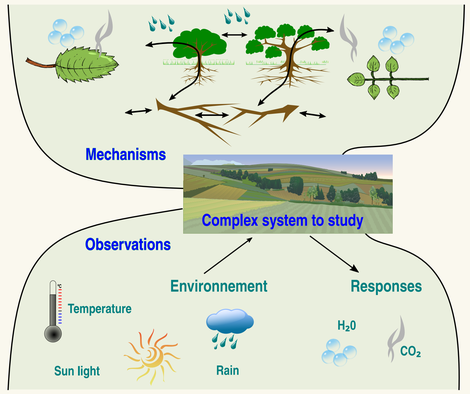
Inferring effective state variables and dynamics from data

Nicolas . Brodu @ Inria . fr

Information theory as a bridge across the Geosciences and Modeling Sciences, workshop at the Schneefernerhaus, Zugspitze, Germany, 11-13 September 2023

1. Modeling complex systems



Complex system — Elements in interaction — Collective behavior → Observations

Ex: ecosystem, geosciences...

Mechanistic models → difficult to aggregate

Difficult to interpret ← Machine learning

Comparison

Statistical Physics — Elements in interaction (Ex: molecules in a gaz) — config. same energy → State variables (Ex: temperature, pressure…) → Global Description (Ex: perfect gaz law)

Proposed approach — config. same pred. info. → State variables (Ex: seasonal cycle) → Predictive model at large scale

Properties

– **No new observation can distinguish two past** sequences in the same causal state
⇒ **equivalent for all modeling purposes**

– **States do not depend on the frame of reference** distribution shapes change but not the classes
⇒ **intrinsic property of the observed process**

Reconstruction of state variables

– **No reason to find back the original frame**
Work in progress for the correspondance

2. Causal States

Temperature Humidity CO_2 Flux …

Ensemble of observations past → future

P(future | past) in distribution

Equivalence classes [Crutchfield 1988] → **Causal state** = {same P(future | past)}

Same causes, same consequences

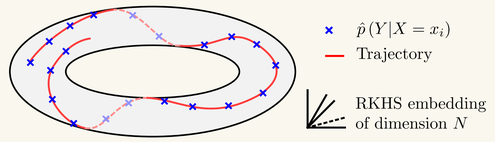
3. State variables

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

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



Choice of coordinates = eigenbasis of Laplace-Beltrami operator Δ [Brodu & Crutchfield 2022]

– Estimated with diffusion maps ⇒ geometry robust to the sampling density

– Eigenbasis ⇔ generalized Fourier modes of the manifold

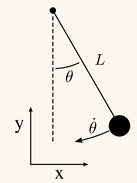
– Each added coordinate best refines the (diffusion) distance between states
⇒ hence better encodes causal states, hence predictive info.

⇒ can be thought of as similar to a non-linear PCA but for predictive info.

3. & 4. ⇒ Interpretation : « as if » macroscopic system driven by hidden state variables + SDE

4. ODE as a special case

Dissipative pendulum

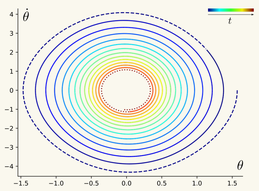


$$\frac{d\theta}{dt} = \dot{\theta}$$
$$\frac{d\dot{\theta}}{dt} = -\frac{b}{m}\dot{\theta} - \frac{g}{L}\sin\theta$$

Each point in phase space is a causal state

– Unique trajectory = future

– No dependency on the past

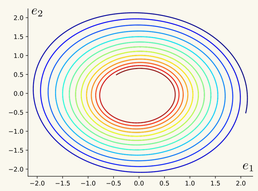


State space: $\theta, \dot{\theta}$

Reconstruction from data

– Positions (x,y) for 1 period

Causal states encoded in a set of coordinates ⇒ state variables



Effective state variables

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

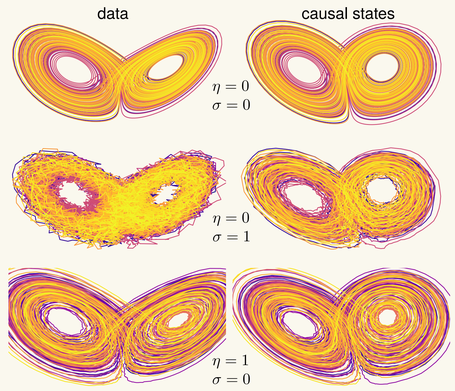
– Simulation of trajectories

– Gaussian noise added to data, var. σ^2

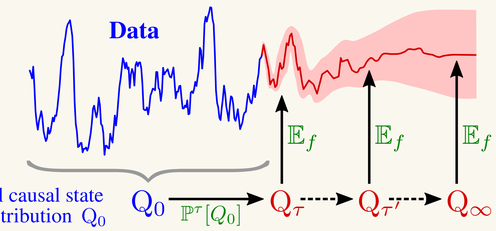
Causal states

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

– Intrinsic noise: SDE details preserved



5. Predictive model



Predictions Individual trajectories

Ensemble expectation converges in the limit to the data mean

$\mathbb{E}_f$ = expectation of a function f

Initial information (latest mesurements) completely lost

$\mathbb{P}$ = Perron-Frobenius operator

SDE ⇔ Diffusion : distribution of states

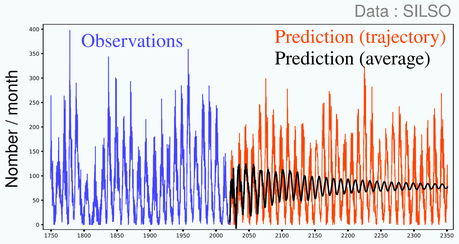
Limit distribution of the SDE

Analoguous to the thermodynamics limit at equilibrium

This model specifies *how* predictive information diffuses through time!

6. Sunspots

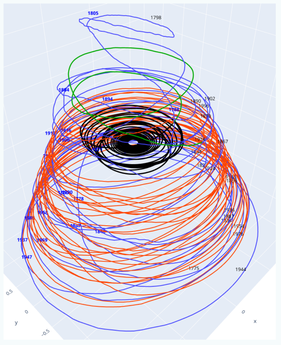
Data : SILSO



Inferred state variables

– 11-years cycle (x,y)

– Amplitude modulations (z)



Structure resembling an attractor embedding

Predictions

– Trajectory constrained on the structure

– **Linear operator converging to the average**

7. El Niño / La Niña (ENSO)

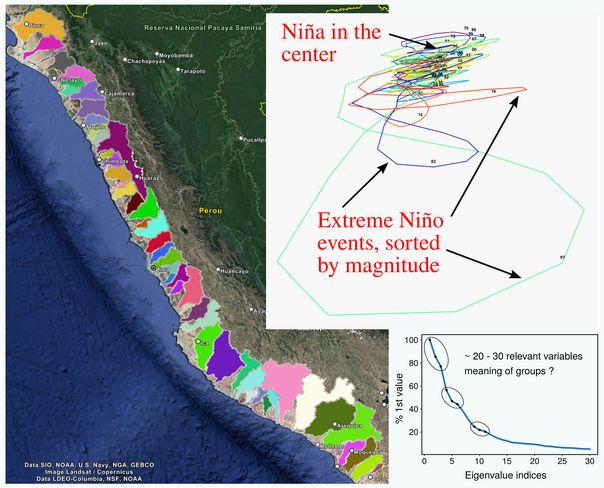
– 50 years of measurements, very high quality

– 49 watersheds, Peruvian coast

– Pacific Ocean : 4 indices of sea surface temperature

– Per watershed: precipitations, runoff, evapotranspiration, temperature

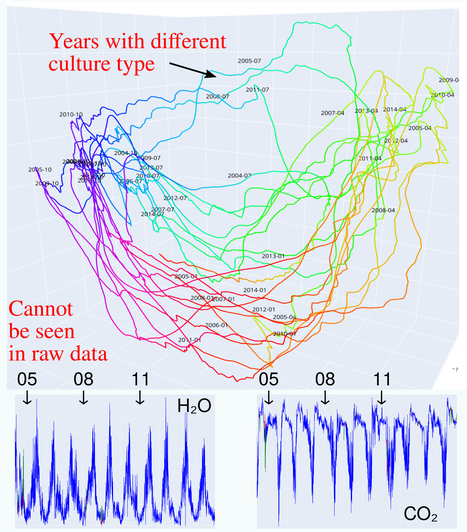
Data: Luc Bourrel, Pedro Rau



8. Grignon field (fr)

– Temperature, soil humidity, sun illumination, evapotranspiration, precipitations

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



9. n-Butane conformations

Data by Stefan Klus

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

– Local frame of reference

Results

– Clusters = molecular conformations (carbon)

– Fast transitions go through eclipsed conformations

– Sub-clusters for hydrogen atom positions (chemically equiv. but distinct in data)

– Shows structure other than cyclic attractor-like examples