Distribution of model parameters and causality

IA pour la Science

January 23th, 2026

Distribution of model parameters and causality

The research focuses on the causal inference for the model selection in machine learning. To select a model we need to estimate the distribution of model parameters.
PyTorch provides a well-developed automated differentiation framework to estimate the parameters of the model. The nearest goal is to propose a method to build Causal Structural Model through the distribution of model parameters and use it for model selection purposes.
Below:

1. Causal machine and Control Problem
2. Clarification of the CI for time series
3. Applications of the Causal Discovery

Causal Inference and Control Problem

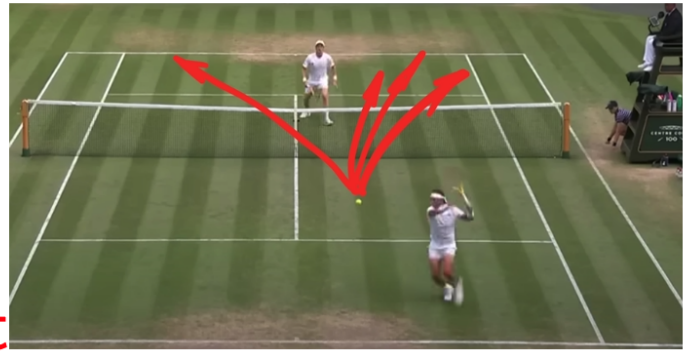
Control

Stimulus

$$
\begin{tikzcd}
 & & z \arrow[d, "\text{Exogenous}"] \\
u = C(\varepsilon) \arrow[rr, "\text{Control}"] & & y = f(x, y, u, z) \arrow[d, "\text{Measurement}"] \\
\hat{f}, \varepsilon \arrow[u, "\text{Decision}"] & & \hat{x}, \hat{y}, \hat{z} \arrow[ll, "\text{Modelling}"']
\end{tikzcd}
$$

Action

Game with opponent



Game with nature

Proportional integral derivative PID controller

$$u(t) = C_{\mathrm{p}}\varepsilon(t) + C_{\mathrm{i}} \int_0^t \varepsilon(\tau)\, d\tau + C_{\mathrm{d}} \frac{d\varepsilon(t)}{dt}$$

Time-delay embedding in Singula Spectrum Analysis

1. Construct the Hankel matrix with time series,

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \boldsymbol{x}_3 \\ \vdots \\ \boldsymbol{x}_t \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ x_2 & x_3 & x_4 & \cdots & x_{n+1} \\ x_3 & x_4 & x_5 & \cdots & x_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_t & x_{t+1} & x_{t+2} & \cdots & x_{t+n} \end{bmatrix}.$$

2. Decompose the matrix and take the k first components,

$$\boldsymbol{X} = \boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n = \boldsymbol{U}\Lambda\boldsymbol{V}^{\mathsf{T}} = \sum_{k=1}^{n} \lambda_k \boldsymbol{u}_k \boldsymbol{v}_k^{\mathsf{T}}.$$

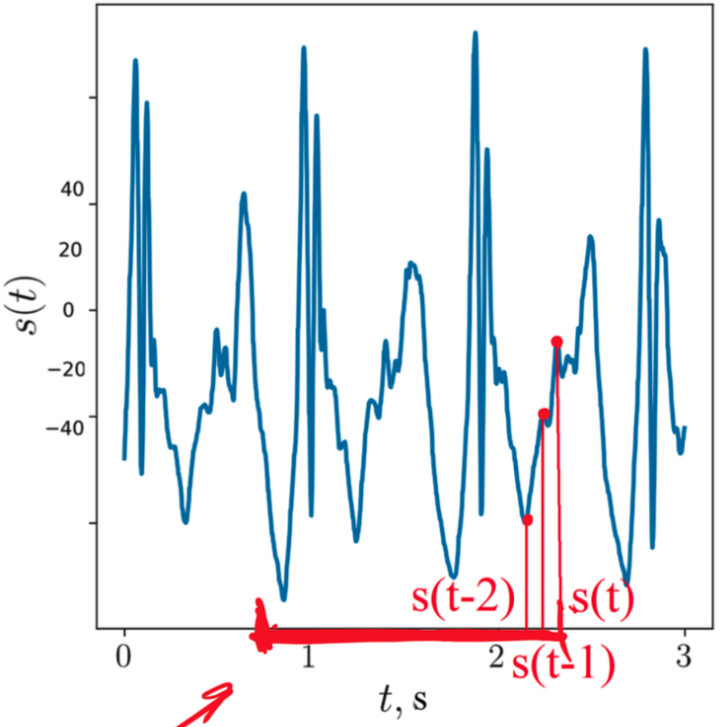
3. Reconstruct the time series by anti-diagonal average

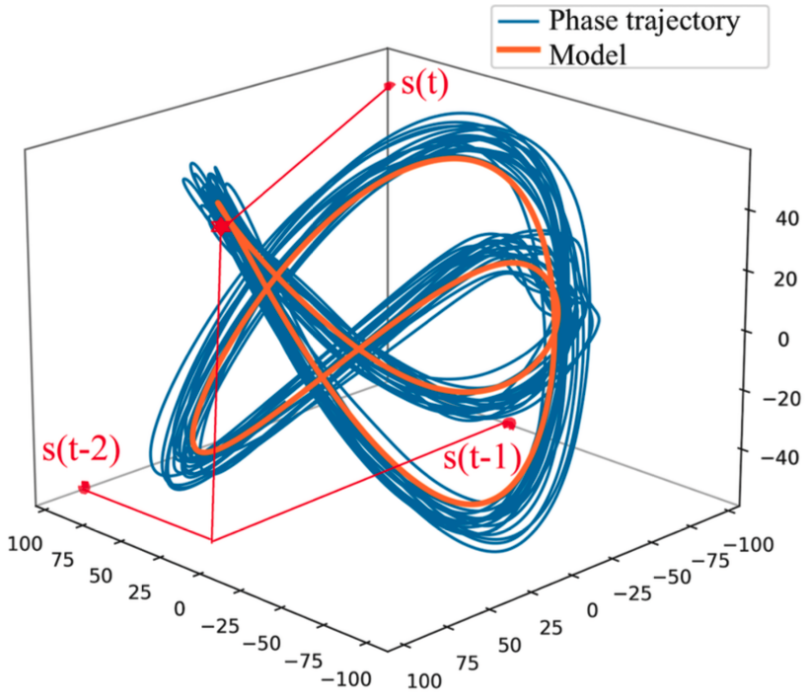
$$\hat{x}_t = \text{mean}\boldsymbol{X}(i,j), \quad i + j = t - 1.$$

State space models describe the change of state $\boldsymbol{x}$ in time $\frac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{z})$

Define the state space by vectors $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t\}$, the phase trajectory $\boldsymbol{X}$.

Model of the phase trajectory in the state space





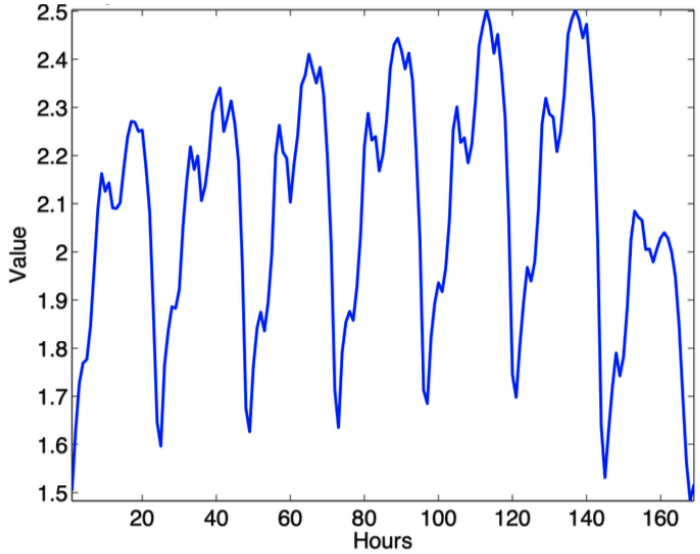
t - 1000

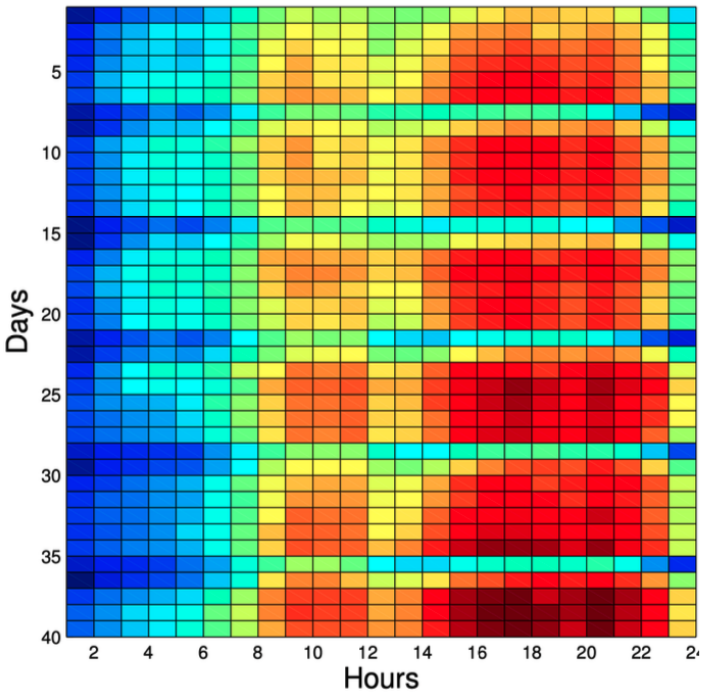
dim(s) $\approx$ 1000 $\longrightarrow$ dim(x) $= 4$

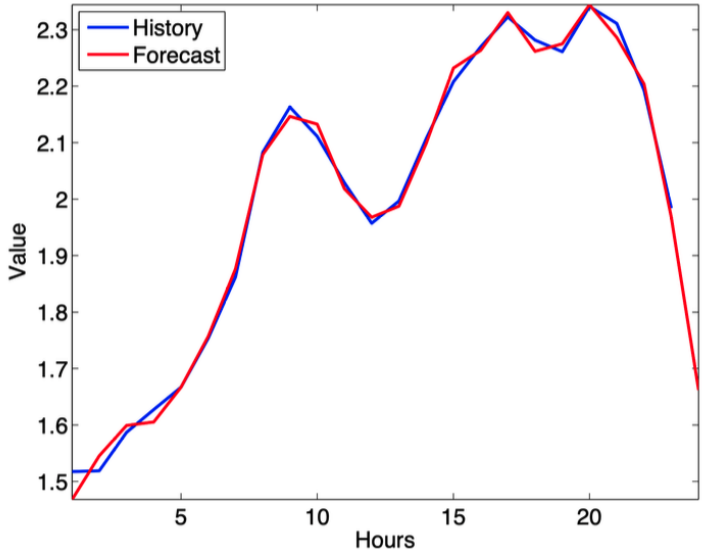
Reduce dimensionality with the principal component analysis, autoencoder $\boldsymbol{z} = \boldsymbol{W}^{\mathsf{T}}\boldsymbol{x}$ where $\boldsymbol{W}$ is an orthogonal (rotation) matrix. The first principal components are given by Singular Values Decomposition

$$\sqrt{\lambda_k}\boldsymbol{V_k} = \boldsymbol{X}^{\mathsf{T}}\boldsymbol{U}_k \quad \text{the SVD is} \quad \boldsymbol{X} = \boldsymbol{U}\Lambda\boldsymbol{V}^{\mathsf{T}}$$

Linear model for time series forecasting







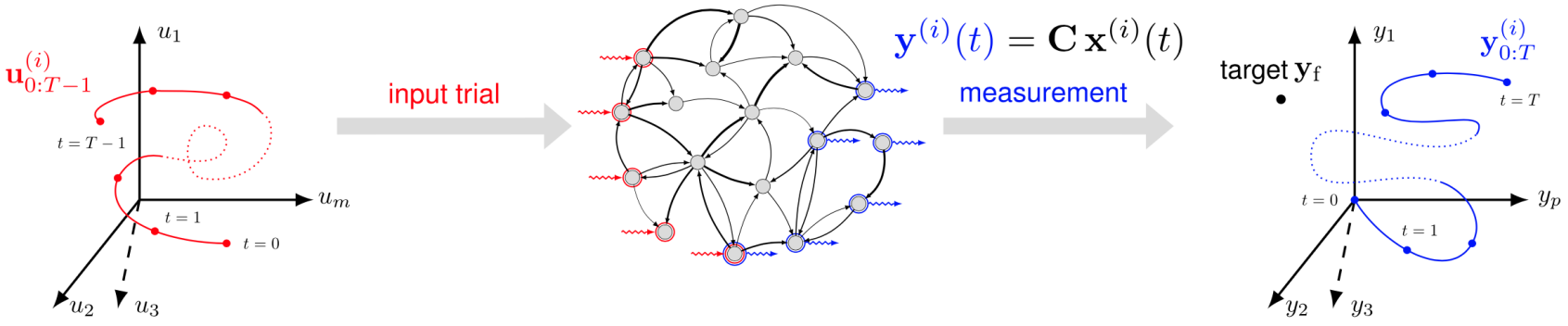
The design matrix is

$$\begin{array}{c|c} \underset{1\times 1}{y_{t+1}} & \underset{1\times n}{\boldsymbol{x}_t} \\ \hline \underset{t\times 1}{\boldsymbol{y}} & \underset{t\times n}{\boldsymbol{X}} \end{array},$$

the forecasting model $\hat{\boldsymbol{y}} = \boldsymbol{w}^\top \boldsymbol{X}$,
the forecast $\hat{y}_{t+1} = \boldsymbol{w}^\top \boldsymbol{x}_t$.

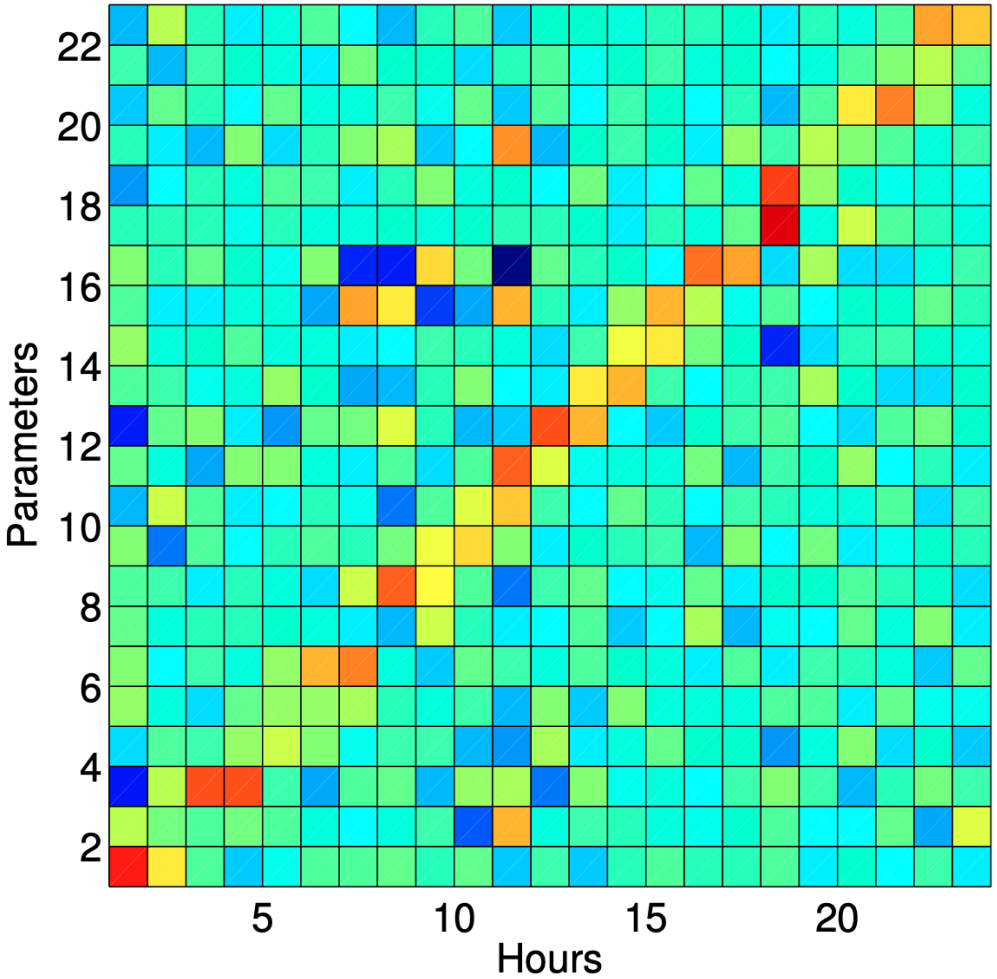
Conrol problem for time series forecasting

$$\mathbf{x}^{(i)}(t+1) = \mathbf{A}\,\mathbf{x}^{(i)}(t) + \mathbf{B}\,\mathbf{u}^{(i)}(t)$$



Covariance matrix of forecasting model parameters

Estimate parameters $\mathrm{w}(\tau) = (\mathrm{X}^{\mathsf{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathsf{T}}\mathrm{y}$, then calculate the sample $s(\tau) = \mathrm{w}^{\mathsf{T}}(\tau)\mathrm{X}_{m+1}$ for each τ of the next $m+1$-th period.



The color shows the value of a parameter for each hour.

Probabilistic model selection

Bayesian inference delivers the error function $S(\mathbf{w})$

Posterior Likelihood Prior

$$p(\mathbf{w}|\mathfrak{D}, \mathbf{A}, \mathbf{B}, \mathbf{f}) = \frac{p(\mathfrak{D}|\mathbf{w}, \mathbf{B}, \mathbf{f})p(\mathbf{w}|\mathbf{A}, \mathbf{f})}{p(\mathfrak{D}|\mathbf{A}, \mathbf{B}, \mathbf{f})}.$$

Evidence
(to select a model)

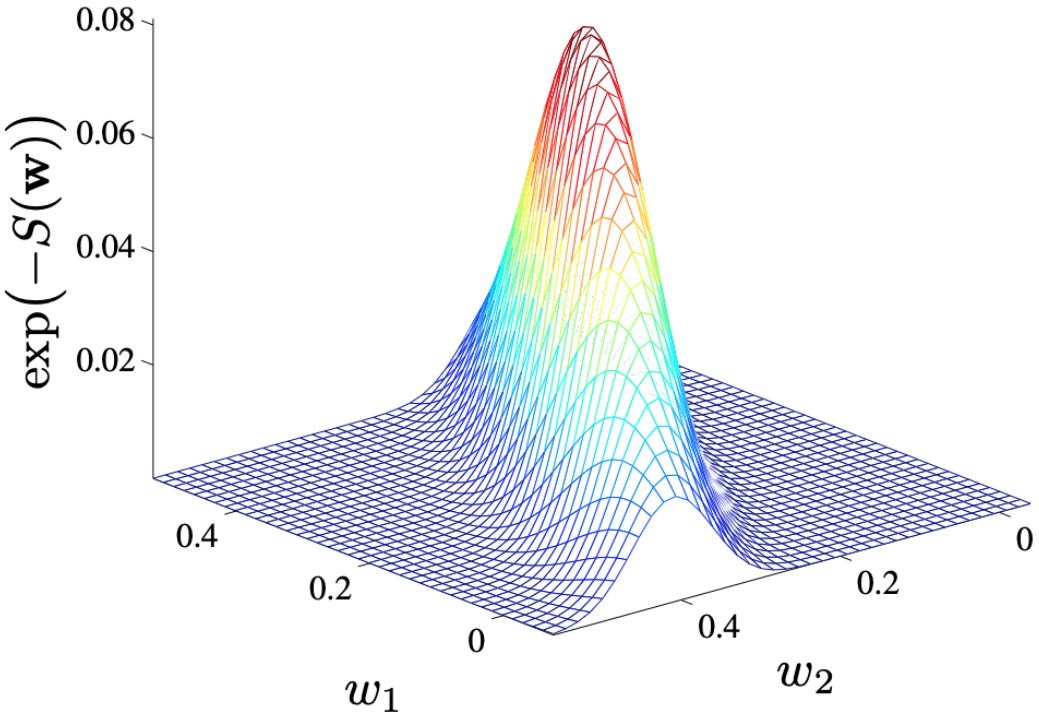
Write the error function given hyperparameters $\mathbf{A}, \mathbf{B}$

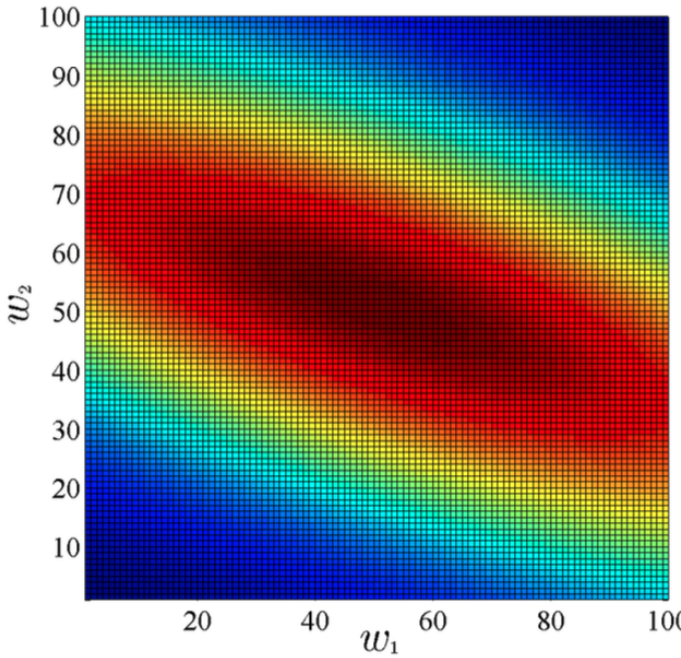
$$S(\mathbf{w}) = \underbrace{\frac{1}{2}(\mathbf{y} - \mathbf{f})^\mathsf{T}\mathbf{B}(\mathbf{y} - \mathbf{f})}_{\text{approximation error}} + \underbrace{\frac{1}{2}(\mathbf{w} - \hat{\mathbf{w}})^\mathsf{T}\mathbf{A}(\mathbf{w} - \hat{\mathbf{w}})}_{\text{regularisation error}},$$

$$S = E_D + E_\mathbf{w} = \boldsymbol{\lambda}^\mathsf{T}\mathbf{s}, \qquad \text{metaparameters } \lambda = \frac{1}{2}.$$

Empirical distribution of model parameters

The value of error function $S(\mathbf{w}|\mathfrak{D}, f)$ depends on parameters.

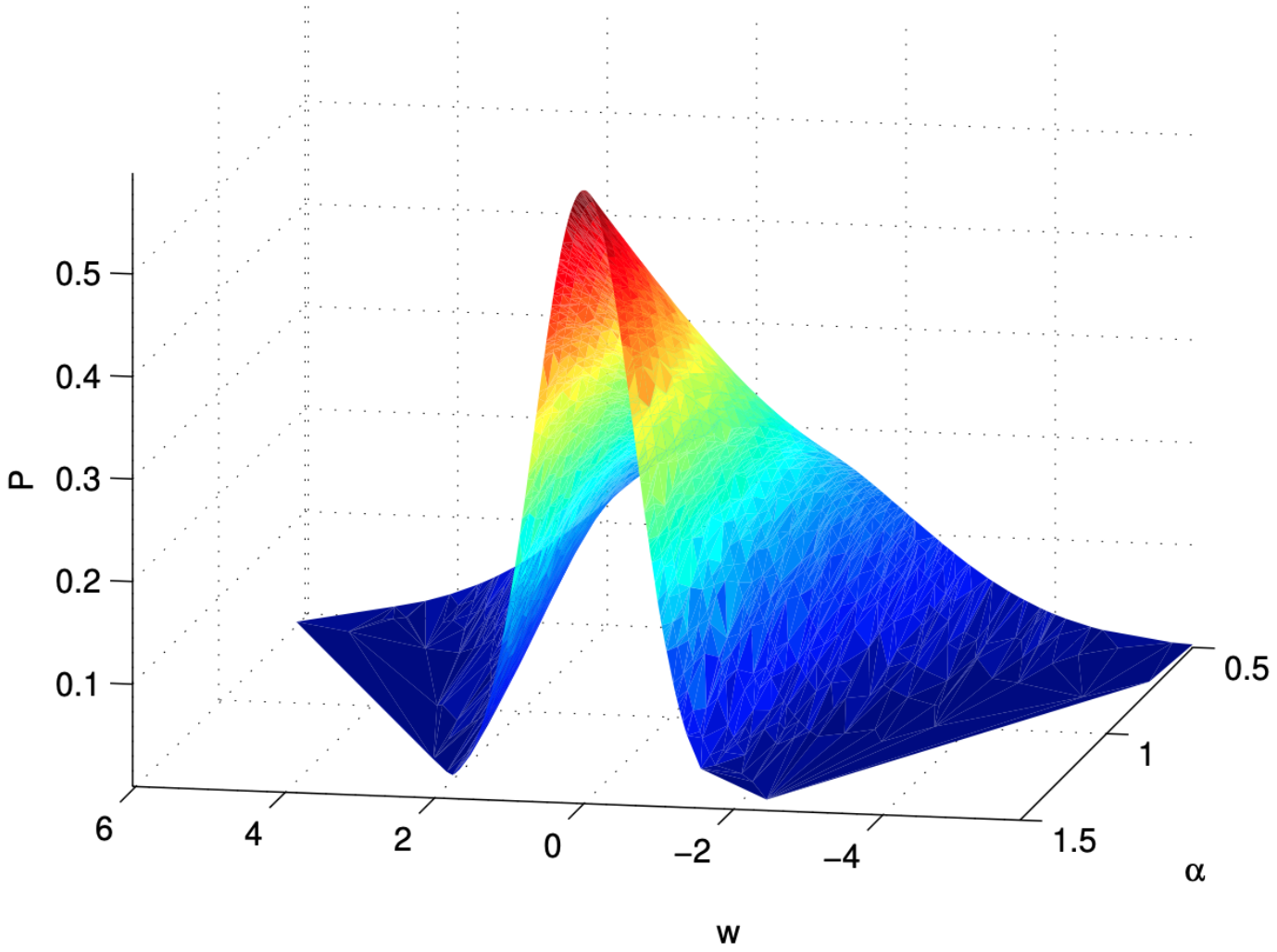




x-axis and y-axis: parameters $\mathbf{w}$, z-axis: $\exp(-S(\mathbf{w}))$

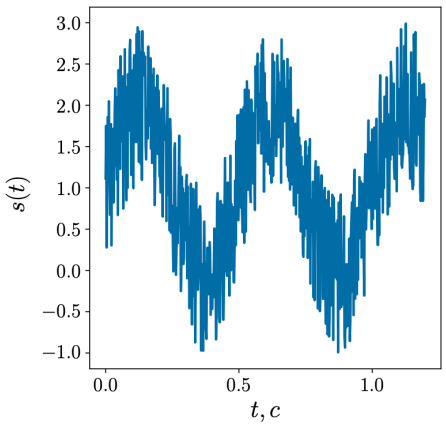
Evidence of the model

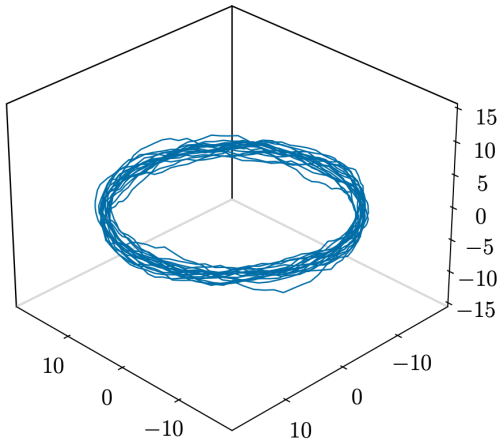
depends on both, error E_D (likelihood) and regularisation $E_{\mathbf{w}}$ (prior).



Parameters $\mathbf{w}$, variance α^{-2}, and $p(\mathbf{w}|\mathfrak{D}, \alpha)$ is the evidence.

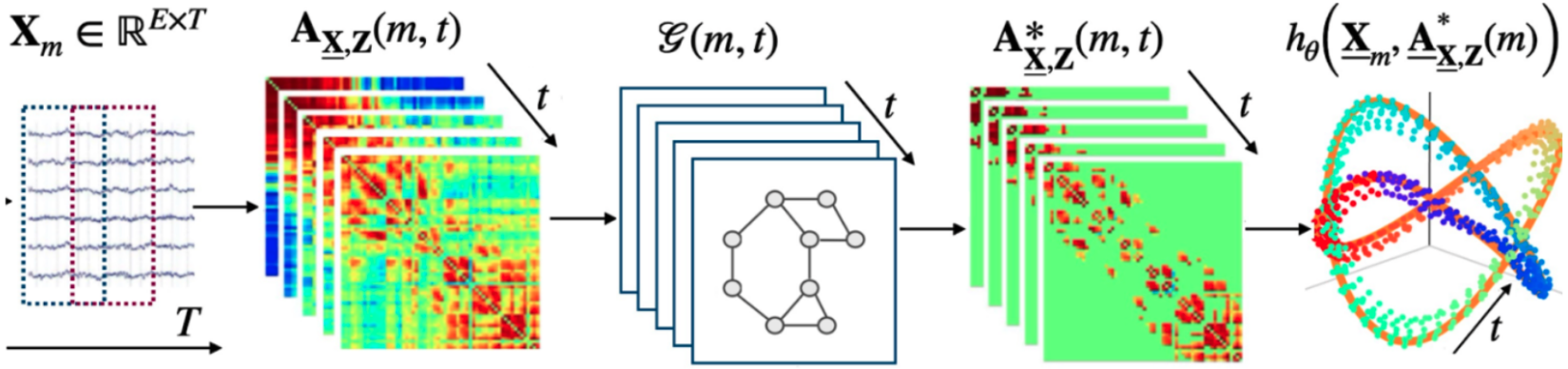
Do-operations and tests





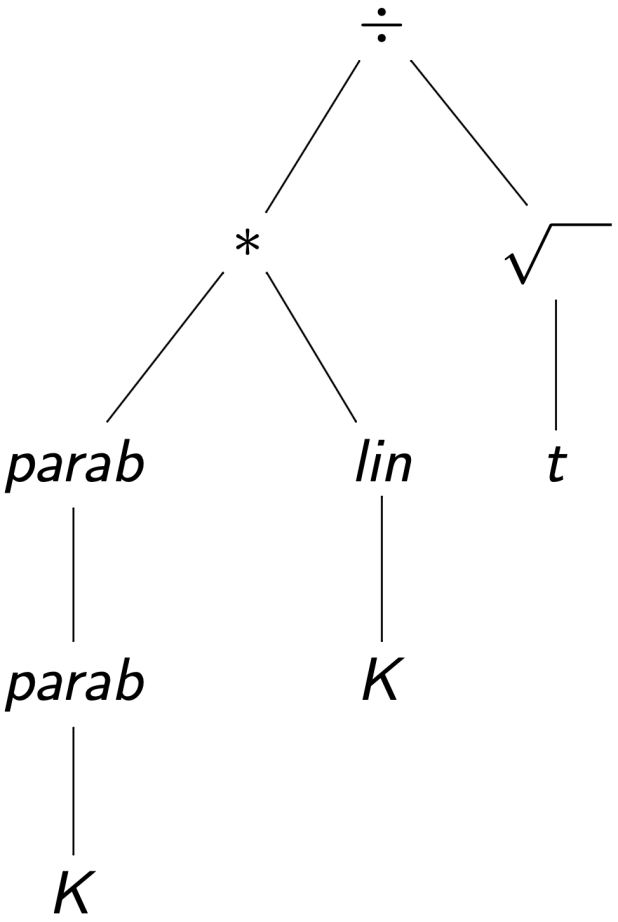
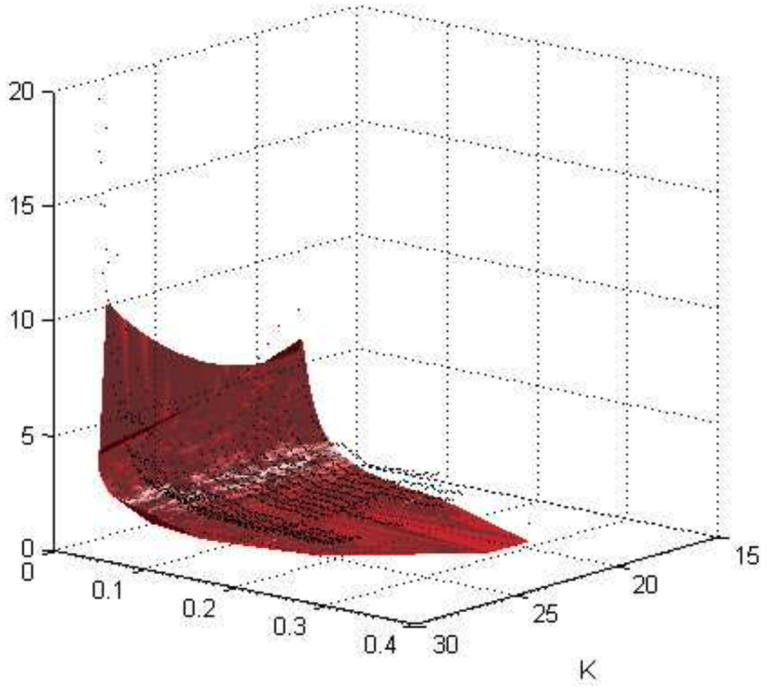
High variance and high co-variance in time series

Dynamic graph reflects dependencies between the time series.



To reconstruct the dependencies

1. Reconstruct phase trajectories through time-delay embedding.
2. Define distance between points of the phase trajectories.
3. Make low-rank decomposition, prune the dependency graph.
4. Reconstruct time series.

Causal discovery for syblolic regression modeling

$$\sigma = \frac{(w_1 K + w_2)(w_1 K^2 + w_2 K + w_3)}{\sqrt{t}}$$