# WHEN SHOULD A WORLD MODEL TRUST ITS IMAGINATION? LOCAL EMPIRICAL SUPPORT PREDICTS RECURSIVE ROLLOUT FAILURE

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

World-model planners recursively evaluate imagined states that may drift into regions weakly supported by the data used to train the model. We ask whether local empirical support at an imagined state predicts rollout-prediction failure several steps later, beyond evidence that the model is already struggling at the current step. Using average $k$-nearest-neighbor distance in learned latent space as a simple support probe, we find that local support adds predictive information beyond current one-step error, rollout position, and latent norm across multiple future offsets. The relationship transfers across five state-based MuJoCo tasks, visual inputs, deterministic residual dynamics, and a Dreamer-style recurrent state-space model (RSSM). It is not captured by the tested measures of global shift magnitude and remains partially informative after conditioning on transition disagreement and Mahalanobis distance, although the residual gain is much smaller in Hopper than in HalfCheetah. The most consistent pattern is geometric. Imagined states associated with future failure lie farther from training support across all dense state-based audits and every RSSM seed. Control experiments show that this diagnostic does not automatically yield a useful intervention. A support score can become more predictive while a fixed support penalty becomes worse, and percentile calibration can weaken an effective raw penalty. Overall, local empirical support is a reproducible forward-looking diagnostic of recursive rollout-prediction failure, while the appropriate control response must be validated separately.

## 1 INTRODUCTION

World-model agents increasingly make decisions by reasoning through predictions that are never directly observed. Planners score candidate trajectories in a learned latent dynamics model, while imagination-based agents train behavior from recursively generated latent futures (Hafner et al., 2025; Hansen et al., 2024). This creates a basic reliability problem. Recursive prediction can carry the model into states that are locally unsupported by its training experience, and optimization can preferentially select precisely those trajectories when model error is exploitable (Yu et al., 2020; Kidambi et al., 2020).

The key question in this paper is therefore not merely whether a test environment is globally out of distribution, nor whether a particular uncertainty estimator can improve return. We ask a more local, conditional, and temporal question: after conditioning on strong information about current model difficulty, does weak empirical support still forecast recursive rollout-prediction failure several steps later? For a planner, this distinction matters because current error can look acceptable while recursive imagination is already entering a region associated with later failure. Our baseline already includes rollout position, current one-step prediction error, and latent norm. We deliberately include realized one-step error as an oracle conditioning variable to make the scientific test conservative. It is not assumed available to a planner, whereas the support statistic itself requires only the imagined latent state and the training support bank. We then ask whether distance to world-model training support adds information about rollout-prediction failure at $t + h$. Average $k$-nearest-neighbor distance is intentionally a simple probe rather than the novelty claim. Nearest-neighbor uncertainty and pessimistic control have substantial precedent.



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.

We test four simpler explanations: contemporaneous error, global OOD, standard uncertainty, and architecture/horizon specificity. If support merely restated current model error, its increment should disappear after conditioning on current one-step error. If it measured only environment-level OOD, global shift or separability should explain it. If standard epistemic uncertainty subsumed it, disagreement or Mahalanobis risk should absorb it. Finally, if the phenomenon were idiosyncratic, it should disappear across future offsets or model families.

Two claims organize the paper. **First, local support is forward-looking beyond current error:** the support–failure relationship persists across future offsets, tasks, visual representations, deterministic residual models, and Dreamer-style recurrent state-space models (RSSMs), while global OOD and standard uncertainty do not fully absorb it. **Second, diagnostic transfer does not imply intervention transfer:** the same reliability signal can remain predictive while a fixed penalty derived from it becomes ineffective or harmful. The clearest counterexample is temporal visual stacking, which increases the IID diagnostic gain in area under the receiver operating characteristic curve ($\Delta$AUROC) from $0.044$ to $0.095$ while the frozen support penalty shifts from $+26.3$ to $-67.9$ return.

This separation matters because world-model systems often couple detection and response. We therefore treat control as a stress test of whether a valid diagnostic automatically induces a useful planning cost. The central empirical thesis is:

> **Local empirical support can warn of future recursive rollout-prediction failure beyond current-error information, but the control response to that warning must be validated separately.**

**What is new and what is not.** Nearest-neighbor uncertainty and pessimistic model-based control are established ideas. Our contribution is the conditional temporal relationship $d_{\mathrm{kNN}}(z_t) \rightarrow$ rollout-prediction failure at $t + h$ after accounting for current-difficulty signals, together with evidence that this diagnostic relationship transfers more reliably than fixed interventions derived from it. The resulting methodological principle is to validate failure forecasting before treating a reliability score as a transferable control cost.

Our contributions are:

- **Forward-looking support beyond current error.** Support at $t$ predicts rollout-prediction failure at $t+h$ after conditioning on current one-step error, rollout position, and latent norm. This relationship persists across future offsets, five successfully evaluated state-based tasks, visual inputs, deterministic residual dynamics, and recurrent stochastic RSSMs. It is not reduced to global OOD and remains partially complementary to disagreement and Mahalanobis risk. The most stable invariant is geometric: failure-associated imagined states are farther from training support across the dense audits and all RSSM seeds.
- **Diagnostic transfer does not imply intervention transfer.** Temporal visual stacking more than doubles the IID incremental diagnostic gain ($0.044 \rightarrow 0.095$) while the same frozen support penalty flips from beneficial ($+26.3$) to harmful ($-67.9$). A pre-specified percentile calibration likewise weakens an effective raw penalty. This motivates a system-design principle: validate reliability estimation and the response policy built around it as separate components.

## 2 Related Work

**World models and latent-space planning.** Modern model-based RL learns compact predictive representations and uses them either to generate imagined training experience or to optimize actions directly. PlaNet established latent-space planning with learned dynamics (Hafner et al., 2019). DreamerV3 demonstrates the scalability of learning behavior through latent imagination across diverse domains (Hafner et al., 2025). TD-MPC2 performs local trajectory optimization through an implicit latent world model across a broad continuous-control suite (Hansen et al., 2024). Our setting is deliberately smaller: rather than proposing a stronger world-model architecture, we ask when a fixed learned model should be trusted during planning.

**Model exploitation, local support, and conservative planning.** A long line of model-based RL work addresses compounding model error and policy exploitation. MBPO limits model usage with

short rollouts branched from real data (Janner et al., 2019). In offline RL, MOPO penalizes rewards using model uncertainty (Yu et al., 2020), while MOReL constructs a pessimistic MDP that discourages behavior in regions judged unreliable (Kidambi et al., 2020). The closest precedent is GELATO, which estimates model error using distance to a learned data manifold and uses nonparametric nearest-neighbor uncertainty as pessimism in offline model-based RL. It evaluates both its geometry-aware metric and a Euclidean latent-distance variant on continuous-control benchmarks (Tennenholtz & Mannor, 2022). Our endpoint differs in that future recursive model-prediction failure is measured directly before asking how a controller should respond.

**Related reliability and imagination diagnostics.** Knowledge Boundary uses ensemble disagreement to identify a region of world-model competence, guiding exploration and filtering unreliable imagined rollouts (Liu et al., 2026). LP-Dreamer detects anomalous imagined transitions through latent linear-prior violations and suppresses their contribution to value learning (Ikram et al., 2026). Recent diagnostic work also asks what kind of error compounds, proposing a kinematic-versus-dynamic characterization of long-horizon DreamerV3 rollouts (Schäfer et al., 2026). Foresight uses action-conditioned world-model latents to detect task-level failure in long-horizon manipulation, including real-robot evaluation (Zhang et al., 2026), while RENEW targets model exploitation in thin-data regions by using epistemic uncertainty to focus model repair from preferences (Bhamidipaty et al., 2026). These works are adjacent but use different endpoints, including task-failure monitoring, anomaly rejection, and model repair rather than conditional prediction of future recursive model error. Together they establish that uncertainty, anomaly detection, and trust boundaries for imagination are important and precede our paper.

**How our question differs.** Our target is not a new detector or filtering rule. We ask whether empirical support at an intermediate imagined state is a forward-looking predictor of recursive rollout-prediction failure at $t + h$, conditional on current one-step error and other baseline features, and whether that relationship transfers more reliably than interventions built from the same signal. This changes the primary endpoint from policy return or anomaly rejection to future rollout-prediction and failure-conditioned geometry. We additionally test alternatives that could make the result trivial: global OOD severity, learned transition disagreement, Mahalanobis distance, a single future offset, and a single latent architecture. In this sense the kNN statistic is a deliberately simple probe used to characterize a temporal reliability regularity, not the claimed algorithmic contribution.

**Uncertainty-aware world models.** Recent work continues to make uncertainty an explicit part of world-model learning and imagination. WIMLE combines multimodal world models with ensemble-based uncertainty and downweights uncertain synthetic transitions (Aghabozorgi et al., 2026). Recent surveys likewise identify compounding error, planner exploitation, uncertainty calibration, and rollout management as central tensions in world-model control (Kirchner et al., 2026). Our evidence complements these approaches by isolating diagnosis from response. We ask which aspects of a support–failure relationship remain stable when uncertainty estimators, representations, horizons, and model families change.

**Dynamics shift and world-model adaptation.** World models may also fail when deployment dynamics differ from training. ReDRAW, for example, adapts pretrained world models through residual corrections in latent dynamics under sim-to-real mismatch (Lanier et al., 2026). Parthasarathy et al. attack a related planning train–test gap by synthesizing planner-relevant training data so that the world model better supports gradient-based planning (Parthasarathy et al., 2025). We take the complementary fixed-model view. Controlled mass and friction shifts test whether existing local support predicts where recursive imagination is unreliable, rather than providing an opportunity to adapt the model itself.

Overall, the contribution is temporal and conditional. Empirical support at $t$ is a reproducible forward-looking diagnostic of recursive rollout-prediction failure at $t+h$ after accounting for current error-related features, while conversion of that diagnostic into a fixed planning intervention transfers less reliably. The experiments characterize the robustness, alternatives, and boundaries of that relationship.

# 3 SETUP AND METHOD

## 3.1 WORLD MODEL

Let $s_t$ denote the environment state and $a_t$ the action. We train an encoder $E$, latent dynamics model $F$, and decoder $D$,

$$z_t = E(s_t), \tag{1}$$
$$\hat{z}_{t+1} = F(z_t, a_t), \tag{2}$$
$$\hat{s}_{t+1} = D(\hat{z}_{t+1}). \tag{3}$$

The state-based models use a 32-dimensional latent representation and lightweight multilayer perceptrons. Training uses nominal random-policy data, as evaluation recursively rolls the model forward using held-out action sequences.

## 3.2 FUTURE ROLLOUT-PREDICTION FAILURE

For an imagined latent state at rollout position $t$, we define a future prediction-error target $e_{t+h}$ at a fixed look-ahead offset $h = 5$. In the grouped IID replication, a rollout state is labeled as a future failure when this error exceeds the 80th percentile of recursive rollout root mean squared error (RMSE) in the held-out IID rollout table. Prediction models are evaluated with grouped splits by rollout start so timesteps from the same imagined rollout do not cross classifier train/test partitions. In the dense dynamics-shift sweeps, the corresponding IID threshold is fixed once for each environment, seed, and perturbation axis and reused at every shifted scale.

The baseline predictor uses quantities available without support geometry, including rollout horizon, current one-step prediction error, and latent norm. The realized one-step error is an oracle conditioning variable used to make the scientific comparison conservative. It is not assumed observable during open-loop imagination. The kNN support statistic itself uses only the imagined latent state and the training support bank. We test whether geometry adds information beyond this baseline.

## 3.3 LOCAL LATENT SUPPORT RISK

Let $\mathcal{M}_{\text{train}} = \{z_i\}_{i=1}^{N}$ be encoded latent states from world-model training data. We define

$$d_{\text{kNN}}(z) = \frac{1}{k} \sum_{j \in \mathcal{N}_k(z; \mathcal{M}_{\text{train}})} \|z - z_j\|_2. \tag{4}$$

Large $d_{\text{kNN}}$ indicates that a latent state lies in a locally weakly supported region of the learned representation.

## 3.4 GEOMETRY-AWARE MPC

We use random-shooting model predictive control (MPC) with a learned reward model. Vanilla planning scores an imagined action sequence $\tau$ using

$$J_{\text{vanilla}}(\tau) = \sum_{t=0}^{H-1} \hat{r}_t. \tag{5}$$

Our soft geometry-aware planner instead uses

$$J_{\text{geo}}(\tau) = \sum_{t=0}^{H-1} \left[\hat{r}_t - \lambda d_{\text{kNN}}(\hat{z}_t)\right]. \tag{6}$$

Unless otherwise stated, canonical planning uses horizon $H = 12$ and 256 sampled candidate action sequences per decision. Earlier hard-truncation, matched-replay, and penalty-sensitivity analyses are retained as exploratory appendix evidence because they predate the training-support correction. They are not used for the confirmatory control claim.

# 4 EXPERIMENTAL SETUP

The state-based diagnostic studies use five independently trained world-model seeds per environment; the confirmatory control study adds five new models, giving $n = 10$ for the five reported control conditions. Each nominal dataset contains 150,000 random-policy transitions. The canonical world-model pipeline uses a chronological 80/20 transition split. Models are fit only on the first 80% and diagnostic rollouts use held-out transitions or fresh simulator rollouts. Because a transition-level boundary can in principle divide an episode, we additionally run a frozen chronological whole-episode audit that enforces zero train/test episode overlap while preserving approximately the same 80/20 transition fraction. The canonical diagnostic remains positive in all five seeds of both original environments (Appendix B). We evaluate nominal dynamics and controlled mass/friction perturbations. All replication claims treat the independently trained world-model seed as the experimental unit. Rollout rows, episodes, and probe observations are not treated as independent model replications.

**Predictive evaluation.** We measure AUROC and area under the precision-recall curve (AUPRC) for predicting future rollout failure.

**Failure-conditioned geometry.** Across dense rollout files, we compare $d_{\mathrm{kNN}}$ for imagined states associated with future failure against states associated with future success.

**Control evaluation.** The canonical confirmatory comparison uses vanilla MPC, soft geometry-penalty MPC, and the matched disagreement penalty. The kNN support reference is constructed only from the first 80% nominal partition used to fit the world model. All main-text control numbers and comparisons use this training-only definition. Episode returns are averaged within each trained model seed before paired differences are computed, and uncertainty is estimated by bootstrapping the 10 seed-level differences. The original cohort reevaluated under the training-only support definition and the five independently trained follow-up models are analyzed under the same support definition.

**Matched ensemble-disagreement comparator.** To compare local support with a conventional epistemic signal, we train three bootstrap latent-dynamics heads while keeping the encoder, base world model, reward model, and planner fixed. For $(z, a)$, disagreement is the RMS deviation of predicted next latents across heads,

$$u(z,a) = \sqrt{\frac{1}{ED}\sum_{e,d}\left(F_e(z,a)_d - \bar{F}(z,a)_d\right)^2}, \qquad E = 3.$$

We calibrate its training-distribution scale so that $Q_{0.9}(u_{\mathrm{train}}) = Q_{0.9}(d_{\mathrm{kNN,train}})$ and use the same $\lambda = 0.50$ in $J_{\mathrm{unc}} = \sum_t [\hat{r}_t - \lambda u_t]$. No return-based tuning of the disagreement coefficient is performed.

# 5 RESULTS

## 5.1 LOCAL SUPPORT ADDS FORWARD-LOOKING INFORMATION BEYOND CURRENT ERROR

The baseline already contains rollout position, current one-step prediction error, and latent norm. Adding kNN geometry nevertheless improves future-failure AUROC in all five HalfCheetah seeds and all five Hopper seeds. A stricter train-group-only target calibration leaves all 10/10 gains positive, changing the mean $\Delta$AUROC only from $0.06143$ to $0.06117$ in HalfCheetah and from $0.01186$ to $0.01176$ in Hopper (Appendix B). Thus the result is not simply "large current error is far from data." Support contributes information about later recursive rollout-prediction failure after current one-step error is already in the predictor. The incremental magnitude is not uniformly large and is notably smaller in Hopper. Hopper therefore illustrates that standalone failure discrimination and residual incremental value need not align. Support can be strongly failure-associated while much of its information is already present in the baseline. Coordinate robustness is also favorable. Orthogonal rotation preserves kNN distance to numerical precision, and standardization/whitening do not systematically reverse the effect (Appendix F).

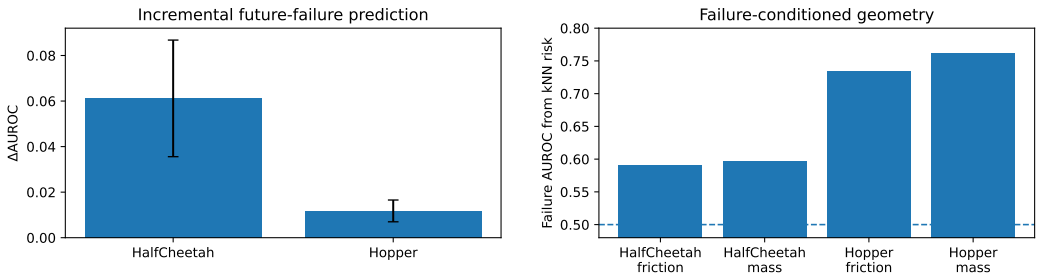


Figure 1: **Local support is forward-looking and failure-associated.** Left: adding kNN geometry to a baseline that already includes current one-step error improves future rollout-prediction AUROC in both environments (mean ± SD; 5/5 positive seeds each). Right: kNN distance predicts future rollout-prediction failure above chance in all four dense perturbation families; per-file AUROC exceeds 0.5 in 160/160 files.

### 5.2 The warning signal persists into the future and across stress tests

The horizon sweep directly tests whether the canonical $h = 5$ result is merely a contemporaneous OOD/error correlation. Relabeling the same rollout tables at $h \in \{1, 2, 3, 5, 8, 10, 12, 16\}$ leaves baseline+kNN $\Delta$AUROC positive in 5/5 seeds at every horizon in both original environments. The profile is task dependent. It is broad and persistent in HalfCheetah, but smaller and decaying in Hopper. We therefore do not claim a fixed warning time. The repeated positive increments nevertheless show that support contains forward-looking information over multiple recursive offsets (Appendix C).

The relationship also survives orthogonal stress tests. Under the same five-step grouped protocol, Walker2d, Ant, and Reacher each show positive kNN increments in 5/5 independently trained models (15/15 total; Appendix E). A residual multi-step model preserves all three diagnostic criteria in 5/5 seeds of both original environments. A compact Dreamer-style RSSM also preserves the underlying support–failure relationship under recurrent stochastic dynamics. In the pre-specified RSSM seeds 0–4, HalfCheetah meets the pre-specified transfer criterion, while Hopper has positive mean $\Delta$AUROC but positive increments in only 3/5 seeds. A separate seeds 5–9 robustness extension meets the same criterion in both environments (Appendix D). Finally, under the strongest conditioning set, which includes the baseline features together with learned transition disagreement and global Mahalanobis distance, kNN still adds mean $\Delta$AUROC of $+0.0372$ in HalfCheetah and $+0.00354$ in Hopper. The increment is positive in 5/5 seeds of each environment (Appendix N). The small Hopper residual reinforces that geometric association and conditional uniqueness are distinct quantities.

Table 1: **Architecture and task transfer at the canonical five-step target.** $\Delta$AUROC is baseline+kNN minus baseline; "fail>succ." counts seeds whose failure-associated states have larger mean kNN distance.

| Extension | Environment | $\Delta$AUROC | Pos. | kNN AUROC | Fail>succ. |
|---|---|---|---|---|---|
| Residual multi-step | HalfCheetah | +0.0108 | 5/5 | 0.581 | 5/5 |
| Residual multi-step | Hopper | +0.0303 | 5/5 | 0.814 | 5/5 |
| RSSM primary | HalfCheetah | +0.0315 | 4/5 | 0.579 | 5/5 |
| RSSM primary | Hopper | +0.0222 | 3/5 | 0.671 | 5/5 |
| Broader task | Walker2d | +0.0179 | 5/5 | 0.690 | 5/5 |
| Broader task | Ant | +0.0361 | 5/5 | 0.576 | 5/5 |
| Broader task | Reacher | +0.0270 | 5/5 | 0.749 | 5/5 |

### 5.3 The most stable invariant is failure-conditioned support geometry

Across 160 dense rollout files, mean kNN support distance is larger for states associated with future failure than for states associated with future success in 160/160 cases. Likewise, kNN distance

has future-failure AUROC above chance in 160/160 files. The effect is moderate but consistent in HalfCheetah and substantially larger in Hopper.

Table 2: Failure-conditioned local geometry. Values are family-level averages from the dense rollout analysis.

| Family | Mean $\Delta$kNN | Cohen's $d$ | Failure AUROC |
|---|---|---|---|
| HalfCheetah friction | 0.090 | 0.334 | 0.590 |
| HalfCheetah mass | 0.094 | 0.351 | 0.596 |
| Hopper friction | 0.398 | 0.836 | 0.734 |
| Hopper mass | 0.442 | 0.920 | 0.761 |

### 5.4 Global shift magnitude is not a universal mechanism

We test two natural explanations for why geometry becomes more useful under some dynamics shifts. First, raw parameter distance $|\delta - 1|$ does not explain the cross-environment response. Second, classifier-based global latent-distribution separability is strongly associated with geometry gain in pooled conditions but weakens after controlling for failure rate, environment, perturbation axis, and physical shift magnitude, and fails leave-one-family-out transfer. These negative results argue against the simple claim that "more OOD" universally implies greater geometry value and against the interpretation that the observed predictive effect is merely a proxy for global OOD severity.

### 5.5 Intervention transfer is environment dependent

Canonical control uses training-only support and $n = 10$ independently trained models per reported condition; Appendix K documents the support-reference correction and cohort construction. The resulting picture is strongly environment dependent. Geometry improves over vanilla in all three Hopper conditions with intervals excluding zero: IID $+86.5$ (9/10 positive; 95% CI $[58.6, 111.2]$), high mass $+70.1$ (10/10; $[52.5, 90.3]$), and high friction $+51.6$ (9/10; $[26.9, 79.1]$). HalfCheetah IID remains positive in the pooled analysis ($+69.9$, 8/10; $[12.7, 155.7]$), but high friction is heterogeneous ($+57.6$, 6/10; $[-5.4, 170.4]$). The independent follow-up HalfCheetah cohort itself has much smaller gains than the reevaluated original cohort, so we do not describe the pooled means as homogeneous replication.

Figure 2 shows the seed-level paired effects. Full bootstrap intervals and cohort splits are in Appendix K. HalfCheetah remains heterogeneous, whereas Hopper shows much tighter replication. Control is therefore evidence about intervention conditionality, not a universal planning-method claim.

### 5.6 Local support and learned disagreement are not interchangeable

The $n = 10$ comparator reinforces the same boundary. Geometry beats disagreement in Hopper IID ($+40.2$, CI $[8.0, 73.6]$) and high friction ($+53.3$, $[24.5, 83.5]$), while high mass is inconclusive. In HalfCheetah IID the methods are not decisively separated, and under high friction disagreement outperforms geometry ($-30.2$, CI $[-57.8, -7.8]$). Thus empirical support and transition disagreement are useful but non-interchangeable control signals.

### 5.7 Temporal visual context strengthens diagnosis but not intervention

The visual Pendulum extension tests whether control weakness is explained by single-frame partial observability. Paired models use either one grayscale frame or a four-frame stack ("stack4") from the same 60k-transition dataset. Under IID dynamics, stack4 raises mean incremental kNN AUROC from $0.044$ to $0.095$ (paired gain $+0.051$), yet the same frozen geometry penalty moves from $+26.3$ return to $-67.9$. Under high gravity, the stack4 diagnostic remains positive on average ($+0.053$; 4/5 seeds), while both penalties are harmful. Better temporal diagnosis therefore does not rescue the fixed intervention (Appendix G).

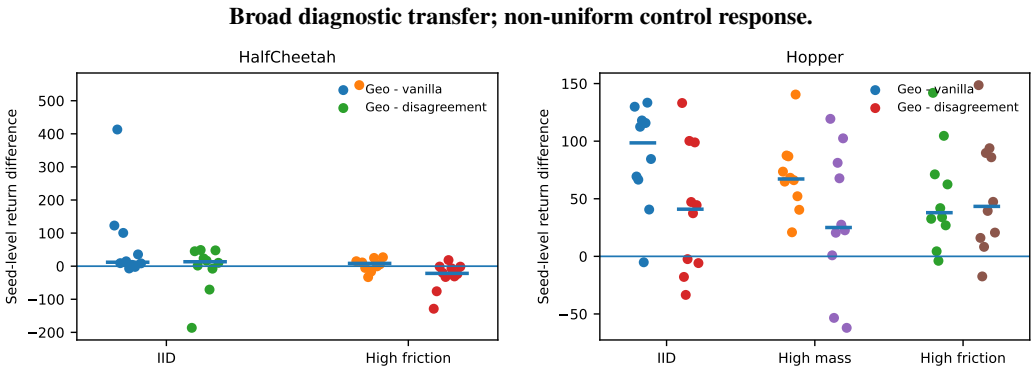


Figure 2: **A reliable diagnostic does not imply a uniform intervention.** Seed-level return differences use environment-specific scales; horizontal segments are medians. Hopper shows tight positive geometry-versus-vanilla replication, while HalfCheetah is substantially more heterogeneous and disagreement can outperform geometry under high friction.

## 6 Discussion

The main empirical principle is that trajectory-local support remains forward-looking after current difficulty is already observed. Immediate error signals need not reveal that recursive imagination is entering weakly supported regions associated with later failure. The relationship survives changes in horizon, task, visual context, deterministic architecture, and recurrent stochastic architecture, while simple global notions of "more OOD" do not provide a transferable mechanism.

Consistency should not be confused with magnitude. HalfCheetah retains a substantial conditional kNN increment after both disagreement and Mahalanobis risk, whereas Hopper's residual increment is small despite strong standalone failure discrimination. Hopper is therefore informative rather than merely weaker. It shows that standalone discrimination and incremental predictive value need not align. Architecture and environment change how much support information is already captured by baseline/error-based predictors, so we claim partial complementarity rather than statistical independence or uniformly large gains.

The negative global-mechanism experiments narrow the claim from "geometry detects OOD environments" to the operational statement that geometry identifies imagined states with weaker local empirical support. This also explains why intervention can matter under nominal dynamics. Optimization can push imagined trajectories toward unsupported regions even without an environment-level shift.

This local-support view is complementary to ensemble and probabilistic uncertainty methods. Diagnostic kNN retains positive incremental information after transition disagreement and global Mahalanobis distance in every seed of both original environments, yet control does not have a universal winner. The visual-stack counterexample makes the separation especially direct. Diagnosis improves while the frozen intervention becomes substantially worse. A pre-specified 90th-percentile calibration also underperforms the raw geometry penalty on average in all five canonical conditions (Appendix O). Across these tests, diagnosis transfers more reliably than intervention. The system-design implication is to validate reliability estimation and the response to that estimate as separate components. Improving the first is not evidence that a fixed control response will improve the second. Alternatives to an additive penalty, such as adaptive imagination-horizon truncation or gating synthetic rollouts when support risk is high, are plausible future responses, but we do not evaluate them here. As a secondary engineering audit, support-bank thinning yields about $2.3\times$ faster queries at 25% bank size while meeting the pre-specified preservation tolerances (Appendix P).

**Scope of the claim.** The evidence supports a reproducible diagnostic relationship across the tested settings, not representation invariance or a universal planning rule. Euclidean kNN remains coordinate-dependent, and the effectiveness of penalties derived from it is conditional on environment and representation.

## 7 LIMITATIONS

The diagnostic evidence remains concentrated in continuous-control settings. The state-based evaluation covers HalfCheetah, Hopper, Walker2d, Ant, and Reacher. One pre-specified environment, InvertedDoublePendulum, could not be evaluated under the frozen 32-step rollout protocol because the random-policy data did not provide valid rollout windows. Architecture transfer includes a deterministic residual multi-step model and a compact Dreamer-style recurrent state-space model (RSSM). The RSSM isolates world-model architecture rather than reproducing the full Dreamer reinforcement-learning stack. It does not include actor or critic learning or online data collection. Its 384-D state is also mapped through a frozen outcome-independent 32-D orthogonal projection before kNN. The visual evaluation is limited to $48 \times 48$ grayscale Pendulum with single-frame and four-frame inputs. Euclidean kNN remains coordinate-dependent. Our checks cover standardization, whitening, orthogonal rotations, and changed learned representations, but not arbitrary invertible reparameterizations.

The intervention results are narrower than the diagnostic results. Geometry-aware MPC performs consistently in Hopper but is more heterogeneous in HalfCheetah, and disagreement performs better in some conditions. Temporal stacking does not rescue the fixed visual penalty, and the tested 90th-percentile calibration weakens the raw geometry controller. These results do not rule out other adaptive mappings from support estimates to control decisions. Support-bank thinning improves runtime but changes the absolute kNN scale, so the reduced-bank controller is not an exact approximation to the full objective. The canonical world-model pipeline uses a chronological transition-level split, although a whole-episode audit with zero train/test episode overlap preserves the core diagnostic in all five seeds of both original environments (Appendix B). Finally, diagnostic rollout rows are correlated, so independently trained world-model seeds, rather than individual rollout rows or episodes, are the inferential replication unit.

## 8 CONCLUSION

Local empirical support is a reproducible forward-looking diagnostic of recursive rollout-prediction failure after conditioning on current error-related features. The relationship persists across future offsets, additional tasks, deterministic residual dynamics, recurrent stochastic RSSMs, and conditioning on disagreement and Mahalanobis risk. Temporal visual context can strengthen it further. Intervention is less stable. Fixed penalties remain environment- and representation-dependent, and a simple training-relative calibration does not improve the raw controller. The practical lesson is not to treat a reliability score as a control objective by default. First validate whether it forecasts failure, then separately validate the response policy built around it.

## REPRODUCIBILITY STATEMENT

The main replication unit, dataset sizes, model architectures, grouped failure protocol, planning hyperparameters, audit corrections, and all extension-specific frozen configurations are documented in the main text and appendix. The supplementary artifacts include frozen extension bundles covering deterministic and recurrent-stochastic architecture transfer, broader environments, temporal visual input, uncertainty complementarity, horizon robustness, calibrated intervention, support-bank scalability, and the episode-disjoint split audit. The archive also includes seed-level and aggregate CSV artifacts, frozen configurations, run manifests where applicable, and plotting and analysis code. For the recurrent stochastic RSSM extension, the pre-specified seeds 0–4 are reported separately from the independent seeds 5–9 robustness extension.

## AI USE STATEMENT

Generative AI tools were used during this project to help identify and survey related research, refine hypotheses and experimental methodology, generate and edit research code, assist with result analysis and interpretation, create plotting and audit utilities, and draft and edit portions of the manuscript. We take responsibility for the final content of this work.

## REFERENCES


Mehran Aghabozorgi, Alireza Moazeni, Yanshu Zhang, and Ke Li. WIMLE: Uncertainty-aware world models with IMLE for sample-efficient continuous control. In *International Conference on Learning Representations*, 2026.

Logan Mondal Bhamidipaty, Mykel Kochenderfer, and Subramanian Ramamoorthy. RENEW: Towards learning world models and repairing model exploitation from preferences. *arXiv preprint arXiv:2607.14180*, 2026.

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 of *Proceedings of Machine Learning Research*, pp. 2555–2565, 2019.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse control tasks through world models. *Nature*, 640(8059):647–653, 2025. doi: 10.1038/s41586-025-08744-2.

Nicklas Hansen, Hao Su, and Xiaolong Wang. TD-MPC2: Scalable, robust world models for continuous control. In *International Conference on Learning Representations*, 2024.

Zarif Ikram, Harry Zhao, Ling Pan, Alex Lamb, and Dianbo Liu. World model anomaly detection with a latent linear prior. *Transactions on Machine Learning Research*, 2026.

Michael Janner, Justin Fu, Marvin Zhang, and Sergey Levine. When to trust your model: Model-based policy optimization. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Rahul Kidambi, Aravind Rajeswaran, Praneeth Netrapalli, and Thorsten Joachims. MOReL: Model-based offline reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Sven Kirchner, Nils Purschke, and Alois Knoll. A survey of world models for physical AI with uncertainty representation and control. *Discover Artificial Intelligence*, 6:1037, 2026. doi: 10.1007/s44163-026-02122-1.

JB Lanier, Kyungmin Kim, Armin Karamzade, Yifei Liu, Ankita Sinha, Kat He, Davide Corsi, and Roy Fox. Adapting world models with latent-state dynamics residuals. In *Proceedings of the 8th Annual Learning for Dynamics and Control Conference*, volume 331 of *Proceedings of Machine Learning Research*, pp. 117–144, 2026.

Zhenxian Liu, Peixi Peng, Yangru Huang, and Yonghong Tian. Perceiving the knowledge boundary: Uncertainty-guided exploration and imagination for world models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(28):23990–23998, 2026. doi: 10.1609/aaai.v40i28.39576.

Arjun Parthasarathy, Nimit Kalra, Rohun Agrawal, Yann LeCun, Oumayma Bounou, Pavel Izmailov, and Micah Goldblum. Closing the train–test gap in world models for gradient-based planning. *CoRR*, abs/2512.09929, 2025. doi: 10.48550/arXiv.2512.09929.

Finn Rasmus Schäfer, Korbinian Moller, Yuan Gao, Christian Oefinger, Sebastian Schmidt, and Johannes Betz. Imagined rollouts are kinematic, not dynamic: A diagnosis of long-horizon world-model failure. *CoRR*, abs/2607.05966, 2026. doi: 10.48550/arXiv.2607.05966.

Guy Tennenholtz and Shie Mannor. Uncertainty estimation using Riemannian model dynamics for offline reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 35, 2022. doi: 10.52202/068431-1381.

Tianhe Yu, Garrett Thomas, Lantao Yu, Stefano Ermon, James Zou, Sergey Levine, Chelsea Finn, and Tengyu Ma. MOPO: Model-based offline policy optimization. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Haoran Zhang, Yifu Lu, Boyang Wang, Xuhui Kang, Yen-Ling Kuo, Zezhou Cheng, Mengdi Wang, and Odest Chadwicke Jenkins. Foresight: Failure detection for long-horizon robotic manipulation with action-conditioned world model latents. *arXiv preprint arXiv:2606.23085*, 2026.

# A EXPERIMENTAL DETAILS

The state-based diagnostic experiments use five random seeds in HalfCheetah-v5 and Hopper-v5. The canonical confirmatory control comparison uses ten independently trained models per reported environment–condition pair. Nominal datasets contain 150,000 random-policy transitions. The archived pipeline uses a chronological 80/20 transition split, with world models trained on the first 80%. The encoder, latent dynamics function, and decoder are separate two-hidden-layer MLPs with SiLU activations and hidden width 128. The latent dimension is 32. Observations and actions are standardized with training-set statistics. The world-model objective is the sum of state reconstruction, latent one-step prediction, and decoded next-state prediction MSEs, with unit weights. The target next latent is stop-gradient. Training uses AdamW with learning rate $3 \times 10^{-4}$, weight decay $10^{-4}$, batch size 512, and 40 epochs.

Recursive diagnostic rollouts use horizon 32. The grouped IID study samples at most 2,500 valid held-out rollout starts per seed. Dense shift studies use 1,500 fresh simulator rollouts per condition. The future-failure target looks five steps ahead. In the grouped IID study, the failure cutoff is the 80th percentile of recursive rollout RMSE in the full held-out rollout table before the grouped classifier split. Dense shift experiments compute an IID cutoff once for each environment, seed, and perturbation axis, then reuse it at all scales. The grouped classifier uses a 30% rollout-group test fraction and $k = 10$ for latent kNN geometry.

For dense shift experiments, mass scales are $\{0.80, 0.90, 1.00, 1.05, 1.10, 1.15, 1.25, 1.35, 1.50\}$ and friction scales are $\{0.70, 0.85, 1.00, 1.10, 1.20, 1.35, 1.50\}$. The control intervention uses a learned reward model with hidden width 128, learning rate $3 \times 10^{-4}$, batch size 512, and 20 epochs. Random-shooting MPC uses horizon 12, 256 candidates, action-noise standard deviation 0.65, and replans every environment step. The default geometry penalty is $\lambda = 0.50$. Hard truncation uses the 95th percentile of training-set kNN risk. Main control evaluation uses five episodes per condition and at most 500 environment steps.

The exploitation diagnostic uses three episodes per condition, at most 300 controller steps, diagnostic stride 10, and a 12-step matched open-loop probe. The penalty sweep uses $\lambda \in \{0, 0.25, 0.50, 1.00\}$ and three episodes per condition. Exploratory diagnostic and sensitivity intervals use the original five trained-model seeds. Canonical control intervals use 10,000 paired bootstrap resamples over ten independently trained model seeds.

For the matched ensemble-disagreement comparator, each trained world-model seed receives three bootstrap latent-dynamics heads. The encoder, base world model, reward model, and MPC candidate generation remain fixed. Disagreement is calibrated per seed by matching its training-set 90th percentile to the 90th percentile of kNN support distance. It is then penalized with the same $\lambda = 0.50$. The uncertainty controller uses the same candidate stream and five evaluation episodes per condition as the geometry controller. For canonical control reporting, five new models are combined with five original models reevaluated under the training-only support definition after harmonizing the geometry support reference to the training partition. Paired method comparisons are computed after averaging episodes within each trained model seed.

# B PREDICTION PROTOCOL AND LEAKAGE CONTROLS

The prediction classifier is split by rollout-start group rather than by individual timestep rows. This prevents timesteps from the same imagined rollout from appearing in both classifier train and test sets. Feature standardization is fitted on classifier-training rows only. For shifted-dynamics experiments, the future-failure threshold is fixed from the IID rollout table rather than recomputed under each shift.

In the original grouped IID replication, the 80th-percentile target threshold is computed from the full held-out rollout table before the group split. Using the archived row-level rollout files, we performed a one-variable robustness audit. The archived protocol and all reported scores were first reproduced to machine precision. We then formed the same group split before target calibration, recomputed the threshold using only classifier-training rollout groups, and regenerated both train and test labels from that fixed training-group threshold. No world model, feature definition, split, standardization,

or classifier setting was changed. The baseline includes rollout horizon, current one-step prediction error, and latent norm. The main geometric comparison adds kNN support distance.

Table 3 gives the directly recomputed seed-level kNN gains from the grouped predictive replication used for the left panel of Figure 1. These rows are derived directly from the archived aggregate prediction scores.

Table 3: Grouped predictive replication. Per-seed improvement of baseline+kNN over baseline.

| Environment | Seed | ΔAUROC | ΔAUPRC |
|---|---|---|---|
| HalfCheetah | 0 | 0.0459 | 0.0539 |
| HalfCheetah | 1 | 0.0366 | 0.0548 |
| HalfCheetah | 2 | 0.0564 | 0.0695 |
| HalfCheetah | 3 | 0.1039 | 0.0812 |
| HalfCheetah | 4 | 0.0644 | 0.0910 |
| Hopper | 0 | 0.0096 | 0.0191 |
| Hopper | 1 | 0.0094 | 0.0126 |
| Hopper | 2 | 0.0160 | 0.0242 |
| Hopper | 3 | 0.0061 | 0.0166 |
| Hopper | 4 | 0.0182 | 0.0212 |

Table 4: Robustness to train-group-only failure-threshold calibration. Values are mean ± sample SD of the per-seed baseline+kNN minus baseline gain. All gains are positive in 5/5 seeds for each environment under both protocols.

| Environment | Original ΔAUROC | Train-group threshold ΔAUROC |
|---|---|---|
| HalfCheetah-v5 | $0.06143 \pm 0.02598$ | $0.06117 \pm 0.02560$ |
| Hopper-v5 | $0.01186 \pm 0.00504$ | $0.01176 \pm 0.00478$ |

The threshold itself changes only slightly across seeds, with a strict-minus-original range of $[-0.00729, 0.00646]$. The corresponding mean ΔAUPRC is also stable. It changes from $0.07007$ to $0.06967$ in HalfCheetah and from $0.01873$ to $0.01871$ in Hopper. Thus the grouped predictive result does not depend materially on using test-group rollout errors to set the binary cutoff.

**Episode-disjoint world-model split audit.** The canonical world-model dataset split is chronological by transition and can in principle place different portions of one episode on opposite sides of the 80/20 boundary. We therefore reran the canonical five-step diagnostic with a frozen chronological whole-episode split chosen nearest the same transition fraction while requiring zero episode identity overlap. All 10 environment–seed runs have zero overlap.

HalfCheetah's original boundary was already episode-aligned, so its result is unchanged in substance. Hopper provides the substantive audit because episode lengths vary and the resulting whole-episode partition genuinely differs. Under the episode-disjoint split, baseline+kNN ΔAUROC remains positive in 5/5 seeds of both environments, with means of $0.06117$ for HalfCheetah and $0.01554$ for Hopper. Mean ΔAUPRC is $0.06967$ and $0.02593$, respectively. kNN-only failure AUROC remains above chance, and failure-associated states remain farther from training support in all 10 models.

Table 5 summarizes the audit. The audit addresses episode identity overlap at the world-model train/test boundary. It does not imply independence of rollout rows, which are still handled by grouped classifier splits and seed-level inference.

## C Early-warning horizon sweep

We reuse each frozen rollout table and relabel future failure at offsets $h \in \{1, 2, 3, 5, 8, 10, 12, 16\}$ using the strict train-group-only threshold protocol. The world model, rollout groups, classifier split, and features are unchanged across horizons. Because horizons share the same trained models and

Table 5: Episode-disjoint split audit. The whole-episode split enforces zero train/test episode overlap while preserving approximately the same 80/20 transition fraction. All positive counts are out of five independently trained models.

| Environment | Mean ΔAUROC | Pos. | kNN AUROC | Fail>succ. |
|---|---|---|---|---|
| HalfCheetah-v5 | 0.06117 | 5/5 | 0.5872 | 5/5 |
| Hopper-v5 | 0.01554 | 5/5 | 0.8374 | 5/5 |

underlying rollout data, the 80 environment–seed–horizon comparisons are not treated as independent replications. The relevant result is that, at every pre-specified horizon, the kNN increment is positive in all five independently trained seeds of both environments.

Table 6: Early-warning horizon sweep. ΔAUROC/AUPRC are baseline+kNN minus baseline; positive counts are independently trained seeds out of five.

| Env. | $h$ | Mean ΔAUROC | Pos. | Mean ΔAUPRC | Pos. |
|---|---|---|---|---|---|
| HalfCheetah | 1 | +0.0534 | 5/5 | +0.0627 | 5/5 |
| HalfCheetah | 2 | +0.0597 | 5/5 | +0.0724 | 5/5 |
| HalfCheetah | 3 | +0.0614 | 5/5 | +0.0722 | 5/5 |
| HalfCheetah | 5 | +0.0612 | 5/5 | +0.0697 | 5/5 |
| HalfCheetah | 8 | +0.0647 | 5/5 | +0.0622 | 5/5 |
| HalfCheetah | 10 | +0.0677 | 5/5 | +0.0600 | 5/5 |
| HalfCheetah | 12 | +0.0650 | 5/5 | +0.0536 | 5/5 |
| HalfCheetah | 16 | +0.0503 | 5/5 | +0.0423 | 5/5 |
| Hopper | 1 | +0.0091 | 5/5 | +0.0194 | 5/5 |
| Hopper | 2 | +0.0110 | 5/5 | +0.0201 | 5/5 |
| Hopper | 3 | +0.0126 | 5/5 | +0.0210 | 5/5 |
| Hopper | 5 | +0.0118 | 5/5 | +0.0187 | 5/5 |
| Hopper | 8 | +0.0081 | 5/5 | +0.0134 | 5/5 |
| Hopper | 10 | +0.0064 | 5/5 | +0.0110 | 5/5 |
| Hopper | 12 | +0.0058 | 5/5 | +0.0093 | 5/5 |
| Hopper | 16 | +0.0051 | 5/5 | +0.0073 | 5/5 |

# D ARCHITECTURE TRANSFER

We train a paired residual multi-step model on the same random-policy dataset used by the baseline model within each environment and seed. The residual model retains latent dimension 32 and width 128 but adds LayerNorm, two action-conditioned residual transition blocks, and recursive three-step latent and decoded-prediction training. The pre-specified architecture-transfer criterion requires, within each environment, positive ΔAUROC, failure-minus-success kNN distance, and kNN-only AUROC above 0.5 in at least 4/5 seeds. The residual model achieves 5/5 on all three criteria in both environments. Effect magnitude is not architecture invariant. Residual-minus-baseline ΔAUROC is negative in all five HalfCheetah pairs and positive in all five Hopper pairs.

Table 7: Architecture transfer at the canonical $h = 5$ target. Raw failure-minus-success kNN distances are architecture-specific and are shown only for within-representation sign/effect context.

| Env. | Architecture | ΔAUROC | Pos. | kNN AUROC | Fail–succ. kNN |
|---|---|---|---|---|---|
| HalfCheetah | Baseline | +0.0576 | 5/5 | 0.5788 | +0.3679 |
| HalfCheetah | Residual multi-step | +0.0108 | 5/5 | 0.5806 | +0.0812 |
| Hopper | Baseline | +0.0135 | 5/5 | 0.8471 | +1.3437 |
| Hopper | Residual multi-step | +0.0303 | 5/5 | 0.8138 | +0.4778 |

## D.1 RECURRENT STOCHASTIC RSSM TRANSFER

We further test a compact Dreamer-style recurrent state-space model, without actor/critic learning or online RL, so that the fixed dataset and diagnostic protocol remain unchanged. The RSSM uses a 128-D GRU deterministic state, 16 categorical stochastic variables with 16 classes each, learned

prior/posterior dynamics, observation and reward prediction, and prior-only recursive imagination after posterior filtering of the real history. The raw 384-D feature $[h_t, z_t]$ is mapped through a frozen, outcome-independent 32-D orthogonal random projection before exact kNN.

The original pre-specified seeds 0–4 remain the primary evaluation set. The pre-specified transfer criterion requires positive $\Delta$AUROC in at least 4/5 seeds, kNN-only AUROC above 0.5 in at least 4/5 seeds, and positive failure-minus-success support distance in at least 4/5 seeds. HalfCheetah meets this criterion, with 4/5 positive $\Delta$AUROC, 5/5 kNN-only AUROC above 0.5, and 5/5 positive failure-minus-success distance. Hopper has positive mean $\Delta$AUROC ($+0.0222$) but only 3/5 positive increments, so it does not meet the pre-specified criterion. Its two negative increments are $-0.00051$ and $-0.00012$, while kNN-only AUROC, failure-support distance, and $\Delta$AUPRC are positive in all five seeds. We therefore describe the primary Hopper result as heterogeneous transfer.

After observing this borderline primary result, we ran a separate five-seed robustness extension in both environments under the identical frozen protocol. Seeds 5–9 meet the same pre-specified criterion in both HalfCheetah (4/5 positive $\Delta$AUROC) and Hopper (5/5). Descriptively across all ten seeds per environment, mean $\Delta$AUROC is $+0.0240$ in HalfCheetah and $+0.0169$ in Hopper, with 8/10 positive in each. All 20 RSSM models have kNN-only failure AUROC above chance and larger mean support distance for future failures. The follow-up is reported as a robustness extension rather than redefining the pre-specified primary experiment.

Table 8: Dreamer-style RSSM architecture transfer. Primary seeds 0–4 and the independent seeds 5–9 robustness extension are reported separately.

| Cohort | Env. | $\Delta$AUROC | Pos. | kNN AUROC $> .5$ | Fail>succ. |
|---|---|---|---|---|---|
| Primary 0–4 | HalfCheetah | +0.0315 | 4/5 | 5/5 | 5/5 |
| Primary 0–4 | Hopper | +0.0222 | 3/5 | 5/5 | 5/5 |
| Follow-up 5–9 | HalfCheetah | +0.0165 | 4/5 | 5/5 | 5/5 |
| Follow-up 5–9 | Hopper | +0.0116 | 5/5 | 5/5 | 5/5 |

## E BROADER-ENVIRONMENT REPLICATION

Four tasks were pre-specified: Walker2d-v5, Ant-v5, Reacher-v5, and InvertedDoublePendulum-v5. The first three pass the frozen technical viability screen and are evaluated without changing the canonical 32-step rollout or five-step failure target. InvertedDoublePendulum is excluded before scientific outcome evaluation because the random-policy dataset produces no valid 32-step held-out rollout windows (the longest held-out fragment is shorter than the frozen horizon). We therefore report four pre-specified tasks, three technically viable tasks, and no post-hoc replacement. Across the viable tasks, baseline+kNN $\Delta$AUROC and $\Delta$AUPRC are positive in all 15 independently trained models; kNN alone is above chance and failure-associated states are farther from support in all 15.

Table 9: Broader-environment replication under the canonical $h = 5$ protocol.

| Env. | $\Delta$AUROC | Pos. | $\Delta$AUPRC | Pos. | kNN AUROC |
|---|---|---|---|---|---|
| Walker2d | +0.0179 | 5/5 | +0.0466 | 5/5 | 0.690 |
| Ant | +0.0361 | 5/5 | +0.0271 | 5/5 | 0.576 |
| Reacher | +0.0270 | 5/5 | +0.0763 | 5/5 | 0.749 |

## F REPRESENTATION ROBUSTNESS

We recompute local geometry for the frozen state-based diagnostic data under raw coordinates, per-dimension standardization, covariance whitening, and a deterministic orthogonal rotation. Data-dependent transforms are fitted only on the nominal training/reference latent set. Orthogonal rotation changes kNN distances by at most approximately $8.58 \times 10^{-7}$, consistent with numerical error. Standardization and whitening can change effect magnitude and occasionally one seed's sign, but

do not systematically reverse the predictive effect. More directly, failure-minus-success support distance remains positive in 5/5 seeds for every evaluated condition and transform.

Table 10: Representation robustness. ΔAUROC is baseline+kNN minus baseline. $\Delta d$ is failure-minus-success kNN support distance.

| Env. | Condition | Transform | ΔAUROC | Pos. | Mean $\Delta d$ | Pos. |
|---|---|---|---|---|---|---|
| HalfCheetah | IID | Raw | $0.0102 \pm 0.0063$ | 5/5 | 0.096 | 5/5 |
| HalfCheetah | IID | Standardized | $0.0081 \pm 0.0049$ | 5/5 | 0.091 | 5/5 |
| HalfCheetah | IID | Whitened | $0.0045 \pm 0.0032$ | 4/5 | 0.165 | 5/5 |
| HalfCheetah | IID | Orthogonal | $0.0102 \pm 0.0063$ | 5/5 | 0.096 | 5/5 |
| HalfCheetah | High friction | Raw | $0.0089 \pm 0.0031$ | 5/5 | 0.086 | 5/5 |
| HalfCheetah | High friction | Standardized | $0.0070 \pm 0.0018$ | 5/5 | 0.082 | 5/5 |
| HalfCheetah | High friction | Whitened | $0.0032 \pm 0.0015$ | 5/5 | 0.150 | 5/5 |
| HalfCheetah | High friction | Orthogonal | $0.0089 \pm 0.0031$ | 5/5 | 0.086 | 5/5 |
| Hopper | IID | Raw | $0.0154 \pm 0.0056$ | 5/5 | 0.469 | 5/5 |
| Hopper | IID | Standardized | $0.0158 \pm 0.0058$ | 5/5 | 0.524 | 5/5 |
| Hopper | IID | Whitened | $0.0178 \pm 0.0117$ | 5/5 | 1.040 | 5/5 |
| Hopper | IID | Orthogonal | $0.0154 \pm 0.0056$ | 5/5 | 0.469 | 5/5 |
| Hopper | High friction | Raw | $0.0342 \pm 0.0151$ | 5/5 | 0.351 | 5/5 |
| Hopper | High mass | Raw | $0.0092 \pm 0.0110$ | 4/5 | 0.471 | 5/5 |
| Hopper | High friction | Standardized | $0.0346 \pm 0.0140$ | 5/5 | 0.391 | 5/5 |
| Hopper | High mass | Standardized | $0.0100 \pm 0.0104$ | 4/5 | 0.527 | 5/5 |
| Hopper | High friction | Whitened | $0.0328 \pm 0.0215$ | 5/5 | 0.725 | 5/5 |
| Hopper | High mass | Whitened | $0.0089 \pm 0.0112$ | 4/5 | 1.054 | 5/5 |
| Hopper | High friction | Orthogonal | $0.0342 \pm 0.0151$ | 5/5 | 0.351 | 5/5 |
| Hopper | High mass | Orthogonal | $0.0092 \pm 0.0110$ | 4/5 | 0.471 | 5/5 |

## G IMAGE-DERIVED AND TEMPORAL VISUAL REPLICATION

The original auxiliary visual experiment uses Pendulum-v1 with synthetic $48 \times 48$ grayscale frames and a 32-D convolutional latent model. A single frame encodes angle but not angular velocity, so the observation is partially observable. The original single-frame visual experiment establishes diagnostic transfer under IID and gravity-shift conditions but inconsistent fixed-penalty control.

A separate paired temporal extension tests whether the missing short-term motion information explains that intervention weakness. For each of five seeds, single-frame and four-frame-stack models are trained from the same 60,000-transition random-policy dataset. The stack is $[I_{t-3}, I_{t-2}, I_{t-1}, I_t]$, with initial-frame repetition at episode starts. The decoder predicts only the newest next frame. The extension uses a 24-step diagnostic rollout, $h = 5$, strict train-group-only 80th-percentile failure calibration, $k = 10$, IID and gravity$\times 1.35$ conditions, and a frozen MPC penalty $\lambda = 0.50$.

Table 11: Paired temporal visual extension. $\Delta R$ is geometry-penalized minus vanilla MPC return.

| Input | Condition | ΔAUROC | Pos. | $\Delta R$ | Pos. |
|---|---|---|---|---|---|
| single | IID | +0.0441 | 5/5 | +26.3 | 5/5 |
| single | High gravity | +0.0393 | 5/5 | -37.4 | 0/5 |
| stack4 | IID | +0.0954 | 5/5 | -67.9 | 1/5 |
| stack4 | High gravity | +0.0531 | 4/5 | -21.5 | 0/5 |

Under IID dynamics, stack4 increases the mean incremental diagnostic from $0.0441$ to $0.0954$. The paired stack4-minus-single gain is $+0.0513$ (4/5 positive; 95% CI $[0.0090, 0.0997]$). The corresponding intervention moves in the opposite direction. Single-frame geometry improves return by $+26.3$ on average (5/5), whereas stack4 reduces return by $-67.9$ (1/5), giving a paired change of $-94.2$. Under high gravity, stack4 has a higher mean diagnostic, but the paired diagnostic interval crosses zero. Both fixed penalties remain harmful, although stack4 is less harmful. Temporal information therefore does not provide a sufficient explanation for the fixed intervention's fragility.

## H DENSE DYNAMICS-SHIFT SWEEPS

The dense sweeps show that geometry's incremental value depends on the environment–perturbation family rather than on a universal scalar notion of physical shift magnitude. Tables 12–13 report mean ΔAUROC for baseline+geometry minus baseline, recomputed from the per-seed dense score table.

Table 12: HalfCheetah dense dynamics sweeps (mean ± SD over five seeds).

| Axis | Scale | Mean ΔAUROC | SD | Positive seeds |
|---|---|---|---|---|
| mass | 0.80 | 0.0134 | 0.0039 | 5/5 |
| mass | 0.90 | 0.0150 | 0.0034 | 5/5 |
| mass | 1.00 | 0.0141 | 0.0033 | 5/5 |
| mass | 1.05 | 0.0158 | 0.0048 | 5/5 |
| mass | 1.10 | 0.0137 | 0.0042 | 5/5 |
| mass | 1.15 | 0.0150 | 0.0035 | 5/5 |
| mass | 1.25 | 0.0159 | 0.0051 | 5/5 |
| mass | 1.35 | 0.0184 | 0.0059 | 5/5 |
| mass | 1.50 | 0.0185 | 0.0043 | 5/5 |
| friction | 0.70 | 0.0146 | 0.0046 | 5/5 |
| friction | 0.85 | 0.0143 | 0.0049 | 5/5 |
| friction | 1.00 | 0.0141 | 0.0033 | 5/5 |
| friction | 1.10 | 0.0143 | 0.0065 | 5/5 |
| friction | 1.20 | 0.0152 | 0.0044 | 5/5 |
| friction | 1.35 | 0.0140 | 0.0037 | 5/5 |
| friction | 1.50 | 0.0133 | 0.0031 | 5/5 |

Table 13: Hopper dense dynamics sweeps (mean ± SD over five seeds).

| Axis | Scale | Mean ΔAUROC | SD | Positive seeds |
|---|---|---|---|---|
| mass | 0.80 | 0.0110 | 0.0035 | 5/5 |
| mass | 0.90 | 0.0132 | 0.0075 | 5/5 |
| mass | 1.00 | 0.0094 | 0.0094 | 4/5 |
| mass | 1.05 | 0.0137 | 0.0083 | 5/5 |
| mass | 1.10 | 0.0096 | 0.0044 | 5/5 |
| mass | 1.15 | 0.0108 | 0.0065 | 5/5 |
| mass | 1.25 | 0.0101 | 0.0067 | 5/5 |
| mass | 1.35 | 0.0094 | 0.0046 | 5/5 |
| mass | 1.50 | 0.0054 | 0.0051 | 4/5 |
| friction | 0.70 | 0.0194 | 0.0094 | 5/5 |
| friction | 0.85 | 0.0137 | 0.0062 | 5/5 |
| friction | 1.00 | 0.0094 | 0.0094 | 4/5 |
| friction | 1.10 | 0.0182 | 0.0080 | 5/5 |
| friction | 1.20 | 0.0246 | 0.0096 | 5/5 |
| friction | 1.35 | 0.0517 | 0.0103 | 5/5 |
| friction | 1.50 | 0.0494 | 0.0294 | 5/5 |

## I GLOBAL SUPPORT-SHIFT FALSIFICATION

The pooled global-support analysis initially suggested a strong relationship between classifier-based latent-distribution separability and geometry gain. Table 14 reproduces that descriptive result from the condition-level means. The subsequent controlled analysis and held-out-family test show why it is not used as the final mechanism claim.

Table 14: Pooled condition-level regressions from the global-support analysis. $\text{MMD}^2$ denotes squared maximum mean discrepancy.

| Predictor | $R^2$ | Pearson $r$ | Spearman $\rho$ |
|---|---|---|---|
| Two-sample AUROC | 0.551 | 0.742 | 0.582 |
| Parameter distance | 0.125 | 0.353 | 0.241 |
| $\text{MMD}^2$ | 0.007 | 0.083 | 0.552 |
| Sliced Wasserstein | 0.016 | 0.126 | 0.488 |
| Cross-kNN | 0.004 | 0.062 | 0.418 |

Table 15: Controlled and seed-bootstrap result for two-sample AUROC in the global-support analysis. The controlled model additionally includes parameter distance, failure rate, environment, and perturbation axis. The reported $p$-value and confidence interval use HC3 heteroskedasticity-robust standard errors.

| Metric | Coef. | $p$ | HC3 95% CI | Boot. positive frac. | Boot. 95% CI |
|---|---|---|---|---|---|
| Two-sample AUROC | 0.0222 | 0.316 | [-0.0212, 0.0657] | 0.937 | [-0.0066, 0.0751] |

## J FAILURE-CONDITIONED LOCAL GEOMETRY

The local analysis contains 160 dense rollout files: 35 HalfCheetah-friction, 45 HalfCheetah-mass, 35 Hopper-friction, and 45 Hopper-mass files. In every file, failure-associated states have larger mean kNN support distance than success-associated states, and kNN failure AUROC is above 0.5. The family-level means appear in Table 2. The audit recomputation preserves all 160 file-level rows in the frozen file-level audit table.

## K CANONICAL CONTROL WITH TRAINING-ONLY SUPPORT

The canonical geometry controller uses only the first 80% training partition as the kNN support reference. Original seeds 0–4 were reevaluated under this training-only definition without retraining the models, and their vanilla episode returns matched the original evaluations exactly. Five additional independently trained models, seeds 5–9, were then evaluated under the same protocol. Table 17 reports the final paired $n = 10$ comparisons.

The cohort split is itself informative. Hopper effects are similar in the reevaluated original and independent follow-up cohorts, whereas HalfCheetah's independent follow-up cohort is much weaker than the original one.

Table 16: Leave-one-family-out transfer for two-sample AUROC. Negative held-out correlations, especially Hopper friction, argue against a universal scalar support-shift mechanism.

| Held-out family | Linear $r$ | Spline $r$ |
|---|---|---|
| HalfCheetah / friction | -0.147 | -0.140 |
| HalfCheetah / mass | -0.086 | 0.102 |
| Hopper / friction | -0.719 | -0.708 |
| Hopper / mass | -0.273 | -0.273 |

Table 17: Canonical $n = 10$ paired control summaries. Positive counts are trained-model seeds out of ten.

| Env. | Condition | Comparison | Mean | Median | Pos. | 95% CI |
|---|---|---|---|---|---|---|
| HalfCheetah | IID | Geo.-Van. | 69.94 | 11.95 | 8/10 | [12.73, 155.70] |
| HalfCheetah | IID | Unc.-Van. | 77.03 | 6.39 | 6/10 | [3.44, 169.63] |
| HalfCheetah | IID | Geo.-Unc. | -7.09 | 13.41 | 7/10 | [-55.29, 28.10] |
| HalfCheetah | High friction | Geo.-Van. | 57.57 | 8.43 | 6/10 | [-5.42, 170.37] |
| HalfCheetah | High friction | Unc.-Van. | 87.80 | 28.87 | 8/10 | [9.09, 215.59] |
| HalfCheetah | High friction | Geo.-Unc. | -30.23 | -21.75 | 1/10 | [-57.81, -7.82] |
| Hopper | IID | Geo.-Van. | 86.55 | 98.51 | 9/10 | [58.64, 111.20] |
| Hopper | IID | Unc.-Van. | 46.37 | 56.43 | 7/10 | [12.79, 77.52] |
| Hopper | IID | Geo.-Unc. | 40.18 | 40.94 | 6/10 | [8.05, 73.59] |
| Hopper | High mass | Geo.-Van. | 70.11 | 67.17 | 10/10 | [52.45, 90.30] |
| Hopper | High mass | Unc.-Van. | 37.43 | 38.34 | 7/10 | [2.39, 73.50] |
| Hopper | High mass | Geo.-Unc. | 32.68 | 25.05 | 8/10 | [-3.54, 67.71] |
| Hopper | High friction | Geo.-Van. | 51.61 | 37.91 | 9/10 | [26.87, 79.10] |
| Hopper | High friction | Unc.-Van. | -1.64 | -0.72 | 5/10 | [-45.79, 43.72] |
| Hopper | High friction | Geo.-Unc. | 53.25 | 43.38 | 9/10 | [24.51, 83.50] |

Table 18: Geometry-minus-vanilla return by cohort. The original cohort is reevaluated using the training-only support reference.

| Env. | Condition | Cohort | Mean | Median | Pos. |
|---|---|---|---|---|---|
| HalfCheetah | IID | Original reevaluated | 132.05 | 100.54 | 5/5 |
| HalfCheetah | IID | Independent follow-up | 7.82 | 4.05 | 3/5 |
| HalfCheetah | High friction | Original reevaluated | 106.99 | 11.16 | 3/5 |
| HalfCheetah | High friction | Independent follow-up | 8.15 | 5.69 | 3/5 |
| Hopper | IID | Original reevaluated | 99.26 | 112.51 | 5/5 |
| Hopper | IID | Independent follow-up | 73.84 | 84.51 | 4/5 |
| Hopper | High mass | Original reevaluated | 76.20 | 73.56 | 5/5 |
| Hopper | High mass | Independent follow-up | 64.02 | 52.18 | 5/5 |
| Hopper | High friction | Original reevaluated | 49.25 | 32.61 | 4/5 |
| Hopper | High friction | Independent follow-up | 53.97 | 41.87 | 5/5 |

## L PLANNER-EXPLOITATION PROBES

These exploratory probes predate the training-only support-reference correction and use the broader collected nominal reference. They are retained as descriptive mechanistic evidence rather than canonical intervention validation. At each diagnostic point, we snapshot the exact MuJoCo position and velocity state (`qpos`/`qvel`), replay the planner-selected action sequence open-loop in an identically shifted probe environment, and compare imagined predictions against the real trajectory. Hopper probe counts differ across methods because episodes can terminate at different times; HalfCheetah has 450 probes for each method and condition. Probe-row correlations are descriptive mechanistic evidence and are not treated as independent model replications.

Table 19: Audited planner-exploitation summary recomputed from all stored diagnostic rows.

| Env. | Condition | Mode | Probes | kNN risk | State RMSE | Reward err. | Optimism |
|---|---|---|---|---|---|---|---|
| HC | friction | penalty | 450 | 1.942 | 0.664 | 0.347 | 0.984 |
| HC | friction | vanilla | 450 | 2.529 | 0.870 | 0.435 | 1.989 |
| HC | IID | penalty | 450 | 1.947 | 0.657 | 0.348 | 0.965 |
| HC | IID | vanilla | 450 | 2.185 | 0.797 | 0.402 | 1.668 |
| Hop | friction | penalty | 142 | 3.303 | 0.393 | 0.171 | 1.120 |
| Hop | friction | vanilla | 118 | 4.604 | 0.542 | 0.256 | 1.956 |
| Hop | mass | penalty | 152 | 3.195 | 0.346 | 0.181 | 0.571 |
| Hop | mass | vanilla | 97 | 5.284 | 0.659 | 0.239 | 1.428 |
| Hop | IID | penalty | 130 | 2.942 | 0.309 | 0.154 | 0.084 |
| Hop | IID | vanilla | 85 | 5.106 | 0.651 | 0.222 | 1.021 |

Table 20: Pearson correlations between mean kNN risk and model-error quantities on vanilla-selected plans. Probe rows are temporally correlated, so effect size and cross-condition consistency are emphasized rather than probe-level significance.

| Env. | Condition | State RMSE | Reward error | Return optimism |
|---|---|---|---|---|
| HC | IID | 0.816 | 0.645 | 0.443 |
| HC | friction | 0.901 | 0.759 | 0.478 |
| Hop | IID | 0.659 | 0.722 | 0.255 |
| Hop | mass | 0.652 | 0.734 | 0.239 |
| Hop | friction | 0.330 | 0.572 | -0.028 |

## M PENALTY SENSITIVITY

This exploratory Hopper sweep predates the training-only support-reference correction and is not part of the canonical $n = 10$ control claim. It remains useful as a sensitivity analysis of the original controller. It is not an optimization study. Table 21 reports every nonzero penalty's seed-level return change relative to that seed's $\lambda = 0$ run together with the selected-plan mean kNN risk.

Table 21: Hopper penalty sweep by trained-model seed.

| Condition | Seed | $\lambda$ | $\Delta R$ vs. $\lambda = 0$ | Mean selected kNN risk |
|---|---|---|---|---|
| friction | 0 | 0.25 | 36.9 | 3.096 |
| friction | 0 | 0.50 | 42.0 | 2.802 |
| friction | 0 | 1.00 | 1.6 | 2.311 |
| friction | 1 | 0.25 | 139.4 | 3.977 |
| friction | 1 | 0.50 | 168.6 | 4.342 |
| friction | 1 | 1.00 | 170.8 | 3.788 |
| friction | 2 | 0.25 | 21.3 | 3.037 |
| friction | 2 | 0.50 | 22.8 | 2.756 |
| friction | 2 | 1.00 | 42.2 | 2.579 |
| friction | 3 | 0.25 | -1.0 | 4.469 |
| friction | 3 | 0.50 | 4.2 | 3.889 |
| friction | 3 | 1.00 | 10.1 | 3.676 |
| friction | 4 | 0.25 | -13.7 | 4.141 |
| friction | 4 | 0.50 | -13.4 | 3.457 |
| friction | 4 | 1.00 | -19.5 | 2.911 |
| mass | 0 | 0.25 | 54.5 | 2.971 |
| mass | 0 | 0.50 | 73.0 | 2.885 |
| mass | 0 | 1.00 | 30.8 | 2.285 |
| mass | 1 | 0.25 | 24.0 | 3.222 |
| mass | 1 | 0.50 | 85.8 | 3.633 |
| mass | 1 | 1.00 | 52.9 | 2.989 |
| mass | 2 | 0.25 | 6.3 | 2.121 |
| mass | 2 | 0.50 | 122.2 | 2.787 |
| mass | 2 | 1.00 | 142.2 | 2.562 |
| mass | 3 | 0.25 | 44.7 | 3.477 |
| mass | 3 | 0.50 | 105.9 | 3.397 |
| mass | 3 | 1.00 | 93.3 | 2.953 |
| mass | 4 | 0.25 | 56.8 | 4.536 |
| mass | 4 | 0.50 | 47.9 | 3.534 |
| mass | 4 | 1.00 | -24.1 | 2.024 |
| IID | 0 | 0.25 | 60.7 | 2.867 |
| IID | 0 | 0.50 | 68.3 | 2.490 |
| IID | 0 | 1.00 | 125.0 | 2.661 |
| IID | 1 | 0.25 | 62.2 | 3.734 |
| IID | 1 | 0.50 | 64.2 | 3.211 |
| IID | 1 | 1.00 | 34.1 | 2.860 |
| IID | 2 | 0.25 | 34.6 | 2.855 |
| IID | 2 | 0.50 | 42.3 | 2.453 |
| IID | 2 | 1.00 | 77.4 | 2.489 |
| IID | 3 | 0.25 | 116.2 | 4.252 |
| IID | 3 | 0.50 | 116.3 | 3.467 |
| IID | 3 | 1.00 | 138.6 | 3.981 |
| IID | 4 | 0.25 | 103.2 | 4.005 |
| IID | 4 | 0.50 | 104.5 | 3.373 |
| IID | 4 | 1.00 | 96.1 | 2.729 |

## N UNCERTAINTY COMPLEMENTARITY

For the canonical control comparator, each trained world-model seed receives three bootstrap latent-dynamics heads, training-distribution 90th-percentile scale matching, and the same $\lambda = 0.50$ as geometry, with no return-based tuning. At $n = 10$, geometry exceeds disagreement in Hopper IID and high friction, is inconclusive in Hopper high mass and HalfCheetah IID, and is worse under HalfCheetah high friction. Thus neither control penalty is universally superior.

A separate diagnostic-complementarity experiment asks whether local support contributes information after disagreement and a global Mahalanobis distance in the learned latent space. Three bootstrap dynamics heads are trained per seed under a frozen protocol. The primary comparison is the incremental kNN gain after adding disagreement to the baseline. kNN remains positive in all five seeds of both environments, and remains positive even after conditioning jointly on disagreement and Mahalanobis distance.

Table 22: Diagnostic complementarity at $h = 5$. Rows report the incremental AUROC from adding kNN to the indicated predictor set. CIs bootstrap independently trained seeds.

| Env. | Added comparison | Mean ΔAUROC | Pos. | 95% CI |
|---|---|---|---|---|
| HalfCheetah | kNN \| baseline | +0.0612 | 5/5 | [+0.0439, +0.0837] |
| HalfCheetah | kNN \| baseline+disagr. | +0.0433 | 5/5 | [+0.0311, +0.0552] |
| HalfCheetah | kNN \| baseline+Mahalanobis | +0.0415 | 5/5 | [+0.0314, +0.0506] |
| HalfCheetah | kNN \| baseline+disagr.+Mahalanobis | +0.0372 | 5/5 | [+0.0265, +0.0474] |
| Hopper | kNN \| baseline | +0.0118 | 5/5 | [+0.0082, +0.0155] |
| Hopper | kNN \| baseline+disagr. | +0.0109 | 5/5 | [+0.0064, +0.0157] |
| Hopper | kNN \| baseline+Mahalanobis | +0.0036 | 5/5 | [+0.0017, +0.0055] |
| Hopper | kNN \| baseline+disagr.+Mahalanobis | +0.0035 | 5/5 | [+0.0016, +0.0055] |

The strongest conditioning set (baseline+disagreement+Mahalanobis) still yields mean kNN increments of $+0.0372$ in HalfCheetah and $+0.00354$ in Hopper, positive in 5/5 seeds each. The smaller Hopper increment is consistent with greater redundancy: kNN is more strongly correlated with both Mahalanobis distance and disagreement there. We therefore claim partial complementarity, not independence or universal superiority.

## O CALIBRATED INTERVENTION

We test one pre-specified attempt to map support risk into a more model-relative control cost. For each imagined latent, raw kNN distance is converted to its empirical percentile under the training-support distribution. The risk weight is zero through the 90th percentile and then increases linearly to one at the maximum percentile. The resulting weight is multiplied by the median absolute training reward and the same frozen $\lambda = 0.50$. No return-based tuning is performed. Ten independently trained models are evaluated in the five canonical state-control conditions.

Table 23: Calibrated intervention. Returns are seed-level paired mean differences. "Cal." uses the frozen 90th-percentile training-relative tail penalty.

| Env. | Condition | Raw–Van. | Cal.–Van. | Cal.–Raw |
|---|---|---|---|---|
| HalfCheetah | IID | +69.0 | +59.0 | -10.0 |
| HalfCheetah | High friction | +41.6 | -6.1 | -47.7 |
| Hopper | IID | +86.4 | +19.5 | -66.9 |
| Hopper | High mass | +74.4 | +12.4 | -61.9 |
| Hopper | High friction | +53.2 | +4.3 | -48.8 |

Calibration underperforms raw geometry on average in all five conditions. The calibrated-minus-raw 95% CI lies entirely below zero in four conditions, with HalfCheetah IID as the only inconclusive case. This does not imply that calibration is generally harmful. It shows only that the tested frozen mapping does not improve control performance. More broadly, training-relative normalization is not a neutral preprocessing step and does not by itself close the gap between diagnosis and intervention.

## P RUNTIME AND SUPPORT-BANK SCALABILITY

We evaluate nested deterministic support subsets $5\% \subset 10\% \subset 25\% \subset 100\%$ of the 120,000-state training support bank. The end-to-end audit measures kNN query time, held-out support-score agreement, matched-candidate planner agreement, and control-return preservation using 10 independently trained models per environment. Runtime evaluation uses 40 timed queries after warmup. Support fidelity uses 3,000 query states. Planner probes use 16 observations per seed with identical candidate action sequences. Control evaluation uses the five canonical conditions, five episodes per controller and seed, horizon 12, 256 candidates, and fixed $\lambda = 0.50$.

Control preservation is summarized using the pre-specified strong, moderate, and weak engineering categories relative to full-bank seed-level return variability. These categories describe preservation under the reduced support bank and are not formal equivalence tests.

Table 24: Support-bank compute and fidelity tradeoff. Spearman and selection agreement are reported as HalfCheetah/Hopper. Control preservation reports the number of conditions in each pre-specified engineering category.

| Bank | Speedup | Support $\rho$ | Planner selection | Control preservation |
|---|---|---|---|---|
| 25% | 2.28× | 0.965/0.986 | 0.863/0.794 | 4 strong, 1 moderate, 0 weak |
| 10% | 3.10× | 0.946/0.978 | 0.794/0.656 | 3 strong, 1 moderate, 1 weak |
| 5% | 3.80× | 0.932/0.972 | 0.719/0.613 | 2 strong, 1 moderate, 2 weak |

At 25%, mean query time drops from 203.2 to 89.6 ms in HalfCheetah and from 197.1 to 86.1 ms in Hopper, corresponding to about $2.3\times$ speedup. Support-score Spearman is 0.965/0.986, candidate-score Spearman is 0.995/0.996, and selected-candidate agreement is 86.3%/79.4%. Four of five control conditions are classified as strong, while Hopper IID is moderate. None are weak. At 10% and 5%, speed reaches about $3.1\times$ and $3.8\times$, but selected-candidate agreement and Hopper control preservation degrade.

Thinning is not a pure computational approximation. Mean raw kNN distance rises as the bank shrinks. At 25%, it reaches about $1.12\times$ the full-bank scale in HalfCheetah and $1.21\times$ in Hopper. With fixed $\lambda$, the effective penalty therefore changes as well. We consequently describe 25% as the strongest tested compute and fidelity compromise rather than as an exact or universally optimal substitute.

The audit also shows that score correlation alone is insufficient. At 5%, candidate-score Spearman remains near 0.99, while exact planner selection falls to 71.9% in HalfCheetah and 61.3% in Hopper. Two Hopper conditions are classified as weak under the pre-specified preservation criteria.

## Q ARTIFACT AUDIT AND REPRODUCIBILITY

The frozen artifact set contains the manuscript supplement together with extension bundles covering residual multi-step architecture transfer, broader environments with a documented technical exclusion, paired temporal visual input, uncertainty complementarity, future-horizon robustness, calibrated intervention, end-to-end support-bank scalability, recurrent stochastic RSSM architecture transfer, and the episode-disjoint split-integrity audit. Each extension preserves its frozen configuration, seed-level or raw result tables, environment and package information or manifests, and analysis and plotting code.

For the recurrent stochastic RSSM extension, the primary seeds 0–4 and the independent seeds 5–9 robustness extension are preserved as separate cohorts.

All reported numerical claims and tables are derived from the corresponding frozen CSV artifacts. Replication counts use independently trained world-model seeds as the inferential unit. Repeated horizons, rollout rows, control episodes, and conditions sharing a trained model are not treated as independent model replications.