Tractable Dendritic RNNs for Reconstructing Nonlinear Dynamical Systems

Manuel Brenner *1 2 Florian Hess * 1 2 Jonas M. Mikhaeil 1 2 Leonard Bereska 3 Zahra Monfared 1
Po-Chen Kuo 4 Daniel Durstewitz 1 2

*contributed equally
1 Dept. of Theoretical Neuroscience, Central Institute of Mental Health, Mannheim, Germany
2 Faculty of Physics and Astronomy, Heidelberg University, Germany
3 University of Amsterdam, Netherlands
National Taiwan University, Taiwan.

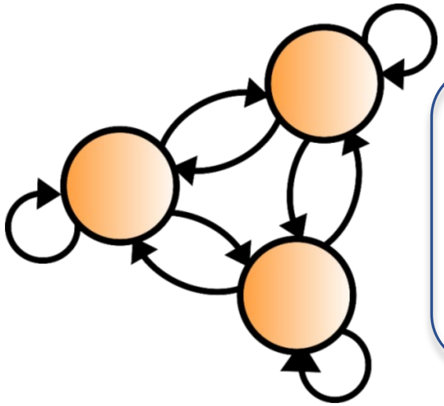
ICML
International Conference On Machine Learning

Dynamical Systems Reconstructions

Lorenz Eqns.

$$\dot{x}_1 = s(x_2 - x_1)$$
$$\dot{x}_2 = rx_1 - x_2 - x_1x_3$$
$$\dot{x}_3 = x_1x_2 - bx_3$$

Inference



Generic reconstruction model

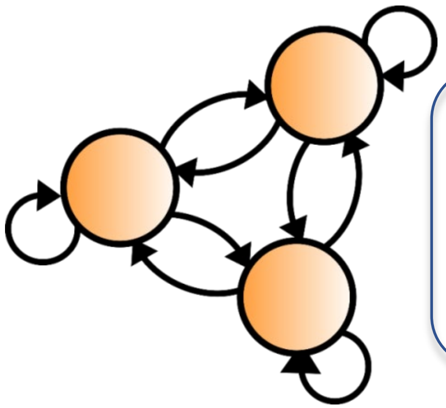
$$\boldsymbol{z}_t = \boldsymbol{F}_{\boldsymbol{\theta}}(\boldsymbol{z}_{t-1}, \boldsymbol{s}_t)$$
$$\mathbf{x}_t = g(\mathbf{z}_t)$$

Dynamical Systems Reconstructions

Lorenz Eqns.

$$\dot{x}_1 = s(x_2 - x_1)$$
$$\dot{x}_2 = rx_1 - x_2 - x_1x_3$$
$$\dot{x}_3 = x_1x_2 - bx_3$$

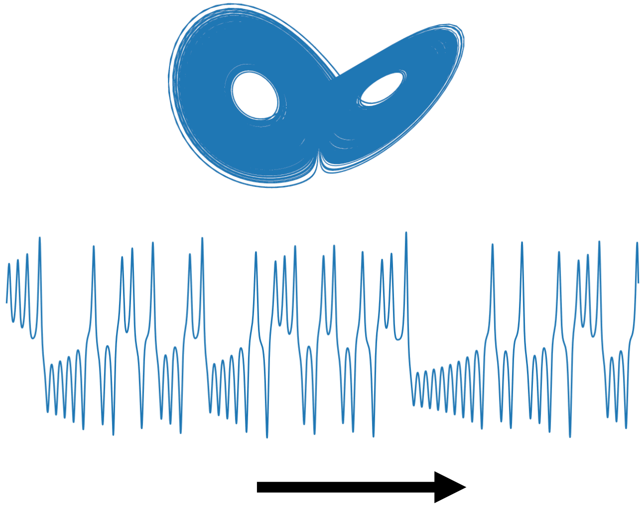
Inference



Generic reconstruction model

$$\mathbf{z}_t = \boldsymbol{F}_{\boldsymbol{\theta}}(\mathbf{z}_{t-1}, \boldsymbol{s}_t)$$
$$\mathbf{x}_t = g(\mathbf{z}_t)$$

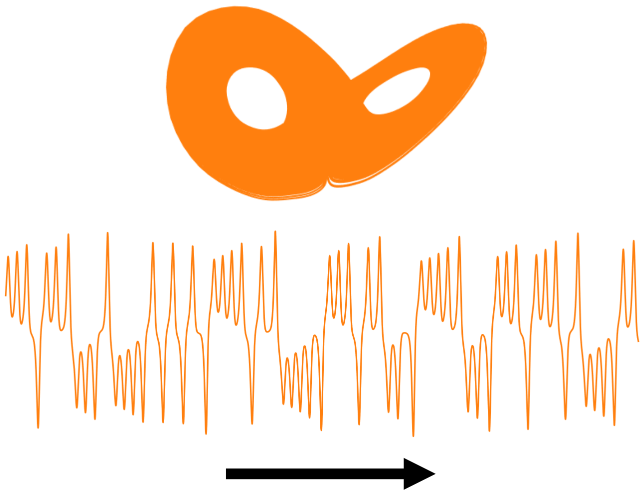
True trajectory



Time

Agreement in geometrical and temporal structure

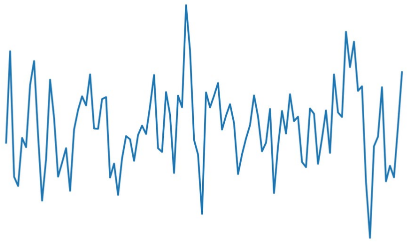
Simulated trajectory



Time

Challenges of Dynamical Systems Reconstructions

Noisy time series



Inference → **Model**

Mathematical tractability
Models often too complex for efficient DS analysis

Latent space size
Often high dimensional spaces required for successful reconstructions

Model assumptions
Models may require structural knowledge or are not dynamically universal

Piecewise Linear Recurrent Neural Network

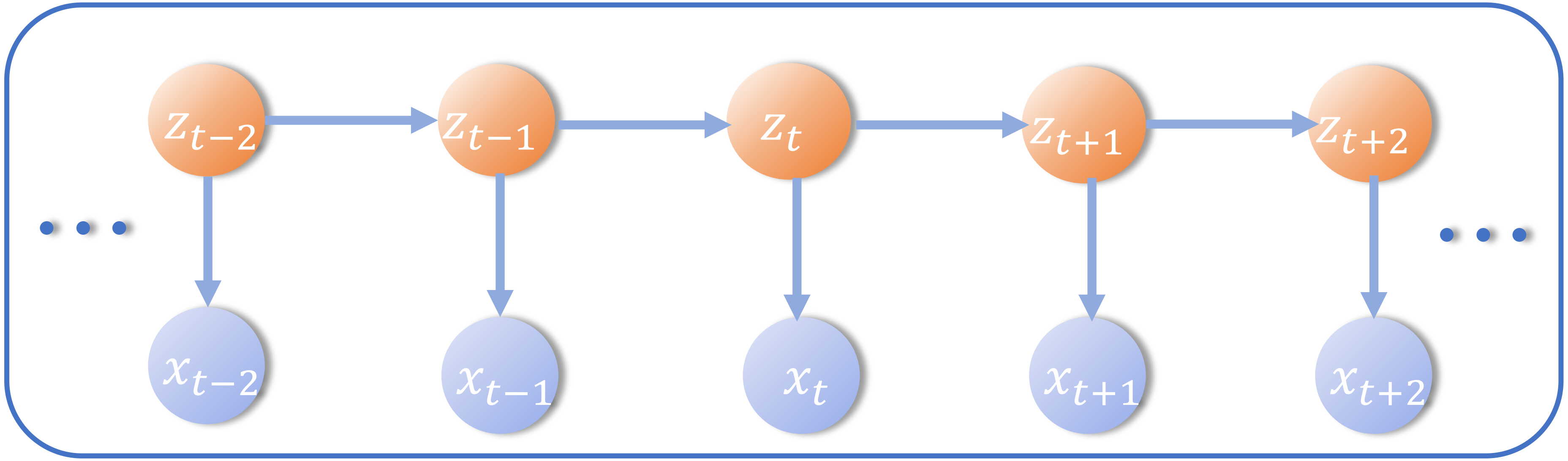
Latent Process

$$\boldsymbol{z}_t = \boldsymbol{A}\boldsymbol{z}_{t-1} + \boldsymbol{W}\phi(\boldsymbol{z}_{t-1}) + \mathcal{C}\boldsymbol{s}_t + \boldsymbol{h} + \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

Observations

$$\boldsymbol{x}_t = \boldsymbol{B}\boldsymbol{z}_t + \boldsymbol{\eta}_t, \boldsymbol{\eta}_t \sim \mathcal{N}(0, \boldsymbol{\Gamma})$$

$$\phi(\boldsymbol{z}_{t-1}) = \max(0, \boldsymbol{z}_{t-1})$$



Mathematical Tractability

Analytical computation of fixed points and k-cycles

Fixed points

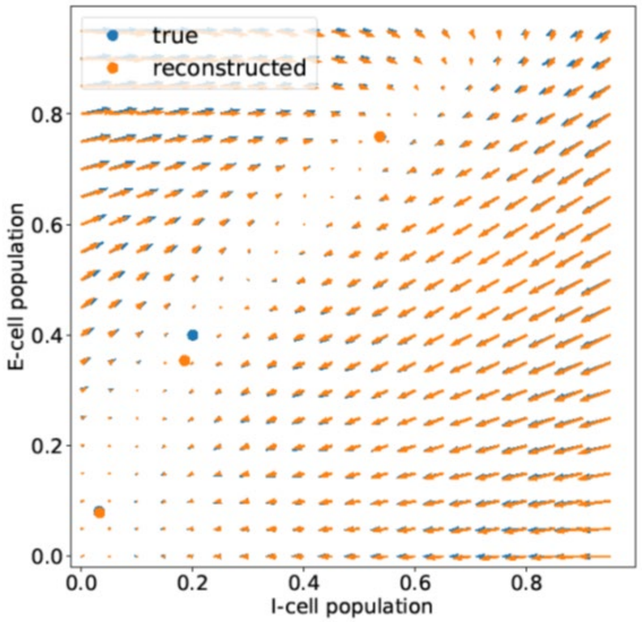
$$\boldsymbol{z}^{*1} = \left(\boldsymbol{I} - \boldsymbol{W}^{B}_{\Omega(t^{*1})}\right)^{-1}\left[\boldsymbol{W}\,\boldsymbol{h}^{B}_{\Omega(t^{*1})} + \boldsymbol{h}_0\right]$$

k-cycles

$$\boldsymbol{z}^{*n} = \left(\boldsymbol{I} - \prod_{i=1}^{n}\boldsymbol{W}^{B}_{\Omega(t^{*n-i})}\right)^{-1}\left(\sum_{j=2}^{n}\left[\prod_{i=1}^{n-j+1}\boldsymbol{W}^{B}_{\Omega(t^{*n-i})}\boldsymbol{W}\boldsymbol{h}^{B}_{\Omega(t^{*n-n+j-2})}\right] + \boldsymbol{W}\boldsymbol{h}^{B}_{\Omega(t^{*n-1})} + \left(\sum_{j=2}^{n}\prod_{i=1}^{n-j+1}\boldsymbol{W}^{B}_{\Omega(t^{*n-i})} + \boldsymbol{I}\right)\boldsymbol{h}_0\right),$$

Tractability

Wilson-Cowan model (neural population dynamics)



Mathematical Tractability

Analytical computation of fixed points and k-cycles

Fixed points

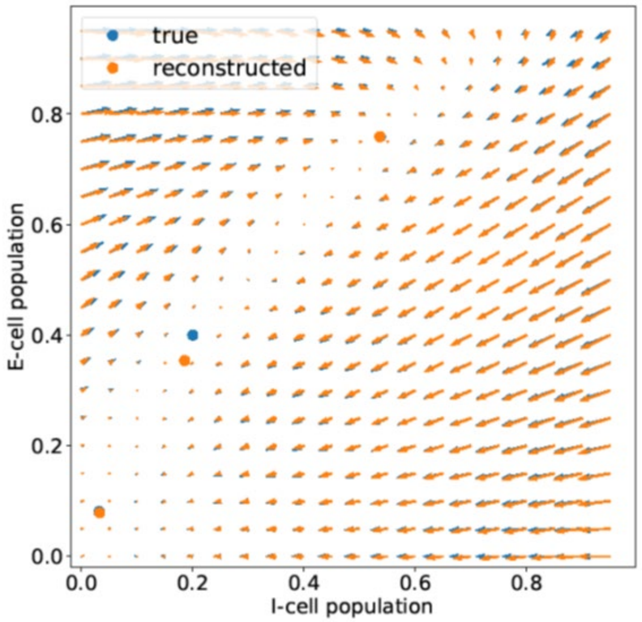
$$\boldsymbol{z}^{*1} = \left(\boldsymbol{I} - \boldsymbol{W}^{B}_{\Omega(t^{*1})}\right)^{-1} \left[\boldsymbol{W}\, \boldsymbol{h}^{B}_{\Omega(t^{*1})} + \boldsymbol{h}_0\right]$$

k-cycles

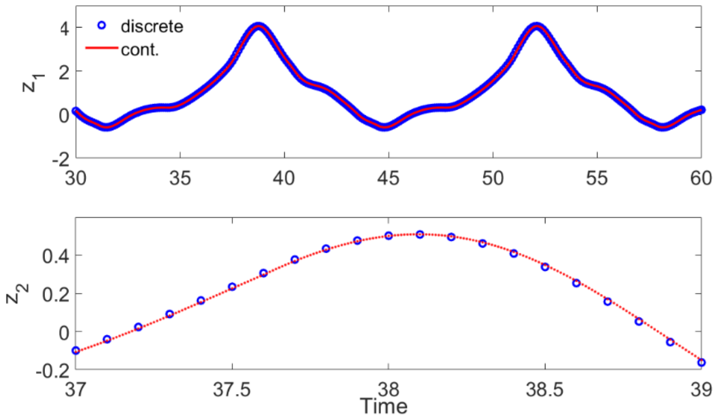
$$\boldsymbol{z}^{*n} = \left(\boldsymbol{I} - \prod_{i=1}^{n} \boldsymbol{W}^{B}_{\Omega(t^{*n-i})}\right)^{-1} \left(\sum_{j=2}^{n} \left[\prod_{i=1}^{n-j+1} \boldsymbol{W}^{B}_{\Omega(t^{*n-i})} \boldsymbol{W} \boldsymbol{h}^{B}_{\Omega(t^{*n-n+j-2})} \right] + \boldsymbol{W} \boldsymbol{h}^{B}_{\Omega(t^{*n-1})} + \left(\sum_{j=2}^{n} \prod_{i=1}^{n-j+1} \boldsymbol{W}^{B}_{\Omega(t^{*n-i})} + \boldsymbol{I} \right) \boldsymbol{h}_0 \right),$$

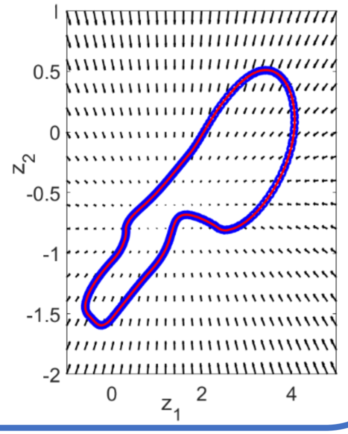
Tractability

Wilson-Cowan model (neural population dynamics)



Transformation into equivalent continuous time ODE systems (Monfared & Durstewitz, ICML 2020)

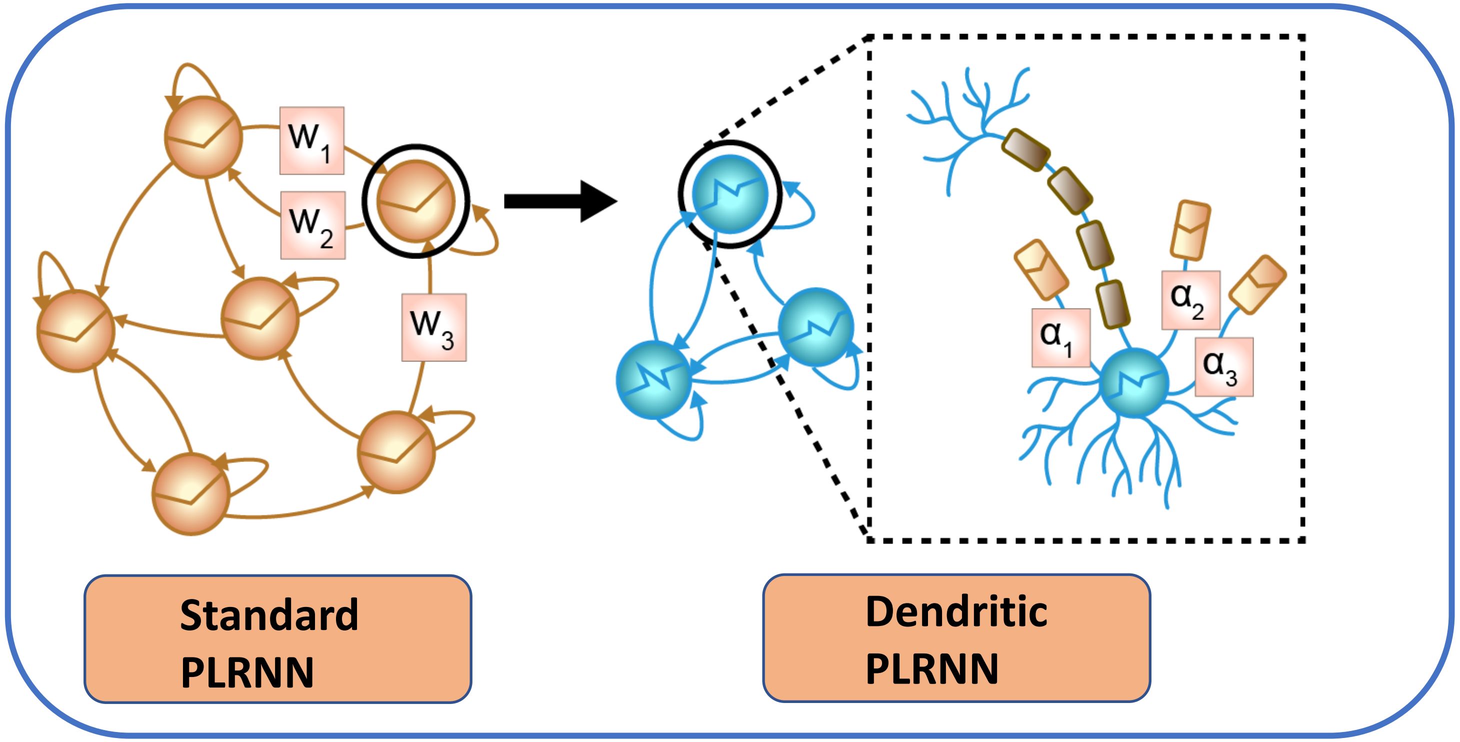




Dendritic PLRNN

Dendritic computation as linear spline basis expansion

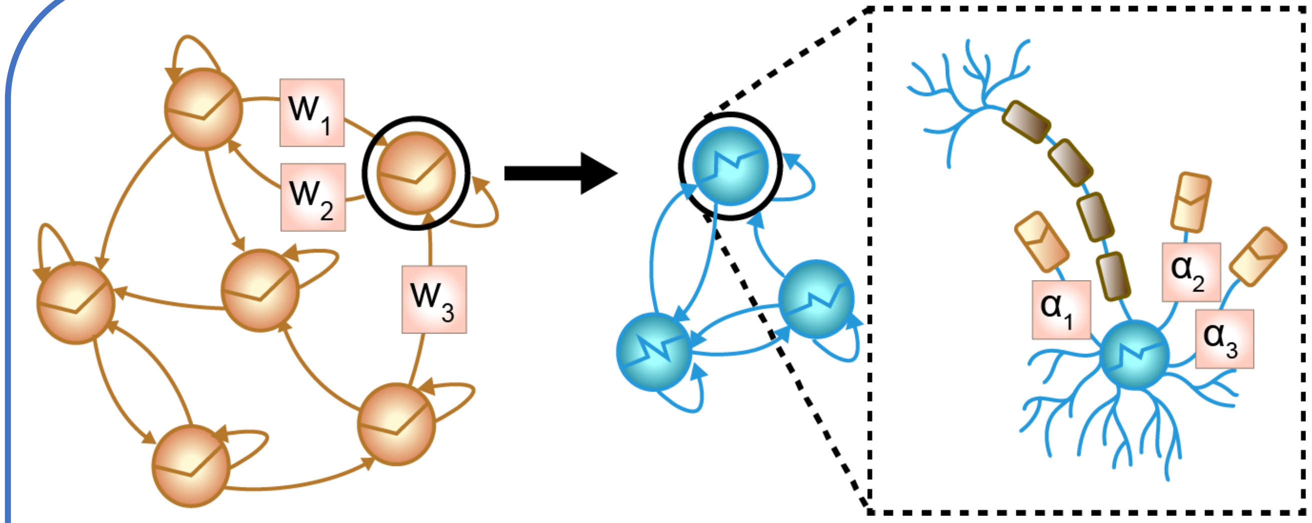
$$\phi(\boldsymbol{z}_{t-1}) = \max(0, \boldsymbol{z}_{t-1}) \longrightarrow \phi(\boldsymbol{z}_{t-1}) = \sum_{b=1}^{B} \alpha_b \max(0, \boldsymbol{z}_{t-1} - \boldsymbol{h}_b),$$



Dendritic PLRNN

Dendritic computation as linear spline basis expansion

$$\phi(\boldsymbol{z}_{t-1}) = \max(0, \boldsymbol{z}_{t-1}) \longrightarrow \phi(\boldsymbol{z}_{t-1}) = \sum_{b=1}^{B} \alpha_b \max(0, \boldsymbol{z}_{t-1} - \boldsymbol{h}_b),$$

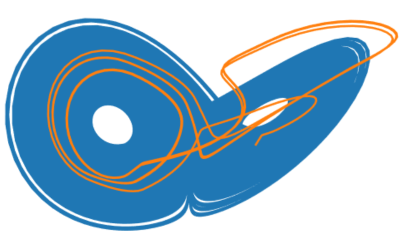


Standard PLRNN

Dendritic PLRNN

Successful reconstructions in much lower dimensions

M=5, PLRNN



M=10, PLRNN

M=5, B=20, dendPLRNN

Theoretical Results

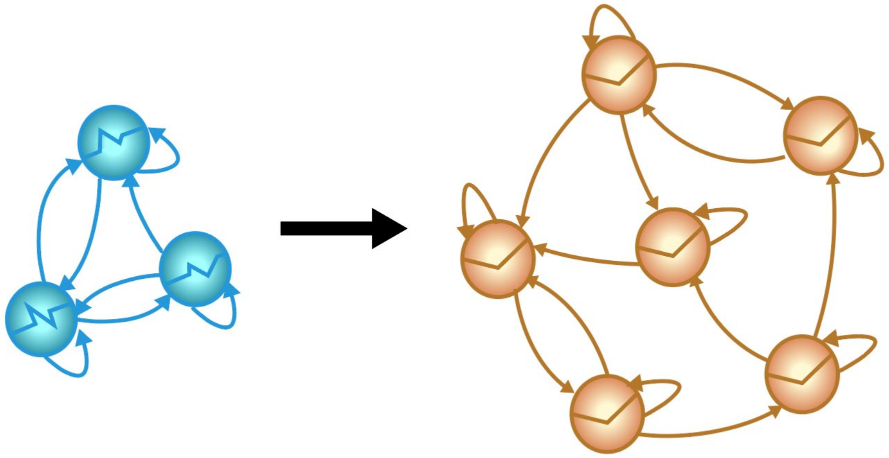
Dimensionality reduction & preserved tractability

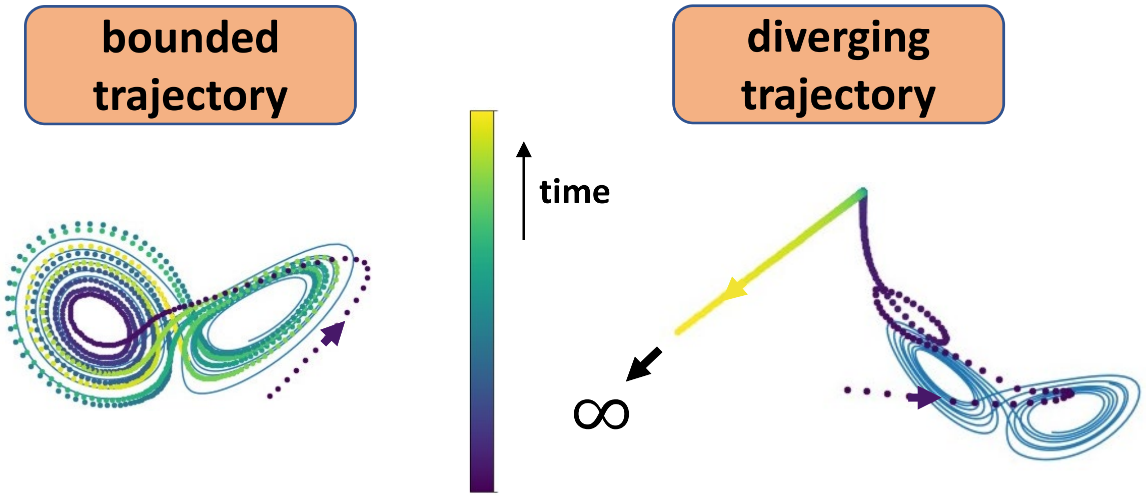
Bounded orbits

Theorem 1. *Any M-dimensional dendPLRNN as defined in Eqs. 1, 4, can always be rewritten as a $M \times B$-dimensional "conventional" PLRNN of the form*

$$\hat{\boldsymbol{z}}_t = \tilde{\boldsymbol{A}} \hat{\boldsymbol{z}}_{t-1} + \tilde{\boldsymbol{W}} \max(0, \hat{\boldsymbol{z}}_{t-1}) + \hat{\boldsymbol{h}}_0 + \tilde{\boldsymbol{C}} \boldsymbol{s}_t + \tilde{\boldsymbol{\epsilon}}_t.$$

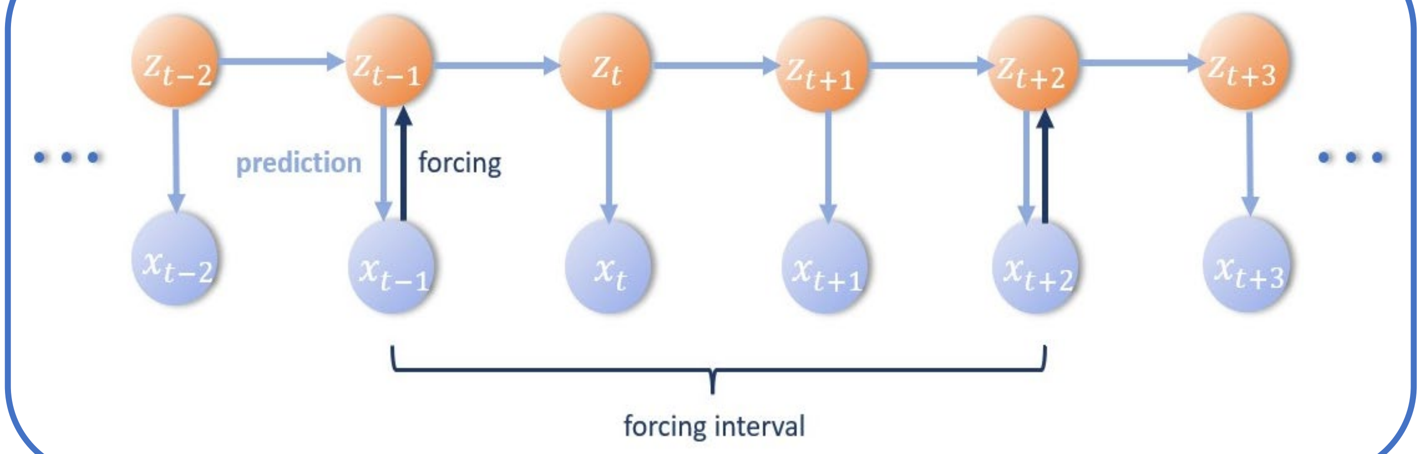
Theorem 2. *For each basis $\{\alpha_b, \boldsymbol{h}_b\}$ in Eq. 4 of a dendPLRNN let us add another basis $\{\alpha_b^*, \boldsymbol{h}_b^*\}$ with $\alpha_b^* = -\alpha_b$ and $\boldsymbol{h}_b^* = \boldsymbol{0}$. Then, for $\sigma_{\max}(\boldsymbol{A}) < 1$, any orbit of this "clipped" dendPLRNN (Eq. 10) will remain bounded.*



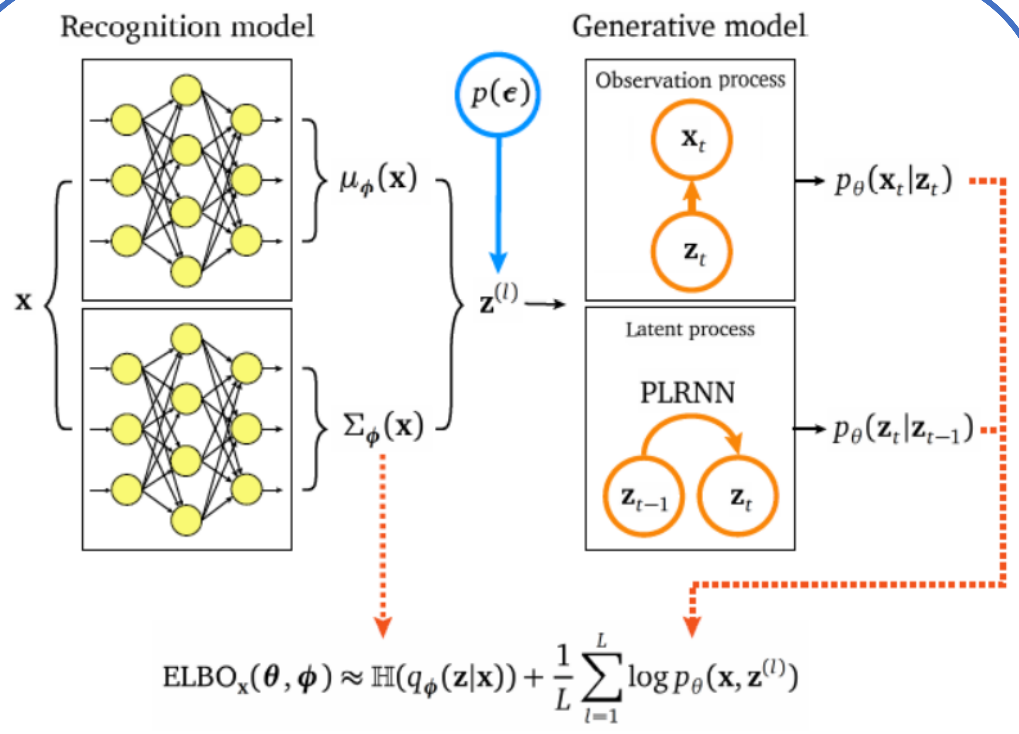


Training Methods

Backpropagation through time with sparse teacher forcing (TF)



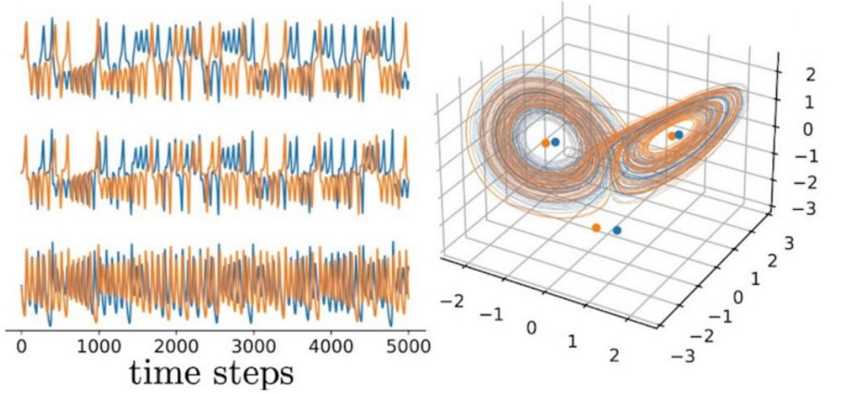
Sequential Variational Autoencoder



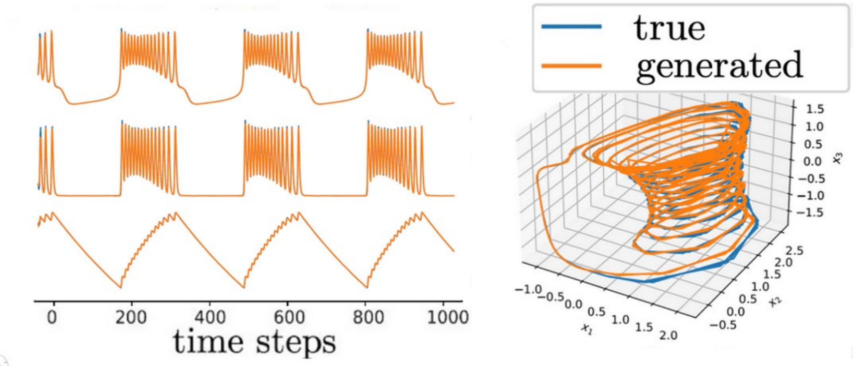
(fig. by Unai Fischer Abaigar)

Simulated Benchmark Systems

Lorenz-63 (atmospheric convection)

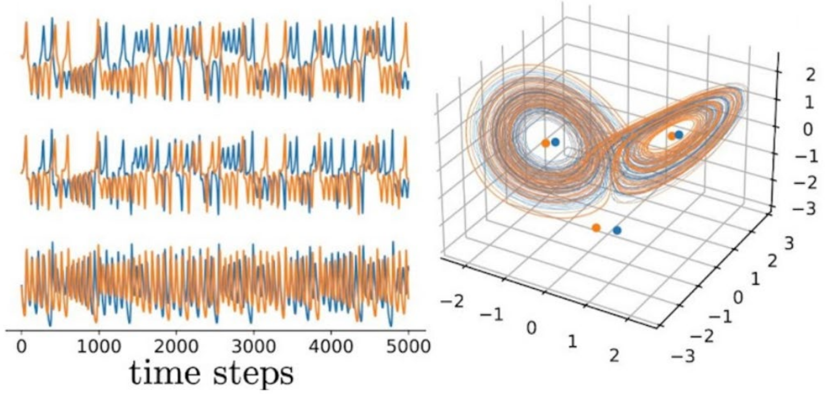


Biophysical bursting neuron model

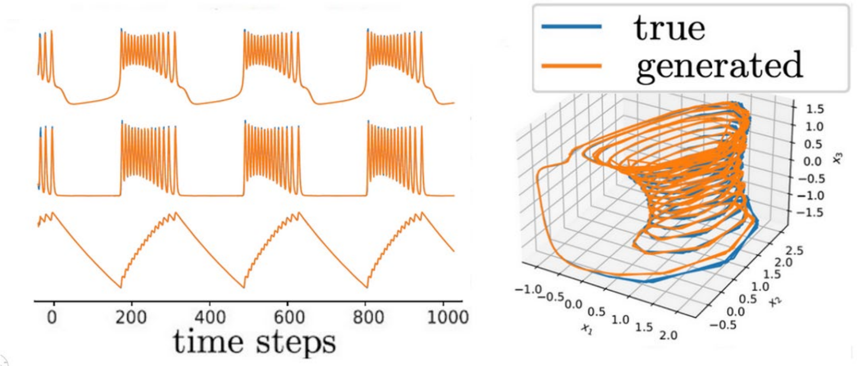


Simulated Benchmark Systems

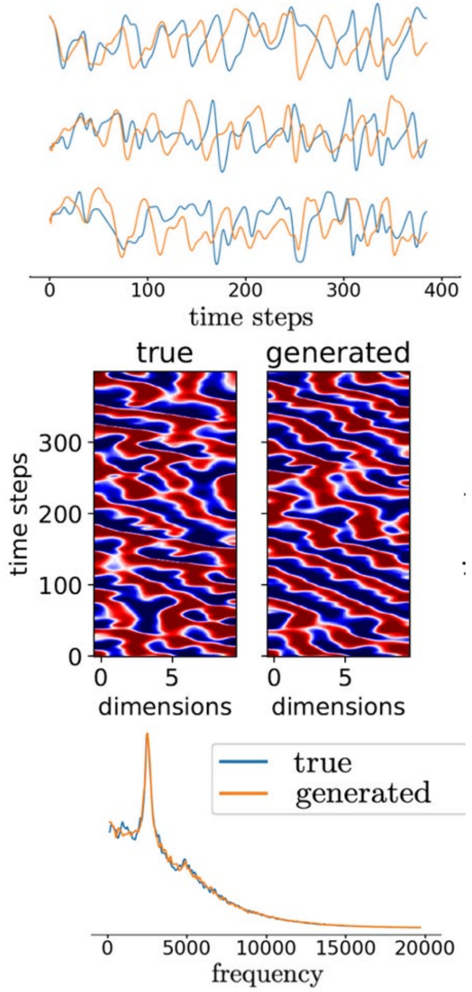
Lorenz-63 (atmospheric convection)



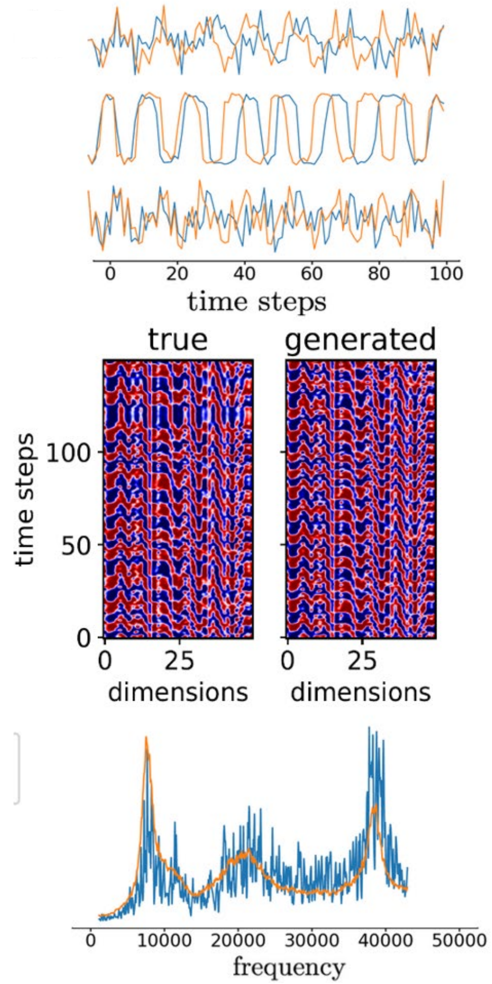
Biophysical bursting neuron model



Lorenz-96 weather model

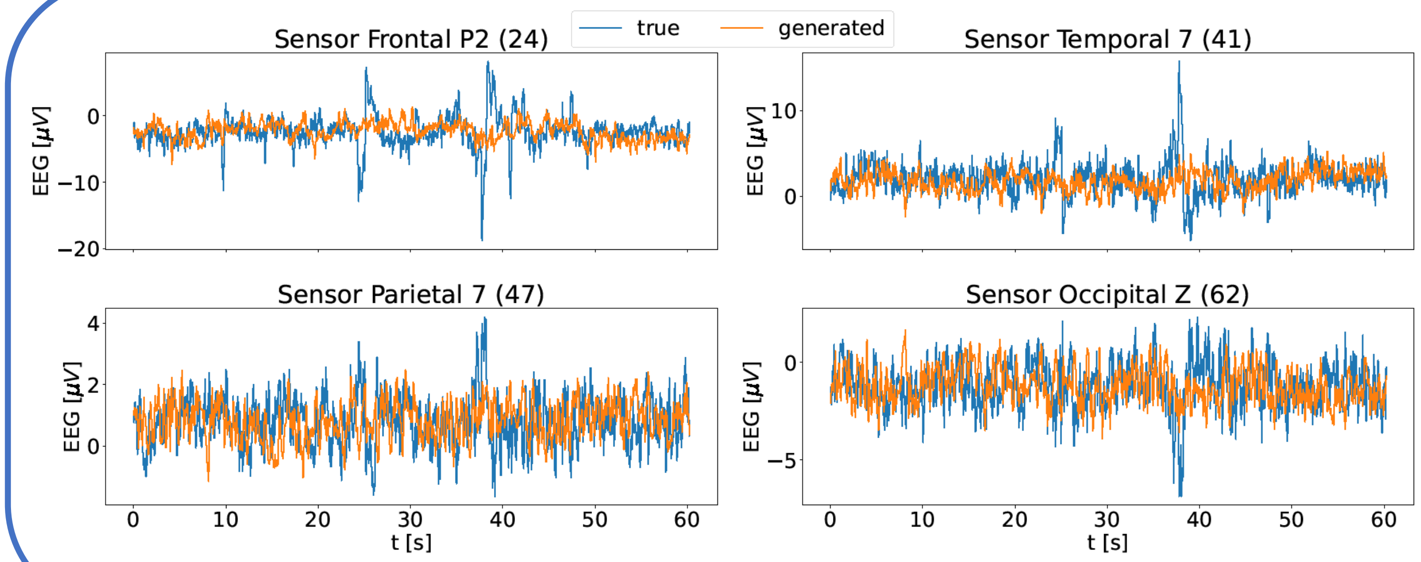


Neural population dynamics

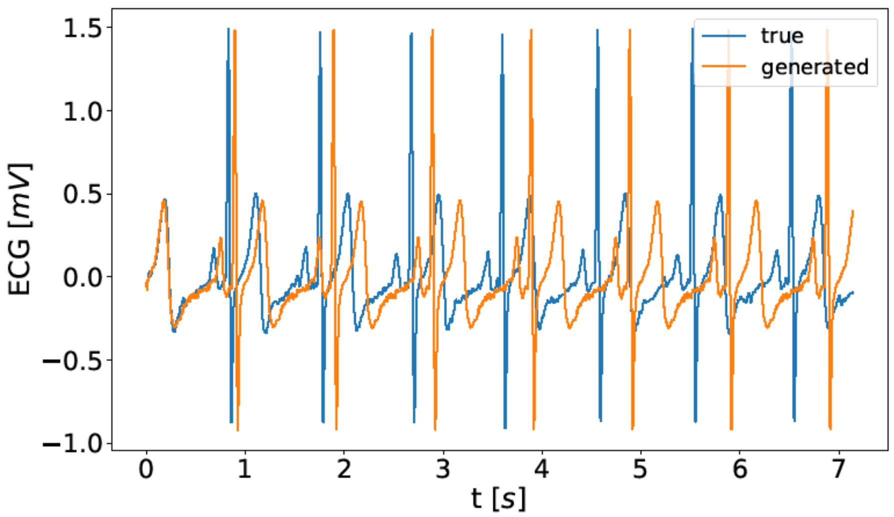


Empirical Data

Experimental EEG Data



Experimental ECG Data



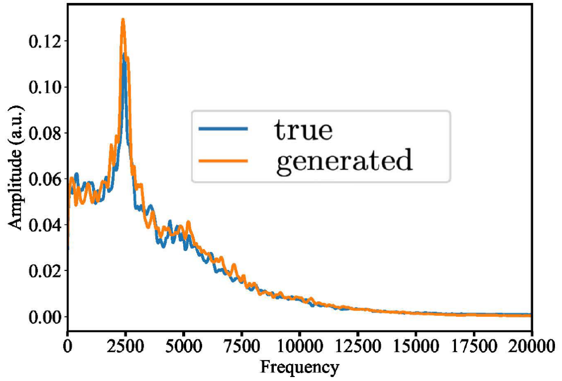
Comparisons

Highly competitive performance

Dataset	Method	PSC	D_{stsp}	20-step PE	Dyn.var.	#parameters
Lorenz-63	dendPLRNN TF	0.997 ± 0.002	0.13 ± 0.18	9.2e−5 ± 2.8e−5	22	1032
	RC	0.991 ± 0.001	0.24 ± 0.05	1.2e−2 ± 0.1e−3	345	1053
	LSTM-MSM	0.985 ± 0.004	0.85 ± 0.07	1.2e−2 ± 0.1e−3	29	1035
	SINDy	**0.998 ± 0.0003**	**0.04 ± 0.01**	**6.8e−5 ± 0.2e−5**	3	252
	Neural ODE	0.992 ± 0.001	0.149 ± 0.014	1.1e−3 ± 4.1e−5	3	1011
Bursting Neuron	dendPLRNN TF	**0.76 ± 0.04**	**0.61 ± 0.09**	6.1e−2 ± 2.2e−2	26	2040
	RC	0.51 ± 0.01	5.1 ± 0.6	8.6e−2 ± 0.1e−2	711	2133
	LSTM-MSM	0.54 ± 0.02	2.83 ± 0.36	**3.9e−2 ± 0.1e−2**	45	2166
	SINDy	0.25 ± 0.01	6.36 ± 0.02	5.4e−1 ± 0.1e−2	3	252
	Neural ODE	0.65 ± 0.017	3.85 ± 0.1	2.1e−1 ± 0.5e−2	3	2073
Lorenz-96	dendPLRNN TF	**0.998 ± 0.0001**	**0.04 ± 0.01**	**4.1e−2 ± 0.8e−2**	50	4480
	RC	0.986 ± 0.008	0.25 ± 0.17	7.1e−1 ± 0.1e−2	440	4400
	LSTM-MSM	0.993 ± 0.002	0.23 ± 0.03	8.2e−1 ± 0.3e−2	62	4384
	SINDy	0.997 ± 0.001	0.06 ± 0.003	6.3e−2 ± 0.1e−3	10	27410
	Neural ODE	0.985 ± 0.001	0.21 ± 0.02	4.4e−2 ± 4.5e−3	10	4130
Neural Population Model	dendPLRNN TF	**0.52 ± 0.01**	0.37 ± 0.05	1.43 ± 0.01	75	9990
	RC	0.34 ± 0.03	2.8 ± 0.4	1.64 ± 0.07	200	10000
	LSTM-MSM	0.51 ± 0.02	**0.29 ± 0.04**	1.56 ± 0.01	56	10298
	SINDy	diverging	diverging	diverging	50	66300
	Neural ODE	0.47 ± 0.03	9.56 ± 0.86	**0.58 ± 0.006**	50	10200
EEG	dendPLRNN TF	**0.923 ± 0.012**	**1.96 ± 0.18**	**0.202 ± 0.007**	128	27058
	RC	0.782 ± 0.002	8.8 ± 0.8	0.78 ± 0.02	448	28672
	LSTM-MSM	0.827 ± 0.002	8.3 ± 0.3	0.708 ± 0.003	168	27728
	SINDy	diverging	diverging	diverging	64	133120
	Neural ODE	0.82 ± 0.002	21.72 ± 0.71	0.31 ± 0.005	64	30559
ECG	dendPLRNN TF	**0.929 ± 0.014**	**0.4 ± 0.6**	0.23 ± 0.03	30	2641
	RC	0.880 ± 0.013	1.78 ± 0.44	0.571 ± 0.013	378	2646
	LSTM-MSM	0.926 ± 0.007	0.59 ± 0.08	**7.0e−2 ± 0.6e−2**	51	2801
	SINDy	diverging	diverging	diverging	7	4424
	Neural ODE	0.90 ± 0.011	1.18 ± 0.02	0.61 ± 0.01	7	2599

Agreement in temporal structure

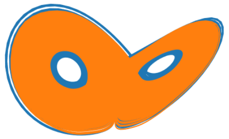
Power spectral correlation



Geometrical agreement

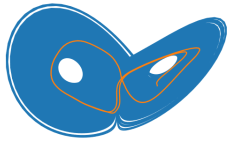
Kullback Leibler divergence in state space

Low KL



Medium KL

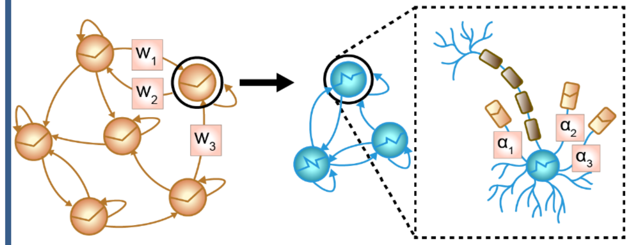
High KL



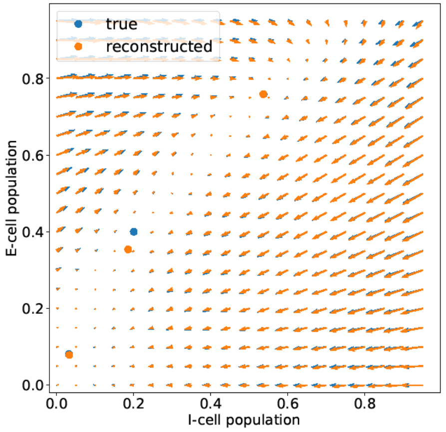
$$D_{\text{stsp}}\left(p_{\text{true}}(\boldsymbol{x}), p_{\text{gen}}(\boldsymbol{x} \mid \boldsymbol{z})\right) \approx \sum_{k=1}^{K} \hat{p}_{\text{true}}^{(k)}(\boldsymbol{x}) \log\left(\frac{\hat{p}_{\text{true}}^{(k)}(\boldsymbol{x})}{\hat{p}_{\text{gen}}^{(k)}(\boldsymbol{x} \mid \boldsymbol{z})}\right)$$

Conclusions

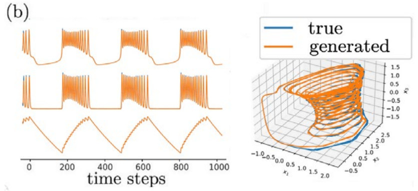
Reduced dimensionality through dendritic basis expansion

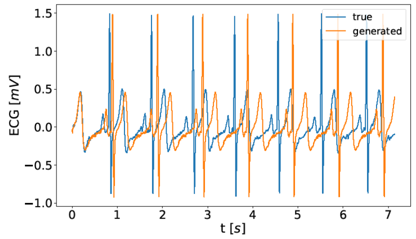


Analytical tractability of important DS properties preserved in dendPLRNN



State of the art performance on benchmark & experimental datasets





Code: https://github.com/DurstewitzLab/dendPLRNN

STRUCTURES CLUSTER OF EXCELLENCE

Deutsche Forschungsgemeinschaft DFG

IMMERSE

This work was funded by the German Research Foundation (DFG) within Germany's Excellence Strategy – EXC-2181 – 390900948 ('Structures'), by DFG grant Du354/10-1 to DD, and the European Union Horizon-2020 consortium SC1-DTH-13-2020 ('IMMERSE').