Deep Kalman filtering and generative diffusion models for chaotic dynamical systems

Haotian Hang, Jingyi Liu, and Morgan Jones

Department of Aerospace and Mechanical Engineering, University of Southern California, Los Angeles, CA

Motivation

Accurate prediction and parameter estimation of nonlinear dynamical systems remain essential challenges in modeling complex physical processes. In this project, we focus on the Lorenz-63 system [1] with fixed-parameter dynamics and evaluate several approaches—Kalman Filter [2], Deep Kalman Filter [3], and a score-based unconditional diffusion model [4,5]—for predicting future system states from time series data.

Dynamical system

The Lorenz-63 system is a canonical low-dimensional chaotic dynamical system defined by three coupled nonlinear ODEs [1]. Originally developed to model atmospheric convection, it captures key features of nonlinear systems, including deterministic chaos. The behavior of its solutions depends sensitively on the system parameters σ, ρ, and β. For standard values (σ=10, β=8/3), the system exhibits four distinct regimes:

$$\frac{dx}{dt} = \sigma(y - x),\quad \frac{dy}{dt} = x(\rho - z) - y,\quad \frac{dz}{dt} = xy - \beta z.$$

- **Stable origin** (ρ<1), where all trajectories decay to (0,0,0).
- **Two symmetric fixed points** (emerge for ρ>1), becoming stable for 1<ρ<ρc.
- **Hopf bifurcation** (ρc≈24.74), marking the onset of instability.
- **Chaotic attractor** (ρ>ρc), where trajectories are aperiodic and highly sensitive to initial conditions.

In our work, we focus on the chaotic regime (ρ=28, σ=10, β=8/3), where traditional prediction methods struggle due to the system's inherent sensitivity and long-term unpredictability.

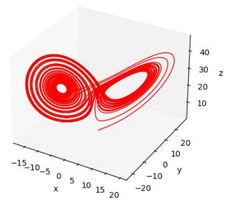

Figure 1. Trajectory for Lorenz-63 system with parameter set ρ=28, σ=10, β=8/3.

Deep Kalman filter

The Deep Kalman Filter (DKF) extends classical Kalman filtering by modeling nonlinear dynamics and observations using deep neural networks [2,3]. It combines generative modeling with variational inference to estimate hidden states xt from observations yt, even when the system dynamics and observation models are highly nonlinear. The process is:

$$x_t \sim p_\theta(x_t \mid x_{t-1}), \quad y_t \sim p_\theta(y_t \mid x_t)$$

where xt is the latent system state at time t, and yt is the corresponding observation. The transition and observation models pθ(·) are parameterized by neural networks. Inference is performed via a variational distribution:

$$q_\phi(x_t \mid y_{\leq t})$$

which approximates the true posterior over latent states. This framework enables learning complex temporal dependencies and model uncertainty beyond what traditional filters can capture.

In Fig. 2 , we showed that, with accurate physical model, extended Kalman filter and particle filter can accurately capture the time evolution of the Lorenzo system. Deep Kalman filter, on the other hand, learns the physical model from noisy data, can capture the strange attractor in the system, while cannot accurately predict the time evolution of the system.

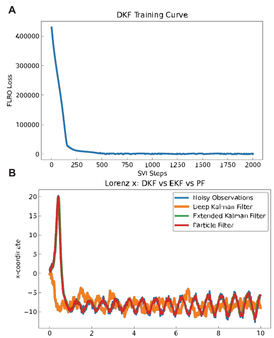

Figure 2. a) Learning curve of Deep Kalman filter. b) Comparison of prediction results via Extended Kalman filter, particle filter and Deep Kalman filter..

Diffusion model

Diffusion models aim to generate data samples directly from pure noise [4,5]. These models can learn to denoise data with or without using measurements to infer the solution. In case, we consider an unconditional diffusion model which only uses lorenz data samples to generate more lorenz trajectories. We also consider the case where the state trajectories are noisy measurements, that is:

$$x, y, z \sim \mathcal{N}(0, \sigma_s)$$

Where the hyperparameter σ_s controls the scale of the noise in the trajectories. The noisy dynamical system for relatively small σ_s maintain the dynamics of the ideal system. We use a score-based diffusion model, which aims to learn the gradient of the probability density of the data's distribution p_x:

$$s(x_i, t_i) = \nabla \ln\, p_x(x_i, t_i)$$

Where x_i are the samples, t_i is the Langevin time, and s is considered the score function. Once the score function is learned, samples can be integrated using the Euler-Maruyama equation:

$$x_i^{n+1} = x_i^n + \tfrac{1}{2}\Delta\tau \cdot s(x_i^n, t_i^n) + \sqrt{\Delta\tau} \cdot w$$

Where w is Gaussian noise and $\Delta\tau$ is the langevin time step. A multi-layer perceptron was used to approximate the score function. As shown in Figure 4, the score function is able to approximate the general shape of the Lorenz dynamics from sampled noise. However, the model does not recover the exact trajectory of the Lorenz system, which is expected—since diffusion modeling is designed to learn the underlying data distribution rather than reproduce deterministic paths.

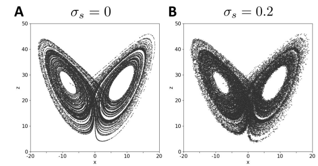

Figure 3. Samples of (A) the idea Lorenz63 system (x-z dimensions) and (B) the noisy system

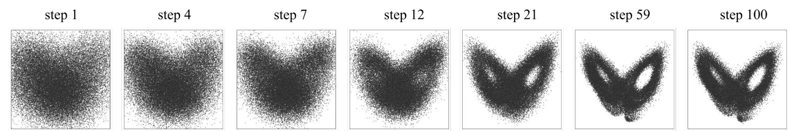

Figure 4. Inverse diffusion Langevin steps of random samples converging to the Lorenz model.

Table 1. Summary of general model performance

	Kalman Filter	Deep Kalman Filter	Unconditional Diffusion Model	Ideal Model
Predictive Speed	✓	✓	✗	✓
Generalization	✗	✓	✓	✓
Physics informed	✓	✗	✗	✓

Conclusion and Future Work

Both the deep Kalman Filter and Diffusion model can capture the general distribution of the Lorenz63 model, however the presented models do not rigorously capture the trajectory dynamics. The Particle Filter in comparison effectively captures these dynamics because explicit knowledge of the system is given to the model to estimate the states. Additional future avenues on diffusion models can include predicting Lorenz parameters given the state dynamics, which may be a similar approach to SINDY.

References Cited

[1] Lorenz, EN40214341417. "E. 1963. Deterministic nonperiodic flow." J. Atmos. Sci 20.1 (1963): 30-41.
[2] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems.
[3] Chinellato, E., & Marcuzzi, F. (2025). State, parameters and hidden dynamics estimation with the Deep Kalman Filter: Regularization strategies. Journal of Computational Science, 87, 102569..
[4] Song, Y., & Ermon, S. (2019). Generative modeling by estimating gradients of the data distribution. Advances in neural information processing systems, 32.
[5] Dasgupta, A., da Cunha, A. M., Fardisi, A., Aminy, M., Binder, B., Shaddy, B., & Oberai, A. A. (2025). Unifying and extending diffusion models through pdes for solving inverse problems. arXiv preprint arXiv:2504.07437.