Selection of dynamic model using analog data assimilation

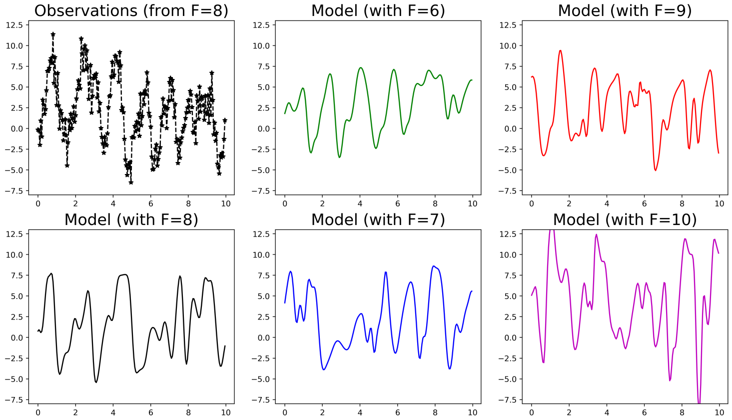
P. Ailliot[1,6], T.T.T. Chau[2], V. Monbet[3], P. Naveau[2,4], J. Ruiz[5], F. Sévellec[1,4], P. Tandeo[6]

Univ. Brest, France[1]
LSCE-IPSL, France[2]
Univ. Rennes I, France[3]
CNRS[4]
Univ. Buenos Aires, Argentina[5]
IMT Atlantique, France[6]

The Second IMT-Atlantique & RIKEN Joint Workshop:
"Statistical Modeling and Machine Learning in Meteorology and Oceanography"
February 10-11, 2020
Brest, France

Context, notation and goal

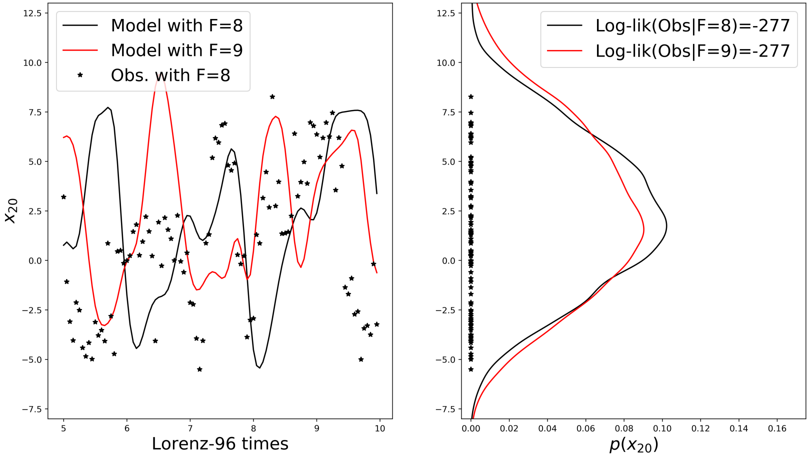
- Given a set of observations $\mathbf{y}$
- And p different dynamic models $\{\mathcal{M}_{(i)}\}_{i=1,\dots,p}$
- With independent realizations $\{\mathbf{x}_{(i)}\}_{i=1,\dots,p}$



- Here, we use the Lorenz-96 and different forcing terms F
- **How can we say which model "match" the observations?**

Solution 1: comparing climatological distributions

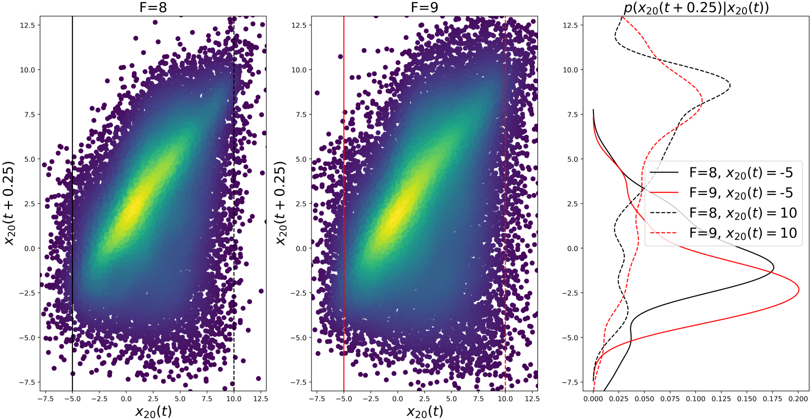
- Compare climatological (marginal) distributions
- $p\left(\mathbf{x}_{(i)}\right)$ $\underline{\text{VS}}$ $p\left(\mathbf{y}\right),\ \forall\ i = 1, \ldots, p$



- Work well if bias or different range of values
- Unable to detect models that are closed to the observations

Solution 2: comparing model dynamics

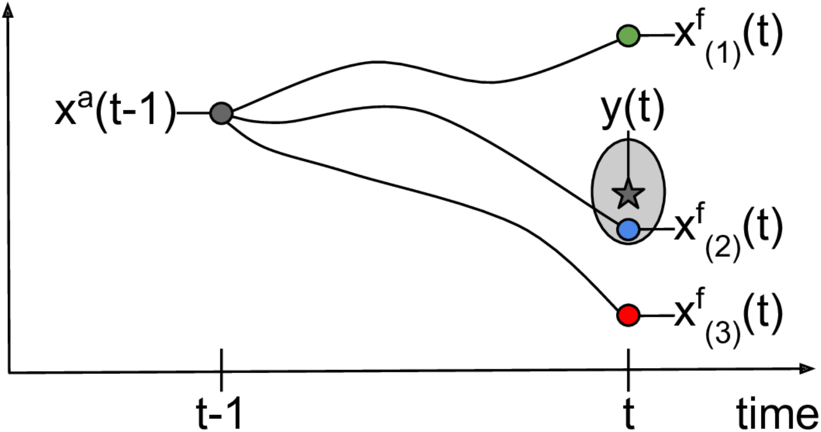
- Compare conditional distributions between consecutive times
- $p\left(\mathbf{x}_{(i)}(t)|\mathbf{x}_{(i)}(t-1)\right)$ <u>VS</u> $p\left(\mathbf{y}(t)\right)$, $\forall\ i = 1, \ldots, p$



- Dynamics are different depending on the models
- Differences appear in the extreme values
- Solution 2 (conditional) is preferred to solution 1 (marginal)

Comparing model dynamics using data assimilation

- Need to start from the best initial condition
- Need to deal with observation uncertainties
- Data Assimilation (DA) is the perfect candidate



- The idea is to evaluate, at each assimilation cycle, $p\left(\mathbf{x}^f_{(i)}(t)|\mathbf{x}^a(t-1)\right)$ <u>VS</u> $p\left(\mathbf{y}(t)\right)$, $\forall\ i = 1, \ldots, p$

Computing model evidence in DA

- Find a metric to compare model dynamics
- Contextual Model Evidence (CME) is a possible one
- Introduced in DA by [Carrassi et al., 2017, Metref et al., 2019]

In the nonlinear and Gaussian DA case:

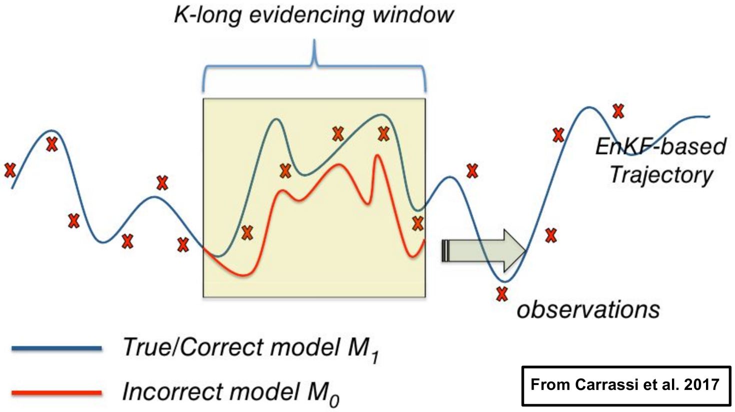
$$\mathrm{CME}_{(i)} = \prod_{t=1}^{T} \mathcal{L}\left(\mathbf{y}(t)|\mathcal{M}_{(i)}\right) \tag{1}$$

with the innovation likelihood given by:

$$\mathcal{L}\left(\mathbf{y}(t)|\mathcal{M}_{(i)}\right) \propto \exp\left(-\mathbf{d}_{(i)}(t)^{\top}\mathbf{\Sigma}_{(i)}(t)^{-1}\mathbf{d}(t)\right) \tag{2}$$

where $\mathbf{d}_{(i)}(t) = \mathbf{y}(t) - \mathbf{H}\mathbf{x}^{f}_{(i)}(t)$ and $\mathbf{\Sigma}_{(i)}(t) = \mathbf{H}\mathbf{P}^{f}_{(i)}(t)\mathbf{H}^{\top} + \mathbf{R}$

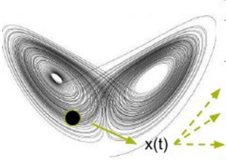
Pros and cons of model evidence in DA



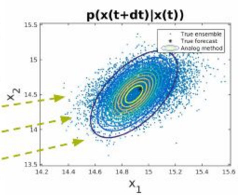
- Pros of CME in DA:
  - use observation ($\mathbf{R}$) and model error ($\mathbf{P}^f_{(i)}$) covariances
  - easy to compute at each DA cycle
- Cons of CME in DA:
  - need to run $\mathcal{M}_{(i)}$ to get $\mathbf{x}^f_{(i)}(t)$ and $\mathbf{P}^f_{(i)}(t)$, $\forall i$ and $\forall t$
  - need to run global model (potentially large)

Getting $p\left(\mathbf{x}_{(i)}(t) | \mathbf{x}^a(t-1)\right)$ using analogs

- Instead of running a model $\mathcal{M}_{(i)}$, use analog forecasting
- Analog forecasts naturally capture $\mathbf{x}^f_{(i)}$ and $\mathbf{P}^f_{(i)}$



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

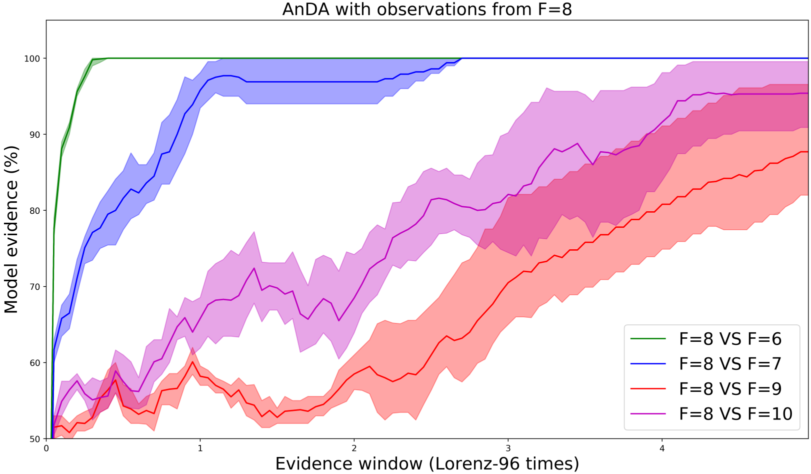


Catalog
(historical observations or numerical simulations)

- Analog forecasting can be easily plugged into DA algorithms
- The Analog Data Assimilation (AnDA) [Tandeo et al., 2015, Lguensat et al., 2017]
- Other forecasting methods can be considered (e.g., neural nets, kernel methods)

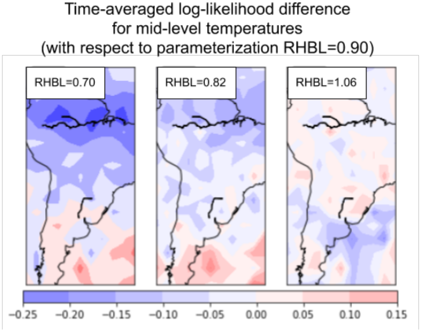
The computation of model evidence using AnDA

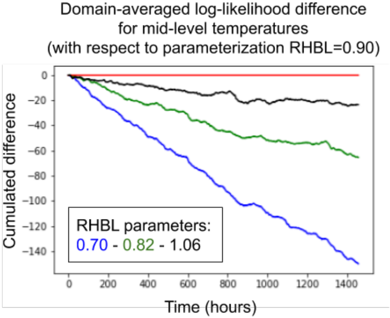
- Need sufficient catalog size to get good performance
- Results similar as the true DA (using model integration)
- Details given in [Chau, 2019]



The interest of AnDA: the locality

- AnDA can be applied to a part of the state
- Thus, AnDA is able to compute CME locally

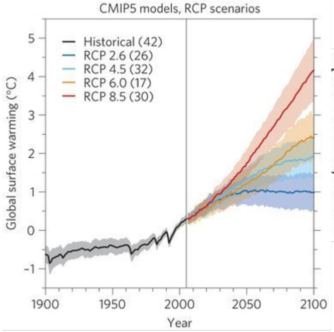




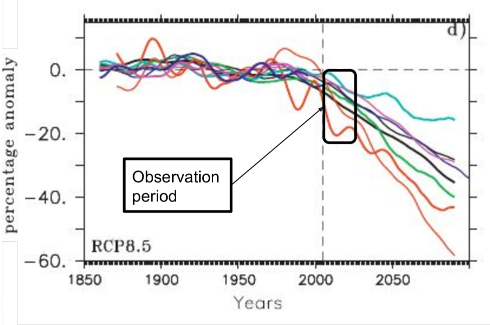
- Has been tested on a simple climate model (SPEEDY)
- 30 years or catalogs with different parameterizations
- Relative Humidity threshold in the Boundary Layer (RHBL = 0.70, 0.82, 0.90, 1.06)

Next step: application to climate simulations

- CMIP contains climate simulation runs for the future
- Different models (20) and scenarios (4) are considered
- For each scenario, each model has several members



RCP scenarios in CMIP simulations



Atlantic Meridional Overturning Circulation (AMOC) simulations from different climate models

- Goal 1 → create weighted projections of climate metrics
- Goal 2 → reduce the uncertainty of climate projections
- Data → compare current observations to climate simulations
- Method → use AnDA and the model evidence metric

Thank you for your attention!

Carrassi, A., Bocquet, M., Hannart, A., and Ghil, M. (2017).
Estimating model evidence using data assimilation.
Quarterly Journal of the Royal Meteorological Society, 143(703):866–880.

Chau, T. T. T. (2019).
Non-parametric methodologies for reconstruction and estimation in nonlinear state-space
PhD thesis.

Lguensat, R., Tandeo, P., Ailliot, P., Pulido, M., and Fablet, R. (2017).
The Analog Data Assimilation.
Monthly Weather Review, 145(10):4093–4107.

Metref, S., Hannart, A., Ruiz, J., Bocquet, M., Carrassi, A., and Ghil, M. (2019).
Estimating model evidence using ensemble-based data assimilation with localization–The model selection problem.
Quarterly Journal of the Royal Meteorological Society, 145(721):1571–1588.

Tandeo, P., Ailliot, P., Ruiz, J. J., Hannart, A., Chapron, B., Easton, R., and Fablet, R. (2015).
Combining analog method and ensemble data assimilation: application to the Lorenz-63 chaotic system.
In Machine Learning and Data Mining Approaches to Climate Science, pages 3–12.