Beyond DNNs: Noise-Robust Occupation-Kernel Digital Twin

Haowe(Alice) Chen [*] **Rushikesh L Kamalapurkar** [**]
Joel Rosenfeld [***]

[*] *University of Florida, Gainesville FL (e-mail: haowei.chen@ufl.edu).*
[**] *University of Florida, Gainesville FL (e-mail: rkamalapurkar@ufl.edu)*
[***] *University of South Florida, Tampa FL (e-mail: rosenfeldj@usf.edu)*

Abstract: We present an occupation-kernel digital twin (OKDT) that combines the Mori–Zwanzig projection with kernel ridge regression. We discretize the continuous occupation kernel using an exponential envelope, resulting in a training problem that is convex and admits a closed-form solution. The resulting model suppresses sensor noise with an $O(N^{-1/2})$ finite-sample bound, and propagates that bound to long prediction horizons. Benchmarks on cylinder-wake flow, the Van der Pol oscillator, and a noisy linear system show that our approach remains stable where vector kernels and deep neural networks diverge.

Keywords: Digital Twin, Safe Learning, Occupation Kernel, Reduced Mori–Zwanzig Identity

1. INTRODUCTION

Digital twins—virtual, physics-informed models of physical systems—form the backbone of modern health-aware and fault-resilient control architectures Rasheed et al. (2020); Gomez et al. (2025). These models must deliver accurate long-horizon forecasts and quantify uncertainty from noisy and incomplete sensor streams Alcaraz et al. (2023). However, standard deep-learning surrogates lack robustness guarantees: for example, sensor noise can destabilize recurrent neural network (RNN) forecasts Gopalakrishnan et al. (2024), and recent studies show that DeepONet variants can oversmooth high-frequency dynamics and eventually diverge during autoregressive rollouts Wang et al. (2025).

The Mori–Zwanzig formalism reveals that reduced-order models of high-dimensional dynamics inherently follow non-Markovian evolution with exponentially decaying memory kernels Chorin et al. (2017). Classical flow-map learning—which assumes a Markovian update of the form $x_{t+1} = F(x_t)$—ignores this intrinsic memory structure, and conventional deep neural networks provide no systematic mechanism for incorporating decaying-memory effects.

Kernel-based approaches such as the occupation-kernel regression Rosenfeld et al. (2022) and dynamic mode decomposition Rosenfeld et al. (2019) embed trajectories in a reproducing-kernel Hilbert space but remain fundamentally offline and offer no finite-sample error guarantees.

Taken together, these gaps motivate the development of a digital-twin surrogate that: (i) incorporates Mori–Zwanzig-style memory, (ii) is robust to sensor noise, (iii) allows convex, closed-form training, and (iv) provides provable finite-sample error certificates.

In this work, we address this gap by developing a novel Occupation Kernel Digital Twin. Our primary contributions are:

- We derive a discrete weighted occupation kernel by applying numerical quadrature to the continuous Mori–Zwanzig memory integral, directly embedding the system's memory structure into the kernel design.
- We formally establish the kernel's variance-shrinkage properties and prove a rigorous $O(N^{-1/2})$ generalization bound. This is supplemented by a Grönwall-based multi-step error bound that certifies long-horizon predictive accuracy.
- We provide an efficient batch training algorithm (see Algorithm 1) with a complexity of $O(N^2 n_M)$.
- We demonstrate that our method achieves state-of-the-art robustness and accuracy on three challenging benchmarks, significantly outperforming vector-Kernel and DNNs baselines in fault detection tasks.

The developed algorithm provides robust, uncertainty-quantified digital twins that are critically required by next-generation safe control architectures such as safe model predictive control, control barrier function filters, and advanced prognostics-aware controllers. Our work provides a practical and theoretically-grounded tool for building reliable autonomous systems.

The remainder of this paper is organized as follows. Section 2 reviews the necessary preliminaries from the Mori-Zwanzig formalism. Section 3 details the derivation of our memory kernel. Sections 4 and 5 present case studies validating our model with training algorithms. Section 6 reports the experimental results on our benchmark systems. Finally, Section 7 concludes the paper and discusses future work.

2. PROBLEM FORMULATION

2.1 Problem Formulation

Consider a high-dimensional dynamical system with full state $x(t) \in \mathbb{R}^N$. We assume access only to a lower-dimensional projection of the system, $z(t) \in \mathbb{R}^d$, sampled at discrete time intervals Δt. Our goal is to learn the evolution of these resolved variables $z_n \approx z(n\Delta t)$ from noisy observations $\hat{z}_n$, predicting future states z_{n+h} while accounting for the effects of the unresolved (hidden) dynamics.

2.2 Notation

We denote the full state by $x(t) \in \mathbb{R}^N$ and assume the dynamics are governed by a globally Lipschitz map $F \in C^1$, such that $\|F(x) - F(y)\| \leq L_F\|x - y\|$. We define P and Q as constant orthogonal projectors such that $P + Q = I$ and $PQ = 0$. The variable $z \in \mathbb{R}^d$ represents the resolved (measured) part of the state, while Qx represents the unresolved (orthogonal) dynamics. The range of P is a d–dimensional subspace $\mathcal{R}(P) \subset \mathbb{R}^N$. Throughout, we identify this subspace with $\mathbb{R}^d$ via an orthonormal basis $U \in \mathbb{R}^{N\times d}$ such that $P = UU^\top$. Hence $z := Px \in \mathcal{R}(P)$ is represented by the reduced coordinate vector $U^\top z \in \mathbb{R}^d$, even though P and Q remain $N \times N$ and satisfy $P + Q = I_N$. The sensor sampling period is $\Delta t > 0$, with discrete time index $n \in \mathbb{Z}_+$. The memory length is denoted by n_M. We construct a lag vector $V_n := (z_n, \ldots, z_{n-n_M}) \in \mathbb{R}^{d(n_M+1)}$ to capture history. The noisy observations are given by $\hat{z}_n := z_n + \xi_n$, where the noise terms ξ_n are independent, identically distributed (i.i.d.), zero-mean, and σ^2-sub-Gaussian.

2.3 Exact Mori-Zwanzig Identity

Let $\mathcal{L}$ be the Liouville operator, defined by its action on a scalar observable $g(x)$ as $\mathcal{L}g(x) := \nabla g(x) \cdot F(x)$. We define the associated semigroup $e^{t\mathcal{L}}$, which describes the time-evolution of observables along the flow of the system such that $e^{t\mathcal{L}}g(x(0)) = g(x(t))$.

The dynamics of the resolved variables, $z(t) = Px(t)$, are governed by the exact Mori-Zwanzig identity (Hijón et al., 2010)

$$\frac{d}{dt}z(t) = R(z(t)) \qquad \text{(Markov term)} \tag{1}$$

$$+ \int_0^t K(z(t-s), s)\, ds \qquad \text{(Memory term)} \tag{2}$$

$$+ \eta(t), \qquad \text{(Orthogonal dynamics)}$$

where the constituent terms are defined as

$$R(z) = P\mathcal{L}z, K(z, s) = P\mathcal{L}e^{\,sQ\mathcal{L}}Q\mathcal{L}z, \eta(t) = e^{\,tQ\mathcal{L}}Q\mathcal{L}x(0).$$

Here, $R(z)$ represents the Markovian contribution dependent only on the instantaneous state, while the memory kernel $K(z, s)$ captures the history-dependent effects mediated by the orthogonal subspace.

Remark on Intractability: While (1)–(2) is exact, the memory kernel K and the noise term η are generally computationally intractable. Evaluating $K(z, s)$ requires computing the propagator $e^{sQ\mathcal{L}}$, which involves solving the dynamics of the orthogonal variables Qx. Since Qx lives in the high-dimensional null space of the projection P, computing these terms explicitly requires solving the full high-dimensional system, negating the computational advantage of reduced-order modeling.

2.4 Reduced Mori-Zwanzig

To make the problem tractable, we employ the reduced Mori-Zwanzig formulation from work of Fu et al. (2020).

Assumption 1. (Exponential memory decay). There exist positive constants C_K and γ such that for every $z \in \mathbb{R}^d$ and all $s \geq 0$,

$$\|K(z, s)\| \;\leq\; C_K\, e^{-\gamma s}.$$

With this assumption, we obtain the *reduced* equation by dropping the orthogonal dynamics and truncating the memory integral as

$$\frac{d}{dt}z(t) = R(z(t)) + \int_0^{T_M} K(z(t-s), s)ds + \varepsilon_{\text{rem}}(t). \tag{3}$$

We choose $T_M := \gamma^{-1}\log(C_K/\varepsilon_{\text{mem}})$ so that the truncation error satisfies $\|\varepsilon_{\text{rem}}\|_\infty \leq \varepsilon_{\text{mem}}$.

2.5 Discrete Flow with Memory

By integrating (3) over $[n\Delta t, (n + 1)\Delta t]$ and applying a first-order quadrature (mid-point), we get

$$z_{n+1} = z_n + G^*(V_n) + \varepsilon_{\Delta t,n}, \tag{4}$$

where the error $\|\varepsilon_{\Delta t,n}\| = \mathcal{O}(\Delta t^2)$ is due to the quadrature, and the target flow map defined in Churchill and Xiu (2023) is given by

$$G^*(V_n) := \Delta t R(z_n) + \Delta t \sum_{\ell=0}^{n_M} \omega_\ell \tilde{K}(z_{n-\ell}),$$

$$\tilde{K}(z_{n-\ell}) := K(z_{n-\ell}, \ell\Delta t),$$

$$\omega_\ell := (1 - \rho)\rho^\ell, \qquad \text{and} \qquad \rho := e^{-\gamma\Delta t}.$$

3. OCCUPATION-KERNEL TWIN

Let $\mathcal{H}$ be a vector-valued RKHS with an *I-separable* kernel $K(x, x') = k(x, x')\, I_d$, where I_d denotes the $d \times d$ identity matrix. For a C^1 trajectory $x : [a, b] \to \mathbb{R}^d$ (where $x(t) \in \mathbb{R}^d$ denotes the resolved state; discretisation converts the history of x into lag vectors $V_n \in \mathbb{R}^{d(n_M+1)}$), and for every $F \in \mathcal{H}$, we define the linear functional $L_x F := \int_a^b F(x(t))\, dt$. By the Riesz representation theorem, there exists a unique element $L^*_{x,v} \in \mathcal{H}$ such that

$$v^\top L_x F = \langle F, L^*_{x,v}\rangle_{\mathcal{H}} \quad \forall v \in \mathbb{R}^d. \tag{5}$$

This construction follows Rosenfeld et al. (2019, 2022).

Because the reduced MZ equation (3) contains the memory envelope $e^{-\gamma s}$, we multiply the integrand in (5) by the same factor and then apply midpoint quadrature. This procedure yields the discrete scalar kernel

$$k_{\text{occ}}(V_n, V'_n) := \sum_{\ell=0}^{n_M} w_\ell\, G_\sigma(z_{n-\ell}, z'_{n-\ell}), \quad w_\ell = (1 - \rho)\rho^\ell, \tag{6}$$

where $k = G_\sigma$ is the Gaussian kernel $G_\sigma(u, v) = \exp[-\|u - v\|^2/(2\sigma^2)]$ and $V_n = (z_n, \ldots, z_{n-n_M})$ is the lag vector.

Because k is positive-definite (p.d.), $L^*_{x,v} \in \mathcal{H}$. Hence, the map $V \mapsto k_{\text{occ}}(V, \cdot)$ acts as an explicit feature map with values in the same vector-valued RKHS.

3.1 RKHS Regression Problem

We observe noisy samples $\hat{z}_n = z_n + \xi_n$; only $\hat{z}_n$ are available for learning. For notational convenience, we learn the increment $Y_n := \hat{z}_{n+1} - \hat{z}_n$ instead of the state itself; this is equivalent to learning the flow map G^* in (4).

Define the training set $S_N = \{(V_n, Y_n)\}_{n=1}^N$. The occupation-RKHS $\mathcal{H}$ has norm $\|f\|^2_{\mathcal{H}} = \alpha^\top K \alpha$ with Gram matrix $K_{ij} = k_{\text{occ}}(V_i, V_j)$. Note that every $f \in \mathcal{H}$ maps $\mathbb{R}^d$ to $\mathbb{R}^d$; after discretisation we evaluate f on lag vectors through $f(V_n)$. We estimate $\hat{G}$ by solving the regularized least-squares problem

$$\hat{G} = \arg\min_{f \in \mathcal{H}} \frac{1}{N} \sum_{n=1}^{N} \|Y_n - f(V_n)\|^2 + \lambda \|f\|^2_{\mathcal{H}} \qquad (7)$$

with regularization parameter $\lambda > 0$. The closed-form solution is $\alpha = (K + \lambda N I)^{-1} Y$, with $Y = [Y_{n_M}, \ldots, Y_{N-1}]^\top \in \mathbb{R}^{(N-n_M)\times d}$.

3.2 Noise-Robust Error Bound

Proposition 1. (Variance Shrinkage). Let ξ_n be σ^2_{noise}-sub-Gaussian. Then for any V,

$$\text{Var}[f(V)] \le \frac{(\sum w_\ell^2)\sigma^2_{\text{noise}}}{2\sigma^2} \qquad (8)$$

where $f(V) = k_{\text{occ}}(V, \cdot)$. For geometric weights $w_\ell = (1-\rho)\rho^\ell$,

$$\sum w_\ell^2 = \frac{(1-\rho)^2}{1-\rho^2} = \frac{1-\rho}{1+\rho},$$

which is asymptotically $\frac{1}{2}(1-\rho)$ as $\rho \to 1$. Thus, the effective noise variance affecting the kernel is scaled by a factor of $O(1-\rho)$, whereas a standard vector Gaussian kernel (single exponential) is exposed to the full noise variance σ^2_{noise}.

Theorem 1. (RKHS Generalisation with Correlated Data). Assume the stochastic process $\{(V_n, Y_n)\}_{n\ge 0}$ is strictly stationary and β-mixing (Bradley, 2005) with coefficient $\beta(m) \le C_\beta \rho_\beta^m$ for some $0 < \rho_\beta < 1$. Let $\lambda = N^{-1/2}$. Then, for any $0 < \delta < 1$,

$$\Pr\Big(\|\hat{G} - G^*\|_{L^2(\mu)} \le C \, \frac{\kappa}{\sqrt{N}} \sqrt{\log(1/\delta)} \Big) \;\ge\; 1 - \delta,$$

where $\kappa^2 := \sup_V k_{\text{occ}}(V, V)$.

Proof 1. (Sketch). Let $\mathcal{F} = \{V \mapsto \langle \alpha, k_{\text{occ}}(V, \cdot)\rangle_{\mathcal{H}} : \|\alpha\|_{\mathcal{H}} \le 1\}$. By Mohri et al. (2018), the empirical Rademacher complexity under geometric β-mixing satisfies $\mathfrak{R}_N(\mathcal{F}) \le \kappa/\sqrt{N}$. Standard kernel-ridge excess-risk analysis (Steinwart and Christmann, 2008) yields $R(\hat{G}) - R(G^*) \le 2\mathfrak{R}_N(\mathcal{F}) + \lambda \|G^*\|^2_{\mathcal{H}}$. Setting $\lambda = N^{-1/2}$ and applying Bernstein's inequality for β-mixing sequences (Merlevède et al., 2009) proves the claim.

Assumption 2. (Dissipative Flow Map). The true flow map G^* in (4) is globally Lipschitz with constant $L_G < 1$; this is satisfied, for example, by any dissipative system after linear rescaling of the state.

Lemma 1. (Single-Step Error Propagation). Writing the one-step state error as $e_{n+1} = \|z_{n+1} - \hat{z}_{n+1}\|$ and adding–subtracting G^* gives $e_{n+1} \le (1 + L_G) e_n + \hat{\varepsilon}_n$ with $\hat{\varepsilon}_n = \|\hat{G}(\hat{V}_n) - G^*(\hat{V}_n)\|$.

Theorem 2. (Multi-Step Bound). For any horizon H and for the same $0 < \delta < 1$ as in Theorem 1, with probability at least $1 - \delta$,

$$\max_{0 \le h \le H} e_{n+h} \;\le\; \frac{1}{1 - L_G} \max_{0 \le h \le H-1} \hat{\varepsilon}_{n+h}.$$

Proof 2. (Sketch). Iterate Lemma 1 and sum the geometric series; Grönwall's discrete lemma (Grönwall, 1919) yields the stated inequality. Substituting the high-probability estimate $\hat{\varepsilon}_n = O_p(N^{-1/2})$ from Theorem 1 produces the closed, N-explicit long-horizon guarantee.

Algorithm 1 Occupation–Kernel Digital Twin (batch version)-OKDT

Input: time series $\{\hat{z}_n\}_{n=0}^N$; memory n_M; bandwidth σ; decay γ ($\rho := e^{-\gamma \Delta t}$); regularisation λ.
Output: trained flow map $\hat{G}(\cdot)$ and predictor $\hat{z}_{n+1} = \hat{z}_n + \hat{G}(\hat{V}_n)$

Build lag vectors $V_n = (\hat{z}_n, \ldots, \hat{z}_{n-n_M})$ for $n = n_M, \ldots, N-1$.
Build training targets $Y_n := \hat{z}_{n+1} - \hat{z}_n$.
Compute geometric weights $w_\ell = (1-\rho)\rho^\ell$.
function KERNEL(V_i, V_j)
    return $k_{\text{occ}}(V_i, V_j) = \sum_{\ell=0}^{n_M} w_\ell \exp\big[-\|\hat{z}_{i-\ell} - \hat{z}_{j-\ell}\|^2/(2\sigma^2)\big]$
end function
Assemble Gram matrix $K_{ij} = $ KERNEL(V_i, V_j).
Solve for coefficients $\alpha = (K + \lambda N I)^{-1} Y$.
function $\hat{G}(V)$ ▷ predictor handle
    return $\sum_{n=n_M}^{N-1} \alpha_n$ KERNEL(V, V_n)
end function

4. CASE STUDY: CYLINDER WAKE FORECASTING

We further demonstrate the capability of the OKDT surrogate on a canonical fluid dynamics problem: the laminar flow past a cylinder, which produces the characteristic von Kármán vortex street. The data are from the dataset provided by Brunton and Kutz (2022). This dataset comprises 151 snapshots of the 2D vorticity field, $\mathbf{V} \in \mathbb{R}^{449\times 199}$, as it evolves.

Experimental Setup We set the model memory $n_M = 25$, regularization $\lambda = 10^{-7}$, and kernel bandwidth $\sigma = 5.0$. For each sensor, the model is trained on the first $T_{\text{train}} = 60$ time steps. It is then seeded with the last $n_M + 1$ points of this training data and tasked with predicting the remaining $T_{\text{test}} \approx 90$ steps.

We test this framework at four physically distinct locations in the flow field, as depicted in **Fig. 1** (right column):

- **Sensor 1 (Wake Center):** $(x, y) = (350, 100)$, placed directly in the primary von Kármán vortex street.
- **Sensor 2 (Wake Edge):** $(x, y) = (350, 130)$, at the shear layer boundary.
- **Sensor 3 (Near Cylinder):** $(x, y) = (250, 100)$, in the calm region immediately behind the cylinder.
- **Sensor 4 (Freestream):** $(x, y) = (350, 180)$, far from the wake dynamics.

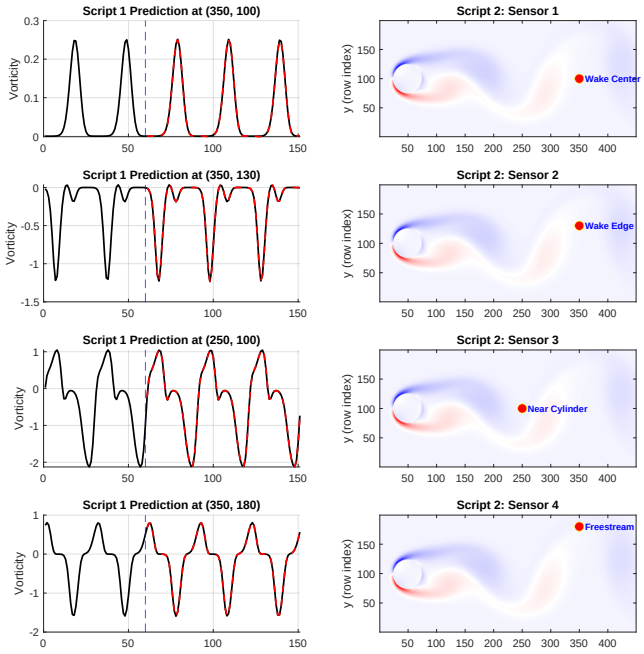

Fig. 1. **Autoregressive forecasting of cylinder flow vorticity using the OKDT model at four distinct sensor locations. (Right Column)** A snapshot of the vorticity field at $t = 80$ shows the physical locations of the four sensors , which are placed in the 'Wake Center' , 'Wake Edge' , 'Near Cylinder' , and 'Freestream'. **(Left Column)** Time-series prediction results for each sensor. The model is trained on data up to the vertical blue line ($t = 60$). The ground truth vorticity is the black solid line, and the OKDT's autoregressive prediction is the red dashed line. The model successfully captures the complex, periodic dynamics in the wake center, the asymmetric oscillations at the wake edge and the quiescent behavior in the other regions.

Results and Discussion The results, presented in **Fig. 1** (left column), demonstrate remarkable accuracy and stability.

In the highly dynamic "Wake Center" , the OKDT (proposed) model perfectly captures both the frequency and amplitude of the periodic vortex shedding after the training cutoff. At the "Wake Edge" , the model reproduces the more complex, asymmetric oscillatory behavior, correctly learning the non-zero-mean dynamics. In both the "Near Cylinder" and "Freestream" locations, where the flow is quiescent, the model correctly identifies and predicts the low-magnitude, non-vortical behavior.

This experiment highlights that the OKDT surrogate, despite being trained on a relatively short time-series ($T = 60$), can learn the underlying dynamics of a complex fluid system from a single point and provide stable, accurate, long-horizon forecasts.

5. CASE STUDY: NONLINEAR ATTRACTOR (VAN DER POL)

We also tested the OKDT's ability to learn a nonlinear attractor. We select the Van der Pol oscillator, a classic benchmark for self-oscillating, nonlinear dynamics.

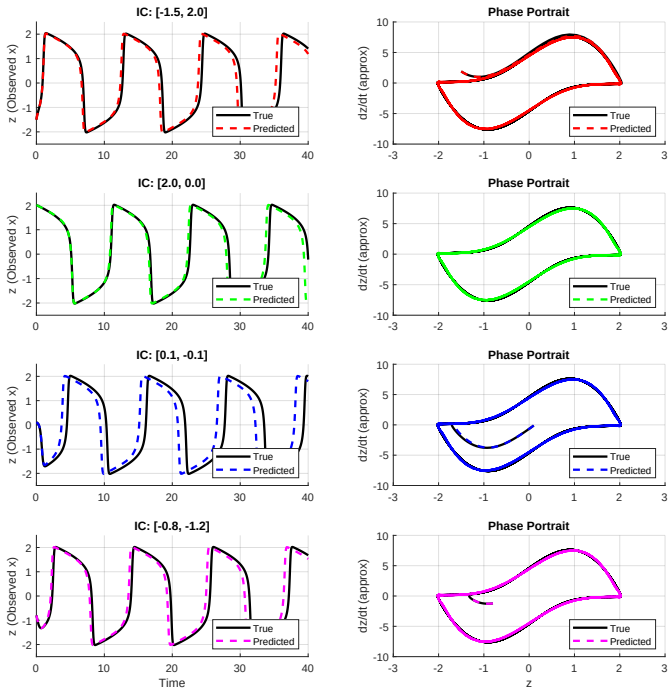

Fig. 2. **Autoregressive prediction of the nonlinear Van der Pol oscillator from four distinct, unseen initial conditions.** The vector-based KRR-M model was trained on a *single* trajectory (from IC $[0.5, 0.5]$). These plots show its generalization to new ICs. **(Left Column)** Time series comparison. The model prediction (colored, dashed) rapidly converges to the ground truth (black, solid). **(Right Column)** Phase portrait comparison. This view confirms that regardless of the starting point (e.g., near the origin or far from the cycle), the model's predicted trajectory correctly converges to the true limit cycle, demonstrating it has learned the attractor's vector field.

Experimental Setup The Van der Pol oscillator is governed by:

$$\ddot{z} - \mu(1 - z^2)\dot{z} + z = 0$$

We use a strong nonlinearity parameter, $\mu = 5.0$, which results in a spiky relaxation oscillator with a distinct limit cycle. Data is generated with $\Delta t = 0.02$. Key hyperparameters are: memory $n_M = 40$, regularization $\lambda = 10^{-6}$, and kernel bandwidth $\sigma = 5.0$.

The central challenge is to learn the system's attractor from limited data. The model is trained on a single trajectory ($T = 80$s) originating from $z_0 = [0.5, 0.5]$. Only the first state coordinate z_1 is measured and used for training; the second coordinate z_2 is treated as unobserved. We therefore take $x = [z_1 \ z_2]^\top$ and the projection $P = [1 \ 0]$, so that the resolved variable is $z = Px = z_1$. The model never sees data from any other part of the state space during training. We test the model's ability to generalize and reproduce the attractor from four distinct, unseen initial conditions, including points inside the limit cycle ($[0.1, -0.1]$) and points on or outside it. For each test, the model is seeded with the first $n_M + 1 = 41$ true points and then runs purely autoregressively.

Results and Discussion The results, presented in **Fig. 2**, show that the OKDT model successfully learns the nonlinear attractor.

- **Time Series (Left Column):** For all four initial conditions, the predicted time series (colored, dashed) rapidly locks onto the true system dynamics (black, solid). Even when starting from the unstable fixed point at the origin (IC: $[0.1, -0.1]$), the model correctly predicts the trajectory as it "spirals out" to the stable limit cycle.
- **Phase Portraits (Right Column):** The phase-space view provides the clearest confirmation. The ground truth (black) shows the precise limit cycle. The predicted trajectories (colored, dashed) all converge to this exact same attractor, regardless of their disparate starting points.

This experiment demonstrates that the memory-based kernel surrogate does not merely "memorize" the training trajectory. Instead, it learns a high-fidelity approximation of the underlying **vector field** on the attractor, enabling it to generalize and make stable, long-horizon predictions for any trajectory that converges to that attractor.

6. BEYOND DNNS

To validate the theoretical guarantees of noise robustness and long-horizon stability, we design a numerical experiment to compare the proposed occupation-kernel digital twin (OKDT) against two standard benchmarks: a standard vector-based kernel ridge regression (KRR-vec) and a deep neural network (DNNs).

6.1 Experimental Setup

System Dynamics. We demonstrate our method on a 2D linear system whose dynamics are governed by $\dot{z} = Az$:

$$\frac{d}{dt}z(t) = \begin{pmatrix} 1 & -4 \\ 4 & -\alpha \end{pmatrix} z(t), \tag{9}$$

with the parameter set to $\alpha = 1.1$. The central challenge of this experiment is to learn a predictive model for the first component, $z_1(t)$, using only its own past measurements. The second component, $z_2(t)$, is treated as an unobserved or "hidden" variable whose influence must be captured implicitly through the model's memory.

Training Data Generation. We generate a single ground-truth trajectory by numerically integrating (9) from the initial condition $z_0 = [-1.5, 1.5]^\top$ over a time horizon of $T = 40$s, with a sampling period of $\Delta t = 0.02$. The resulting clean trajectory $\{z_n\}$ is then corrupted with i.i.d. Gaussian noise to produce the observation set $\{\hat{z}_n\}$:

$$\hat{z}_n = z_n + \xi_n, \quad \text{where} \quad \xi_n \sim \mathcal{N}(0, \sigma^2_{\text{noise}} I),$$

with a noise level of $\sigma_{\text{noise}} = 0.1$.

Model Training and Evaluation. From the 2D noisy trajectory $\{\hat{z}_n\}$, we isolate the first component to create a 1D time series $\{\hat{z}_{1,n}\}$. This 1D series is used to train three competing models: our proposed OKDT, a standard vector-kernel KRR (KRR-vec), and a feedforward Deep Neural Network (DNN). All models are trained to predict the next increment, $\hat{z}_{1,n+1} - \hat{z}_{1,n}$, based on a lag vector of previous observations, $V_n = (\hat{z}_{1,n}, \ldots, \hat{z}_{1,n-n_M})$.

Key hyperparameters are set as follows: memory length $n_M = 30$, regularization $\lambda = 10^{-6}$, and kernel bandwidth $\sigma = 6.0$. The DNN is a Multi-Layer Perceptron (MLP) with two hidden layers of 20 neurons each, using ReLU activation functions, and is trained for 300 epochs used in work Churchill and Xiu (2023) for comparison.

For evaluation, the trained models are tested on four new initial conditions. Each model is seeded with the first $n_M + 1$ noisy observations of z_1 from a new trajectory and must then predict the subsequent evolution autoregressively for $T = 20$s.

6.2 Results and Discussion

The results in **Fig. 3** demonstrate the clear superiority of the proposed OKDT. Across all four initial conditions, it accurately predicts the damped oscillations of the true system from noisy, partial observations. In contrast, the baseline models fail decisively: the standard KRR-vec is unstable and diverges immediately, while the DNN correctly learns the frequency but fails to model the amplitude decay, resulting in a qualitatively incorrect forecast of an undamped oscillator. This highlights the critical role of inductive bias. The OKDT model succeeds because its exponentially weighted memory, derived from Mori-Zwanzig theory, is perfectly suited to capturing the influence of hidden dynamics from limited, noisy data. The generic architectures of the KRR-vec and DNN lack this necessary structure, preventing them from learning the true dynamics in this challenging scenario.

7. CONCLUSION AND FUTURE WORK

This paper proposes an occupation-kernel digital twin that fuses the Mori–Zwanzig projection with kernel ridge regression. The resulting surrogate inherits the memory structure of the reduced dynamics, suppresses sensor noise by a factor proportional to $1-\rho$, and admits a finite-sample generalisation bound of order $\mathcal{O}(N^{-1/2})$ that propagates to any prediction horizon. Experiments on cylinder-wake flow, the Van der Pol oscillator, and a noisy linear system show that the model delivers long-horizon accuracy and stability where vector-kernel methods and deep neural networks diverge, providing a dependable modelling layer for fault-resilient and health-aware control.

Future work will extend the method to streaming data and closed-loop operation. An online Nyström implementation will allow the digital twin to update in real time, while structure-preserving kernels and learned decay rates could embed physical invariants and adapt the memory horizon automatically. Integrating the multi-step error envelope into chance-constrained MPC or control-barrier-function filters promises end-to-end safety guarantees. Finally, large-scale studies on aircraft prognostics and battery ageing will evaluate the approach under industrial conditions.

REFERENCES

Alcaraz, S. et al. (2023). Uncertainty-aware deep learning for digital twin-driven monitoring: Application to fault detection in power lines. *arXiv preprint*. URL `https://arxiv.org/abs/2303.10954`.

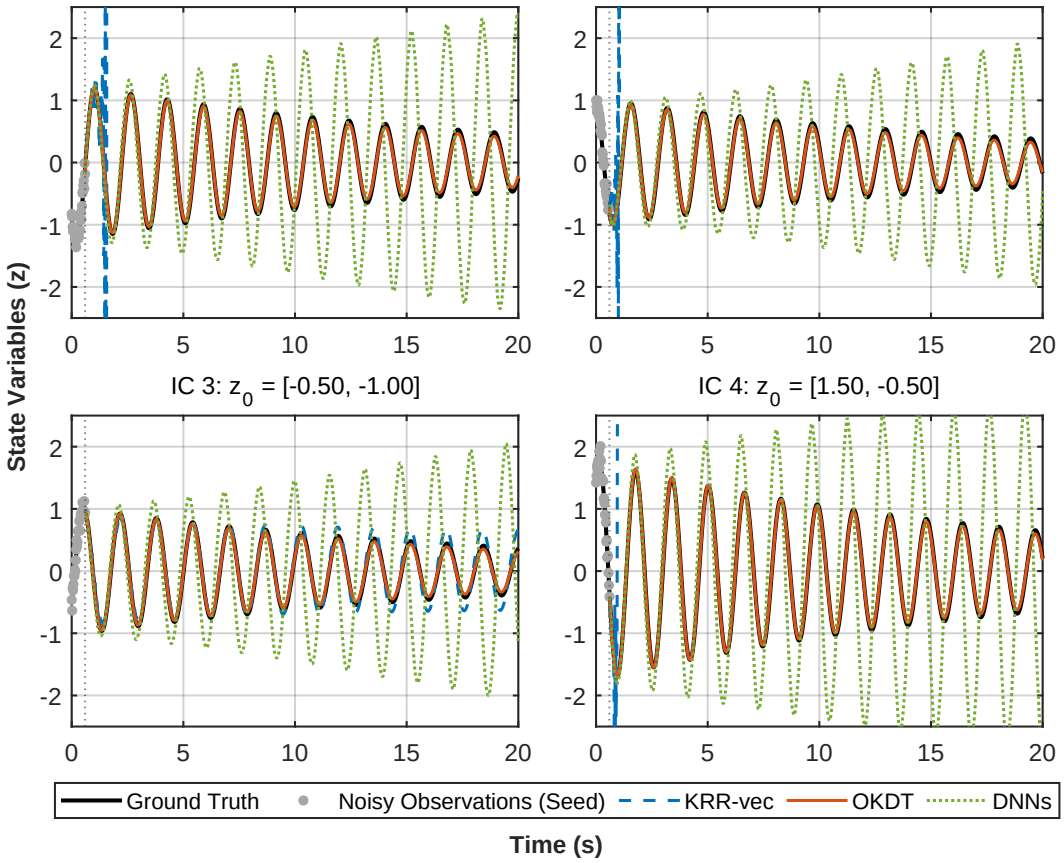

Fig. 3. **Autoregressive roll-out comparison from four distinct initial conditions (ICs).** Models are seeded with $n_M + 1 = 31$ noisy observations (gray circles, shown up to the vertical dashed line). They then predict autoregressively for $T = 20$s. The proposed **OKDT (orange, solid)** remains stable and accurately tracks the ground truth (black, solid). In contrast, both the **KRR-vec (blue, dashed)** and **DNNs (green, dotted)** benchmarks, trained on the same noisy data, diverge catastrophically.

Bradley, R.C. (2005). *Introduction to Strong Mixing Conditions*. Metric Press, Utah.

Chorin, A.J., Hald, O.H., and Kupferman, R. (2017). Model reduction and memory effects via the mori–zwanzig formalism. *arXiv preprint*. URL `https://arxiv.org/abs/1708.02235`.

Churchill, V. and Xiu, D. (2023). Flow map learning for unknown dynamical systems: Overview, implementation, and benchmarks. *Journal of Machine Learning for Modeling and Computing*, 4(2).

Fu, X., Chang, L.B., and Xiu, D. (2020). Learning reduced systems via deep neural networks with memory. *Journal of Machine Learning for Modeling and Computing*, 1(2).

Gomez, L. et al. (2025). Innovative framework for fault detection and system resilience in hydropower operations using digital twins and deep learning. *Scientific Reports*, 15, 1–14. doi:10.1038/s41598-025-98235-1.

Gopalakrishnan, R. et al. (2024). Convergence and stability analysis of recurrent neural networks for rapid structural damage assessment under seismic loads. *PLOS One*, 19(10), e0336101. doi: 10.1371/journal.pone.0336101.

Grönwall, T.H. (1919). Note on differential and integral inequalities. *Annals of Mathematics*, 20(2), 292–296.

Hijón, C., Español, P., Vanden-Eijnden, E., and Delgado-Buscalioni, R. (2010). Mori–zwanzig formalism as a practical computational tool. *Faraday discussions*, 144, 301–322.

Merlevède, F., Peligrad, M., and Rio, E. (2009). Bernstein inequality and moderate deviations under strong mixing conditions. *IMS Lecture Notes Monograph Series*, 56, 273–292.

Mohri, M., Rostamizadeh, A., and Talwalkar, A. (2018). *Foundations of Machine Learning*. MIT Press.

Rasheed, A., San, O., and Kvamsdal, T. (2020). Digital twin: Values, challenges and enablers from a modeling perspective. *IEEE Access*, 8, 21980–22012. doi: 10.1109/ACCESS.2020.2966491.

Rosenfeld, J.A., Kamalapurkar, R., Russo, B., and Johnson, T.T. (2019). Occupation kernels and densely defined liouville operators for system identification. In *Proceeding of IEEE Conference on Decision and Control*, 6455–6460. doi:10.1109/CDC40024.2019.9029337.

Rosenfeld, J.A., Kamalapurkar, R., Gruss, L.F., and Rosenfeld, J. (2022). Dynamic mode decomposition for continuous time systems with the liouville operator. *Journal of Nonlinear Science*, 32(5), 1–39. doi: 10.1007/s00332-021-09746-w.

Steinwart, I. and Christmann, A. (2008). *Support Vector Machines*. Springer.

Wang, B. et al. (2025). Fedonet: Fourier-embedded deeponet for spectrally accurate operator learning. *arXiv preprint*. URL `https://arxiv.org/abs/2509.12344`.