A Kuramoto Model Approach to Predicting Chaotic Systems with Echo State Networks

Sophie Wu,[1,2,3] Jackson Howe,[1,2,3] and Lyle Muller (supervisor)[1,2,3]

[1] *Department of Mathematics, Western University*
[2] *Western Institute for Neuroscience*
[3] *Western Academy for Advanced Research*

An Echo State Network (ESN) with an activation function based on the Kuramoto model (Kuramoto ESN) is implemented, which can successfully predict the logistic map for a non-trivial number of time steps. The reservoir in the prediction stage exhibits binary dynamics when a good prediction is made, but the oscillators in the reservoir display a larger variability in states as the ESN's prediction becomes worse. Analytical approaches to quantify how the Kuramoto ESN's dynamics relate to its prediction are explored, as well as how the dynamics of the Kuramoto ESN relate to another widely studied physical model, the Ising model.

I. INTRODUCTION

Applications of machine learning in applied mathematics and physics are widespread and successful. Using machine learning to approximate functions in high dimensions has opened up many new avenues for investigating nonlinear dynamics in the past three decades. Additionally, the practice of predicting and analyzing chaotic systems with neural networks offers the opportunity to learn more about both the dynamics of the neural network as well as the dynamics of the chosen chaotic system [1].

Why predict the logistic map with a Kuramoto ESN? While the complex dynamics of Echo State Networks, the Kuramoto model, and the logistic map have been investigated extensively in separate pieces of literature, this project broadly aims to combine previously disconnected analytical approaches to all three of these systems. While insights are gathered from a variety of angles, we are especially interested in analyzing the dynamics of our Kuramoto ESN as it learns the logistic map, and we emphasize our implementation of the Kuramoto model in the ESN as a novel approach to analyzing the properties of the logistic map.

II. METHODS

A. Echo State Networks

Recurrent Neural Networks (RNNS), which allow for recurrent connections between their artificial neuron layers, have a unique ability to learn high-dimensional dynamical systems. In some cases this learning ability is much better than that of their feed-forward counterparts, which feature single-layer connections between their artificial neurons (without recurrence) [2]. However, feed-forward neural networks have been used predominantly for the last 25 years in machine learning research. This is largely because recurrent connections within RNNs produce nonlinear dynamics, and as such, training these networks is significantly more challenging than simple back propagation [3].

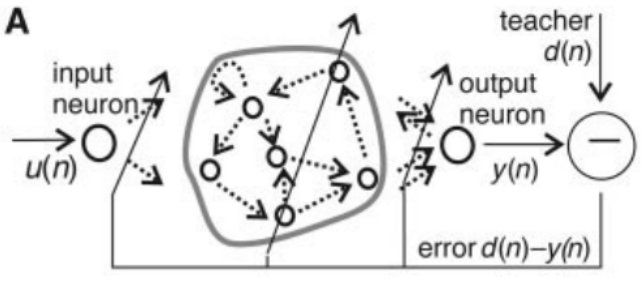


Figure 1: Echo State Network

The neural network discussed in this report—the Echo State Network—is a type of RNN. ESNs fall under a paradigm known as reservoir computing (RC). An ESN (Figure 1) is composed of an input layer, a middle layer (the reservoir), and an output layer. The reservoir is usually sparsely connected, and the majority of the reservoir weights

are unchanged from their initial conditions throughout the training procedure. Only the weights connected to the output layer are modified in training, allowing ESNs to circumvent many of the traditional difficulties in training RNNs [4]. The universal function approximation property possessed by ESNs allows them to produce high-dimensional non-linear dynamics and predict chaotic time series, such as the logistic map, remarkably well [5].

A classic ESN with input data dimension L, network size N, and output size P has an input matrix $\boldsymbol{U} \in M_{N\text{x}L}(\mathbb{R})$, a reservoir matrix $\boldsymbol{W} \in M_{N\text{x}N}(\mathbb{R})$, and an output matrix $\boldsymbol{V} \in M_{P\text{x}N}(\mathbb{R})$. For each node in $\boldsymbol{W}$, its state at time t, $\boldsymbol{x}[t]$ is updated by:

$$\boldsymbol{x}[t+1] = \boldsymbol{f}(\eta \boldsymbol{U} \cdot (\boldsymbol{a}[t+1] - \beta) + \varepsilon \boldsymbol{W} \cdot \boldsymbol{x}[t]) \tag{1}$$

for input data at time t, $\boldsymbol{a}[t+1]$ and a nonlinear activation function $\boldsymbol{f}$. There are 4 hyperparameters in the network, η, the input matrix scaling, β, the input bias term, ε, the network coupling strength, and κ, the number of connections from each oscillator in the network. After some transient time t_1 and training time t_2, the output matrix $\boldsymbol{V}$ is trained using linear least squares,

$$\boldsymbol{V} = \boldsymbol{D}\boldsymbol{X}^H(\boldsymbol{X}\boldsymbol{X}^H)^{-1} \tag{2}$$

$$\boldsymbol{y}[t+1] = \boldsymbol{V}\boldsymbol{x}[t] \tag{3}$$

where $\boldsymbol{X} \in M_{N\text{x}(t_1+t_2)}(\mathbb{R})$ is a matrix of oscillator states, $\boldsymbol{D} \in M_{P\text{x}(t_1+t_2)}(\mathbb{R})$ is a matrix of the ground truth system, and H denotes the Hermitian transpose of a matrix. Thus for the testing time, t_3, the state update is given by equation (3). The predicted output from the model, $\boldsymbol{y}[t+1]$ is then used in equation (1) in place of $\boldsymbol{a}[t+1]$ to create a free-running prediction. Model performance is quantified via the Correlation Coefficient between $\boldsymbol{D'} \in M_{P\text{x}t_3}(\mathbb{R})$ and $\boldsymbol{Y'} \in M_{N\text{x}t_3}(\mathbb{R})$, a matrix of the free-running predicted output.

B. Kuramoto Model

The Kuramoto model is a canonical model for describing synchrony in nonlinear oscillator networks, represented by

$$\dot{\theta}_i = \omega + \varepsilon \sum_{j=1}^{N} A_{ij} \sin(\theta_j - \theta_i) \tag{4}$$

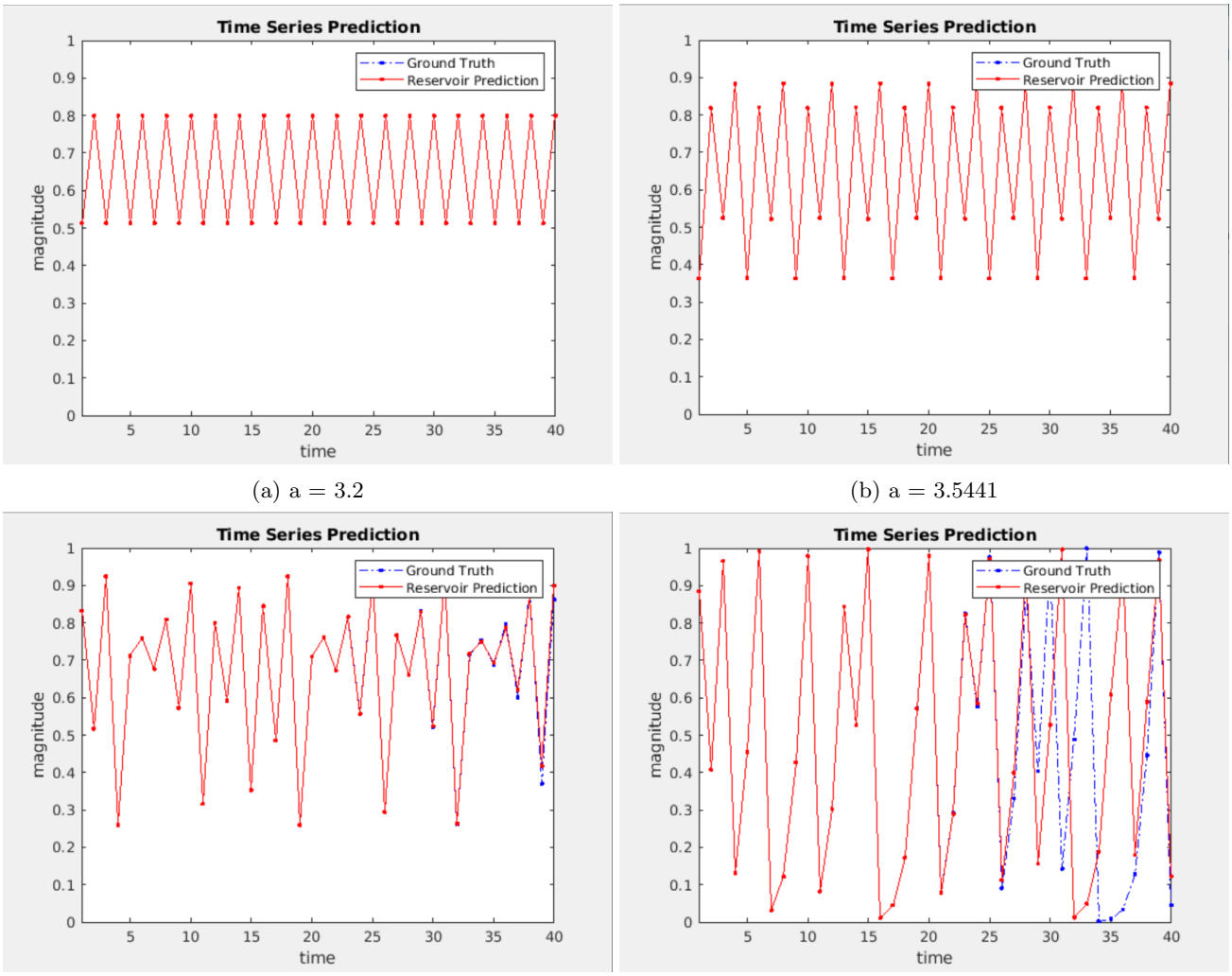
Where θ_i represents the state of oscillator i, ω is the external input, ε is the coupling strength between oscillators, and $\boldsymbol{A}$ is the adjacency matrix representing the network of connections between oscillators. The Kuramoto model is widely used in computational and theoretical neuroscience to describe traveling waves in the brain [6] [7]. While the original Kuramoto model has no analytical solution, an algebraic model with coincidental argument to the original Kuramoto model has recently been developed, which provides many novel analytical insights to the original Kuramoto model [8]. Both the original and algebraic approach to the Kuramoto model can serve as activation functions in an artificial neural network. Since the Kuramoto model has been widely studied as a model for synchrony, and the algebraic approach can be solved with complex-valued solutions, studying the application of these activation functions may allow for tantalizing insights into the dynamics of such a neural network.

In the classic ESN, the activation function, $\boldsymbol{f}$ in equation (1) is commonly chosen to be *tanh* [9]. In our implementation of the Kuramoto ESN, we use an activation function modeled after the Kuramoto model. Equation (4) represents the original Kuramoto model. Note that in equation (4) external input, ω, is constant to all oscillators, but a solution exists for heterogeneous input to oscillators that is constant over time [8]. We consider a similar system to equation (4) that is complex-valued; it is introduced in [8] with a shifted coordinate frame:

$$\dot{\psi}_i = \varepsilon \sum_{j=1}^{N} W_{ij} \left[\sin(\psi_j - \psi_i) + i\cos(\psi_j - \psi_i)\right] \tag{5}$$

which implies $\psi_i \in \mathbb{C}$. Multiplying both sides by of the equation by i and using Euler's formula, the equation becomes:

$$i\dot{\psi}_i = \gamma e^{-i\psi_i} \sum_{j=1}^{N} W_{ij} e^{i\psi_j} \tag{6}$$



(a) a = 3.2

(b) a = 3.5441

(c) a = 3.7

(d) a = 4

Figure 2: Discrete logistic map prediction

Multiplying both sides of the equation by $e^{i\psi_i}$, we get:

$$i\dot{\psi}_i e^{i\psi_i} = \gamma \sum_{j=1}^{N} W_{ij} e^{i\psi_i} \tag{7}$$

Now, noticing that $i\dot{\psi}_i e^{i\psi_i} = \frac{d}{dt} e^{i\psi_i}$ and making the substitution $x_i = e^{i\psi_i}$, the equation simplifies to

$$\dot{x}_i = \gamma \sum_{j=1}^{N} W_{ij} x_j \tag{8}$$

Which, in matrix form is:

$$\dot{\boldsymbol{x}} = \gamma \boldsymbol{W} \boldsymbol{x} \tag{9}$$

Whose general solution is

$$\boldsymbol{x}[t] = e^{\gamma t \boldsymbol{W}} \boldsymbol{x}[0] \tag{10}$$

This update equation is the one used in the Kuramoto ESN. The constraint that the Kuramoto ESN must have constant heterogeneous external input over time is mentioned earlier in this section, but since the external input

follows this rule:

$$\boldsymbol{\omega}[t] = \boldsymbol{U} \cdot \boldsymbol{a}[t] \tag{11}$$

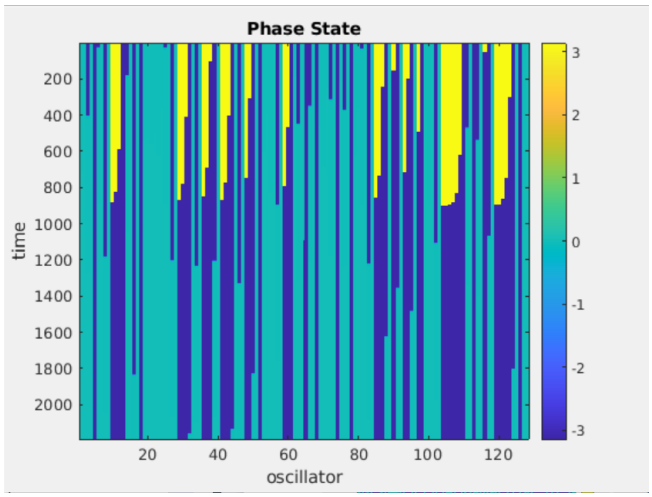
the input is not constant over time. We use a windowed approach to the time dimension, with a time window of one and a time step of one, effectively making the update rule recursive:

$$\boldsymbol{x}[t] = e^{\gamma t \boldsymbol{W}} \boldsymbol{x}[t-1] \tag{12}$$

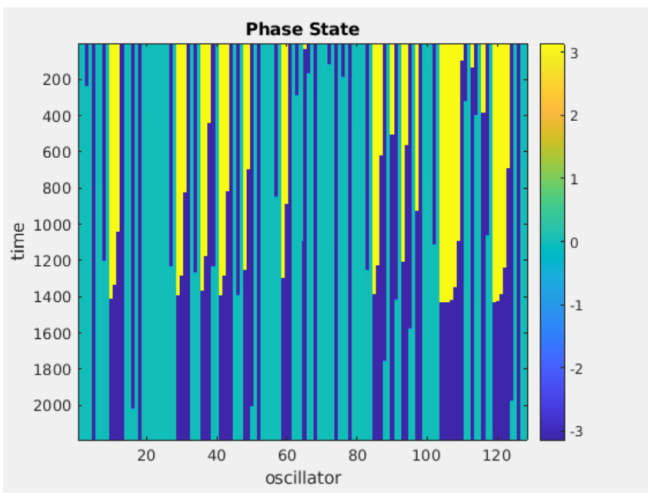
In addition to altering the activation function from the *tanh* function, we make use of a unique reservoir topology. The network for our Kuramoto ESN is that of a ring, or a one-dimensional torus. The matrix representation of our ring is circulant, and thus the Circulant Diagonalization Theorem (CDT) states that the eigenvectors of the reservoir are simply its fourier modes, and can be determined analytically. Thus, equation (12) can also be written as

$$\boldsymbol{x}[t] = \boldsymbol{V} e^{\gamma t \boldsymbol{D}} \boldsymbol{V'} \boldsymbol{x}[t-1] \tag{13}$$

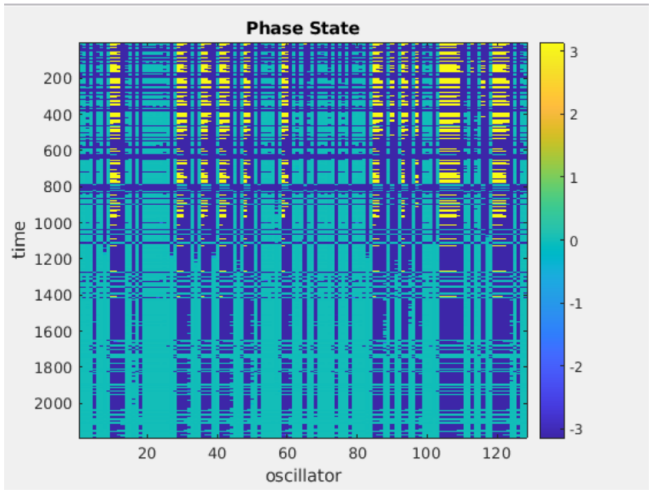
where $\boldsymbol{V}$ is a matrix of the eigenvalues of $\boldsymbol{W}$and $\boldsymbol{D}$ is a diagonal matrix of the eigenvalues of $\boldsymbol{W}$. The analytical description of eigenvalues of the system allows us to further study the Kuramoto ESN in an analytical framework through the eigenmodes of the matrix $\boldsymbol{W}$, given by the inner product $\mu_i[t] = \langle \vec{x[t]}, \vec{v_i} \rangle$, where μ_i denotes the ith eigenmode associated with the ith eigenvalue v_i.



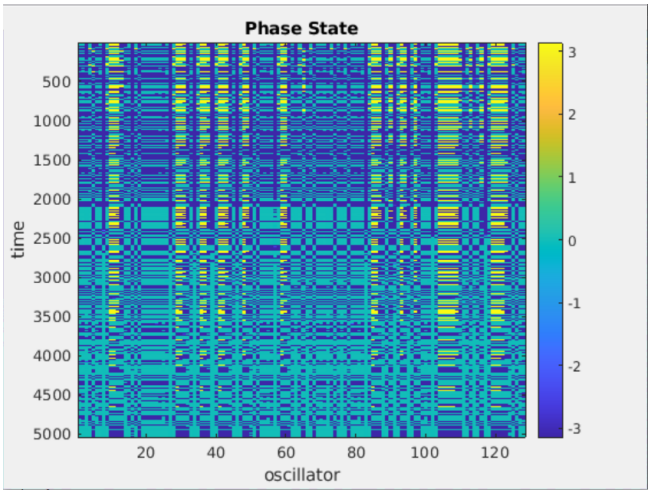
(a) a = 3.2



(b) a = 3.5441

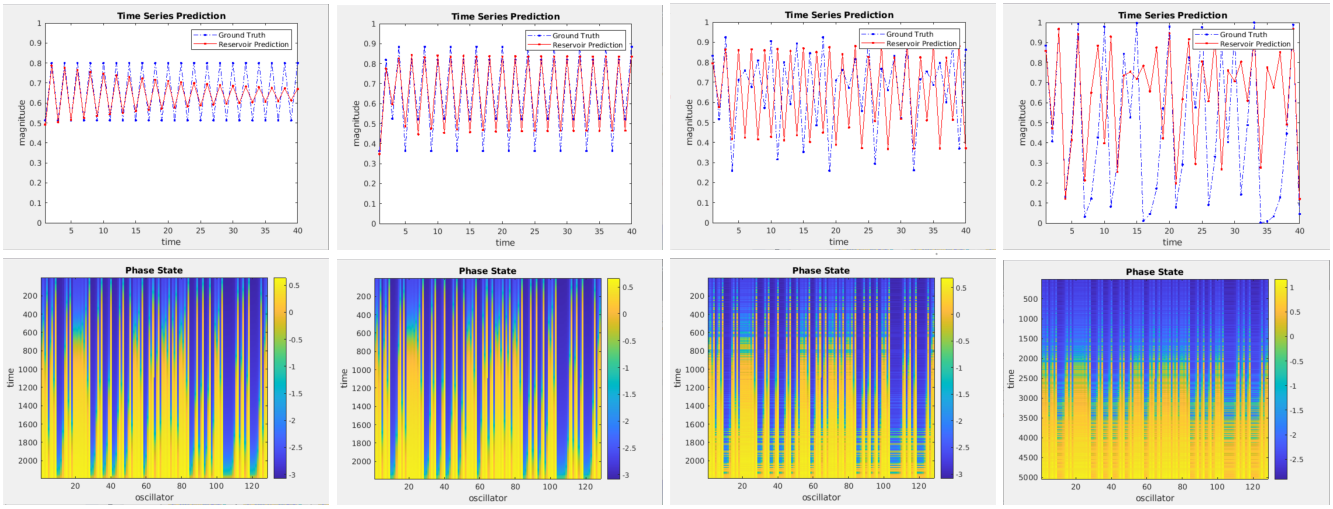


(c) a = 3.7



(d) a = 4

Figure 3: Binary reservoir dynamics (teal and dark blue) for well-predicting Kuramoto ESNs



(a) a = 3.2 (b) a = 3.5441 (c) a = 3.7 (d) a = 4

Figure 4: Poor network predictions and their corresponding reservoir dynamics

C. The logistic map

The logistic map is a canonical system in chaos literature. Although it is the result of a very simple recurrence relation, it can produce complex and chaotic behavior. The discrete logistic map is defined recursively as follows:

$$x_{n+1} = a \cdot x_n(1 - x_n) \tag{14}$$

with $a \in (0, 4]$ and $x_n \in [0, 1]$

Past $a =$ 3.56995..., a number related to the Fiegenbaum constant, the logistic map demonstrates chaotic behaviour. Interestingly, although equation (14) fully characterizes the logistic map, two closed-form analytical solutions have been discovered — a solution for $a = 2$, and one for $a = 4$ (which is of the most interest to our research on chaotic systems):

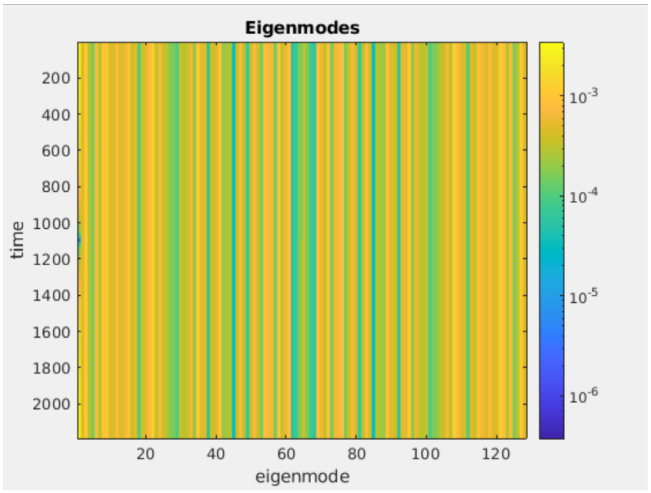
$$f(n) = \frac{1}{2}(1 - \cos(2^n \arccos(1 - 2x_0))) \tag{15}$$

III. RESULTS

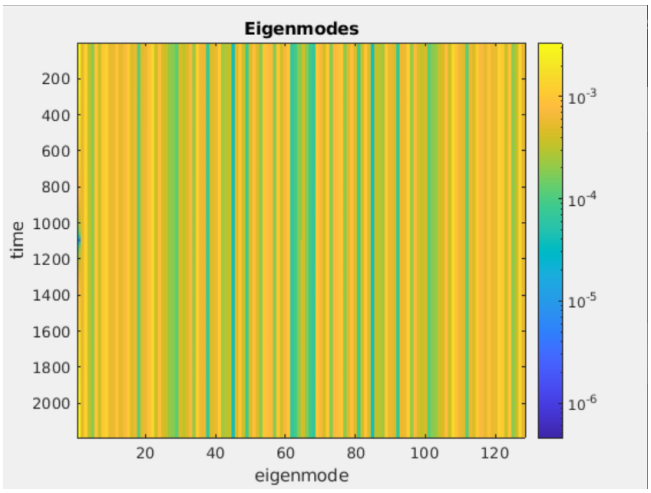
A. Analyzing network performance

Firstly, we classify what is deemed a good prediction from our network. The Correlation Coefficient, r, is used to compare the correlation between the Kuramoto ESN's free-running prediction and the logistic map testing data. We say that when $r > 0.95$ for t time steps, that our network is performing well. When $r <= 0.95$, we say that the network is performing poorly.

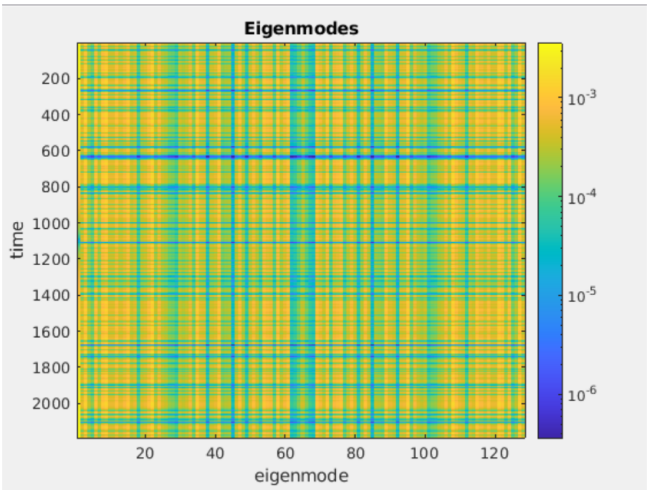
As shown in Figure 2, the Kuramoto ESN makes an accurate free-running prediction for 30 time steps, with $r > 0.96$ for each of the 4 cases tested. The hyperparameter selections η, ε, κ, and β have been found through randomly scanning. It can be seen in the contrast between the groups of Figure 2(a-b) and Figure 2(c-d) that the Kuramoto ESN predicts non-chaotic regimes of the logistic map better than it predicts chaotic regimes, and further within those chaotic regimes, the model does much better in prediction when the Maximum Lyapunov Exponent (MLE) is low, such as at $a = 3.7$, than it does when the MLE is higher, such as at $a = 4$.



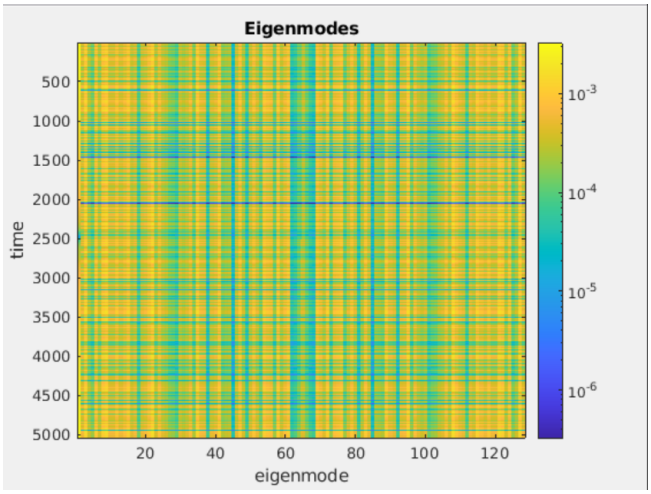
(a) a = 3.2



(b) a = 3.5441



(c) a = 3.7



(d) a = 4

Figure 5: Eigenmodes of the reservoir evaluated on a log scale

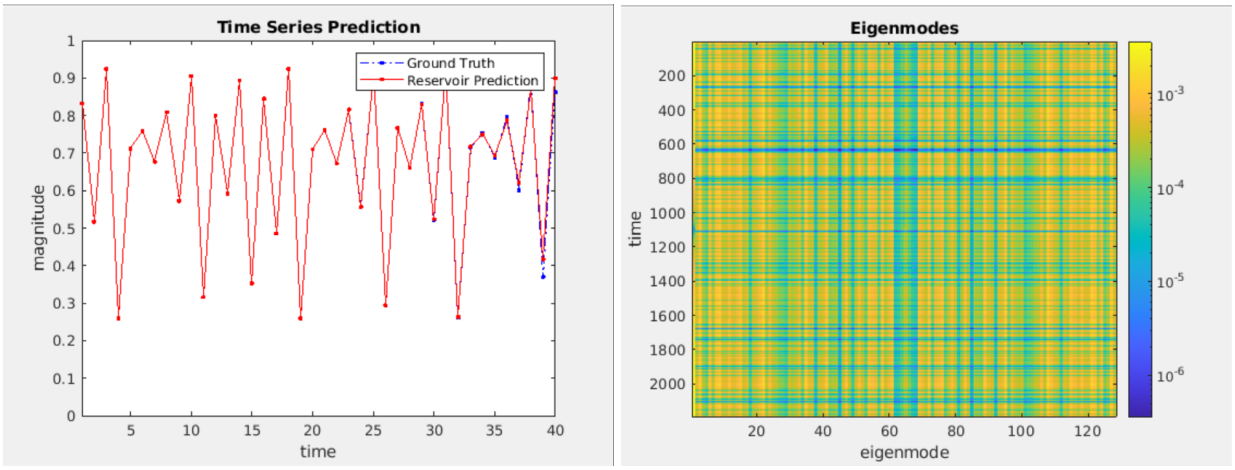
B. Network dynamics

Figure 3 shows the plots of arguments (mean-subtracted) of the oscillators in the Kuramoto ESN as a function of increasing time. One can see an initial transient (yellow and teal), followed by the emergence of binary phase states in the network (dark blue and teal). It can be noted that these binary dynamics fail to arise in a bad prediction of the logistic map dynamics (Figure 4).

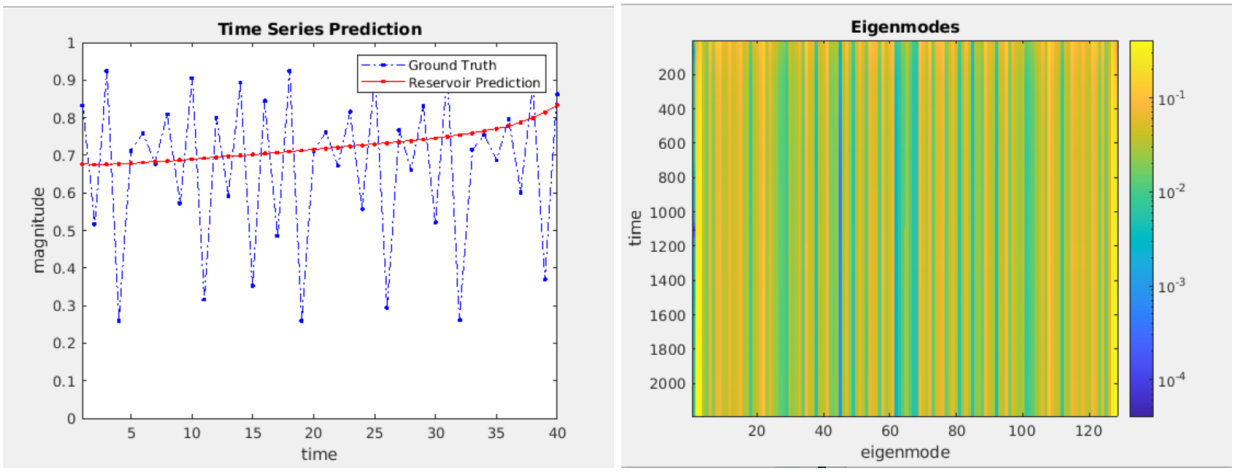
In addition to generating the reservoir dynamics, we also analyze the eigenmode contributions to the dynamics of the reservoir (Figure 5). The dynamics of the reservoir are clearly not governed by the dynamics of any eigenmodes, as they are in fully synchronized systems. It is noted that for a coupling strength $0.001 \leq \varepsilon \leq 0.0001$, dominating eigenmodes begin to emerge (Figure 6). Unfortunately, no scenario in which these dominating eigenmodes coincide with a good prediction has been found.

IV. DISCUSSION

As shown in Figure 3, the phase of the oscillators in the reservoir demonstrate binary dynamics in a good prediction. Because these binary dynamics are so apparent in the simulations, there may be some connection between the binary



(a) $\varepsilon = 3.1673$



(b) $\varepsilon = 3.1673e^{-4}$

Figure 6: Eigenmode contributions in coincidence with predictions

dynamics of a Kuramoto ESN during strong prediction and the binary dynamics of another known physical model, the Ising model. If some connection is found between the reservoir of the Kuramoto ESN and the Ising model, one might be able to nicely apply predictions to physical problems, such as magnetism and gaseous dynamics. In addition to the physical meaning of this connection, there is opportunity to better understand the Ising model in 2 or 3 dimensions, as n-dimensional analogs for the Kuramoto ESN exist. The Ising model is a cannonical model in nonlinear dynamics, and is applied to many areas of research from biology to physics, to engineering.

In addition to gathering more data about the dynamics of the system, a major limitation of the eigenmode analysis currently being done on the system is that the analysis is linear, while the system we analyze is nonlinear. This can explain the complexity of the eigenmode contribution in Figure 5, although eigenmode analysis still may be useful yet, as our system does exhibit areas of local linearity.

Another factor that makes analysis of our Kuramoto ESN more challenging is the fact that there is an absence of traveling waves in our network. Since the Kuramoto ESN echoes the dynamics of the chaotic system it aims to predict, it is reasonable that there are no waves in the network. While the lack of waves is to be expected, it makes analysis in terms of eigenmodes and geometry of synchrony less likely. As such, we must also look for alternative avenues to proceed with network dynamics analysis.

In future work, the authors hope to improve the capabilities of the Kuramoto ESN to predict the logistic map for more time steps, as analysis of longer term free-running dynamics may be more fruitful. We also want to continue formalizing the algebraic analysis of the network, and continue to explore the connection between the Kuramoto ESN

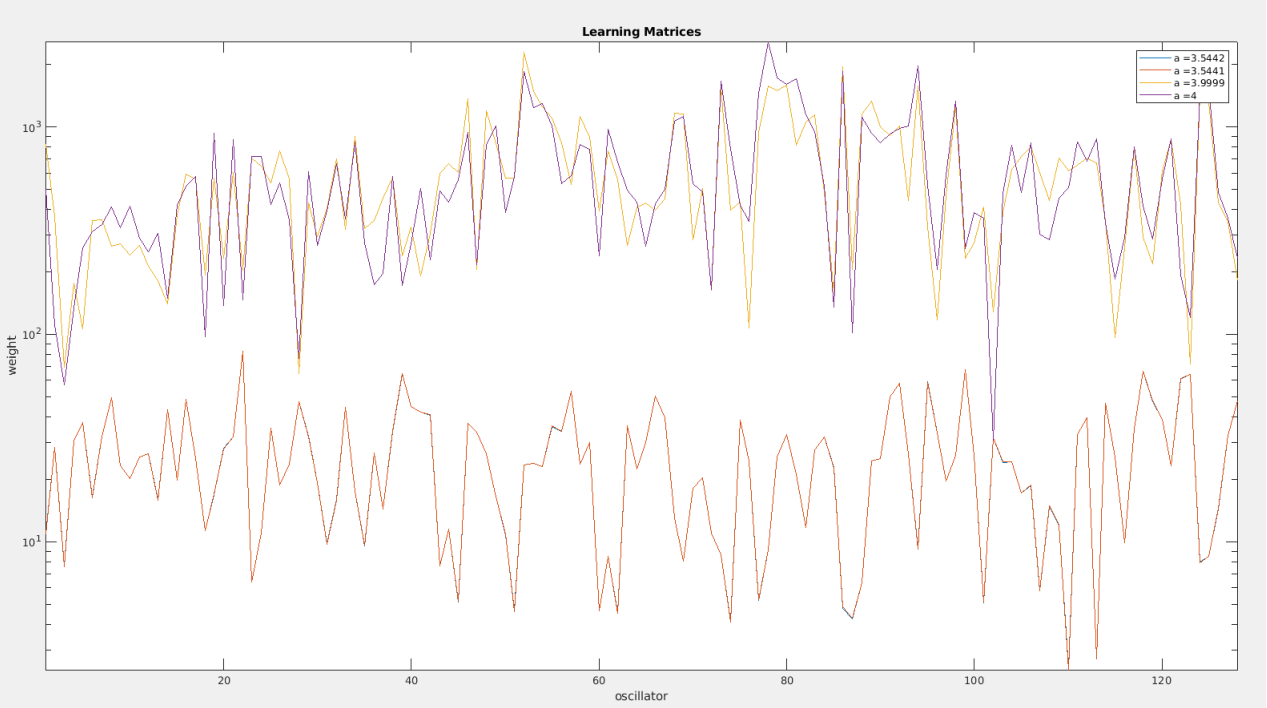


Figure 7: Oscillators of output matrices for different values of a. The weighting for each oscillator (vertical axis) is log-scaled.

and the Ising model.

V. CONCLUSION

The Kuramoto ESN introduced in this paper uses a complex-valued state update rule and alterations to the reservoir topology to provide novel analysis about reservoir computing. Specifically, our observations offer some insight into how ESNs learn and predict nonlinear systems. Using the complex-valued approach to the Kuramoto activation function, we predict the logistic map at a variety of chaotic and non-chaotic regimes successfully for 30 time steps. The Kuramoto ESN reservoir exhibits binary dynamics, but quite indiscernable eigenmode signatures. The authors look forward to continuing to analyze past results from this project, exploring new approaches for analyzing on the Kuramoto ESN reservoir, and looking to use the analytical description of the Kuramoto ESN to formally link it to the Ising model. Analytical descriptions of the ESN, both during and after training, may provide valuable insights into both reservoir computing and nonlinear oscillator networks.

VI. SUPPLEMENTARY

This section offers analytical approaches to this system that may warrant future exploration, but currently remain disconnected to our main conclusions.

A. Learning Matrix Correlation

In most implementations of the Kuramoto ESN, the input matrix weights are sampled randomly from a uniform distribution [-0.5 0.5]. To eliminate some of the randomness from the model, mean-subtracted values generated from the logistic map at $a = 4$ from time $t = 800$ to time $t = 800 + N$ were used in place of the random input vector, $\boldsymbol{U}$, where N is the size of the input vector. With this deterministic model, we noticed high correlation in the output

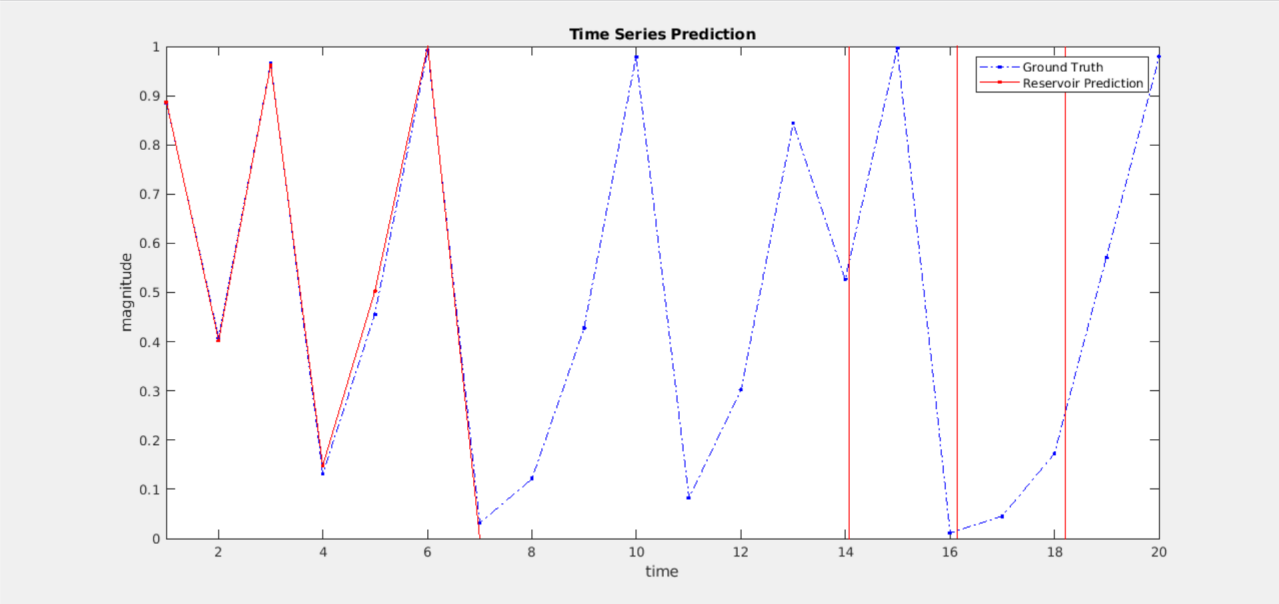


Figure 8: Prediction of the logistic map at a = 3.9999 using a readout matrix trained on a = 4

matrices of trained ESNs on $a = 4$ and $a = 4 - \varepsilon$ (Figure 7). We can also observe decent prediction (Figure 8) of the logistic map at $a = 4 - \varepsilon$ using an output matrix trained for $a = 4$. The logistic map at $a = 4$ is highly chaotic and sensitive to initial conditions, thus the fact that there is any correlation between the learning matrices at $a = 4$ and $a = 4 - \varepsilon$ is noteworthy. This inference indicate that the analytical solution for the logistic map at $a = 4$ may provide insight into the logistic map at other similar a values.

B. Tchebyshev Polynomials

Functions on finite intervals can be approximated using linear combinations of Tchebyshev polynomials. First, choose n to be the number of Tchebyshev nodes on the interval $[a, b]$. The nodes are calculated using the following formula:

$$x_i = cos(\frac{2i-1}{2n}\pi) \tag{16}$$

with $i \in 1, 2, ..., n$. This equation returns Tchebyshev nodes from $[-1, 1]$ that can be re scaled to values from $[a, b]$. Through Tchebyshev approximation, we can find a polynomial which has equal values to the function of interest at the Tchebyshev nodes. Traditionally, Tchebyshev polynomials are used to approximate continuous functions. However, rescaling the nodes allows the Tchebyshev polynomials to be used to analyze discrete functions through continuous approximations. This may be an especially worthwhile approach for analyzing the logistic map, since it is of special interest to this project to extend our understanding of the continuous analytical solution to the logistic map at $a = 4$ to other a values for which we have only discrete and recursive representations. Currently, we can transform equation (14) to an evaluated Tchebyshev polynomial as follows. Considering the substitution:

$$\boldsymbol{X}(t) = 1 - 2\boldsymbol{x}(t) \tag{17}$$

Thus equation (14) becomes:

$$\boldsymbol{X}(t+1) = 2\boldsymbol{X}^2(t) - 1 \tag{18}$$

Then equation (15) evaluates to:

$$\boldsymbol{X}(t) = \cos(2^t \arccos(\boldsymbol{X}(0))) \tag{19}$$

We can now express this transformed discrete logistic map as a direct evaluation of a Tchebyshev polynomial:

$$\boldsymbol{x}(t+1) \equiv T_n(\boldsymbol{x}(t)) \tag{20}$$

Or

$$\boldsymbol{x}(t+1) \equiv \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2k} (\boldsymbol{x}[t]^2 - 1)^k \boldsymbol{x}[t]^{n-2k} \tag{21}$$

Now we draw the reader's attention to the Taylor series expansion (for matrices) for equation (12), the Kuramoto ESN activation function:

$$\boldsymbol{x}[t+1] = \boldsymbol{x}[\boldsymbol{t}] \sum_{i=0}^{n} \frac{(\gamma t \boldsymbol{W})^i}{i!} \tag{22}$$

Attempts have been made to truncate equation (22) at a suitable n, and equate terms in (21) with terms in (22). The use case of Tchebyshev polynomials to interpolate functions may lead to a formal mathematical description of how the ESN learns the logistic map for $a \in (0, 4]$.

VII. ACKNOWLEDGEMENTS

The authors would like to acknowledge Gabriel Benigno, Dr. Roberto Budzinski, Molly, and the members of the Muller Lab for insightful discussion and mentorship, and for providing a welcoming and exciting atmosphere in which we were able to explore this project. The authors extend a special thank you to Dr. Lyle Muller for all the mentorship, supervision, and expertise provided throughout the summer. This work is supported by the USRI program at Western and the NSERC USRA program.

[1] E. Weinan, Notices of the American Mathematical Society **68**, 565 (2021).
[2] O. Barak, Current opinion in neurobiology **46**, 1 (2017).
[3] P. J. Werbos, Proceedings of the IEEE **78**, 1550 (1990).
[4] H. Jaeger, scholarpedia **2**, 2330 (2007).
[5] L. Grigoryeva and J.-P. Ortega, Neural Networks **108**, 495 (2018).
[6] J. A. Acebrón, L. L. Bonilla, C. J. P. Vicente, F. Ritort, and R. Spigler, Reviews of modern physics **77**, 137 (2005).
[7] M. Breakspear, S. Heitmann, and A. Daffertshofer, Frontiers in human neuroscience **4**, 190 (2010).
[8] L. Muller, J. Mináč, and T. T. Nguyen, Physical Review E **104**, L022201 (2021).
[9] M. Lukoševičius, in *Neural networks: Tricks of the trade* (Springer, 2012), pp. 659–686.