

# Emulators and DA

An introduction

Julien Brajard

# Scope of this presentation

Objective: foster the discussion on the links between data assimilation ML-based emulators

## Outline

- Quick introduction on weather forecast emulators
- What could be the near future of the integrated forecasting system
- What is the link with DA and what does it mean for the DA community?

# Back in time...2018

## Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Machine Learning Approach

Jaideep Pathak,<sup>1,2</sup> Brian Hunt,<sup>3,4</sup> Michelle Girvan,<sup>1,3,2</sup> Zhixin Lu,<sup>1,3</sup> and Edward Ott<sup>1,2,5</sup>

<sup>1</sup>*Institute for Research in Electronics and Applied Physics,  
University of Maryland, College Park, Maryland 20742, USA*

<sup>2</sup>*Department of Physics, University of Maryland, College Park, Maryland 20742, USA.*

<sup>3</sup>*Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA*

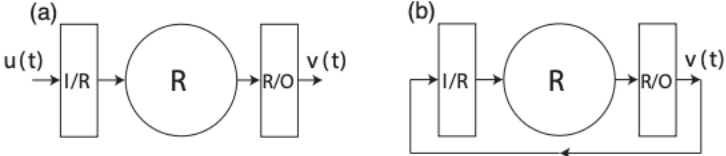
<sup>4</sup>*Department of Mathematics, University of Maryland, College Park, Maryland 20742, USA.*

<sup>5</sup>*Department of Electrical and Computer Engineering, University of Maryland, Maryland 20742, USA.*

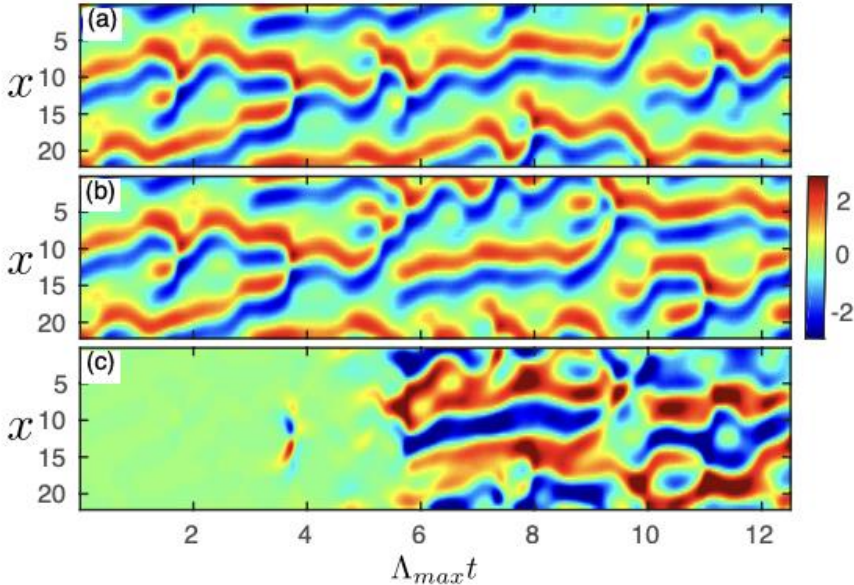
(Dated: December 12, 2017)

We demonstrate the effectiveness of using machine learning for model-free prediction of spatiotemporally chaotic systems of *arbitrarily large spatial extent and attractor dimension* purely from observations of the system's past evolution. We present a parallel scheme with an example implementation based on the reservoir computing paradigm and demonstrate the scalability of our scheme using the Kuramoto-Sivashinsky equation as an example of a spatiotemporally chaotic system.

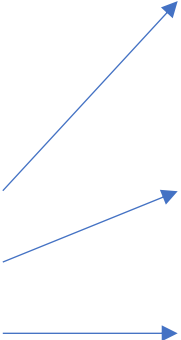
### Reservoir computing



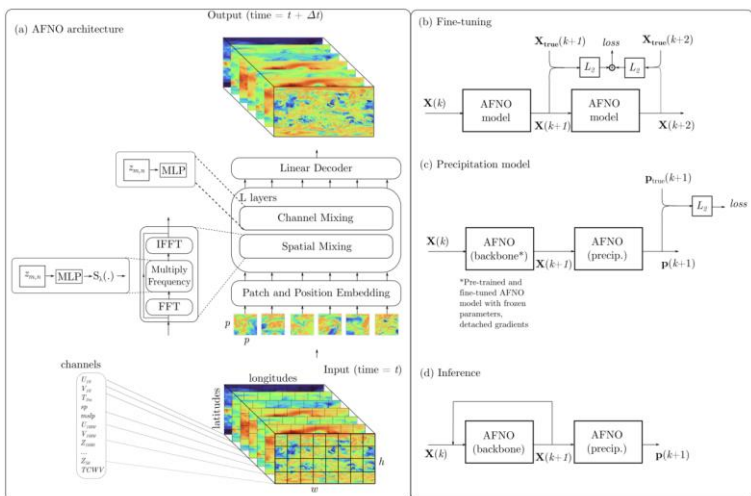
### Kuramoto-Sivashinsky (KS) equation



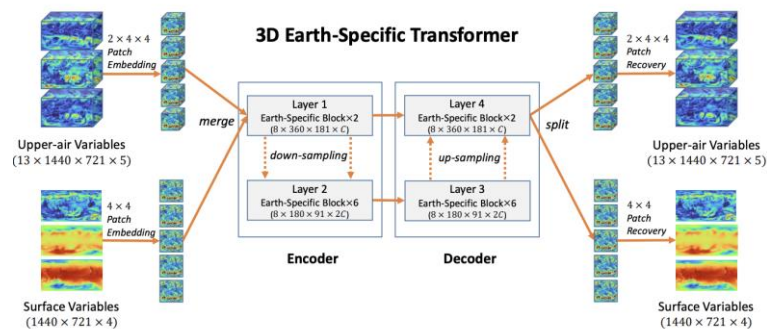
Physical-based model  
ML learning emulator  
Difference



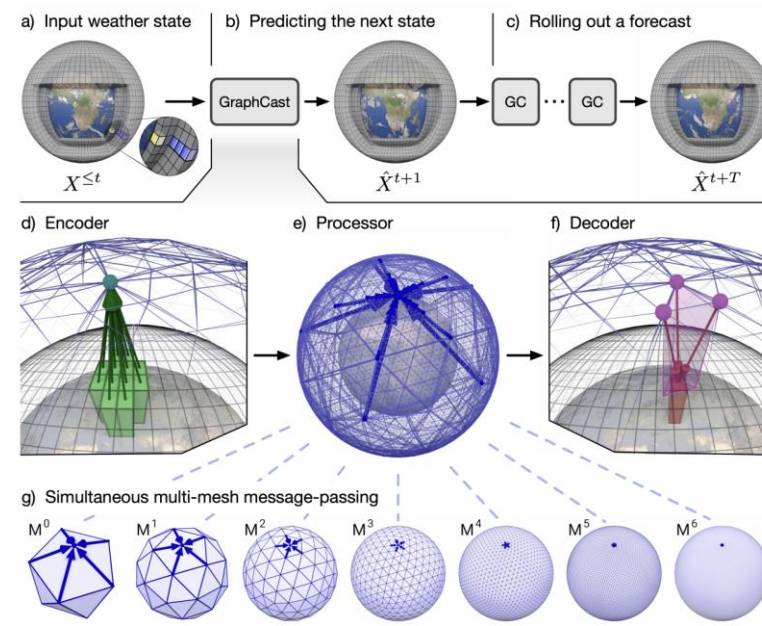
# Four years later: 2022



Pathak et al. (NVIDIA, ...)  
**Fourcastnet**



Bi et al. (Huawei Cloud Computing)  
**Pangu-Weather**



Lam et al. (Deepmind, ...)  
**Graphcast**

Feb

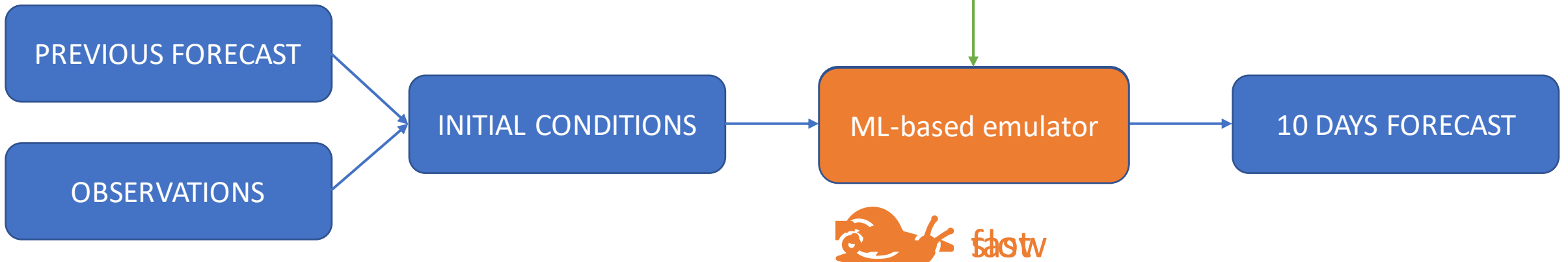
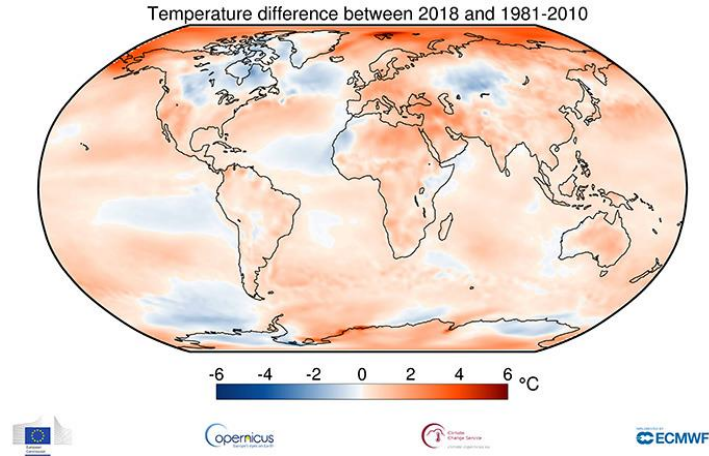
Nov

24th Dec

2022

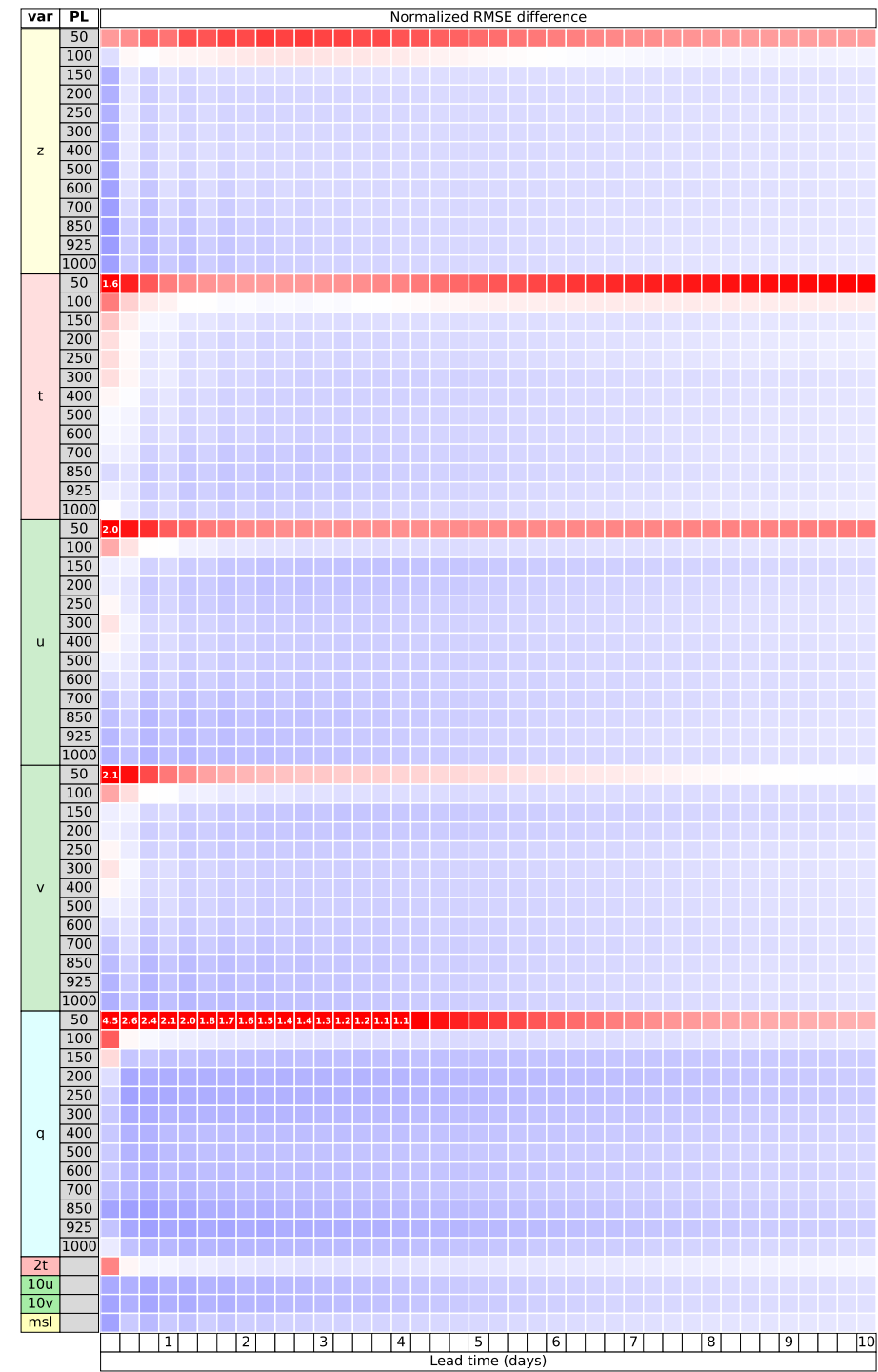
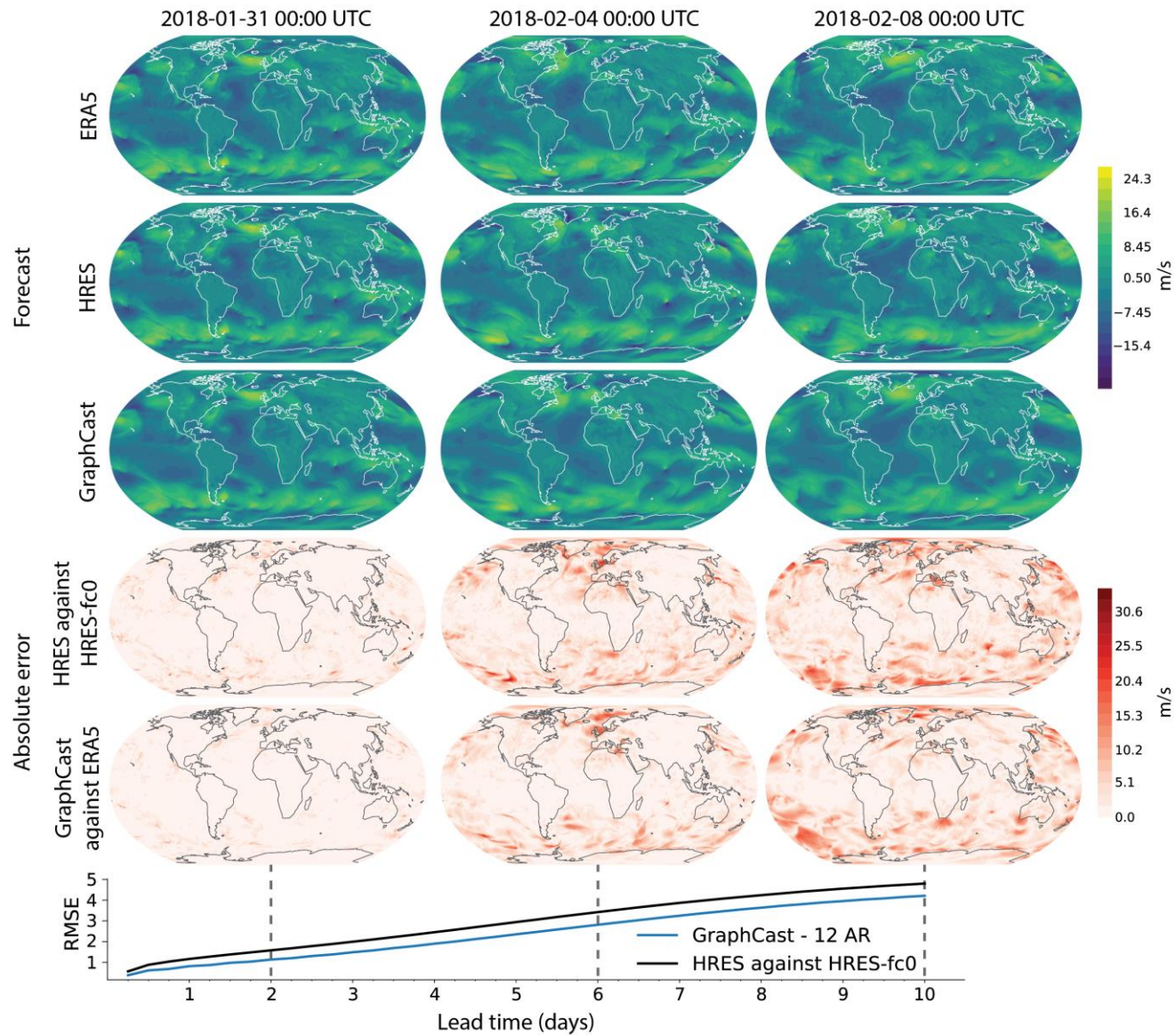
# What is the objective of a ML based emulator?

ERA5 1979-2018



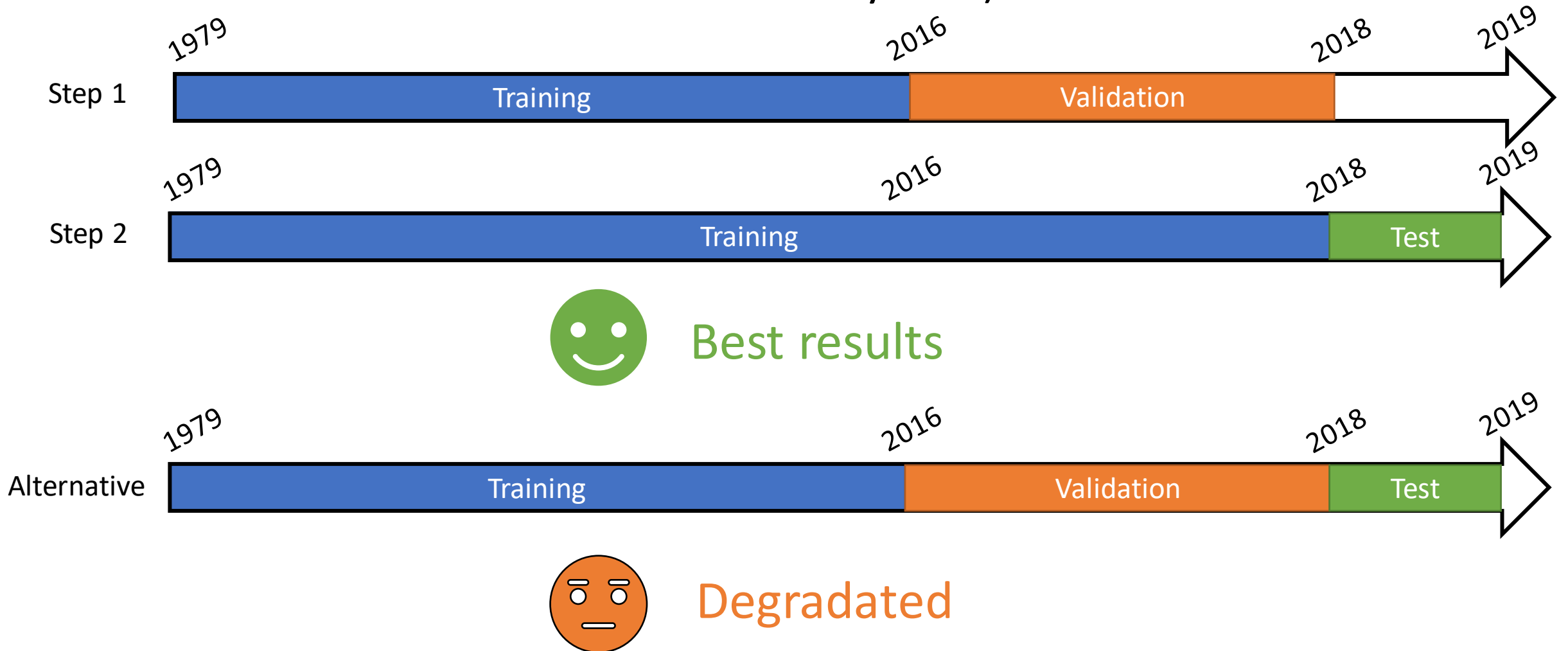
# Skill of Graphcast

10u: Initialization time 2018-01-29 00:00 UTC

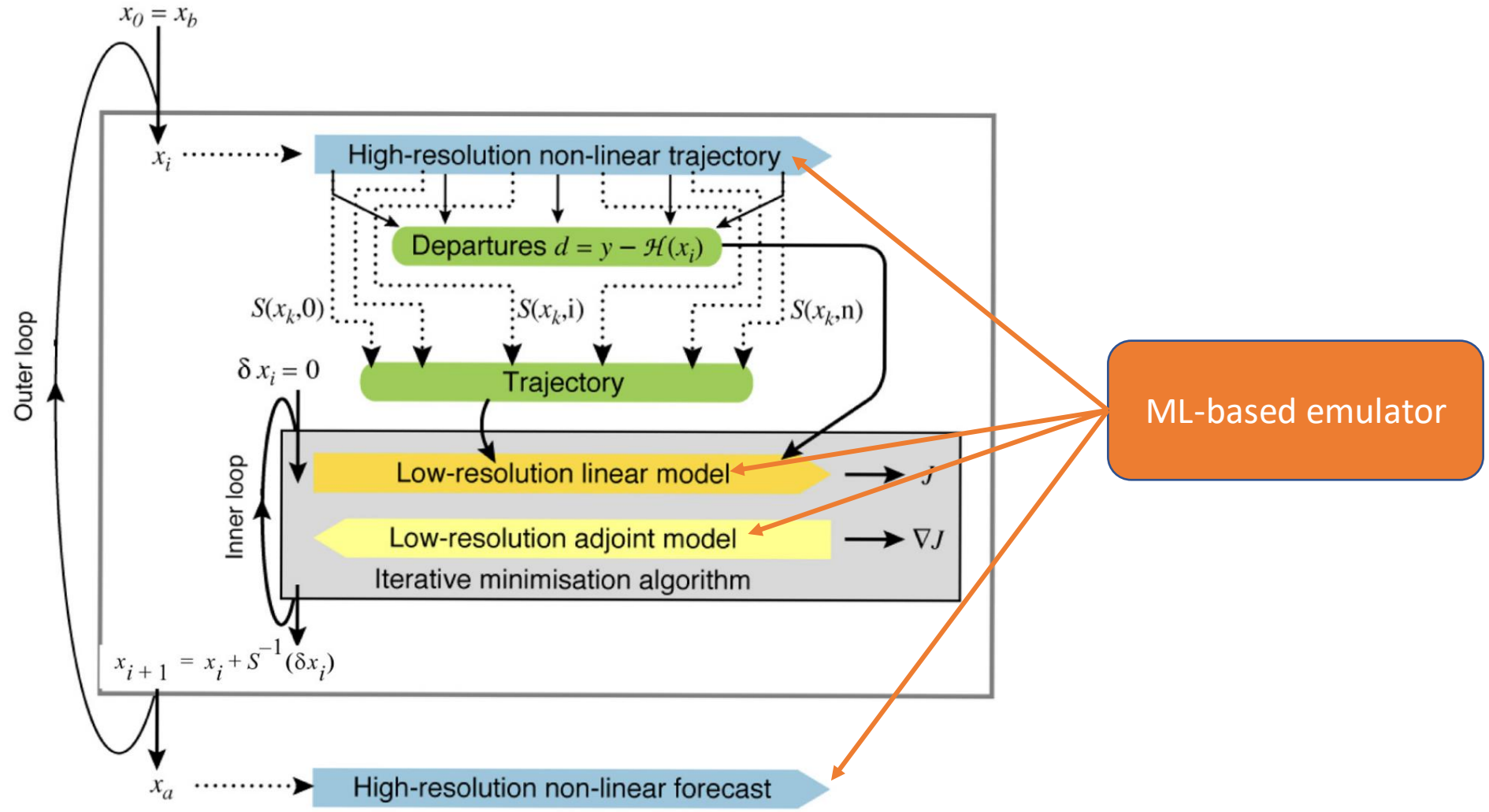


# Connexion with DA

- The training is done using reanalysis (obtained using both a physical based model and a data assimilation system)



# Consequence for the DA system



Schematic of ECMWF 4dVar



# Beneficial for DA?

## 4dVar

For each minimization iteration (outer loop), it is necessary to:

	Physical-based model	Neural network
Run the forward model	Costly	cheap
	Need developments	Need training
Run the adjoint model	Costly	cheap
	Many Development efforts	Comes for free

# Beneficial for DA?

## EnKF

For each assimilation cycle:

	Physical-based model	Neural network
Run an ensemble of forward models	Costly	cheap
	Need developments	Need training

# Preliminary conclusion

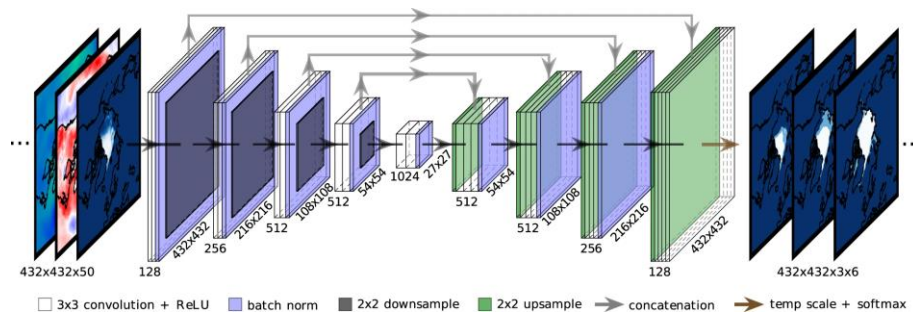
It is realistic to think that soon (in the next months?), the forecasting model in operational numerical weather forecast could be replaced by a neural network

But...

So far, physics based reanalysis (ERA5) are still instrumental to train the neural network

# Next steps?

- Ocean or sea ice emulators:  
Regional applications, or process based (e.g. surface mesoscale activity)

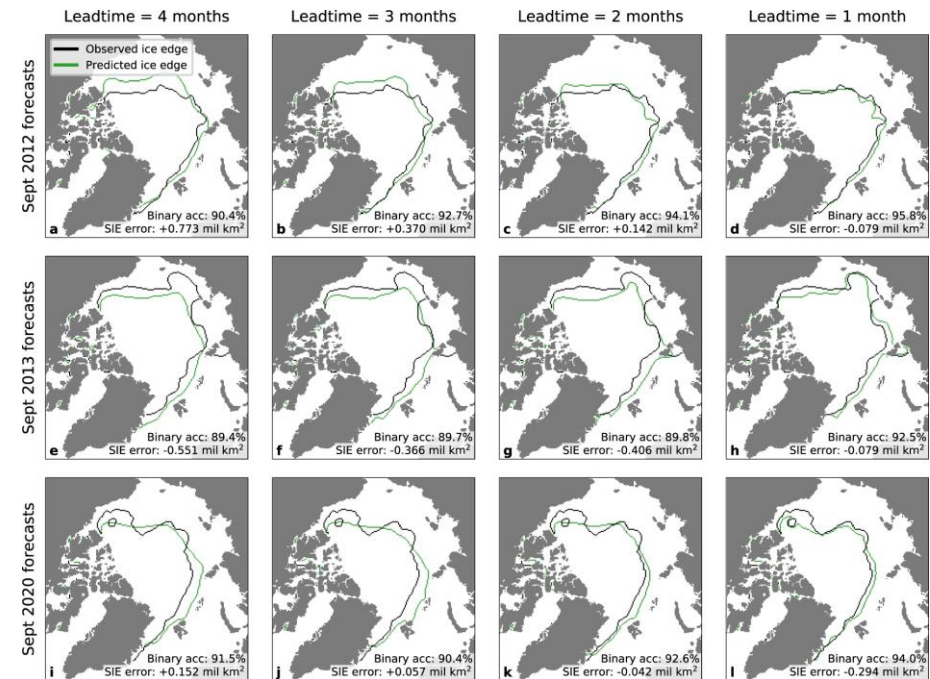


IceNet: Andersson et al. 2021

OceanNet (in preparation): Chattopadhyay et al.

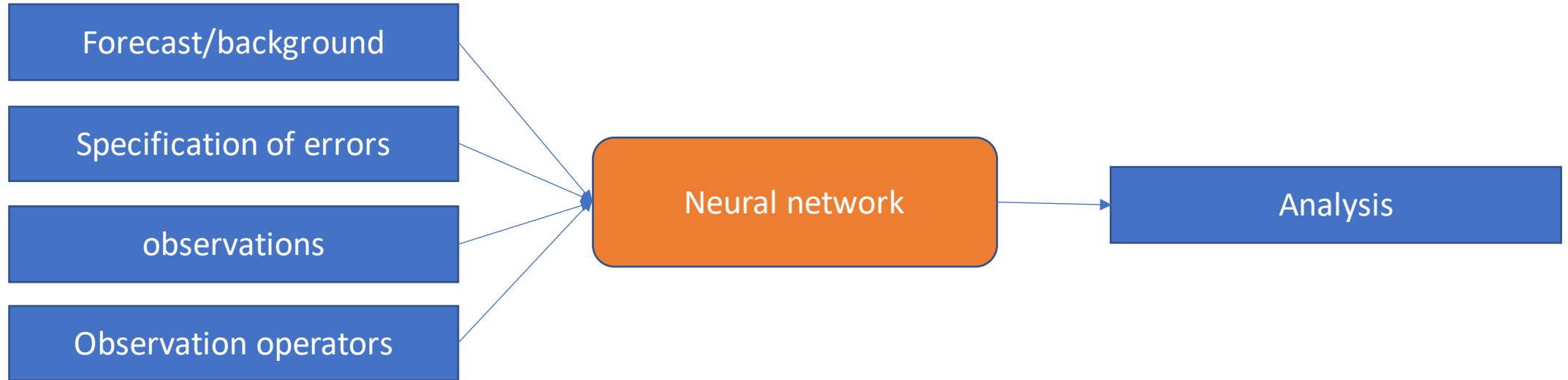
## Comments:

- Ocean is slower and less observed: current reanalysis could be insufficient for training
- Big tech companies (Nvidia, deepmind) are less interested in the ocean



# Next steps?

- Emulation of DA?



Demonstrated on sub-system (regional, only ocean surface) or specific types of observations (interpolation of gridded satellite data)

Sauzède, 2015, Abardenel 2018, Beauchamps et al. 2023

## Comments:

- Could relax strong hypothesis of DA (gaussianity, linearity)
- How would it be trained?