

Data-driven reconstruction of partially observed dynamical systems

Dr. Pierre Tandeo, associate professor

IMT Atlantique, Brest, France

August 23, 2023

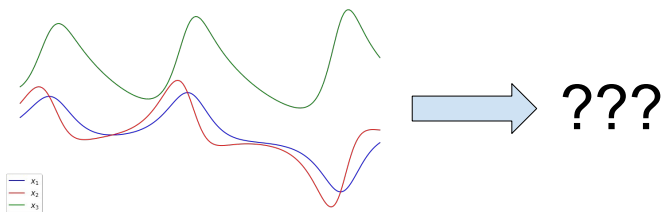


10th International Congress on Industrial and Applied Mathematics

ICIAM 2023 TOKYO

Context presented using the Lorenz-63 system

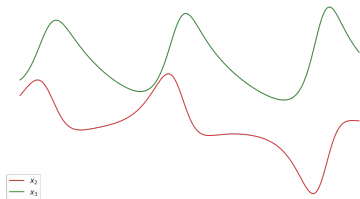
Make predictions of multivariate time series:



Constraints:

- no dynamical model (i.e., equations are not available)
- only partial observations available (e.g., one variable is missing)

~~$$\dot{x}_1 = \sigma(x_2 - x_1)$$
$$\dot{x}_2 = x_1(\rho - x_3) - x_2$$
$$\dot{x}_3 = x_1x_2 - \beta x_3$$~~



Proposed approach and scientific questions

True dynamical model f :

$$\dot{\mathbf{x}} = f(\mathbf{x})$$

with:

- $\mathbf{x} = [x_1, x_2, x_3]$
- f is the Lorenz model

Proposed dynamical model \tilde{f} :

$$\dot{\mathbf{x}} = \tilde{f}(\mathbf{x})$$

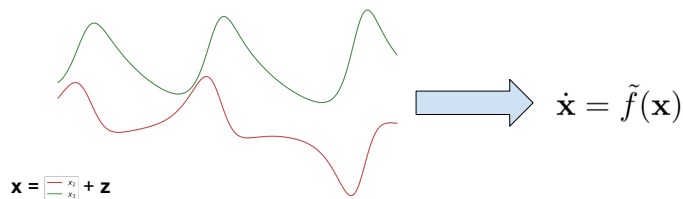
with:

- $\mathbf{x} = [\cancel{x_1}, x_2, x_3, \mathbf{z}]$
- \tilde{f} is a statistical emulator

Questions:

- can we learn \tilde{f} without observing x_1 ?
- are predictions of \tilde{f} accurate and reliable?
- can we learn \mathbf{z} and what is its meaning?

Mathematical formulation



State-space model (with emulated model + augmented state):

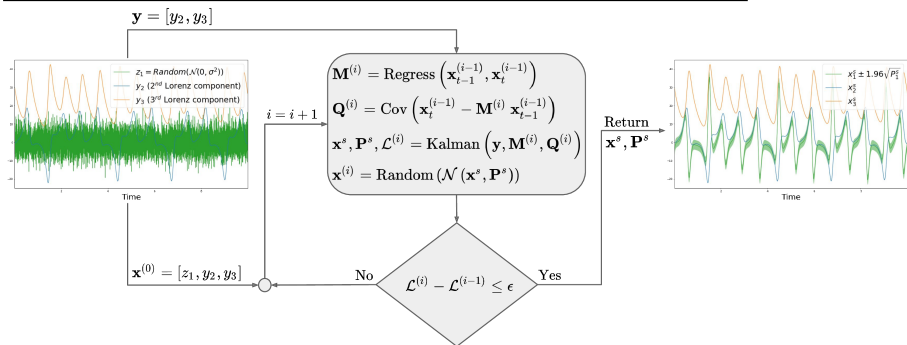
$$\dot{\mathbf{x}} = \tilde{f}(\mathbf{x}) + \boldsymbol{\eta}$$

$$\mathbf{y} = h(\mathbf{x}) + \boldsymbol{\epsilon}$$

- $\mathbf{y} = [y_2, y_3]$ → **partial observations** of the system
- $\mathbf{x} = [x_2, x_3, \mathbf{z}]$ → **augmented state** with latent variables \mathbf{z}
- \tilde{f} → learned using **regression methods**

Proposed algorithm

Schematic representation of the proposed iterative algorithm:



- latent variable z_1 is initialized as a Gaussian white noise
- \tilde{f} is approximated using a constant and linear operator \mathbf{M}
- everything is linear and Gaussian so we use the Kalman smoother
- likelihood \mathcal{L} is the optimization criterion

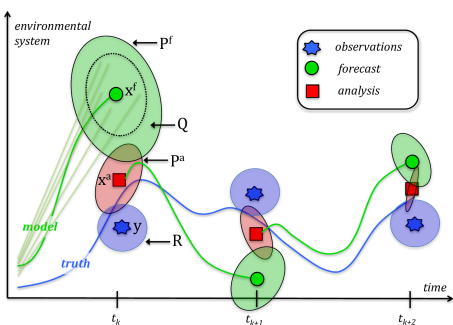
Optimization criterion

Likelihood of the observations:

$$\mathcal{L}(\mathbf{y}) \propto \prod_{t=1}^T \exp\left(-\mathbf{d}(t)^\top \boldsymbol{\Sigma}(t)^{-1} \mathbf{d}(t)\right)$$

with the **innovation** defined by its mean and covariance:

$$\mathbf{d}(t) = \mathbf{y}(t) - \mathbf{H}\mathbf{x}^f(t) \text{ and } \boldsymbol{\Sigma}(t) = \mathbf{H}\mathbf{P}^f(t)\mathbf{H}^\top + \mathbf{R}.$$



Interest of the likelihood criteria:

- EM-like algorithm, thus **likelihood increases** until convergence
- quantitative criteria to select the **optimal number** of latent components

Example of the Lorenz system:

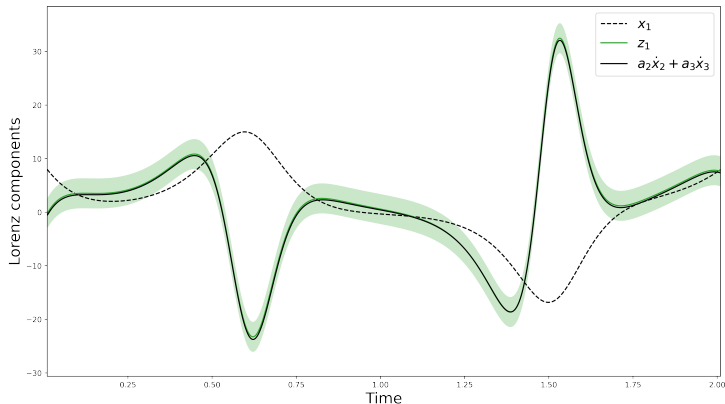
- x_2 and x_3 are observed, but never x_1
- $\mathbf{x} = [x_2, x_3, z_1]$

Meaning of z_1

Meaning of z_1 :

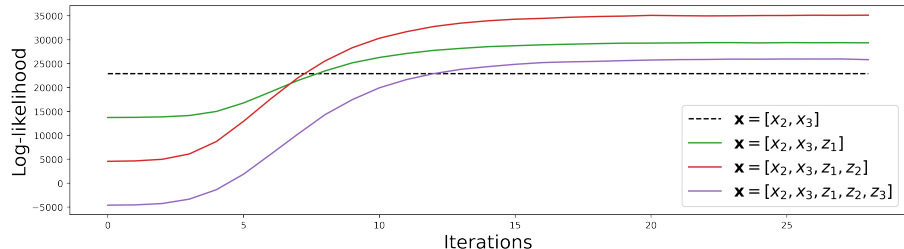
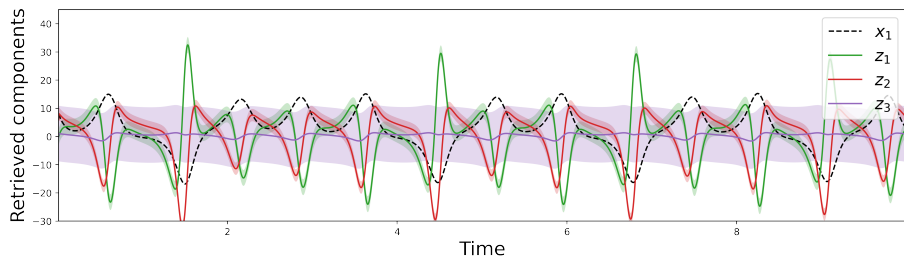
- revealed by **symbolic regression**
- between z_1 and usual transformations of x_2 and x_3

z_1	x_2	x_3	x_2x_3	x_2^2	x_3^2	$\sqrt{x_2}$	$\sqrt{x_3}$...	\dot{x}_2	\dot{x}_3
-------	-------	-------	----------	---------	---------	--------------	--------------	-----	-------------	-------------



Estimation of z_2 and z_3

Estimation of z_2 and z_3 :



$\Rightarrow z_3$ is **useless** (almost constant and lower likelihood)

Meaning of z_1 and z_2

Meaning of z_1 and z_2 :

$$z_1 = a_2 \dot{x}_2 + a_3 \dot{x}_3$$

$$z_2 = b_1 \dot{z}_1 + b_2 \dot{x}_2 + b_3 \dot{x}_3$$

$$= b_2 \dot{x}_2 + b_3 \dot{x}_3 + b_1 a_2 \ddot{x}_2 + b_1 a_3 \ddot{x}_3$$

⇒ z_1 and z_2 take into account **time delays** of x_2 and x_3

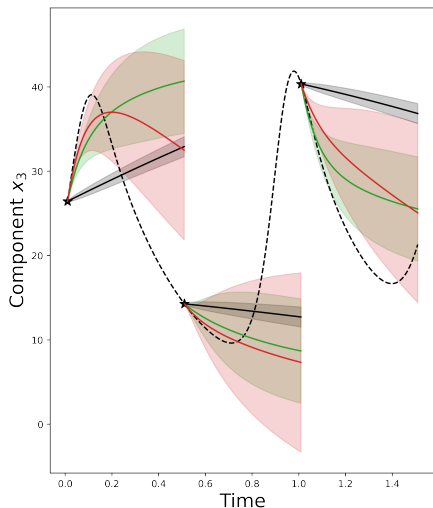
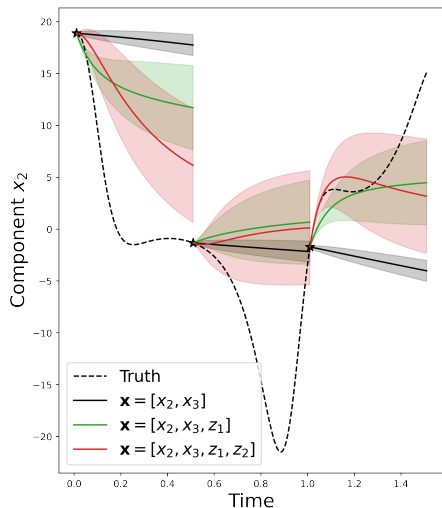
⇒ strong connections with **Takens' and Taylor's theorems**

Linear surrogate model of the Lorenz system:

- f is approximated using a **linear regression**
- it corresponds to a constant 4×4 matrix **M**
- it can be used to **make predictions** of x_2 and x_3

Example of statistical forecasts using linear operators

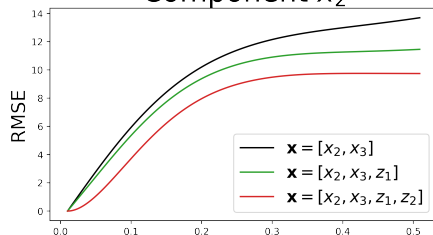
Example of statistical forecasts using linear operators:



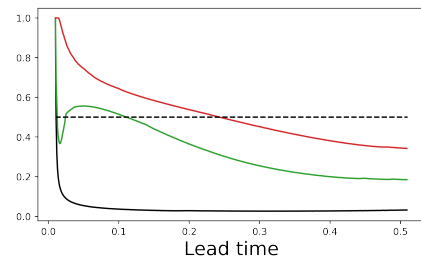
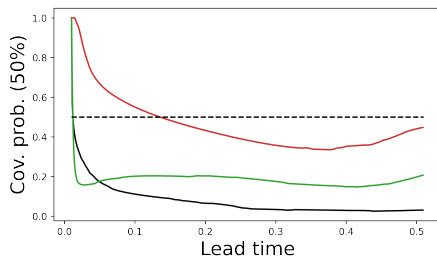
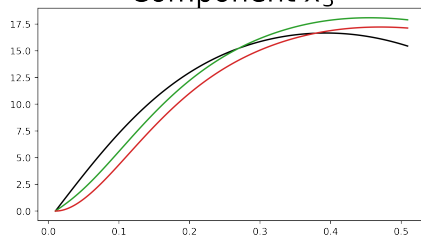
Performance metrics of the linear operators

Performance metrics of the linear operators:

Component x_2



Component x_3



Conclusions and perspectives

Context:

- data-driven predictions (i.e., without physical equations)
- missing components of the system (e.g., x_1 of Lorenz)

Conclusions:

- combination of data assimilation and machine learning
- inference of latent variables, interesting physical meaning

Perspectives:

- apply the methodology on climate indices
- learn time dependent and linear dynamical operators $\mathbf{M}(t)$

ありがとうございます (arigatō gozaimasu)!

Tandeo et al. 2023 [NPG]:

Nonlin. Processes Geophys., 30, 129–137, 2023
<https://doi.org/10.5194/npg-30-129-2023>
© Author(s) 2023. This work is distributed under
the Creative Commons Attribution 4.0 License.



Nonlinear Processes
in Geophysics



Data-driven reconstruction of partially observed dynamical systems

Pierre Tandeo^{1,2,3}, Pierre Ailliot⁴, and Florian Sévellec^{5,2}

Contact:

- pierre.tandeo@imt-atlantique.fr
- www.tandeo.wordpress.com