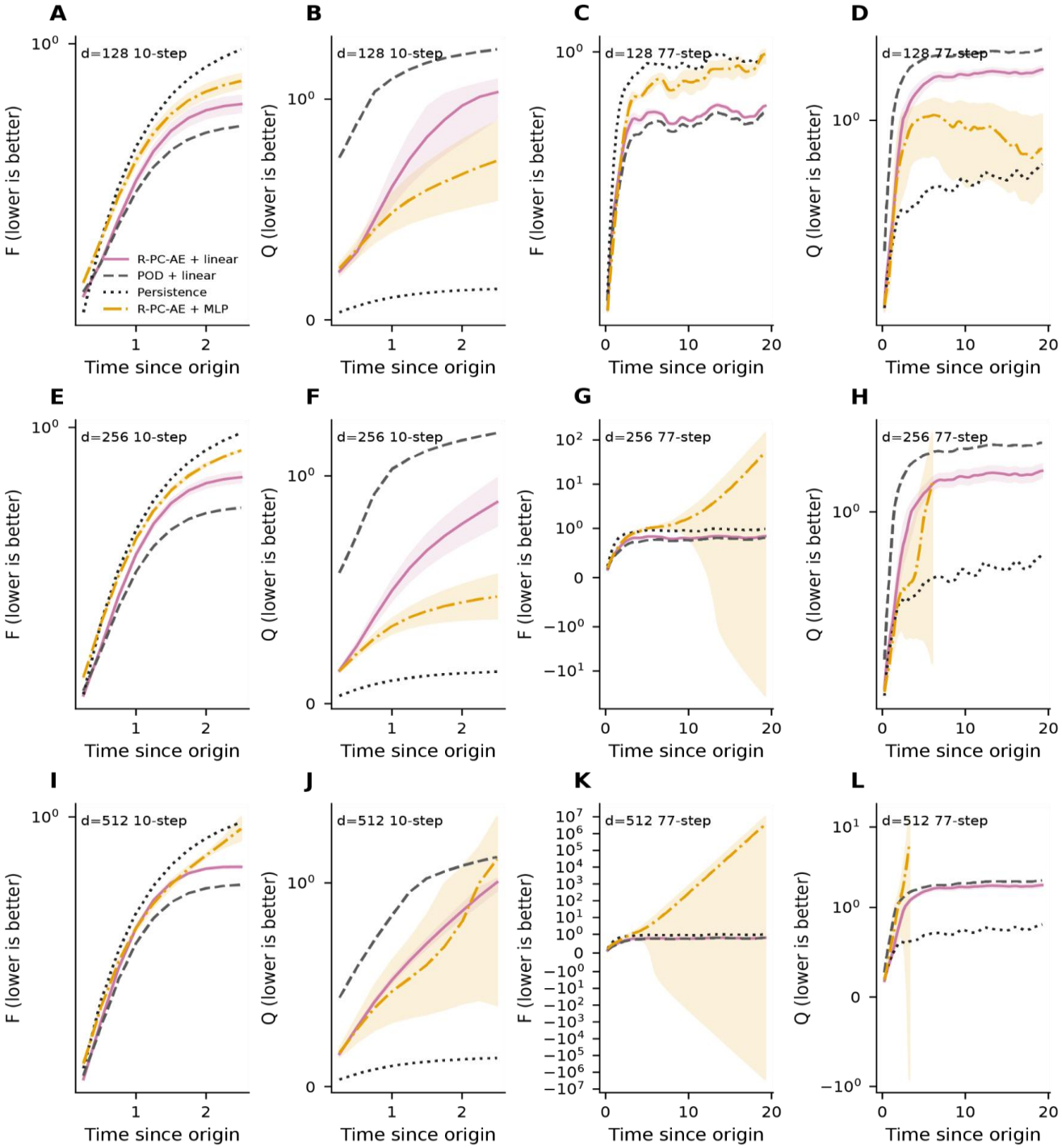
Supplementary Figure 4. Spatial reconstruction diagnostics. Rows show latent dimensions $d = 128,256,512$; columns show velocity shell energy and amplitude-weighted nematic correlation. Curves use the designated primary decoder variants, with a single reference curve in each panel. Neural curves show the mean and shading indicates ±1 sample standard deviation across three fitted seeds; POD is deterministic. The energy display shows k up to 40 for legibility; complete bins remain in the source file. Correlation is unconnected and normalized at zero lag. These diagnostics evaluate reconstructed snapshots rather than spatial forecasts.

Supplementary Figure 5



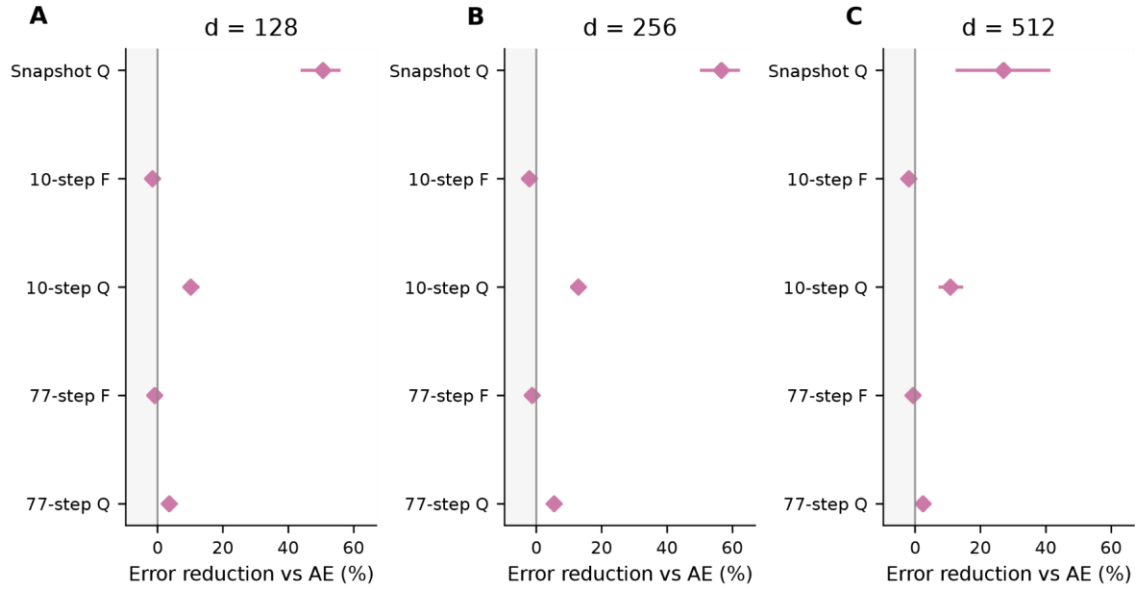
Supplementary Figure 5. Task-specific decoder arithmetic diagnostics. Rows show snapshot reconstruction (**A–C**), 77-step linear rollouts (**D–F**) and 77-step MLP rollouts (**G–I**). Columns show maximum absolute legacy-versus-direct decoding differences, direct-versus-higher-precision-reference differences and concentration changes introduced by the normalization round trip. Colors identify R-AE (green) and R-PC-AE (magenta); circles, squares and triangles identify training seeds 1729, 2718 and 31415, respectively. Each point shows the recorded maximum for one method, latent dimension, seed and task. Horizontal offsets separate methods and seeds and have no numerical meaning. Panels with maxima at or below 10^{-6} use linear axes with scientific notation; panels g and h use symmetric logarithmic axes with a linear region below 10^{-9}. All axes start at zero and show nonnegative absolute differences; ranges differ between panels. Comparisons use identical raw network outputs and do not represent differences between independently evolved latent trajectories. Maxima are arithmetic diagnostics in physical channel units, rather than normalized prediction errors or uncertainty intervals.

Supplementary Figure 6



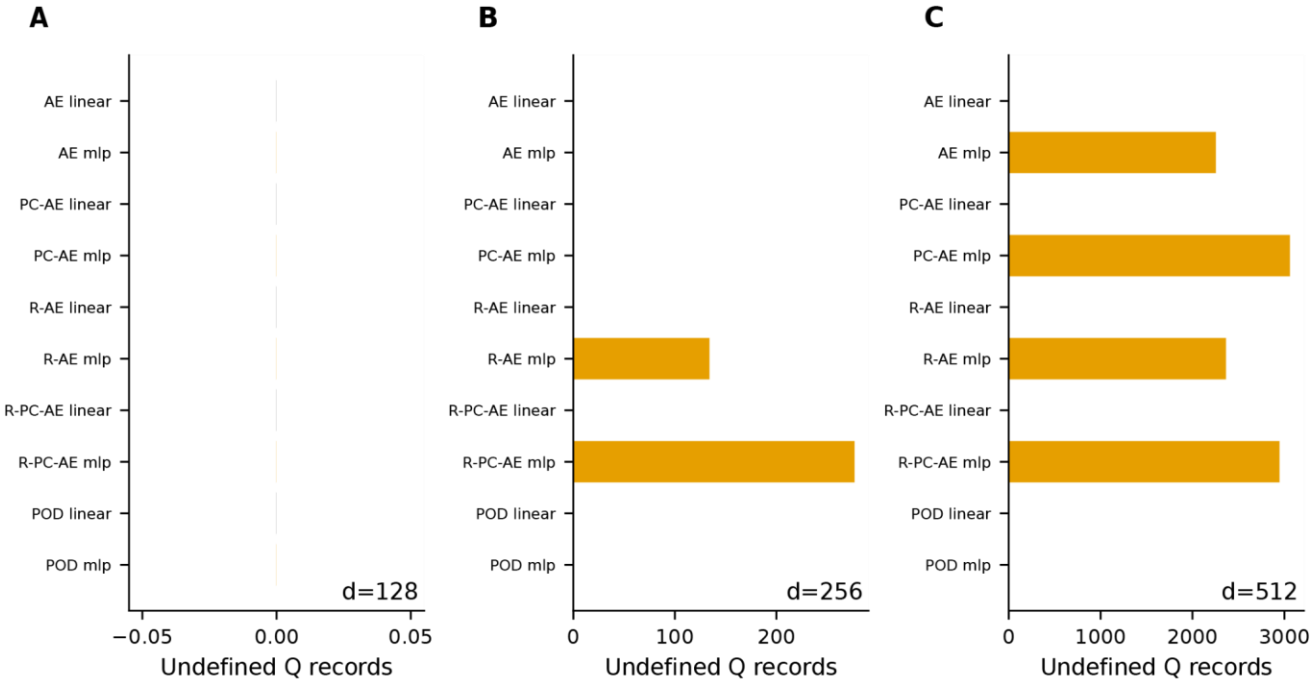
Supplementary Figure 6. Forecast accuracy and long-rollout stability of linear and nonlinear latent dynamics. A–D), E–H), I–L) Results at $d = 128,256,512$, respectively. Columns show ten-step F, ten-step Q, 77-step F and 77-step Q. R-PC-AE with linear autoregression or a residual MLP uses direct corrected decoding and is compared with projected POD/linear and physical persistence. Horizontal axes show elapsed simulation time since the forecast origin. Vertical axes use symmetric logarithmic scales to display large finite errors. Neural curves show means across three fitted seeds; shading indicates ±1 sample standard deviation where all required scores are defined. Symmetric SD bands can extend below zero, although the error metrics are nonnegative. Undefined aggregates remain gaps rather than averages over surviving seeds. Deterministic controls have no seed-variability band, and no calibrated prediction intervals are shown.

Supplementary Figure 7



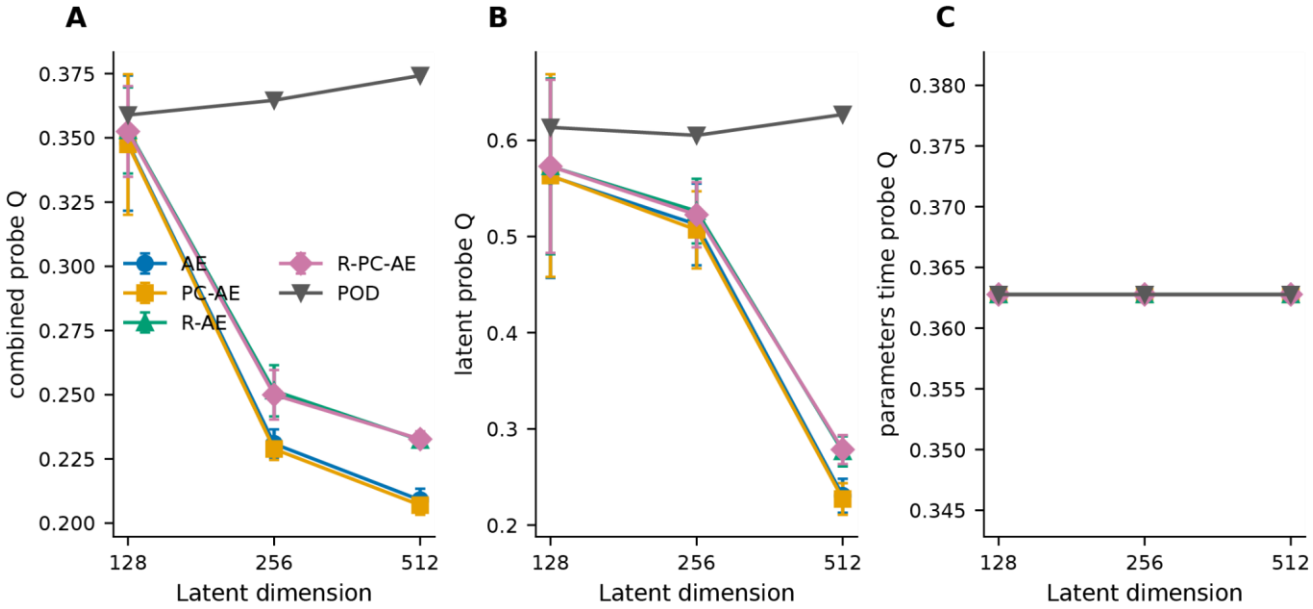
Supplementary Figure 7. Observable reconstruction gains and forecasting gains differ in magnitude. A), B), C) R-PC-AE relative to AE at $d = 128,256,512$, respectively. Each panel compares snapshot reconstruction Q, ten-step forecast F and Q, and 77-step rollout F and Q. Percentage error reduction is $100(1 - \underline{e}_{R-PC-AE}/\underline{e}_{AE})$, where $\underline{e}$ is the overall equal-group mean averaged across three fitted seeds. Forecast comparisons use linear autoregression fitted separately to each representation. Positive values favor R-PC-AE; negative values favor AE. Whiskers are percentile 95% matched parameter-group bootstrap intervals based on 21 groups, 2,000 resamples and bootstrap seed 1729. Intervals are conditional on the fitted models. Axes are shared across latent dimensions. Percentages describe relative changes in distinct error metrics, rather than equivalent physical quantities.

Supplementary Figure 8



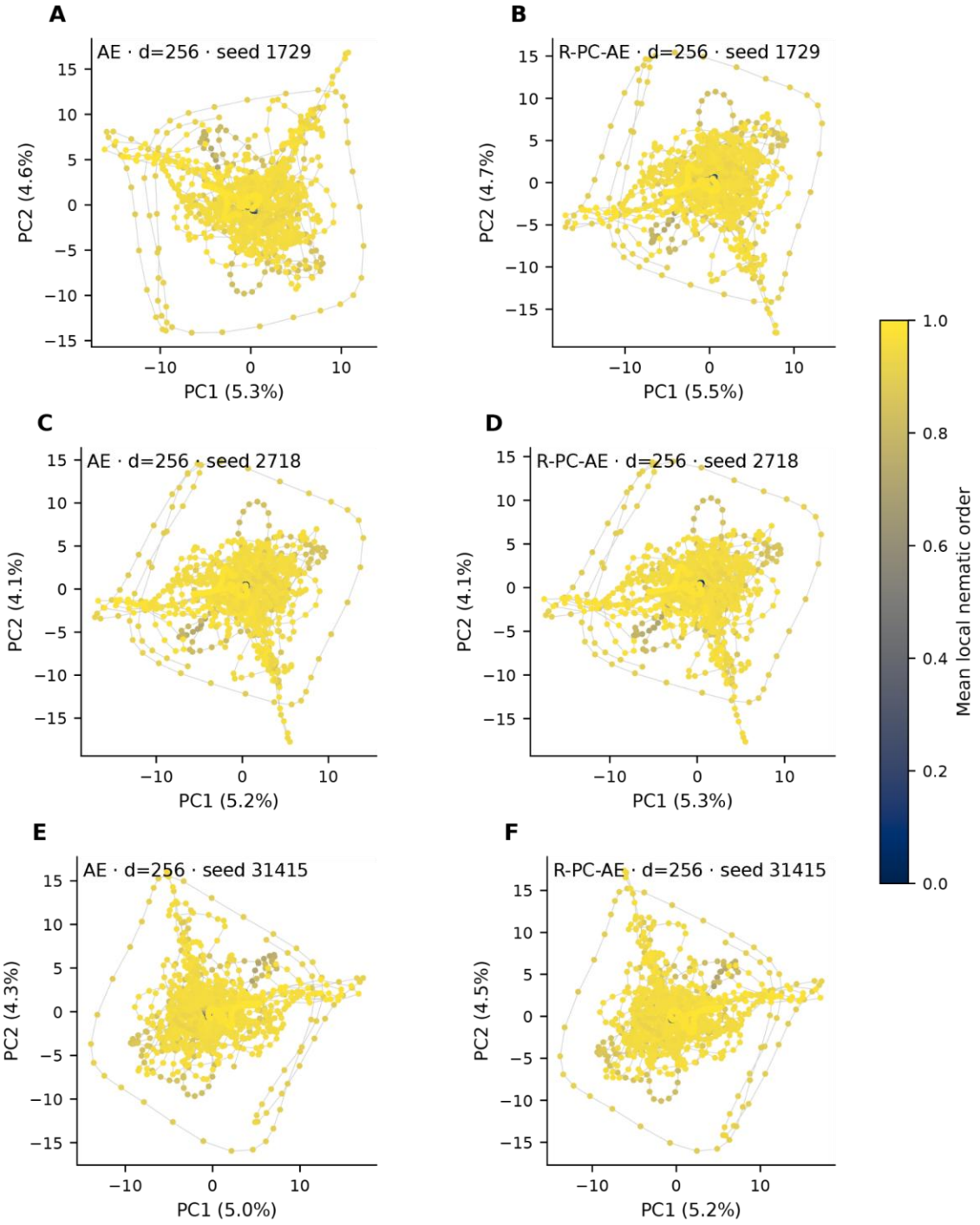
Supplementary Figure 8. Undefined physical-observable scores during long rollouts. A), B), C) Numbers of undefined Q records at $d = 128,256,512$, respectively, summed across three seed-labelled evaluations for each representation and temporal predictor. Each record corresponds to a trajectory and forecast lead in the 77-step rollout. All displayed counts are zero at $d = 128$. Only the designated primary decoder variants are included. Counts describe metric availability rather than independent sample sizes or forecast accuracy. The POD basis is shared across seeds; POD/linear evaluations repeat a deterministic control, whereas POD/MLP transitions are separately fitted.

Supplementary Figure 9



Supplementary Figure 9. Observable estimation from latent variables, control parameters and time. Validation-set probe error (Q) using **A)** latent variables combined with control parameters and time, **B)** latent variables alone, and **C)** control parameters and time alone. Lower values indicate better agreement with the reference observables. Markers and whiskers show the mean ±1 sample standard deviation across three fitted neural seeds; POD is deterministic. The parameter-and-time control is identical across representations and latent dimensions, so its markers overlap in C). Probe coefficients and preprocessing are fitted on training data, with ridge regularization selected on the validation set used for the displayed scores.

Supplementary Figure 10



Supplementary Figure 10. Validation trajectories in principal-component projections of learned representations. A, B) Seed 1729; **C, D)** seed 2718; **E, F)** seed 31415. Left and right columns show AE and R-PC-AE, respectively, at d=256. Each panel contains all 24 validation trajectories with 81 stored frames each (1,944 points). PCA is fitted separately to standardized training latents and applied to the validation latents. Colors indicate the reference spatially averaged local nematic order, and lines connect successive frames within each trajectory. Axis labels give the explained training variance. The projections are descriptive, and their independently fitted axes are not aligned across representations or seeds.

Supplementary data legends

Supplementary Data 1. Complete reconstruction and forecasting summaries. The file Supplementary_Data_1.csv reports results separately by latent dimension ($d = 128,256,512$), training seed, representation, evaluation task, temporal predictor and decoder variant. It includes block-balanced reconstruction error R_{rec}, velocity–orientation field error F, physical-observable error Q, and the individual velocity-intensity, enstrophy and mean-local-order errors. Physical diagnostics include nonpositive and near-floor concentration fractions, negative orientation-moment eigenvalue fractions, local-order exceedance fractions, trace–concentration discrepancy, velocity divergence and strain mismatch. Scores average observations within trajectories, trajectories within parameter groups, and parameter groups equally; fitted seeds remain separate. Primary decoder variants are unprojected AE and PC-AE, direct corrected R-AE and R-PC-AE, and projected POD. Additional variants are retained for comparison. Evaluated record and trajectory counts and metric-specific invalid-record counts accompany the scores. Empty metric entries indicate undefined results, not zero error. Repeated deterministic POD/linear and physical-persistence evaluations under different seed labels are not independent fits.

Supplementary Methods

Fields and physical conventions

The input is a two-dimensional periodic field with channel ordering c, vx, vy, Dxx, Dxy, Dyx, Dyy, Exx, Exy, Eyx and Eyy. D is the concentration-weighted second orientational moment, not the unit-trace orientation tensor. Its realizable conditions are symmetry, positive semidefiniteness and trace(D)=c. Dividing its anisotropy magnitude by positive c gives local scalar nematic order. E is the symmetric velocity gradient, not an independently defined orientation variable.

Physical spacing is domain period divided by grid size, even when stored coordinates include both endpoint labels. Spectral derivatives preserve the even-grid real first derivative by setting the Nyquist multiplier to zero. The stored interval of 0.25 is the separation between data frames, rather than the simulation integrator step. The nominal scalar-count compression ratios refer to the 11-channel stored tensor and do not remove redundant symmetric components.

Error metrics

For each of the c, v, D and E blocks, the mean squared physical-field discrepancy is divided by the square of that block's uncentered training RMS scale. R_rec is the average of these four normalized block mean squared errors. F is one half of the sum of the square roots of the normalized v-block and D-block mean squared errors. Thus F does not aggregate all 11 channels and R_rec is not its square. The short-forecast version averages all leads 1-10 at all complete origins; the long-rollout version averages one 77-lead recursive trajectory.

Q is the mean standardized absolute discrepancy in spatially averaged velocity intensity, enstrophy and local nematic order. Observable standard deviations are training-derived. A near-constant component has an inactive scale in the frozen observable_stats file and is excluded by the evaluator. Velocity intensity and enstrophy discrepancies are separately retained, as is order error, so the composite score is auditable.

Concentration must be strictly above the configured near-floor threshold for Q to be defined. The direct evaluator distinguishes a nonpositive concentration fraction from a near-floor concentration fraction. The negative-D fraction uses minimum eigenvalue below −0.000001; the order exceedance fraction uses S greater than 1.000001. Trace-c discrepancy is an RMS metric. There is no binary trace-violation column in the supplied summary, so a zero trace-violation count is not inferred from it.

Decoder arithmetic and controls

The hard output map begins with a smooth positive-concentration transformation and the raw anisotropy pair. At fixed transformed concentration, the anisotropy is radially limited to a disk of radius c. Symmetric diagonal and off-diagonal D entries are reconstructed from this bounded pair. This enforces the moment cone, but is not a full Euclidean projection jointly optimizing c and D. The v and E channels are unchanged by this map.

The frozen primary realizable variant is direct_corrected. It uses the same raw network output as the historical inference path and applies the map directly with float64 intermediates. The legacy variant includes the historical normalized round trip. AE and PC-AE primary results use their legacy raw decode; POD primary results use projected physical fields. Projection-only AE and PC-AE variants are retained as controls. Persistence has variant legacy because it copies the observed physical field; choosing direct_corrected for persistence would omit the actual records. Supplementary Figure 5 separates all three recorded decoder tasks: reconstructed snapshots, 77-step linear rollouts and 77-step MLP rollouts. Their recorded maxima remain distinct, including large finite changes during unstable MLP rollouts.

Hierarchical aggregation and uncertainty

Within each fitted seed, the evaluator averages observations within each trajectory, trajectories within each alpha-zeta parameter group, and the parameter groups equally. The 26 test trajectories represent 21 parameter

groups. This avoids weighting a group more heavily simply because it contains more initializations. It does not make groups independent random samples from a broader physical population. The fitted seeds 1729, 2718 and 31415 are reported individually.

The unchanged paired table reports comparator-minus-R-PC-AE contrasts. Its percentile 95% intervals average the matched seeds first, then bootstrap parameter groups with replacement for 2,000 draws and random seed 1729. Positive contrasts favor R-PC-AE because scores measure errors. This uncertainty is conditional on the three fits and the represented group design. No frame bootstrap, hypothesis-test P values or significance stars are used.

The publication-only percentage Q reduction is 100 times one minus the ratio of the R-PC-AE and AE group-mean scores. It is a ratio of overall means, not the mean of within-group percentage reductions. The matched bootstrap resamples numerator and denominator together. Comparisons against deterministic persistence use the same matched-group unit. All these additional display-level quantities are recorded separately from the frozen paired table in publication_group_contrasts.csv.

Supplementary Fig. 7 applies the matched ratio-of-means construction to R-PC-AE versus AE for snapshot Q and linear forecast F and Q at both horizons. Each endpoint and dimension is processed separately; scores and fitted models are unchanged. Signed reconstruction error maps (Supplementary Fig. 3) share a full-range symmetric scale across models within each field.

Seed SD bands describe how the three fitted neural models differ. They are not confidence intervals for a population of neural fits, and are not calibrated uncertainty intervals on a predicted field. Deterministic POD with linear dynamics and physical persistence have repeated evaluations under seed labels, not three independent fitted replicas. Their repetitions are not used to manufacture narrow uncertainty intervals.

Failure and coverage accounting

A metric's primary aggregate is undefined if any required record is undefined. The publication layer retains this rule across fitted seeds. It does not discard failed models, nonfinite records or unfavorable finite values. Metric-specific invalid counters, trajectory coverage and first invalid leads accompany the source data. The coverage status tracks undefined observables or overflows and can differ from the counter for a particular metric. For example, Q can be undefined when a field error is still finite.

Supplementary Fig. 6 uses a symmetric logarithmic vertical scale where large finite MLP errors would obscure other curves. Missing segments indicate an undefined aggregate, not a hidden favorable value. Supplementary Fig. 8 reports the Q-specific number of invalid records summed over labelled fitted/evaluated runs; this is an accounting total rather than an independent sample size. Long-rollout and short-forecast tasks remain separate.

Spatial reconstruction diagnostics

Shell energy sums the velocity Fourier energy in radial wavenumber bins, including the mean mode. The sum integrates to one half of mean squared speed. Nematic correlation is an unconnected amplitude-weighted correlation of $q=(D_{xx}-D_{yy},D_{xy}+D_{yx})/c$, normalized by its zero-lag value. It is not a correlation of unit directors. Zero-order or near-floor states can make this diagnostic undefined. The supplied summaries average frames, trajectories and groups before reporting each seed curve.

Publication plots retain each compression dimension and its designated decoder variant. Reference curves use a single logical truth copy rather than pooling duplicated truth records across representation families. These spatial diagnostics evaluate reconstructed snapshots; they do not demonstrate forecast spatial fidelity or a correct governing closure.

Descriptive field views and latent PCA

Field examples use the minimum, median and maximum ordered alpha-zeta parameter groups, the first trajectory ID in each group and the midpoint frame. Reconstruction examples use the median group at seed 1729 and d=256. Shared color limits allow direct visual comparison within each field.

PCA visualizations fit a two-component linear projection on standardized training latents and apply it to all 24 validation trajectories, with 81 stored frames each (1,944 points per panel), at d=256. Figure 5 uses fixed seed 1729; Supplementary Figure 10 retains all three fitted seeds. These are already selected model checkpoints, not a best-ranked or d=512 selection. The supplied export contains d=256 latent arrays only. The complete fitted training mean, scale and PCA components accompany the visual export. Each representation and fitted seed has its own basis; axes are not aligned across models or seeds. Validation trajectories are connected in temporal order. Main Figure 5 colors the fixed-seed projections by normalized stored time; Supplementary Figure 10 colors all seeds by measured mean local order. The first two training components capture only a small fraction of variance. No embedding is refitted for the new color display. Such views can reveal progression or overlap in the displayed subspace, but do not identify latent axes as physical variables, locate a phase boundary or establish a mechanism.

Supplementary video legends

Supplementary Video 1. Reference active-fluid dynamics. Speed, vorticity and local nematic order are shown for all 81 stored frames of the first trajectory in the median ordered parameter group ($\alpha = -3$, $\zeta = 9$). Spatial derivatives are calculated on the original 256×256 grid; fields are displayed at 128×128 resolution. Coordinates span a periodic domain of side length 10 in simulation units. Each field has a separate color bar with limits held fixed throughout the animation. The video shows reference simulation fields rather than model predictions.

Supplementary Video 2. Evolution of aggregate field error during long-term forecasting. The animation progressively reveals the lead-wise field error F for the 77-step rollout at $d = 256$, initialized from the first four observed frames. Curves compare R-PC-AE with direct physical-space decoding and linear autoregression, projected POD with linear autoregression, and physical persistence. Scores are averaged within parameter groups and then equally across groups. The neural curve shows the mean across three fitted seeds, with shading indicating ± 1 sample standard deviation; deterministic controls have no seed-variation band. The horizontal axis gives elapsed simulation time from the last observed frame. The animation displays aggregate errors, not spatial predictions, and the shading does not represent calibrated predictive uncertainty.

Supplementary Video 3. Evolution of aggregate observable error during short-term forecasting. The animation progressively reveals the lead-wise physical-observable error Q for ten-step forecasts at $d = 256$, using all 68 complete forecast origins per test trajectory. Models and seed-variation conventions match Supplementary Video 2. At each lead, errors are averaged over origins within each trajectory, trajectories within each parameter group, and parameter groups with equal weight. The horizontal axis gives elapsed simulation time from the last observed frame of each forecast. Shading describes variation across three fitted neural seeds; overlapping forecast origins are not treated as independent replicates.