Bridging observations and numerical modelling of the ocean and climate using machine learning

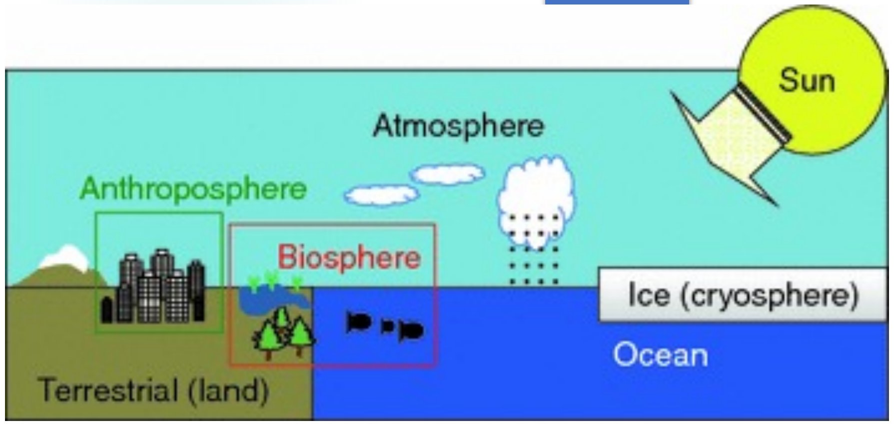
JULIEN BRAJARD, NERSC

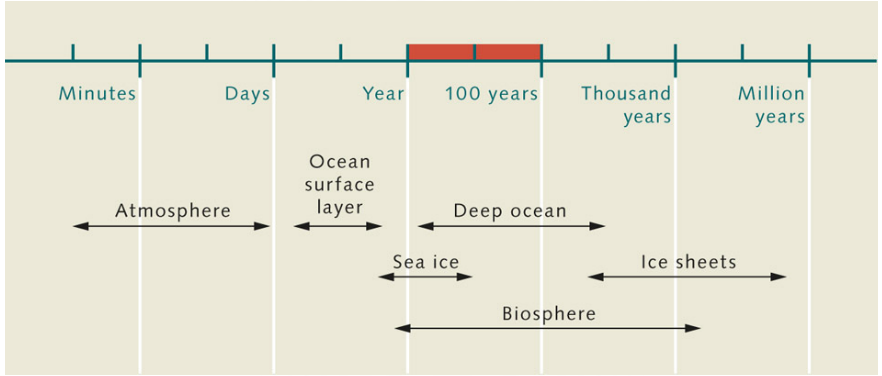
COLLABORATIONS:

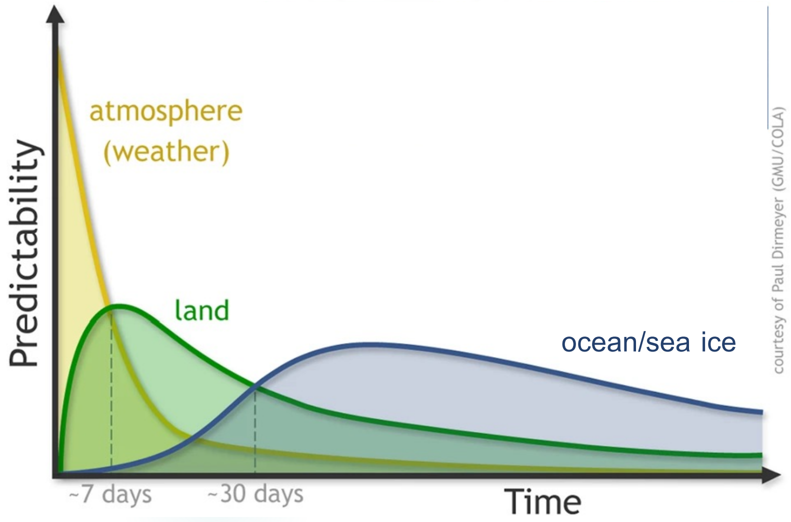
S. BARTELEMY (UIB), L. BERTINO (NERSC), M. BOCQUET (E. PONT), A. CARRASSI (U READING), F. COUNILLON (NERSC) A. KOROSOV (NERSC), E. ÓLASON (NERSC)

Role of the ocean and the sea-ice in the climate system

- The ocean and the sea-ice (O+SI) are 2 components of the **climate system**.
- O+SI contain **large spatial and temporal scales** (>YEAR) that are particularly relevant for climate studies.
- O+SI are a **source of predictability** for the next decade.

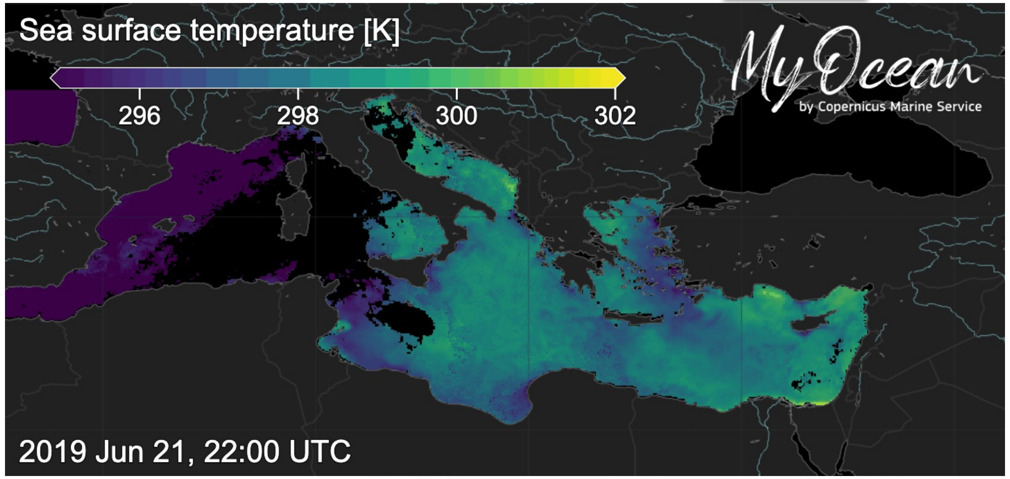


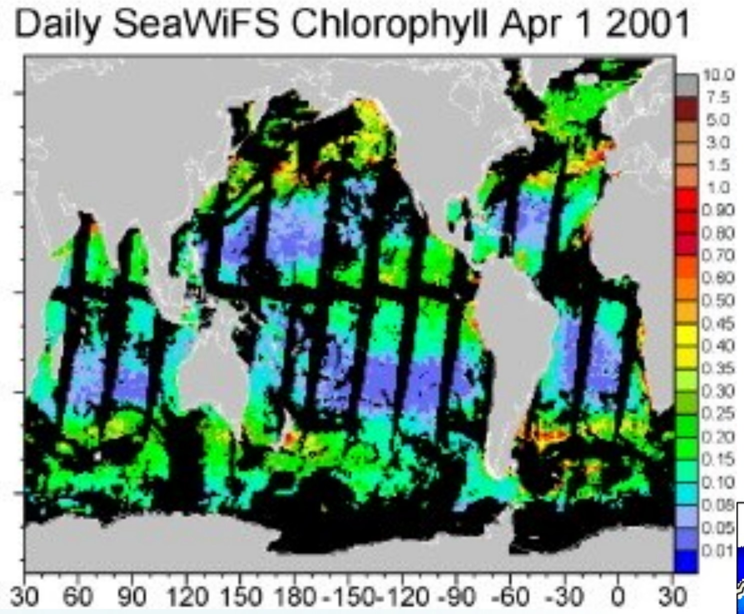


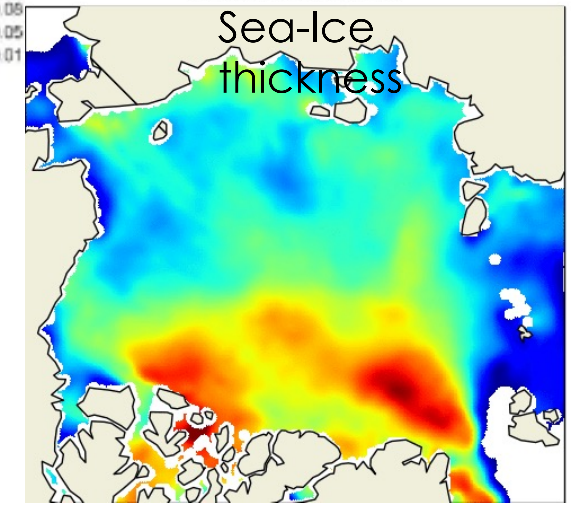


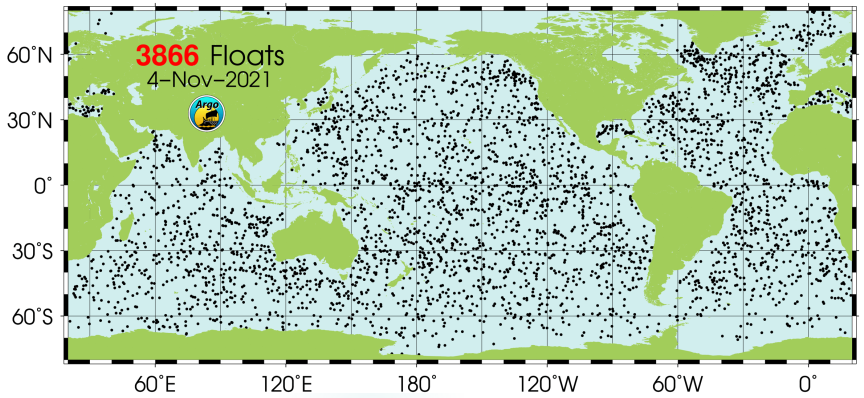
Credits: Gettelman and Rood, 2016, Meincke und Latif, 1995, Dirmeyer

First source of knowledge: observations

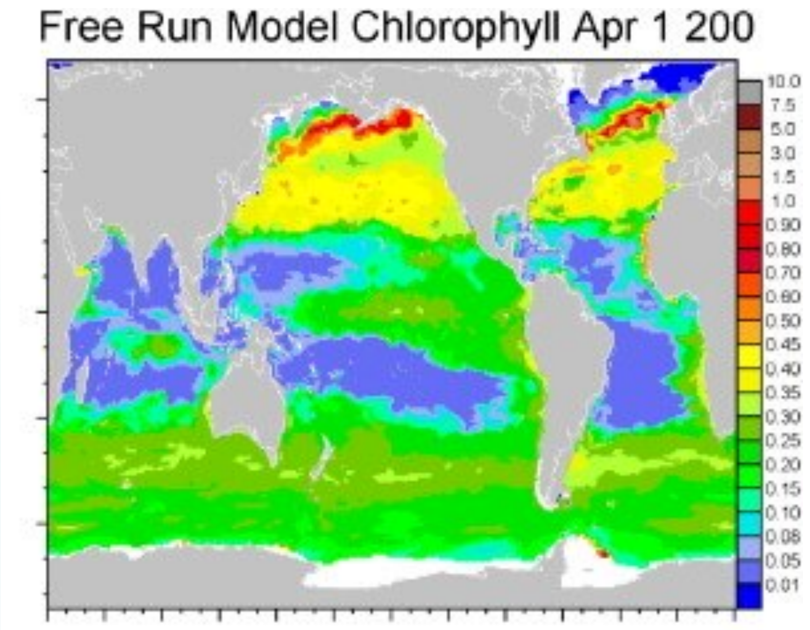


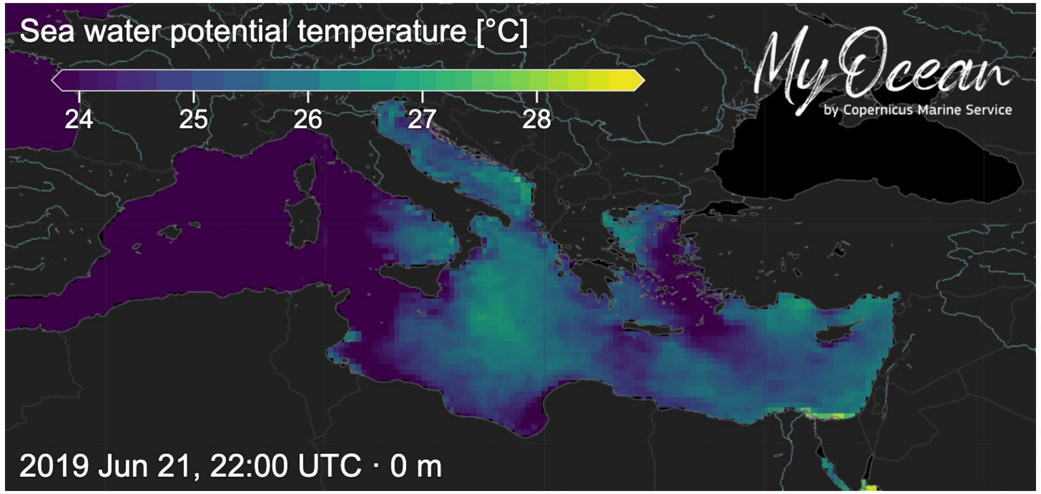


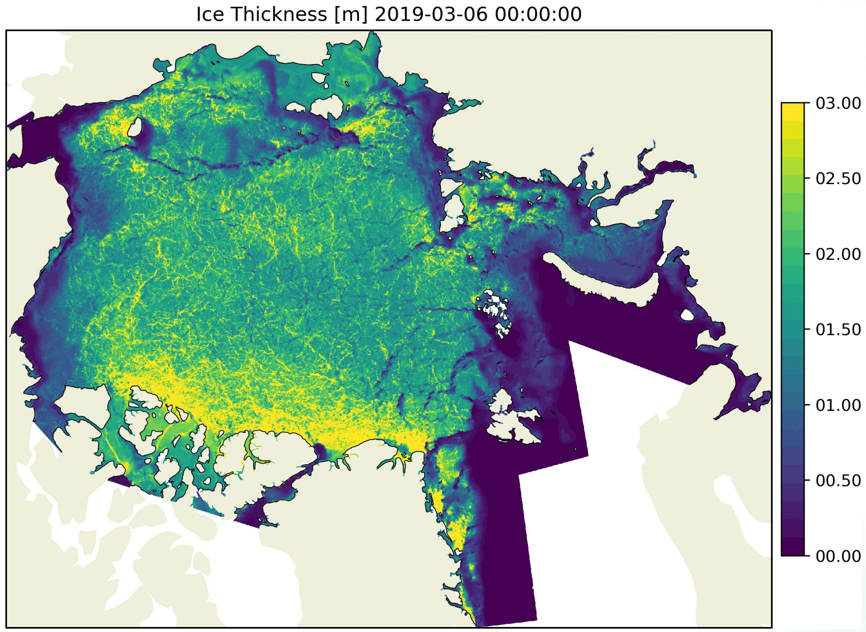




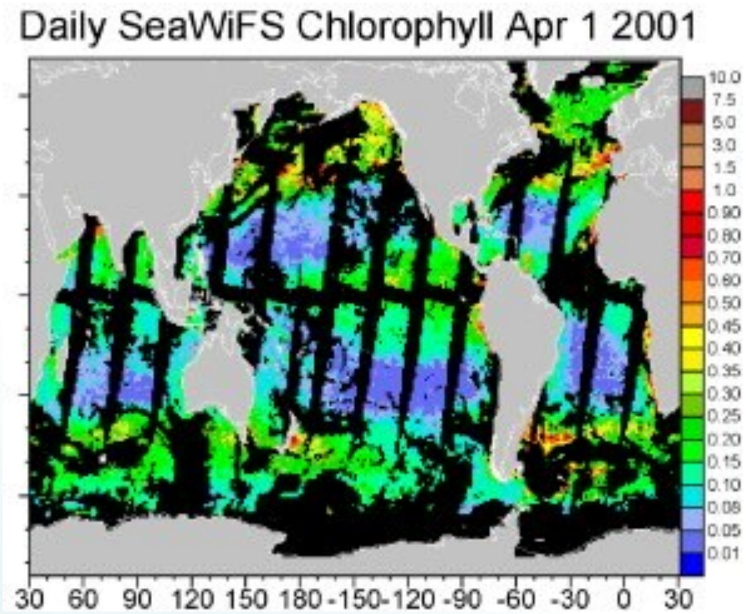
First source of knnowledge: numerical models

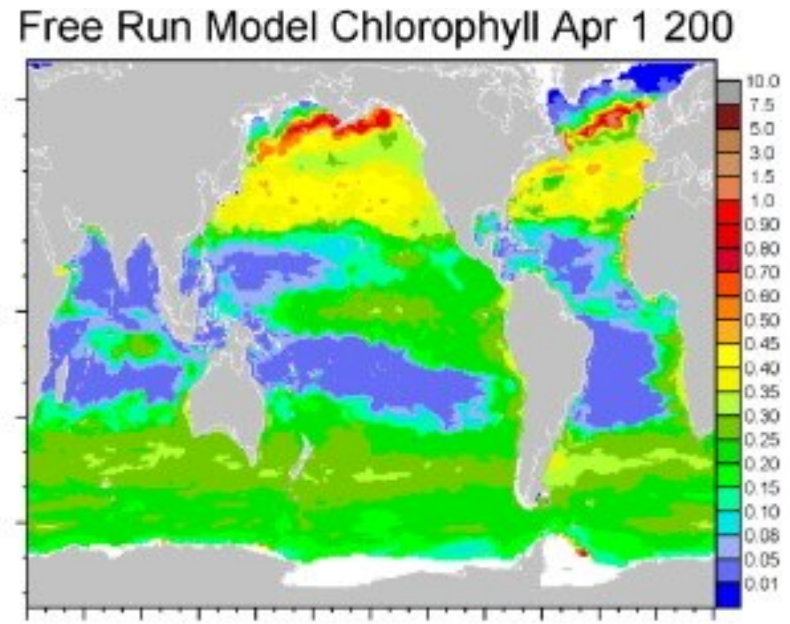






Comparison





Observations

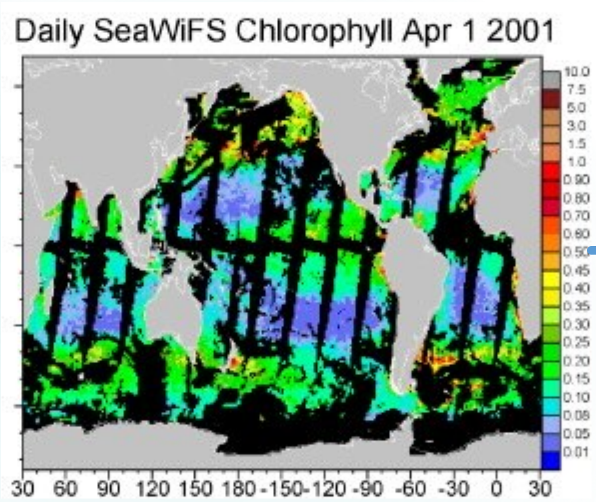
- **Sparse**
- **Noisy**

Numerical model

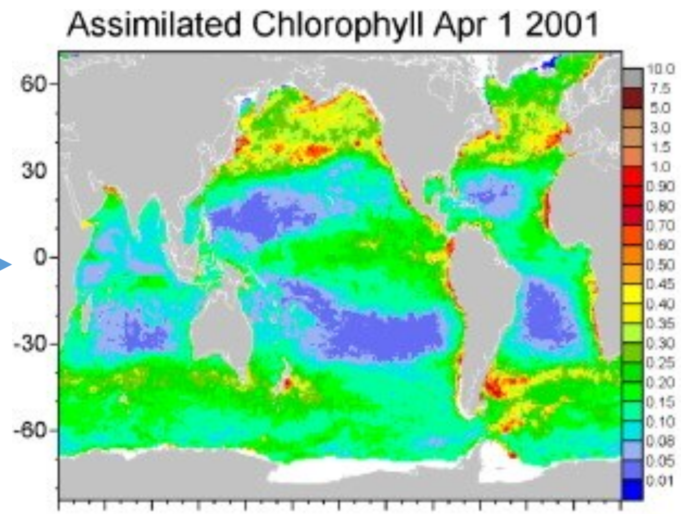
- **Biased**
- **Low-resolution** (for climate studies)
- Due to the chaotic nature, model will **diverge from the reality** after some time

Merging model/observation: data assimilation

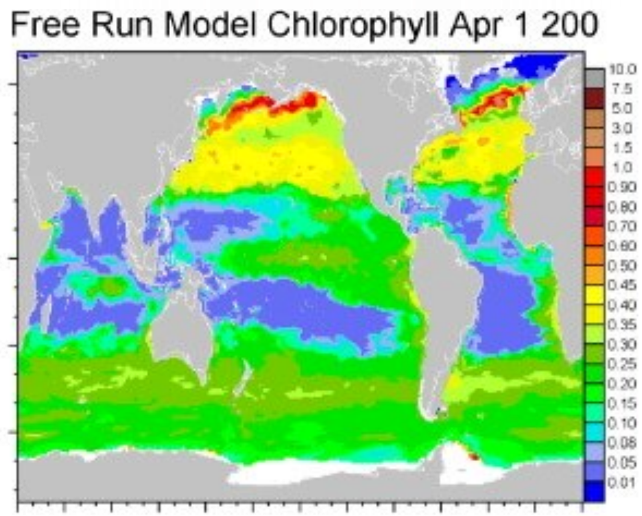
"The very best way to make a forecast out of a numerical model and a set of observations."



Observations



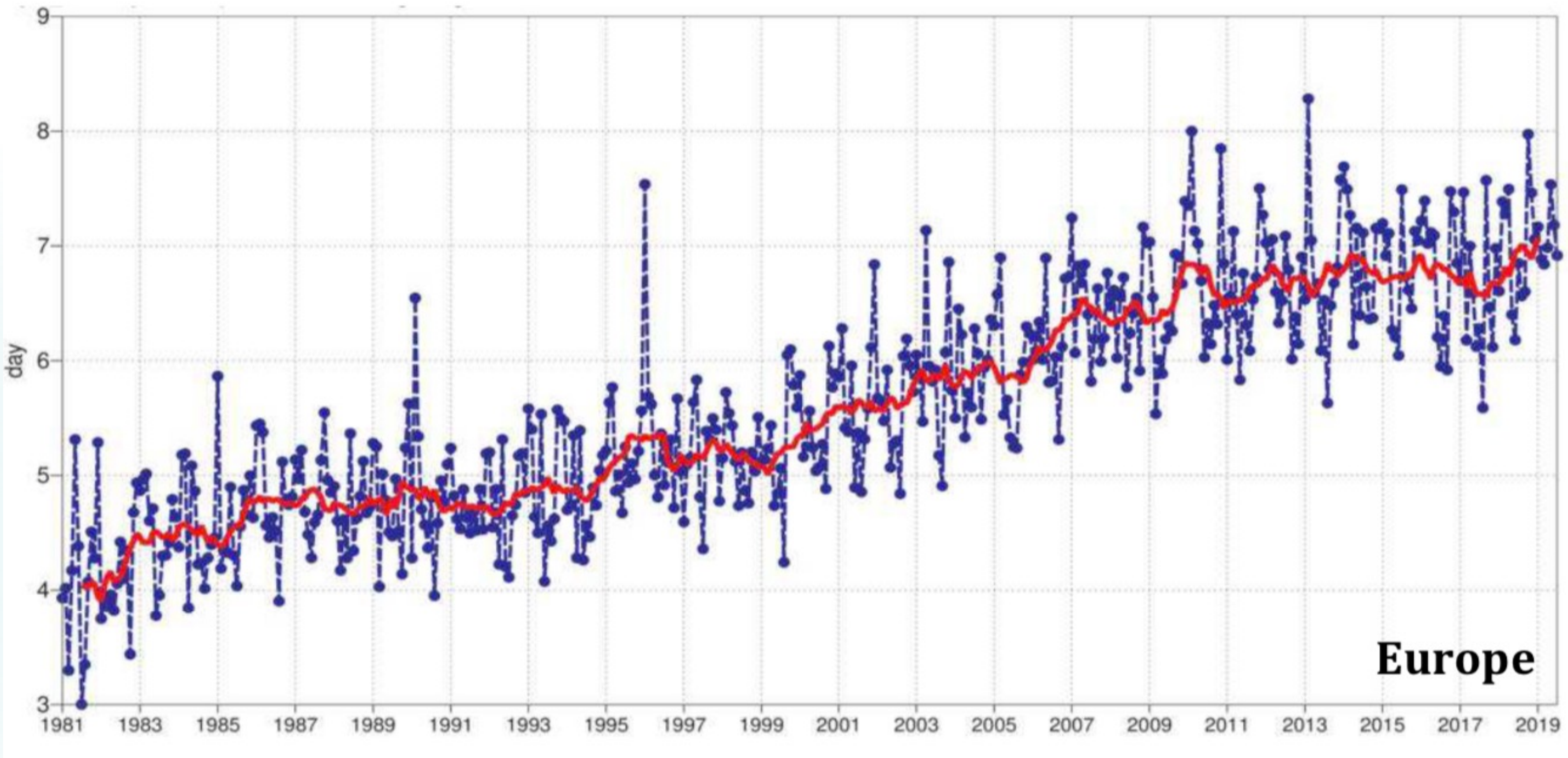
"Analysis" from data assimilation



Numerical model

Some achievements of DA

Horizon of prediction (in days) of the weather from ECMWF



What data assimilation does and does not?

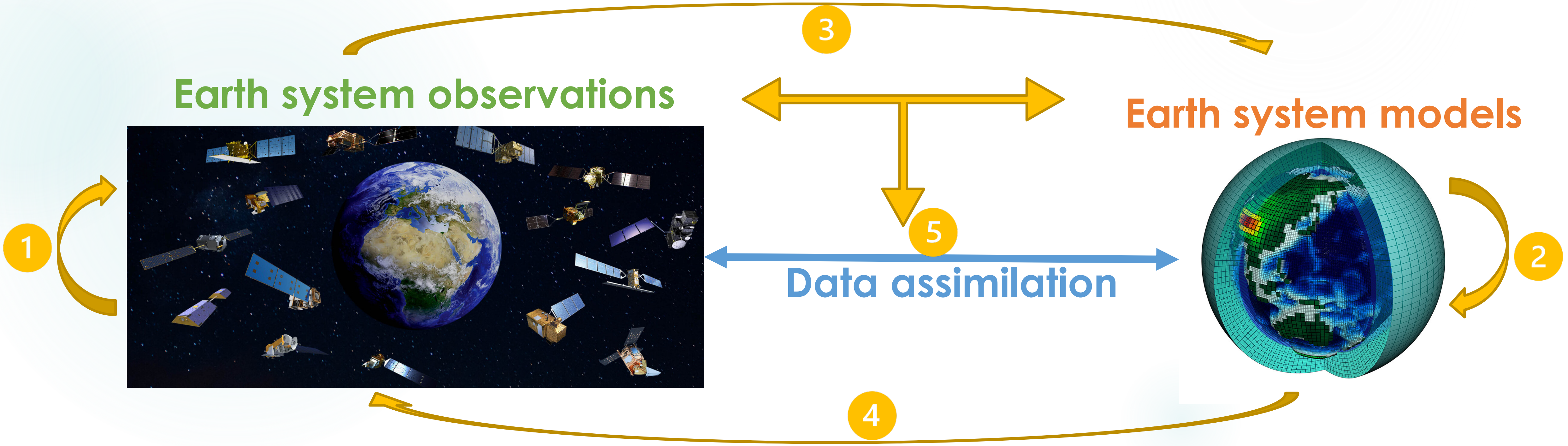
- Good at reconstructing **initial conditions** for forecast
- Handle **noisy and sparse observations**
- Give an accurate estimation of the **uncertainty**

BUT:

- Does **not "learn"**. The model is not "smarter" from the observations
- Some information contained in the observations is **lost**. (e.g. if observations are at a higher resolution than the model).

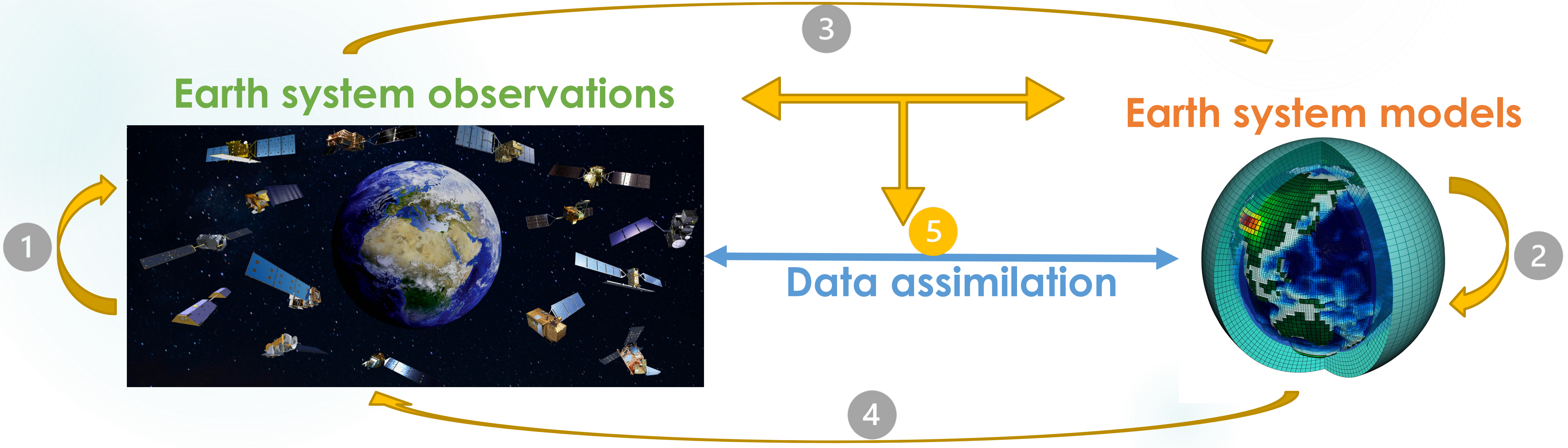
Machine learning to the rescue?

1. Learn from from observations and apply to observations (**data-driven model, now-casting, inference, interpolation, …**)
2. Learn from model and apply to model (**emulators, improved parametrization, …**)
3. Learn from observations and apply to model (**post-processing bias correction, …**)
4. Learn from model and apply to observations (**inference of new variables, downscaling, …**)
5. Learn at the intersection between observation and model via data assimilation (**improved parametrization, extend data assimilation, …**)

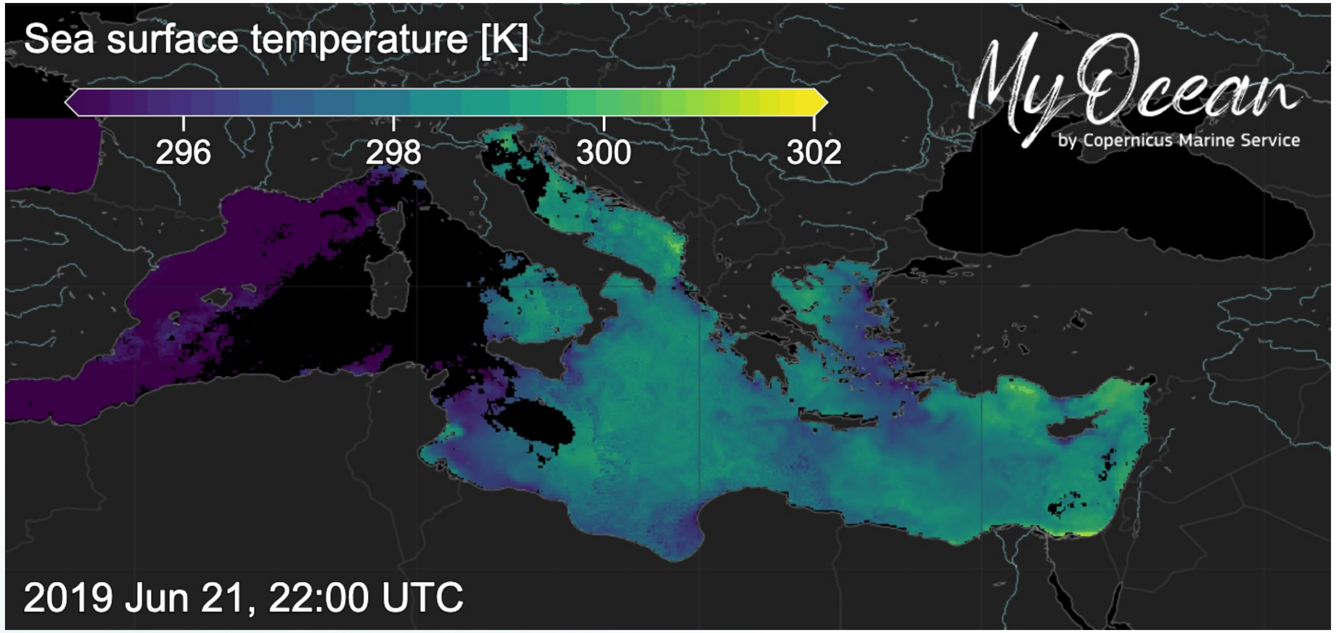


Machine learning to the rescue?

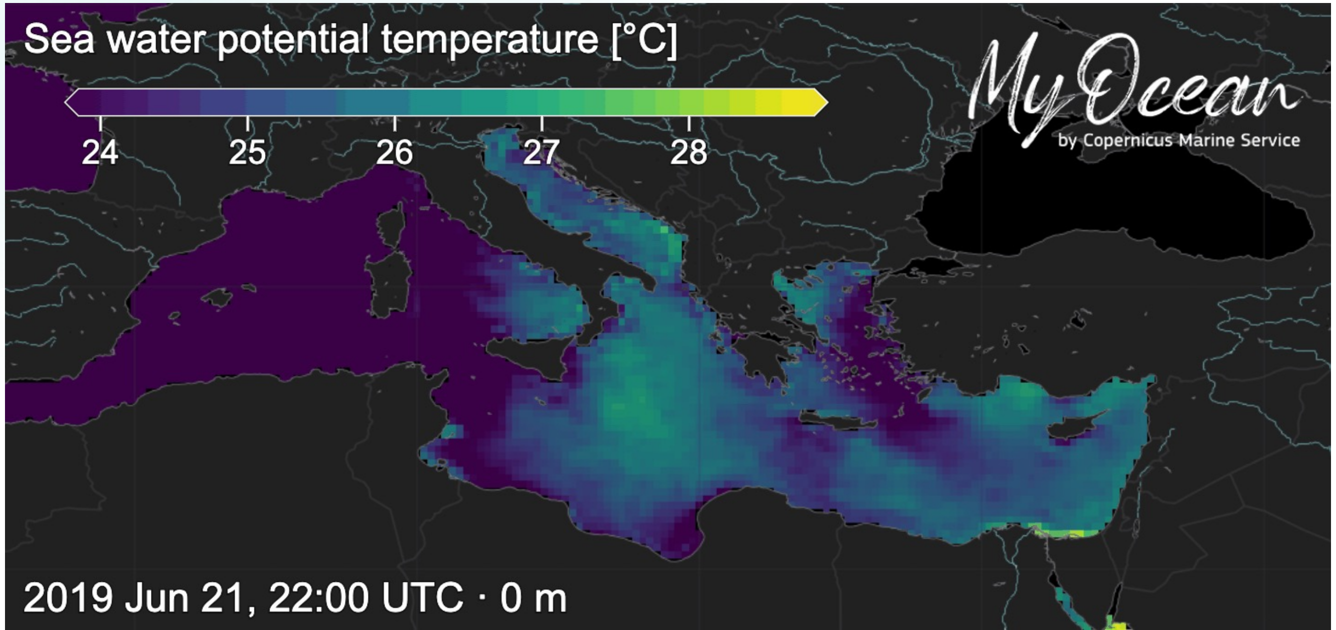
1. Learn from from observations and apply to observations (data-driven model, now-casting, inference, interpolation, …)
2. Learn from model and apply to model (emulators, improved parametrization, …)
3. Learn from observations and apply to model (post-processing bias correction, …)
4. Learn from model and apply to observations (inference of new variables, downscaling, …)
5. Learn at the intersection between observation and model via data assimilation (improved parametrization, extend data assimilation, …)



One application: extend data assimilation



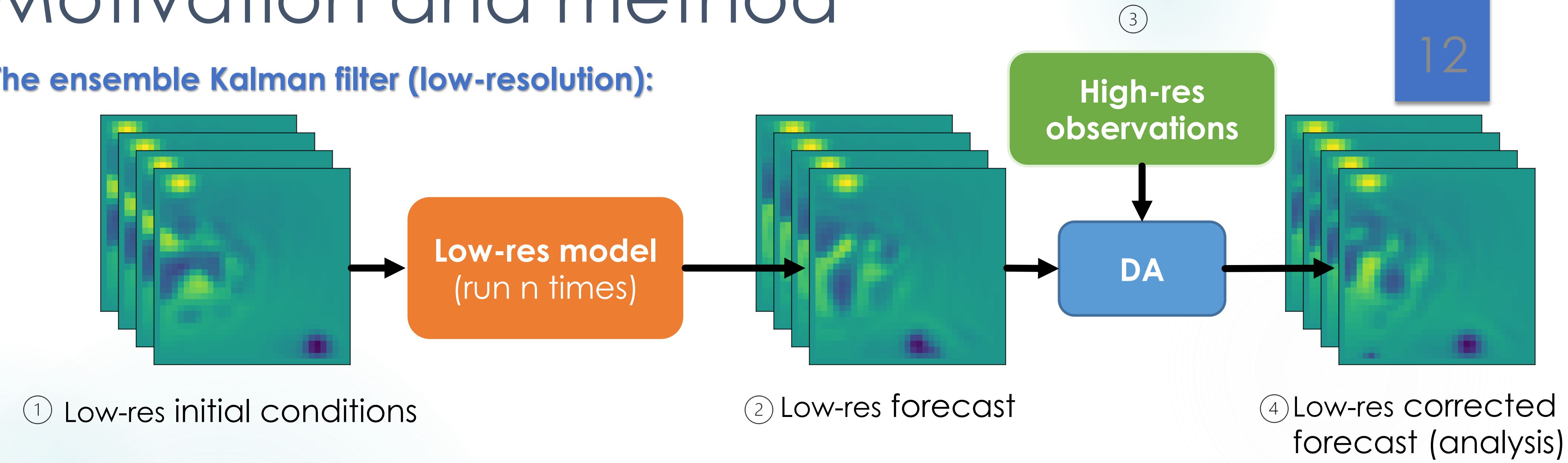
Observation (high-resolution)



Model output (lower resolution)

Motivation and method

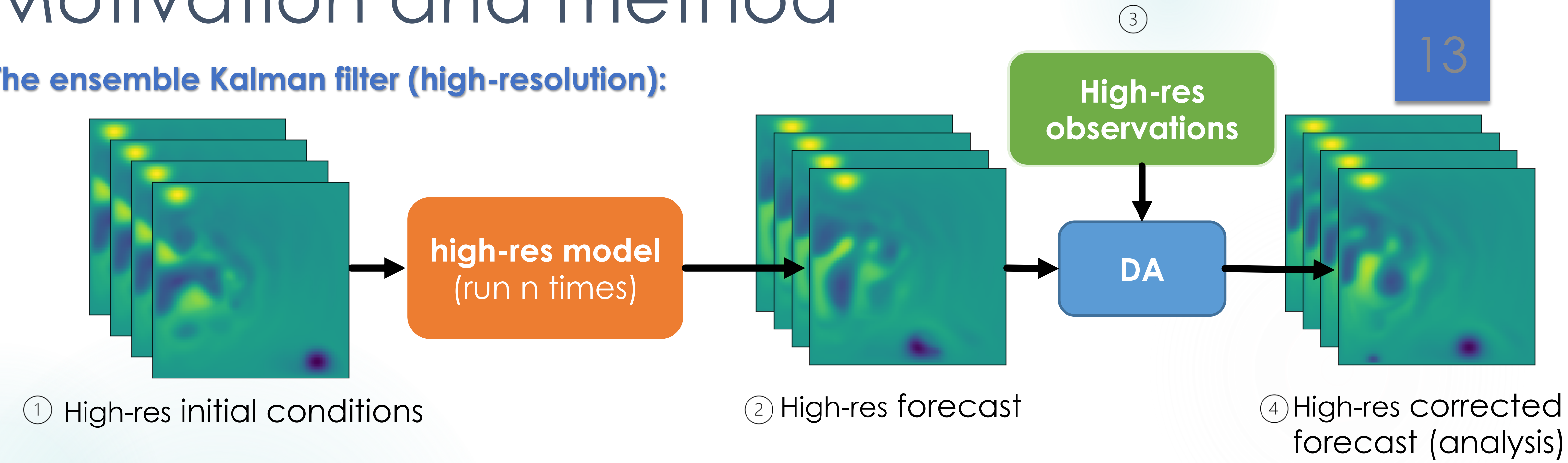
The ensemble Kalman filter (low-resolution):



	EnKF low-res
Computational cost	Low ✓
Ensemble size	Big ✓
Observation error	High ✗
High-res processes	Poorly resolved ✗

Motivation and method

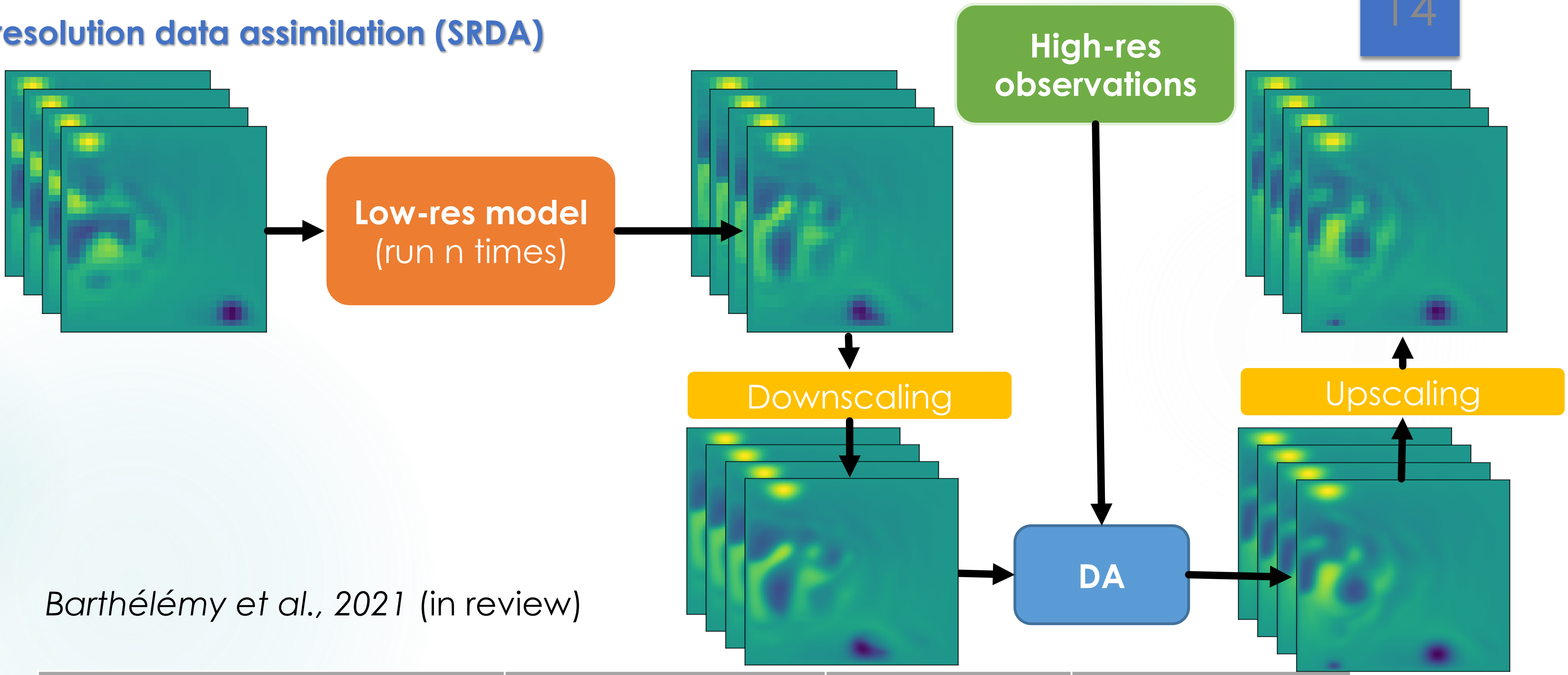
The ensemble Kalman filter (high-resolution):



	EnKF low-res	EnKF high-res
Computational cost	Low ✓	High ✗
Ensemble size	Big ✓	Small ✗
Observation error	High ✗	Low ✓
High-res processes	Poorly resolved ✗	Resolved ✓

Motivation and method

Super-resolution data assimilation (SRDA)



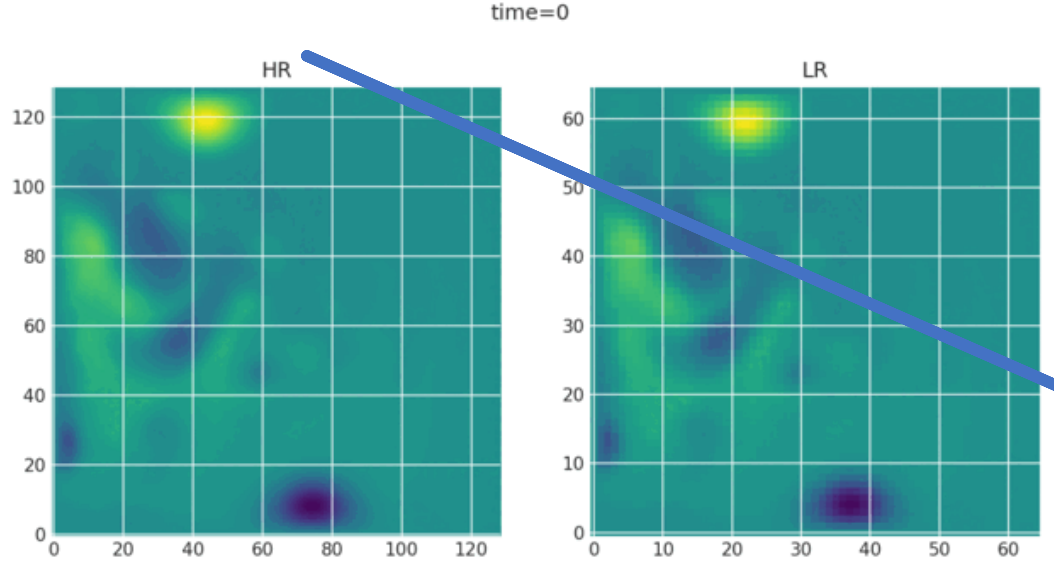
Barthélémy et al., 2021 (in review)

	EnKF low-res	EnKF high-res	SRDA
Computational cost	Low ✓	High ✗	Low ✓
Ensemble size	Big ✓	Small ✗	Big ✓
Observation error	High ✗	Low ✓	Low ✓
High-res processes	Poorly resolved ✗	Resolved ✓	Emulated –

Setup a numerical experiment

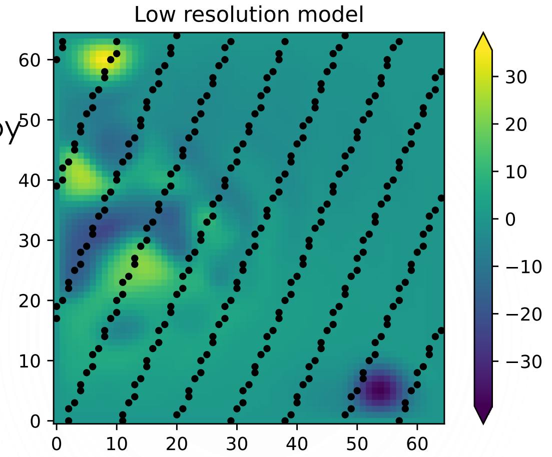
The model

A quasi-geostrophic model (represents the motion of the ocean surface)



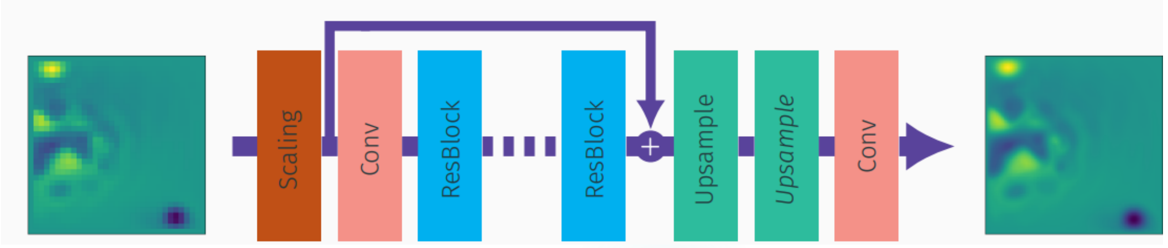
The observations

- **Synthetic observations:** Produce by high-resolution model (we know the expected result!)
- Mimic satellite altimeter observations
- **Sparse** and **noisy**

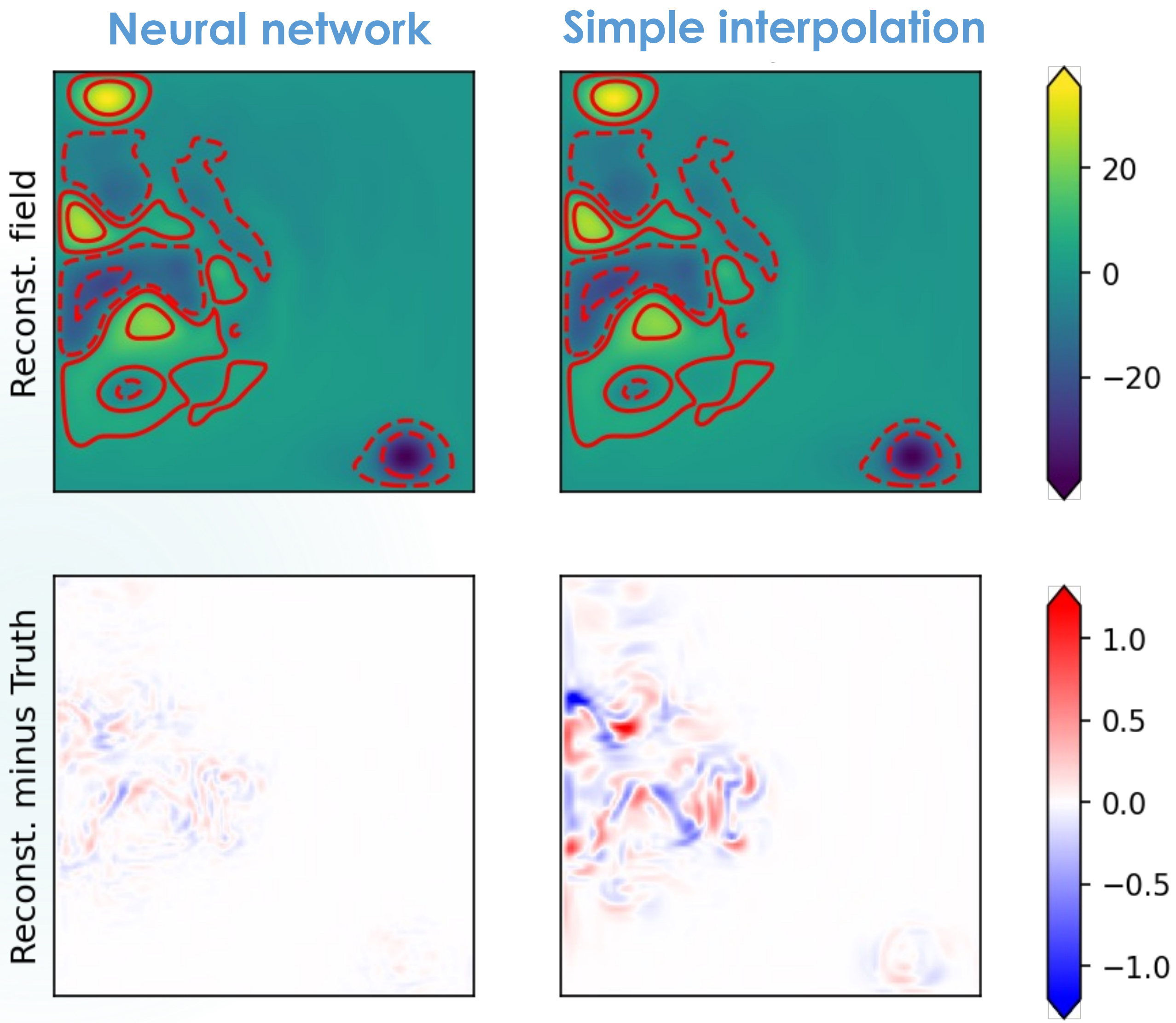


Downscaling

- A simple cubic spline **interpolation**
- A **neural network** trained using a free high-res simulation

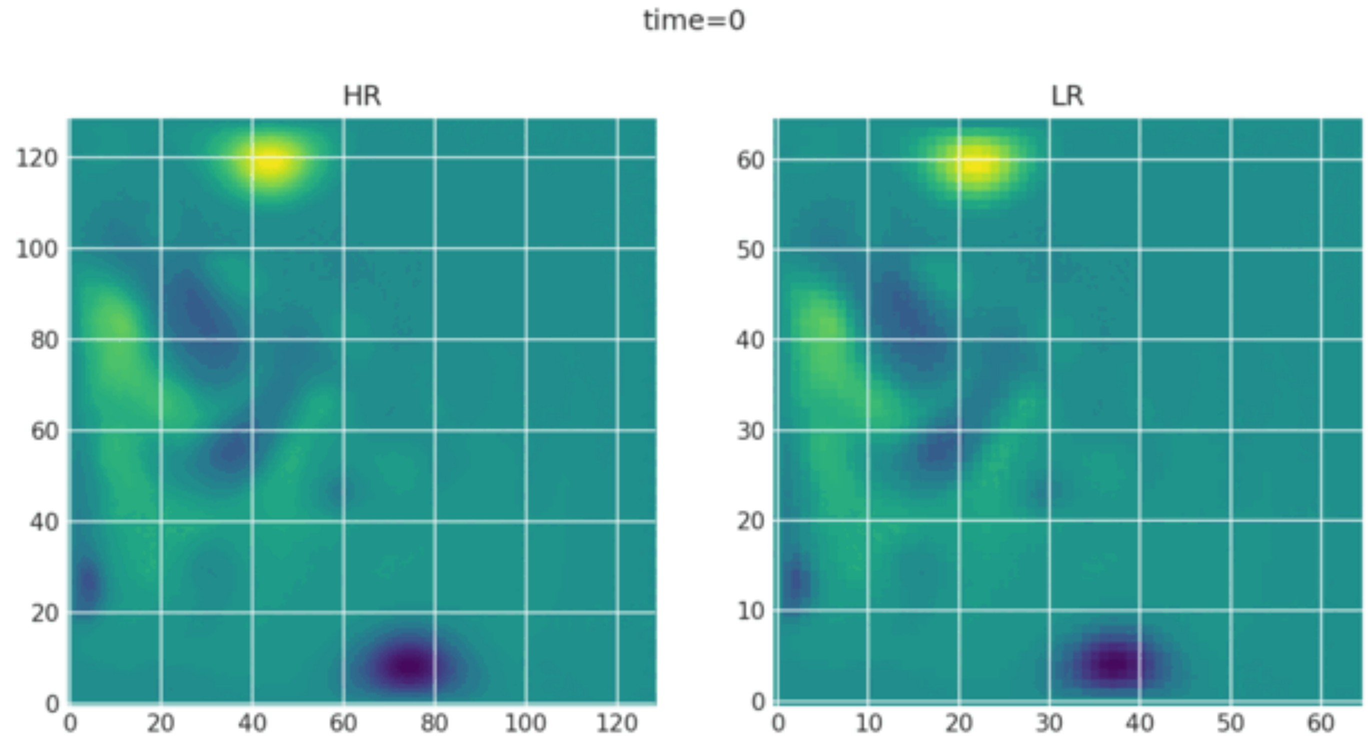


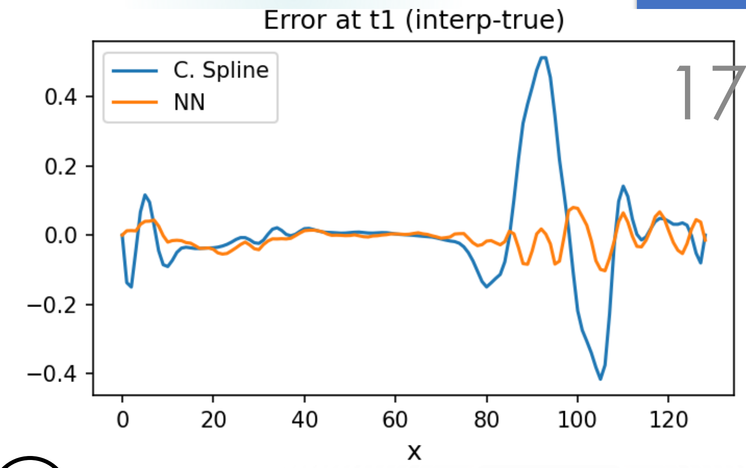
Downscaling performance

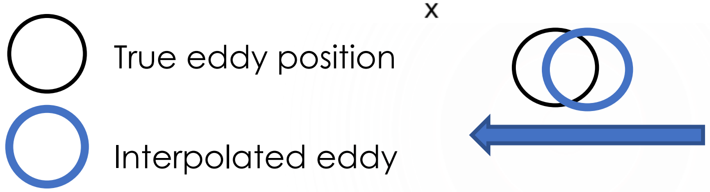


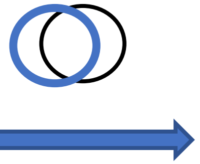
Red contours: true high-resolution field

Correction of model error









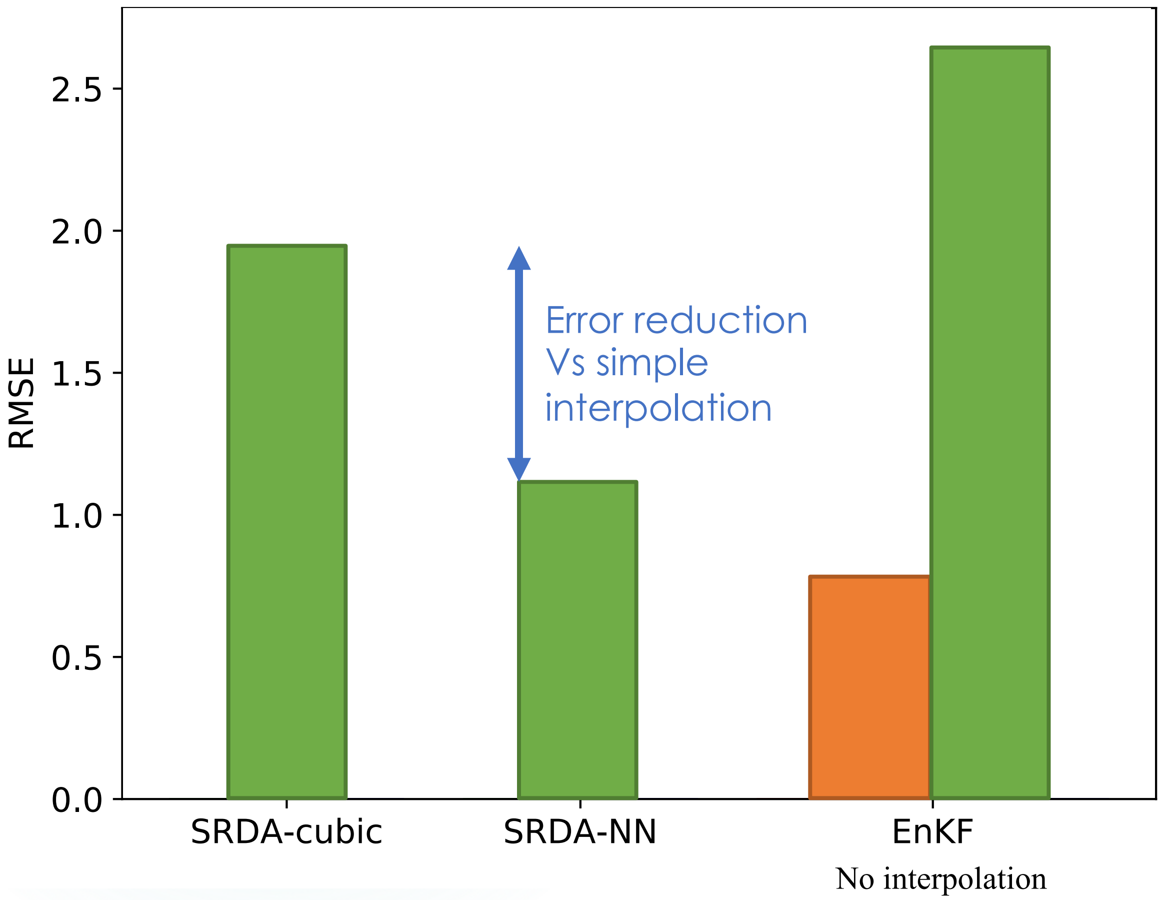
Error of the low-res. model

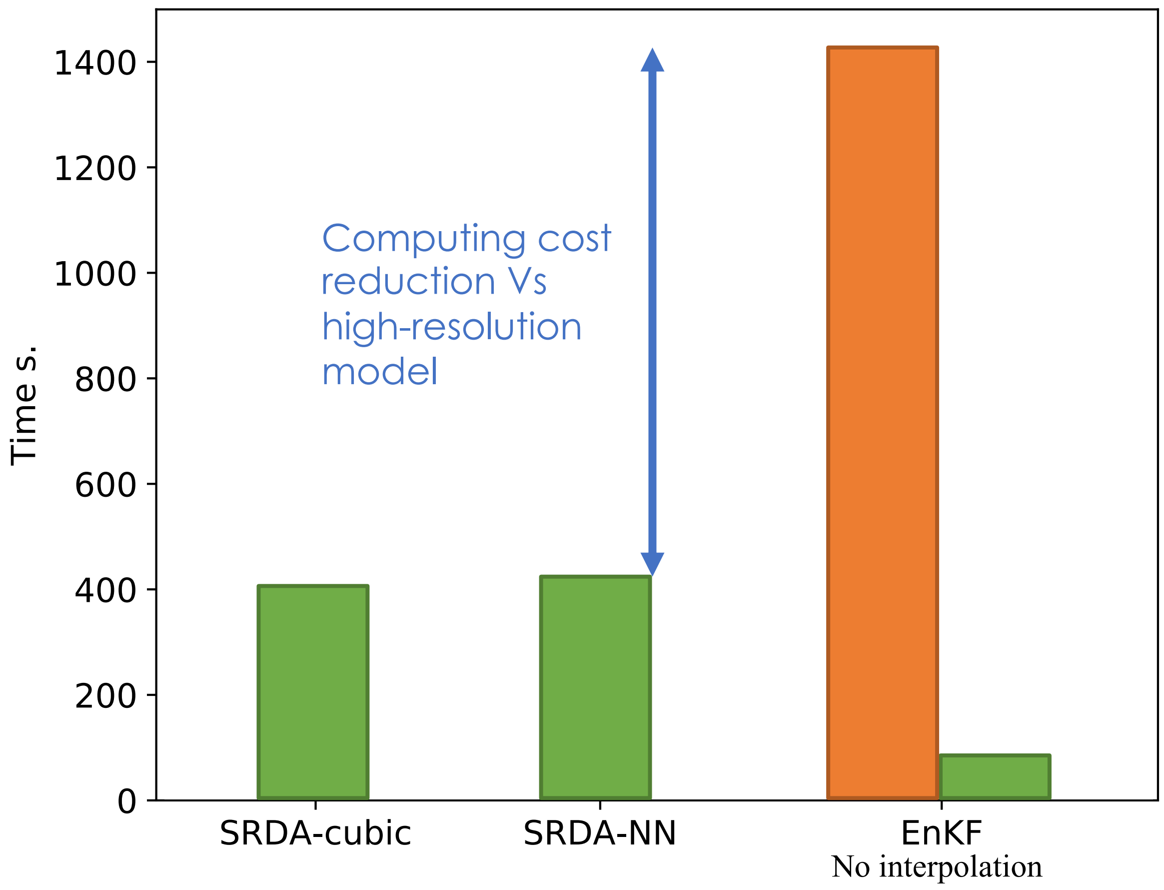
- Eddy propagation are **too slow** in the low-res model
- The neural network has **learnt this displacement error.**

Performance of SRDA

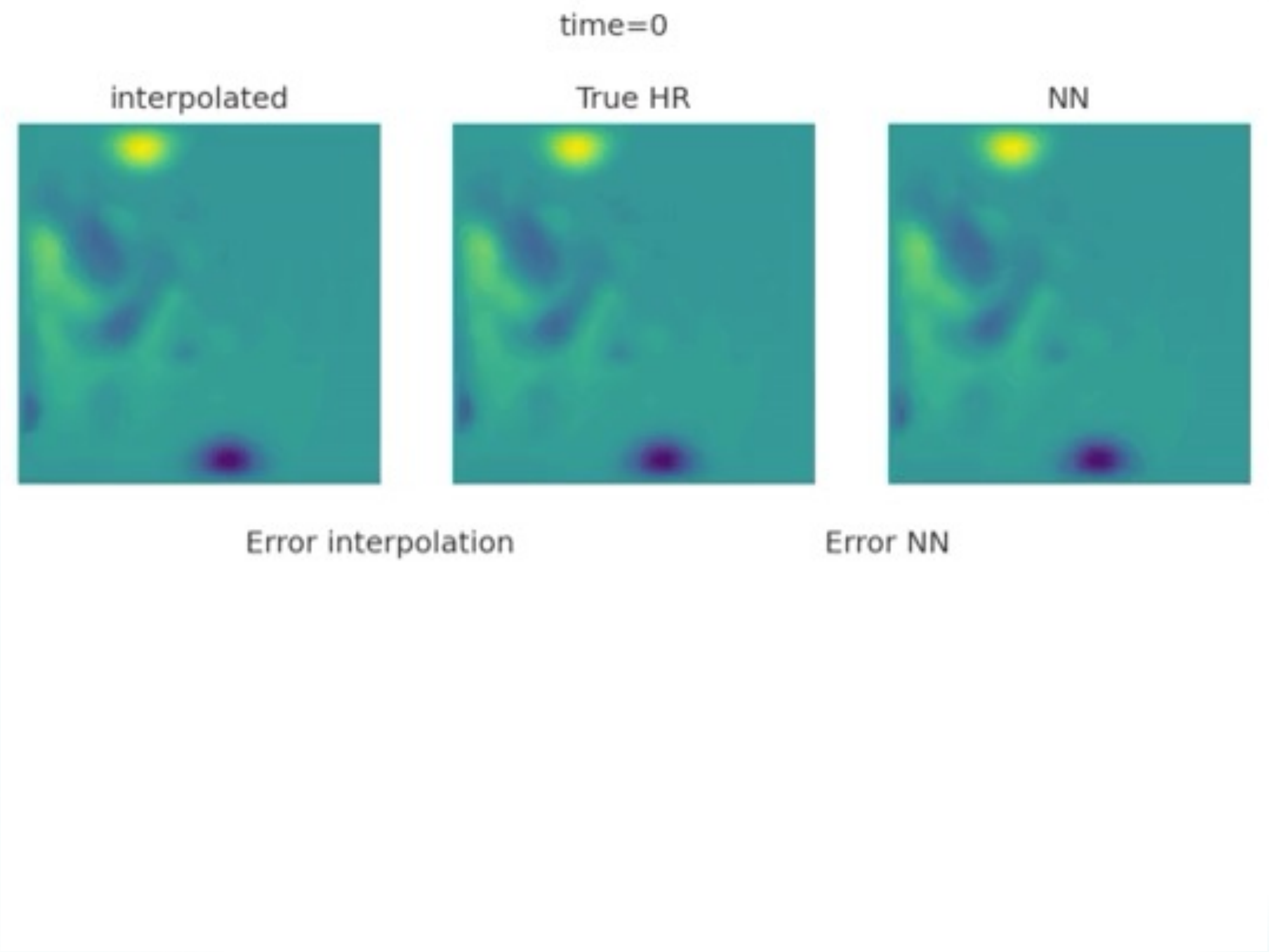
Using a high-resolution model

Using a low-resolution model



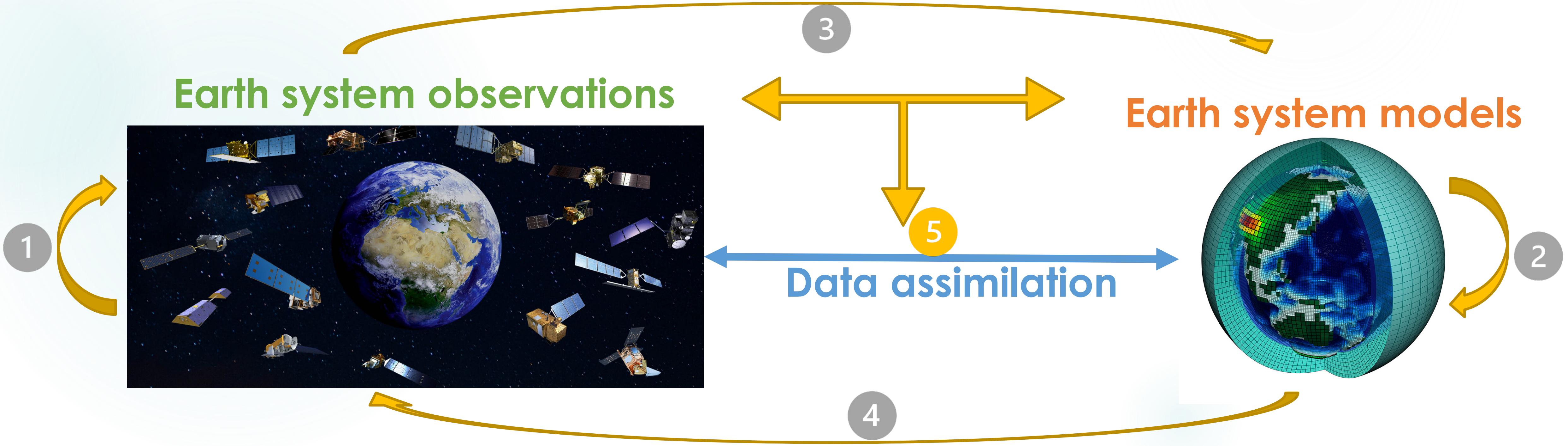


Have we learned something?

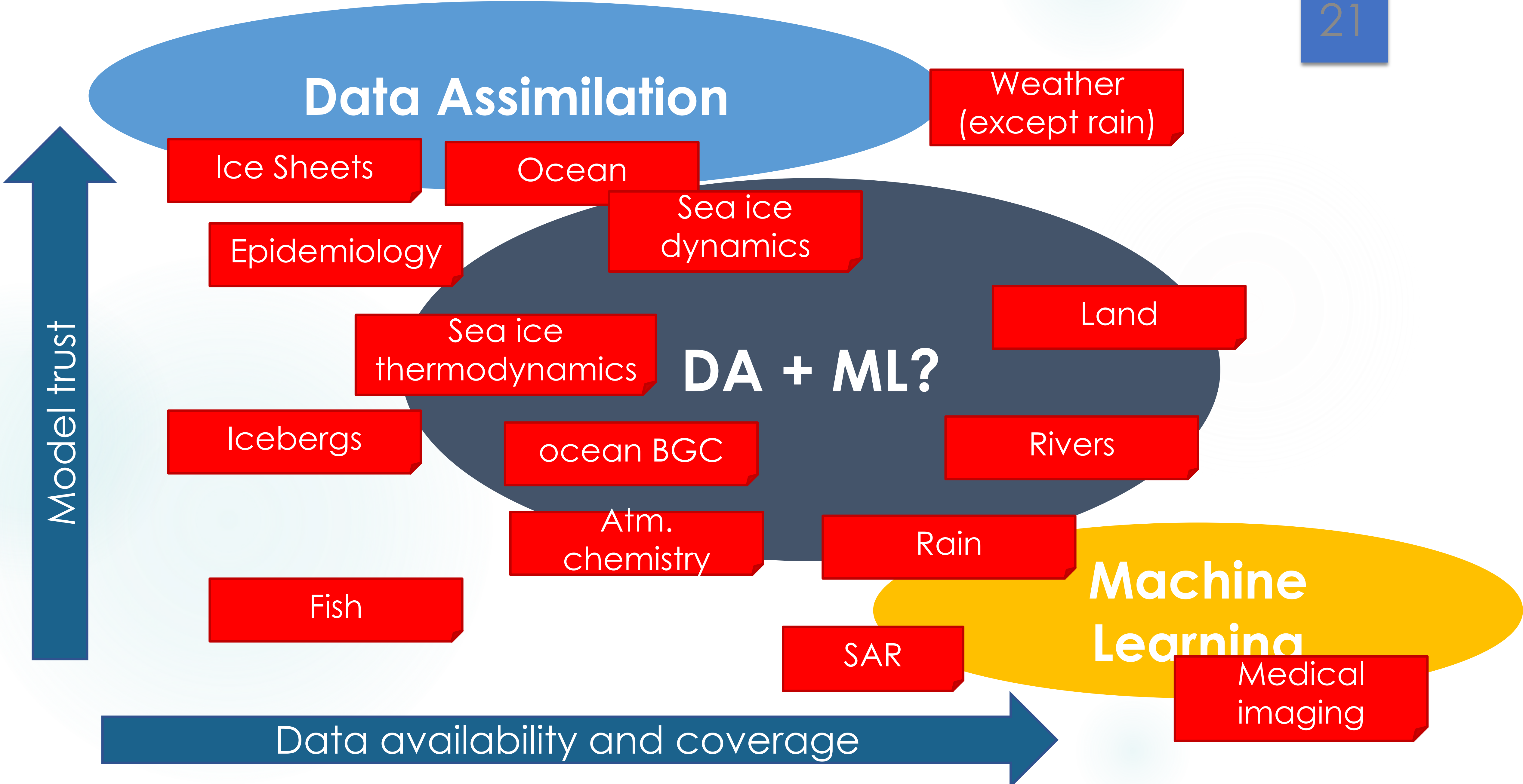


Machine learning to the rescue?

1. Learn from from observations and apply to observations (**data-driven model, now-casting, inference, interpolation, …**)
2. Learn from model and apply to model (**emulators, improved parametrization, …**)
3. Learn from observations and apply to model (**post-processing bias correction, …**)
4. Learn from model and apply to observations (**inference of new variables, downscaling, …**)
5. Learn at the intersection between observation and model via data assimilation (**improved parametrization, extend data assimilation, …**)



Cases of applications

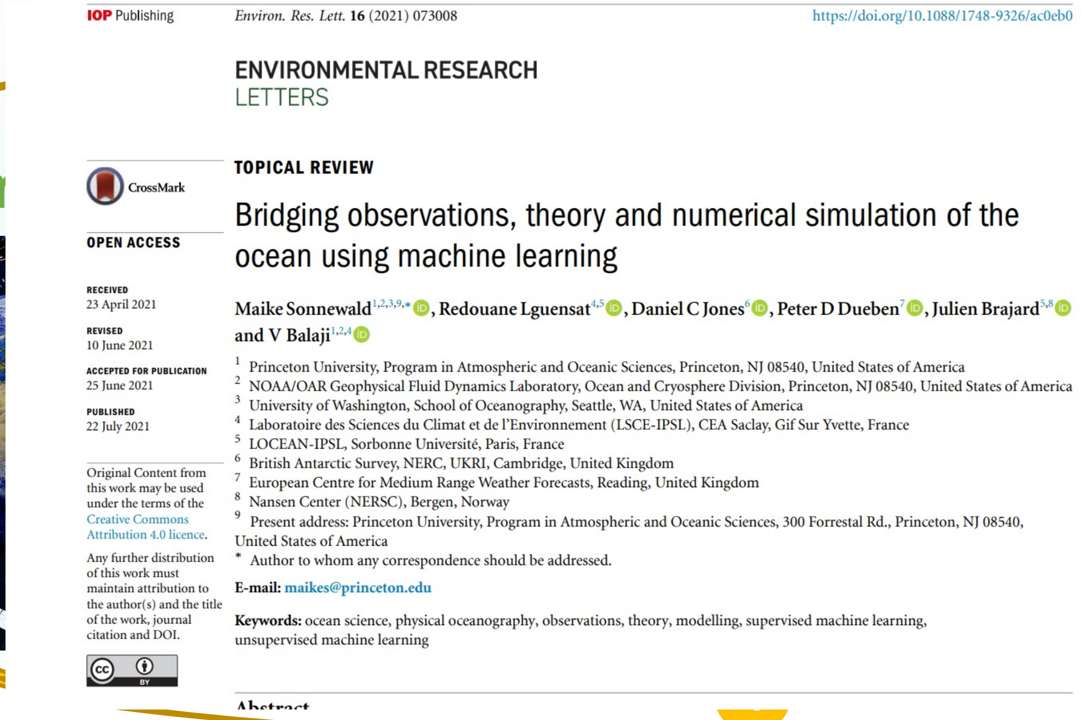


Conclusions

- Machine learning can **complement** model and observations to improve our understanding and the prediction of the ocean/sea-ice system.
- In particular, it can be very efficient to **enhance under-represented scales** in model and observations.
- More application in the review article: “Bridging observation, theory and numerical simulation of the ocean using Machine Learning”. https://doi.org/10.1088/1748-9326/ac0eb0

Earth system

1



IOP Publishing Environ. Res. Lett. 16 (2021) 073008 https://doi.org/10.1088/1748-9326/ac0eb0

ENVIRONMENTAL RESEARCH
LETTERS

CrossMark

OPEN ACCESS

RECEIVED
23 April 2021

REVISED
10 June 2021

ACCEPTED FOR PUBLICATION
25 June 2021

PUBLISHED
22 July 2021

TOPICAL REVIEW

Bridging observations, theory and numerical simulation of the ocean using machine learning

Maike Sonnewald[1,2,3,9,*], Redouane Lguensat[4,5], Daniel C Jones[6], Peter D Dueben[7], Julien Brajard[5,8] and V Balaji[1,2,4]

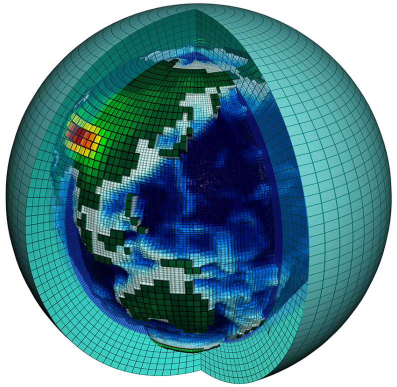
1 Princeton University, Program in Atmospheric and Oceanic Sciences, Princeton, NJ 08540, United States of America
2 NOAA/OAR Geophysical Fluid Dynamics Laboratory, Ocean and Cryosphere Division, Princeton, NJ 08540, United States of America
3 University of Washington, School of Oceanography, Seattle, WA, United States of America
4 Laboratoire des Sciences du Climat et de l’Environnement (LSCE-IPSL), CEA Saclay, Gif Sur Yvette, France
5 LOCEAN-IPSL, Sorbonne Université, Paris, France
6 British Antarctic Survey, NERC, UKRI, Cambridge, United Kingdom
7 European Centre for Medium Range Weather Forecasts, Reading, United Kingdom
8 Nansen Center (NERSC), Bergen, Norway
9 Present address: Princeton University, Program in Atmospheric and Oceanic Sciences, 300 Forrestal Rd., Princeton, NJ 08540, United States of America
* Author to whom any correspondence should be addressed.

E-mail: maikes@princeton.edu

Keywords: ocean science, physical oceanography, observations, theory, modelling, supervised machine learning, unsupervised machine learning

Abstract

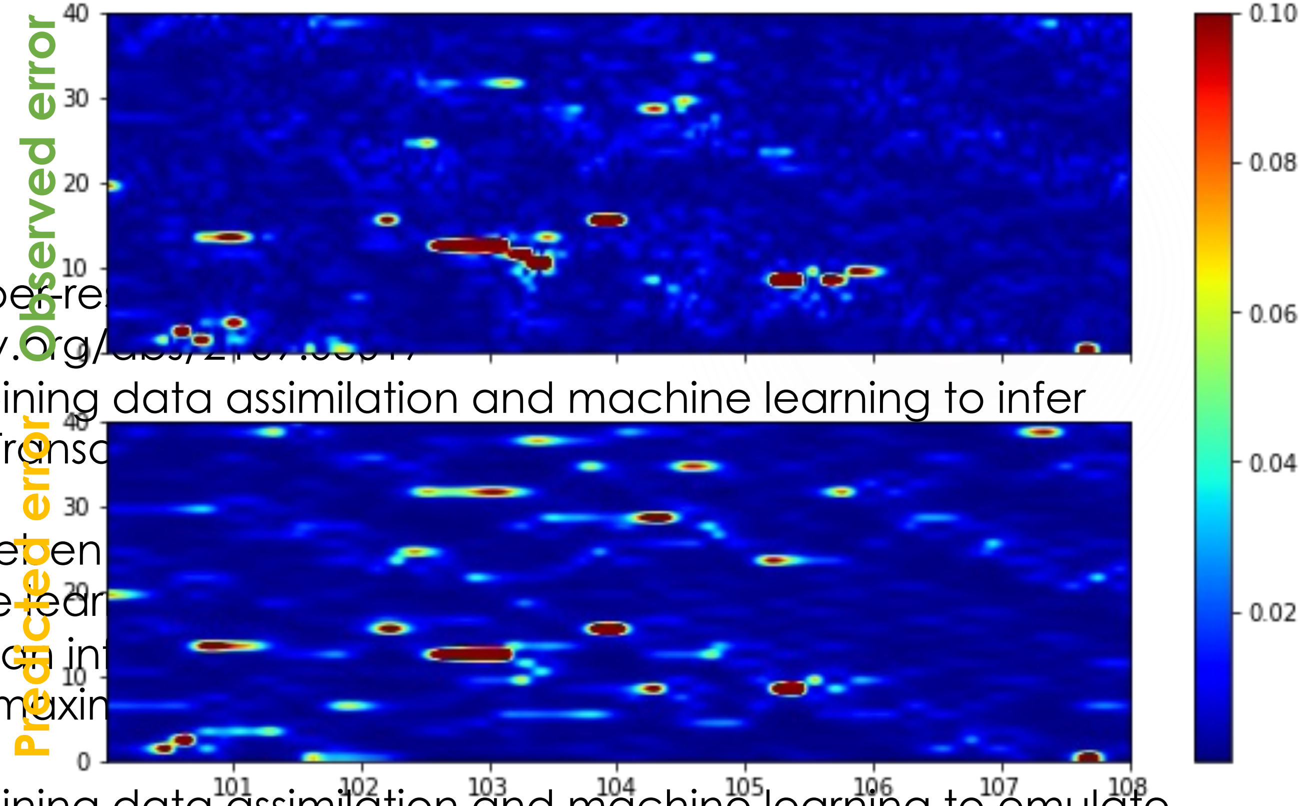
Earth system models



2

Challenging questions

- Uncertainty estimate (natural in any DA framework)
- Non-autonomous systems (e.g. in a global warming context)



https://github.com/nansencenter/ml-crashcourse/blob/main/L96_demonstrator.ipynb

References

- Barthélémy S, Brajard J, Bertino L, Counillon F. Super-re
- Submitted to ocean modelling, 2021. https://arxiv.org/
- Brajard J, Carrassi A, Bocquet M, Bertino L. Combining data assimilation and machine learning to infer unresolved scale parametrization. Philosophical Transa Physical and Engineering Sciences. 2021;
- Sonnewald M, Lguensat R, Jones D, Brajard J, Due en numerical simulation of the ocean using machine lear
- Bocquet M, Brajard J, Carrassi A, Bertino L. Bayesian in assimilation, machine learning and expectation-maxim 2020;2(1).
- Brajard J, Carrassi A, Bocquet M, Bertino L. Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the Lorenz 96 model. Journal of Computational Science. 2020;44.

Contact me: julien.brajard@nersc.no

NERSC

References

- Barthélémy S, Brajard J, Bertino L, Counillon F. Super-resolution data assimilation
- Submitted to ocean modelling, 2021. https://arxiv.org/abs/2109.08017
- Brajard J, Carrassi A, Bocquet M, Bertino L. Combining data assimilation and machine learning to infer unresolved scale parametrization. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2021;
- Sonnewald M, Lguensat R, Jones D, Brajard J, Dueben P, Balaji V. Bridging observations, theory and numerical simulation of the ocean using machine learning. Environmental Research Letters. 2021;
- Bocquet M, Brajard J, Carrassi A, Bertino L. Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization. Foundations of Data Science (FoDS). 2020;2(1).
- Brajard J, Carrassi A, Bocquet M, Bertino L. Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the Lorenz 96 model. Journal of Computational Science. 2020;44.