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.

Predictive Structure and Realized Forecast Gain in Nonlinear Fields

Rongzhen Dai

dairongzhen3@gmail.com

Research Article

Keywords: Predictive representation, nonlinear dynamics, readout functions, physical mode decomposition, conditional information, field prediction

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11566735/v1

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

Additional Declarations: No competing interests reported.

Predictive Structure and Realized Forecast Gain in Nonlinear Fields

Abstract

We measure forecast accuracy recovered by changing the readout of an unchanged representation. In a controlled known-mode Brusselator ensemble, replacing affine with quadratic readout lowers mean future-field relative L2 error by 26.83% on 720 new initial states generated after method freezing. The prespecified paired bootstrap interval is positive, conditional on three fixed encoders trained with affine decoding heads. A secondary control retains a 26.69% reduction within the same training-derived four-dimensional output space. This measures a realized deficit of the selected affine reader, rather than the code's optimal decodable risk or total retained information. Physical-component diagnostics locate most of the squared-loss reduction in uniform and remaining-field components. Under this same controlled generation law, a richer known-mode quadratic dictionary has lower mean relative L2 but less accurate critical amplitudes; prior and parameter budgets differ. In a complementary driven lattice, frozen sensor codes and predictors are evaluated on new conditional futures. Conservative lower bounds certify additional information at four ages and fixed-predictor score gains at two, although all four score means are positive. This difference describes certificate resolution at the allocated continuation budget, not an information–utility separation. The contribution is a prospective fixed-code readout intervention, common-output-space control and physical localization of the recovered accuracy.

Keywords: Predictive representation; nonlinear dynamics; readout functions; physical mode decomposition; conditional information; field prediction.

1 Introduction

A low-dimensional representation is evaluated through a prediction map, or readout. We test how changing that map changes future-field accuracy while the encoded state remains fixed. We hold three learned neural encoders fixed and quantify the full-field accuracy recovered by replacing affine readout with quadratic readout. These encoders were themselves trained with affine decoding heads: the intervention tests a remaining limitation of the selected post-training affine reader without changing the learned state coordinates. A new prospective validation freezes the methods before generating initial states and futures, and a common-output-space control tests whether this improvement requires a larger output-space allowance. Known physical amplitudes localize the forecast errors. A complementary lattice study fixes a sensor refinement and its predictors before generating new random futures, then measures additional information and realized score gain separately. The shared question is which predictive distinctions a state description makes operationally usable for a specified future and reader.

A concrete example motivates the distinction. Near a pattern-forming threshold, a compact code can predict the critical amplitude accurately while leaving larger errors in the uniform part of the field. A predictor with a richer physical feature map can then forecast the whole field better, even with a less accurate critical amplitude. Similarly, resolving a coarse sensor bin can retain information about a future local order parameter, while the score improvement of an already trained probability lookup remains statistically unresolved. The two examples call for different tests: physical error localization in the first case, and separate information and score bounds in the second.

Predictive compression has established foundations in past–future information bottlenecks and causal-state approximations [2, 3, 4]. Time-lagged autoencoders compress current observations to predict later observations [1]. Their relationship to slow modes depends on the covariance metric and loss [7]; nonlinear projection and reduced dynamics are also developed research directions [8, 9]. Our neural predictor implements the existing time-lagged regression objective. We use a fixed-code readout intervention to determine which forecast improvement is accessible without changing that encoder, then compare it with explicit physical references and full-field error components.

Related coarse-graining work already selects stochastic mesoscopic models by predictive performance [14], studies predictive information near nonequilibrium criticality [15], and learns predictive partitions or coarse stochastic dynamics [16, 18, 19]. Predictor-constrained information has a broader formal theory [20]. These precedents motivate separating information in an observation from the performance of a particular trained model. Our fixed pair of probability lookups does not optimize a model-class information quantity, and the field regressions do not estimate conditional mutual information.

The main evidence is the forecast improvement recovered from fixed neural codes, its retention in a shared four-dimensional output space, and its localization in physical field components. In the Brusselator, the unique prespecified comparison is supported on newly generated initial states after reader freezing. Physical quadratic products still give lower full-field error but less accurate critical amplitudes than the neural quadratic reader, locating the remaining difference in uniform and remaining-field components. The lattice separately certifies retained information and a fixed predictor's score gain; differing positivity decisions describe the resolution of conservative bounds under the allocated continuation budget. These are complementary measurements of the state-description–reader–future relation, with separate losses and probability statements. Predictive compression and model-dependent information already have established foundations; the added result is this controlled prospective comparison under explicit reader and output-space changes.

Analyses of fixed-encoder decoding distinguish optimal prediction from decoder approximation error under squared reconstruction loss [7]. Our controlled comparison measures a concrete instance in a dynamical field: the accuracy recovered from unchanged codes, its retention in a common output space, and the physical components in which the predictions differ. The main text follows that shared question. Section 2 defines the evaluation objects. Sections 3 and 4 report the physical diagnostic and prospective refinement study. Section 5 relates their evidence and limitations. Supporting horizon, transfer and molecular comparisons, the separate active-sensing experiments, and finite-task numerical proofs remain in the supplements. A measured-signal audit additionally distinguishes separately fitted models from historical-input dependence within one unchanged predictor.

2 Evaluation objects and resource accounting

Let s_t be a physical state, o_t an observation and $z = e(o_t) \in \mathbb{R}^d$ a representation. A fixed reader g predicts a declared target $Y_\tau = H(s_{t+\tau})$ at horizon τ. Its risk under an evaluation law P is

$$R_P(e, g; \ell) = \mathbb{E}_P[\ell(g(e(o_t)), Y_\tau)]. \tag{1}$$

The loss ℓ, target map H, normalization, reader and evaluation law are part of the question. A lower loss for one component does not order losses for the full field. A changed reader can expose structure inaccessible to the previous reader without adding an independent observation.

For the deterministic field forecasts we report the saved mean per-state relative L2 error

$$\overline{E} = \frac{1}{n} \sum_{i=1}^{n} \frac{\|\widehat{Y}_i - Y_i\|_2}{\|Y_i\|_2 + 10^{-10}}, \tag{2}$$

where Y_i is the 128-component future deviation field. Coordinate R^2 uses its corresponding test variance denominator. These are forecast diagnostics rather than information estimates. For the stochastic categorical task we instead use a conditional information target and the log score of fixed probability forecasts. Natural logarithms give units of nats.

Resource dimensions are kept separate throughout. A four-component real code is not a four-bit message. Six physical amplitudes and their 21 quadratic products have six base measured coordinates but 27 feature columns. An affine field reader has an output-rank allowance determined by those columns. A refinement of a finite sensor code increases resolution and message cost even when it uses the same sensors. These differences matter before assigning a forecasting gain to representation quality [1, 4, 20].

	Physical-field diagnostic	Prospective refinement
State system	Brusselator, 64 nodes/two species	Driven 64×64 two-field lattice
Observation	Full present deviation field	Four local observables, coarse or refined bins
Target	Full future field, primarily $\tau = 10$	Six-class local terminal mean, lag 16
Comparison	Affine/quadratic readers of fixed neural or physical coordinates	Nested fixed codes $E, Z4$ and frozen lookups
Evidence unit	720 prospective initial states; original test and 360 external cached states	Fixed bank; new conditional future streams
Main evidence	Prospective fixed-code reader control, shared output space, physical errors	Preallocated conditional information/score bounds
Resource difference	Prior, feature/output rank and fitted coefficient count	Additional within-bin sensor resolution

Table 1: Two evaluations of predictive state descriptions. They have separate losses and evaluation laws; no combined effect size or cross-system confirmation test is formed.

2.1 AI-assisted preparation and verification

Retained field-study records identify OpenAI ChatGPT 6.1, Codex (gpt-6.1-sol), Anthropic Claude 5.5, Consensus Deep, and PaperQA using the DeepSeek official API (historical deepseek-chat and later deepseek-flash). Their uses included formulation, code drafting and inspection, numerical cross-checks, literature discovery, citation screening, external critique and writing. The lattice study records ChatGPT-assisted planning, code inspection, numerical checking and manuscript preparation, without a complete retained model-version history. The present integration uses Codex to organize and edit existing evidence. Original sources and calculations remain auditable; AI tools are not authors, and the human author retains responsibility for the final scientific claims.

3 Physical structure in a same-field forecasting comparison

3.1 Controlled reaction–diffusion ensemble

The Brusselator equations on $x \in [0, 2\pi]$, with no-flux boundaries, are [6]

$$\begin{aligned} \partial_t X &= \partial_{xx} X + 1 - (B+1)X + X^2 Y, \\ \partial_t Y &= 16\partial_{xx} Y + BX - X^2 Y. \end{aligned} \tag{3}$$

We use 64 physical nodes and deviations $u = X - 1$, $v = Y - B$. The critical parameter is $B_c = 25/16$. Its first spatial mode is $\cos(x/2)$, with critical species direction $(1, -5/16)^T$ and fast direction $(1, -5)^T$. Under $(u, v)^T = a(1, -5/16)^T \cos(x/2) + O(a^2)$, the related physical reference gives $\dot{a} = (64/75)(B - B_c)a - a^3/5 + \cdots$ [10]. Supplement A1 derives the linear

modal normalization directly from the species eigenvectors and the DCT-I convention. The near-threshold amplitude equation supplies physical context; its cubic coefficient is not used to compute the field predictions, component projections or readout gains.

The generator deliberately uses those first-mode directions. Before a common positivity-preserving shrinkage, critical and fast amplitudes have standard deviations 0.018 and 0.055, the two uniform species perturbations each have standard deviation 0.035, and independent species coefficients for spatial modes $2, \ldots, 10$ have standard deviation $0.035/\sqrt{n}$. Each regime has 1,800 independent initial states split into 1,080 training, 360 validation and 360 test states. Three parameter values, $1.5, 25/16, 1.625$, are used for the representation comparison; the added same-field modal diagnostic is at B_c only.

The corrected numerical pipeline applies exact spectral-diffusion half steps around a Heun reaction step, with $\Delta t = 0.0025$, using the physical-node DCT-I convention. It approximates a specified discretized flow. The complete generator, update and available step/reference checks are in Supplement A1, A14 and A19. The initial-state construction makes this a known-structure benchmark; successful recovery does not establish discovery of modes in an arbitrary unknown system.

3.2 Predictors and physical feature maps

The predictive autoencoder (PRED) follows the time-lagged objective [1]: its present-field encoder predicts a scaled future field. The corresponding reconstruction autoencoder (RECON) predicts the present. Encoders have widths $128 \to 128 \to 64 \to d$, SiLU hidden activations and a linear training decoder. The retained runs use three model seeds, at most 50 epochs, batch size 128, learning rate 0.001, and validation patience 8. Evaluation freezes the encoder and fits an affine ridge reader on training states to the raw future field. Later same-recipe recovery models supply the saved unquantized critical-regime forecasts; their input-to-forecast replay agrees at the reported precision with the historical table. They are not recovered historical weights or new-data replications (Supplement A17).

The linear comparator, rank-truncated ridge predictive regression (RTRP), first fits the scaled training future, truncates the fitted-target singular vectors and then fits a separate ridge reader from its standardized code to the raw field. Both fixed ridge coefficients are 0.001 in a sum-of-squares convention; the reader intercept is unpenalized. This name specifies the actual two-stage implementation, without equating it to a standard reduced-rank ridge optimum. Featurewise target scaling in the baseline comparison differs from the single global future RMS used in the horizon study (Supplement A9).

The physical reader starts from six fixed present amplitudes. On the grid $x_j = 2\pi j/63$, trapezoidal endpoint weights define cosine coefficients U_n, V_n of u, v for $n = 0, 1, 2$. Set

$$a = \frac{16}{75}(5U_1 + V_1), \qquad s = U_1 - a, \qquad f = (a, s, U_0, V_0, U_2, V_2). \tag{4}$$

The quadratic map is $F_{27} = (f, (f_p f_q)_{p \le q})$. Appending an intercept, training-only ordinary least squares maps six or 27 features to the same raw 128-component future. Validation is unused in this original physical diagnostic; the subsequent common-reader control uses validation as specified below. Its features and coefficients were fixed before its fit, but the original test set had already been inspected in earlier work; the added comparison is exploratory. PRED metrics average the three individual seed losses, not an ensemble forecast. Supplement A20 gives coefficients, forecast identities and independent arithmetic checks.

3.3 Linear selection, nonlinear structure and component-resolved errors

At $d = 1$ and $\tau = 10$, PCA and RECON have field errors 0.997145 and 0.997179; RTRP gives 0.523774 and PRED 0.497125. Thus a predictive linear coordinate already captures much of the

gain over present reconstruction. At $d = 4$, PRED gives 0.038779, compared with 0.152072 for RTRP and 0.152088 for the tested full-state linear reference. The latter is an empirical reference, not a lower bound over all linear algorithms. Supplement A2/A9 and the retained source tables give the full dimension and three-regime comparisons.

Reader input	Features	Field error	Critical R^2	Uniform R^2
Full current field	128	0.152088	0.977386	0.988384
Six known amplitudes	6	0.150637	0.977350	0.988687
Amplitudes + products	27	0.022884	0.999547	0.999806
PRED recovery code	4	0.038779	0.999673	0.996637

Table 2: Same full-field target at B_c, $\tau = 10$, original 360 test states. Critical and joint-uniform R^2 are projections of each field forecast. Neural entries average individual-model metrics. The feature count, physical prior, reader and output-rank allowances differ.

Table 2 places all predictors on the same future field. The six-feature linear reader is close to the full linear reference. Quadratic products expose substantially more predictability, giving error 0.022884. Each saved PRED seed has a larger field mean than that quadratic reader. Comparing the latter with the per-state mean of the three PRED losses favors the quadratic reader on 288 states and PRED on 72, with no ties or assigned inferential test.

The physical reactions contain $Bu^2+2uv+u^2v$, with opposite signs in the two species. Products of uniform and first-mode terms, or of first and second spatial modes, can project onto the first mode. In the scalar diagnostic, six linear amplitudes give critical-amplitude $R^2 = 0.977350$; quadratic and cubic expansions give 0.999547 and 0.999786. The quadratic augmentation removes 98.0% of residual squared error left by the six-linear-feature fit (Supplement A11). This supplies a physically specified predictive reference, while leaving the neural computation unidentified.

Crucially, the full-field ordering and critical-amplitude ordering differ. PRED has the higher critical-amplitude R^2; the quadratic reader has the lower whole-field error and smaller uniform and remaining-field squared errors. To locate these differences, decompose each error $e_i = \widehat{Y}_i - Y_i$ as

$$e_i = e_i^c + e_i^s + e_i^u + e_i^r, \tag{5}$$

using the critical, fast, uniform and remaining-field maps. With $d_i = \|Y_i\|_2 + 10^{-10}$ and component set $\mathcal{J} = \{c, s, u, r\}$,

$$\frac{1}{n}\sum_i \frac{\|e_i\|_2^2}{d_i^2} = \sum_{j\in\mathcal{J}} \frac{1}{n}\sum_i \frac{\|e_i^j\|_2^2}{d_i^2} + 2\sum_{j<k} \frac{1}{n}\sum_i \frac{\langle e_i^j, e_i^k\rangle}{d_i^2}. \tag{6}$$

The species directions are nonorthogonal and projection/evaluation norms differ, so all six cross terms remain. Equation 6 gives the exact squared-error decomposition with nonorthogonal cross terms. Its contributions refer to the squared loss, whereas the mean relative L2 in Eq. 2 uses the unsquared norm.

The original 360-state component comparison remains in Supplement A20; the prospective three-reader comparison below applies the same decomposition to the 720 new states.

The original comparison motivates a direct control: change the reader while retaining each neural code. The physical readers have 896/3584 coefficients for 6/27 columns; an affine four-code reader has 640 coefficients in addition to its 25,028-parameter encoder. Physical prior, feature expansion, output rank and fitting conventions differ. The control below standardizes the reader-fitting convention and tests a richer reader of exactly the same stored code; it does not equalize all resources.

3.4 Frozen-code readout control and external-cache evaluation

We compare affine and quadratic readers of each of three saved four-dimensional PRED codes. The latter has four linear terms and ten deterministic products, hence 14 columns. The physical references use six amplitudes or their 27-column expansion; a 128-column full-current affine reader is retained. Each predicts the same raw future field. Columns are centered and standardized only on the 1,080 training states, with a standard-deviation floor of 10^{-12}. Targets use their training mean and one global centered root-mean-square scale. An unpenalized intercept and ridge grid $\alpha \in \{0, 10^{-6}, 10^{-4}, 10^{-2}, 1\}$ are common to all nine reader instances. The 360 validation states select the unrounded mean relative-L2 minimum, with exact ties favoring larger α. All nine choices and coefficients are frozen before the new scoring. Thus there are 45 distinct configurations, without training or rerunning an encoder (Supplement A23).

The original 360-state test has already been inspected and supplies exploratory evidence. A separate cache contains 360 unique initial states, with no input overlap with the original 1,800 states, generated from the same initial-state law and numerical solver. Its stored unquantized PRED latents allow evaluation without neural execution or refitting. This cache was generated and analyzed for a different frozen-model comparison before the present readouts were selected. It is external to fitting, but a complete outcome-blind design history is not established. We therefore use it as an external-cache robustness check, distinct from both a newly prospective confirmation and the lattice's later-generated outcomes.

Reader	Columns	Original-test error	External-cache error
Full-current affine	128	0.151973	0.172576
Physical affine	6	0.150564	0.168710
PRED affine	4	0.038800	0.041683
PRED quadratic	14	0.026861	0.029014
Physical quadratic	27	0.022884	0.024310

Table 3: Common-reader control: mean per-state relative L2 on the same future-field target. PRED rows average three separate seed losses, not ensemble predictions. These newly selected readers are distinct from the historical coefficients in Table 2; the original-test and external columns use identical frozen readouts.

On the external cache, the quadratic PRED reader lowers mean error by 30.39%, with improvement in all three seed means. Its paired mean gain is 0.012669, with a descriptive paired-bootstrap 95% interval $[0.010602, 0.014919]$. It improves 274 states and worsens 86 after seed-loss averaging. The physical quadratic reader remains lower in mean error than every quadratic PRED seed; the seed-averaged paired gap is 0.004704, interval $[0.002482, 0.006910]$, with 233 states favoring the physical reader and 127 the neural quadratic reader. Supplement A23 reports all three prelisted contrasts and nominal multiplicity-adjusted endpoints. These approximate percentile intervals resample initial states, conditional on frozen pipelines; they exclude training, selection, design-history and solver uncertainty and are not finite-sample certificates.

The reader intervention measures an achieved forecast improvement relative to the selected affine reader. It provides a concrete witness of task-relevant structure accessible to the quadratic reader, while leaving the best attainable risk over other readers and the total information retained by the code undetermined. Original encoder training and post-training reader fitting also differ in target scaling and regularization; a gain does not establish that no affine training or fitting protocol could recover it. Four raw neural coordinates expanded to 14 columns and six physical amplitudes expanded to 27 remain different budgets; their reader coefficient counts are 1,920 and 3,584, with centered output-rank upper bounds 14 and 27. The physical construction also uses known modes, whereas the neural representation includes its separately trained encoder. These comparisons do not isolate a unique cause of the remaining gap or establish equal-resource superiority.

The component contrast also persists on the external cache: quadratic PRED has critical-amplitude $R^2 = 0.999585$, compared with 0.999444 for the physical quadratic reader, whereas uniform-component R^2 is 0.998944 versus 0.999798. Relative to the affine PRED reader, the richer reader lowers uniform and remaining-field squared errors, but its critical-amplitude R^2 is slightly lower on average. Loss rankings have their own boundary: on the original test, quadratic PRED has slightly smaller mean squared relative loss than physical quadratic readout; on the external cache this secondary ranking reverses. Lower mean relative L2 therefore does not imply a uniform advantage across states, components or losses. All cross terms, absolute losses and quantiles are retained in A23.

3.5 Prospective validation and a shared output-space control

We perform a control experiment by holding each encoder fixed and comparing affine and quadratic reader functions on the same initial states. We next fixed the nine selected readers, the three encoder checkpoints and the evaluation protocol before generating 720 new initial states from the same controlled law at B_c, $\tau = 10$. The new states are unique and have no input overlap with the original 1,800 states or the earlier external 360. Encoders perform only a frozen forward pass; no neural weights, reader choices or normalization parameters are fitted on these new states. The solver, grid and main loss in Eq. 2 are unchanged. This prospective bank is distinct from the earlier, previously examined external cache.

Before generating the new states, we also formed one proper orthogonal decomposition (POD) from centered, globally scaled training futures. Its nested bases $Q_r \in \mathbb{R}^{128\times r}$, $r \in \{4, 14\}$, define the same output subspace for every reader at each r. Each predicts training target coefficients and reconstructs the field using the training mean, scale and Q_r. Nine reader instances, two output spaces and the same five ridge values give 90 additional configurations; validation alone selects 18 readers. These and the original nine readers were sealed before the new generation. At $r = 4$, all nine centered coefficient maps have numerical rank four. At $r = 14$, an affine four-code reader still has rank at most four. These are preplanned secondary diagnostics, not additional confirmatory tests.

The single prespecified primary compares the original unrestricted affine and quadratic readers. With L^A_{ij} and L^Q_{ij} denoting Eq. 2's per-state loss for fixed model j on state i, define

$$\Delta_i = \frac{1}{3}\sum_{j=1}^{3}(L^A_{ij} - L^Q_{ij}), \qquad \overline{\Delta} = \frac{1}{720}\sum_{i=1}^{720}\Delta_i. \tag{7}$$

We resample the 720 independent initial states as pairs for 10,000 bootstrap replicates. Seed losses are averaged within state; the 2,160 model–state entries are not independent physical observations. The prespecified criterion is a positive lower endpoint of the approximate 95% percentile interval. This inference conditions on the three fixed encoders and readers and does not integrate training, selection or solver uncertainty.

Output restriction	PRED affine	PRED quadratic	Physical quadratic
Unrestricted	0.036970	0.027050	0.023878
Common POD4	0.036970	0.027101	0.023940
Common POD14	0.036970	0.027050	0.023878

Table 4: Mean relative L2 on all 720 newly generated states. PRED entries average the individual losses of the same three frozen encoders. Only the unrestricted neural affine–quadratic comparison is the primary; POD and physical-reader comparisons are descriptive. The physical dictionary has a different prior and coefficient budget.

Projecting these 720 true future fields onto the training-derived POD14 affine space has mean relative L2 truncation error 2.33×10^{-8} (Supplement A25.5). This small output-truncation

residual helps explain the agreement with unrestricted results at the displayed precision; it is not a prediction error or an additional validation.

The primary mean difference is 0.009920, with approximate 95% interval [0.008859, 0.011014]. Mean field error decreases by 26.83%; the quadratic reader improves 544 states and worsens 176 after seed-loss averaging, and each model seed's mean improves. Under the common POD4 restriction, the decrease remains 26.69%, with 543 states improved and 177 worsened (Fig. 1). Thus the readout improvement can occur without enlarging the shared output subspace or its measured rank: both neural coefficient maps have rank four. The comparison changes how coordinates within that space depend on the same code. It does not separate function-class approximation, regularization and finite-sample fitting effects. In particular, the primary mean relative L2 loss is not squared loss, so this observation is not a claim of closer approximation to a conditional mean.

Squared relative-error component	PRED affine	PRED quadratic	Physical quadratic
Critical	0.0006953	0.0006909	0.0011486
Uniform	0.0008694	0.0004188	0.0000536
Remaining field	0.0006376	0.0001177	0.0000246
Total, including fast and all cross terms	0.0022307	0.0012327	0.0012252

Table 5: Physical error localization on the 720 prospective states, unrestricted readers. Terms decompose mean squared relative loss in Eq. 6, not mean relative L2 or percentages of its improvement. Neural columns average the three individual-model squared-loss terms. The selected diagonal rows do not sum to the total; Supplement A25.4 retains the fast component and all six cross terms. These component comparisons are descriptive.

Table 5 locates the accuracy recovered by changing the reader of the same neural codes. The uniform and remaining-field squared errors decrease, while the critical contribution changes little. The physical quadratic reader retains still smaller uniform and remaining-field errors but a larger critical error. The controlled readout change therefore recovers task-relevant structure from the existing code; the component diagnostic identifies where the forecasts differ without assigning the remaining gap to a unique physical or neural mechanism.

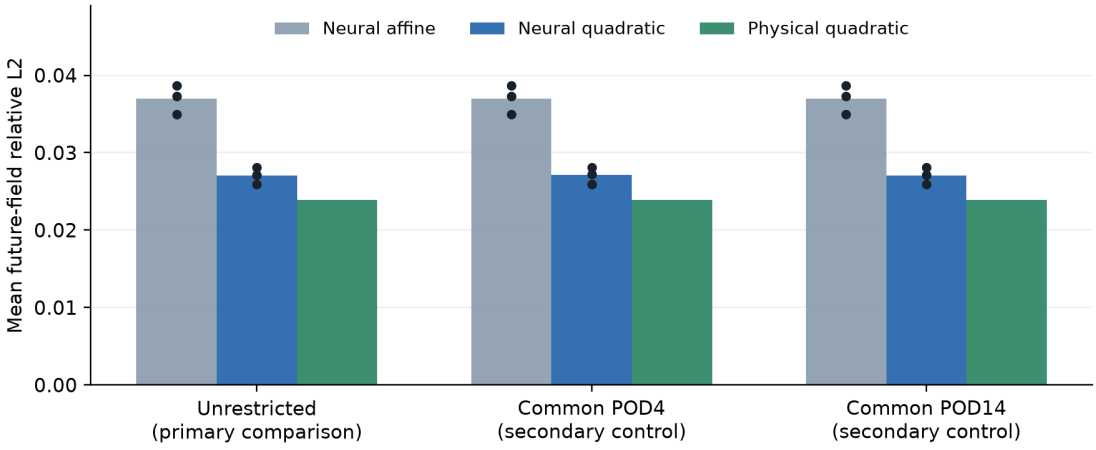


Figure 1: The same 720 states evaluated with unrestricted and shared-output readers. Bars show mean relative L2; dots show the individual fixed neural-model means, not confidence limits. POD4 and POD14 are secondary controls on the same bank. A common output space leaves differences in feature expansion, reader parameters, physical prior and encoding cost.

The physical quadratic reader remains lower in full-field mean than every neural quadratic

reader, while its critical-amplitude R^2 is 0.999399 versus the neural average 0.999560; joint-uniform R^2 is 0.999792 versus 0.999041. The component contrast therefore persists in new states. Even at POD4, affine and quadratic neural readers have 20 and 60 fitted coefficients per model; the physical quadratic reader has 112, with a shared basis counted separately. These controls match the output subspace rather than the entire resource budget. Supplement A25 retains every pipeline, all reverse-direction states, absolute errors, quantiles, numerical ranks and nonorthogonal cross terms. A fixed first-36-state half-step check preserves the primary direction; it is a numerical sensitivity check, not a strict ODE certificate for the 720 states.

3.6 Prediction horizon and physical interpretation

The sampled Brusselator horizons $\{0.1, 0.5, 2, 10\}$ expose different contributions. At 0.1, the one-dimensional neural code has current fast-coordinate $R^2 = 0.996362$; at 10 its current critical-coordinate $R^2 = 0.972129$. At 0.5 and 2, the uniform future-component squared centered norm is 77.07% and 91.26% of the total training-field norm; these ratios are not an orthogonal additive variance partition. Both uniform current species amplitudes predict the one-dimensional RTRP code with test $R^2 = 0.999692$ and 0.999954 (Supplement A9).

The critical and fast first-mode eigenvalues are $0, -75/16$, whereas the uniform block has $-7/32 \pm i\sqrt{975}/32$. Thus the finite-horizon observations are consistent with fast decay, longer-lived uniform oscillation, and a critical contribution. Intermediate-horizon historical neural latents were not retained, so the uniform interpretation rests on fields and linear coordinates, not a direct probe of those neural encoders. Loss geometry matters for this interpretation [7]; no unique slow eigenfunction or autonomous reduced law is identified.

4 Prospective refinement: information and fixed-predictor gain

4.1 Driven lattice and the finite-bank task

A real field q and auxiliary field y occupy a periodic 64×64 lattice. With unit spacing and nearest-neighbor pairs $\langle i, j\rangle$, the energy is

$$\mathcal{F}[q, y; a] = \sum_i \left(\frac{a q_i^2}{2} + \frac{u_4 q_i^4}{4} + \frac{\kappa y_i^2}{2} - g q_i y_i \right) + \frac{K}{2} \sum_{\langle i,j \rangle} (q_i - q_j)^2. \tag{8}$$

Here $u_4 = \kappa = g = K = T = 1$, with order/auxiliary mobilities $\Gamma = \nu = 1$. The symbol u_4 denotes the quartic coefficient, distinct from the Brusselator deviation u. The executable dynamics use OU half steps for y around a Metropolis-adjusted q update, with $\Delta t = 0.02$; rejected proposals consume a step. The protocol is

$$a(t) = a_c + \mathrm{clip}(-t/\tau_Q, -1, 1), \quad \tau_Q = 64, \quad a_c = -0.274. \tag{9}$$

The reference a_c comes from finite-size static calibration, not a certified thermodynamic-limit critical point. Standard Metropolized-integrator theory provides context [17]; the claims here concern the recorded split kernel, without a driven-continuum convergence assertion (Supplement B1/B2).

Complete formation states were saved at ages $t = -8, 0, 16, 20$. The original formation bank used finite static chains, followed by the driven evolution from $t = -64$, with observation blocks fixed before formation; its initialization and all source identities are retained in B4. The same 512 path identities recur across ages. For every age–parent state, 16 independently assigned conditional random futures advance for lag 16, yielding 32,768 six-class terminal outcomes. The target O_{16} bins the selected 8×8 block's terminal q mean at cutpoints $(-1, -0.5, 0, 0.5, 1)$. No new outcomes are generated for the present combined manuscript.

The four future windows are $[-8, 8], [0, 16], [16, 32], [20, 36]$. Their future forcing remains active and differs. Age-specific lookups and class distributions also differ. The results are conditional on the fixed realized complete states, with new future streams supplying continuation uncertainty. They are not population guarantees for new formation states or a controlled causal age comparison.

4.2 Nested sensor codes and frozen probability forecasts

The unquantized observation $X4$ contains the chosen block's q mean, y mean, population variance of q, and internal sign-wall fraction. The coarse code E bins these four quantities. The refinement $Z4$ retains E and adds a predetermined midpoint bit where a finite interval is available, with unchanged tail rules. Hence $E = h(Z4)$ and both are deterministic functions of the same four sensors. The refinement adds resolution and message cost, rather than new independent sensor channels. Supplement B3 gives all thresholds and exact tail/equality conventions.

An earlier, disjoint 256-parent training bank supplies age-specific count lookups. It was inspected during method development; the prospective claim begins with the subsequent freezing of the codes, predictor and evaluator before new evaluation labels existed. For six classes $k = 0, \ldots, 5$, define training prior $\pi(k) = (c_{\text{global}}(k) + 1)/(\sum_j c_{\text{global}}(j) + 6)$ and frozen probabilities

$$\begin{aligned} q_E(k) &= \frac{c_E(k) + \lambda_s \pi(k)}{\sum_j c_E(j) + \lambda_s}, \\ q_Z(k) &= \frac{c_Z(k) + \lambda_s q_E(k)}{\sum_j c_Z(j) + \lambda_s}, \qquad \lambda_s = 64. \end{aligned} \tag{10}$$

The smoothing weight λ_s is denoted τ in the original lattice records, and differs from drive time τ_Q and forecast horizon. An unseen refined training cell falls back to q_E; all probabilities are positive. No lookup was refitted to the prospective outcomes. The fourteen-component descriptor called S14 in the source contains $X4$ as its first four entries; its other ten coordinates are unused by E and $Z4$.

4.3 Two distinct finite-simulation certificates

At one age, let J be uniform on the $n = 512$ saved complete states, and let $p_i(k)$ be the true class probability under the prescribed continuation kernel from state i. The finite-bank law is $P_B(J = i, O = k) = p_i(k)/n$. The information endpoint is $I_B(O_{16}; Z4 \mid E)$, not a plug-in estimate from sixteen counts.

Form the fixed vector $w_i(k) = q_Z(k \mid i) - q_E(k \mid i)$ and subtract its uniform-parent mean within each E cell to obtain $v_i(k)$. Then $\mathbb{E}_B[v_J(k) \mid E] = 0$. Write $A = n^{-1} \sum_i \|v_i\|_2^2$ and

$$\widehat{\mu} = \frac{1}{nm} \sum_{i,r} v_i(O_{ir}), \quad r_v = \sqrt{\frac{\log(1/\alpha) \sum_i [\max_k v_i(k) - \min_k v_i(k)]^2}{2n^2 m}}, \tag{11}$$

where $m = 16$. Cell centering, Cauchy–Schwarz and Pinsker give $|\mu|^2 \le 2AI_B$; Hoeffding bounds the sampled moment [12, 13]. Thus a valid one-sided lower certificate is

$$I_{\text{lower}} = \frac{[\max(0, \widehat{\mu} - r_v)]^2}{2A}, \tag{12}$$

with zero reported when $A = 0$. Supplement B8 supplies the full proof, conditioning and independence assumptions; these are applications of standard inequalities.

For operational performance, define the fixed score difference $g_i(k) = \log(q_Z(k \mid i)/q_E(k \mid i))$. Its mean and lower bound are

$$\begin{aligned} \widehat{G} &= \frac{1}{nm} \sum_{i,r} g_i(O_{ir}), \\ G_{\text{lower}} &= \widehat{G} - \sqrt{\frac{\log(1/\alpha) \sum_i [\max_k g_i(k) - \min_k g_i(k)]^2}{2n^2 m}}. \end{aligned} \tag{13}$$

Positive gain favors the refinement. Two bounds at each of four ages give eight preallocated one-sided endpoints, each at $\alpha = 0.00625$, for simultaneous lower-coverage probability at least 0.95 by the union bound. It requires the specified conditional independence of future streams; unique seed numbers alone are not a proof of that assumption. The coverage does not include formation-population or training uncertainty. No direct age-difference test was allocated.

4.4 Prospective outcomes and their interpretation

Age	Information lower	Mean score gain	Score lower	Positive lower bounds
−8	0.0000487294	0.00608743	−0.00999280	Information
0	0.0000535577	0.00965340	−0.00286194	Information
16	0.0003815571	0.02439372	0.01308309	Information and score
20	0.0002984871	0.03554475	0.02455406	Information and score

Table 6: Frozen lattice endpoints. Information is in nats and score gain in nats per outcome. All four point score gains are positive. Positive lower bounds at only two ages do not demonstrate absent early gains or a significant cross-age difference.

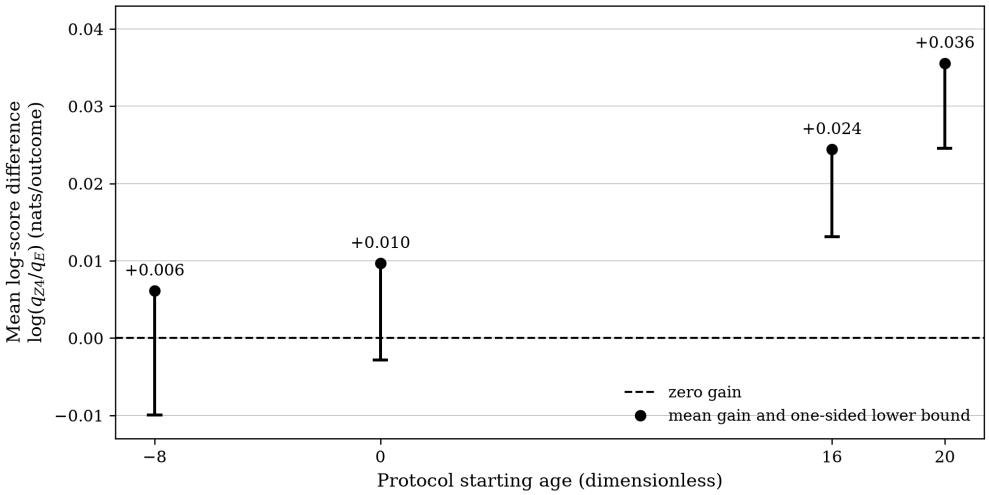


Figure 2: Previously measured lattice score gains and preallocated one-sided lower limits. The plotted endpoints are retained from the lattice source. The horizontal zero line is the no-gain reference; the four points correspond to different future drive windows.

The refinement has strictly positive conditional information certificates at all four ages, including the earliest two. Fixed-predictor mean score gains are also positive throughout, but their conservative lower bounds resolve positivity only at ages 16 and 20 (Table 6). At the earlier ages, information is certified while operational advantage remains unresolved at the allocated error level. Finite continuation budget and the conservatism of the range-based Hoeffding bound can limit this resolution; the experiment does not identify their separate contributions. Increasing

the number of conditional futures would shrink the bound radius for fixed parent states and predictors, but also changes the sampled mean and need not produce a positive lower endpoint. This is a difference in certification outcomes under the present budget. It neither establishes a true information–utility separation nor tests a difference between ages. Six positive lower bounds among eight endpoints retain their separate endpoint meanings; they do not define an additional omnibus test.

The descriptive class and prediction diagnostics help contextualize the experiment. Empirical six-class entropy is 1.3762, 1.4940, 1.7648 and 1.7730 nats in age order; extreme-class proportions are 0.39%, 2.86%, 29.08% and 37.19%. Brier-score reductions are positive, while refined training-cell coverage decreases from 77.93% to 68.36% (Supplement B7). These post-outcome observations do not isolate a physical mechanism. The cross-entropy difference is algebraically the same score gain, not an independent replication. Smoothing 256 is a descriptive sensitivity, outside the eight-endpoint family.

Because $Z4$ is determined by $X4$, which is included in S14, conditional data processing gives

$$I_B(O; \mathrm{S14} \mid E) \geq I_B(O; X4 \mid E) \geq I_B(O; Z4 \mid E). \tag{14}$$

The positive certificates therefore also bound these richer observations relative to the same coarse E. They provide no evidence that the extra ten S14 coordinates add information beyond continuous $X4$, that S14 is sufficient, or that a predictive state is born at a later age.

5 What the combined evidence resolves

5.1 One evaluation question, complementary evidence

The principal empirical contribution is the prospective fixed-code readout intervention, its shared-output-space control and its physical error localization. On 720 new same-distribution initial states generated after method freezing, quadratic readout of unchanged neural codes lowers full-field error, and the reduction remains in a common four-dimensional output space. Within the same known-mode generation law, the physical dictionary has lower mean relative L2 but less accurate critical amplitudes, locating differences in uniform and remaining-field components. This is a diagnostic reference with different prior and coefficient budgets. An affine-reader deficit therefore cannot by itself locate the whole forecast deficit in encoding loss. Nor does lower mean relative L2 imply a ranking under every loss: the exploratory squared-relative-loss ranking reverses between the original test and external cache (A23). In the lattice, a fixed refinement changes only the resolution of four sensors, and the future experiment separately bounds additional information and lookup score gain. The two systems retain their own losses, conditioning and error control rather than forming a common significance claim. The supplementary measured-signal case distinguishes model refitting from input dependence in a fixed model.

The evidence laws remain distinct. The 32,768 lattice outcomes support probability assertions conditional on a fixed complete-state bank and specified independent continuation streams. The new field primary uses paired bootstrap over 720 initial states, conditional on three already trained encoders. Its interval is approximate; the common-output-space and physical comparisons remain descriptive. The three seeds document these fixed trained models, not the distribution of outcomes from retraining. In the lattice, the four positive information bounds and two positive score bounds describe resolution of different conservative certificates, not different times at which information and utility first exist. The exploratory original test and previously examined external cache keep their earlier evidence roles. Agreement with them strengthens the specific readout finding without creating new neural-training replicates or a distribution-free guarantee.

5.2 Relation to the closest methods

Time-lagged autoencoders already implement present-to-future neural compression [1], and variance/timescale issues were analyzed in this architecture [7]. The field contribution here is a fixed-code reader comparison tested on newly generated initial states, a training-only common-output-space control, and explicit physical products with complete component/cross-term diagnostics. It evaluates the contribution of readout choice in this benchmark rather than introducing an autoencoder or optimizing over all possible readers. Predictive bottlenecks and information-efficient measurements already treat future relevance under constraints [2, 4, 16, 5].

Taverniers et al. selected stochastic coarse models by prediction tests, including one-step and autonomous forecasts [14]. Our lattice task fixes the observational map and a terminal categorical outcome, with an unchanged simulator. Predictive critical information [15] and learned stochastic coarse dynamics [19] address different inference targets. Here the additional experiment is a frozen, prospective pair of conditional information and fixed-lookup score certificates, with all four drive windows retained. Standard inequalities and this combined presentation do not establish a new general information theorem or priority over every coarse-graining method.

5.3 Supporting cases and boundaries

A Coupled Electric Drives recorded-signal audit separates model-building gains from dependence on a historical input [21, 22]. The history-augmented ridge predictor H has lower RMSE than the separately fitted shared-input predictor S in the five-step task. Replacing only H's history difference by its development mean, with coefficients and other inputs fixed, lowers test-record RMSE from 0.666120 to 0.615456, while raising validation RMSE from 0.297689 to 0.317889. Persistence already has lower test RMSE than H (0.643946). This post-hoc diagnostic uses recorded future controls. It measures the varying history coordinate's effect in that fixed predictor, rather than intrinsic historical information or a physical memory. A24 retains both records, all error metrics, the strong simple references and observation budgets. Refitted-model comparisons and fixed-model input dependence therefore receive separate empirical descriptions.

The KS horizon comparison supports target-time dependence on a separate chaotic system: matched and mismatched representations give errors 0.550600 and 0.698865 at target horizon 4. Eighteen whole test-trajectory mean differences are positive; the saved exploratory trajectory-bootstrap interval is $[0.133347, 0.161123]$ for their absolute difference, conditional on the retained models. It is a reanalysis on the original physical cache, not fresh KS replication (Supplement A13). The molecular comparison retains TICA, supervised linear and torsion-feature references; its prespecified target-specific cross-ranking is unconfirmed. It supplies a boundary case, without a molecular mechanism or equivalence claim (A21).

Additional sensing is retained as a separate supporting application (A3–A7, A10, A12, A16), not a second principal result. The corrected Brusselator ambiguity primary remains directionally favorable by approximately 1.68%, but unconfirmed, with one-sided randomization $p = 0.084983$. Ten independently regenerated blocks do not resolve a positive mean effect, and the short-horizon relative-error grid is adverse. KS has favorable fixed-pool evidence without matching fresh-data repetition. These outcomes do not support a two-system independent replication or a universal uncertainty-to-value rule. The finite-bit and true-ODE proofs concern separate fixed quantized-model comparisons, specified finite sets and stated arithmetic assumptions; they do not certify the exploratory unquantized quadratic reader (A8/A18 and the separate proof attachment).

The field initial generator uses known modes and one discretization. The prospective validation retains that law and the three frozen encoders; common POD spaces leave unequal function classes, parameters, physical priors and encoding costs. The lattice information law conditions on a fixed realized bank, with finite-chain formation history and age-dependent configurations, forcing and lookups. Neither study isolates a unique physical mechanism of neural computation or a causal age effect. New formation-population guarantees, fully matched resource budgets

and driven-kernel continuum limits remain outside the demonstrated scope.

6 Conclusion

Our principal result quantifies forecast accuracy recovered by changing only the reader of frozen four-coordinate neural codes. On 720 new initial states generated after protocol and reader freezing, quadratic readout lowers mean relative L2 from 0.036970 to 0.027050; the prespecified paired mean difference has a positive approximate 95% bootstrap interval. A common four-dimensional output-space control retains a 26.69% reduction as a secondary diagnostic. Under the same controlled generation law, the richer known-mode dictionary has lower mean relative L2 and smaller uniform and remaining-field squared errors, while the neural code predicts critical amplitudes more accurately. These are diagnostic comparisons with unequal resources, rather than a general ordering of physical and learned representations. The fixed-code intervention measures achieved readout improvement, not optimal decoding or training-seed generality. In the complementary lattice study, conservative certificates at the allocated conditional-future budget resolve positive information bounds at four ages and positive fixed-predictor score bounds at two; all four score means are positive. The differing decisions do not demonstrate an information–utility separation. Declaring the reader, field component and sampling assumptions makes these controlled forecast comparisons interpretable within their stated budgets.

Supplementary materials

The combined Supplementary Information contains Part A, the complete field, molecular and sensing evidence with source section identifiers A1–A22 plus the readout-control and external-cache record A23 and the recorded-signal history diagnostic A24, the prospective 720-state and common-output-space record A25, and Part B, the lattice methods, provenance and full conditional-information proof B1–B9. A new overview maps their roles in this manuscript. Supplementary Proofs preserve the complete finite-task and true-ODE arguments for the separate coded-model comparisons. Detailed supplemental references and original result identities are retained; they are not additional independent experiments.

Statements and Declarations

Required author details and declarations are provided through the submission system.

The fixed public repository version [11] contains selected materials from the original field/sensing study, including original physical-readout forecasts and a no-fit arithmetic replay. It does not deposit this combined manuscript, the common-reader controls, either newer field-validation bank, the lattice study or the CED diagnostic. The retained field/lattice evaluation companion, `Evaluation_Replay.zip`, reconstructs the two earlier 360-state field comparisons and the eight lattice endpoints without new fitting, neural execution or outcome generation. The separate CED companion supplies saved record-level predictions and arithmetic replay (A24). The accompanying anonymized `Prospective720_Evaluation_Replay.zip` (Online Resource 5) supplies the frozen protocol, reader parameters, original fitting cache, three retained checkpoints, new initial states and endpoints, all 27 forecasts, losses and bootstrap draws. It is a reviewable copy of the existing result archive with identifying metadata and local paths sanitized; scientific arrays and frozen coefficients are unchanged. Its portable evaluation-only replay verifies saved forecasts and the primary comparison without training, fitting, encoder execution or integration. It does not reconstruct historical neural training. A25 specifies the members and replay coverage. This accompanying file is separate from the fixed older public deposit and does not claim a new public accession. The larger historical molecular, interval and

trajectory payloads retain their stated boundaries in A15/A18/A21/B9. Manuscript integration uses these already completed calculations and generates no additional experimental result.

References

[1] Wehmeyer, C., & Noé, F. (2018). Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *The Journal of Chemical Physics*, 148, 241703. doi:10.1063/1.5011399.

[2] Creutzig, F., Globerson, A., & Tishby, N. (2009). Past-future information bottleneck in dynamical systems. *Physical Review E*, 79, 041925. doi:10.1103/PhysRevE.79.041925.

[3] Still, S., Crutchfield, J. P., & Ellison, C. J. (2010). Optimal causal inference: Estimating stored information and approximating causal architecture. *Chaos*, 20, 037111. doi:10.1063/1.3489885.

[4] Sachdeva, V., Mora, T., Walczak, A. M., & Palmer, S. E. (2021). Optimal prediction with resource constraints using the information bottleneck. *PLoS Computational Biology*, 17(3), e1008743. doi:10.1371/journal.pcbi.1008743.

[5] Murphy, K. A., & Bassett, D. S. (2024). Machine-Learning Optimized Measurements of Chaotic Dynamical Systems via the Information Bottleneck. *Physical Review Letters*, 132, 197201. doi:10.1103/PhysRevLett.132.197201.

[6] Prigogine, I., & Lefever, R. (1968). Symmetry Breaking Instabilities in Dissipative Systems. II. *The Journal of Chemical Physics*, 48, 1695–1700. doi:10.1063/1.1668896.

[7] Chen, W., Sidky, H., & Ferguson, A. L. (2019). Capabilities and limitations of time-lagged autoencoders for slow mode discovery in dynamical systems. *The Journal of Chemical Physics*, 151(6), 064123. doi:10.1063/1.5112048.

[8] Otto, S. E., Macchio, G. R., & Rowley, C. W. (2023). Learning nonlinear projections for reduced-order modeling of dynamical systems using constrained autoencoders. Chaos, 33, 113130. doi:10.1063/5.0169688.

[9] Linot, A. J., & Graham, M. D. (2022). Data-driven reduced-order modeling of spatiotemporal chaos with neural ordinary differential equations. Chaos, 32, 073110. doi:10.1063/5.0069536.

[10] Dai, R. (2026). Internal compression, collective state formation, and event memory. Author preprint, Zenodo. `https://doi.org/10.5281/zenodo.23211854`.

[11] Dai, R. (2026). Future-Conditioned Representation and Active Sensing in Dynamical Systems. Version 1, Zenodo. `https://doi.org/10.5281/zenodo.23599714`.

[12] Hoeffding, W. Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.* **58**, 13–30 (1963). https://doi.org/10.1080/01621459.1963.10500830

[13] Cover, T. M. and Thomas, J. A. *Elements of Information Theory*, 2nd ed. Wiley (2006). https://doi.org/10.1002/047174882X

[14] Taverniers, S., Haut, T. S., Barros, K., Alexander, F. J. and Lookman, T. Physics-based statistical learning approach to mesoscopic model selection. *Phys. Rev. E* **92**, 053301 (2015). https://doi.org/10.1103/PhysRevE.92.053301

[15] Tchernookov, M. and Nemenman, I. Predictive information in a nonequilibrium critical model. *J. Stat. Phys.* **153**, 442–459 (2013). https://doi.org/10.1007/s10955-013-0833-6

[16] Wang, D. and Tiwary, P. State predictive information bottleneck. *J. Chem. Phys.* **154**, 134111 (2021). https://doi.org/10.1063/5.0038198

[17] Bou-Rabee, N. and Vanden-Eijnden, E. Pathwise accuracy and ergodicity of Metropolized integrators for SDEs. *Comm. Pure Appl. Math.* **63**, 655–696 (2010). https://doi.org/10.1002/cpa.20306

[18] Wang, D., Qiu, Y., Beyerle, E. R., Huang, X. and Tiwary, P. An information bottleneck approach for Markov model construction. *J. Chem. Theory Comput.* **20**, 5352–5367 (2024). https://doi.org/10.1021/acs.jctc.4c00449

[19] Christians, L. F., Wojnar, A. and Pak, A. J. Probabilistic forecasting for coarse-grained molecular dynamics. *J. Chem. Theory Comput.* **22**, 4231–4246 (2026). https://doi.org/10.1021/acs.jctc.5c02131

[20] Xu, Y., Zhao, S., Song, J., Stewart, R. and Ermon, S. A theory of usable information under computational constraints. *International Conference on Learning Representations* (2020). https://arxiv.org/abs/2002.10689

[21] Wigren, T. and Schoukens, M. Coupled Electric Drives Data Set and Reference Models. Technical Report 2017-024, Department of Information Technology, Uppsala University (2017). `https://www.nonlinearbenchmark.org/benchmarks/coupled-electric-drives`.

[22] Fisher, A., Rudin, C. and Dominici, F. All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *Journal of Machine Learning Research* **20**(177), 1–81 (2019). `https://www.jmlr.org/papers/v20/18-760.html`.

Supplementary Files

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

- supplement.pdf
- supplementaryproofs.pdf
- EvaluationReplay.zip
- CEDEvaluationReplay.zip
- Prospective720EvaluationReplay.zip