Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Physical fidelity and forecasting in compressed active matter representations

**Leonel Guaman**
Yachay Tech
**Franklin Limaico**
Yachay Tech
**Hernán Morales-Navarrete**
hernan.morales@udla.edu.ec

Universidad de Las Américas

**Article**

**Keywords:**

**Posted Date:** November 6th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11611326/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

**Additional Declarations:** No competing interests reported.

# Abstract

Reduced descriptions of active fluids must capture physically meaningful states and support prediction of their evolution, yet the relationship between these requirements remains unclear. Here, we investigate how physical constraints affect reconstruction and forecasting in compressed representations of a simulated active suspension. We compare autoencoders incorporating soft cross-field penalties, hard concentration–orientation moment realizability, or both, with unconstrained autoencoders and weighted proper orthogonal decomposition at matched latent dimensions. Combining the constraints reduces the aggregate standardized reconstruction error across spatially averaged velocity intensity, enstrophy and local nematic order by 51–57% at latent dimensions of 128 and 256, with approximately a 1% increase in balanced field error. The combined model also eliminates detected violations of the enforced admissibility conditions and outperforms post-inference projection in observable reconstruction. These benefits extend selectively to forecasting. With linear latent dynamics, the combined model reduces ten-step observable error by 10–13% relative to the unconstrained autoencoder, while slightly increasing field error. Against persistence, field and observable metrics yield opposite rankings at the reference latent dimension. Linear-readout analyses further distinguish observable accessibility from decoder reconstruction accuracy. Together, these results demonstrate that physical constraints can substantially improve observable preservation, while their predictive benefits depend on the temporal model and quantity evaluated. They motivate assessing physical admissibility, retained information and temporal accuracy as complementary requirements for reduced active-fluid models.


# Introduction

Active matter consists of constituents that consume energy to generate motion or mechanical stresses, sustaining collective behavior away from thermodynamic equilibrium. Examples include reconstituted motor–microtubule assemblies[1], bacterial suspensions[2], cells and tissues[3], collective animal motion[4] and synthetic active colloids[5]. In active suspensions, interactions between particle orientation and the surrounding fluid can produce spontaneous flows, nematic order and complex spatiotemporal patterns[6,7]. Continuum theories describe these phenomena through coupled concentration, velocity and orientational fields. Although such descriptions connect microscopic activity to collective dynamics, their spatially resolved states contain many degrees of freedom, making it challenging to identify compact descriptions that retain the relevant physical information.

Low-dimensional representations offer a way to organize these high-dimensional states and construct tractable models of their evolution[8,9]. By mapping coupled fields to a smaller set of variables, they can facilitate comparisons between states and the assessment of which information is needed to reconstruct physical observables or predict subsequent dynamics. Their scientific value depends on retaining the information required for the intended task, beyond simply reducing the number of variables. For active suspensions, this raises the question of how preserving physical structure in a compressed state relates to its usefulness for temporal prediction.

Machine learning provides several routes to incorporating physical structure into models of active fluids, including inference of hydrodynamic equations from field data[10] and learning closures for kinetic theories[11]. In reduced models, physical structure can enter through constraints on reconstructed fields or through the equations governing the evolution of reduced variables[8,9]. These choices address different requirements. Whereas constraints on reconstructed fields can promote consistency between physical quantities or enforce admissibility of individual states, constraints on reduced dynamics act on how those states evolve. This distinction motivates evaluating physical reconstruction and temporal prediction separately when assessing a learned representation[12].

Here, we investigate how physical constraints shape the information retained in compressed representations of an active fluid using simulations from The Well[11,13]. We compared autoencoders that encourage consistency between coupled fields, enforce physically admissible concentration and orientation moments, or combine both approaches. Comparisons at matched representation sizes, together with a linear dimensional reduction baseline, allow us to assess how these choices affect the reconstruction of physical observables and the prediction of subsequent dynamics.

Combining these constraints improves the reconstruction of selected physical observables at all three latent dimensions examined, with the largest relative gains at the two smaller latent dimensions and a small increase in overall field reconstruction error. The forecasting comparisons reveal a more task-dependent relationship, with benefits that differ between field accuracy and physical observables. Together, these results clarify the capabilities and limitations of physically constrained representations and provide a basis for assessing how admissibility, observable fidelity and predictive performance contribute to useful reduced descriptions of active matter.

## Results

# Active-fluid states and representation comparison

To investigate how physical constraints affect the reconstruction and predictive performance of compressed representations, we used active-fluid simulations from The Well's *active_matter* dataset[13]. These simulations describe rod-like particles coupled through orientational alignment and the flow of an incompressible viscous fluid. The simulated states comprise concentration $c$, velocity $v$, the concentration-weighted second orientational moment $D$, and the strain-rate tensor $E$. Figure 1A illustrates these fields, with additional examples across the parameter range in Supplementary Fig. 1 and the reference trajectory shown in Supplementary Video 1. The official training, validation and test partitions contain 175, 24 and 26 trajectories, respectively. Each trajectory comprises 81 frames on a periodic $256 \times 256$ grid. Retaining all stored vector and tensor components gives 11 channels and 720,896 scalar values per frame. We compared representations with latent dimensions $d = 128,256,512$, corresponding to nominal reductions in scalar count by factors of 5632, 2816 and 1408, respectively. These factors characterize the bottleneck size rather than model memory or compressed file size.

To distinguish the contributions of physical penalties and enforced admissibility, we compared four autoencoder configurations with the same underlying neural architecture at each latent dimension (Fig. 1B,C). The unconstrained autoencoder (AE) minimizes reconstruction error, while the physics-constrained autoencoder (PC-AE) additionally penalizes tensor asymmetry, velocity divergence and inconsistency between the strain-rate tensor and the symmetric velocity gradient. The realizability-constrained autoencoder (R-AE) instead applies a hard decoder map that enforces positive concentration and a symmetric, positive-semidefinite orientation moment satisfying $tr\left(D\right) = c$. The combined model (R-PC-AE) uses this map together with the remaining soft penalties. Weighted proper orthogonal decomposition (POD) provides a linear dimensional reduction baseline with field weighting matched to the neural reconstruction objective. Its primary decoded outputs are projected onto the same admissible concentration–moment conditions. Each neural configuration was fitted using three training seeds. For forecasting, POD with linear autoregression and physical persistence, which repeats the last observed field, provide deterministic comparisons. Table 1 summarizes the study design.

Table 1

**Dataset and design of the representation and forecasting comparisons.** Autoencoder configurations separate unconstrained reconstruction (AE), soft physical penalties (PC-AE), hard realizability constraints (R-AE) and their combination (R-PC-AE). Weighted proper orthogonal decomposition (POD) provides a linear representation baseline. Comparisons are performed separately at each latent dimension and forecast horizon. Neural training variability is described by three fitted seeds; the POD basis with linear autoregression and physical persistence are deterministic controls. Paired bootstrap intervals resample parameter groups after averaging matched seed scores within each group. The 26 test trajectories represent 21 parameter groups; frames and overlapping forecast origins are not treated as independent replicates. These intervals describe comparison uncertainty conditional on the fitted models, rather than uncertainty in individual predicted fields.

| Component | Study design | Evaluation details |
|---|---|---|
| Dataset | The Well *active_matter* | 175 training, 24 validation and 26 test trajectories; 81 frames per trajectory; periodic $256 \times 256$ grid; 11 stored channels |
| Representations | AE, PC-AE, R-AE, R-PC-AE and weighted POD | Latent dimensions $d = 128,256,512$; three training seeds per neural configuration |
| Latent dynamics | Linear autoregression; residual multilayer perceptron (MLP) as a secondary comparison | Four observed states and known control parameters; recursive prediction without teacher forcing |
| Reference forecast | Physical persistence | Repeats the last observed physical field |
| Short forecast | All 68 complete forecast origins per trajectory | Ten predicted frames following each four-frame history |
| Long rollout | One origin per trajectory | First four frames initialize prediction of the remaining 77 frames |
| Training variability | Three fitted neural models per configuration | Individual seed scores and their sample standard deviation |
| Paired comparison intervals | Matched parameter-group bootstrap | 21 test parameter groups; 2,000 resamples; bootstrap seed 1729; conditional on the fitted models |

We assessed these representations separately through reconstruction, physical admissibility, accessibility of physical information and recursive forecasting (Fig. 1D). Balanced reconstruction error, $R_{rec}$, averages the mean squared errors of the concentration, velocity, orientation-moment and strain-rate blocks after scaling by their respective training RMS values. Physical-observable error, $Q$, measures the mean standardized absolute discrepancy in three spatially averaged quantities: velocity intensity, enstrophy and mean local nematic order. Forecast field error, $F$, averages the normalized root mean squared errors of velocity and the orientation moment. The scores therefore assess complementary aspects of accuracy, with lower values indicating closer agreement with the reference simulation. Finally, physical admissibility is evaluated separately, and these measures do not establish satisfaction of the complete governing equations.

## Physical constraints improve observable reconstruction at similar field error

Physical constraints substantially changed the accuracy of reconstructed observables despite producing only small differences in aggregate field reconstruction error. Relative to the unconstrained AE, R-PC-AE reduced physical-observable error $Q$ by 50.6%, 56.6% and 27.1% at latent dimensions $d = 128,256,512$, respectively (Fig. 2A,C; Table 2). These improvements were accompanied by increases of approximately 1.0–1.1% in block-balanced reconstruction error $R_{rec}$ (Fig. 2B). Therefore, similar accuracy under the field reconstruction objective did not imply similar preservation of the selected physical observables.

Table 2

**Physical-observable reconstruction gains and associated field-error costs.** Physical-observable error $Q$ aggregates standardized absolute discrepancies in spatially averaged velocity intensity, enstrophy and local nematic order. Scores average observations within trajectories, trajectories within parameter groups, and groups equally, followed by averaging across three fitted neural seeds. Absolute $Q$ reduction is $Q_{AE} - Q_{R-PC-AE}$; percentage reduction is $100(1 - Q_{R-PC-AE}/Q_{AE})$. Brackets indicate percentile 95% matched parameter-group bootstrap intervals. Matched seed scores are averaged within each group before resampling the 21 groups with replacement 2,000 times using random seed 1729. Percentage intervals resample numerator and denominator together. All intervals are conditional on the fitted models. Relative increase in block-balanced field error is $100(R_{rec,R-PC-AE}/R_{rec,AE} - 1)$, with positive values indicating increased field error. Percentages use unrounded overall means. The test set contains 26 trajectories in 21 parameter groups.

| Latent dimension $\varvec{d}$ | AE $\varvec{Q}$ | R-PC-AE $\varvec{Q}$ | Absolute $\varvec{Q}$ reduction [95% interval] | $\varvec{Q}$ reduction (%) [95% interval] | $\varvec{R}_{\varvec{rec}}$ increase (%) |
|---|---|---|---|---|---|
| 128 | 0.402 | 0.199 | 0.203 [0.159, 0.247] | 50.6 [43.7, 55.9] | 1.14 |
| 256 | 0.251 | 0.109 | 0.142 [0.115, 0.167] | 56.6 [50.0, 62.2] | 1.01 |
| 512 | 0.131 | 0.096 | 0.036 [0.016, 0.055] | 27.1 [12.4, 41.3] | 1.06 |

The matched comparisons supported a positive average benefit at each latent dimension. AE-minus-R-PC-AE differences in $Q$ were 0.203 (95% parameter-group bootstrap interval 0.159–0.247), 0.142 (0.115–0.167) and 0.036 (0.016–0.055) at $d = 128,256,512$, respectively (Fig. 3A; Table 2). The corresponding

percentage reductions and intervals are reported in Table 2. After averaging the three fitted seeds, R-PC-AE had lower $Q$ in all 21 test parameter groups at $d = 128$ and $256$, and in 17 of 21 groups at $d = 512$ (Fig. 2D). The improvement therefore extended across the sampled parameter groups at the two smaller latent dimensions, while becoming less uniform at the largest dimension. Bootstrap intervals describe variation across the represented parameter groups conditional on the fitted models.

Increasing latent dimension reduced $Q$ for both AE and R-PC-AE, but narrowed their absolute difference. The relative benefit of the combined constraints was largest at $d = 256$, rather than increasing monotonically with representation size. The individual constraint configurations also showed different responses. R-PC-AE achieved the lowest mean $Q$ among the compared methods at $d = 128$ and $256$, whereas hard-only R-AE achieved the lowest mean at $d = 512$ (Fig. 2A). Combining soft penalties with hard realizability therefore did not provide a uniform advantage over either constraint strategy alone.

These results demonstrate that aggregate field error and physical-observable accuracy capture distinct properties of a reconstruction. This distinction is physically relevant since enstrophy depends on velocity gradients, while local nematic order depends nonlinearly on the orientation moment and concentration. Consequently, a small change in aggregate field error can accompany a substantial change in observable accuracy. The observed benefit concerns the selected spatially averaged quantities; local discrepancies and spatial organization are assessed separately through reconstruction examples and signed error maps (Supplementary Figs. 2 and 3), together with velocity shell-energy spectra and nematic correlations (Supplementary Fig. 4).

## Hard realizability and soft penalties control different physical properties

The observable improvements above were accompanied by distinct changes in moment admissibility and consistency between fields. The hard decoder map limits the orientation-moment anisotropy relative to positive concentration, ensuring a symmetric, positive-semidefinite moment with trace equal to concentration and local nematic order bounded by one. R-PC-AE had no detected nonpositive concentrations, negative orientation-moment eigenvalues or order values above one in the reconstructed test snapshots across all three latent dimensions and training seeds. At $d = 256$, both realizable neural models eliminated detected negative eigenvalues, whereas the mean negative-eigenvalue fractions were 0.124 for AE and 0.160 for PC-AE (Fig. 3B). Soft penalties alone therefore did not ensure moment admissibility.

Enforcing admissibility after training was also insufficient to recover the observable accuracy of R-PC-AE. Projecting the unconstrained AE output leaves its encoder and latent representation unchanged, providing a comparison between post hoc output correction and training with constraints. R-PC-AE retained lower $Q$ than projected AE at all three latent dimensions (Fig. 3A). At $d = 256$, the matched difference was 0.151, with a 95% parameter-group bootstrap interval of 0.123–0.178. Under the tested projection, repairing admissibility alone did not reproduce the benefit of the combined training approach. R-PC-AE also improved

on the soft-only and hard-only models at $d = 128$ and $256$; at $d = 512$, the interval for its comparison with R-AE included zero.

The cross-field diagnostics revealed a different ordering of the models. At $d = 256$, soft penalties reduced velocity divergence and inconsistency between the stored strain-rate field and the symmetric velocity gradient, while hard-only R-AE retained residuals similar to AE (Fig. 3C,D). PC-AE achieved smaller residuals than R-PC-AE, and POD retained substantially smaller residuals than either neural model. Thus, enforcing admissible orientation moments and improving kinematic consistency address separate properties of the reconstructed state. The combined model improved the selected observables without minimizing every physical residual. Complete diagnostics are provided in Supplementary Data 1.

These comparisons use direct physical-space decoding for the realizable models, avoiding the historical normalization round trip without changing network parameters or latent predictions. Maximum legacy-versus-direct field differences were below $1.6 \times 10^{-7}$ for reconstructed snapshots and the evaluated linear rollouts (Supplementary Fig. 5). Larger arithmetic differences occurred during unstable MLP rollouts. These diagnostics compare decoding implementations for identical raw network outputs; they do not establish the cause of rollout instability or demonstrate that changing the decoding arithmetic restores accurate forecasts.

## Forecast accuracy depends on the quantity being evaluated

Having established that physical constraints improve selected reconstructed observables, we next asked whether these benefits extend to temporal prediction. We fitted latent transition models while keeping the representations fixed, using four observed states and the known control parameters to initialize recursive forecasts. Linear autoregression provided the primary comparison, with a residual multilayer perceptron (MLP) testing a nonlinear alternative. We assessed ten-step forecasts from all 68 complete origins per trajectory and longer rollouts predicting the remaining 77 frames from the first four observations.

At $d = 256$, linear dynamics in the R-PC-AE representation improved velocity–orientation field accuracy relative to physical persistence, but produced larger errors in the selected physical observables (Fig. 4; Table 3). For ten-step forecasting, field error $F$ decreased from 0.749 for persistence to 0.675 for R-PC-AE, whereas observable error $Q$ increased from 0.106 to 0.579. The same ordering held over the 77-step rollout, and matched parameter-group bootstrap intervals supported these opposite directions at both horizons (Fig. 4C,F). POD with linear autoregression achieved still lower field errors, but higher observable errors than R-PC-AE. These comparisons demonstrate that closer agreement between predicted fields does not necessarily imply more accurate evolution of their spatially averaged physical statistics.

Table 3

**Field and physical-observable forecasting errors at latent dimension $\varvec{d} = 256$. $F$ averages training-scale-normalized velocity and orientation-moment RMSE; $Q$ aggregates standardized errors in the three spatially averaged physical observables. Both scores are dimensionless, and lower values indicate better agreement. Ten-step scores average all 68 complete origins per trajectory and ten predicted leads. The 77-step rollout uses the first four observed frames and averages all remaining leads. Scores weight parameter groups equally. R-PC-AE results average three fitted seeds; POD/linear and physical persistence are deterministic controls. R-PC-AE uses direct corrected decoding and POD includes moment projection. Conditional matched parameter-group intervals for comparisons against persistence are shown in Fig. 4C,F.**

| Task | Model | $\varvec{F}$ | $\varvec{Q}$ |
|---|---|---|---|
| 10-step forecast | R-PC-AE + linear | 0.675 | 0.579 |
| 10-step forecast | POD + linear | 0.621 | 1.195 |
| 10-step forecast | Physical persistence | 0.749 | 0.106 |
| 77-step rollout | R-PC-AE + linear | 0.778 | 1.313 |
| 77-step rollout | POD + linear | 0.731 | 1.881 |
| 77-step rollout | Physical persistence | 0.926 | 0.694 |

A matched comparison between AE and R-PC-AE showed that part of the observable reconstruction benefit persisted during forecasting. With linear autoregression fitted separately to each representation, R-PC-AE reduced ten-step $Q$ by 10.2–12.9% across the three latent dimensions, while increasing field error $F$ by 1.5–2.1% (Supplementary Fig. 7). Over 77 steps, the relative $Q$ benefit was smaller, ranging from 2.4% to 5.5%, and field error again favored AE. Physical constraints therefore improved prediction of the selected observables relative to the unconstrained representation, although these gains were smaller than the reconstruction gains and did not surpass persistence on $Q$. Persistence's low observable error alongside larger field error illustrates how retaining global statistics and reproducing local field evolution can yield different model rankings.

The nonlinear transition produced a further dependence on the forecast horizon. For R-PC-AE at $d = 128$ and $256$, the MLP achieved lower ten-step $Q$ than linear autoregression, but higher field error. At $d = 256$ and $512$, its longer rollouts developed very large field errors and undefined observable scores (Supplementary Figs. 6 and 8). Undefined records were retained in the failure accounting, and affected aggregates were left undefined rather than calculated from surviving finite results (Supplementary Data 1). Thus, improved short-horizon observable prediction did not ensure stable long-horizon evolution under the tested nonlinear model. Together, these results show that the predictive value of a representation must be assessed jointly with the transition model, forecast horizon and physical quantity of interest. Supplementary Videos 2 and 3 show the evolution of aggregate forecast errors, rather than spatial predictions.

## Latent states retain physical information beyond a linear parameter-and-time baseline

The forecasting results raise a further question about the information retained by the compressed states. Even when a latent transition predicts poorly, the representation may still contain information about the current physical observables. To examine this possibility without relying on the reconstruction decoder or temporal model, we fitted ridge-regression probes to the frozen representations. We compared predictions based on latent variables alone, control parameters and time alone, and their combination. The combined probe tests whether latent variables improve observable estimation beyond a linear parameter-and-time baseline. Probe coefficients were fitted on training data, with regularization selected on the validation set; the reported validation scores therefore provide a descriptive assessment.

Adding neural latent variables produced larger and more consistent gains at $d = 256$ and 512 than at $d = 128$ (Fig. 5C–E). For R-PC-AE, the reduction in $Q$ relative to the parameter-and-time control was $0.113 \pm 0.010$ at $d = 256$ and $0.130 \pm 0.003$ at $d = 512$, where the spread is one sample standard deviation across three fitted seeds. POD provided no corresponding improvement at these dimensions. However, AE and PC-AE achieved lower combined-probe errors than R-PC-AE (Supplementary Fig. 9). Thus, the advantage of R-PC-AE in decoded observable reconstruction does not imply that those observables are more accurately recovered by a linear readout of its latent state.

Principal component analysis (PCA) provides a complementary view of how validation trajectories occupy the learned representations (Fig. 5A,B). At $d = 256$ and seed 1729, the first two components explain 9.9% of standardized training-latent variance for AE and 10.2% for R-PC-AE. These projections show only a small part of each representation and cannot establish physical coordinates or rank representation quality. Supplementary Fig. 10 shows the corresponding views across all three fitted seeds, colored by the measured mean local nematic order.

Together with the reconstruction and forecasting results, these probes distinguish three aspects of a reduced description. Physical observables can be preserved by the reconstruction decoder, recovered through a separate readout, and predicted through a latent transition, with different representation rankings in each task. The compressed states retain information useful for estimating the instantaneous physical state, while successful reduced dynamics additionally require a transition that propagates the relevant information accurately.

# Discussion

This study shows that incorporating physical structure into compressed representations can improve the reconstruction of selected observables without a comparable improvement in every downstream task. Combining soft cross-field penalties with hard moment realizability reduced observable reconstruction error by approximately 51–57% at latent dimensions of 128 and 256, while increasing balanced field error by approximately 1%. These gains exceeded those obtained by projecting an ordinary autoencoder's outputs onto the admissible set after inference. The benefit therefore extends beyond repairing invalid decoded moments. At the same time, its magnitude and the ranking of constrained models depend on latent

dimension, with hard realizability alone achieving a lower mean observable error than the combined model at $d = 512$. Physical constraints are thus useful design choices whose effects must be assessed for the quantities and representation sizes of interest.

The contrast between observable and field errors helps explain why reconstruction should be evaluated at more than one level. Balanced field error measures average discrepancies in the stored fields, whereas the selected observables depend on nonlinear combinations of those fields and, for enstrophy, spatial derivatives. Similar aggregate field errors can therefore accompany substantially different observable errors. The improvement in $Q$ demonstrates better preservation of these selected quantities, but does not imply uniformly better local structure or agreement at every spatial scale. The spatial diagnostics complement the aggregate scores by exposing discrepancies that a single observable or field metric can conceal. Together, these measurements provide a more informative assessment of a compressed physical state than reconstruction error alone.

Forecasting introduces a further distinction because it requires the representation to support accurate temporal evolution. At $d = 256$, R-PC-AE with linear autoregression improved field agreement relative to persistence but produced larger errors in the spatially averaged observables. Projected POD with linear dynamics achieved still lower field errors, despite its weaker observable reconstruction and larger forecast observable errors. Conversely, R-PC-AE improved observable forecasts relative to AE under the same linear transition family, while slightly worsening field forecasts. These comparisons show that the reconstruction benefit can extend to temporal prediction, but its value depends on the endpoint and comparator. Persistence provides a particularly informative baseline because global statistics can remain similar even as local structures evolve. Its strong $Q$ performance is consistent with this possibility, although the aggregate scores alone do not establish the mechanism.

The linear-readout analysis further distinguishes information retained in a representation from its use by a particular decoder or transition. At the two larger latent dimensions, adding neural latent variables improved observable estimation relative to a linear parameter-and-time control. However, AE and PC-AE achieved lower errors with the combined features than R-PC-AE, despite the latter's stronger decoded observable reconstruction. This difference shows that the accuracy with which observables can be recovered depends on how the latent information is read out. Because regularization was selected on the validation set used to report these scores, the comparison is descriptive. The gains demonstrate complementary predictive value relative to the fitted linear control, without establishing independence from simulation parameters and time. Together, these findings show that compressed states can retain accessible information about instantaneous physical observables even when forecasting performance varies across metrics and temporal models.

These findings are consistent with the different requirements imposed by constraints on individual states and constraints on their evolution. The hard decoder map enforces positive concentration and admissible orientation moments, but does not enforce the governing transport equations or the temporal relationships between successive states. Approaches that learn coordinates together with an evolution law[8], or impose thermodynamic structure on reduced dynamics[9], address additional requirements. Our experiments do not

identify which change would most improve forecasting. Representation content, temporal history, transition expressiveness, conditioning and finite-data estimation may all contribute. In particular, the poor long-rollout behavior of the configured multilayer perceptron does not establish a general limitation of nonlinear temporal models. The supported conclusion is that snapshot constraints alone do not ensure accurate evolution under the transitions examined here.

The controlled comparison developed here provides a foundation for investigating how physical constraints shape reduced descriptions of active fluids. Extending this analysis across architectures, latent dimensions and physical regimes would help identify when improvements in observable reconstruction also support accurate temporal evolution. A particularly relevant direction is to combine admissible decoding with objectives that constrain the evolution of the reduced state, and to assess whether their benefits persist across forecast horizons and observables. Applications to three-dimensional flows and experimental measurements would further test how these relationships depend on spatial complexity and measurement uncertainty. Such extensions would build on the distinction established here between preserving physically meaningful states and learning their dynamics.

Taken together, our results show that physical constraints can substantially improve the reconstruction of selected observables, while their benefits for linear readout and forecasting depend on the task. Evaluating admissibility, observable reconstruction, accessible information and temporal prediction together makes these differences explicit. This approach provides a basis for developing reduced models in which physically meaningful representations and accurate dynamics are assessed as complementary objectives.

## Methods

# Dataset and protocol

We used The Well active_matter simulations of a kinetic active fluid[11,13]. The study's pinned manifest contains 225 trajectories with official train, validation and test membership. Each frame has 11 channels in the order c, $v_x$, $v_y$, $D_{xx}$, $D_{xy}$, $D_{yx}$, $D_{yy}$, $E_{xx}$, Exy, $E_{yx}$, $E_{yy}$. The periodic domain has side length 10; the stored interval is 0.25, distinct from the integration step. Derivatives use physical spacing 10/256 rather than adjacent differences between endpoint-inclusive coordinate labels. Data hashes, split membership, selected checkpoints and analysis inputs are recorded in the frozen provenance. The evaluator fits no encoder, transition, probe or normalization on test data. The publication renderer reads its existing metrics and produces display items. Optional raw-field and latent visualizations are explicitly publication-only descriptive analyses; they do not change the final-test endpoints or model selections.

# Normalization and model fitting

Component means and uncentered root mean square scales were estimated on training data. Predictions are inverse-transformed before physical quantities are evaluated. There is no per-frame normalization. Neural encoders have four circularly padded convolutional stages with widths 32, 64, 128 and 256, ReLU activation and two-by-two pooling, followed by the dense latent bottleneck. The decoder expands that bottleneck through a dense layer and transposed convolutions back to the 11 channels.

Neural branches share a 20-epoch data-only warm-up within each seed. Subsequent branches train for up to 60 epochs. At d = 256, soft-only weights were screened at 0.01, 0.1 and 1; combined realizable weights at 0.001, 0.01 and 0.1. The selected weights were 0.01 and 0.001, respectively, and the d = 128 and d = 512 configurations used those fixed values rather than repeating the weight sweep. Soft-only candidates were eligible when their validation R_rec did not exceed 1.05 times that of AE; combined realizable candidates were assessed against R-AE using the same threshold. Eligible candidates were selected by validation Q, with R_rec and smaller weight as tie-breakers. The d = 256 first-seed pilot selected the physical weights, which were fixed for the remaining seeds and compression budgets. Adam used learning rate 0.0001, weight decay 0.00000001 and gradient clipping at norm one. The resolved configurations and selected checkpoint identities are authoritative.

# Physical constraints and observables

PC-AE penalizes nondimensional squared residuals of D antisymmetry, E antisymmetry, velocity divergence and E minus the symmetric velocity gradient. The realizable branches omit the D symmetry penalty because the hard map already imposes it. The moment map restricts the anisotropy pair (Dxx − Dyy, Dxy + Dyx) radially at fixed predicted positive concentration. It reconstructs symmetric D with trace c; it is not claimed to be the Euclidean projection of the full joint concentration-tensor vector. The direct corrected variant performs this operation in physical space with float64 intermediates and smooth positive-concentration mapping, without a final normalization round trip.

$$c > 0, \quad D = D^{\top}, \quad tr\,(D) = c, \quad \lambda_{min}\,(D) \geq 0, \quad 0 \leq S \leq 1$$

The spatially averaged velocity intensity is one half of mean squared speed. Enstrophy is one half of mean squared vorticity. Local nematic order is the norm of the orientation anisotropy divided by concentration. Q averages absolute discrepancies in these three spatially averaged quantities, standardized by their training-set standard deviations. A component with nonfinite training scale is inactive, following the frozen implementation.

$$K_v = \tfrac{1}{2}\langle v_x^2 + v_y^2 \rangle, \quad \Omega = \tfrac{1}{2}\langle (\partial_x v_y - \partial_y v_x)^2 \rangle$$

$$S = \sqrt{\left[ (D_{xx} - D_{yy})^2 + (D_{xy} + D_{yx})^2 \right]} / c$$

$$R_{rec} = \tfrac{1}{4} \sum_b \langle [(\hat{x}_b - x_b)/s_b]^2 \rangle$$

$$F = \tfrac{1}{2}(\sqrt{MSE_v} + \sqrt{MSE_D})$$

$$Q = mean_j \left| O_j(\hat{x}) - O_j(x) \right| / \sigma_j$$

Here, $b \in \{c, v, D, E\}$, and $s_b$ is the uncentered training RMS shared by all components of block $b$. Block mean squared errors average over spatial positions and stored components. $MSE_v$ and $MSE_D$ denote these errors divided by $s_v^2$ and $s_D^2$, respectively. The observables are $O_1 = K_v$, $O_2 = \Omega$, with training standard deviations $\sigma_j$. All three observables contributed to Q in this study.

Spectral first derivatives use periodic grids and zero the even-grid Nyquist multiplier to preserve real-valued first derivatives. An eigenvalue below − 0.000001 counts as negative; order above 1.000001 counts as exceeding the bound. Q is undefined when predicted concentration is at or near the 0.00000001 floor, or an observable overflows. These tolerance-based checks should not be described as strict floating-point equality tests. Supplementary Methods give the metric and arithmetic conventions in detail.

# Baselines and downstream models

POD uses the same normalized fields with block weighting matched to balanced reconstruction. Its basis is fitted incrementally on training snapshots. Primary POD results include post-decoding moment projection; unprojected and projection-only controls remain in the source tables. A single deterministic POD basis and linear transition are reused across seed labels.

Latent dynamics use four previous states and alpha and zeta. The primary transition is ridge-regularized linear autoregression. The secondary residual MLP has two hidden layers of width 256 and predicts a latent increment. Representation parameters stay frozen during transition fitting; decoded-field supervision can propagate through the fixed decoder to the transition. Forecasts are recursive latent predictions, without teacher forcing or re-encoding predicted fields. Physical persistence repeats the last observed physical field.

Probe feature scalers and regression coefficients were fitted on training data. Ridge regularization was selected using validation error, and the reported probe scores were calculated on that same validation set. These comparisons therefore assess linear accessibility descriptively rather than through an independent test evaluation. The probe endpoint measures linear accessibility of the three observables rather than physical interpretability of individual coordinates. Linear autoregression was fitted to latent increments, with ridge regularization selected by validation ten-step field error. The residual MLP was trained using one-step decoded reconstruction error, and its checkpoint was selected by validation ten-step field error.

# Statistical Analyses

For each metric, observations are averaged within trajectories, trajectories within parameter groups, and groups equally within each fitted seed. The three neural seed values and their sample standard deviation describe fitting variation. They do not represent independent test datasets. The 26 test trajectories cover 21 parameter groups; frames and overlapping origins are not treated as independent replicates. POD reconstruction, POD with linear autoregression and physical persistence are deterministic controls; repeated seed labels do not represent independent fits. MLP transitions are fitted separately for each seed.

That evaluator averages matched seeds within each parameter group, then resamples the groups with replacement for 2,000 bootstrap draws using random seed 1729. Percentile limits at 2.5% and 97.5% give descriptive 95% intervals conditional on the fitted seeds. The publication renderer uses the same resampling unit and settings for percentage reductions and comparisons against persistence that are not present in the original paired table. These derived contrasts are marked publication-only in their source file. No hypothesis-test P values or significance stars are reported; no multiplicity correction or seed-population confidence claim is implied.

If any required contributing observation is undefined, the associated primary mean remains undefined. The renderer propagates those nulls across fitted seeds and never averages only the surviving finite scores. Coverage and metric-specific invalid counters are retained alongside the figures. Finite large errors are not winsorized or suppressed. Curve bands show ± 1 sample standard deviation over the fitted neural seeds at each lead, not bootstrap uncertainty across test groups or calibrated predictive uncertainty.

# Reproducibility

Quantitative Results statements were checked against the frozen output tables during independent manuscript editing; the Python renderer produces display items and numerical source data, without generating manuscript prose or Word documents. The renderer records source hashes, software versions and output checksums, and verifies that input files are unchanged. A separate read-only exporter verifies the existing protocol lock, source checkpoints and data hashes before obtaining qualitative field views and cached latents. PCA is fitted on standardized training latents and applied to validation latents separately for each representation and seed; axes are not aligned across models. The accompanying software (https://github.com/BioDecodingLab/Physically_Constrained_Representations_of_Active_Matter/) includes installation, rendering and test commands. Supplementary Data 1 contains the complete endpoint summaries.

# Declarations

Data availability

The underlying simulation data are publicly available through The Well[13]. Results, trained models and configurations used in this study are archived on Zenodo athttps://zenodo.org/records/19544135.

Code availability

The code for data preparation, representation learning, temporal modeling, evaluation and generation of figures and tables is available at https://github.com/BioDecodingLab/Physically_Constrained_Representations_of_Active_Matter. The code version used in this study is archived together with the trained models and results on Zenodo athttps://zenodo.org/records/19544135.

Author contributions

H.M.-N. conceived the study. H.M.-N., L.G. and F.L. developed and evaluated the computational pipeline. L.G. independently reran the code and verified the consistency of the results. All authors analyzed and interpreted the results. H.M.-N. drafted the manuscript with input from all authors. All authors reviewed, edited, and approved the final manuscript.

Declaration of generative AI and AI-assisted technologies

The authors used ChatGPT (OpenAI) to assist with language editing, rephrasing, and improving the clarity and style of the manuscript, as well as troubleshooting and code refactoring. All AI-assisted text and code changes were reviewed, checked and revised by the authors. Scientific judgment and interpretation remained the responsibility of the authors, who take full responsibility for the accuracy, reproducibility and integrity of the work.

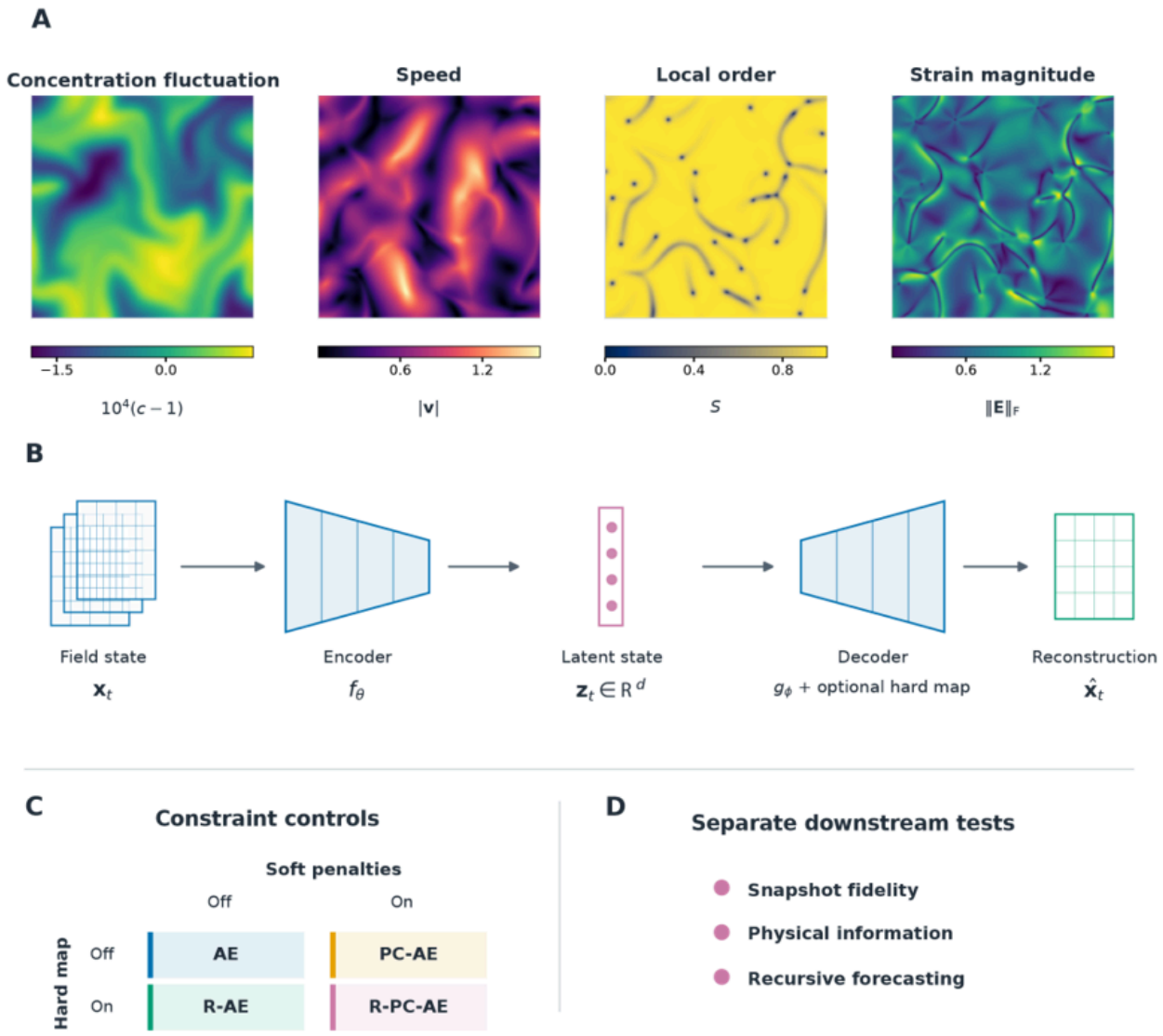
Competing interests

The authors declare no competing interests.

# References


1. Surrey, T., Nédélec, F., Leibler, S. & Karsenti, E. Physical Properties Determining Self-Organization of Motors and Microtubules. *Science* **292**, 1167–1171 (2001).
2. Aranson, I. S. Bacterial active matter. *Rep. Prog. Phys.* **85**, 076601 (2022).
3. Brückner, D. B. & Hannezo, E. Tissue Active Matter: Integrating Mechanics and Signaling into Dynamical Models. *Cold Spring Harb Perspect Biol* **17**, a041653 (2025).
4. Iyer, P., Fedosov, D. A. & Gompper, G. Dynamics of active swarms at the edge of disorder. *Commun Phys* **9**, 37 (2026).
5. Bishop, K. J. M., Biswal, S. L. & Bharti, B. Active Colloids as Models, Materials, and Machines. *Annual Review of Chemical and Biomolecular Engineering* **14**, 1–30 (2023).
6. Marchetti, M. C. *et al.* Hydrodynamics of soft active matter. *Rev. Mod. Phys.* **85**, 1143–1189 (2013).
7. Doostmohammadi, A., Ignés-Mullol, J., Yeomans, J. M. & Sagués, F. Active nematics. *Nat Commun* **9**, 3246 (2018).
8. Champion, K., Lusch, B., Kutz, J. N. & Brunton, S. L. Data-driven discovery of coordinates and governing equations. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 22445–22451 (2019).
9. Park, J. S. R., Cheung, S. W., Choi, Y. & Shin, Y. tLaSDI: Thermodynamics-informed latent space dynamics identification. *Computer Methods in Applied Mechanics and Engineering* **429**, 117144 (2024).
10. Colen, J. *et al.* Machine learning active-nematic hydrodynamics. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2016708118 (2021).
11. Maddu, S., Weady, S. & Shelley, M. J. Learning fast, accurate, and stable closures of a kinetic theory of an active fluid. *Journal of Computational Physics* **504**, 112869 (2024).
12. Qu, H. *et al.* Representation Learning for Spatiotemporal Physical Systems. Preprint at https://doi.org/10.48550/ARXIV.2603.13227 (2026).
13. Ohana, R. *et al.* The Well: a Large-Scale Collection of Diverse Physics Simulations for Machine Learning. Preprint at https://doi.org/10.48550/ARXIV.2412.00568 (2024).


# Figures




## Figure 1

Physical fields and study design: A) Reference simulation fields at α=-3, ζ=9 and t=10 (frame 40), showing concentration deviations from c=1, speed |v|, local nematic order S and the Frobenius norm of the stored strain-rate tensor |(|E|)|_F. Concentration deviations are displayed as 10^4 (c-1) to resolve their small amplitude; negative displayed values indicate c<1, not negative concentration. The snapshot is selected from the first trajectory in the median ordered parameter group. The periodic domain has side length 10, and quantities are expressed in simulation units. B) Autoencoder representation of a field state x_t by a latent z_t ϵ R^d, followed by decoding to a reconstructed state x̂_t, with an optional hard output map. Models are compared at latent dimensions d=128,256,512, using three training seeds and a separate weighted proper orthogonal decomposition (POD) baseline. C) The four autoencoder configurations distinguish unconstrained reconstruction (AE), soft physical penalties (PC-AE), hard realizability constraints (R-AE) and their combination (R-PC-AE). The hard map enforces positive concentration and a symmetric, positive-semidefinite orientation moment with trace equal to concentration. D) Separate evaluations assess snapshot fidelity, physical information accessible through linear probes of frozen representations, and recursive forecasting. Ten-step forecasts used all 68 complete origins per trajectory; 77-step rollouts used

the first four frames to predict the remaining trajectory. Panel A shows simulation data; panels B–D are schematic.

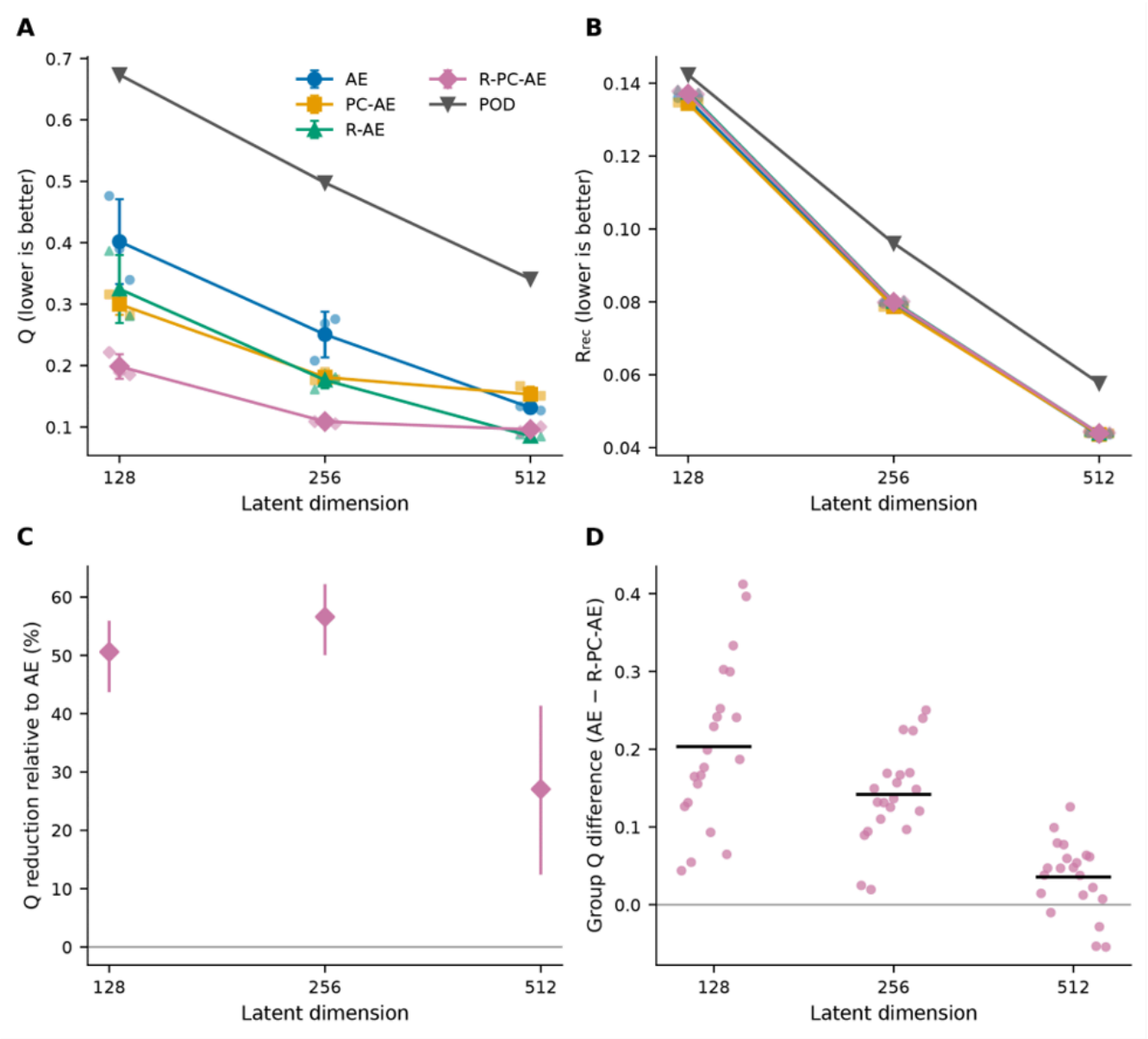



**Figure 2**

Physical-observable and field reconstruction errors across latent dimensions. A) Physical-observable error Q, which aggregates standardized discrepancies in spatially averaged velocity intensity, enstrophy and local nematic order. B) Block-balanced field reconstruction error R_rec. Lower values indicate better agreement in both panels. Small markers show individual training-seed scores; large markers and whiskers show the mean ±1 sample standard deviation across three neural fits. Weighted POD is deterministic and has no seed-variability error bar. C) Percentage reduction in Q for R-PC-AE relative to AE, calculated as 100(1-__Q_(R-PC-AE)/__Q_AE). Whiskers show percentile 95% matched parameter-group bootstrap intervals. D) Parameter-group differences Q_AE-Q_(R-PC-AE), averaged over three fitted seeds. Each point represents one

of 21 test parameter groups; black horizontal lines indicate the equal-group mean. Positive values favor R-PC-AE. Bootstrap intervals use 2000 matched resamples of parameter groups, with random seed 1729, and are conditional on the fitted models. They differ from the seed standard deviations in A and B and do not represent prediction intervals. The 21 groups contain 26 test trajectories. R-AE and R-PC-AE use direct corrected physical-space decoding, POD includes moment projection, and AE and PC-AE use unprojected decoder outputs.

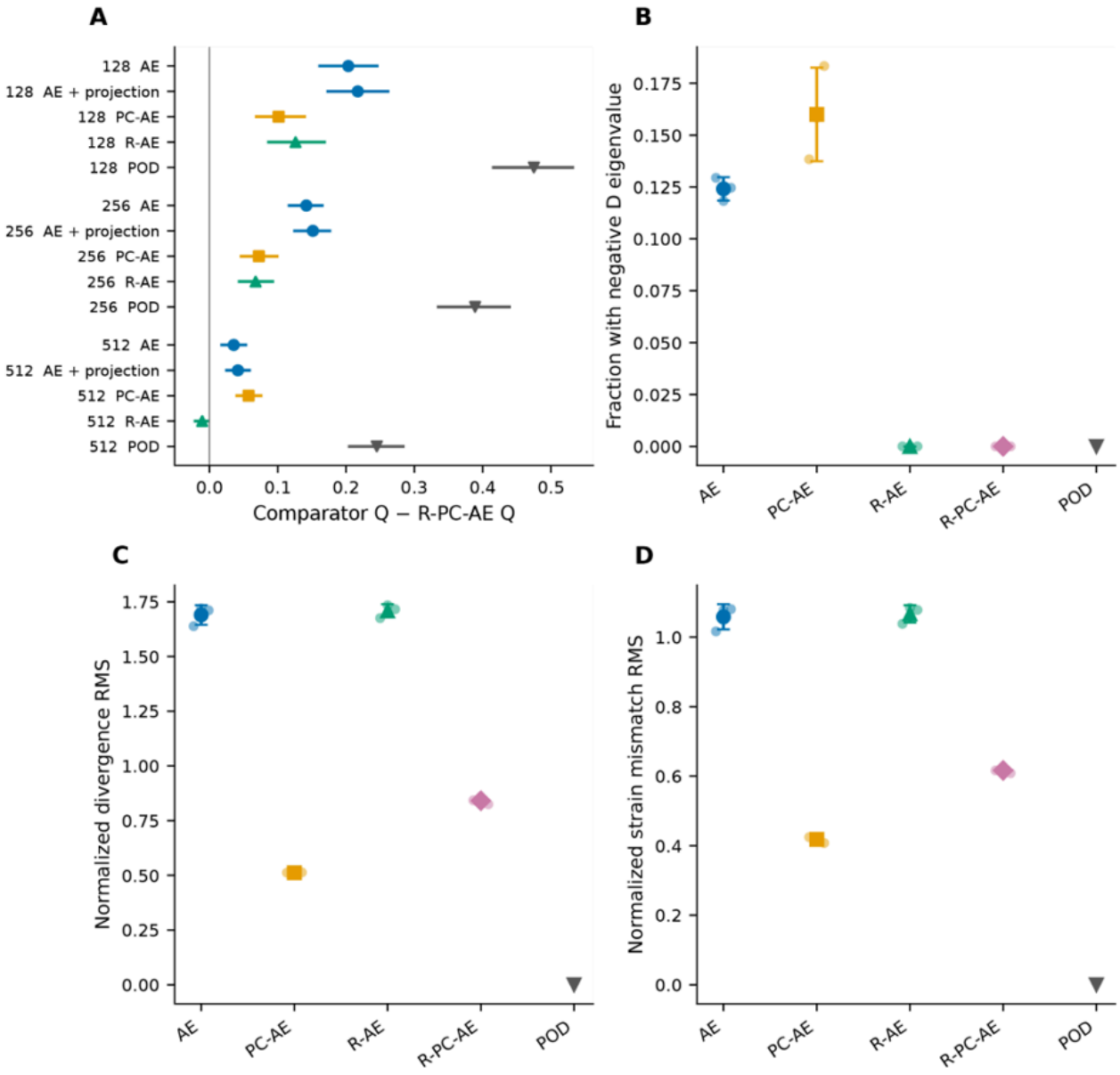



**Figure 3**

Distinct effects of physical constraints on observable accuracy, moment admissibility and kinematic consistency. A) Comparator-minus-R-PC-AE reconstruction Q at each latent dimension. Positive differences favor R-PC-AE. AE plus projection applies moment-admissibility correction after decoding without retraining the representation. Markers show matched mean contrasts, and horizontal intervals are percentile 95%

parameter-group bootstrap intervals based on 21 groups, three fitted neural seeds and 2,000 resamples with bootstrap seed 1729. Intervals are conditional on the fitted models. B–D) Reconstruction diagnostics at d=256, showing the fraction of spatial points with minimum D eigenvalue below -10^(-6) (B), normalized velocity-divergence RMS (C) and normalized strain-mismatch RMS (D). Strain mismatch measures the discrepancy between E and the symmetric velocity gradient. Small markers show individual neural fits; large markers and whiskers show the mean ±1 sample standard deviation across three fits. POD is deterministic and has no seed-variability error bar. R-AE and R-PC-AE use direct corrected physical-space decoding; POD includes moment projection. These diagnostics assess reconstructed snapshots rather than satisfaction of the complete governing equations.

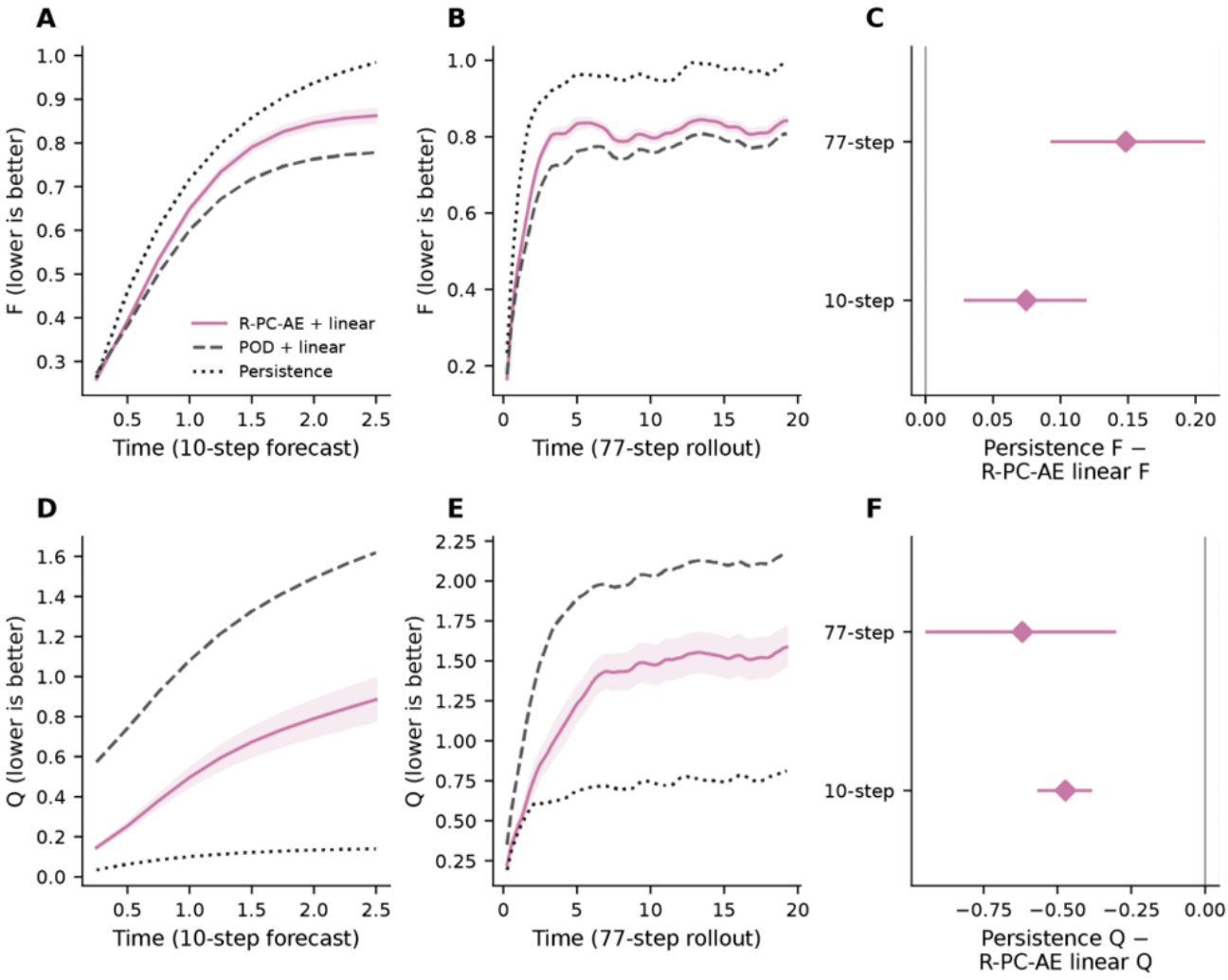



**Figure 4**

Field accuracy and physical-observable accuracy yield different forecast comparisons. Results are shown at latent dimension d=256. A), D) Ten-step field error F and physical-observable error Q, respectively, averaged over all 68 complete forecast origins per trajectory. B), E) Corresponding errors during the 77-step rollout initialized from the first four observed frames. Time denotes elapsed simulation time since the forecast origin. R-PC-AE uses linear autoregression and direct corrected decoding; POD uses linear autoregression

and moment-projected outputs. Physical persistence repeats the last observed field. Neural curves show means across three fitted seeds, with shading indicating ±1 sample standard deviation. POD/linear and persistence are deterministic controls. C), F) Persistence-minus-R-PC-AE contrasts in aggregate F and Q, respectively. Positive values favor R-PC-AE. Markers show mean contrasts; horizontal intervals are percentile 95% matched parameter-group bootstrap intervals based on 21 groups containing 26 trajectories, with 2000 resamples and bootstrap seed 1729. Intervals are conditional on the fitted models. Curve shading describes seed variability, not calibrated uncertainty in predicted fields.

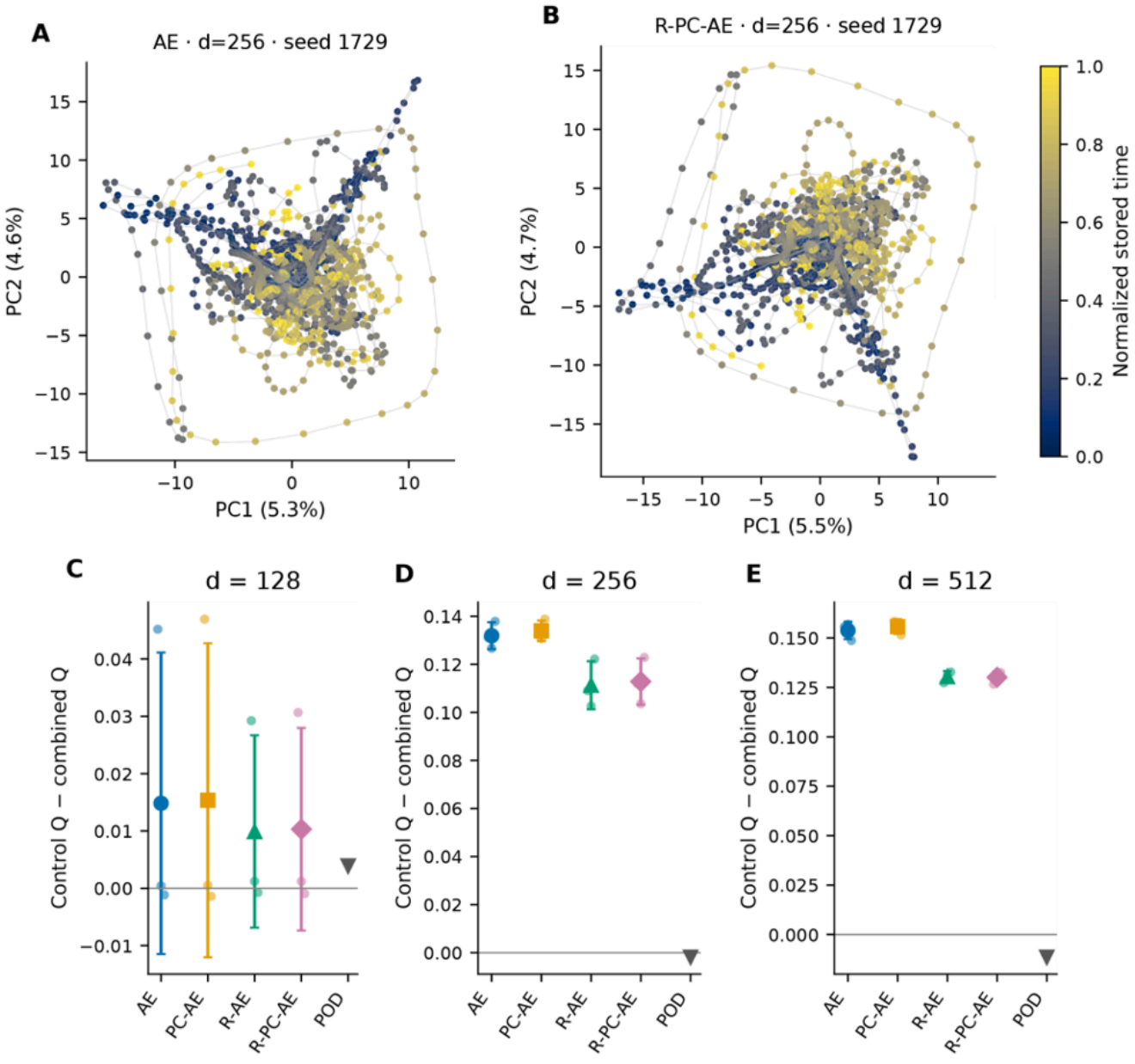



**Figure 5**

Latent trajectories and linear accessibility of physical observables. A, B) Validation trajectories projected onto the first two principal components of standardized training latents for AE and R-PC-AE, respectively, at d=256 and seed 1729. Each panel contains 24 trajectories with 81 stored frames each (1,944 points). Colors

indicate normalized stored time, and lines connect successive frames within each trajectory. PCA is fitted separately for each representation; axis labels give the explained training variance, and axes are not aligned between panels. C–E) Reduction in observable error obtained by adding latent variables to a ridge probe using control parameters and time, at d=128, 256 and 512, respectively. The plotted difference is (Qcontrol-Qcombined), so positive values indicate improvement. Small markers show individual neural fits; large markers and whiskers show the mean ±1 sample standard deviation across three fitted seeds. POD is deterministic. Probe coefficients and preprocessing are fitted on training data; regularization is selected on the validation set used for the displayed scores.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryVideo3aggregateQ.mp4
- SupplementaryVideo2aggregateF.mp4
- SupplementaryVideo1referencefields.mp4
- SupplementaryData1.csv
- ActiveMatterSupplement.pdf