

# Comment rapprocher connaissance physique et apprentissage statistique pour les données spatio-temporelles ?

Thomas Romary

STIM, Mines Paris PSL

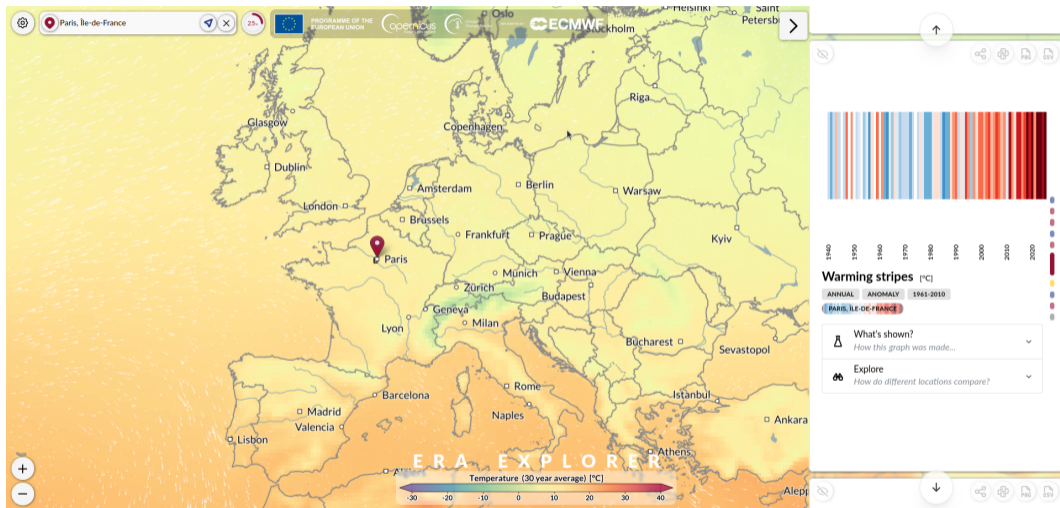
Soirée scientifique Geolearning, 8 avril 2026



**GEOLEARNING**  
CHAIRE /// Data Science for the Environment

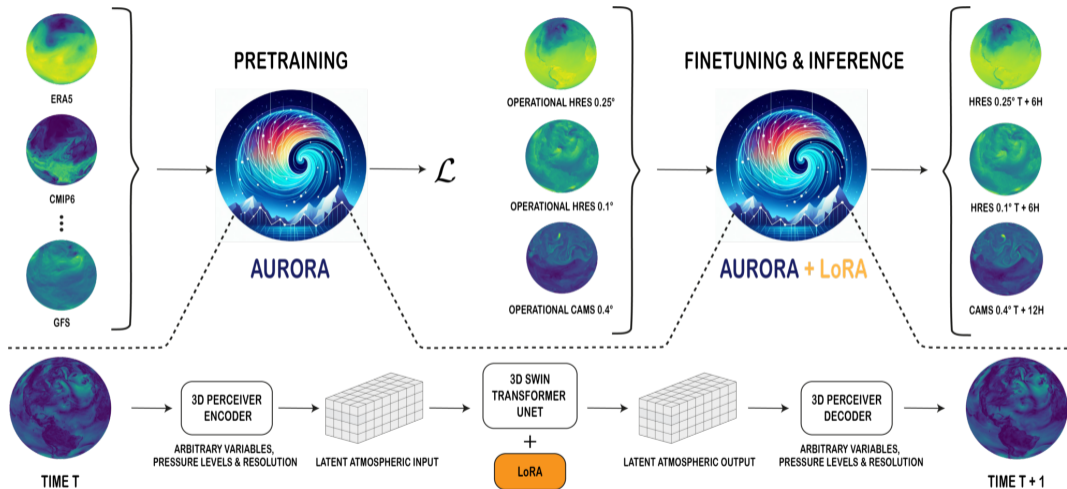


# Du global au local : le mur des données en environnement



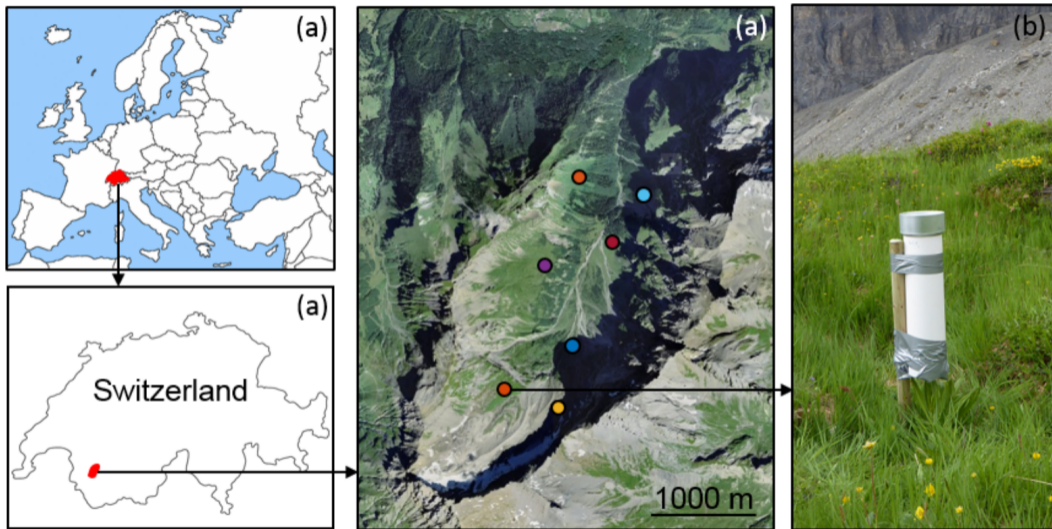
Source : Copernicus era-explorer ERA5 reanalysis, (ECMWF)

# Du global au local : le mur des données en environnement



Source : Aurora, Microsoft Research

## Du global au local : le mur des données en environnement



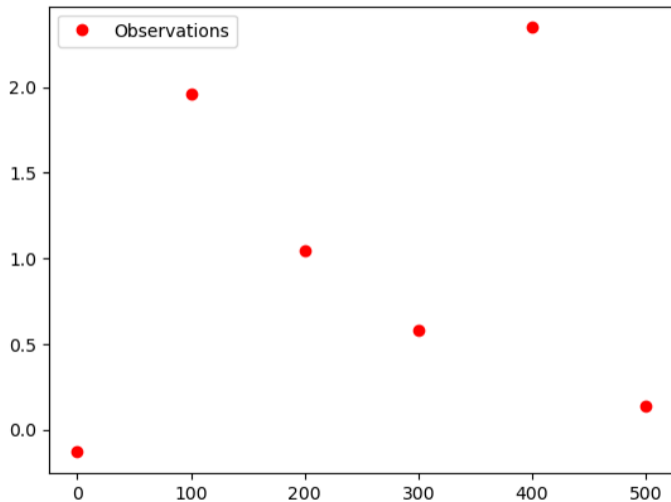
Source : Benoit et al. (2018), WRR

# Du global au local : le mur des données en environnement

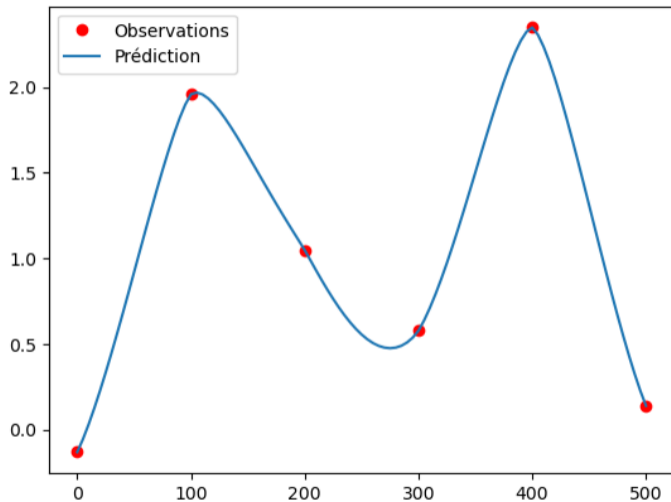
Limite des approches purement data-driven :

- Echelle de résolution des données globale souvent trop grossière pour les applications locales
- Mauvaise généralisation à des conditions extrêmes
- Pas de quantification de l'incertitude

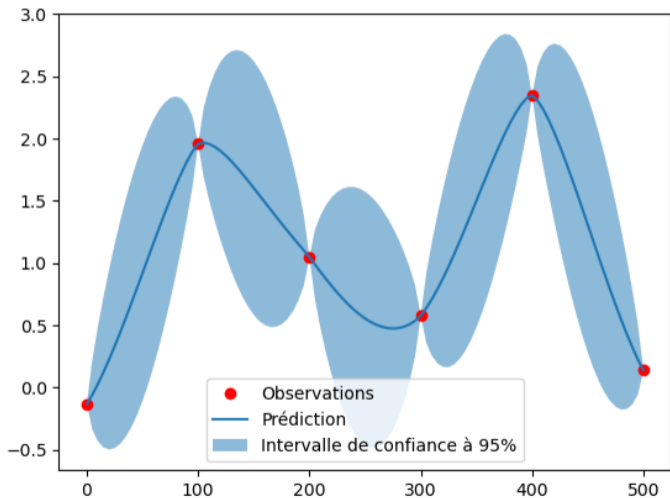
# Le processus gaussien : un pont entre statistique et physique



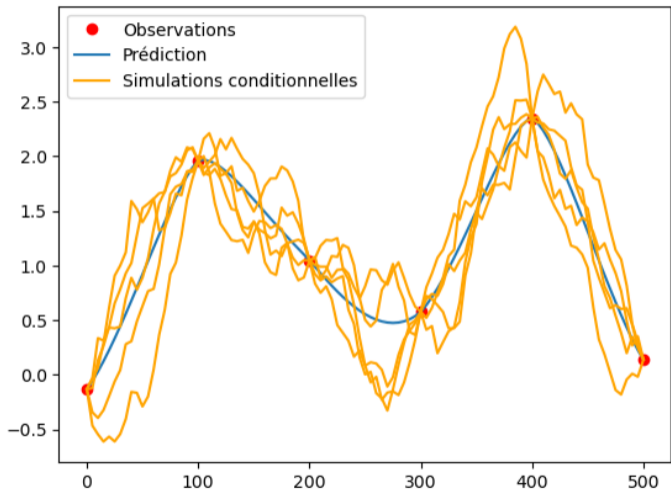
# Le processus gaussien : un pont entre statistique et physique



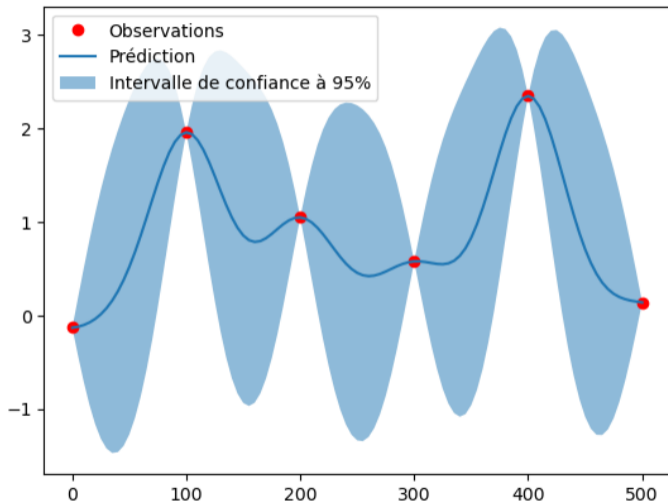
# Le processus gaussien : un pont entre statistique et physique



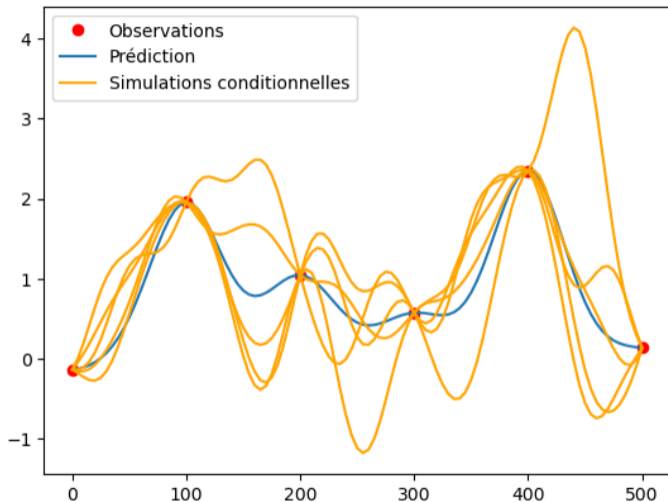
# Le processus gaussien : un pont entre statistique et physique



# Le processus gaussien : un pont entre statistique et physique



# Le processus gaussien : un pont entre statistique et physique



# Le processus gaussien : un pont entre statistique et physique

Un processus gaussien est caractérisé par :

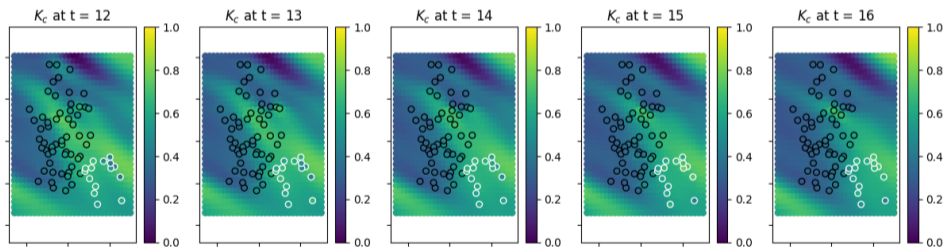
- une fonction de moyenne  $\mu(\mathbf{x})$
- une fonction de covariance  $\mathbf{k}(\mathbf{x}, \mathbf{x}')$

En conditionnant le processus gaussien sur des observations, on peut faire des prédictions à des points non observés, avec une quantification de l'incertitude

# Intégrer la connaissance physique dans les processus gaussiens : l'approche SPDE

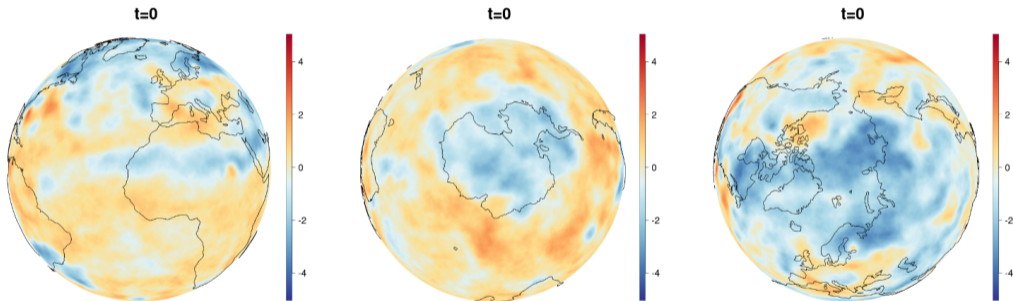
(Stochastic Partial Differential Equation)

$$\left[ \frac{\partial}{\partial t} + \frac{1}{c}(\kappa^2 - \nabla \cdot \mathbf{H} \nabla)^\alpha + \frac{1}{c} \boldsymbol{\gamma} \cdot \nabla \right] X(\mathbf{s}, t) = \frac{\tau}{\sqrt{c}} Z(\mathbf{s}, t)$$



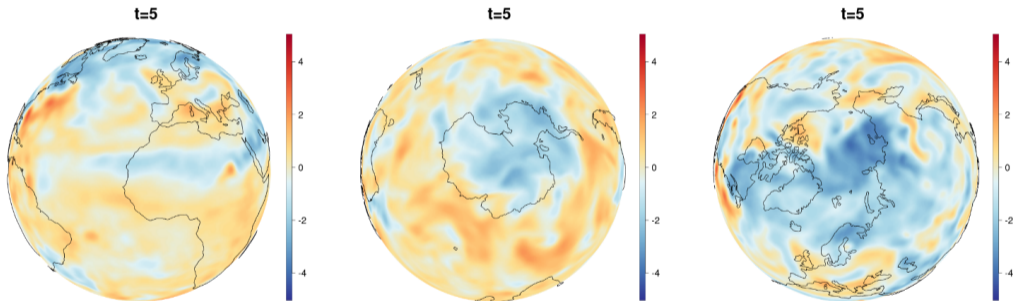
Prédiction du clear sky index en 5 pas de temps, source : Clarotto et al. (2024), Spatial Statistics

# Intégrer la connaissance physique dans les processus gaussiens : l'approche SPDE



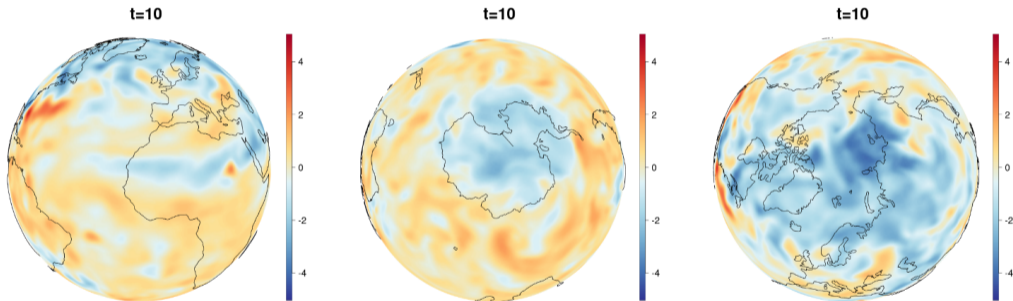
source : Pereira M., Clarotto L., Desassis N. (2025)

# Intégrer la connaissance physique dans les processus gaussiens : l'approche SPDE



source : Pereira M., Clarotto L., Desassis N. (2025)

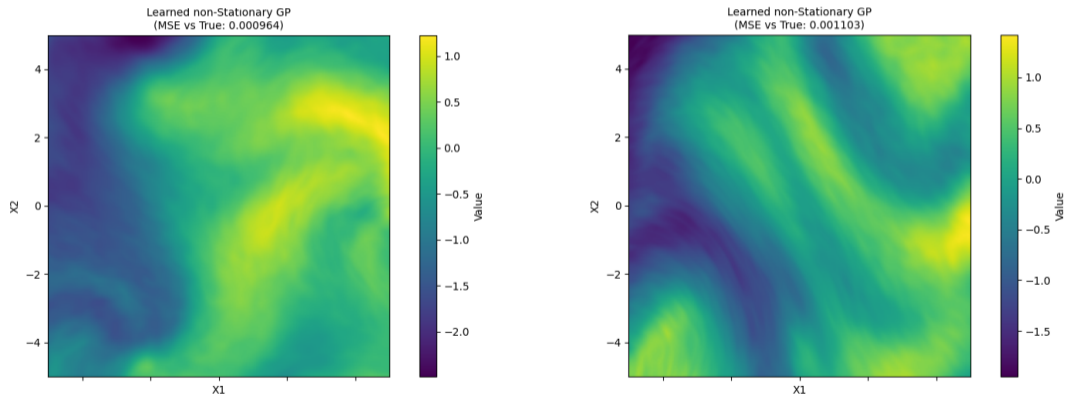
# Intégrer la connaissance physique dans les processus gaussiens : l'approche SPDE



source : Pereira M., Clarotto L., Desassis N. (2025)

# Deep Kernel Learning : allier la flexibilité du deep learning à la rigueur statistique

$$\mathbf{k}_\theta(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{k}_0(|\mathbf{f}_\theta(\mathbf{x}_i) - \mathbf{f}_\theta(\mathbf{x}_j)|)$$



source : Romary et al. (2025)

## Des applications variées

- Météorologie : simulation de conditions météorologiques à haute résolution
- Modélisation de la qualité de l'air : prédiction de la concentration en polluants
- Pollution des sols : modélisation de la distribution spatiale de contaminants
- Energie : prédiction de la production d'énergie renouvelable (photovoltaïque, éolienne, houle)
- ...

## Des défis méthodologiques et computationnels à relever

- Scalabilité
- Intégration du deep kernel learning dans l'approche SPDE
- Physique plus complexe : prise en compte de processus non linéaires, de couplages entre variables, etc.
- Relâcher l'hypothèse de normalité
- ...