Is the Brain Macroscopically Linear?

A System Identification of Resting State Dynamics

Erfan Nozari

University of California, Riverside

Joint work with Jeni Stiso, Lorenzo Caciagli, Eli Cornblath,
Xiaosong He, Max Bertolero, Arun Mahadevan,
George Pappas, and Dani Bassett

Summer School 2021
Current and Future Applications of Network and Control Sciences for Psychiatry
September 30, 2021

" Truth ... is much too complicated to allow anything but approximation "

— von Neumann

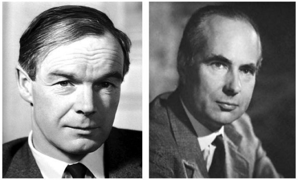
Common points of **tradeoff**:

- **Linear vs. nonlinear**
- Small vs. large dimensional
- Deterministic vs. stochastic
- Stationary vs. time-varying
- ⋮

Nonlinearity: Essential at Microscale

- Fundamental to the HH model

$$
\begin{aligned}
\dot{V}_m &= g_l(V_l - V_m) + g_{\mathrm{K}} n^4 (V_{\mathrm{K}} - V_m) + g_{\mathrm{Na}} m^3 h (V_{\mathrm{Na}} - V_m) + I \\
\dot{n} &= \frac{0.01(10 - V_m)}{e^{1 - V_m/10} - 1}(1 - n) + \frac{e^{-V_m/80}}{8} n \\
\dot{m} &= \frac{0.1(25 - V_m)}{e^{2.5 - V_m/10} - 1}(1 - m) + 4e^{-V_m/18} m \\
\dot{h} &= 0.07 e^{-V_m/20}(1 - h) + \frac{1}{e^{3 - V_m/10} + 1} h
\end{aligned}
$$

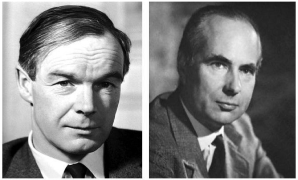


A. L. Hodgkin & A. F. Huxley

Nonlinearity: Essential at Microscale

- Fundamental to the HH model

$$
\begin{aligned}
\dot{V}_m &= g_l(V_l - V_m) + g_{\mathrm{K}} n^4 (V_{\mathrm{K}} - V_m) + g_{\mathrm{Na}} m^3 h (V_{\mathrm{Na}} - V_m) + I \\
\dot{n} &= \frac{0.01(10 - V_m)}{e^{1 - V_m/10} - 1}(1 - n) + \frac{e^{-V_m/80}}{8} n \\
\dot{m} &= \frac{0.1(25 - V_m)}{e^{2.5 - V_m/10} - 1}(1 - m) + 4e^{-V_m/18} m \\
\dot{h} &= 0.07 e^{-V_m/20}(1 - h) + \frac{1}{e^{3 - V_m/10} + 1} h
\end{aligned}
$$



A. L. Hodgkin & A. F. Huxley

✓ **Data-driven**

✓ Functionally, nonlinearity is **essential** for

- Excitable behavior (spiking)
- Limit cycles (rhythmic spiking)
- Logical operations (and, or)
- Bistability
- ⋮

- Seemingly, a network interconnection of billions of nonlinear neurons only explodes in complexity and repertoire of nonlinear behavior!

$\Rightarrow$ Resulting **assumption**: **accurate models** of macroscopic neurodynamics **must be nonlinear**

Nonlinearity: Present at Macroscale?

- Seemingly, a network interconnection of billions of nonlinear neurons only explodes in complexity and repertoire of nonlinear behavior!

$\Rightarrow$ Resulting **assumption**: **accurate models** of macroscopic neurodynamics **must be nonlinear**

$$\tau \frac{dE}{dt} = -E(t) + \big(1 - rE(t)\big) f_E\big[w_{EE}E - w_{EI}I + h_E(t)\big],$$
$$\tau \frac{dI}{dt} = -I(t) + \big(1 - rI(t)\big) f_I\big[w_{IE}E - w_{II}I + h_I(t)\big],$$

(Wilson & Cowan, 1972)

Nonlinearity: Present at Macroscale?

- Seemingly, a network interconnection of billions of nonlinear neurons only explodes in complexity and repertoire of nonlinear behavior!

⇒ Resulting **assumption**: **accurate models** of macroscopic neurodynamics **must be nonlinear**

$$
\begin{aligned}
\dot{y}_0(t) &= y_3(t)\\
\dot{y}_3(t) &= Aa\,\mathrm{Sigm}[y_1(t) - y_2(t)] - 2ay_3(t) - a^2 y_0(t)\\
\dot{y}_1(t) &= y_4(t)\\
\dot{y}_4(t) &= Aa\{p(t) + C_2\,\mathrm{Sigm}[C_1 y_0(t)]\}\\
&\quad - 2ay_4(t) - a^2 y_1(t)\\
\dot{y}_2(t) &= y_5(t)\\
\dot{y}_5(t) &= Bb\{C_4\,\mathrm{Sigm}[C_3 y_0(t)]\} - 2by_5(t) - b^2 y_2(t)
\end{aligned}
$$

(Jansen & Rit, 1995)

Nonlinearity: Present at Macroscale?

- Seemingly, a network interconnection of billions of nonlinear neurons only explodes in complexity and repertoire of nonlinear behavior!

⇒ Resulting **assumption**: **accurate models** of macroscopic neurodynamics **must be nonlinear**

$$\dot{\theta}_n = \omega_n + \frac{1}{N}\sum_{m=1}^{N} K_{mn}\sin(\theta_m - \theta_n)$$

(Kuramoto, 1984)

Nonlinearity: Present at Macroscale?

- Seemingly, a network interconnection of billions of nonlinear neurons only explodes in complexity and repertoire of nonlinear behavior!

⇒ Resulting **assumption**: **accurate models** of macroscopic neurodynamics **must be nonlinear**

$$\dot{x}_i = y_i - a x_i^3 + b x_i^2 - z_i + \left[K_{11}\left(X_1 - x_i\right) - K_{12}\left(X_2 - x_i\right)\right] + IE_i$$
$$\dot{y}_i = c_i - d x_i^2 - y_i$$
$$\dot{z}_i = rsx_i - rz_i - m_i$$

(Stefanescu-Jirsa, 2008)

Nonlinearity: Present at Macroscale?

- Seemingly, a network interconnection of billions of nonlinear neurons only explodes in complexity and repertoire of nonlinear behavior!

⇒ Resulting **assumption**: **accurate models** of macroscopic neurodynamics **must be nonlinear**

$$\dot{x}_i = y_i - a x_i^3 + b x_i^2 - z_i + \left[K_{11}\left(X_1 - x_i\right) - K_{12}\left(X_2 - x_i\right)\right] + IE_i$$
$$\dot{y}_i = c_i - d x_i^2 - y_i$$
$$\dot{z}_i = r s x_i - r z_i - m_i$$

(Stefanescu-Jirsa, 2008)

- **Is this true?**

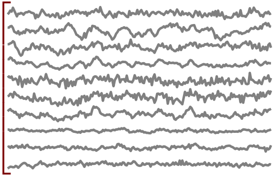
System Identification/Learning of Resting-State Dynamics

Generally-nonlinear ODE models are widespread in comp neuro:

$$\mathcal{M}: \quad \begin{aligned} \dot{\mathbf{x}}(t) &= f(\mathbf{x}(t)) + \mathbf{e}_1(t), \qquad \mathbf{x}(t_0) = \mathbf{x}_0 \\ \mathbf{y}(t) &= h(\mathbf{x}(t)) + \mathbf{e}_2(t) \end{aligned}$$

- $\mathbf{y}(t)$: neuroimaging time series
- $\mathbf{x}(t)$: "internal" state vector (any dimension, could be $= \mathbf{y}(t)$)
- $\mathbf{e}_1(t)$: process noise
- $\mathbf{e}_2(t)$: scanner noise

$$\mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{bmatrix} =$$



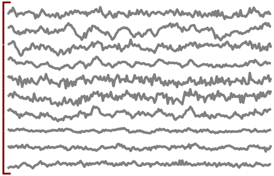
Generally-nonlinear ODE models are widespread in comp neuro:

$$\mathcal{M}: \quad \begin{cases} \dot{\mathbf{x}}(t) = f(\mathbf{x}(t)) + \mathbf{e}_1(t), & \mathbf{x}(t_0) = \mathbf{x}_0 \\ \mathbf{y}(t) = h(\mathbf{x}(t)) + \mathbf{e}_2(t) & \end{cases}$$

- $\mathbf{y}(t)$: neuroimaging time series
- $\mathbf{x}(t)$:
- $\mathbf{e}_1(t)$:
- $\mathbf{e}_2(t)$: scanner noise

Problem: find the model that best fits the data!

$$\mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{bmatrix} =$$



System Identification: Prediction Error Framework

- Philosophy: the better a model can **predict** the future of time series from its past ("initial conditions"), the more **accurate** it is

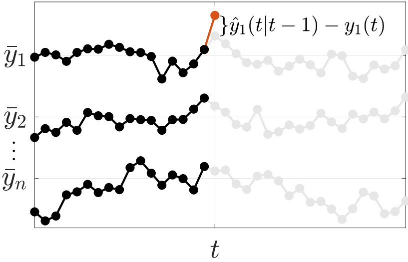
- Simplest: one-step ahead prediction (OSAP)

$$\min_{\mathcal{M}} \sum_{i,t} \left[\hat{y}_i(t|t-1) - y_i(t)\right]^2$$

- Model comparison:

  1. $R_i^2 = 1 - \dfrac{\sum_t [\hat{y}_i(t|t-1) - y_i(t)]^2}{\sum_t [\bar{y}_i - y_i(t)]^2} \leq 1$

  2. Whiteness of residuals (χ^2 test)

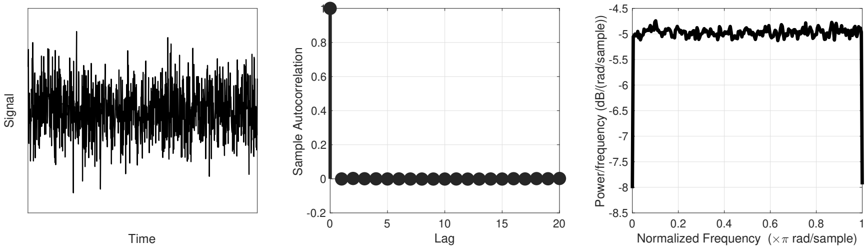


Whiteness of Residuals

- Prediction error (out of sample):

$$\boldsymbol{\varepsilon}(t) = \hat{\mathbf{y}}(t|t-1) - \mathbf{y}(t)$$

- Existence of **"dynamics"** in a signal $\equiv$ **non-white spectrum**



- Can be statistically verified via a χ^2 test: **the smaller the test statistic Q, the better**

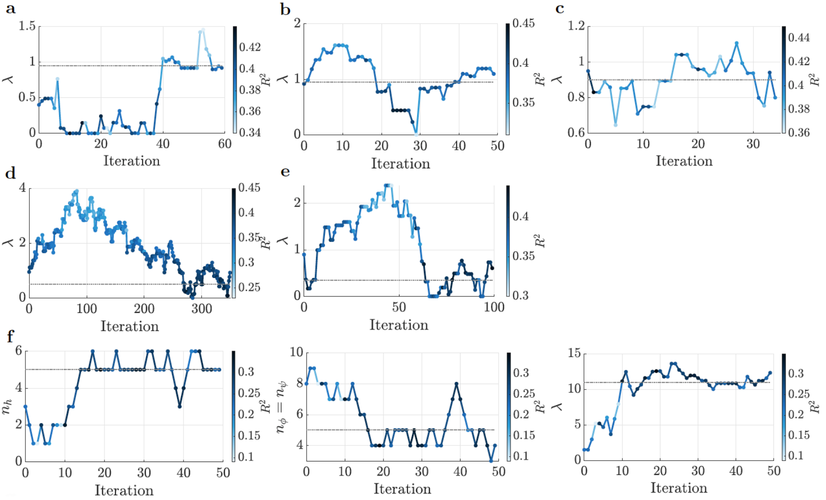
Linear Models

Label	Title	Equation		Parameters
Linear (dense)	Linear models with states at the BOLD/LFP level	$\mathbf{y}(t) - \mathbf{y}(t-1) = \mathbf{W}\mathbf{y}(t-1) + \mathbf{e}(t)$	(4)	None
Linear (sparse)				$\lambda = 0.95$ (fMRI) $\lambda = 1.2$ (iEEG)
AR-2 (sparse)†	Linear autoregressive models	$\mathbf{y}(t) - \mathbf{y}(t-1) = \mathbf{W}\mathbf{y}(t-1) + \mathbf{D}_2\mathbf{y}(t-2) + \mathbf{D}_3\mathbf{y}(t-3) + \cdots + \mathbf{D}_d\mathbf{y}(t-d) + \mathbf{e}(t)$	(5)	$d = 2, \lambda = 0.95$, diagonal $\mathbf{D}_2$
VAR-2 (sparse)†				$d = 2, \lambda = 0.9$
AR-3 (sparse)†				$d=3, \lambda=0.5$, diagonal $\mathbf{D}_2, \mathbf{D}_3$
VAR-3 (sparse)†				$d = 3, \lambda = 0.35$
AR-100 (sparse)‡				$d = 100, \lambda = 1.5$
AR-100 (scalar)‡				$d = 102$
Linear w/ HRF†	Linear models with states at the neural level	$\mathbf{x}(t) - \mathbf{x}(t-1) = \mathbf{W}\mathbf{x}(t-1) + \mathcal{G}_1(q)\hat{\mathbf{e}}_1(t)$ $\mathbf{y}(t) = \mathcal{H}(q)\mathbf{x}(t) + \mathcal{G}_2(q)\hat{\mathbf{e}}_2(t)$	(6a)	$n_h = n_\phi = n_\psi = 5, \lambda = 11$
		$\mathcal{H}(q) = \sum_{p=1}^{n_h} \mathrm{diag}(\mathbf{H}_{:,p})q^{-p}$	(6b)	
		$\mathcal{F}_1(q) = \mathbf{I} - \mathcal{G}_1^{-1}(q) = \sum_{p=1}^{n_\phi} \mathrm{diag}(\mathbf{\Phi}_{:,p})q^{-p}$	(6c)	
		$\mathcal{F}_2(q) = \mathbf{I} - \mathcal{G}_2^{-1}(q) = \sum_{p=1}^{n_\psi} \mathrm{diag}(\mathbf{\Psi}_{:,p})q^{-p}$	(6d)	
Subspace	Linear models with abstract data-driven states	$\mathbf{x}(t) - \mathbf{x}(t-1) = \mathbf{W}\mathbf{x}(t-1) + \mathbf{e}_1(t)$ $\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{e}_2(t)$ $\mathrm{Cov}\left(\begin{bmatrix}\mathbf{e}_1(t)\\ \mathbf{e}_2(t)\end{bmatrix}\right) = \begin{bmatrix}\mathbf{Q} & \mathbf{M}\\ \mathbf{M}^T & \mathbf{R}\end{bmatrix}$	(7)	s = 1, r = 3, n = 25 (fMRI) s = 11, r = 49, n = 436 (iEEG)
Zero	Zero model	$\mathbf{y}(t) - \mathbf{y}(t-1) = \mathbf{e}(t)$		None

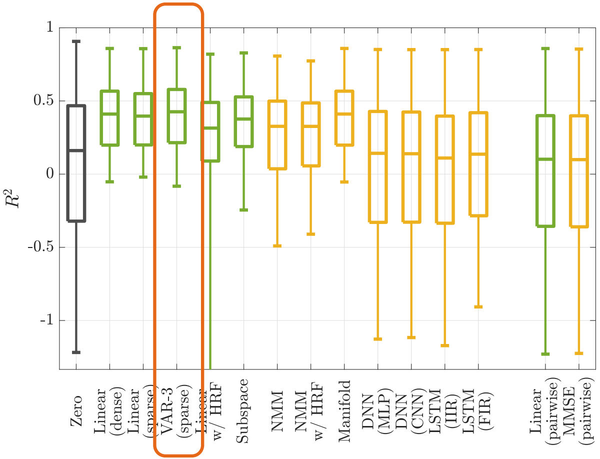
Nonlinear Models

Label	Title	Equation	Parameters
NMM	Nonlinear neural mass models	$\mathbf{y}(t)-\mathbf{y}(t-1)=(\mathbf{W}\psi_{\boldsymbol{\alpha}}(\mathbf{y}(t-1))-\mathbf{D}\mathbf{y}(t-1))\Delta_T+\mathbf{e}(t)$	MINDy default (fMRI) $\lambda_1=\lambda_2=0.2, \lambda_3=2$, $\lambda_4=0.5$ (iEEG)
NMM w/ HRF†		$\mathbf{x}(t)-\mathbf{x}(t-1)=(\mathbf{W}\psi_{\boldsymbol{\alpha}}(\mathbf{x}(t-1))-\mathbf{D}\mathbf{x}(t-1))\Delta_T+\mathbf{e}_1(t)$ (9) $\mathbf{y}(t)=\mathcal{H}(q)\mathbf{x}(t)+\mathbf{e}_2(t)$	MINDy default
DNN (MLP)	Nonlinear models via multi-layer perceptron deep neural networks	$\mathbf{y}(t)-\mathbf{y}(t-1)=f(\mathbf{y}(t-1),\dots,\mathbf{y}(t-d))+\mathbf{e}(t)$ (10)	$d=1, D=6, W=2$ (fMRI) $d=6, D=4, W=26$ (iEEG)
DNN (CNN)	Nonlinear models via convolutional deep neural networks	$\mathbf{y}(t)-\mathbf{y}(t-1)=f(\mathbf{y}(t-1),\dots,\mathbf{y}(t-d))+\mathbf{e}(t)$	$d=17, D=2, l_{\text{filt}}=7$, $n_{\text{filt}}=11, n_{\text{pool}}=4$, $p_{\text{drop}}=0.4$ (fMRI) $d=11, D=7, l_{\text{filt}}=2$, $n_{\text{filt}}=13, n_{\text{pool}}=1$, $p_{\text{drop}}=0.5$ (iEEG)
LSTM (IIR)	Nonlinear models via long short-term memory recurrent neural networks	$\mathbf{y}(t)-\mathbf{y}(t-1)=f(\mathbf{y}(t-1),\dots,\mathbf{y}(0))+\mathbf{e}(t)$	$W=12$ (fMRI) $W=1$ (iEEG)
LSTM (FIR)		$\mathbf{y}(t)-\mathbf{y}(t-1)=f(\mathbf{y}(t-1),\dots,\mathbf{y}(t-d))+\mathbf{e}(t)$	$d=1, W=16$ (fMRI) $d=7, W=1$ (iEEG)
Manifold	Nonlinear manifold-based models	$\mathbf{y}(t)-\mathbf{y}(t-1)=f(\mathbf{y}(t-1),\dots,\mathbf{y}(t-d))+\mathbf{e}(t)$	$d=1, h=830$ (fMRI) $d=7, h=1.2\times10^4$ (iEEG)
MMSE (pairwise)†	Nonlinear minimum mean squared error models (optimal)	$y_i(t)-y_i(t-1)=\mathrm{E}[y_i(t)-y_i(t-1)\mid y_j(t-1)], i,j=1,\dots,n$	$N=280, \beta=0.156$
MMSE (scalar)‡		$y_i(t)-y_i(t-1)=\mathrm{E}[y_i(t)-y_i(t-1)\mid y_i(t-1),\dots, y_i(t-d)], i=1,\dots,n$	$d=15, N=300, \beta=0.007$

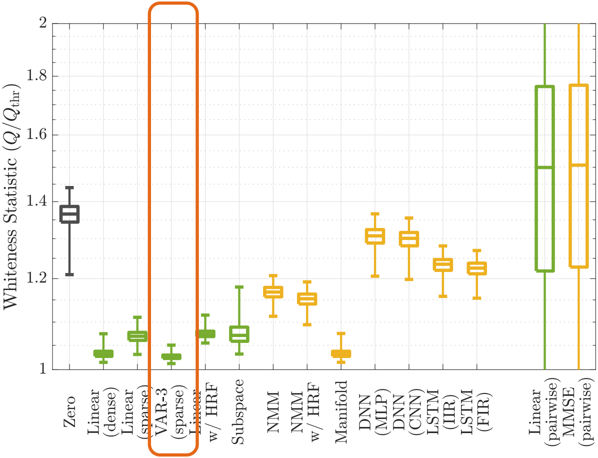
- Simultaneous tuning of each model's hyper-parameters via SGD, until reaching **steady state**



- $R_i^2 = 1 - \frac{\sum_t [\hat{y}_i(t|t-1) - y_i(t)]^2}{\sum_t [\bar{y}_i - y_i(t)]^2} \leq 1$, for any brain region i

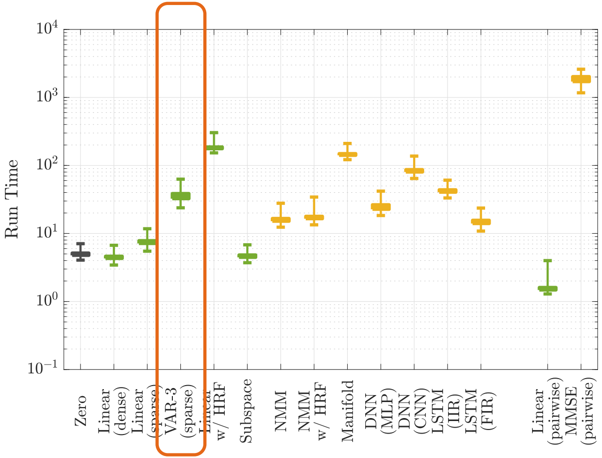


- Same linear VAR-3 model gives the whitest residuals



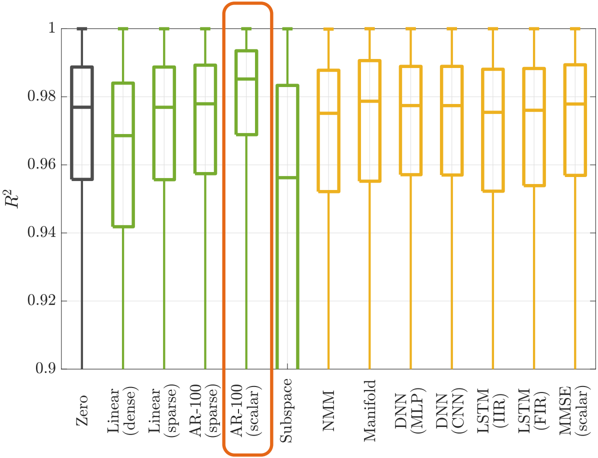
Comparison: Linear Models Are Faster (fMRI)

- Linear models are also faster (training + cross validation time)



Same Story for iEEG

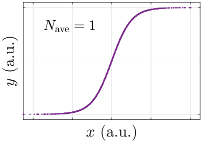
- Same holds true for iEEG



? Why?!

1. Averaging across space

$$y_i(t) = \sigma(x_i(t)), i = 1, \ldots, N$$

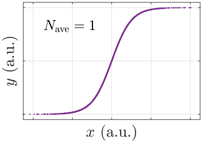


Reasons for Linearity

1. Averaging across space

$$y_i(t) = \sigma(x_i(t)), i = 1, \ldots, N$$

$$\langle x_i \rangle(t) = \frac{1}{N} \sum_{i=1}^{N} x_i(t), \qquad \langle y_i \rangle(t) = \frac{1}{N} \sum_{i=1}^{N} y_i(t)$$

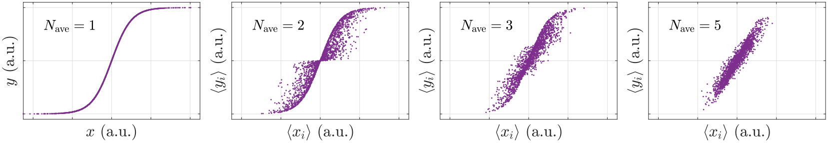


Reasons for Linearity

1. Averaging across space

$$y_i(t) = \sigma(x_i(t)), i = 1, \ldots, N$$

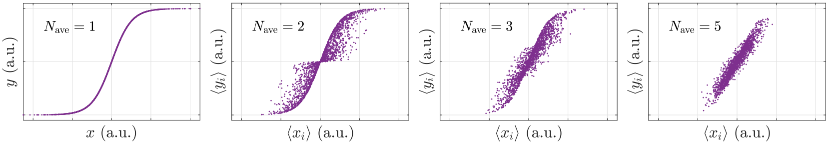
$$\langle x_i \rangle(t) = \frac{1}{N} \sum\nolimits_{i=1}^{N} x_i(t), \qquad \langle y_i \rangle(t) = \frac{1}{N} \sum\nolimits_{i=1}^{N} y_i(t)$$



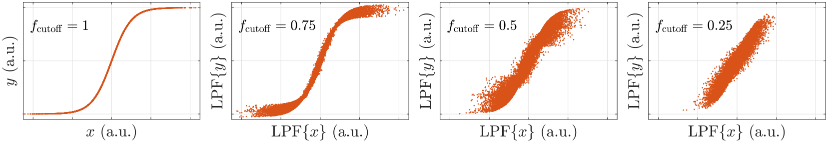
Reasons for Linearity

1. Averaging across space

$$y_i(t) = \sigma(x_i(t)), i = 1, \ldots, N$$
$$\langle x_i \rangle(t) = \frac{1}{N} \sum_{i=1}^{N} x_i(t), \qquad \langle y_i \rangle(t) = \frac{1}{N} \sum_{i=1}^{N} y_i(t)$$

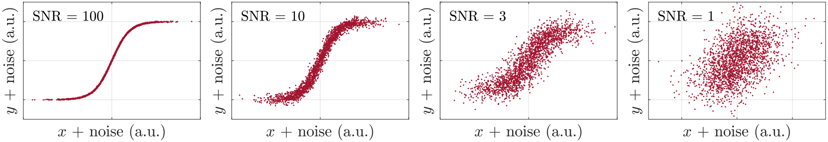


2. Averaging across time (low-pass filtering)

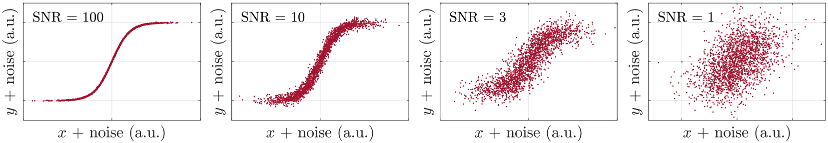


Reasons for Linearity – cont'd

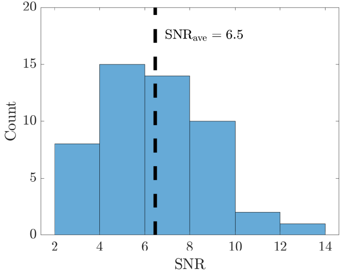
3. Observation noise



3. Observation noise

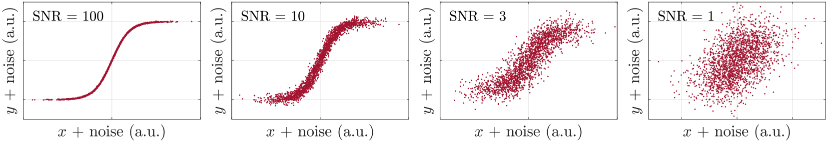


- SNR estimates for our fMRI dataset (upper bound)

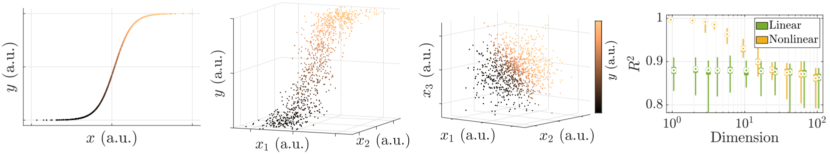


Reasons for Linearity – cont'd

3. Observation noise



4. Sample scarcity



⋆ What we have is a **combined effect** of 1-4

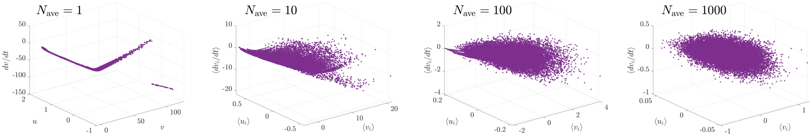
Replication on a Spiking Model

The Izhikevic model:

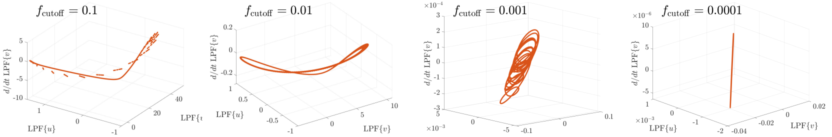
$$\dot{v}_i = 0.04v_i^2 + 5v_i + 140 - u_i + I, \qquad \text{if} \;\; v_i \geq 30mV : \begin{cases} v_i \leftarrow c \\ u_i \leftarrow u_i + d \end{cases}$$

$$\dot{u}_i = a(bv_i - u_i)$$

1. Averaging across space



2. Averaging across time (low-pass filtering)



Summary

- ✓ The critical role of scale in computational modeling of neural dynamics
- ✓ The brain is certainly nonlinear at microscale, but apparently linear at the macroscale at rest
- ✓ Potential reasons for (apparent) linearity: spatial averaging, temporal averaging, observation noise, dimensionality

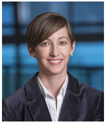
- ? Expanding the families of nonlinear models
- ? Extension of identified models for control
- ? Extensions to include time delays
- ? Are linear models accurate enough?


Scale of Abstraction
Macroscale
Mesoscale
Microscale
Task
Perception
Attention
Decision Making
Rest
Identification
Analysis
Control
Problem

Thank You!

Joint work with

- Prof. Danielle Bassett
- Prof. George Pappas
- Dr. Max Bertolero
- Dr. Jeni Stiso
- Dr. Lorenzo Caciagli
- Dr. Eli Cornblath
- Dr. Xiosong He
- Dr. Arun Mahadevan





and available at

Is the brain macroscopically linear? A system identification of resting state dynamics
E. Nozari, M. A. Bertolero, J. Stiso, L. Caciagli, E. J. Cornblath, X. He, A. S. Mahadevan, G. J. Pappas, D. S. Bassett
bioRxiv