

# Dynamical models of fMRI

Papers by the Durstewitz lab. Specifically ones by:  
Dr. Durstewitz, Dr. Georgia Koppe, Dr. Jiarui Chen, and Eric Volkmann

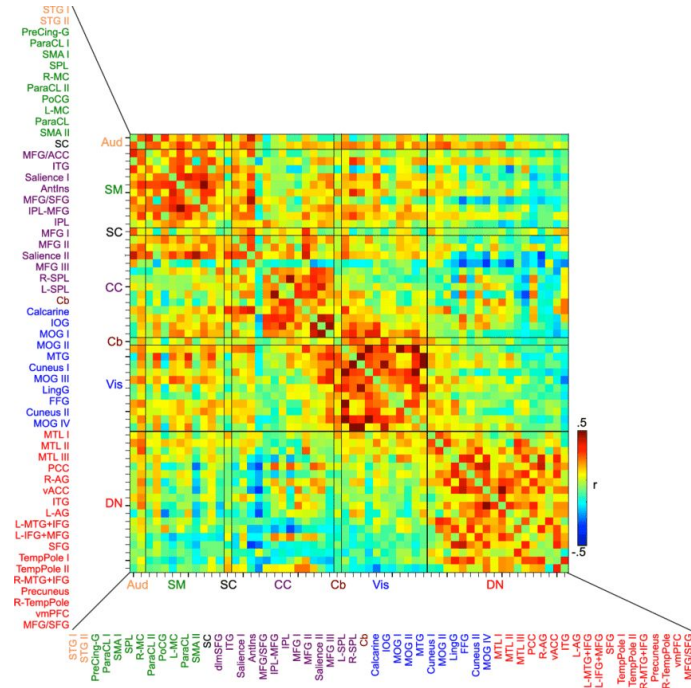
# What is a dynamical model?

$$\begin{array}{c} \left[ \begin{array}{c} \square \\ \frac{dx}{dt} \\ \square \end{array} \right] \\ n \times 1 \end{array} = \begin{array}{c} \text{System Matrix} \\ \left[ \begin{array}{cccc} \square & \square & \square & \square \\ \square & \square & \square & \square \\ \square & \square & \square & \square \\ \square & \square & \square & \square \end{array} \right] \\ A \\ n \times n \end{array} \begin{array}{c} \text{State Vector} \\ \left[ \begin{array}{c} \square \\ x(t) \\ \square \end{array} \right] \\ n \times 1 \end{array} + \begin{array}{c} \text{Input Matrix} \\ \left[ \begin{array}{ccc} \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \end{array} \right] \\ B \\ n \times r \end{array} \begin{array}{c} \text{Input Vector} \\ \left[ \begin{array}{c} \square \\ u(t) \\ \square \end{array} \right] \\ r \times 1 \end{array}$$

Figure from: <https://cookierobotics.com/018/>

# Why dynamical models?

$$\begin{array}{c}
 \left[ \begin{array}{c} \frac{dx}{dt} \\ \vdots \end{array} \right] = \begin{array}{c} \text{System Matrix} \\ \left[ \begin{array}{cccc} & & & \\ & A & & \\ & & & \\ & & & \end{array} \right] \end{array} \begin{array}{c} \text{State Vector} \\ \left[ \begin{array}{c} x(t) \\ \vdots \end{array} \right] \end{array} \\
 n \times 1 \qquad \qquad \qquad n \times n \qquad \qquad \qquad n \times 1
 \end{array}$$



Leftmost figure from: <https://cookiebotics.com/018/> Rightmost figure from: [https://www.researchgate.net/figure/Static-mean-functional-network-connectivity-correlation-matrix-of-ICNs-across-task-and\\_fig2\\_275667579](https://www.researchgate.net/figure/Static-mean-functional-network-connectivity-correlation-matrix-of-ICNs-across-task-and_fig2_275667579)

# Why dynamical models? Previous approaches

- Dynamic causal modeling (DCM) [1]
- Criticality in neuroscience (when a state collapses to another) [2]
- Stability of brain dynamics [3]

[1] Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *Neuroimage*, 19(4), 1273-1302.

[2] O'Byrne, J., & Jerbi, K. (2022). How critical is brain criticality?. *Trends in Neurosciences*, 45(11), 820-837.

[3] Kelso, J. S. (2012). Multistability and metastability: understanding dynamic coordination in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1591), 906-918.

# Why dynamical models? Previous approaches

- Dynamic causal modeling (DCM) [1]
- Criticality in neuroscience (when a state collapses to another) [2]
- Stability of brain dynamics [3]
- Brain attractors [4]

[1] Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *Neuroimage*, 19(4), 1273-1302.

[2] O'Byrne, J., & Jerbi, K. (2022). How critical is brain criticality?. *Trends in Neurosciences*, 45(11), 820-837.

[3] Kelso, J. S. (2012). Multistability and metastability: understanding dynamic coordination in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1591), 906-918.

[4] Knierim, J. J., & Zhang, K. (2012). Attractor dynamics of spatially correlated neural activity in the limbic system. *Annual review of neuroscience*, 35(1), 267-285.

# What are attractors?

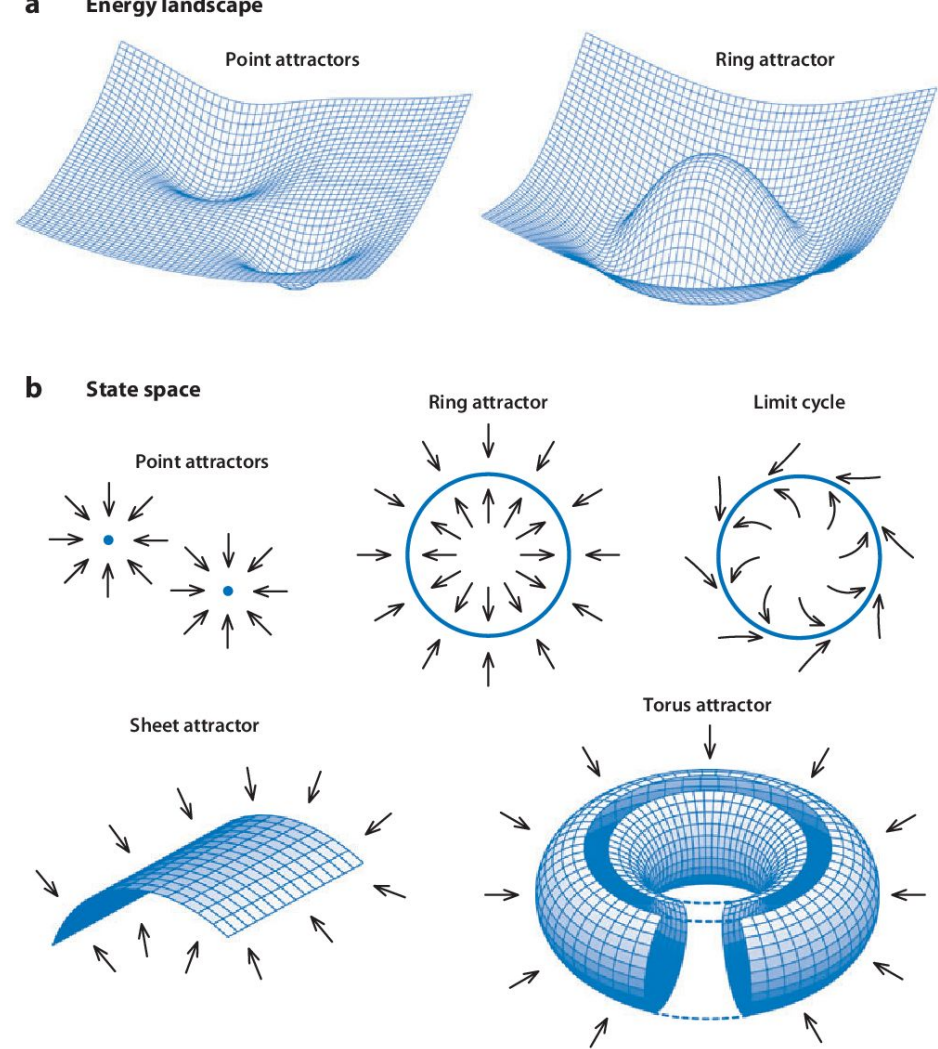


Figure from: Knierim, J. J., & Zhang, K. (2012). Attractor dynamics of spatially correlated neural activity in the limbic system. *Annual review of neuroscience*, 35(1), 267-285.

# What is stability?

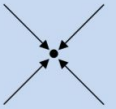

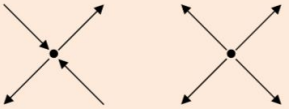
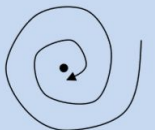
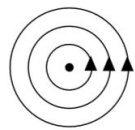
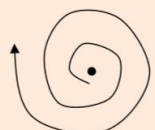
	Stable $\text{Re}(\lambda_d) < 0$	Lyapunov stable* $\text{Re}(\lambda_d) = 0$	Unstable $\text{Re}(\lambda_d) > 0$
Real eigenvalues	Stable point $\text{Re}(\lambda_1) < 0, \text{Re}(\lambda_2) < 0$ 	Neutral point $\text{Re}(\lambda_1) = 0, \text{Re}(\lambda_2) < 0$ $\text{Re}(\lambda_1) = 0, \text{Re}(\lambda_2) = 0$ 	Saddle point   Unstable point $\text{Re}(\lambda_1) > 0, \text{Re}(\lambda_2) < 0$ $\text{Re}(\lambda_1) > 0, \text{Re}(\lambda_2) > 0$ 
Complex conjugate eigenvalues	 $\text{Re}(\lambda_1) = \text{Re}(\lambda_2) < 0$ Stable spiral focus	 $\text{Re}(\lambda_1) = \text{Re}(\lambda_2) = 0$ Neutral center	 $\text{Re}(\lambda_1) = \text{Re}(\lambda_2) > 0$ Unstable spiral focus

Figure from:

[https://math.libretexts.org/Bookshelves/Scientific\\_Computing\\_Simulations\\_and\\_Modeling/Introduction\\_to\\_the\\_Modeling\\_and\\_Analysis\\_of\\_Complex\\_Systems\\_%28Sayama%29/07%3A\\_ContinuousTime\\_Models\\_II\\_Analysis/7.05%3A\\_Linear\\_Stability\\_Analysis\\_of\\_Nonlinear\\_Dynamical\\_Systems](https://math.libretexts.org/Bookshelves/Scientific_Computing_Simulations_and_Modeling/Introduction_to_the_Modeling_and_Analysis_of_Complex_Systems_%28Sayama%29/07%3A_ContinuousTime_Models_II_Analysis/7.05%3A_Linear_Stability_Analysis_of_Nonlinear_Dynamical_Systems)

# PLRNNs: dynamic model

This article considers simple discrete-time piecewise-linear (PL) recurrent neural networks (RNN) of the form

$$\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \mathbf{W} \max\{\mathbf{0}, \mathbf{z}_{t-1} - \boldsymbol{\theta}\} + \mathbf{C}\mathbf{s}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \quad (1)$$

A is the system matrix (together with W)

C is the input matrix

The system 'affects' itself through a ReLU (the max function)

→ This makes the dynamics piecewise linear



# Why piecewise linear?

A particular advantage of the PLRNN model is that all its fixed points can be obtained easily analytically by solving (in the absence of external input) the  $2^M$  linear equations

$$\mathbf{z}_* = (\mathbf{A} + \mathbf{W}_\Omega - \mathbf{I})^{-1} \mathbf{W}_\Omega \boldsymbol{\theta}, \quad (2)$$

where  $\Omega$  is to denote the set of indices of units for which we assume  $z_m \leq \theta_m$ , and  $\mathbf{W}_\Omega$  the respective connectivity matrix in which all columns from  $\mathbf{W}$  corresponding to units in  $\Omega$  are set to 0. Obviously, to make  $\mathbf{z}_*$  a true fixed point of (1), the solution to (2) has to be consistent with the defined set  $\Omega$ , that is  $z_m \leq \theta_m$  has to hold for all  $m \in \Omega$  and  $z_m > \theta_m$  for all  $m \notin \Omega$ . For networks of moderate size (say  $M < 30$ ) it is thus computationally feasible to explicitly check for all fixed points and their stability.

# PLRNNs: Decoder model

For estimation from experimental data, latent state model (1) is then connected to some  $N$ -dimensional observed vector time series  $\mathbf{X} = \{\mathbf{x}_t\}$  via a simple linear-Gaussian model,

$$\mathbf{x}_t = \mathbf{B}\phi(\mathbf{z}_t) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{\Gamma}), \quad (3)$$

where  $\phi(\mathbf{z}_t) = \max\{\mathbf{0}, \mathbf{z}_t - \boldsymbol{\theta}\}$ ,  $\{\boldsymbol{\eta}_t\}$  is the (white Gaussian) observation noise series with diagonal covariance matrix  $\boldsymbol{\Gamma} = \text{diag}([\gamma_{11}^2 \dots \gamma_{NN}^2])$ , and  $\mathbf{B}$  an  $N \times M$  matrix of regression weights. Thus, the idea is that only the PL-transformed activation  $\phi(\mathbf{z}_t)$  reaches the ‘observation surface’ as, e.g., with spiking activity when the underlying membrane dynamics itself is not visible. We further assume for the initial state,

$$\mathbf{z}_1 \sim N(\boldsymbol{\mu}_0 + \mathbf{s}_1, \boldsymbol{\Sigma}), \quad (4)$$

# Applying this framework to fMRI data

A decoder model that relates the neuronal processes given as latent time series  $\{z_t\}$  to measured BOLD time series  $\{x_t\}$  may be formulated as in [40],

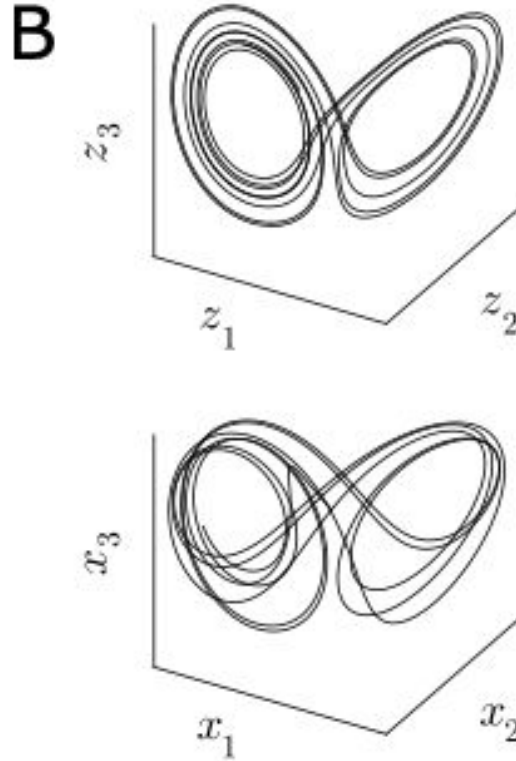
$$x_t = \mathbf{B} ((hrf * z)_t) + \mathbf{J} r_t + \eta_t, \quad \eta_t \sim N(0, \mathbf{\Gamma}) \quad (5)$$

with regression coefficient matrices  $\mathbf{B} \in \mathbb{R}^{N \times M}$  and  $\mathbf{J} \in \mathbb{R}^{N \times P}$ , nuisance variables  $r_t \in \mathbb{R}^P$  (such as movement or respiratory artifacts) and a Gaussian observation noise term  $\eta_t$  (with usually diagonal covariance  $\mathbf{\Gamma} \in \mathbb{R}^{N \times N}$ ). Here,  $*$  denotes the convolution operation and  $z$  is a history of states  $z_{t-\tau:t}$ ,

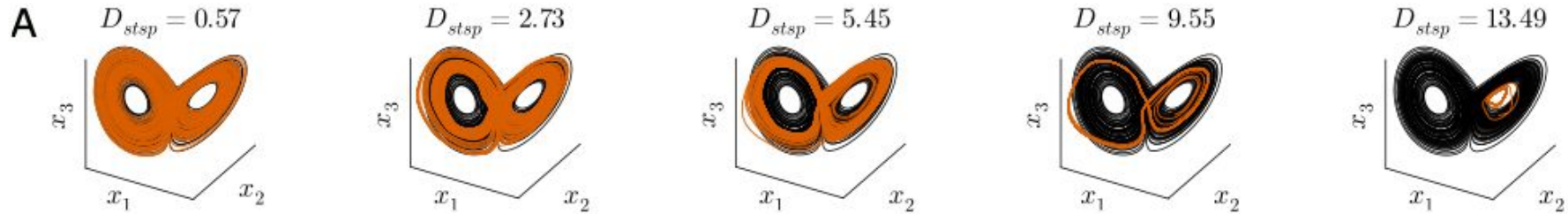
# Simulations

Lorenz 3D dynamic system

Lorenz 3D dynamic system with HRF



# Measuring similarity to underlying system

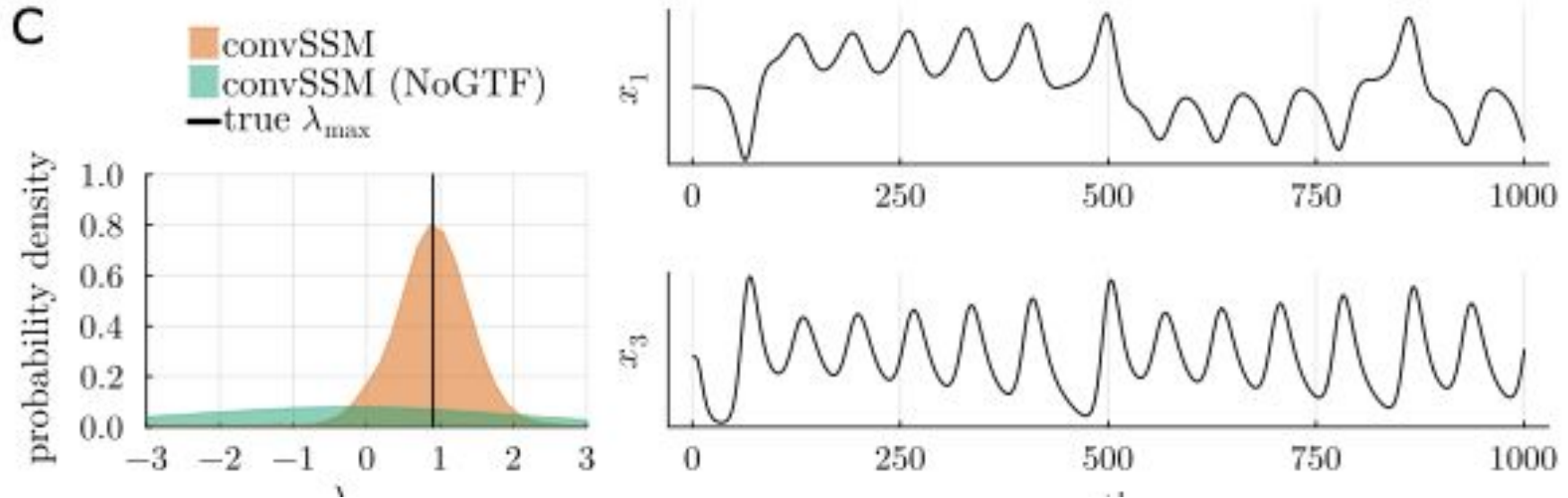


They use the KL-divergence between their own freely simulated model, and the actual underlying system instead of prediction MSE.

→ Tiny changes in initial conditions for **chaotic** dynamical systems like the Lorenz system can cause large deviations between **predicted states in the future**.

→ KL-divergence cares more about ‘**geometrical similarity**’

# Performance on chaotic system

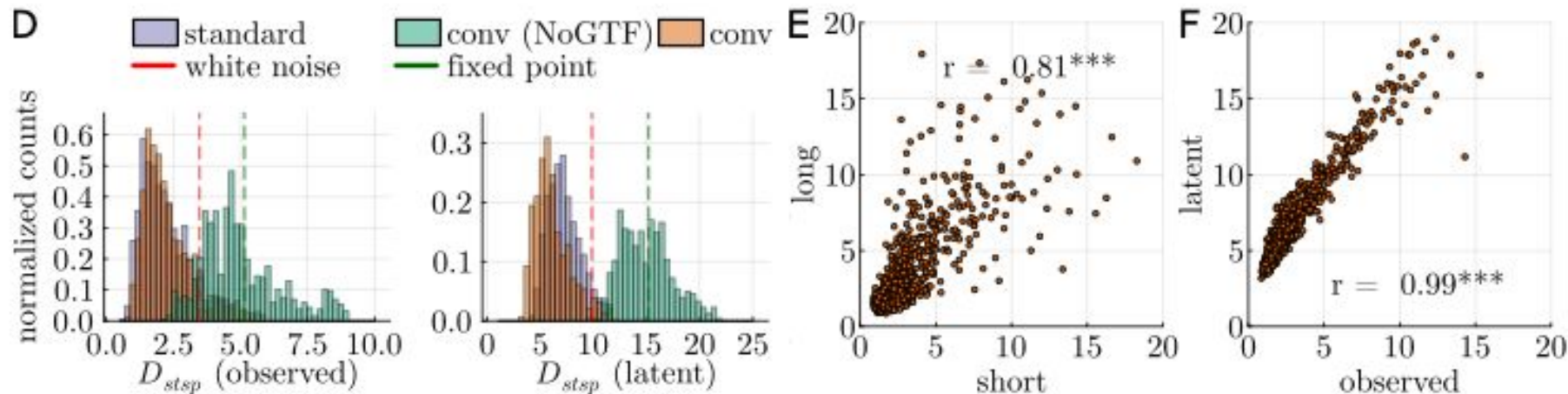


C: We can calculate the fitted model's eigenvalues and compare it to the ground-truth model's eigenvalues!

# fMRI simulated data (ALN)

The adaptive linear-nonlinear (ALN) cascade model is a population model of spiking neural networks. The dynamical variables of the ALN model describe the average firing rate and other macroscopic variables of a randomly connected, delay-coupled network of excitatory and inhibitory adaptive exponential integrate-and-fire neurons (AdEx) with non-linear synaptic currents [3].

# Performance on fMRI simulated data



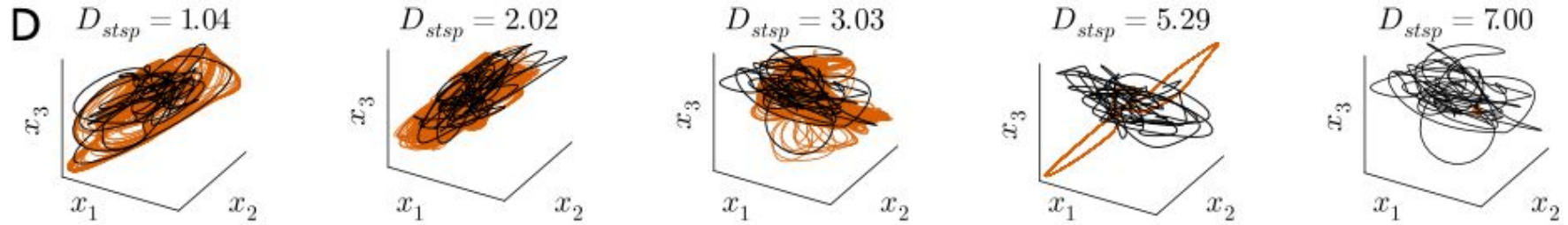
D: their HRF-based model is Conv, Standard is their standard model (no HRF)

E:  $D_{stsp}$  on a long vs short test set (because fMRI timeseries are short)

F:  $D_{stsp}$  for observed and latent timeseries

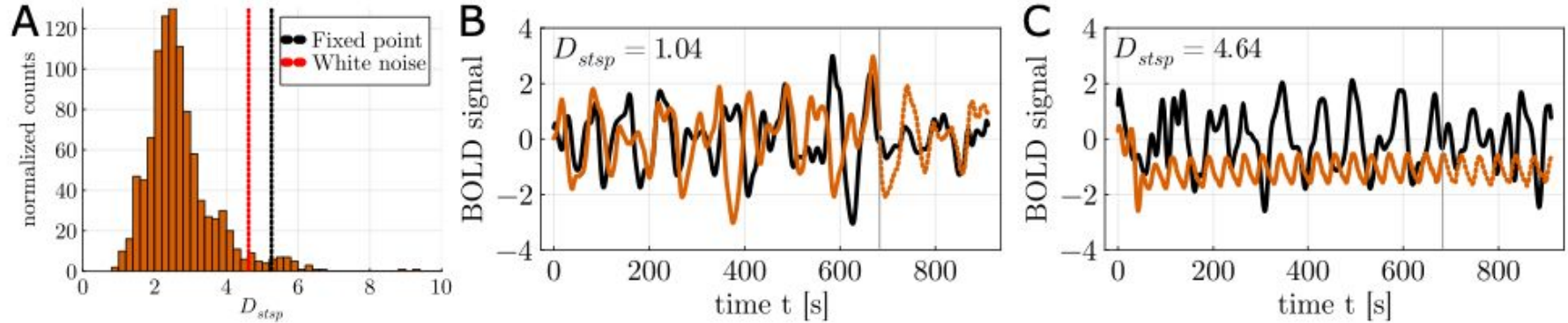


# Performance on real fMRI data



These are examples of goodness of geometric fit on real fMRI data

# Performance on real fMRI data

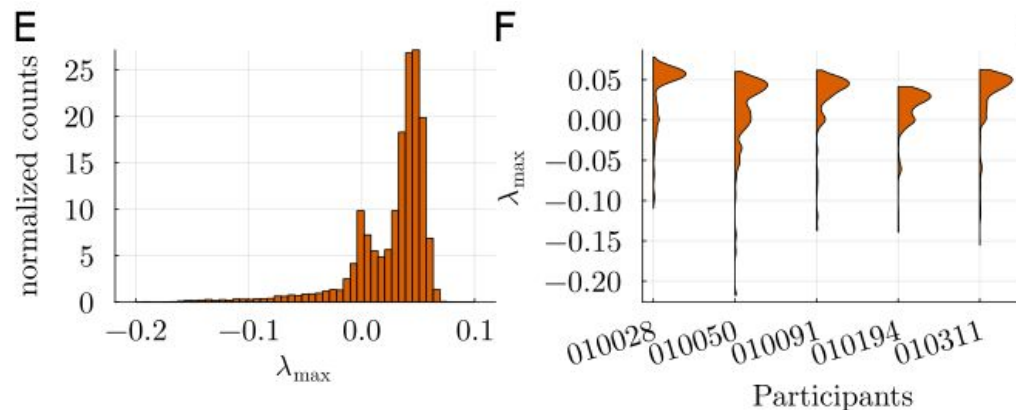


A:  $D_{stsp}$  for inferred models (across different random seeds)

B: A good fit of the data (note that it's more about the geometric fit)

C: A relatively bad fit of the data (again, it's about the geometric not exact fit)

# Performance on real fMRI data



E: Distribution of maximum dynamic eigenvalues across different model fits (different random seeds, and maybe? subjects)

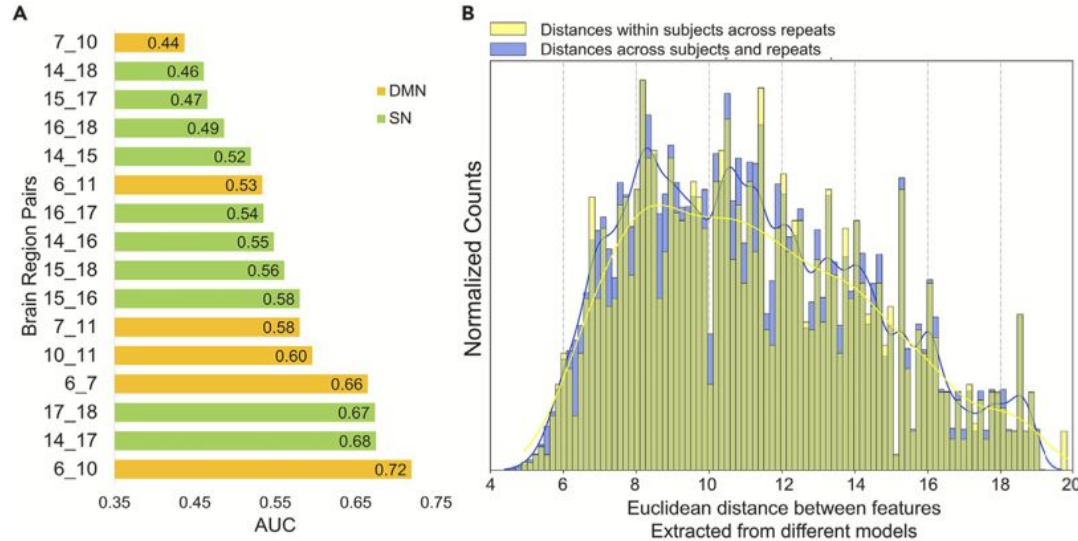
# Psychiatric disorder identification with dynamic features

**Table S3.** Description of the 18 extracted dynamical features<sup>1,2</sup>, related to “PLRNN Model for Dynamics Reconstruction and Feature Extraction” section in STAR Methods.

Index	Feature Description	Data Type
F01	Total number of fixed points in system	Integer
F02	The number of unstable fixed points in system	Integer
F03	The average of absolute imaginary eigenvalues of all fixed points	Continuous
F04	The average of maximum absolute eigenvalues of all fixed points	Continuous
F05/6/7	The variance of the parameters in both/regularized/non-regularized part of transition matrices (A and W)	Continuous
F08/9	The average of the parameters in regularized/non-regularized part of bias matrix (h)	Continuous
F10	The number of stable cycles	Integer
F11/12	The average of the parameters in regularized/non-regularized part of weight matrix (W)	Continuous
F13	The sum of absolute model weights averaged across latent states	Continuous
F14	The average Euclidean distance between the inferred latent states over time (speed)	Continuous
F15	The average variance of the latent states over time	Continuous
F16	The frequency of system switched between different orthants in state space	Continuous
F17/18	The average and variance over columns of regression coefficient matrix (B)	Continuous

Reference: Chen, J., Benedyk, A., Moldavski, A., Tost, H., Meyer-Lindenberg, A., Braun, U., ... & Schwarz, E. (2024). Quantifying brain-functional dynamics using deep dynamical systems: Technical considerations. *Iscience*, 27(8).

# Performance for psychiatric disorders: hard to repeat

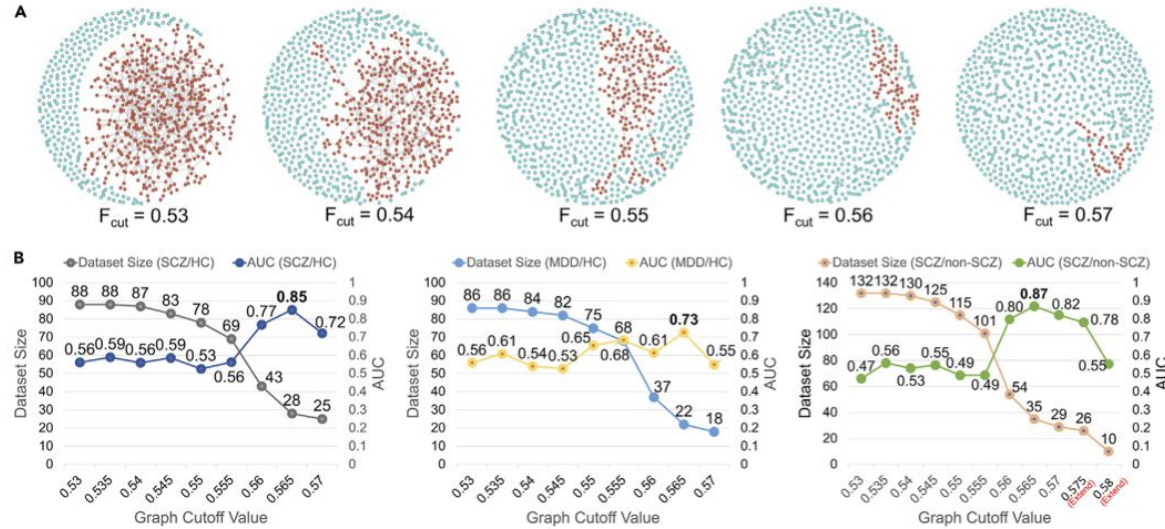


**Figure 2. Performance of RF classifiers using features from different brain regions and demonstration of weak feature robustness**

(A) Average AUC scores of RF classifiers for the SCZ/HC classification task in 5-fold cross validation using features extracted from region pairs belonging to the DMN (yellow) and SN (green). The complete results can be found in the [Table S5](#). The full list of region indices/names can be found in [Table S2](#).

(B) The distribution of the Euclidean distances between features extracted from different PLRNN models (blue: distance across participants and repeats; yellow: distance within participants across repeats).

# Performance for psychiatric disorders: graph clustering helps



**Figure 3. Graph clustering approach and impact on classification performance**

(A) Illustration of graph-filtering cut-off ( $F_{cut}$ ) on the graph where each node represents an individual per model building repeat. The largest cluster is highlighted in red, while other nodes or smaller clusters are marked in cyan.

(B) The LOOCV performances of the RF classifiers using the graph clustering method in the SCZ/HC, MDD/HC, and SCZ/non-SCZ (HC + MDD) classification tasks.

The detail of model performance can be found in the [Table S9](#).