PREDICTIVE CLOSURE IN COARSE-GRAINED WORLD MODELS: MEMORY AND RESOLUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Coarse-grained 3D scene simulators—including volumetric latent world models, BEV occupancy representations, and voxelised scene dynamics—replace high-resolution scene states with compact macro representations. Predictive insufficiency in such representations can arise from two distinct sources: inadequate temporal context or loss of predictive information during spatial aggregation.

We introduce a **two-axis predictive closure audit** that separates these effects. TRACK A evaluates *memory closure* by measuring whether extending a length-L macro history reduces out-of-sample prediction error. TRACK B evaluates *feature-relative resolution closure* by testing whether a declared micro-level feature provides additional predictive information beyond the same macro context, using a strictly nested, cross-fitted residual-augmentation procedure.

We further establish two theoretical boundaries for predictive closure: the same observed macro law can arise from both strongly lumpable and non-lumpable micro processes, and finite recursive-state dimension does not imply finite observation-window memory. These results motivate testing predictive closure directly rather than inferring it from macro prediction accuracy or state dimension.

1 INTRODUCTION

Modern world models routinely compress high-dimensional visual or spatial states into coarse latent, occupancy, or voxel representations and rely on temporal memory to recover predictive structure lost by that compression. Predictive accuracy is often treated as evidence that a coarse-grained representation is adequate. This inference is generally unjustified. Coarse graining can preserve useful predictive structure while also inducing non-Markovian memory through unresolved degrees of freedom (Chorin et al., 2000; Geiger, 2026). A coarse state may therefore predict well because it is intrinsically sufficient, or because predictive information discarded by spatial aggregation is recovered indirectly from additional temporal context.

This ambiguity matters because the two cases require different interventions. If predictive information remains in earlier macro history, increasing temporal context may be sufficient. If it remains in discarded micro-scale structure, longer memory cannot certify the representation itself; the aggregation map or retained spatial features must instead be reconsidered. Classical lumpability theory formalises when an aggregated Markov process remains exactly Markovian (Kemeny & Snell, 1976), while more recent information-theoretic work studies coarse-graining and reduced Markov descriptions beyond exact lumpability (Geiger, 2026).

The central question is therefore not simply whether a coarse-grained simulator predicts well, but where any remaining predictive information resides. Relative to the same macro context, it can remain either in additional macro history or in discarded micro-scale structure. Predictive-state and state-space formulations further show that a compact recursive predictive state need not coincide with a finite window of raw observations (Littman & Sutton, 2001; Van Overschee & De Moor, 1994). Distinguishing these sources of predictive information is the focus of this work.

This motivates a two-axis predictive closure audit that separates these possibilities relative to the same macro context. Track A asks whether additional macro history improves prediction beyond a fixed length-L context, thereby testing whether predictive information remains in past macro states.

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.

Track B asks whether a declared micro-level feature improves prediction beyond that same context, using a strictly nested, time-ordered residual-augmentation procedure to test whether predictive information remains in spatial detail discarded by the latent, voxel, or BEV representation. The two tracks therefore differ only in the source from which additional predictive information is allowed to enter: past macro history for Track A, and discarded micro-scale structure for Track B. The audit is deliberately predictor- and feature-relative: it measures usable predictive information under a declared observation protocol rather than claiming to recover unobserved kernel-level structure.

Our contributions are threefold.

(1) We formalise memory closure and feature-relative resolution closure as distinct predictive quantities, separating temporal insufficiency from information loss caused by coarse spatial representation.

(2) We establish theoretical boundaries showing that macro observations alone do not identify strong lumpability and that finite recursive-state dimension does not imply finite observation memory.

(3) We develop a two-track predictive audit and evaluate it on a controlled paired 3D voxel simulator with exact, weak, approximate, and non-lumpable coarse-graining regimes.

These results provide a diagnostic framework for determining whether prediction failure should be addressed by extending temporal context or revising spatial representation.

2 Related Work

2.1 Lumpability and aggregation theory

A projection g of a finite-state Markov chain is *strongly lumpable* when micro states assigned to the same aggregation block induce identical transition probabilities between macro blocks (**?**). This kernel-level notion provides a classical criterion for exact Markovian aggregation. Related work has extended Markov-chain reduction beyond exact lumpability, including information-theoretic formulations of coarse graining and state-space reduction (**?**). Recent studies of imperfect measurements in quantum systems further show how coarse observation can induce apparent memory even when the underlying dynamics are closed (Tabanera-Bravo & Godec, 2025). Together, these results distinguish properties of the underlying transition process from those visible through an aggregated observation.

2.2 Predictive memory and reduced dynamics

Predictive state representations describe dynamical systems through statistics of future observations rather than through an explicitly specified latent physical state (**?**). State-space and subspace-identification methods, such as N4SID, instead recover finite-dimensional recursive realisations from observed input–output sequences (**?**). These approaches distinguish a recursive predictive state from a finite window of raw observations. In particular, finite-dimensional recursive dynamics need not imply a finite-order autoregressive representation, making recursive-state dimension and observation-window length conceptually different measures of predictive memory.

2.3 Mori–Zwanzig memory learning

The Mori–Zwanzig formalism characterises coarse-grained dynamics through a projected drift, a history-dependent memory term, and an orthogonal fluctuation. Both the memory kernel and the fluctuation depend on the chosen projection, so temporal dependence in reduced dynamics is closely tied to the information removed during coarse graining.

Recent learning-based approaches incorporate this structure into data-driven reduced models. DiffGLE uses a differentiable generalised-Langevin formulation with coloured noise to reproduce coarse-grained dynamical statistics (Jeong et al., 2024). Graph-based coarse-grained modelling has also used the structure of leading Mori–Zwanzig memory terms to relate interaction range to message-passing depth and memory length (Yu et al., 2025). Complementary spectral approaches analyse non-Markovian reduction errors through projection or commutation defects (López Villanueva, 2026). These studies collectively emphasise that memory can arise as a structural consequence of coarse graining rather than simply as an architectural property of the predictive model.

2.4 MEMORY-AUGMENTED 3D AND WORLD MODELS

Recent spatio-temporal foundation models apply large Transformer and diffusion architectures to heterogeneous urban dynamics. UrbanDiT scales diffusion transformers to multiple urban spatio-temporal data types and tasks (Yuan et al., 2024), while OpenCity develops transferable spatio-temporal representations for traffic prediction across cities (Li et al., 2025). More recent diffusion-based urban models extend this direction to conditioned forecasting and cross-city transfer (Imanov et al., 2026).

A related line of work introduces explicit memory for long-horizon visual world generation. DecMem combines sparse global memory with anchored local memory to maintain long-term consistency (Yang et al., 2026), whereas AnchorWeave retrieves multiple local geometric memories to preserve spatial consistency across extended camera trajectories (Wang et al., 2026). These developments make temporal memory an increasingly explicit component of long-horizon world modelling.

Across these literatures, temporal dependence after coarse graining and information loss caused by aggregation are closely related but are usually characterised through different theoretical and modelling frameworks. This distinction motivates treating temporal memory and spatial-resolution sufficiency as separate properties of a coarse-grained dynamical representation. "'

3 PROBLEM FORMULATION

The remaining predictive information relative to a fixed macro context can enter from two distinct sources. It may be contained in earlier macro history that lies outside the current length-L context, or in micro-scale structure that has been discarded by the coarse representation. These two sources need not vary together: additional history may improve prediction even when the tested micro-level feature does not, while discarded micro-scale information may remain predictive even after the macro history is sufficient. We therefore treat temporal memory and representational resolution as two separate dimensions of predictive closure.

Let $X_t \in \mathbb{R}^{d_X}$ denote the micro state at time t, and let $Y_t = g(X_t) \in \mathbb{R}^{d_Y}$ be the declared macro observation under an aggregation map g, where d_X and d_Y need not be equal. Let $H_t = (Y_1, \ldots, Y_t)$ denote the complete observed macro history, and $H_t^{(L)} = Y_{t-L+1:t}$ the most recent length-L macro context. When paired micro–macro observations are available, let $\phi_t = \phi(X_t, g(X_t))$ denote a predeclared micro-level feature derived from the paired state.

We consider squared prediction loss throughout the formulation. The two closure quantities differ only in the source of additional information compared against the same declared macro context $H_t^{(L)}$: earlier macro history for memory closure, and discarded micro-scale structure for feature-relative resolution closure.

3.1 MEMORY CLOSURE

Memory closure asks whether macro history preceding the most recent L observations contains additional predictive information about Y_{t+1}.

Define the Bayes risks $R_A^\star(L) \inf_f \mathbb{E}\left[\left|Y_{t+1} - f(H_t^{(L)})\right|^2\right]$,

and $R_A^\star(\infty) = \inf_f \mathbb{E}\left[|Y_{t+1} - f(H_t)|^2\right]$.

The population memory gap is

$$D_A(L) = R_A^\star(L) - R_A^\star(\infty) \geq 0. \tag{1}$$

Under squared loss,

$$D_A(L) = \mathbb{E}\left[\left|\mathbb{E}[Y_{t+1} \mid H_t] - \mathbb{E}[Y_{t+1} \mid H_t^{(L)}]\right|^2\right]. \tag{2}$$

Definition 1 (Memory closure)**.** *The declared macro representation satisfies memory closure at context length L and tolerance ϵ_A when $D_A(L) \leq \epsilon_A$.*

Thus, memory closure is relative to a declared context length and predictive tolerance. It concerns information retained in earlier macro history, not information hidden at the micro resolution.

3.2 Feature-Relative Resolution Closure

Resolution closure asks whether a declared micro-level feature provides predictive information beyond the same macro context $H_t^{(L)}$. This question requires paired micro–macro observations.

For a predeclared feature ϕ_t, define $R_B^\star(L,\phi) = \inf_q \mathbb{E}\left[\left|Y_{t+1} - q(H_t^{(L)}, \phi_t)\right|^2\right]$.

The population feature-relative resolution gap is

$$D_B(L,\phi) = R_A^\star(L) - R_B^\star(L,\phi) \geq 0. \tag{3}$$

Definition 2 (Feature-relative resolution closure)**.** *The declared micro feature ϕ satisfies resolution closure relative to context length L and tolerance ϵ_B when $D_B(L,\phi) \leq \epsilon_B$.*

Conversely, $D_B(L,\phi) > \epsilon_B$ indicates that the declared micro feature carries predictive information about Y_{t+1} beyond the length-L macro history.

The definition is explicitly feature-relative. Failure to detect incremental value from one declared feature ϕ_t does not imply that all micro-scale information has been removed by the aggregation map.

4 Theoretical Analysis

4.1 Non-Identifiability of Strong Lumpability

Proposition 1 (Macro observations do not identify strong lumpability)**.** *Two Markov chains can share the same four-state micro state space, the same aggregation map, and the same projected macro law under a common initialization, while differing in lumpability: one is strongly lumpable and the other is not.*

Proof. Let the micro state be $X_t = (Y_t, H_t) \in \{0,1\}^2$, with aggregation map $g(Y_t, H_t) = Y_t$. After sampling Y_{t+1}, let H_{t+1} be resampled independently from $\mathrm{Bernoulli}(1/2)$, and assume $H_0 \mid Y_0 \sim \mathrm{Bernoulli}(1/2)$.

For the first chain, define $\widetilde{q}_{y,h} = \Pr(Y_{t+1} = 1 \mid Y_t = y, H_t = h)$ independently of h: $\widetilde{q}_{0,0} = \widetilde{q}_{0,1} = 0.3, \widetilde{q}_{1,0} = \widetilde{q}_{1,1} = 0.7$. All micro states within the same macro block therefore induce identical next-block distributions, so the chain is strongly lumpable.

For the second chain, let $\widehat{q}_{0,0} = 0.1, \widehat{q}_{0,1} = 0.5, \widehat{q}_{1,0} = 0.5, \widehat{q}_{1,1} = 0.9$. The next-block law depends on the hidden coordinate H_t, so this chain is not strongly lumpable. Its within-block total-variation defect is $\delta_{\mathrm{lump}}(\widehat{P}, g) = \max\{|0.5 - 0.1|, |0.9 - 0.5|\} = 0.4$.

Because $H_t \mid Y_t$ remains uniform under the stated resampling rule, $\Pr(Y_{t+1} = 1 \mid Y_t = 0) = \frac{0.1+0.5}{2} = 0.3$, and $\Pr(Y_{t+1} = 1 \mid Y_t = 1) = \frac{0.5+0.9}{2} = 0.7$. Hence both chains induce the same projected macro transition matrix, $P_Y = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}$, despite having incompatible strong-lumpability properties. □

Proposition 1 shows that the observed macro law alone cannot determine whether the underlying pair (g, P) is strongly lumpable. Therefore, neither $D_A(L)$ nor $D_B(L,\phi)$ is interpreted as a kernel-level lumpability certificate.

4.2 Finite Recursive State vs. Observation Memory

A second distinction concerns the dimension of a recursive predictive state and the length of the observed history required for prediction.

Proposition 2 (Finite recursive state does not imply finite observation memory)**.** *Consider the invertible ARMA*$(1,1)$ *process*

$$(1-\varphi B)Y_t = (1+\vartheta B)e_t, \tag{4}$$

with $|\varphi| < 1, |\vartheta| < 1, \vartheta \neq 0, \varphi + \vartheta \neq 0$*. The process admits a finite-dimensional recursive innovations realisation, while its autoregressive innovation representation is*

$$Y_t = \sum_{k\geq 1} (\varphi+\vartheta)(-\vartheta)^{k-1} Y_{t-k} + e_t, \tag{5}$$

whose lag coefficients are non-zero for all $k \geq 1$*. Thus finite recursive-state dimension does not imply a finite observation-window length sufficient for exact autoregressive prediction.*

For $|\vartheta| < 1$, the autoregressive coefficients decay geometrically, but remain non-zero at every finite lag. Under stable dynamics, the contribution of increasingly distant lags can decay geometrically. It instead establishes that recursive-state dimension and finite-window predictive sufficiency are distinct quantities.

5 Predictive Closure Audit

For TRACK A, let $\mathcal{F}_L$ denote the predictor class using $H_t^{(L)}$, and let $\mathcal{F}_{\text{ref}}$ denote a nested reference class using $H_t^{(L_{\text{ref}})}$, with $L_{\text{ref}} > L$. Define $R_A^{\mathcal{F}}(L) = \inf_{f\in\mathcal{F}_L} \mathbb{E}\left[\left\|Y_{t+1} - f(H_t^{(L)})\right\|^2\right]$, and $R_A^{\mathcal{F}}(L_{\text{ref}}) = \inf_{f\in\mathcal{F}_{\text{ref}}} \mathbb{E}\left[\left\|Y_{t+1} - f(H_t^{(L_{\text{ref}})})\right\|^2\right]$. The operational memory gap is

$$D_A^{\mathcal{F}}(L, L_{\text{ref}}) = R_A^{\mathcal{F}}(L) - R_A^{\mathcal{F}}(L_{\text{ref}}). \tag{6}$$

For TRACK B, let $\mathcal{M}$ denote the declared class of micro-augmentation functions and assume that $M(\phi) \equiv 0$ belongs to $\mathcal{M}$. Define $R_B^{\mathcal{F},\mathcal{M}}(L,\phi) = \inf_{\substack{f\in\mathcal{F}_L \\ m\in\mathcal{M}}} \mathbb{E}\left[\left\|Y_{t+1} - f(H_t^{(L)}) - m(\phi_t)\right\|^2\right]$, and the operational resolution gap

$$D_B^{\mathcal{F},\mathcal{M}}(L,\phi) = R_A^{\mathcal{F}}(L) - R_B^{\mathcal{F},\mathcal{M}}(L,\phi). \tag{7}$$

The operational gaps provide predictor-class-relative empirical counterparts to $D_A(L)$ and $D_B(L,\phi)$, linking the population definitions to the finite predictive models used in the audit.

5.1 Statistical Decision Rules for Predictive Closure

Predictive closure is assessed from the uncertainty around the estimated gap rather than from the point estimate alone. Let ϵ_A and ϵ_B denote the predeclared tolerances for memory and feature-relative resolution closure, respectively. Each track then admits three possible outcomes: closure, detectable predictive insufficiency, or an indeterminate region in which the available evidence does not separate the two.

For TRACK A, let $C_A^{\mathcal{F}}(L, L_{\text{ref}}) = \left[\underline{D}_A^{\mathcal{F}}(L, L_{\text{ref}}), \overline{D}_A^{\mathcal{F}}(L, L_{\text{ref}})\right]$ denote the confidence interval for $\widehat{D}_A^{\mathcal{F}}(L, L_{\text{ref}})$. Memory closure is supported when $\overline{D}_A^{\mathcal{F}}(L, L_{\text{ref}}) \leq \epsilon_A$, while memory insufficiency is supported when $\underline{D}_A^{\mathcal{F}}(L, L_{\text{ref}}) > \epsilon_A$. Intervals crossing ϵ_A remain indeterminate.

For TRACK B, let $C_B^{\mathcal{F},\mathcal{M}}(L,\phi) = \left[\underline{D}_B^{\mathcal{F},\mathcal{M}}(L,\phi), \overline{D}_B^{\mathcal{F},\mathcal{M}}(L,\phi)\right]$ denote the confidence interval for $\widehat{D}_B^{\mathcal{F},\mathcal{M}}(L,\phi)$. Feature-relative resolution closure is supported when $\overline{D}_B^{\mathcal{F},\mathcal{M}}(L,\phi) \leq \epsilon_B$, while incremental micro-scale predictive information is supported when $\underline{D}_B^{\mathcal{F},\mathcal{M}}(L,\phi) > \epsilon_B$. Intervals crossing ϵ_B are likewise left indeterminate.

This three-way decision rule separates practical closure from detectable predictive gain while retaining an explicit region of statistical uncertainty. The resulting conclusions inherit the information available to each track: TRACK A characterizes predictive information accessible through additional macro history, whereas TRACK B characterizes predictive information accessible through the declared micro feature beyond the same macro context. Together, the two decisions form the empirical readout of the closure audit used in the experiments that follow.

5.2 TRACK A: ESTIMATING MEMORY CLOSURE

TRACK A estimates the predictor-class-relative memory gap $D_A^{\mathcal{F}}(L, L_{\mathrm{ref}})$ by comparing prediction from the tested context $H_t^{(L)}$ with a nested longer-history reference $H_t^{(L_{\mathrm{ref}})}$. In the empirical audit, both predictors are regularized linear autoregressive models selected by forward chaining.

For each rolling forecast origin t, fit $\widehat{f}_{L,t} \in \mathcal{F}_L$ and $\widehat{f}_{\mathrm{ref},t} \in \mathcal{F}_{\mathrm{ref}}$ using only observations available before the forecast origin.

Their out-of-sample squared losses are $\ell_{L,t} = \left\| Y_{t+1} - \widehat{f}_{L,t}(H_t^{(L)}) \right\|^2$, and $\ell_{\mathrm{ref},t} = \left\| Y_{t+1} - \widehat{f}_{\mathrm{ref},t}(H_t^{(L_{\mathrm{ref}})}) \right\|^2$.

The empirical memory-gap estimate is

$$\widehat{D}_A^{\mathcal{F}}(L, L_{\mathrm{ref}}) = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} (\ell_{L,t} - \ell_{\mathrm{ref},t}) \,. \tag{8}$$

where $\mathcal{T}$ denotes the set of rolling forecast origins.

A positive $\widehat{D}_A^{\mathcal{F}}(L, L_{\mathrm{ref}})$ indicates that, under the declared predictor classes, the longer macro context improves out-of-sample prediction relative to the tested length-L context. TRACK A evaluates a predeclared grid of context lengths and identifies the smallest tested L for which the operational memory gap is compatible with the tolerance ϵ_A.

The closure quantities can be evaluated with different predictor classes. For Track A, we use regularized linear autoregressive predictors with forward-chaining model selection, together with residual-dependence and recursive-state diagnostics.

5.3 TRACK B: ESTIMATING FEATURE-RELATIVE RESOLUTION CLOSURE

TRACK B evaluates whether the declared micro feature ϕ_t contains predictive information beyond the fixed macro context $H_t^{(L)}$. Paired micro–macro observations are required.

For each rolling forecast origin t, the procedure is as follows.

(1) Fit a macro-only predictor $\widehat{A}_t : H_i^{(L)} \mapsto Y_{i+1}$ on the preceding training window, with hyperparameters selected using forward-chaining inner validation.

(2) Generate time-ordered out-of-fold residuals $e_i = Y_{i+1} - \widehat{A}_t^{(-i)}(H_i^{(L)})$, where $\widehat{A}_t^{(-i)}$ denotes a prediction for observation i obtained without using that observation in fitting.

(3) Fit a micro-augmentation model $\widehat{M}_t : \phi_i \mapsto e_i$. The candidate augmentation class explicitly includes $M(\phi) \equiv 0$, so that the augmented predictor can recover the macro-only predictor exactly.

(4) Refit the selected macro-only predictor on the complete training window and form $\widehat{Y}_{t+1}^{A} = \widehat{A}_t(H_t^{(L)})$, and $\widehat{Y}_{t+1}^{A+M} = \widehat{A}_t(H_t^{(L)}) + \widehat{M}_t(\phi_t)$.

(5) Record the paired out-of-sample loss difference $d_t = \ell\left(Y_{t+1}, \widehat{Y}_{t+1}^{A}\right) - \ell\left(Y_{t+1}, \widehat{Y}_{t+1}^{A+M}\right)$.

The empirical resolution-gap estimate is

$$\widehat{D}_B^{\mathcal{F},\mathcal{M}}(L, \phi) = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} d_t. \tag{9}$$

A positive $\widehat{D}_B^{\mathcal{F},\mathcal{M}}(L, \phi)$ indicates that the declared micro feature provides incremental predictive value beyond the fixed macro context under the tested predictor classes.

The explicit null augmentation is essential. It ensures that the augmented predictor class contains the macro-only predictor and avoids forcing an unhelpful micro component into the forecast.

As in TRACK A, the audit is not tied to a specific regression model. The experiments use regularized linear predictors, while the closure interpretation remains conditional on the declared predictor class and micro feature.

6 EXPERIMENTS

A central question is whether temporal dependence and information lost through coarse spatial aggregation can be distinguished empirically in a world model. Controlled paired voxel simulations provide a setting in which both the underlying micro state and the declared macro representation are observable, making it possible to examine these two sources of predictive insufficiency under the same predictive context.

TRACK A characterizes how the memory-gap profile changes as the coarse-graining structure is systematically perturbed, while TRACK B examines whether the declared micro-scale feature contributes predictive information beyond the fixed macro context. The resulting experiments therefore connect the two closure quantities to distinct empirical signatures of temporal and representational insufficiency.

6.1 CONTROLLED VOXEL BENCHMARK

NewSceneCity v2 is a controlled paired 3D voxel generator in which both the micro state and its declared macro aggregation are observed. The micro state is represented on a $G \times G \times G$ voxel grid with n_{state} channels per voxel. The grid is partitioned into $B \times B \times B$ macro cells, and the declared aggregation map is the block mean.

The experimental configuration uses $G = 4, B = 2, n_{\mathrm{state}} = 4$, yielding a 32-dimensional macro observation. Four controlled coarse-graining regimes are considered: *exact*, *weak*, *approximate*, and *non-lumpable*. Each regime is evaluated over the defect grid $\{0, 0.2, 0.4, 0.6, 0.8\}$.

For Track A, the reference context is $L_{\mathrm{ref}} = 8$, with candidate contexts $L \in \{1, \ldots, 7\}$. Each regime–defect configuration is evaluated using five random seeds and 20 rolling forecast origins per seed, giving 100 paired out-of-sample comparisons per tested context. Regularized linear autoregressive predictors are selected using forward-chaining validation.

For Track B, the declared micro feature is the vectorised current micro state, $\phi_t = \mathrm{vec}(X_t)$. The macro predictor and micro-augmentation model are regularized linear models, with the explicit null augmentation $M(\phi) \equiv 0$ included in the admissible augmentation class.

The per-voxel dynamics are

$$X_i^{(t+1)} = a_i X_i^{(t)} + b\, x_t^{(B(i))} + c \sum_{j \in N(i)} w_{ij} X_j^{(t)} + \sigma_{\mathrm{noise}} \xi_i^{(t)}, \tag{10}$$

where $B(i)$ denotes the macro cell containing voxel i, $x_t^{(B(i))}$ is the block-level driving input broadcast within that cell, and $N(i)$ is the 6-face 3D neighbourhood.

Each trajectory contains 400 evaluated steps after a 200-step warm-up. Five random seeds are used for each synthetic configuration. Uncertainty is quantified using block-bootstrap 95% confidence intervals over rolling forecast origins.

Autoregressive orders selected by MSE–BIC are reported only as auxiliary diagnostics; Track A conclusions are based on the paired operational memory gap rather than on the selected order itself.

6.2 TRACK A: REGIME-SENSITIVE MEMORY-GAP PROFILES

The empirical Track A profile exhibits markedly different behaviour across the four coarse-graining regimes. Table 1 reports the paired loss difference at the longest tested context, $L = 7$, relative to the $L_{\mathrm{ref}} = 8$ reference predictor.

In contrast, the approximate regime exhibits a systematic shift in the empirical memory-gap profile. At $L = 7$, the magnitude of the paired loss difference increases from 5.28×10^{-3} at defect 0 to 2.42×10^{-2} at defect 0.8, corresponding to a $4.6\times$ amplification. The weak regime shows a milder but qualitatively similar trend, while the non-lumpable regime changes only slightly across the same defect grid.

Across the tested context range, the main empirical signal is therefore the regime-dependent shape and displacement of the memory-gap profile. The profile remains stable in the exact construction

Table 1: Track A empirical gap at $L = 7$ under increasing synthetic defect. Values are paired out-of-sample loss differences relative to the $L_{\mathrm{ref}} = 8$ reference predictor.

Regime	0.0	0.2	0.4	0.6	0.8
Exact	-6.18×10^{-3}	-6.18×10^{-3}	-6.18×10^{-3}	-6.18×10^{-3}	-6.18×10^{-3}
Weak	-4.43×10^{-3}	-5.75×10^{-3}	-6.70×10^{-3}	-7.71×10^{-3}	-1.37×10^{-2}
Approximate	-5.28×10^{-3}	-8.45×10^{-3}	-1.05×10^{-2}	-1.27×10^{-2}	-2.42×10^{-2}
Non-lumpable	-8.56×10^{-3}	-8.68×10^{-3}	-8.96×10^{-3}	-9.44×10^{-3}	-1.00×10^{-2}

but changes systematically under weak and approximate coarse-graining perturbations. These results show that Track A captures changes in the predictive role of temporal context induced by the structure of the coarse-grained dynamics.

6.3 TRACK B: SYNTHETIC NULL BEHAVIOUR

The empirical Track B estimate is non-positive across all four coarse-graining regimes, all five defect levels, and all three random seeds per cell. The observed values lie approximately in the range $-3.2 \times 10^{-2} \leq \widehat{D}_B \leq -1.8 \times 10^{-2}$.

No positive incremental predictive contribution from the declared micro feature is detected in these configurations. Under the tested predictor class, the observed pattern is therefore consistent with a feature-relative null: the declared micro-scale information does not produce an out-of-sample predictive gain beyond the fixed macro context. This empirical outcome is compatible with several underlying situations, including a sufficiently informative macro representation, micro-scale information that is not exploitable by the chosen predictor class, or an effect too weak to be resolved at the available sample size.

Importantly, the same non-positive pattern persists despite providing the augmentation with a high-dimensional micro-scale state. The residual-augmentation procedure therefore remains stable in settings where additional micro-scale information yields no detectable incremental predictive value.

6.4 JOINT MEMORY–RESOLUTION BEHAVIOUR

The four controlled settings examine the joint behaviour of the memory and resolution axes. Across these settings, both TRACK A and TRACK B yield non-positive empirical gaps under the current construction.

For Track B, the four estimates are approximately $-6.9 \times 10^{-3}, -1.1 \times 10^{-2}, -1.2 \times 10^{-1}, -2.2 \times 10^{-1}$, with corresponding confidence intervals remaining below zero. Track A shows the same qualitative pattern across the tested context grid. These settings occupy a restricted region of the joint memory–resolution space rather than all possible closure configurations, but provide a stable joint baseline for interpreting the two axes.

The Track A memory-gap profile varies systematically with coarse-graining structure: the exact regime remains invariant to the defect parameter, whereas the weak and approximate regimes exhibit clear shifts at longer tested contexts. Temporal-context dependence is therefore shaped by how the underlying state is spatially aggregated.

Track B shows a different pattern. No positive incremental contribution from the declared micro feature is detected under the tested linear predictor class, even when a richer micro-scale state is available. Across the four-setting calibration, both axes remain free of artificial positive closure signals, although the tested configurations cover only part of the joint memory–resolution space.

Taken together, these results show that predictive sufficiency in a coarse world model cannot be reduced to a single notion of prediction quality. Sensitivity to additional temporal context and predictive information remaining outside the coarse representation capture different aspects of the model state. The two-axis audit separates these effects under a common predictive context, making it possible to distinguish changes associated with temporal context from those associated with state representation.

7 CONCLUSION

This work introduced a two-axis predictive closure framework for coarse-grained world models, separating predictive information that remains in additional macro history from information discarded by the coarse spatial representation. This distinction is formalised through memory closure and feature-relative resolution closure, providing two complementary views of predictive sufficiency under the same macro context. The accompanying theoretical results clarify the limits of what can be inferred from coarse observations alone: an observed macro law does not identify strong lumpability, and the existence of a finite-dimensional recursive predictive state does not imply finite observation memory. The resulting audit translates these distinctions into operational quantities that can be evaluated from paired predictive comparisons, with controlled voxel experiments showing systematic responses to changes in coarse-graining structure and stable behaviour under the tested feature-relative null settings.

Together, these results clarify why predictive accuracy alone cannot determine where the remaining predictive information resides. It may lie in additional temporal context or in structure discarded by the coarse state, and these two cases call for different interventions. Predictive closure therefore provides a principled way to distinguish whether a coarse world model should be improved through additional memory or through a richer state representation.

AI USE STATEMENT

Generative AI tools were used to assist with language editing, clarity, and structural refinement of the manuscript. All mathematical arguments, theoretical claims, experimental design, numerical results, and cited sources were reviewed and verified by the authors. The authors take full responsibility for the content of the paper.

REPRODUCIBILITY STATEMENT

The theoretical results are reproducible from the formal assumptions, definitions, and derivations provided in the paper and appendices. The coarse-graining maps, predictive-closure criteria, memory classes, resolution parameters, and prediction losses are specified explicitly, together with the conditions under which the stated results hold. The examples and counterexamples are defined by complete stochastic models, allowing the corresponding predictive risks and memory requirements to be recomputed directly. Where numerical evaluations are used, the underlying processes, parameter ranges, coarse-graining rules, and evaluation quantities are reported so that the figures and tables can be regenerated independently. Extended proofs and derivations are included in the supplementary material, with no dependence of the main theoretical conclusions on proprietary datasets or inaccessible models.

REFERENCES

Alexandre J. Chorin, Ole H. Hald, and Raz Kupferman. Optimal prediction and the mori–zwanzig representation of irreversible processes. *Proceedings of the National Academy of Sciences*, 97(7): 2968–2973, 2000. doi: 10.1073/pnas.97.7.2968.

Bernhard C. Geiger. Information-theoretic reduction of markov chains. *Computer Science Review*, 59:100802, 2026. doi: 10.1016/j.cosrev.2025.100802.

Olaf Yunus Laitinen Imanov, Derya Umut Kulali, and Taner Yilmaz. Urban spatio-temporal foundation models for climate-resilient housing: Scaling diffusion transformers for disaster risk prediction, 2026.

Jinu Jeong, Ishan Nadkarni, and Narayana R. Aluru. DiffGLE: Differentiable coarse-grained dynamics using generalized langevin equation, 2024.

John G. Kemeny and J. Laurie Snell. *Finite Markov Chains*. Springer-Verlag, New York, 1976. ISBN 978-0-387-90192-3.

Zhonghang Li, Long Xia, Lei Shi, Yong Xu, Dawei Yin, and Chao Huang. Open spatio-temporal foundation models for traffic prediction. *ACM Transactions on Intelligent Systems and Technology*, 2025. doi: 10.1145/3773912.

Michael L. Littman and Richard S. Sutton. Predictive representations of state. In *Advances in Neural Information Processing Systems*, volume 14, pp. 1555–1561, 2001.

Ignacio López Villanueva. A spectral energy framework for non-markovian model reduction error via mori–zwanzig commutation defects. PhilArchive, 2026. Preprint.

Jorge Tabanera-Bravo and Aljaž Godec. Purely quantum memory in closed systems observed via imperfect measurements. *Quantum*, 9:1938, 2025. doi: 10.22331/q-2025-12-11-1938.

Peter Van Overschee and Bart De Moor. N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30(1):75–93, 1994. doi: 10.1016/0005-1098(94)90230-5.

Zun Wang, Han Lin, Jaehong Yoon, Jaemin Cho, Yue Zhang, and Mohit Bansal. AnchorWeave: World-consistent video generation with retrieved local spatial memories, 2026.

Zhenhao Yang, Xiaoshi Wu, Zhengyao Lv, Xiaoyu Shi, Xintao Wang, Pengfei Wan, Kun Gai, and Kwan-Yee K. Wong. DecMem: Towards minute-long consistent world generation with decoupled memory, 2026.

Yin Yu, John Harlim, Daning Huang, and Yan Li. Learning coarse-grained dynamics on graph. *Physica D: Nonlinear Phenomena*, 481:134801, 2025. doi: 10.1016/j.physd.2025.134801.

Yuan Yuan, Chonghua Han, Jingtao Ding, Depeng Jin, and Yong Li. UrbanDiT: A foundation model for open-world urban spatio-temporal learning, 2024.

A Appendix

A.0 Non-identifiability motivation: Proposition 1

The construction in Proposition 1 is reproduced here for completeness; it does not enter the empirical audit. Its purpose is to clarify what can and cannot be identified from the declared macro law alone.

Let the micro state be

$$X_t = (Y_t, H_t) \in \{0,1\}^2,$$

with aggregation map

$$g(Y_t, H_t) = Y_t.$$

After sampling Y_{t+1}, let

$$H_{t+1} \sim \text{Bernoulli}(1/2)$$

be resampled independently, and assume

$$H_0 \mid Y_0 \sim \text{Bernoulli}(1/2).$$

We construct two micro-level Markov chains that induce the same projected macro law under this common initialization, although only one is strongly lumpable.

Model $\tilde{P}$ (strongly lumpable). Define

$$\tilde{q}_{y,h} = (Y_{t+1} = 1 \mid Y_t = y, H_t = h)$$

independently of h, with

$$\tilde{q}_{0,0} = \tilde{q}_{0,1} = 0.3, \qquad \tilde{q}_{1,0} = \tilde{q}_{1,1} = 0.7.$$

Hence all micro states within the same aggregation block induce the same next-block distribution. Therefore the pair $(\tilde{P}, g)$ is strongly lumpable, with projected transition matrix

$$P_Y = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}.$$

Model $\hat{P}$ (non-lumpable). Now define

$$\hat{q}_{0,0} = 0.1, \qquad \hat{q}_{0,1} = 0.5, \qquad \hat{q}_{1,0} = 0.5, \qquad \hat{q}_{1,1} = 0.9.$$

The next-block law therefore depends on the hidden coordinate H_t. For block $Y_t = 0$, the two next-block distributions are

$$(0.9, 0.1) \qquad \text{and} \qquad (0.5, 0.5),$$

while for block $Y_t = 1$ they are

$$(0.5, 0.5) \qquad \text{and} \qquad (0.1, 0.9).$$

Thus the within-block total-variation defect is

$$(\hat{P}, g) = \max\{|0.5 - 0.1|,\ |0.9 - 0.5|\} = 0.4,$$

so $(\hat{P}, g)$ is not strongly lumpable.

Nevertheless, because $H_t \mid Y_t$ remains uniform under the independent resampling rule,

$$(Y_{t+1} = 1 \mid Y_t = 0) = \frac{0.1 + 0.5}{2} = 0.3,$$

and

$$(Y_{t+1} = 1 \mid Y_t = 1) = \frac{0.5 + 0.9}{2} = 0.7.$$

Therefore $\hat{P}$ induces exactly the same projected macro transition matrix

$$P_Y = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$$

as $\tilde{P}$ under the stated initialization.

Implication. The observed macro law alone cannot determine whether the underlying pair (g, P) is strongly lumpable. Consequently, neither the memory closure quantity nor the feature-relative resolution closure quantity is interpreted as a kernel-level lumpability certificate. The two-axis audit instead characterises predictive closure under the declared observation protocol.

A.1 AR(∞) REPRESENTATION FOR PROPOSITION 2

Consider the invertible ARMA$(1, 1)$ process from Proposition 2,

$$(1 - \phi B)Y_t = (1 + \vartheta B)e_t,$$

with

$$|\phi| < 1, \qquad |\vartheta| < 1, \qquad \vartheta \neq 0, \qquad \phi + \vartheta \neq 0.$$

Because $|\vartheta| < 1$, the inverse moving-average operator admits the convergent expansion

$$(1 + \vartheta B)^{-1} = \sum_{k=0}^{\infty} (-\vartheta)^k B^k.$$

Applying this inverse to both sides gives

$$e_t = (1 + \vartheta B)^{-1}(1 - \phi B)Y_t.$$

Expanding,

$$\begin{aligned} e_t &= \sum_{k=0}^{\infty} (-\vartheta)^k Y_{t-k} - \phi \sum_{k=0}^{\infty} (-\vartheta)^k Y_{t-1-k} \\ &= Y_t - (\phi + \vartheta) \sum_{k=1}^{\infty} (-\vartheta)^{k-1} Y_{t-k}. \end{aligned}$$

Rearranging yields the autoregressive innovation representation

$$Y_t = \sum_{k=1}^{\infty} (\phi + \vartheta)(-\vartheta)^{k-1} Y_{t-k} + e_t.$$

Since

$$\vartheta \neq 0 \qquad \text{and} \qquad \phi + \vartheta \neq 0,$$

the coefficient

$$(\phi + \vartheta)(-\vartheta)^{k-1}$$

is non-zero for every finite $k \geq 1$. Hence no finite observation window gives the exact autoregressive predictor, even though the ARMA$(1, 1)$ process admits a finite-dimensional recursive innovations realisation.

For $|\vartheta| < 1$, these coefficients decay geometrically with lag, so increasingly distant observations may have rapidly diminishing predictive effect. The proposition does not claim that arbitrarily distant observations remain practically important; it establishes only that finite recursive-state dimension and finite observation-window sufficiency are distinct notions.

A.2 Generative process details for v2

For completeness, the per-step update at voxel i in macro cell $B(i) \in \{1, \ldots, B\}^3$ is

$$\overset{(t+1)}{i} = a_i \overset{(t)}{i} + b\, \boldsymbol{x}_t^{(B(i))} + c \sum_{j \in N(i)} w_{ij} \overset{(t)}{j} + \sigma_{\text{noise}}\, \xi_i^{(t)},$$

where $B(i)$ denotes the macro cell containing voxel i, $\boldsymbol{x}_t^{(B(i))}$ is the block-level driving input broadcast to voxels in that cell, and $N(i)$ is the 6-face 3D neighbourhood of voxel i.

The micro state is defined on a $G \times G \times G$ voxel grid with n_{state} channels. The grid is partitioned into $B \times B \times B$ macro cells, and the declared aggregation map is the block mean. The resulting macro observation therefore has dimension

$$B^3 n_{\text{state}},$$

whereas the full micro state has dimension

$$G^3 n_{\text{state}}.$$

The controlled coarse-graining regimes vary the degree to which within-block dynamics and spatial coupling are compatible with the declared aggregation. All channels evolve synchronously, and the relaxation, driving, coupling, and noise parameters are restricted to their predeclared ranges.

The experiments reported in the main text use

$$G = 4, \qquad B = 2, \qquad n_{\text{state}} = 4,$$

giving a 32-dimensional macro time series. Each trajectory contains 400 evaluated steps after a 200-step warm-up, with five random seeds and defect levels

$$\{0, 0.2, 0.4, 0.6, 0.8\}.$$

A.3 Detailed Track A synthetic results

Table 2 reports the empirical Track A gap for three representative context lengths over all four coarse-graining regimes and all five defect levels. The reference context is $L_{\text{ref}} = 8$.

The exact regime is invariant over the complete defect grid. At the longest tested context, the weak and approximate regimes show a systematic increase in the magnitude of the empirical paired loss difference as defect increases. The approximate regime changes from -5.28×10^{-3} to -2.42×10^{-2} between defect 0 and 0.8, whereas the non-lumpable construction changes only from -8.56×10^{-3} to -1.00×10^{-2}.

These values are interpreted as predictor-class-relative empirical profiles only. They are not used to identify a minimum sufficient context length, and their signs are not interpreted as population memory-gap signs.

Table 2: Empirical Track A paired loss differences on NewSceneCity v2.

Regime	Defect	$L = 1$	$L = 4$	$L = 7$
Exact	0.0	-5.22×10^{-2}	-2.81×10^{-2}	-6.18×10^{-3}
Exact	0.2	-5.22×10^{-2}	-2.81×10^{-2}	-6.18×10^{-3}
Exact	0.4	-5.22×10^{-2}	-2.81×10^{-2}	-6.18×10^{-3}
Exact	0.6	-5.22×10^{-2}	-2.81×10^{-2}	-6.18×10^{-3}
Exact	0.8	-5.22×10^{-2}	-2.81×10^{-2}	-6.18×10^{-3}
Weak	0.0	-5.20×10^{-2}	-2.26×10^{-2}	-4.43×10^{-3}
Weak	0.2	-4.91×10^{-2}	-2.45×10^{-2}	-5.75×10^{-3}
Weak	0.4	-4.76×10^{-2}	-2.82×10^{-2}	-6.70×10^{-3}
Weak	0.6	-4.50×10^{-2}	-2.92×10^{-2}	-7.71×10^{-3}
Weak	0.8	-4.53×10^{-2}	-3.04×10^{-2}	-1.37×10^{-2}
Approximate	0.0	-5.48×10^{-2}	-2.72×10^{-2}	-5.28×10^{-3}
Approximate	0.2	-5.33×10^{-2}	-3.06×10^{-2}	-8.45×10^{-3}
Approximate	0.4	-4.91×10^{-2}	-3.09×10^{-2}	-1.05×10^{-2}
Approximate	0.6	-4.55×10^{-2}	-3.19×10^{-2}	-1.27×10^{-2}
Approximate	0.8	-5.00×10^{-2}	-3.29×10^{-2}	-2.42×10^{-2}
Non-lumpable	0.0	-5.10×10^{-2}	-2.53×10^{-2}	-8.56×10^{-3}
Non-lumpable	0.2	-5.13×10^{-2}	-2.56×10^{-2}	-8.68×10^{-3}
Non-lumpable	0.4	-5.16×10^{-2}	-2.62×10^{-2}	-8.96×10^{-3}
Non-lumpable	0.6	-5.20×10^{-2}	-2.69×10^{-2}	-9.44×10^{-3}
Non-lumpable	0.8	-5.24×10^{-2}	-2.77×10^{-2}	-1.00×10^{-2}

Table 3: Four-setting calibration experiment. Because all settings produce closure-like behaviour under the declared predictor and feature classes, this experiment is treated as a null calibration rather than a two-by-two validation of the four possible closure states.

Setting	$\widehat{D}_B$	95% CI	$\widehat{D}_A(1)$	$\widehat{D}_A(7)$
1	-6.93×10^{-3}	$[-8.99, -4.87]\times10^{-3}$	-5.67×10^{-2}	-9.87×10^{-3}
2	-1.05×10^{-2}	$[-1.21, -0.889]\times10^{-2}$	-5.26×10^{-2}	-8.24×10^{-3}
3	-1.24×10^{-1}	$[-1.36, -1.12]\times10^{-1}$	-1.14	-2.30×10^{-1}
4	-2.19×10^{-1}	$[-2.22, -2.15]\times10^{-1}$	-1.15	-2.74×10^{-1}

A.4 Four-setting null calibration

We additionally evaluate both audit axes on four controlled sanity settings. Under the current construction, all four settings yield non-positive empirical Track A and Track B estimates.

The purpose of this experiment is limited. It verifies that the current implementation does not generate artificial positive closure gaps in these four null-like settings. It does not establish that the audit separates all four combinations of memory and resolution closure.

A.5 Synthetic Track B null results

On NewSceneCity v1, the empirical Track B estimate is non-positive across all four regimes, all five defect levels, and all three random seeds per cell. The full observed range is approximately

$$-3.19 \times 10^{-2} \leq \widehat{D}_B \leq -1.76 \times 10^{-2}.$$

The exact regime is invariant to the defect parameter. The weak, approximate, and non-lumpable regimes exhibit modest variation, but no cell produces a positive incremental micro-feature contribution.

This is treated as a feature-relative null result. In particular, $\widehat{D}_B \leq 0$ is not interpreted as a population resolution closure certificate.

A.6 Real-data aggregation and stationarity analysis

The real-data case study contains 244 Bournemouth LSOAs observed monthly from January 2016 through June 2025. An exploratory stationarity analysis divides the sequence into three periods and compares block-mean with block-max aggregation.

Table 4: Qualitative sign pattern of the exploratory Track B stationarity analysis.

Period	Aggregation	Track B pattern
2016–2019	block mean	all non-positive
2016–2019	block max	all positive
2020–2022	block mean	all non-positive
2020–2022	block max	mixed
2023–2025	block mean	all non-positive
2023–2025	block max	all non-positive

The pattern is therefore both representation- and regime-dependent: the only uniformly positive exploratory signal occurs for block-max aggregation during 2016–2019.

Because this analysis predates the final strictly nested residual-augmentation protocol, it is reported as exploratory rather than as evidence of feature-relative resolution failure.

A.7 Strictly nested Track B robustness

We repeat the real-data Track B analysis with strict temporal nesting. For each rolling forecast origin, hyperparameters are selected using forward-chaining inner validation, residuals are generated using time-ordered out-of-fold predictions, and the micro augmentation is fit only to those residuals. The null augmentation is retained as an admissible candidate throughout model selection.

Under this procedure, every tested period–spatial-scale cell has a negative empirical Track B estimate. Representative values are:

Table 5: Strictly nested Track B robustness results.

Period	K	$\widehat{D}_B$	95% CI
2016–2025	10	-1.58	$[-2.25, -0.94]$
2016–2025	30	-15.43	$[-21.22, -10.99]$
2016–2025	60	-111.27	$[-144.53, -81.43]$
2016–2019	10	-493.63	$[-598.23, -396.47]$
2020–2022	10	-60.33	$[-63.19, -52.71]$
2023–2025	10	-6.20	$[-8.01, -4.59]$

Across the full set of 12 evaluated cells, 11 block-bootstrap confidence intervals exclude zero. The positive gains observed in earlier, less strictly nested analyses therefore do not survive the final evaluation protocol.

This comparison motivates the strict cross-fitting requirement in Sec. 5.2: without it, residual augmentation can overstate the incremental predictive contribution of the declared micro feature.

A.8 Count-aware Track A robustness

We additionally examine whether the selected macro-memory order is sensitive to the scoring criterion. Three alternatives are considered: squared prediction error, Poisson deviance used as a mean-targeting proper loss, and a negative-binomial criterion allowing additional dispersion.

The modal selected order is therefore $L^\star = 6$ in every reported period–scale configuration and under all three criteria. The negative-binomial selection distribution is broader: at $K = 60$, approximately 22% of LSOAs select an order below six. We interpret this as local heterogeneity in the memory diagnostic rather than as a change in the dominant population-level conclusion.

Table 6: Modal selected Track A order under alternative scoring criteria.

Period	K	MSE	Poisson	Negative binomial
2016–2025	10	6	6	6
2016–2025	30	6	6	6
2016–2025	60	6	6	6
2016–2019	60	6	6	6
2020–2022	60	6	6	6
2023–2025	60	6	6	6

A.9 Reproducibility details

All synthetic experiments use fixed random seeds and identical evaluation protocols within each comparison. Track A results use rolling-origin evaluation with forward-chaining model selection. Track B results use time-ordered residual construction and, in the final protocol, strictly nested cross-fitting.

Synthetic uncertainty and real-data uncertainty are estimated using block-bootstrap intervals with the same declared resampling rules within each experiment. All reported closure statements remain conditional on the declared predictor class, feature construction, aggregation map, and observation horizon.