Inferring effective state variables and dynamics from data

Grenoble AI for Physics workshop
May 29, 2024

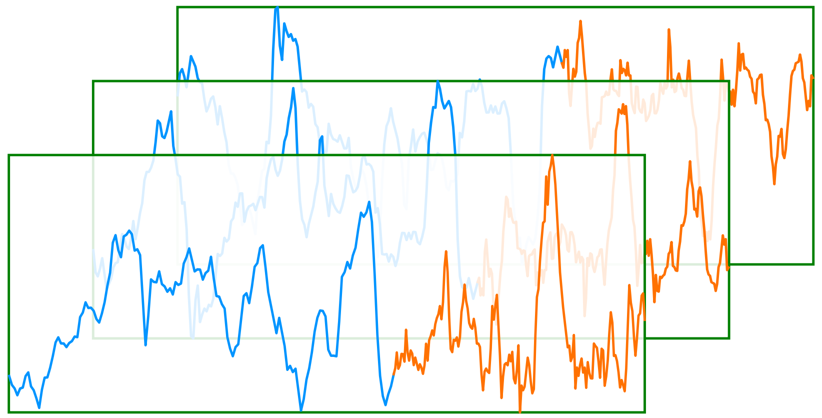
Nicolas Brodu

Inria

Information on the project https://team.inria.fr/comcausa/

Context / Goal: Modeling from observations

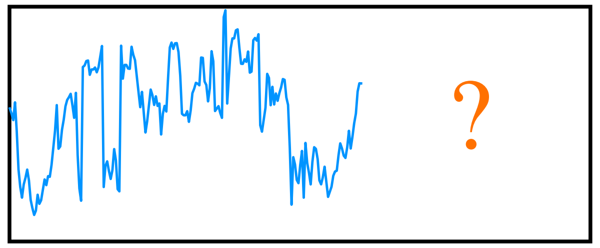
Observations → **Model** → **Predictions**



Ensemble of
past → future

– From data

– Principled construction



Parameters

– Interpretable (= characteristic scales)

– Few meta-parameters

Use case = complex system

– Example : a forest

– No access to all micro. parameters

– No known global states / equations

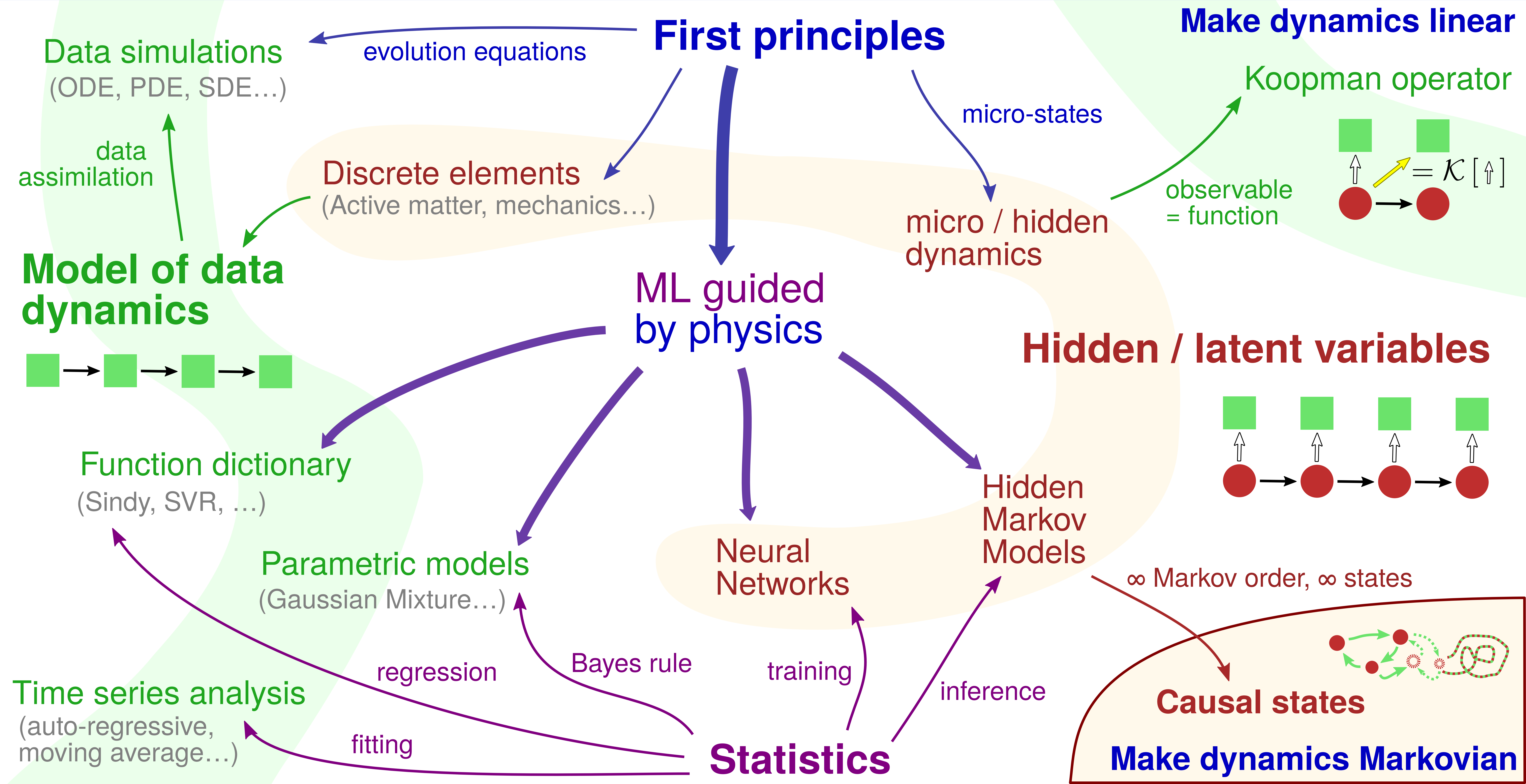
– Macro. data can be measured

Inferred

– Effective state variables

– Dynamics at given scales

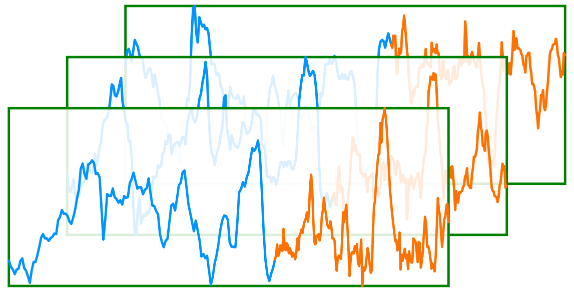
(very incomplete) Modeling panorama



Notion of causal states

Known generative model → analysis

Ensemble of observations → inference



past → future

possible future sequences

same distribution of futures

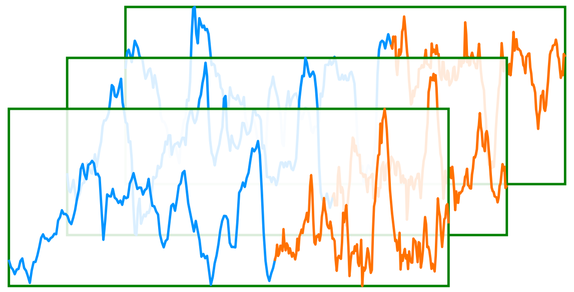
Causal state = {all pasts: same P(future | past)}

Same causes, same consequences

Intrinsic properties

Known generative model → analysis

Ensemble of observations → inference



past → future

Causal state = {all pasts: same P(future | past)}

– States do not depend on the frame of reference (distribution shapes change but not the equiv. classes)

⇒ **intrinsic property of the physical process**

– No new observation can distinguish two past sequences in the same causal state

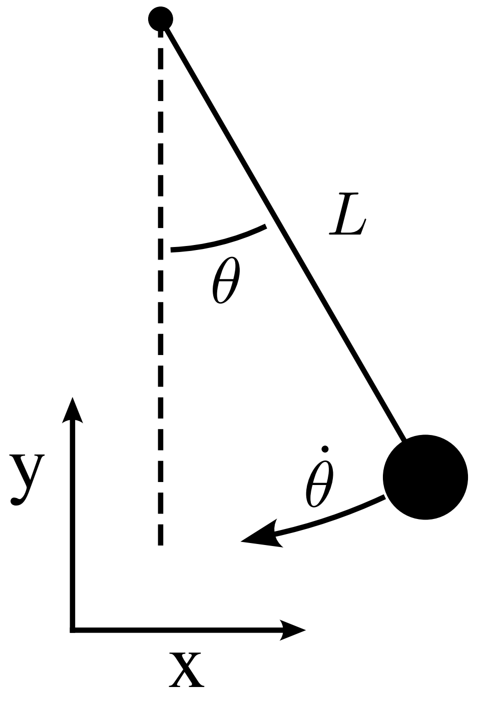
⇒ **finest building blocks for modeling** (computational mechanics [Crutchfield, 1988])

– No dependency left on histories

⇒ **Markovian dynamics** (any process ⇒ Markov process [Knight, 1975])

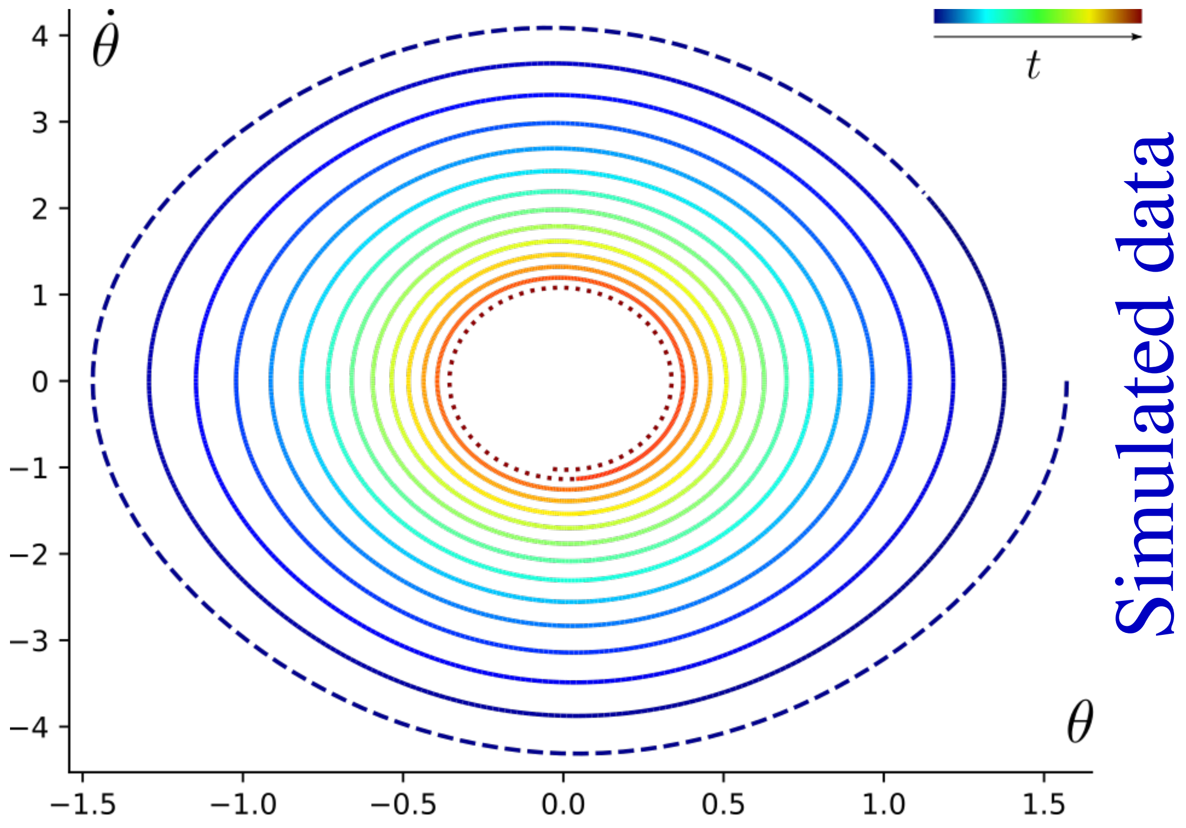
Effective state variables, recovering laws of motion

Dissipative pendulum



$$\frac{d\theta}{dt} = \dot{\theta}$$

$$\frac{d\dot{\theta}}{dt} = -\frac{b}{m}\dot{\theta} - \frac{g}{L}\sin\theta$$



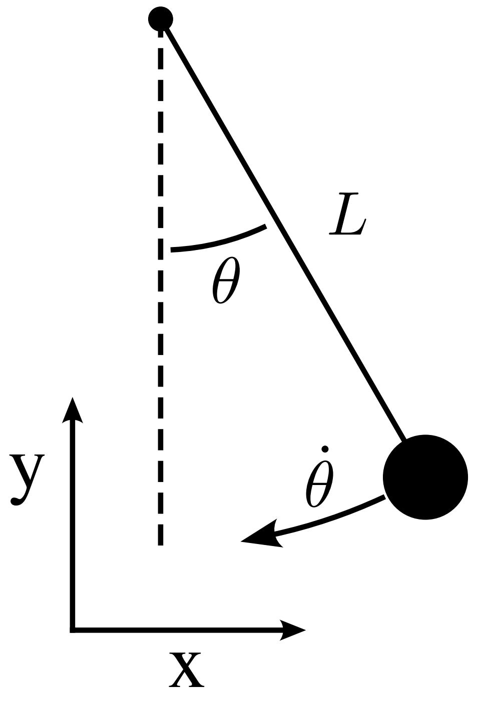
State space: θ, $\dot{\theta}$

Causal states = points in phase space

- Unique trajectory = future
- No dependency on the past

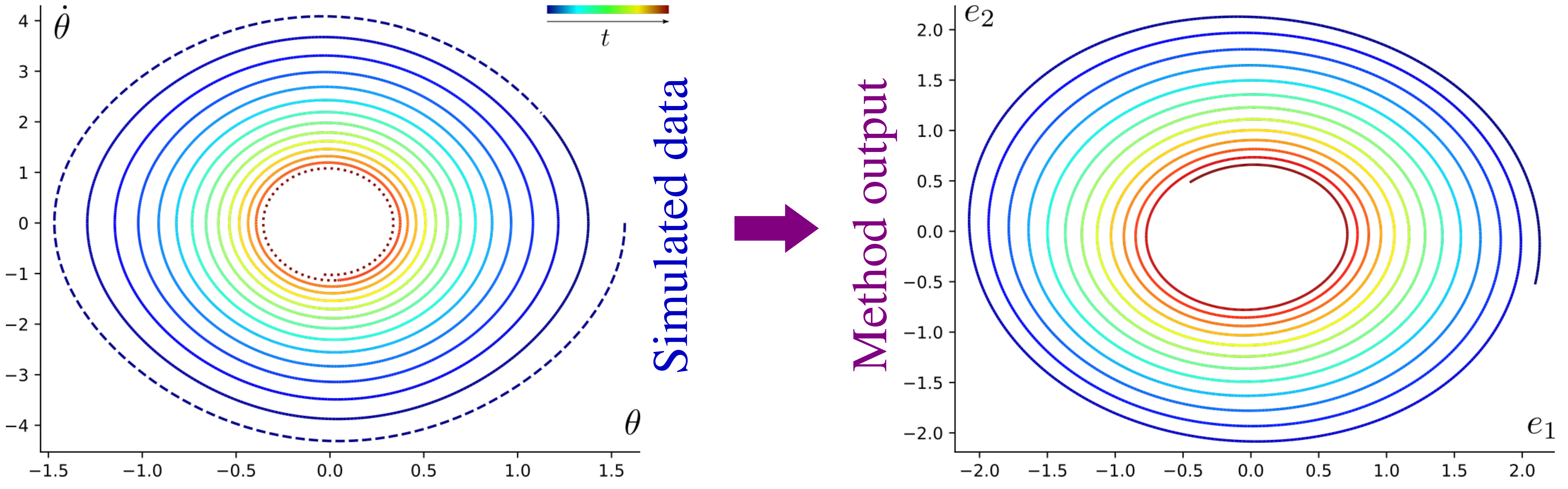
Effective state variables, recovering laws of motion

Dissipative pendulum



$$\frac{d\theta}{dt} = \dot{\theta}$$

$$\frac{d\dot{\theta}}{dt} = -\frac{b}{m}\dot{\theta} - \frac{g}{L}\sin\theta$$



State space: θ, $\dot{\theta}$

Effective state variables

Causal states = points in phase space

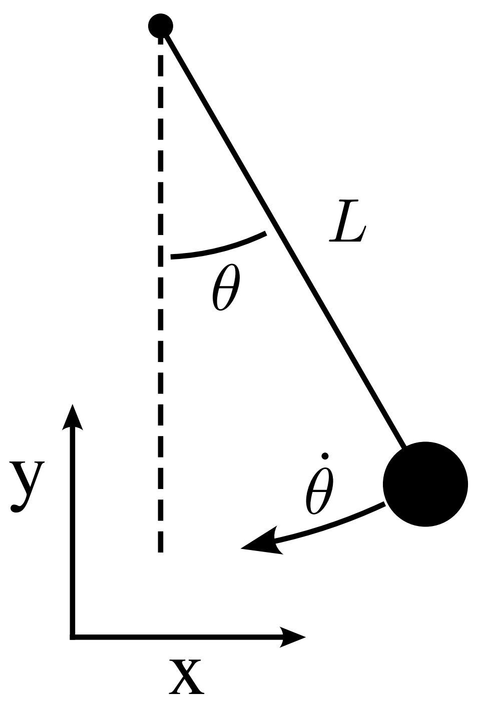
- Unique trajectory = future
- No dependency on the past

Causal states are independent from referential

- Build one from data ⇒ other state variables
- Equivalently valid law of motion

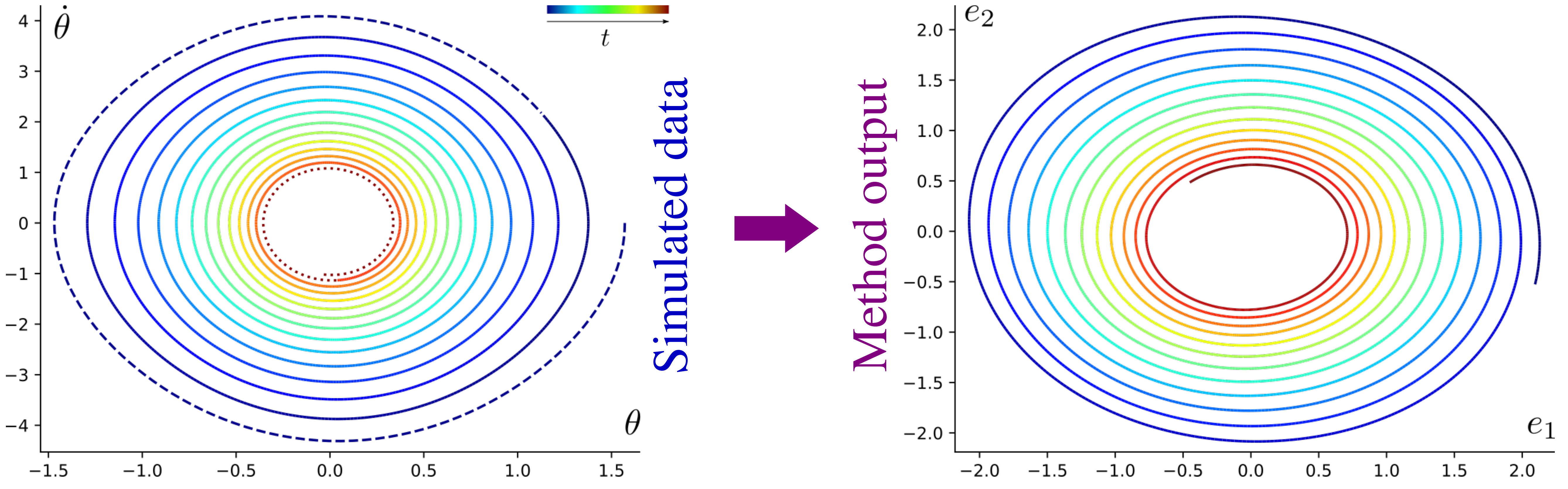
Effective state variables, recovering laws of motion

Dissipative pendulum



$$\frac{d\theta}{dt} = \dot{\theta}$$

$$\frac{d\dot{\theta}}{dt} = -\frac{b}{m}\dot{\theta} - \frac{g}{L}\sin\theta$$



State space: θ, $\dot{\theta}$

Effective state variables

Causal states = points in phase space

- Unique trajectory = future
- No dependency on the past

Causal states are independent from referential

- Build one from data ⇒ other state variables
- Equivalently valid law of motion

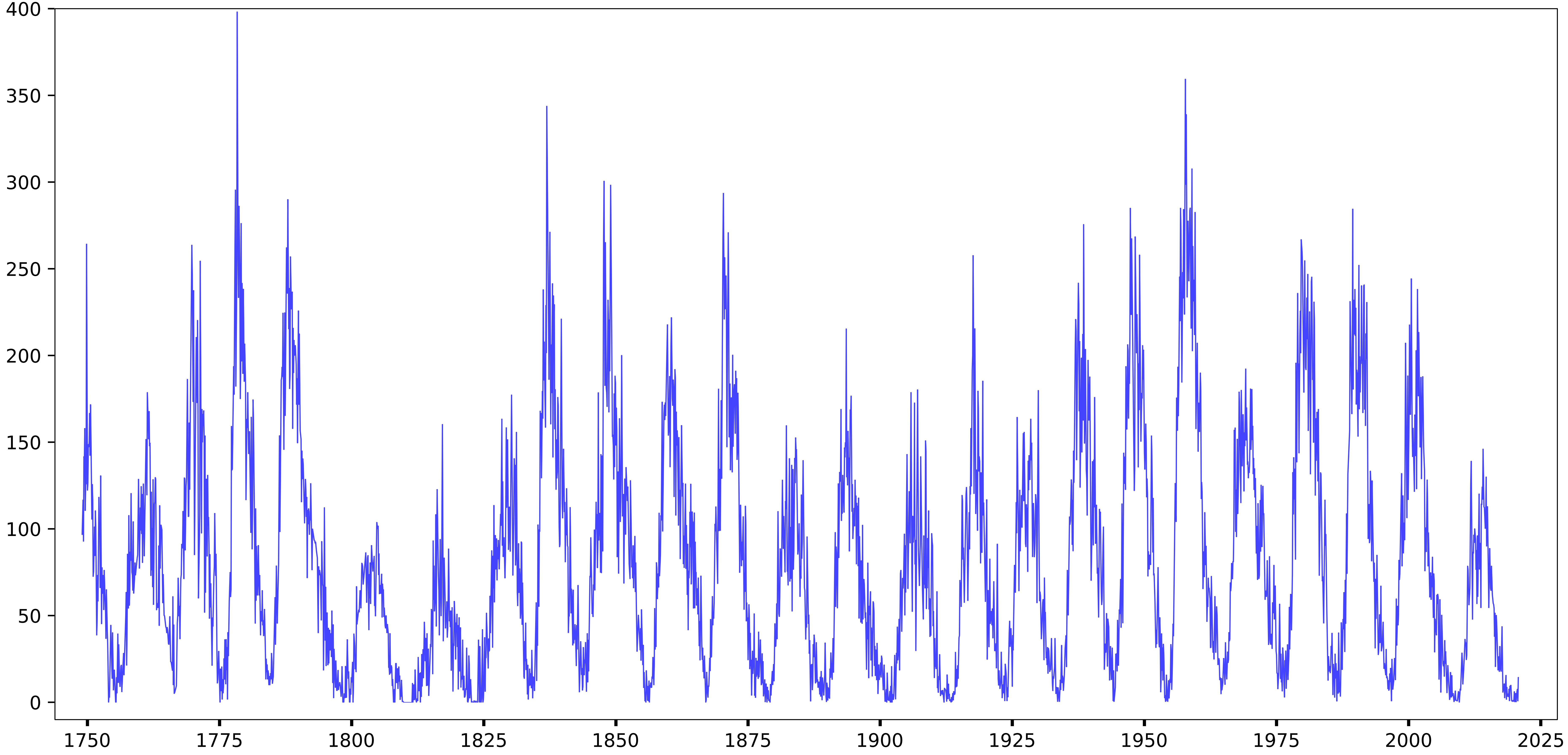
Reconstruction **from measured data**

- Model interpretation = **effective state variables + evolution** at that data scale
- Capturing effective physical law ⇒ **may generalize better** out of observed samples

Case study on real data: monthly sunspots observations

Nomber / month

Data : SILSO

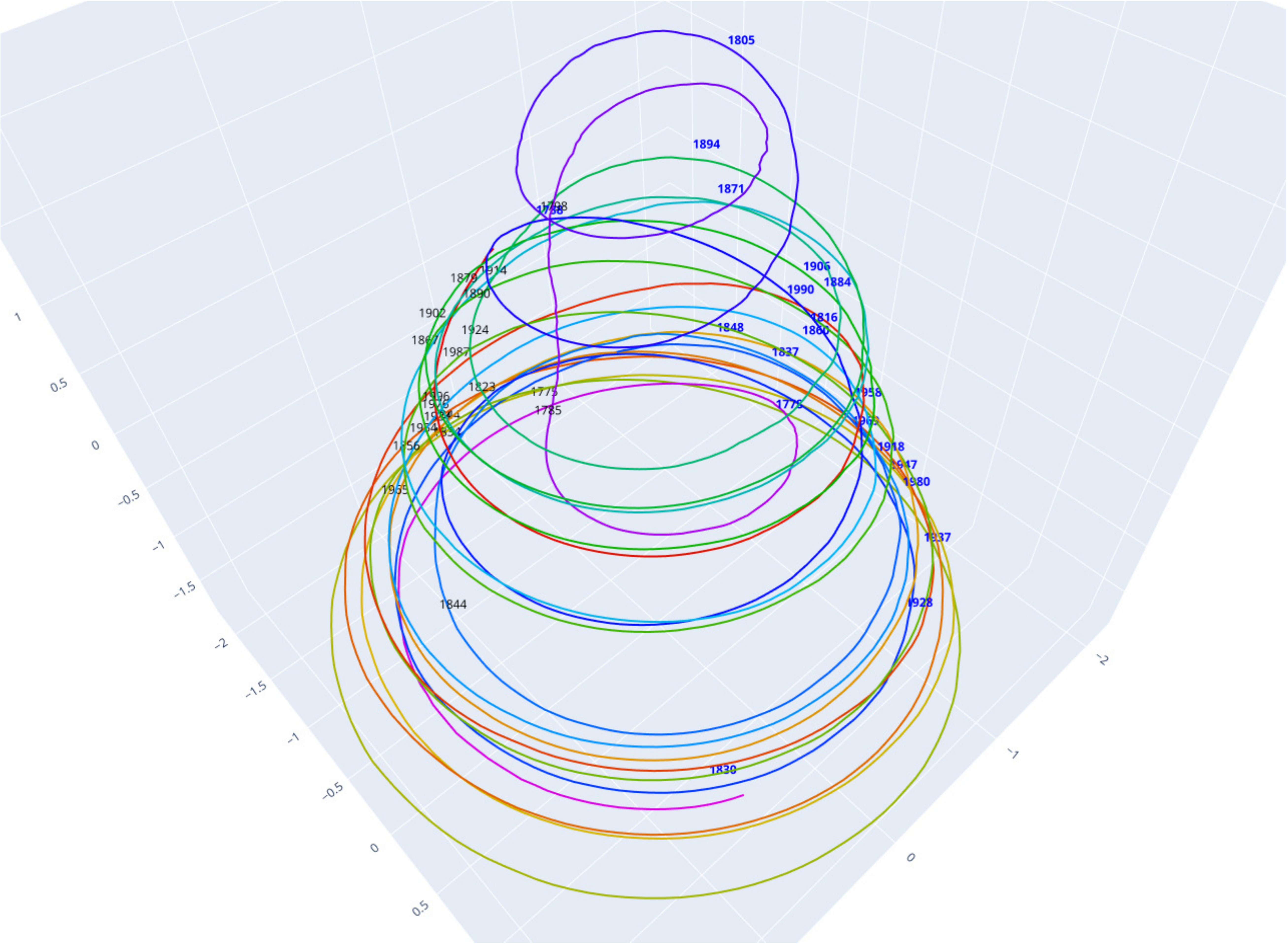


Sunspots – Structure in effective state space

Inferred state variables

– 11-years cycle (x and y)
– Amplitude modulations (z)

Trajectories on a structure *resembling* an attractor embedding



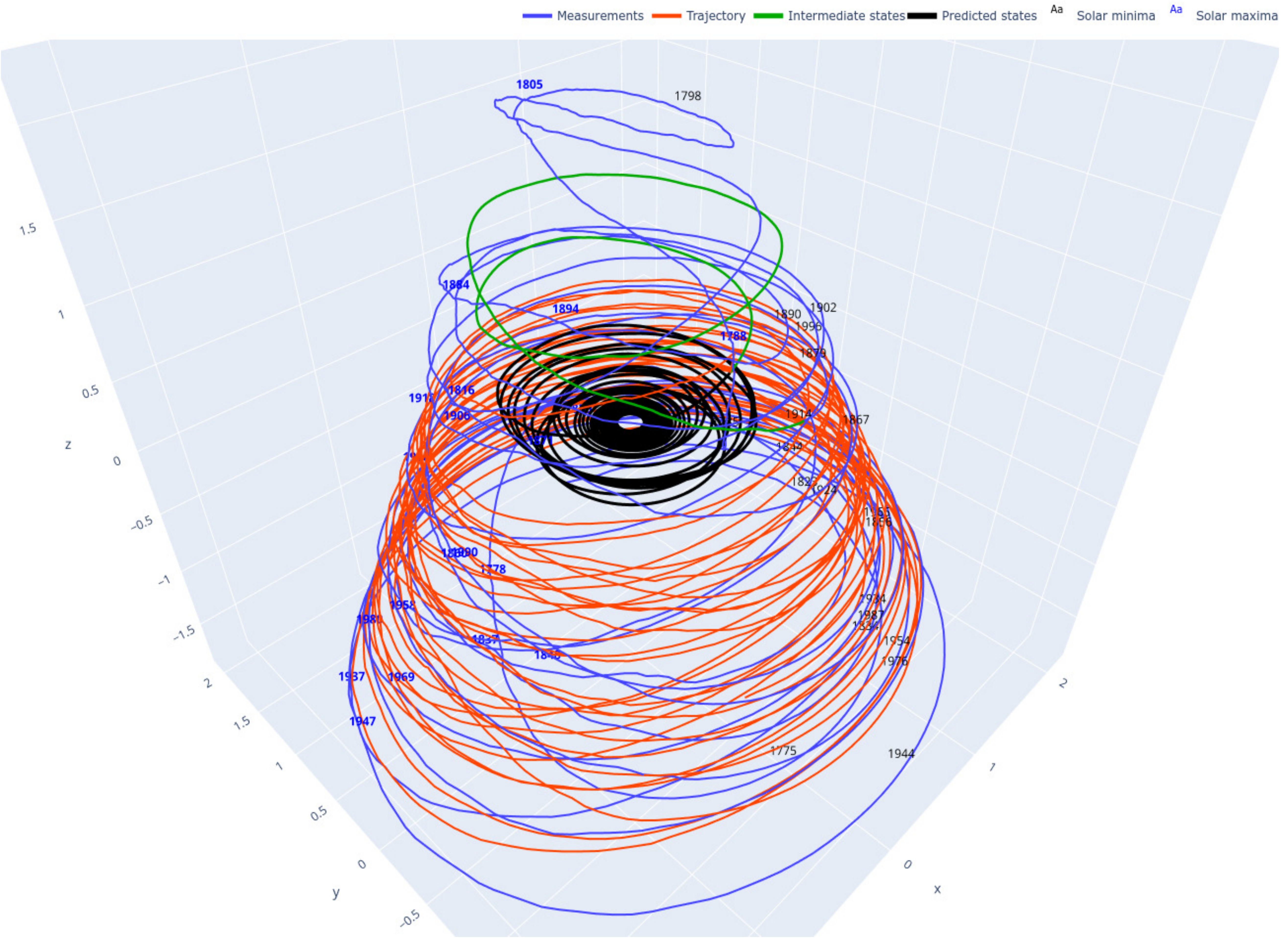
Sunspots – Predictions on the structure

Inferred state variables

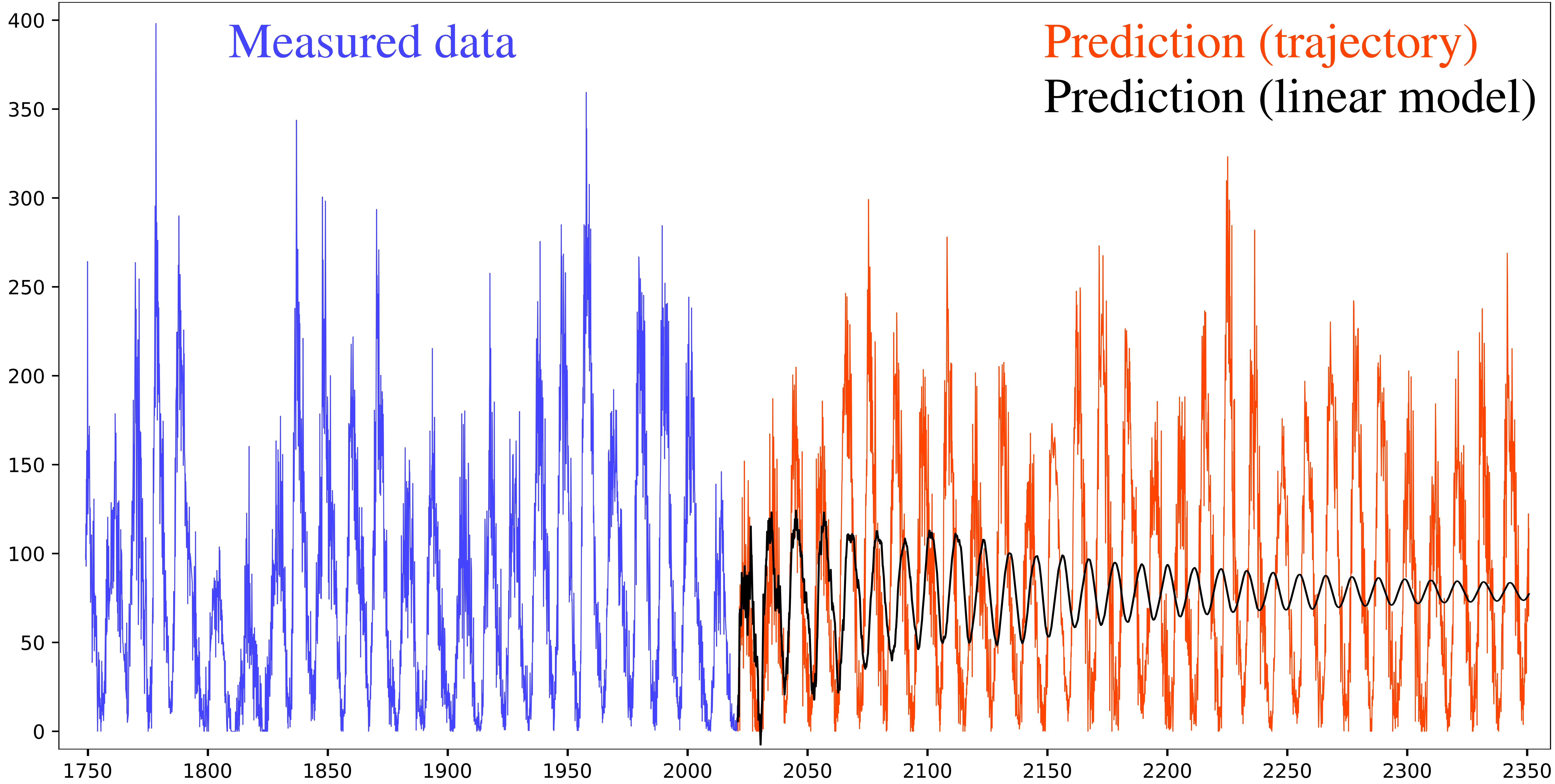
– 11-years cycle (x and y)
– Amplitude modulations (z)

Trajectories on a structure *resembling* an attractor embedding

Predictions
– Trajectory constrained on the structure
– **Linear operator converging to the average state**



Sunspots – Predictions in the data space



Robustness to measurement noise

Noisy Lorenz system (SDE)

$$du = -a\,(u - v)\,dt + \eta dW$$

$$dv = (bu - v - uw)\,dt + \eta dW$$

$$dw = (-cw + uv)\,dt + \eta dW$$

$(a, b, c) = (10, 28, 8/3)$

Data

– Simulation: $V(t) = [u, v, w](t)$

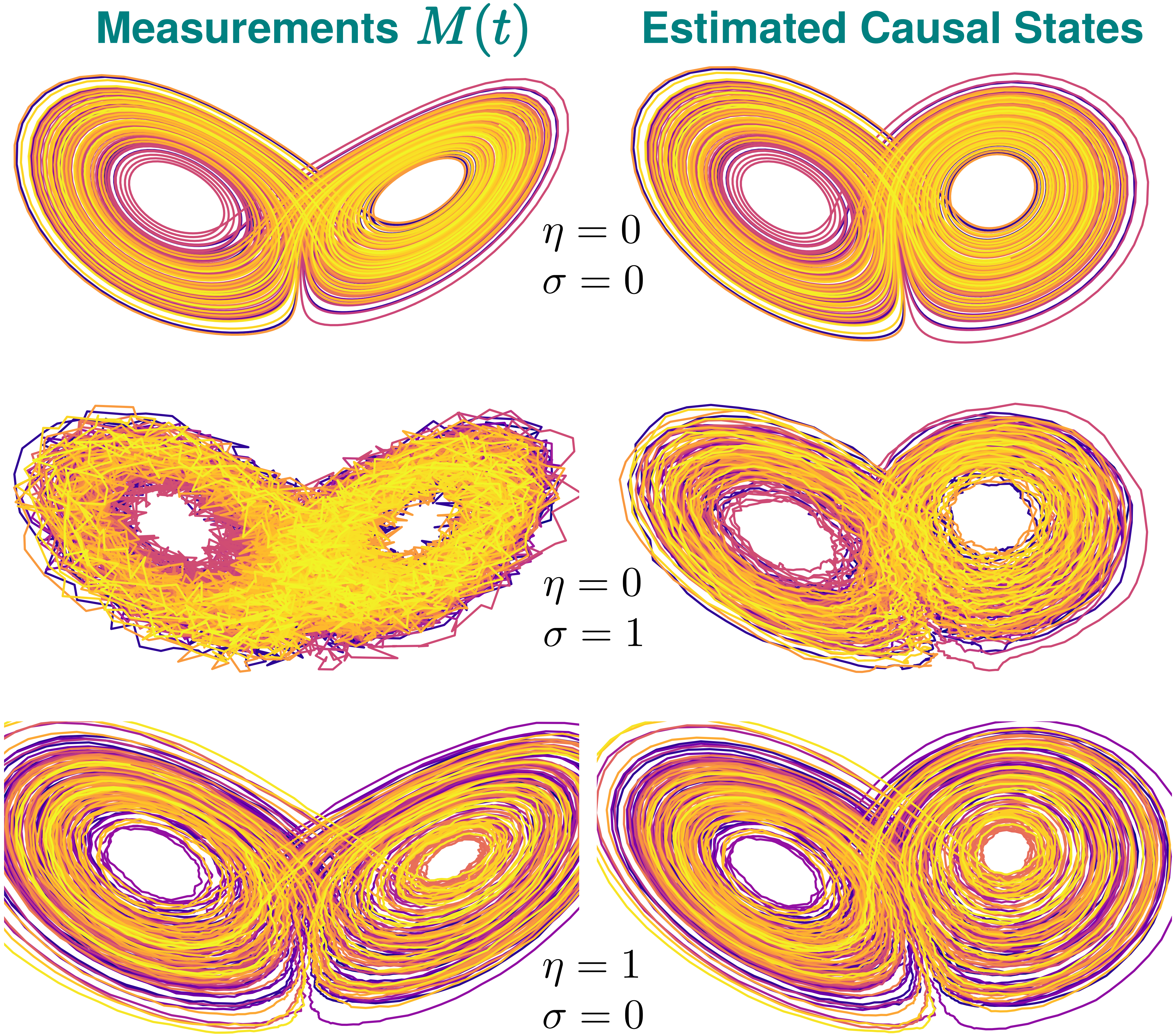
– Addition of Gaussian noise, var. σ^2

– Mesurements: $M(t) = V(t) + G(\sigma^2)$

Causal states

– Adding noise does not change equiv. classes ⇒ robust to measurement noise!

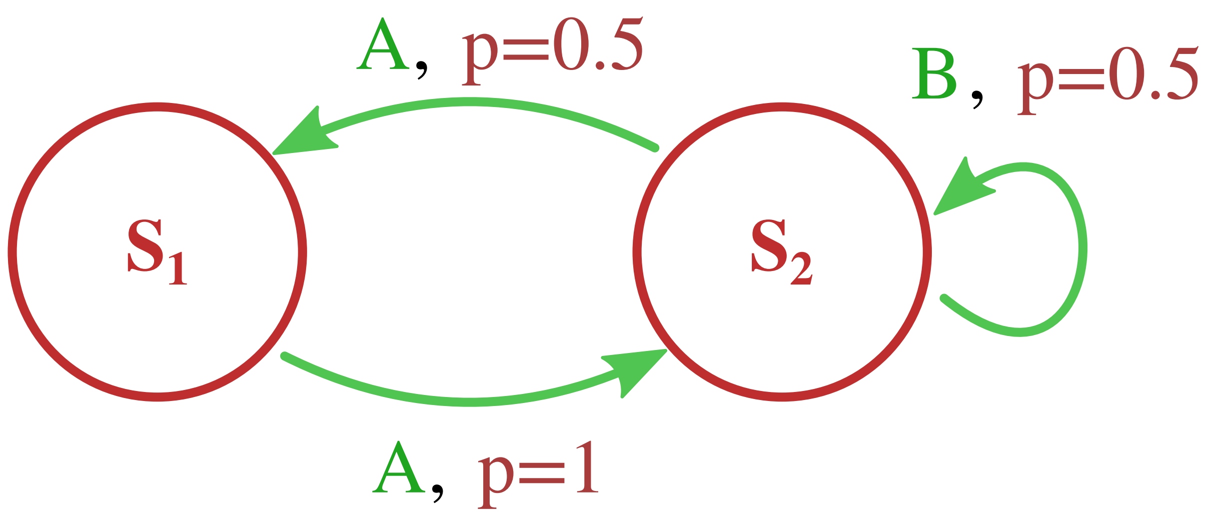
– Intrinsic noise: SDE details preserved



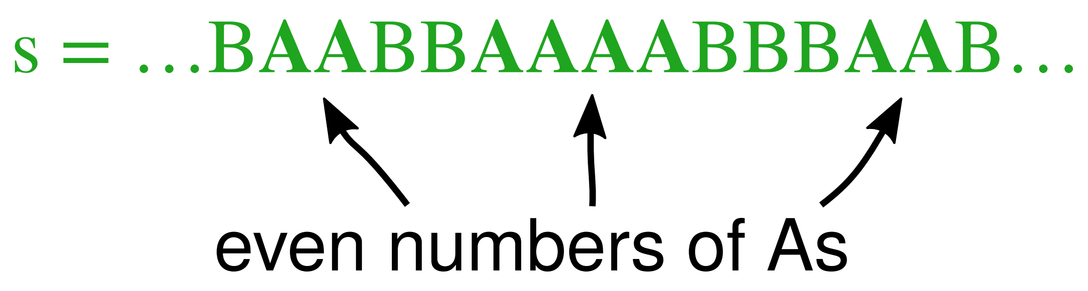
Finite state machines, HMM

Formal system

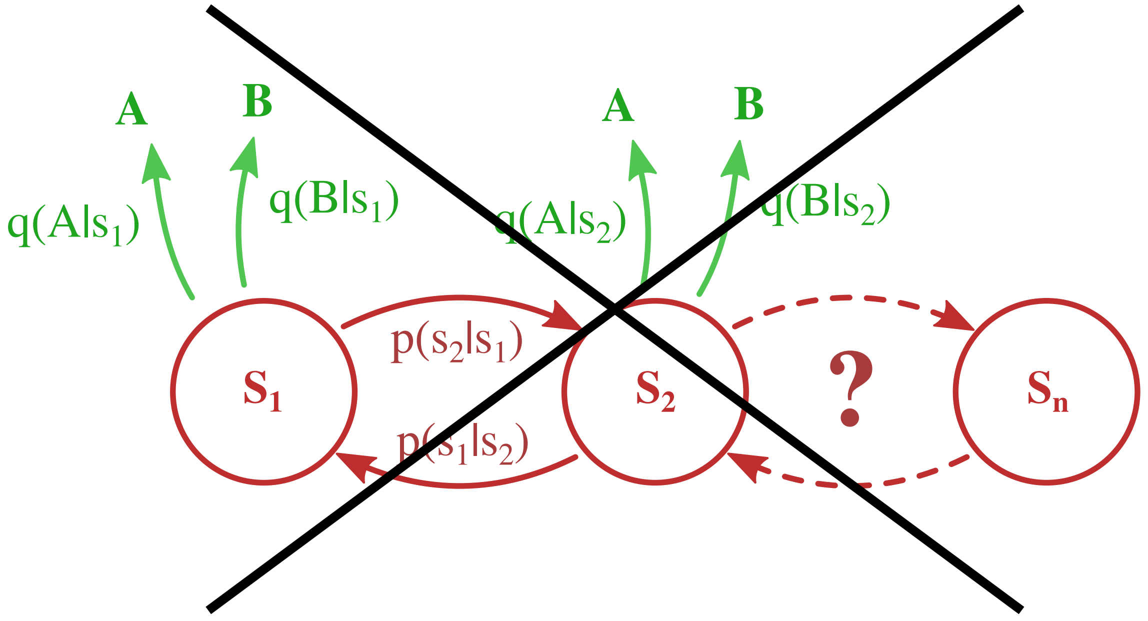
Observables = strings ∈ {A,B}*



Example : Even process



[Crutchfield 1988]



∞ Markov order, ∞ states

Minimal edge-emitting & unifilar HMM ⇒ States are causal states

NOT a state-emitting HMM

Inferred from data

Generate

Ensemble of observations

past → future

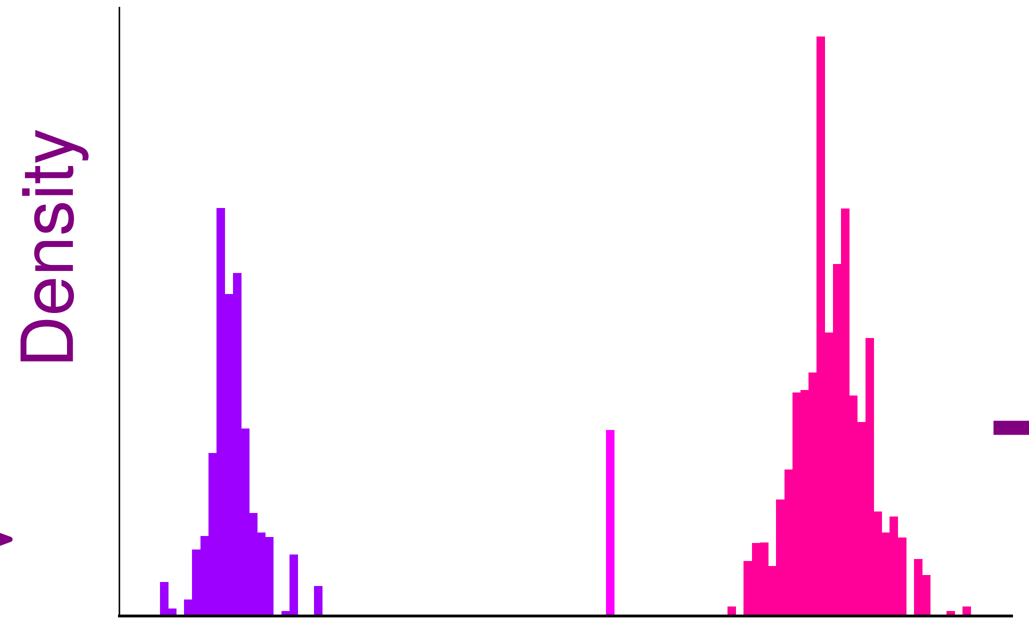
…BAABBAAABAAA…

…AABAAAABBAAB…

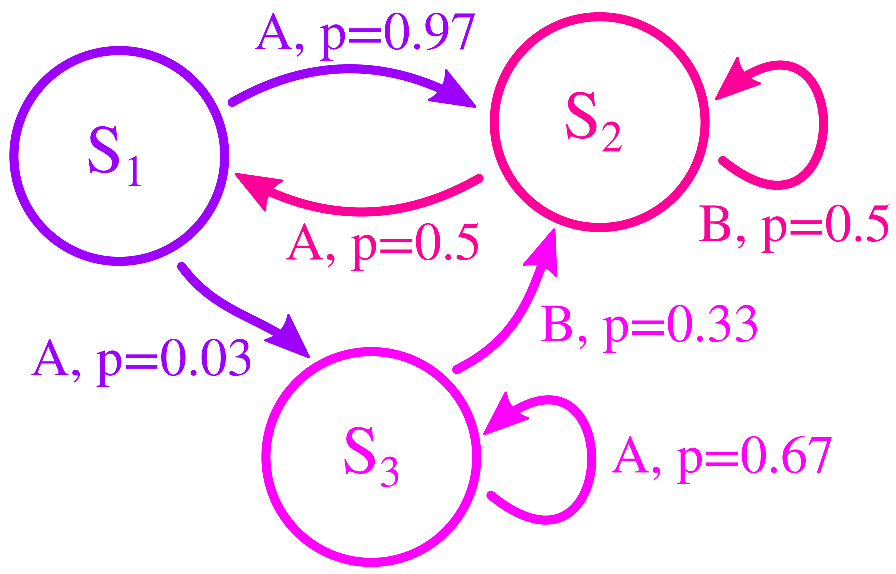
…ABAABBBBAABB…

Infer

[Brodu 2022]



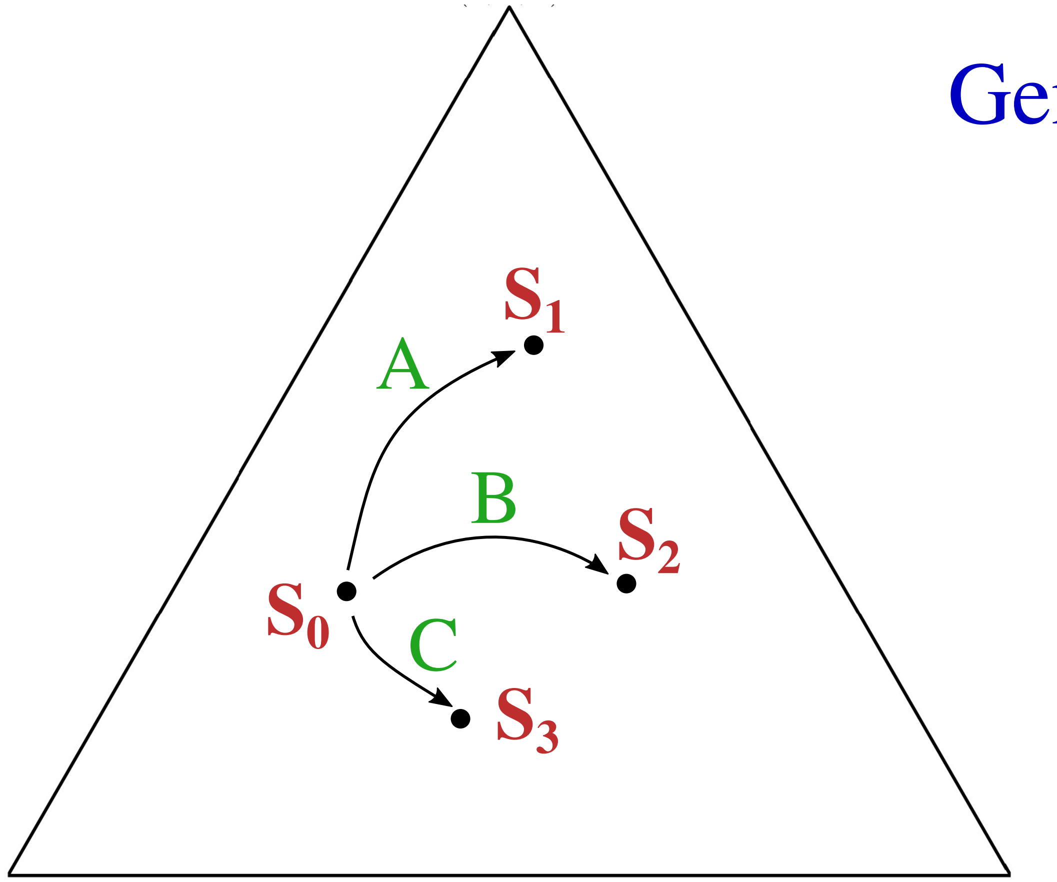
Cluster



Finite length artifact
(sequences of all A)

Generic HMM, IFS

Formal system



Jump dynamics

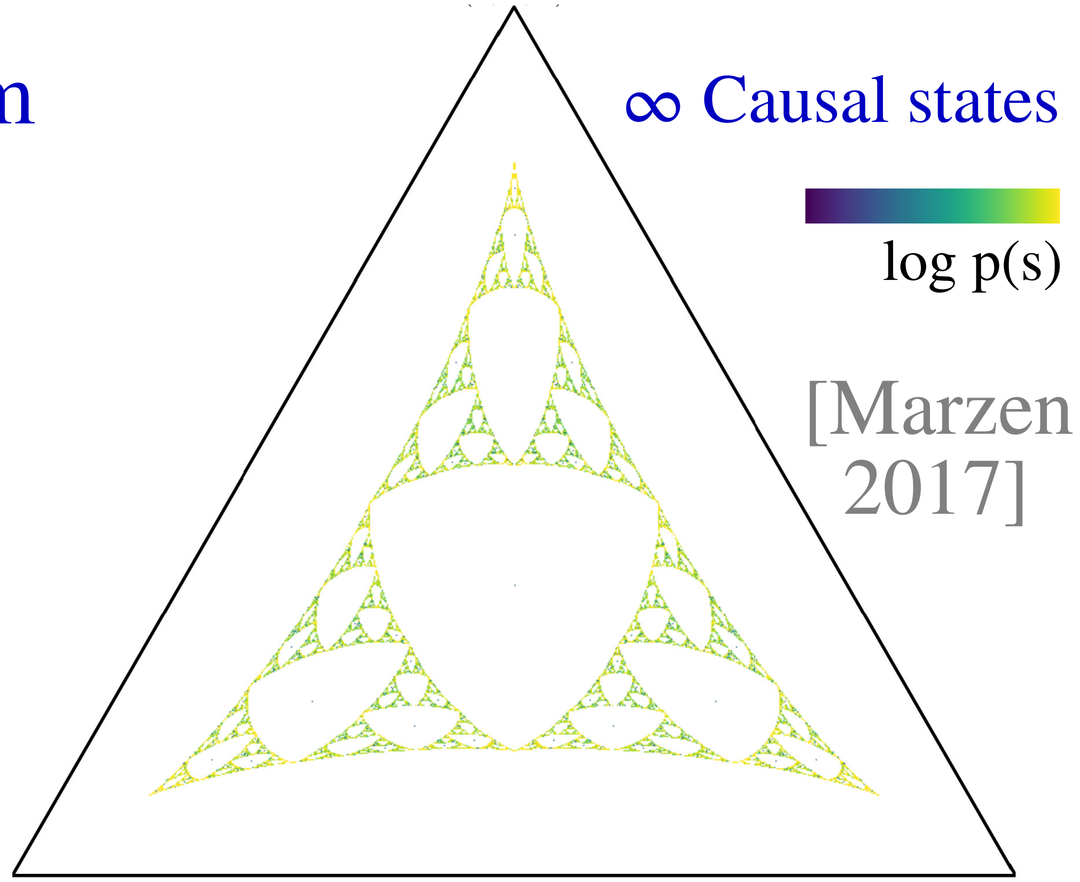
Generic HMM ⇒ Iterated Function System

[Jurgens 2021]

Symbols $i \in \{A, B, C\}$

$$s_{n+1} = f^{i}\left(s_n\right) \quad \text{with} \quad p\left(i|s\right)$$

Observables = strings ∊ {A,B,C}*



[Marzen 2017]

Theoretical limit density

Inferred from data

Generate

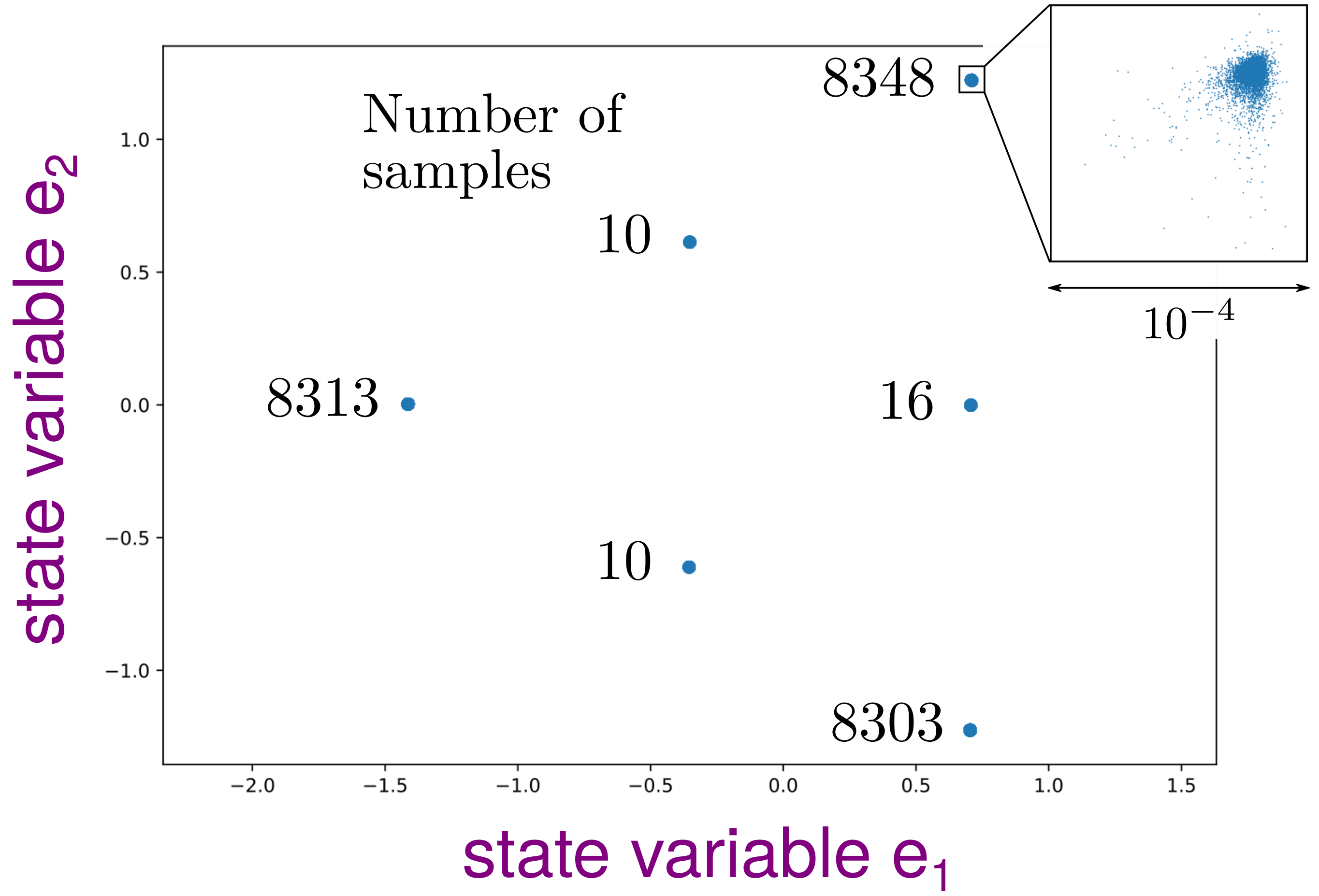
Ensemble of observations
past → future

...BACBBCACBCAB...
...ACABBCBACCAA...
...CBACBAACBABC...

Using 25k samples

Infer

[Brodu 2022]



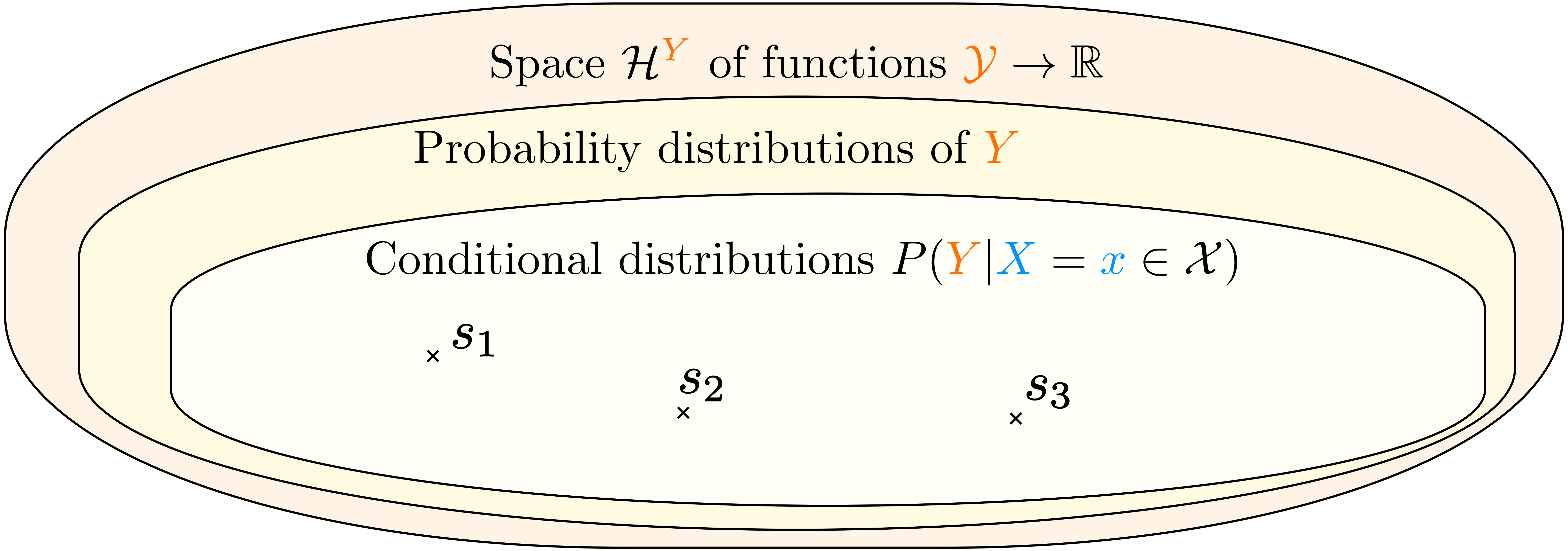
From causal states to state variables - 3. Cooking recipe

Step 1: A functional space for representing conditional distributions

Step 2: Geometric structure of the causal states set

Step 3: Parametrize that set structure ⇒ coordinates = effective variables

A functional space for representing conditional distributions



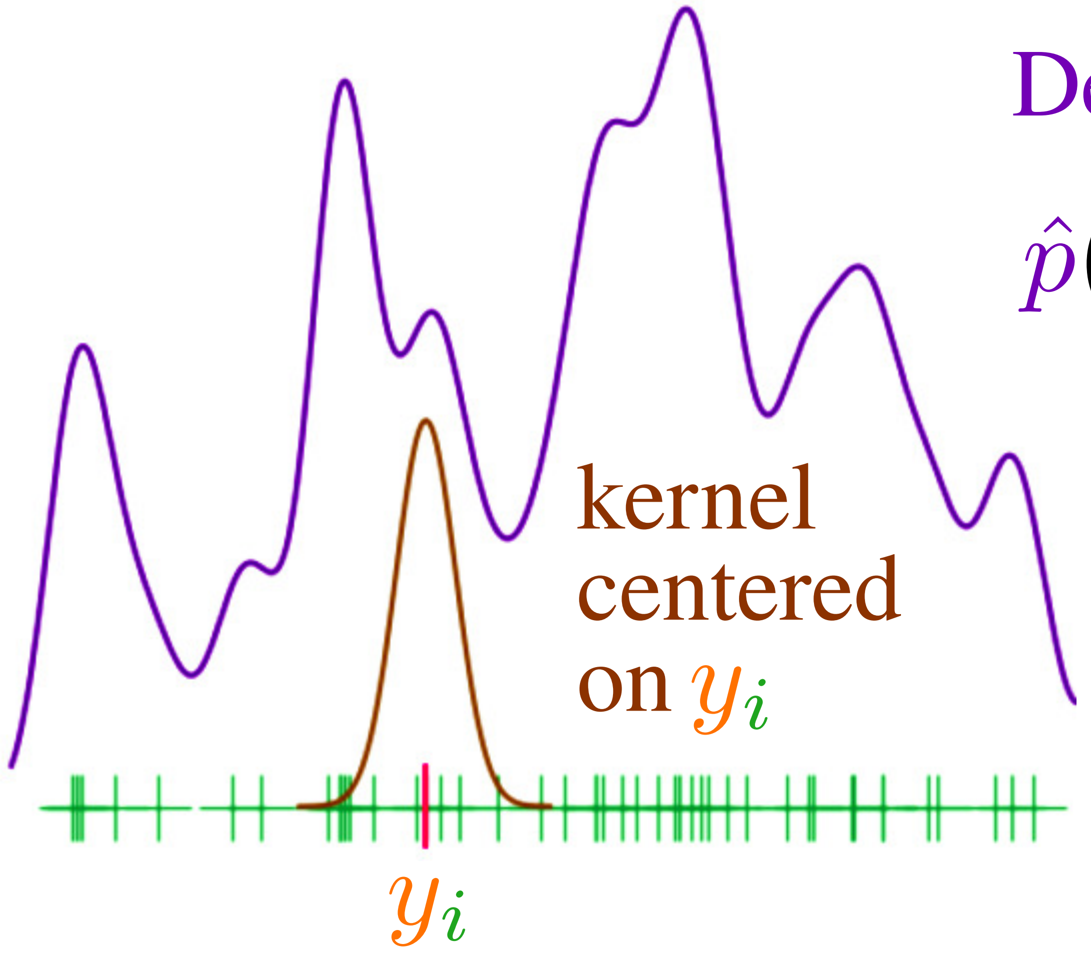
Causal state = {same P(future | past)} $\Rightarrow$ whole class mapped to the same point

Some possibilities : explicit functional basis, neural networks, information geometry…

Easier and convenient : Reproducing Kernel Hilbert Spaces (RKHS)

[Brodu 2022]
[Loomis 2023]

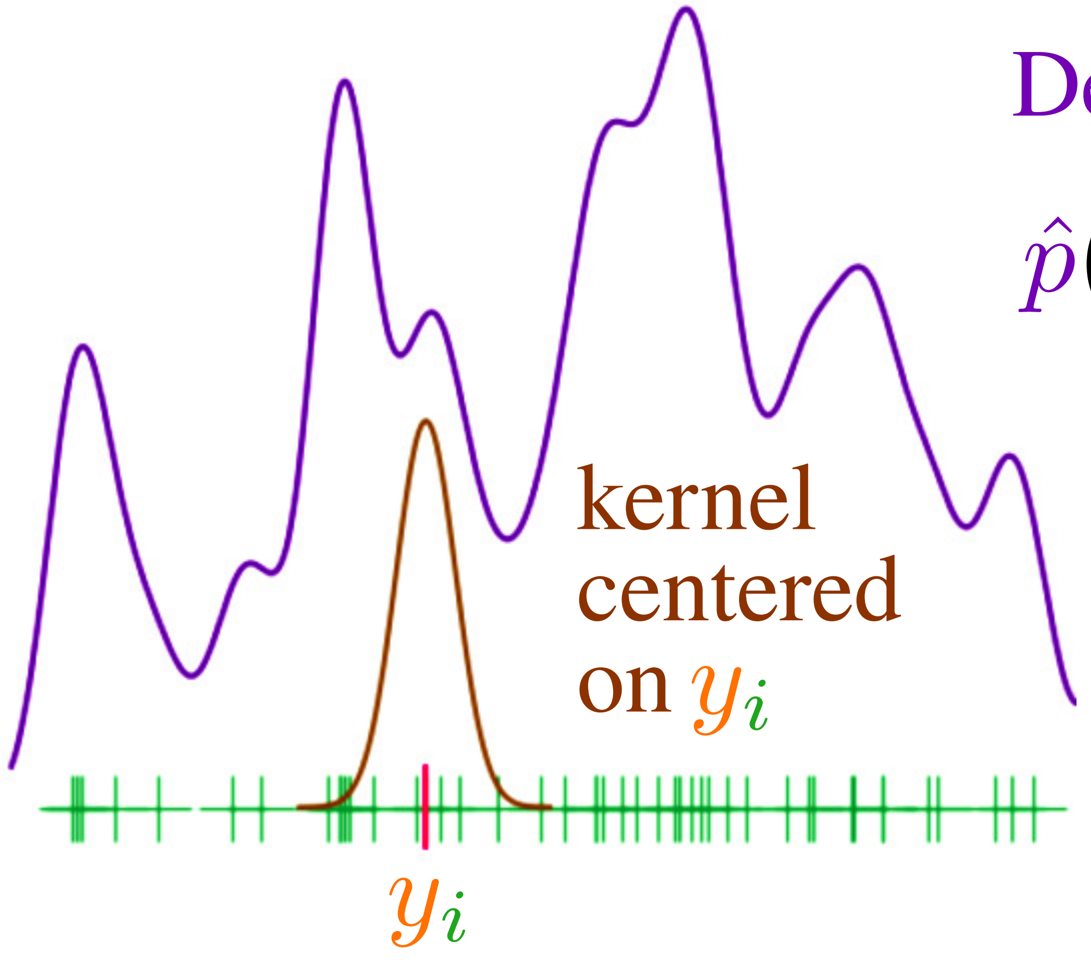
RKHS embedding of distributions - Some intuition



Density estimation

$$\hat{p}(y') = \frac{1}{N} \sum_i k(y_i, y')$$

RKHS embedding of distributions - Some intuition

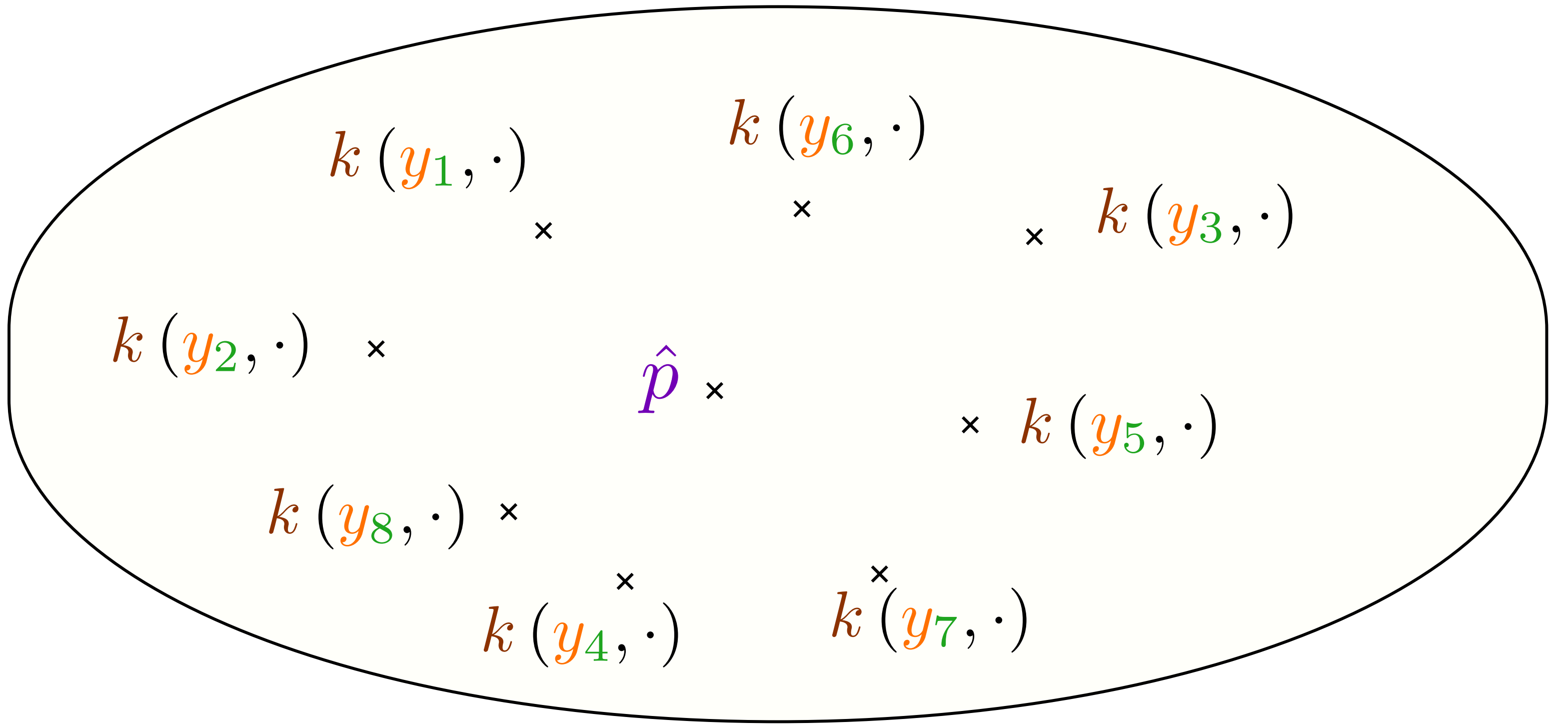


sample values

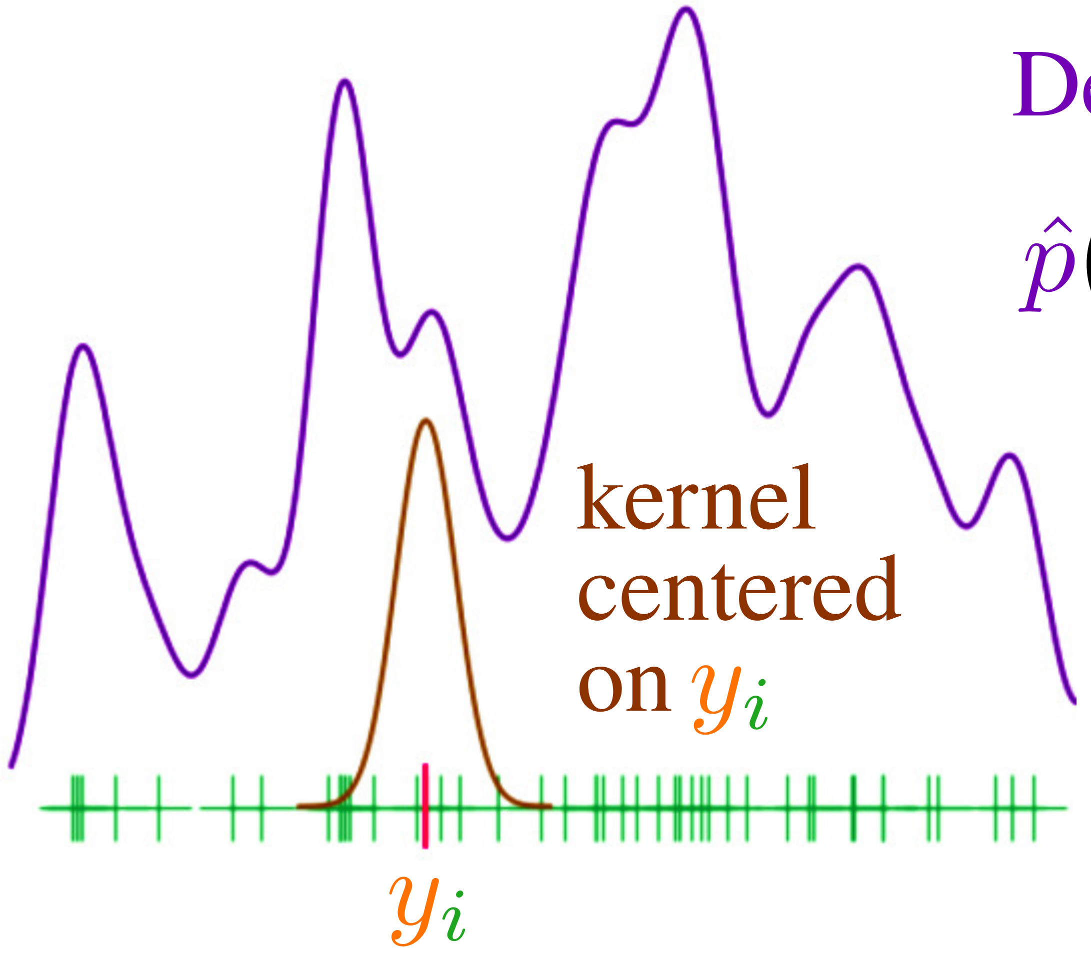
Density estimation

$$\hat{p}(y') = \frac{1}{N} \sum_i k(y_i, y')$$

Geometric view = barycenter



RKHS embedding of distributions - Some intuition

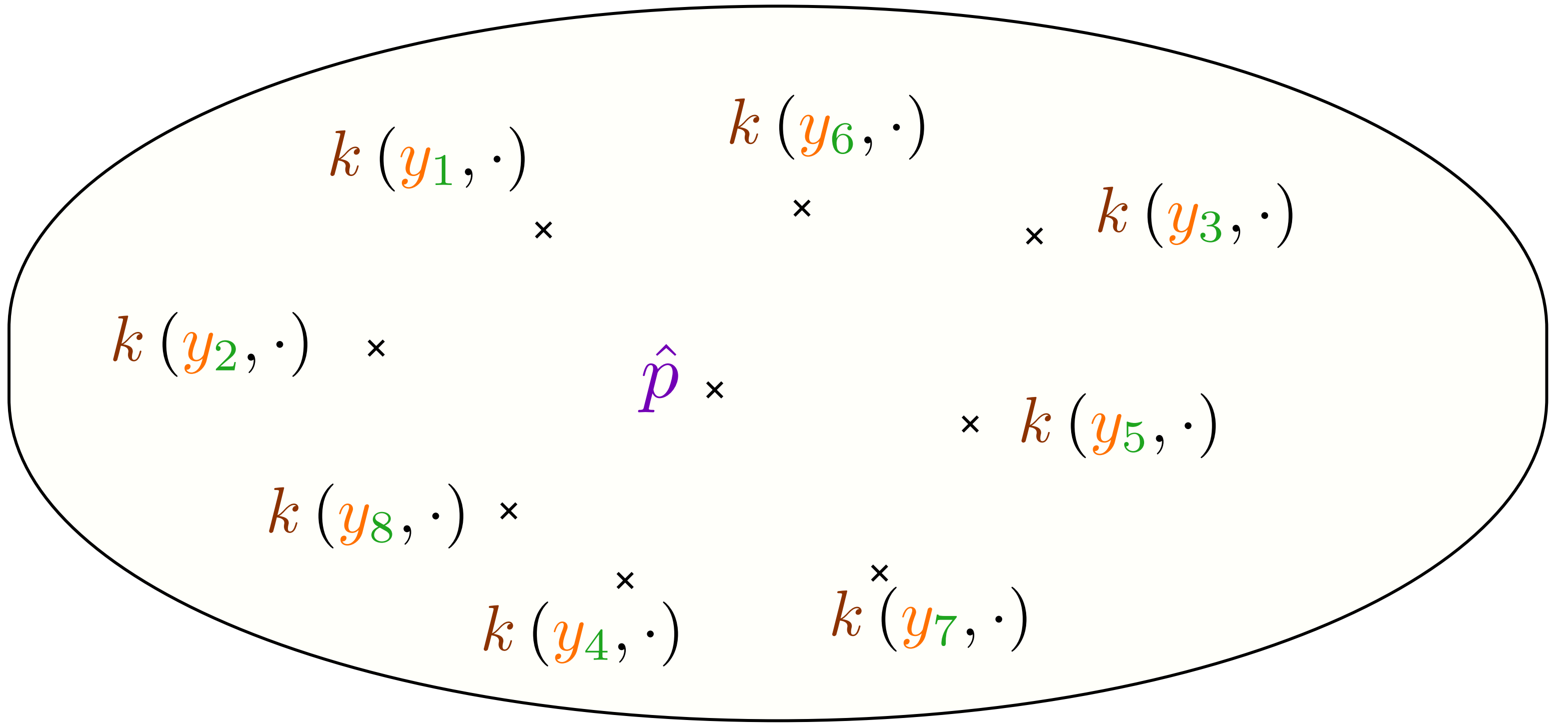


sample values

Density estimation

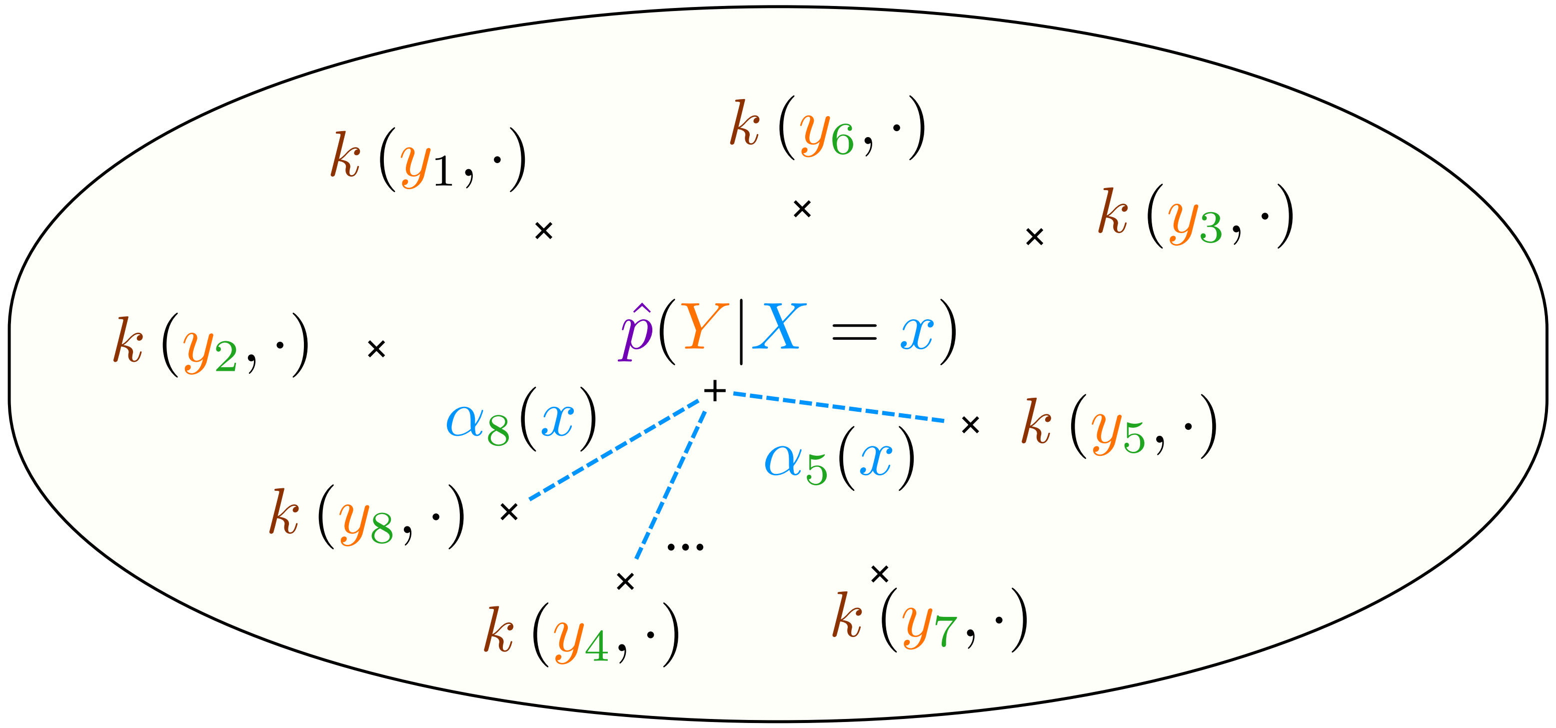
$$\hat{p}(y') = \frac{1}{N} \sum_i k(y_i, y')$$

Geometric view = barycenter



Conditional distribution = weighted average

[Gretton_al_12]



Conditional density estimation

$$\hat{p}(Y|X = x) = \sum_i \alpha_i(x) k(y_i, \cdot)$$

$\alpha_i(x)$ = proximity with x_i, using $k'(x_i, \cdot)$

sample values = pairs (x_i, y_i)

Working with real data: combining kernels

Reproducing kernel property [Aronszajn 1950]

$$\langle k(x,\cdot), f\rangle_{\mathcal{H}} = f(x)$$

k generalizes the δ function $\langle f, \delta_x\rangle_{L^2} = \int f\delta_x dx' = f(x)$

Inner product : k is positive symmetric definite

Many known kernels, many data types (scalars, vectors, graphs, strings…)

Kernels can be combined, scaled ⇒ Consistently merge heterogenous data sources

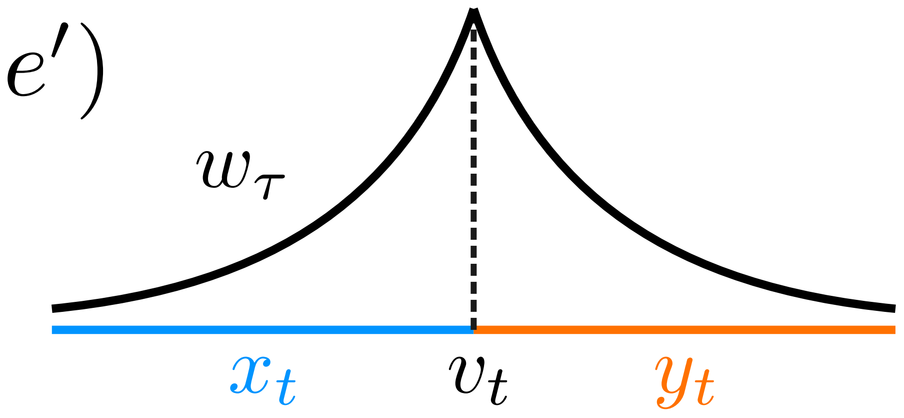
E.g., T = temperature, P = precipitations, E = evapotranspiration

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$$

$$k^V(v,v') = k^T(t,t')\ k^P(p,p')\ k^E(e,e')$$

$$\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$$

$$k_{\text{seq}} = \sum_{\tau=0}^{L} w_\tau k\left(x_{t-\tau},\cdot\right)$$



Kernels act on dimentionless data ⇒ A scale is needed for each source

$$k^V\left(\frac{v}{\lambda},\frac{v'}{\lambda}\right) = \exp\left(-\left\|\frac{v}{\lambda}-\frac{v'}{\lambda}\right\|^2\right)$$

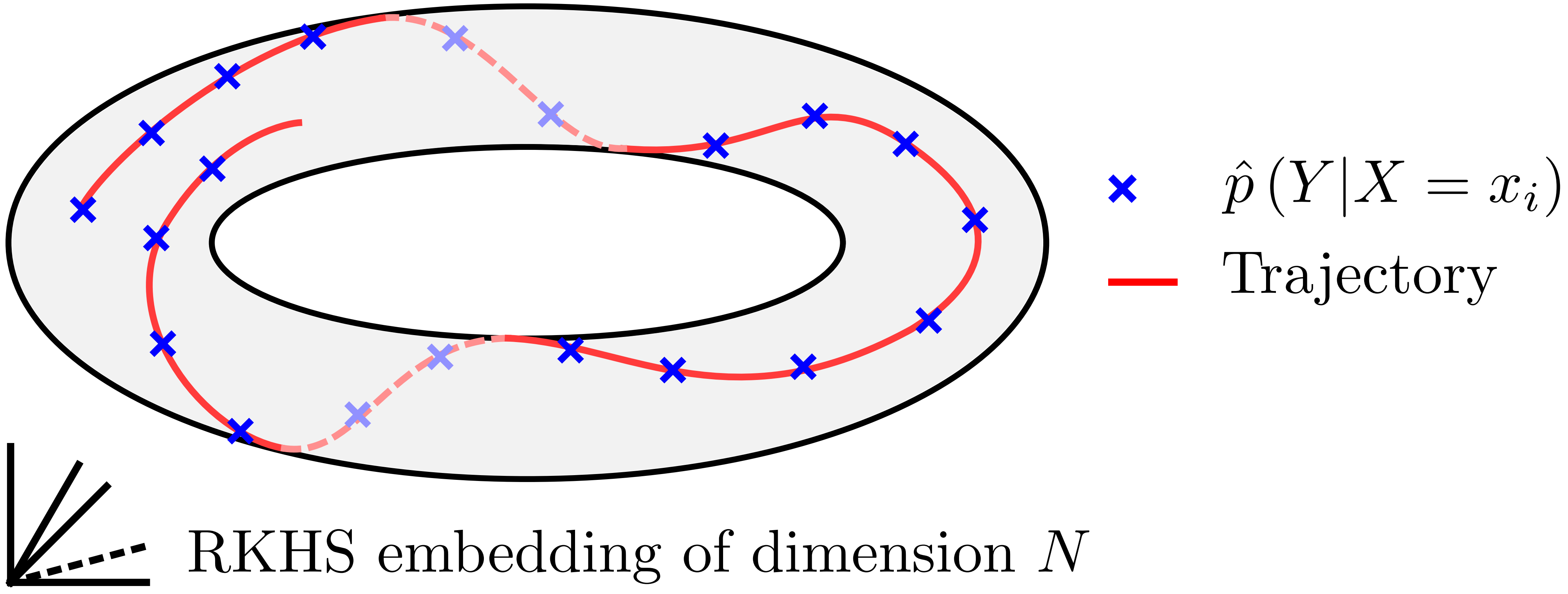
Geometric structure of the causal states set

Hypothesis: system described by M state parameters

⇒ causal states set dimension at most M (manifold, fractal…)

Physical property: does not depend on data size N

States lying on a reduced dimension set $M \ll N$

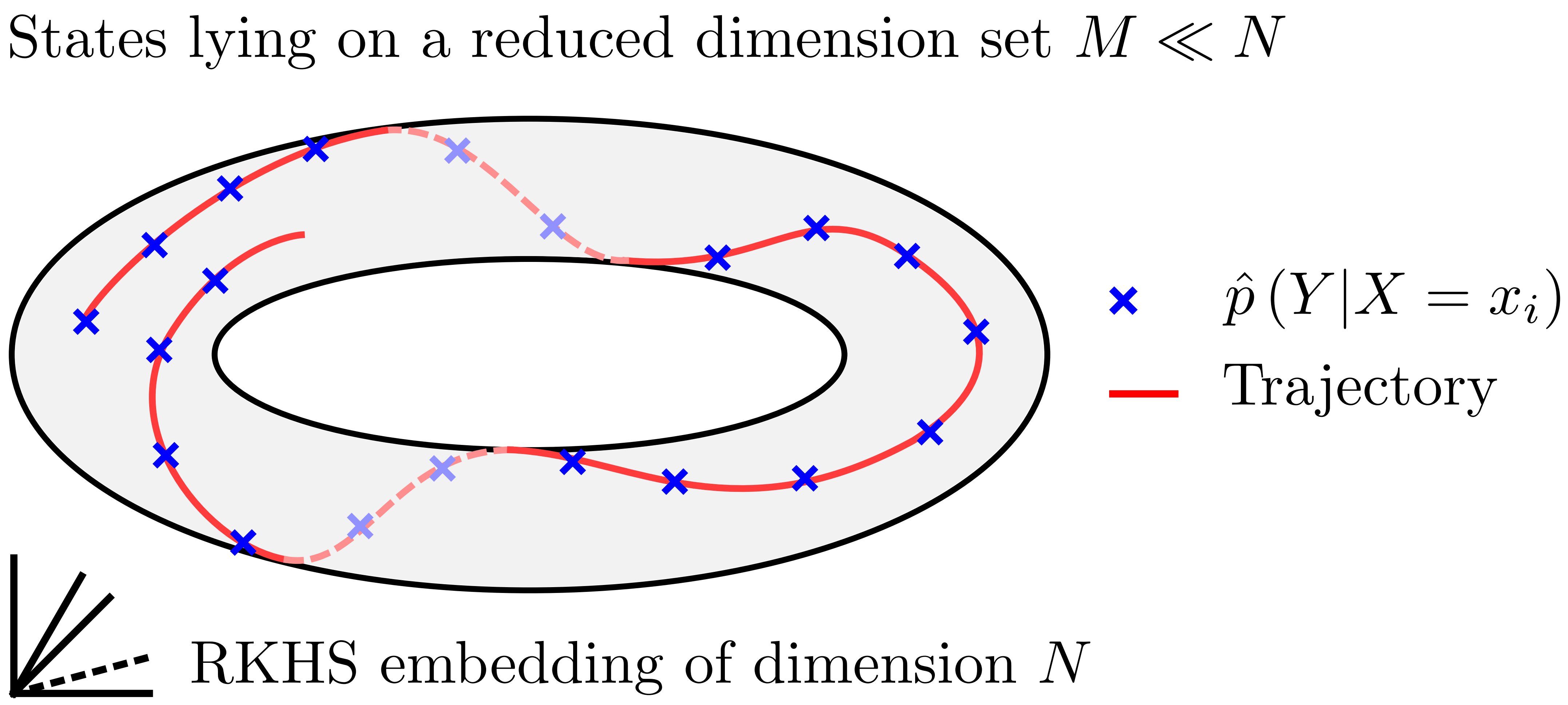


Steps 3: Coordinates ⇒ Effective state variables

Hypothesis: system described by M state parameters

⇒ causal states set dimension at most M (manifold, fractal…)

Physical property: does not depend on data size N



[Brodu 2022 + in prep.]

Choice of basis = Generalized Fourier modes of minimally embedding manifold

– Causal states = distributions over future ⇒ Additional modes refine predictive info

– Estimated with diffusion maps ⇒ M inferred from an eigenspectrum decay profile

[Coifman 2006, Berry 2020]

CO_2 Flux and evapotranspiration : Grignon site (INRAE)

Data

– Experimental field, Grignon FR

– Temperature, soil humidity, sun illumination, évapo-transpiration, precipitations

Photo : bing.com

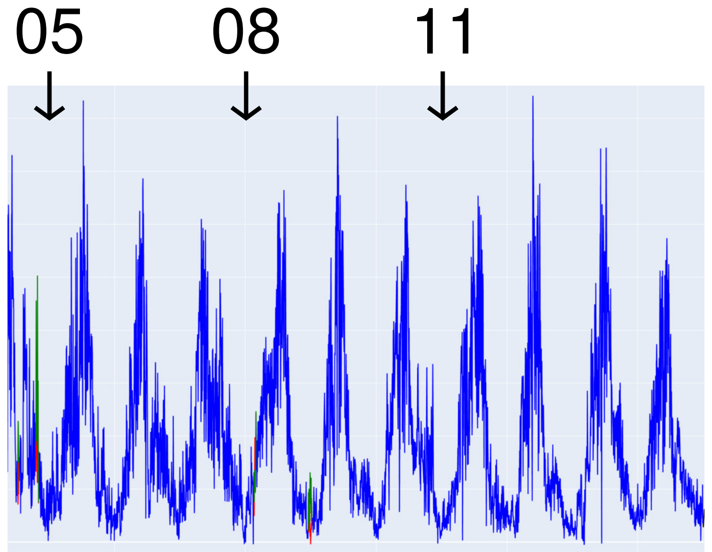
– ICOS data (flux tower + in situ sensors), 11 years of daily observations

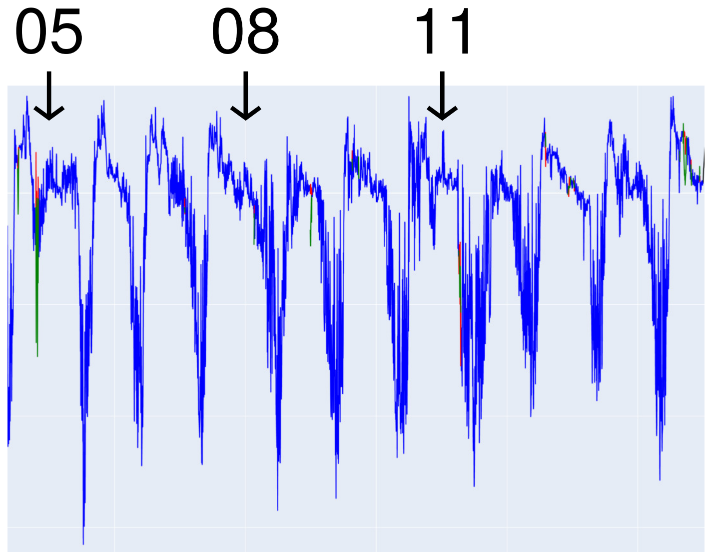
Dynamics

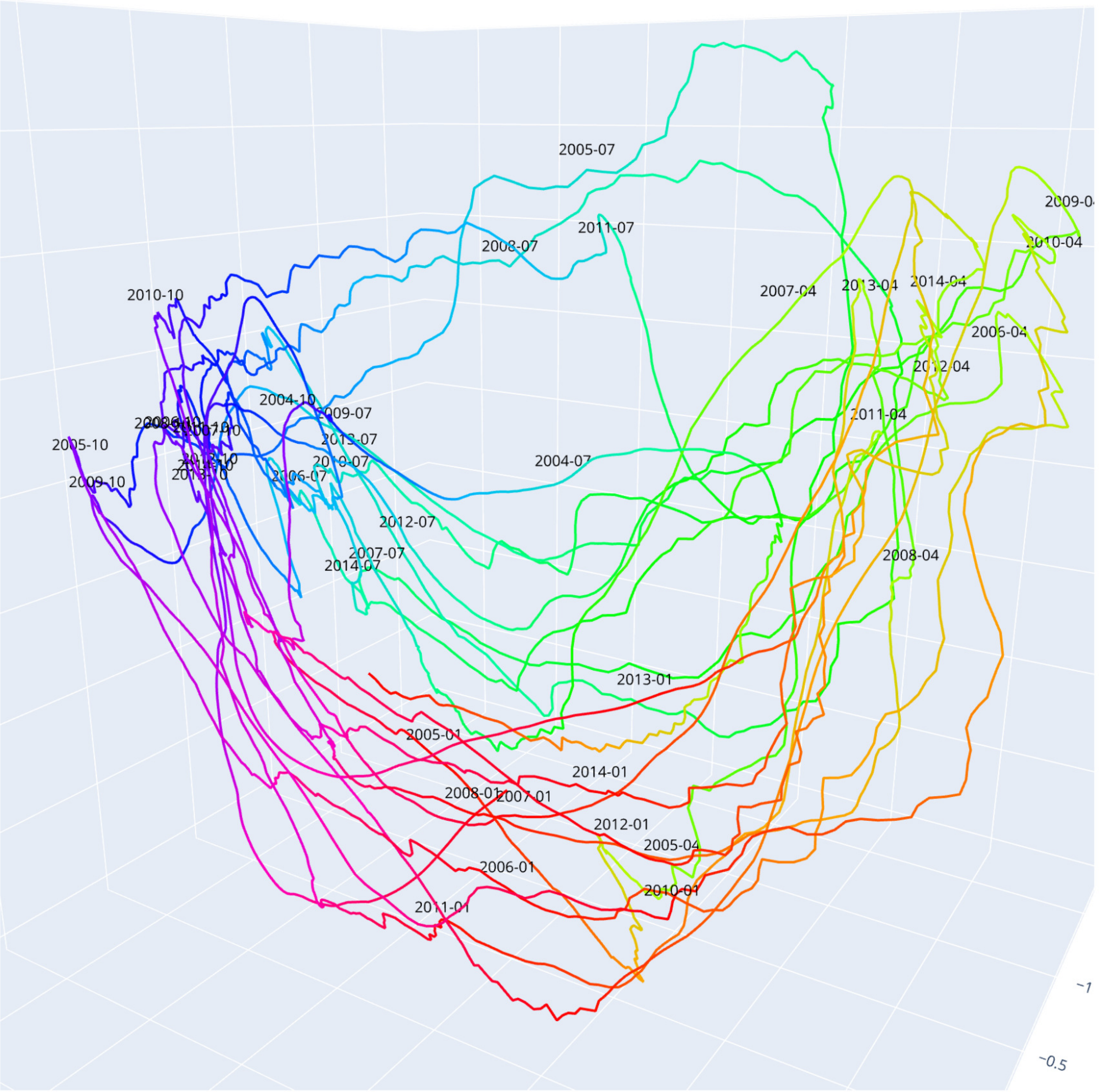
– Variables 1, 2 = seasonal cycle. Var. 4 (visible) : different trajectories in 2005, 2008, 2011

– Culture of corn then ⇒ plant response is ≠

– Impossible to distinguish visually on raw data





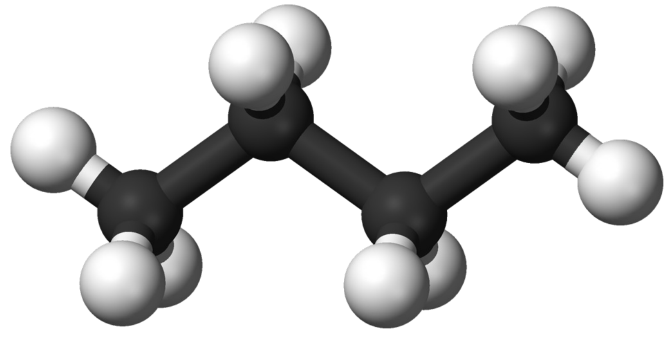


[Soon on Arxiv]

colors = days of the year

Molecular dynamics of n-butane

[Soon on Arxiv]



n-butane molecule (wikipedia)

Inputs Data : Stefan Klus

– Positions x,y,z of atoms sampled every 200 fs

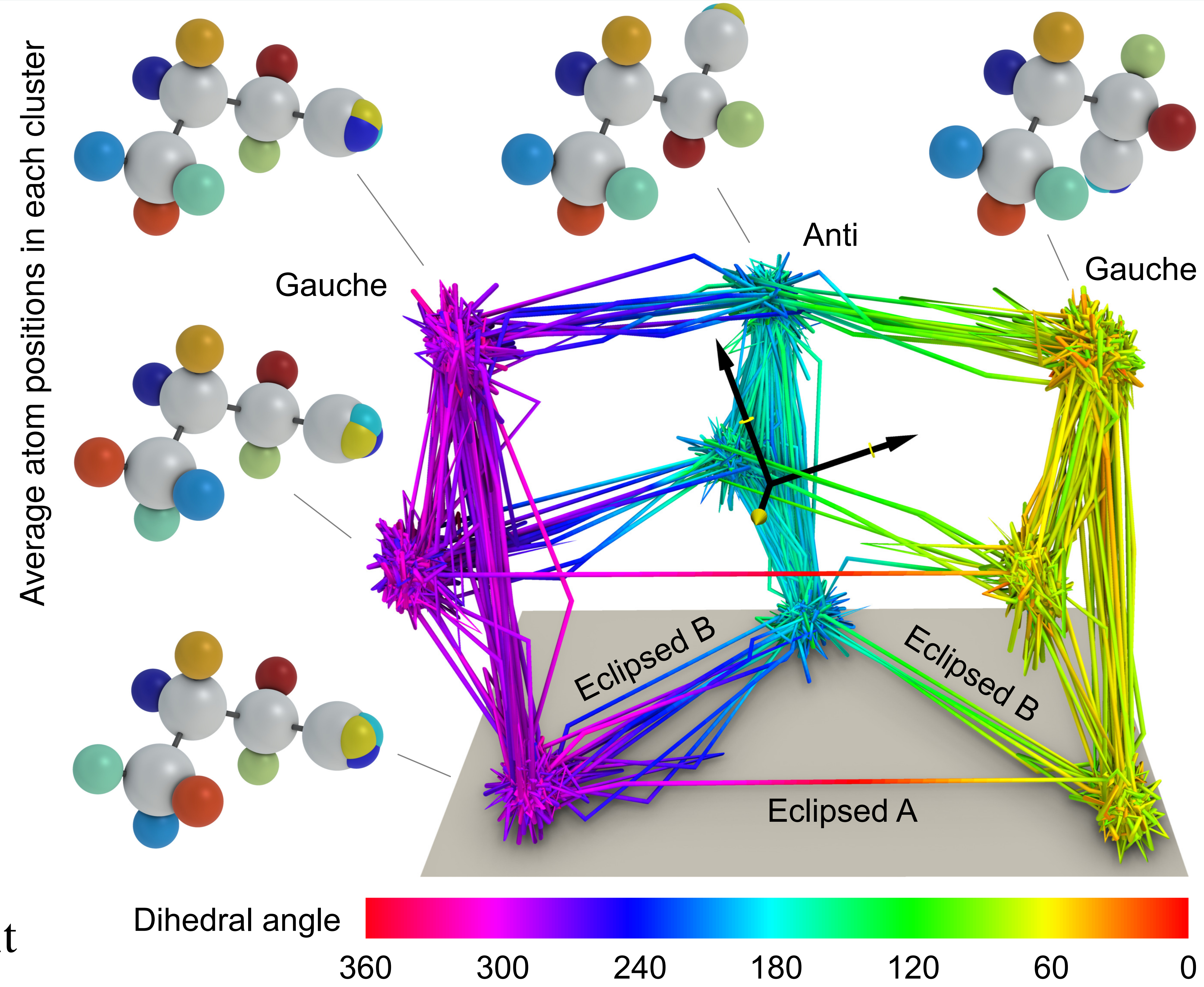
– Local frame of reference

Dynamics

– Clusters = meta-stable conformations

– Sub-clusters = hydrogen position (chemically equivalent, distinct in data)

– Slow clusters + fast transitions ⇒ discrete dynamics at large Δt



Modeling ENSO

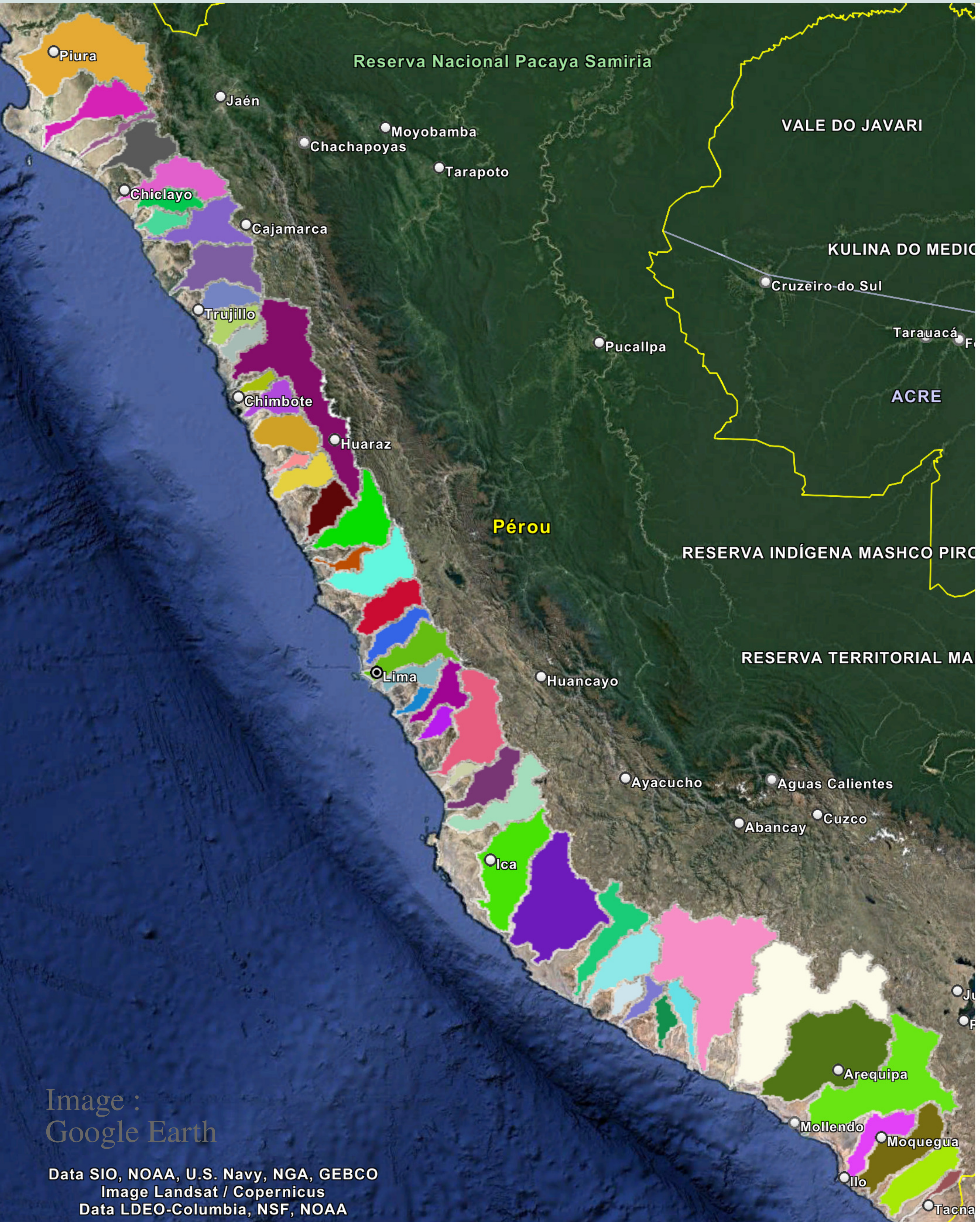


Image : Google Earth

Data

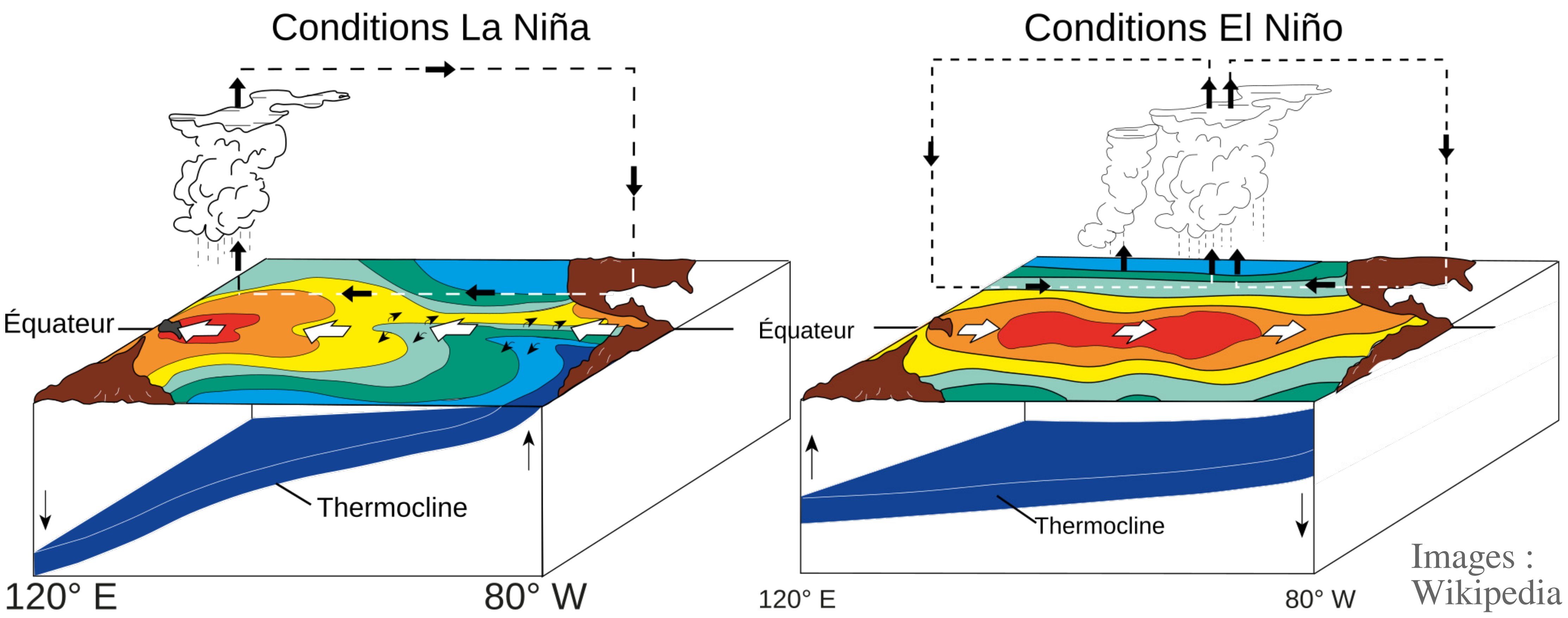
– 50 years of measurements, very high quality

– 49 watersheds, Peruvian coast

– Pacific Ocean : 4 indices of sea surface temperature

– Per watershed: precipitations, runoff, evapotranspiration, temperature

Project with: Luc Bourrel (France), Pedro Rau (Peru)



Images : Wikipedia

Modeling ENSO: preliminary results

Inferred state variables

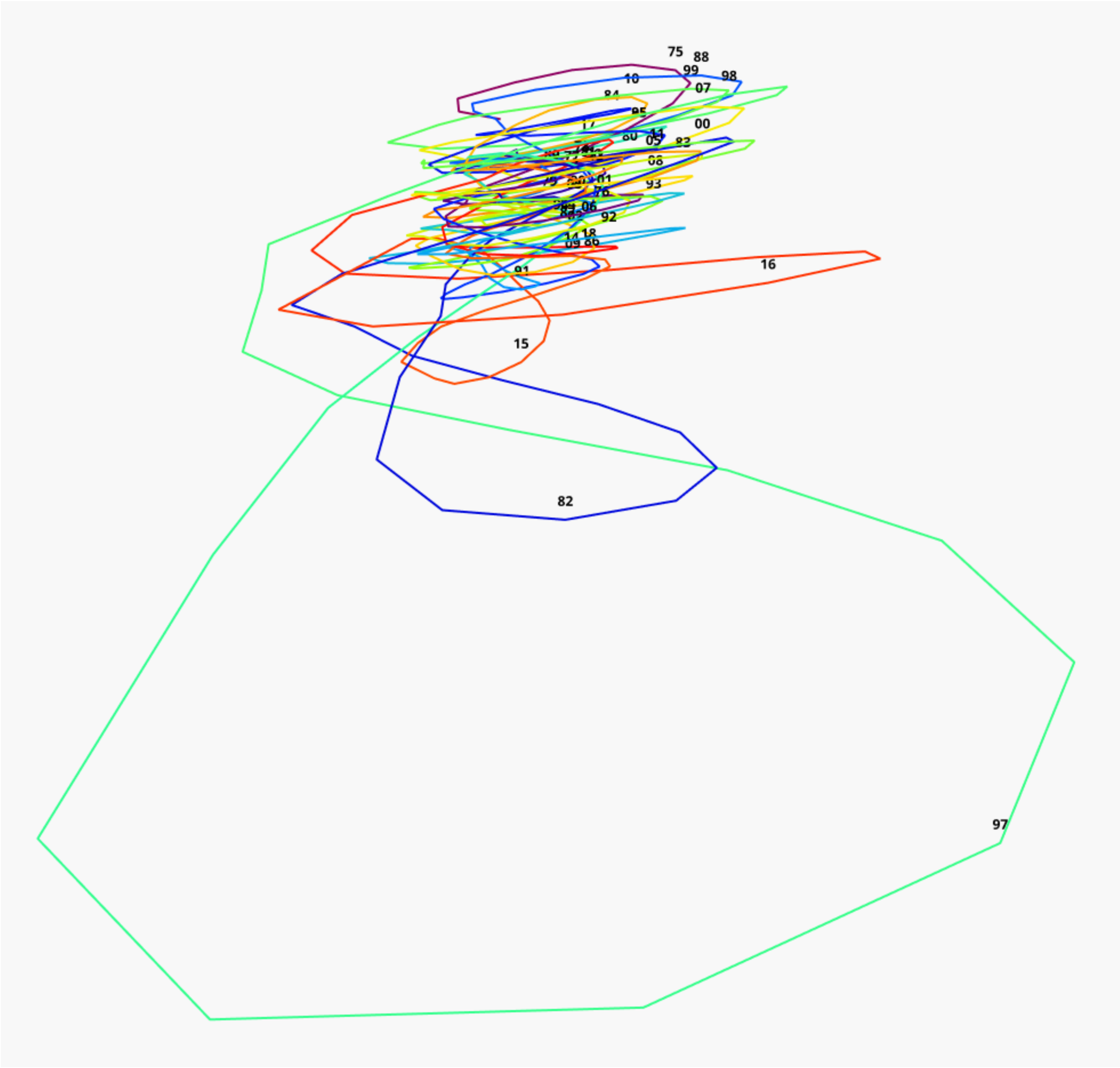
- Seasonal cycle
- General amplitude of the oscillation
- Much more than 3 dimensions… difficult to visualize

Anomaly detection

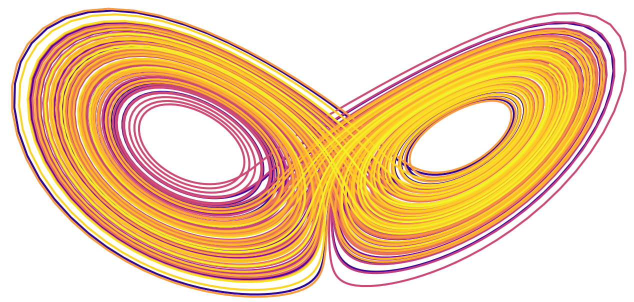
- Niño extremes of 1997, 1982, 2016
- 2016 weaker ⇔ closer to the regular structure
- Conditions Niña a the center of the structure
- Deviations of trajectories monthes before

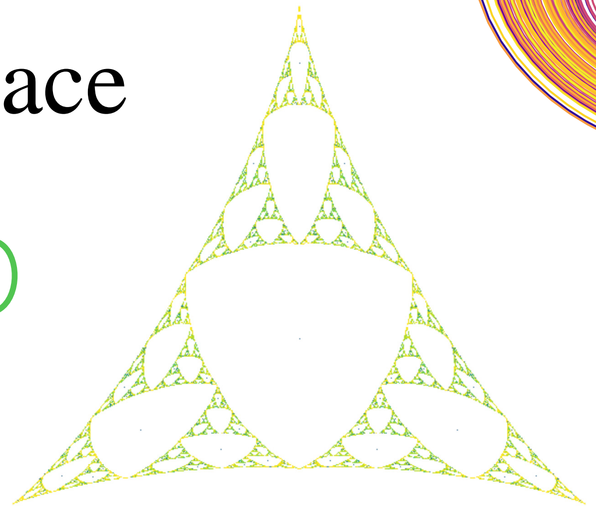
Work in progress - plans

- Trying various predictive models
- Regionalization : inference of local models for each hydroclimatically consistent region



Effective stochastic model: Work in progress

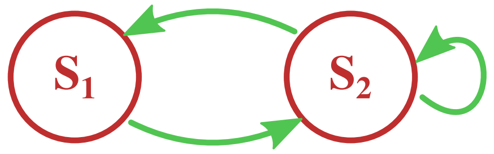




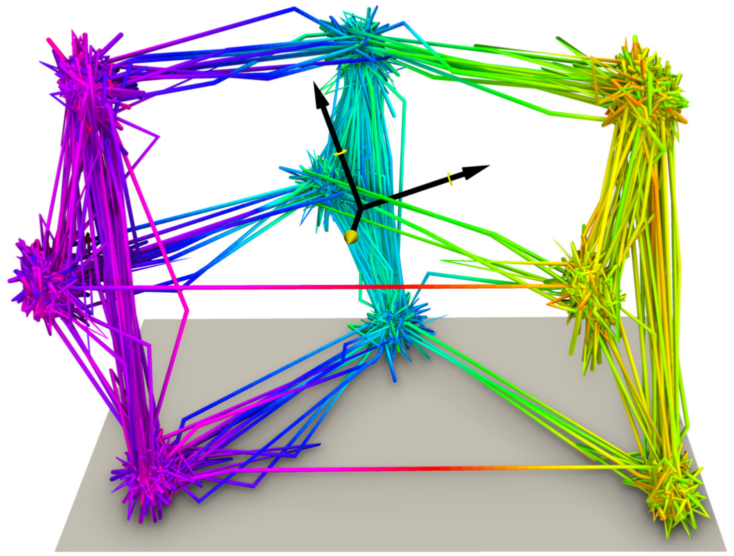
Diversity of cases to model:

– ODE / SDE: Causal states = 1-to-1 mapping with state space

– Finite states, edge-emitting unifilar HMM



– General HMM, IFS, infinite states



Data-based reconstruction could be anything in between!

– Propose an encoding of causal states $s_t \equiv \left(e_t^1, e_t^2, \ldots, e_t^M\right)$

– Fit the transition dynamics with that encoding

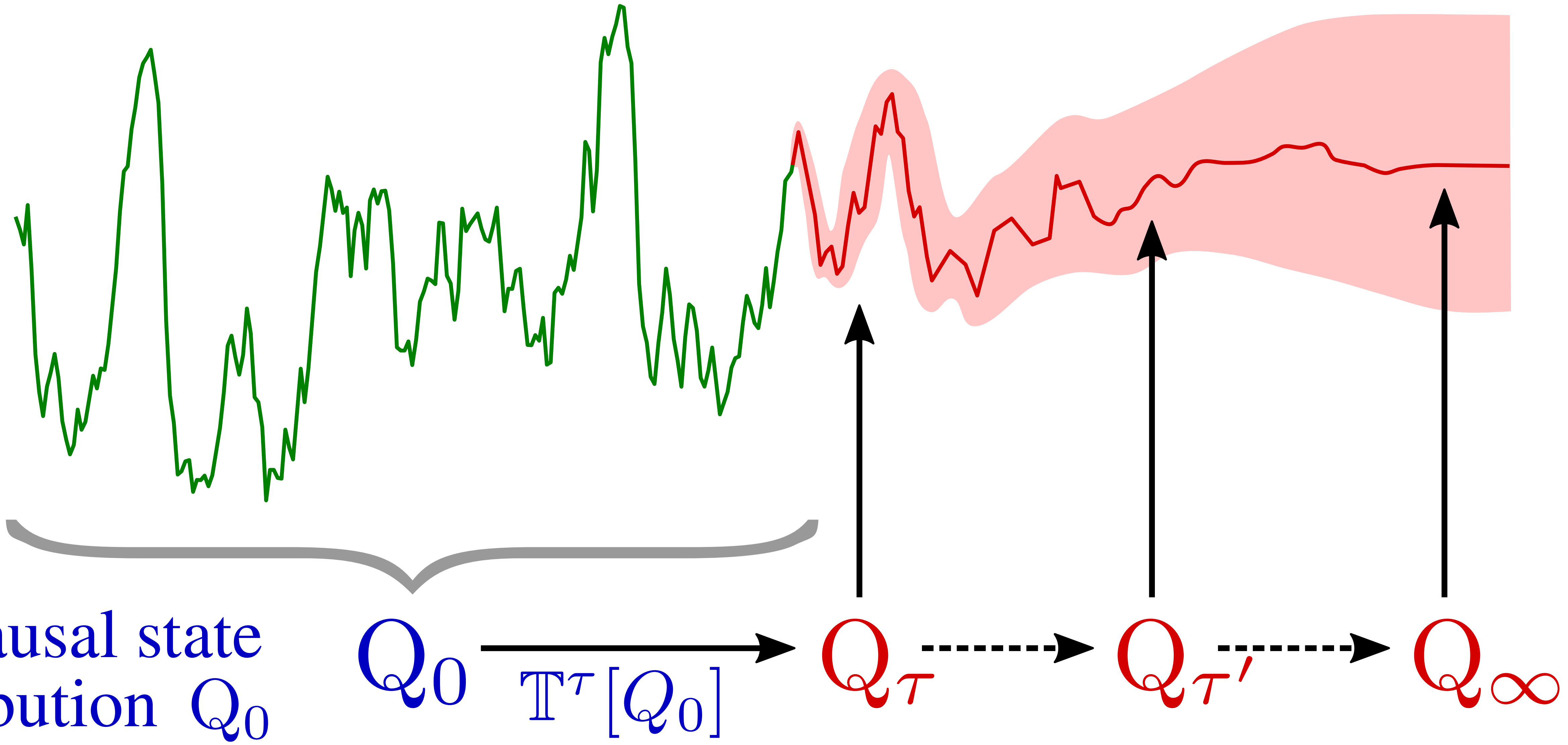
$$s_{t+1} - s_t \sim P\left(\Delta s, a \mid s_t\right) \quad \text{with density} \quad p\left(\Delta s, a\right) = f\left(e_t^1, e_t^2, \ldots, e_t^M\right)$$

$$ds = a(s)dt + b(s)dW \quad \text{if} \quad dJ(s) = 0 \quad \text{otherwise} \quad s' \sim P(s'|s)$$

– Work in progress: many ML tools and possibilities

Quantifying the diffusion of information

Measured data — **Predictions**



Converging in the limit to the data mean

Initial information (latest mesurements) completely lost

Initial causal state or distribution Q_0

$\mathbb{T}$ = transfer operator

SDE ⇔ Diffusion : distribution of states

Limit distribution of the SDE

(in theory…) This model specifies *how* predictive information diffuses through time!

Conclusion: Causal inference / Global modeling

Goals

Modeling
- Complex Systems
- Machine learning approach (≠ neural nets)
- Modeling the dynamics at the scale of data

Understanding
- Principled construction
- Generalizes ODE / SDE
- Effective state variables + dynamics

Inputs

Heterogeneous data
- Temporal data (spatial extension possible)
- (quasi-)arbitrary, (non-numerical possible)
- Multiple sources combined properly (RKHS)

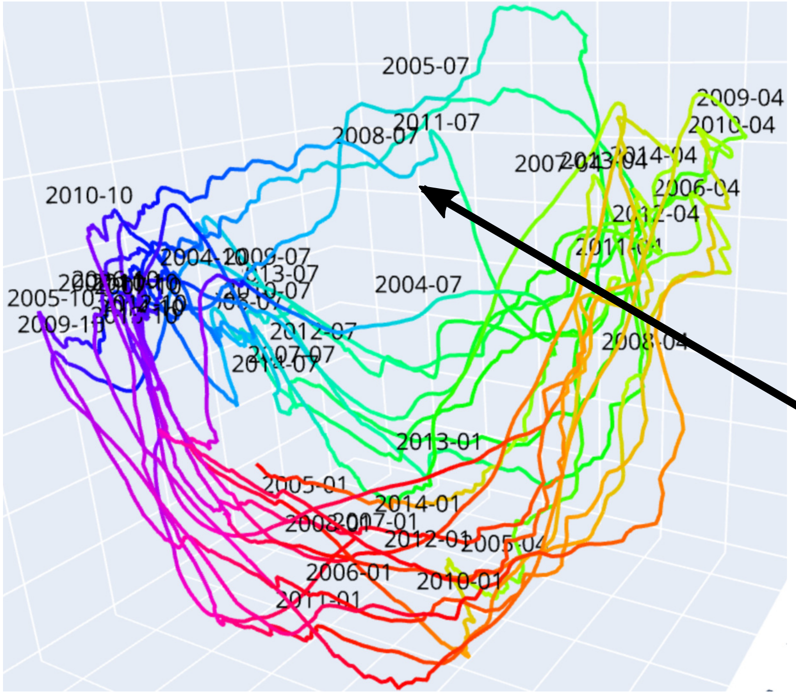
Few parameters, interpretable
- Length of the past / future causal dependencies
- Characteristic scales of each data source
- A few more internal meta-parameters (kernel)

Outputs

Effective state variables
- Condense predictive info. at data scale
- Link with known mechanisms?

Predictive model
- Dynamics in reduced dimensions
- Work in progress…

Application example: Anomaly detection



Inferred dynamics

Anomalies