The analog data assimilation: method, applications and implementation

Pierre Tandeo
pierre.tandeo@imt-atlantique.fr

IMT-Atlantique, Brest, France

16th of February 2018

RIKEN AICS 理化学研究所 計算科学研究機構
RIKEN Advanced Institute for Computational Science

Before starting

Works in collaboration with:

P. Ailliot	R. Lguensat	R. Fablet	M. Pulido	J. Ruiz
Univ. Brest	Univ. Grenoble	IMT-Atlantique	Univ. Reading	Univ. Buenos Aires
France	France	France	UK	Argentina

Who am I:

Brest, France

Main oceanographic institute in France:

Ifremer

My engineering school:

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

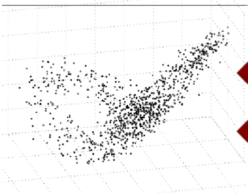
- Master in applied mathematics
- PhD in spatial oceanography (IFREMER, France)
- Postdoc in atmospheric sciences (Univ. Corrientes, Argentina)
- Postdoc and Associate Professor (IMT-Atlantique, France)

General Concept

Classic Data Assimilation

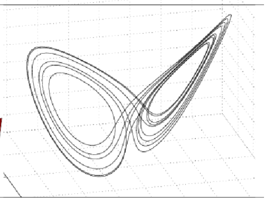
$$\frac{dx_1}{dt} = \sigma(x_2 - x_1)$$
$$\frac{dx_2}{dt} = x_1(\rho - x_3) - x_2$$
$$\frac{dx_3}{dt} = x_1 x_2 - \beta x_3$$

→ model-driven
→ model integration needed
→ high computational cost

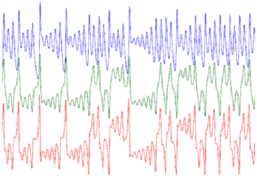


Equations

Catalog



→ data-driven
→ statistical emulator
→ low computational cost



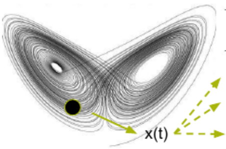
Analog Data Assimilation

Outline

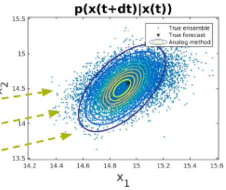
Analog forecasting (general formulation)

- State of the art:
  - introduced by [Lorenz, 1969]
  - destroyed by [Van Den Dool, 1994]
  - revival (large datasets + machine learning) by [Zhao and Giannakis, 2016]

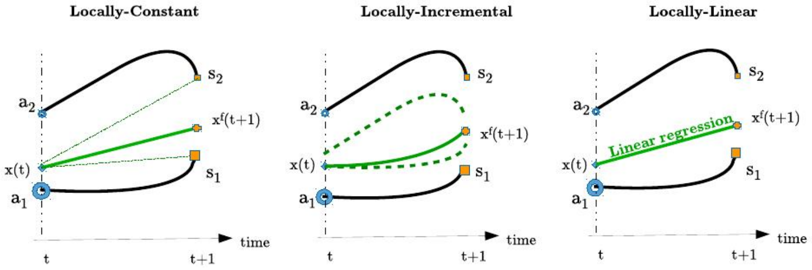
- Analog forecasting:
  - use historical datasets (observations, simulations)
  - use the k-nearest neighbors
  - emulate the dynamical model



Analogs (a_k): x_1,x_2,x_3 (time t)	Successors (s_k): x_1,x_2,x_3 (time $t+\mathrm{d}t$)
$(-0.33, +3.26, +25.51)$	$(+0.01, +3.23, +24.84)$
$(+0.01, +3.23, +24.84)$	$(+0.32, +3.20, +24.19)$
...	...
$(-2.76, -4.50, +19.18)$	$(-2.93, -4.71, +18.80)$
$(-2.93, -4.71, +18.80)$	$(-3.11, -4.95, +18.45)$

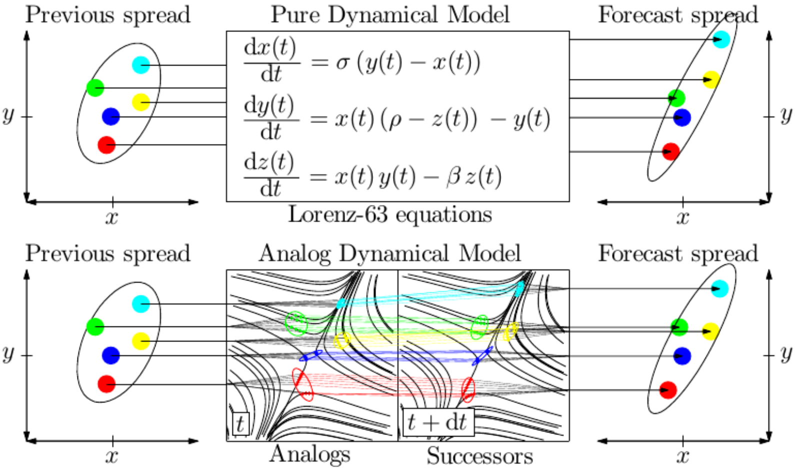


Analog forecasting (different implementations)



- Weighted mean of the successors (locally constant):
  $x^f(t+1) = \sum_{k=1}^{K} \omega_k[x(t)] s_k[x(t)]$
- Weighted mean of the increments (locally incremental):
  $x^f(t+1) = x(t) + \sum_{k=1}^{K} \omega_k[x(t)] \left(s_k[x(t)] - a_k[x(t)]\right)$
- Regression between analogs and successors (locally linear):
  $x^f(t+1) = \beta_0[x(t)] + \beta_1[x(t)] x(t)$

Analog data assimilation (model-driven VS data-driven)



$\Rightarrow$ Share the same sequential framework (Kalman/particle filter)

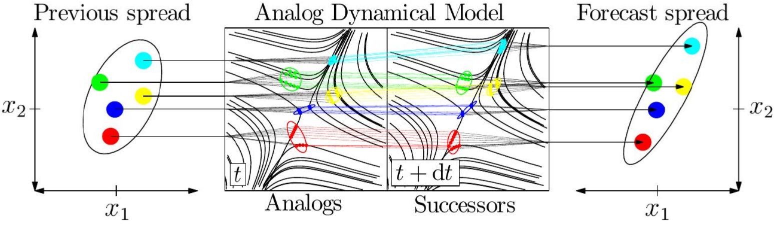
Analog data assimilation (general implementation)

Nonlinear state-space model (analog forecasting operator $\mathcal{A}$):

$$\mathbf{x}(t) = \mathcal{A}\left(\mathbf{x}(t - \mathrm{d}t), \boldsymbol{\eta}(t)\right) \quad (1)$$

$$\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) + \boldsymbol{\epsilon}(t) \quad (2)$$

Sequential implementation:



Step 0:

- start from last analysis step (time t)
- N members with same weights (EnKF)
- N particles with different weights (PF)

Step 1:

- find K nearest analogs (a_1, ..., a_K)
- compute their weights (w_1, ..., w_K)
- use appropriate distance/kernel

Step 2:

- combine the K successors (s_1, ..., s_K) using (w_1, ..., w_K)
- different analog forecasting strategies (constant, incremental, linear)

Step 3:

- different sampling schemes (Gaussian, multinomial)
- compare forecasts to noisy observations
- recombine N members (EnKF) or particles (PF)

t ← t+1

Experimental settings (model-driven VS data-driven)

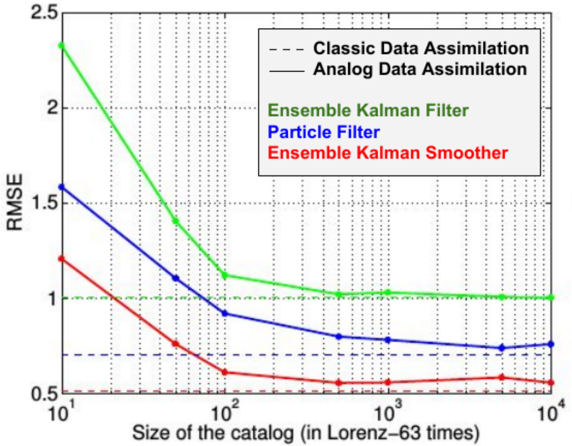


Figure 1: Effect of filtering method and catalog size on state reconstruction

- simulated data (Lorenz-63)
- 1 obs. variable (x_1 with R=2)
- partial obs. (8 time steps)

$\Rightarrow$ Equivalence for large enough catalog size

Experimental settings (model evidence)

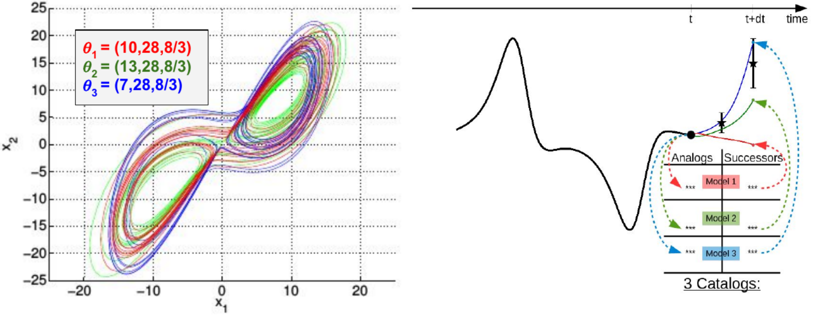


Figure 2: Scheme of the Analog Data Assimilation with various catalogs

- same previous experiment
- 3 catalogs with different parameters (θ_1, θ_2, θ_3)
- obs. generated with θ_1

$\Rightarrow$ Able to retrieve the good parameterization:
θ_1 (61%), θ_2 (27%), θ_3 (12%)

Experimental settings (global VS local analogs)

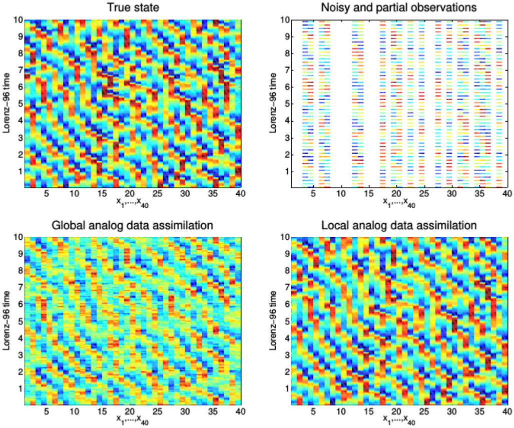


Figure 3: Effect of local and global analogs on state reconstruction

- simulated data (Lorenz-96)
- 20 obs. variables (noise R=2)
- partial obs. (4 time steps)

$\Rightarrow$ Local analog strategy outperforms the global one

Applications (use of satellite historical datasets)

- Daily and mesoscale datasets
- Synergy between satellite sources
- Large number of already seen situations (e.g., eddy motion)

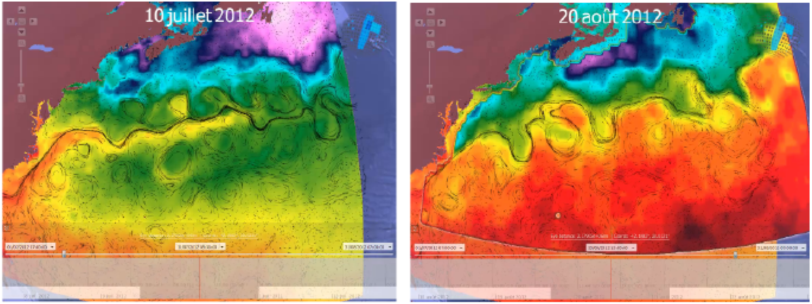


Figure 4: Surface observations of temperature (left, 40 years) or salinity (right, 10 years) with oceanic currents (both, 25 years). Full animation: `https://www.youtube.com/watch?v=Wn5grSFPQFA`.

Applications (results exploiting satellite historical datasets)

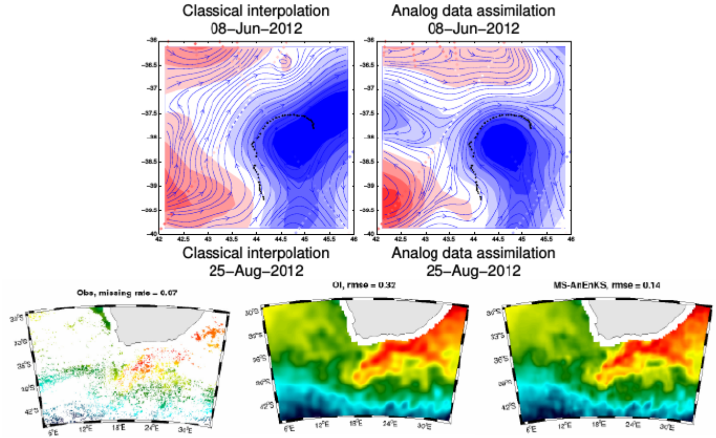


Figure 5: Comparison between classic interpolation and interpolation based on analogs for ocean currents (top) and temperatures (bottom).

$\Rightarrow$ Analog-based interpolations learn adaptively: advection, diffusion and spatial correlation lengths.

Applications (use of ensemble simulations)

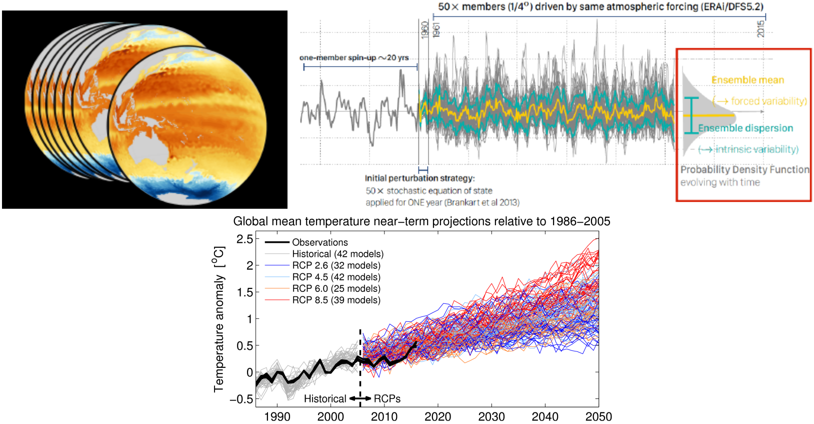
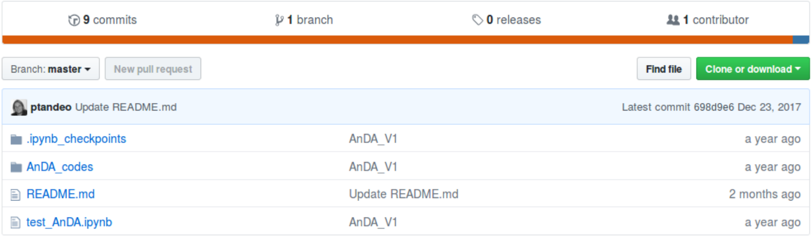


Figure 6: Examples of ensemble simulations in oceanography (OCCIPUT dataset, top) and in climate (CMIP dataset, bottom).

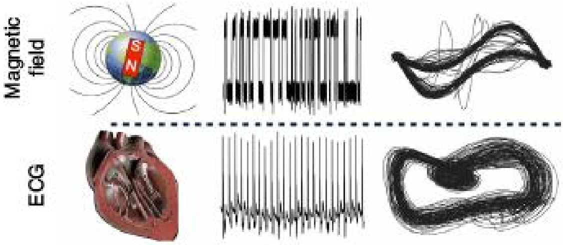
⇒ 2 ongoing projects in oceanography (comparison with classic data assimilation) and meteorology/climate (model evidence)

Conclusions

- Data-driven assimilation
  (exploit historical datasets: observations, simulations)
- Various implementations
  (global/local analogs, AnEnKF/AnPF/AnEnKS)
- Easy, fast and flexible, especially for local/partial analysis
  (compared to model-driven data assimilation)
- Python library on GitHub
  (`https://github.com/ptandeo/AnDA`)



Perspectives



- Methodology:
  - transform raw data into attractor (see [Brunton et al., 2017])
  - automatic distance learning
  - other possible forecast operators (e.g., neural nets)
- Applications:
  - complex dynamical systems
    (ecology, meteorology/climate, medicine, etc...)
  - no available models and lot of data
- Collaborations are welcome!

AnDA References

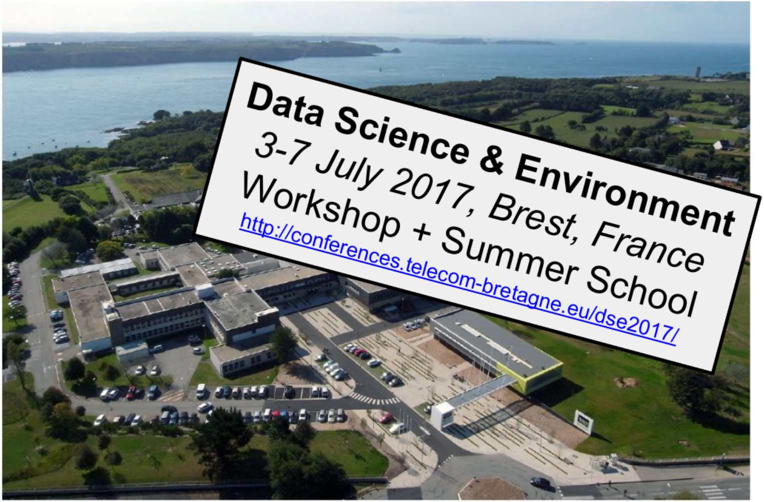
Methodological publications:

- Lguensat et al. 2017, **The Analog Data Assimilation**, *Monthly Weather Review*
- Tandeo et al. 2015, **Combining Analog Method and Ensemble Data Assimilation: Application to the Lorenz-63 Chaotic System**, *Machine Learning and Data Mining Approaches to Climate Science (Springer)*
- Tandeo et al. 2014, **The Analog Ensemble Kalman Filter and Smoother**, *"Climate Informatics" workshop, Boulder (CO)*

Application publications:

- Fablet et al. 2017, **Data-driven Methods for Spatio-Temporal Interpolation of Sea Surface Temperature Images**, *IEEE Transactions on Computational Imaging*
- Tandeo et al. 2015, **The Analog Data Assimilation: Application to 20 Years of Altimetric Data**, *"Climate Informatics" workshop, Boulder (CO)*

IMT-Atlantique: Birthplace of AnDA



Thank you for your attention!
Any questions?

pierre.tandeo@imt-atlantique.fr

Cited references

Brunton, S., Brunton, B., Proctor, J., Kaiser, E., and Kutz, J. (2017).
Chaos as an intermittently forced linear system.
Nature communications, 8(1):19.

Lorenz, E. N. (1969).
Atmospheric predictability as revealed by naturally occurring analogues.
Journal of the Atmospheric Sciences, 26(7):636–646.

Van Den Dool, H. M. (1994).
Searching for analogues, how long must we wait?
Tellus A, 46(3):314–324.

Zhao, Z. and Giannakis, D. (2016).
Analog forecasting with dynamics-adapted kernels.
Nonlinearity, 29(9):2888–2939.