Predicting Latent States of Dynamical Systems With State-Space Reconstruction and Gaussian Processes

Kurt Butler,† Guanchao Feng,† Charles B. Mikell,* Sima Mofakham,* and Petar M. Djurić,†
†*Department of Electrical and Computer Engineering*
**Department of Neurosurgery*
Stony Brook University
Stony Brook, NY 11794, USA

***Abstract*—Predicting future observations of a system is a classical task in signal processing. However the effects of nonlinear dynamics, unobserved variables and observation noise make this task difficult in practice. We propose a data-driven non-parametric approach to model systems with latent dynamics using state-space reconstruction and Gaussian processes. With this approach, both latent states and future observations can be predicted together. When applicable, this method is efficient even with short time series. We demonstrate the method on synthetic data and then showcase its efficacy and accuracy in predicting brain dynamics on a data set obtained from traumatic brain injury patients.**

***Index Terms*—state space reconstruction, nonlinear dynamics, Gaussian processes, traumatic brain injury, attractors**

I. Introduction

Physical systems are often described by systems of interacting states or modes. In many systems, these interactions are deterministic and so the evolution of the system in time is not random. Even in the absence of stochastic effects and external influences on the system, deterministic systems can exhibit complicated phenomena such as bifurcations, chaos, and phase transitions [1], [5]. These phenomena can complicate modeling, but even worse is that in many applications not all state variables of a system are observable. In this case, latent variable models are required to model the dynamics. Latent variable models generally invoke complicated tools for inference such as variational approximations which may be costly to train [3], [17]. Thus, there remains an appeal in low complexity approaches to modeling latent state spaces.

Possibly the simplest approach to latent state modeling is provided by state space reconstruction (SSR) [4], which is sometimes called attractor reconstruction in the nonlinear dynamics literature [11], [21] or delay embedding within the neuroscience community [12], [22]. SSR allows an experimenter to construct a portrait of the latent state space that generated an observed signal. If the system contains an attractor, a distinguished geometric structure in the state space that attracts nearby trajectories, then SSR will reconstruct the attractor in the latent space. The conditions for this reconstruction were first given by Floris Takens when the attractor is a compact smooth manifold [21], but there have been several generalizations of Takens' work to systems with strange attractors and fractal structures [4], [16]. Although not all systems have attractors, the presence of an attractor can be exploited for many difficult scientific tasks, including the analysis of neurological activity [2] and causal discovery in ecological systems [20].

States in an attractor remain inside the attractor for all future times [19], and so one may use the reconstructed attractor as a proxy for the latent state space [18]. To learn the dynamics on the reconstructed attractor, we propose to use Gaussian process regression (GPR). Existing Gaussian process (GP) approaches to modeling latent dynamics rely on learning a mapping into the latent space [3], [8], [10], [15], [17], [24], but SSR provides easy to implement and compelling alternative to other latent variable models.

Our contribution in this work is a non-parametric and data-driven approach to predicting latent states of nonlinear systems by employing GPR to model the dynamics on the reconstructed latent states. Under this approach, we show that predicting latent states and future observations can be done concurrently. We argue that GPR has numerous advantages, especially for signals with short or noisy observation windows.

II. Problem formulation

The typical discrete-time state-space model consists of a vector-valued latent state $\mathbf{x}_n$, an observation signal y_n and mathematical relationships that describe how they evolve in time. Abstractly, we may write

$$\mathbf{x}_{n+1} = F(\mathbf{x}_n), \tag{1}$$

$$y_n = h(\mathbf{x}_n), \tag{2}$$

where $\mathbf{x}_n \in \mathbb{R}^d$, $y_n \in \mathbb{R}$, and F and h are smooth functions. The traditional state-space model includes additive noise in the equations, but we exclude the noise while we are developing the theory for our approach.

An attractor is a subset $A \subset \mathbb{R}^d$ of the state space that will attract and trap nearby states [19]. Given an observation signal y_n whose corresponding latent state vectors $\mathbf{x}_n$ reside in an attractor A, SSR can reconstruct the latent attractor by concatenating delayed observations as follows.

We first fix positive integer parameters Q and τ, which are called the *embedding dimension* and *embedding delay*,

The authors thank the support of NSF under Award 2021002, and the U.S. Department of Education under the GAANN Fellowship.

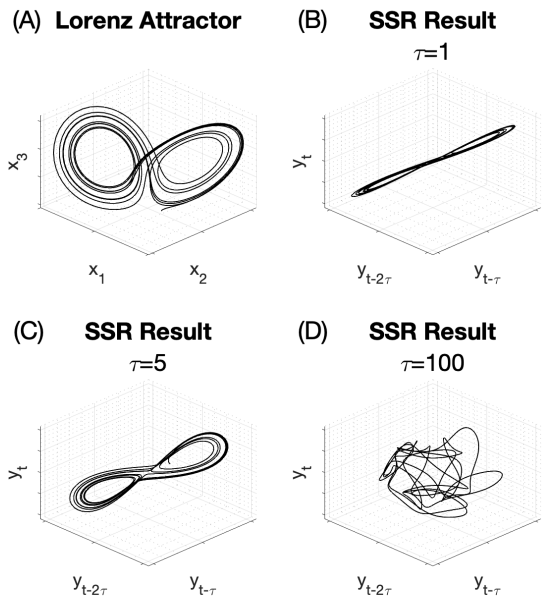

Fig. 1. Three dimensional SSR results for a Lorenz system, which is defined in (9). **(A)** The Lorenz attractor in the latent state space. **(B)** Shadow manifold when $\tau = 1$. Since adjacent samples in time are highly correlated, the shadow manifold is compressed to a line. **(C)** Shadow manifold when $\tau = 5$. The shape of the Lorenz attractor can be recognized as a nonlinear transformation of the original in (A). **(D)** Shadow manifold when $\tau = 100$. The Lorenz system is chaotic and large τ results in very complicated embeddings.

respectively. We then define the Q-dimensional SSR vector by

$$\mathbf{m}_n^y = \begin{bmatrix} y_{n-(Q-1)\tau} & \cdots & y_{n-\tau} & y_n \end{bmatrix} \tag{3}$$

for any $n \geq 1 + (Q-1)\tau$. When we choose Q to be large enough, each SSR vector $\mathbf{m}_n^y$ represents a point on the reconstructed attractor. If we choose Q to satisfy $Q > 2\dim(A)$, then reconstruction is mathematically guaranteed [16], although for many systems smaller Q will suffice. The parameter τ is used to tune the quality of the reconstruction by controlling the correlation between coordinates in the SSR vector. The set of all points $\mathbf{m}_n^y$ is called the *shadow manifold*, denoted by $\mathcal{M}_y$. In Fig. 1, we visualize several three-dimensional reconstructions of a Lorenz attractor to show how the parameter τ influences the result.

The mathematical justification for SSR is given by Takens' theorem [21], which asserts that the mapping in (3) provides a diffeomorphism[1] between the latent attractor A and the shadow manifold $\mathcal{M}_y$ for most observation functions h in (2). A probabilistic version of Takens' theorem says that if we randomly choose the observation function, then we reconstruct the attractor with probability 1 [16]. The practical result here is that there is a diffeomorphism D such that $D(\mathbf{x}_n) = \mathbf{m}_n^y$ for all $\mathbf{x}_n$ inside the attractor.

[1]A diffeomorphism is a differentiable transformation with a differentiable inverse transformation.

SSR will reconstruct the basic topological and differential properties of an attractor, but it also permits us to model the latent dynamics as trajectories on the shadow manifold. Using the diffeomorphism D and the function F from (1), we may define a function $\tilde{F}$ on $\mathcal{M}_y$ by the rule $\tilde{F}(\mathbf{m}) = D(F(D^{-1}(\mathbf{m})))$. This map reconstructs the latent dynamics on the shadow manifold, i.e.

$$\tilde{F}(\mathbf{m}_n^y) = D(F(\mathbf{x}_n)) = D(\mathbf{x}_{n+1}) = \mathbf{m}_{n+1}^y. \tag{4}$$

Since the last coordinate of $\mathbf{m}_n^y$ is y_n, learning the function $\tilde{F}$ implies we can forecast the observation signal.

In general, to model $\tilde{F}$ we would need to model a vector-valued function. However, if we instead model the τ-th iterate of the function $\tilde{F}$,

$$\tilde{F}^\tau(\mathbf{m}_n^y) = \underbrace{\tilde{F}(\cdots \tilde{F}(\mathbf{m}_n^y))}_{\tau \text{ times}} = \mathbf{m}_{n+\tau}^y, \tag{5}$$

then we only need to learn to predict $y_{n+\tau}$,

$$\begin{aligned} \tilde{F}^\tau(\mathbf{m}_n^y) &= \tilde{F}^\tau\left(\begin{bmatrix} y_{n-(Q-1)\tau} & \cdots & y_{n-\tau} & y_n \end{bmatrix}\right) \\ &= \begin{bmatrix} y_{n-(Q-2)\tau} & \cdots & y_{n+\tau} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{m}_n^y \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} & y_{n+\tau} \end{bmatrix}, \end{aligned} \tag{6}$$

where the second equality follows from (5). Thus, if we learn a function f such that

$$y_{n+\tau} = f(\mathbf{m}_n^y), \tag{7}$$

then we can reconstruct the latent dynamics.

By learning the function $\tilde{F}$, we can both predict the observation signal as well as predict latent states. If it is acceptable to downsample by τ to do prediction, then the modeling problem becomes computationally inexpensive. Although $\mathbf{m}_n^y$ and $\mathbf{x}_n$ can be numerically quite different quantities, for practical purposes $\mathbf{m}_n^y$ is a useful proxy for the latent state.

Since Takens' theorem assumes a noiseless system, we omitted noise from the state update (1) and the observation equation (2) while developing the theory. There exist stochastic generalizations of Takens' theorem [18], but they are technical and beyond the scope of this work. However, in the stochastic case one may obtain a probabilistic version of (7) [18].

III. Proposed solution

A. Gaussian process regression

To model the dynamics of the latent state, we propose to use Gaussian process regression (GPR) [14] to fit the function f in (7). Abstractly, GPR fits a function f to a set of inputs $\mathbf{x}_n$ and outputs y_n such that $y_n = f(\mathbf{x}_n)$. GPR supposes a Gaussian process prior over the space of functions, which means that for any set of input points $\mathbf{x}_1, ..., \mathbf{x}_n, \mathbf{x}_*$, the corresponding output points have a Gaussian distribution:

$$(y_1, ..., y_n, y_*) \sim \mathcal{N}(\mathbf{0}, \mathbf{K}),$$

where the covariance $\mathbf{K}_{ij} = k_\theta(\mathbf{x}_i, \mathbf{x}_j)$ is specified by a kernel function k_θ with parameter vector θ. For a point $\mathbf{x}_*$ which we want to know $f(\mathbf{x}_*)$, we simply compute the conditional expected value

$$\hat{f}(\mathbf{x}_*) = \mathbb{E}(y_*|y_1, ...y_n) = \mathbf{k}_*^\top \mathbf{K}^{-1}\mathbf{y},$$

where $\mathbf{k}_*$ and $\mathbf{y}$ are vectors in $\mathbb{R}^n$ defined by $(\mathbf{k}_*)_i = k_\theta(\mathbf{x}_*, \mathbf{x}_i)$ and $\mathbf{y}_i = y_i$. Training the GPR predictor $\hat{f}$ usually means we optimize the value of θ as to maximize the likelihood of the observed data set. GPR is non-parametric, robust to noise and efficient with small sample sizes [14].

GPR can also provide a probabilistic prediction by considering $p(y_*|y_1, ..., y_n)$, which is a Gaussian distribution, and from this we may compute a 95% confidence interval around any estimate. This confidence interval may model our uncertainty due to perturbations in the latent state. Accounting for observation noise could be accomplished by incorporating denoising into SSR, as in [16], but we do not perform denoising in this paper.

B. Latent state prediction

We will assume in this section that there exists a latent attractor which may be reconstructed by SSR. Checking this assumption is nontrivial [21], and the problem of detecting attractors is left for future work.

Before training the GPR predictor, one must also learn the parameters Q and τ used in SSR. There are several possible options, but we provide a heuristic approach. We first select τ by use of an autocorrelation function where τ is taken to be the first lag in which the autocorrelation function drops below 0.5. Next, we select the embedding dimension Q using the false-nearest-neighbors approach of [9]. A false neighborhood is a pair of points $\mathbf{m}_s^y, \mathbf{m}_t^y$ that are nearest neighbors in space but not time; i.e., $\mathbf{m}_s^y$ is the nearest neighbor to $\mathbf{m}_t^y$ in our recorded data set, but $s \neq t \pm 1$. In false nearest neighborhoods, we iteratively increase Q until the number of false neighborhoods is below 5%. With these considerations in mind, we outline a procedure to perform τ-step latent state prediction in Algorithm 1.

Algorithm 1 τ-step Latent State Prediction

Input: Observation signal $y_1, ..., y_N$
Normalize the signal y_n
Compute the autocorrelation function $r_{yy}(k)$ of y
$\tau = \min\{k > 0; r_{yy}(k) < 0.5\}$
Select Q using false-nearest-neighbors
$\mathbf{M}_{rc} = y_{r+c\tau}$ for $r = 1, ..., N - Q\tau$ and $c = 1, ..., Q - 1$
$\mathbf{Y}_r = y_{r+Q\tau}$ for $r = 1, ..., N - Q\tau$
Train a GPR to predict each $\mathbf{Y}_r$ given $\mathbf{M}_{r:}$
for each $y_{t+\tau}$ to be predicted, **do**
    $\mathbf{m}_t^y := \begin{bmatrix} y_{t-(Q-1)\tau} & \cdots & y_t \end{bmatrix}$
    $\hat{y}_{t+\tau} := \text{GPR}(\mathbf{m}_t^y)$
    $\hat{\mathbf{m}}_{t+\tau}^y := \begin{bmatrix} y_{t-(Q-2)\tau} & \cdots & y_t & \hat{y}_{t+\tau} \end{bmatrix}$
end for

IV. Results

We considered a number of explicit examples to test our approach. In each example, the GPR model used a Gaussian kernel function with a linear basis function [14]. We compared our method to other SSR-based predictors of the observation signal from ecology and finance. The first method used simplex projection to make predictions [23]. The second method was the Chaos-SVM predictor [7]. We also considered a nonlinear autoregressive model (NARX) produced by a neural network. The NARX model used the same input dimension as the SSR models, but had unit delays. We considered other neural predictors, including recurrent and long short-term memory networks (LSTM) [6], but their performance was generally similar to the NARX model. For each system, we normalized the signals before processing. We defined the signal-to-noise ratio (SNR) as $20\log_{10}(P_s/P_n)$ where P_s and P_n are the signal and noise powers, respectively.

A. Oscillator system

If a signal is a superposition of K sinusoids with random frequencies, then its latent attractor is a K-dimensional torus [5]. One way to produce such a signal is to observe multiple oscillating latent systems. If $\mathbf{x}_t \in \mathbb{R}^2$, then the differential equation

$$\dot{\mathbf{x}}_t = \begin{bmatrix} 0 & -f \\ f & 0 \end{bmatrix} \mathbf{x}_t$$

has the solution $\mathbf{x}_t = [\cos(ft), \sin(ft)]^\top$. Therefore, sinusoidal signals may be interpreted as state trajectories in some state space. If we now observe a signal that is a function of multiple sinusoids, we may interpret it as an observation of a state-space composed of oscillators. Let us observe

$$y_t = p\left(2\sin(f_1 t) + \sin(f_2 t) + \sin(f_3 t)\right) + w_t, \qquad (8)$$

where $p(x) = -0.01x^4 + 0.3x^2 - x$ is a polynomial used to add nonlinearity. Since the signal observes three oscillators, the system's attractor is three-dimensional and we need Q=7 dimensions or less to embed it. White Gaussian noise (WGN) w_t is added to $p(\cdot)$ such that the SNR is 20 dB.

In the simulations, we set f_1, f_2, f_3 to $\sqrt{1/30}, 1/3, \sqrt{5}/3$. Each predictor is trained on the initial 100 samples, and then sequentially validated on an additional 400 samples. The results are shown in Fig. 2.

B. Lorenz system

We observe a Lorenz system [19] through the x_1-coordinate

$$\begin{aligned} \dot{x_1} &= a(x_2 - x_1) \\ \dot{x_2} &= x_1(c - x_3) - x_1 \\ \dot{x_3} &= x_1 x_2 - b x_3 \\ y_t &= x_{1,t} + w_t \end{aligned} \qquad (9)$$

where $(a, b, c) = (10, 8/3, 28)$. We simulated the model using a Runge-Kutta method and observed the process with sampling period 0.05. We add WGN w_t to the observation such that the SNR is 20 dB. We use the initial 250 samples for training, and we performed a 7-step prediction on an additional 900 samples, shown in Fig. 3.

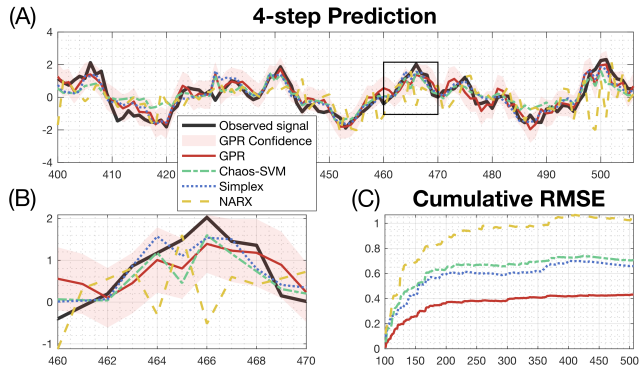

Fig. 2. Four-step prediction result for the oscillator system. Our procedure chose $\tau = 4$ and $Q = 6$ for SSR. **(A)** Some of the predictions made by each model. We also show the 95% confidence interval of the GPR predictions. **(B)** We zoom in on the boxed region in (A) to see the predictions more clearly. **(C)** Cumulative root mean squared error (CRMSE) for each point predictor, which is defined to be $\text{CRMSE}(T) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}$. The limit of CRMSE as $T \to \infty$ is the root-mean-squared error (RMSE) of each predictor, which is shown in Table 1.

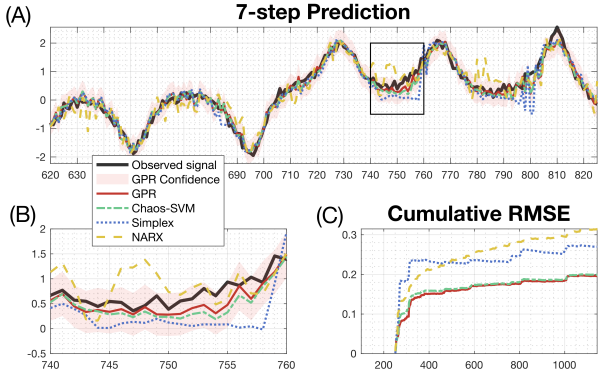

Fig. 3. Seven-step prediction results for the Lorenz system. Our method selected $\tau = 7$ and $Q = 3$ for SSR. **(A)** A portion of the observation window showing predictions and the observed signal, similar to Fig. 2. **(B)** We zoom in on the boxed region in (A) to see the predictions more clearly. **(C)** CRMSE plots for each predictor.

C. Brain Electrophysiological Signal

We applied the predictors to electroencephalography (EEG) signals recorded from comatose traumatic brain injury patients at Stony Brook University Hospital. The EEG signals were recorded from the scalp with a 18 contact standard 10-20 system and a sampling frequency of 256 Hz. We identified a minimum of forty minutes of resting data for each patient with minimal sedation and artifact. Before the analysis, we preprocessed the EEG recordings by screening for artifacts, bipolar re-referencing, mean subtraction, bandpass filtering (0.5-30 Hz), down-sampling, and detrending using a Savitzky-Golay filter [13].

In Fig. 4, we show the results of a 10-step prediction using each of the four predictors. We successfully reconstructed the underlying attractor supporting the bulk of the temporal evolution of these brain dynamics, evident from the accurate

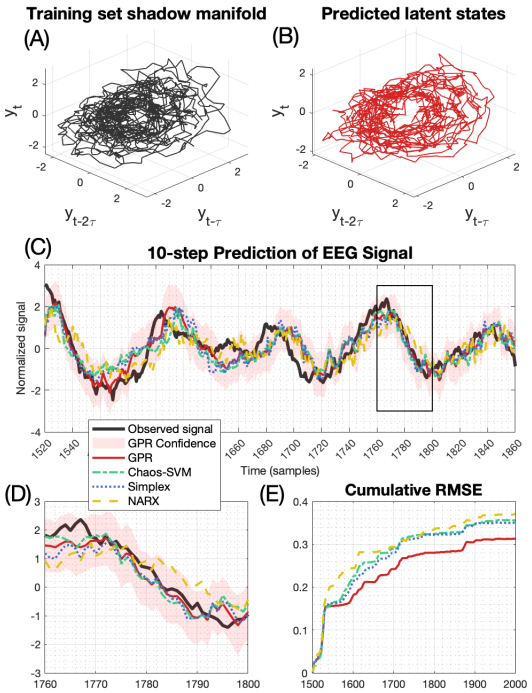

Fig. 4. Prediction result for the EEG signal. **(A)** Shadow manifold of the training data set. Our method selected $\tau = 10$ and $Q = 5$ for SSR. **(B)** Latent trajectory predicted by the proposed method. The shadow manifold produced by the predicted latent states resembles the manifold learned during training. **(C)** Out-of-sample predictions of the EEG signal. We show each of the four predictors under study as well as the GPR 95% confidence interval. The width of a grid square represents the 10-step prediction interval. The plot is zoomed to examine one portion of the predictions. **(D)** We zoom in on the boxed region in (C) to see the predictions more clearly. **(E)** CRMSE plots for each predictor.

prediction of EEG values compared to the observations.

V. Discussion

In each system, every estimator made reasonable predictions most of the time. For synthetic data, the NARX model typically performed worse than the other predictors, indicating that SSR, viewed as feature selection, can be valuable to improve prediction accuracy. This intuitively makes sense, since the NARX model implicitly performs SSR with $\tau = 1$, and optimizing τ allows us to better visualize the latent attractor (Fig. 1). We anecdotally found that these results were consistent as we decreased the observation SNR, but we did not include a figure for this.

We may also highlight the connection to autoregressive models here. Our results suggest that a nonlinear autoregressive model can implicitly perform SSR to reconstruct latent structure, and GPR provides a competitive and flexible tool for modeling the latent dynamics of the reconstructed latent states. The techniques that we used

for selecting SSR parameters as well as our selected GPR model were all somewhat standard, but this indicates that an awareness of the presence of a latent attractor may be exploited for making more accurate models without invoking complicated techniques.

Model	Oscillator	Lorenz	EEG
GPR	0.481	0.221	0.616
Chaos-SVM	0.787	0.225	0.712
Simplex	0.734	0.304	0.697
Nonlinear ARX	1.174	0.356	0.729

TABLE I
FINAL ROOT-MEAN-SQUARED ERROR (RMSE) FOR EACH MODEL AND SYSTEM. FOR EACH OF THE THREE SYSTEMS, THE GPR PREDICTORS YIELDED THE LOWEST PREDICTION ERROR.

Analysis of the EEG signal result is more challenging because we do not know the ground truth. Namely, we do not know if there is truly a latent attractor. In viewing the SSR result we can visually suspect that there is structure that our predictor may exploit. In Fig. 4, we show the latent states both from the predictions as well the training data. Since the GPR predictor can accurately model the dynamics of the observed EEG signal, we decide that the learned shadow manifold is a decent proxy for the latent system. While this provides some evidence for the existence of a low dimensional attractor in the neural state-space, additional analyses are required to make any statements of scientific value. Additionally, not all features of the signals could be predicted by the method, which indicates that the latent attractor is not the only information being recorded by the observation signal. A full discussion of the dynamical and neurological significance of our result is beyond the scope of this paper.

VI. CONCLUSIONS

In this work we detailed the SSR approach to reconstructing latent states, and we showed that GPR is a competitive and effective tool for building a dynamical model on the reconstructed state-space. The effectiveness of this approach relies on the nontrivial assumption that there are attractors in the latent state-space. We will discuss the detection of attractors in future work.

REFERENCES

[1] P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality. *Physical Review A*, 38(1):364, 1988.

[2] A. M. Bruno, W. N. Frost, and M. D. Humphries. A spiral attractor network drives rhythmic locomotion. *eLife*, 6:e27342, 2017.

[3] A. Damianou, M. Titsias, and N. Lawrence. Variational Gaussian Process Dynamical Systems. *Advances in Neural Information Processing Systems*, 24, 2011.

[4] E. R. Deyle and G. Sugihara. Generalized Theorems for Nonlinear State Space Reconstruction. *PLOS One*, 6(3):e18295, 2011.

[5] J.-P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *The Theory of Chaotic Attractors*, pages 273–312, 1985.

[6] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[7] S.-C. Huang, P.-J. Chuang, C.-F. Wu, and H.-J. Lai. Chaos-based support vector regressions for exchange rate forecasting. *Expert Systems with Applications*, 37(12):8590–8598, 2010.

[8] M. Karl, M. Soelch, J. Bayer, and P. Van der Smagt. Deep Variational Bayes Filters: Unsupervised learning of state space models from raw data. *arXiv preprint arXiv:1605.06432*, 2016.

[9] M. B. Kennel and H. D. Abarbanel. False neighbors and false strands: A reliable minimum embedding dimension algorithm. *Physical Review E*, 66(2):026209, 2002.

[10] N. Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. *Advances in Neural Information Processing systems*, 16, 2003.

[11] Z. Lu, B. R. Hunt, and E. Ott. Attractor reconstruction by machine learning. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(6):061104, 2018.

[12] N. Marrouch, J. Slawinska, D. Giannakis, and H. L. Read. Data-driven Koopman operator approach for computational neuroscience. *Annals of Mathematics and Artificial Intelligence*, 88(11):1155–1173, 2020.

[13] S. Mofakham, Y. Liu, A. Hensley, J. R. Saadon, T. Gammel, M. E. Cosgrove, J. Adachi, S. Mohammad, C. Huang, P. M. Djurić, et al. Injury to thalamocortical projections following traumatic brain injury results in attractor dynamics for cortical networks. *Progress in Neurobiology*, page 102215, 2022.

[14] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[15] V. Rutten, A. Bernacchia, M. Sahani, and G. Hennequin. Non-reversible Gaussian processes for identifying latent dynamical structure in neural data. *Advances in Neural Information Processing Systems*, 33:9622–9632, 2020.

[16] T. Sauer, J. A. Yorke, and M. Casdagli. Embedology. *Journal of Statistical Physics*, 65(3):579–616, 1991.

[17] Q. She and A. Wu. Neural Dynamics Discovery via Gaussian Process Recurrent Neural Networks. In *Uncertainty in Artificial Intelligence*, pages 454–464. PMLR, 2020.

[18] J. Stark, D. S. Broomhead, M. E. Davies, and J. Huke. Delay Embeddings for Forced Systems. II. Stochastic Forcing. *Journal of Nonlinear Science*, 13(6):519–577, 2003.

[19] S. H. Strogatz. *Nonlinear Dynamics and Chaos: With applications to Physics, Biology, Chemistry, and Engineering*. CRC Press, 2018.

[20] G. Sugihara, R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty, and S. Munch. Detecting Causality in Complex Ecosystems. *Science*, 338(6106):496–500, 2012.

[21] F. Takens. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence*, pages 366–381. Springer, 1981.

[22] R. Vicente, M. Wibral, M. Lindner, and G. Pipa. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *Journal of Computational Neuroscience*, 30(1):45–67, 2011.

[23] H. Ye and G. Sugihara. Information leverage in interconnected ecosystems: Overcoming the curse of dimensionality. *Science*, 353(6302):922–925, 2016.

[24] B. M. Yu, J. P. Cunningham, G. Santhanam, S. Ryu, K. V. Shenoy, and M. Sahani. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Advances in Neural Information Processing Systems*, 21, 2008.