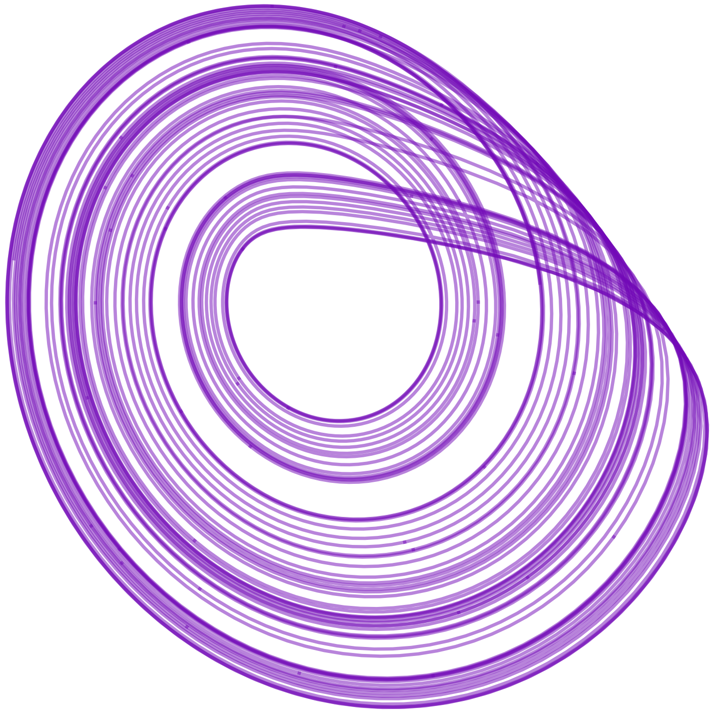
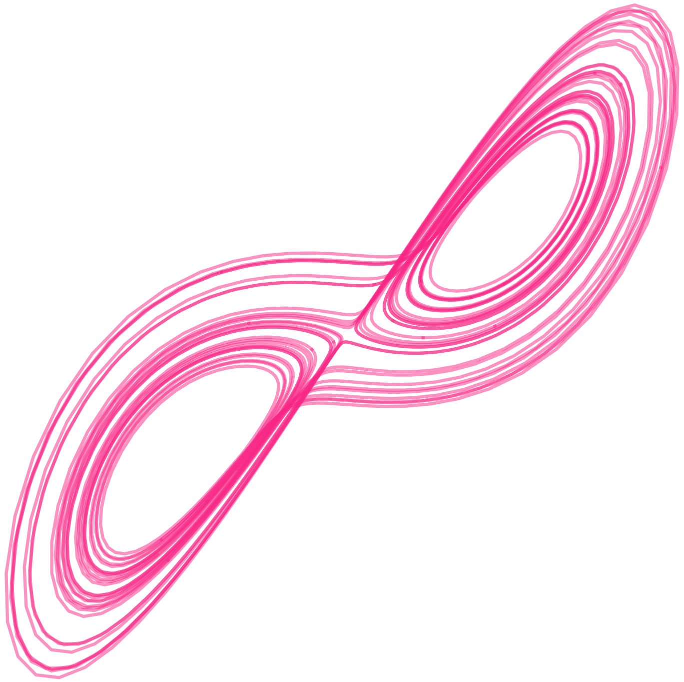
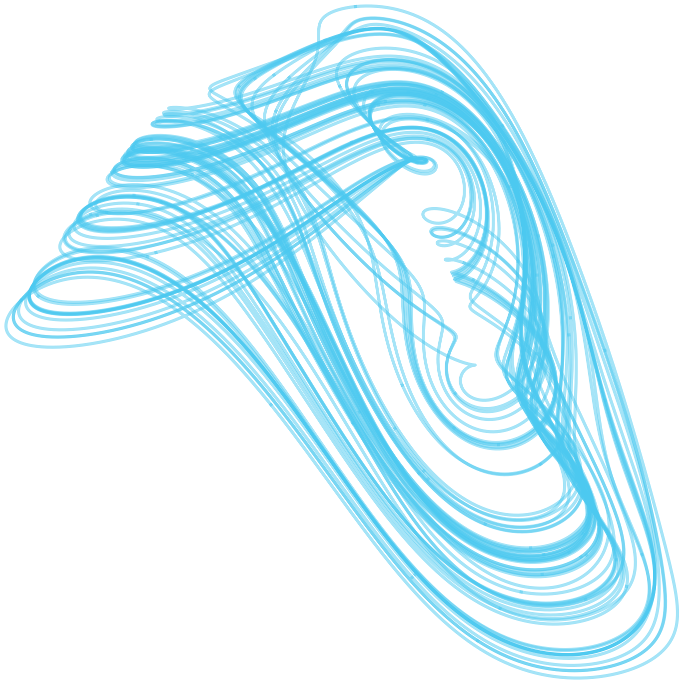
Model-free inference and zero-shot forecast in complex systems







HES-SO
June 2025

SANTA FE INSTITUTE

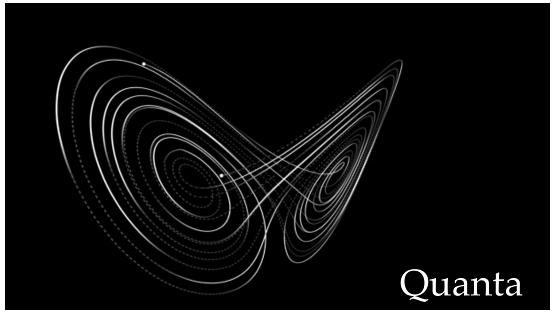
Yuanzhao Zhang
y-zhang.com

Simple

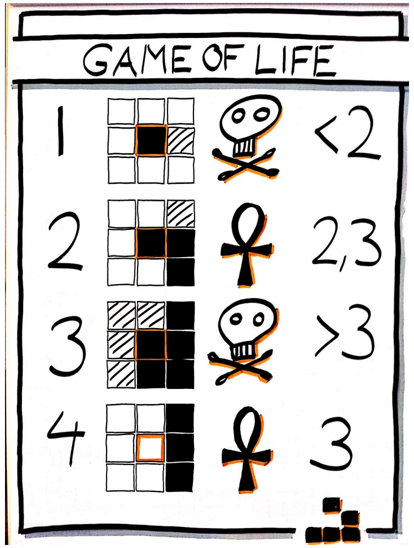
Complex

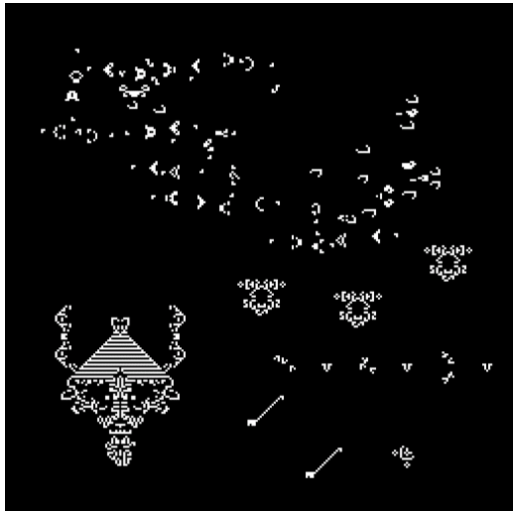
Lorenz system

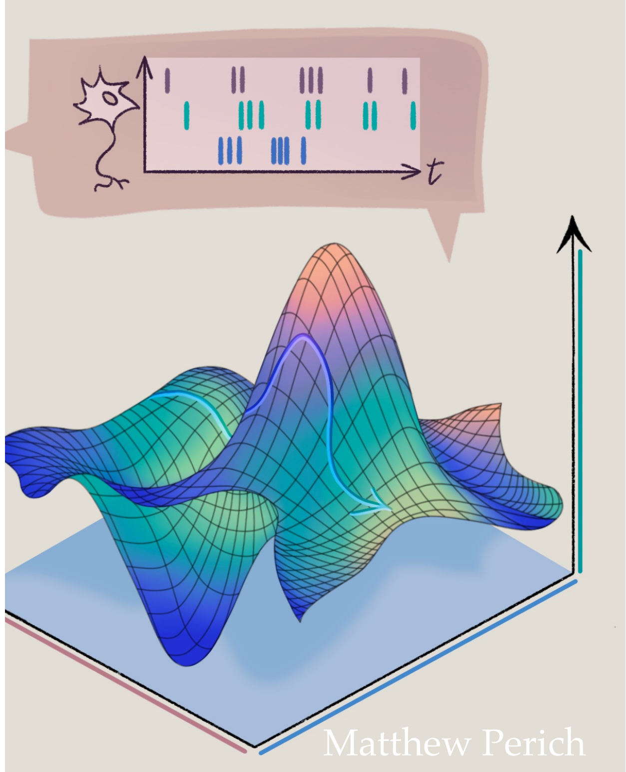
$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y - x)$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = x(\rho - z) - y$$
$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z$$



Game of Life



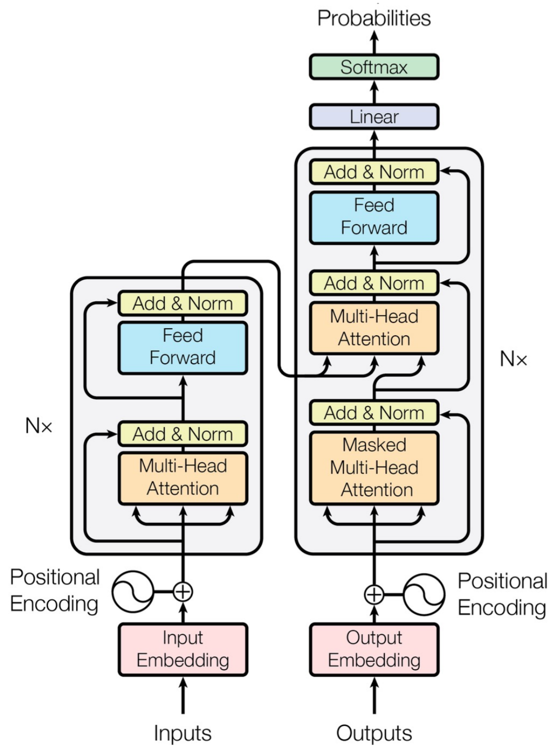




Neural manifolds

Complex

Simple

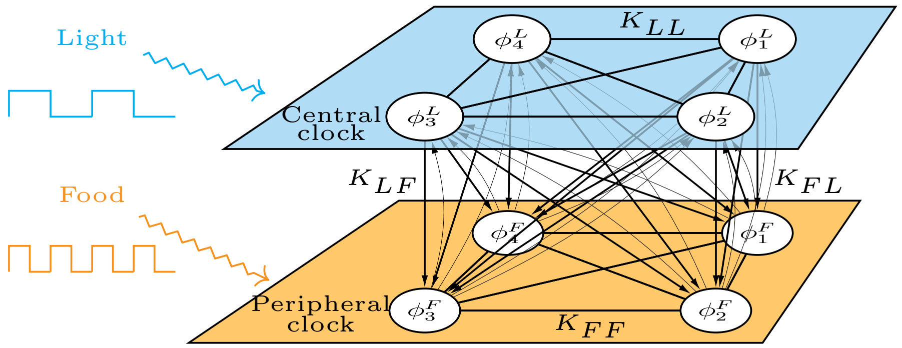


[A] [B] … [A] → [B]

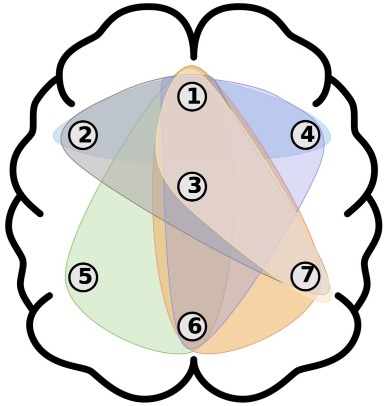
Induction heads

Brain as a dynamical system

Dynamics on neuronal networks



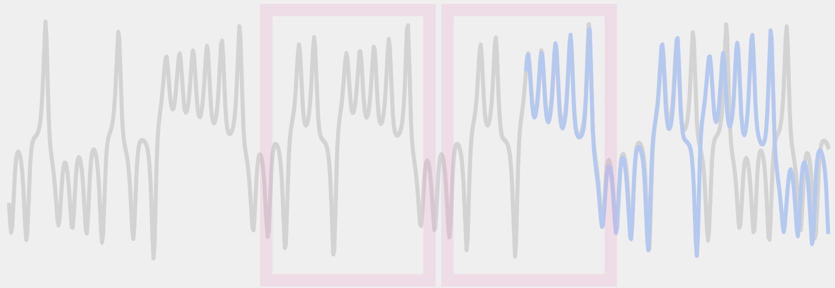
Networks from neuronal dynamics



Zero-shot forecasting of chaotic dynamics

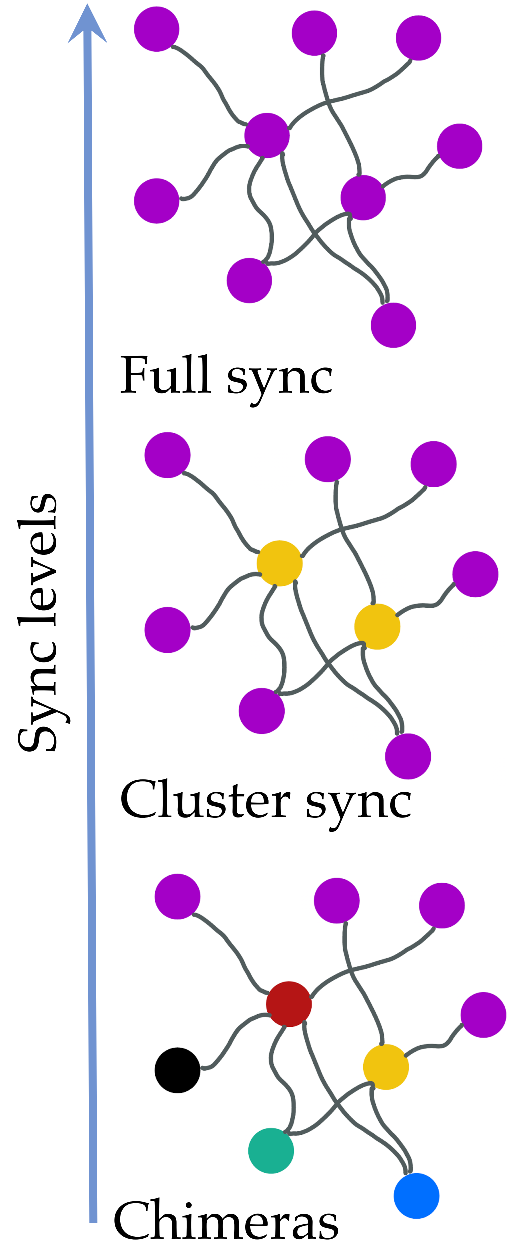
Foundation model as a tool for forecasting previously unseen dynamics

Foundation model as a "model organism" for learning from limited data




Veritasium

Brain as a dynamical system: Dynamics on neuronal networks



Zhang & Strogatz, Nat. Commun. 2021.

Sugitani, Zhang & Motter, PRL 2021.

Zhang, Lucas & Battiston, Nat. Commun. 2023.

Zhang & Motter, SIAM Rev. 2020.

Zhang & Strogatz, PRL 2021.

Zhang et al., Commun. Phys. 2021.

Zhang et al., PRX 2020.

Zhang & Motter, PRL 2021.

Zhang et al., Sci. Adv. 2024.

Circadian rhythm

Hannay, Forger & Booth, Sci. Adv. 2018.

Zhang et al., PNAS 2021.

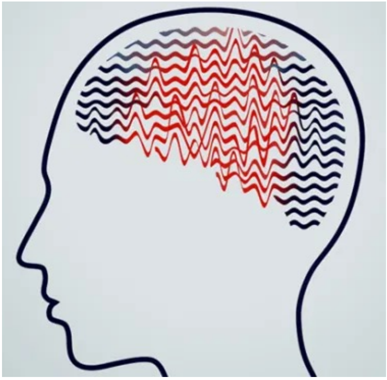
Huang, Zhang & Braun, Chaos 2023.

Memory

Salazar et al., Science 2012.

Jacob, Hähnke & Nieder, Neuron 2018.

Reinhart & Nguyen, Nat. Neurosci. 2019.



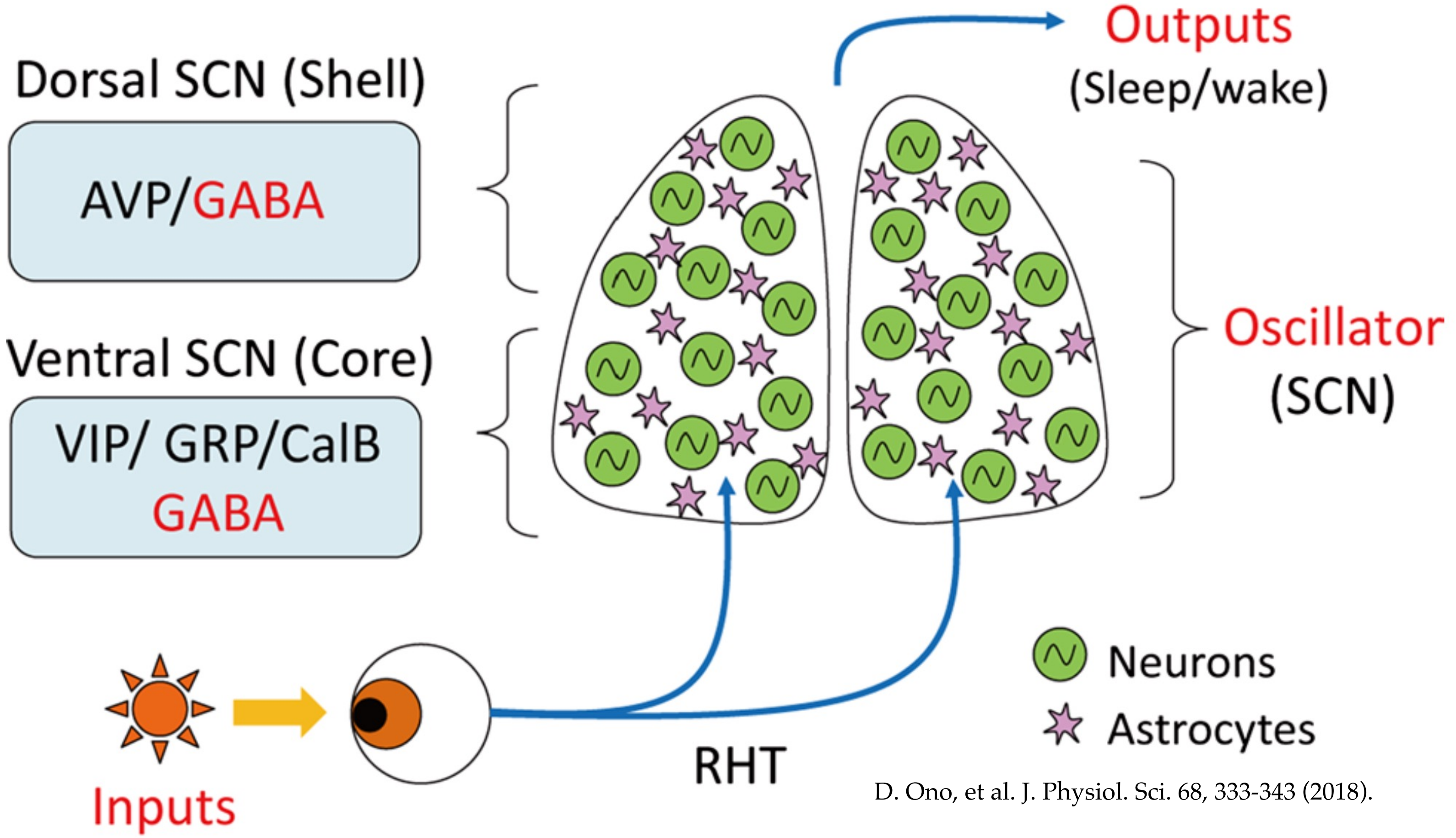
Seizure

Jirsa et al., Brain 2014.

Andrzejak et al., Sci. Rep. 2016.

Kuhlmann et al., Nat. Rev. Neurol 2018.

Synchronization: A bridge from the microscopic to the macroscopic



D. Ono, et al. J. Physiol. Sci. 68, 333-343 (2018).

Circadian rhythm is produced by the synchronized activity of about 20,000 neurons in the suprachiasmatic nucleus

Mathematical model of the central circadian clock

Phase of the oscillator (neuron)

Coupling term

Light stimulus

$$\frac{d\phi_k}{dt} = \omega_k + \frac{K}{N}\sum_{j=1}^{N} \sin(\phi_j - \phi_k) + L(t)Q(\phi_k)$$

Natural frequency

Phase response function
(inferred from data)

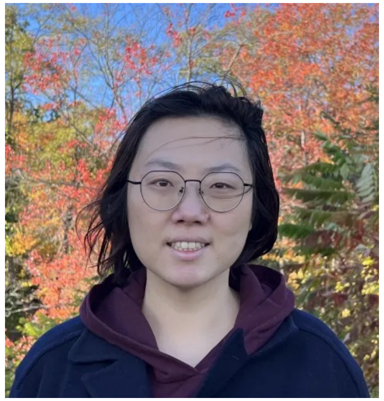
Although simple, it can explain a lot of the clinical/experimental observations

Lu et al., Chaos 2016.

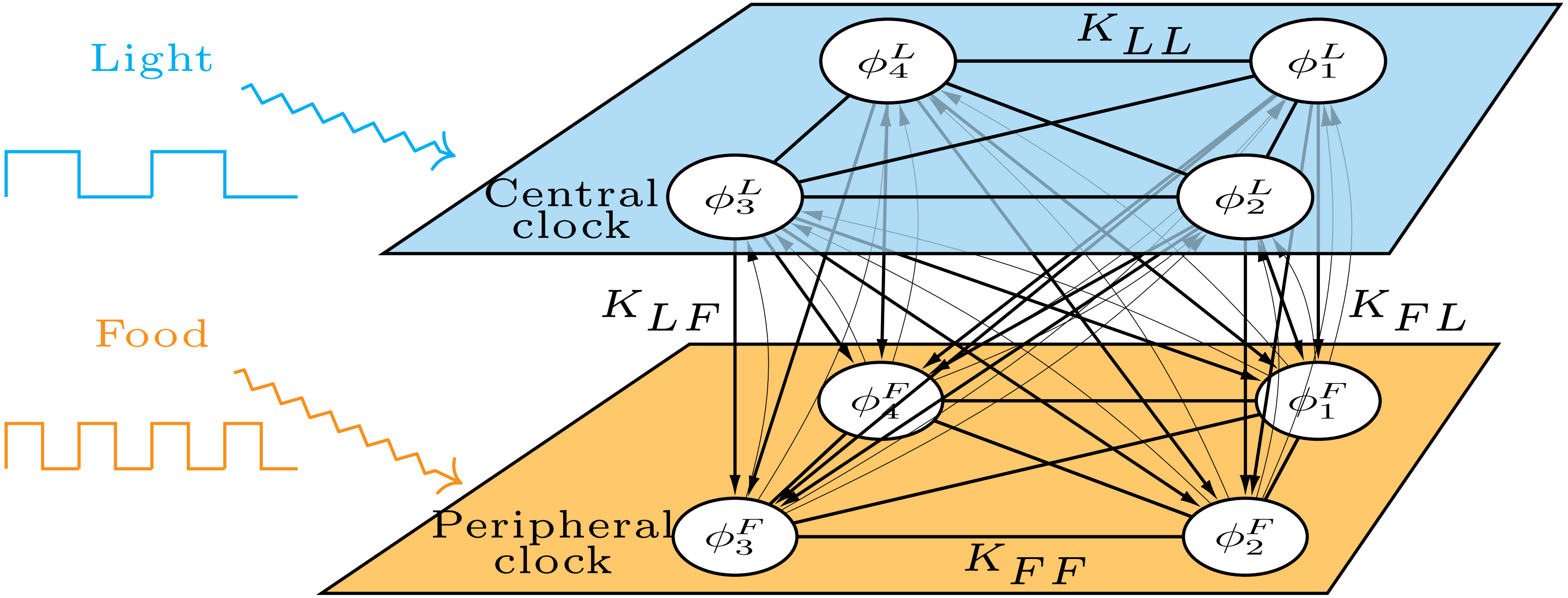
Hannay, Booth, and Forger, Sci. Adv. 2018.

Incorporating peripheral clocks

Rosemary Braun
Northwestern



Pepper Huang
Smith College

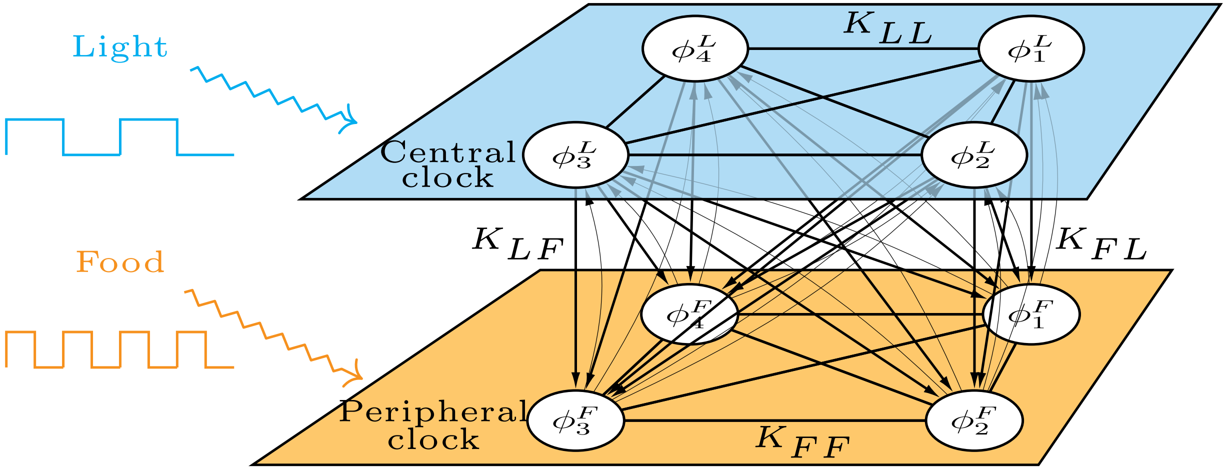


$$\frac{d\phi_k^F}{dt} = \omega_k^F + \underbrace{\frac{K_{FF}}{N_F}\sum_{j=1}^{N_F}\sin(\phi_j^F - \phi_k^F)}_{\text{intralayer coupling}} + \underbrace{\frac{K_{LF}}{N_L}\sum_{j=1}^{N_L}\sin(\phi_j^L - \phi_k^F + \alpha)}_{\text{interlayer coupling}} + \underbrace{F(t)M(\phi_k^F)}_{\text{food stimulus}},$$

$$\frac{d\phi_k^L}{dt} = \omega_k^L + \underbrace{\frac{K_{LL}}{N_L}\sum_{j=1}^{N_L}\sin(\phi_j^L - \phi_k^L)}_{\text{intralayer coupling}} + \underbrace{\frac{K_{FL}}{N_F}\sum_{j=1}^{N_F}\sin(\phi_j^F - \phi_k^L - \alpha)}_{\text{interlayer coupling}} + \underbrace{L(t)Q(\phi_k^L)}_{\text{light stimulus}}.$$

Huang, **Zhang**, and Braun, Chaos 2023.

Reduced model on invariant manifold answers new questions



$$\frac{d\phi_k^F}{dt} = \omega_k^F + \underbrace{\frac{K_{FF}}{N_F}\sum_{j=1}^{N_F}\sin(\phi_j^F - \phi_k^F)}_{\text{intralayer coupling}} + \underbrace{\frac{K_{LF}}{N_L}\sum_{j=1}^{N_L}\sin(\phi_j^L - \phi_k^F + \alpha)}_{\text{interlayer coupling}} + \underbrace{F(t)M(\phi_k^F)}_{\text{food stimulus}},$$

$$\frac{d\phi_k^L}{dt} = \omega_k^L + \underbrace{\frac{K_{LL}}{N_L}\sum_{j=1}^{N_L}\sin(\phi_j^L - \phi_k^L)}_{\text{intralayer coupling}} + \underbrace{\frac{K_{FL}}{N_F}\sum_{j=1}^{N_F}\sin(\phi_j^F - \phi_k^L - \alpha)}_{\text{interlayer coupling}} + \underbrace{L(t)Q(\phi_k^L)}_{\text{light stimulus}}.$$

There is a hidden four-dimensional manifold that is **invariant** and **attracting** under the dynamics

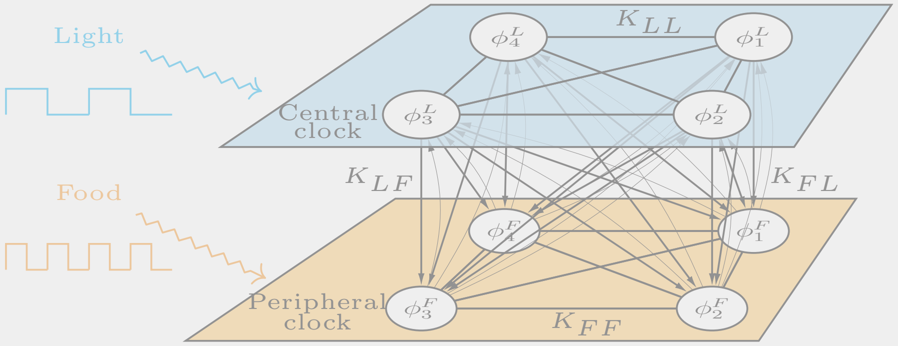
Reduce the model from 20,000+ coupled ODEs to 4 coupled ODEs with physiologically meaningful macroscopic variables

The model allows us to ask interesting new questions about the effect of competing stimuli

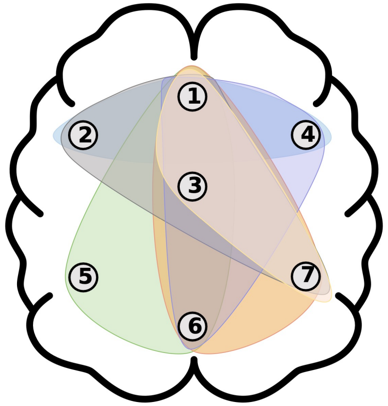
E.g., use food to combat jet lag

Brain as a dynamical system

Dynamics on neuronal networks

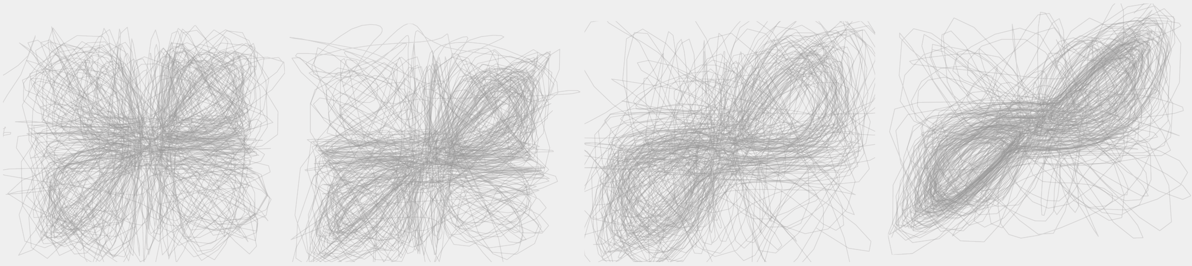


Networks from neuronal dynamics

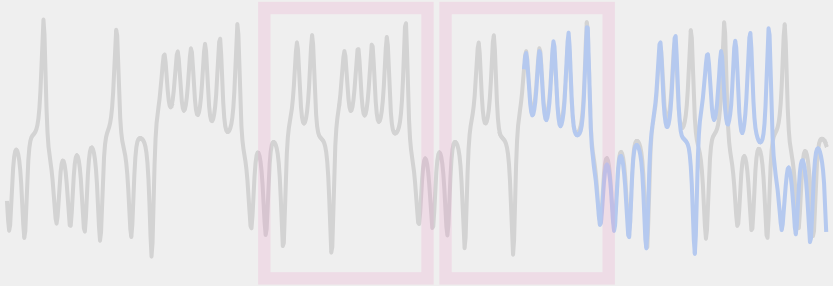


Zero-shot forecasting of chaotic dynamics

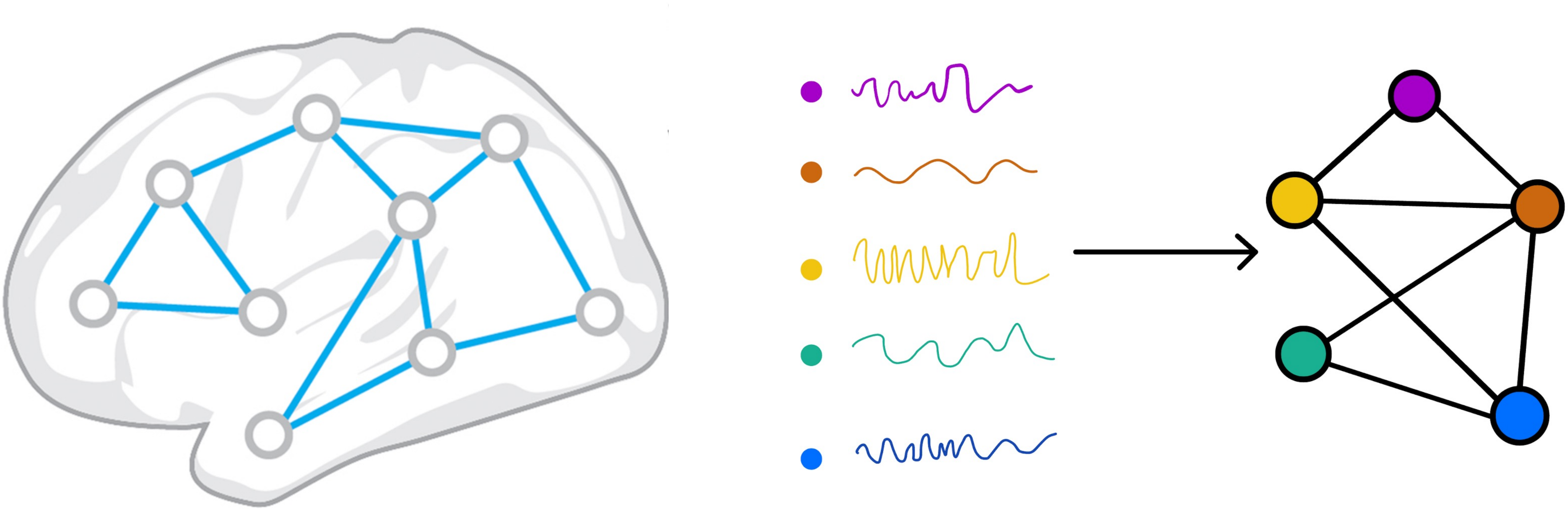
Foundation model as a tool for forecasting previously unseen dynamics



Foundation model as a “model organism” for learning from limited data

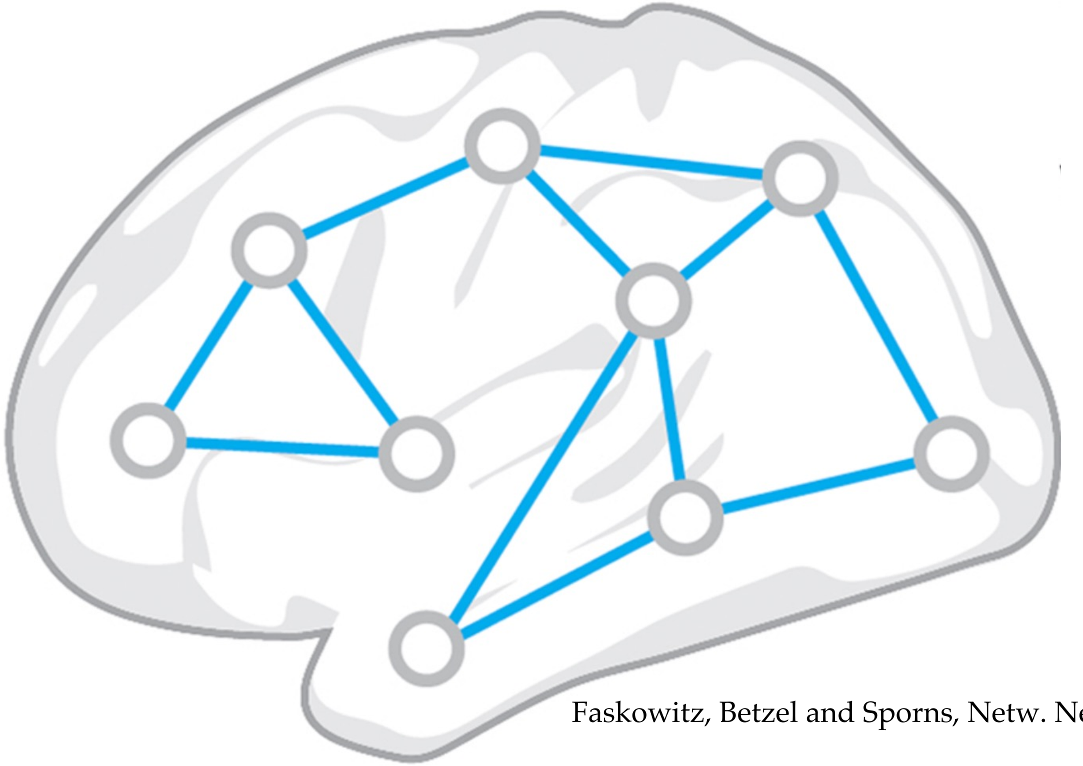


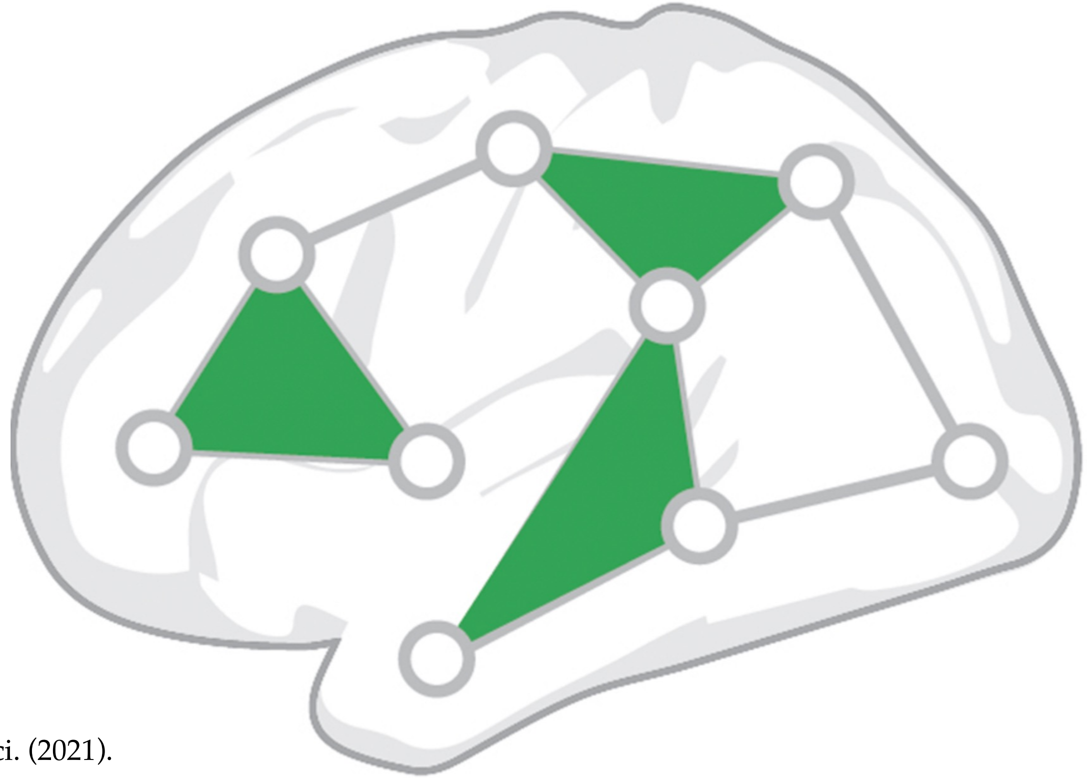
Brain as a dynamical system: Networks from neuronal dynamics



Faskowitz, Betzel and Sporns, Netw. Neurosci. (2021).

How important are higher-order interactions in the brain?



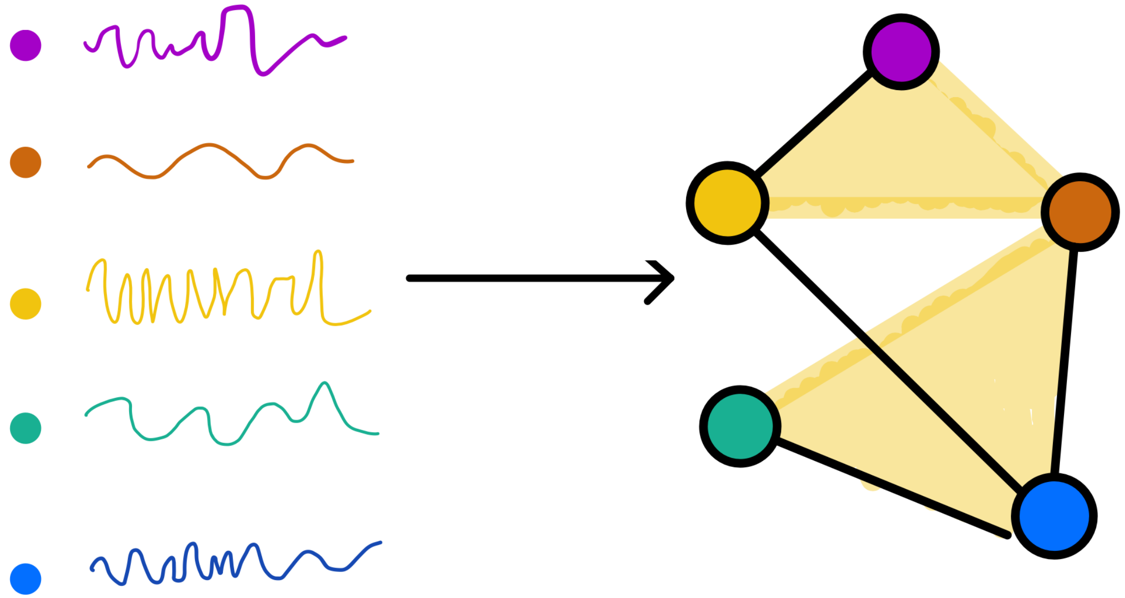


Faskowitz, Betzel and Sporns, Netw. Neurosci. (2021).

Is the brain more like this?

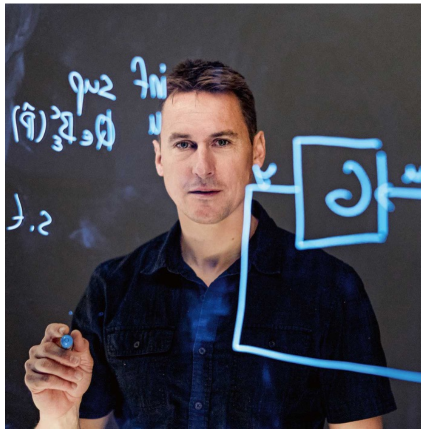
Or that?

We need a method that can infer higher-order interactions from time-series data



Robin Delabays
HES-SO

Giulia De Pasquale
TU Eindhoven



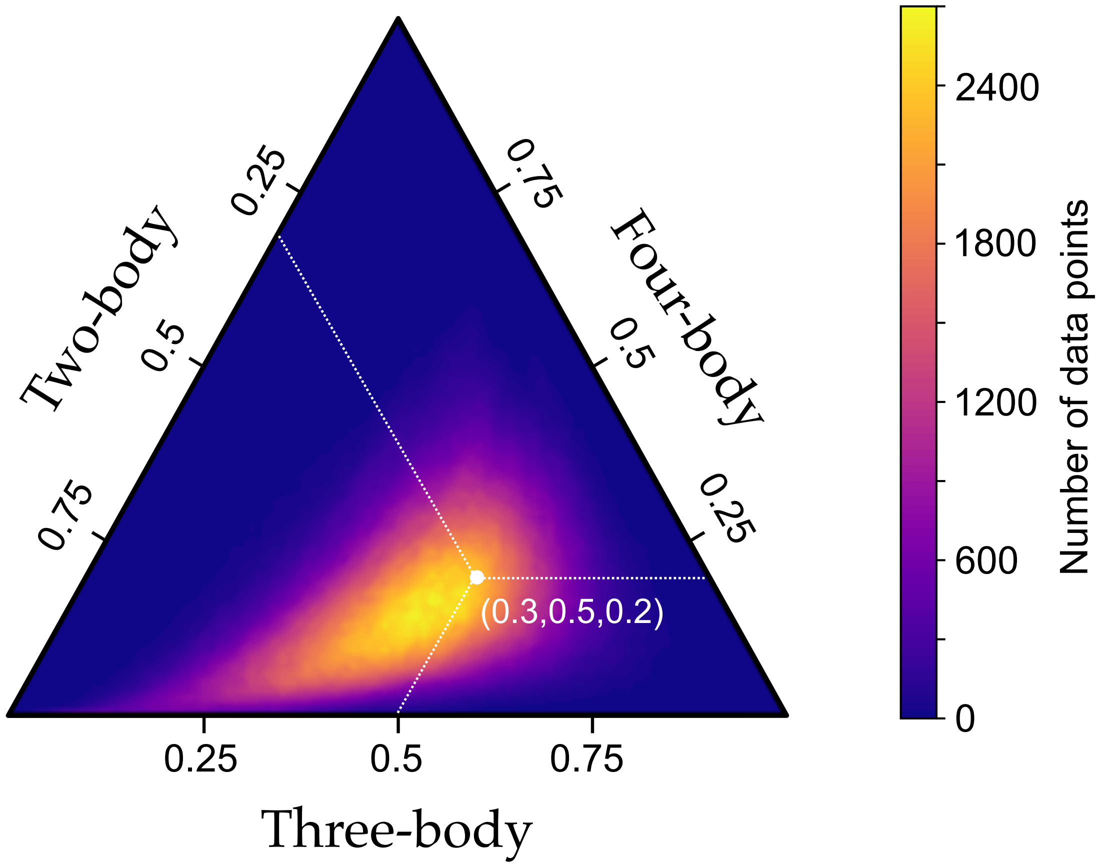
Florian Dörfler
ETH Zürich

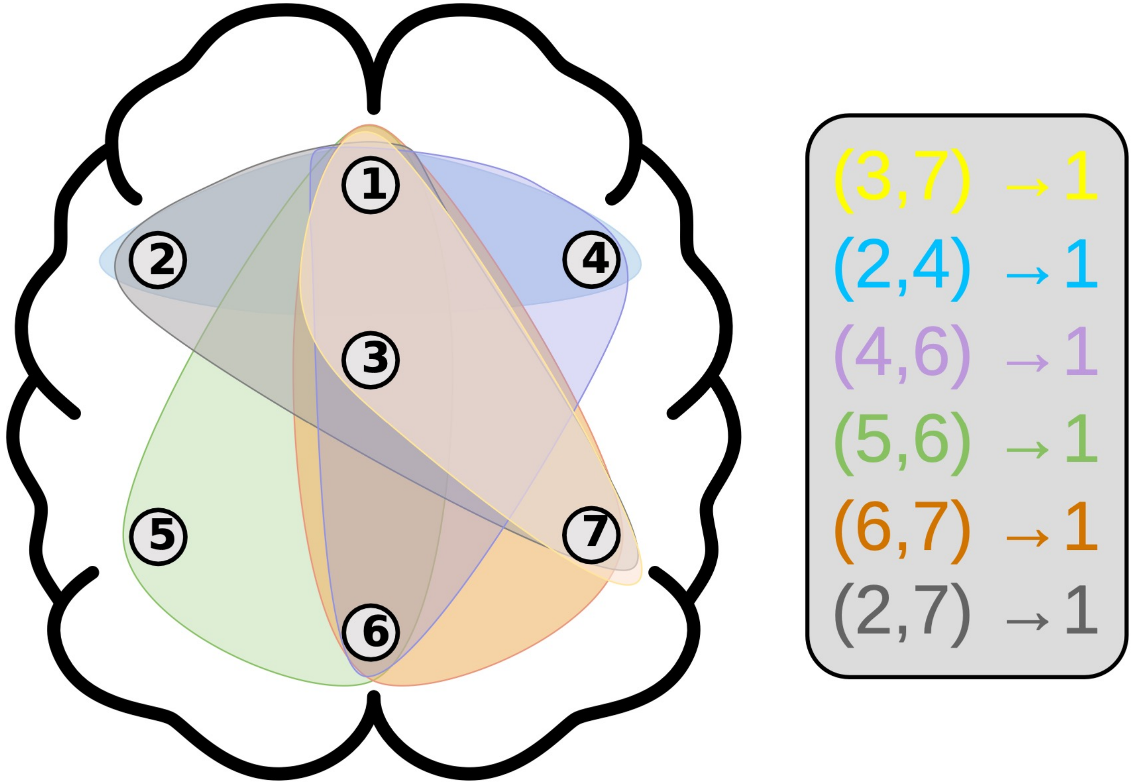
- A generalization of the (causal) network inference problem
- Must be model free: Because we don’t have a reliable model for brain dynamics!
- Key challenge: How to distinguish a triangle and a 2-simplex from dynamics?
- Key idea: Taylor expansion and sparse regression

Delabays, De Pasquale, Dörfler, and **Zhang**, Nat. Commun. 2025

Higher-order interactions shape macroscopic brain dynamics

Relative contribution from each order of interaction

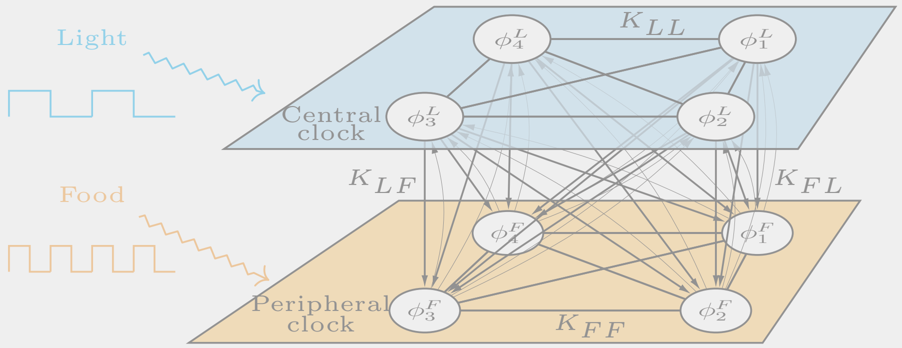




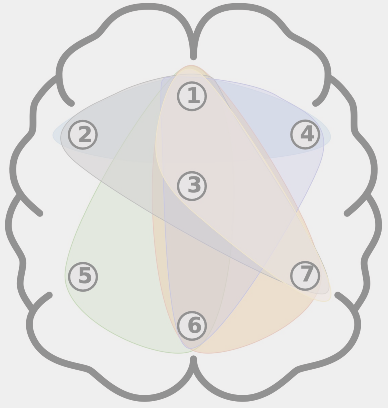
- Resting-state EEG data from 109 subjects
- Divide the brain into 7 regions, infer up to the fourth-order interactions
- Around 60% of the dynamics are explained by nonpairwise interactions
- The six most prominent (directed) hyperedges all point toward area 1 (roughly the prefrontal cortex)!

Brain as a dynamical system

Dynamics on neuronal networks

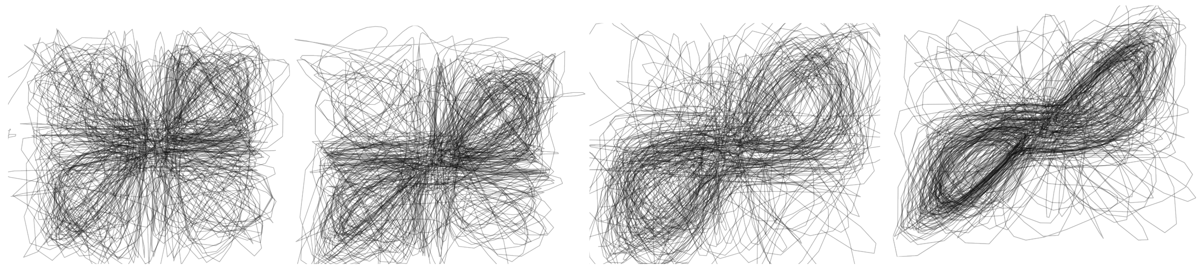


Networks from neuronal dynamics

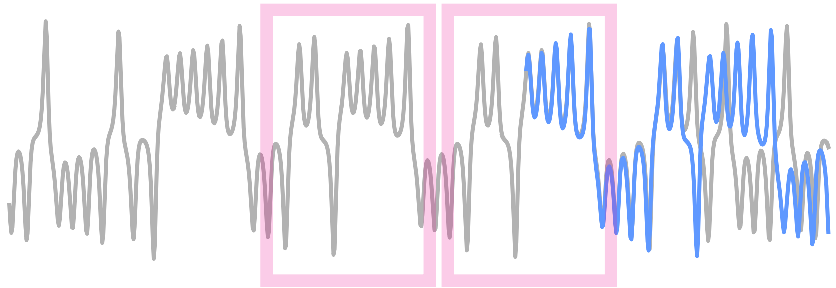


Zero-shot forecasting of chaotic dynamics

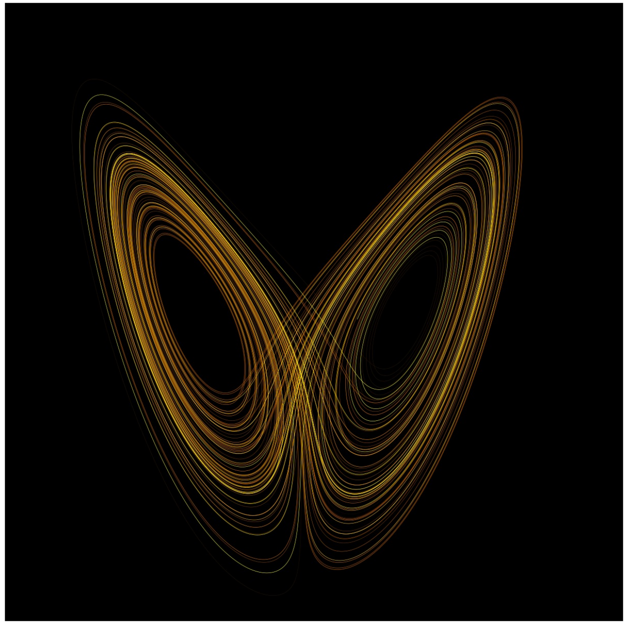
Foundation model as a tool for forecasting previously unseen dynamics



Foundation model as a "model organism" for learning from limited data



Learning dynamical systems from data: Equation discovery

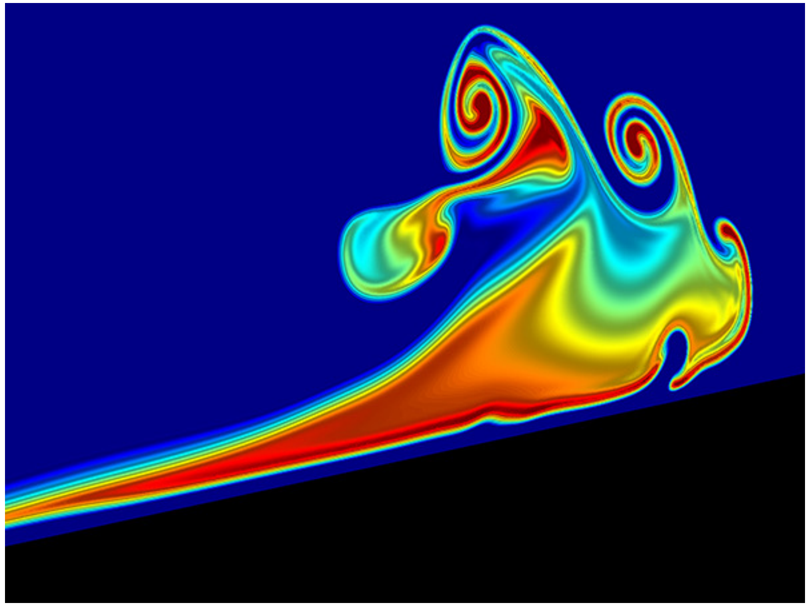


$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y - x)$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = x(\rho - z) - y$$
$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z$$

Sparse regression (e.g., SINDy)

Genetic programming / Symbolic regression

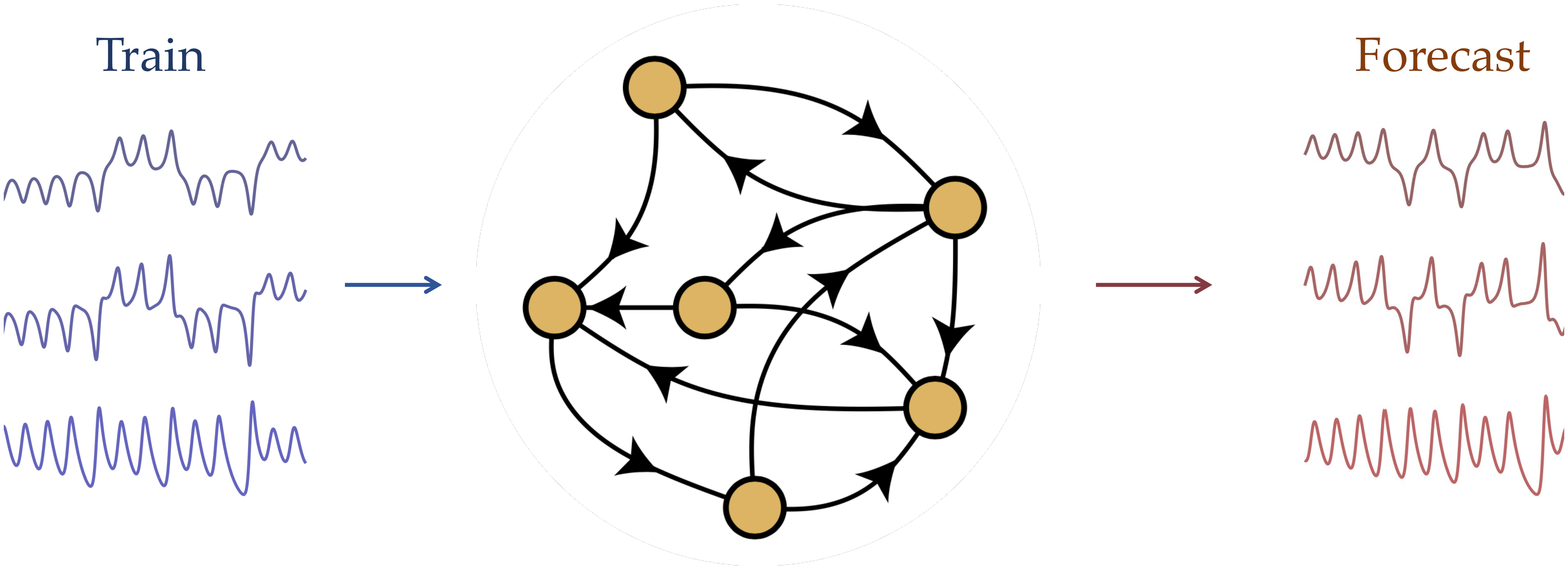
Koopman / DMD



$$\rho\left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}\right) = -p + \nabla\cdot\mathsf{T} + \mathbf{f}$$

…

Learning dynamical systems from data: Forecasting



Reservoir computing

Neural ODE

Recurrent neural nets

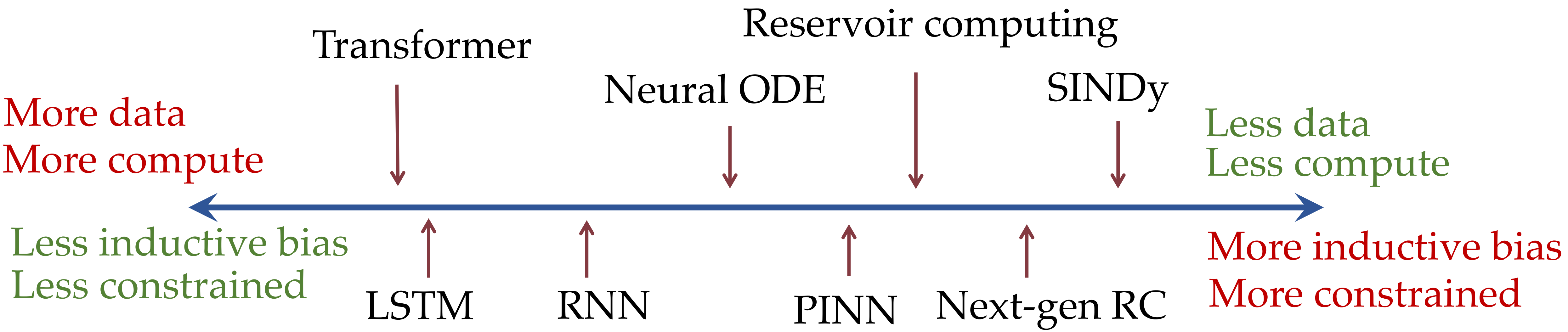
Neural operators

Transformers

Physics-informed neural nets

...

Learning dynamical systems from data: There is usually a tradeoff

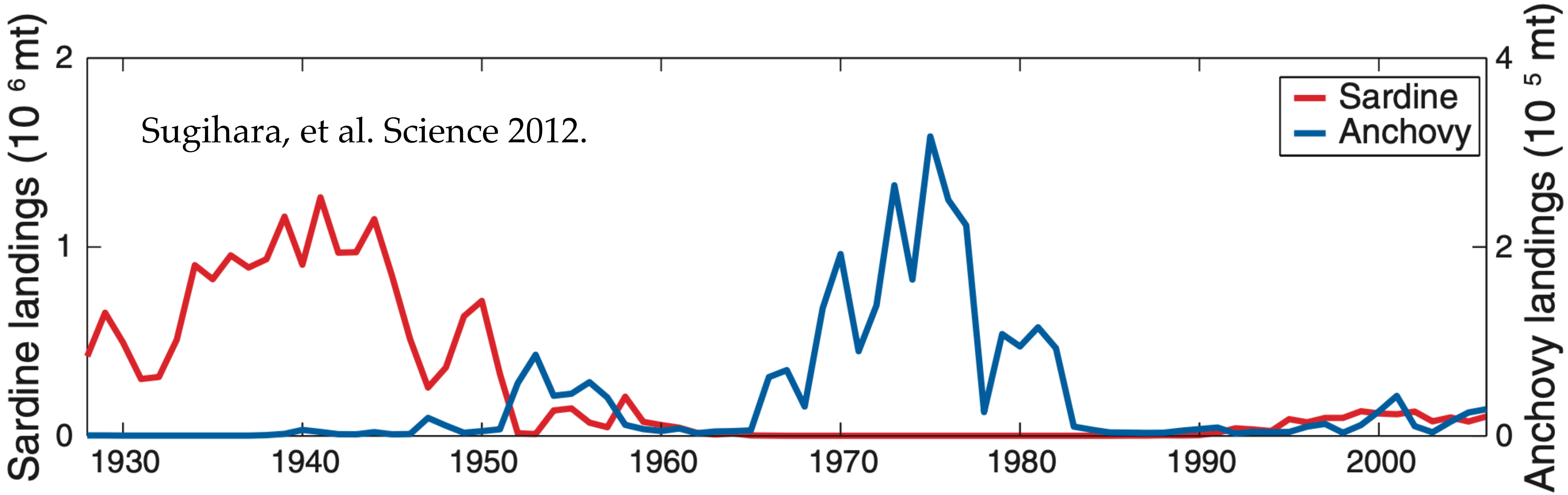


You either need a good **model**, or a lot of **data**

Zhang and Cornelius, PRR 2023.

What if we lack both model and data?

In many applications, we don't have a good **model**,
and high-quality **data** are not easy to come by



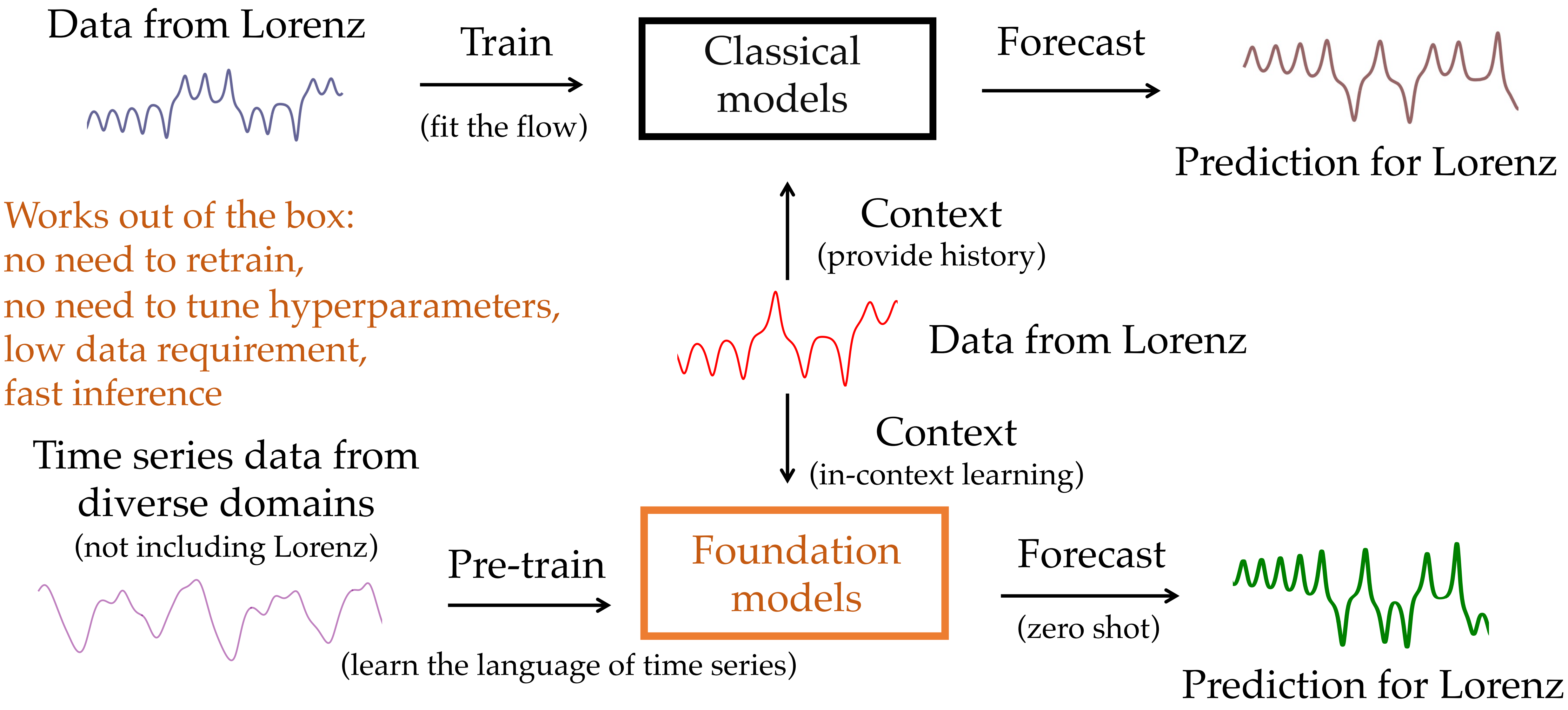
Can we forecast what happens next solely based on a short context time series?

This is a task that many living systems solve everyday (e.g., crossing the street)

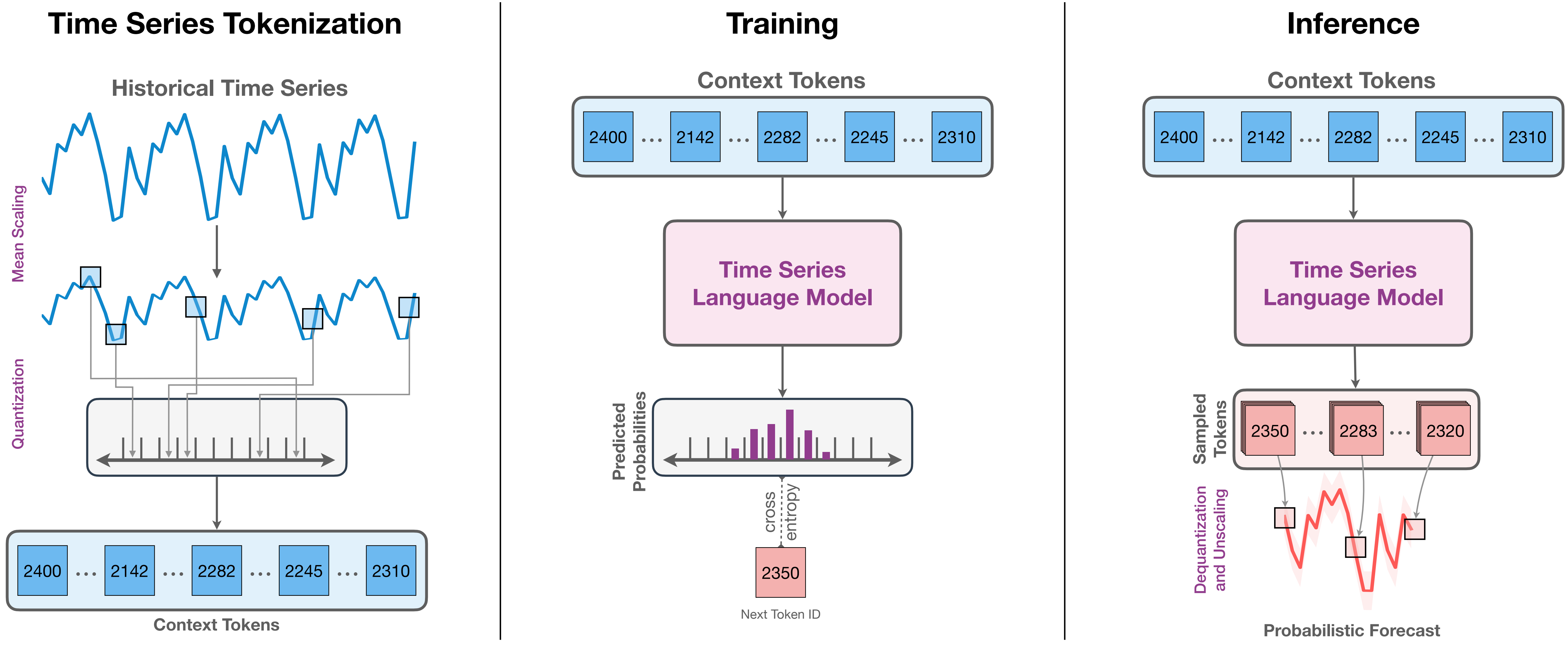
Can we use pre-trained transformers (**foundation models**) for this task?

What strategies do they use to make **zero-shot forecasts**?

Foundation models vs classical models



Chronos: ChatGPT for time series



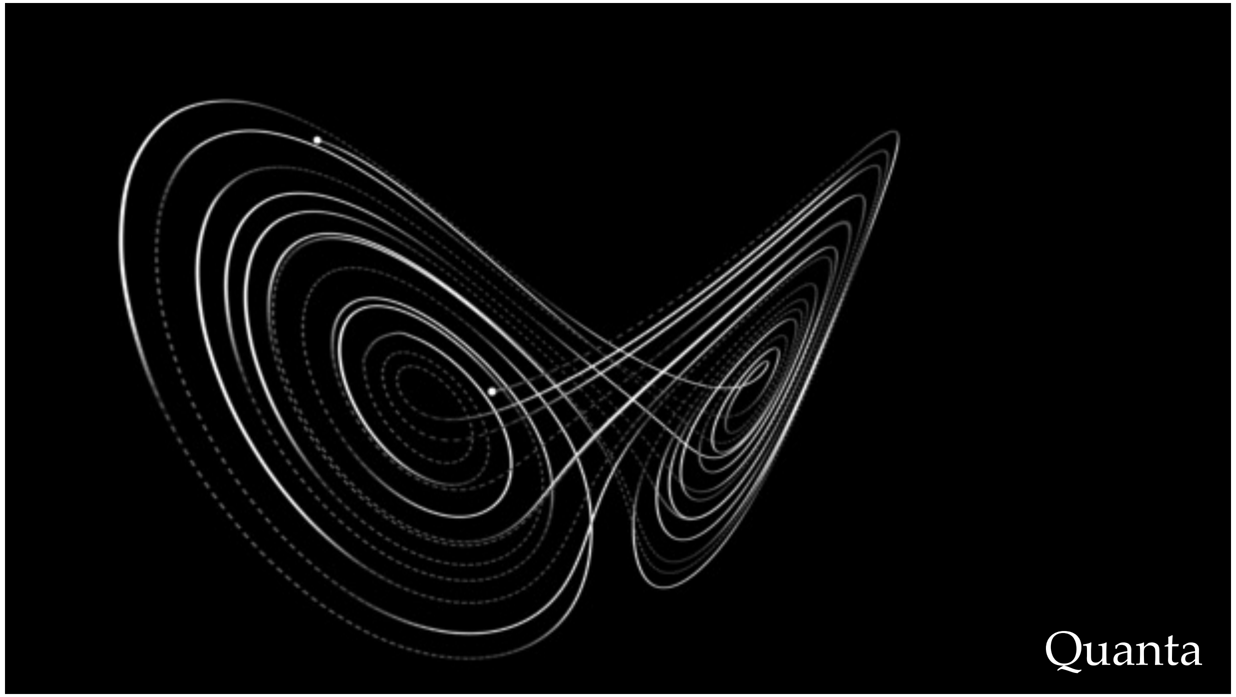
Trained on both real-world time-series data (weather, finance, traffic, etc.) and synthetic data

Works for both encoder-decoder and decoder-only models

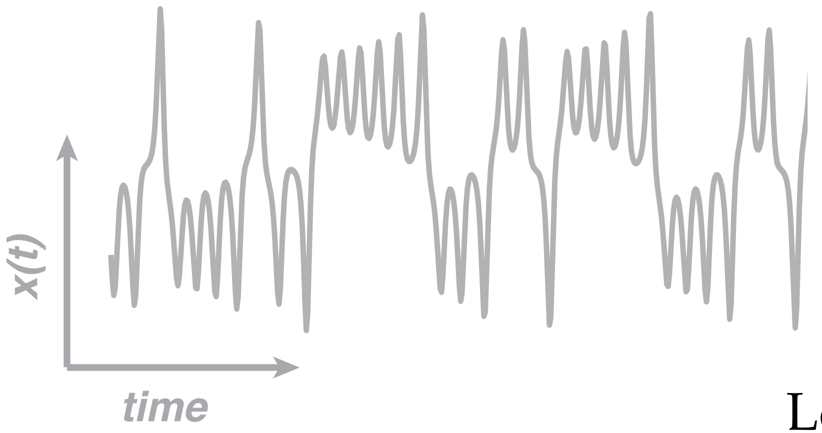
Ansari, et al. TMLR 2024.

Why applying foundation models to chaotic dynamical systems?

- Test generalization (Chronos wasn't designed to forecast chaotic systems)
- Not just short-term "weather," but also long-term "climate"
- Machine learning of dynamical systems still very much in the old paradigm of "training on the same system you want to predict"

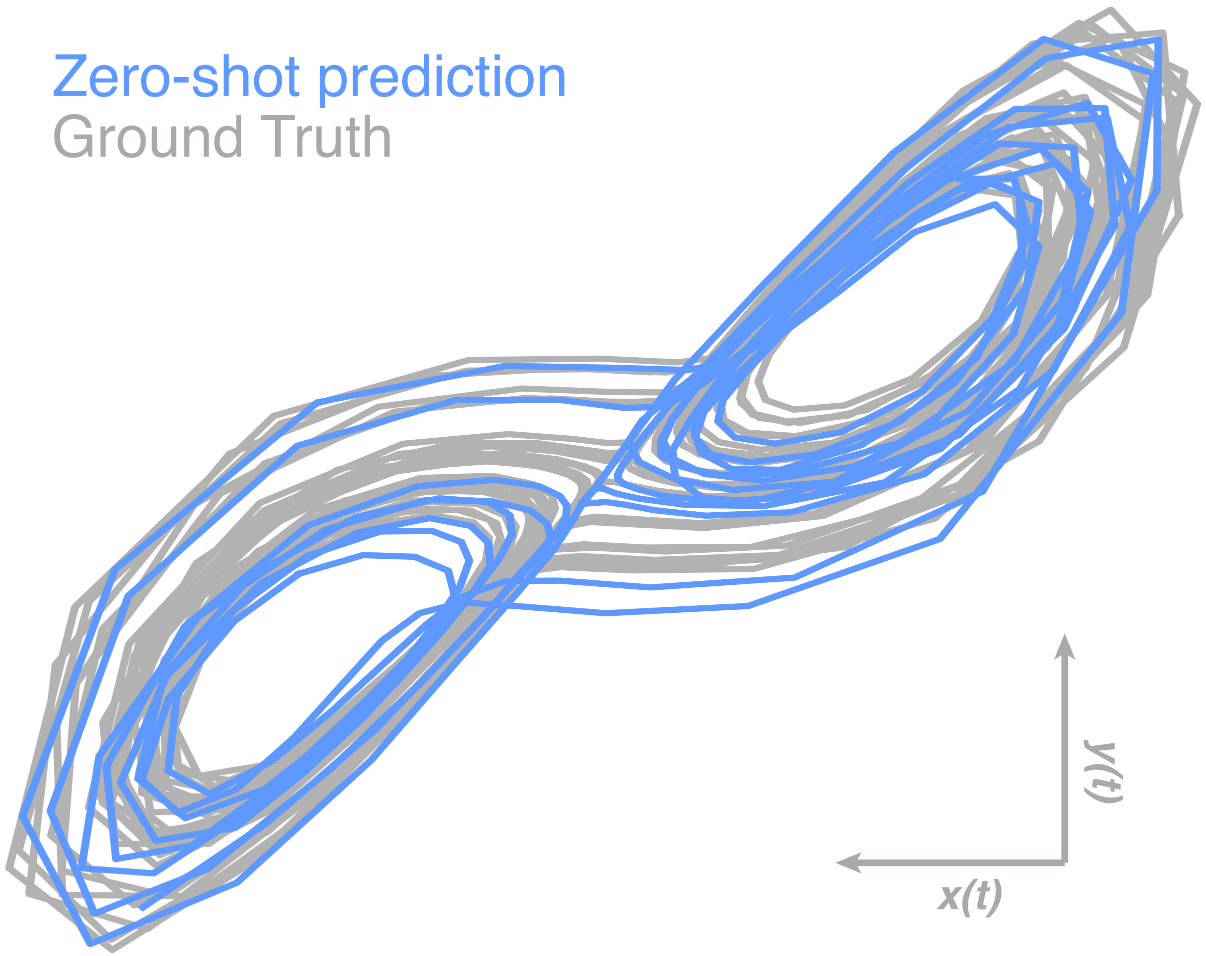


One example and two surprises



Lorenz system

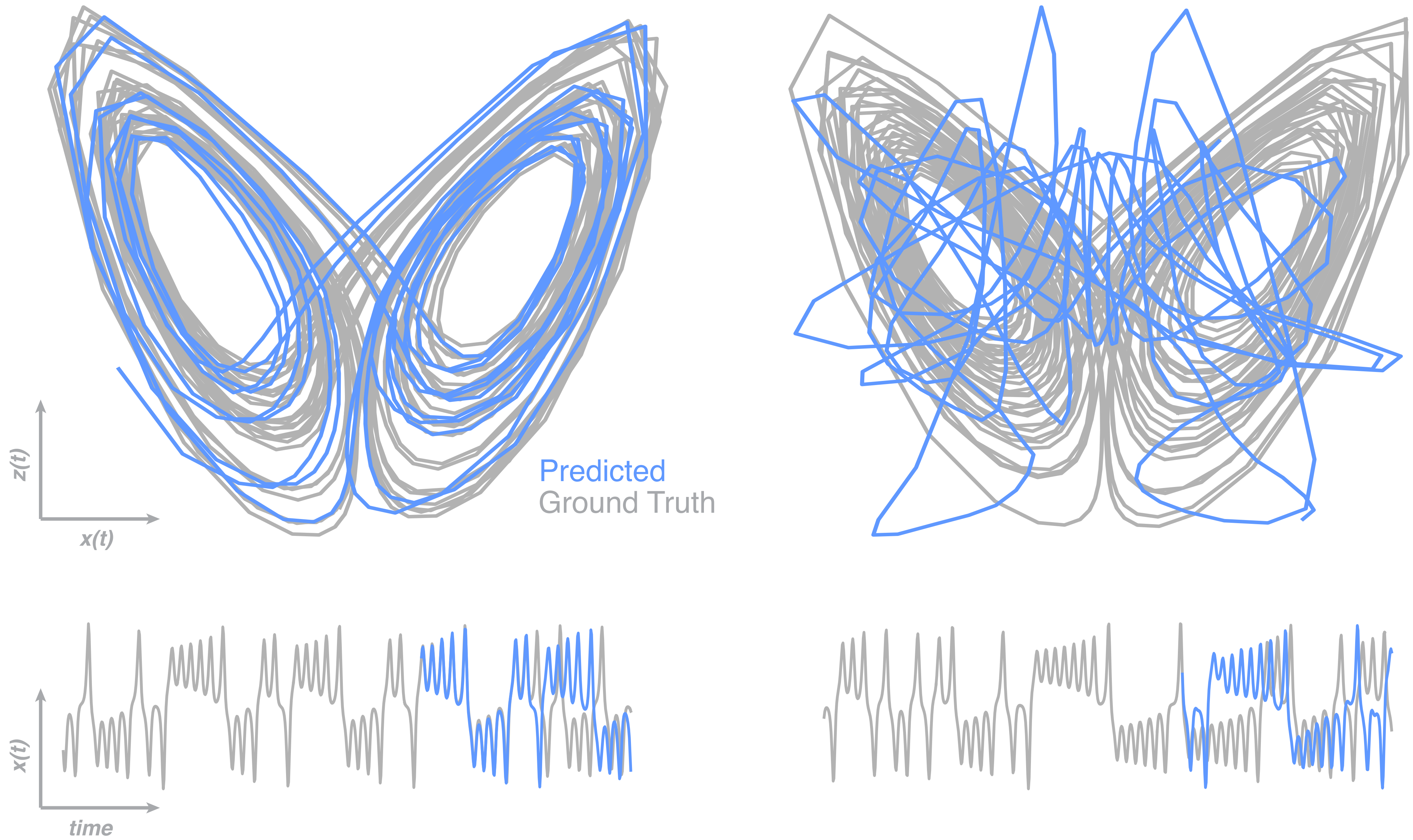
$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y - x)$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = x(\rho - z) - y$$
$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z$$



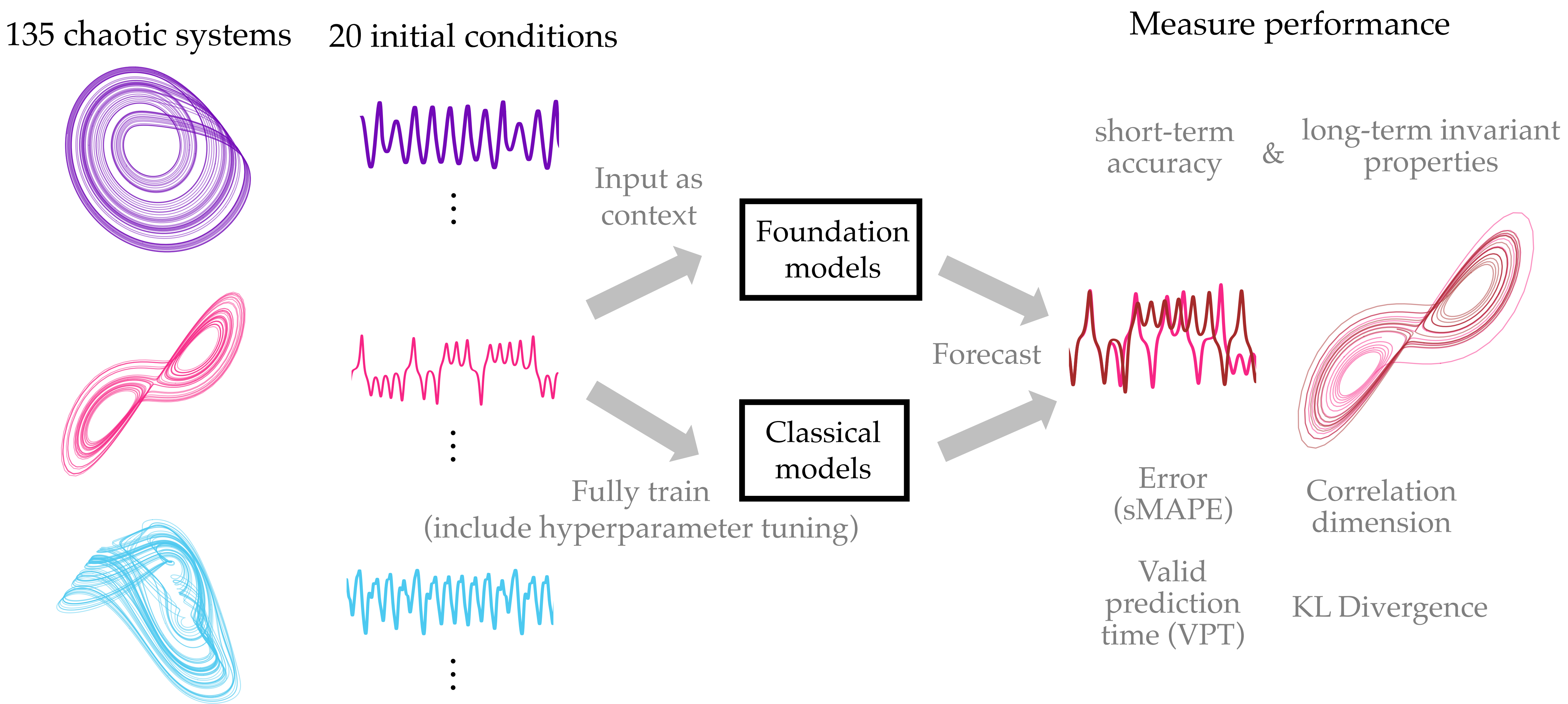
William Gilpin
UT Austin

Zhang and Gilpin, ICLR 2025

Chronos performance can be sensitive to initial conditions

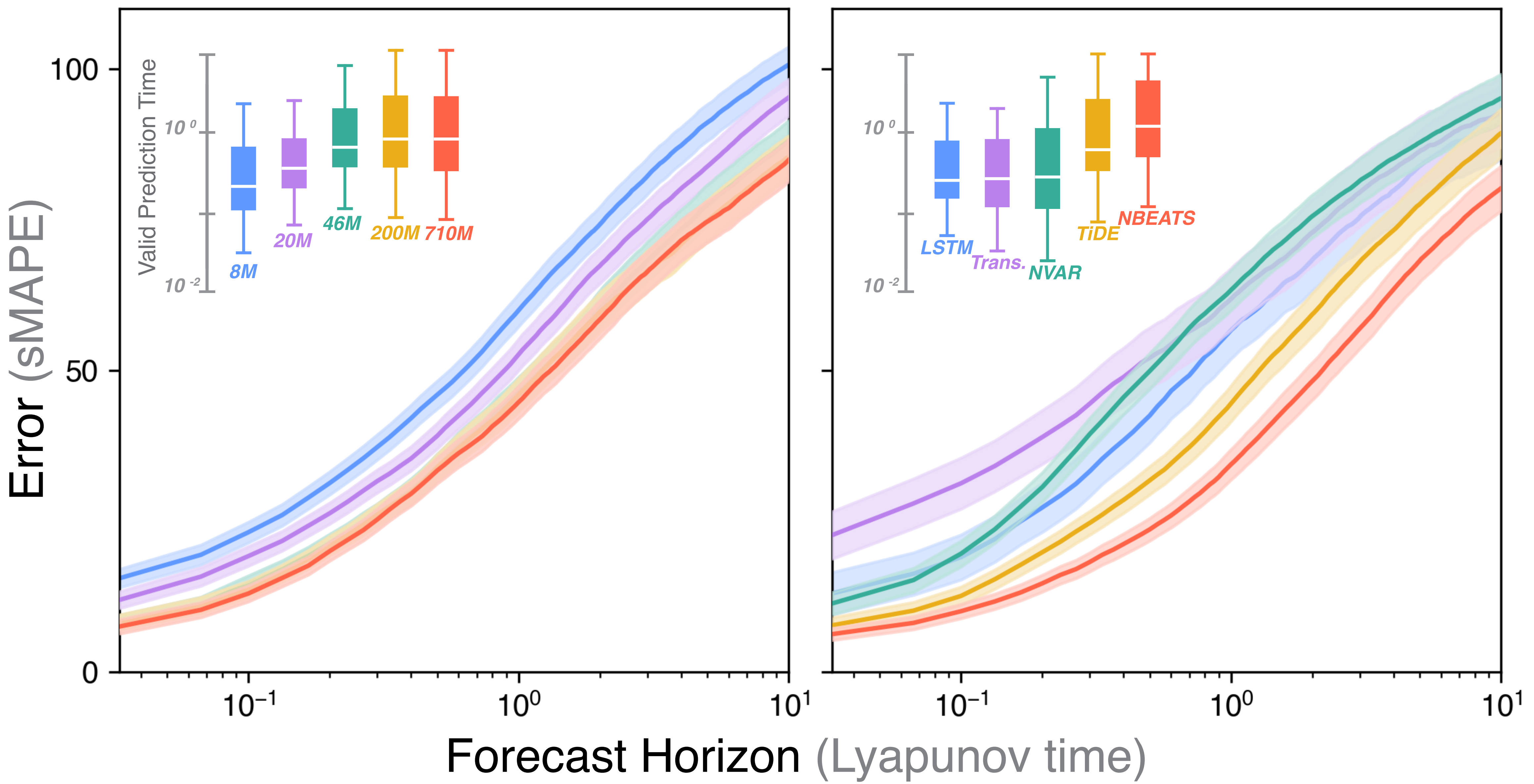


Chaos as a benchmark for zero-shot forecasting of time series

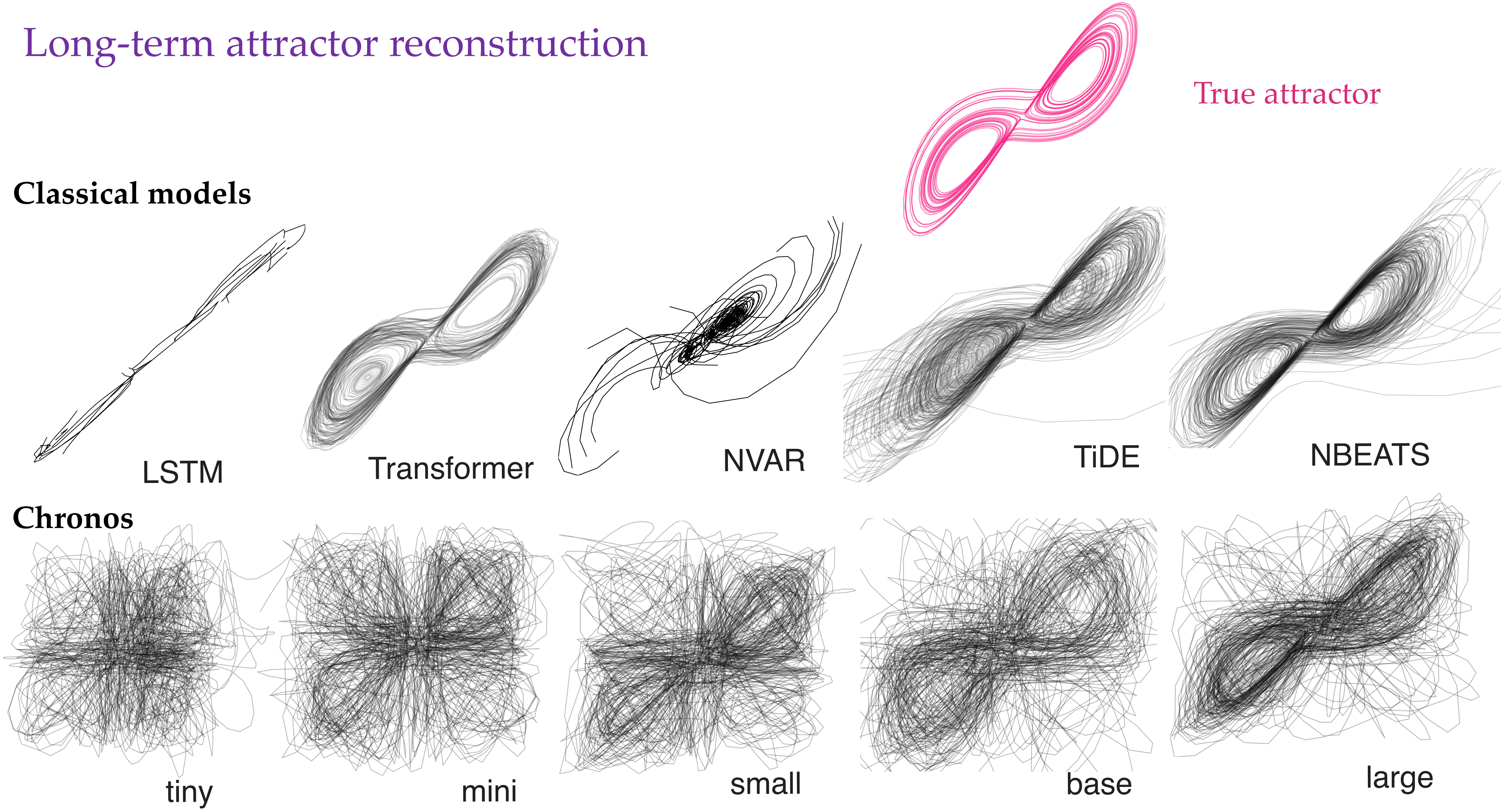


None of the chaotic trajectories are used to tune the weights of foundation models

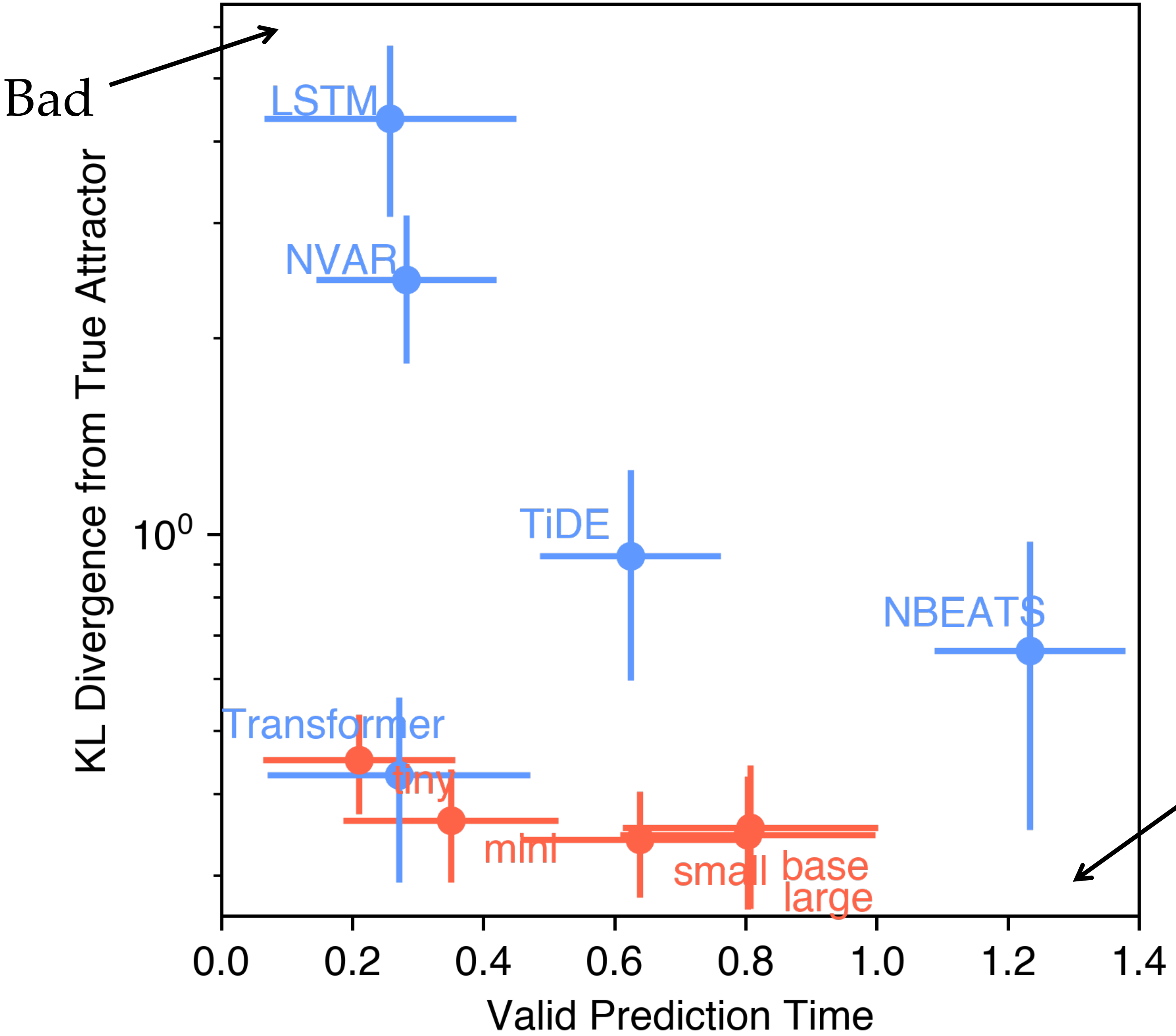
Short-term forecast accuracy



Long-term attractor reconstruction

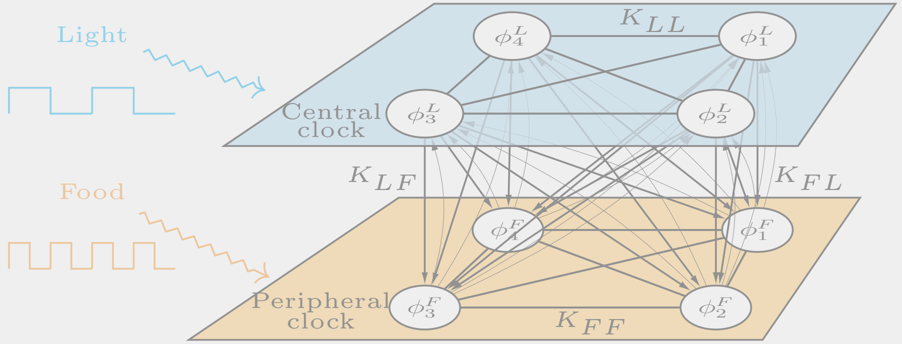


Foundation models effectively forecast previously unseen dynamics

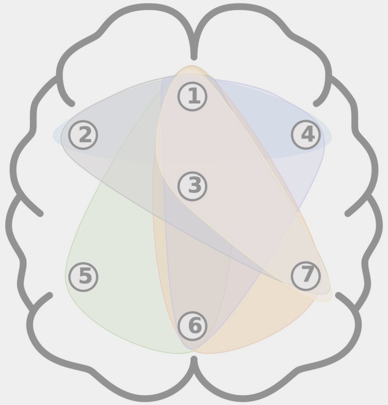


Brain as a dynamical system

Dynamics on neuronal networks



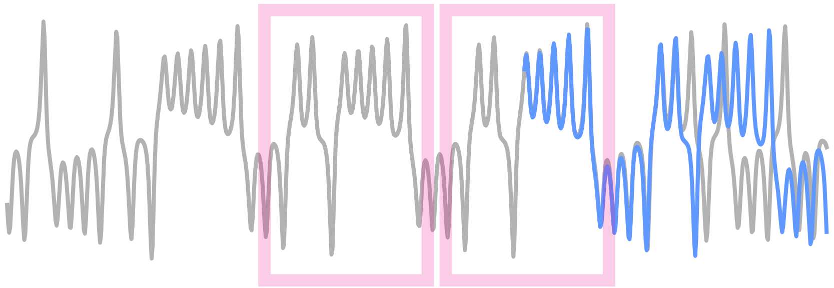
Networks from neuronal dynamics



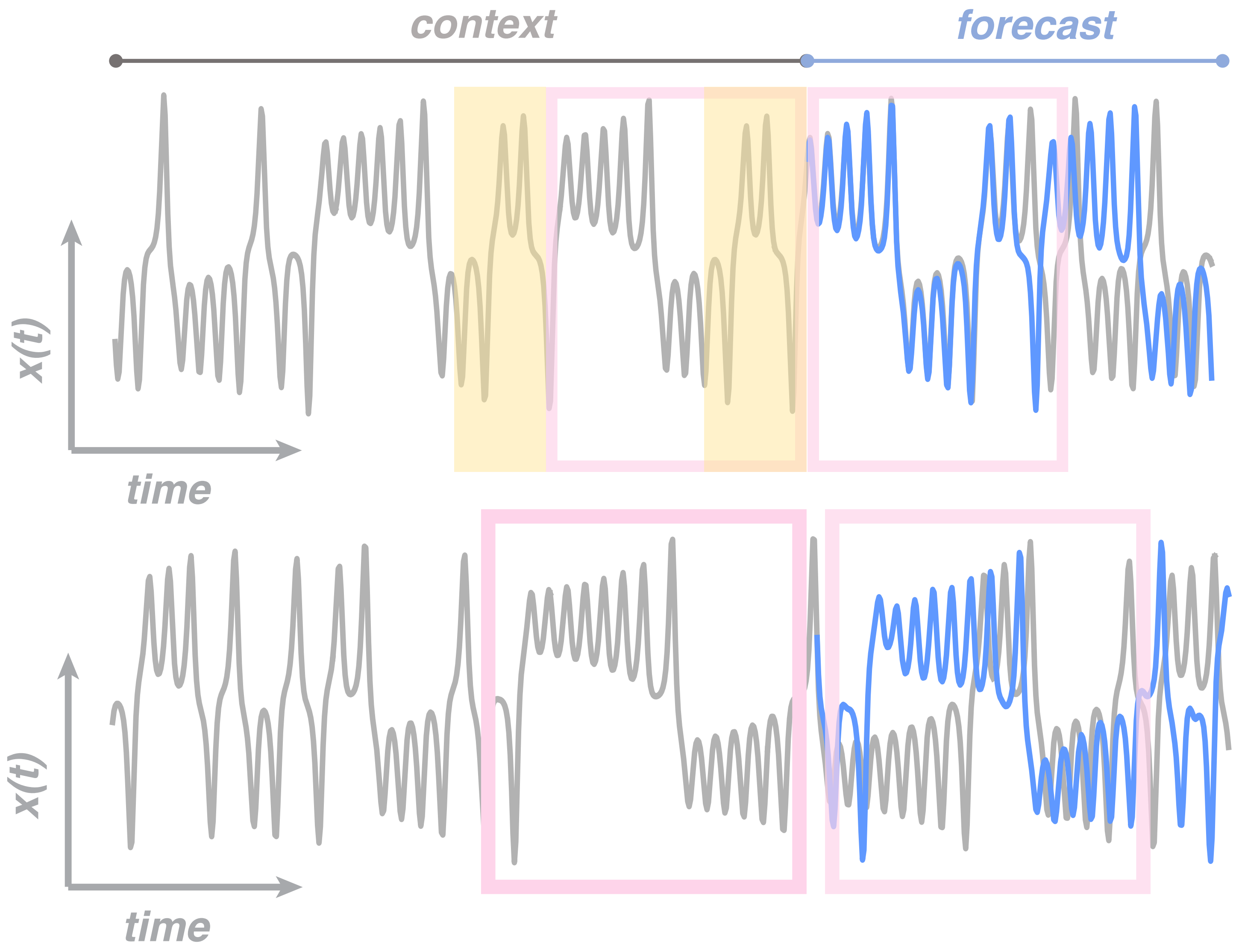
Zero-shot forecasting of chaotic dynamics

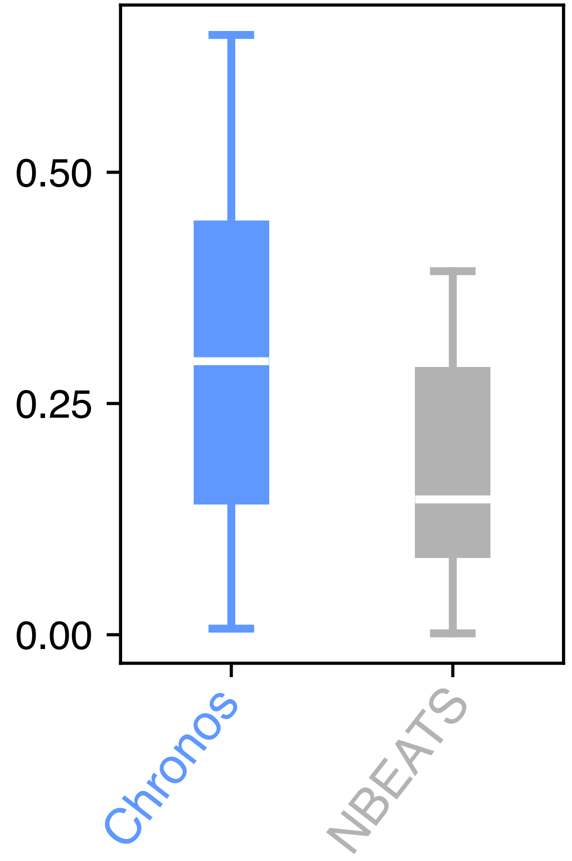
Foundation model as a tool for forecasting previously unseen dynamics

Foundation model as a "model organism" for learning from limited data



Foundation models use simple strategies for zero-shot forecasting





Chronos basically does **context parroting**!

Zhang and Gilpin, ICLR 2025

Chronos rediscovered a classical strategy from nonlinear forecasting on its own

Article | Published: 19 April 1990

Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series

George Sugihara & Robert M. May

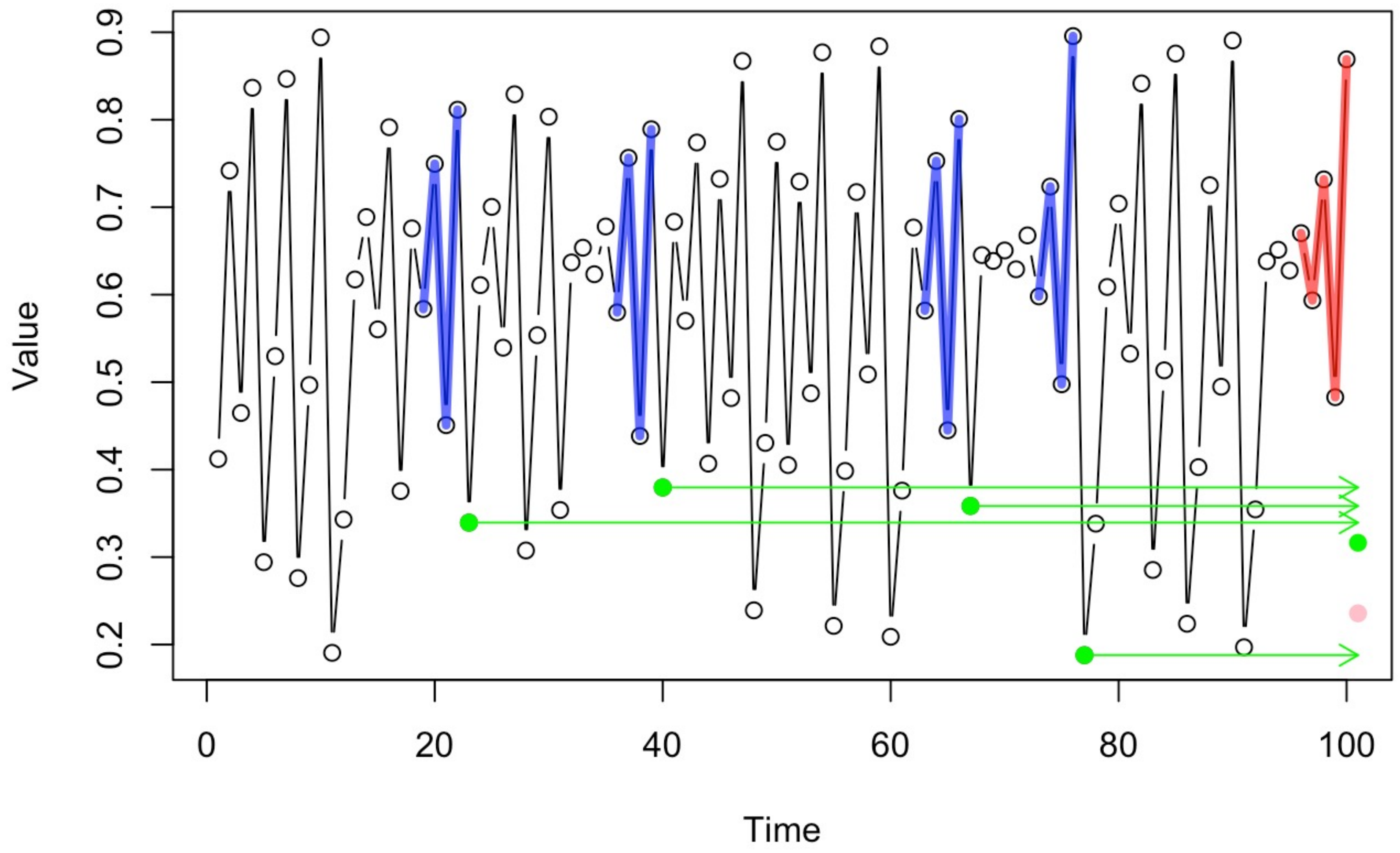
Nature **344**, 734–741 (1990) | Cite this article

Abstract

An approach is presented for making short-term predictions about the trajectories of chaotic dynamical systems. The method is applied to data on measles, chickenpox, and marine phytoplankton populations, to show how apparent noise associated with deterministic chaos can be distinguished from sampling error and other sources of externally induced environmental noise.

Simplex projection vs context parroting

Owen Petchey, 10.5281/zenodo.57081



Chronos basically rediscovered the **simplex projection** idea in *Sugihara & May, Nature (1990)*, but with a higher embedding dimension and no averaging

How did Chronos discover context parroting?

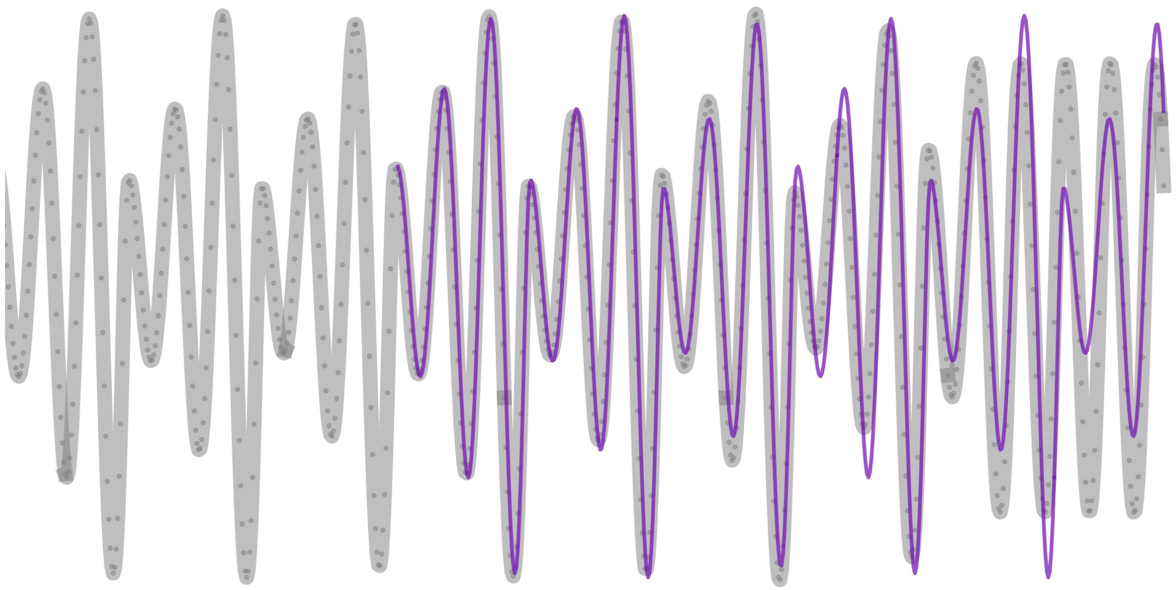
[A] [B] ... [A] → [B]

Context parroting could come from **induction heads**, which underlies a lot of in-context learning in simple transformers

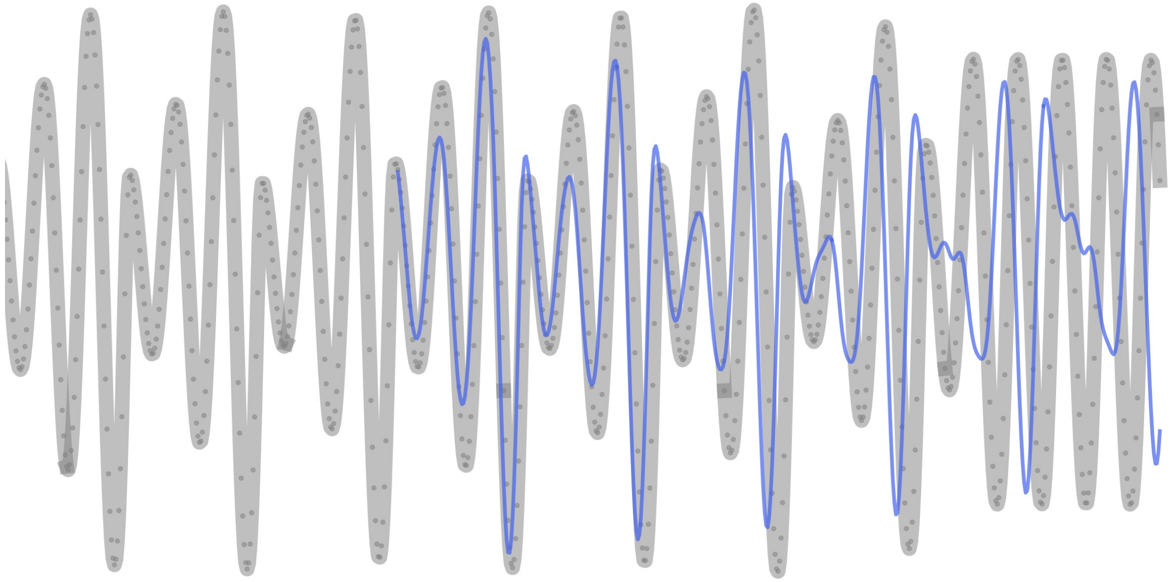
Olsson, et al., Transformer Circuits Thread, 2022.
Reddy, ICLR 2024

Context parroting as a mechanism for zero-shot forecasting

Parroting



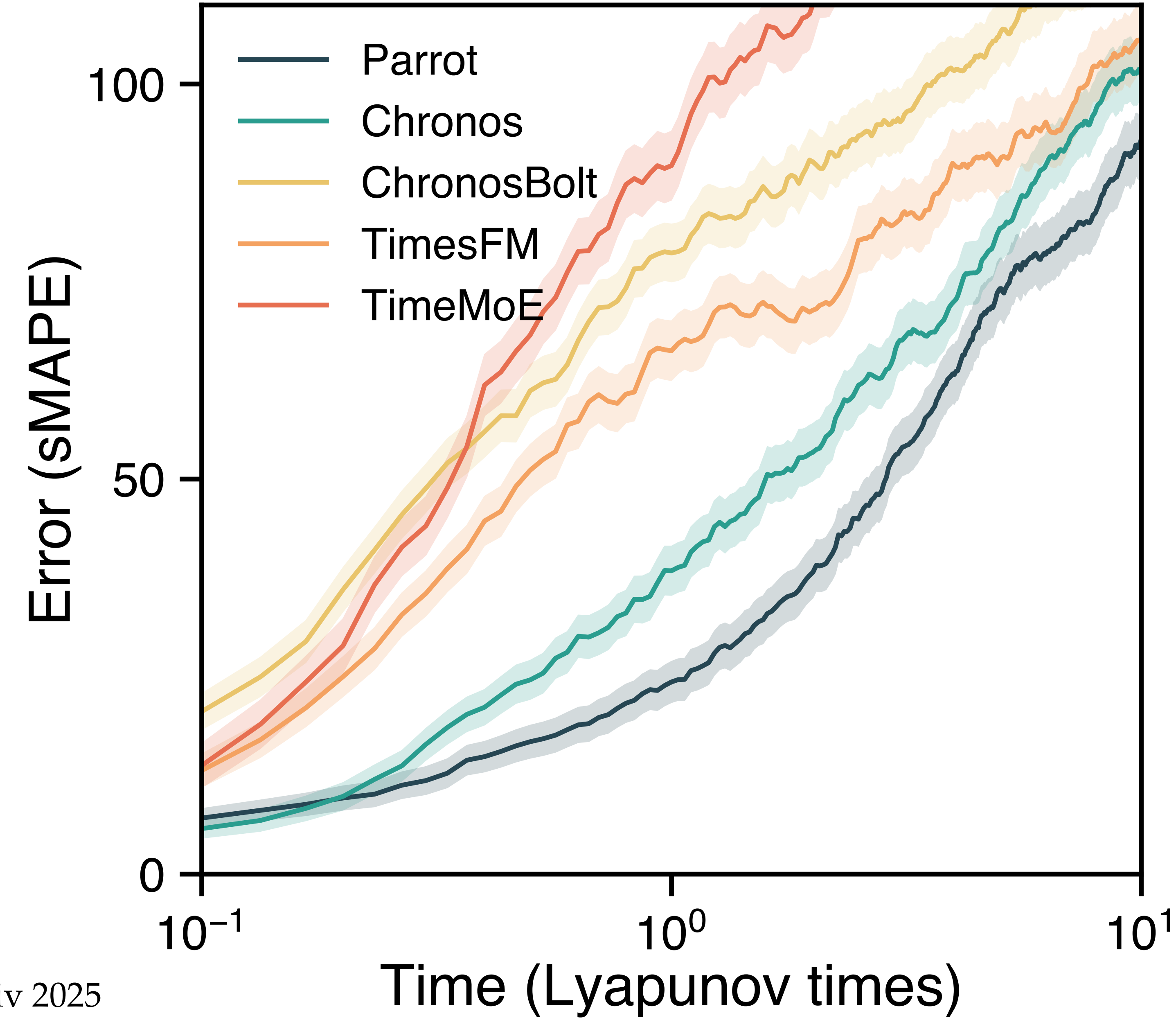
Chronos



It’s not just zero-shot forecasting, context parroting has **zero parameter** and requires **zero training**!

Can it outperform Chronos?

Context parroting vs foundation models

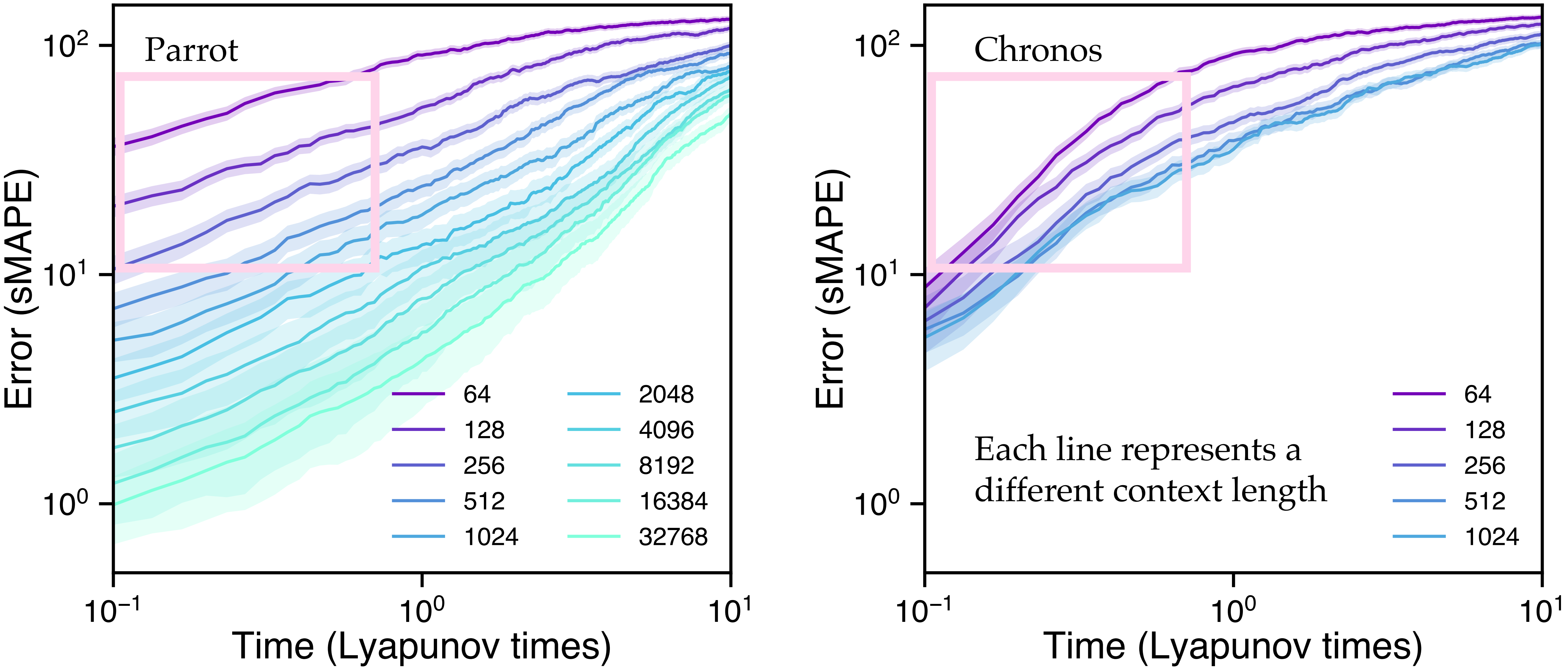


Zhang and Gilpin, arXiv 2025

Context lengths matter

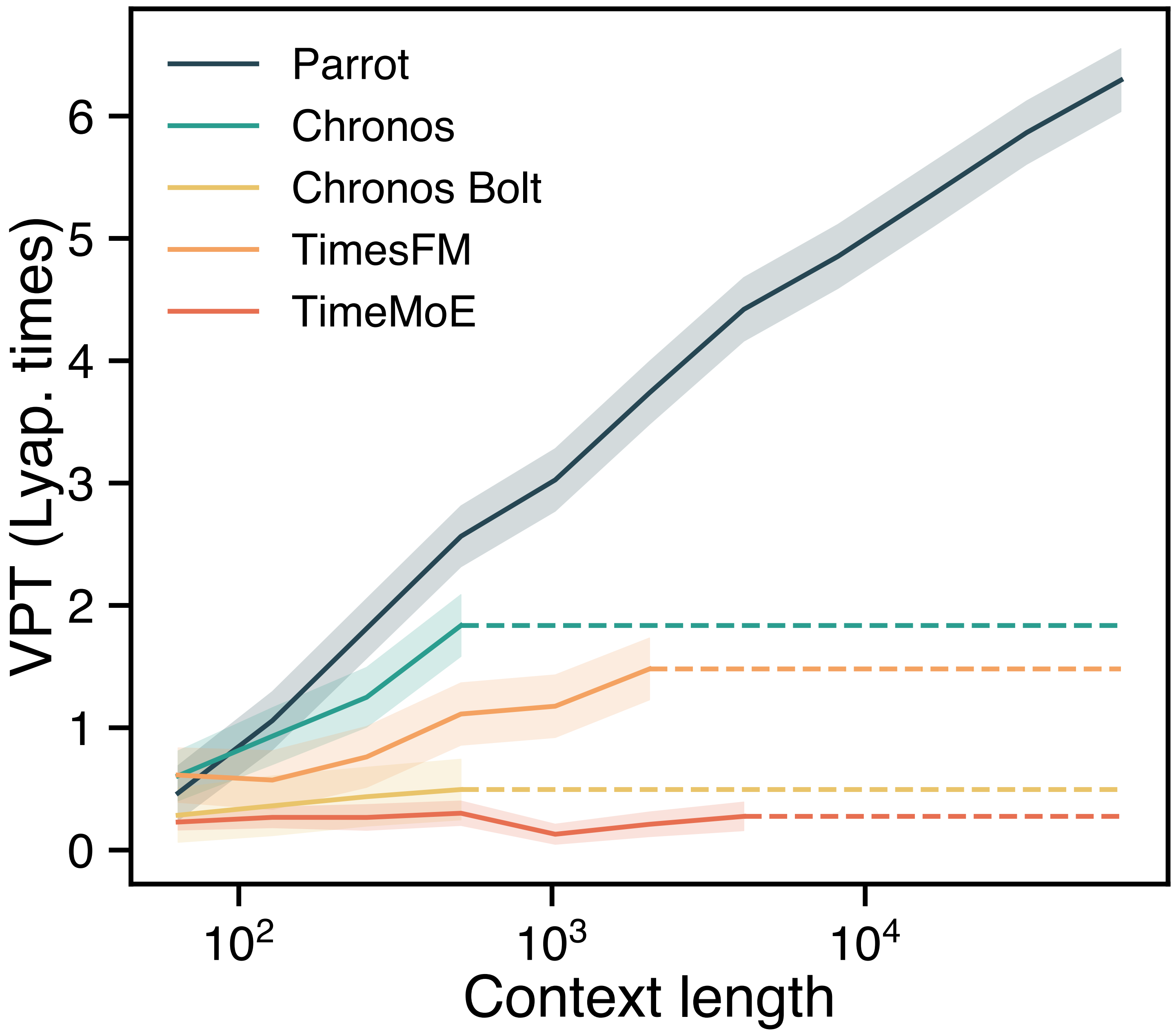
Context parroting can better utilize longer context data

Chronos do better than parroting for short contexts. How?

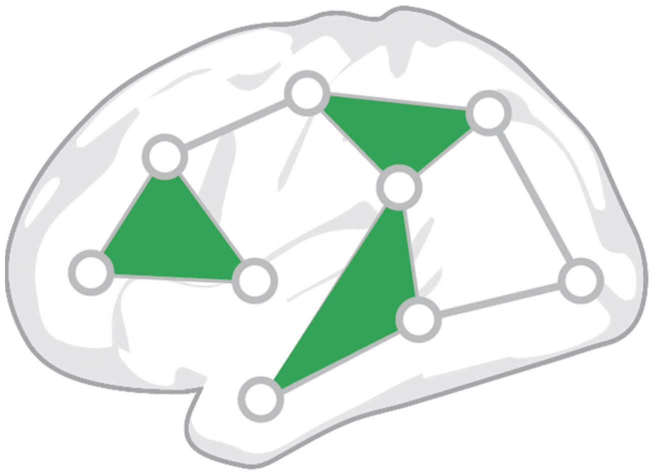


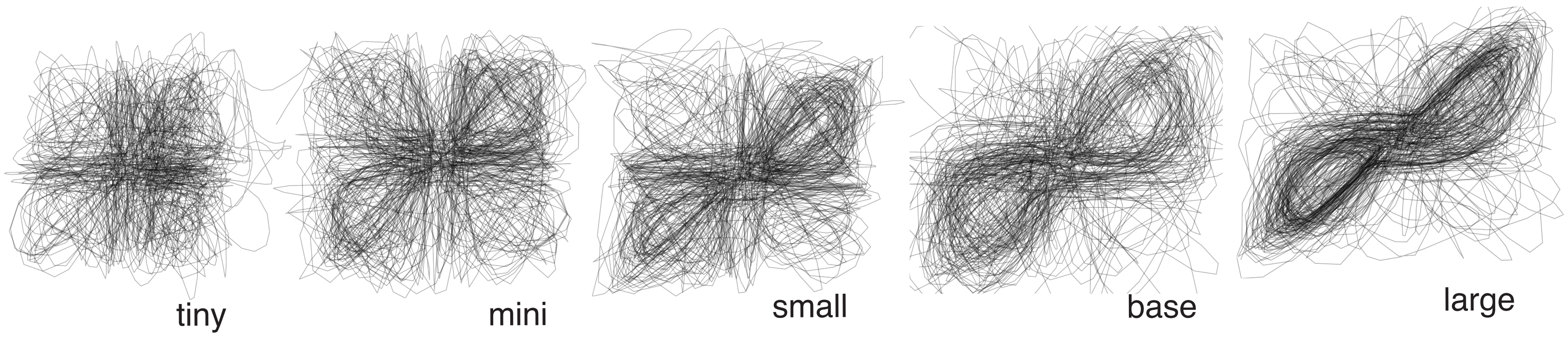
Context parroting vs foundation models

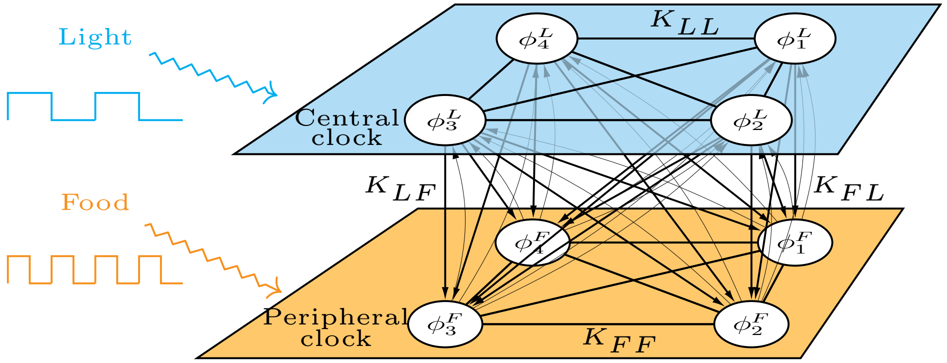
- Now we have come full circle…
- Inference cost of context parroting is negligible compared to foundation models
- Foundation models do have tricks beyond context parroting and can deal with nonstationary time series

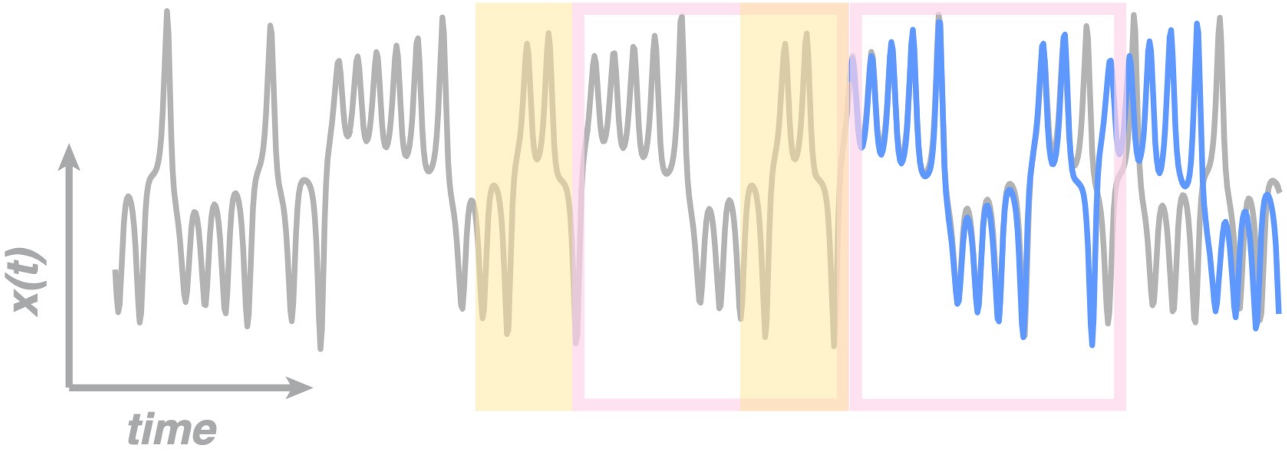


Thanks!









Funding