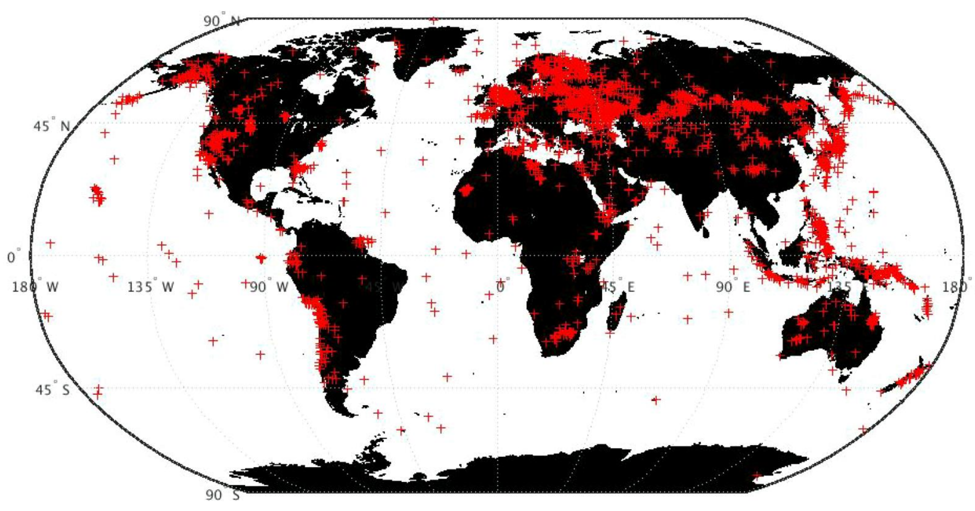
On using self-sustained events for stochastic waveform modelling with deep neural networks

Xavier Cassagnou[1] , Christophe Millet[1,2] , Mathilde Mougeot[1,3],
Cyril Nefzaoui Blanchard[1,2]

[1] Centre Borelli, ENS Paris-Saclay, F-91190 Gif-sur-Yvette, France
[2] CEA, DAM, DIF, 91297 Arpajon, France
[3] ENSIIE, F-91000 Évry, France

Context

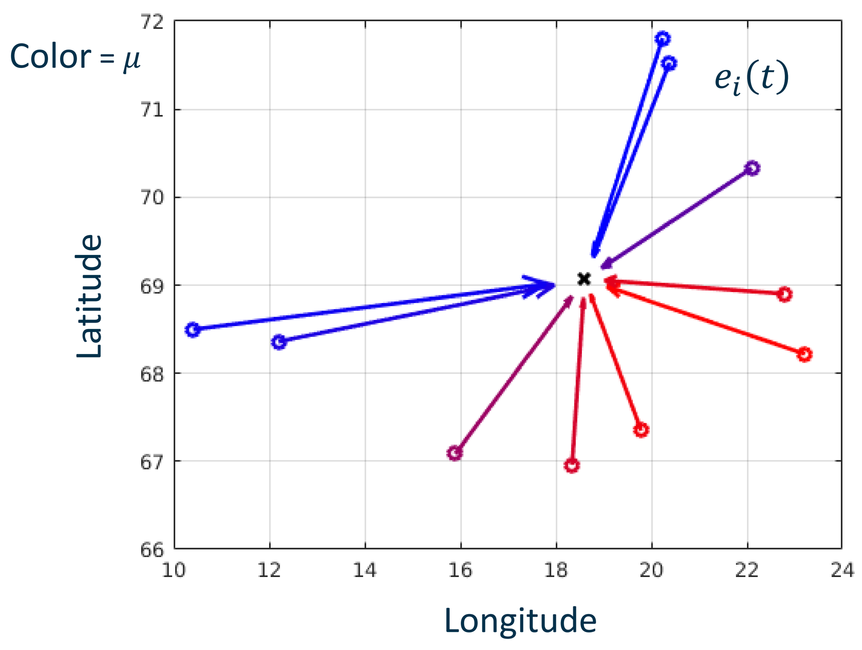
- Waveform sensor stations of the International Monitoring System (IMS).
- Self sustained perturbators are “mixed“ with the **event of interest**. Such perturbators are known to cause false alarms.
- Need to produce automated data-driven processes to reduce the false alarms rate.



Localization of events in the database REB that have generated infrasound detections during the period 2011-2019.

Usual approach:

- Global Association, NET-VISA, etc.
- Case-by-case study that relies on expert judgments, propagation models and prior on the distribution of perturbators.



Scientific question.

*Can we predict the effect of **perturbators** from past data recorded at a given infrasound station?*

- Usual data include time series, azimuths, metadata and eventually knowledge obtained from post-processing of signals.
- Background noise is due to **many localized events** distributed in the near-to-far field, meteorological effects. All these effects are difficult to represent numerically.

A simplified approach to represent a signal generated from n sources is using the linear superposition.

A station records a signal $s(t) = \sum_{i=1}^{n} s_i(t)$ where each function $s_i(t)$ satisfies an ordinary differential equation (ODE).

The ODE depends on two elements:

- The source $e_i(t)$;
- The **atmospherical context μ** (sound speed profile, etc.).

Mathematical approach

- We assume that the signal is related to a linear ODE $x'(t) = f(x(t)) = Ax(t)$. For self-sustained events, we get

$$\begin{pmatrix} \dot{s_i} \\ \ddot{s_i} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\boldsymbol{\omega_i^2} & 0 \end{pmatrix} \begin{pmatrix} s_i \\ \dot{s_i} \end{pmatrix}.$$

- Summing up the contributions leads to

$$x = \begin{pmatrix} s_1 \\ \dot{s_1} \\ \cdots \\ s_n \\ \dot{s_n} \end{pmatrix} \text{ and } A = \begin{pmatrix} 0 & 1 & \cdots & 0 & 0 \\ -\boldsymbol{\omega_1^2} & 0 & & \ddots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \ddots & & 0 & 1 \\ 0 & 0 & \cdots & -\boldsymbol{\omega_n^2} & 0 \end{pmatrix}.$$

Learning matrix A

Through a time series $y^k = \varphi(x^k)$, learn f or its flow $F(x^k) = x^{k+1}$ with a Neural Network (NN).

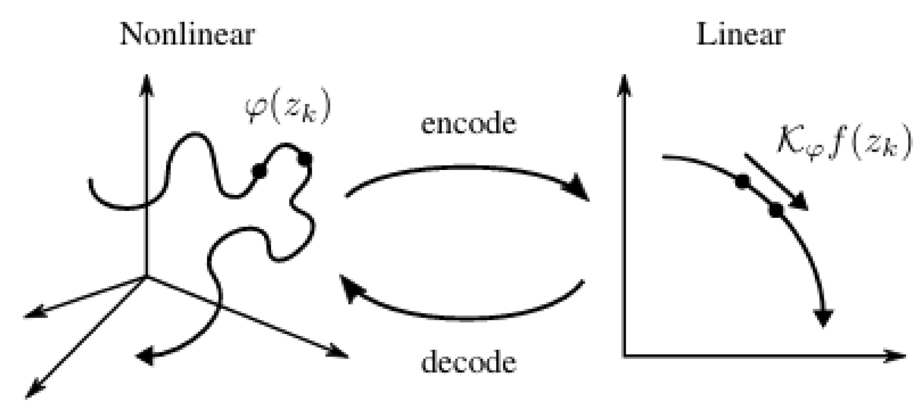
Koopman theory

The Koopman operator $\mathcal{K}_F$ defined for any $\varphi : \mathbb{R}^n \to \mathbb{R}$ by

$$\mathcal{K}_F \varphi(x^k) = \varphi \circ F(x^k) = \varphi(x^{k+1})$$

contains all the information of F (as well as A).

Thus, the information of the time series can be encoded in an observable space where the evolution operator is linear.



Approach #1 : estimate of f

Hypothesis : $\varphi = \mathrm{id}$.
We minimize the following loss function for parameters θ and ν^k:

$$\underbrace{\sum_{k=q+1}^{m-q} \sum_{i=-q}^{q} \omega_i \left\| F_\theta^i(x^k) - x^{k+i} \right\|^2}_{\text{with } F_\theta = \mathrm{RK}_4 f_\theta \text{ and } x^k = y^k + \nu^k} + \underbrace{\omega_\nu \sum_{t=1}^{m} \left\| \nu^k \right\|^2 + \omega_W \sum_{i=1}^{l} \| W_i \|_F^2}_{\text{penalization}}$$

W_i is the weights matrix between the layers i and $i+1$ of the network, and $\omega_\nu, \omega_W, (\omega_i)_i$ are hyperparameters to tune.
$(\nu^k)_{1 \leq k \leq m}$ is a gaussian noise.

Advantage : simultaneous learning of the weights θ of the network and of the noise’s characteristics.

Approach #2 : learning of F

Hypothesis : φ is any function $\mathbb{R}^n \to \mathbb{R}$.
The Koopman operator is learnt by an autoencoder architecture

$$F = \chi_d \circ C \circ \chi_e$$

where the operator C is linear.

The loss function takes into account the estimate of $D = C^{-1}$ to ensure the reversibility of the dynamics:

$$\overbrace{\frac{\omega_{\mathrm{id}}}{m} \sum_{k=1}^{m} \| x^k - \chi_d \circ \chi_e(x^k) \|^2}^{\text{auto-encoding}} + \overbrace{\frac{1}{mq} \sum_{k=1}^{m} \sum_{i=1}^{q} \omega_+ \varepsilon_{ik}^+ + \omega_- \varepsilon_{ik}^-}^{\text{temporal evolution}} + \omega_c \overbrace{[\| DC - I_\kappa \|^2 + \| CD - I_\kappa \|^2]}^{\text{consistance}}$$

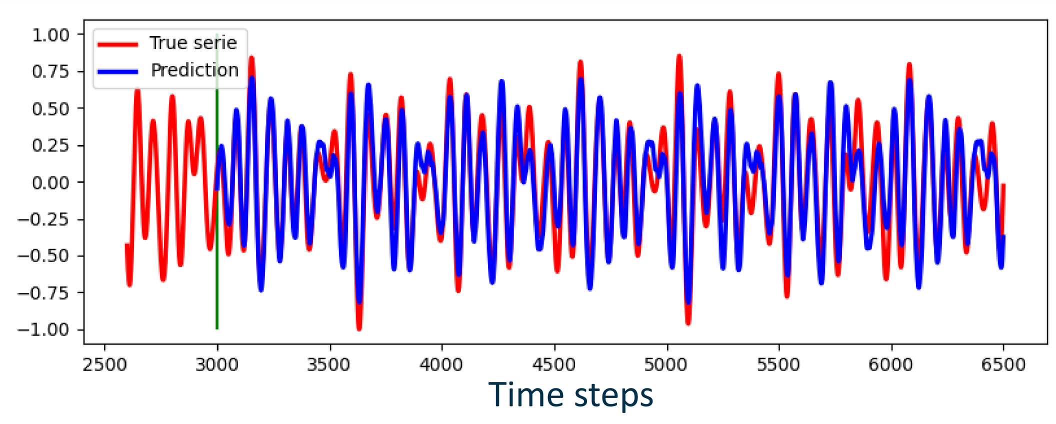
with $\varepsilon_{ik}^{\pm} = \left\| x^{k \pm i} - \chi_d \circ C^{\pm i} \circ \chi_e(x^k) \right\|^2$.

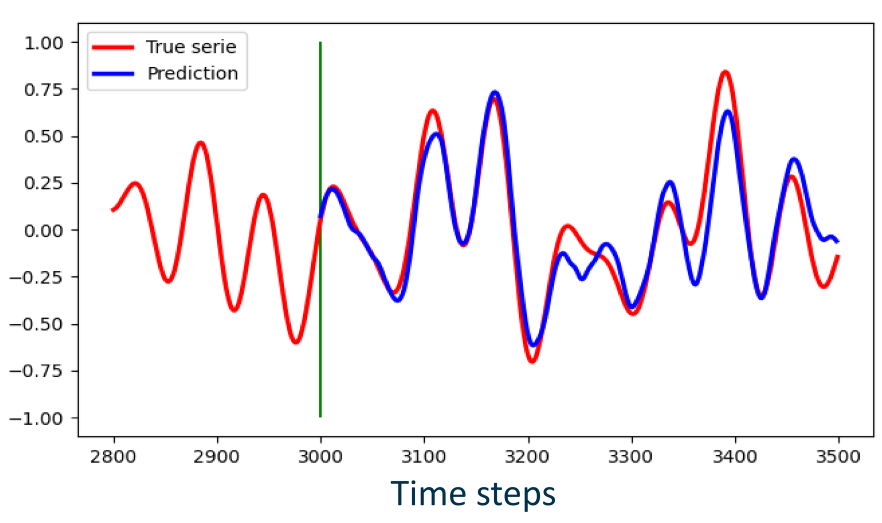
Advantage : learning the evolution operator for the observable.

Results

Preliminary numerical results with “toy models”:

- ODE: 15-dimensional linear system of coupled ODEs.
- The matrix A is chosen randomly and skew-symmetric.





Approach #2 with a linear ODE:

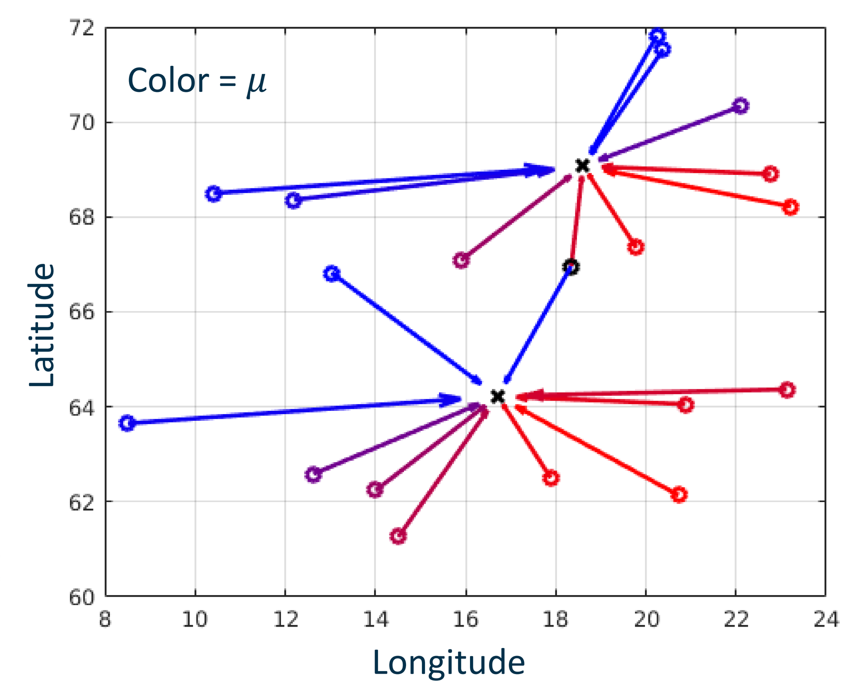
- **Predictions** are in very good overall agreement with the **truth**.

Architecture of the NN:

- The input layer has 260 units corresponding to the 260 previous time steps.
- The encoder and the decoder are both made of 3 hidden layers of 80 units.
- The evolution operator C is a matrix 27x27.

Future work

- Use of strategies #1 and #2 for extracting and predicting features which are due to recurrent and localized events (human activities, etc.) in realistic noisy signals, without knowledge on the source locations.
- Identify the role of the **context parameter μ** (that codes the effect of atmosphere) in the learning strategy
- Generalize strategies #1 and #2 to multiple infrasound stations using a connected network.



References

[1] Rudy, S. H., Kutz, J. N., and Brunton, S. L. (2019). *Deep learning of dynamics and signal-noise decomposition with time-stepping constraints.* Journal of Computational Physics, 396:483–506.
[2] Lusch, B., Kutz, J. N., and Brunton, S. (2018). *Deep learning for universal linear embeddings of nonlinear dynamics.* Nature Communications, 9.
[3] Azencot, O., Erichson, N., Lin, V., and Mahoney, M. W. (2020). *Forecasting Sequential Data using Consistent Koopman Autoencoders.* International Conference on Machine Learning.