Observing spatio-temporal dynamics of excitable media using reservoir computing

Roland S. Zimmermann\textsuperscript{1,2} and Ulrich Parlitz\textsuperscript{1,2,3}

\textsuperscript{1)} Max Planck Institute for Dynamics and Self-Organization, Am Fassberg 17, 37077 Göttingen, Germany
\textsuperscript{2)} Institute for Nonlinear Dynamics, Georg-August-Universität Göttingen, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany
\textsuperscript{3)} German Center for Cardiovascular Research (DZHK, partnersite Göttingen), Robert-Koch-Straße 42a, 37075 Göttingen, Germany

(Dated: 19 April 2018)

We present a dynamical observer for two dimensional PDE models describing excitable media, where the required cross prediction from observed time series to not measured state variables is provided by Echo State Networks receiving input from local regions in space, only. The efficacy of this approach is demonstrated for (noisy) data from a (cubic) Barkley model and the Bueno-Orovio-Cherry-Fenton model describing chaotic electrical wave propagation in cardiac tissue.

To appear in: Chaos, Vol 28 issue 4, April 2018
If for some dynamical process of interest a mathematical model exists but not all of its state variables can be measured an observer may be employed for reconstructing dynamical variables from available time series. This approach constitutes the core of any data assimilation scheme and it is usually implemented by incorporating the model equations into the algorithm. As an alternative, machine learning methods can be applied to approximate the functional relation between observed time series and (unknown) state variables. With this approach, the model of the process is only used to generate training data for the machine learning algorithm but it is not required for the following reconstruction of dynamical variables from new observations. Therefore, this kind of learning observer can also be trained with experimental data, where during an extended measurement not only the observed time series but also other dynamical variables of interest are measured. We demonstrate the efficacy of this approach for observing spatio-temporal chaotic dynamics in excitable media (e.g., cardiac tissue) using Echo State Networks for learning and approximating the required relations between observed and not observed quantities. As a crucial concept for coping with the high dimensionality of the underlying dynamics of the type of spatially extended systems considered here, only time series from nearby sampling sites are fed into the network for cross predicting unobserved variables at some reference location.

I. INTRODUCTION

With the increase in available computational power and novel theoretical concepts the usage of neural networks for machine learning gained exceeding popularity in recent years\textsuperscript{1}. A particular class called recurrent neural networks represents dynamical systems which are able to develop temporal memory and therefore qualify to analyze and process temporal data as needed in text or speech analysis\textsuperscript{2,3}, for example. However, due to their more involved structure recurrent networks are harder to train compared to feed forward networks. Current algorithms like backpropagation through time are occasionally unstable, as the gradients of the cost functions may vanish or diverge\textsuperscript{3,4} complicating the training of recurrent networks.
To solve this issue *Echo State Networks* (ESNs) have been introduced by H. Jaeger\(^5\) in 2001. Almost at the same time the concept of *Liquid State Machines* was published by W. Maass et al.\(^6\). Even though both methods have a different origin they are very similar\(^7\) and therefore, nowadays commonly known as the *Reservoir Computing* framework\(^8\). Driven by some input signal ESNs provide a unique nonlinear transformation of this signal as their output which can be used for prediction or classification \(^8–10\). This approach has also been implemented in hardware\(^11,12\) providing very efficient solutions for particular applications. From a dynamical systems point of view ESNs are driven dynamical systems exhibiting generalized synchronization\(^13–16\) resulting in a (nonlinear) functional relation between the state of the driving system and the state of the driven system which can be exploited for modeling and prediction\(^17–19\). Feeding back the output of a (previously trained) reservoir to its input results in autonomously running ESNs that can be used for prediction and for estimating Lyapunov exponents\(^9,20\). Lu et al.\(^21\) demonstrated how ESNs can be employed to implement dynamical state observers where the underlying dynamical equations of the system of interest are only required for generating suitable training data, but are not part of the observer. This approach is demonstrated here with spatio-temporal data generated by two PDE models describing two-dimensional excitable media making use of the concept of local states\(^22\) that was recently also used by Pathak et al.\(^23\) for predicting complex spatio-temporal dynamics employing Echo State Networks.

### II. EXCITABLE MEDIA

Many physical, chemical, or biological systems are excitable\(^24–26\). They possess a stable equilibrium, but perturbations above some specific threshold may result in a trajectory which returns to the fixed only after some large excursion in state space. For neurons or cardiac cells, for example, this dynamical behavior leads to characteristic pulses of the cell membrane voltage called action potentials. Spatially extended excited systems, also called *excitable media*, very often generate spiral or scroll waves, which can be stable, or meandering, or may break-up in spatio-temporal chaos. An important field where excitable dynamics plays a major role are electro-mechanical excitation waves in the heart muscle\(^27\), because its direct relation to different forms of cardiac arrhythmias\(^28\).
**A. Barkley Model**

The Barkley model has been developed by D. Barkley in 1991 as a qualitative model for excitable media. The model consists of a fast changing variable $u$ and a slowly changing, inhibitory variable $v$ and it can be used, for example, to describe electrical excitation waves in cardiac tissue where variable $u$ represents the membrane voltage of the cells. To simulate chaotic wave breakup (corresponding to cardiac arrhythmias) we consider here a modified version \(^{30}\) (1) of the standard Barkley model with a cubic nonlinearity in the differential equation describing the dynamics of the variable $v$.

\[
\begin{align*}
\frac{\partial u}{\partial t} &= D \cdot \nabla^2 u + \frac{1}{\epsilon}(1-u) \left( u - \frac{v+b}{a} \right) \\
\frac{\partial v}{\partial t} &= u^3 - v.
\end{align*}
\] (1)

The diffusion constant $D$ describes the spatial coupling and is set to $D = 1/25$. The three remaining parameters are chosen to be\(^ {30}\)

\[
a = 0.8 \quad b = 0.01 \quad \epsilon = 0.02.
\] (2)

The system has been simulated on a grid of $150 \times 150$ points with a grid constant of $\Delta x = 0.1$ space units. An explicit Euler stepping in time with $\Delta t = 0.01$ and 5 point approximation of the Laplace operator were used in combination with no-flux boundary conditions. A typical snapshot of the $u$-pattern is shown in Fig. 1(a).

**B. Bueno-Orovio-Cherry-Fenton Model**

The Bueno-Orovio-Cherry-Fenton (BOCF) model provides a more realistic description of excitable cardiac dynamics\(^ {31}\). The evolution of the four system variables is given by four partial differential equations (PDEs)

\[
\begin{align*}
\frac{\partial u}{\partial t} &= D \cdot \nabla^2 u - (J_{si} + J_{fi} + J_{so}) \\
\frac{\partial v}{\partial t} &= \frac{1}{\tau_v} (1 - H(u - \theta_v))(v_{\infty} - v) - \frac{1}{\tau_v^+} H(u - \theta_v)v \\
\frac{\partial w}{\partial t} &= \frac{1}{\tau_w^-} (1 - H(u - \theta_w))(w_{\infty} - w) - \frac{1}{\tau_w^+} H(u - \theta_w)w \\
\frac{\partial s}{\partial t} &= \frac{1}{2\tau_s} (1 + \tanh(k_s(u - u_s))) - 2s,
\end{align*}
\] (3)
Fig. 1. Snapshots of the dynamics of the $u$-variable of (a) the Barkley model (1) and (b) the BOCF model (3) at time step $n = 1000$ of the test data set.

with $H(\cdot)$ being the Heaviside function\textsuperscript{31}. The three currents $J_{si}$, $J_{fi}$ and $J_{so}$ are given by the equations

$$J_{si} = -\frac{1}{\tau_{si}}H(u - \theta_w)ws$$

$$J_{fi} = -\frac{1}{\tau_{fi}}vH(u - \theta_v)(u - \theta_v)(u_a - u)$$

$$J_{so} = \frac{1}{\tau_o}(u - u_o)(1 - H(u - \theta_v)) + \frac{1}{\tau_{so}}H(u - \theta_w).$$

Furthermore, seven voltage dependent variables

$$\tau_v^- = (1 - H(u - \theta_v^-))\tau_{v1}^- + H(u - \theta_v^-)\tau_{v2}^-$$

$$\tau_w^- = \tau_{w1}^- + \frac{1}{2}(\tau_{w2}^- - \tau_{w1}^-)(1 + \tanh(k_w(u - u_w^-)))$$

$$\tau_o^- = \tau_{so1}^- + \frac{1}{2}(\tau_{so2}^- - \tau_{so1}^-)(1 + \tanh(k_{so}(u - u_{so}^-)))$$

$$\tau_s = (1 - H(u - \theta_w))\tau_{s1} + H(u - \theta_w)\tau_{s2}$$

$$\tau_o = (1 - H(u - \theta_o))\tau_{o1} + H(u - \theta_o)\tau_{o2}$$

(5)

$$v_\infty = \begin{cases} 
1, & \text{if } u \leq \theta_v^- \\
0, & \text{if } u \geq \theta_v^- \end{cases}$$

$$w_\infty = (1 - H(u - \theta_o))(1 - \frac{u}{\tau_{w\infty}}) + H(u - \theta_o)w_\infty^*$$
The characteristic behavior is determined through 28 parameters. In our simulations we used the set of parameters given in Table I for which the BOCF model exhibits chaotic excitation wave dynamics similar to the Ten Tusscher-Noble-Noble-Panfilov (TNNP) model. The chaotic dynamics of this system is actually transient chaos whose lifetime grows exponentially with system size. To obtain chaotic dynamics with a sufficiently long lifetime the system has been simulated on a domain of 500 × 500 grid points with a grid constant of \( \Delta x = 1.0 \) space units and a diffusion constant \( D = 0.2 \). Furthermore, an explicit Euler stepping in time with \( \Delta t = 0.1 \), a 5 point approximation of the Laplace operator, and no-flux boundary conditions were used. A snapshot of the \( u \)-pattern is shown in Fig. 1(b). For further processing the \( u, v, w \) and \( s \) fields have been mapped to a grid of size of 150 × 150 points using a bilinear interpolation of OpenCV.

| \( u_0 \) | 0 | \( \tau_{v2}^- \) | 1150 | \( \tau_{fi}^- \) | 0.11 | \( \tau_{s1}^- \) | 2.7342 |
| \( u_1 \) | 1.58 | \( \tau_{v2}^+ \) | 1.4506 | \( \tau_{o1}^- \) | 6 | \( \tau_{s2}^- \) | 3 |
| \( \theta_v \) | 0.3 | \( \tau_{w1}^- \) | 70 | \( \tau_{o2}^- \) | 6 | \( k_s \) | 2.0994 |
| \( \theta_w \) | 0.015 | \( \tau_{w2}^- \) | 20 | \( \tau_{so1}^- \) | 43 | \( u_s \) | 0.9087 |
| \( \theta_v^- \) | 0.015 | \( \tau_{w2}^- \) | 65 | \( \tau_{so2}^- \) | 0.2 | \( \tau_{si}^- \) | 2.8723 |
| \( \theta_o \) | 0.006 | \( \tau_{u_v}^- \) | 0.03 | \( k_{so}^- \) | 2 | \( \tau_{w\infty}^- \) | 0.07 |
| \( \tau_{w1}^- \) | 60 | \( \tau_{w_*}^- \) | 280 | \( u_{so}^- \) | 0.65 | \( w_{\infty}^* \) | 0.94 |

### TABLE I. TNNP model parameter values for the BOCF model

### III. RESERVOIR COMPUTING

ESNs are a simplified type of recurrent neural networks, in which the majority of weights describing the strength of the connections is not changed during the training process. This solves the previously addressed issues of instability during the training of recurrent networks. In general, a network \( E \) maps a temporal signal \( \vec{u}_n \in \mathbb{R}^{N_u} \) to the temporal output \( \vec{y}_n \in \mathbb{R}^{N_y} \) where \( n \) denotes discrete time.

Following Jaeger’s design, we use networks consisting of a reservoir of \( N \) nonlinear units. The connections inside the reservoir are described by the weight matrix \( \mathbf{W} \in \mathbb{R}^{N \times N} \). Together with a bias \( b_{in} \) the input signal \( \vec{u} \) is injected to the reservoir’s units weighted by the input matrix \( \mathbf{W}_{in} \in \mathbb{R}^{N \times (N_u+1)} \). A schematic illustration of the ESN’s structure with the
FIG. 2. General structure of the ESN. From left to right the input signal $\vec{u}_n = \vec{u}(n)$ passes first through $N_u$ input units, afterwards through a reservoir of $N$ units until it reaches $N_y$ output units where the output signal $\vec{y}_n = \vec{y}(n)$ is calculated.

matrices $\mathbf{W}$ and $\mathbf{W}_{in}$ is given in Fig. 2. The dynamics of the network is mainly defined by the input signal and the weight matrix $\mathbf{W}$.

The state of the network is described by these reservoir’s units and will be called $\vec{s}_n$. Its dynamics is given by the update equation

$$\vec{s}_n = (1 - \alpha)\vec{s}_{n-1} + \alpha f_{in}(\mathbf{W}_{in}[b_{in}; \vec{u}_n] + \mathbf{W}\vec{s}_{n-1}),$$  \hspace{1cm} (6)

with the leaking rate $\alpha \in (0, 1]$ and $b_{in} = 1$. We chose the nonlinear transfer function $f_{in}$ to be $\tanh(\cdot)$.

Using the internal states one can define the extended inner states

$$\vec{x}_n = [b_{out}; \vec{s}_n; \vec{u}_n] \in \mathbb{R}^{1+N+N_u},$$  \hspace{1cm} (7)

with an output bias $b_{out} = 1$. Linear superposition of these extended internal states is used to construct the output $\vec{y}_n$ of the reservoir

$$\vec{y}_n = \mathbf{W}_{out}\vec{x}_n = \mathbf{W}_{out}[b_{out}; \vec{s}_n; \vec{u}_n],$$  \hspace{1cm} (8)

with output matrix $\mathbf{W}_{out} \in \mathbb{R}^{N_y \times (1+N+N_u)}$. The output matrix is determined by minimizing the cost function

$$C(\mathbf{W}_{out}) = \sum_n ||\vec{y}_n - \mathbf{W}_{out}\vec{x}_n||^2 + \lambda \text{Tr} (\mathbf{W}_{out} \mathbf{W}_{out}^T)$$  \hspace{1cm} (9)
with respect to $W_{out}$. The first term measures the $l^2$ error of the prediction, while the second one is a regularization term with strength $\lambda > 0$ which is used to avoid overfitting of $W_{out}$.

During the training process the input signal is first applied for $T_0$ time steps to let some transient of the network decay. Afterwards, for $T$ time steps in the range $T_0 < n \leq T_0 + T$ the input is applied and the extended states $\tilde{x}_n$ are observed and collected. Using a matrix notation one can introduce the state matrix $X \in \mathbb{R}^{(1+N_u+N) \times T}$ whose columns consist of the extended states $\tilde{x}_n$. In analogy, the target outputs $\tilde{y}_n$ are the columns of the output matrix $Y \in \mathbb{R}^{N_y \times T}$. With this notation Eq. (8) can be rewritten as

$$Y = W_{out}X.$$ (10)

The training process is finalized by computing the minimum of the cost function (9), which is given by

$$W_{out} = YX^T (XX^T + \lambda I)^{-1}.$$ (11)

To stabilize the training procedure and make the ESN more resistant to noisy input data, we added normally distributed noise $\bar{n}$ ($\mu = 0, \nu = \nu_{max}$) to the argument of the nonlinear transfer function $f_{in}$ and treated $\nu_{max}$ as a hyperparameter that was optimized. In order to predict the temporal signal $\tilde{y}_n$ for $n > T_0 + T$, the input signal $\bar{u}_n$ is successively fed into the network observing the extended states. Using these states the predicted output can be calculated using Eq. (8).

The weight matrix $W$ is initialized with uniformly distributed random numbers $\in [-0.5, 0.5]$ and turned into a sparse matrix with only $\epsilon = 5\%$, $\epsilon = 10\%$ or $\epsilon = 20\%$ non-vanishing elements by randomly eliminating non-zero values. Then all elements of $W$ are rescaled in a grid search to find an optimal value of the spectral radius $\rho(W) = \max\{|\mu_1|, |\mu_2|, \ldots, |\mu_N|\}$ where $\mu_n (n = 1, \ldots, N)$ are the eigenvalues of $W$. The elements of the (dense) matrix $W_{in}$ are sampled from a uniform distribution of random numbers in $[-0.5, 0.5]$.

The performance of the reservoir depends strongly on the hyperparameters listed in Table II. Since the state of the art gradient descent optimization method for ESN hyperparameters is prone to instabilities we used a rigorous grid search instead to find optimal values for the hyperparameters. The optimal values have been determined by means of a cross validation.
Hyperparameter Values

$N \in \{50, 100, 200, 400\}$

$\rho \in \{0.1, 0.5, 0.8, 0.95, 1.1, 1.5, 3.0\}$

$\alpha \in \{0.05, 0.2, 0.5, 0.9, 0.95\}$

$\epsilon \in \{0.05, 0.1, 0.2\}$

$\nu_{\text{max}} \in \{10^{-4}, 10^{-5}\}$

$\lambda \in \{5 \cdot 10^{-2}, 5 \cdot 10^{-3}, 5 \cdot 10^{-4}, 5 \cdot 10^{-5}, 5 \cdot 10^{-6}\}$

TABLE II. The examined range of hyperparameters.

procedure. For a given set of hyperparameter values (after transients of the network decayed) the first $T = 10000$ time steps were used to estimate the corresponding output matrix $W_{\text{out}}$. Then a validation error based on $T_{\text{val}} = 2000$ consecutive time steps was computed and used to evaluate the performance of this set of hyperparameter values (Mean Squared Error (12) and Normalized Root Mean Squared Error (13)). Once all combinations of hyperparameter values had been evaluated the selection of the optimal set was made and this set was finally evaluated with another (consecutive) input sequence of length $T_{\text{test}} = 2000$. The examined parameter values are presented in Table II and the optimal values for the examples investigated will be given in Tables IV and V (below).

By feeding the output signal back to the input of the ESN one can implement an autonomous dynamical model which can be used, for example, to estimate Lyapunov exponents of the process of interest.

IV. MODEL LOCALITY

In a straightforward implementation of the ESN concept the input would consist of the entire spatio-temporal signal provided by the PDEs describing the excitable medium. For the examples considered here this multivariate input would be given by $150 \times 150 = 22500$ variables and the ESN should possess (at least) a similar number of nonlinear units (nodes). As a result, one would have to deal with very large matrices and very high computational costs. To avoid this particular type of 'curse of dimensionality' we exploit the fact that the underlying dynamics is governed by local interactions (of cells) and for predicting dynamics at a given spatial location only information about its surroundings are required. Technically,
this assumption of locality can be implemented by means of local states\textsuperscript{22,23}. This allows us to use independent but identical ESNs for each pixel. The local states contain information about the dynamics at one specific pixel \((i, j)\) and its local neighbours in a \(\sigma \times \sigma\) sized square with a total area of \(\sigma^2\) pixels. The construction of local states is illustrated in Fig. 3. As immediate neighbours might have a high cross correlation and therefore, contain mostly similar (and redundant) information we introduced a further hyperparameter \(\Delta\sigma\) which describes the distance between included neighbours: A value \(\Delta\sigma = 1\) means that all \(\sigma^2\) pixels are included, whereby from \(\Delta\sigma = 2\) follows that only every second pixel in each direction is used. This proceed yields for each (reference) pixel a \(\lceil \frac{\sigma^2}{\Delta\sigma} \rceil\) dimensional vector, which will be used as the input vector for the corresponding ESN. We treated \(\sigma\) and \(\Delta\sigma\) as additional hyperparameters and examined the values listed in Table III.

\[
\begin{array}{cccc}
\sigma & 1 & 3 & 5 & 7 \\
\Delta\sigma & 1 & 1 & 2 & 3 \\
\end{array}
\]

TABLE III. Examined values of \(\sigma\) and \(\Delta\sigma\) for the generation of the local states.

V. CROSS PREDICTION

The Barkley model and the BOCF model have both been simulated for 14000 time steps. The first \(T_{\text{training}} = 10000\) steps are used to train the networks while the next
\(T_{\text{validation}} = 2000\) steps are used to evaluate the different hyperparameters during the grid search. Finally, the models’ performance are tested over the last \(T_{\text{test}} = 2000\) time steps. For both models the ESNs are driven by time series sampled with step size \(\Delta t = 0.01\) and \(\Delta t = 0.1\) for the Barkley model and the BOCF model, respectively.

For the Barkley model the cross predictions \(u \rightarrow v\) and \(v \rightarrow u\) have been analyzed, and for the BOCF model predictions \(u \rightarrow (v, w, s), v \rightarrow (u, w, s), w \rightarrow (u, v, s),\) and \(s \rightarrow (u, v, w)\), have been examined. To test the developed approach under more realistic conditions, we added normally distributed noise \((\mu = 0, \sigma = 0.075)\) to the corresponding source (input) field generated by the Barkley and the BOCF model. For the \(u\) and \(v\) variables of the Barkley model this resulted in signal-to-noise ratios (SNRs) of 23.7dB and 17.2dB, respectively. The noisy ESN input from the BOCF provided by the \(u, v, w,\) or \(s\) signals had SNRs of 23.7dB, 17.2dB, 21.8dB, and 19.0dB, respectively. Figure 4 show snapshots of the resulting noisy input of the ESN.

To perform the cross prediction we distinguished between inner pixels and those near to a border of the field. This distinction is necessary as pixels near the borders do not have enough neighboring pixels to construct a complete local state. Furthermore, due to the impact of the no-flux boundary conditions we expect the effective dynamics close to the boundaries to be slightly different from the dynamics inside the simulation region. For the boundary pixels, we used distinct but identically constructed ESNs for each pixel. Those were trained for each pixel on its own to obtain the prediction.

For the inner pixels a two staged method has been applied: At first an ESN has been created as described above. Next, for each pixel a copy of this network has been created and trained. Finally, we used the calculated output matrices \(W_{\text{out},(i,j)}\) for all inner pixels \((i,j)\) to determine the averaged output matrix \(W_{\text{out}} = \langle W_{\text{out},(i,j)}\rangle\). This matrix then has been used as the output matrix for each pixel’s ESN to calculate the prediction. Following this procedure we were able to yield better results than by leaving out the averaging step and predicting each pixel with its own output matrix \(W_{\text{out},(i,j)}\).

Instead of averaging the output matrix, one could also summarize the state matrices \(X\) of each pixel’s ESN in a single huge matrix, to calculate one common output matrix \(W_{\text{out}}\). As this method requires much more memory, we skipped further analysis and only used the averaging approach.
FIG. 4. Snapshots of the dynamics of the $u$-variable (a) and the $v$-variable (b) the Barkley model, and the $u$-variable (c) and $w$ variable (d) of the BOCF model at time step $n = 1000$ of the test data set, all with the added Gaussian noise $\in \mathcal{N}(0, 0.075)$.

A. Barkley Model

For the Barkley model cross predictions in both possible directions, $u \rightarrow v$ and $v \rightarrow u$, have been performed. The hyperparameters of the ESN have been searched in the previously described way. Furthermore, an optimization of the parameters $\sigma$ und $\Delta \sigma$ has been conducted. The results of both are listed in Table IV.

Since spatial coupling is present only in the differential equation for $u$ (Eq. (1)) the local dynamics of $v$ is governed by local values of $u$ and $v$, only. Therefore, the optimal value of $\sigma$ for the cross prediction $u \rightarrow v$ equals one, i.e. no information from neighboring dynamics is required. In contrast, predictions of $u$ from $v$ benefit from samples around the reference site with optimal values of $\sigma = 5$ and $\Delta \sigma = 2$ (see Table IV). Figures 5 and 6 show snapshots from cross predictions $u \rightarrow v$ and $v \rightarrow u$, respectively. Figure 5(a) shows the (original)
noiseless \( v \) variable of the Barkley system at the time step \( n = 1000 \) of the test data set. Next to it in Fig. 5(b) is the resulting \( v \) variable of the cross prediction from the noisy \( u \) variable (snapshot shown in Fig. 4(a)). Figure 5(c) shows the absolute error, i.e. the magnitude of the difference between Figs. 5(a) and 5(b). For comparison, Fig. 5(d) shows the absolute error if a noiseless input signal \( u \) is used to predict the \( v \) variable. As can be seen, even with noisy input the wave patterns of the \( v \) variable are very well predicted (Fig. 5(b)), with only minor errors (Fig. 5(c)). Table IV shows the hyperparameter values used for this cross prediction and the Mean Squared Error

\[
\text{MSE}(v) = \frac{1}{m^2 T_{\text{Test}}} \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{t=1}^{T_{\text{Test}}} (v_{ij}^{\text{true}}(t) - v_{ij}(t))^2
\]

(12)
as well as the Normalized Root Mean Squared Error

\[
\text{NRMSE}(v) = \sqrt{\frac{\text{MSE}(v)}{\text{MSE}(\bar{v})}}
\]

(13)

where \( \bar{v} \) denotes the temporal and spatial mean values of the training sequence \( v_{ij}(t) \) and \( m = 150 \).

Figure 6(a) shows the (original) noiseless \( u \) variable of the Barkley system at the time step \( n = 1000 \) of the test data set and Fig. 6(b) displays the resulting \( u \) variable of the cross prediction from the noisy \( v \) variable (snapshot shown in Fig. 4(b)). Figure 6(c) shows the absolute error, i.e. the magnitude of the difference between Figs. 6(a) and 6(b). For comparison, Fig. 6(d) shows the absolute error if a noiseless input signal \( v \) is used to predict the \( u \) variable. This result shows that variable \( u \) of the Barkley model can be predicted from noisy \( v \) time series with relatively low errors (Fig. 6(c)) if local dynamics is exploited by means of local states representing a region of \( 5 \times 5 \) pixels (sampled with \( \Delta \sigma = 2 \), see Fig. 3).

B. BOCF Model

For the BOCF system (3) we computed cross predictions from noisy \( u, v, w \) or \( s \) time series. Similar to the Barkley model, only the differential equation for \( u \) contains the diffusion term which provides local spatial coupling. Therefore, cross predictions from \( u \) to the other variables would for noiseless data in principle not benefit from additional information.
FIG. 5. Cross prediction of the $v$ variable of the Barkley model Eq. (1) from noisy $u$ data (see Fig. 4(a)) at time step $n = 1000$ of the test data set. (a) True $v$ values, (b) predicted $v$ values, (c) magnitude of difference between true and predicted $v$ values, (d) magnitude of difference between true and predicted $v$ values if noiseless $u$ data are used as input of the ESN. Note the different ranges of the colorbars in figures (c) and (d).

obtained by including neighboring samples in the local states ($\sigma > 1$). In the following we shall present prediction results for the direction $u \rightarrow (v, w, s)$ and $w \rightarrow (u, v, s)$. Similar results have been obtained for $v \rightarrow (u, w, s)$ and $s \rightarrow (u, v, w)$ (see MSE and NRMSE values in Table V). Figure 7 shows cross prediction results for the case $u \rightarrow (v, w, s)$. Snapshots of the original $v$, $w$ and $s$ variable of the BOCF system at the time step $n = 1000$ of the test data set are displayed in Fig. 7(a), 7(e) and 7(i). The corresponding predicted variables are shown next to them in Figs. 7(b), 7(f), 7(j) and the absolute errors of the prediction are displayed in Figs. 7(c), 7(g) and 7(k). The last column shows patterns of absolute errors for predictions based on noiseless $u$ data which are slightly better. As can be seen all essential features of the non-measured dynamics ($v$, $w$ and $s$) can be recovered from the measure $u$.
FIG. 6. Cross prediction of the $u$ variable of the Barkley model Eq. (1) from noisy $v$ data (see Fig. 4(b)) at time step $n = 1000$ of the test data set. (a) True $u$ values, (b) predicted $u$ values, (c) magnitude of difference between true and predicted $u$ values, (d) magnitude of difference between true and predicted $u$ values if noiseless $v$ data are used as input of the ESN.

time series even in the presence of noise. Similar but slightly worse results have been obtained for the cross prediction $w \rightarrow (u, v, s)$ as shown in Fig. 8. The hyperparameter values used for the different cross prediction tasks and the global MSE and NRMSE are listed in Table V. These results show that in all four cases the observed (noisy) variable contains all relevant information about the (chaotic) spatio-temporal dynamics of the BOCF model and that ESNs based on local states can exploit this information to recover the values and the dynamics of all three remaining dynamical variables of the model Eq. (3).
TABLE IV. Hyperparameters and errors for cross predictions of variables of the Barkley model Eq. (1).

<table>
<thead>
<tr>
<th></th>
<th>$u \rightarrow v$</th>
<th>$v \rightarrow u$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>$\Delta \sigma$</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>$N$</td>
<td>200</td>
<td>400</td>
</tr>
<tr>
<td>$\rho(\lvert W \rvert)$</td>
<td>1.10</td>
<td>1.50</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.05</td>
<td>0.20</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>$\nu_{\text{max}}$</td>
<td>$10^{-4}$</td>
<td>$10^{-5}$</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>$5 \cdot 10^{-6}$</td>
<td>$5 \cdot 10^{-6}$</td>
</tr>
</tbody>
</table>

MSE 0.000039 0.0036
NRMSE 0.021 0.15

VI. CONCLUSION

By combining the concepts of Reservoir Computing (Echo State Networks) and local states an efficient method has been devised for cross-predicting dynamical variables of PDEs describing chaotic two dimensional excitable media. Decomposing the (cross) prediction task into many local applications of ESNs (driven by spatially localized input) turned out to render the prediction task feasible and efficient (and might be exploited in future parallel implementations on dedicated hardware). This approach for (cross) prediction can not only be applied with noise free signals but proves to be quite robust and efficient also for noisy input driving the Echo State Networks. In contrast to conventional observers or data assimilation schemes model equations (here: the Barkley model and the BOCF model) are only required for generating training data but not for solving the cross prediction task. Therefore, this approach could also be used for applications where instead of a mathematical model only measured data are available. In this case, training data would be obtained during an extended measurement where in addition to the input data (for example, a membrane voltage) also target time series are measured (for example, some ionic currents using patch clamp methods). Once the relation between input and target has been learned the ESN
\[
\begin{array}{cccccccccccc}
\sigma & 1 & 3 & 3 & 5 & 1 & 5 & 7 & 7 & 7 & 3 & 3 & 3 \\
\Delta \sigma & 1 & 1 & 1 & 1 & 1 & 1 & 3 & 1 & 3 & 1 & 1 & 1 \\
N & 50 & 400 & 200 & 400 & 50 & 400 & 200 & 400 & 400 & 200 & 400 & 400 \\
\rho(\|W\|) & 1.10 & 0.95 & 0.95 & 1.1 & 0.95 & 1.1 & 1.5 & 3.0 & 1.5 & 0.95 & 0.95 & 1.1 \\
\alpha & 0.95 & 0.05 & 0.20 & 0.2 & 0.2 & 0.2 & 0.7 & 0.5 & 0.5 & 0.95 & 0.2 & 0.2 \\
\epsilon & 0.1 & 0.1 & 0.2 & 0.1 & 0.2 & 0.1 & 0.05 & 0.1 & 0.05 & 0.1 & 0.05 & 0.05 \\
\nu_{\text{max}} & 10^{-5} & 10^{-5} & 10^{-5} & 10^{-5} & 10^{-4} & 10^{-4} & 10^{-4} & 10^{-4} & 10^{-4} & 10^{-4} & 10^{-4} & 10^{-5} \\
\lambda & 5 \cdot 10^{-6} & 5 \cdot 10^{-3} & 5 \cdot 10^{-3} & 5 \cdot 10^{-2} & 5 \cdot 10^{-2} & 5 \cdot 10^{-6} & 5 \cdot 10^{-5} & 5 \cdot 10^{-5} & 5 \cdot 10^{-5} & 5 \cdot 10^{-5} & 5 \cdot 10^{-4} & 5 \cdot 10^{-3} \\
\end{array}
\]

TABLE V. Hyperparameters and errors for the cross prediction of the \(v\), \(w\) and the \(s\) variable of the BOCF model Eq. (3) using the \(u\) variable.

can predict (future) values of the target from corresponding input.

ACKNOWLEDGMENTS

We thank T. Lilienkamp, L. Thiede, and S. Luther for inspiring scientific discussions on excitable media, Reservoir Computing, and continuous support. UP also acknowledges support by the German Research Foundation (DFG) (Collaborative Research Center SFB 937 Project A18) and BMBF (grant no. FKZ031A147, GO-Bio).
FIG. 7. Cross prediction of the variables of the $v$, $w$ and $s$ of the BOCF model from a noisy $u$ time series (see Fig. 4(c)) with SNR = 23.7dB. Shown are the results at time step $n = 1000$ of the test data set. The first, second, and third row show results for $v$, $w$ and $s$, respectively. In the first column the true values are displayed. The second column shows the corresponding predictions and the third column the magnitude of the error (difference between first and second column). The last column shows for comparison prediction errors obtained with noiseless input $u$.

REFERENCES

FIG. 8. Cross prediction of the variables of the $u$, $w$ and $s$ of the BOCF model from a noisy $w$ time series (see Fig. 4(d)) with SNR = 21.8 dB. Shown are the results at time step $n = 1000$ of the test data set. The first, second, and third row show results for $u$, $v$ and $s$, respectively. In the first column the true values are displayed. The second column shows the corresponding predictions and the third column the magnitude of the error (difference between first and second column). The last column shows for comparison prediction errors obtained with noiseless input $w$.


