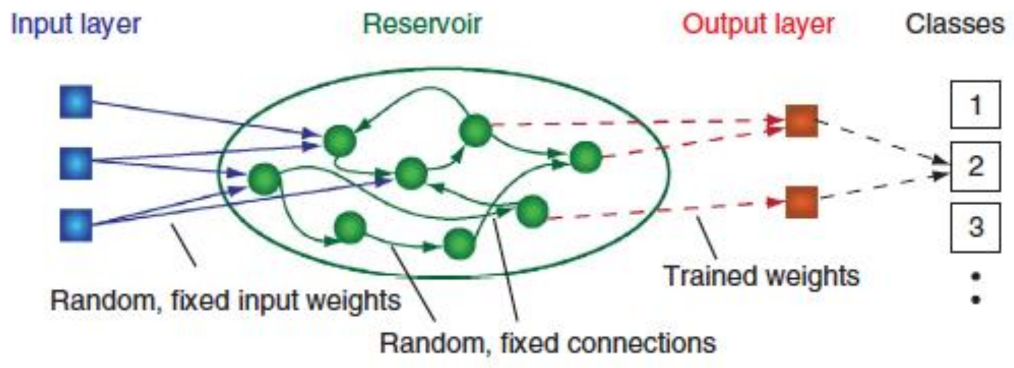
Statistics of Attractor Embeddings in Reservoir Computing

Louis Pecora, US Naval Research Laboratory, Washington, DC, US

Thomas Carroll, US Naval Research Laboratory, Washington, DC, US



from the AI world → Nonlinear Dynamics world

Nonlinear Dynamics techniques and concepts → the AI world

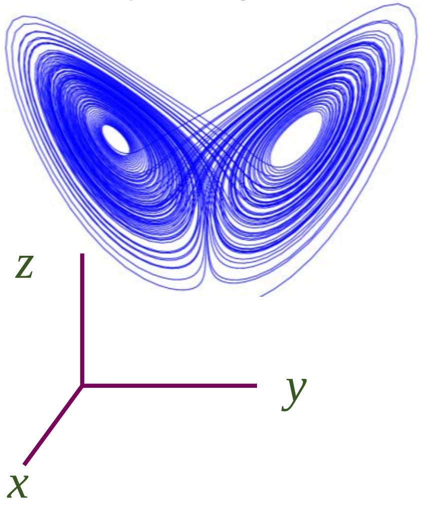
Third Symposium on Machine Learning and Dynamical Systems
Fields Institute, September 2022

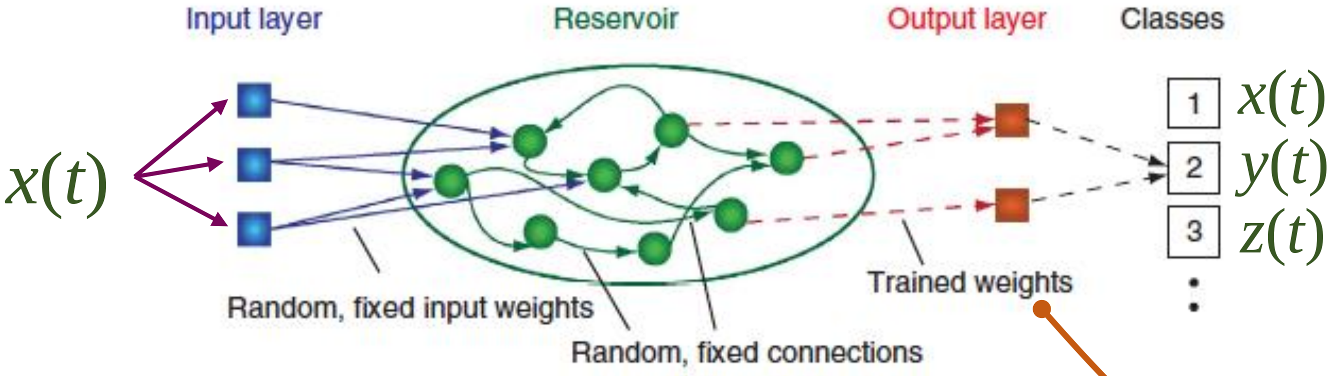
Introduction to
Reservoir Computers (RC)

What can a reservoir computer do? (1)

$\mathbf{r}(t) = (r_1, r_1, \ldots, r_N)$

Lorenz chaotic trajectory





$$z(t) = \sum_{j=1}^{N} W_{z,j} r_j$$

FAST training

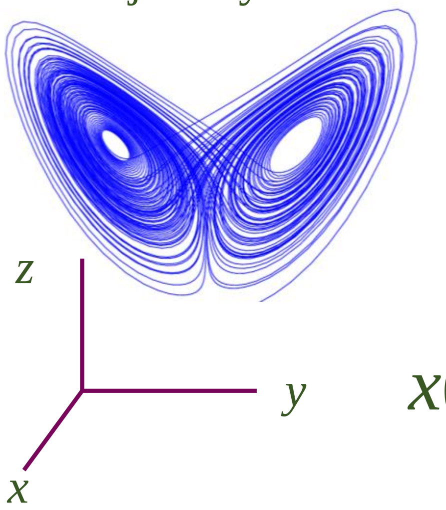
Only train output weights

Reservoir is unchanged

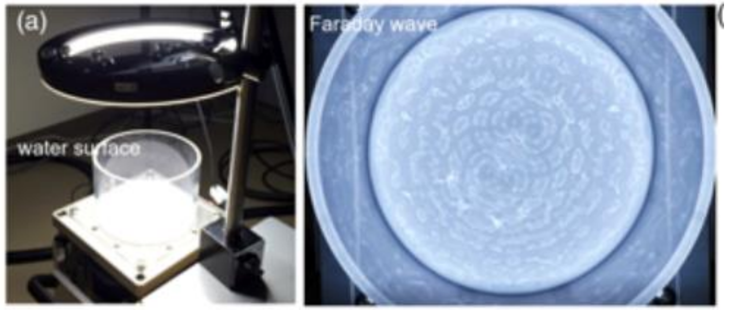
FAST operation

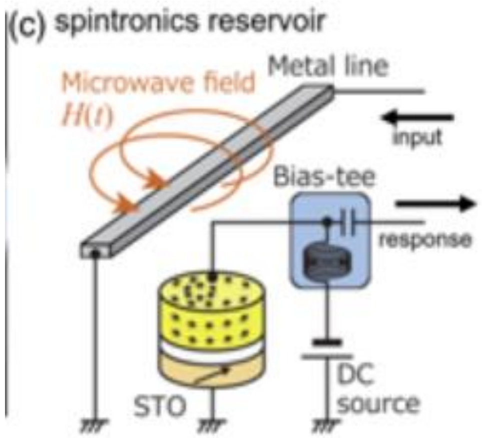
What can a reservoir computer do? (2)

Lorenz chaotic trajectory



$x(t)$



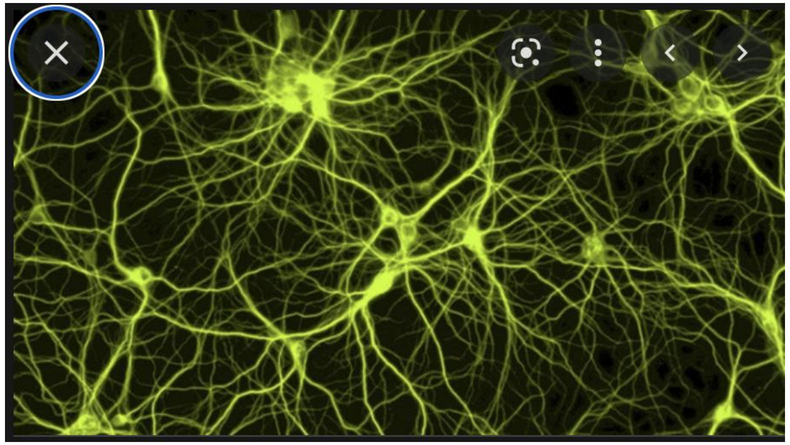


Drive with one signal from a Dynamical system and reproduce The other signals, i.e. the whole trajectory !

$x(t)$
$y(t)$
$z(t)$

$$z(t) = \sum_{j=1}^{N} W_{z,j} r_j$$

Neuronal Networks



FAST training

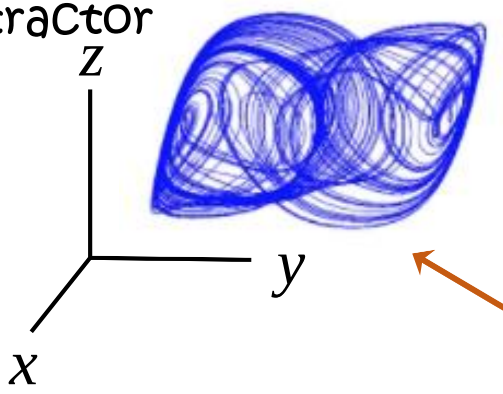
FAST operation

- RC can be physical systems. <u>Dynamical Systems</u>

RC embeddings

Takens theorem (1981)

Original attractor



pick a time delay (τ) and dimension (d)

$\rightarrow x(t)$

$\mathbf{v}_1=[x(t), x(t-\tau), x(t-2\tau),\dots, x(t-(d-1)\tau)],$

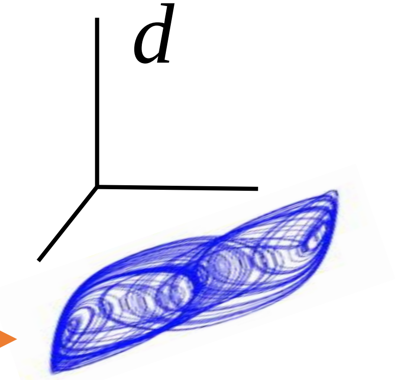
$\mathbf{v}_2=[x(t+\tau), x(t), x(t-2\tau),\dots, x(t-(d)\tau)],$

$\mathbf{v}_3=[x(t+2\tau), x(t+3\tau), x(t+4\tau),\dots, x(t-(d+1)\tau)],$

…

diffeomorphism
(continuous, differentiable, inverse)
Whitney Embedding Theorem

φ



Reconstructed attractor

Dynamical and geometric properties of Original attractor are also the same in the Reconstructed attractor.

What time delay (τ) and dimension (d) to use?
Still not fully worked out.

Develop a mathematical model that will expose the nonlinear dynamics of RC and the underlying geometric structure.

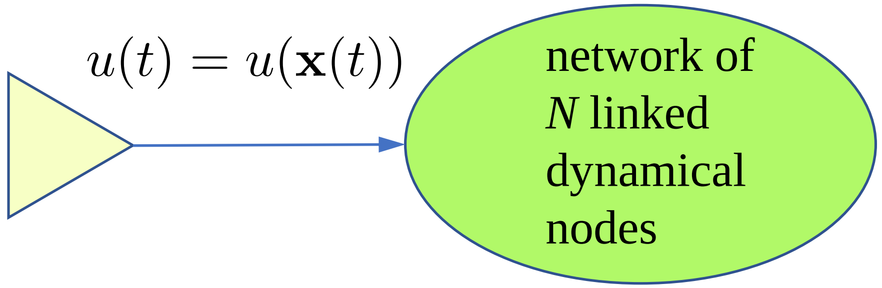
drive

RC

map

$\mathbf{x} \in \mathbb{R}^K$ $\quad \mathbf{x}(t+1) = \mathbf{h}(\mathbf{x}(t))$

$u(t) = u(\mathbf{x}(t))$



$\mathbf{r}(t) = \mathbf{f}[u(t\text{-}1), \mathbf{r}(t\text{-}1)]$

$\mathbf{r}(t) = \mathbf{f}[u(t\text{-}1), \mathbf{r}(t\text{-}1)]$

map $\quad \mathbf{r}_n(t) = \mathbf{f}[u(t-1), \mathbf{f}[u(t-2), \mathbf{f}[u(t-3), ..., \mathbf{f}[u(t-n), \mathbf{r}_0]...]]] \equiv \mathbf{g}_n(u, \mathbf{r}_0)$

We want the sequence $\{\mathbf{r}_n(t)\}$ to converge to the same point as *n* increases since we expect the RC to be in generalized synchronization. Using the Cauchy condition on the initial value r_0 we need to have

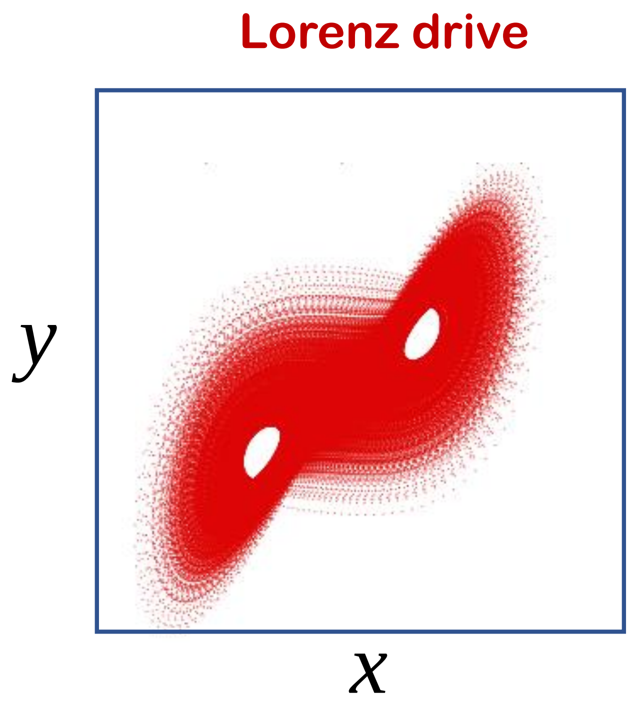
$|\mathbf{g}_k(u, \mathbf{r}_0) - \mathbf{g}_l(u, \mathbf{r}_0)| < \epsilon$ for a choice of ϵ and for *k* and *l* large enough.

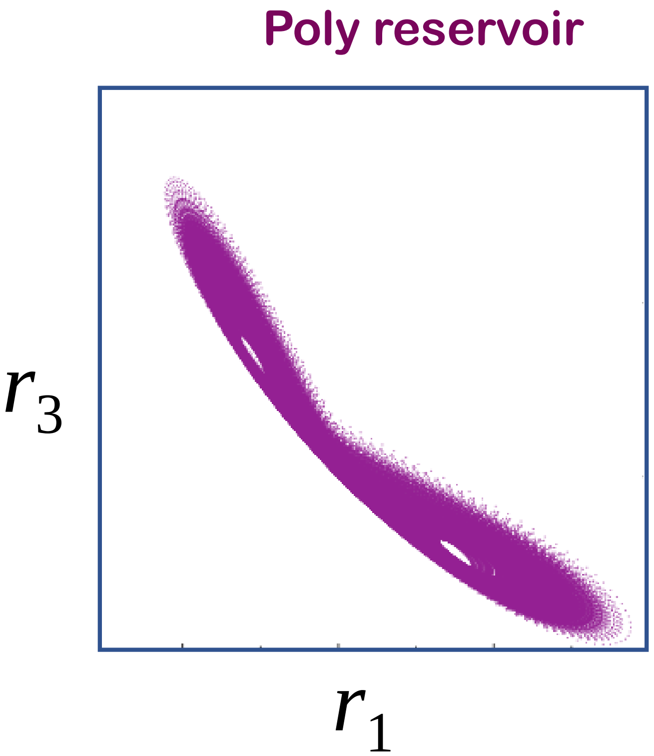
$\mathbf{g}_l(u, \mathbf{r}_0) \to \mathbf{r}(t)$ Uniformly convergent. $\mathbf{r}(t)$ is unique and inherits properties of $\{\mathbf{g}_l\}$

=> dynamically driven RCs can reconstruct the attractor of the drive system

Reconstructing an attractor using RC

Lorenz-Poly system 100 polynomial, 1 dimensional nodes





Theorems don't cover many cases of drive-RC systems. And they don't necessarily give quantitative information for the system.
How can we gauge the relationship between the drive and RC given generated time series or data?

We need statistics to gauge continuity and differentiability and other mathematical properties from data/time series.

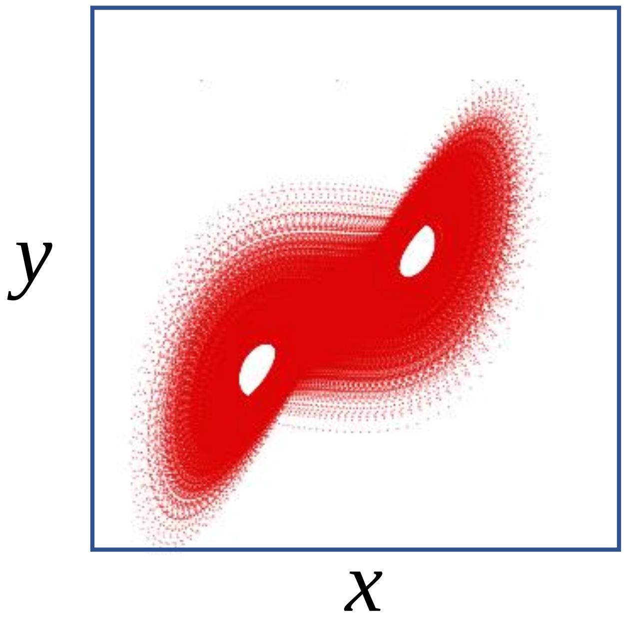
The continuity and differentiability statistics and other measures of and RCs and embeddings

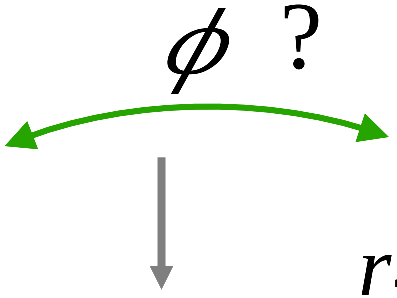
Reconstructing an attractor using RC

$$\frac{dx}{dt}=\sigma(y-x)$$
$$\frac{dy}{dt}=-xz+\rho x-y$$
$$\frac{dz}{dt}=xy-\beta z$$

$$\frac{dr_i}{dt}=\alpha[\kappa(p_1 r_i+p_2 r_i^2)+p_3 r_i^3+\sum_{j=1}^{N}A_{ij}r_j+w_i x(t)]$$

Lorenz drive

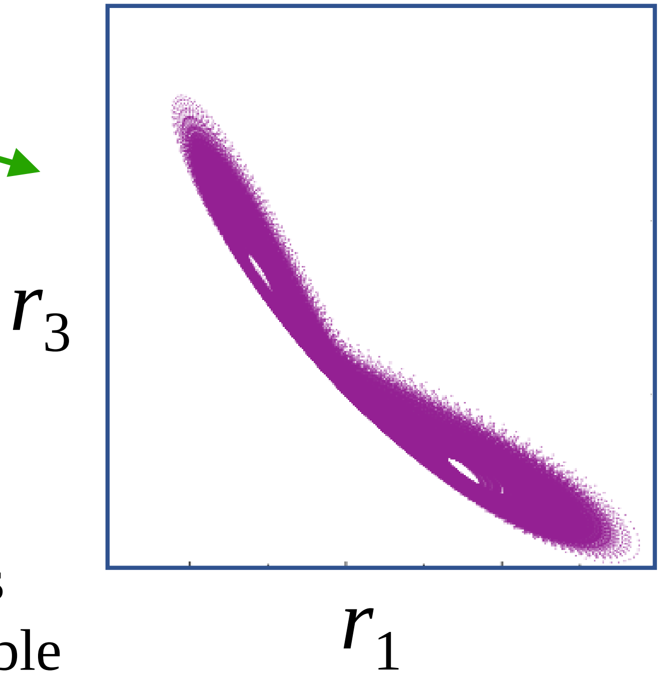




continuous
differentiable
invertible
inv. continuous
inv. differentiable

diffeomorphism

Poly reservoir



A Continuity Statistic

A function $f(x)$ is continuous at a point x_0

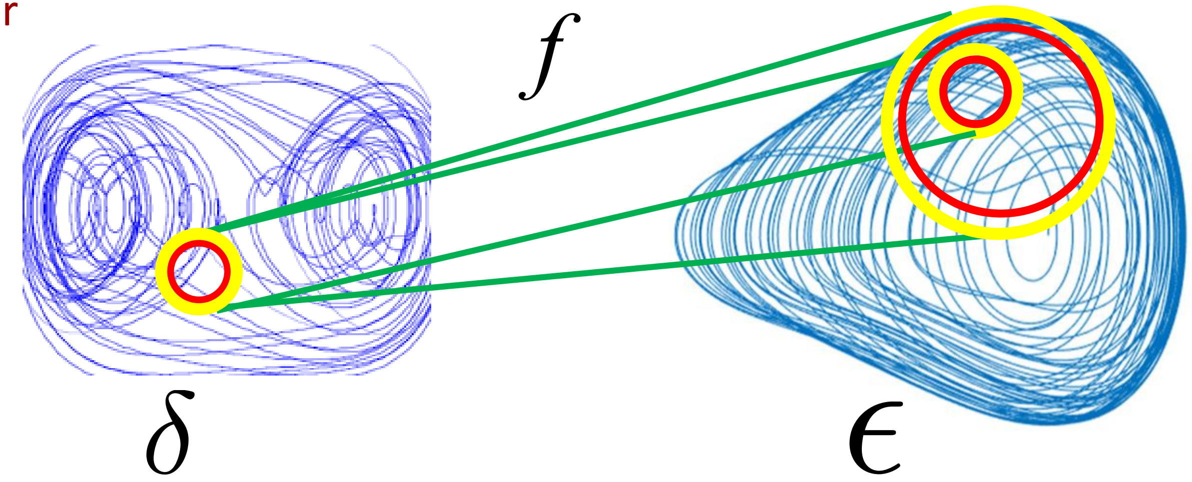
$$\forall\, \epsilon > 0 \;\; \exists\, \delta > 0 : \text{whenever} \;\; |x - x_0| < \delta \;\; \Rightarrow |f(x) - f(x_0)| < \epsilon$$

Not functions, but two simultaneous vector data sets (time series) {**y**(t)} and {**r**(t)} t=1,2,3,... from drive $\mathcal{D}$ and from reservoir $\mathcal{R}$

drive reservoir

$$\begin{array}{ccc} \mathbf{y}(t_1) & \longleftrightarrow & \mathbf{r}(t_1) \\ \mathbf{y}(t_2) & \longleftrightarrow & \mathbf{r}(t_2) \\ \mathbf{y}(t_3) & \longleftrightarrow & \mathbf{r}(t_3) \\ \vdots & & \vdots \end{array}$$

drive reservoir



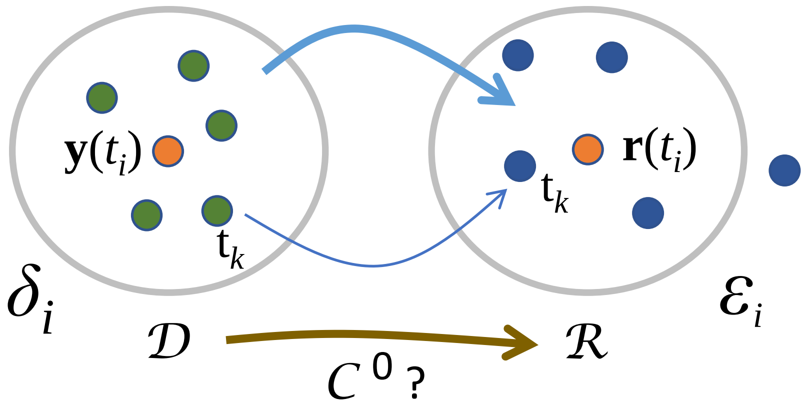
drive

RC

A Continuity Statistic

A function $f(x)$ is continuous at a point x_0

$$\forall \epsilon > 0 \ \exists \delta > 0 : \text{whenever } |x - x_0| < \delta \ \Rightarrow |f(x) - f(x_0)| < \epsilon$$



Null Hypothesis: points are mapped into ε set with prob. p_ε

$p_\varepsilon = 0.5$ a coin flip 

$n_\varepsilon = 6$ to reject Null at 0.98

$< \varepsilon_i > = \varepsilon^*$ $\varepsilon^* / \varepsilon_{min}$ or $\varepsilon^* / \sigma_{std}$: continuity statistic

These statistics depend on the amount of data. We cannot let $\varepsilon \rightarrow 0$.

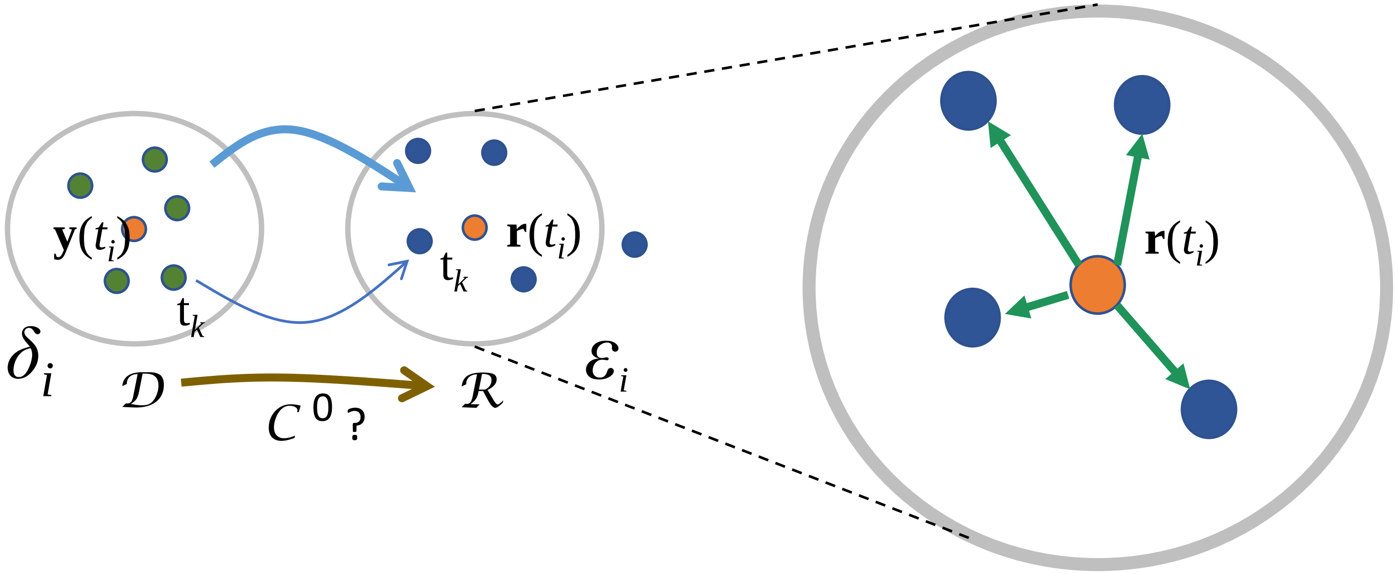
A Continuity Statistic

- Statistics for Mathematical Properties of Maps between Time-Series Embeddings, L.M. Pecora, T.L. Carroll, and J.F. Heagy,, Physical Review E, 54, 3420 (1995)
- Detecting Drive-Response Geometry in Generalized Synchronization, L.M. Pecora and T.L. Carroll, International Journal of Bifurcations and Chaos, 10, 875-890 (Apr, 2000)
- A Unified Approach to Attractor Reconstruction, L.Pecora, L. Moniz, J. Nichols, and T. Carroll, CHAOS 17, 013110 (2007)
- Kraemer, Datseris, Kurths, I Z Kiss , Ocampo-Espindola and Marwan, New J. Phys. **23**, 033017 (2021)

A Differentiability Statistic

A function $f(x)$ is differentiable at a point x_0 if local points are approximated by a linear map from x_0 ,
i.e. there is a tangent space.

Use local points from the continuity statistic to see what dimension the the Singular Values of the differences from x_0 are.



SVdim
differentiability
statistic

$\varepsilon^*/\varepsilon_{\text{min}}$ or $\varepsilon^*/\sigma_{\text{std}}$ and SVdim ~ Diffeomorphism

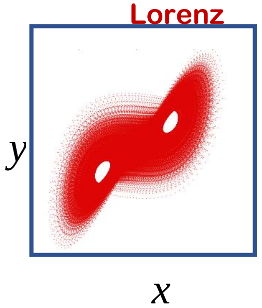
A Continuity Statistic (remarks)

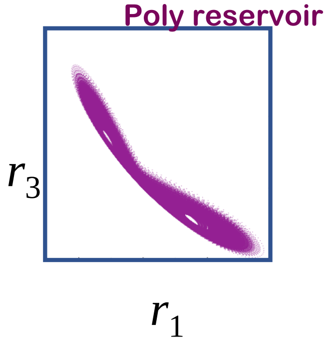
$\varepsilon^*/\varepsilon_{\min}$ or $\varepsilon^*/\sigma_{std}$

- We assume **nothing** about the possible functional relations between the data sets.
- The statistic is for **one** direction only ($\mathcal{D} \rightarrow \mathcal{R}$). It says nothing about the inverse.
- The inverse is a **separate independent** statistic, ($\mathcal{R} \rightarrow \mathcal{D}$)
- The statistic is inherently **local**.
- The statistic is dependent on the number of points in the data set.
- $\varepsilon^*/\sigma_{std}$ is approximately the relative size of the **smallest** discontinuity we can detect.
- If ε^* scales with $\varepsilon_{\min}$, then this is further evidence of a continuous function.
- This is a **statistic**= evidence (or not) of a continuous function. Not a proof.

The continuity and differentiability statistics and other measures and RCs and embeddings

A Continuity Statistic (simple test)





$$\frac{dx}{dt}=\sigma(y-x)$$
$$\frac{dy}{dt}=-xz+\rho x-y$$
$$\frac{dz}{dt}=xy-\beta z$$

$\mathcal{D}$

$p_1 < 0,\ p_2 < 0$

$$\frac{dr_i}{dt} = \alpha[\kappa(p_1 r_i + p_2 r_i^2) + p_3 r_i^3 + \sum_{j=1}^{N} A_{ij} r_j + w_i x(t)]$$

damping factor

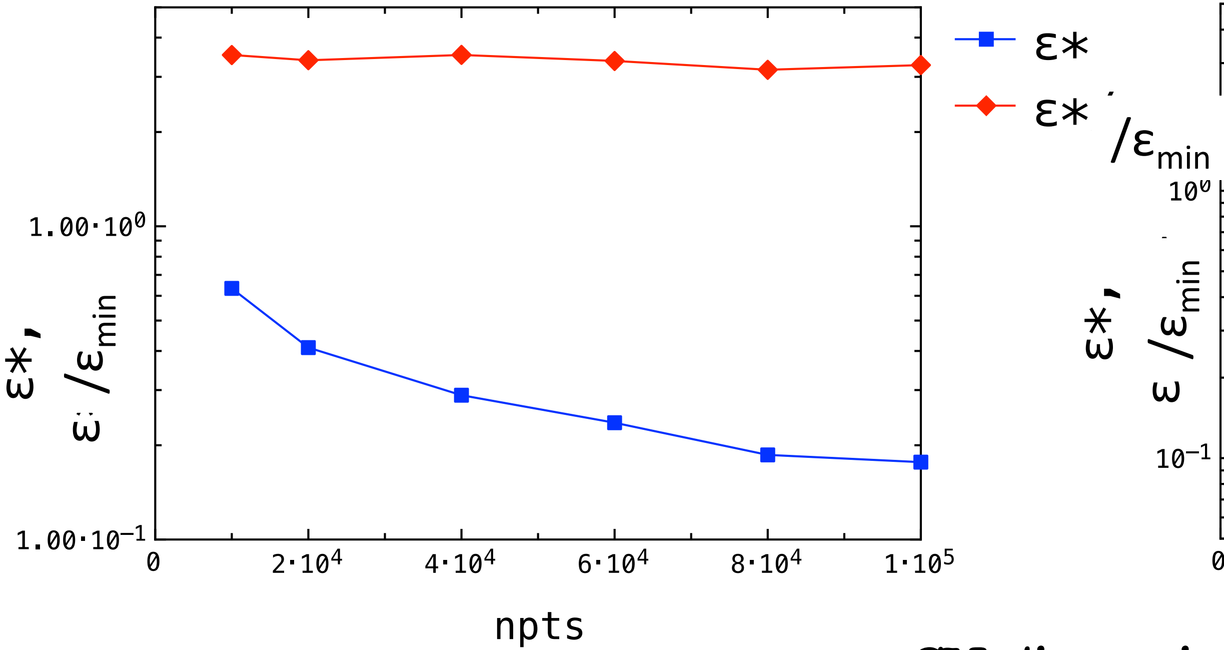
driving term

$\mathcal{R}$

κ=1.0

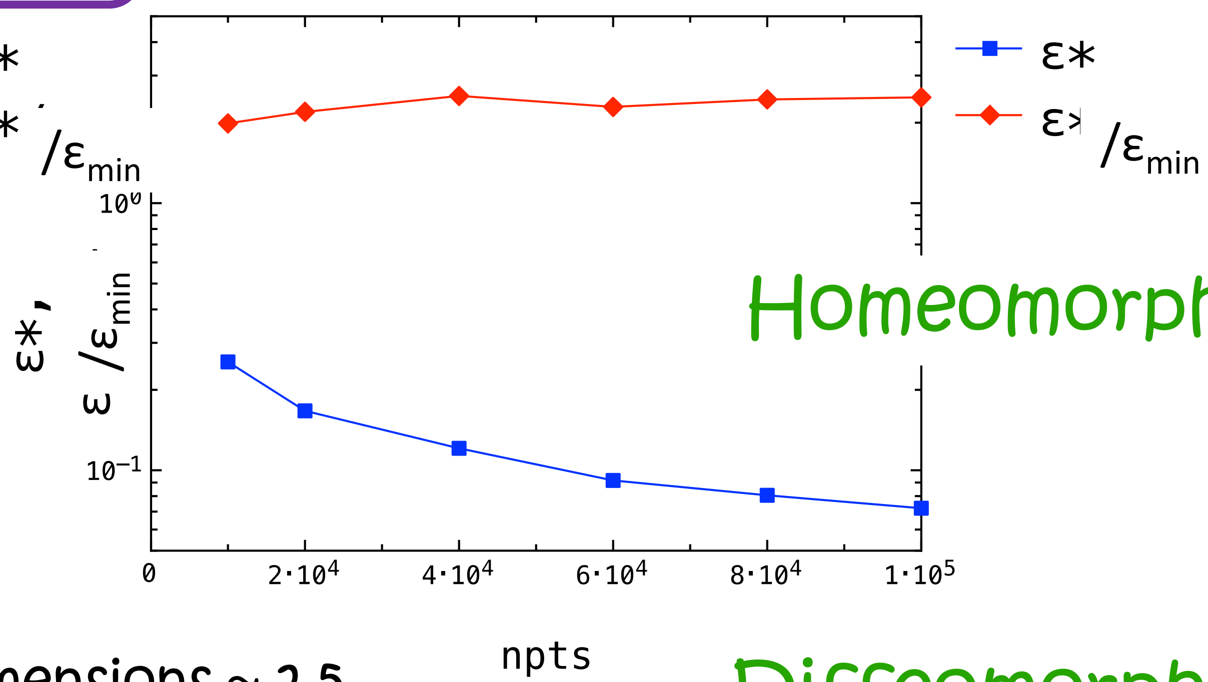
$\mathcal{D} \longrightarrow \mathcal{R}$

Drive -> Reservoir (ε*)



$\mathcal{R} \longrightarrow \mathcal{D}$

Reservoir -> Drive (ε*)



Homeomorphism

SV dimensions ~ 2.5

Diffeomorphism

Training and Testing errors and continuity statistic (40 K points)

x drive

$$\frac{dx}{dt} = \sigma(y - x)$$

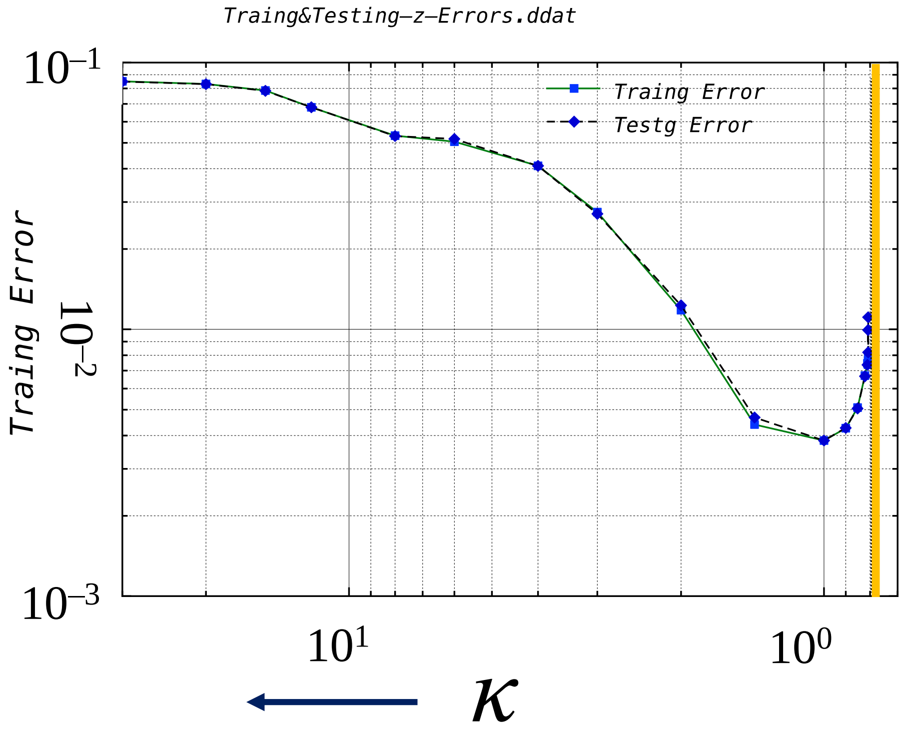
$$\frac{dy}{dt} = -xz + \rho x - y$$

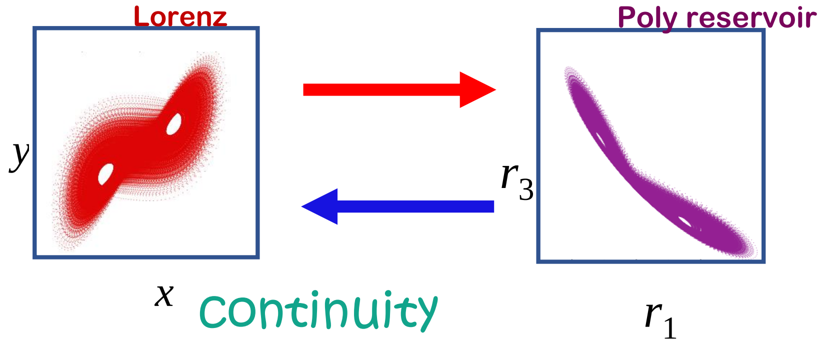
output *z*

$$\frac{dz}{dt} = xy - \beta z$$

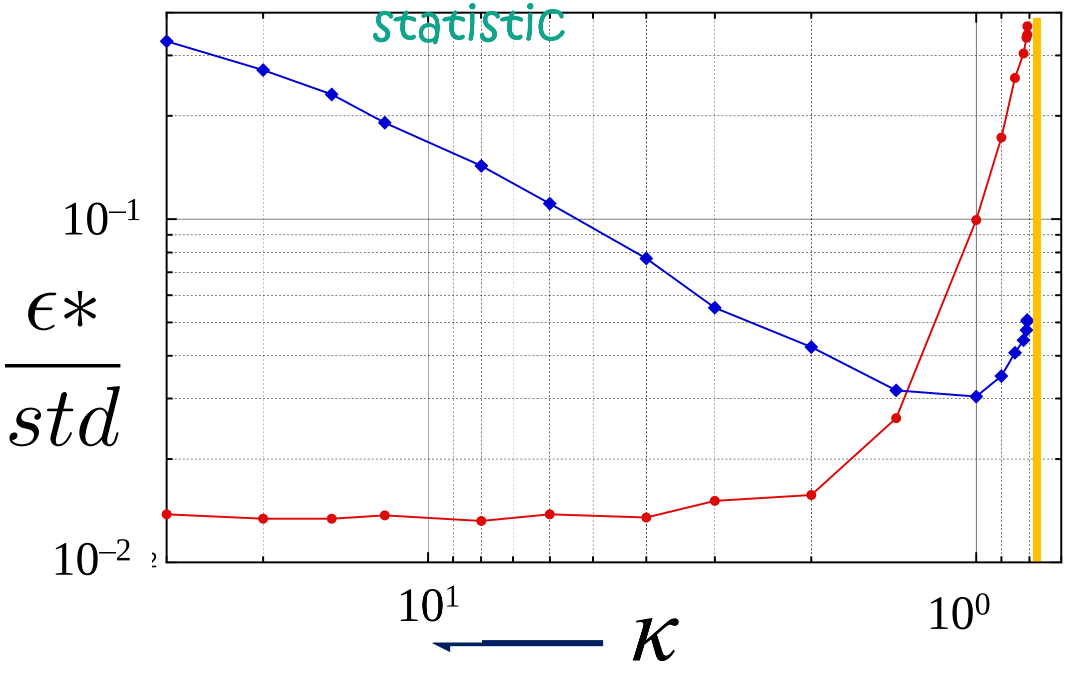
$$\frac{dr_i}{dt} = \alpha[\kappa(p_1 r_i + p_2 r_i^2) + p_3 r_i^3 + \sum_{j=1}^{N} A_{ij} r_j + w_i x(t)]$$

κ

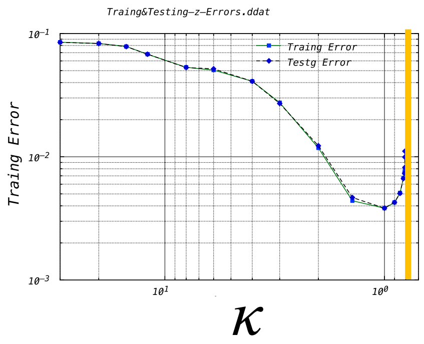


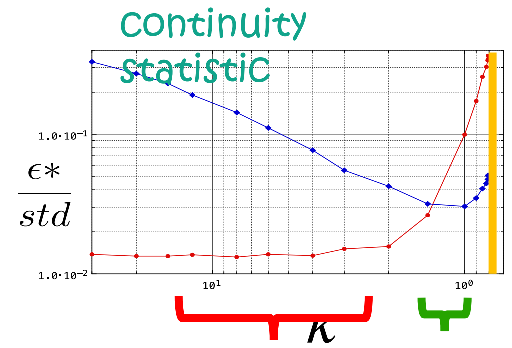


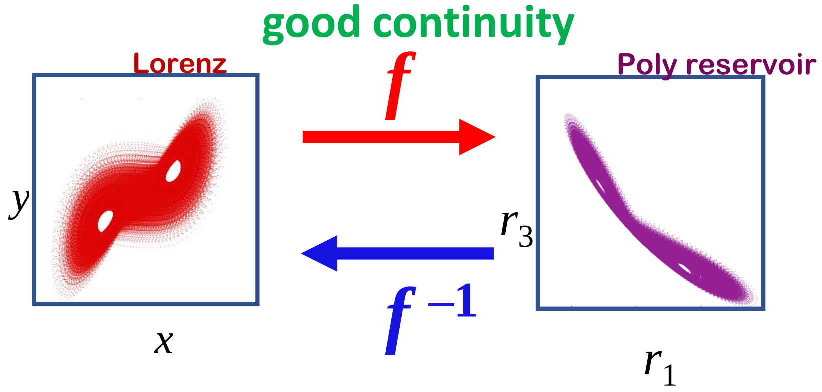
Continuity statistic



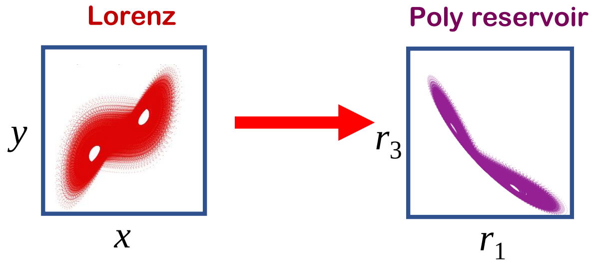
continuity and dynamics







Large dissipation

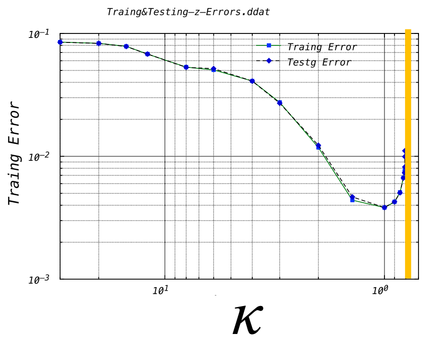


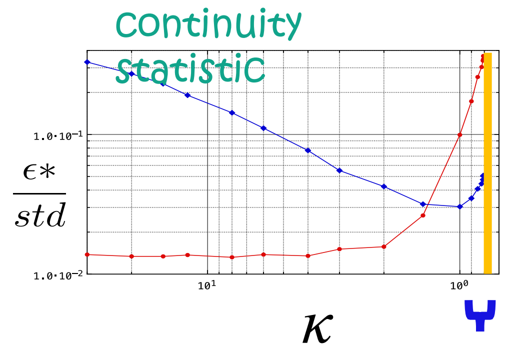
RC points are squeezed down to small region.

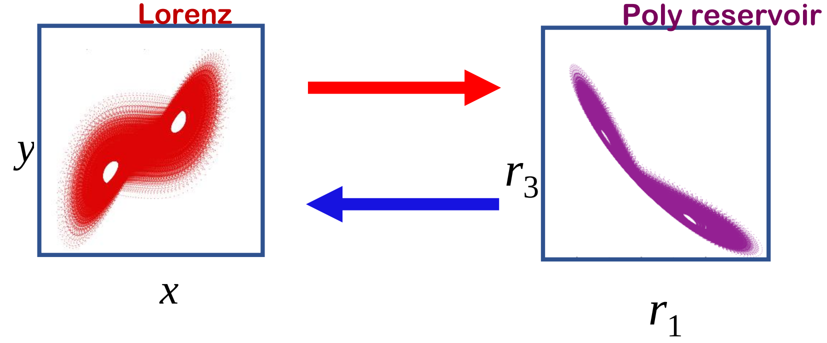
$$\mathbf{r}(t) = \Phi_\tau[\Psi_\tau(\mathbf{x}(t-\tau)), \Phi_\tau[\Psi_\tau(\mathbf{x}(t-2\tau)), \ldots \Phi_\tau[\Psi_\tau(\mathbf{x}(t-n\tau))]\ldots]]$$

Under-embedding

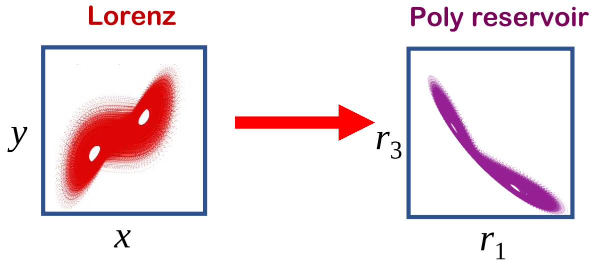
continuity and dynamics





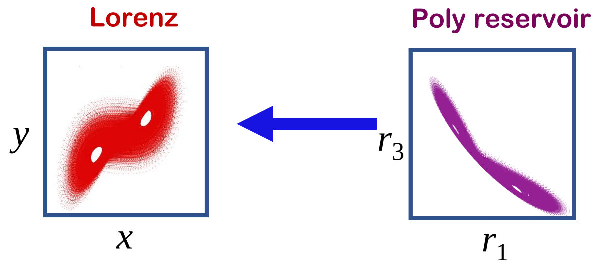


Small dissipation



nearby Lorenz points are spread out on the RC

Inverse= contraction



$$\mathbf{r}(t) = \Phi_\tau[\Psi_\tau(\mathbf{x}(t-\tau)), \Phi_\tau[\Psi_\tau(\mathbf{x}(t-2\tau)), ..., \Phi_\tau[\Psi_\tau(\mathbf{x}(t-n\tau))]...]]$$

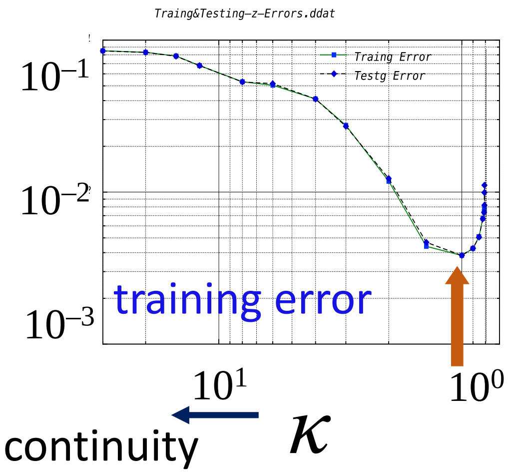
Over-embedding

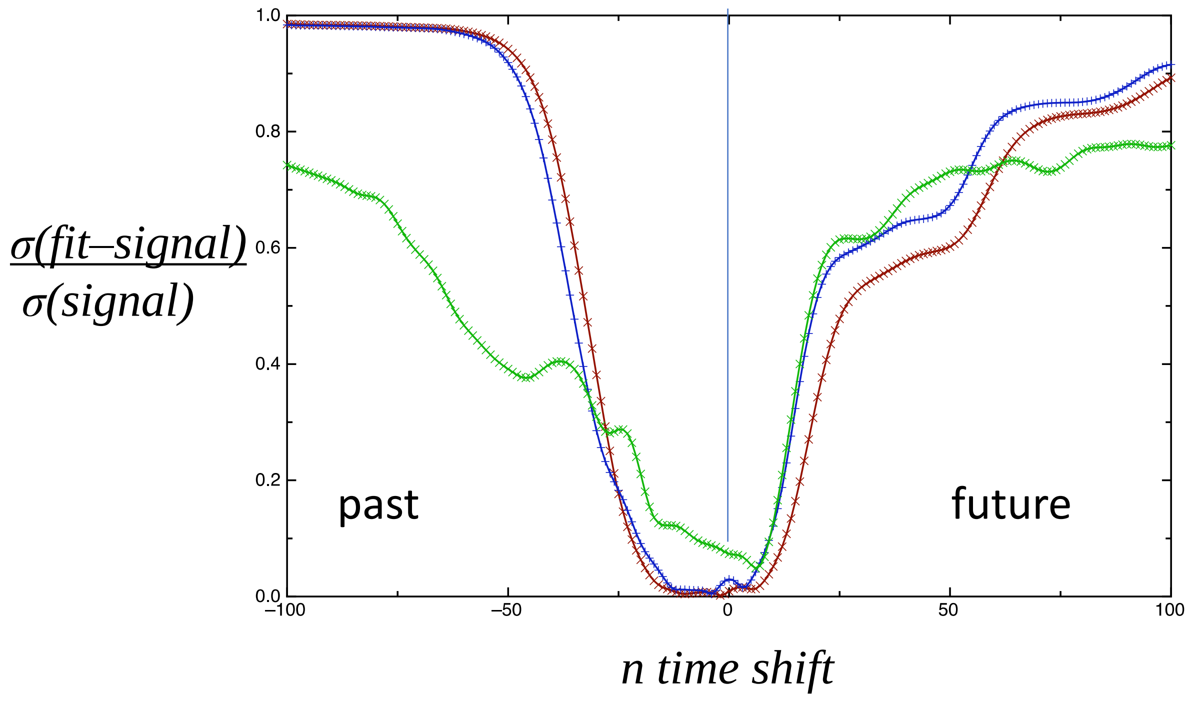
Predicting and Postdicting drive components

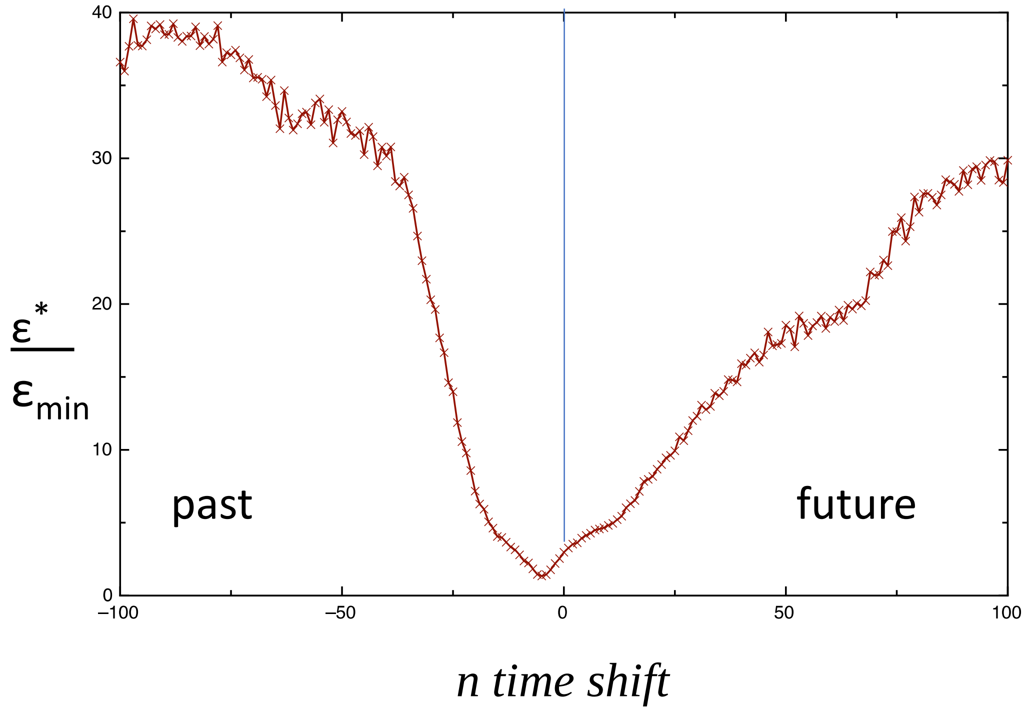
Continuity and training error with time shifts

Predicting and **Postdicting** (predicting into the past)

$\kappa = 1.0$







Fit errors trend matches continuity

Postdiction captures the "fading memory" quantitatively

Detecting Basins of Attraction

Bifurcations and Basins of Attraction

Lorenz -> Polynomial (deg.3) (Lor_Poly)

$$\frac{dx}{dt} = \sigma(y - x)$$
$$\frac{dy}{dt} = -xz + \rho x - y$$
$$\frac{dz}{dt} = xy - \beta z$$

$$\frac{dr_i}{dt} = \alpha\left[p_1 r_i + p_2 r_i^2 + p_3 r_i^3 + \sum_{j=1}^{N} A_{ij} r_j + u_i x\right]$$

driving term

training x,y,z

p_1: -7.0, -6.0, -5.0, -4.0, -3.0, -2.0, -1.0, -0.5

edge of chaos

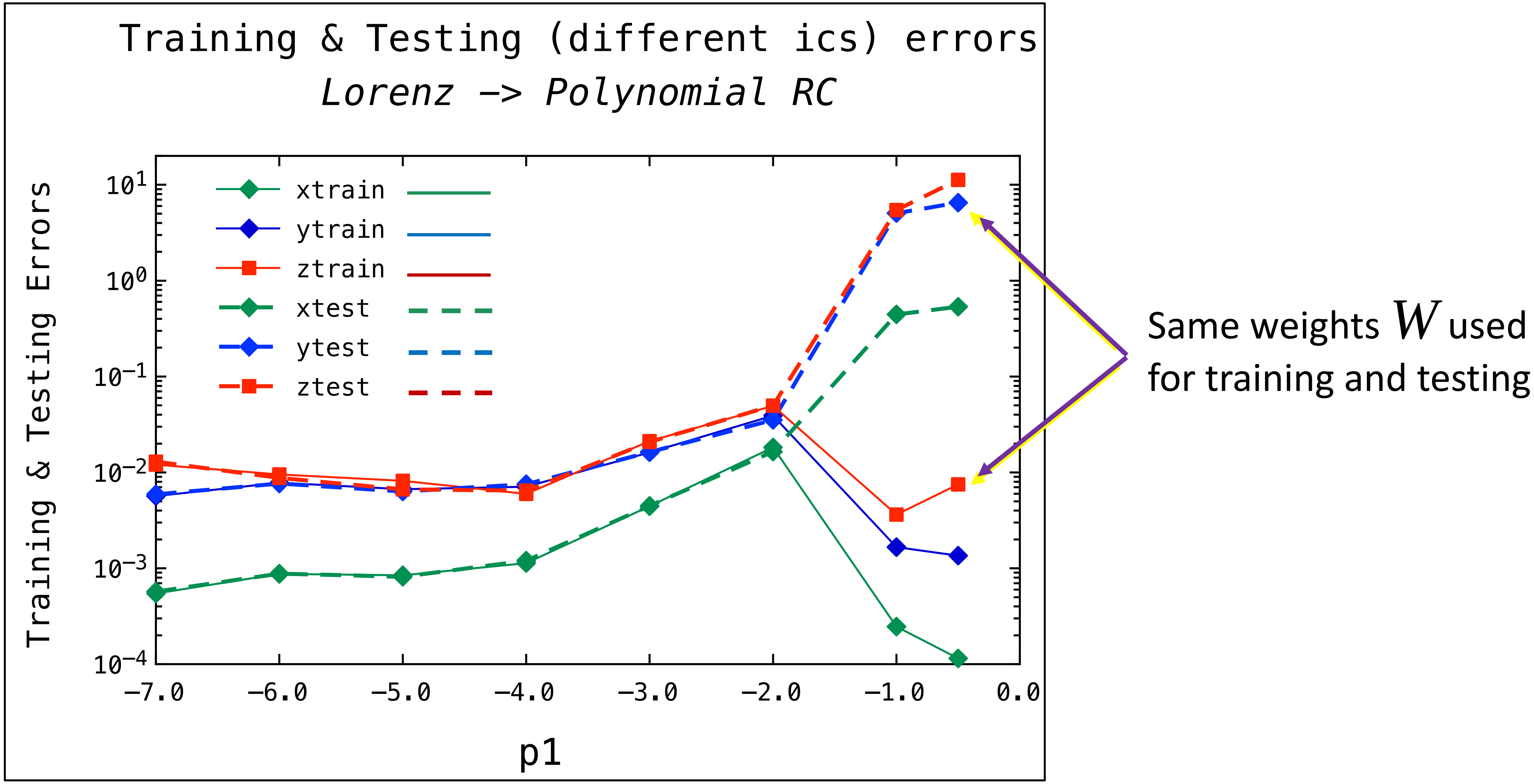
$p_2 = +3.0$

$p_3 = -1.0$

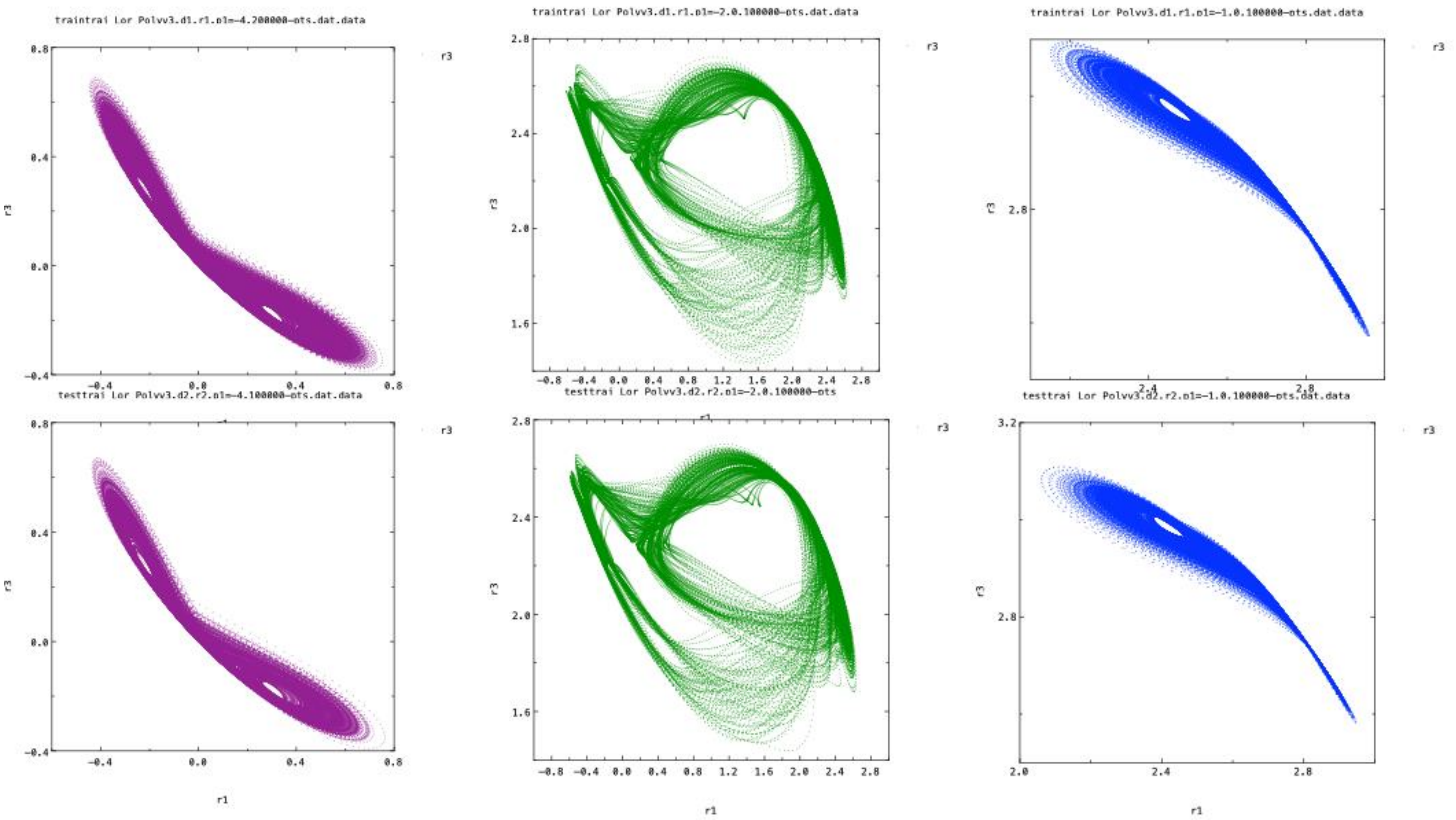
- 40000 points in time series
- training error,
- testing errors – both **time shifted** and **different ics**!
- continuity statistic

Training & Testing errors
Lorenz -> Polynomial (deg.3) (Lor_Poly)

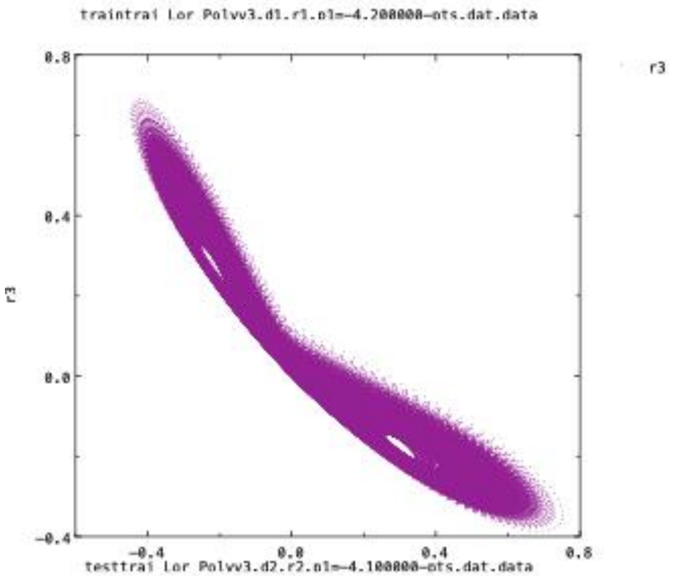
$$\frac{dr_i}{dt} = \alpha \left[p_1 r_i + p_2 r_i^2 + p_3 r_i^3 + \ldots \right]$$

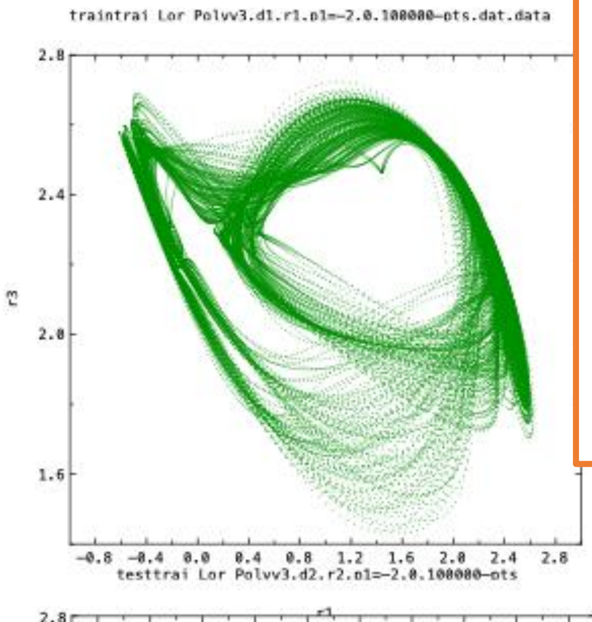


Reservoir Trajectories (Polynomial degree 3)



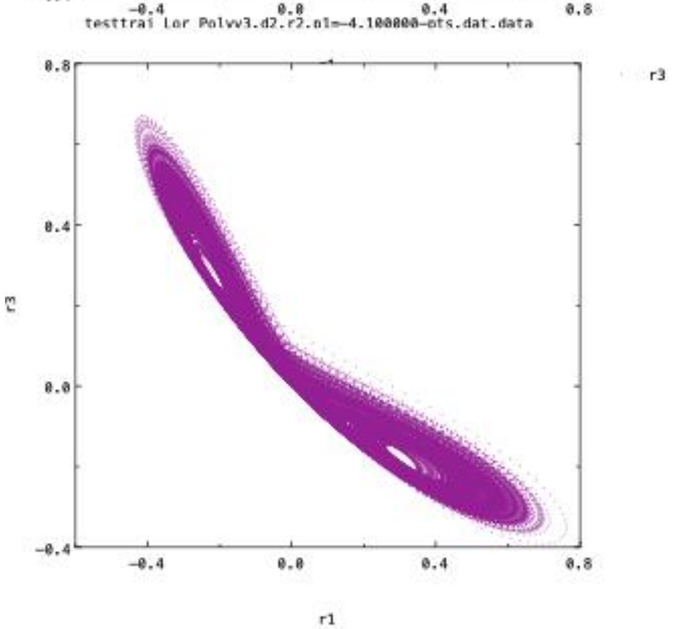
Reservoir Trajectories (Polynomial degree 3)

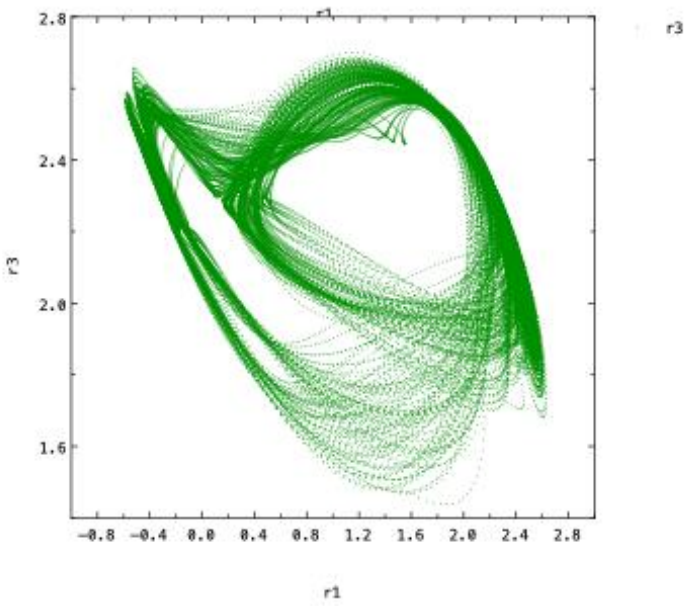


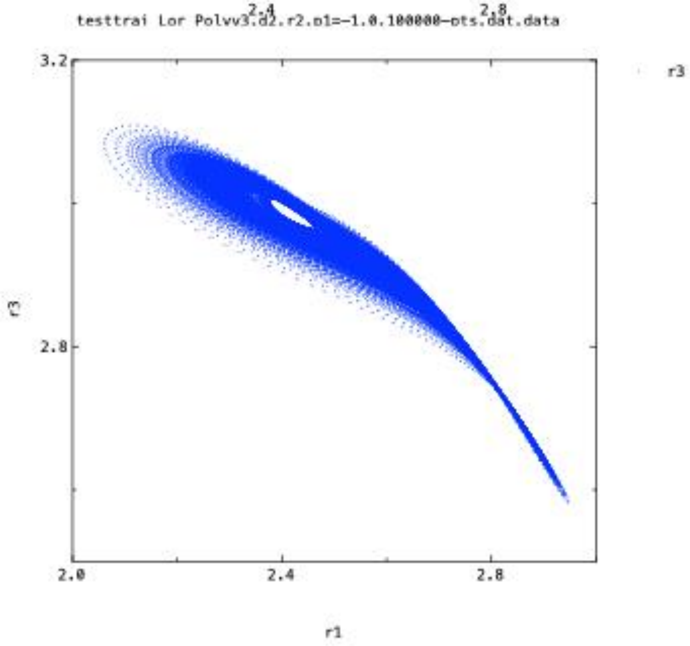


!! for p1=-1.0, appears like attractors are slightly shifted from each other. Maybe very long transients? They are close, but not the same. Or the synchronization invoked by the p3 term is not exact since that term →0 faster than linear leaving the system to wander when close, but not forced to get fully "synchronized" Or, the attractor dimension maybe getting much larger given the smaller Lyapunov exponents (check this) and a lot of them near the same value so perturbations (numerical) can push the trajectory into many directions.

Follow up: Actually at least two of the coordinates did not match at all => basin statistic is correct.







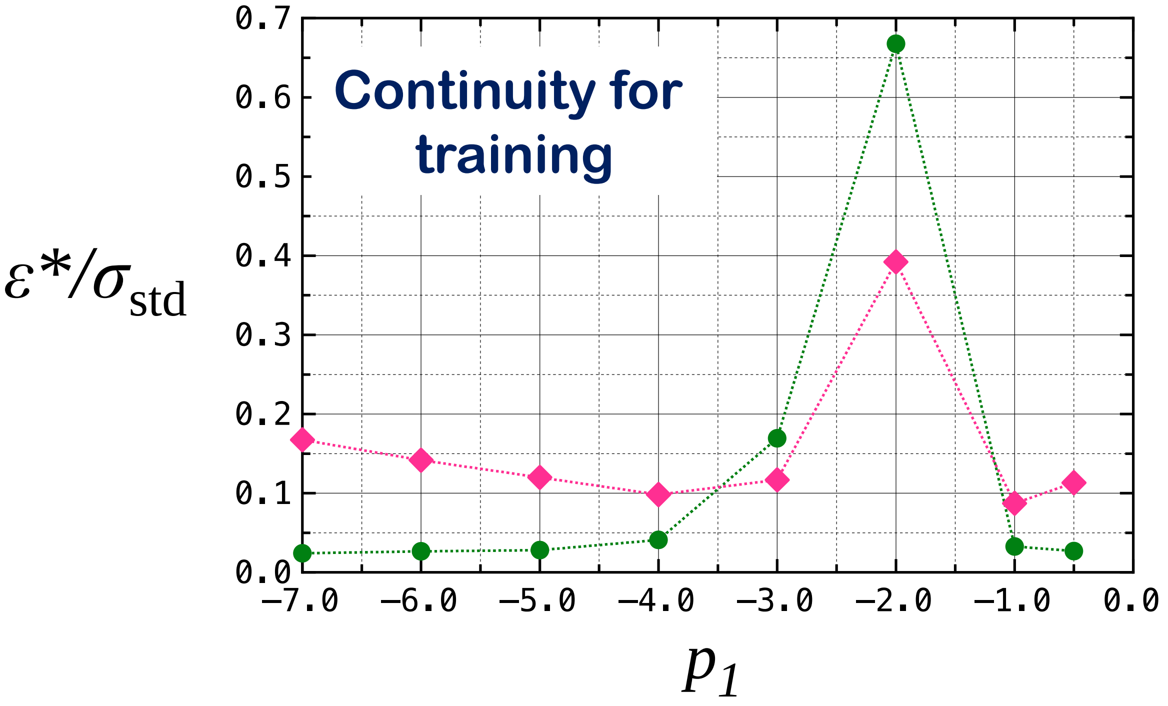
Continuity Statistic

$$\frac{dx}{dt}=\sigma(y-x)$$
$$\frac{dy}{dt}=-xz+\rho x-y$$
$$\frac{dz}{dt}=xy-\beta z$$

$\mathcal{D}$

$$\frac{dr_i}{dt}=\alpha\left[p_1 r_i+p_2 r_i^2+p_3 r_i^3+\sum_{j=1}^{N} A_{ij} r_j+u_i x\right]$$

$\mathcal{R}$



● $\mathcal{D}\longrightarrow\mathcal{R}$

◆ $\mathcal{D}\longleftarrow\mathcal{R}$

SV dim Statistic has same trends

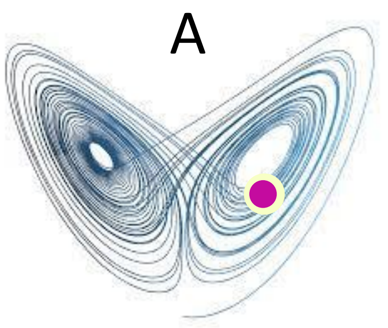
Same continuity results for testing

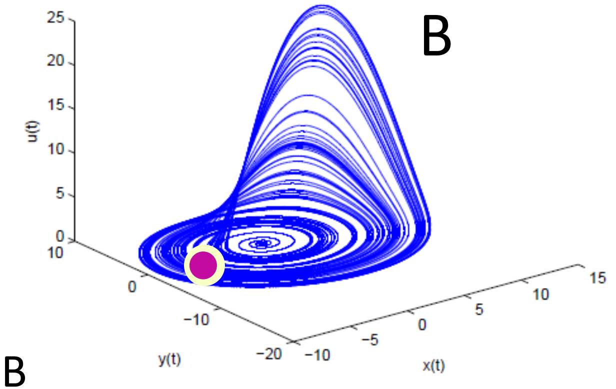
Are the training and testing time series on the same attractor?

Attractor Comparison Statistic

(including different basins of attraction-
same dynamics, same parameters, <u>different</u> ics.)

- 100 dimensional systems
- Do NOT de-mean, shift or rescale (std, etc.)





A → B

Get nearest neighbor(s) on B to point on A
and calc. distances from B point to A point

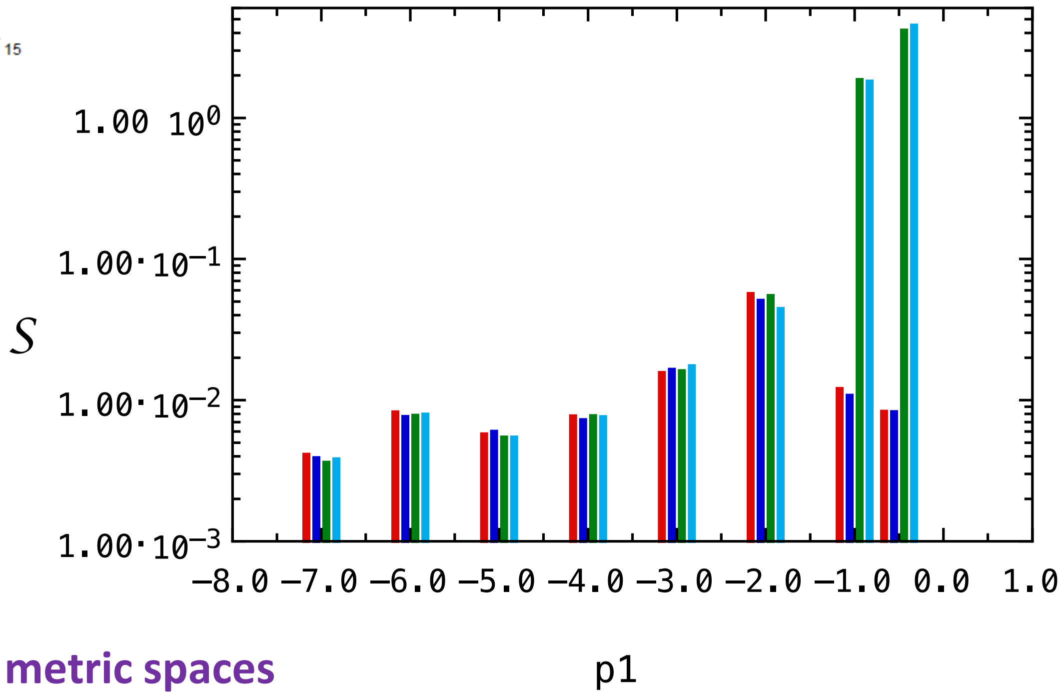
Do this for several points (1000)
and calc. average distance= S

Do this for B → A

Do this for train and test reservoirs $\mathcal{R}$

Test $\mathcal{R}$ → *Train* $\mathcal{R}$
Train $\mathcal{R}$ → *Test* $\mathcal{R}$
} different attractors

Train ($\mathcal{R}$→$\mathcal{R}$ time shifted)
Train ($\mathcal{R}$ time shifted→$\mathcal{R}$)
} same attractor

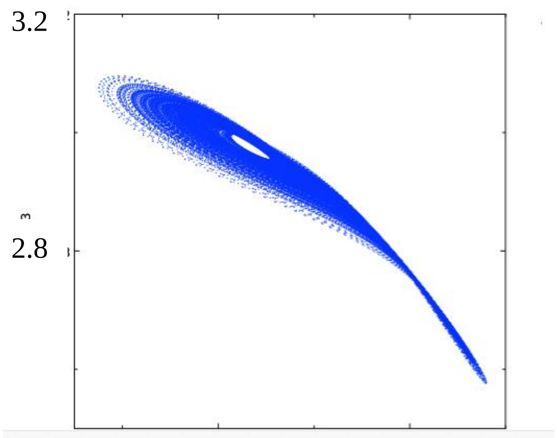


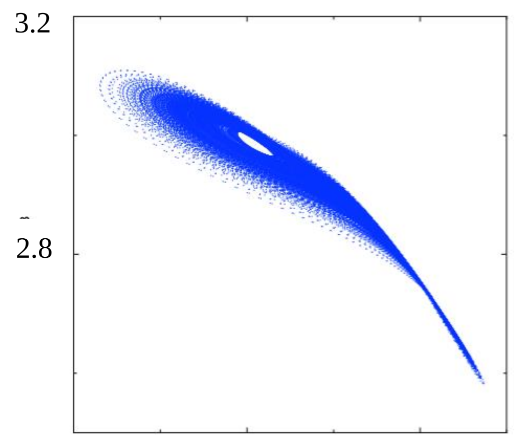
Similar to distances between point sets in metric spaces

Attractor Comparison Statistic

training attractor *testing attractor*

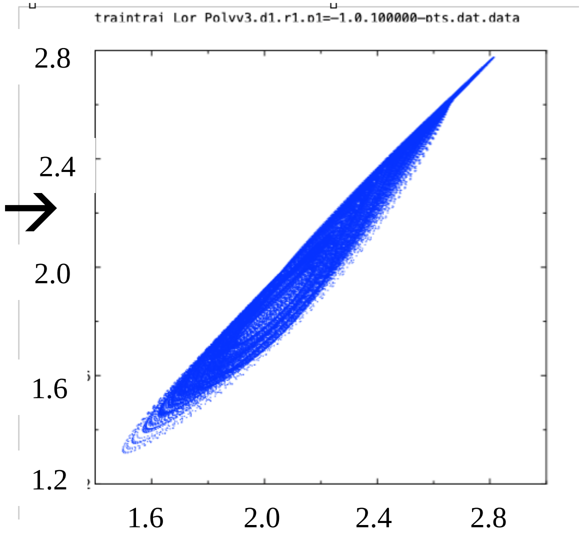
r_3 vs. r_1 →

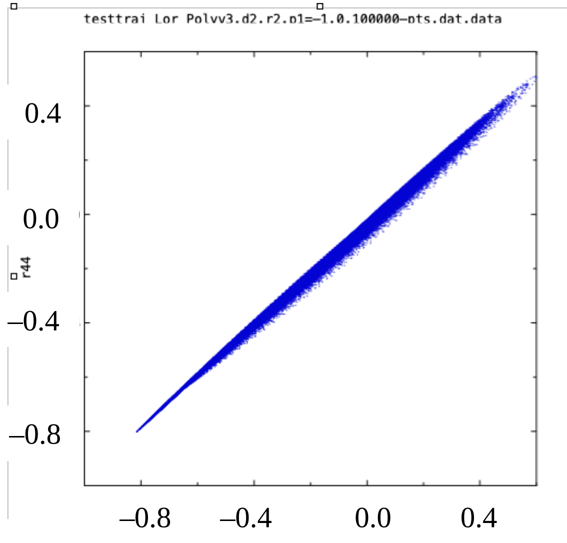




training attractor *testing attractor*

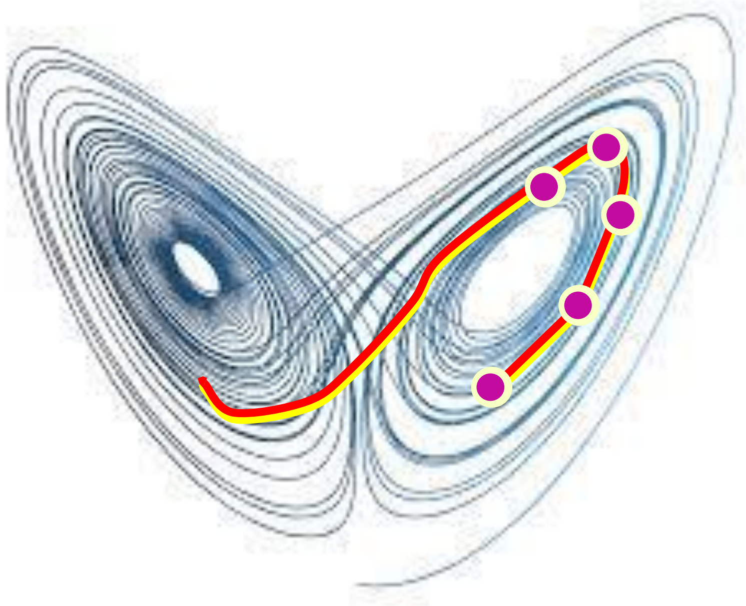
r_{37} vs. r_{44} →



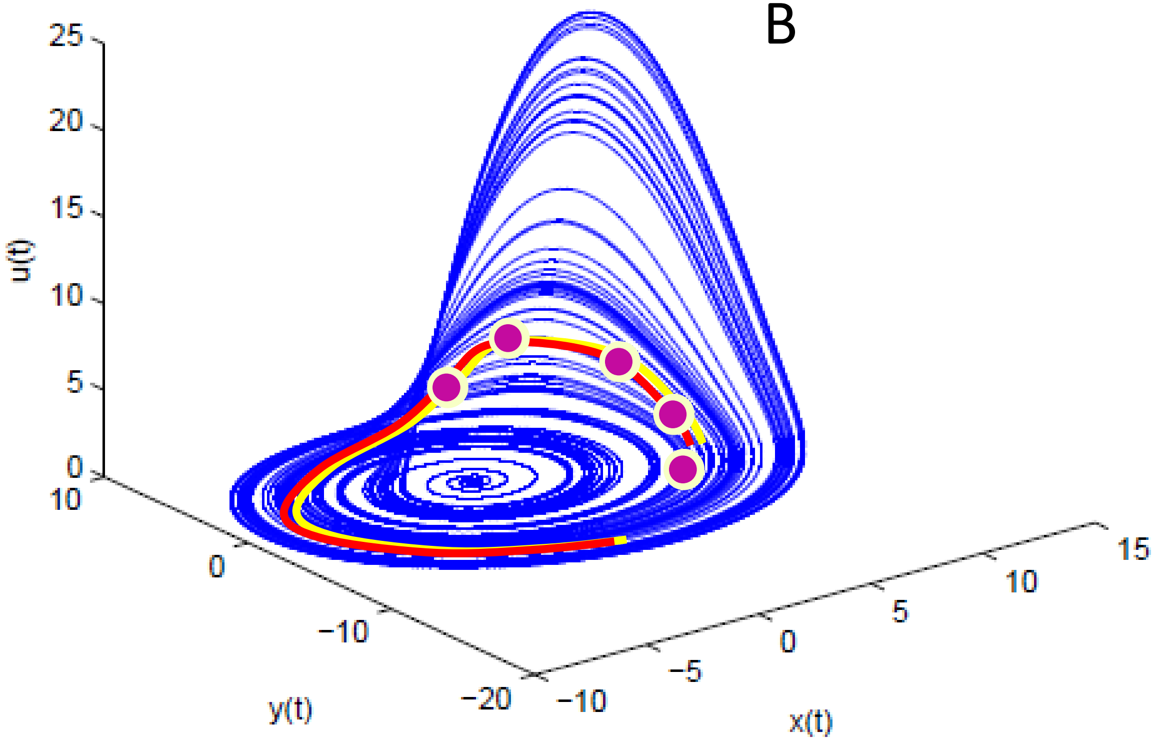


Adding to the robustness of the ACS

A



B



Don't forget the dynamics!

Conclusions

- Even in simple RC systems nonlinear phenomena are important and nonlinear analysis captures the behavior quantitatively.

- The computer science/AI communites have taken network dynamics in an interesting and potentially useful direction, but the analysis of these systems must be informed by nonlinear dynamics.

- We don't always have accurate models or theorems. Need statistics that are modeled on mathematical concepts (continuity and differentiability) and make no more assumptions than necessary.

- Reservoir properties cannot all be measured independent of the drive signals. Dynamical properties (memory, synchronization, attractor embeddings, stability) are all linked to the drive and the RC.

Paper to the arXiv soon

Questions, comments ?

Tom Carroll, US Naval Research Laboratory

T. L. Carroll and L. M. Pecora, Network Structure Effects in Reservoir Computers, Chaos vol. 29, 083130
T. L. Carroll, Dimension of Reservoir Computers, Chaos vol. 30, 013102
T. L. Carroll, Path Length Statistics in Reservoir Computers, Chaos vol. 30, 083130
T. L. Carroll, Adding Filters to Improve Reservoir Computer Performance, Physica D vol. 416, 132798 (January 2021)
T. L. Carroll, Low Dimensional Manifolds in Reservoir Computers, Chaos vol. 31, 043113 March 2021
T. L. Carroll, Optimizing Reservoir Computers for Signal Classification, Frontiers in Physiology 12:685121

- "Low dimensional manifolds in reservoir computers", T. Carroll, **Chaos 31**, 043113 (2021)
- "Optimizing memory in reservoir computers", T. Carroll, **Chaos 32**, 023123 (2022);
- "Do reservoir computers work best at the edge of chaos?", T. Carroll, **Chaos 31**, 043113 (2021)