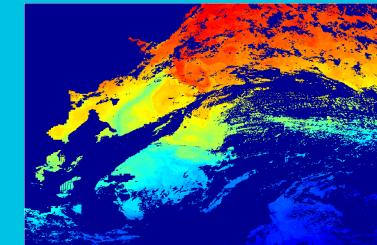




IMT Atlantique

Bretagne-Pays de la Loire
École Mines-Télécom

Learning-based strategies for the modeling and reconstruction of dynamical systems



R. Fablet

ronan.fablet@imt-atlantique.fr

TDMF, 2017

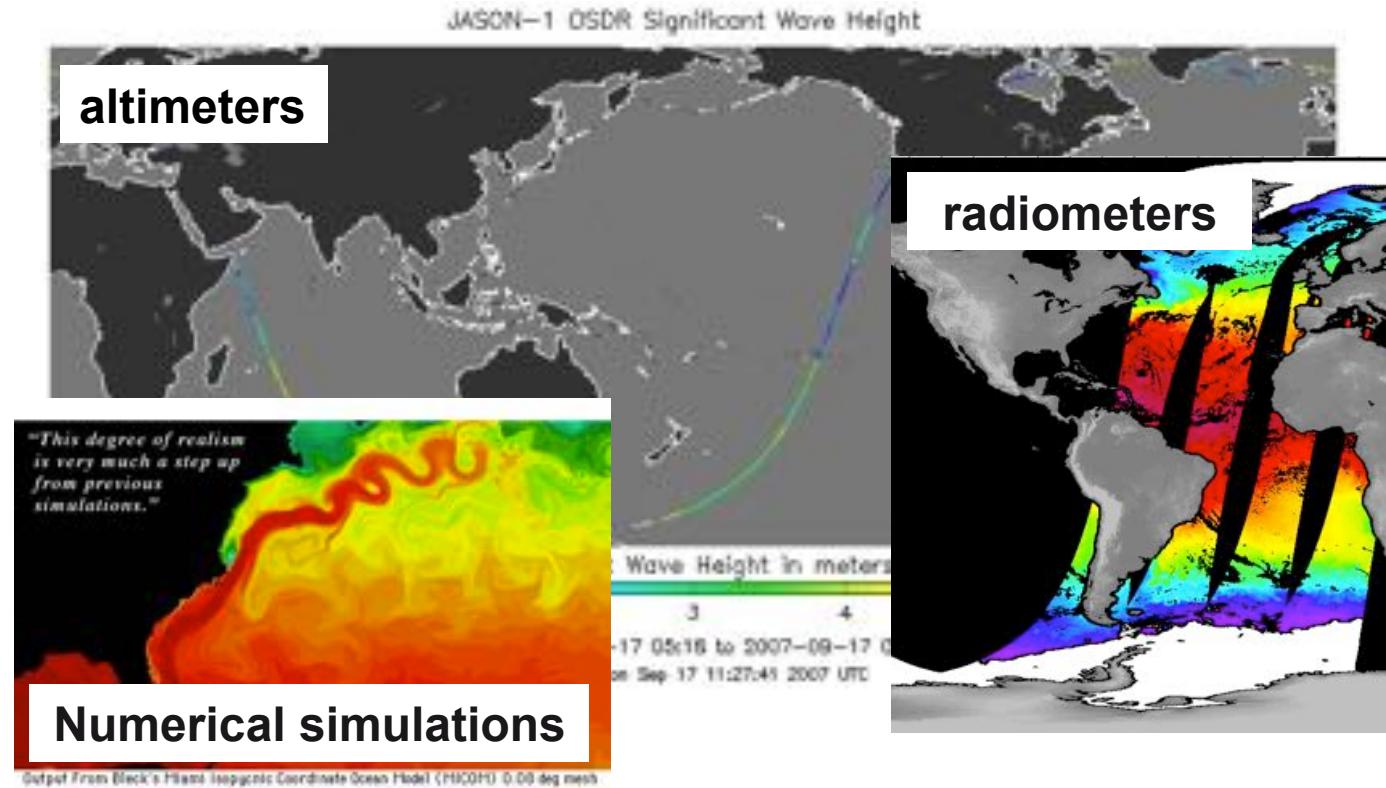
Lab-STICC

OUR APPLICATION CONTEXT: SATELLITE OCEAN SENSING



SATELLITE OCEAN SENSING & MODELING

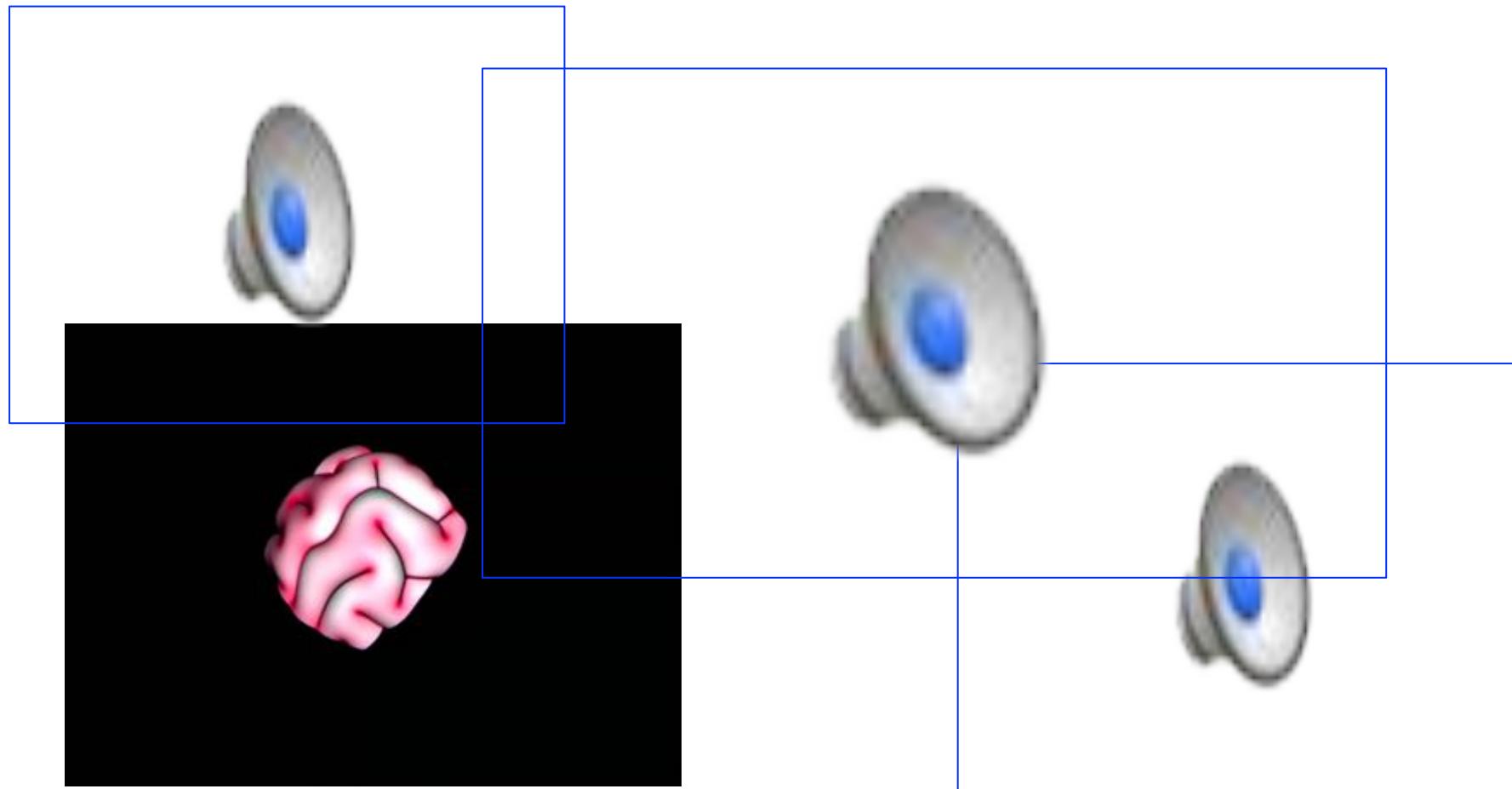
A variety of sensors and models



Key issue

How to characterize and reconstruct high-resolution ocean dynamics anywhere and anywhen from the multi-source

BROADER CONTEXT: DYNAMICAL SYSTEMS



How to infer the dynamical states from observations ? Which dynamical models ?

PHYSICALLY-DRIVEN FRAMEWORK: DATA ASSIMILATION



Known physical
equations
(ODE, PDE,...)

$$\begin{aligned}\frac{dx(t)}{dt} &= \sigma (y(t) - x(t)) \\ \frac{dy(t)}{dt} &= x(t) (\rho - z(t)) - y(t) \\ \frac{dz(t)}{dt} &= x(t) y(t) - \beta z(t)\end{aligned}$$

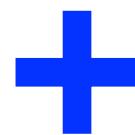
Lorenz-63 equations

Data assimilation (eg, Evensen, 2007)

- State-space setting

Dynamic model

$$\partial_t X = F(X, \xi, t, \theta)$$



Observation model

$$Y_t = H(X, \zeta, t, \phi)$$

- Sequential resolution using stochastic filters (Ensemble Kalman filters, particle filters)



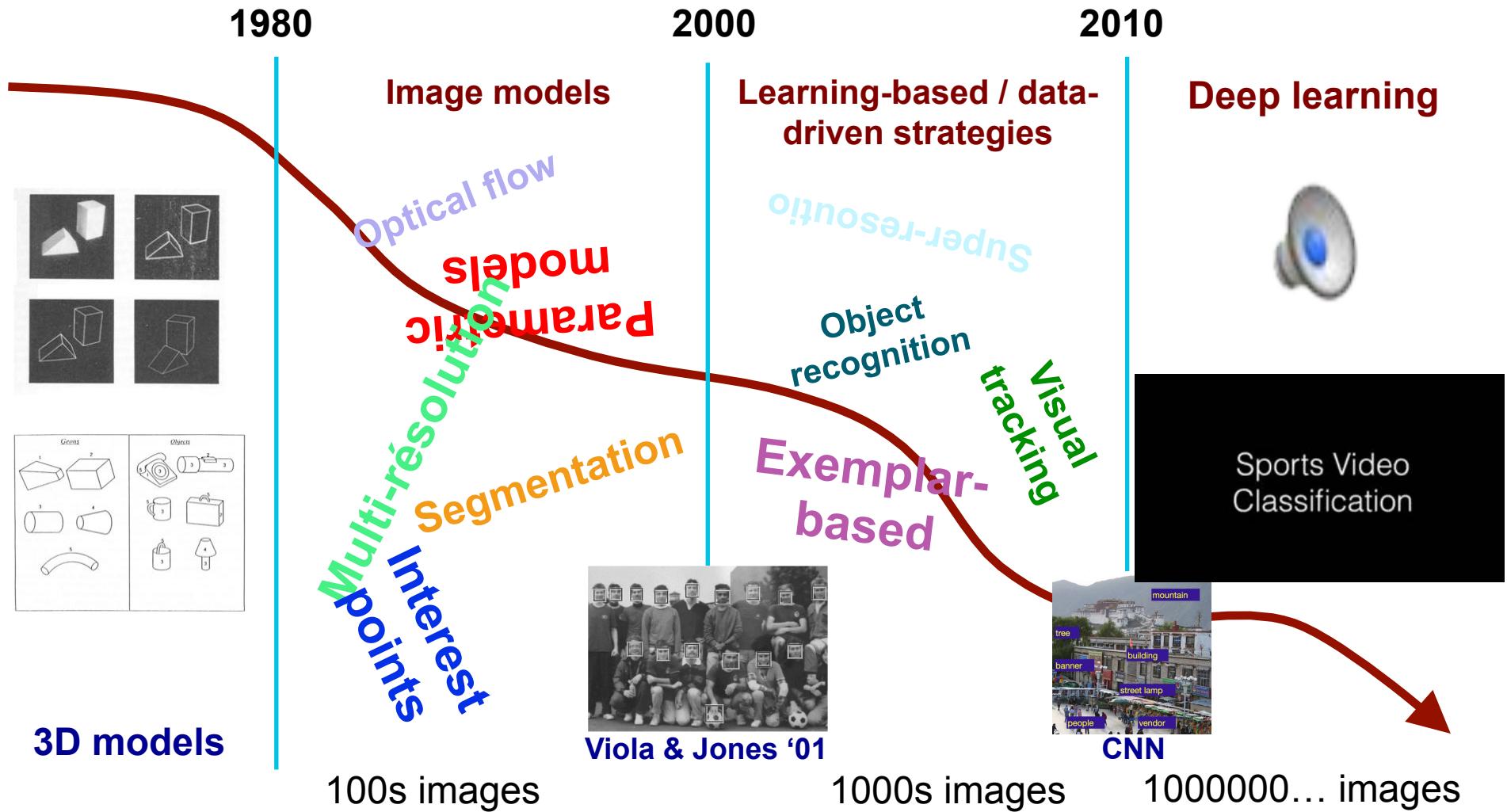
DO WE TRUST OR NEED MODELS ?

CAN WE DISCOVER MODELS ?

**CAN WE REDUCE/OPTIMIZE THE
COMPUTATIONALLY EFFICIENCY?**

The holy grail of DATA ?

COMPUTER VISION/SIGNAL PROCESSING: RECENT LEARNING-BASED BREAKTHROUGHS



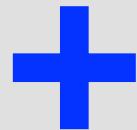
Key issue: evolution towards learn image/signal representations from data to solve complex tasks

MODEL-DRIVEN vs. DATA-DRIVEN APPROACHES

Model-driven paradigm

Dynamical model

$$X_t \rightarrow \text{ODE solver} \quad \partial_t X = F(X, \xi, t, \theta) \rightarrow X_{t+1}$$



Observation model

$$Y_t = H(X, \zeta, t, \phi)$$

Data-driven paradigm

Dynamical model



Observation model

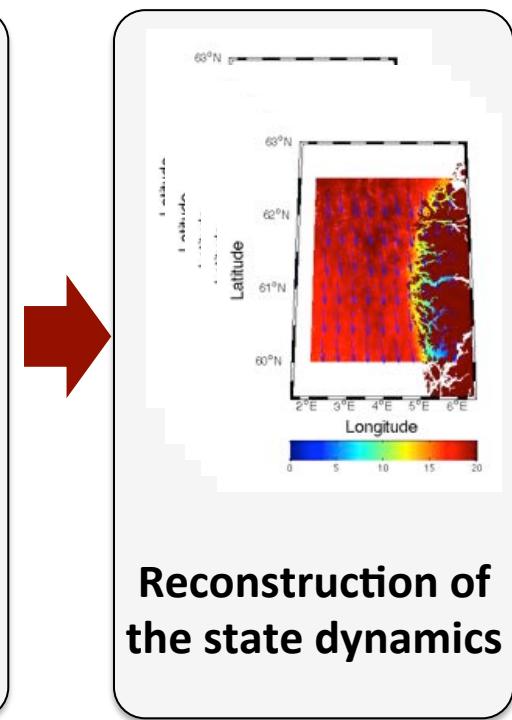
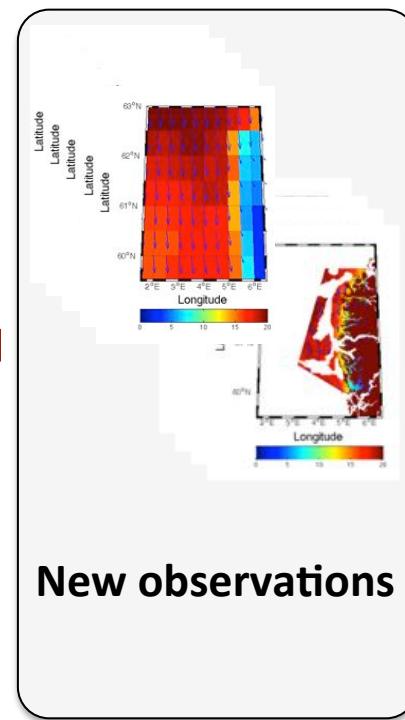
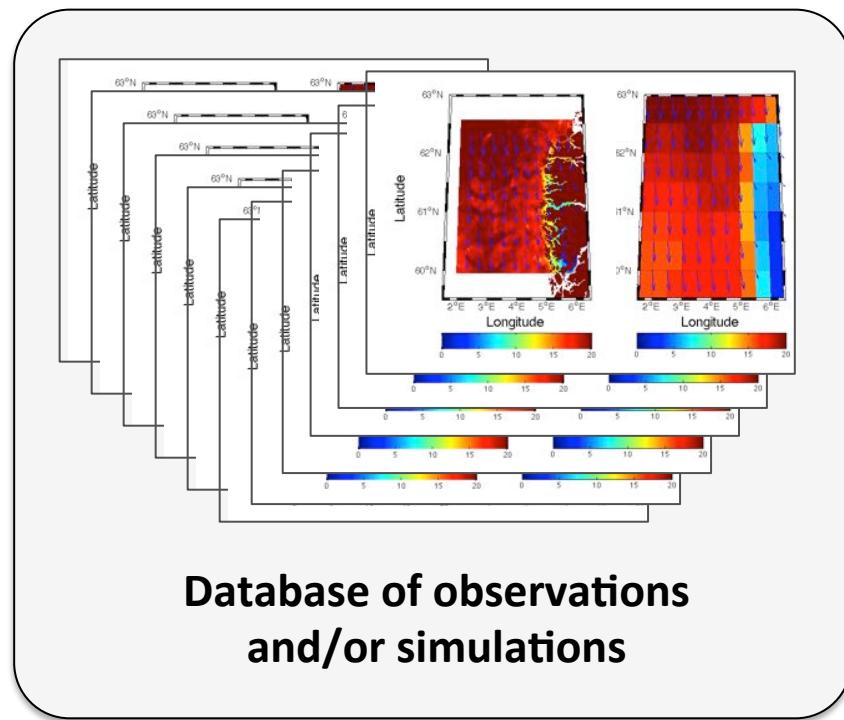
$$Y_t = H(X, \zeta, t, \phi)$$

DATA-DRIVEN ASSIMILATION

Dynamic model

$$\partial_t X = F(X, \xi, t, \theta)$$

Towards data-driven assimilation models ?



Key objective

Building implicit data-driven representation of the state dynamics from available observation/simulation datasets

THE ANALOG DATA ASSIMILATION

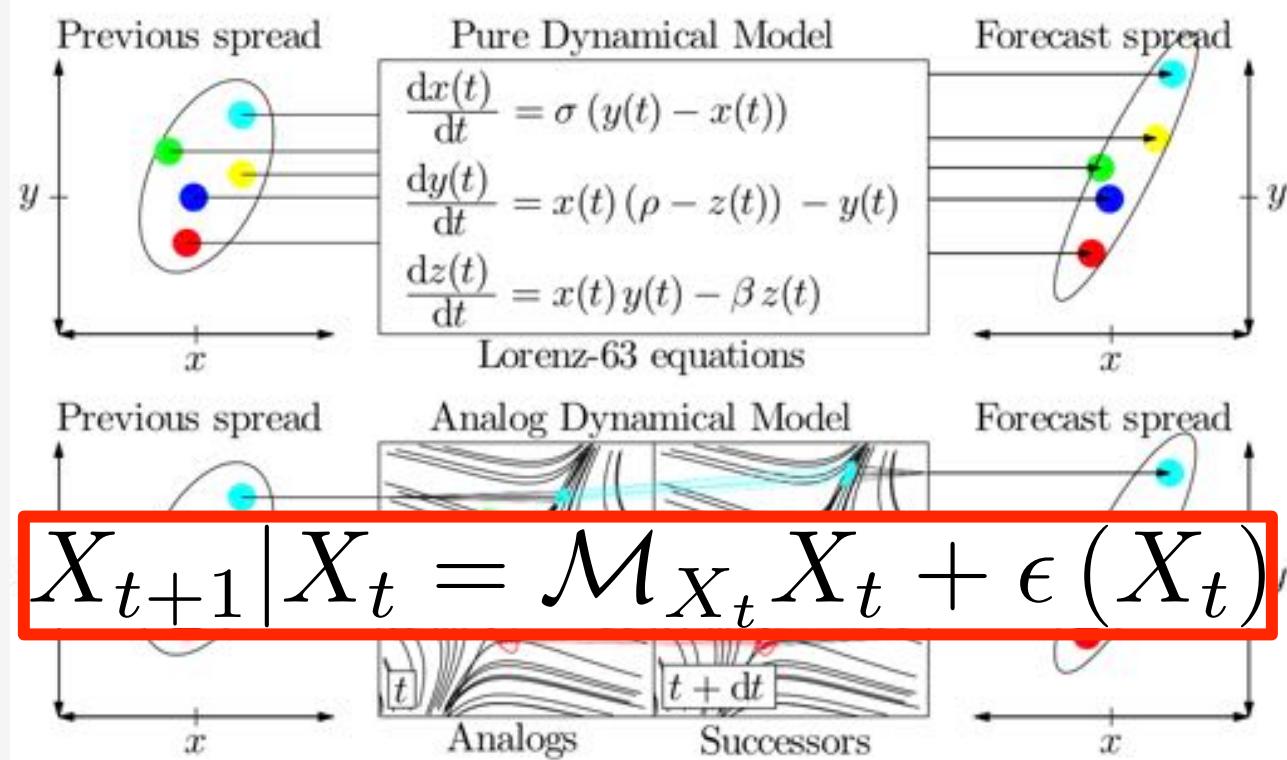
THE ANALOG DATA ASSIMILATION

Stochastic assimilation using analog dynamic models

Key idea

Replacing the explicit dynamic model by an analog forecasting operator

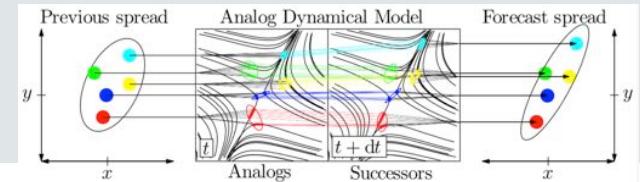
**Plug-and-play application to stochastic filters
(e.g., EnKF, PF)**



<https://github.com/rfablet>

Lguensat et al., 2017

THE ANALOG DATA ASSIMILATION



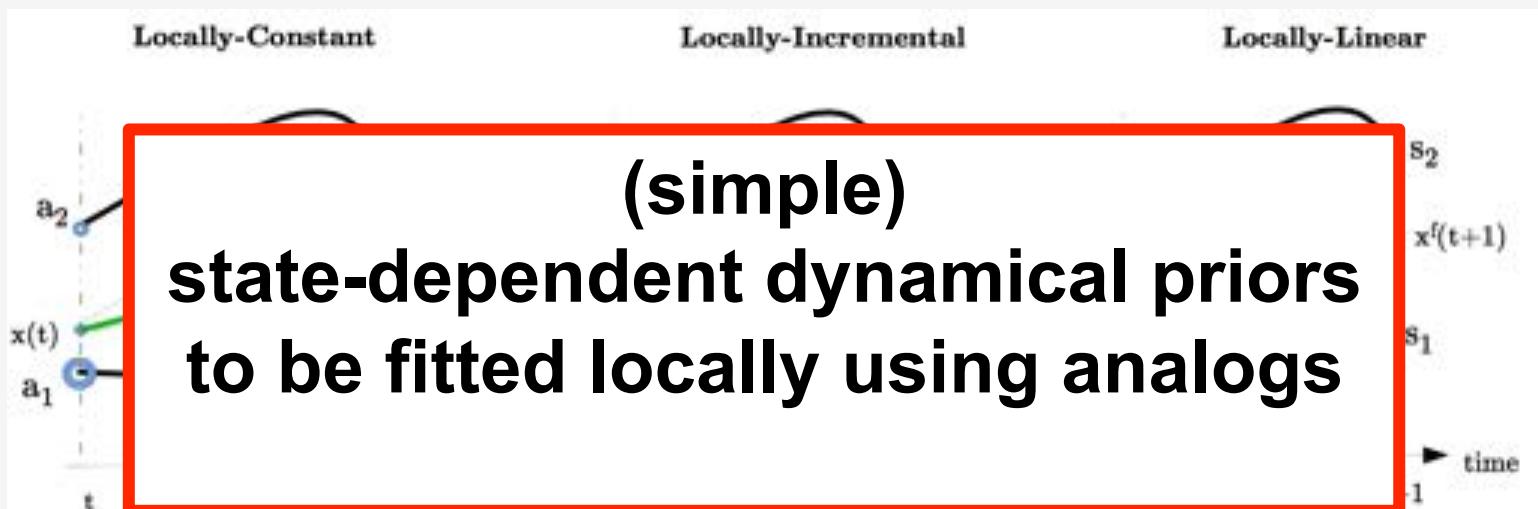
Analog forecasting operators (Lguensat et al., 2017)

Gaussian model with mean and covariance estimated from the nearest-neighbors of the current state (« local surrogate model »):

$$X_t | X_{t+1} \sim \mathcal{N} \left(\mu \left(\{a^k, s^k, \omega^k, X_t\} \right), \Sigma \left(\{a^k, s^k, \omega^k\} \right) \right)$$

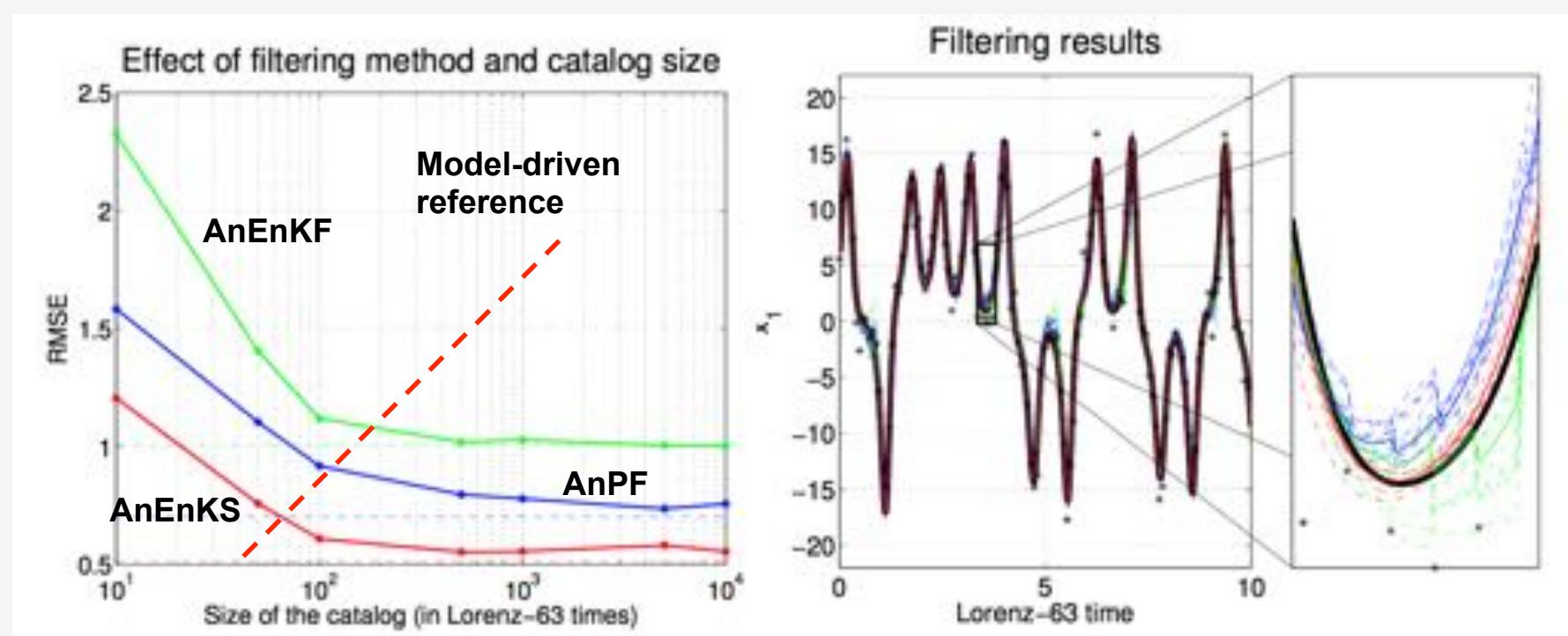
K analogs & successors for x_t similarity (kernel) weights

Different parameterizations:



DATA-DRIVEN ASSIMILATION MODELS

Analog assimilation for low-dimensional systems (eg Lorenz-63)



Best analog forecasting operator:

Locally-linear analog operator

NEURAL-NETWORK-BASED REPRESENTATION FOR DYNAMICAL SYSTEMS

DYNAMICAL SYSTEMS AS RECURRENT NEURAL NETWORKS

ODE solver: Runge-Kutta(4) example

$d_t X =$

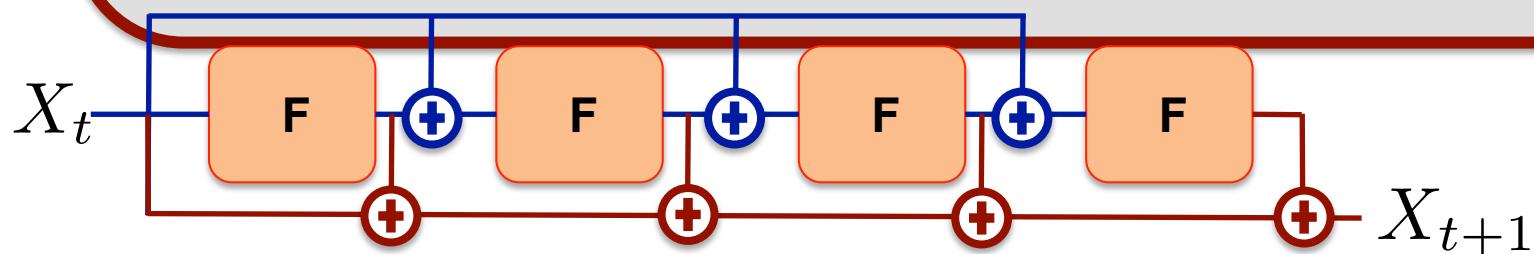
Neural network parameterization for operator F

$X_{t+\Delta t} =$

Learning operator F from training data

Graph

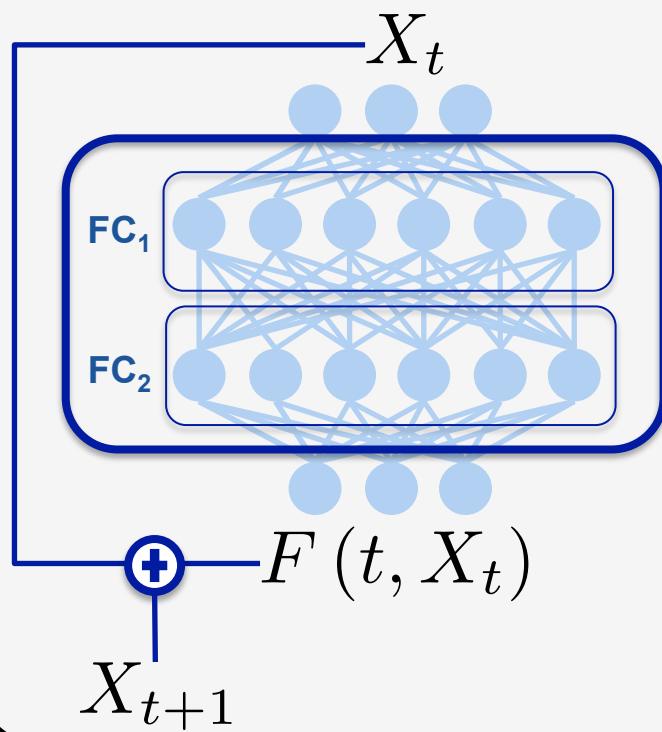
Plug-and-play use in assimilation schemes



DYNAMICAL SYSTEMS AS RECURRENT NEURAL NETWORKS

MLP-based parameterization for dynamical systems

Euler-based
approximation



Limited ability of FC layers to represent bilinear non-linearities

$$\begin{aligned}\frac{dx(t)}{dt} &= \sigma(y(t) - x(t)) \\ \frac{dy(t)}{dt} &= x(t)(\rho - z(t)) - y(t) \\ \frac{dz(t)}{dt} &= x(t)y(t) - \beta z(t)\end{aligned}$$

Lorenz-63 equations

or overly complex
parameterization using multi-
layer / « deep » architectures

DATA-DRIVEN REPRESENTATIONS FOR DYNAMICAL SYSTEMS

Numerical experiments for Lorenz-63 dynamics

Forecasting experiments

Noise-free training data

Forecasting time step	t_0+h	t_0+4h	t_0+8h
Analog forecasting	$<10^{-6}$	0.002	0.005
Sparse regression	$<10^{-6}$	0.002	0.006
MLP	$<10^{-6}$	0.018	0.044
<i>Bi-NN(4)</i>	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$

Noisy training data ($\sigma=0.5$)

Forecasting time step	t_0+h	t_0+4h	t_0+8h
Analog forecasting	$<10^{-6}$	2.01	2.2
<i>Bi-NN(4)</i>	$<10^{-6}$	0.054	0.14



Assimilation experiment

(1 obs. every 8 time steps)

Noise standard deviation in training data	0	0.25	1
<i>True model</i>	<u>0.50</u>	-	-
Analog forecasting	0.65	1.17	1.81
<i>Bi-NN(4)</i>	0.60	0.75	0.86

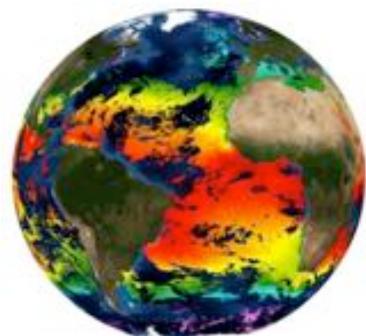
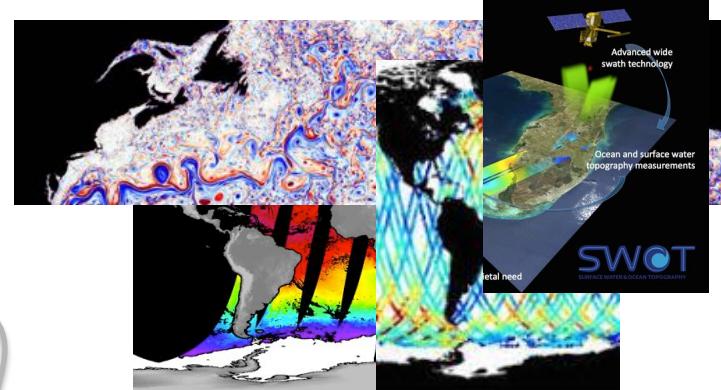
MODEL-DRIVEN VS. DATA-DRIVEN STRATEGIES

Key messages

- Data-driven representations as relevant alternatives to model-driven schemes
- Ability to learn new representations from large-scale observation and/or simulation datasets
- Different yet complementary data-driven representations (e.g., analog vs. neural-net methods)
- Data are great Dig deeper or/and be friend with physicists

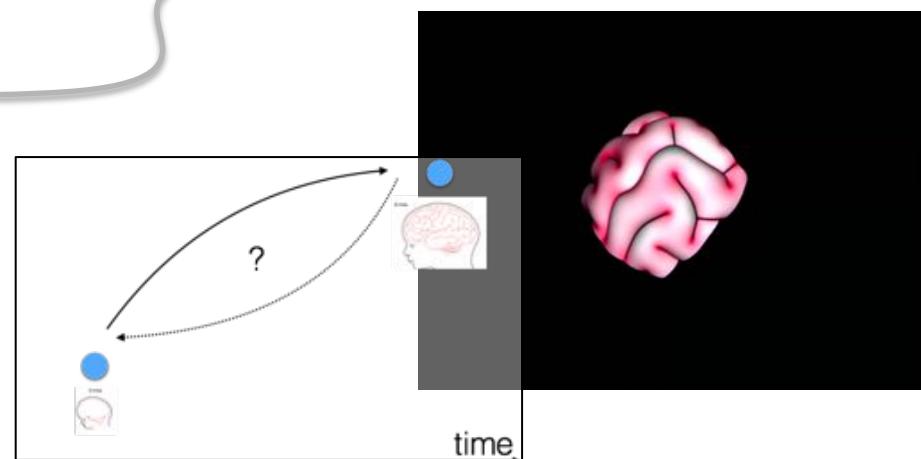
ONGOING & FUTURE WORK

Application to ocean dynamics and satellite ocean sensing



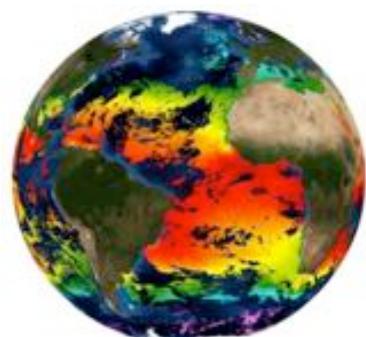
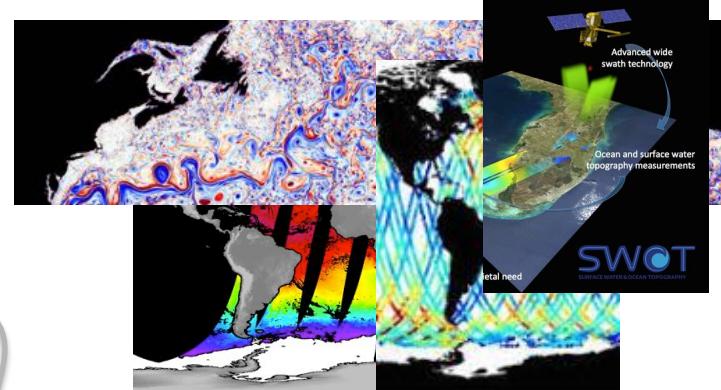
Scaling up to large-scale dataset

Early brain
development



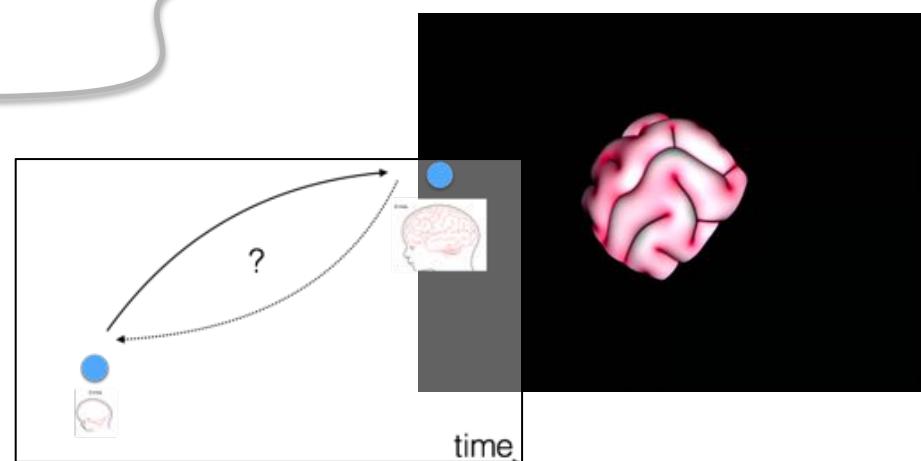
ONGOING & FUTURE WORK

Application to ocean dynamics and satellite ocean sensing



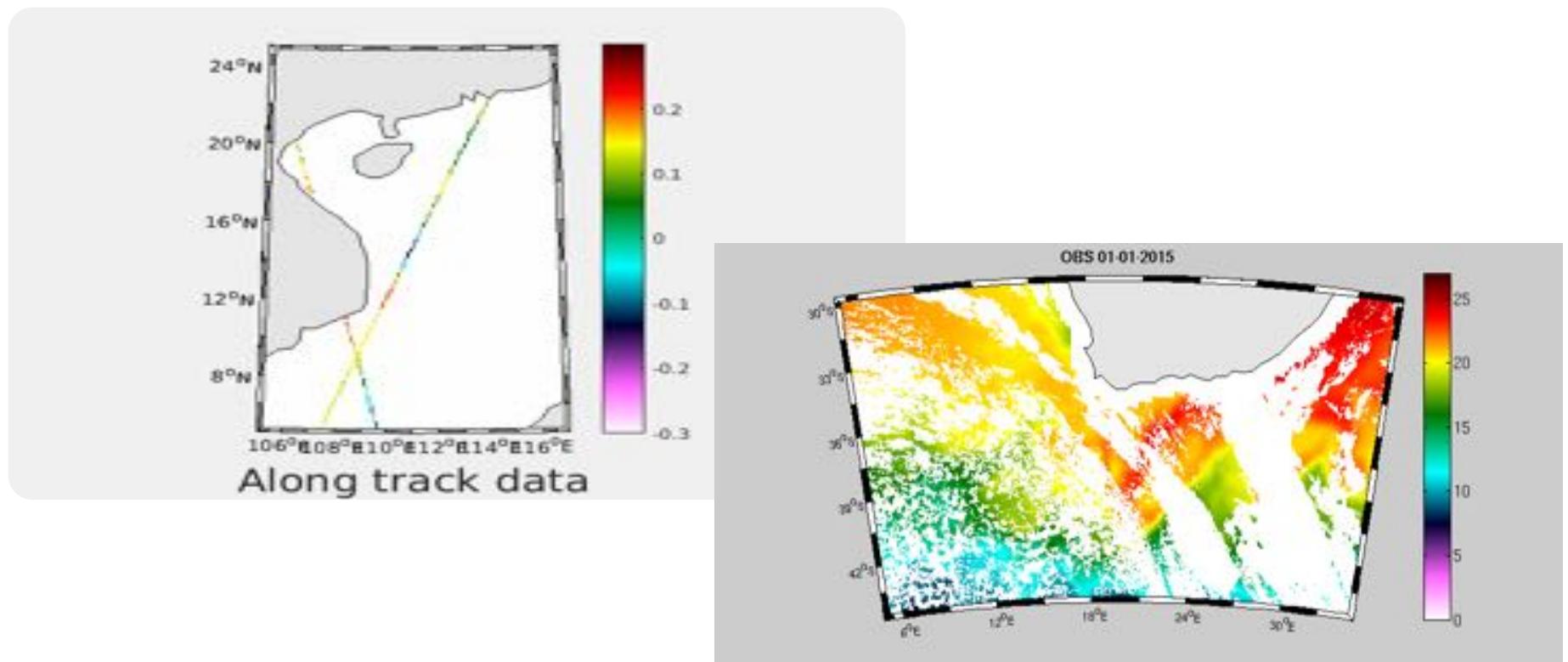
Scaling up to large-scale dataset

Early brain
development

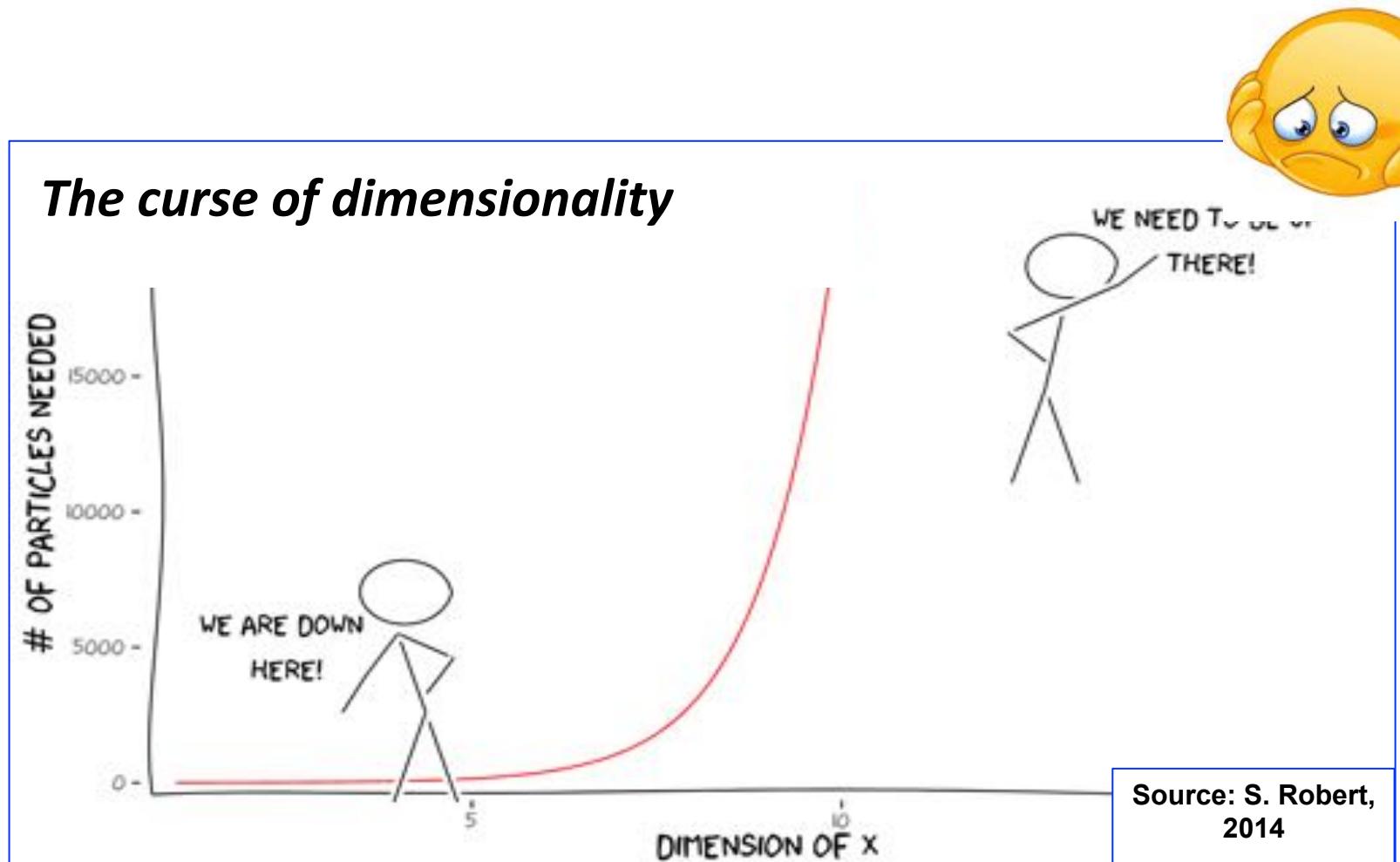


DATA-DRIVEN ASSIMILATION FOR GEOPHYSICAL SYSTEMS

Do data-driven strategies apply to high-dimensional states (e.g., 2D fields)?



DATA-DRIVEN ASSIMILATION & HIGH-DIMENSIONAL SYSTEMS



Should we wait for a few billions of years to implement analog data assimilation for real applications



DATA-DRIVEN ASSIMILATION MODELS

Turn data assimilation into a puzzle game

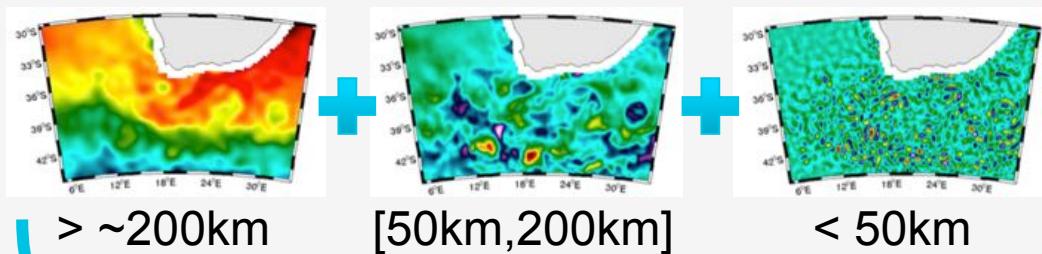


How to break the assimilation issue into a series of (independent) low-dimensional sunproblems ?

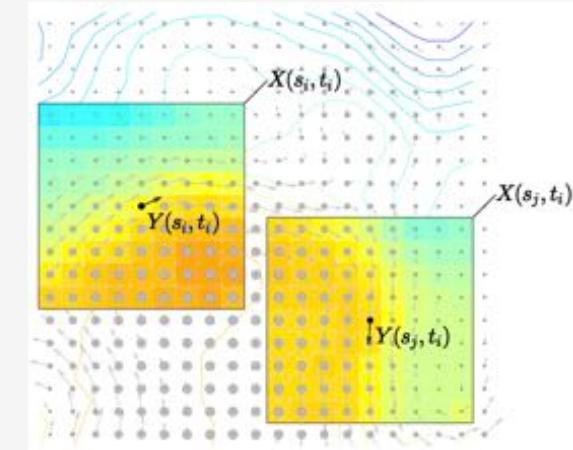
DATA-DRIVEN ASSIMILATION MODELS

Turn analog data assimilation into a puzzle game

Multi-scale (spatial) decomposition



Patch-based representations



EOF-based decomposition

« local » data assimilations at different scales and independently for different patch positions

DATA-DRIVEN ASSIMILATION MODELS

Considered state-space model

$$\left\{ \begin{array}{l} X = \bar{X} + \sum_i dX_i + \xi \quad \text{Multiscale decomposition} \\ \bar{X} \propto \mathcal{G}(\bar{X}^b, \Gamma) \quad \text{Gaussian prior (Optimal interpolation)} \\ dX_i \propto \mathcal{M}_i \quad \text{Patch-level Analog forecasting operators} \\ Y = \mathcal{H}(X, \Omega, \epsilon) \end{array} \right.$$

$dX_i(\mathcal{P}_s, t) = \mathcal{M}_i(dX_i(\mathcal{P}_s, t-1), \nu(t))$

(patch-level) Markov Random fields

Numerical resolution using a coarse-to-fine strategy

Patch-level dynamical operators constrained by EOF expansions to locally encode the spatial structure of the 2D field

DATA-DRIVEN ASSIMILATION MODELS

Numerical experiments

Simulation of of groundtruthed dataset using real sampling patterns (along-track vs. cloud patterns)

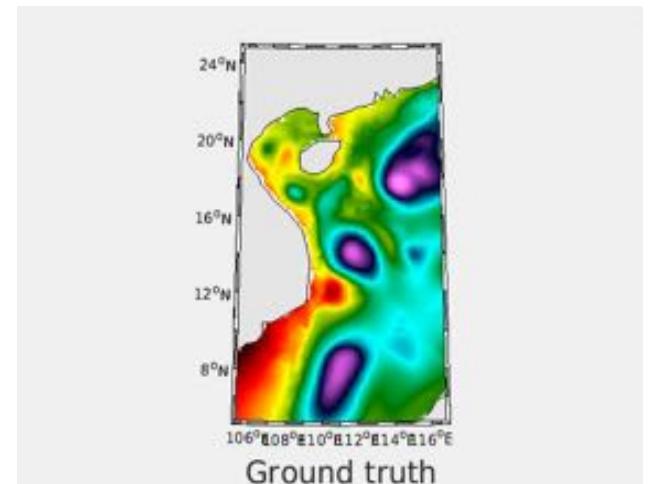
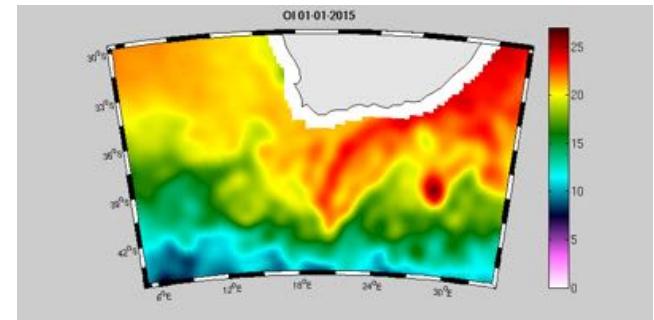
Two case-studies: SLA & SST

SLA dataset: OFES numerical simulations

SST dataset: cloud-free SST product (e.g., OSTIA)

Comparison to state-of-the-art algorithms:

- *Optimal interpolation (OI) (Donlon et al., 2012)*
- *DINEOF (Ping et al., 2016)*



DATA-DRIVEN ASSIMILATION MODELS

Analog data assimilation:

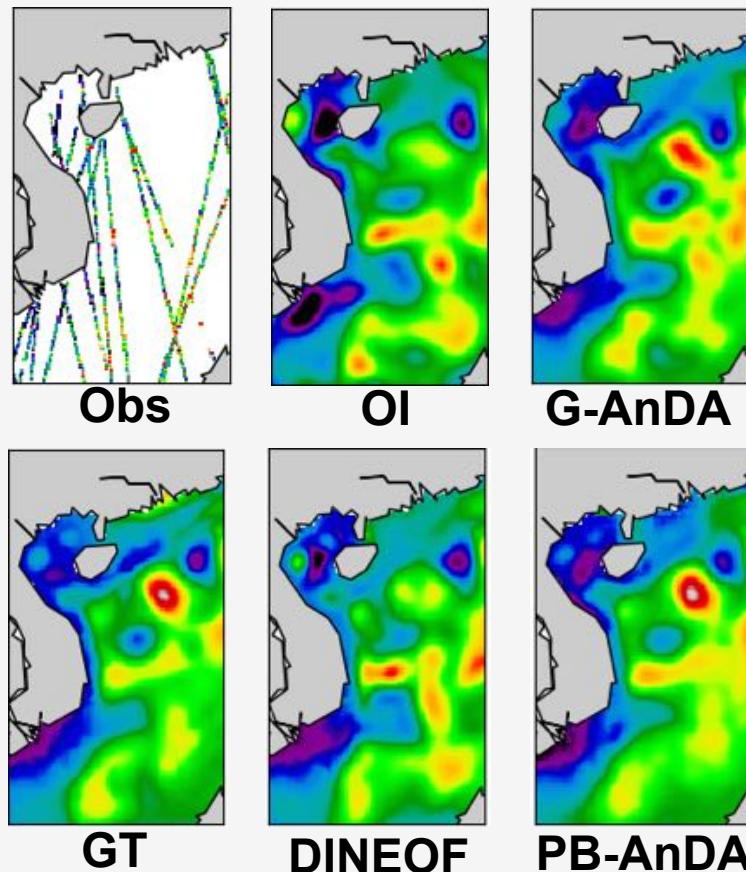
SLA (noise variance $\sigma=0.03$)			SST (AVHRR cloud pattern)		
	RMSE	ρ		RMSE	ρ
OI	0.07	0.41	OI	0.48	0.69
DINEOF	0.06	0.45	DINEOF	0.40	0.79
G-AnDA	0.04	0.67	G-AnDA	0.38	0.81
PB-AnDA	0.03	0.71	PB-AnDA	0.24	0.93

Relative improvement up to 30%/50% w.r.t. OI & DINEOF in terms of RMSE/correlation for the considered OSSE case-studies

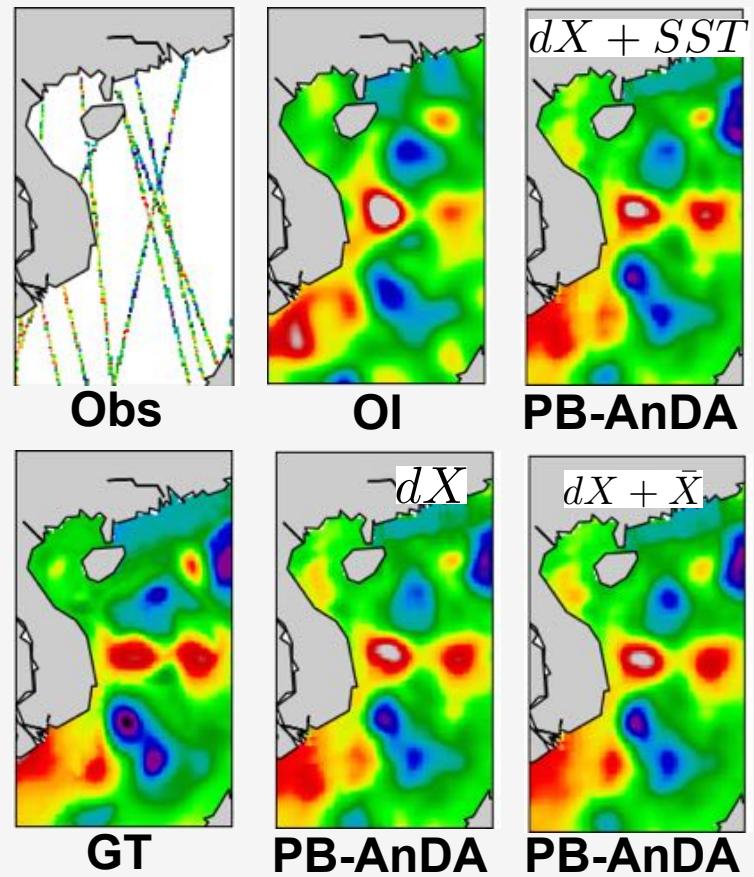
DATA-DRIVEN ASSIMILATION MODELS

Analog data assimilation for SLA (South China Sea)

Reconstruction example
for different schemes

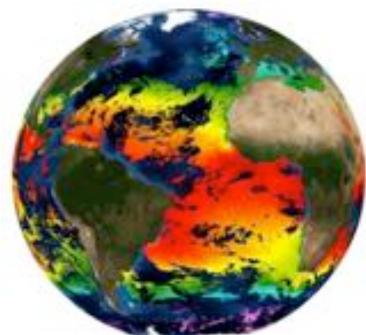
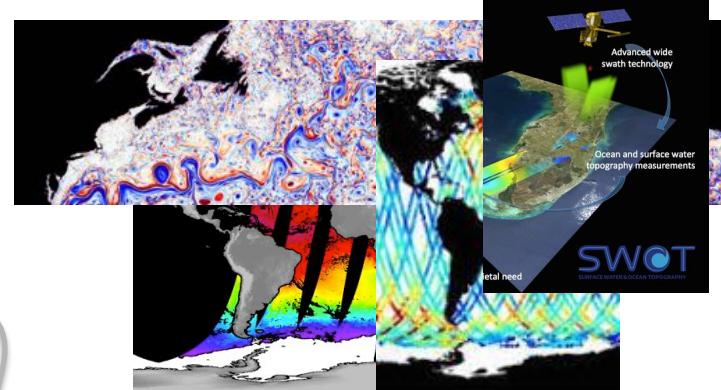


Reconstruction example
with different variables



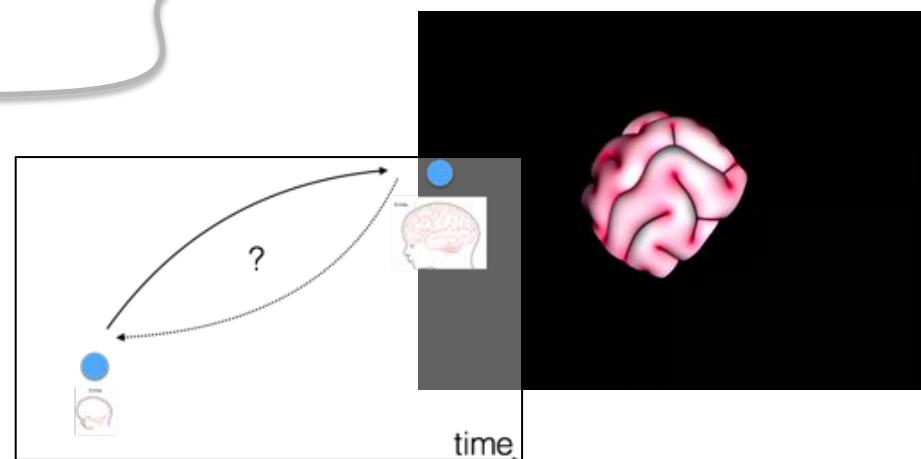
ONGOING & FUTURE WORK

Application to ocean dynamics
and satellite ocean sensing



Scaling up to large-scale dataset

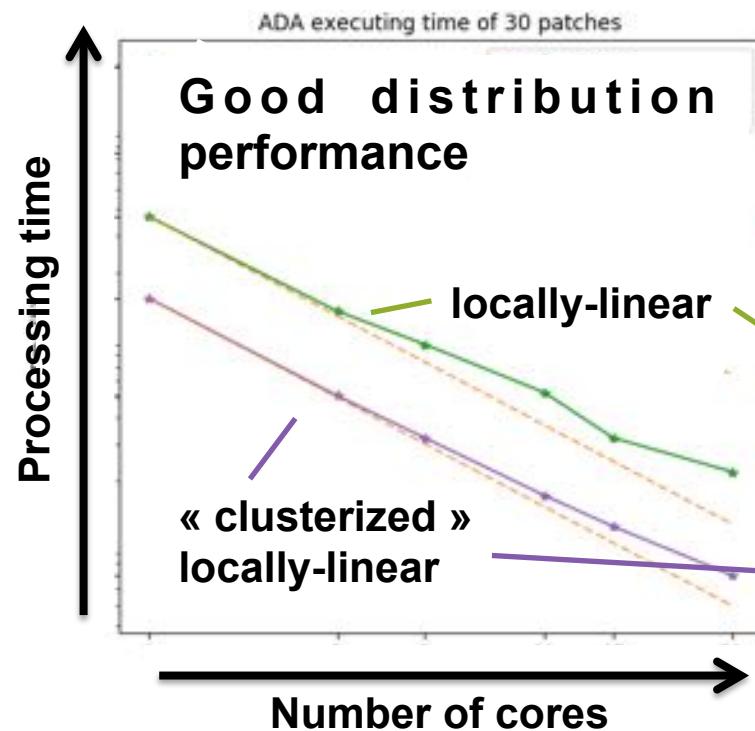
Early brain development



DATA-DRIVEN ASSIMILATION MODELS

Can we scale up to large-scale dataset and global scale ?

- Parallel processing (python library)
- Acceleration of nearest-neighbor search (eg, FLANN)



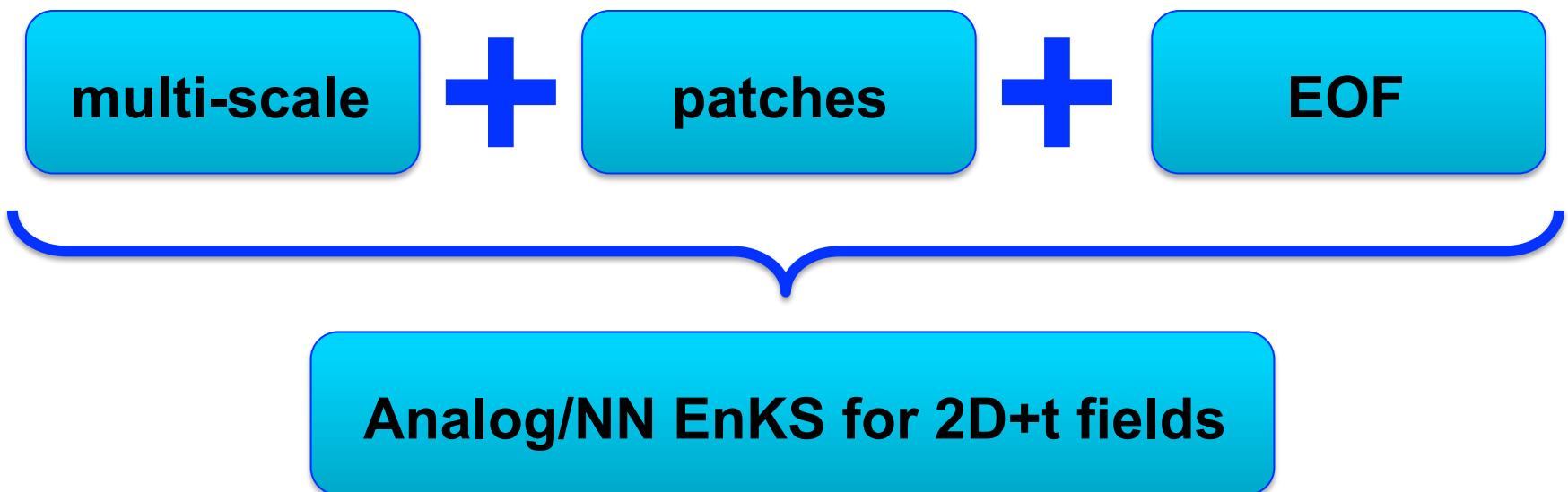
**SST case-study: 150x500 grid
~ 200 20x20 patches**

	RMSE	ρ	Time
OI	0.48	0.69	-
DINEOF	0.40	0.79	4'
G-AnDA	0.38	0.81	82'
PB-AnDA	0.24	0.93	75'
PB-AnDA-CLL	0.24	0.93	23'

MODEL-DRIVEN VS. DATA-DRIVEN STRATEGIES

Summary

- Data-driven representations as relevant alternatives to model-driven schemes, including 2D+t systems
- Ability to learn new representations from large-scale observation and/or simulation datasets
- Different yet complementary data-driven representations



DATA-DRIVEN STRATEGIES: WHAT NEXT ?

Question to be addressed ?

- **Data representativity & generalization capabilities for (geo)physical systems ?**
- **Dealing with extremes ?**
- **Global vs. locally-adapted representations ?**
- **Data are great.... Dig deeper or/and be friend with physicists. Towards physically-sound/interpretable data-driven representations ?**
- **Beyond forecasting performance ? Embedding physically-derived priors ?**

Acknowledgements

Thank you.

Joint work with P. Ailliot, B. Chapron, C. Herzet, R. Lguensat, S. Ouala, F. Rousseau, P. Tandeo, P. Viet,...

Lguensat al., The Analog Data Assimilation, MWR 2017.

Fablet et al., Data-driven Methods for Spatio-Temporal Interpolation of Sea Surface Temperature Images. IEEE TCI, 2017.

More at https://www.researchgate.net/profile/Ronan_Fablet



TERALAB