The University of California, Santa Cruz

AM170B: Mathematical Modeling II

State Estimation of the Multi-Scale Lorenz 96 System Using Echo State Networks, Knowledge-Based Models, Hybrid Approaches, and Ensemble Kalman Filters

May 30, 2025

Maxwell Arcinas, Josiah Aviles, Mohamed Badri, Marcus Chan, Anderson Compalas

Abstract

State estimation in multiscale chaotic systems remains a critical challenge as slow, medium, and fast dynamics are difficult to capture using purely physics-based or data driven methods. Traditional imperfect models that truncate fast-scale interactions often drift, while standalone reservoir networks can diverge under long-term forecasts. In this work, we develop and compare three state-estimation systems for the Lorenz 96 system - (1) an imperfect model, (2) a pure echo state network (ESN), (3) hybrid ESN combining the imperfect model with reservoir corrections, and (4) each of these coupled with an Ensemble Kalman Filter (EnKF) for data assimilation. We generate datasets using the Lorenz 96 system, standardize the resulting time series, and use them to train and test each method. The experiments reveal that the imperfect model maintains accuracy longer than the ESN in isolation, while the hybrid ESNs on the Lorenz 96 fail to capture system dynamics. When working with the EnKF under varying observation intervals and noise levels, the imperfect model with the EnKF consistently retains NRMSE below 0.4, where as the ESN + EnKF shows sharp degradation under temporally sparse observables. Our findings demonstrate the value of combining physics based systems with data assimilation methods to achieve robust state estimation in higher-dimensional chaotic systems.

I. INTRODUCTION

This work investigates different state estimation strategies for the multiscale Lorenz 96 system using three modeling approaches: a knowledge-based imperfect model, a data-driven Echo State Network (ESN), and a hybrid architecture combining both. Our hybrid scheme was directly motivated by the results of [6], who demonstrated that a hybrid ESN significantly outperformed its constituent models on the Kuramoto–Sivashinsky (KS) system. We sought to evaluate whether similar performance gains could be achieved on the more challenging Lorenz 96 system.

We began by evaluating the forecasting ability of each standalone model using normalized root mean square error (NRMSE) and valid time metrics. The imperfect model, despite ignoring the fastest (Z) variables, exhibited better predictive stability than the ESN alone. The hybrid schemes underperformed both individual models, failing to preserve meaningful dynamics or reduce long-term error.

Next, we evaluated the effect of combining each model with an Ensemble Kalman Filter (EnKF) for state estimation under varying observational frequencies and noise

levels. We excluded the hybrid model from this assimilation stage due to its poor forecast quality. The imperfect model demonstrated robust estimation performance across all tested regimes of observation frequency and noise. In contrast, the ESN + EnKF system performed well only when observations were available at every timestep. Its performance degraded significantly under reduced temporal frequency, despite prior evidence showing that ESN + EnKF architectures perform well under added noise and spatial sparsity [4].

In what follows, we describe the underlying system in Section II, outline our different forecasting methods in Section III, assess both forecasting and state estimation capabilities in Section IV, and finally conclude in Section V discussing takeaways, caveats, and future direction.

II. MODEL

We investigate the behavior of a multiscale chaotic system governed by the Lorenz 96 equations. The system consists of three hierarchically coupled levels of dynamics:

$$
\begin{aligned}
X_k &\in \mathbb{R}^K && \text{(large-scale, slow variables)} \\
Y_{j,k} &\in \mathbb{R}^{K\times J} && \text{(intermediate-scale variables coupled to } X_k\text{)} \\
Z_{l,j,k} &\in \mathbb{R}^{K\times J\times L} && \text{(fast, small-scale variables nested within each } Y_{j,k}\text{)}
\end{aligned}
$$

The evolution of the full system is described by the following ordinary differential equations (ODEs):

$$\frac{dX_k}{dt} = X_{k-1}(X_{k+1} - X_{k-2}) - X_k + F - \frac{hc}{b}\sum_{j=1}^{J} Y_{j,k}, \tag{1}$$

$$\frac{dY_{j,k}}{dt} = -cb\,Y_{j+1,k}(Y_{j+2,k} - Y_{j-1,k}) - cY_{j,k} + \frac{hc}{b}X_k - \frac{he}{d}\sum_{l=1}^{L} Z_{l,j,k}, \tag{2}$$

$$\frac{dZ_{l,j,k}}{dt} = ed\,Z_{l-1,j,k}(Z_{l+1,j,k} - Z_{l-2,j,k}) - eZ_{l,j,k} + \frac{he}{d}Y_{j,k}. \tag{3}$$

Each equation includes nonlinear advection, damping, and cross-scale coupling terms. Indices are periodic in space, and constants such as F, h, c, b, e, and d control the strength of forcing and scale separation.

To generate data, we simulate this system using a python script, adapted from [2]. In our setup, we fix $K = 8$, $J = 8$, and $L = 8$. The simulation initializes X_k with random

integers between -5 and 5, and $Y_{j,k}, Z_{l,j,k}$ with Gaussian noise. We evolve the system using a fourth-order Runge–Kutta (RK4) integrator with a timestep of $\Delta t = 0.005$, for a total of 1,500,000 steps (equivalent to a physical duration of 75,000 units).

At each timestep, we store the full X and Y states. After simulation, the time series are standardized (zero mean, unit variance), and the X and Y vectors are concatenated to form a final dataset of shape (1500000, 72), where each row corresponds to a full system state. This data is used as the ground truth for the state estimation experiments in this project.

III. METHODS

A. Imperfect Model

The first method we employ is a simplified, physics-based imperfect model adapted from [2]. This model retains the ODEs for the large-scale X_k variables and their coupling to the intermediate-scale $Y_{j,k}$ variables, but omits the influence of the fast $Z_{l,j,k}$ variables. Specifically, the Y-equation is truncated by removing the final term representing Z-scale coupling.

The imperfect model dynamics are given by:

$$\frac{dX_k}{dt} = X_{k-1}(X_{k+1} - X_{k-2}) - X_k + F - \frac{hc}{b}\sum_{j=1}^{J} Y_{j,k}, \tag{4}$$

$$\frac{dY_{j,k}}{dt} = -cb\, Y_{j+1,k}(Y_{j+2,k} - Y_{j-1,k}) - cY_{j,k} + \frac{hc}{b}X_k. \tag{5}$$

This truncated system preserves the core nonlinear dynamics but lacks the multiscale feedback from the Z variables, making it an "imperfect" yet numerically stable model for the full system. It serves as a baseline dynamical model in our state estimation framework.

To evolve the imperfect model, we sample an initial condition from the normalized dataset described in Section II. We then denormalize the selected initial states using the empirical means and standard deviations computed during data generation:

$$X_0 = \tilde{X}_0 \cdot \sigma_X + \mu_X, \tag{6}$$

$$Y_0 = \tilde{Y}_0 \cdot \sigma_Y + \mu_Y, \tag{7}$$

where $\tilde{X}_0$ and $\tilde{Y}_0$ denote the normalized initial conditions, and μ_X, μ_Y, σ_X, and σ_Y are the saved mean and standard deviation vectors for X and Y, respectively.

The system is then evolved forward in time using a fourth-order Runge–Kutta (RK4) integrator with timestep $\Delta t = 0.005$:

$$[X_{n+1}, Y_{n+1}] = \mathrm{RK4}(X_n, Y_n, \Delta t), \tag{8}$$

After simulation, the predicted X and Y trajectories can be re-normalized to allow comparison against the normalized ground truth:

$$\tilde{X}_t^{(\mathrm{model})} = \frac{X_t - \mu_X}{\sigma_X}, \tag{9}$$

$$\tilde{Y}_t^{(\mathrm{model})} = \frac{Y_t - \mu_Y}{\sigma_Y}. \tag{10}$$

In our results section, we assess the predictive accuracy of the imperfect model on its own and when combined with data assimilation via the Ensemble Kalman Filter (EnKF).

B. Echo State Network

Our second method leverages a data-driven dynamical model known as the Echo State Network (ESN), adapted from [1]. ESNs are a form of reservoir computing in which a fixed, randomly generated, sparse Erdős–Rényi network (the "reservoir") transforms input signals into high-dimensional dynamical features. Only the final output weights are trained, which allows for efficient learning and fast prediction.

The ESN consists of:

- A sparse recurrent matrix $A \in \mathbb{R}^{N \times N}$ with spectral radius ρ,
- An input weight matrix $W_{\mathrm{in}} \in \mathbb{R}^{N \times d}$, where d is the input dimension,
- A nonlinear activation tanh applied at each step.

Given an input sequence $\{u(t)\}$, the reservoir state evolves as:

$$r(t+1) = \tanh(A\, r(t) + W_{\mathrm{in}}\, u(t)), \tag{11}$$

where $r(t) \in \mathbb{R}^N$ is the internal state of the reservoir at time t.

To train the ESN, we collect all reservoir states over a training window of length T to form the matrix $R \in \mathbb{R}^{N\times T}$. We then compute the output weights W_{out} to map the reservoir states to the target signal $Y \in \mathbb{R}^{d\times T}$.

Following [2], we augment the reservoir states with nonlinear cross-product terms for even-indexed units, defined as:

$$r_j^*(t) = \begin{cases} r_{j-1}(t) \cdot r_{j-2}(t), & \text{if } j \text{ is even and } j \geq 2, \\ r_j(t), & \text{otherwise.} \end{cases} \tag{12}$$

We apply Tikhonov regularization (ridge regression) to compute the output weights:

$$W_{\text{out}} = Y R^{*T} \left(R^* R^{*T} + \beta I \right)^{-1}, \tag{13}$$

where R^* is the matrix of augmented states and β is a small regularization constant.

At inference time, the trained reservoir is used to generate predictions in an autoregressive fashion which outputs a prediction and updates the reservoir state. At each step, the output $\hat{u}(t)$ and reservoir state $r(t)$ is computed as:

$$r^*(t) = \text{nonlinear transform of } r(t), \tag{14}$$

$$\hat{u}(t) = W_{\text{out}}\, r^*(t), \tag{15}$$

$$r(t+1) = \tanh(A\, r(t) + W_{\text{in}}\, \hat{u}(t)). \tag{16}$$

Our implementation follows the above structure using sparse matrices and efficient NumPy-based operations. We fix the reservoir size $N = 4992$, with an input dimension of $d = 8$ (corresponding to the X variables), and train over 500,000 time steps. The reservoir matrices A and W_{in} are generated randomly and fixed, while W_{out} is computed analytically using the pseudo-inverse form of ridge regression. Training and inference scripts were adapted from the codebase of [1].

C. Hybrid Echo State Network

We aim to extend the hybrid forecasting framework introduced by [6], which demonstrated that combining a partial physics-based model with a data-driven Echo State Network (ESN) can significantly improve forecast accuracy greater than that of individual models. Their study, conducted on the spatiotemporally chaotic Kuramoto–Sivashinsky

(KS) system, showed that a hybrid ESN outperformed both the standalone reservoir model and the imperfect physics-based model. Motivated by this success, we investigate whether a similar hybrid strategy can enhance prediction in the multiscale Lorenz 96 system, where the dynamics involve both slow and fast variables coupled across multiple scales.

In our project, we adapt this hybrid modeling framework to the multiscale Lorenz 96 system described in Section II. Our goal is to assess whether a similar hybrid approach can improve state estimation performance in a chaotic system where partial physical knowledge, like our truncated imperfect model, is available.

Our hybrid model integrates two components: (1) an imperfect physics-based model $K[X(t), Y(t)]$ that evolves the X and Y variables using a truncated version of the Lorenz 96 equations (omitting Z-coupling), and (2) a reservoir computer (ESN) trained to correct forecast errors from the imperfect model. At each timestep, the hybrid input vector is defined as:

$$u(t) = \begin{bmatrix} K[X(t), Y(t)] \\ X(t) \end{bmatrix}, \quad \text{where } K[X(t), Y(t)] = [X(t + \Delta t), Y(t + \Delta t)].$$

During training, $X(t)$ is the ground truth; during inference, it is the previous hybrid prediction. This structure allows the ESN to incorporate both the imperfect forecast and current system state, updating its internal reservoir accordingly. A learned output matrix W_{out} then maps the reservoir state to a prediction of either $X(t)$ alone or the full state $(X(t), Y(t))$, depending on the hybrid scheme used.

We explore three hybridization schemes, distinguished by how they couple the ESN with the imperfect model. These schemes are summarized in Table I.

TABLE I: Comparison of Hybrid ESN Integration Strategies

Scheme	**ESN Output**	**Used by Imperfect Model**	**Model Coupling**
X-only (Coupled)	$X(t)$ only	$\hat{X}(t) \rightarrow$ evolve $Y(t)$	Partial coupling
Joint (X, Y)	$X(t), Y(t)$	$\hat{X}(t), \hat{Y}(t) \rightarrow$ evolve both	Full coupling
X-only (Uncoupled)	$X(t)$ only	None; imperfect model evolves on its own	Decoupled

Training follows the same procedure as in the standalone ESN (see Section III B), with one key modification: during training, the output of the imperfect model $K[X(t), Y(t)]$

is concatenated to the nonlinearly transformed reservoir state $r^*(t)$ before calculating the output weights. Specifically, the target variables (either X or (X, Y), depending on the scheme) are regressed against the extended state vector:

$$\tilde{r}(t) = \begin{bmatrix} K[X(t), Y(t)] \\ r^*(t) \end{bmatrix}. \tag{17}$$

Tikhonov regularization is then used to compute the output weights W_{out} via:

$$W_{\text{out}} = Y\,\tilde{R}^T \left(\tilde{R}\tilde{R}^T + \beta I\right)^{-1}, \tag{18}$$

where $\tilde{R}$ is the matrix of augmented states collected over the training window.

Theoretically, this hybrid architecture should allow the ESN to implicitly learn residual corrections to the imperfect model's dynamics, leveraging both physical structure and data-driven adaptability. In our experiments, we evaluate how each of the three hybrid integration schemes described above affects the accuracy of forecasting and state estimation.

D. Ensemble Kalman Filter

To further our study of state estimation, we implemented an Ensemble Kalman Filter (EnKF) to perform data assimilation on our model. The goal was to correct the trajectory of an imperfect model by assimilating synthetic observations derived from the ground-truth data. This approach follows a classic twin experiment framework, as described in prior EnKF papers, such as the work on Lorenz's 1963 model by Exploring our Magnetic Earth [5].

Synthetic observations were constructed by sampling the true trajectory at various assimilation intervals, then perturbing the samples states with Gaussian noise. The observation noise was drawn from $\mathcal{N}(0, R)$ where $R = \sigma_{obs}^2 I_K$ and σ_{obs} denotes the standard deviation of the observational uncertainty.

The initial ensemble was sampled from a multivariate normal distribution $\mathcal{N}(x_0, P_0)$, where x_0 corresponds to the first state of the true trajectory and $P_0 = I_K$ represents the initial uncertainty. We propagated each ensemble member x_i^f using the RK4 step and replaced $X(t_a)$ with the ensemble mean at the assimilation time step (t_a) updated using the standard Kalman update formula and Kalman gain:

$$x_i^a = x_i^f + K(y^{obs} + \epsilon_i - Hx_i^f) \tag{19}$$

$$K = P_{xy} P_{yy}^{-1}. \tag{20}$$

Here, H is the observation matrix ($H = I_k$, full state observed), $\epsilon_i \sim \mathcal{N}(0, R)$ is a stochastic perturbation to preserve ensemble spread, and the superscripts f and a denote forecast and analysis ensembles, respectively. The cross-covariance P_{xy} and the observation covariance P_{yy} are calculated using the anomalies in $\mathbf{X}$ and $\mathbf{Y}$.

To evaluate the robustness and efficiency of the EnKF under different observational conditions, we performed a series of trials varying both the assimilation interval and the observational noise level. Specifically, the filter was tested across observation intervals of $1, 5, 10,$ and 20 time steps, and for each interval it was run under four noise levels: $\sigma_{obs} = 0.1, 0.25, 0.5,$ and 1.0. Each configuration was repeated across 10 independent trials to assess stability and variability. Performance was evaluated using normalized root mean square error (NRMSE) between the EnKF-corrected forecast and the true trajectory, allowing a systematic comparison of forecast accuracy under varied observational regimes. The formulation follows the standard procedures described in Evensen's work on ensemble-based filtering methods [3].

E. Metrics

To evaluate the accuracy and stability of each modeling approach, we use two key metrics: the Normalized Root Mean Squared Error (NRMSE) and the Valid Time.

NRMSE measures the relative deviation of the predicted state from the true state over time. At each timestep t, we compute the ℓ_2 norm of the error between the true and predicted state vectors, and normalize it using a global scaling factor derived from the ground truth.

Formally, let $X_{\text{true}}(t) \in \mathbb{R}^K$ and $X_{\text{pred}}(t) \in \mathbb{R}^K$ denote the true and predicted X vectors at time t, respectively. The NRMSE at time t is computed as:

$$\text{NRMSE}(t) = \frac{\|X_{\text{true}}(t) - X_{\text{pred}}(t)\|_2}{\sqrt{\frac{1}{T}\sum_{s=1}^{T} \|X_{\text{true}}(s)\|_2^2}}, \tag{21}$$

where the denominator is the root mean squared norm of the true states across all T timesteps. This global normalization ensures that NRMSE values are comparable across models and runs.

We define the *Valid Time* as the first timestep at which the NRMSE exceeds a specified threshold τ. That is,

$$t_{\text{valid}} = \min\left\{t \in [1, T] \mid \text{NRMSE}(t) > f\right\}, \tag{22}$$

if such a t exists; otherwise, t_{valid} is undefined. In our experiments, we set $f = 0.4$ which was the same threshold used in [6].

The valid time provides a measure of how long a model can produce accurate forecasts or state estimates before diverging significantly from the ground truth.

IV. RESULTS

In this section, we evaluate the performance of each model across two main tasks. First, we assess the standalone forecasting accuracy of the Imperfect Model, Echo State Network (ESN), and Hybrid ESN by comparing their predictions to the ground truth without any data assimilation. This allows us to isolate the predictive capabilities of each modeling approach.

Next, we evaluate state estimation performance when each model is combined with the Ensemble Kalman Filter (EnKF). This tests how well each system can incorporate observational data to correct its internal state and improve long-term accuracy under partial observability.

A. Imperfect Model Forecast

Figure 1 shows the results of a single recursive rollout of the Imperfect Model over 1000 timesteps, starting from $t = 500{,}000$. The contours are plotted in normalized state space, where both the ground truth and model predictions have been standardized to zero mean and unit variance. The first panel displays the ground truth X_k values, the second shows the forecasted trajectory from the imperfect model, and the third panel presents the pointwise error between the two. A vertical black dotted line marks the *valid time* $\tau = 334$, which is the first timestep at which the NRMSE exceeds the threshold value $f = 0.4$. This visualization highlights the initial alignment and eventual divergence between the imperfect model and the true dynamics, illustrating both its short-term predictive power and long-term instability.

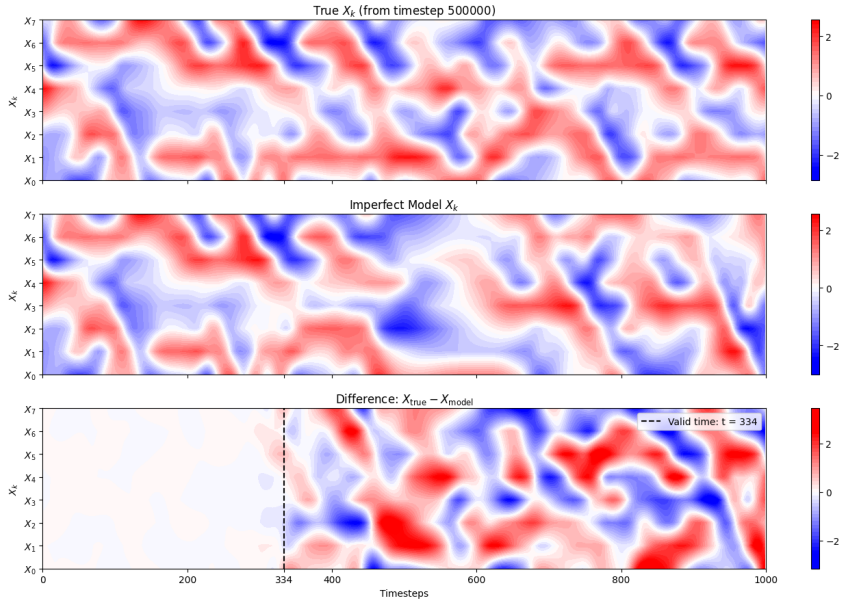

FIG. 1: Countour Plot for one run of the Imperfect model with a recursive rollout for 1000 timesteps from $t = 500,000$. Performance crosses our $f = 0.4$ threshold at a valide time of $\tau = 334$

Figure 2 shows the Normalized Root Mean Squared Error (NRMSE) over time for the same 1000-timestep rollout of the Imperfect Model beginning at t = 500,000. The NRMSE is computed using a global normalization factor based on the RMS magnitude of the ground truth states (as described in Section III E). As the forecast progresses, the error remains low initially but gradually increases due to the model's inability to capture the full dynamics of the system. The point at which the NRMSE exceeds the threshold $f = 0.4$ corresponds to the valid time $\tau = 334$, consistent with the divergence marked in Figure 1.

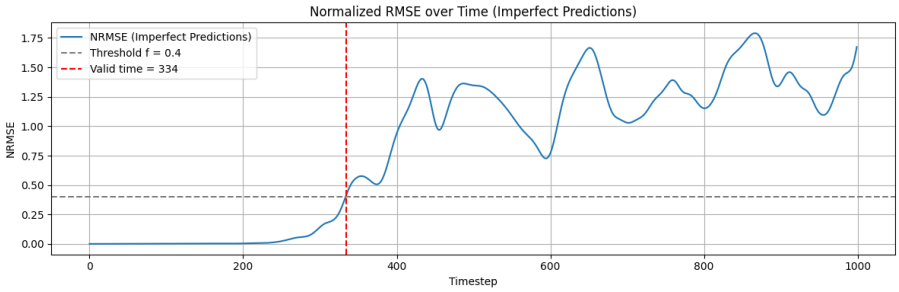

FIG. 2: NRMSE for Imperfect Model prediction from $t = 500,000$ for 1000 timestes.

Figure 3 shows the average NRMSE over time for the Imperfect Model, computed across 10 independent runs with different initial conditions. Each forecast spans 1000

timesteps, and the NRMSE at each time step is averaged across trials. Shaded bands represent one standard deviation around the mean, capturing the variability in model performance due to differing initial conditions. On average, the model remains accurate for approximately 356 timesteps before exceeding the threshold of $f = 0.4$, which we report as the mean valid time $\tau = 356$. This analysis provides a more robust assessment of model performance beyond single-run diagnostics.

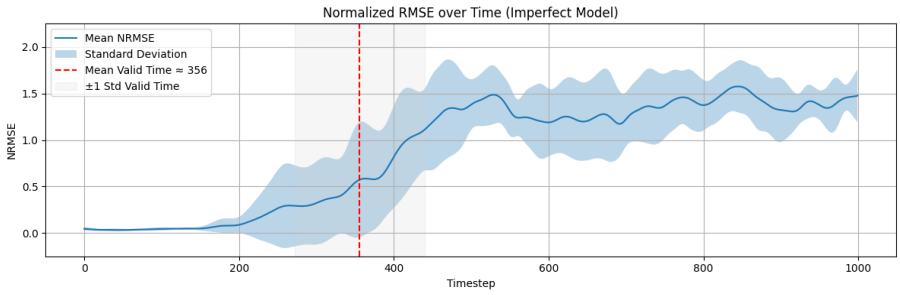

FIG. 3: Mean NRMSE plot averaged over 10 different initial conditions of the imperfect model forecast for 1000 timesteps. Mean valid time for the imperfect model forecast is $\tau = 356$

B. ESN Forecast

Figure 4 shows a single recursive rollout of the Echo State Network (ESN) over 1000 timesteps, starting from $t = 501{,}000$. Since the ESN was trained on the first 500,000 timesteps, we initialize the reservoir by performing a warm-up: the reservoir is updated sequentially using the ground truth inputs until the desired initial condition is reached. This ensures that the internal reservoir state is synchronized with the dynamics before forecasting begins.

As in previous figures, the first panel shows the normalized ground truth X_k values, the second shows the ESN forecast, and the third panel presents the error. A vertical dotted line marks the valid time $\tau = 115$, the first timestep at which the NRMSE exceeds the threshold $f = 0.4$. This result shows earlier divergence compared to the Imperfect Model run shown in Figure 1, where $\tau = 334$. For this specific initial condition, the ESN performs worse than the imperfect model in terms of predictive stability.

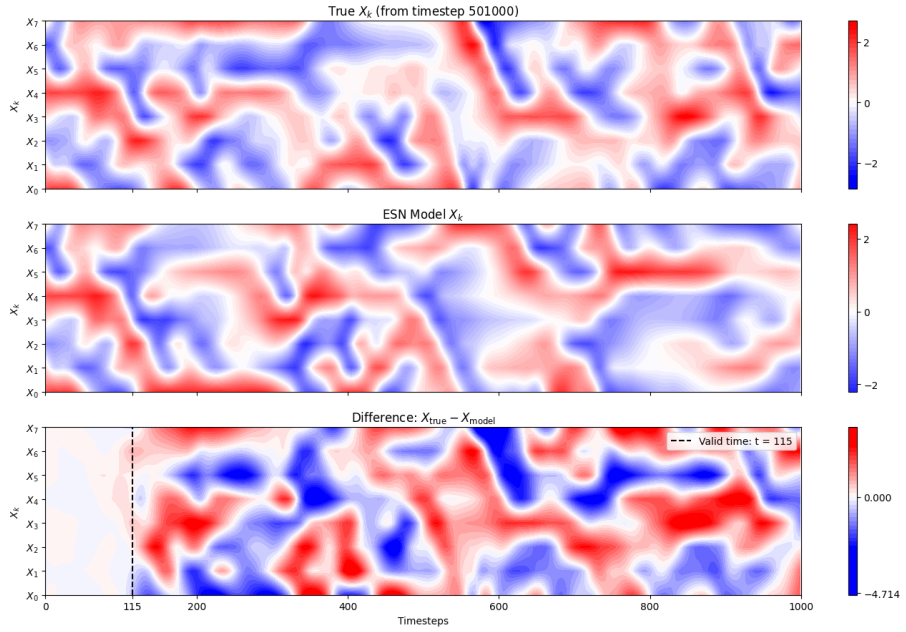

FIG. 4: Countour Plot for one run of the ESN model with a recursive rollout for 1000 timesteps from $t = 501,000$. Performance crosses our $f = 0.4$ threshold at a valid time of $\tau = 115$ for this specific run.

Figure 5 shows the mean NRMSE over time for the ESN model, averaged across 10 independent runs with different initial conditions after the training window ($t = 500,000$). Each rollout spans 1000 timesteps, and the ESN reservoir is warmed up using the ground truth data prior to the initial condition, as described earlier. The shaded region indicates one standard deviation around the mean, capturing the variability in forecast quality across different initializations. On average, the ESN model exceeds the NRMSE threshold of $f = 0.4$ at timestep $\tau = 195$, indicating a shorter valid time compared to the Imperfect Model (which had $\tau = 356$ on average). While the ESN can capture short-term dynamics, it tends to diverge more rapidly than our knowledge-based imperfect model.

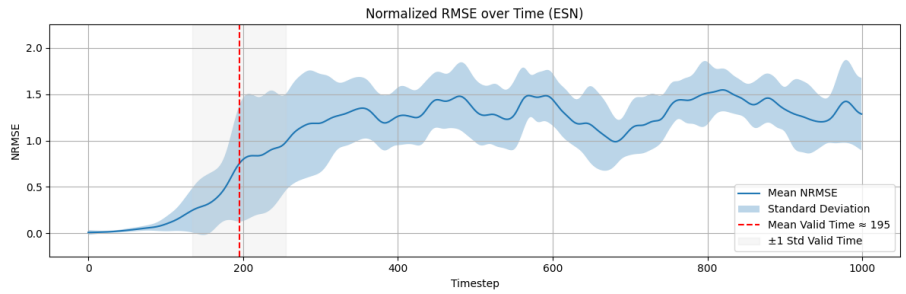

FIG. 5: Mean NRMSE plot averaged over 10 different initial conditions of the ESN model forecast for 1000 timesteps. Mean valid time for the imperfect model forecast is $\tau = 195$

C. Hybrid ESN Forecast

We now evaluate the forecasting performance of the three hybrid integration schemes described in the previous Section III C. Each configuration combines the imperfect model and ESN in a different manner to address potential failure modes in standalone approaches.

Figure 6 shows the result for the X-only coupled scheme. Despite being trained for the same 500,000 timesteps as the standalone ESN, the hybrid model quickly collapses to a nearly static output. The reservoir dynamics become highly regular and fail to track the variability of the ground truth system. This breakdown suggests that feeding predicted $X(t)$ into the imperfect model is insufficient to stabilize or guide the joint evolution when $X(t)$ and $Y(t)$ become misaligned.

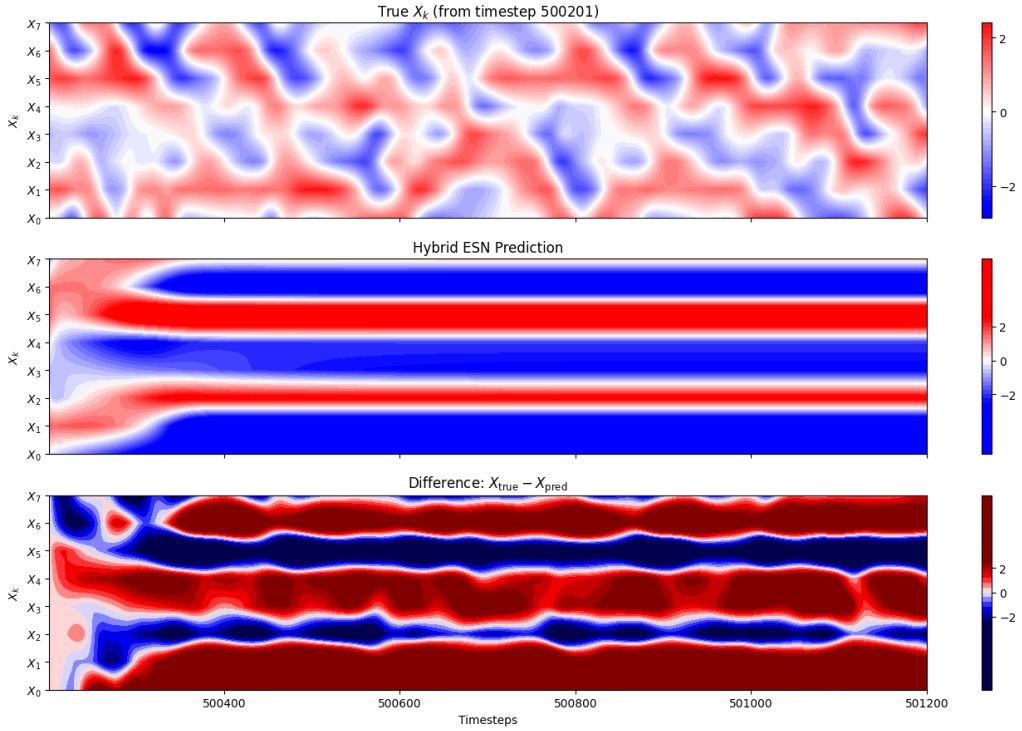

FIG. 6: Forecast performance of the Hybrid ESN with X-only prediction and partial coupling. The ESN predicts only the X variables, which are fed into the imperfect model to evolve $Y(t)$.

Figure 7 examines the uncoupled configuration, where the Hybrid ESN and imperfect model evolve independently after initialization. Although this scheme avoids the misalignment introduced by feedback coupling, it still struggles to produce accurate forecasts. The model appears to resemble structure very early on but quickly diverges. Without coor-

dinated correction, the reservoir and physics-based dynamics can evolve inconsistently, degrading long-term performance.

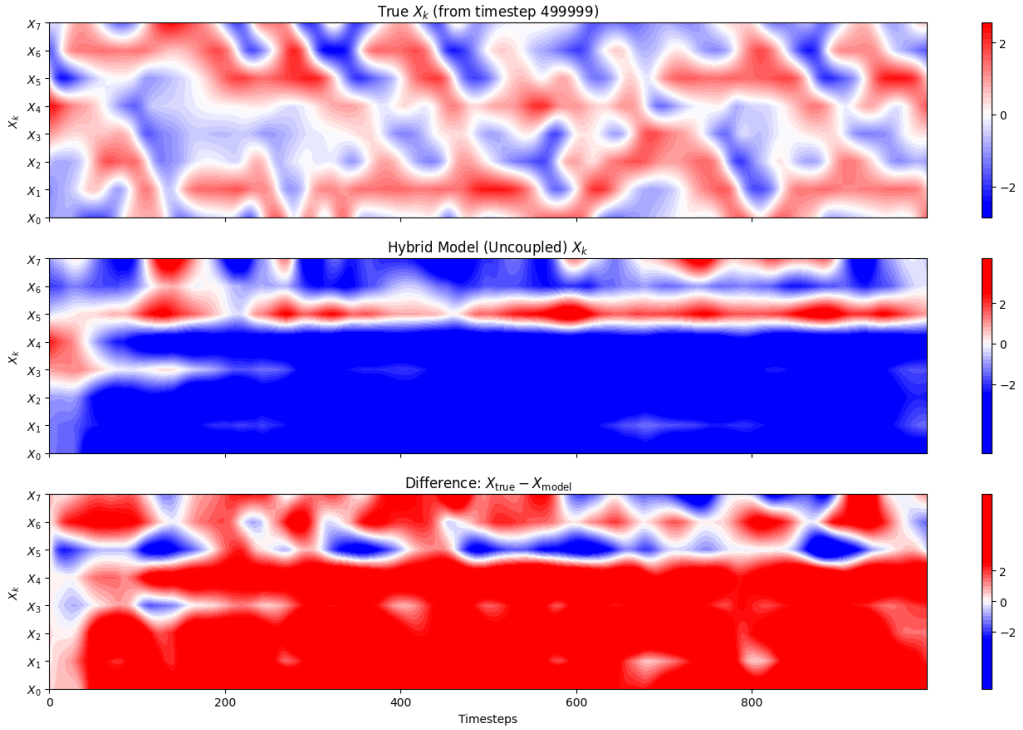

FIG. 7: Forecast performance of the Hybrid ESN with X-only prediction and no coupling. The imperfect model and ESN evolve independently after initialization.

Finally, Figure 8 presents results from the fully coupled XY prediction scheme. This configuration avoids temporal inconsistency by having the ESN predict both $X(t)$ and $Y(t)$ simultaneously. These predictions are then used to update both models. While the resulting dynamics show more internal variation than the X-only variants, the model still fails to capture the true trajectory. Even with synchronized predictions, the system diverges from the correct attractor structure within a short forecast window. (Note: training length here was reduced to 100,000 timesteps due to memory constraints.)

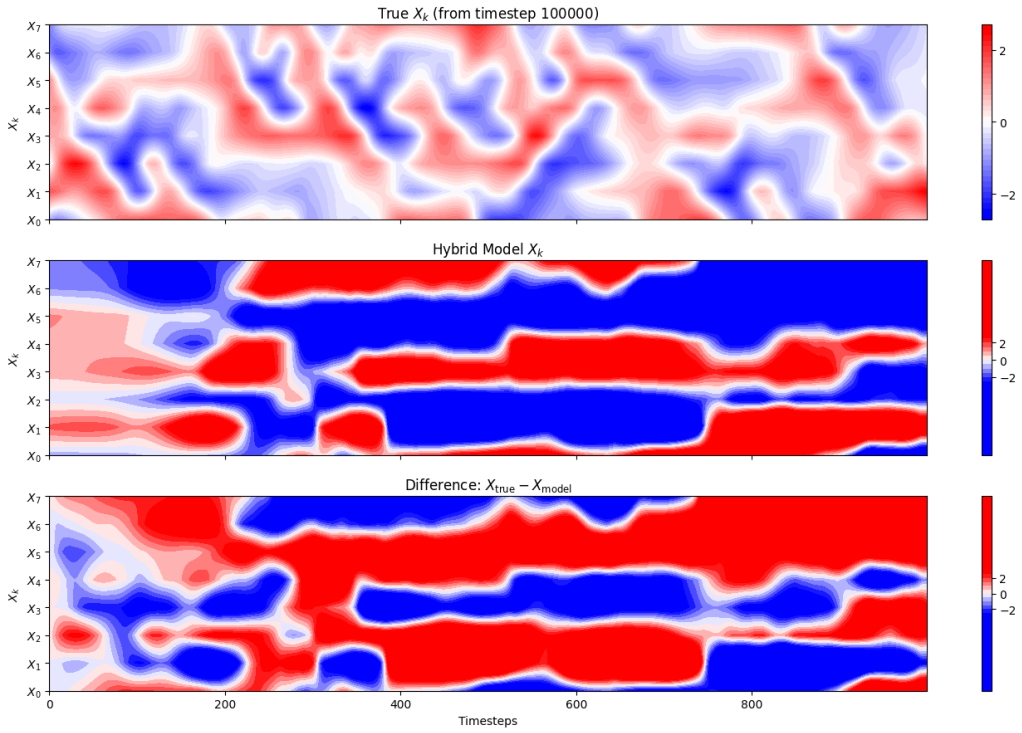

FIG. 8: Forecast performance of the fully coupled Hybrid ESN with joint (X, Y) prediction. Predictions are used to evolve both the reservoir and the imperfect model.

Overall, none of the hybrid schemes produce forecasts that resemble the true system behavior. In contrast to the benchmark study by [6], which demonstrated successful hybridization using the Kuramoto–Sivashinsky (KS) system, our results on the multiscale Lorenz 96 system indicate that the hybrid approach may not generalize well across different dynamical systems. The KS system is univariate and spatially chaotic, whereas Lorenz 96 is multivariate and features strong timescale separation, properties that may challenge hybrid integration.

These findings highlight the need for greater customization when extending hybrid forecasting frameworks to systems with hierarchical coupling. Future directions may include more robust synchronization mechanisms or learned co-evolution schemes between physics-based and data-driven components.

D. Imperfect Model + EnKF State Estimation

To evaluate the effectiveness of data assimilation, we combine the knowledge-based Imperfect Model with the Ensemble Kalman Filter (EnKF) for state estimation. In subsequent experiments, we use 100 ensemble members and simulate 10 different initial

conditions under varying observation noise levels and observation frequencies. Specifically, we test observation noise standard deviations from the set $\{0.1,\ 0.25,\ 0.5,\ 1.0\}$ and observation intervals from the set $\{1,\ 5,\ 10,\ 20\}$. The model noise covariance is fixed at $Q = 0.5$ throughout all trials.

Figure 9 illustrates a single example run of the Imperfect Model with EnKF state estimation over 10,000 timesteps. In this case, observations are assimilated every 10 timesteps, with additive Gaussian noise of standard deviation 0.1. The resulting NRMSE remains below the threshold $f = 0.4$ for the majority of the time series, demonstrating that the EnKF is able to effectively correct for imperfections in the model dynamics and maintain long-term stability. This contrasts sharply with the uncorrected imperfect model forecasts, which typically diverge after the valid time.

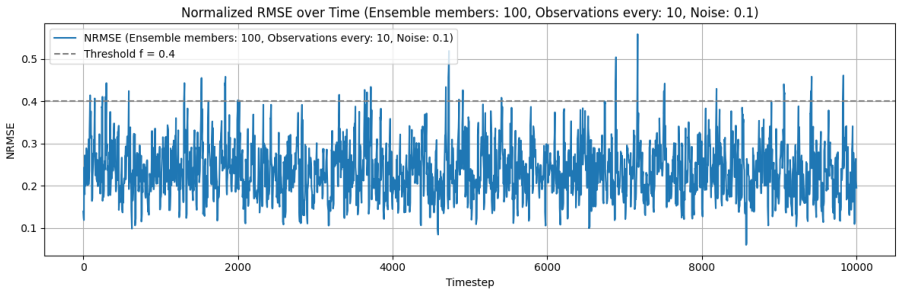

FIG. 9: NRMSE plot for imperfect model state estimation where we assimilate noisy data with the EnKF every 10 timesteps. As demonstrated, state estimation with our imperfect model keeps the NRMSE for the majority of datapoints below our threshold.

To further illustrate the effectiveness of data assimilation in this configuration (observation interval = 10, noise = 0.1), Figure 10 presents time series plots of the odd-indexed slow variables (X_1, X_3, X_5, X_7) over a 1000-timestep window. The black solid lines represent the true system trajectories, while the blue dotted lines show the corresponding estimates produced by the Imperfect Model corrected using the Ensemble Kalman Filter. Despite receiving observations only every 10 timesteps and with added noise, the corrected state estimates closely track the true dynamics.

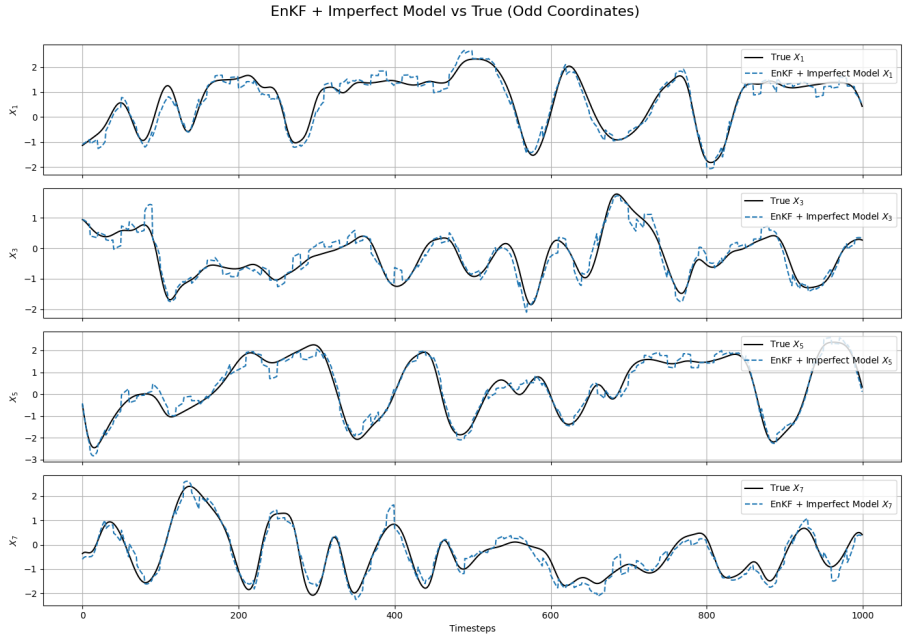

FIG. 10: Time series plots of the odd-indexed slow variables (X_1, X_3, X_5, X_7) over 1000 time steps. The true trajectory (black, solid line) is compared to the forecast from the imperfect model corrected using the EnKF (blue, dotted line). Data assimilation is applied every 10 time steps using noisy observations.

Figure 11 shows the average NRMSE over time for 10 different initial conditions using the Imperfect Model with EnKF state estimation. In this experiment, observations are assimilated every 10 timesteps with Gaussian noise of standard deviation 0.1. Each run spans 1000 timesteps, and the NRMSE is averaged across all runs to quantify the expected performance of the state estimation procedure. The shaded region represents one standard deviation across trials. The results demonstrate that, under this configuration, the expected NRMSE remains consistently below the threshold of $f = 0.4$ for the entire duration, highlighting the stability and effectiveness of EnKF when paired with a partially accurate model.

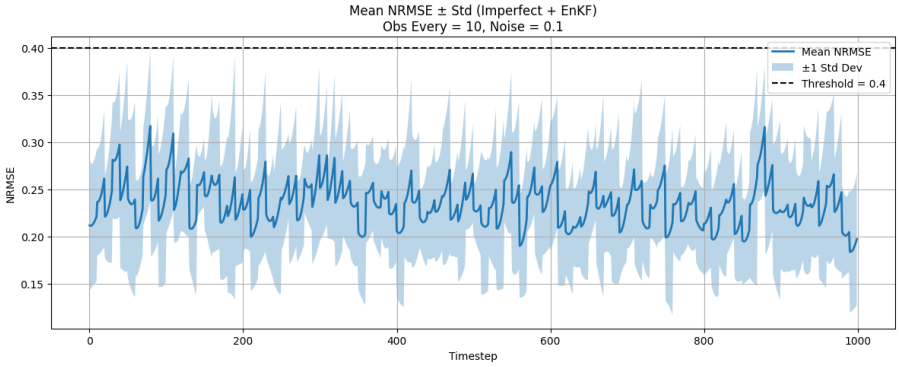

FIG. 11: Average NRMSE over time across 10 initial conditions using the Imperfect Model with EnKF. Observations are assimilated every 10 timesteps with noise level $\sigma = 0.1$. The mean NRMSE remains below the threshold $f = 0.4$ across all 1000 timesteps, demonstrating robust state estimation performance. Shaded band shows ± 1 standard deviation.

To summarize the effectiveness of the Imperfect Model with EnKF across varying assimilation settings, we perform state estimation over 10 different initial conditions for each combination of observation interval and noise level. For each run, we compute the percentage of timesteps where the NRMSE remains below the threshold $f = 0.4$ over a 1000-timestep forecast window. Figure 12 displays the average percentage of accurate timesteps for each configuration.

The results show that the EnKF provides robust correction across a wide range of observation settings. For the lowest noise level ($\sigma = 0.1$), the percentage of accurate timesteps remains above 90% across all observation intervals. Furthermore, when observations are frequent (e.g., every 1 or 5 steps), the EnKF maintains high accuracy even under substantial noise. Notably, even at the highest noise level ($\sigma = 1.0$), the accuracy remains above 60% when observations are assimilated every 1 or 5 timesteps. This demonstrates that even highly noisy data, when assimilated regularly, can significantly improve the performance of a partially informed model like the imperfect Lorenz 96 approximation.

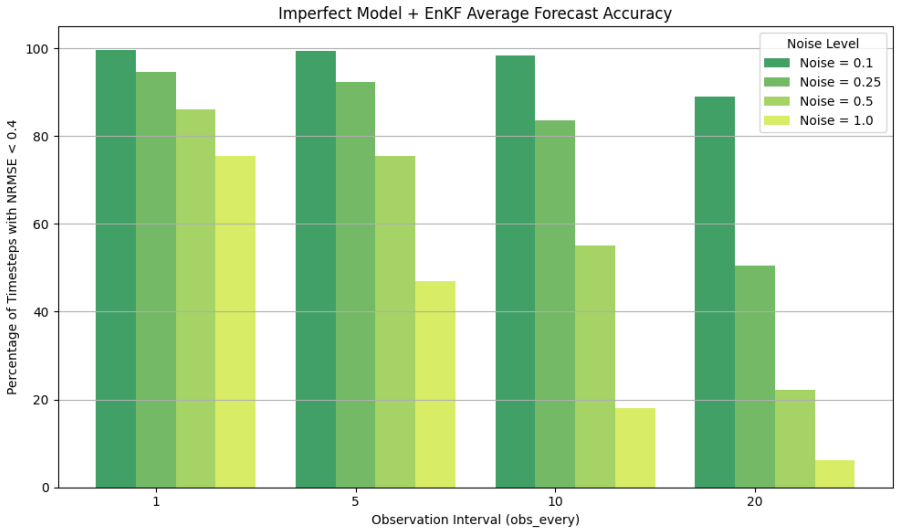

FIG. 12: Average percentage of time steps with NRMSE below 0.4 as a function of observation interval and noise level. The EnKF was run over 10 trials for each combination of assimilation interval $(1, 5, 10, 20)$ and noise level $(0.1, 0.25, 0.5, 1.0)$. Each bar shows the mean forecast accuracy under the error threshold.

E. ESN + EnKF State Estimation

We next conducted a series of state estimation experiments using the Echo State Network (ESN) in conjunction with the Ensemble Kalman Filter (EnKF). The experimental design mirrors that of the Imperfect Model + EnKF setup: we used 100 ensemble members, tested 10 distinct initial conditions, and explored a grid of observation intervals $1, 5, 10, 20$ paired with noise levels $0.1, 0.25, 0.5, 1.0$..

In this configuration, ESN predictions are passed through the EnKF at each timestep and then fed back as input to the reservoir, forming a feedback loop. This design enables assimilation to directly influence the ESN's internal state, ideally guiding its trajectory toward the true system dynamics.

Figure 13 shows the resulting forecast accuracy, defined as the percentage of timesteps with NRMSE below 0.4. As shown, when observations are assimilated at every timestep, the ESN + EnKF system performs comparably to the Imperfect Model + EnKF, even in the presence of moderate noise. However, accuracy declines sharply as the observation frequency drops. For observation intervals greater than 1, the forecast quality deteriorates substantially even when the entire state is available, suggesting poor performance under temporal sparsity.

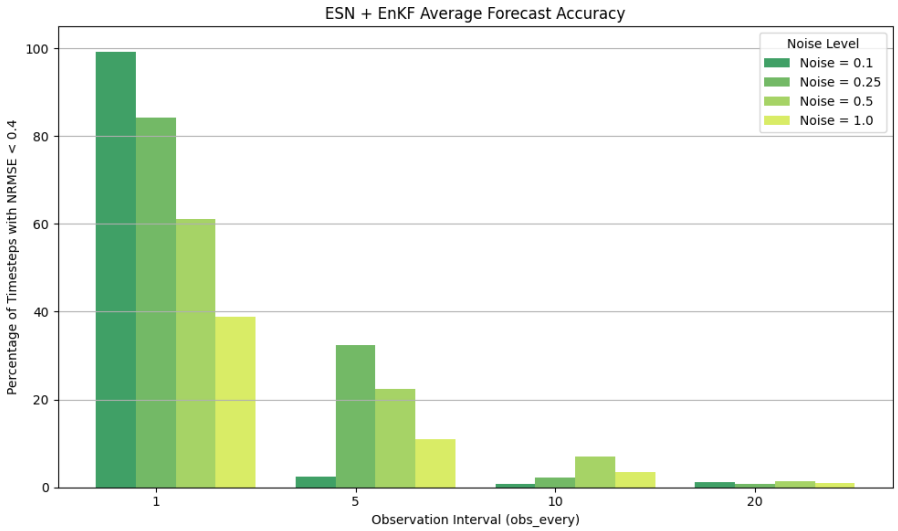

FIG. 13: Bar plot showing the percentage of time steps where the ESN + EnKF hybrid model achieves an NRMSE below 0.4. Each group of bars corresponds to a different observation interval $(1, 5, 10, 20)$, and each color represents a different observation noise level ($\sigma = 0.1, 0.25, 0.5, 1.0$). Results are averaged over 10 trials for each setting. Forecast accuracy clearly declines as observation noise increases and assimilation becomes less frequent.

Prior work has shown that the ESN + EnKF framework can be effective in settings with spatial sparsity, where only a subset of the system state is observed at each timestep [4]. For example, assimilating 50% of the state variables with low Gaussian noise at every timestep yielded strong results. However, their experiments do not address temporal sparsity, a scenario where the full state is observed but only at intermittent intervals.

Our results suggest that the ESN + EnKF performs well when data is assimilated frequently, but struggles when corrections are infrequent, despite full-state access. This divergence implies that while the EnKF can effectively correct ESN dynamics with regular spatial updates, it may be ill-suited for temporally sparse data assimilation where observational updates arrive at a slower rate than the ESN's internal dynamics.

We speculate that this fragility arises from the untrained and high-dimensional nature of the ESN reservoir. Unlike the imperfect model, whose physics-based evolution can be gently corrected by assimilation, the ESN's autonomous dynamics may drift rapidly between updates, making it difficult for the EnKF to exert meaningful corrective influence. In essence, once prediction begins, the ESN may evolve too freely for occasional corrections to remain effective.

These findings suggest that while the ESN + EnKF architecture shows promise when observational updates occur at the same frequency as the ESN's internal timestep, more robust assimilation strategies may be required for realistic applications where data arrives intermittently.

V. CONCLUSIONS

This study set out to evaluate the performance of knowledge-based, data-driven, and hybrid approaches to forecasting and state estimation in the multiscale Lorenz 96 system. Our results lead to several key conclusions and caveats. First, while the Echo State Network (ESN) performed poorly in standalone forecasting, its integration with the Ensemble Kalman Filter (EnKF) enabled competitive state estimation only when observations were assimilated at every timestep. This suggests that ESNs may still serve as useful surrogate models in scenarios with high-frequency, noisy, or spatially sparse observations but show limited capabilities under temporally sparse observations.

Second, our hybrid schemes, motivated by prior successes on the Kuramoto–Sivashinsky (KS) equation, consistently underperformed. Despite combining the imperfect model with an ESN reservoir, none of the hybrid configurations yielded forecasts that resembled the true system behavior. This failure is likely rooted in the more complex nature of the Lorenz 96 system: its multiscale coupling and multivariate dependencies introduce synchronization challenges that naive hybrid architectures cannot resolve.

These findings carry important implications. Hybrid models cannot be assumed to generalize across systems with different structures even when both systems are chaotic. The KS system is univariate and spatially chaotic, whereas Lorenz 96 features hierarchical coupling and faster internal scales. Simply transferring the same hybrid methodologies between them may lead to fragile dynamics, numerical instability, and forecast collapse.

Going forward, successful hybridization in multiscale systems may require more principled design. Possible directions may include learned synchronization modules, or dual-timescale reservoirs. Furthermore, incorporating memory components, could further help compensate for temporal sparsity. Finally, hybrid models may benefit from adaptive trust weighting between the physical and learned components similar to how an EnKF balances model forecasts and noisy observations, allowing the system to shift reliance between the knowledge-based and data-driven model.

In summary, while hybrid modeling remains an appealing strategy, its effectiveness

depends critically on the between system structure and model design. While our results affirm the performance of the individual components for forecasting and state estimation, the rapid degradation of the ESN+EnKF under temporal sparsity and the failure of our hybrid model to capture true dynamics show the need for further development of custom architectures when applying data-physics fusion to chaotic, multiscale systems like Lorenz 96.

Source code can be found at: `https://github.com/mcvlix/lorenz-96-hybrid-esn`

---

[1] Ashesh Chattopadhyay, Pedram Hassanzadeh, and Devika Subramanian. Data-driven predictions of a multiscale lorenz 96 chaotic system using machine-learning methods: reservoir computing, artificial neural network, and long short-term memory network. *Nonlinear Processes in Geophysics*, 27(3):373–389, July 2020.

[2] Ashesh Chattopadhyay, Adam Subel, and Pedram Hassanzadeh. Data-driven superparameterization using deep learning: Experimentation with multiscale lorenz 96 systems and transfer learning. *Journal of Advances in Modeling Earth Systems*, 12(11), November 2020.

[3] Geir Evensen. The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, 2003.

[4] Debdipta Goswami, Artur Wolek, and Derek A. Paley. Data-driven estimation using an echo-state neural network equipped with an ensemble kalman filter. In *2021 American Control Conference (ACC)*, pages 2549–2554, 2021.

[5] NASA/GSFC. Application of the ensemble kalman filter (enkf) to lorenz's 1963 model, 2010. Exploring Our Magnetic Earth.

[6] Jaideep Pathak, Alexander Wikner, Rebeckah Fussell, Sarthak Chandra, Brian R. Hunt, Michelle Girvan, and Edward Ott. Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(4), April 2018.