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

Data-driven methods in geophysics

Pierre Tandeo
RIKEN visitor, June-July 2018

Outline

I. **Covariance estimation in data assimilation**
   A. Nonlinear and Gaussian state-space model
   B. What are Q and R covariances?
   C. Timeline of estimation of Q and R
   D. Methods comparison: preliminary results
   E. Summary about this review paper

II. **A bunch of data-driven methods**
   A. AnDA applied to spatial oceanography
   B. Deep learning on SAR images
   C. Predictability of rogue waves
   D. Machine learning for satellite data
   E. Machine learning for solar energy prediction

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

Covariance estimation in data assimilation

Nonlinear and Gaussian state-space model

The most popular formulation in DA is:

$$\begin{cases} \mathbf{x}(k) = \mathscr{M}\left(k-1, \mathbf{x}(k-1)\right) + \eta(k), \\ \mathbf{y}(k) = \mathscr{H}\left(k, \mathbf{x}(k)\right) + \varepsilon(k), \end{cases}$$

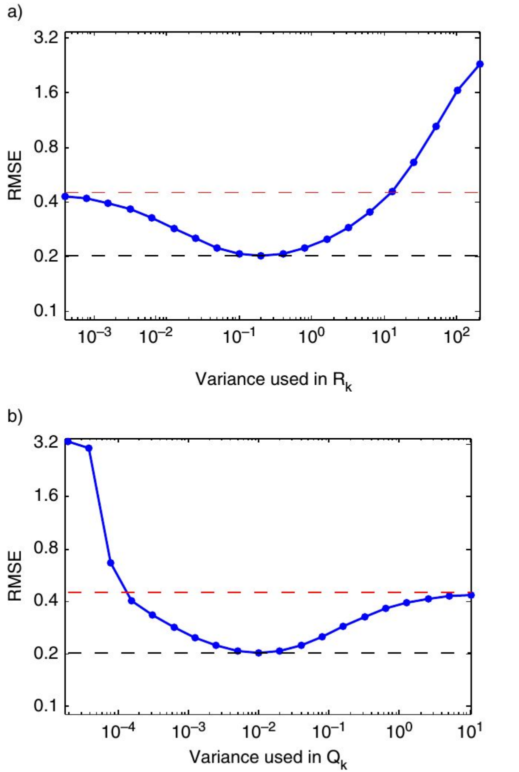
With important unknown covariance matrices **Q** and **R**:

$$\eta(k) \sim \mathscr{N}\left(\mathbf{0}, \mathbf{Q}(k)\right)$$
$$\varepsilon(k) \sim \mathscr{N}\left(\mathbf{0}, \mathbf{R}(k)\right)$$

→ Reconstruction of **x** highly depends **Q** and **R**

→ Joint Estimation of **Q** and **R** is necessary

→ Many authors worked on this topic but...



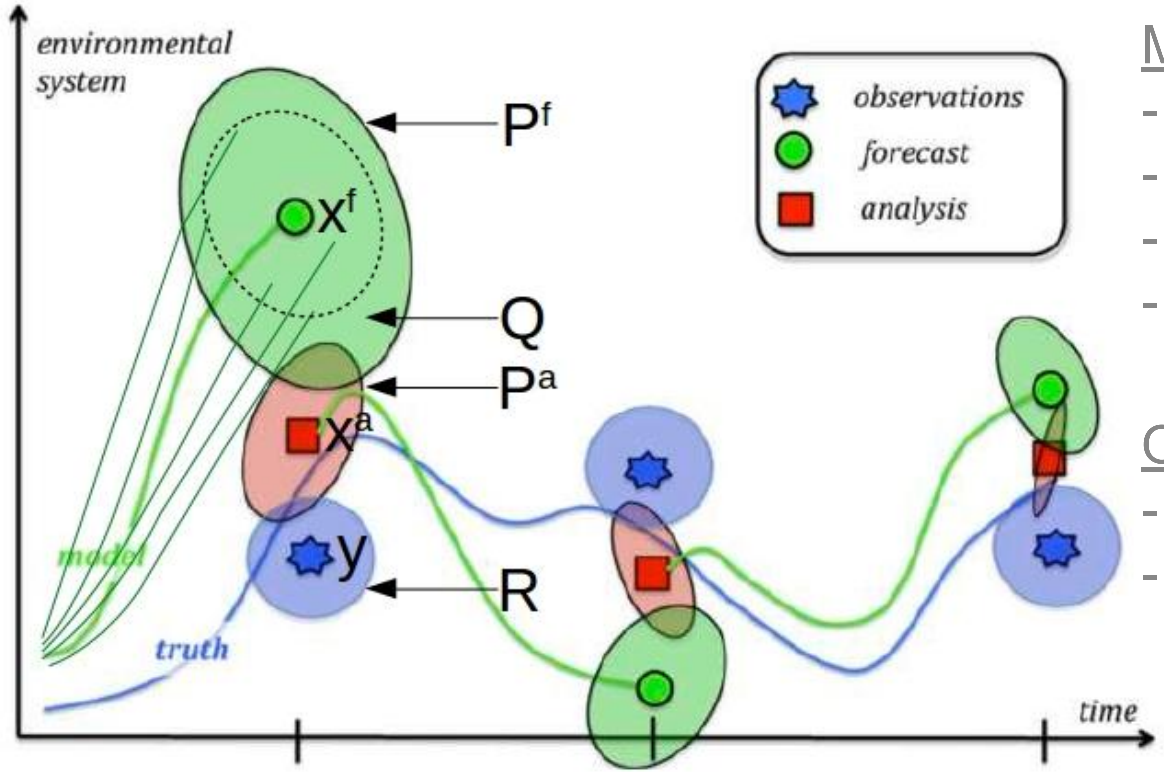
Simulation results on the L96
Berry & Sauer 2013 [Tellus A]

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

Covariance estimation in data assimilation

What are Q and R covariances?

Scheme of the sequential DA:



Model error **Q** represents:

- model deficiencies
- errors in the parameters
- unresolved scales
- errors in numerical schemes

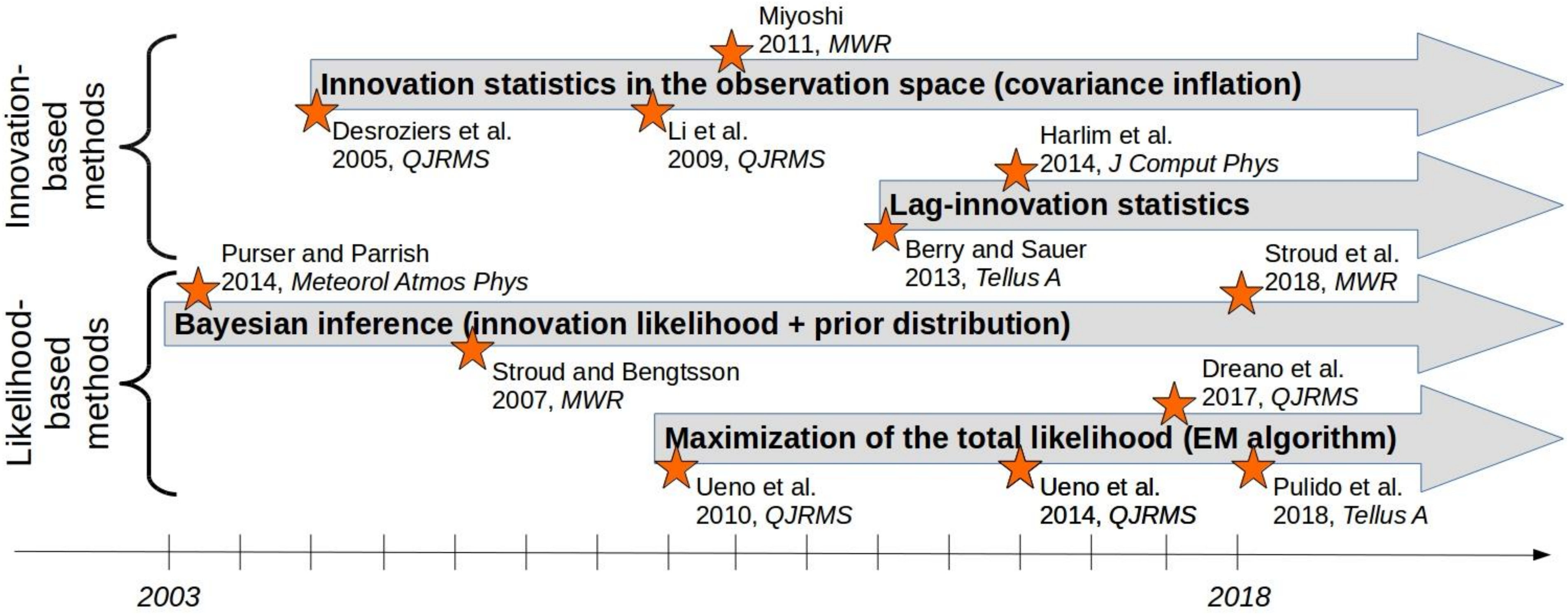
Observation error **R** represents:

- instrument noise
- representativeness error

Covariance estimation in data assimilation

Timeline of estimation of Q and R

Research field initiated by Daley et al. 1992 [MWR] & Dee 1995 [MWR]



Covariance estimation in data assimilation

Methods comparison: preliminary results

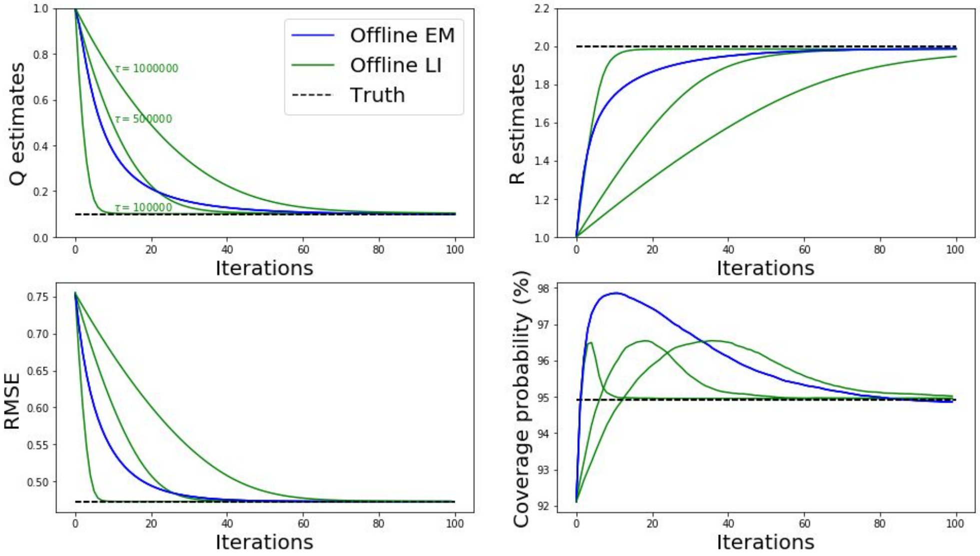
<u>Focus on 2 methods:</u>

- Lag-Innovation (LI) method
- Expectation-Maximization (EM) algorithm
- robust and accurate in practice

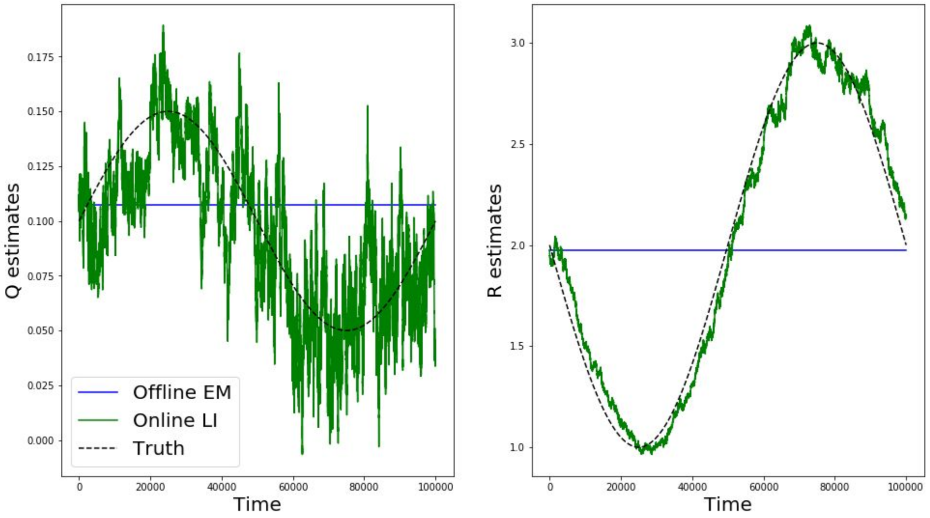
<u>Numerical experiment:</u>

- linear and Gaussian AR(1) model
- constant Q and R variances
- time varying Q and R variances

Offline estimation (constant Q & R):



Online estimation (varying Q & R):



Covariance estimation in data assimilation

Summary about this review paper

Schedule:

- draft available soon on arXiv
- accepted for submission in MWR
- ongoing simulation paper with L96 model & SPEEDY model

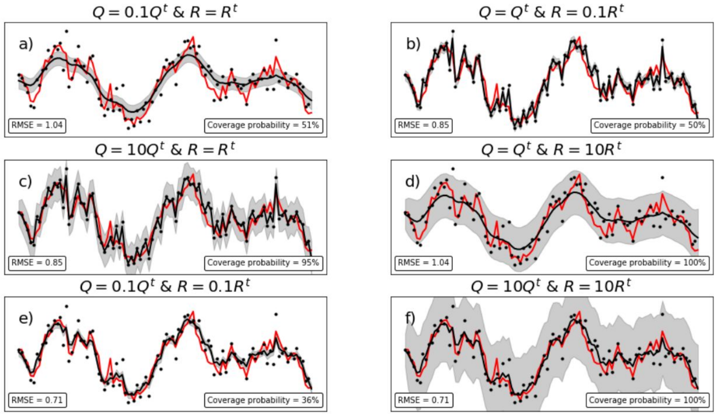


FIG. 1. Example of a univariate AR(1) process generated using Eq. (3) with $Q^t = 1$ (red line), noisy observations using $R^t = 1$ (black dots) and reconstructions using a Kalman smoother (black lines and gray 95% confidence interval) with different values of Q and R.

→ See my presentation at the UQ workshop in February 2018 for more details

AMS Journals Welcome Review Articles

David M. Schultz

Chair, Subcommittee on Reviews, and Chief Editor, Monthly Weather Review

See all authors & affiliations

https://doi.org/10.1175/MWR-D-18-0114.1

Published Online: 23 April 2018

My feedbacks about such review paper:

- very long, lot of reading, need to be honest and exhaustive, sometimes boring…
- but I have now a global view of the methods
- nice to build collaborations
- I hope it will be useful!

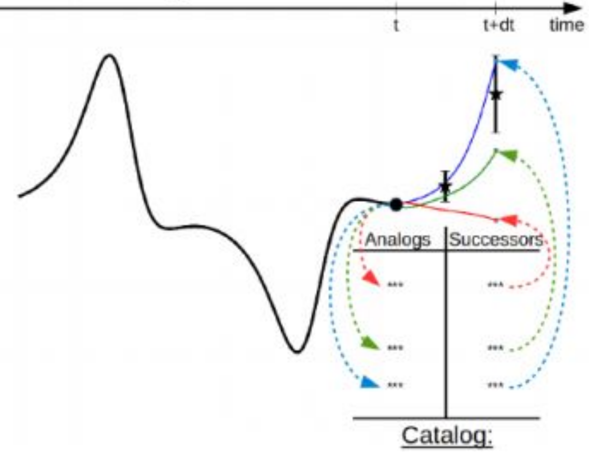
A bunch of data-driven methods

AnDA applied to spatial oceanography (2018-2020, with IFREMER and Univ. Grenoble)

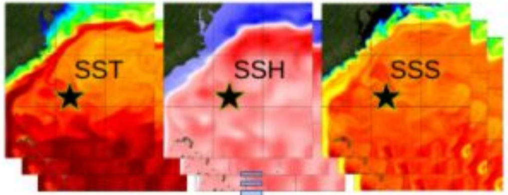
Goal:

- adaptive spatio-temporal interpolator for spatial oceanography
- use an ensemble of numerical simulations (50 ensembles, 55 years)
- apply the Analog Data Assimilation (AnDA, Lguensat et al. 2017 [MWR])

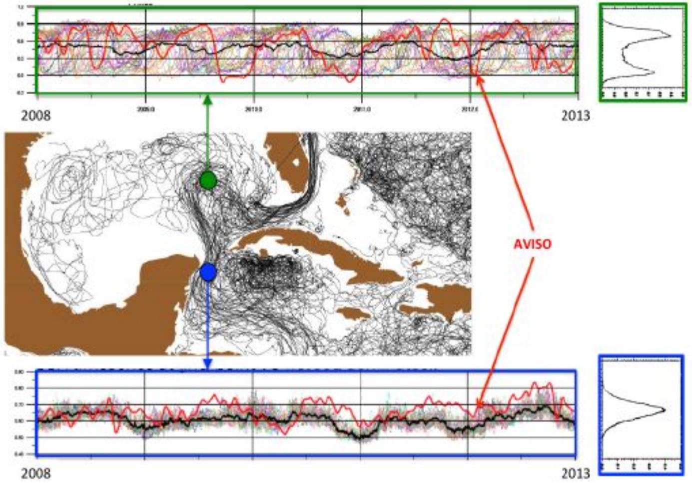
Analog Data Assimilation:



NEMO ensemble simulation:



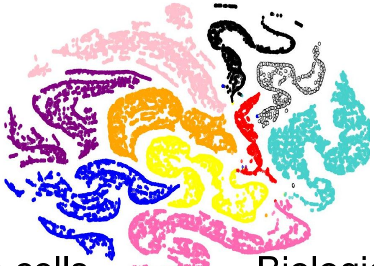
Analogs (a_k): SST, SSH, SSS (time t)	Successors (s_k): SST, SSH, SSS (time $t+dt$)
(23.86, 0.71, 36.14)	(23.97, 0.73, 36.35)
(23.21, 0.43, 36.56)	(22.96, 0.41, 35.78)
...	...
(21.37, 0.15, 32.26)	(20.67, 0.14, 32.01)
(19.18, 0.09, 32.15)	(19.98, 0.11, 32.98)



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

A bunch of data-driven methods

Deep learning on SAR images (2016-2019, with IFREMER and Univ. Seattle & New Hampshire)

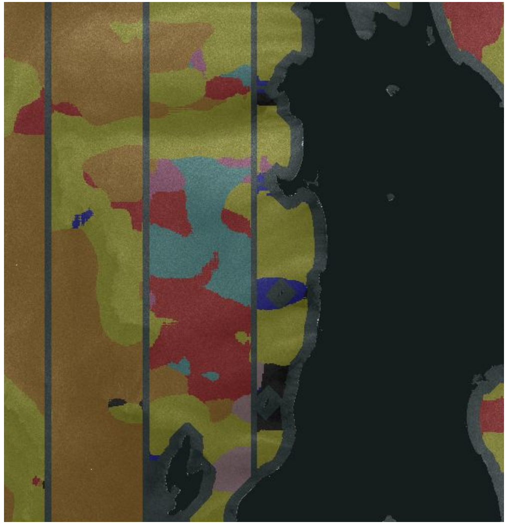


<u>Goal:</u>

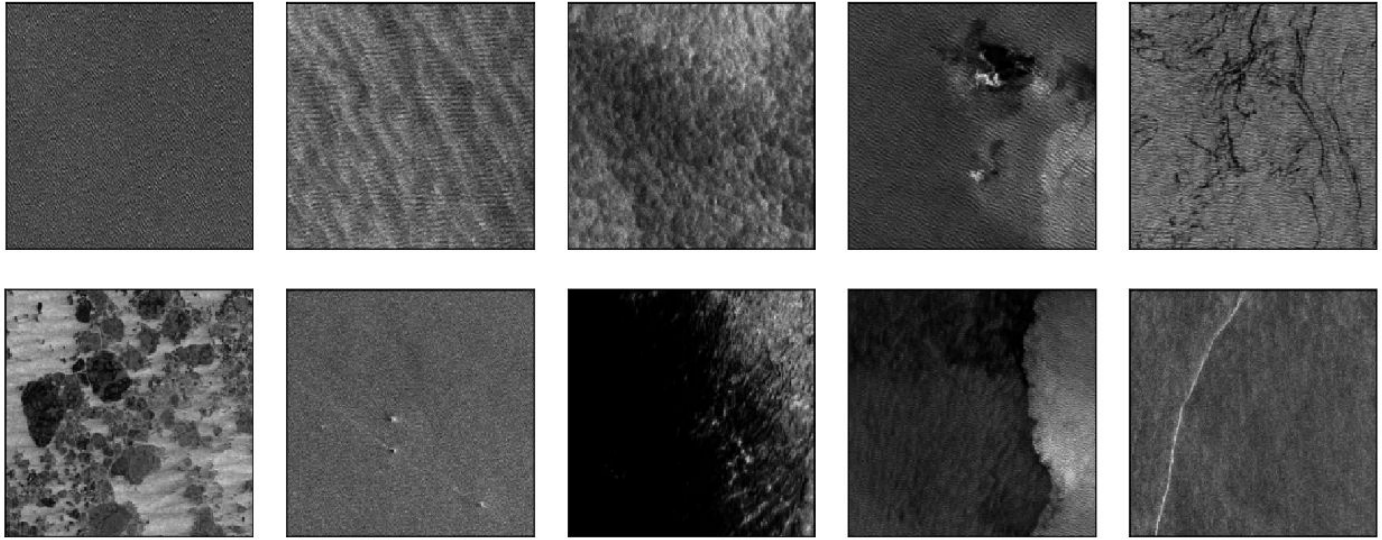
- use 20x20 km annotated SAR images (37,560)
- apply deep learning to classify natural phenomena

<u>Results:</u>

- 98% accuracy!
- lot of applications



Ocean swell   Wind streaks   Convective cells   Biological slicks



Sea ice   Icebergs   Oil sea   Atmospheric & oceanic fronts

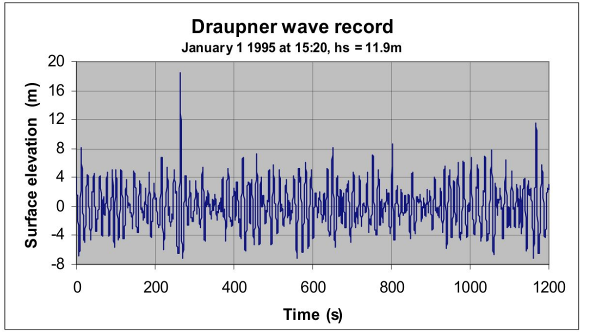
A bunch of data-driven methods

Predictability of rogue waves (2017-2019, with FEM and MIT)

Goal:

- nowcasting of extreme wave events (up to 5 minutes)
- apply a mix of data-driven & model-driven approaches

Mitsuyasu (2009) [J OCEANOGR]



Famous Draupner rogue wave

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

A bunch of data-driven methods

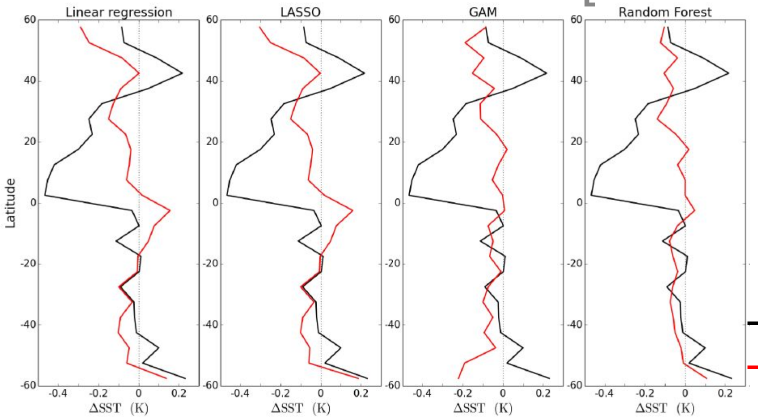
Machine learning for satellite data (with IFREMER and Météo-France)

Goal:

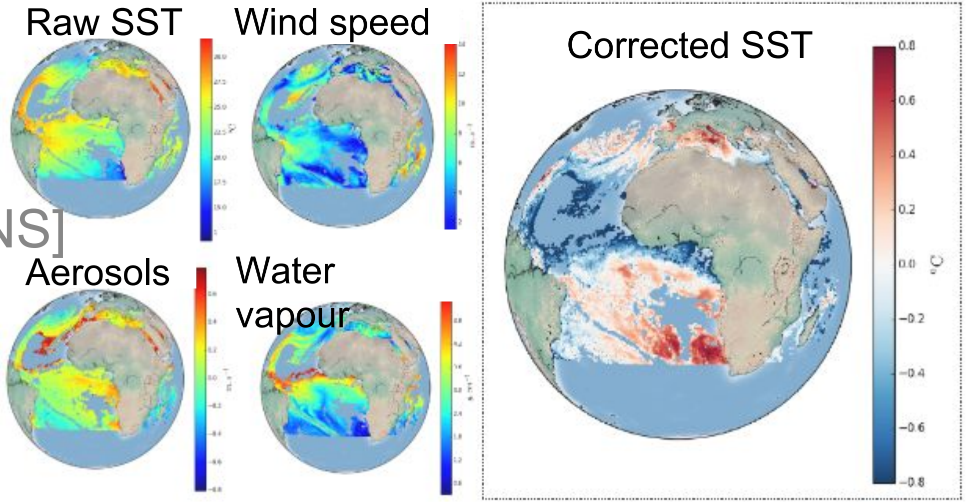
- post-process satellite SST data using atmospheric information
- use geostationary data and *in situ* measurements (485,600 match-ups)
- apply machine learning regression algorithms

Results:

- 31% of the variability is explained
- operationally applied at Météo-France
- Saux Picart et al. 2017 [REMOTE SENS]



- Raw bias
- Bias after correction



A bunch of data-driven methods

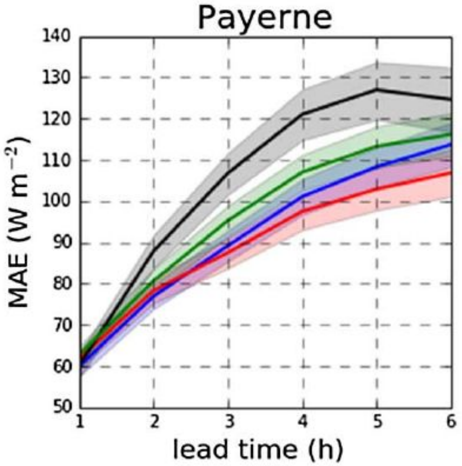
Machine learning for solar energy prediction (with Elum Energy, a French startup)

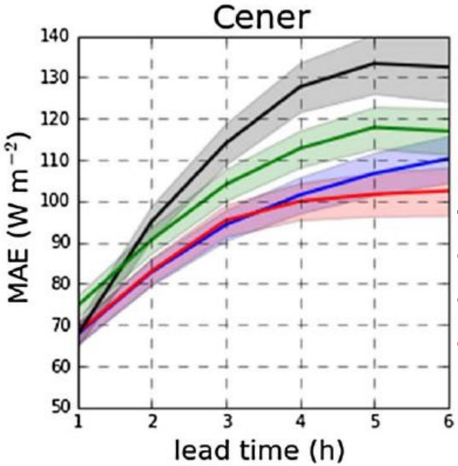
<u>Goal:</u>

- predict solar irradiance 6h ahead for solar panels and energy saving
- use geostationary data (2011-2016, hourly data, 0.05°)
- apply analog forecasting and other statistical methods

<u>Results:</u>

- method can be applied everywhere in Europe and Africa
- Ayet & Tandeo et al. 2018 [SOL ENERGY]





- Persistency
- VAR(1)
- Analog forecasting
- Analog forecasting (with post-processing)