Geometric Closure of Linear Controlled Models via Tangent–Normal Decomposition

Girolamo Oddo

May 2026

Abstract

Linear models remain central in engineering for their simplicity and interpretability, but their local validity often requires data-driven closure to capture nonlinear effects. This paper proposes a geometric closure framework for nominal linear controlled models, where the mismatch with the true dynamics is decomposed into tangential and normal components with respect to a local chart. The correction is learned from data through a bounded dictionary and separate ridge regressions with asymmetric regularization. Unlike generic residual learning, the method preserves the linear backbone while providing an interpretable geometric description of model-form error. A chart-selection strategy is also included, allowing canonical or data-driven coordinates. The approach is validated on a synthetic benchmark and a vehicle-dynamics case, where it improves prediction accuracy and shows that the most suitable geometric representation is problem-dependent.

1 Introduction

Linear state-space models represent a cornerstone of modern engineering, providing a robust foundation for system identification, estimation, and control. Their prevalence is justified by their inherent interpretability and the availability of mature analytical tools for stability and performance verification. However, these models typically function as local approximations; when a system operates far from its nominal linearization point, the discrepancy between linear theory and nonlinear reality often results in a significant loss of fidelity and the accumulation of prediction errors.

To address these limitations, a common strategy involves augmenting nominal models with data-driven residual terms. While such hybrid approaches can substantially reduce prediction error, they frequently treat the model mismatch as an unstructured function approximation problem. In the absence of explicit constraints, these corrections act as "black boxes" that lack physical interpretability and may produce trajectories that violate the intrinsic geometric organization of the system's state space.

In many physical systems, the state does not evolve uniformly across the ambient space. Instead, trajectories are often constrained to low-dimensional manifolds induced by conservation laws, actuator limits, or nonlinear couplings. This observation suggests that model mismatch is not a stochastic disturbance, but a structured geometric deformation. If the nominal linear model is viewed as a first-order approximation of the manifold's tangent bundle, then the prediction error can be formally decomposed into components that are either tangential or normal to this surface. Motivated by this geometric perspective, this work introduces a framework for the geometric closure of linear controlled models via tangent–normal decomposition. The proposed method explicitly partitions the residual correction into tangential components, which govern the flow along the manifold, and normal components, which account for deviations away from it. By integrating data-driven basis selection with asymmetric regularization, the framework yields a structured and interpretable correction that preserves the geometric integrity of the data while maintaining the computational efficiency of linear regression techniques.

2 Literature review

Model closure is the problem of representing the effect of unresolved or neglected dynamics in a reduced description. Across the literature, the available strategies can be grouped into four broad families: analytical closures, hybrid closures, purely neural closures, and data-driven sparse discovery methods. These families differ primarily in how much structure from the underlying physics is retained, how much is learned from data, and whether the correction is local, nonlocal, Markovian, or memory-dependent.

Analytical closures and projection-based reduction. The most classical route is analytical reduction, where the full dynamics are projected onto a resolved subspace and the effect of the unresolved variables is represented explicitly as a memory and noise term. In the Mori–Zwanzig framework, the reduced model is exact in principle, but the resulting closure is typically non-Markovian and difficult to evaluate or approximate in practice [2, 3]. This line of work has been especially influential in turbulence and coarse-grained simulation, where the unresolved contribution is often modeled through finite-memory kernels, renormalization, or dynamic subgrid-scale approximations [5, 6, 7]. The main strength of analytical closure is interpretability and consistency with the underlying projection, but its main limitation is that the resulting operators are often problem-specific and algebraically demanding. Recent reviews emphasize that analytical closure remains important, yet it is frequently too rigid when the reduced coordinates or the closure manifold are not known a priori [8].

Hybrid closures: first-principles models augmented by data. A widely used compromise is hybrid modeling, in which a mechanistic core is kept fixed while a learned correction accounts for model-form error. This paradigm dates back to early process-modeling work combining first-principles equations with neural networks [1] and has since become a standard template in process systems engineering and scientific machine learning [9, 31, 34]. In fluid mechanics and related fields, hybrid closures often learn residual source terms, missing constitutive laws, or correction fields on top of a baseline solver [4, 32, 29, 30]. A particularly important branch is physics-informed learning, where the loss function or architecture enforces known governing laws while the unknown parts are fitted from data [11, 10]. Universal differential equations further generalize this idea by embedding trainable components directly inside differential-equation models [12]. Hybrid methods are attractive because they retain structure and data efficiency, but they can still become opaque when the correction is learned as an unconstrained black box.

Neural closures and operator learning. Another investigated approach replaces the correction model with a neural parameterization. Neural ordinary differential equations represent the dynamics directly through a learned vector field, enabling continuous-depth modeling and adaptive solvers [13]. Latent ODEs extend this idea to irregularly sampled sequences through a latent dynamical system [14]. Hamiltonian neural networks and related structure-preserving models encode conservation structure explicitly, improving long-term rollout behavior in conservative systems [15]. In closure settings, neural networks have been used to model truncated modal effects in reduced-order models, including neural closures for nonlinear model order reduction and more general neural closure models with Markovian and non-Markovian components [22, 23, 24, 25]. Operator-learning methods such as DeepONet push this idea further by approximating maps between function spaces rather than finite-dimensional states, which is appealing for surrogate modeling and PDE closure [16]. These methods are flexible and often accurate, but their learned corrections may be difficult to interpret physically unless additional structure is imposed.

Data-driven sparse discovery and coordinate learning. The last family seeks not only to close the model but to discover the governing equations themselves. SINDy formulates dynamical-system identification as sparse regression over a library of candidate terms, giving compact and interpretable equations when the true dynamics are sparse in a suitable basis [17]. The controlled extension SINDYc includes external inputs and feedback terms [18], while PDE-FIND generalizes sparse regression to spatio-temporal systems [19]. Later developments added latent coordinate discovery through autoencoders, showing that an appropriate coordinate system can be learned jointly with the sparse dynamics [20]. Data-driven closure models extend sparse identification to reduced-order settings with memory effects [21], and more recent variants improve robustness to noise, discontinuities, partial observability, and physical priors [26, 27, 35]. Recent turbulence work also shows that sparse or data-enabled closure can be used to discover both specific and generalizable closures from simulation data [36]. The main appeal of sparse discovery is interpretability, but its performance depends strongly on the chosen library, the coordinate system, and the availability of sufficiently rich data.

Relation to the present work. The present paper belongs to the hybrid-residual family, but it is more specific than generic residual learning because the baseline is an explicit linear controlled model, assumed to provide a local first-order approximation of the nonlinear dynamics. The proposed contribution is not to replace this baseline, but to close its mismatch through a structured decomposition of the residual into tangential and normal components with respect to a local chart. This makes the correction interpretable as a geometry-aware residual model that preserves the linear backbone while distinguishing in-manifold calibration from off-manifold deviation. In this sense, the method combines ideas from hybrid modeling and reduced-order closure, with the additional prior that the nominal linear model already captures the dominant local tangent behavior of the system.

3 Problem formulation

Consider a discrete-time controlled system

$$x_{k+1} = f(x_k, u_k), \qquad x_k \in \mathbb{R}^n,\ u_k \in \mathbb{R}^m, \tag{1}$$

and a nominal linear controlled model

$$x^{\mathrm{lin}}_{k+1} = Ax_k + Bu_k, \tag{2}$$

with $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$. The objective is to approximate the mismatch between the nominal model and the true dynamics through a geometrically constrained closure term.

It is assumed that the observed trajectories are locally organized around a low-dimensional manifold $\mathcal{M} \subset \mathbb{R}^n$. The nominal linear model is treated as a local reference, while the mismatch is decomposed into tangential and normal components with respect to a selected subspace.

4 Proposed solution

Let $V \in \mathbb{R}^{n\times d}$ and $W \in \mathbb{R}^{n\times(n-d)}$ be orthonormal bases of the selected tangent and normal subspaces, respectively, with $V^TV = I_d$, $W^TW = I_{n-d}$, and $W^TV = 0$. After state and input normalization, the corrected model is written as a *linear prediction plus correction*, as clarified in Figure 1.

$$z_{k+1} = z^{\mathrm{lin}}_{k+1} + Vg_t(s_k, \bar{u}_k) + Wg_n(s_k, \bar{u}_k), \tag{3}$$

where

$$z_k = \mathcal{N}_x(x_k), \qquad \bar{u}_k = \mathcal{N}_u(u_k), \qquad z^{\mathrm{lin}}_{k+1} = \mathcal{N}_x(Ax_k + Bu_k), \tag{4}$$

and

$$s_k = V^T z_{k+1}^{\text{lin}}. \tag{5}$$

Here $\mathcal{N}_x$ and $\mathcal{N}_u$ denote affine standardization maps for states and inputs. The closure maps g_t and g_n represent the tangential and normal corrections acting on the nominal linear prediction.

The closure maps are approximated linearly in the parameters through a bounded dictionary $\phi(\cdot)$,

$$g_t(s,u) = C_t\phi(s,u), \qquad g_n(s,u) = C_n\phi(s,u), \tag{6}$$

with $C_t \in \mathbb{R}^{d\times p}$ and $C_n \in \mathbb{R}^{(n-d)\times p}$. In the implementation, the dictionary is built from bounded nonlinear functions to avoid polynomial blow-up and improve numerical robustness. A representative choice is

$$\phi(s,u) = \big[1,\ \psi_s(s)^T,\ \psi_u(u)^T,\ \mathrm{vec}\big(\psi_s(s)\psi_u(u)^T\big)\,,\ \mathrm{vec}\big(\psi_s(s)\psi_s(s)^T\big)\,,\ \mathrm{vec}\big(\psi_u(u)\psi_u(u)^T\big)\quad\big]^T, \tag{7}$$

where, for example

$$\psi_s(s) = \begin{bmatrix}\tanh(s)\\ \sin(s)\\ \cos(s)\end{bmatrix}, \qquad \psi_u(u) = \begin{bmatrix}\tanh(u)\\ \sin(u)\\ \cos(u)\end{bmatrix}. \tag{8}$$

The coefficients are identified from data by ridge regression, with stronger regularization on the tangential term than on the normal one.

Remark 1. *The matrices V and W are chosen as an orthogonal decomposition of the selected coordinate chart. They coincide with true tangent and normal spaces only when the local manifold hypothesis is valid and the chosen chart is well aligned with the data.*

Assumption 1 (Local graph structure)**.** *The true dynamics are locally representable as a graph over the selected tangent subspace, up to a smooth deformation in the normal bundle.*

Assumption 2 (Regularity)**.** *The closure maps are sufficiently smooth and bounded in the operating region so that a finite dictionary approximation is meaningful.*

Assumption 3 (Identifiability)**.** *The available data span the relevant local region of the state space with sufficient richness to estimate the tangent and normal closures.*

5 Algorithm

Given a dataset $\mathcal{D} = \{(x_k, u_k, x_{k+1})\}_{k=1}^N$, the algorithm is:

(1) normalize states and inputs;

(2) select V, W either canonically or via PCA;

(3) choose d manually or by validation;

(4) build the feature matrix $\Phi = [\phi(s_k, \bar{u}_k)]_{k=1}^N$;

(5) compute tangential and normal residuals;

(6) solve two ridge regressions for C_t and C_n;

(7) deploy the corrected model for simulation or prediction.

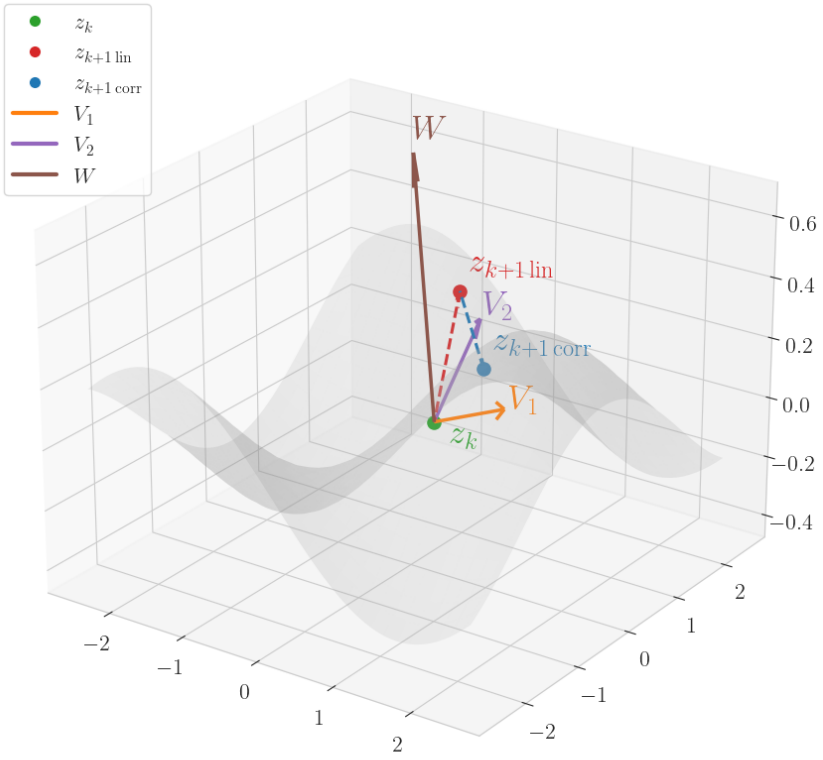

Figure 1: Geometric representation of closure logic.

More precisely, define the linear prediction

$$z_{k+1}^{\mathrm{lin}} = \mathcal{N}_x(Ax_k + Bu_k) \tag{9}$$

and the normalized residual

$$r_k = z_{k+1} - z_{k+1}^{\mathrm{lin}}. \tag{10}$$

Then, for the tangent and normal residuals

$$r_k^{(t)} = V^T r_k, \qquad r_k^{(n)} = W^T r_k, \tag{11}$$

and the ridge problems are

$$\min_{C_t} \|Y_t - C_t\Phi\|_F^2 + \lambda_t \|C_t\|_F^2, \qquad \min_{C_n} \|Y_n - C_n\Phi\|_F^2 + \lambda_n \|C_n\|_F^2, \tag{12}$$

with $\lambda_t > \lambda_n$.

6 Framework validation

The proposed framework is validated through four stress tests: in-distribution prediction, out-of-distribution initial conditions, reduced-data training, and noisy measurements. A fifth test deliberately violates the tangential–normal graph hypothesis by injecting a tangential correction in the true system.

The evaluation metric is the trajectory RMSE, averaged over 50 Monte Carlo runs. The baseline linear model is reported alongside the corrected models in Tables 1 and 2.

6.1 Synthetic case

The synthetic benchmark is built from the nominal linear model

$$x_{k+1}^{\text{lin}} = Ax_k + Bu_k, \tag{13}$$

with

$$A = \begin{bmatrix} 0.92 & 0.08 & 0.00 & 0.00 \\ -0.12 & 0.90 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.85 & 0.15 \\ 0.00 & 0.00 & -0.10 & 0.88 \end{bmatrix}, \qquad B = \begin{bmatrix} 0.05 \\ 0.02 \\ 0.00 \\ 0.03 \end{bmatrix}. \tag{14}$$

The true synthetic data-generating system is obtained by hiding the closure in a rotated coordinate frame. Let

$$Q = \begin{bmatrix} R(0.65) & 0 \\ 0 & R(-0.45) \end{bmatrix}, \qquad R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{15}$$

For each step, the nominal linear prediction is first computed, then transformed as

$$z_{k+1}^{\text{lin}} = Q^T x_{k+1}^{\text{lin}}. \tag{16}$$

The true update is

$$z_{k+1} = z_{k+1}^{\text{lin}} + \begin{bmatrix} 0 \\ 0 \\ c_1(s_k, u_k) \\ c_2(s_k, u_k) \end{bmatrix}, \qquad x_{k+1} = Qz_{k+1}, \tag{17}$$

where

$$s_k = \begin{bmatrix} z_{k+1,1}^{\text{lin}} \\ z_{k+1,2}^{\text{lin}} \end{bmatrix}, \tag{18}$$

and

$$c_1(s, u) = 0.14 \tanh(1.2s_1) + 0.06 \sin(0.8s_2) + 0.03s_1u, \tag{19}$$

$$c_2(s, u) = -0.11 \tanh(0.9s_2) + 0.05 \cos(1.5s_1) + 0.02u^2. \tag{20}$$

The violation test adds an additional tangential component to the first two rotated coordinates, thereby stressing the graph assumption. Additive Gaussian noise is injected into the target next-state observations Y, i.e., $Y_{\text{noisy}} = Y + \sigma\mathcal{N}(0, I)$, while keeping inputs X_k and U_k unchanged. This simulates measurement uncertainty in the learned dynamics without altering the underlying system trajectories.

The synthetic case shows, in Table 1, that the canonical basis remains the best choice, while the PCA basis is systematically worse. The automatic mode remains close to the canonical solution.

Table 1: Synthetic case: mean trajectory RMSE over 50 Monte Carlo runs.

Scenario	Linear	Canonical	PCA	Auto
Baseline	0.157854	0.012699	0.051291	0.012949
OOD	0.203196	0.074990	0.150764	0.075387
Low data	0.157854	0.024327	0.088613	0.032285
Noisy	0.157854	0.014008	0.051097	0.014223
Violation	0.203356	0.041156	0.108022	0.041385

6.2 Vehicle dynamics case

To assess practical relevance, the same framework is applied to a vehicle-dynamics surrogate. The state is

$$x = [e_y, e_\psi, v_y, r]^T, \tag{21}$$

where e_y and e_ψ denote the lateral position and heading errors with respect to a straight reference path, while v_y and r are the lateral velocity and yaw rate of the vehicle. The input is the steering angle δ.

The nominal linear model is the small-slip bicycle approximation obtained by linearizing the tire forces around zero slip angles. In discrete time, the nominal dynamics are

$$x_{k+1}^{\text{lin}} = Ax_k + B\delta_k, \tag{22}$$

with

$$A = \begin{bmatrix} 1 & \Delta t\, v_x & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 + \Delta t\, a_{11} & \Delta t\, a_{12} \\ 0 & 0 & \Delta t\, a_{21} & 1 + \Delta t\, a_{22} \end{bmatrix}, \qquad B = \begin{bmatrix} 0 \\ 0 \\ \Delta t\, b_1 \\ \Delta t\, b_2 \end{bmatrix}, \tag{23}$$

where

$$a_{11} = -\frac{C_f + C_r}{mv_x}, \qquad a_{12} = -v_x + \frac{-l_f C_f + l_r C_r}{mv_x}, \qquad a_{21} = \frac{-l_f C_f + l_r C_r}{I_z v_x}, \qquad a_{22} = -\frac{l_f^2 C_f + l_r^2 C_r}{I_z v_x}, \tag{24}$$

and

$$b_1 = \frac{C_f}{m}, \qquad b_2 = \frac{l_f C_f}{I_z}. \tag{25}$$

This linear model differs from the nonlinear plant because it replaces the tire saturation law with its local small-slip approximation.

The true plant is a nonlinear bicycle model with Pacejka-type tire saturation:

$$\alpha_f = \delta - \arctan\left(\frac{v_y + l_f r}{v_x}\right), \qquad \alpha_r = -\arctan\left(\frac{v_y - l_r r}{v_x}\right), \tag{26}$$

$$F_{y,f} = 2D_f \sin(C_f \arctan(B_f \alpha_f)), \qquad F_{y,r} = 2D_r \sin(C_r \arctan(B_r \alpha_r)), \tag{27}$$

$$e_{y,k+1} = e_{y,k} + \Delta t\,(v_{y,k} + v_x e_{\psi,k}), \qquad e_{\psi,k+1} = e_{\psi,k} + \Delta t\, r_k, \tag{28}$$

$$v_{y,k+1} = v_{y,k} + \Delta t \left(-v_x r_k + \frac{F_{y,f} + F_{y,r}}{m}\right), \qquad r_{k+1} = r_k + \Delta t \left(\frac{l_f F_{y,f} - l_r F_{y,r}}{I_z}\right). \tag{29}$$

Table 2 reports the mean RMSE values for the linear baseline and the corrected models. In this case, PCA and automatic basis selection consistently outperform the canonical basis.

Table 2: Vehicle dynamics case: mean trajectory RMSE over 50 Monte Carlo runs.

Scenario	Linear	Canonical	PCA	Auto
Baseline	0.101419	0.100556	0.033070	0.033070
OOD	0.244637	0.390641	0.144223	0.144223
Low data	0.101419	2.081159	0.887398	0.967464
Noisy	0.101419	0.105771	0.052717	0.052717
Violation	0.159107	0.114361	0.044160	0.044160

The vehicle results in Table 2 and Figure 2 show that PCA and automatic basis selection outperform the canonical chart, indicating that the most effective geometric representation is data-driven rather than tied to the physical coordinates. The residual scatter is consistent with this view, revealing a largely tangential mismatch in the baseline model and confirming that PCA improves the closure in both tangent and normal directions.

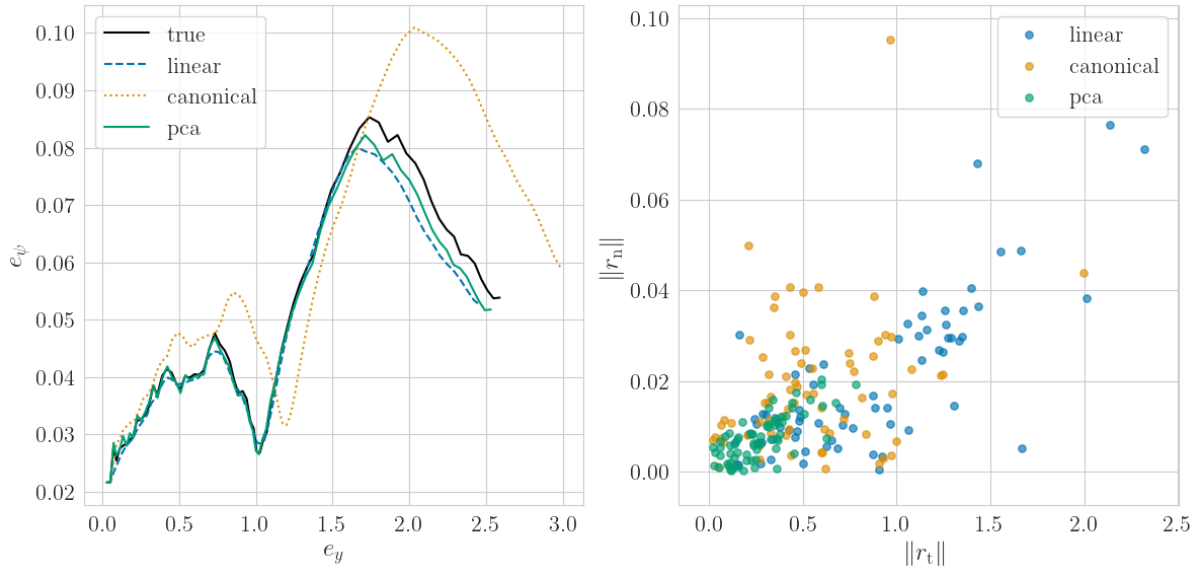

Figure 2: Left. Phase plane of lateral distance from the straight reference path e_y and heading error e_ψ for the first baseline run. Right. Residuals in normal and tangent directions.

7 Discussion

The experiments indicate that the method is effective when the manifold hypothesis is approximately valid and the tangent chart is well chosen. The synthetic case is favorable to the canonical basis, while the vehicle case strongly favors PCA. This confirms that basis selection is problem-dependent and should not be fixed a priori, especially when the physical coordinates are not aligned with the dominant geometric directions of the data.

A further relevant observation is that the correction does not merely reduce the one-step prediction error, but also provides a structured way to interpret model mismatch. The tangential term acts as a geometry-preserving adjustment along the chosen chart, whereas the normal term captures the off-manifold discrepancy. This separation is useful both for prediction and for diagnosis, since it highlights whether the nominal linear model is mainly misaligned along the manifold or whether it fails to represent the manifold itself.

The main strengths of the approach are interpretability, explicit geometric structure, and a clear separation between tangential calibration and normal closure. The main limitations are local validity, sensitivity to the tangent dimension d, and degradation when the manifold is not a graph over the chosen subspace or when data are scarce. In particular, the reduced-data stress test shows that basis selection and regularization can become fragile when the available data do not sufficiently cover the local operating region.

A further limitation is that the current formulation uses a single local chart. When the dynamics exhibit strong regime changes, folds, or multivalued projections, a multi-chart or regime-dependent extension is likely necessary. In that sense, the present framework should be viewed as a local geometric closure method rather than a global manifold reconstruction technique.

Conclusion

This work introduced a geometric closure framework for nominal linear controlled models, designed to reduce the gap between a local linear approximation and the underlying nonlinear dynamics. The central idea is to treat model mismatch not as an unstructured residual, but as a structured correction that can be decomposed into tangential and normal components with respect to a local chart. In this way, the proposed closure preserves the computational simplicity of linear state-space models while adding a geometrically meaningful representation of the neglected nonlinear effects.

The resulting formulation provides two complementary advantages. From a predictive standpoint, it improves trajectory accuracy by learning data-driven corrections on top of the nominal linear model. From an interpretability standpoint, it separates the in-manifold calibration of the dynamics from the off-manifold deformation, making it possible to inspect where and how the linear model fails. The use of bounded dictionaries, ridge regression, and asymmetric regularization further keeps the method lightweight, numerically stable, and suitable for practical implementation.

The numerical experiments support these claims. On the synthetic benchmark, the method successfully recovers the most appropriate chart when the geometry is aligned with the physical coordinates, while the vehicle-dynamics case shows that a data-driven basis such as PCA can be more effective when the natural coordinates are not geometrically optimal. The automatic selection mode provides a practical compromise, remaining close to the best-performing basis in both settings. These results indicate that the success of the closure depends not only on the quality of the residual model, but also on the choice of the local geometric representation.

At the same time, the method has clear limitations. It is inherently local, relies on a meaningful tangent-space approximation, and can degrade when the available data are insufficient or when the true dynamics cannot be represented as a graph over the chosen subspace. For this reason, the present approach should be viewed as a local geometric closure strategy rather than a global nonlinear modeling framework.

Overall, the proposed tangent–normal closure offers a compact and interpretable alternative to generic residual learning for controlled dynamical systems. It is particularly attractive in settings where a linear model is already available and physically meaningful, but requires a structured correction to account for nonlinear effects without abandoning the analytical benefits of the baseline model.

References

[1] D. C. Psichogios and L. H. Ungar, "A hybrid neural network–first principles approach to process modeling," *AIChE Journal*, vol. 38, no. 10, pp. 1499–1511, 1992. DOI: 10.1002/aic.690381003

[2] A. J. Chorin, O. H. Hald, and R. Kupferman, "Optimal prediction and the Mori–Zwanzig representation of irreversible processes," *Proceedings of the National Academy of Sciences*, vol. 97, no. 7, pp. 2968–2973, 2000. DOI: 10.1073/pnas.97.7.2968

[3] P. Stinis, "Renormalized Mori–Zwanzig-reduced models for systems without scale separation," *Proceedings of the Royal Society A*, vol. 471, no. 2176, 20140446, 2015. DOI: 10.1098/rspa.2014.0446

[4] E. J. Parish and K. Duraisamy, "A paradigm for data-driven predictive modeling using field inversion and machine learning," *Journal of Computational Physics*, vol. 305, pp. 758–774, 2016. DOI: 10.1016/j.jcp.2015.11.012

[5] E. J. Parish and K. Duraisamy, "A dynamic subgrid scale model for Large Eddy Simulations based on the Mori–Zwanzig formalism," *Journal of Computational Physics*, vol. 349, pp. 154–175, 2017. DOI: 10.1016/j.jcp.2017.07.053

[6] E. J. Parish and K. Duraisamy, "Non-Markovian closure models for large eddy simulation using the Mori–Zwanzig formalism," *Physical Review Fluids*, vol. 2, 014604, 2017. DOI: 10.1103/PhysRevFluids.2.014604

[7] D. Qi and J. Harlim, "Machine learning-based statistical closure models for turbulent dynamical systems," *Philosophical Transactions of the Royal Society A*, vol. 380, no. 2229, 20210205, 2022. DOI: 10.1098/rsta.2021.0205

[8] B. Sanderse, P. Stinis, R. Maulik, and S. E. Ahmed, "Scientific machine learning for closure models in multiscale problems: A review," *Foundations of Data Science*, vol. 7, no. 1, pp. 298–337, 2025. DOI: 10.3934/fods.2024043

[9] J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar, "Integrating physics-based modeling with machine learning: A survey," *arXiv preprint arXiv:2003.04919*, 2020. URL: https://arxiv.org/abs/2003.04919

[10] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, pp. 422–440, 2021. DOI: 10.1038/s42254-021-00314-5

[11] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019. DOI: 10.1016/j.jcp.2018.10.045

[12] C. Rackauckas, Y. Ma, J. Martensen, C. Warner, K. Zubov, R. Supekar, D. Skinner, A. Ramadhan, and A. Edelman, "Universal differential equations for scientific machine learning," *arXiv preprint arXiv:2001.04385*, 2020. URL: https://arxiv.org/abs/2001.04385

[13] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, "Neural ordinary differential equations," *arXiv preprint arXiv:1806.07366*, 2018. URL: https://arxiv.org/abs/1806.07366

[14] Y. Rubanova, R. T. Q. Chen, and D. K. Duvenaud, "Latent ordinary differential equations for irregularly-sampled time series," in *Advances in Neural Information Processing Systems 32*, 2019. URL: https://arxiv.org/abs/1907.03907

[15] S. Greydanus, M. Dzamba, and J. Yosinski, "Hamiltonian neural networks," *arXiv preprint arXiv:1906.01563*, 2019. URL: https://arxiv.org/abs/1906.01563

[16] L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, "Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators," *Nature Machine Intelligence*, vol. 3, no. 3, pp. 218–229, 2021. DOI: 10.1038/s42256-021-00302-5

[17] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 113, no. 15, pp. 3932–3937, 2016. DOI: 10.1073/pnas.1517384113

[18] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Sparse identification of nonlinear dynamics with control (SINDYc)," *IFAC-PapersOnLine*, vol. 49, no. 18, pp. 710–715, 2016. DOI: 10.1016/j.ifacol.2016.10.249

[19] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Data-driven discovery of partial differential equations," *Science Advances*, vol. 3, no. 4, e1602614, 2017. DOI: 10.1126/sciadv.1602614

[20] K. P. Champion, B. Lusch, J. N. Kutz, and S. L. Brunton, "Data-driven discovery of coordinates and governing equations," *Proceedings of the National Academy of Sciences*, vol. 116, no. 45, pp. 22445–22451, 2019. DOI: 10.1073/pnas.1906995116

[21] S. Pan and K. Duraisamy, "Data-driven discovery of closure models," *SIAM Journal on Applied Dynamical Systems*, vol. 17, no. 4, pp. 2381–2413, 2018. DOI: 10.1137/18M1177263

[22] O. San and R. Maulik, "Neural network closures for nonlinear model order reduction," *Advances in Computational Mathematics*, vol. 44, pp. 1717–1750, 2018. DOI: 10.1007/s10444-018-9590-z

[23] O. San and R. Maulik, "Machine learning closures for model order reduction of thermal fluids," *Applied Mathematical Modelling*, vol. 60, pp. 681–710, 2018. DOI: 10.1016/j.apm.2018.03.037

[24] A. Gupta and P. F. J. Lermusiaux, "Neural closure models for dynamical systems," *Proceedings of the Royal Society A*, vol. 477, no. 2252, 20201004, 2021. DOI: 10.1098/rspa.2020.1004

[25] A. Gupta and P. F. J. Lermusiaux, "Generalized neural closure models with interpretability," *Scientific Reports*, vol. 13, 10634, 2023. DOI: 10.1038/s41598-023-35319-w

[26] B. M. de Silva, K. Champion, M. Quade, J.-C. Loiseau, J. N. Kutz, and S. L. Brunton, "PySINDy: A Python package for the sparse identification of nonlinear dynamics from data," *Journal of Open Source Software*, vol. 5, no. 49, 2104, 2020. DOI: 10.21105/joss.02104

[27] A. A. Kaptanoglu, B. M. de Silva, K. T. Foerster, J. N. Kutz, and S. L. Brunton, "PySINDy: A comprehensive Python package for robust sparse system identification," *Journal of Open Source Software*, vol. 7, no. 69, 3994, 2022. DOI: 10.21105/joss.03994

[28] D. Ghosh, E. Hermonat, P. Mhaskar, S. Snowling, and R. Goel, "Hybrid modeling approach integrating first-principles models with subspace identification," *Industrial & Engineering Chemistry Research*, vol. 58, no. 30, pp. 13533–13543, 2019. DOI: 10.1021/acs.iecr.9b00900

[29] B. Sun, C. Yang, Y. Wang, W. Gui, I. Craig, and L. Olivier, "A comprehensive hybrid first principles/machine learning modeling framework for complex industrial processes," *Journal of Process Control*, vol. 86, 30–43, 2020. DOI: 10.1016/j.jprocont.2019.11.012

[30] H. Narayanan, M. von Stosch, F. Feidl, M. Sokolov, M. Morbidelli, and A. Butté, "Hybrid modeling for biopharmaceutical processes: advantages, opportunities, and implementation," *Frontiers in Chemical Engineering*, vol. 5, 1157889, 2023. DOI: 10.3389/fceng.2023.1157889

[31] L. Rajulapati, S. Chinta, B. Shyamala, and R. Rengaswamy, "Integration of machine learning and first principles models," *AIChE Journal*, vol. 68, no. 6, e17715, 2022. DOI: 10.1002/aic.17715

[32] S. Yang, P. Navarathna, S. Ghosh, and B. W. Bequette, "Hybrid modeling in the era of smart manufacturing," *Computers & Chemical Engineering*, vol. 140, 106874, 2020. DOI: 10.1016/j.compchemeng.2020.106874

[33] W. Bradley, J. Kim, Z. Kilwein, L. Blakely, M. Eydenberg, J. Jalvin, C. Laird, and F. Boukouvala, "Perspectives on the integration between first-principles and data-driven modeling," *Computers & Chemical Engineering*, vol. 166, 107898, 2022. DOI: 10.1016/j.compchemeng.2022.107898

[34] P. Shah, S. Pahari, R. Bhavsar, and J. S. I. Kwon, "Hybrid modeling of first-principles and machine learning: A step-by-step tutorial review for practical implementation," *Computers & Chemical Engineering*, vol. 194, 108926, 2025. DOI: 10.1016/j.compchemeng.2024.108926

[35] C. Lathourakis and A. Cicirello, "Physics enhanced sparse identification of dynamical systems with discontinuous nonlinearities," *Nonlinear Dynamics*, vol. 112, no. 13, pp. 11237–11264, 2024. DOI: 10.1007/s11071-024-09652-2

[36] Z. Yang, X. Shan, X. I. A. Yang, and W. Zhang, "Data-enabled discovery of specific and generalisable turbulence closures," *Journal of Fluid Mechanics*, vol. 1016, R1, 2025. DOI: 10.1017/jfm.2025.10332

[37] D. González, F. Chinesta, and E. Cueto, "Learning corrections for hyperelastic models from data," *Frontiers in Materials*, vol. 6, 14, 2019. DOI: 10.3389/fmats.2019.00014

[38] S. Asgari, S. M. MirhoseiniNejad, H. Moazamigoodarzi, R. Gupta, R. Zheng, and I. K. Puri, "A gray-box model for real-time transient temperature predictions in data centers," *Applied Thermal Engineering*, vol. 185, 116319, 2021. DOI: 10.1016/j.applthermaleng.2020.116319