Large-scale Deep Learning for Weather and Climate Prediction

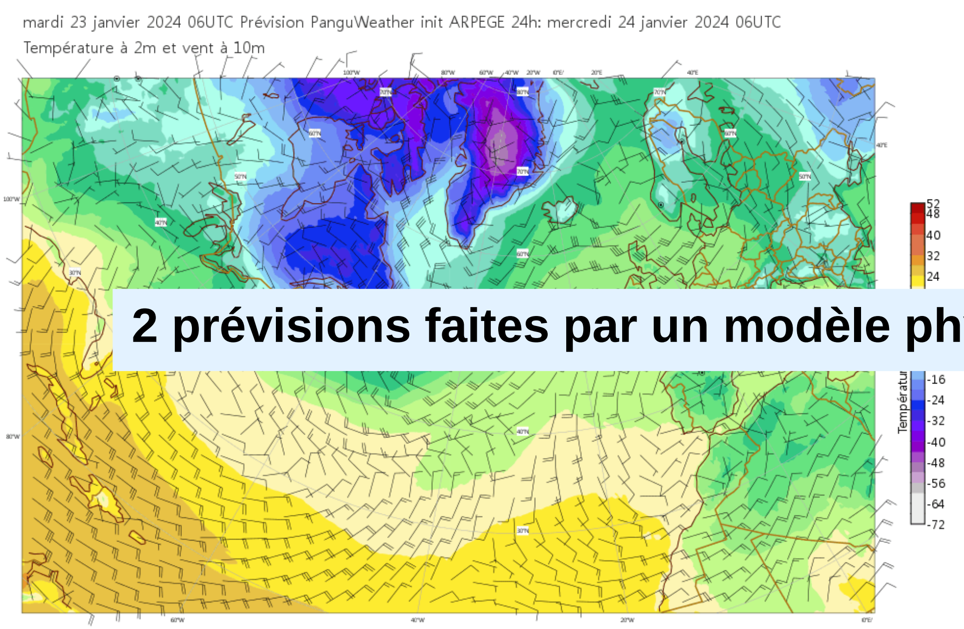
Workshop on Heterogeneous Data and Large Representation Models in Science, Toulouse, 2 October 2024

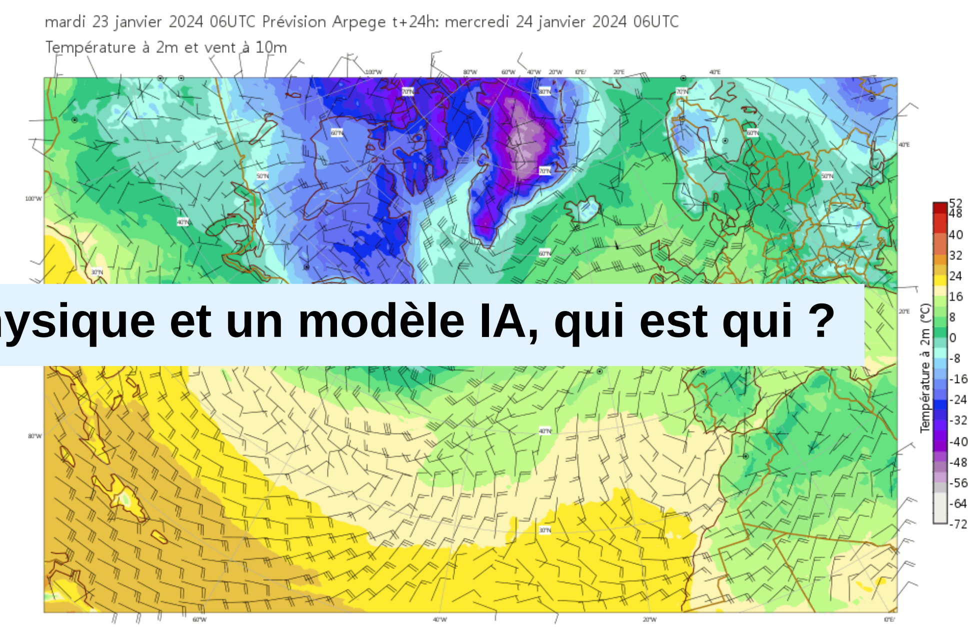
Laure Raynaud

Introduction

- Both weather & climate simulations are produced by numerical models, that integrate **physical knowledge** about the atmosphere (and possibly about the other earth system components)
- ML has recently changed the landscape, opening new perspectives to **enhance the computation, accuracy and processing** of weather & climate simulations

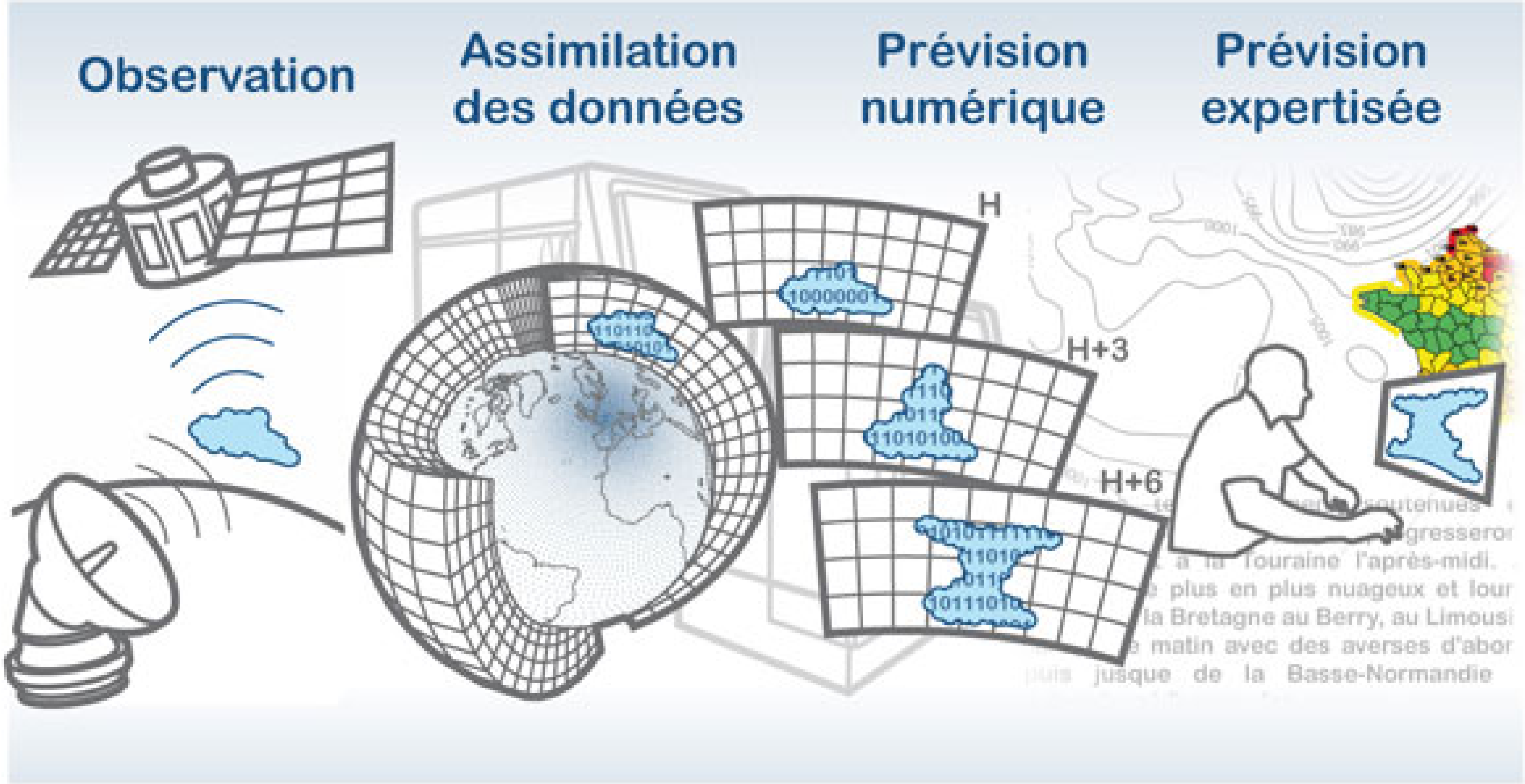
- The most disruptive application is a **partial to complete replacement of physics-based models by data-driven models** → **focus of this presentation**





2 prévisions faites par un modèle physique et un modèle IA, qui est qui ?

The weather prediction processing chain



Acquisition Contrôles Pré-traitements	**Modélisation** Physique de l'atmosphère Méthodes numériques	**Exploitation** Expertise humaine et automatisée

AI can be integrated in the different stages, in particular in the modeling part

Weather prediction : a complex and high-dimensional problem

- State-of-the-art atmospheric models are resolved on spatial grid with resolution ~ 10km at the global scale and ~ 1km at the regional scale : they simulate a wide range of scales, from large-scale flow to very localized phenomena (thunderstorms, turbulence), from minutes to several days ahead.

- They also integrate a large range of heterogeneous observations data for the computation of their initial conditions (through data assimilation techniques)

$$x^{t+1} = \mathcal{M}(x^t) \; ; \; x^0 = \mathcal{F}(\tilde{x}^0, y)$$

$O(10^9)$ $O(10^7)$

- Dimensions keep increasing as model resolution and observations improve

Fundamental physical principles

$\mathcal{M}$

- **Conservation of mass**
- **Conservation of energy**
- **Conservation of momentum**
- **Consider budgets of these quantities for a control volume**

From physics-based models to data-driven models : a range of possible solutions

Initial conditions → **Physics model** → **Forecast**

Well established models, but performances limited by
- understanding of physics
- computing resources

Initial conditions → **Hybrid Physics/AI** → **Forecast**

The 'in-between', ML could help
- uncover physical relationships
- reduce computational burden

while preserving the physical consistency of classical modeling

Initial conditions → **AI model** → **Forecast**

A change of paradigm, with very rapid predictions, and challenges
- costly training
- reliance on availability of high quality datasets
- black box, physical consistency ?
- generalization to out of distrib ?

The rise of data-driven modeling : an unexpected rapid (r)evolution ?

- **A (simplified) timeline**

AI replacing our forecasting model ?
I don't understand what you mean !

2015

2018

Challenges and design choices for global weather and climate models based on machine learning

Peter D. Dueben and Peter Bauer

European Centre for Medium-range Weather Forecasts, Shinfield Rd, Reading, RG2 9AX, UK

2021

Can deep learning beat numerical weather prediction?

M. G. Schultz, C. Betancourt, B. Gong, F. Kleinert, M. Langguth, L. H. Leufen, A. Mozaffari and S. Stadtler

2022/23

L'IA de Google DeepMind vient d'humilier les météorologues humains

AIFS: a new ECMWF forecasting system

Everyone runs its own ML model ?

The rise of data-driven modeling : an unexpected rapid (r)evolution ?

- **A (simplified) timeline**

AI replacing our forecasting model ?
I don't understand what you mean !

Can deep learning beat numerical weather prediction?

M. G. Schultz, C. Betancourt, B. Gong, F. Kleinert, M. Langguth, L. H. Leufen, A. Mozaffari and S. Stadtler

GenCast: Diffusion-based ensemble forecasting for medium-range weather

Ilan Price*[,1], Alvaro Sanchez-Gonzalez*[,1], Ferran Alet[1], Timo Ewalds[1], Andrew El-Kadi[2], Jacklynn Stott[1], Shakir Mohamed[1], Peter Battaglia[1], Remi Lam[1] and Matthew Willson[1]
*Equal contributions, [1]Google DeepMind, [2]Imperial College, London

Google, Huawei, NVIDIA, Microsoft, ECMWF

2015 **2018** **2021**

GraphCast: Learning skillful medium-range global weather forecasting

Remi Lam*[,1], Alvaro Sanchez-Gonzalez*[,1], Matthew Willson*[,1], Peter Wirnsberger*[,1], Meire Fortunato*[,1], ...aton-Rosen[1], Weihua Hu[1], Alexander Merose[2], ...lynn Stott[1], Alexander Pritzel[1], Shakir Mohamed[1] and

Pangu-Weather: A 3D High-Resolution System for Fast and Accurate Global Weather Forecast

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian, *Fellow, IEEE*

Challenges and design choices for global weather and climate models based on machine learning

Peter D. Dueben and Peter Bauer
European Centre for Medium-range Weather Forecasts, Shinfield Rd, Reading, RG2 9AX, UK

The rise of data-driven modeling : an unexpected rapid (r)evolution ?

- **A (simplified) timeline**

AI replacing our forecasting model ?
I don't understand what you mean !

2015

2018

Challenges and design choices for global weather and climate models based on machine learning

Peter D. Dueben and Peter Bauer

European Centre for Medium-range Weather Forecasts, Shinfield Rd, Reading, RG2 9AX, UK

Can deep learning beat numerical weather prediction?

M. G. Schultz, C. Betancourt, B. Gong, F. Kleinert, M. Langguth, L. H. Leufen, A. Mozaffari and S. Stadtler

2021

2022/23

L'IA de Google DeepMind vient d'humilier les météorologues humains

AIFS: a new ECMWF forecasting system

Everyone has its own ML model ?

The rise of data-driven modeling : is it really a surprise ?

- Weather Centers have **very large archives** of observations (satellite, in-situ) and Numerical Weather and Climate Prediction models
- More and more datasets are **under an Open Access licence**
- The **prediction problem** is inherently a good deep learning challenge
- **All the ingredients are here** to make ML a powerful tool for weather & climate prediction (given we have enough GPU resources & ML expertise)

Problem overview

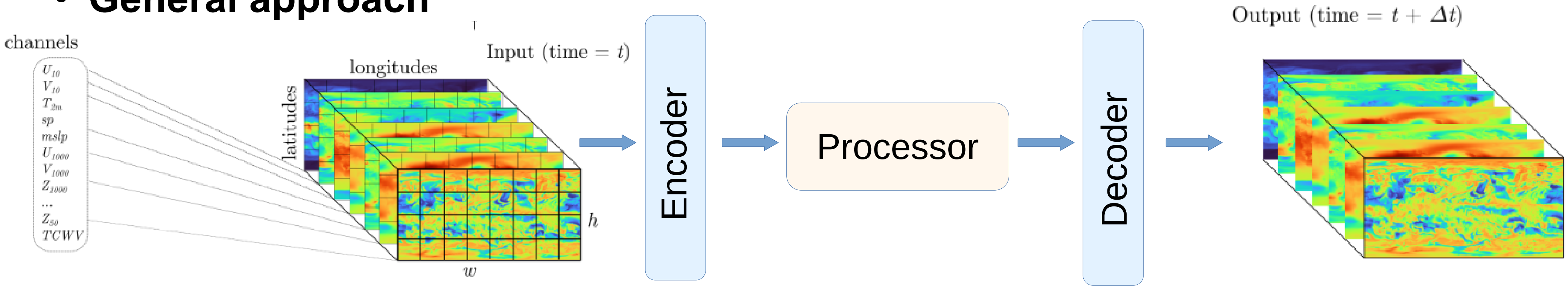
ML emulates the forecasting model $\mathcal{M}$

$$x^{t+dt} = \tilde{\mathcal{M}}(x^t)$$

- Prediction over several time steps obtained with an auto-regressive approach

$$x^t \rightarrow x^{t+dt} = \tilde{\mathcal{M}}(x^t) \rightarrow x^{t+2dt} = \tilde{\mathcal{M}}(x^{t+dt}) \ldots \rightarrow x^{t+ndt} = \tilde{\mathcal{M}}(x^{t+(n-1)dt})$$

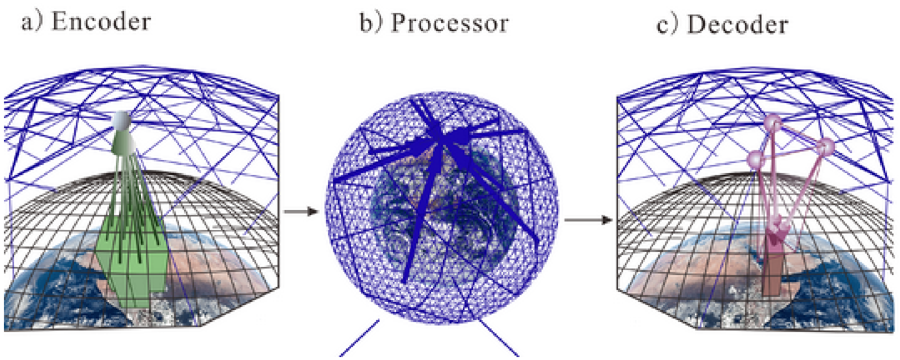
- **General approach**



- **Inputs** : generally gridded data of atmospheric variables at different altitudes

What's behind the most popular models ?

A common training datasets : ERA5 data, one of the most accurate reconstructions of the past weather on a ~30km global mesh, available from 1940s



A diversity of AI architectures

- CNN
- Vision Transformers
- Graph Neural Networks
- Neural operators

Performances close to those of physical models for the medium range, with some known weaknesses

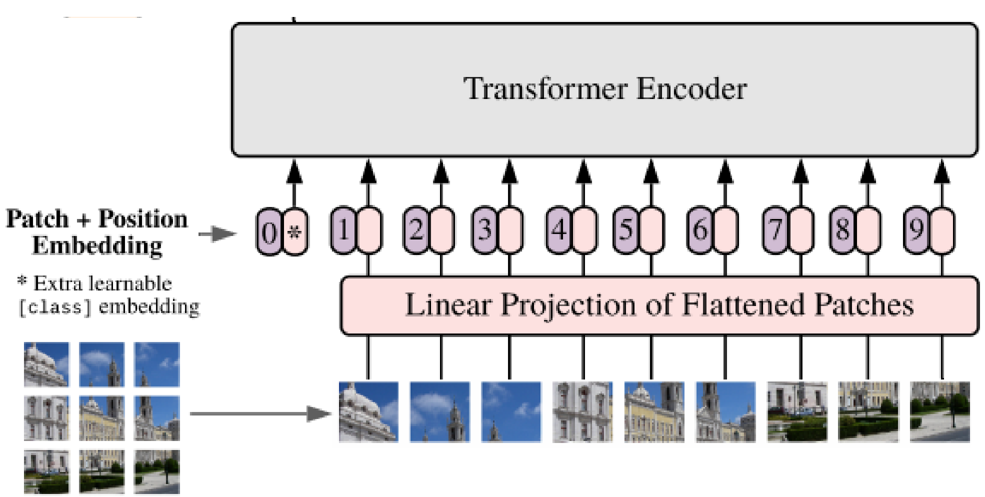
A very rapid inference time : a few sec to min (compared to ~ hour with physical models)

A black box

- Need for interpretability and explainability tools

The next challenges

- Emulators at very high resolution
- Learning from heterogeneous observations



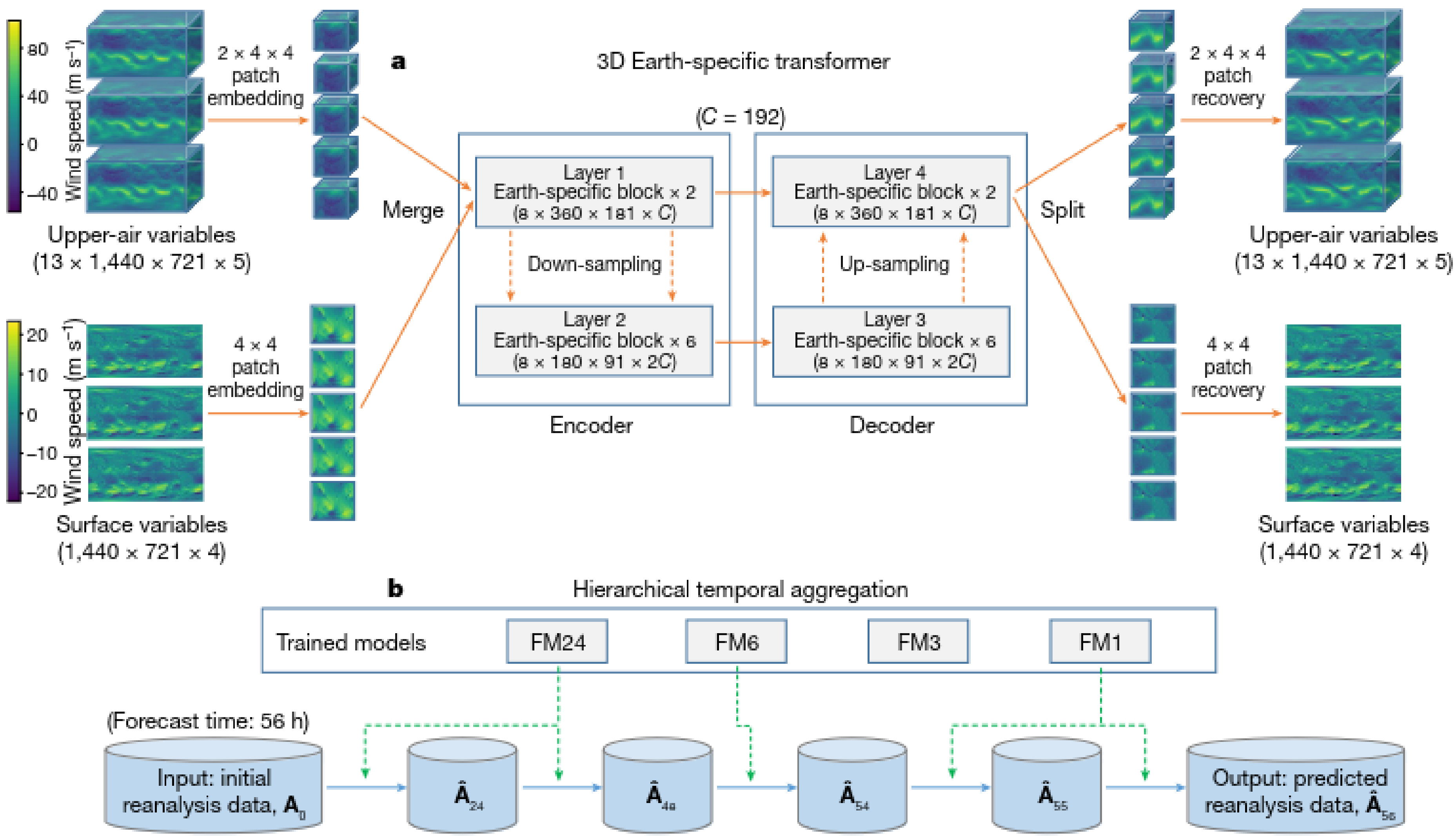
An (incomplete) overview of current models

Basic Model	Description	Method
MLP (Multi-Layer Perceptron)	early neural network suited for nonlinear challenges	Dueben et al. [62]
CNNs (Convolutional Neural Networks)	efficiently processing spatial data and extracting features	Scher et al. [61] Weyn et al. [64] Weyn et al. [78]
ResNet (Residual Network)	enabling deeper networks through efficient residual connections	Rasp et al. [79]
GNN (Graph Neural Network)	capturing spatial and temporal dynamics critical in fluid dynamics	Keisler et al. [80] GraphCast [65]
Transformer	processes input data through self-attention and feedforward layers	FourCastNet [73] FengWu [74] Pangu [24] ClimaX [81] FengWu-GHR [77] Fuxi [75]
EPD (Encode–Process–Decode)	a differentiable model with deep learning	NeuralGCM [76]

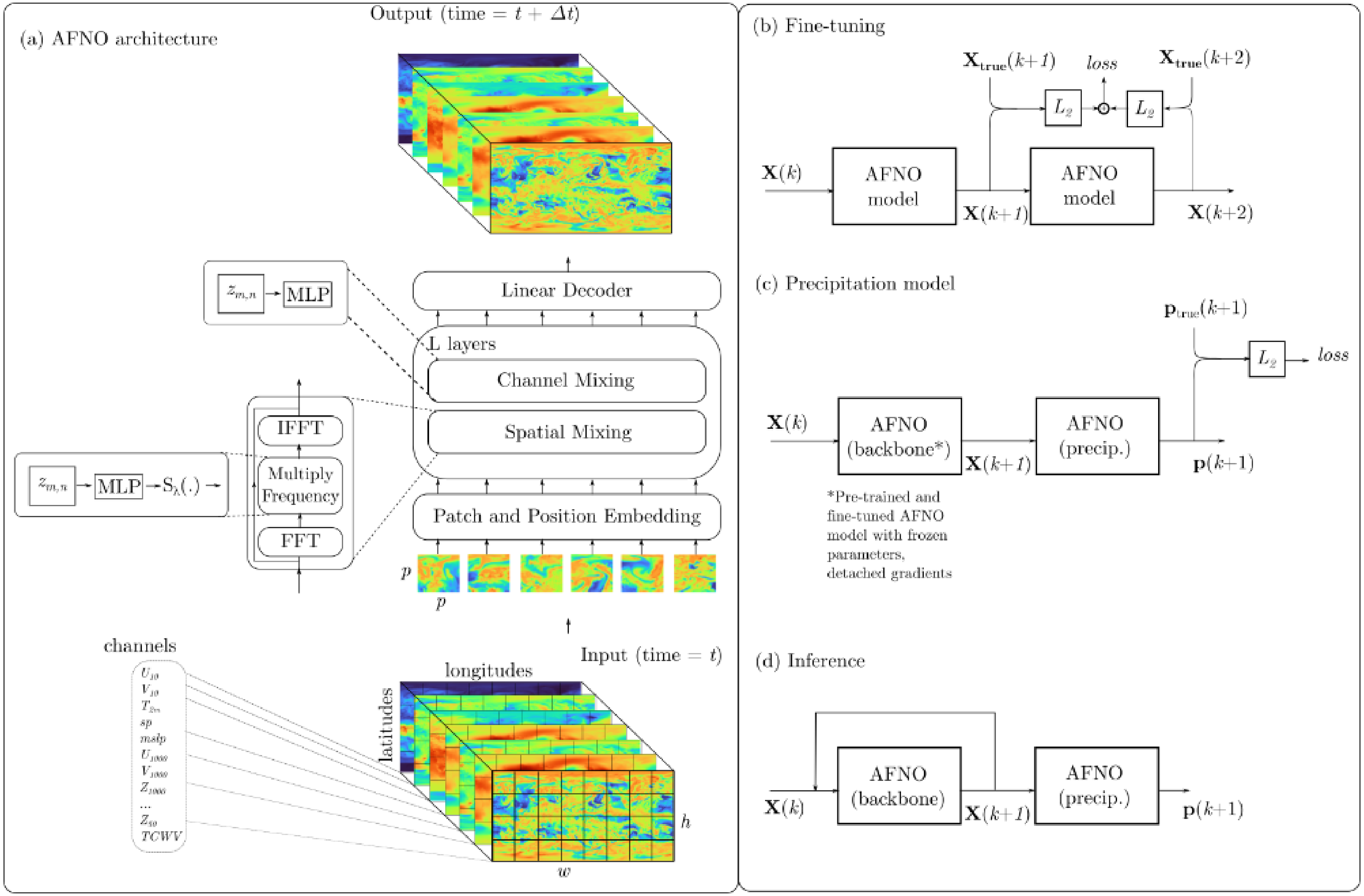
Method.	Hardware	Training Cost	Inference Cost
Weyn et al. [78]	1 NVIDIA V100 GPU	2–3 days	4-week forecast in less than 0.2 s on one GPU
Keisler et al. [80]	1 NVIDIA A100 GPU	5.5 days	5-day forecast takes about 0.8 s on one GPU
FourCastNet [73]	64 NVIDIA A100 GPU	16 h	1 week-long forecast in less than 2 s on one GPU
Pangu [24]	192 NVIDIA V100 GPU	16 days	5-day forecast takes 1.4 s on one GPU
FengWu [74]	32 NVIDIA A100 GPU	17 days	10-day forecast takes 0.6 s for on one GPU
GraphCast [65]	32 Cloud TPU v4	about 4 weeks	10-day forecast takes less than 1 min on one TPU
ACE [99]	4 NVIDIA A100 GPU	63 h	1 day simulate takes 1 s on one GPU
NeuralGCM [76]	16~256 Cloud TPU v4	1 day to 3 weeks	10-day forecast takes from 2.5 s to 119 s in different spatial resolutions on one TPU

From Wu and Xue, 2024 – See also https://github.com/jaychempan/Awesome-LWMs

Example : PanguWeather (Huawei)

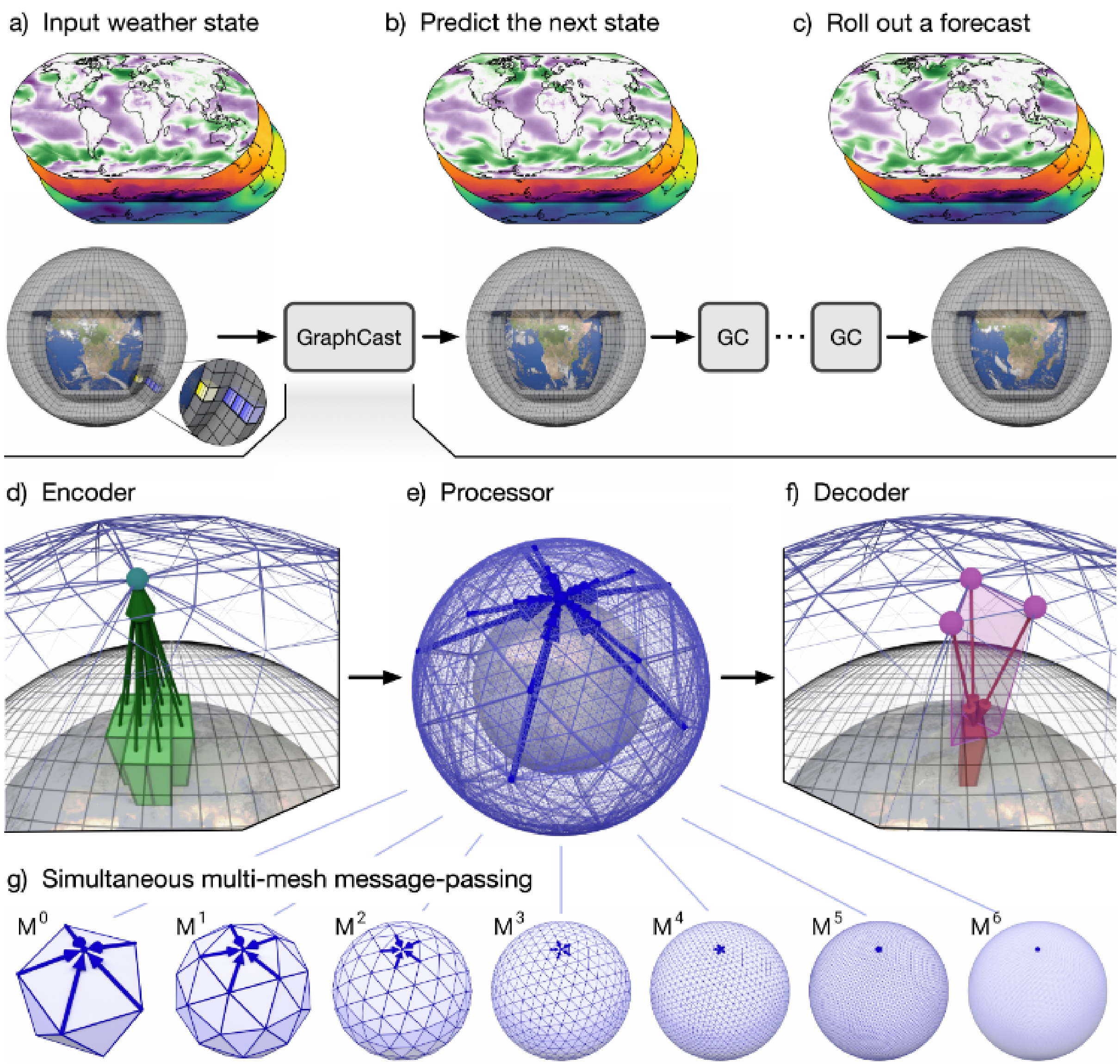


Example : FourCastNet (NVIDIA)



Combines the Fourier Neural Operator (FNO) learning approach with a ViT backbone.

Example : GraphCast (Google DeepMind)



See next presentation

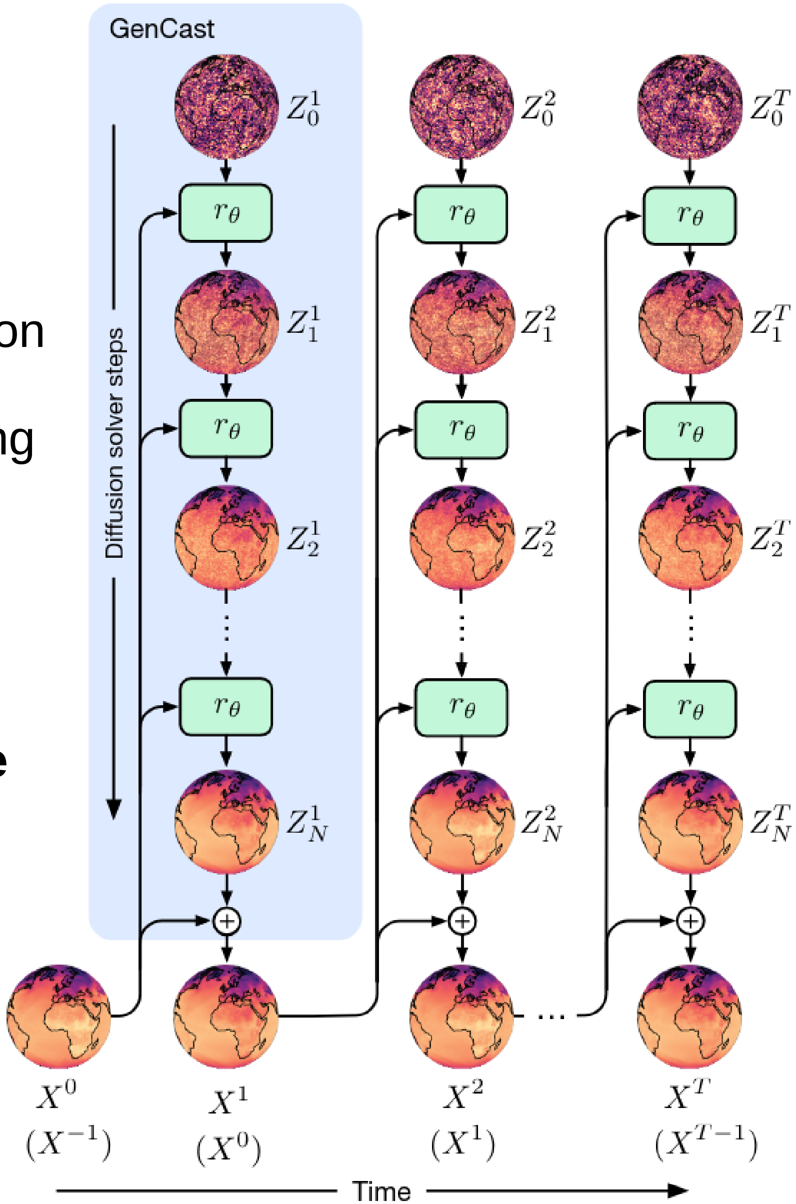
Conclusions

- The problem of data-driven atmospheric modeling is being adressed by both industrial and academics, with a **rapid acceleration since 2022**

- Most implementations focus on the forecast model emulation with an **encoder-processor-decoder architecture and Transformer backbone** (also GNN for flexibility of encoding + decoding). *New models showing up very quickly : only a selection has been presented here !*

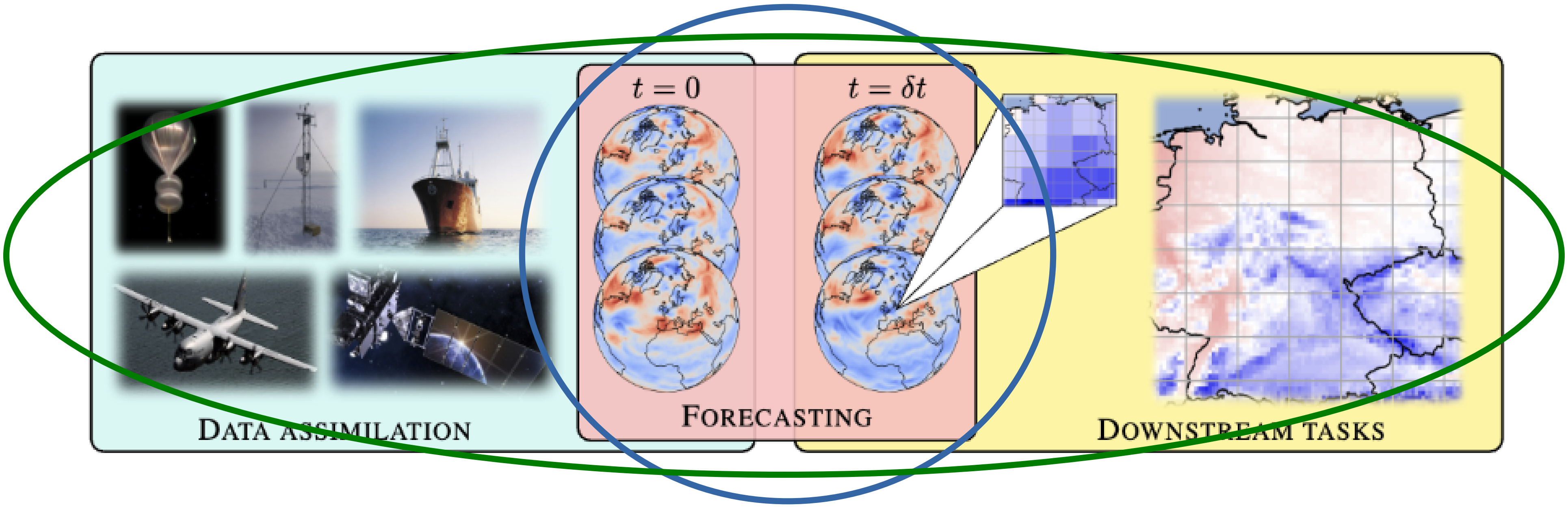
- Current focus
  - **High-resolution** regional models (~ km)
  - **Generative ML** such as diffusion models : they **increase the realism** ('sharpness') of predictions and allow for **uncertainty quantification,** a key issue in weather forecasting (see e.g. GenCast model from Google)

→ **Will be explored in the ANITI Chair EXPLEARTH**



What's next ?

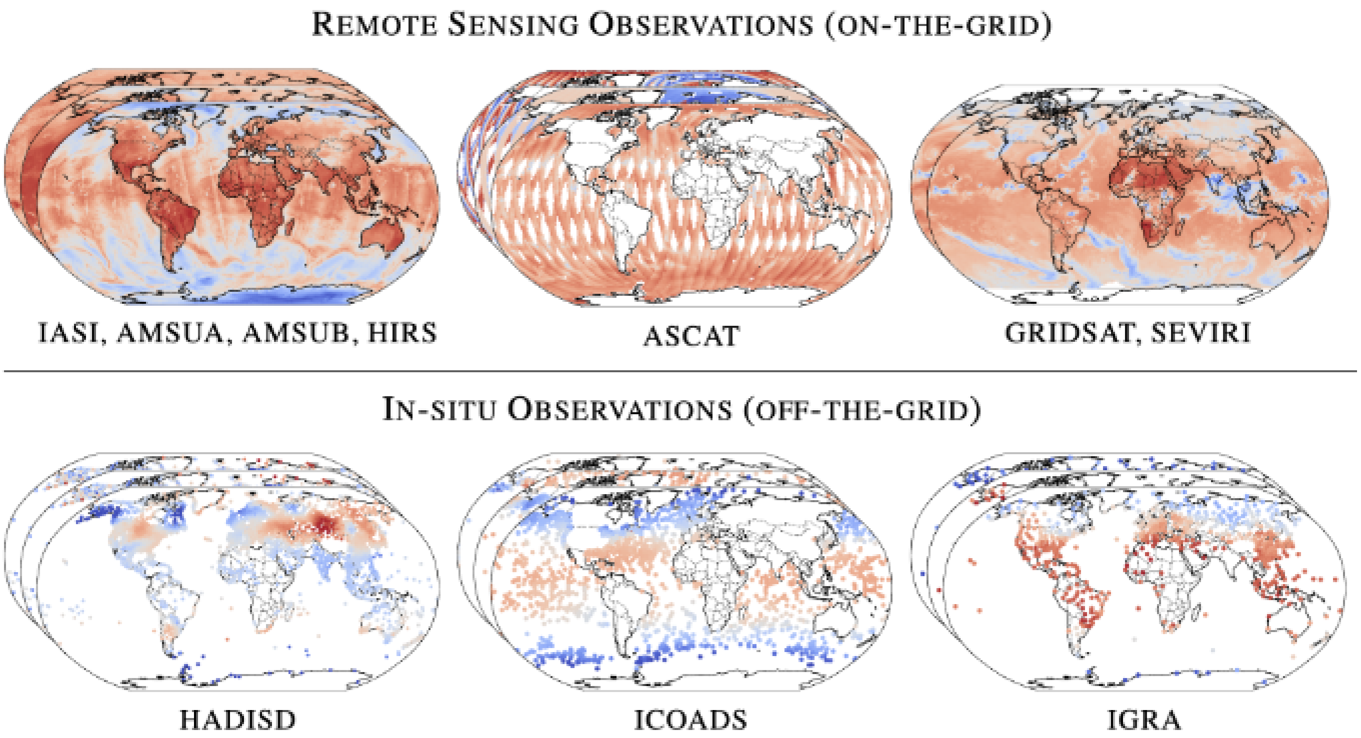
- In current approaches the initial state remains estimated by traditional approaches : **the next challenge is to design systems that directly learn from heterogeneous, sparse and non-static observations**, in order to emulate the entire pipeline.



Current ML models

Next generation ML models : an end-to-end approach

Learning from heterogeneous observations



- A large variety of instruments
- Different quantities mesured
- Spatial-temporal heterogeneity
- Sparse, missing and non-static observations
- A wide range of formats and data structures

Early end-to-end approaches

Assessing the Feasibility of an NWP Satellite Data Assimilation System Entirely Based on AI Techniques

Eric S. Maddy, Sid A. Boukabara, and Flavio Iturbide-Sanchez, *Senior Member, IEEE*

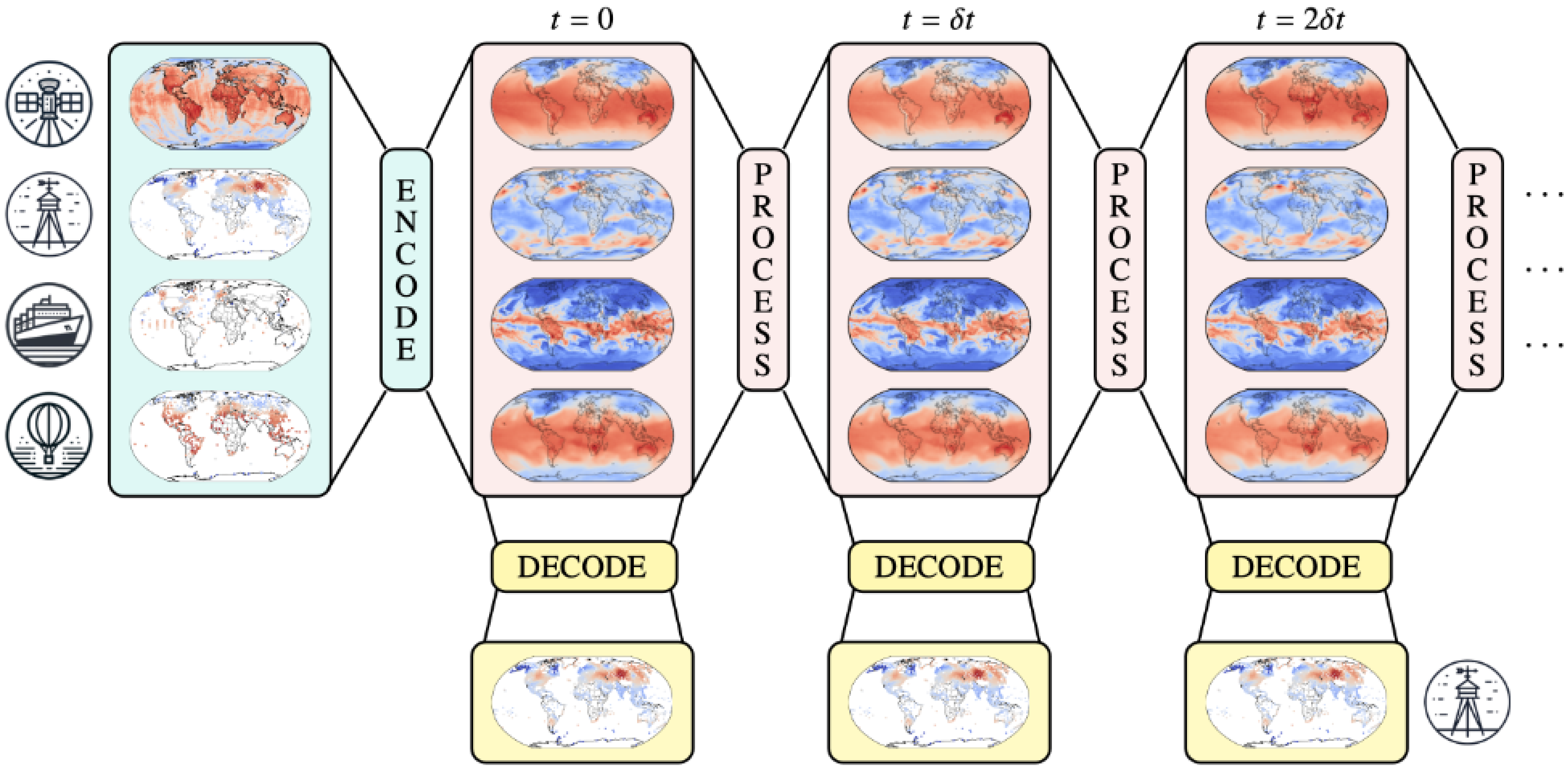
Deep Learning for Day Forecasts from Sparse Observations

DATA DRIVEN WEATHER FORECASTS TRAINED AND INITIALISED DIRECTLY FROM OBSERVATIONS

End-to-end data-driven weather prediction

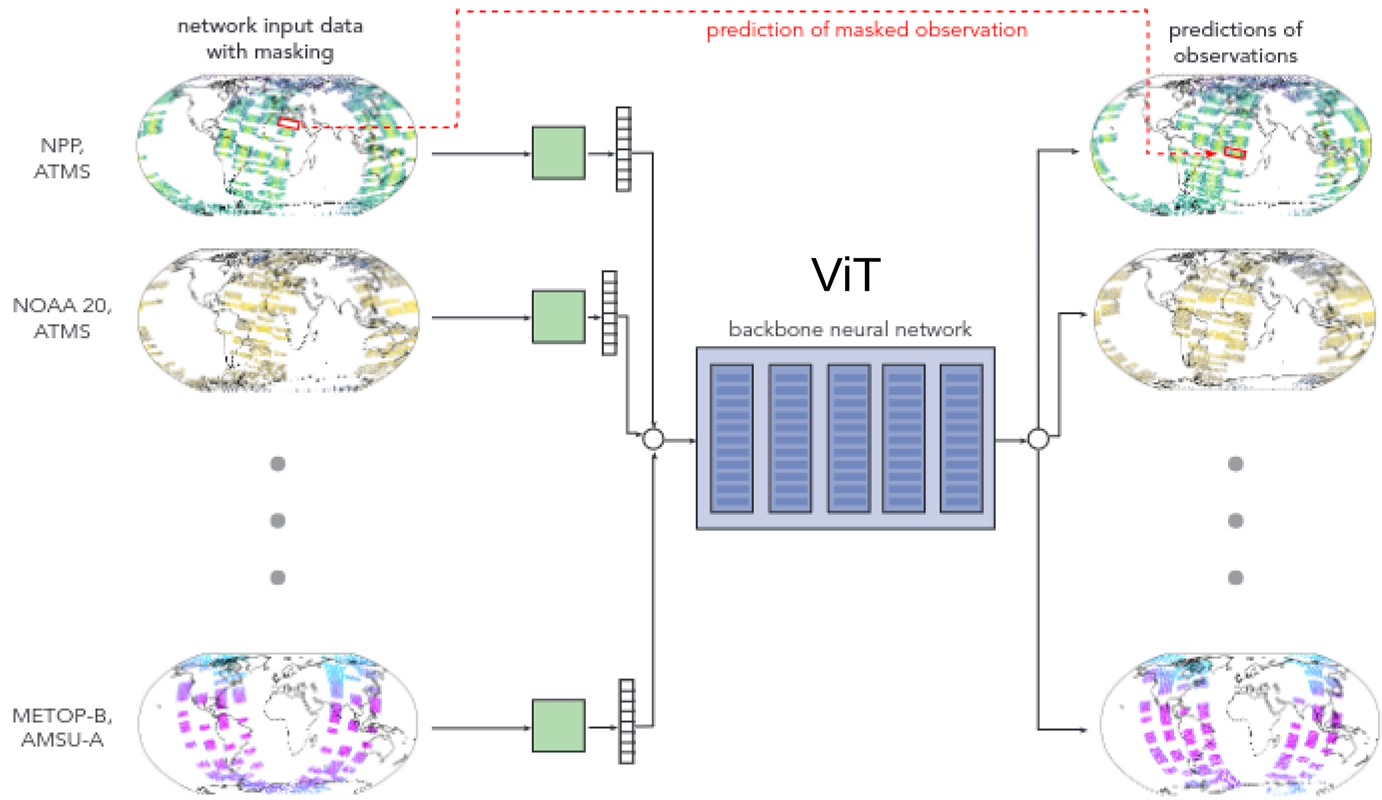
Anna Vaughan[*†1], Stratis Markou[*†2], Will Tebbutt[2], James Requeima[3], Wessel P. Bruinsma[4], Tom R. Andersson[‡9], Michael Herzog[6], Nicholas D. Lane[1], Matthew Chantry[8], J. Scott Hosking[5,7] and Richard E. Turner[*2,4]

Aardvark Weather (2024) : an end-to-end data-driven weather prediction system



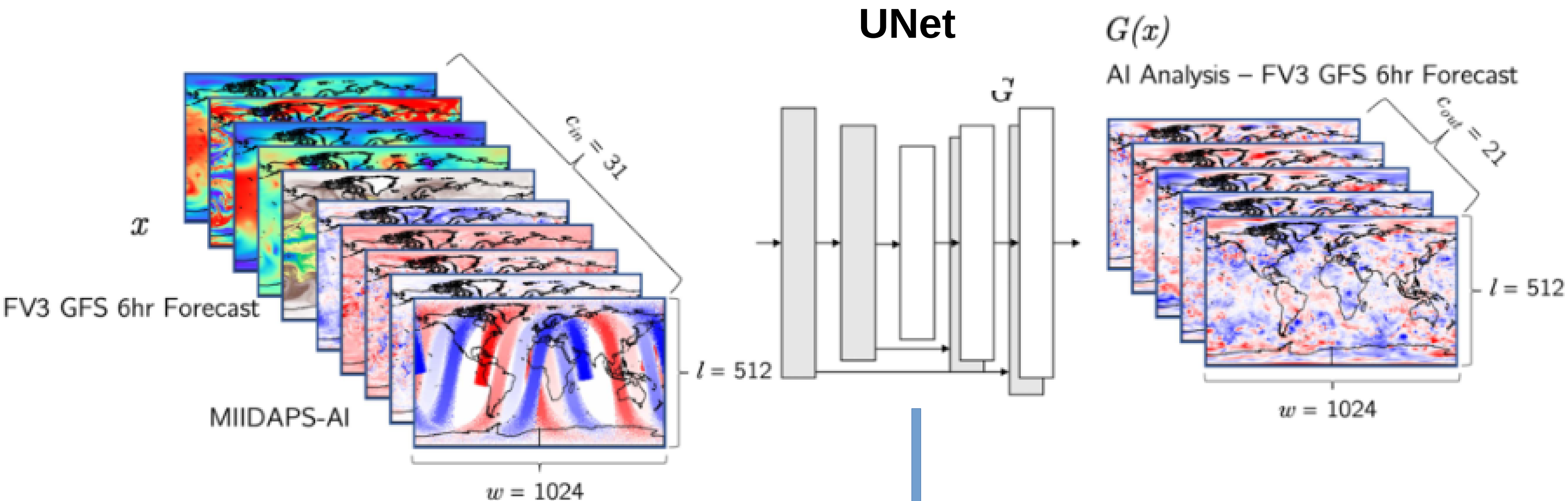
A ViT forms the backbone of the encoder and processor modules

ECMWF (McNally et al., 2024)



« A key challenge is to learn the spatial and temporal correlations that exist within a given observation type, but also between different measurement systems, and encapsulate these within the internal latent space of the machine learning model »

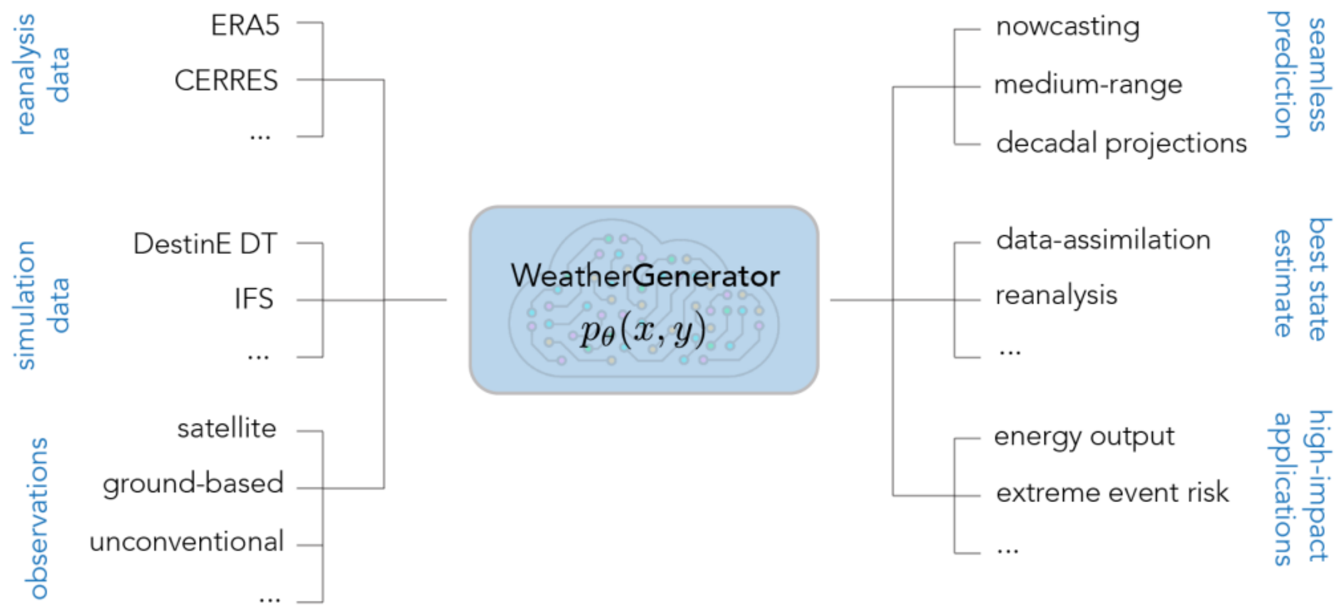
Emulation of data assimilation (Maddy et al., 2024)



$$x^0 = \mathcal{F}(\tilde{x}^0, y)$$

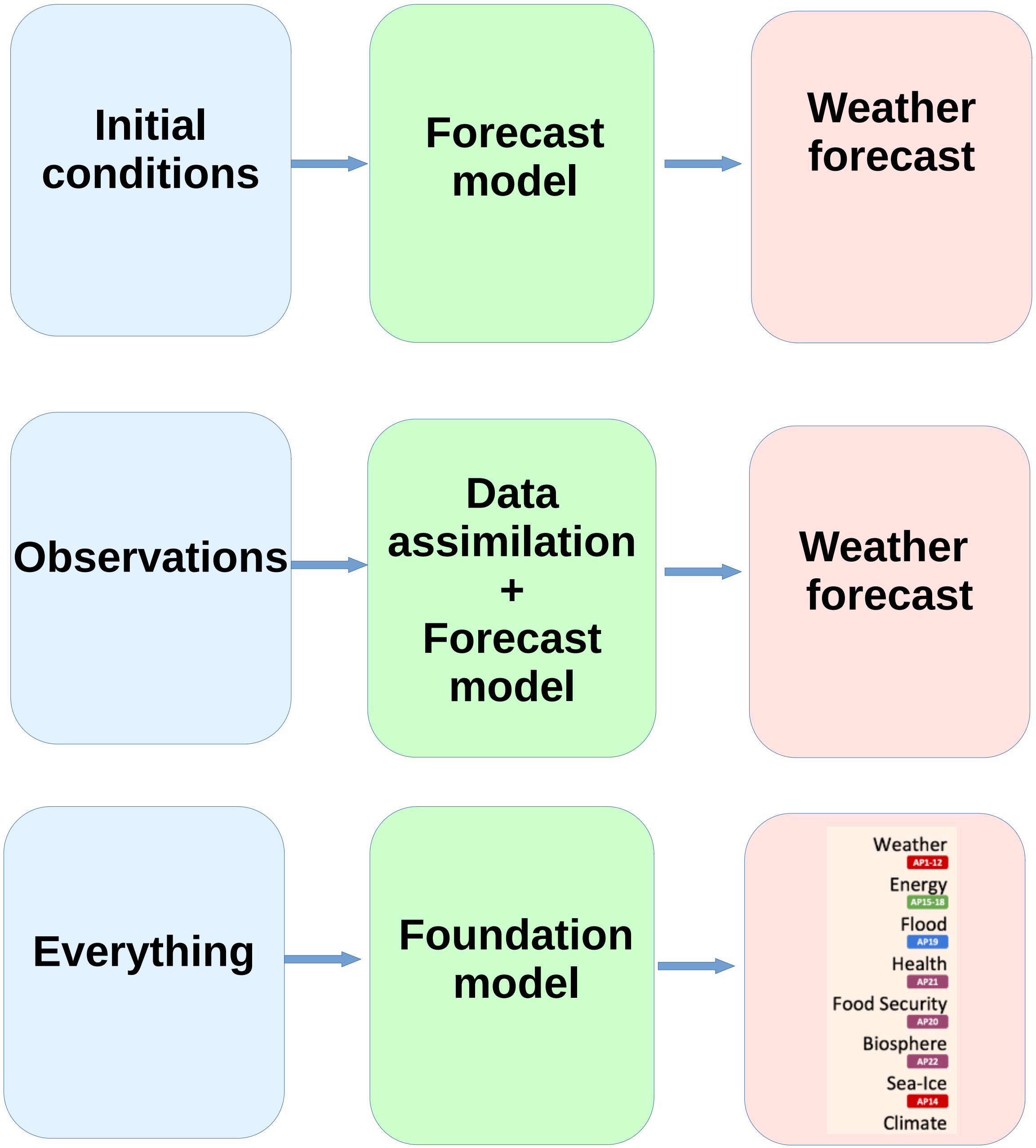
The Graal : a foundation model for the Earth System ?

- Coming soon : **Weather Generator** (Horizon Project 2025-2029)
- Building on early works by Lessig et al. : AtmoRep model
- From heterogeneous data to a wide range of applications …



- *Imagine if … there are off-the-shelf tools for a wide range of applications, including (1) data assimilation, (2) global and limited area ensemble predictions, (3) downscaling, (4) local vegetation, urban, flood, health, and energy models, (5) visualisation, (6) data compression and many more. (From P. Dueben)*

In a few years, a range of possibilities



2022 and +
→ Extend to very high resolution and probabilistic/ensemble prediction

2024 and +
→ Ongoing demonstration of feasibility and stability

2024 and +
→ Learn a rich representation of the Earth system, combined with task-specific decoders to provide a wide range of applications in Earth sciences, dream or reality ?

Cross-cutting challenges

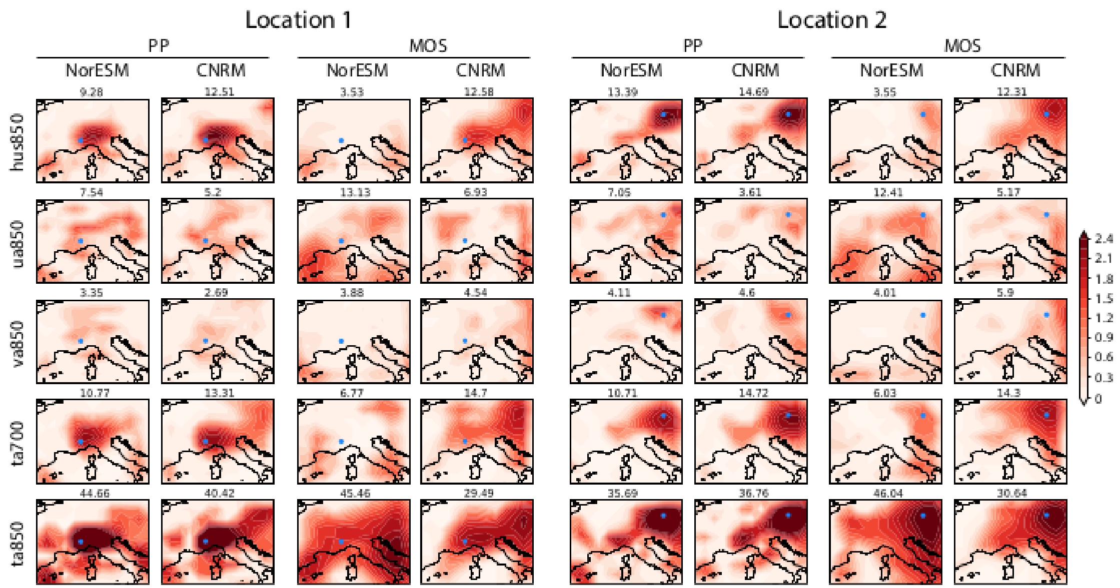
- **Physical consistency** (custom loss, architectures constraints, verification methods)
- **Generalization** on out-of-distribution samples (representation of extreme events)
- **Uncertainty quantification** : probabilistic deep learning approaches
- Gaining insight from **XAI**
  - The applicability of XAI approaches originally proposed for image classification (Grad-CAM, LIME, Shap, …) are now being tested on weather and climate tasks.
  - The sensitivity to the choice of XAI method is still an open question

First steps toward XAI

Finding the Right XAI Method—A Guide for the Evaluation and Ranking of Explainable AI Methods in Climate Science

Interpretable Machine Learning for Weather and Climate Prediction: A Survey

Transferability and explainability of deep learning emulators for regional climate model projections: Perspectives for future applications



Saliency maps (from Bano-Madina et al., 2024)

Will be explored in the ANITI Chair EXPLEARTH with IMT (L. Risser)

Concluding remarks

- **AI is here to stay** and is likely to disrupt the computation and exploitation of weather/climate predictions

- AI has become a new research topic at Météo-France : the 2024-2025 focus is to develop **our own data-driven model for km-scale forecast** (https://github.com/meteofrance/py4cast)

- AI emulators of weather models are likely to come into operations very soon, next challenge is **end-to-end systems** that exploit the large corpus of heterogeneous observations

- We need to **gain more insight into the black box**, integrate more physical constraints and further refine the evaluation framework

- Fully exploiting the potential of AI requires a **pluri-disciplinary approach** : different communities need to work together

"I think that you will all agree that we are living in most interesting times. I never remember myself a time in which our history was so full, in which day by day brought us new objects of interest, and, let me say also, new objects for anxiety."

British statesman Joseph Chamberlain, 1898 (taken from a Canadian colleague)