



**HAL**  
open science

# Data-driven identification and reconstruction of partially observed dynamical systems

Pierre Tandeo

► **To cite this version:**

Pierre Tandeo. Data-driven identification and reconstruction of partially observed dynamical systems. ISDA 2024 - 10th International Symposium on Data Assimilation, RIKEN, Oct 2024, Kobe, Japan. hal-04764759

**HAL Id: hal-04764759**

<https://imt-atlantique.hal.science/hal-04764759v1>

Submitted on 8 Nov 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Data-driven identification and reconstruction of partially observed dynamical systems

Dr. Pierre Tandeo, associate professor

IMT Atlantique, Brest, France  
RIKEN Cluster for Pioneering Research, Kobe, Japan

The 10th International Symposium on Data Assimilation (ISDA)  
October 21-25, 2024, Kobe, Japan

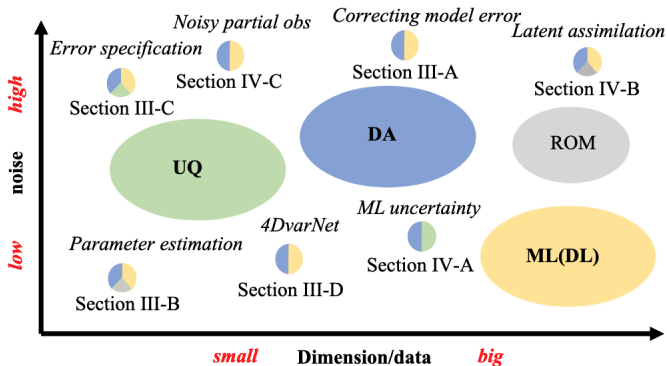




# DA + ML + UQ: a quick overview<sup>1</sup>

## Context:

- rapidly growing field of research
- still lot of developments to do



<sup>1</sup>Cheng et al., *Machine Learning with Data Assimilation and Uncertainty Quantification for Dynamical Systems: a Review*, IEEE/CAA Journal of Automatica Sinica, 2023

## Data assimilation formulation:

$$\mathbf{x}(t + dt) = \mathcal{M}(\mathbf{x}(t), \boldsymbol{\eta}(t))$$

$$\mathbf{y}(t) = \mathcal{H}(\mathbf{x}(t), \boldsymbol{\epsilon}(t))$$

## Machine learning appears everywhere:

- state  $\mathbf{x}$  and observations  $\mathbf{y}$
- operators  $\mathcal{M}$  and  $\mathcal{H}$
- errors  $\boldsymbol{\eta}$  and  $\boldsymbol{\epsilon}$
- inversion of the system  $p(\mathbf{x}|\mathbf{y})$

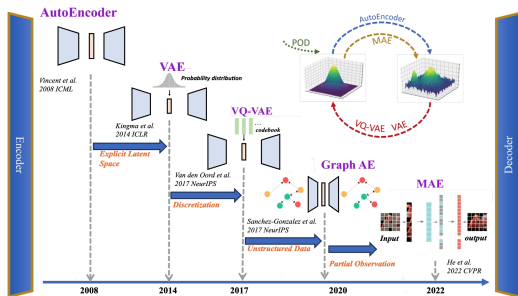
# DA + ML + UQ: a quick overview

$$\mathbf{x}(t + dt) = \mathcal{M}(\mathbf{x}(t), \boldsymbol{\eta}(t))$$

$$\mathbf{y}(t) = \mathcal{H}(\mathbf{x}(t), \boldsymbol{\epsilon}(t))$$

State of the system ( $\mathbf{x}$ ):

- reduced-order modeling (principal components, VAE)
- state augmentation (parameters, latent variables)



Source: Cheng et al. 2023, IEEE/CAA Journal of Automatica Sinica

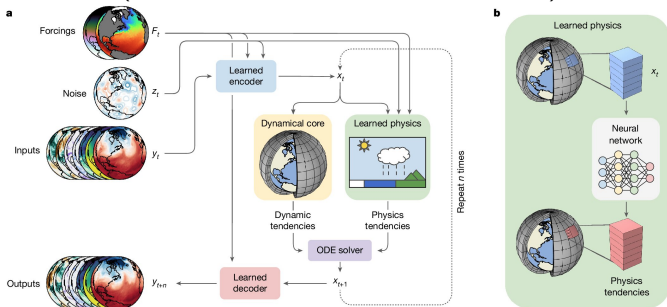
# DA + ML + UQ: a quick overview

$$\mathbf{x}(t + dt) = \mathcal{M}(\mathbf{x}(t), \boldsymbol{\eta}(t))$$

$$\mathbf{y}(t) = \mathcal{H}(\mathbf{x}(t), \boldsymbol{\epsilon}(t))$$

Dynamical operator ( $\mathcal{M}$ ):

- surrogate models (AI, RNN, analogs, AR1 processes)
- hybrid models (GCM core + ML parameterization)



Source: Kochkov et al. 2024, Nature

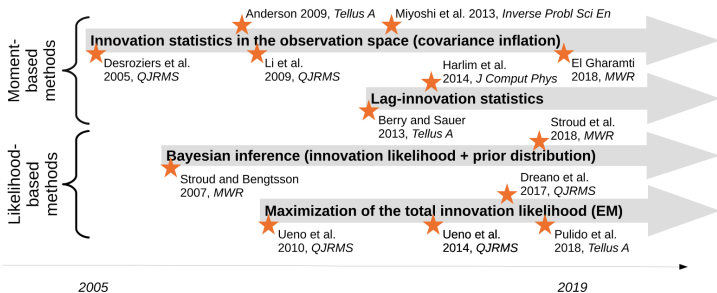
# DA + ML + UQ: a quick overview

$$\mathbf{x}(t + dt) = \mathcal{M}(\mathbf{x}(t), \boldsymbol{\eta}(t))$$

$$\mathbf{y}(t) = \mathcal{H}(\mathbf{x}(t), \boldsymbol{\epsilon}(t))$$

Model and observation errors ( $\boldsymbol{\eta}$  and  $\boldsymbol{\epsilon}$ ):

- innovation statistics, inflation methods, statistical approaches, NN
- joint estimation of error covariances  $\mathbf{Q}$  and  $\mathbf{R}$



Source: Tandeo et al. 2020, Monthly Weather Review

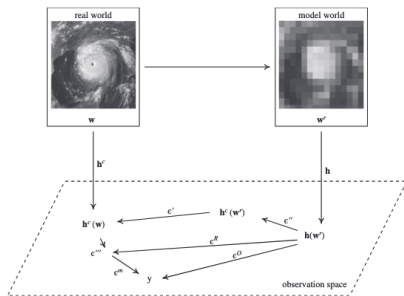
# DA + ML + UQ: a quick overview

$$\mathbf{x}(t + dt) = \mathcal{M}(\mathbf{x}(t), \boldsymbol{\eta}(t))$$

$$\mathbf{y}(t) = \mathcal{H}(\mathbf{x}(t), \boldsymbol{\epsilon}(t))$$

Observation operator ( $\mathcal{H}$ ):

- emulation of transfer functions using ML (e.g., radiative)
- estimation of representativity error



Source: Janjić et al. 2018, QJRMS

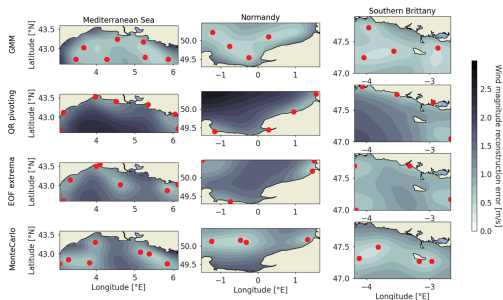
# DA + ML + UQ: a quick overview

$$\mathbf{x}(t + dt) = \mathcal{M}(\mathbf{x}(t), \boldsymbol{\eta}(t))$$

$$\mathbf{y}(t) = \mathcal{H}(\mathbf{x}(t), \boldsymbol{\epsilon}(t))$$

Noisy and partial observations ( $\mathbf{y}$ ):

- aggregation, fusion, and data exclusion
- optimal sampling of the observations



Source: Marcille et al. 2023, Wind Energy Science

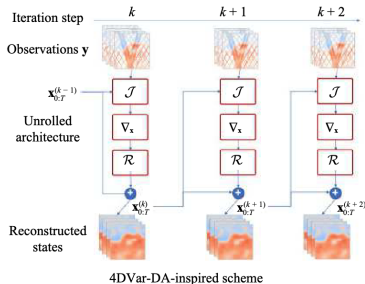
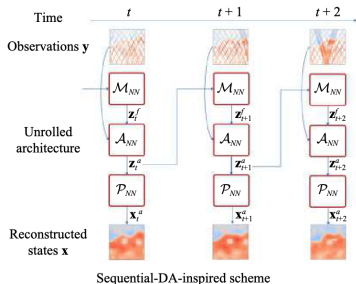
# DA + ML + UQ: a quick overview

$$\mathbf{x}(t + dt) = \mathcal{M}(\mathbf{x}(t), \boldsymbol{\eta}(t))$$

$$\mathbf{y}(t) = \mathcal{H}(\mathbf{x}(t), \boldsymbol{\epsilon}(t))$$

Full system resolution  $p(\mathbf{x}|\mathbf{y})$ :

- optimization problem (link between 4D-Var and NN)
- probabilistic ML using Bayesian approaches



Source: Cheng et al. 2023, IEEE/CAA Journal of Automatica Sinica



# A simple combination of DA + ML + UQ<sup>1</sup>

## General framework:

$$\begin{aligned}\mathbf{x}(t + dt) &= \mathcal{M}(\mathbf{x}(t), \boldsymbol{\eta}(t)) \\ \mathbf{y}(t) &= \mathcal{H}(\mathbf{x}(t), \boldsymbol{\epsilon}(t))\end{aligned}$$

## Proposed framework:

$$\begin{aligned}\mathbf{x}(t + 1) &= \mathbf{M}\mathbf{x}(t) + \mathbf{Q}^{1/2}\boldsymbol{\eta}(t) \\ \mathbf{y}(t) &= \mathbf{H}\mathbf{x}(t) + \mathbf{R}^{1/2}\boldsymbol{\epsilon}(t)\end{aligned}$$

## Assumptions:

- **linear** model and observation operators
- **Gaussian** state, observations, and error terms
- **constant** operators and error covariances

---

<sup>1</sup>Tandeo et al., *Data-driven Reconstruction of Partially Observed Dynamical Systems*, Nonlinear Processes in Geophysics, 2023

# A simple combination of DA + ML + UQ<sup>1</sup>

## General framework:

$$\begin{aligned}\mathbf{x}(t + dt) &= \mathcal{M}(\mathbf{x}(t), \boldsymbol{\eta}(t)) \\ \mathbf{y}(t) &= \mathcal{H}(\mathbf{x}(t), \boldsymbol{\epsilon}(t))\end{aligned}$$

## Proposed framework:

$$\begin{aligned}\mathbf{x}(t + 1) &= \mathbf{M}\mathbf{x}(t) + \mathbf{Q}^{1/2}\boldsymbol{\eta}(t) \\ \mathbf{y}(t) &= \mathbf{H}\mathbf{x}(t) + \mathbf{R}^{1/2}\boldsymbol{\epsilon}(t)\end{aligned}$$

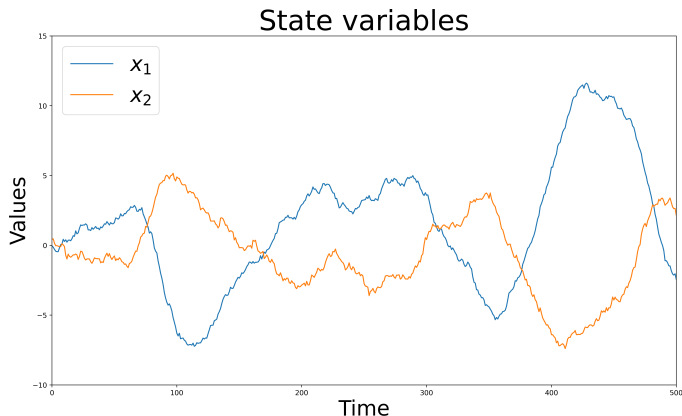
## Goal:

- make **probabilistic forecasts**
- purely **data-driven** approach
- **interpretation** of the retrieved model

---

<sup>1</sup>Tandeo et al., *Data-driven Reconstruction of Partially Observed Dynamical Systems*, Nonlinear Processes in Geophysics, 2023

# Classic approach: Kalman + EM



State-space model:

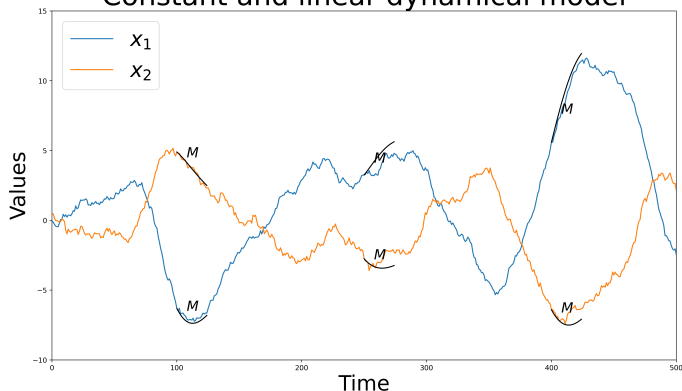
Estimate  $\mathbf{x}(0), \dots, \mathbf{x}(T)$

Variables and matrices:

$\mathbf{x} = [x_1, x_2]$

# Classic approach: Kalman + EM

## Constant and linear dynamical model

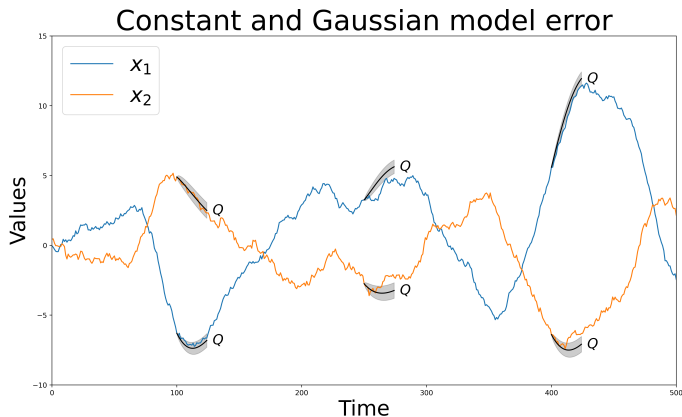


State-space model:  
 $\mathbf{x}(t+1) = \mathbf{M}\mathbf{x}(t)$

Variables and matrices:

$$\mathbf{M} = \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$$

# Classic approach: Kalman + EM



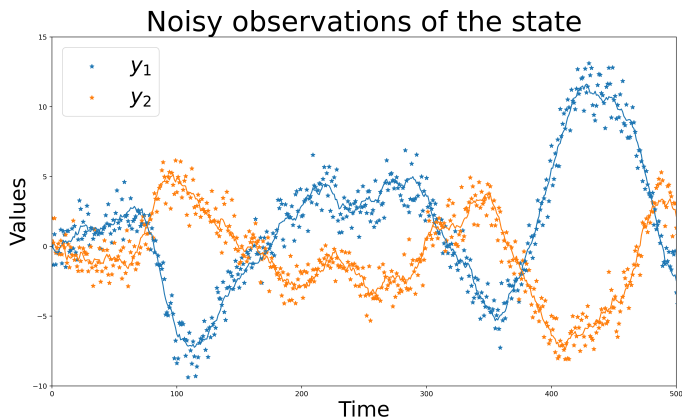
State-space model:

$$\mathbf{x}(t+1) = \mathbf{M}\mathbf{x}(t) + \mathbf{Q}^{1/2}\boldsymbol{\eta}(t)$$

Variables and matrices:

$$\boldsymbol{\eta}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad \mathbf{Q} = \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$$

# Classic approach: Kalman + EM



State-space model:

$$\mathbf{x}(t+1) = \mathbf{M}\mathbf{x}(t) + \mathbf{Q}^{1/2}\boldsymbol{\eta}(t)$$

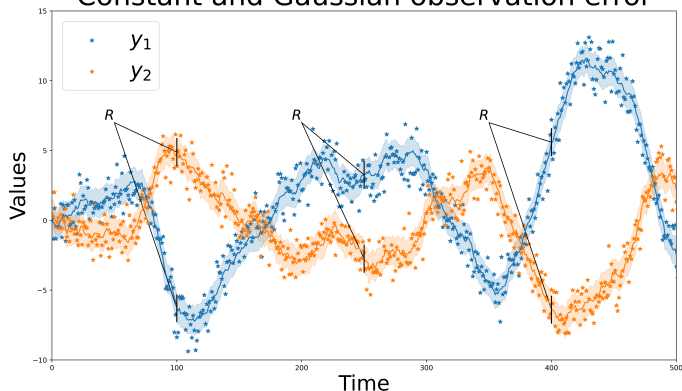
$$\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t)$$

Variables and matrices:

$$\mathbf{y} = [y_1, y_2], \quad \mathbf{H} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

# Classic approach: Kalman + EM

## Constant and Gaussian observation error



State-space model:

$$\mathbf{x}(t+1) = \mathbf{M}\mathbf{x}(t) + \mathbf{Q}^{1/2}\boldsymbol{\eta}(t)$$

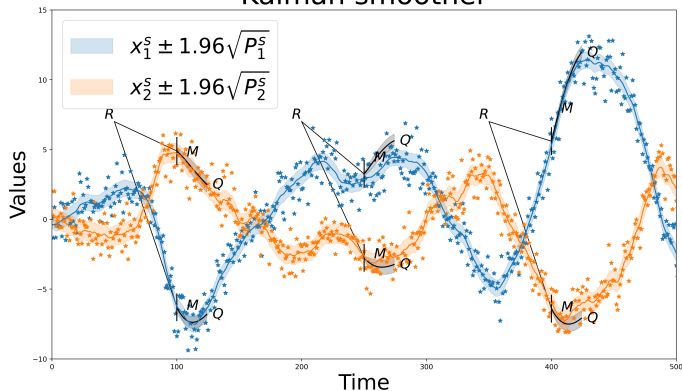
$$\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{R}^{1/2}\boldsymbol{\epsilon}(t)$$

Variables and matrices:

$$\boldsymbol{\epsilon}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad \mathbf{R} = \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$$

# Classic approach: Kalman + EM

## Kalman smoother



State-space model:

$$\mathbf{x}(t+1) = \mathbf{M}\mathbf{x}(t) + \mathbf{Q}^{1/2}\boldsymbol{\eta}(t)$$

$$\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{R}^{1/2}\boldsymbol{\epsilon}(t)$$

⇒ Kalman smoother

Variables and matrices:

$$\hat{\mathbf{M}}, \hat{\mathbf{Q}}, \hat{\mathbf{R}} = \underset{\mathbf{M}, \mathbf{Q}, \mathbf{R}}{\operatorname{argmax}} \mathcal{L}(\mathbf{y})$$

⇒ EM algorithm



# The Expectation-Maximization algorithm<sup>1</sup>

Initialization: **M, Q, R**

---

<sup>1</sup>Shumway and Stoffer, *An approach to time series smoothing and forecasting using the EM algorithm*, Journal of Time Series Analysis, 1982

# The Expectation-Maximization algorithm<sup>1</sup>

Initialization:  $\mathbf{M}$ ,  $\mathbf{Q}$ ,  $\mathbf{R}$

Expectation (DA step):

- fix  $\mathbf{M}$ ,  $\mathbf{Q}$ ,  $\mathbf{R}$
- estimation of  $\mathbf{x}_{0:T}^s$  and  $\mathbf{P}_{0:T}^s$
- $\Rightarrow$  Kalman smoother

---

<sup>1</sup>Shumway and Stoffer, *An approach to time series smoothing and forecasting using the EM algorithm*, Journal of Time Series Analysis, 1982

# The Expectation-Maximization algorithm<sup>1</sup>

Initialization:  $\mathbf{M}$ ,  $\mathbf{Q}$ ,  $\mathbf{R}$

Expectation (DA step):

- fix  $\mathbf{M}$ ,  $\mathbf{Q}$ ,  $\mathbf{R}$
- estimation of  $\mathbf{x}_{0:T}^s$  and  $\mathbf{P}_{0:T}^s$
- $\Rightarrow$  Kalman smoother

Maximization (ML step):

- fix  $\mathbf{x}_{0:T}^s$  and  $\mathbf{P}_{0:T}^s$
- estimation of  $\mathbf{M}$ ,  $\mathbf{Q}$ ,  $\mathbf{R}$
- $\Rightarrow$  analytic solution

---

<sup>1</sup>Shumway and Stoffer, *An approach to time series smoothing and forecasting using the EM algorithm*, Journal of Time Series Analysis, 1982

# The Expectation-Maximization algorithm<sup>1</sup>

Initialization:  $\mathbf{M}$ ,  $\mathbf{Q}$ ,  $\mathbf{R}$

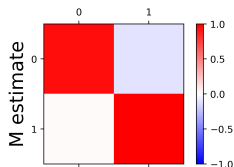
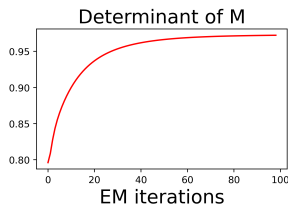
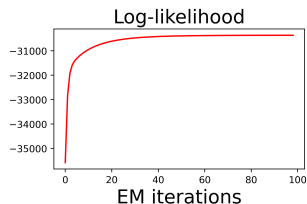
Expectation (DA step):

- fix  $\mathbf{M}$ ,  $\mathbf{Q}$ ,  $\mathbf{R}$
- estimation of  $\mathbf{x}_{0:T}^s$  and  $\mathbf{P}_{0:T}^s$
- $\Rightarrow$  Kalman smoother

Maximization (ML step):

- fix  $\mathbf{x}_{0:T}^s$  and  $\mathbf{P}_{0:T}^s$
- estimation of  $\mathbf{M}$ ,  $\mathbf{Q}$ ,  $\mathbf{R}$
- $\Rightarrow$  analytic solution

Repeat E- and M-steps until convergence of  $\mathcal{L}(\mathbf{y})$



<sup>1</sup>Shumway and Stoffer, *An approach to time series smoothing and forecasting using the EM algorithm*, Journal of Time Series Analysis, 1982

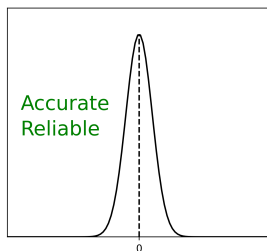
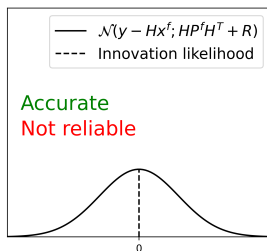
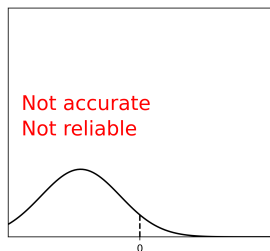
## The innovation likelihood $\mathcal{L}(\mathbf{y})$

$$\mathcal{L}(\mathbf{y}) = \prod_{t=0}^T \mathcal{N}\left(0; \mathbf{y}(t) - \mathbf{H}\mathbf{x}^f(t), \mathbf{H}\mathbf{P}^f(t)\mathbf{H}^\top + \mathbf{R}\right)$$

# The innovation likelihood $\mathcal{L}(\mathbf{y})$

$$\mathcal{L}(\mathbf{y}) = \prod_{t=0}^T \mathcal{N}\left(0; \mathbf{y}(t) - \mathbf{H}\mathbf{x}^f(t), \mathbf{H}\mathbf{P}^f(t)\mathbf{H}^\top + \mathbf{R}\right)$$

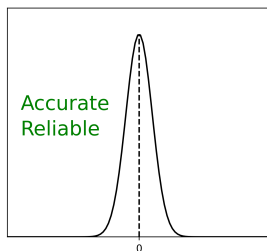
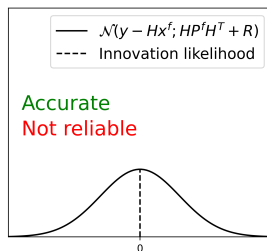
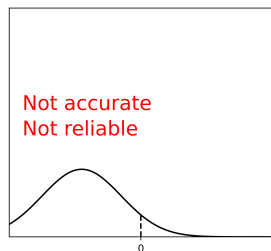
Graphical representation of the innovation likelihood:



# The innovation likelihood $\mathcal{L}(\mathbf{y})$

$$\mathcal{L}(\mathbf{y}) = \prod_{t=0}^T \mathcal{N}\left(0; \mathbf{y}(t) - \mathbf{H}\mathbf{x}^f(t), \mathbf{H}\mathbf{P}^f(t)\mathbf{H}^\top + \mathbf{R}\right)$$

Graphical representation of the innovation likelihood:

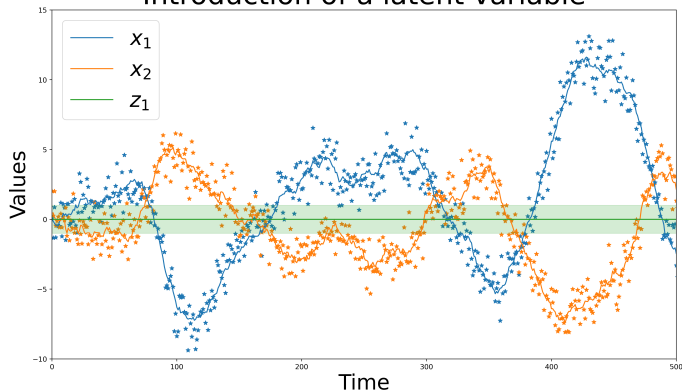


$\mathcal{L}(\mathbf{y})$  depends on  $\mathbf{M}$ ,  $\mathbf{Q}$ ,  $\mathbf{R}$ :

- $\mathbf{x}^f(t) = \mathbf{M}\mathbf{x}^a(t-1) \rightarrow$  forecast precision
- $\mathbf{P}^f(t) = \mathbf{M}\mathbf{P}^a(t-1)\mathbf{M}^\top + \mathbf{Q} \rightarrow$  forecast reliability
- $\mathbf{R} \rightarrow$  observation error

# Proposed approach: Kalman + EM + augmented state

## Introduction of a latent variable



State-space model:

$$\mathbf{x}(t+1) = \mathbf{M}\mathbf{x}(t) + \mathbf{Q}^{1/2}\boldsymbol{\eta}(t)$$

$$\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{R}^{1/2}\boldsymbol{\epsilon}(t)$$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Variables and matrices:

Augmented state  $\mathbf{x} = [x_1, x_2, z_1]$

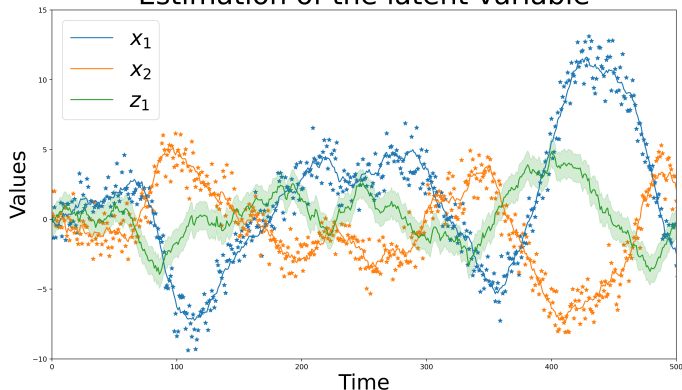
$\mathbf{M}$  and  $\mathbf{Q}$  are  $3 \times 3$

$\mathbf{R}$  is  $2 \times 2$



# Proposed approach: Kalman + EM + augmented state

## Estimation of the latent variable



State-space model:

$$\mathbf{x}(t+1) = \mathbf{M}\mathbf{x}(t) + \mathbf{Q}^{1/2}\boldsymbol{\eta}(t)$$

$$\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{R}^{1/2}\boldsymbol{\epsilon}(t)$$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

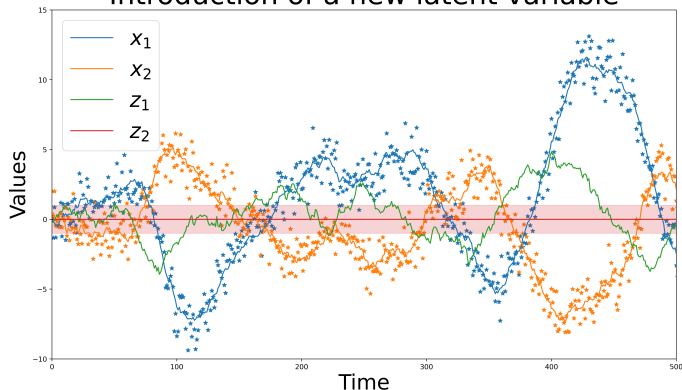
Variables and matrices:

Estimate  $\hat{\mathbf{M}}, \hat{\mathbf{Q}}, \hat{\mathbf{R}}$

Estimate  $z_1(0), \dots, z_1(T)$

# Proposed approach: Kalman + EM + augmented state

## Introduction of a new latent variable



State-space model:

$$\mathbf{x}(t+1) = \mathbf{M}\mathbf{x}(t) + \mathbf{Q}^{1/2}\boldsymbol{\eta}(t)$$

$$\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{R}^{1/2}\boldsymbol{\epsilon}(t)$$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Variables and matrices:

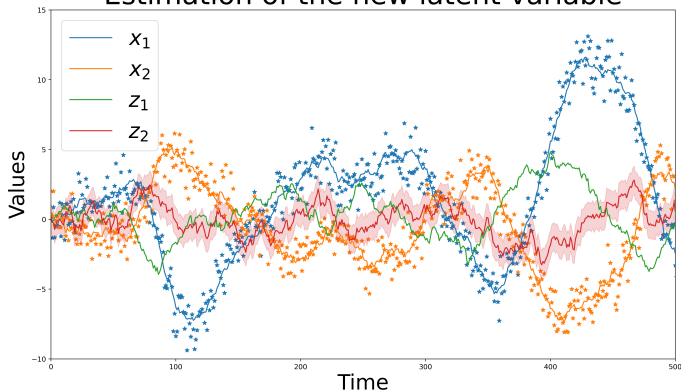
New state  $\mathbf{x} = [x_1, x_2, z_1, z_2]$

$\mathbf{M}$  and  $\mathbf{Q}$  are  $4 \times 4$

$\mathbf{R}$  is  $2 \times 2$

# Proposed approach: Kalman + EM + augmented state

## Estimation of the new latent variable



State-space model:

$$\mathbf{x}(t+1) = \mathbf{M}\mathbf{x}(t) + \mathbf{Q}^{1/2}\boldsymbol{\eta}(t)$$

$$\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{R}^{1/2}\boldsymbol{\epsilon}(t)$$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Variables and matrices:

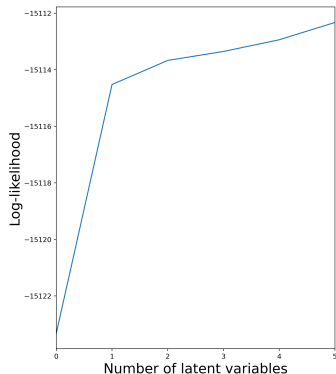
Estimate  $\hat{\mathbf{M}}, \hat{\mathbf{Q}}, \hat{\mathbf{R}}$

Estimate  $z_2(0), \dots, z_2(T)$

# Main scientific questions

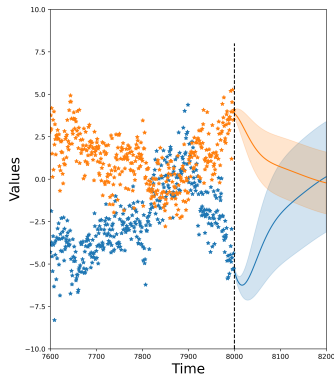
## Latent variables:

- how many are needed?
- what is their meaning?
- why do we need them?

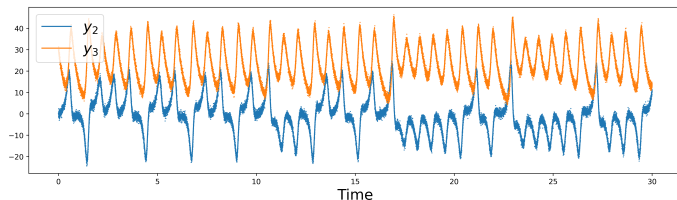


## Quality of the forecasts:

- are they precise?
- are they reliable?
- how do they compare?



# Application to Lorenz-63: formulation and OSSE



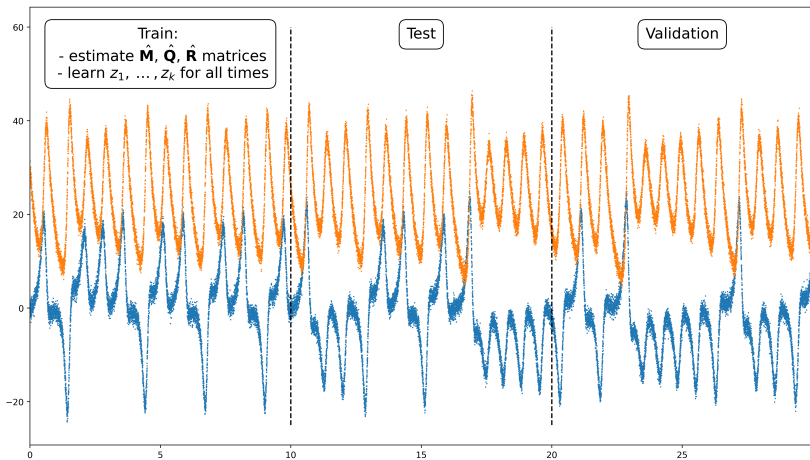
## Observations, state, and latent variables:

- $\mathbf{y} = [y_2, y_3]$  → **partial observations** of the system
- $\mathbf{x} = [x_2, x_3, \mathbf{z}]$  → **augmented state** with latent variables  $\mathbf{z}$
- $\mathbf{z} = [z_1, \dots, z_k]$  → **latent variables**

## OSSE:

- $dt = 0.001$  → observation sampling
- $\text{Var}(\epsilon) = 1$  → observation error variance

# Estimation of $\hat{\mathbf{M}}$ , $\hat{\mathbf{Q}}$ , $\hat{\mathbf{R}}$ and $z_1, \dots, z_k$ using training data

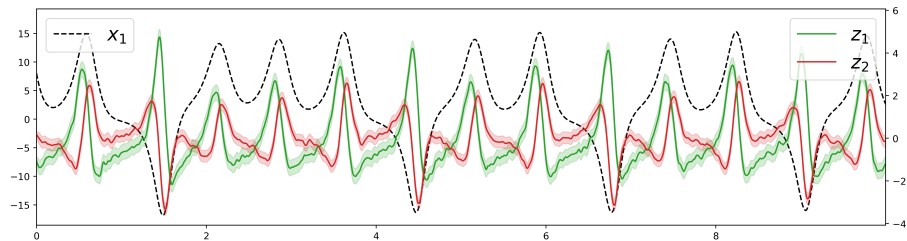


Time series are divided in three parts:

- train, test, and validation datasets
- EM algorithm is applied to the training data

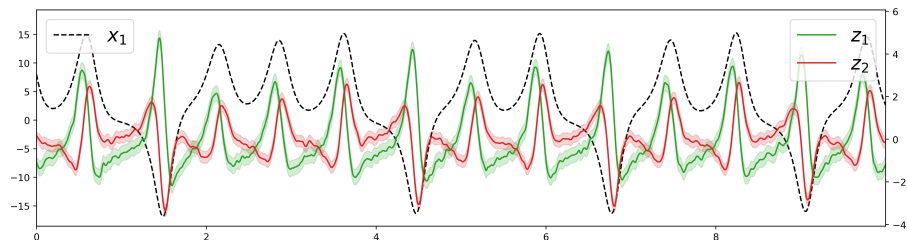
# Estimation and meaning of the latent variables

## Estimation of $z_1$ and $z_2$ :



# Estimation and meaning of the latent variables

## Estimation of $z_1$ and $z_2$ :



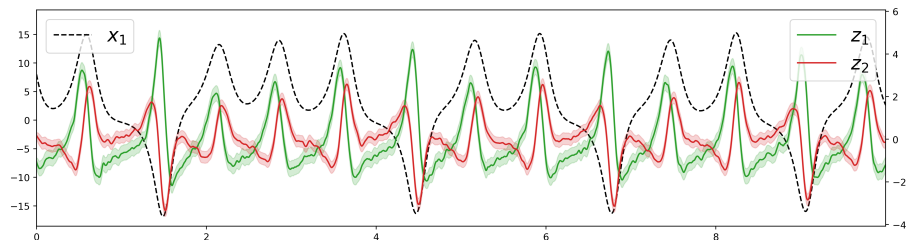
## Meaning of $z_1$ :

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



# Estimation and meaning of the latent variables

## Estimation of $z_1$ and $z_2$ :



## Meaning of $z_1$ :

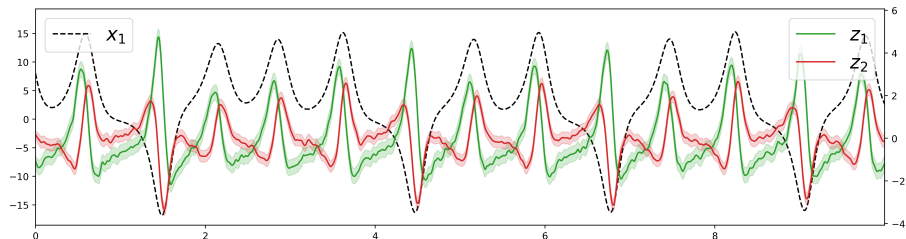
$$z_1 = a_2 \dot{x}_2 + a_3 \dot{x}_3$$

## Meaning of $z_2$ :

$$\begin{aligned} z_2 &= b_1 \dot{z}_1 + b_2 \dot{x}_2 + b_3 \dot{x}_3 \\ &= b_2 \dot{x}_2 + b_3 \dot{x}_3 + b_1 a_2 \ddot{x}_2 + b_1 a_3 \ddot{x}_3 \end{aligned}$$

# Estimation and meaning of the latent variables

## Estimation of $z_1$ and $z_2$ :



## Meaning of $z_1$ :

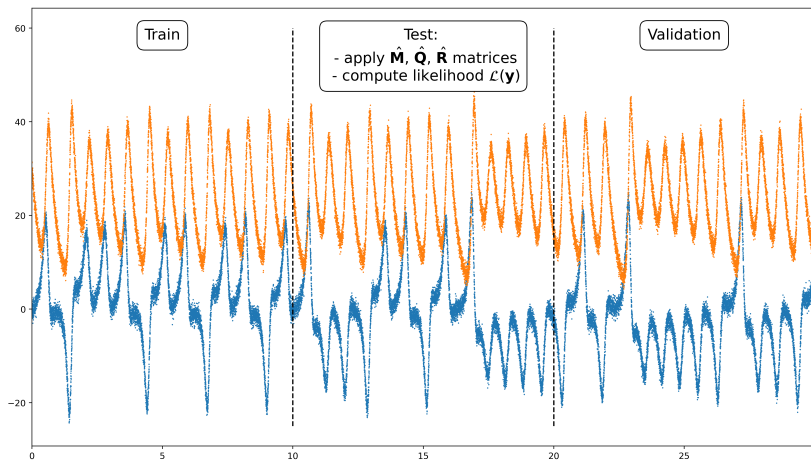
$$z_1 = a_2 \dot{x}_2 + a_3 \dot{x}_3$$

## Meaning of $z_2$ :

$$\begin{aligned} z_2 &= b_1 \dot{z}_1 + b_2 \dot{x}_2 + b_3 \dot{x}_3 \\ &= b_2 \dot{x}_2 + b_3 \dot{x}_3 + b_1 a_2 \ddot{x}_2 + b_1 a_3 \ddot{x}_3 \end{aligned}$$

⇒ latent variables take into account **time delays** of  $x_2$  and  $x_3$

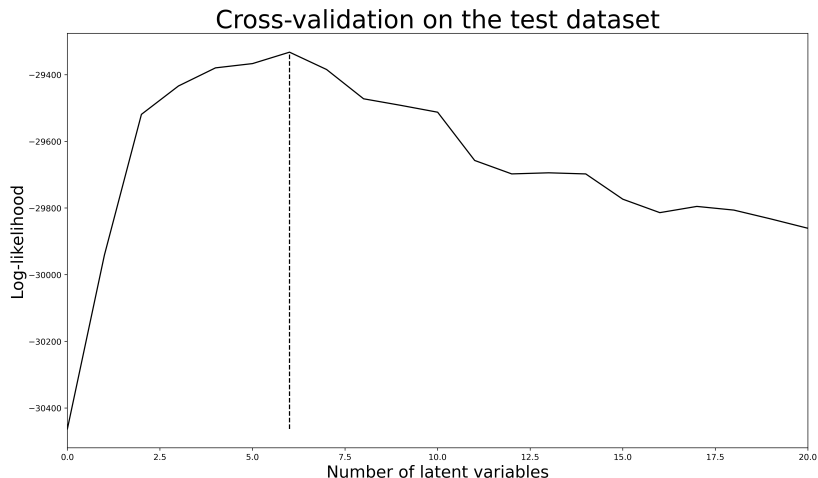
# Number of latent variables to retain



Different models are tested:

- from  $\mathbf{x} = [\mathbf{x}_2, \mathbf{x}_3, \mathbf{z}_1]$
- to  $\mathbf{x} = [\mathbf{x}_2, \mathbf{x}_3, \mathbf{z}_1, \dots, \mathbf{z}_{20}]$

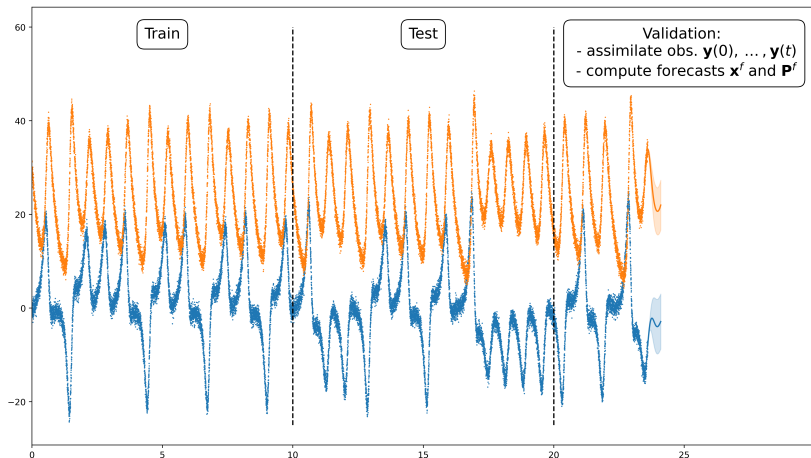
# Number of latent variables to retain



Optimal model in terms of likelihood:

- with  $\mathbf{x} = [\mathbf{x}_2, \mathbf{x}_3, \mathbf{z}_1, \dots, \mathbf{z}_6]$
- six latent variables are necessary

# Probabilistic forecasts using testing data



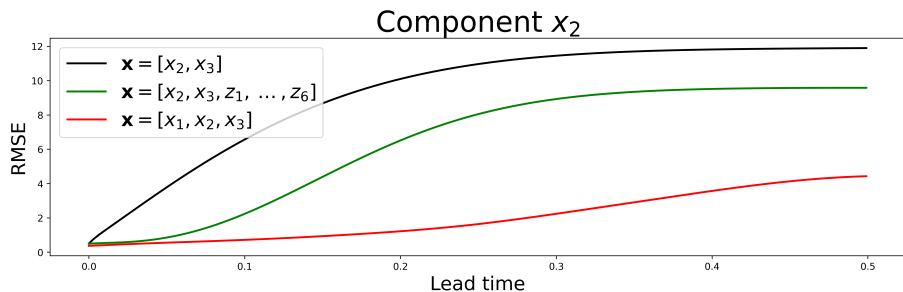
Recursive predictions, starting from an initial time  $t$ :

$$\mathbf{E}[\mathbf{x}_{t+1} | \mathbf{y}_1, \dots, \mathbf{y}_t] = \mathbf{M}\mathbf{E}[\mathbf{x}_t | \mathbf{y}_1, \dots, \mathbf{y}_t]$$

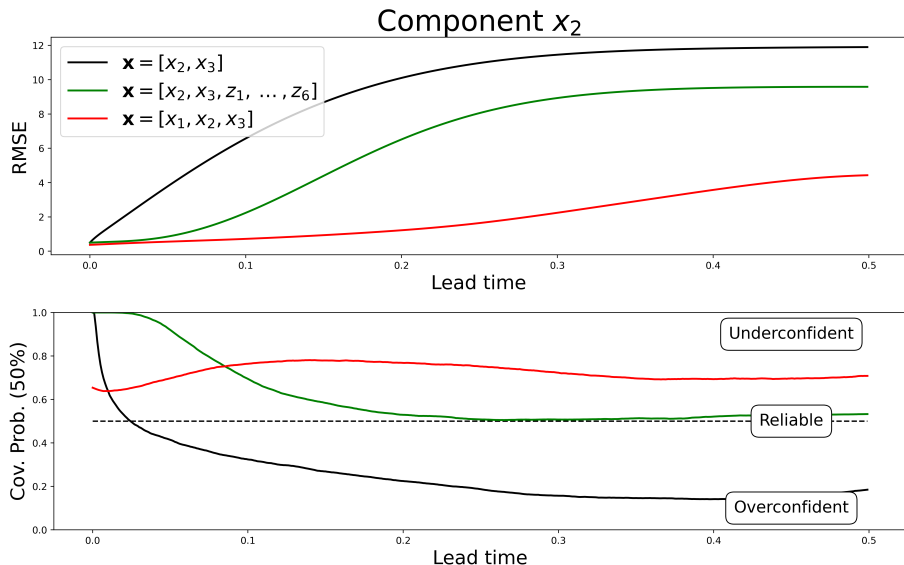
$$\text{Cov}[\mathbf{x}_{t+1} | \mathbf{y}_1, \dots, \mathbf{y}_t] = \mathbf{M}\text{Cov}[\mathbf{x}_t | \mathbf{y}_1, \dots, \mathbf{y}_t]\mathbf{M}^T + \mathbf{Q}$$

# Example of probabilistic forecasts (data- VS model-driven)

# Performance metrics (data- VS model-driven)



# Performance metrics (data- VS model-driven)

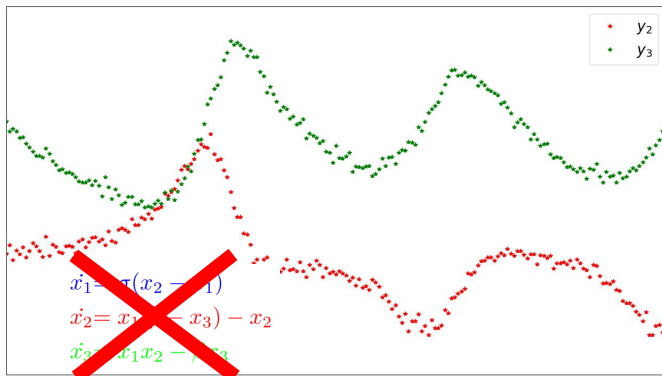




# Conclusions and perspectives

## Context:

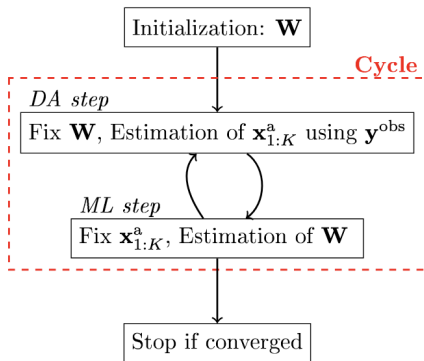
- pure data-driven predictions (i.e., without physical equations)
- missing components of the system (e.g.,  $x_1$  of Lorenz-63)
- simple assumptions (linearity, Gaussianity, constant operators)



# Conclusions and perspectives

## Key messages:

- fruitful combination of DA + ML
- inference of latent variables (interesting physical meaning)
- precise and reliable probabilistic forecasts

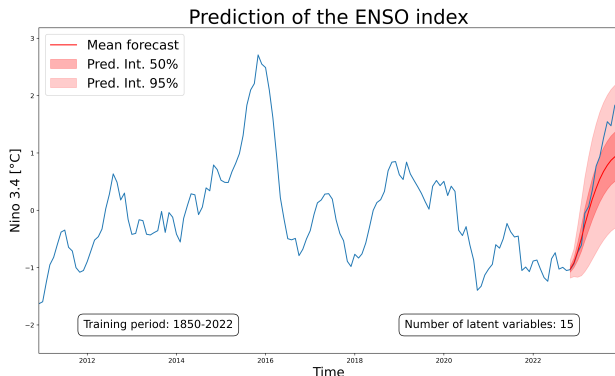


Source: Brajard et al. 2020, Journal of Computational Science

# Conclusions and perspectives

## Perspectives:

- application to climate indices (e.g., ENSO, AMO, NAO, etc.)
- application to high-dimensional data
- learn time-dependent  $\mathbf{M}(t)$  and  $\mathbf{Q}(t)$  operators

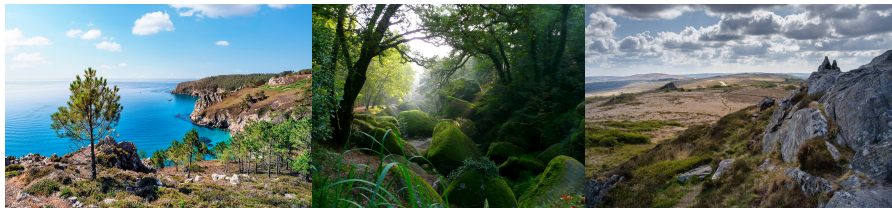


ありがとうございます (arigatō gozaimasu)!

**Interested in this topic? Collaborations are welcome!**

Contact:

- [pierre.tandeo@imt-atlantique.fr](mailto:pierre.tandeo@imt-atlantique.fr)
- [www.tandeo.wordpress.com](http://www.tandeo.wordpress.com)



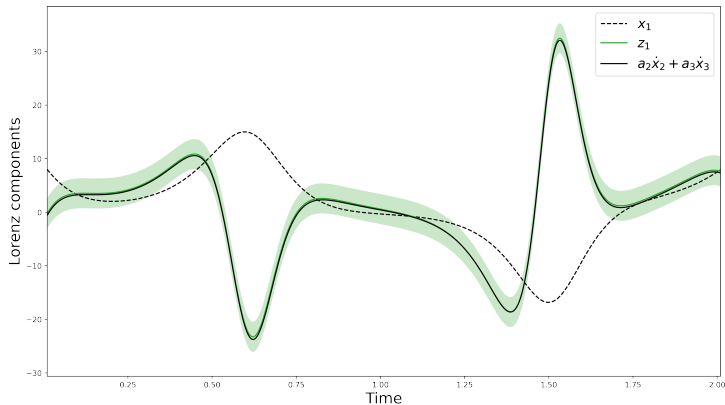
**Figure:** Nice places to visit near Brest, Bretagne, France

# Meaning of $z_1$

## Meaning of $z_1$ :

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

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



# Link between latent variables and $x_1$

Expression of  $z_1$ :

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

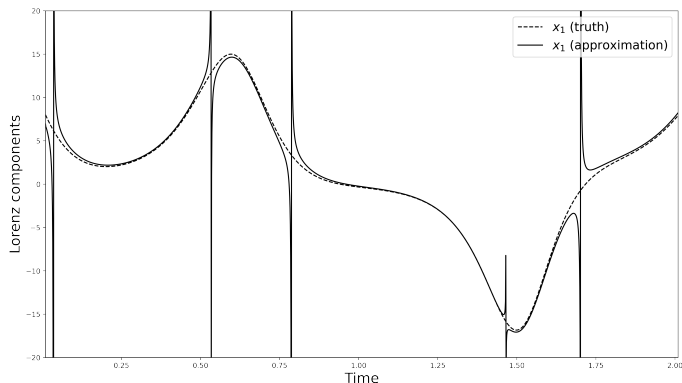
Lorenz-63 system:

$$\dot{x}_2 = x_1(\rho - x_3) - x_2$$

$$\dot{x}_3 = x_1 x_2 - \beta x_3$$

Approximation of  $x_1$ :

$$\hat{x}_1 = \frac{z_1 + a_2 x_2 + a_3 \beta x_3}{a_2 \rho - a_2 x_3 + a_3 x_2}$$

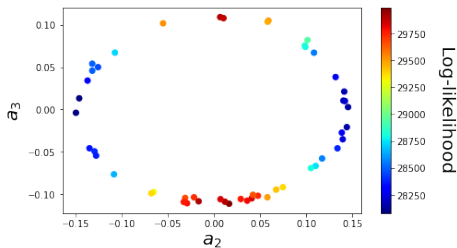
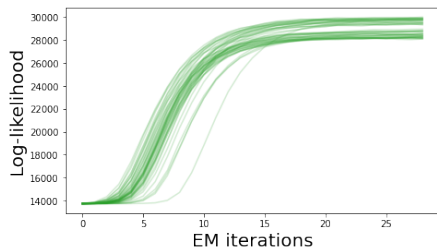


$\Rightarrow x_1$  is partially retrieved (division by zero in specific regions)

# Non-unique solution

Meaning of  $z_1$ :

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



⇒ **two clusters** of solutions ( $a_2 \approx 0$  and  $a_3 \approx 0$ )

⇒  $x_3$  **more important** than  $x_2$  in the Lorenz-63 system