Data-Model Coupling for SST-DA

Workshop

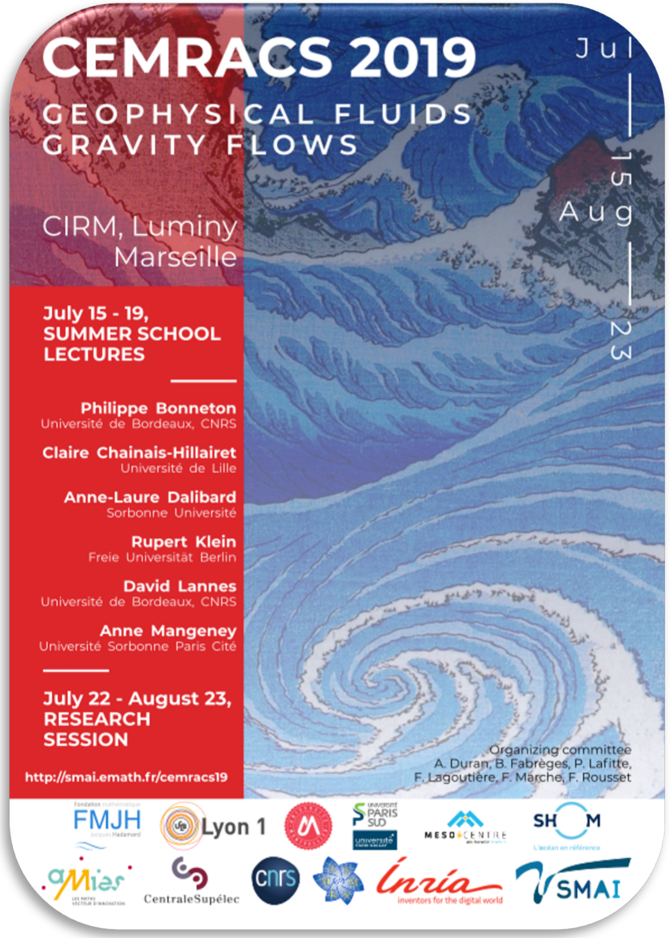
Statistical Modeling and Machine Learning in Meteorology and Oceanography

Brest, february 10-11, 2020

Jules Guillot, LMBA-UBS

PADOCARI project:

PArtial Differential equations for OCeanic ARtificial Intelligence



Supervised by:

- Emmanuel FRENOD
- Hélène FLOURENT
- Julien BRAJARD
- Frédérique CHARLES

Participants:

- Jules GUILLOT
- Guillaume KOENIG
- Hadi MINBASHIAN

Outline

- Data-Model Coupling
- SST-Data Assimilation
- Numerical results
- Conclusion
- References

Data-Model Coupling

DATA ASSIMILATION

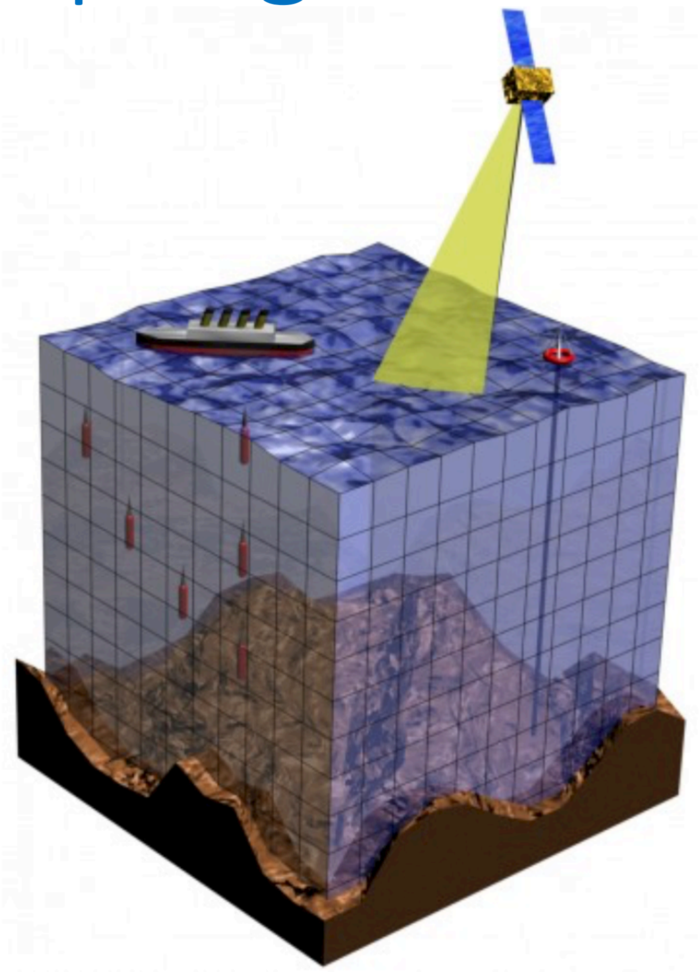
Reality

Data flow

Observations

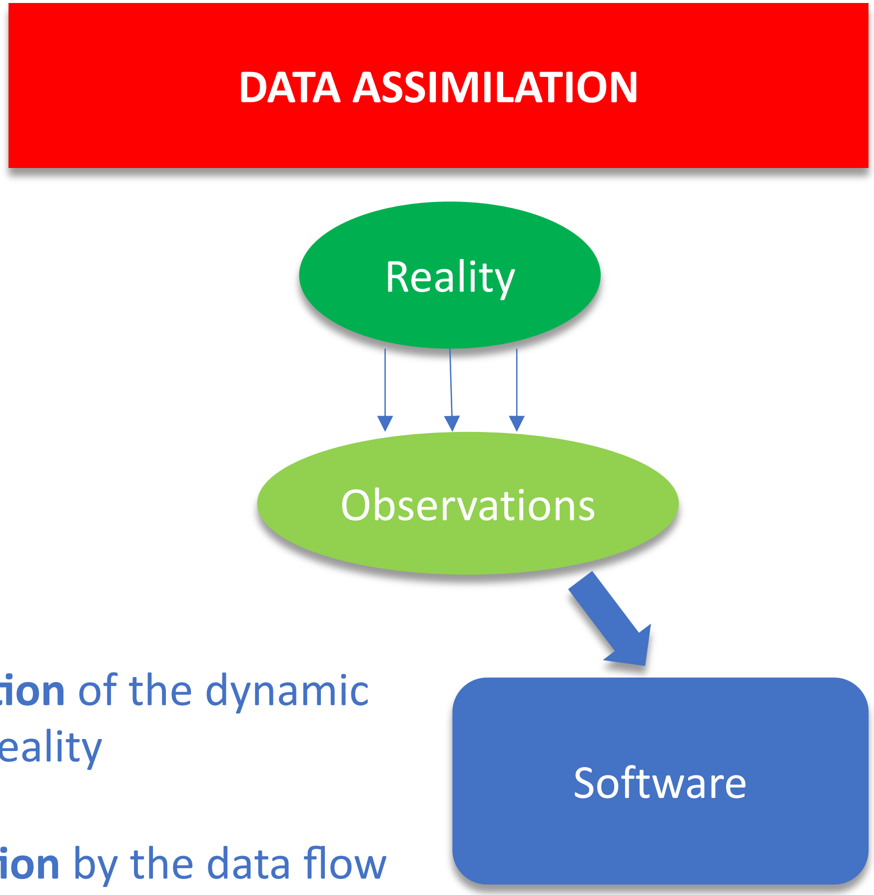
?

Real state at each time t

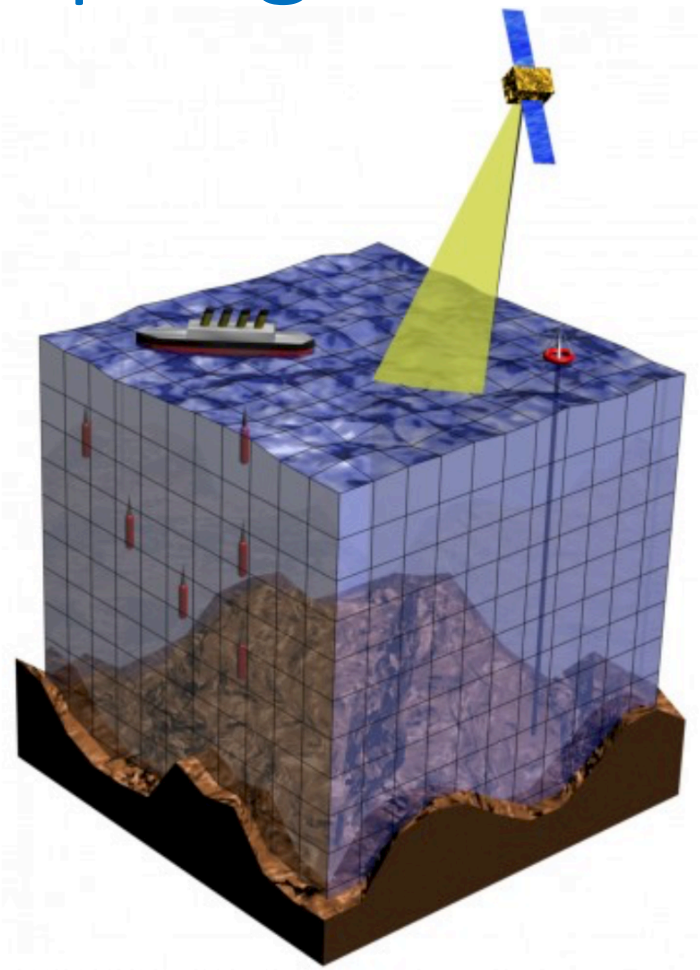


from CERFACS

Data-Model Coupling

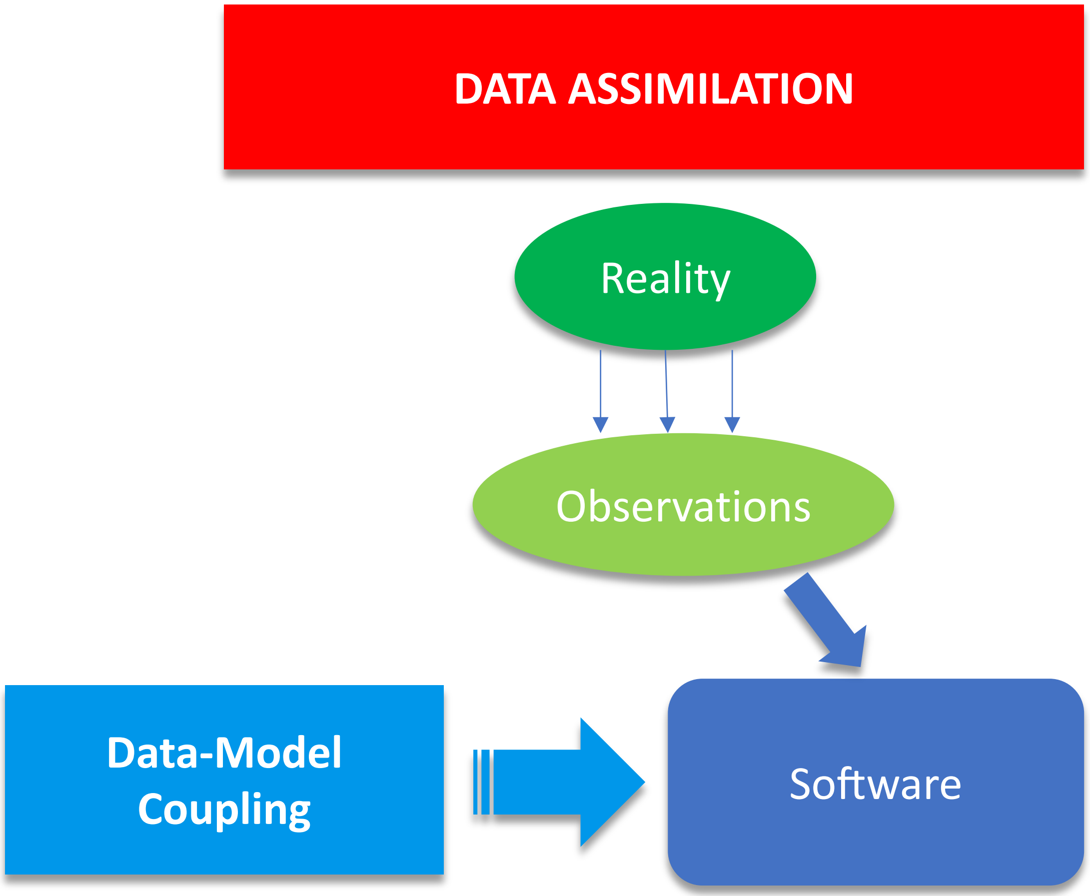


- **Simulation** of the dynamic of the reality
- **Correction** by the data flow

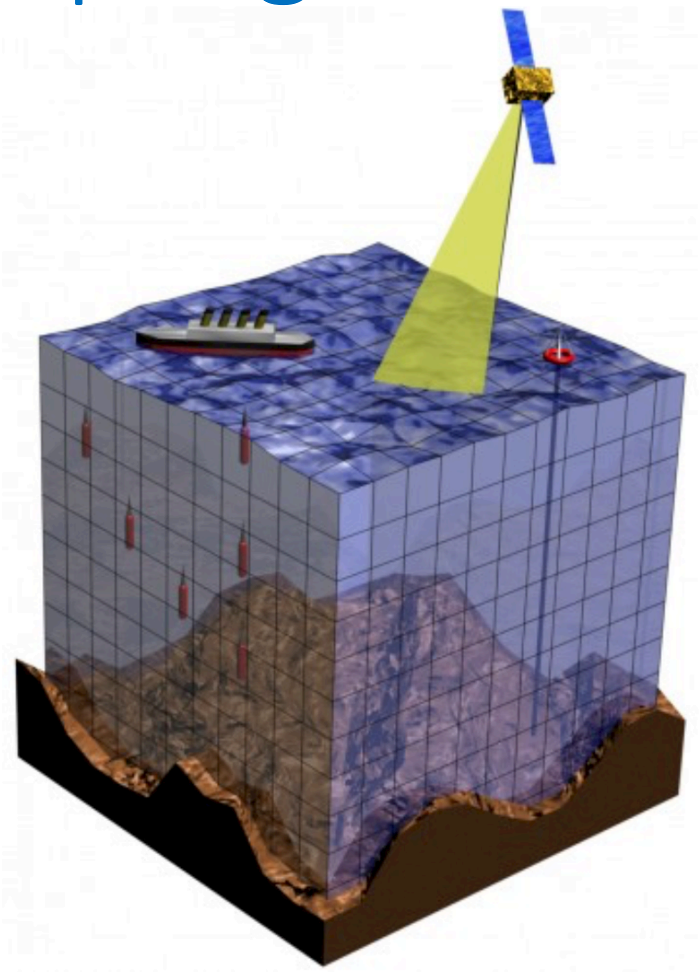


from CERFACS

Data-Model Coupling

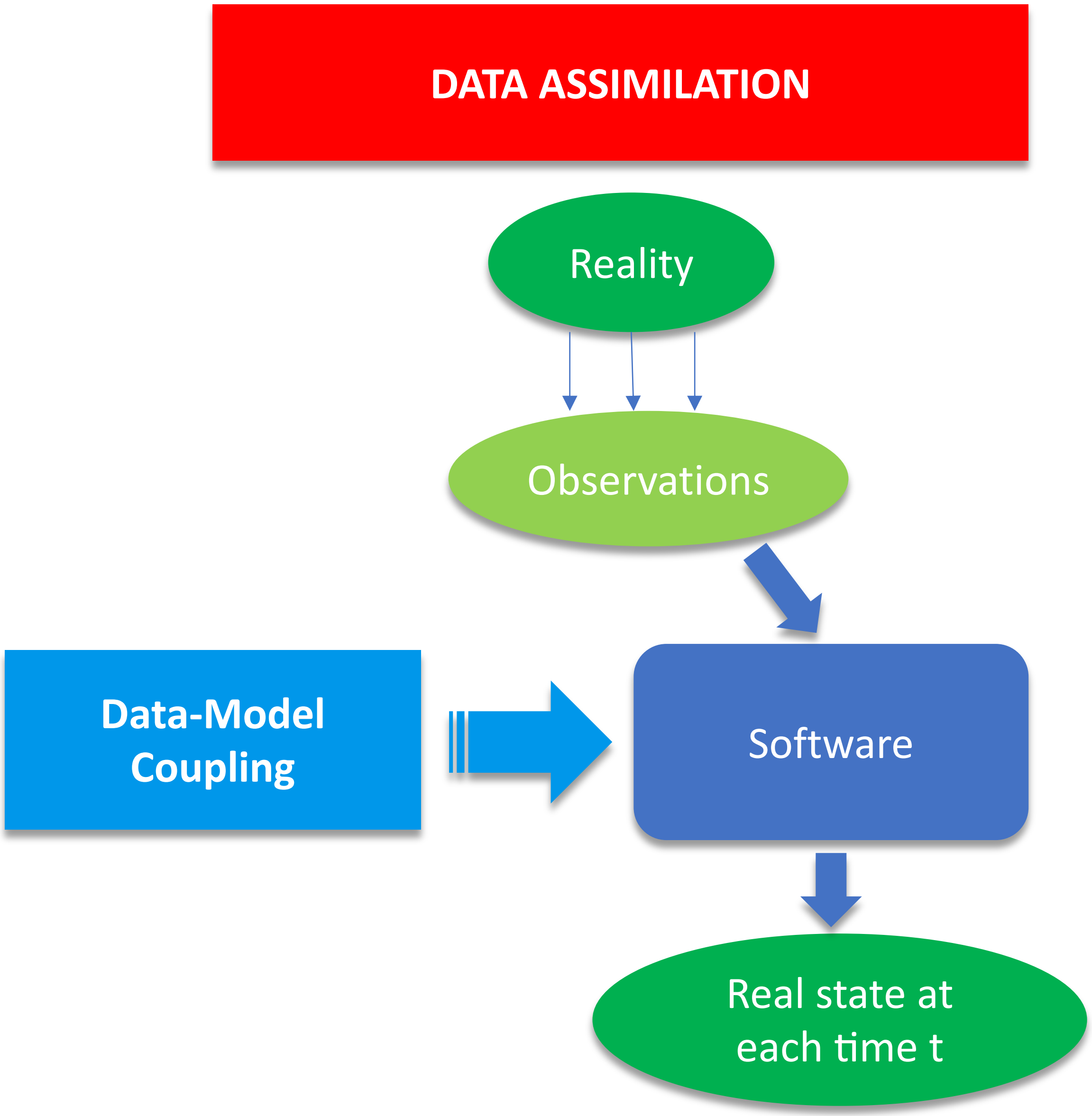


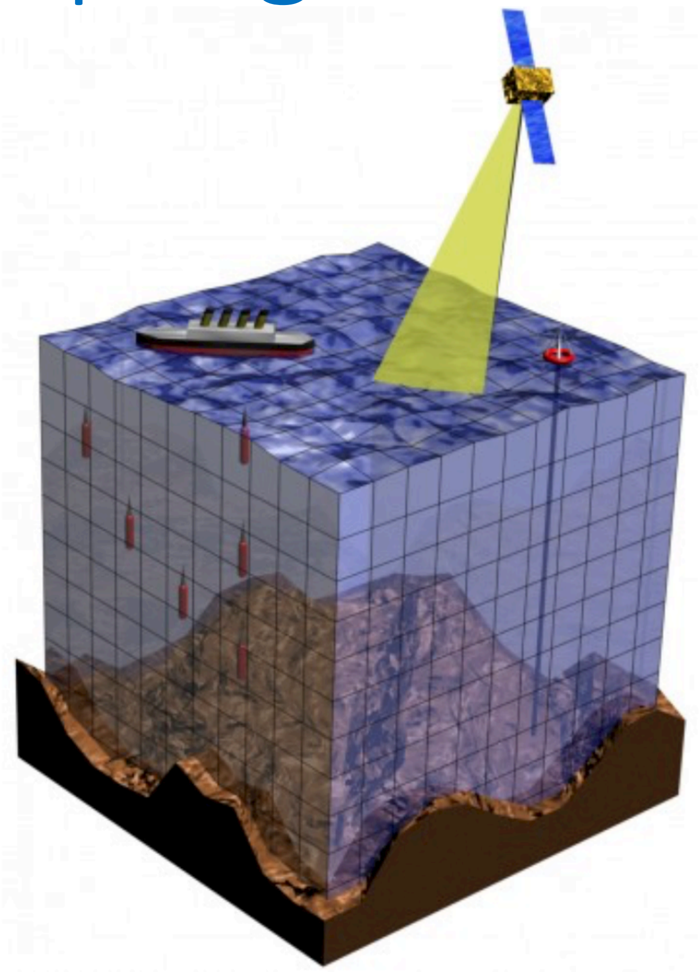
- Parameterized PDEs, ODEs …
- Database



from CERFACS

Data-Model Coupling





from CERFACS

Data-Model Coupling

Other definition of Data Assimilation :

$$\begin{cases} x(t) = M[x(t-1)] + \eta(t) \\ y(t) = H[x(t)] + \varepsilon(t) \end{cases}$$

$$\eta(t) \sim N(0, Q(t))$$

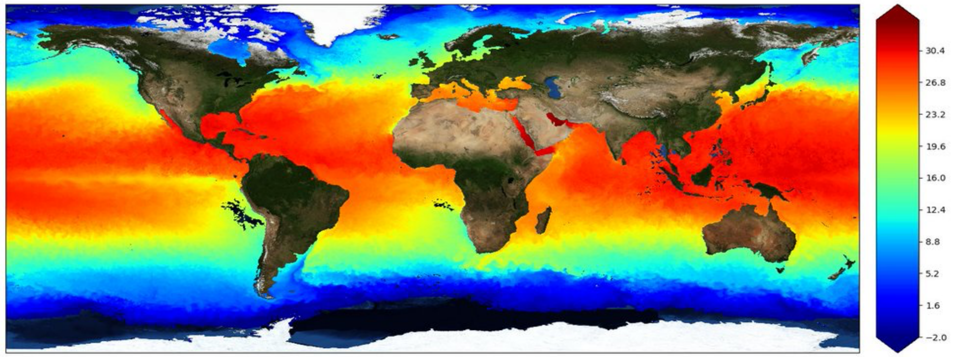
$$\varepsilon(t) \sim N(0, R(t))$$

- **t** the discrete time
- **x** the real state
- **y** the associated observation
- **M** the model
- **H** the observation operator
- **η** the model uncertainty
- **ε** the observation error

SST-Data Assimilation

- SST plays a significant role in analyzing and assessing the dynamics of weather and other biological systems.
- Various applications : weather forecasting, or planning of coastal activities.
- Weather satellites make huge quantities of very high resolution SST data available.

Sentinel 3A SLSTR sea surface temperature (S3A_SL_2_WST) - September 2016



EUMETSAT Copernicus

SST-Data Assimilation

- Standard physical methods for forecasting SST use coupled ocean-atmosphere prediction systems, based on the **Navier-Stokes equations**. These models rely on multiple physical hypotheses and do not optimally exploit the information available in the data.

- Despite the availability of large amounts of data, direct applications of **machine learning** methods do not lead to competitive state of the art results.

SST-Data Assimilation

The model of Flourent et al.

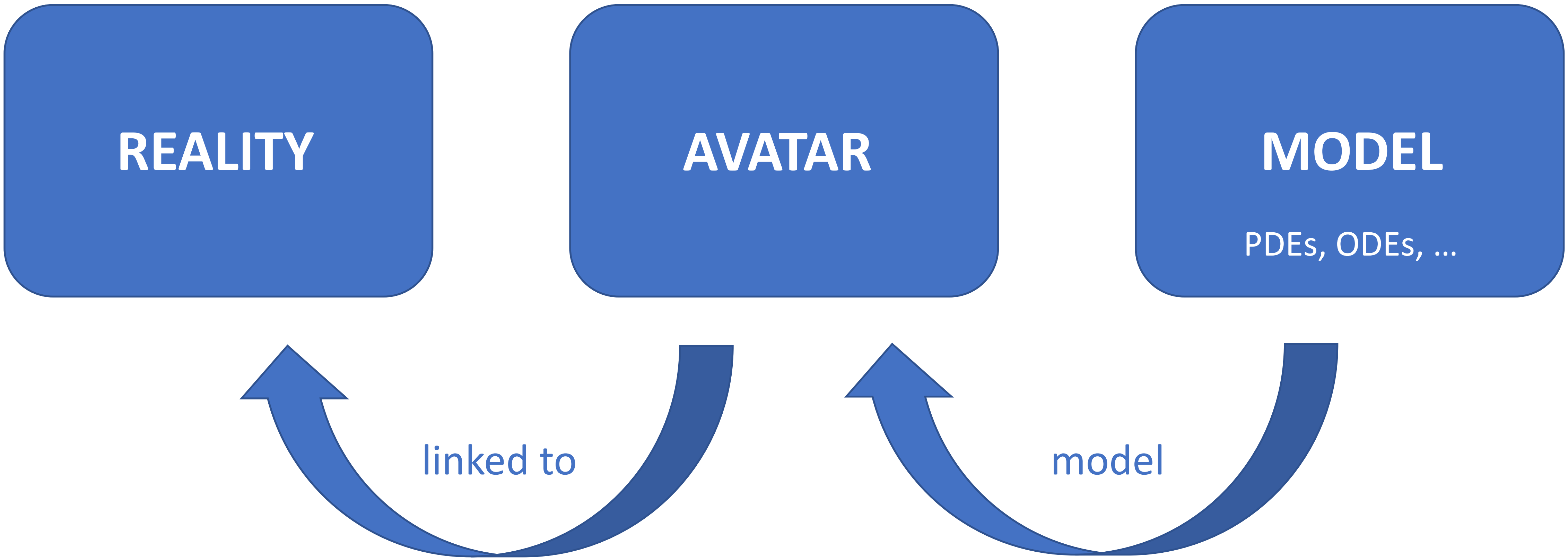
- **H. Flourent, E. Frénod & V. Sincholle** (2019). An Innovating Statistical Learning Tool Based on Partial Differential Equations, Intending Livestock Data Assimilation.

- Model for biological variables with **convection-diffusion phenomena**.

- **Data-model coupling** approach : Model based on PDEs and ODEs with parameters learnt by data.

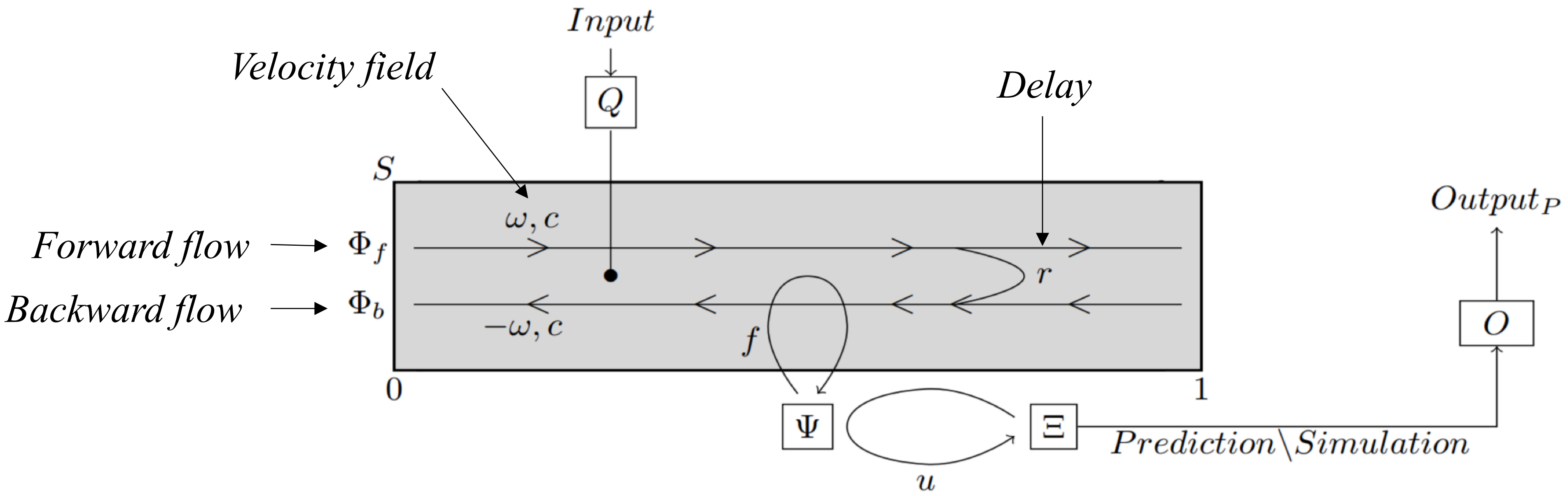
SST-Data Assimilation

The model of Flourent et al.



SST-Data Assimilation

The model of Flourent et al.



Avatar from H.Flourent et al.

SST-Data Assimilation

The model of Flourent et al.

$$\frac{\partial\{\Phi_f(\mathtt{d})\}}{\partial t}(t,x) + \omega_{\mathtt{d}}(x)\frac{\partial\{\Phi_f(\mathtt{d})\}}{\partial x}(t,x) - c_{\mathtt{d}}\frac{\partial\left[\chi\frac{\partial\left[\{\Phi_f(\mathtt{d})\}\right]}{\partial x}\right]}{\partial x}(t,x)$$
$$= \{Q(\mathtt{d})\}(t,x) - f_{\mathtt{d}}\{F(\mathtt{d})\}(x)\{\Phi_f(\mathtt{d})\}(t,x), \quad (1)$$

$$\frac{\partial\{\Psi(\mathtt{d})\}}{\partial t}(t,x) = f_{\mathtt{d}}\{F(\mathtt{d})\}(x)\{\Phi_b(\mathtt{d})\}(t,x) - u_{\mathtt{d}}\{\Psi(\mathtt{d})\}(t,x), \quad (2)$$

$$\frac{\partial\{\Xi(\mathtt{d})\}}{\partial t}(t,x) = u_{\mathtt{d}}\{\Psi(\mathtt{d})\}(t,x)\left(\frac{L_{\mathtt{d}} - \{s(\mathtt{d})\}(t)}{L_{\mathtt{d}}}\right). \quad (3)$$

$$\{s(\mathtt{d})\}(t) = \int_{\Omega(\mathtt{d})}\{\Xi(\mathtt{d})\}(t,x)\,dx, \quad (4)$$

SST-Data Assimilation

The model of Flourent et al.

$$\frac{\partial\{\Phi_f(\mathtt{d})\}}{\partial t}(t,x) + \omega_{\mathtt{d}}(x)\frac{\partial\{\Phi_f(\mathtt{d})\}}{\partial x}(t,x) - c_{\mathtt{d}}\frac{\partial\left[\chi\frac{\partial\left[\{\Phi_f(\mathtt{d})\}\right]}{\partial x}\right]}{\partial x}(t,x)$$
$$= \{Q(\mathtt{d})\}(t,x) - f_{\mathtt{d}}\{F(\mathtt{d})\}(x)\{\Phi_f(\mathtt{d})\}(t,x), \quad (1)$$

$$\frac{\partial\{\Psi(\mathtt{d})\}}{\partial t}(t,x) = f_{\mathtt{d}}\{F(\mathtt{d})\}(x)\{\Phi_b(\mathtt{d})\}(t,x) - u_{\mathtt{d}}\{\Psi(\mathtt{d})\}(t,x), \quad (2)$$

$$\frac{\partial\{\Xi(\mathtt{d})\}}{\partial t}(t,x) = u_{\mathtt{d}}\{\Psi(\mathtt{d})\}(t,x)\left(\frac{L_{\mathtt{d}} - \{s(\mathtt{d})\}(t)}{L_{\mathtt{d}}}\right). \quad (3)$$

$$\{s(\mathtt{d})\}(t) = \int_{\Omega(\mathtt{d})} \{\Xi(\mathtt{d})\}(t,x)\,dx, \quad (4)$$

SST-Data Assimilation

<u>Learning of the parameters for the model of Flourent et al.</u>

Minimisation at each time t of the objective function, using the DIRECT algorithm on R:

$$f_{obj}(\omega_{\mathtt{d}}, r_{\mathtt{d}}, f_{\mathtt{d}}, u_{\mathtt{d}}, L_{\mathtt{d}}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(s_{i_{obs}}(1) - s_{j_{pred}}(1))}{s_{i_{obs}}(1)}\right)^2$$

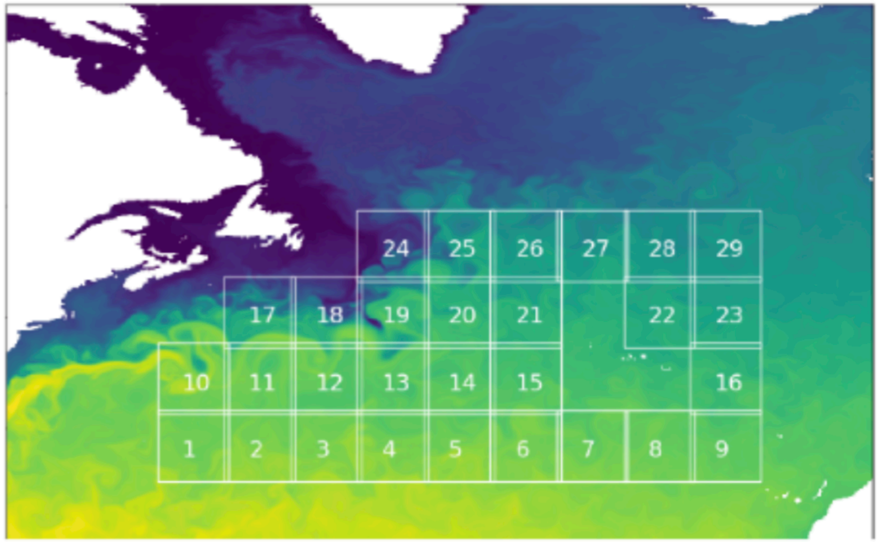
Here t=1, n is the size of the training dataset, $s_{i_{obs}}$ is the observed value, and $s_{j_{pred}}$ the predicted value.

SST-Data Assimilation

Our modified model

Goal: **simulate and predict the SST**

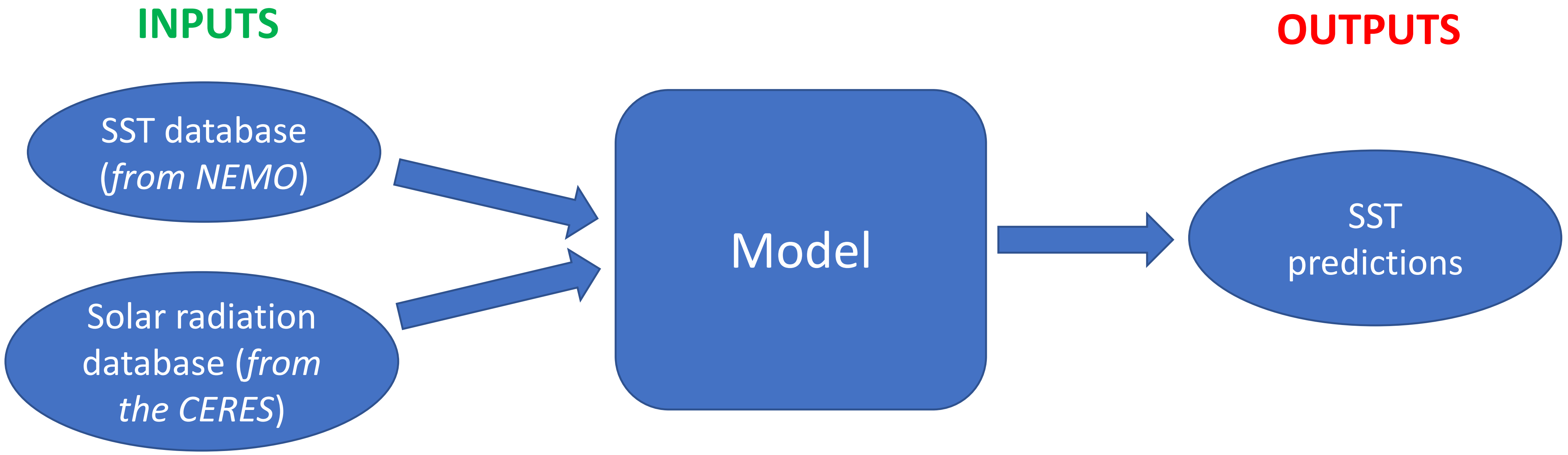
- Advection-diffusion phenomena
- Use of SST database
- Use of solar radiation (at the top of the atmosphere) database



Localisation of the dataset, from E. de Bézenac et al.

SST-Data Assimilation

Our modified model



SST-Data Assimilation

Our modified model

$$\frac{\partial \Phi_f}{\partial t}(t,x) + \omega \frac{\partial \Phi_f}{\partial x}(t,x) - c\frac{\partial \left[\chi \frac{\partial \Phi_f}{\partial x}\right]}{\partial x}(t,x) = Q(t,x) - fF(x)\Phi_f(t,x) \qquad (1)$$

$$\frac{\partial \Psi}{\partial t}(t,x) = fF(x)(t,x) - u\Psi(t,x) \qquad (2)$$

$$\frac{\partial \Xi}{\partial t}(t,x) = -u\Psi(t,x)\Big(\Xi(t,x) - sup\Big) - P\Big(\Xi(t,x) - inf\Big) \qquad (3)$$

$$s(t) = \int_{\Omega} \Xi(t,x)\,dx \qquad (4)$$

SST-Data Assimilation

Our modified model

Source term (solar radiation)

Flow

$$\frac{\partial \boxed{\Phi_f}}{\partial t}(t,x) + \boxed{\omega}\frac{\partial \Phi_f}{\partial x}(t,x) - \boxed{c}\frac{\partial\left[\chi \dfrac{\partial \Phi_f}{\partial x}\right]}{\partial x}(t,x) = \boxed{Q}(t,x) - fF(x)\Phi_f(t,x) \qquad (1)$$

Advection velocity

Diffusion coefficient

$$\frac{\partial \boxed{\Psi}}{\partial t}(t,x) = fF(x)(t,x) - u\Psi(t,x) \qquad (2)$$

Hidden variables

$$\frac{\partial \boxed{\Xi}}{\partial t}(t,x) = -u\Psi(t,x)\Big(\Xi(t,x) - sup\Big) - P\Big(\Xi(t,x) - inf\Big) \qquad (3)$$

Output

$$\boxed{s(t)} = \int_{\Omega} \Xi(t,x)\,dx \qquad (4)$$

SST-Data Assimilation

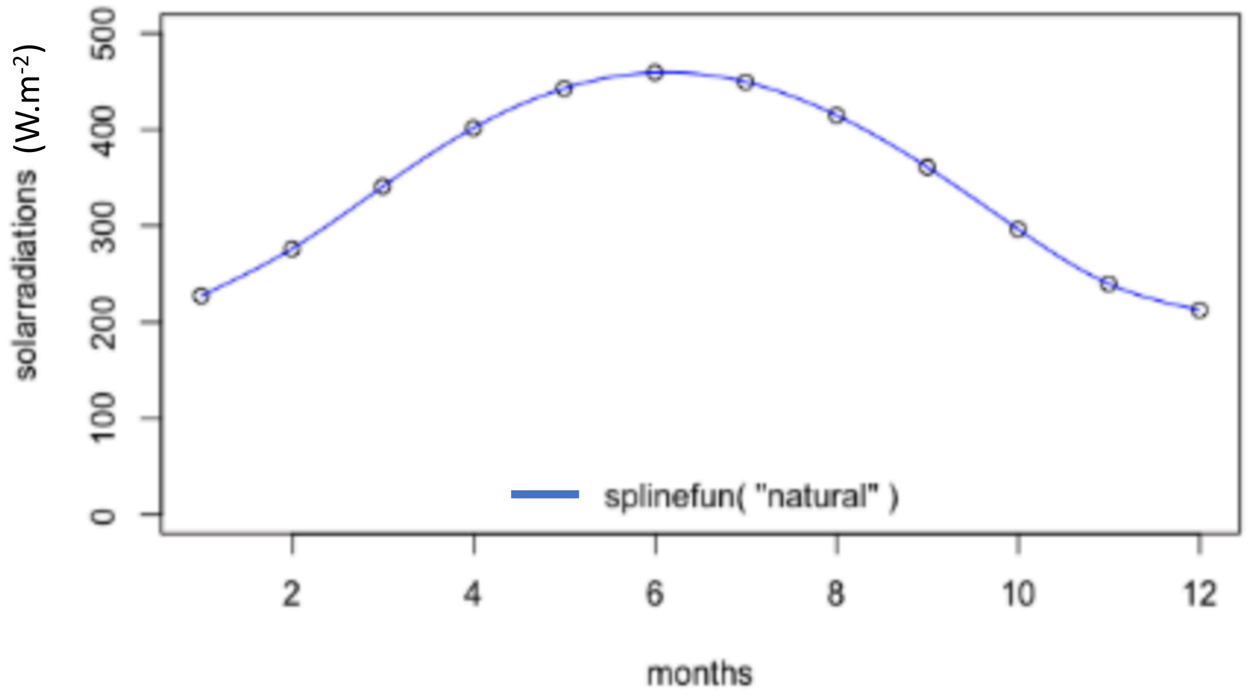
Our modified model

- We have to determine **u**, **inf** and **supp**.

- To that, we use the **same learning tool** as the previous model, using the available data.

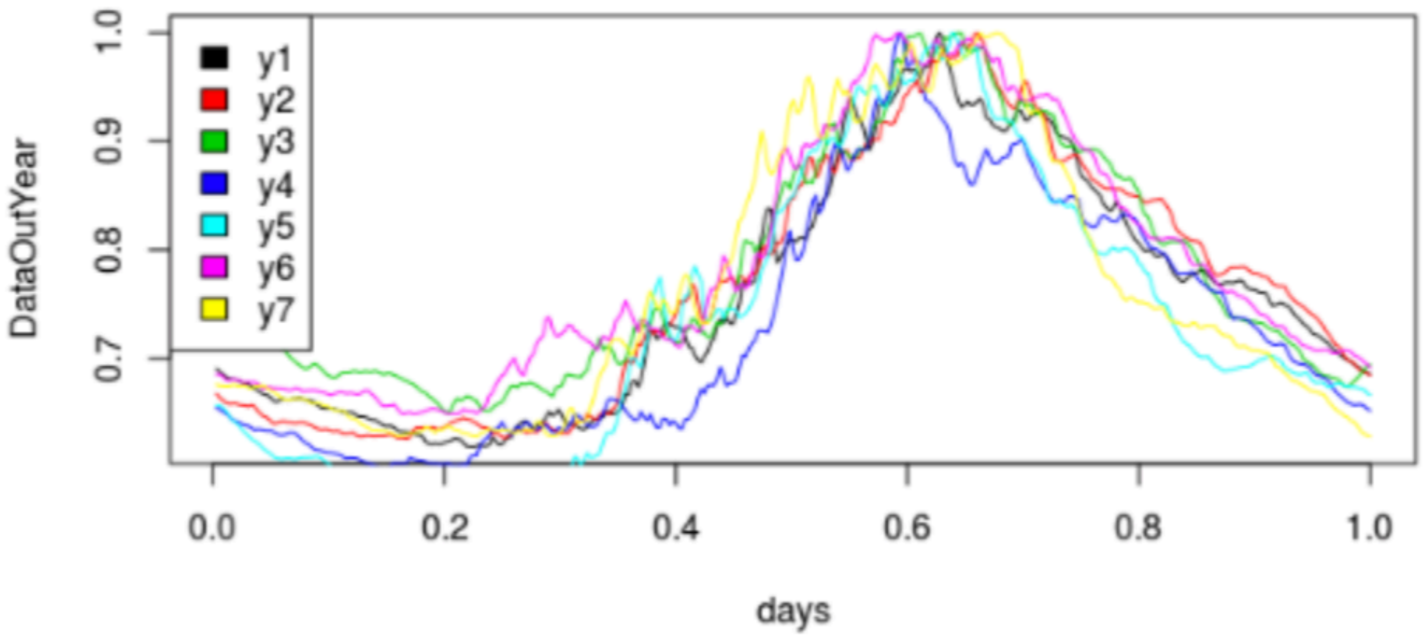
Numerical results

Solar radiation simulated by a spline



Numerical results

Real SST for different years



Numerical results

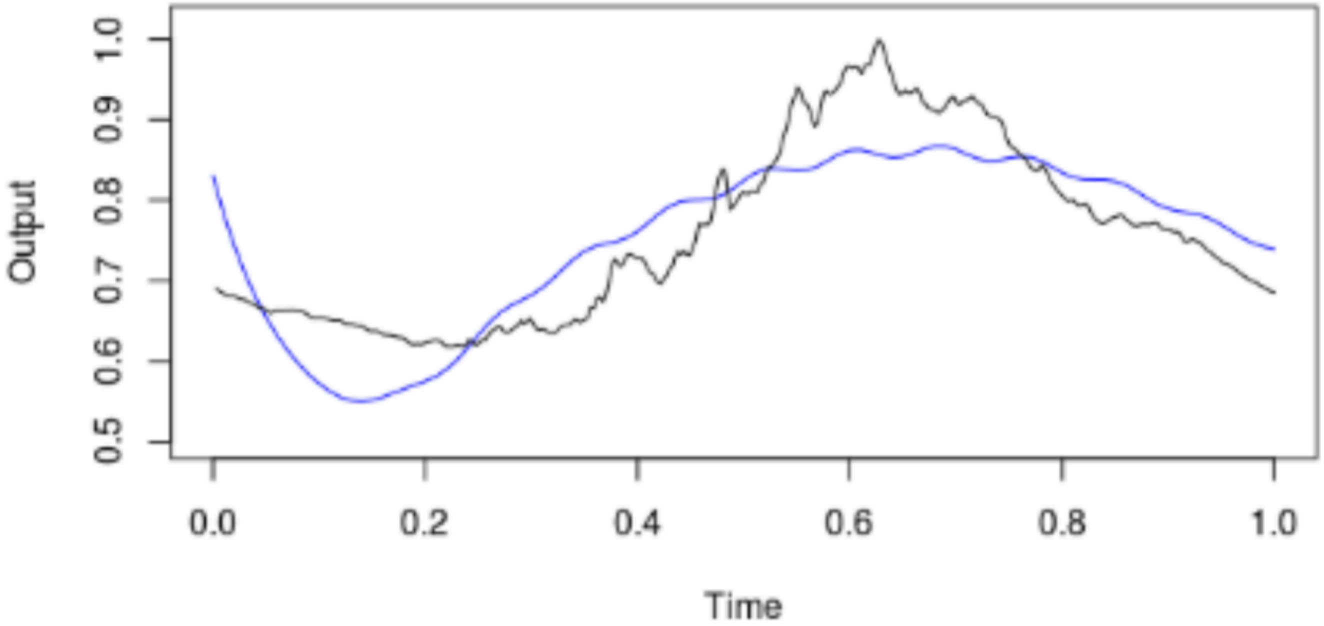
Learning of the parameters for each year

Year	sup	inf	u	obj func
1	56.6872427983148	0.5037952	199.999999999961	0.0488598607471862
2	68.4254272816738	0.4976314	199.999993030828	0.0461400247339915
3	51.7736625511663	0.5224552	199.999999999882	0.0579528378552093
4	62.5269300303108	0.4888889	199.999999999908	0.0554632691635029
5	68.8096166836958	0.4444428	199.999811832358	0.0640272984624972
6	52.5925925925795	0.5375095	199.999999999961	0.0575910849643628
7	54.1394604472804	0.5199167	199.999999999961	0.0733631601644184

- That indicates the stability of the parameters

Numerical results

Our modified model (in blue) compared to real SST of year 1 (in black)

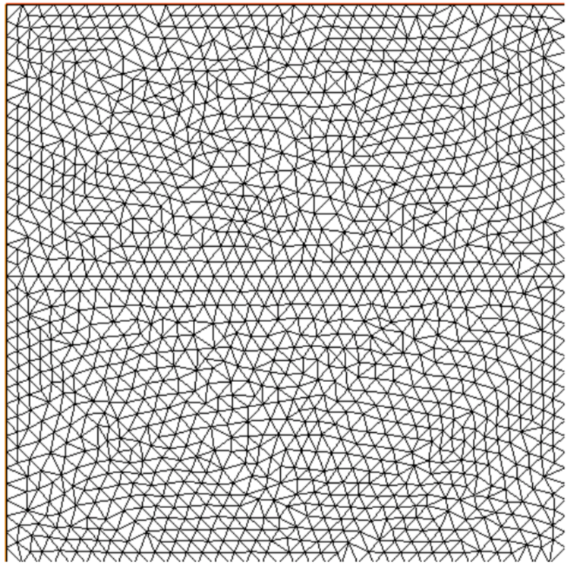


- Same seasonal trend
- No high frequency variations in our model

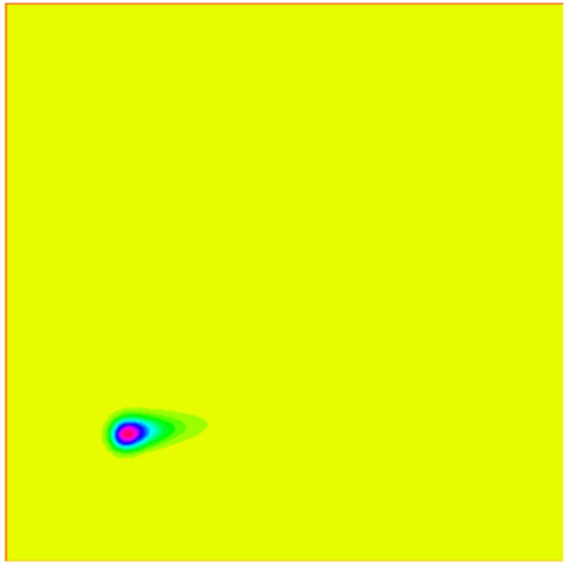
Numerical results

2D-extension of the model :

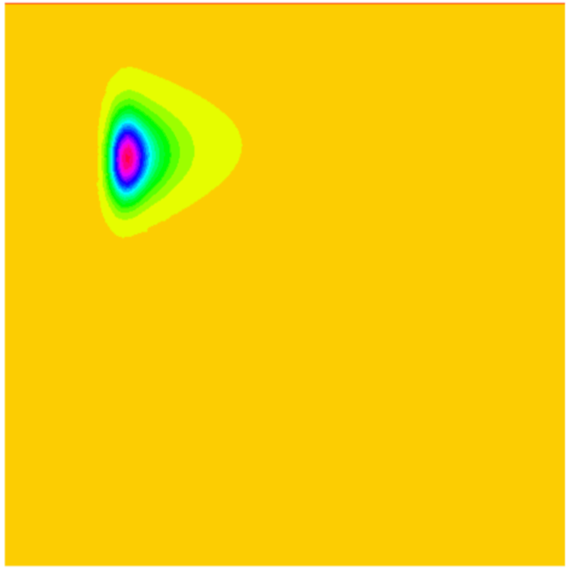
Grid



Initial condition



Result



Simulation of the advection-diffusion phenomena, using a Finite Element Method on Freefem++

Numerical results

Which modeling software ?

FREEFEM

- ✓ Fast
- ✓ Easy to use
- ✗ Hard to interface with structured data
- ✗ Hard to interface with higher level language (e.g Python)

FEniCS PROJECT

- ✓ Python compatible
- ✓ Easy interfacing with structured data
- ✓ Wide variety of tutorials
- ✗ Slower

Conclusion

- We adapted a biological model to simulate SST, with the same underlying phenomena, using the available data.
- Some difficulties to well simulate the diffusion phenomenon with Freefem++.

<u>To go on with our model:</u>

- ➢ Simulating the advection-diffusion phenomena using Galerkin discontinuous functions on Fenics.
- ➢ Using a Neural Network instead of the learning tool of the model of Flourent et al.

References

- **L. Bertino, M. Bocquet, J. Brajard, A. Carrassi** (2019). Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models.
- **E. de Bézenac, P. Gallinari, A. Pajot** (2018). Deep Learning for Physical Processes: Incorporating Prior Scientific Knowledge.
- **S. Brunton, J. Kutz, J. Proctor** (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems.

References

- **H. Flourent, E. Frénod, V. Sincholle** (2019). An Innovating Statistical Learning Tool Based on Partial Differential Equations, Intending Livestock Data Assimilation.
- **G. Lin, S. Zhang** (2018). Robust data-driven discovery of governing physical laws with error bars.
- **R. Monson, W. Sacks, D. Schimel** (2006). Coupling between carbon cycling and climate in a high-elevation, subalpine forest: a model-data fusion analysis.

Thank you for your attention !