NS-WM: NEURAL-SYMBOLIC WORLD MODELING UNDER PARTIAL OBSERVABILITY

Anonymous authors
Paper under double-blind review

ABSTRACT

World models under partial observability must generalize to combination-OOD settings, where familiar sensor views and operating regimes reappear only in withheld combinations. We propose NS-WM, a neuro-symbolic framework that factorizes view-dependent observation mappings from sparse state–action–regime dynamics. This separation enables the same transition model to operate with different sensor maps, while observation-aware state estimation links masked histories to the latent state required for rollout. For controlled systems, NS-WM learns symbolic latent dynamics via an encoder route and jointly identifies dynamics and sensor maps in reference-sensor coordinates, estimating states by matching visible histories. Both routes support multi-action rollouts by reusing the current state across alternative action plans, and a validation-selected fusion combines observation predictions. Shared temporal and sensor experts extend the prediction interface to heterogeneous sequences. Across six benchmark suites and an additional controlled population, NS-WM performs structured world modeling under partial observation: infer states from masked histories, generate future trajectories under proposed actions, and decode them through changing sensor views. In addition, coefficient and state-aliasing diagnostics make learned transitions and failure modes inspectable. We further evaluate forecasting extensions on heterogeneous sequence suites. Results indicate that structured state estimation and reusable dynamics support compositional prediction, while absolute forecasts remain sensitive to periodic state ambiguity and validation-selected fusion is not uniformly beneficial.

1 INTRODUCTION

World models predict how a dynamical system evolves and how its future changes under alternative actions (Ha & Schmidhuber, 2018; Hafner et al., 2019; 2020). Under partial observability, they must infer a state from incomplete histories before simulating future observations (Kaelbling et al., 1998; Littman et al., 2001). Varying sensors and operating conditions make this difficult: sensor views determine how the state is measured, whereas regimes and actions influence its evolution. Collecting training trajectories for every pairing is costly. We therefore study *combination out-of-distribution (combination-OOD) prediction*: a sensor view and an operating regime each occur in training, but their evaluated pairing is withheld. The goal is to reuse what was learned about these factors to predict their withheld pairing, including how alternative action plans change the future.

Sparse identification provides a foundation for reusable dynamics: SINDy selects compact equations from a prescribed library (Brunton et al., 2016a), and SINDYc incorporates control inputs (Brunton et al., 2016b). SINDy autoencoders learn coordinates jointly with sparse dynamics (Champion et al., 2019; Lusch et al., 2018), while encoder-based symbolic models reconstruct hidden states from partial observations (Lu et al., 2022). Multi-view state-space models address varying sensing configurations (Hwang et al., 2023), and structured or context-conditioned dynamics investigate reuse across environments (Zhao et al., 2022; Kirchmeyer et al., 2022; Lei et al., 2022; Wang & Huang, 2025). For combination-OOD prediction, these capabilities must work together: a transition learned across regimes must be initialized from the available masked history and decoded through the appropriate sensor mapping. A transition alone does not specify how to make this connection under a withheld view-regime pairing.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

NS-WM addresses this connection by separating view-dependent sensing from state-action-regime evolution. A shared symbolic transition represents how actions and regimes modify the state, while sensor-specific mappings determine how that state appears in each view. This supplies a concrete inductive bias for combination-OOD transfer: a familiar regime can drive the same transition and a familiar sensor mapping can decode its trajectory without fitting a separate transition for their pairing. Sparse coefficients restrict the transition to selected terms in a prescribed library, making the reused action and regime dependencies explicit. This design motivates transfer; its predictive benefit is assessed by the withheld-pair experiments.

The second requirement is to initialize these reusable dynamics from partial observations. NS-WM uses an encoder route that maps masked histories to a learned state, alongside an identification route that fits dynamics and sensor maps in reference-sensor coordinates and estimates an initial state by matching simulated visible histories. At inference, the latter changes the state estimate while keeping dynamics and sensor maps fixed. Each route reuses its current state across alternative future action plans, linking historical evidence to action-conditioned forecasts. The routes retain separate coordinates and combine predictions in observation space; shared temporal and sensor experts extend the interface to heterogeneous sequences. State estimation remains consequential: our controlled diagnostics find candidates separated by approximately 2π with similar visible-history residuals but different hidden-channel futures. Thus, equation recovery, initialization, and action-effect accuracy must be evaluated separately.

Our contributions are threefold:

- Factorized symbolic prediction for combination-OOD transfer. NS-WM separates view-dependent observation mappings from shared sparse state-action-regime dynamics. This structure allows learned sensing and evolution factors to be recombined under withheld view-regime pairings, with explicit action and regime dependencies instead of a separately fitted transition for each pairing.
- Observation-aware initialization of reusable dynamics. We connect masked histories to symbolic rollout through encoder-based inference and reference-coordinate history fitting. The fitting route estimates the historical initial state under fixed dynamics and sensor maps; each route then reuses its current state across action plans. This supplies the state-inference step needed to apply the shared transition under a combination-OOD pairing without refitting its coefficients on test data.
- Evaluation of combination-OOD prediction and its failure modes. Across six suites and an additional controlled population, NS-WM improves combination-OOD MSE over both baselines in $13/14$ condition-population settings and improves both forecasting and action-effect skill in all nine controlled settings. Equation-recovery, component, and state-aliasing diagnostics distinguish the predictive roles of the fitted dynamics, state initialization, and validation-selected fusion.

2 Related Work

Sparse dynamics and state reconstruction. SINDy identifies sparse governing equations from a prescribed feature library, and SINDYc extends identification to controlled systems (Brunton et al., 2016a;b). SINDy autoencoders jointly learn coordinates and sparse dynamics (Champion et al., 2019); Lu et al. (2022) reconstruct hidden states and discover symbolic equations from partial observations by matching time derivatives. NS-WM targets combination-OOD forecasting through masked-history initialization; its reference-coordinate identification route uses paired full-channel training targets.

World models and observation-based inference. Latent-state, token-based, and diffusion-based world models support action-conditioned simulation (Ha & Schmidhuber, 2018; Hafner et al., 2019; 2020; 2023; Micheli et al., 2023; Alonso et al., 2024). Fuse2Control combines multi-view sequences in a state-space model and supports control with missing views (Hwang et al., 2023). Deep Kalman Filters and Deep Variational Bayes Filters learn state-space models with approximate inference (Krishnan et al., 2015; Karl et al., 2017); predictive-state representations and moving-horizon estimation provide complementary views of state recovery from finite histories (Littman et al., 2001; Rao et al.,

2003; Kaelbling et al., 1998). NS-WM forecasts the input view, including its hidden channels, under combination-OOD pairings, with paired-view supervision during training.

Compositional generalization and dynamics reuse. C-SWM learns object- and relation-based transitions, while HOWM studies recombination in object-oriented world modeling (Kipf et al., 2020; Zhao et al., 2022). CoDA adapts context-conditioned dynamics to new physical systems (Kirchmeyer et al., 2022); Variational Causal Dynamics learns reusable latent mechanisms across environments (Lei et al., 2022); WM3C uses language-guided composable causal components for task generalization (Wang & Huang, 2025). NS-WM tests combination-OOD sensing-regime pairings with fixed identified dynamics and sensor maps; matched-action forecasts assess predictive effects.

3 NS-WM

NS-WM predicts future observations from masked histories by combining state estimation, symbolic dynamics, and view-dependent decoding. For controlled systems, an encoder route and an identification route maintain separate state coordinates and combine their forecasts in observation space (Figure 1).

3.1 Prediction Task

Let $y_\tau^{(v)} \in \mathbb{R}^C$ be an observation in view v, $m_\tau \in \{0,1\}^C$ its availability mask, and r a known regime descriptor. Given

$$\mathcal{H}_t = (m_\tau \odot y_\tau^{(v)}, m_\tau, u_\tau)_{\tau=t-L+1}^{t}, \qquad \mathcal{A} = (u_{t+1}, \ldots, u_{t+H}), \tag{1}$$

where $\mathcal{H}_t$ is the available history at current time t, $\mathcal{A}$ is a future action plan, τ indexes observation times, L is the history length, and H is the forecast horizon. The vector $y_\tau^{(v)}$ contains C real-valued observation channels in sensor view v; m_τ is its binary visibility mask, u_τ is the incoming action, and $\odot$ denotes elementwise multiplication. We predict H future observations. The incoming action $u_{\tau+1}$ drives the transition to time $\tau+1$, separating evolution from sensing:

$$s_{\tau+1} = F(s_\tau, u_{\tau+1}, r), \qquad y_\tau^{(v)} = G_v(s_\tau) + \epsilon_\tau. \tag{2}$$

where s_τ is the underlying physical state at time τ, $u_{\tau+1}$ acts over the transition to $\tau+1$, r is the supplied regime descriptor, and ϵ_τ is observation noise. The functions F and G_v denote physical state evolution and sensing in view v, respectively; learned predictors approximate these processes from observations. Predictors do not receive the underlying physical state. Combination-OOD withholds a view-regime pair while retaining both factors elsewhere in training; parameter-OOD tests regime values outside the training range. For heterogeneous sequences, the condition can instead describe a source or event.

3.2 Identification and History-Based State Estimation

The first two reference-view training channels define coordinates $x \in \mathbb{R}^2$. Paired full-channel training targets fit view-specific sensor maps and a shared controlled vector field:

$$G_v^{\mathrm{id}}(x) = \Phi(x) B_v, \qquad \dot{x} = \Theta(x, u, r)\Xi. \tag{3}$$

where $x = (x_1, x_2) \in \mathbb{R}^2$ is the reference-sensor state and $\dot{x}$ is its time derivative. The row vectors $\Phi(x)$ and $\Theta(x, u, r)$ collect sensor-map and dynamics features; B_v maps sensor features to the C channels of view v, and Ξ maps dynamics features to the two state derivatives. Here u is the control input, r is the regime descriptor, and the superscript id denotes the identified route. The libraries contain polynomial and trigonometric features, augmented with action, regime, and interaction terms for dynamics. Smoothed training trajectories supply derivative targets; RMS normalization, ridge regression, and iterative thresholding estimate sparse coefficients. An affine bridge from paired view-2 and reference-view observations supplies additional reference-coordinate trajectories. Fitting uses training observations, without simulator equations or hidden states.

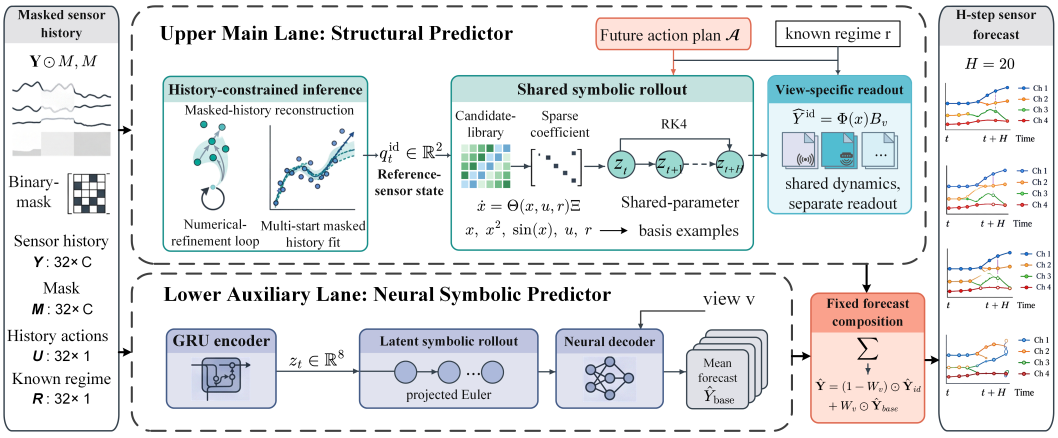

Figure 1: Controlled NS-WM combines encoder-based state inference with history fitting under identified dynamics and sensor maps. Each route reuses its state estimate across future action plans, and validation-fitted weights combine the decoded forecasts in observation space. The schematic illustrates three plans and evaluation uses four; any belief notation denotes state estimates or candidates, not a posterior distribution.

At inference, the fitted dynamics and sensor maps remain fixed. Starting at $t_0 = t - L + 1$, let $x_\tau(\xi)$ integrate historical actions from candidate state $x_{t_0} = \xi$. The data-fit objective is

$$\xi^\star \approx \arg\min_{\xi} \sum_{\tau=t_0}^{t} \left\| m_\tau \odot \left[G_v^{\mathrm{id}}(x_\tau(\xi)) - y_\tau^{(v)} \right] \right\|_2^2, \qquad q_t^{\mathrm{id}} = x_t(\xi^\star). \tag{4}$$

where $t_0 = t - L + 1$ is the history-start time, $\xi \in \mathbb{R}^2$ is a candidate state at t_0, and $x_\tau(\xi)$ is its simulated state under the historical actions. The selected candidate $\xi^\star$ gives the current-state estimate q_t^{id}; m_τ, $y_\tau^{(v)}$, and v are the history mask, observation, and view. The squared Euclidean norm $\|\cdot\|_2^2$ sums visible-channel residuals. The approximation sign reflects finite candidate search and the regularization and ranking rules below. Multiple candidates are refined by regularized Gauss-Newton updates and ranked by historical fit with a small state-norm penalty. A validation-selected stable variant adds adaptive damping, backtracking, and a minimum-norm tie-break. Historical integration and future rollout use RK4. State estimation accesses neither future observations nor future action plans; candidate and numerical settings appear in Appendix A.

3.3 Encoder-Based Symbolic Prediction

A GRU processes masked observations, mask indicators, historical actions, and the regime descriptor; an MLP produces an eight-dimensional current state. A fixed feature library with trainable coefficients defines the latent dynamics:

$$z_t = E_\theta(\mathcal{H}_t, r), \qquad f_\phi(z, u, r) = \Psi(z, u, r) C_\phi. \tag{5}$$

where $z_t \in \mathbb{R}^8$ is the learned current state, θ denotes encoder parameters, and ϕ denotes vector-field parameters. The encoder E_θ receives history $\mathcal{H}_t$ and regime r; $\Psi(z, u, r)$ is a row of features of a learned state z, action u, and regime r, and C_ϕ is the coefficient matrix producing the eight-dimensional drift f_ϕ. The library includes constant, linear, quadratic, sine, action, regime, and interaction features, with no neural residual drift. Projected Euler updates bound the drift and state, and a view-conditioned MLP decodes future observations. These learned coordinates differ from the identification route's two-dimensional reference coordinates; both routes produce point estimates.

The training objective combines observation prediction, paired-view and all-view supervision, action-effect prediction, last-observation reconstruction, temporal consistency, and coefficient sparsity:

$$\begin{aligned} \mathcal{L} = \mathcal{L}_{\mathrm{pred}} + \lambda_{\mathrm{cv}}\mathcal{L}_{\mathrm{cv}} + \lambda_{\mathrm{av}}\mathcal{L}_{\mathrm{av}} + \lambda_{\mathrm{ae}}\mathcal{L}_{\mathrm{ae}} \\ + \lambda_{\mathrm{rec}}\mathcal{L}_{\mathrm{rec}} + \lambda_{\mathrm{temp}}\mathcal{L}_{\mathrm{temp}} + \lambda_1 \|C_\phi\|_{1,\mathrm{mean}}. \end{aligned} \tag{6}$$

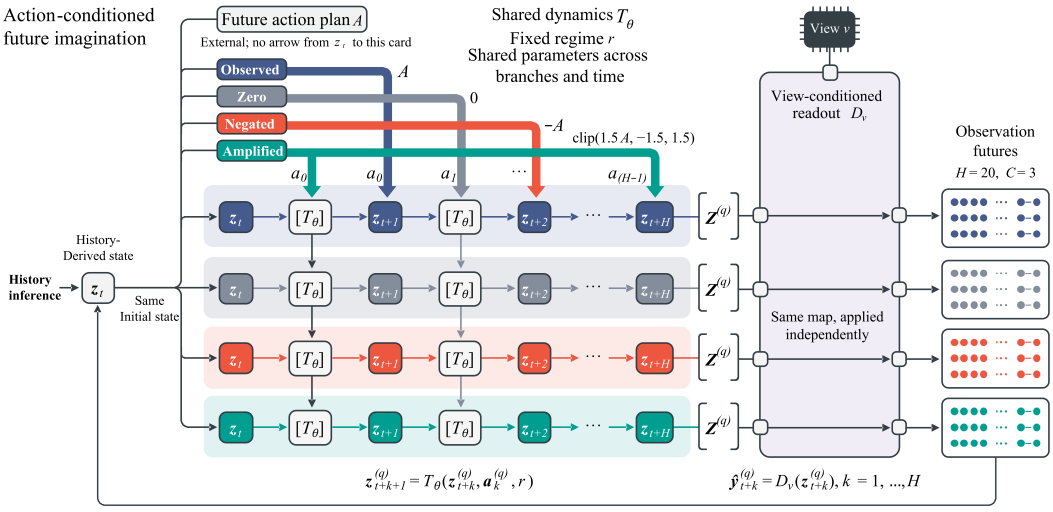

Figure 2: The observed, zero, negated, and amplified action plans start from the same history-derived state within each NS-WM route. Route-specific transitions and decoders produce one forecast per plan, after which fusion combines corresponding predictions across routes. Different action branches remain separate.

where $\mathcal{L}$ is the total training loss. The terms $\mathcal{L}_{\text{pred}}$, $\mathcal{L}_{\text{cv}}$, $\mathcal{L}_{\text{av}}$, $\mathcal{L}_{\text{ae}}$, $\mathcal{L}_{\text{rec}}$, and $\mathcal{L}_{\text{temp}}$ measure requested-view prediction, paired-view prediction, all-available-view prediction, action-effect prediction, last-observation reconstruction, and temporal consistency, respectively. Each corresponding λ is a nonnegative loss weight; λ_1 weights sparsity, and $\|C_\phi\|_{1,\text{mean}}$ is the mean absolute entry of the drift coefficient matrix. Target-availability masks exclude unavailable supervision. Matched action and zero-action branches supervise trajectory differences from the same history. Temporal consistency compares imagined states with a stop-gradient EMA encoder on observation-informed training windows; future observations enter only training supervision. Appendix A gives the transition, loss definitions, and training settings.

3.4 Action-Conditioned Rollouts and Prediction Fusion

Each route infers its current state once and reuses it across alternative future action plans (Figure 2). Route-specific transitions and decoders produce observation forecasts. The final forecast is

$$\widehat{Y}_{h,c}^{(b,v)} = (1 - w_{v,c})\widehat{Y}_{h,c}^{\text{id},(b,v)} + w_{v,c}\overline{Y}_{h,c}^{\text{enc},(b,v)}. \tag{7}$$

where $\widehat{Y}$ is the final forecast, $\widehat{Y}^{\text{id}}$ is the identified-route forecast, and $\overline{Y}^{\text{enc}}$ is the arithmetic mean of K frozen encoder-based symbolic forecasts. The index b identifies the future action branch, v identifies the sensor view, $h \in \{1, \ldots, H\}$ is the offset from current time t, and $c \in \{0, \ldots, C-1\}$ is an observation channel. The scalar $w_{v,c} \in [0, 1]$ weights the encoder ensemble; $1 - w_{v,c}$ weights identification. Convex least squares on validation targets fits $w_{v,c} \in [0, 1]$ once per view and channel. Weights remain fixed across test histories, horizons, and action plans; fusion combines corresponding forecasts without aligning latent coordinates or averaging action branches. Model and recipe selection use validation macro-domain MSE. Primary mixtures use $K = 3$ and additional-population mixtures use $K = 1$, with one supplementary condition selecting stable identification alone.

3.5 Extension to Heterogeneous Sequences

For shared-sensor sequences, channel-wise ridge experts predict future deviations from the last observation, using training-fitted conditional moments for missing channels and temporal coefficients shared across views. Regime-specific fits fall back to pooled coefficients for regimes without a regime-specific training fit; static validation weights indexed by regime, view, and channel combine these forecasts with the symbolic ensemble. Dysts instead uses a view-conditioned ridge expert with

history and action descriptors. Selected shared-sensor backbones use observed-state anchoring and an inverse-hyperbolic-sine (asinh) feature transform. These extensions include direct multi-horizon forecasts rather than assuming the controlled reference-state model; action-effect evaluation requires matched simulated branches. Expert details are in Appendix A.

4 EXPERIMENTS

4.1 EVALUATION PROTOCOL

Benchmarks and shifts. We evaluate eight development conditions across six suites: three controlled-pendulum conditions and one condition each from application sequences, Dysts, Tennessee Eastman Process (TEP), V2-all, and V2-synthetic. The controlled conditions withhold one of $(v, r) = (2, 1.0), (0, 0.7), (1, 1.0)$ while retaining both individual factors elsewhere in training. An additional pendulum population (data seed 2028) evaluates all six choices of withheld pair among three views and two training regimes. Each choice trains/selects its own predictor; this is not one model trained without all six pairs. Controlled histories have $L = 32$ steps, prediction horizon $H = 20$, and sampling interval $\Delta t = 0.05$. Each view has three channels, of which two are visible to the encoder and state fitter. Training regimes are $r \in \{0.7, 1.0\}$; parameter-OOD uses $r = 1.3$. Sensor noise has standard deviation 0.02. Training uses 14 episodes per regime and evaluation uses five. In-distribution (IID) evaluation retains the training view-regime pairings with independently generated episodes. Episode seeds differ across training, validation, IID, combination-OOD, and parameter-OOD splits; overlapping windows within an episode are not independent replicates. For each history, the simulator generates four plans: the observed plan $\mathcal{A}$, zero action, $-\mathcal{A}$, and $\mathrm{clip}(1.5\mathcal{A}, -1.5, 1.5)$, where scaling and clipping act on each action in the sequence, with clipping limits -1.5 and 1.5. The other suites retain the masked-history/future-observation interface. Sequence suites withhold view-source pairs; TEP withholds a manipulation-pattern-target pairing, using disjoint phase blocks for IID and separate manipulation files for OOD. TEP action descriptors encode scheduled manipulation metadata rather than measured actuator controls. Additional OOD tests change the dataset, system, or target and are distinguished from controlled parameter extrapolation (Appendix B).

Comparators and selection. Throughout the text, tables, and figures, S denotes the final NS-WM predictor, N denotes the neural world model, and D denotes the direct forecaster. N uses a recurrent state encoder, neural transition, and view decoder. D maps a recurrent history representation and the full future action plan to all horizon outputs. Both controlled baselines therefore receive action information. Primary N/D results average seeds 11, 22, and 33; S is one selected composite, generally averaging three symbolic checkpoints. Candidate selection uses macro-domain MSE averaged over the designated validation domains; mixture weights use validation targets. Architecture, ensemble size, identification priors, and search budgets differ, so this is a system-level comparison rather than a compute-matched architectural ablation. The additional population provides development evidence, not a prospectively blinded test; Appendix B details reporting and seed variation.

Metrics. Forecast MSE averages squared error over eligible target entries, including channels hidden in the input. For action effects, each nonzero plan is paired with the zero-action branch from the same history. We report

$$E_{\mathrm{effect}} = \mathrm{mean}\Big[(\Delta\widehat{Y}^{\mathcal{A}} - \Delta Y^{\mathcal{A}})^2\Big], \qquad \mathrm{skill} = 1 - \frac{E_{\mathrm{effect}}}{\mathrm{mean}[(\Delta Y^{\mathcal{A}})^2]}. \tag{8}$$

where $Y^{\mathcal{A}}$ and Y^0 are matched true future trajectories under nonzero plan $\mathcal{A}$ and the zero-action plan, and hats denote predictions from the same history. The differences $\Delta Y^{\mathcal{A}} = Y^{\mathcal{A}} - Y^0$ and $\Delta\widehat{Y}^{\mathcal{A}} = \widehat{Y}^{\mathcal{A}} - \widehat{Y}^0$ are true and predicted action effects. Squaring is elementwise, and mean averages over eligible branch pairs, forecast steps, and target channels. Thus E_{effect} is action-effect MSE, and skill normalizes it by the mean squared true effect, which must exceed the evaluation threshold 10^{-12}; otherwise skill is left undefined. Skill 1 is perfect; skill 0 equals predicting no action effect. The metric measures agreement with matched simulated branches under alternative action plans, rather than causal identification from observational data. The main composition test contains 240 histories and 720 dependent action comparisons. These are not independent statistical trials, and raw MSE is not pooled across differently scaled suites.

Table 1: IID and combination-OOD MSE across eight development conditions. S: NS-WM; N: neural world model; D: direct forecaster. Primary N/D entries are three-seed means; S is one selected composite. Bold denotes lowest MSE within a condition and split. Full precision and available standard deviations are provided in the supplementary ledger.

	IID			Combination-OOD		
Condition	S	N	D	S	N	D
Application	**0.1023**	0.1365	0.1358	**5.55e−8**	0.6762	0.0028
Pend. (2,1.0)	**0.0403**	0.2109	0.2704	**0.0797**	0.3342	0.1890
Pend. (0,0.7)	**0.0354**	0.1118	0.1284	**4.10e−4**	0.0267	0.0522
Pend. (1,1.0)	**0.0529**	0.1157	0.1018	**4.53e−4**	0.0337	0.0591
Dysts	**0.7190**	1.3943	1.2183	0.0116	0.0141	**0.0112**
TEP	**0.1953**	0.7960	0.3940	**0.1674**	0.1817	0.1991
V2-all	**0.0867**	0.1037	0.1033	**4.12e−7**	2.63e−4	1.53e−4
V2-synthetic	**0.2091**	0.2422	0.2462	**0.3717**	0.6626	0.5649

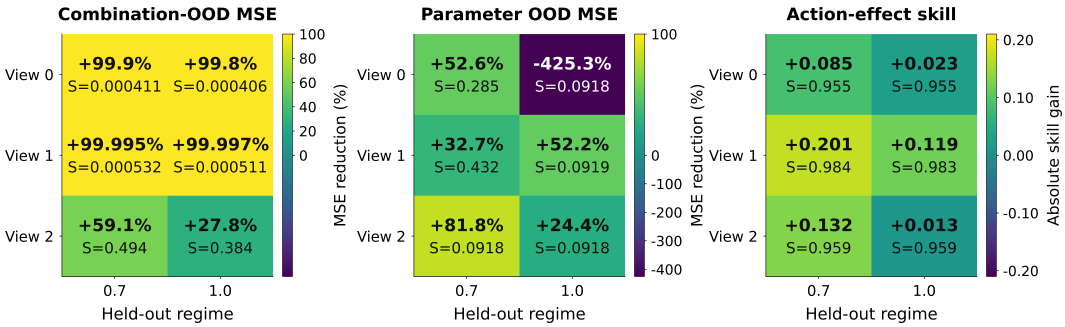

Figure 3: Generalization across all six held-out view-regime pairs in the additional population. Panels show combination-OOD MSE reduction, parameter-OOD MSE reduction, and action-effect skill gain relative to the stronger baseline. Positive values favor NS-WM.

4.2 Combination-OOD Generalization

Table 1 reports IID and combination-OOD MSE. NS-WM improves over both baselines in seven of eight primary composition conditions (87.5%) and all eight primary IID conditions. NS-WM reduces combination-OOD MSE by approximately 58% relative to the stronger baseline in the main pendulum condition, and by approximately 98% and 99% in the other two primary pendulum conditions, respectively. Together, these system-level comparisons support the proposed combination-OOD prediction framework; they do not isolate the contribution of each design choice. Heterogeneous suites extend this prediction interface through direct forecasting experts. Dysts is the exception: its combination-OOD MSE is approximately 4% higher than the direct baseline.

Across the additional six held-out pendulum pairs, NS-WM has the lowest combination-OOD MSE in all six, with reductions ranging from approximately 28% to more than 99.9% relative to the stronger baseline (Figure 3). Together, the primary and additional evaluations show lower combination-OOD MSE in 13/14 condition-population settings (92.9%), lower IID MSE in 14/14, and lower controlled parameter-OOD MSE in 8/9 (88.9%). These are evaluated settings, including repeated pairings under different populations, rather than fourteen distinct combinations. Appendix C reports every endpoint.

The main composition result also extends across individual histories. Averaging each history's error over four action branches, twenty steps, and all three channels, NS-WM outperforms both baselines on 194/240 histories (80.8%). Its median history MSE is approximately 86% lower than N and 73% lower than D. Appendix E shows all histories and prediction-target pairs, distinguishing this descriptive window-level analysis from the condition-level comparisons above.

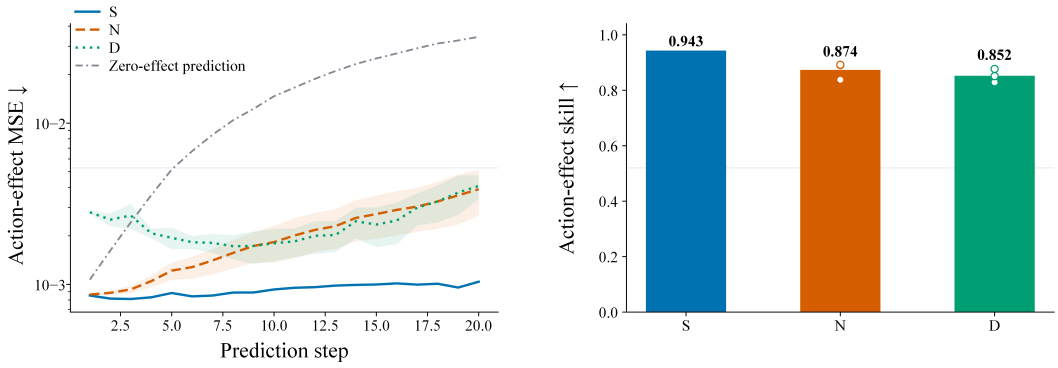

Figure 4: Main controlled-pendulum combination-OOD action effects over 240 histories and 720 nonzero-versus-zero branch pairs. Left: action-difference error by horizon. Right: aggregate skill. N/D shading and error bars reflect variation over three model seeds, not confidence intervals; S is one composite.

4.3 Predicting Changes Caused by Alternative Action Plans

In the main controlled condition, NS-WM reduces action-effect MSE by approximately 55% relative to N and 62% relative to D (Figure 4). Its action-effect skill is approximately 8% higher than N and 11% higher than D. These results show improved prediction of trajectory changes across prescribed actions, beyond the absolute accuracy of one nominal forecast. Figure 3 extends this result to the additional population: skill is highest in all six composition settings, with relative skill improvements of approximately 1%-26% over the stronger baseline. Including the three primary controlled settings, NS-WM improves both combination-OOD MSE and action-effect skill in all nine controlled condition-population settings. Within each route the history and inferred state are identical across plans; only the future action sequence changes. This observation is consistent with action-conditioned prediction using the recovered state. These evaluations concern predictive simulation and action effects; downstream policy learning and closed-loop control remain to be evaluated separately.

4.4 Recovered Dynamics and Their Predictive Roles

In the representative controlled identification fit for the main held-out pendulum condition ($v = 2, r = 1.0$) at seed-11, the prescribed library recovers four active terms and no additional terms:

$$\dot{z}_0 \approx z_1, \qquad \dot{z}_1 \approx -0.150 z_1 + 0.700 u - 0.999 r \sin z_0. \tag{9}$$

where $z_0 = x_1$ and $z_1 = x_2$ label the two reference-sensor coordinates used in the diagnostic figures, rather than components of the encoder state z_t. Dots denote time derivatives, u is the applied control, and r is the pendulum regime parameter. The four fitted terms represent kinematic evolution, damping, action input, and restoring dynamics, respectively. Coefficients are rounded for readability; the unrounded fit has relative coefficient errors below 0.14% (Figure 8) and is reported in Appendix D. This is recovery in the chosen reference-sensor coordinates and candidate library. To examine predictive dependence on these terms, we remove one term at a time while freezing all other coefficients, sensor maps, and recovered states. Removing the action term eliminates action-effect skill, a 100% reduction. Retaining the restoring term reduces combination-OOD MSE by approximately 99% relative to its removal. These frozen-component diagnostics show predictive reliance on the recovered terms; they are not training ablations of the final mixture (Appendix D).

Conditional vector fields. Figure 5 visualizes the identified dynamics under different actions and operating regimes while keeping the fitted coefficients fixed. Changing the action shifts the flow and the zero-acceleration contour, whereas changing the regime modifies the restoring dynamics. These conditional slices illustrate how a single symbolic transition represents action- and regime-dependent evolution, complementing the trajectory-based evaluations of compositional generalization and action-effect prediction. Appendix D further decomposes the transition into signed term contributions (Figure 7).

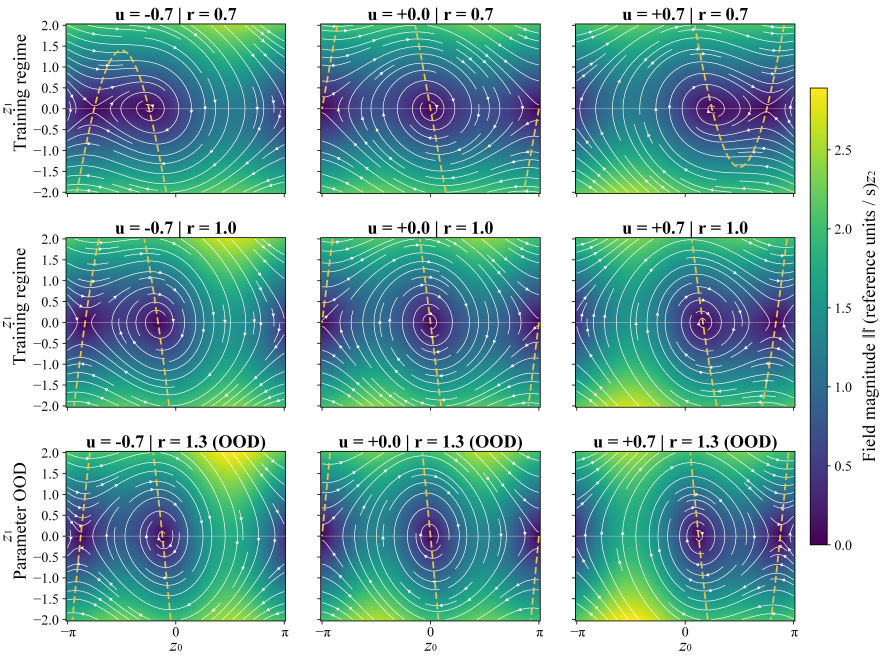

Figure 5: **Action- and regime-conditioned dynamics of the identified model.** Columns correspond to actions $u \in \{-0.7, 0, 0.7\}$ and rows to regimes $r \in \{0.7, 1.0, 1.3\}$, including the held-out parameter value $r = 1.3$. Colors indicate the magnitude of the predicted state derivative, white streamlines show the induced flow, and yellow curves mark zero predicted acceleration.

Historical fit and predictive ambiguity. Accurate equations do not by themselves resolve ambiguity in history-based state estimation. In a diagnosed parameter-OOD case, two candidate initial states separated by approximately 2π have nearly identical visible-history errors, yet one yields more than 99.9% lower hidden-future MSE than the other. This illustrates that historical consistency need not determine accurate future observations in a hidden sensor channel. Future targets are used only for post-hoc diagnosis, not for state selection. Appendix D provides the history-objective surfaces and phase-offset analysis.

For the main held-out sensor view under the observed-action branch, the final predictor also achieves lower average error than both baselines on each of the three channels. Appendix E.1 reports the corresponding errors across histories and forecast steps (Figure 17).

Component tradeoffs. Validation selection reduces macro-domain MSE by approximately 17% relative to identification alone. However, identification alone performs better on all three main controlled test splits, reducing combination-OOD MSE by approximately 99% relative to the final mixture (Table 4). Validation gains therefore do not guarantee gains on each test distribution, and the controlled results do not establish that fusion is necessary. These findings should be read under the supervision and reference-coordinate priors in Section 3; ensemble sizes and computational budgets also differ across systems.

5 Conclusion

NS-WM separates view-dependent sensing from shared symbolic dynamics and initializes these dynamics from partial histories for combination-OOD prediction. It improves combination-OOD MSE in $13/14$ evaluated condition-population settings and IID MSE in all fourteen, while improving both combination prediction and action-effect skill in all nine controlled settings. These results support reusable, action-conditioned prediction across the tested recombinations. Accurate equation recovery alone does not resolve state ambiguity, and validation-selected fusion does not consistently outperform its best constituent on test data.

AI USE STATEMENT

In this work, we used generative AI tools for manuscript language polishing, formula typesetting, and figure caption refinement. We did not use generative AI tools for the development of the core methodology, theoretical or mathematical claims, experimental design or implementation, or result analysis and interpretation; other research tasks requiring disclosure, such as synthetic data generation and qualitative data analysis, are not applicable to this work. Additionally, we used generative AI tools for preliminary literature-search assistance. We reviewed all AI-assisted work and manually cross-checked AI-polished text against our original arguments. All experiment-related descriptions were independently verified by two authors. We take full responsibility for the final content of this work, including any text, claims, or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This study evaluates NS-WM on simulated, synthetic, and publicly available benchmark trajectories. No new data are collected from human participants, and no experiments involving human subjects or deployed decision-making systems are conducted. The datasets are used for model training, validation, and evaluation within the experimental protocols described in this paper.

REPRODUCIBILITY STATEMENT

Section 3 specifies the two controlled prediction routes and validation mixtures. Appendix A provides optimization settings and inference steps; Appendix B describes splitting and reporting units. Appendix C reports the complete evaluation ledger, including adverse endpoints and available seed variation.

REFERENCES

Eloi Alonso, Adam Jelley, Vincent Micheli, Anssi Kanervisto, Amos Storkey, Tim Pearce, and François Fleuret. Diffusion for world modeling: Visual details matter in Atari. In *Advances in Neural Information Processing Systems*, 2024.

Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016a. doi: 10.1073/pnas.1517384113.

Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Sparse identification of nonlinear dynamics with control (SINDYc). *arXiv preprint arXiv:1605.06682*, 2016b.

Kathleen Champion, Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116 (45):22445–22451, 2019. doi: 10.1073/pnas.1906995116.

David Ha and Jürgen Schmidhuber. World models. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pp. 2555–2565. PMLR, 2019.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations*, 2020.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*, 2023.

Hyeongjoo Hwang, Seokin Seo, Youngsoo Jang, Sungyoon Kim, Geon-Hyeong Kim, Seunghoon Hong, and Kee-Eung Kim. Information-theoretic state space model for multi-view reinforcement learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 14249–14282. PMLR, 2023.

Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1–2):99–134, 1998. doi: 10.1016/S0004-3702(98)00023-X.

Maximilian Karl, Maximilian Soelch, Justin Bayer, and Patrick van der Smagt. Deep variational Bayes filters: Unsupervised learning of state space models from raw data. In *International Conference on Learning Representations*, 2017.

Thomas Kipf, Elise van der Pol, and Max Welling. Contrastive learning of structured world models. In *International Conference on Learning Representations*, 2020.

Matthieu Kirchmeyer, Yuan Yin, Jeremie Dona, Nicolas Baskiotis, Alain Rakotomamonjy, and Patrick Gallinari. Generalizing to new physical systems via context-informed dynamics model. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 11283–11301. PMLR, 2022.

Rahul G. Krishnan, Uri Shalit, and David Sontag. Deep Kalman filters. *arXiv preprint arXiv:1511.05121*, 2015.

Anson Lei, Bernhard Schölkopf, and Ingmar Posner. Variational causal dynamics: Discovering modular world models from interventions. *arXiv preprint arXiv:2206.11131*, 2022.

Michael L. Littman, Richard S. Sutton, and Satinder Singh. Predictive representations of state. In *Advances in Neural Information Processing Systems*, volume 14, 2001.

Peter Y. Lu, Joan Ariño Bernad, and Marin Soljačić. Discovering sparse interpretable dynamics from partial observations. *Communications Physics*, 5:206, 2022. doi: 10.1038/s42005-022-00987-z.

Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9:4950, 2018. doi: 10.1038/s41467-018-07210-0.

Vincent Micheli, Eloi Alonso, and François Fleuret. Transformers are sample-efficient world models. In *International Conference on Learning Representations*, 2023.

C. V. Rao, J. B. Rawlings, and D. Q. Mayne. Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. *IEEE Transactions on Automatic Control*, 48(2):246–258, 2003. doi: 10.1109/TAC.2002.808470.

Xinyue Wang and Biwei Huang. Modeling unseen environments with language-guided composable causal components in reinforcement learning. In *International Conference on Learning Representations*, 2025.

Linfeng Zhao, Lingzhi Kong, Robin Walters, and Lawson L. S. Wong. Toward compositional generalization in object-oriented world modeling. *arXiv preprint arXiv:2204.13661*, 2022.

A Implementation and Inference

Controlled neural backbone. The main encoder uses a GRU with hidden width 128 and an eight-dimensional state, followed by view-specific neural readout. Inputs concatenate masked channel values, mask bits, incoming actions, and regime values. AdamW uses learning rate 10^{-3}, weight decay 10^{-5}, batch size 512, and gradient-norm clipping at 5. The main training limit is 160 epochs with patience 35. Temporal training uses five warmup epochs followed by a ten-epoch ramp, four imagined-state anchors, and EMA decay 0.99. Reconstruction concerns the last history observation, not reconstruction of every historical sample. The temporal teacher's observation-informed windows are used during training only. Other suite backbones use the configurations identified in the accompanying selection manifest; the main pendulum weights are not asserted to apply to every suite. The main loss weights are $(\lambda_{\mathrm{cv}}, \lambda_{\mathrm{av}}, \lambda_{\mathrm{ae}}, \lambda_{\mathrm{rec}}, \lambda_1) = (0.2, 0.2, 0.2, 0.2, 3 \times 10^{-4})$, with temporal weight ramping to 0.03. Direct latent alignment and functional-consistency weights are zero. Temporal normalization computes each teacher coordinate's standard deviation jointly over the batch and horizon-anchor dimensions, floored at 0.1.

Sparse regression and coordinate access. Only training observation targets enter coefficient estimation. The first two reference-view channels define two coordinates; this is a stronger observation prior than an unconstrained latent state. Paired observations fit a view-2-to-reference affine bridge where that view supplies trajectories for coordinate estimation. Observation features comprise ten constant, linear, quadratic, sine, and cosine terms. Dynamics add seven action/regime features, giving seventeen candidates per derivative coordinate. A nine-sample degree-three local polynomial provides smoothing and derivatives; four boundary samples are omitted at each end. Features are normalized by RMS, ridge regularization is 10^{-6}, and sequential thresholding uses at most five refits. Candidate relative thresholds include 0.03 and 0.1; the representative primary fit selects 0.03. A coefficient's contribution is compared to target scale, rather than thresholding unnormalized coefficients indiscriminately.

Initial-state optimization. The original fitter scans the first coordinate on $[-20, 20]$ with spacing 0.25. The second coordinate is initialized from a local history slope and clipped to $[-8, 8]$. Candidate initialization uses the first three history observations, retains eight separated starts, and applies ten Gauss-Newton updates. The finite-difference step is 0.002, damping is 10^{-3}, and each coordinate update is clipped to $[-0.5, 0.5]$. Candidate ranking uses the squared masked residual averaged over all history/channel positions, including masked zeros, plus $10^{-7}\|\xi\|_2^2$, where ξ is the history-start candidate and $\|\cdot\|_2$ is the Euclidean norm. RK4 uses the known sampling interval, with the supplied action held constant within each step.

The stable variant uses sixteen starts, adaptive damping, and line-search scales $1, 0.5, 0.25$. Among candidates within 10^{-5} of the best history residual, it prefers smaller state norm. The selected additional-population $(v, r) = (0, 1.0)$ model is this standalone variant; the primary ambiguity composites use the original fitter. Neither fitter uses future observation error to select a state.

Controlled prediction procedure.

1. Train symbolic encoders/decoders on available paired targets. Independently fit reference-coordinate dynamics and sensor maps using training observations.
2. Freeze each fitted predictor. On validation histories, obtain forecasts, estimate convex view/channel mixture weights, and select a candidate by macro-domain validation MSE.
3. For a test history, encode the current state in each neural backbone. In the identified route, initialize multiple history-start candidates, integrate historical controls, optimize their masked residuals, and advance the selected fit to the current time.
4. For each of the four future plans, reuse the respective current states, evolve them under that plan, and decode the states into the requested sensor view.
5. Average the frozen backbone forecasts and combine them with the identified forecast using the fixed selected weight. Evaluate future targets only after predictions have been formed.

Sequence experts. For channel c, a channel expert constructs features from the last value and the preceding values centered at that last value. Ridge regression predicts the future change, which is

added back to the last value. With standardized features and an intercept,

$$\widehat{y}_{t+1:t+H,c} = y_{t,c}\mathbf{1}_H^\top + a_c^\top B_{r,c}, \tag{10}$$

$$B_{r,c} = \arg\min_B n_{r,c}^{-1}\|X_{r,c}B - T_{r,c}\|_F^2 + \alpha\|B_{\text{slope}}\|_F^2 + 10^{-9}\|B_{\text{intercept}}\|_F^2. \tag{11}$$

where $\widehat{y}_{t+1:t+H,c}$ is the row of H future predictions for channel c, $y_{t,c}$ is its last value used by the expert, and $\mathbf{1}_H$ is a length-H column of ones. The column a_c contains the standardized history features and intercept; $\top$ denotes transpose. For regime r and channel c, $X_{r,c}$ stacks $n_{r,c}$ training feature rows, $T_{r,c}$ stacks their future deviations from the last value, and $B_{r,c}$ is the fitted coefficient matrix with H output columns. In the optimization, B is a candidate matrix, B_{slope} and $B_{\text{intercept}}$ are its feature and intercept rows, α is the ridge penalty, and $\|\cdot\|_F$ is the Frobenius norm. Missing channel values are imputed as described below. Regimes with training fits use their fitted coefficients; other regimes use pooled coefficients. Training-fitted conditional moments of channel levels and five-step differences impute missing sensor channels. The coefficients are shared across views. Validation mixtures are fitted by regime, view, and channel, with pooled support when a group has fewer than five examples. The selected penalties are 0.01 for application, 10^{-4} for TEP, 10^{-4} for V2-all, and 0.01 for V2-synthetic.

Dysts instead uses a view-conditioned ridge predictor with $\alpha = 10^{-4}$. Its features include the last masked observation, flattened centered history, and supplied past/future action descriptors; per-regime fits fall back to a pooled fit. These are direct multi-horizon corrections, not additional identified physical transitions. The final selection manifest records which candidate won for each condition.

A.1 Symbolic Transitions and Training Objectives

The identification observation library $\Phi(x)$ contains constant, linear, quadratic, sine, and cosine terms. Its controlled dynamics library is

$$\Theta(x, u, r) = [\Phi(x), u, r, rx_1, rx_2, r\sin x_1, r\sin x_2, ru]. \tag{12}$$

where $x = (x_1, x_2)$ is the reference-sensor state, u is the control input, and r is the regime descriptor. The brackets concatenate the sensor features $\Phi(x)$ with seven scalar action, regime, and interaction features to form the dynamics feature row $\Theta(x, u, r)$; juxtaposition denotes multiplication and $\sin$ acts on the indicated coordinate. The encoder library Ψ contains constant, linear, quadratic, and sine features, actions and the regime descriptor, and state-action, state-regime, sine-regime, and action-regime interactions. The controlled update is

$$z_{\tau+1} = \Pi_{[-10,10]}\left[z_\tau + \Delta t\,\beta\tanh\left(f_\phi(z_\tau, u_{\tau+1}, r)/\beta\right)\right], \qquad \beta = 1/\Delta t. \tag{13}$$

where z_τ is the learned state at time τ, $u_{\tau+1}$ is the incoming action, r is the regime descriptor, and $\Delta t > 0$ is the sampling interval. The scalar β sets the drift bound; $\tanh$ acts elementwise, and $\Pi_{[-10,10]}$ clips each updated state coordinate to $[-10, 10]$. The vector field f_ϕ is defined in Equation 5; β is distinct from action-branch index b. A view embedding and MLP implement $\widehat{y}_{\tau+1}^{\text{enc},(v)} = D_{\psi,v}(z_{\tau+1})$, where $D_{\psi,v}$ is the decoder with parameters ψ conditioned on sensor view v, and $\widehat{y}_{\tau+1}^{\text{enc},(v)}$ is its predicted observation from state $z_{\tau+1}$.

For target-availability mask Q, distinct from history-input mask m, the masked prediction loss is

$$\ell_Q(\widehat{Y}, Y) = \frac{\sum Q \odot (\widehat{Y} - Y)^2}{\sum Q}. \tag{14}$$

where $\widehat{Y}$ and Y are matching arrays of predicted and target observations, Q is a binary mask with one for an available target and zero otherwise, and ℓ_Q is the corresponding mean squared error. The operator $\odot$ and the square act elementwise, both sums run over all scored entries, and the denominator counts available entries; the expression applies when this count is positive. Prediction supervision scores the requested view; cross-view and all-view terms score paired and all available view targets, respectively. Unavailable view-regime targets are masked out. Reconstruction scores the final history observation in the input and paired views. For matched action and zero-action targets,

$$\mathcal{L}_{\text{ae}} = \ell_Q(\widehat{Y}^{\mathcal{A}} - \widehat{Y}^0, Y^{\mathcal{A}} - Y^0). \tag{15}$$

where $\mathcal{L}_{\mathrm{ae}}$ is the action-effect training loss, $\widehat{Y}^{\mathcal{A}}$ and $\widehat{Y}^{0}$ are predicted trajectories under an action plan $\mathcal{A}$ and zero action, and $Y^{\mathcal{A}}$ and Y^{0} are their paired targets from the same history. The mask Q selects available paired target entries in the previously defined loss ℓ_Q. Both branches originate from the same history. Temporal consistency uses coordinate-normalized Smooth-L_1 loss between imagined states at selected anchors and a stop-gradient EMA encoder. Its observation-informed training windows preserve the input sensor mask.

For validation examples in view v, the static fusion weight is

$$w_{v,c} = \mathrm{clip}_{[0,1]}\left(\frac{\sum_{i,h}\delta_{i,h,c}e_{i,h,c}}{\max(\sum_{i,h}\delta_{i,h,c}^2, 10^{-12})}\right), \tag{16}$$

where i indexes validation forecast examples in view v, h indexes future steps, and c indexes channels. The quantities $\delta_{i,h,c} = \overline{Y}^{\mathrm{enc}}_{i,h,c} - \widehat{Y}^{\mathrm{id}}_{i,h,c}$ and $e_{i,h,c} = Y_{i,h,c} - \widehat{Y}^{\mathrm{id}}_{i,h,c}$ are the ensemble-minus-identification prediction difference and identification residual, respectively; Y denotes the validation target. The operator $\mathrm{clip}_{[0,1]}$ restricts the scalar weight $w_{v,c}$ to the convex interval, and $\max$ floors its denominator at 10^{-12}. For a positive unfloored denominator, minimizing $\sum_{i,h}(e_{i,h,c} - w_{v,c}\delta_{i,h,c})^2$ gives the displayed numerator-to-denominator ratio; clipping imposes the interval constraint. The numerical floor stabilizes the ratio when the predictors nearly coincide. The same fitted weight applies to every action branch.

B Data Protocols and Reporting

Controlled episodes have rollout length 240. With history 32, horizon 20, and stride 4, the implemented window loop yields 48 histories per episode. Five view-regime pairs are available for training and IID evaluation, and one is reserved for composition evaluation. Four action branches produce 13,440 training rows, 4,800 validation/IID rows per split, 960 composition rows, and 2,880 parameter-OOD rows. These row counts include repeated histories under different action branches. Training and evaluation use independent episode generation, with seed offsets 0, 101, 202, 303, and 404 for training, validation, IID, composition, and parameter splits, respectively.

The six suites correspond to the repository's application, controlled-pendulum, Dysts, TEP, V2-all, and V2-synthetic builders. Application and V2-all hold out view 2 with PEMS-SF traffic, with NASA C-MAPSS FD001 as dataset OOD. V2-synthetic holds out view 2 with Van der Pol ($\mu = 3$), where μ is the nonlinear damping parameter of the Van der Pol oscillator, and uses the Lorenz63 hidden-z sequence as dataset OOD. Here z names the third physical coordinate of Lorenz63, not the encoder state z_t. Dysts withholds view 2 with ForcedVanDerPol and tests DoublePendulum as system OOD; its supplied controls are zero, not recorded interventions. Continuous sequence splits use chronological fractions 0.65 for training and 0.15 for validation, respecting episode boundaries where defined.

TEP differs: composition excludes the line-pattern manipulation of target s1, and target OOD excludes target a11. The event protocol divides each available episode into pre-event, active, and recovery phases and allocates disjoint chronological blocks within each eligible phase to training, validation, and IID testing. Short phases that cannot support three nonoverlapping windows are reserved for testing. Composition and target OOD use the excluded manipulation files. TEP action descriptors encode known scheduled manipulations, not measured actuator controls. Thus its compositional factors are manipulation pattern and target, not a withheld sensor-regime pair. These suite names denote implemented data protocols, not six identical physical intervention environments.

Primary N/D entries are means over three model seeds, with the recorded standard deviations in the complete tables below. S is one final composite, which may itself average three backbones. A baseline standard deviation therefore cannot be assigned to S, and the difference between entries is not a significance test. Additional-population N/D entries use the recorded single-seed comparison. No independent full-pipeline uncertainty estimate is available. The new population differs from the primary data seed, but was inspected during development and is not claimed as a blind confirmatory study.

C Complete Evaluation Results

Tables 2 and 3 list all 51 recorded endpoints from the final comparison ledger. MSE is lower-is-better; action-effect skill is higher-is-better. Other OOD follows the suite's dataset, system, or event split and is not automatically parameter OOD. The main text table and these tables are generated from the same structured source. Primary baseline uncertainty is variation over model seeds, not uncertainty over sampled physical systems.

Table 2: Complete primary evaluation. N and D show mean ± standard deviation over three model seeds; S is one composite. Bold marks the best entry in each row; lower is better for MSE and higher is better for action-effect skill.

Condition	Split / metric	S	N	D
Application	IID	**0.1023**	0.1365 ± 0.0088	0.1358 ± 0.0015
Application	Combination	**5.55e−8**	0.6762 ± 1.0464	0.0028 ± 0.0015
Application	Dataset	**37.6248**	46.4360 ± 3.7692	47.7619 ± 0.4701
Pend. (2,1.0)	IID	**0.0403**	0.2109 ± 0.0901	0.2704 ± 0.1366
Pend. (2,1.0)	Combination	**0.0797**	0.3342 ± 0.1476	0.1890 ± 0.1390
Pend. (2,1.0)	Comb. skill	**0.9432**	0.8735 ± 0.0308	0.8523 ± 0.0241
Pend. (2,1.0)	Parameter	**0.1685**	0.2383 ± 0.0382	0.2395 ± 0.0100
Pend. (0,0.7)	IID	**0.0354**	0.1118 ± 0.0455	0.1284 ± 0.0191
Pend. (0,0.7)	Combination	**4.10e−4**	0.0267 ± 0.0050	0.0522 ± 0.0142
Pend. (0,0.7)	Comb. skill	**0.9514**	0.8823 ± 0.0488	0.6884 ± 0.0377
Pend. (0,0.7)	Parameter	**0.0331**	0.0985 ± 0.0480	0.0987 ± 0.0198
Pend. (1,1.0)	IID	**0.0529**	0.1157 ± 0.0140	0.1018 ± 0.0134
Pend. (1,1.0)	Combination	**4.53e−4**	0.0337 ± 0.0143	0.0591 ± 0.0143
Pend. (1,1.0)	Comb. skill	**0.9622**	0.9196 ± 0.0194	0.8322 ± 0.0591
Pend. (1,1.0)	Parameter	**0.0036**	0.0301 ± 0.0138	0.0384 ± 0.0027
Dysts	IID	**0.7190**	1.3943 ± 0.7180	1.2183 ± 0.4635
Dysts	Combination	0.0116	0.0141 ± 0.0022	**0.0112 ± 2.64e−4**
Dysts	System	0.0039	0.0067 ± 0.0022	**0.0033 ± 6.21e−5**
TEP	IID	**0.1953**	0.7960 ± 0.1339	0.3940 ± 0.0621
TEP	Combination	**0.1674**	0.1817 ± 0.0016	0.1991 ± 0.0143
TEP	Target	**2.1012**	2.9662 ± 0.1249	2.6196 ± 0.0955
V2-all	IID	**0.0867**	0.1037 ± 0.0037	$0.1033 \pm 9.75\text{e}{-4}$
V2-all	Combination	**4.12e−7**	$2.63\text{e}{-4} \pm 2.43\text{e}{-4}$	$1.53\text{e}{-4} \pm 6.37\text{e}{-5}$
V2-all	Dataset	**275.4577**	297.5955 ± 14.1140	313.7709 ± 4.6048
V2-synthetic	IID	**0.2091**	0.2422 ± 0.0374	0.2462 ± 0.0399
V2-synthetic	Combination	**0.3717**	0.6626 ± 0.3381	0.5649 ± 0.1780
V2-synthetic	Dataset	32.6610	**30.7119 ± 0.9988**	31.4071 ± 0.8818

Table 3: Complete additional-population evaluation (data seed 2028). Each condition is a separately selected model; no full-pipeline standard deviations are available. Bold marks the best entry in each row; lower is better for MSE and higher is better for action-effect skill.

Condition	Split / metric	S	N	D
Pend. (2,1.0)	IID	**0.0555**	3.8316	4.3608
Pend. (2,1.0)	Combination	**0.3843**	0.5325	0.5714
Pend. (2,1.0)	Comb. skill	**0.9594**	0.9464	0.8639
Pend. (2,1.0)	Parameter	**0.0918**	0.1581	0.1216
Pend. (0,0.7)	IID	**0.0621**	0.1200	0.1274
Pend. (0,0.7)	Combination	**4.11e−4**	1.1604	0.4435
Pend. (0,0.7)	Comb. skill	**0.9555**	0.8701	0.4845

Table 4: Main controlled-condition component MSE. Bold marks the best component in each split. Identification alone is best on these test splits; the validation-selected mixture improves on the symbolic ensemble.

Split	Identification	Symbolic ensemble	Final mixture
IID	**0.000498**	0.13833	0.04035
Combination-OOD	**0.000430**	0.09302	0.07971
Parameter OOD	**0.000419**	0.18415	0.16849

Table 5: Frozen term removal from the identified component, evaluated on the main combination split. Bold marks the best value in each metric column. No coefficient refitting or state re-estimation follows removal.

Variant	Forecast MSE	Effect MSE	Effect skill
Intact	**4.30e−4**	**7.93e−4**	**0.9514**
No kinematic	0.1124	0.0043	0.7375
No damping	0.0020	8.53e−4	0.9477
No action	0.0121	0.0163	0.0000
No restoring	0.0504	0.0010	0.9380

Condition	Split / metric	S	N	D
Pend. (0,0.7)	Parameter	**0.2853**	0.7816	0.6015
Pend. (0,1.0)	IID	**0.0555**	4.7136	4.4267
Pend. (0,1.0)	Combination	**4.06e−4**	0.2448	0.2454
Pend. (0,1.0)	Comb. skill	**0.9554**	0.9320	0.9301
Pend. (0,1.0)	Parameter	0.0918	**0.0175**	0.0279
Pend. (1,0.7)	IID	**0.0574**	0.0618	0.0644
Pend. (1,0.7)	Combination	**5.32e−4**	20.8471	11.7606
Pend. (1,0.7)	Comb. skill	**0.9839**	0.7826	0.6697
Pend. (1,0.7)	Parameter	**0.4319**	0.6414	0.8272
Pend. (1,1.0)	IID	**0.0555**	4.4327	5.5930
Pend. (1,1.0)	Combination	**5.11e−4**	17.8987	24.0548
Pend. (1,1.0)	Comb. skill	**0.9833**	0.7971	0.8642
Pend. (1,1.0)	Parameter	**0.0919**	0.1924	0.2669
Pend. (2,0.7)	IID	**4.20e−4**	0.0394	0.0123
Pend. (2,0.7)	Combination	**0.4937**	1.3643	1.2072
Pend. (2,0.7)	Comb. skill	**0.9593**	0.8268	0.7069
Pend. (2,0.7)	Parameter	**0.0918**	0.5034	0.5078

D Component and Failure Analyses

The unrounded representative identified transition is

$$\dot{z}_0 = 0.999914 z_1, \qquad \dot{z}_1 = -0.149792 z_1 + 0.699646 u - 0.999132 r \sin z_0. \tag{17}$$

where $z_0 = x_1$ and $z_1 = x_2$ are the first and second reference-sensor coordinates, dots denote time derivatives, u is the control input, and r is the pendulum regime parameter. The displayed numbers are fitted kinematic, damping, action, and restoring coefficients in the prescribed library.

Validation mixture. The main mixture reduces validation macro MSE by approximately 17% relative to identification alone, but does not improve each test split over the best constituent. On combination-OOD, the mixture reduces MSE by approximately 14% relative to the symbolic ensemble, while identification alone reduces MSE by approximately 99% relative to the mixture. Thus the full-system comparison against N/D does not establish that fusion is necessary for the controlled

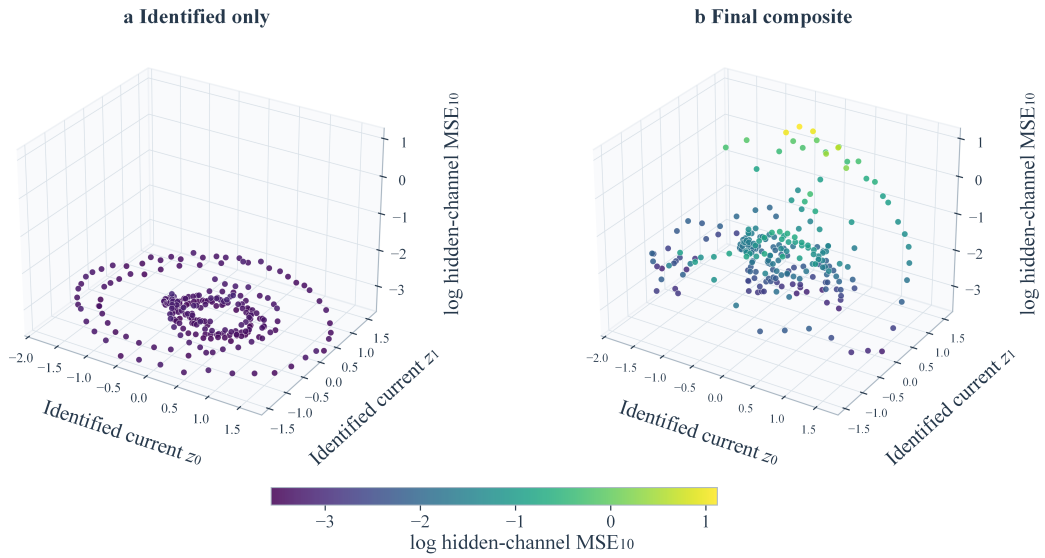

Figure 6: **Hidden-channel forecast errors across reference-state coordinates.** Each panel retains all 240 histories and uses common coordinate limits and a shared error color scale. Horizontal coordinates are the current-state estimates from the identified component in both panels. Height and color encode $\log_{10}$ hidden-channel forecast MSE, averaged over four action plans and twenty steps. This compares the identified component with the final composite, not with the neural and direct baselines. Overlapping windows are not independent replicates.

gains. For the main controlled model, the symbolic-backbone weights by view and channel are

$$W_{\text{enc}} = \begin{bmatrix} 0 & 0.0066995 & 0 \\ 0 & 0 & 0 \\ 1 & 0.0069495 & 0 \end{bmatrix}, \qquad W_{\text{id}} = \mathbf{1}_{3\times 3} - W_{\text{enc}}. \tag{18}$$

where W_{enc} contains the encoder-ensemble weights $w_{v,c}$, W_{id} contains the complementary identification weights, and $\mathbf{1}_{3\times 3}$ is the all-ones matrix. Rows follow views 0, 1, and 2, and columns follow channels 0, 1, and 2; subtraction is elementwise. These are static fitted weights; they are not an input-dependent gating network. Their validation benefit does not imply that the final mixture dominates the identified component on the test population. Table 4 makes that distinction explicit.

Prediction error across reference-state coordinates. Figure 6 locates hidden-channel forecast errors in the identified component's current-state coordinates for all 240 histories in the main combination-OOD split. Each error averages four action plans and twenty forecast steps. The identified component has lower, more concentrated errors in this cohort, whereas the final composite exhibits a wider error range, consistent with Table 4. Both panels use the same identified coordinates for comparison; these are not latent coordinates of the final composite. The plot is a descriptive association, not a causal attribution of prediction error to state coordinates.

Frozen term removal. The coefficient recovery and term-removal analyses concern the identified component. A removed term is set to zero while all other parameters and the already recovered current states are kept fixed. No retraining or new state fitting follows removal. Table 5 thus measures local reliance of this fitted predictor on a term; it is not a training ablation of the full NS-WM system.

Signed term contributions. Figure 7 shows how the recovered terms act along a rollout. At prediction intervals 1, 10, and 20, the action term contributes 0.239, 0.263, and 0.279 to the second-coordinate derivative, while the restoring term contributes 0.198, 0.175, and 0.054. The damping contribution changes sign as the coordinate velocity changes, and the signed terms sum to the displayed derivative. This gives an inspectable account of the fitted transition at concrete rollout states, complementing the coefficient-recovery and frozen-removal results.

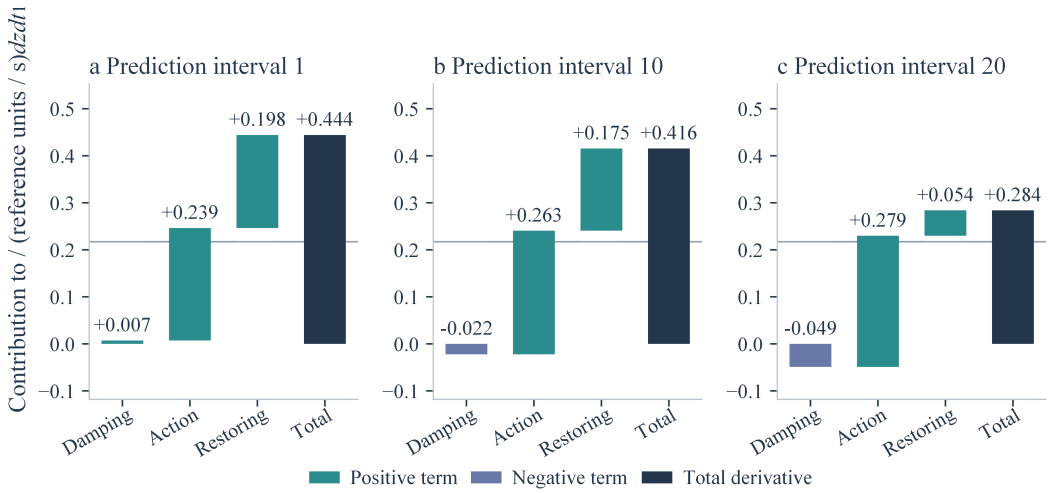

Figure 7: Signed contributions of recovered dynamics terms at three prediction intervals. Each contribution is coefficient times basis value at the interval start, and the total is the instantaneous derivative, not an RK4 increment. The fixed example is selected by median true action-effect energy without using model errors. This explains the identified component's transition; it is not a training ablation.

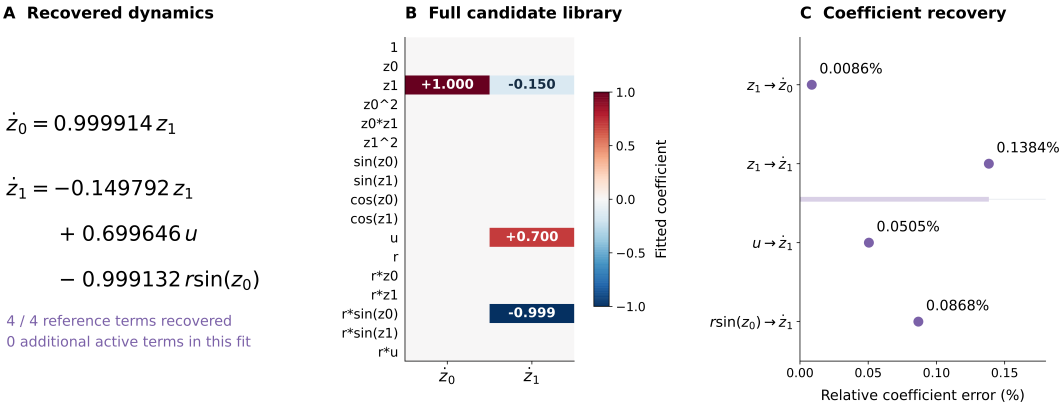

Figure 8: Recovered controlled dynamics in the prescribed reference-coordinate library. All four active terms are recovered, without additional active terms, and relative coefficient errors are below 0.14%.

Geometry of the history objective. Figure 11 gives a three-dimensional rendering of the history objectives for neighboring windows starting at 120 and 124 in the diagnosed parameter-OOD episode. Repeated low-loss valleys reveal the approximately periodic ambiguity in the history-start coordinate. The vertical axis is $\log_{10}$ visible-history MSE under fixed fitted dynamics and sensor maps, not a neural training loss or a posterior probability. The surfaces complement the candidate-error comparison and phase-offset diagnostic reported below.

Periodic ambiguity. In a diagnosed parameter-OOD failure, five neighboring windows from one episode account for more than 99.9% of view-2/channel-0 squared error. Their recovered coordinate differs by approximately -2π. A post-hoc alias comparison gives nearly equal historical residuals, but shifting the candidate by 2π reduces future observation MSE by more than 99.9%. This uses future targets only to diagnose the failure; it is not the inference selection rule. It demonstrates that a small masked-history residual need not uniquely determine the future in the requested sensor coordinates.

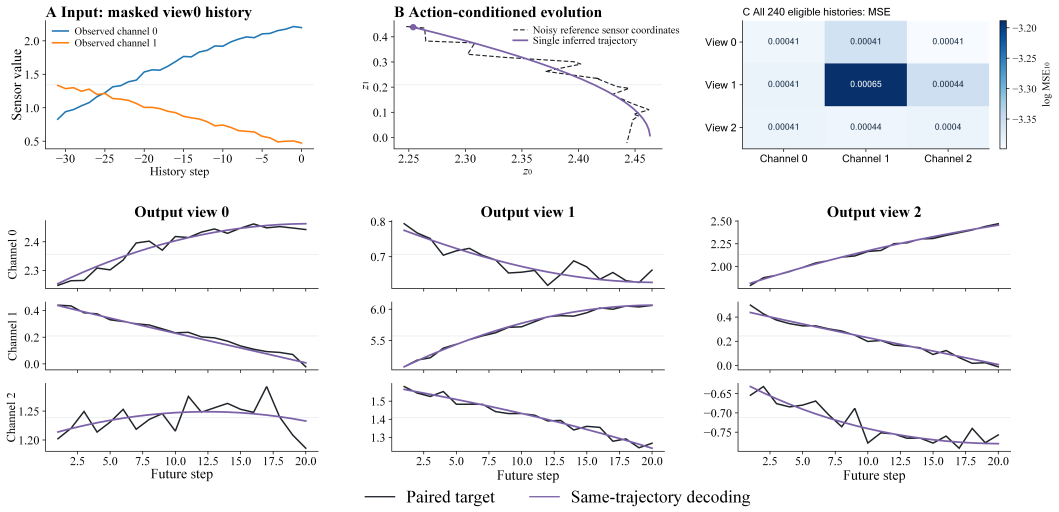

Figure 9: Selected trajectory visualization across sensor views. The plotted history/trajectory illustrates view-dependent decoding; aggregate performance is given in the numerical tables.

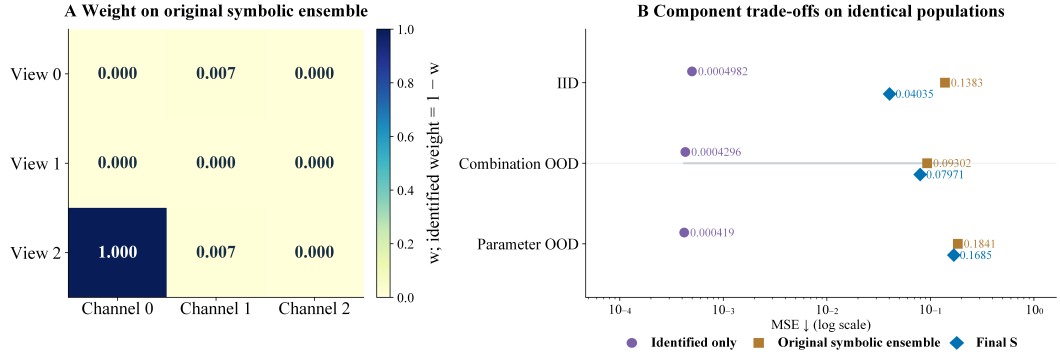

Figure 10: Component predictions and validation-selected weights. The identified component, backbone ensemble, and full mixture are distinct estimators; improved validation error does not ensure improvement over each component on every test split.

E Additional Evidence for Predictive Generalization

The additional plots reuse frozen predictions and the final evaluation ledger; they introduce no new training or independent test population. Across all recorded condition/split/metric endpoints, NS-WM is better than both baselines on $47/51$ (92.2%): $24/27$ primary and $23/24$ additional-population endpoints. An endpoint is a condition-specific metric, not a unique sensor-regime combination. The main text therefore emphasizes the separately computed $13/14$ composition-setting result and $9/9$ controlled action-effect result.

Complete endpoint effects. Figures 13 and 14 retain every recorded endpoint, including negative gains. On the primary Dysts composition test, S has approximately 4% higher MSE than D. Dysts system OOD and V2-synthetic dataset OOD also favor a baseline. In the additional $(v, r) = (0, 1.0)$ condition, S has approximately 425% higher parameter-OOD MSE than N. These values qualify extrapolation claims while leaving the composition and action-effect success counts unchanged.

History-level accuracy. Figure 15 shows all 240 histories in the main composition split. Each error averages four branches and twenty steps over the indicated channels. N/D average seed-specific errors over seeds 11, 22, and 33; they do not average predictions before scoring. The final S predictor

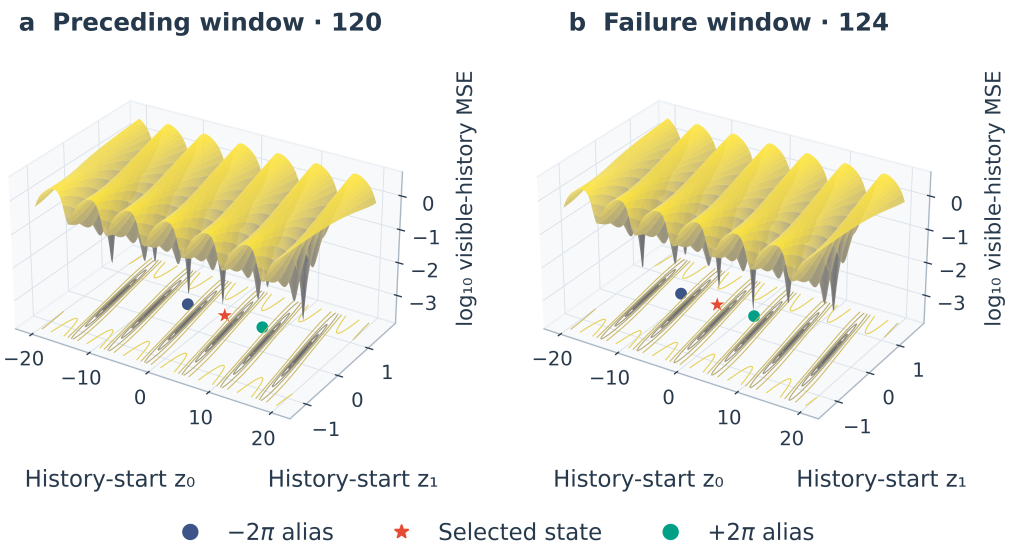

Figure 11: **Geometry of history-state ambiguity.** The preceding and failure windows use the same fixed-grid history-objective evaluation. Surface rendering subsamples the grid to 101×91 vertices; projected contours use the full 401×181 grid. Markers locate the selected candidate and its $\pm 2\pi$ aliases at their evaluated historical losses. These are diagnostic candidates, not samples from a state posterior.

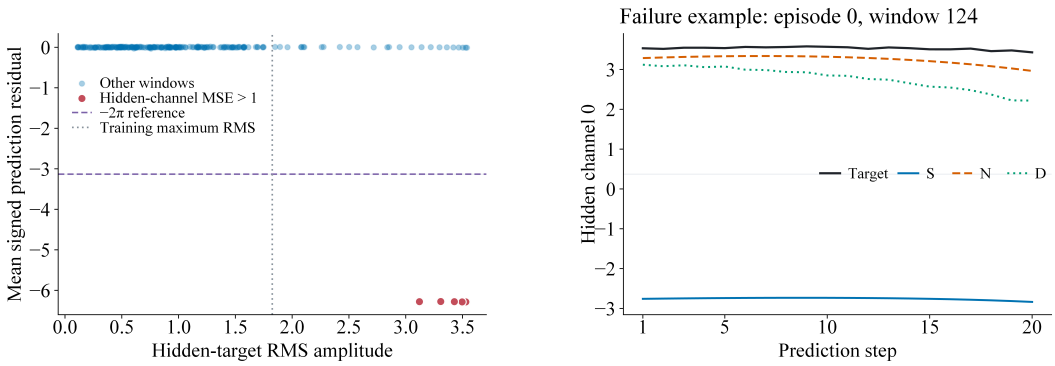

Figure 12: Phase-offset diagnostic for a parameter extrapolation failure. Concentrated errors and an approximately periodic state offset are consistent with a state-aliasing failure; the visualization does not replace the aggregate adverse endpoint in Table 3.

has lower all-channel error than both baselines in 194 histories (80.8%). This is a paired descriptive comparison within five episodes with overlapping windows. The identified component is included to distinguish its performance from the final mixture.

Prediction-target agreement. Figure 16 examines all three channels of the main held-out view under the observed action plan. Every panel contains 4,800 prediction-target pairs from 240 histories and twenty future steps. The visible-channel predictions of S lie close to the identity line, and its MSE is below both seed-11 baselines on all three channels. This diagnostic fixes the baseline seed and action branch, whereas the main aggregate comparison averages the declared seeds and branches.

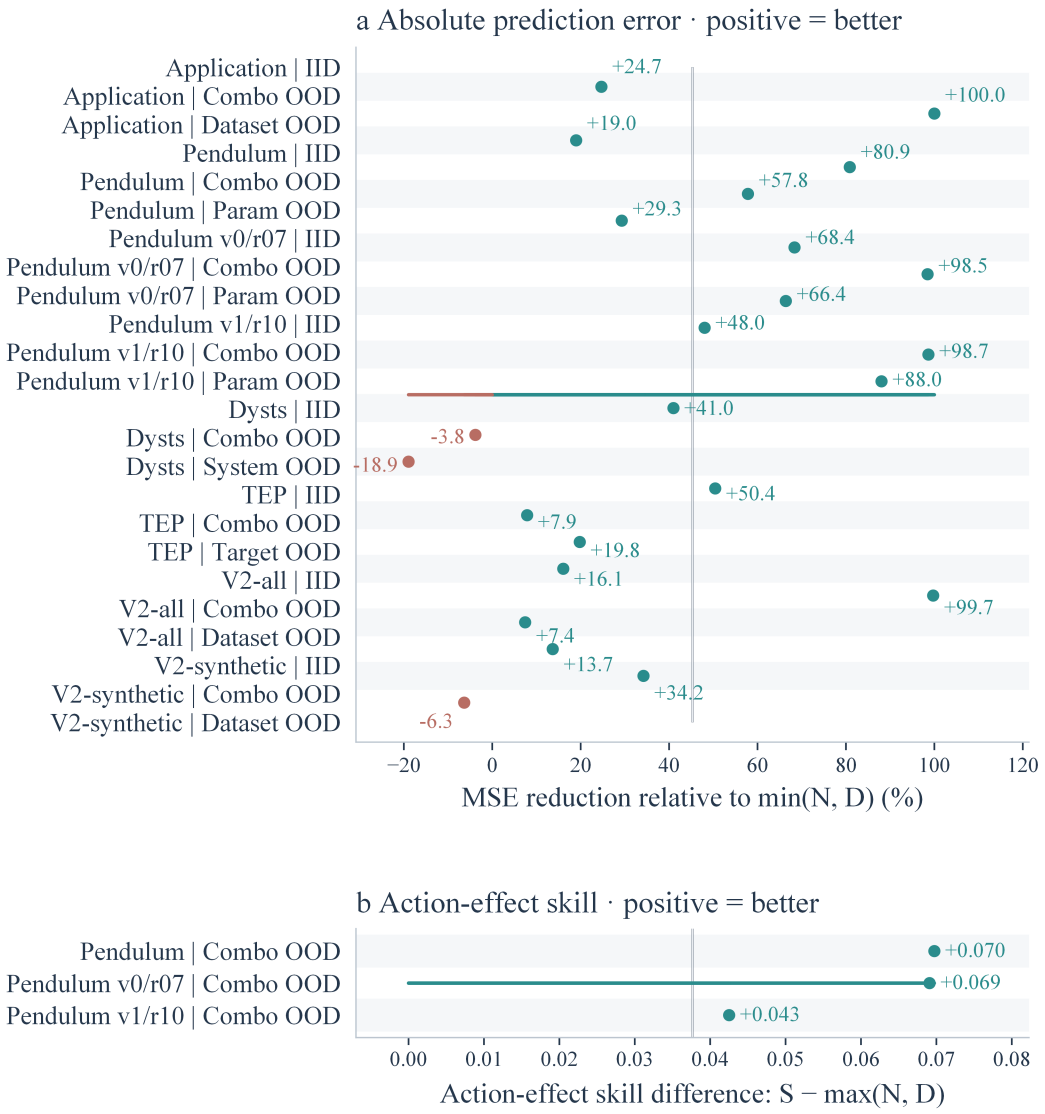

Figure 13: All 27 primary endpoints, with gains over the stronger baseline. MSE reductions and action-effect skill differences are shown on separate scales. NS-WM improves 24 endpoints. Each endpoint is one condition/split/metric result; no independence or confidence interval is inferred from the endpoint count.

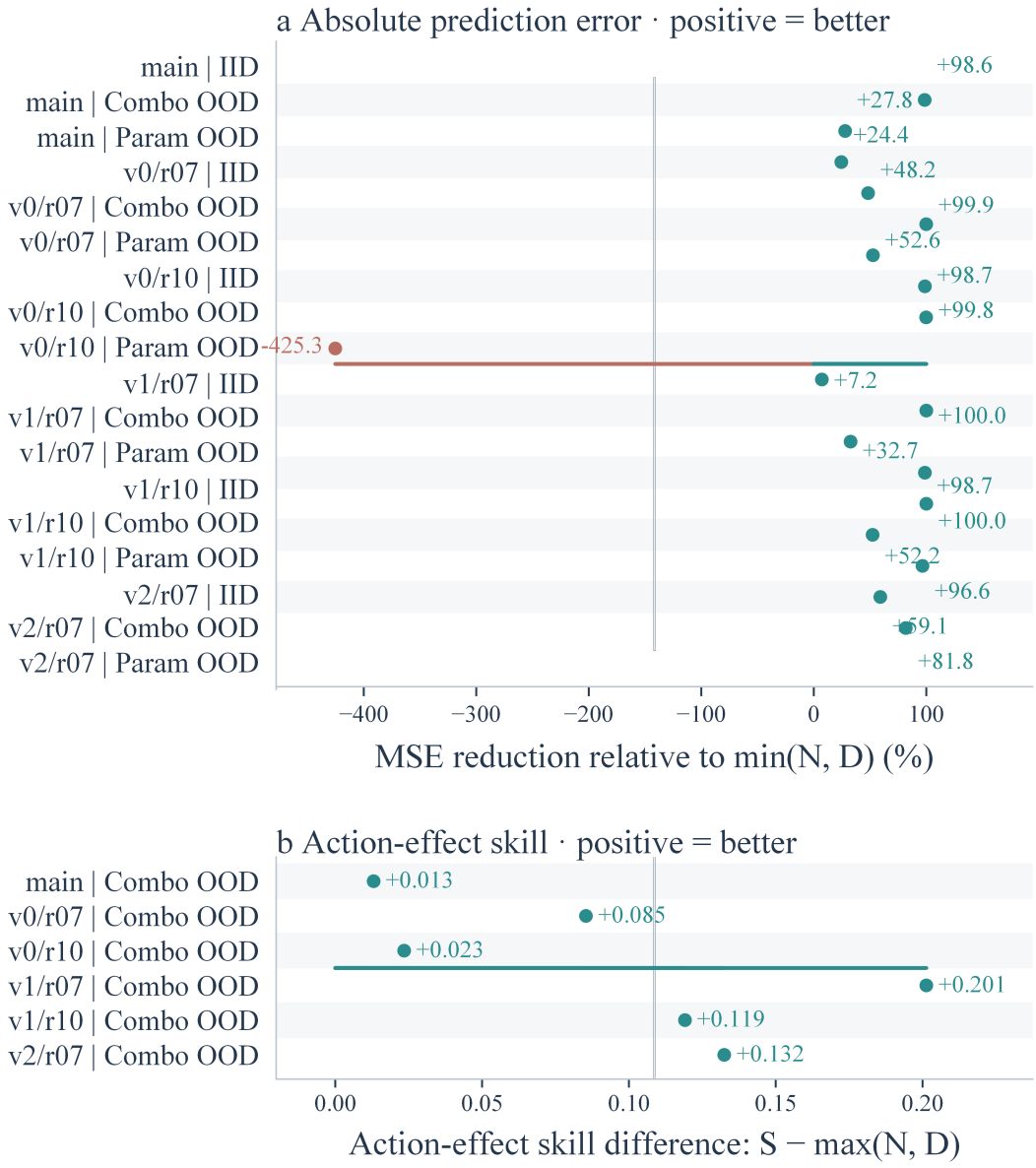

Figure 14: All 24 additional-population endpoints. NS-WM improves 23 endpoints, including combination-OOD MSE and action-effect skill in all six conditions. The complete parameter-OOD range is retained. Baselines here use the recorded single-seed evaluation.

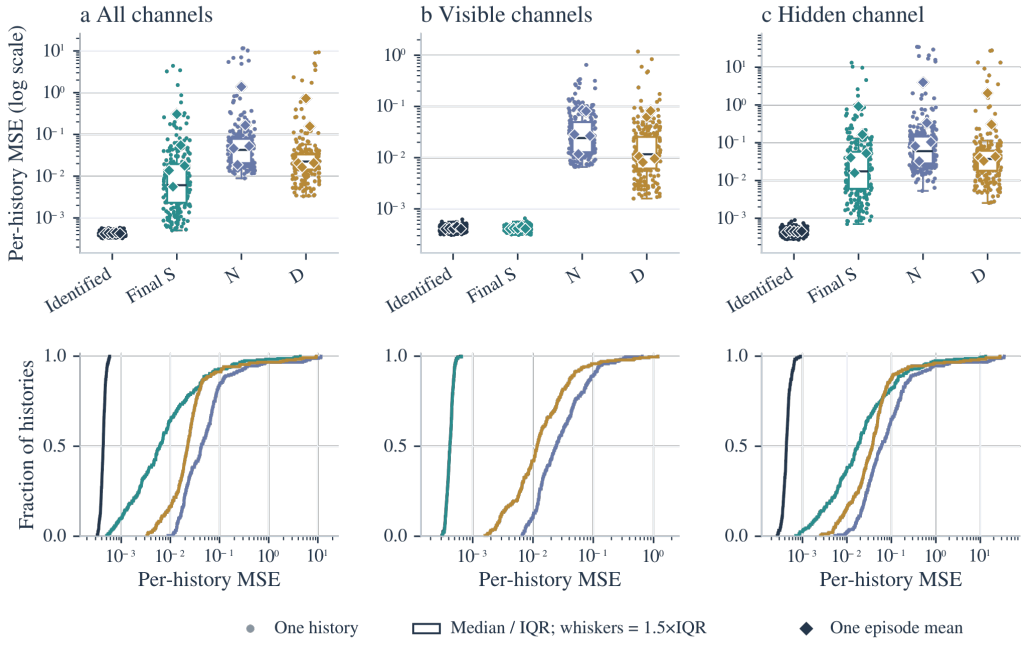

Figure 15: Main combination-OOD history-error distributions and empirical cumulative distributions. All histories are shown. Box plots display median, interquartile range, and 1.5-IQR whiskers; diamonds are episode means. The individual identified component and final mixture are different estimators.

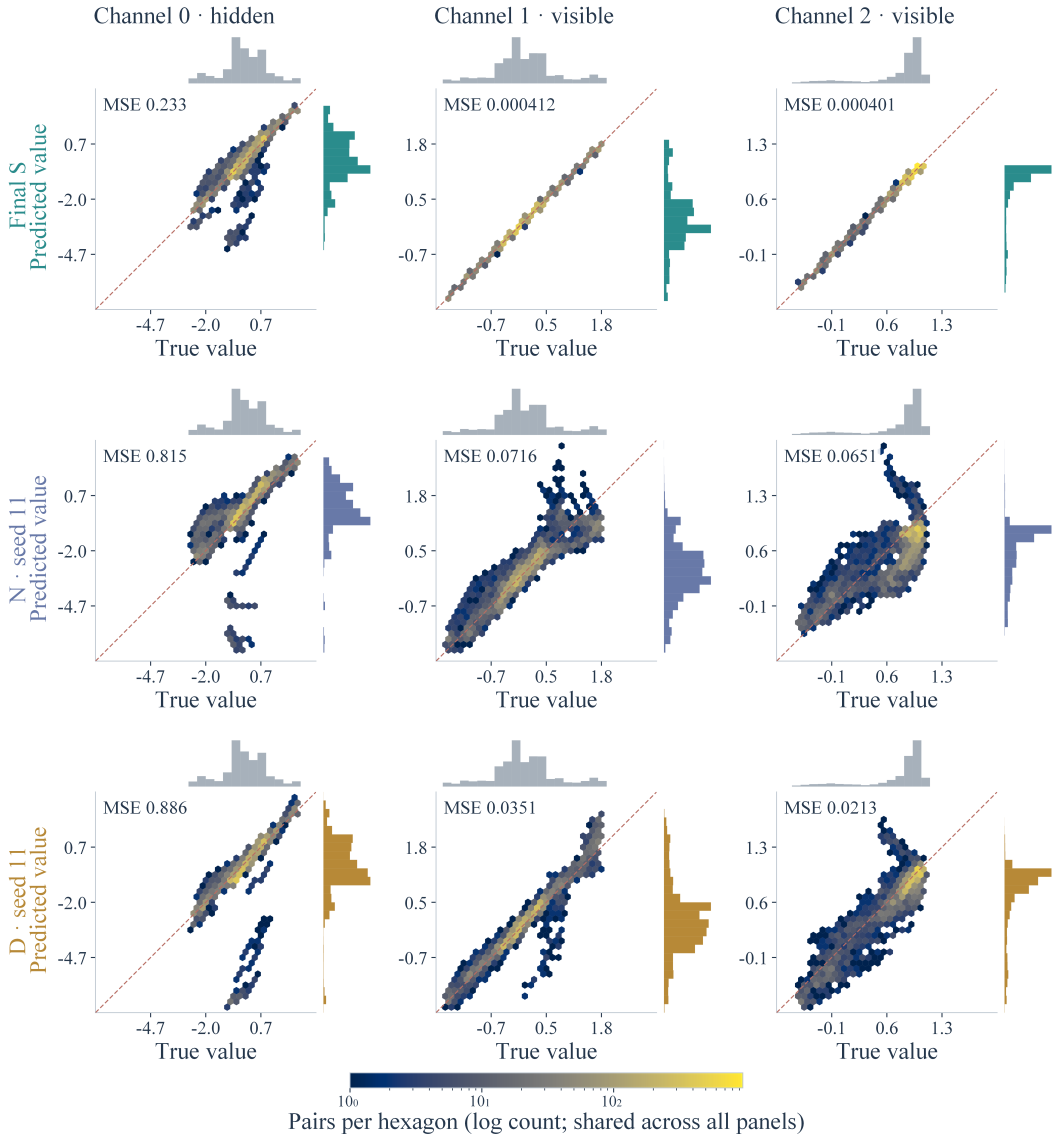

Figure 16: Prediction-target agreement for the main held-out view and observed action branch. S is the final composite; N/D use seed 11. Hexagons aggregate all 4,800 pairs per panel with a shared logarithmic count scale; axes are matched within each channel. Marginal histograms show distribution shape. The dashed line is perfect agreement.

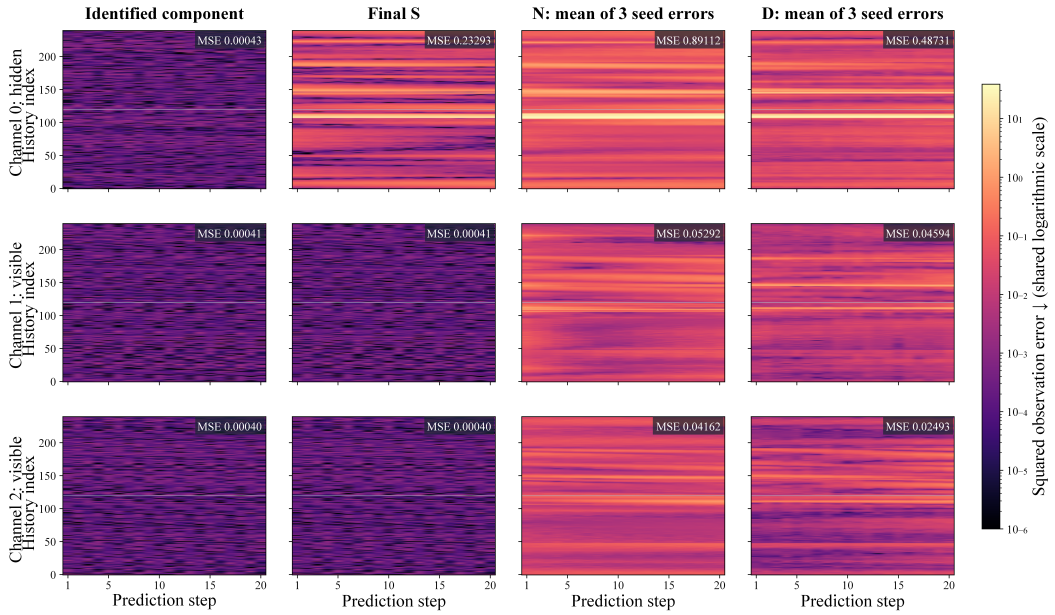

Figure 17: **Prediction errors across channels and forecast steps.** Rows correspond to observation channels, and columns compare the identified component, final NS-WM predictor, neural world model, and direct forecaster. Each panel contains squared errors for 240 histories over 20 forecast steps in sensor view 2 under the observed-action branch. Histories are ordered by episode and history start; separators mark episode boundaries. All panels share a logarithmic color scale with a display floor of 10^{-6} and no upper clipping. Neural-world-model and direct-forecaster errors are averaged across the evaluated model seeds.

E.1 Prediction Errors Across Channels and Forecast Steps

Figure 17 reports squared prediction errors for 240 histories over 20 forecast steps in the main held-out sensor view (view 2), under the observed-action branch. Rows distinguish the hidden channel from the two visible channels, while columns distinguish the identified component, the final NS-WM predictor, and the two baselines. In this diagnostic cohort, the final predictor achieves lower average MSE than both baselines on each of the three channels. The shared logarithmic scale shows how errors are distributed across histories and forecast steps, complementing the aggregate forecast and action-effect metrics. N/D entries average seed-specific squared errors rather than scoring averaged predictions; the overlapping history windows are not independent replicates.