HOMER DURAND

# RECALAGE D'HISTORIQUE ET MACHINE LEARNING POUR LA RECHERCHE DE PARAMÈTRES DE MODÈLES D'OCÉAN/ATMOSPHÈRE

**Encadrement :** R. Lguensat (LOCEAN-LSCE), J. Deshayes (LOCEAN), V. Balaji (Princeton - IPSL)

## Subject

Numerical models are central to climate science.

The estimation of the parameters of these models is very important because it greatly influences their predictive quality.

Numerical simulations are extremely expensive in terms of computing time. Developing statistical emulators of low cost and fast computation time is a research direction that is very active in the machine learning community.

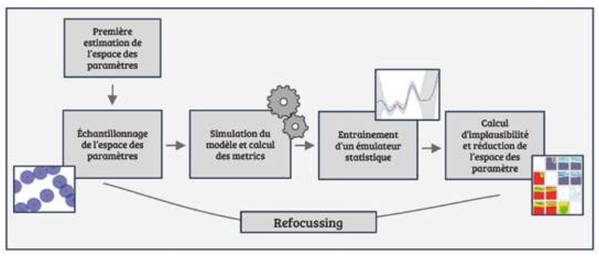
History Matching (**HM**) based on Gaussian Process Regression (**GPR**) has recently attracted the interest of the climate science community for parameter tuning (Hourdin et al. 2020, Couvreux et al. 2020, Williamsson et al. 2017)

In this work, we are interested in studying:

1. The possibility (or not) of using **HM** for multiscale models
2. See if other models from the machine learning community (random forest, bayesian neural network, etc..) would allow to obtain similar or even more interesting results.

## History Matching

- **Statistical method for the calibration of numerical models based on observed data and simulation results**
- Sampling of the parameter space and calculation of metrics by model simulation
- Use of a statistical model to predict the metrics and the variability of the prediction from a point in the parameter space
- Comparison between observed data and emulator predictions by a distance measure
- Reduction of the model parameter space by rejecting the most unlikely areas




- **Sampling method :** Latin Hypercube Sampling Minimax
- **Emulator :** Gaussian Process (or Linear Regression)
- **Implausibility :** $I_r(\mathbf{x}) = \frac{|z_r - \mathrm{E}^*[g_r(\mathbf{x})]|}{(V_{o,r} + V_{m,r} + V_{c,r})^{1/2}}$

Takes into account the variance of the emulator predictions, the uncertainty on the observations and the divergence between the model and the "physical reality"

## Two layers Lorenz96

- System of partial differential equations introduced in 1996 by E. Lorenz to study the predictability of meteorological systems
- Toy model regularly used as a "test case" for data assimilation and the study of chaotic systems
- Models three characteristics of climate systems: advection, diffusion and external forcing

$$\frac{dX_k}{dt} = \underbrace{-X_{k-1}(X_{k-2} - X_{k+1})}_{\text{Advection}} \underbrace{-X_k}_{\text{Diffusion}} \underbrace{+F}_{\text{Forcing}} \underbrace{-hc\bar{Y}_k}_{\text{Coupling}}$$

$$\frac{1}{c}\frac{dY_{j,k}}{dt} = \underbrace{-bY_{j+1,k}(Y_{j+2,k} - Y_{j-1,k})}_{\text{Advection}} \underbrace{-Y_{j,k}}_{\text{Diffusion}} \underbrace{+\frac{h}{j}X_k}_{\text{Coupling}}$$

- A slow component (X) coupled to a fast component (Y). Interest for the study of coupled ocean-atmosphere models.
- Four parameters: h, F, c, b.

| Params | Prior | True |
|---|---|---|
| F | [-20,20] | 10 |
| h | [-2,2] | 1 |
| c | [0,20] | 10 |
| b | [-20,20] | 10 |

$$f(X,Y) = \begin{pmatrix} X \\ \bar{Y} \\ X^2 \\ X\bar{Y} \\ \bar{Y}^2 \end{pmatrix}$$

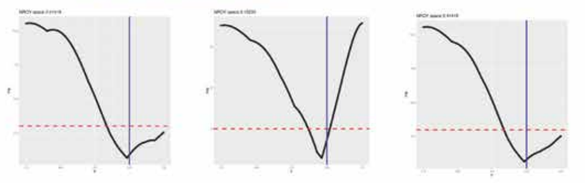


## First approach

- Handling of a tool for history recalibration and for the creation of emulators of the Gaussian process type
  - Debugging
  - Application to two layers Lorenz96
- Parallelization of some sections of the code
- Modification of the code to use different emulators (Gpytorch for GPs)

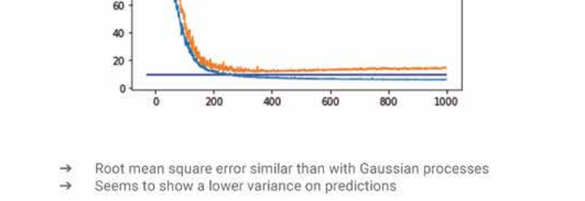
## Experiments

- **One Layer Lorenz96 model**



→ Seems to show better performance without the X_sq metric

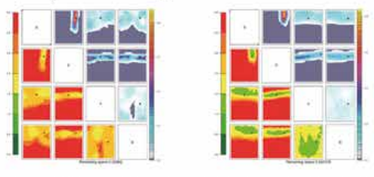
- **Number of samples for emulator training** (40, 120, 200)



→ A reduced number of samples allows to significantly reduce the parameter space

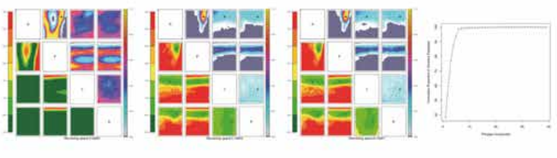
## Results

- **Comparison of integration scheme (Euler vs RK4) to study model divergence**



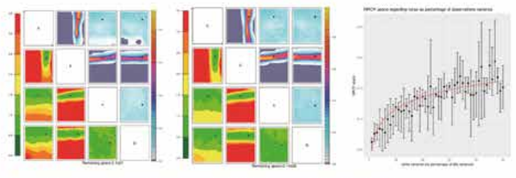
→ The observations are in less likely areas with a lower quality integration scheme (euler explicit)

- **Reduction of the dimension of the metrics space (PCA or EOF)**



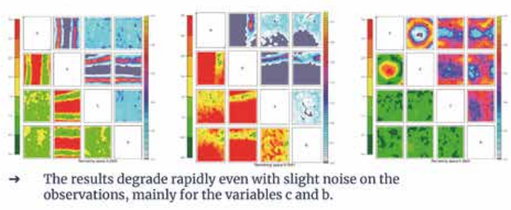
→ 97% of variance explained from 7 variables

- **Impact of a white Gaussian noise on the observations on the NROY space**



→ The results degrade rapidly even with slight noise on the observations, mainly for the variables c and b.

- **Impact of the choice of metrics**

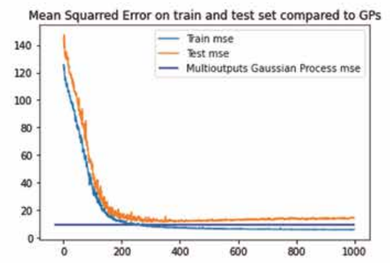


→ The results degrade rapidly even with slight noise on the observations, mainly for the variables c and b.

## Future work

- New emulators from Machine Learning :
  - Bayesian neural networks
  - Decision tree forest
- Application to the NEMO ? ocean model

- **Comparison of the predictions of different emulators:**
  - Kullback-Leibler divergence
  - Root mean square error
  - Mean variability




→ Root mean square error similar than with Gaussian processes
→ Seems to show a lower variance on predictions

Laboratoire d'Océanographie et du Climat
Expérimentations et Approches Numériques