Le réseau de transport d'électricité

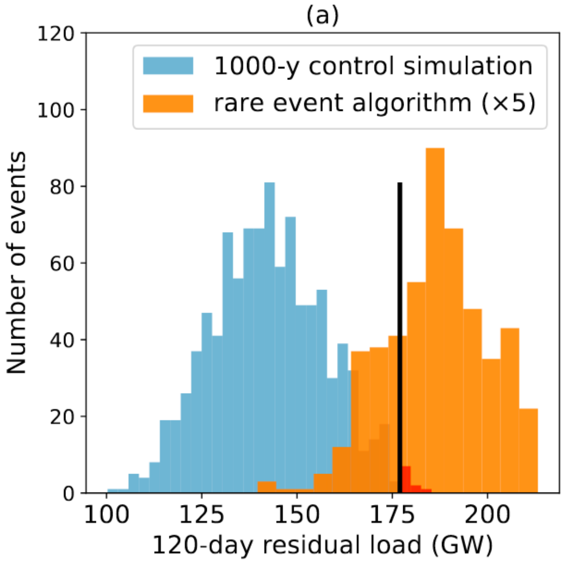
Leveraging Deep Learning and Rare Event Sampling to study climate extremes impacting the power system

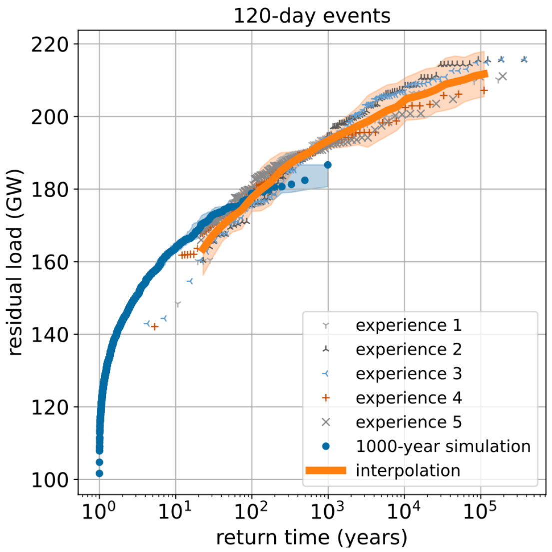
PhD Candidate: **Amaury Lancelin**
Supervison: **Freddy Bouchet (ENS), Laurent Dubus (RTE)**

Workshop "Energy, Mathematics, Theoretical challenges", October 2nd, 2024

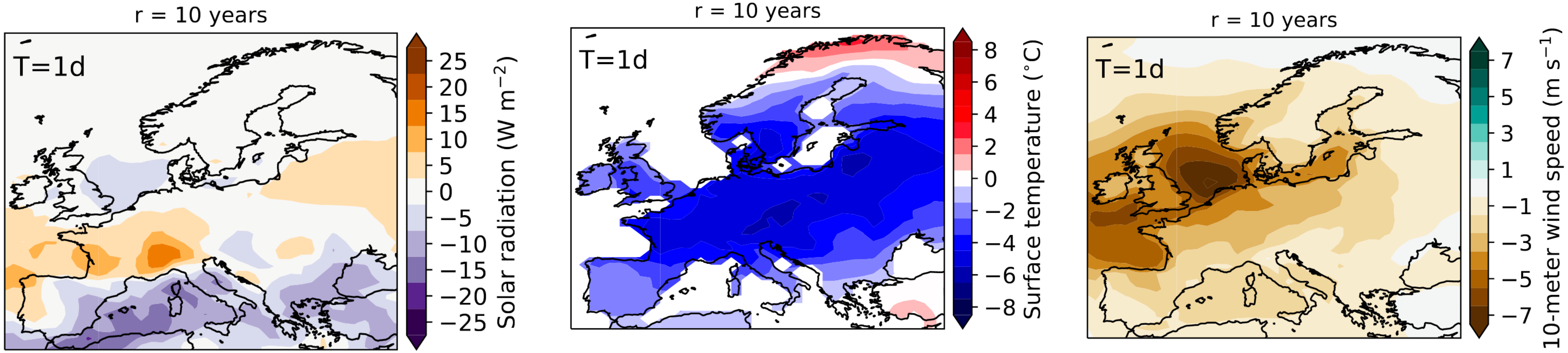
Application of RES for extreme climate events stressful for power systems – *PhD B. Cozian*

- Improved sampling of long-lasting rare events
- Allows to study the associated meteorological conditions that causes the events
- **Fail to sample shorter events**



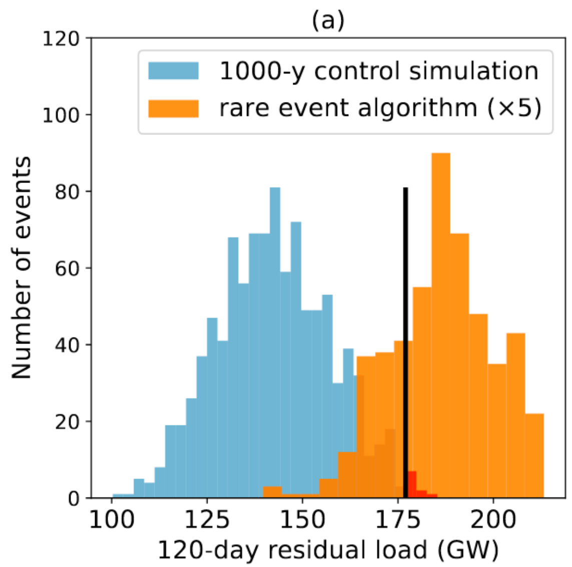


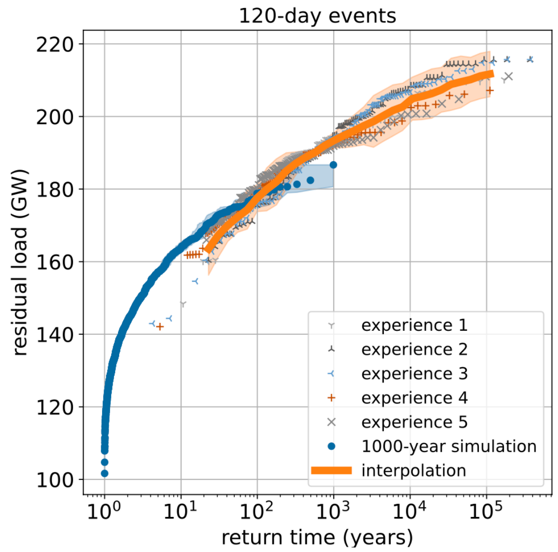
Composite maps of high residual load events



Application of RES for extreme climate events stressful for power systems – *PhD B. Cozian*

- Improved sampling of long-lasting rare events
- Allows to study the associated meteorological conditions that causes the events
- **Fail to sample shorter events**

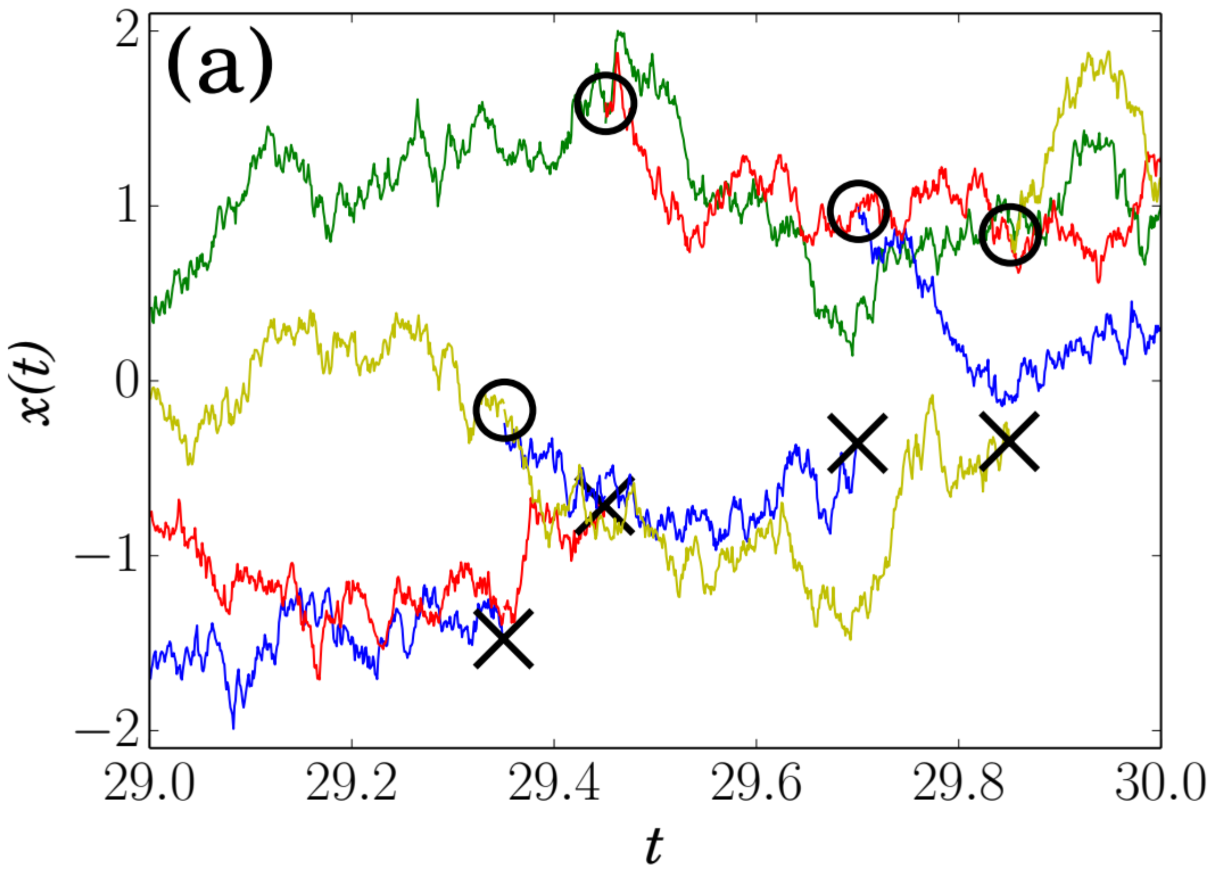




We need to develop a new methodology to sample shorter events (e.g. a few days or weeks) critical for power network stability

Genealogical particle analysis algorithm: selecting, killing and cloning trajectories - *(Del Moral, Garnier, 2006)*

We regularly stop simulations and resample to weight the ensemble toward the rare event of interest but track the unbiased probabilities.

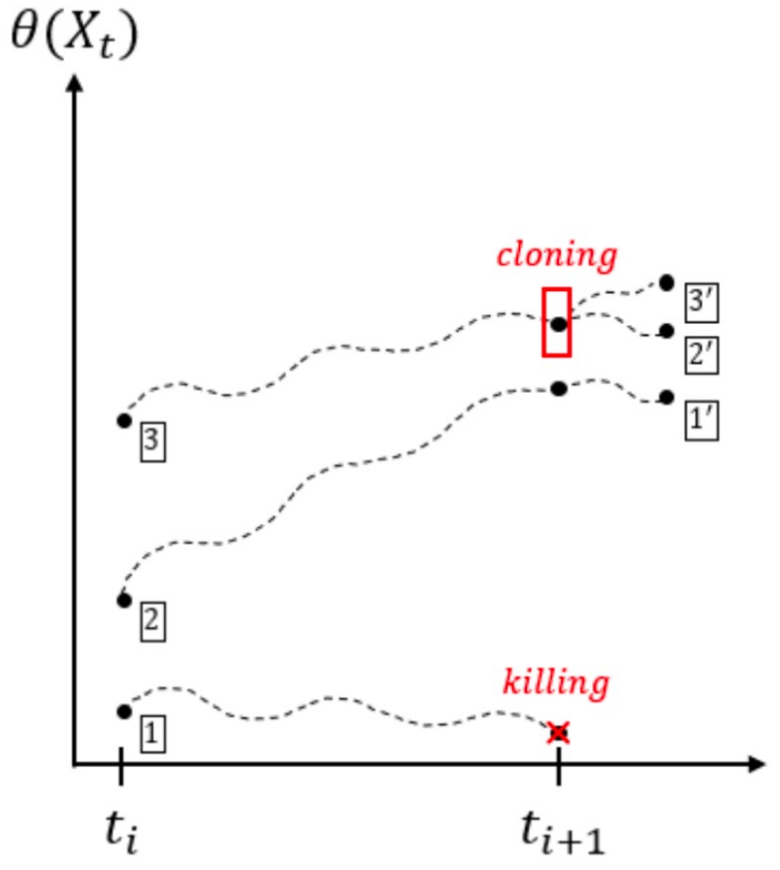


Sample paths of the Giardina-Kurchan algorithm from *(Bouchet, Jack, Lecomte, Nemoto, 2016)*

Weight trajectory *i* with:

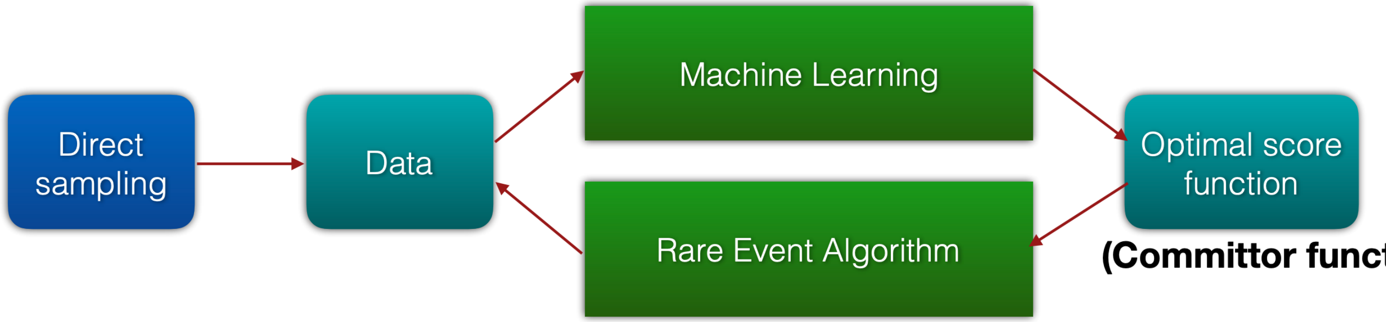
$$W_n^i = \exp\left(k \int_{t_{i-1}}^{t_i} O\left(X_n(t)\right) dt\right) / Z_i\,.$$

Idea for tackling shorter events: Coupling ML with Rare event sampling



From Noyelle, R., *PhD Thesis (2024)*

Coupling rare event algorithms with data based learning of committor functions

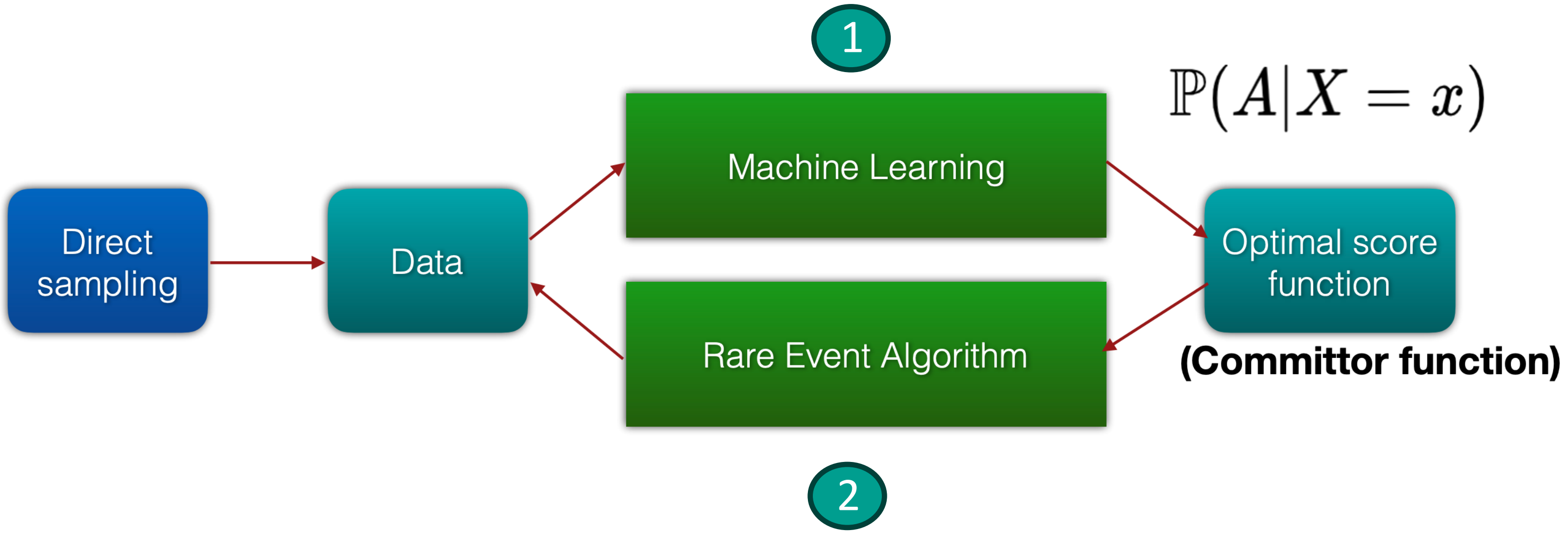


(Committor function)

$$\mathbb{P}(A|X = x)$$

- One example: Bouchet, Jack, Lecomte, Nemoto, PRE, 2016 (Committor function) (For X in dimension 1)

Idea for tackling shorter events: Coupling ML with Rare event sampling



Two options for 1 :

A. Direct prediction of the event of interest

Lovo & Lancelin et al. (2024)

B. Use an AI-emulator of Climate model

1

Direct prediction: Tackling the Accuracy-Interpretability trade-off

Lovo & Lancelin et al. (2024)

https://arxiv.org/abs/2410.00984

Tackling the Accuracy-Interpretability Trade-off in a Hierarchy of Machine Learning Models for the Prediction of Extreme Heatwaves

- Application: prediction of heatwaves over France

Lovo & Lancelin et al. (2024)

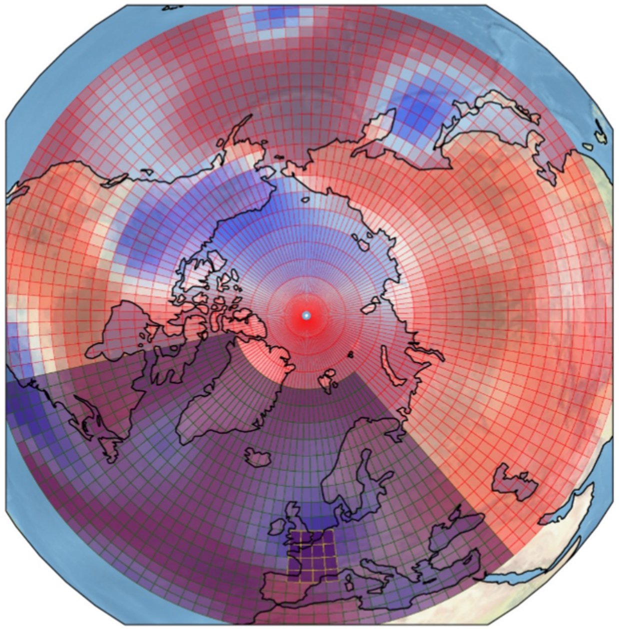
Target: Predict the distribution of heatwaves defined with

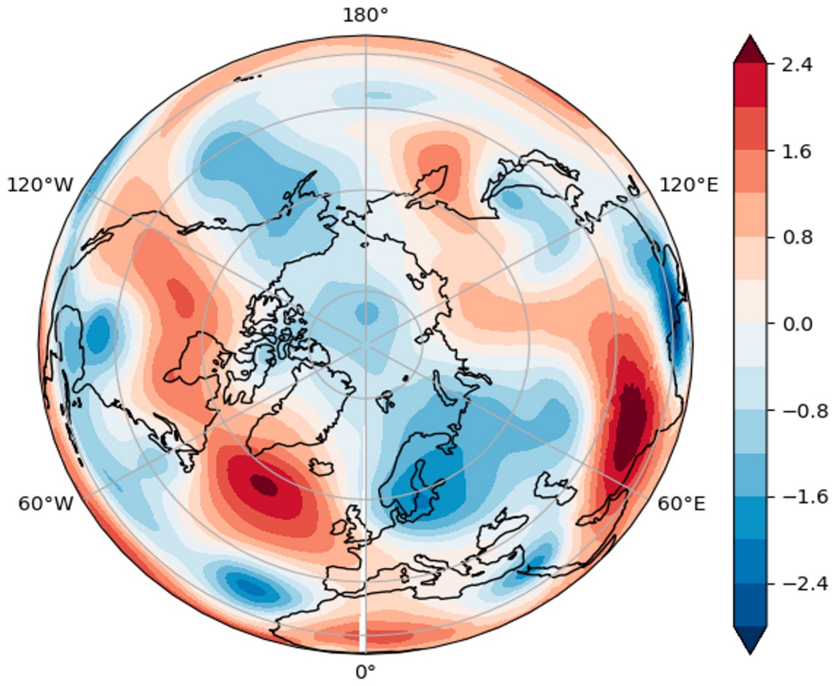
$$A(t) = \frac{1}{T}\int_{t}^{t+T} \frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} [T_{2\ m} - \mathbb{E}(T_{2\ m})](\vec{r}, t')d\vec{r}dt'$$

- T is the duration of the heatwave, we set $\mathrm{T} = 14$ days here.
- $\mathcal{D}$ is the fixed region of interest. Here, $\mathcal{D} =$ France.

Heatwave occurs if $A(t) > a_q$ with a_q an arbitrary high percentile.

We used simulated data from a state-of-the art Climate model





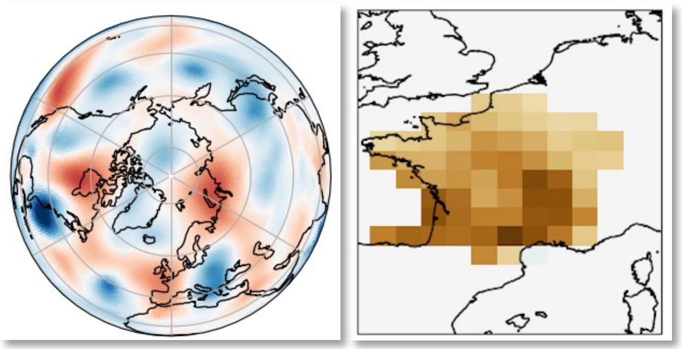
- 1000 years of simulated data of the CESM model at 1° resolution
- No atmosphere/ocean coupling
- Stationary climate of the years 2000

A probabilistic regression task from initial conditions

Predictors:

Geopotential Height at 500hPa

Soil Moisture over France



$\mathbf{X}(t)$ Snapshot of weather fields at time t

Target:

Predict $\mathbb{P}(\ A(t+\tau) \mid \mathbf{X}(\mathrm{t})\ )$ with a focus on extremes

- τ is the lead time at which we want to predict. $\tau = 0$,5,10,30, ... days.

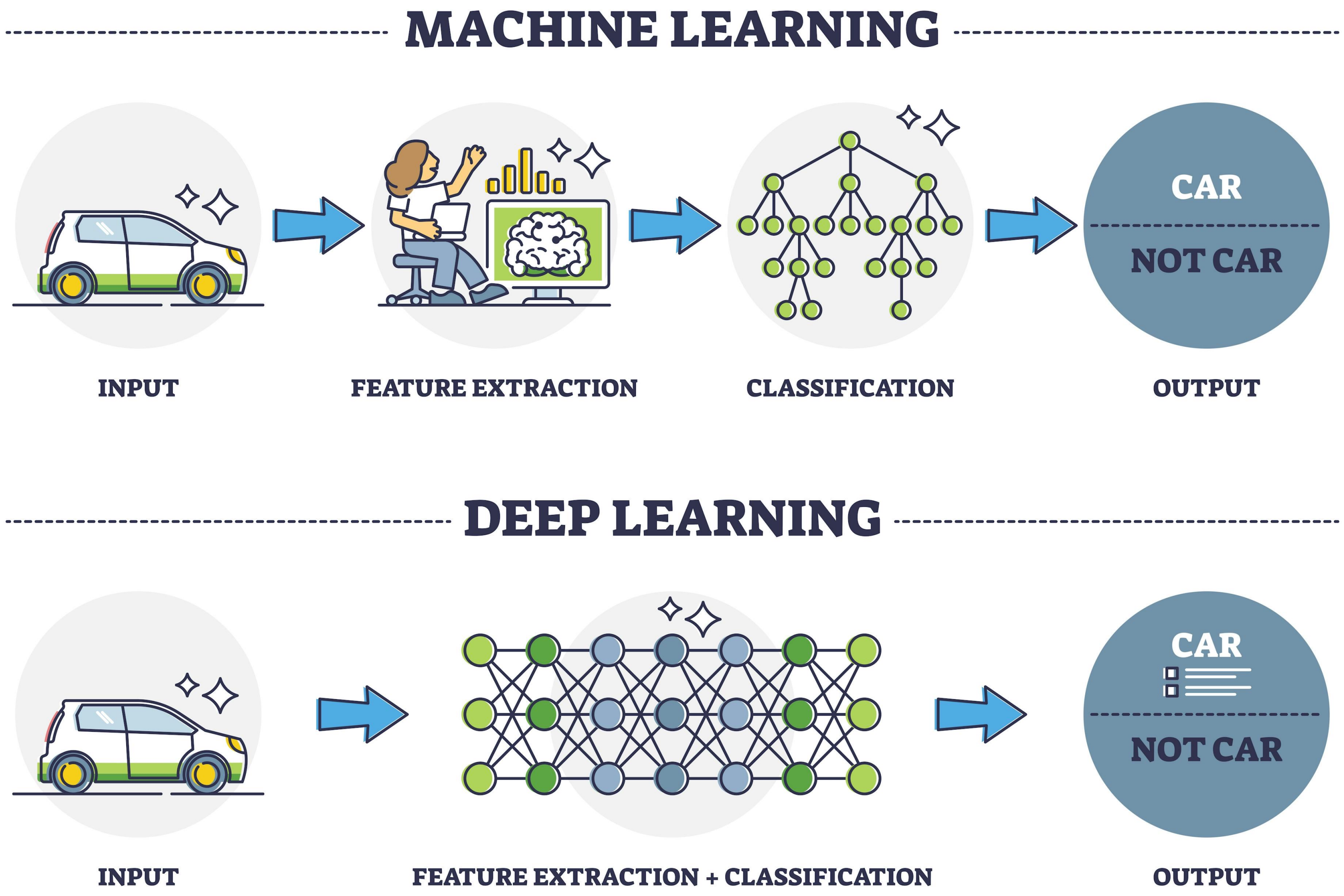
Parametric approximation: $\boxed{\mathbb{P}(A|X=x) \sim \mathcal{N}(\mu(x), \sigma(x))}$

Can we find an optimum between accuracy and interpretability?

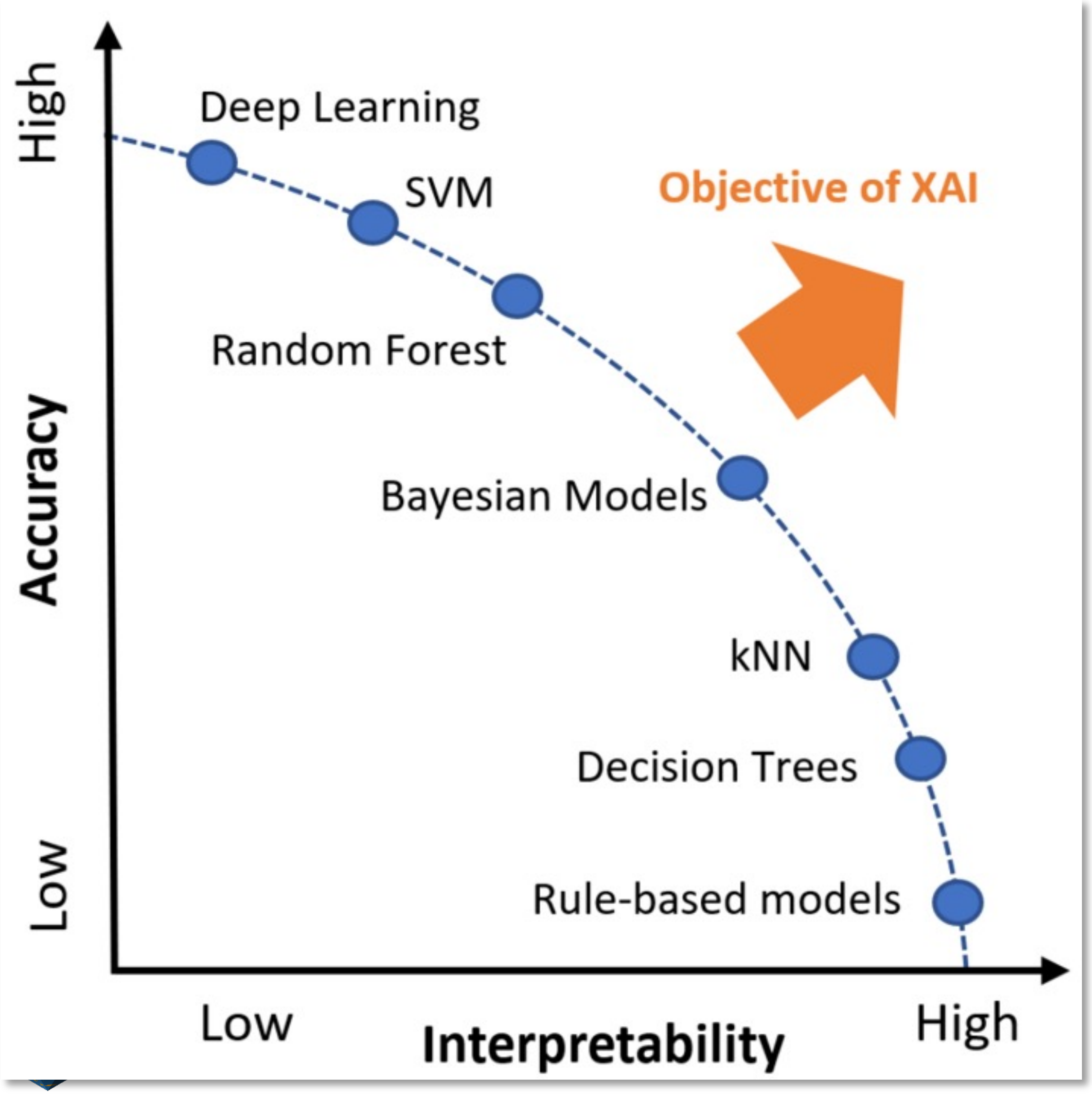
Method	$\hat{\mu}(X;\theta)$	$\hat{\sigma}(X;\theta)$	trainable parameters	non-trainable parameters	hyperparameters
GA	$M \cdot X$	σ	27 425	0	1
IINN	$g_\mu(M \cdot X)$	$s(g_\sigma(M \cdot X))$	55 058*	0	10
ScatNet	$\beta_\mu \cdot \phi(X)$	$s(\beta_\sigma \cdot \phi(X))$	19 930*	656 640*	5
CNN	$g_\mu(X)$	$s(g_\sigma(X))$	684 000*	0	10

To predict extreme heatwaves, we used a hierarchy of models from the simplest to the more complex ones.

An important distinction between "classic" ML and Deep Learning



Can we find a optimum between accuracy and interpretability?



- Specifically, predicting rare events with high societal impacts requires an understanding of how predictions are generated.
- Could also inform our physical understanding of phenomena

$$\mathbb{P}(A|X=x) \sim \mathcal{N}(\mu(x), \sigma(x))$$

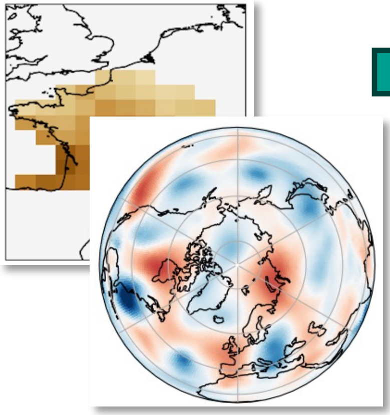
A simple model: The Gaussian Approximation

We assume $\boxed{X, A \sim \mathcal{N}(0, \Sigma)}$ Then, it comes $\boxed{A \mid X \sim \mathcal{N}\left(\mu_{A|X}, \Sigma_{A|X}\right)}$

where $\mu_{A|X} = \boxed{\Sigma_{AX}\Sigma_{XX}^{-1}} X$ and $\sigma_{A|X} = \Sigma_{AA} - \Sigma_{AX}\Sigma_{XX}^{-1}\Sigma_{XA}$ (does not depend on the current observation **X**)

Input

Snapshot of weather fields **X**

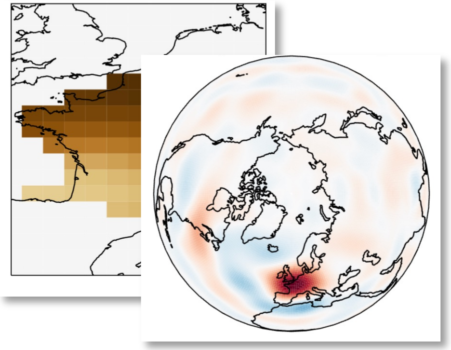


Prediction model: The Gaussian Approximation

Dot product

$*$

Optimal projection patterns



Output

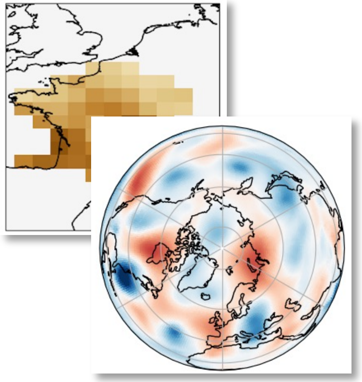
$\hat{\mu}(X)$ and constant $\hat{\sigma}$

(Mascolo & Lovo et al., 2024)

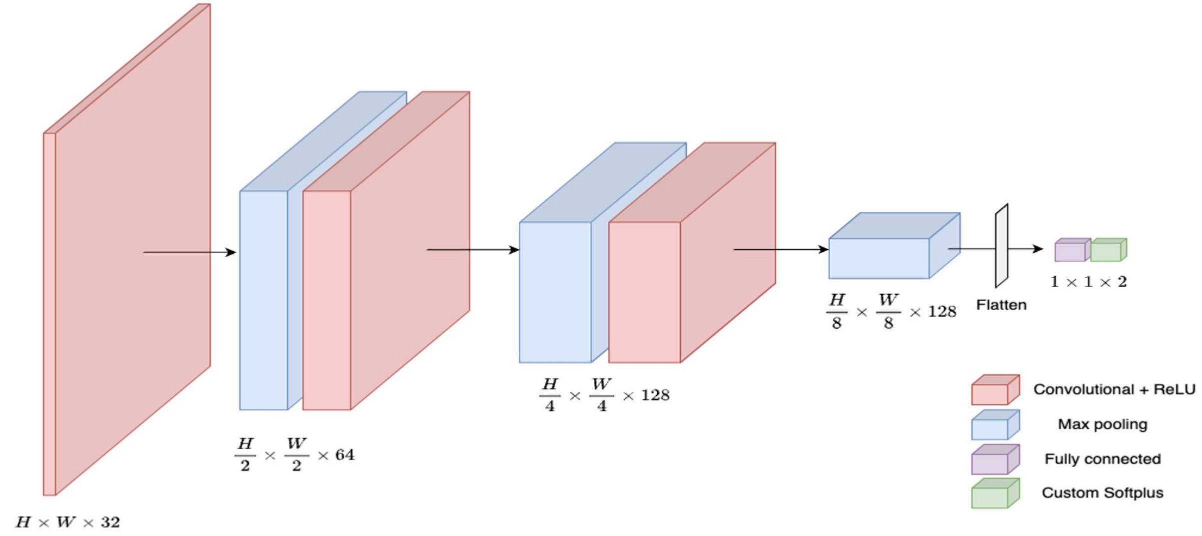
A more complex model: Convolutional Neural Networks

Input

Snapshot of weather fields **X**



Predictive model: CNN



Output

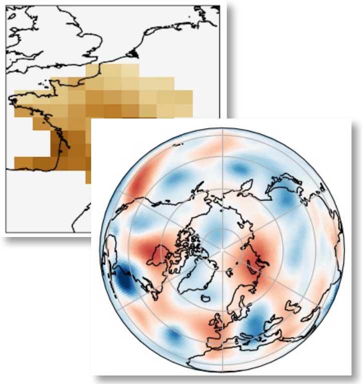
$\hat{\mu}(X), \hat{\sigma}(X)$

- Pros: Particularly performant on these kind of tasks (with enough training data)
- Cons: A lot of parameters to learn & behave as a **black-box**

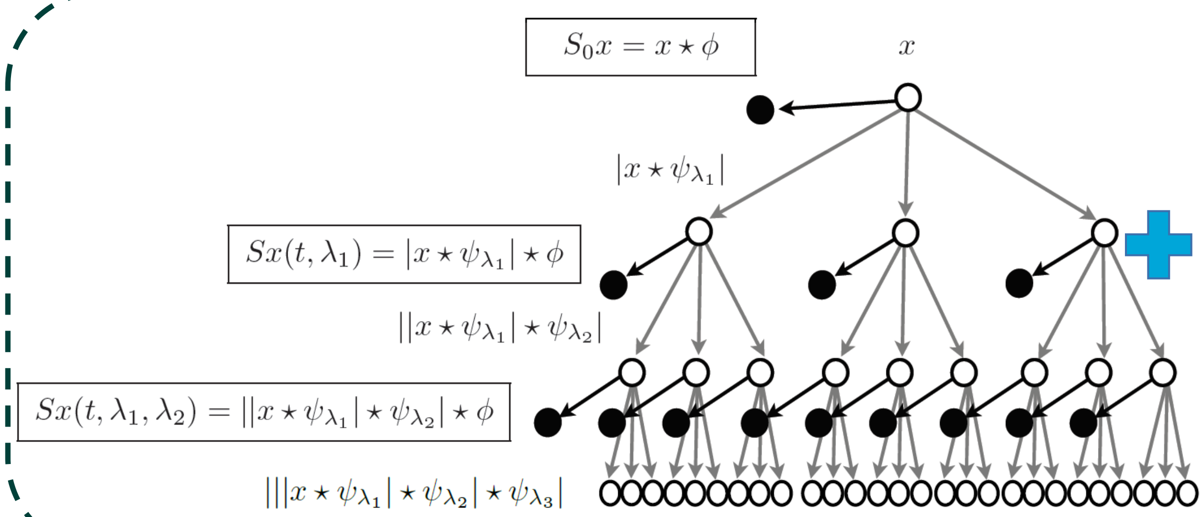
An intermediate model with a convolutional architecture: Scattering Networks

Input

Snapshot of weather fields **X**



Predictive model: Scattering Network



A **linear projection**

(Bruna & Mallat, 2012)

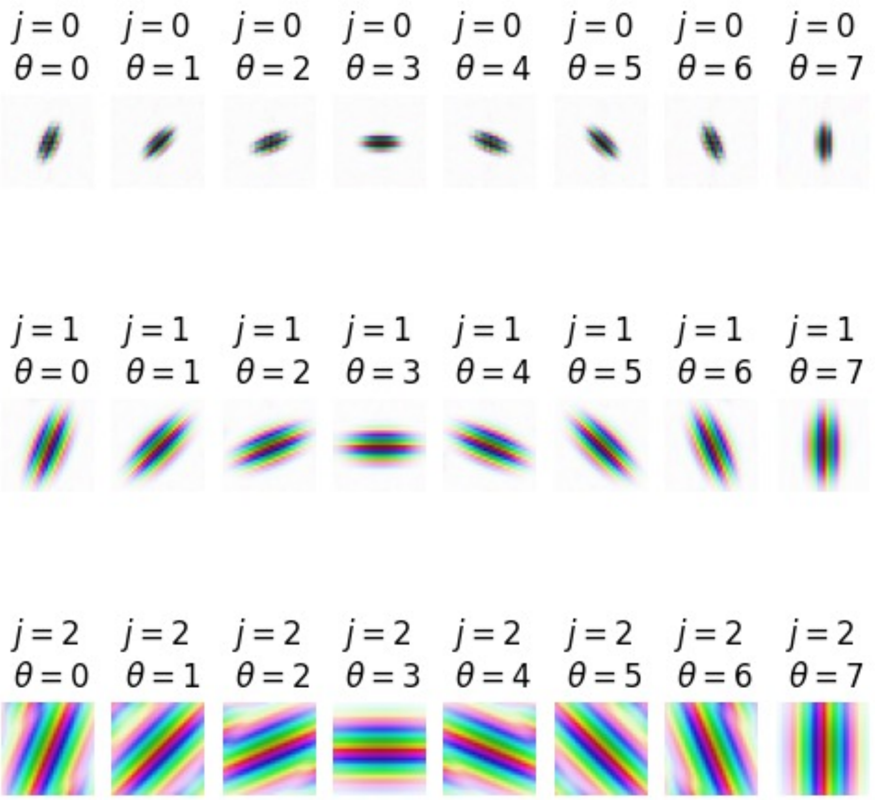
Output

$\hat{\mu}(X), \hat{\sigma}(X)$

- Idea: Filter images with wavelets at various orientations and scales to extract frequency information, apply non-linearity and pooling, then project the features linearly.
- Advantages: Very few learnable parameters & interpretable by design

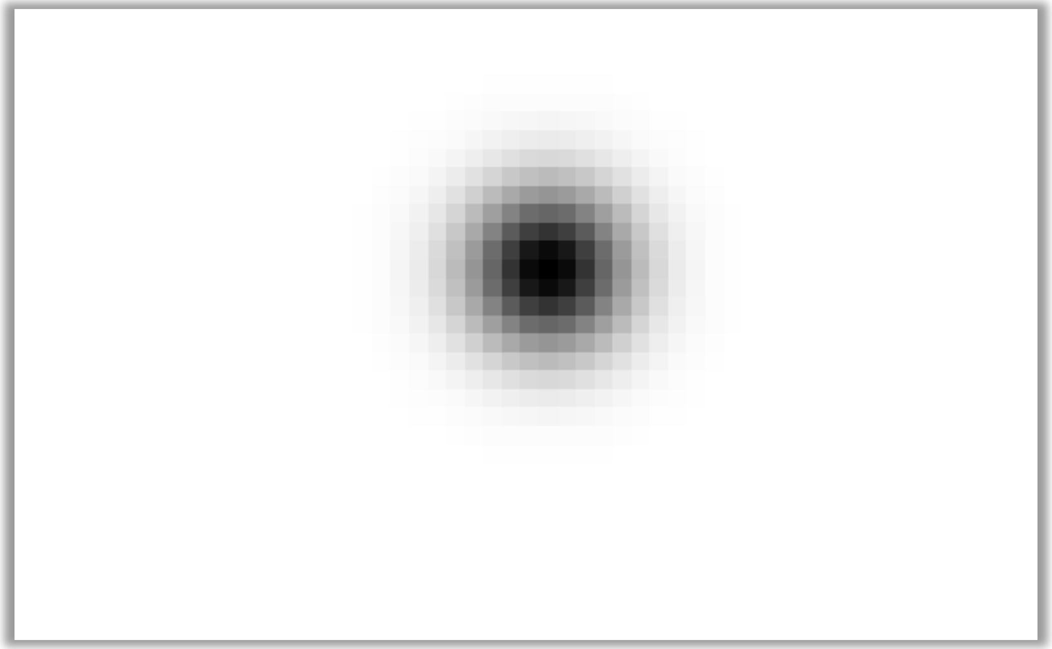
An intermediate model with a convolutional architecture: Scattering Networks

Wavelets filters for different scales and orientations



(Bruna & Mallat, 2012)

Gaussian low-pass filter



2 hyperparameters: J and L

$$0 \leq j < J$$
$$0 \leq \theta < L$$

Can we find an optimum between accuracy and interpretability?

Method	$\hat{\mu}(X;\theta)$	$\hat{\sigma}(X;\theta)$	trainable parameters	non-trainable parameters	hyperparameters
GA	$M \cdot X$	σ	27 425	0	1
IINN	$g_\mu(M \cdot X)$	$s(g_\sigma(M \cdot X))$	55 058*	0	10
ScatNet	$\beta_\mu \cdot \phi(X)$	$s(\beta_\sigma \cdot \phi(X))$	19 930*	656 640*	5
CNN	$g_\mu(X)$	$s(g_\sigma(X))$	684 000*	0	10

To predict extreme heatwaves, we used a hierarchy of models from the simplest to the more complex ones.

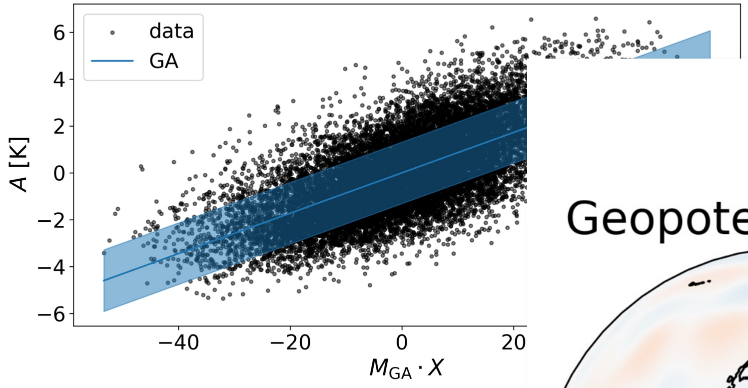
Who needs CNNs when you have wavelets?

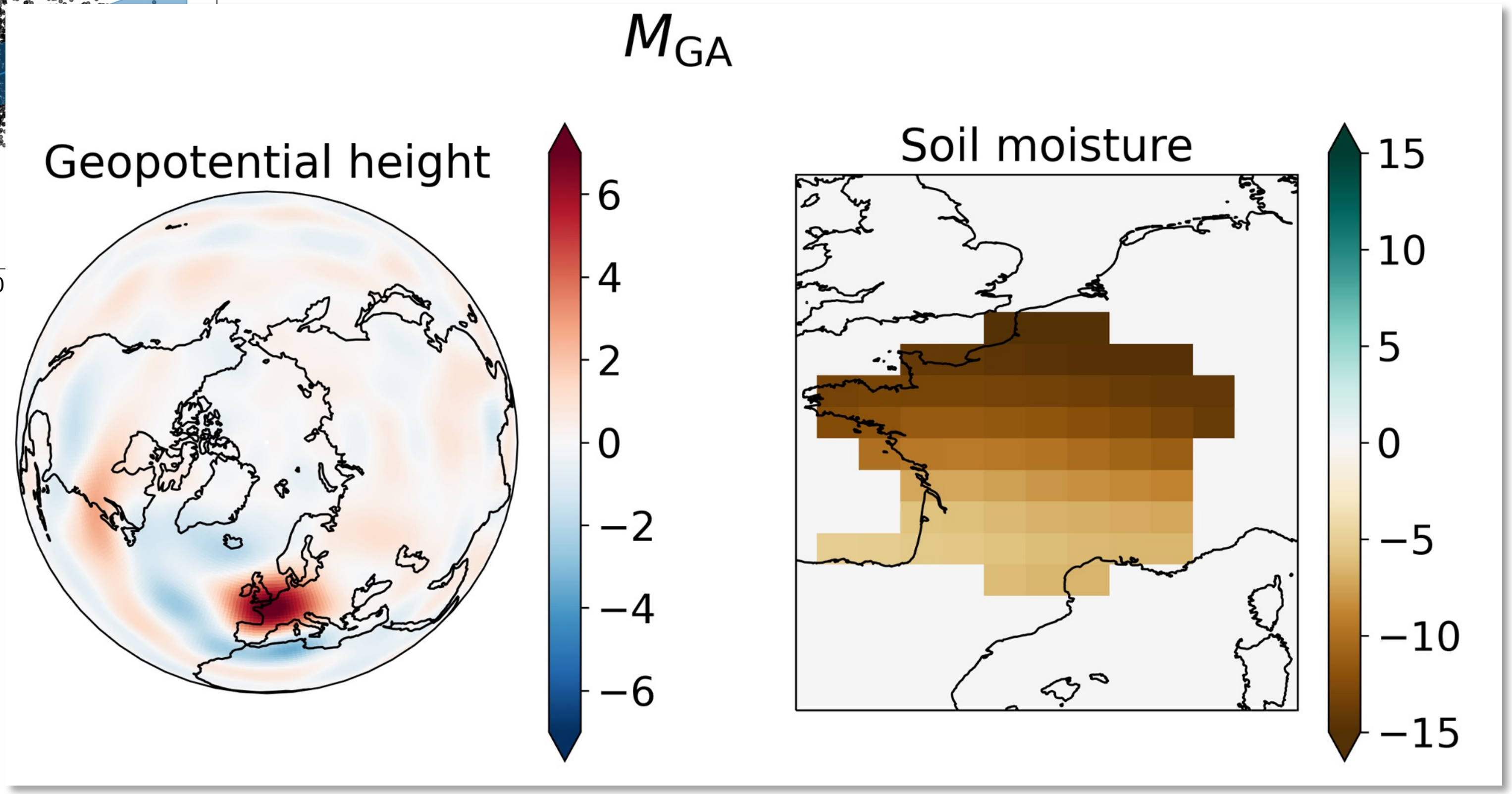
The higher, the better

Model	Metric: CRPSS	NLLS	BCES
GA	0.2864 ± 0.0009	0.2169 ± 0.0009	0.293 ± 0.001
IINN	0.287 ± 0.002	0.217 ± 0.002	0.291 ± 0.003
ScatNet	0.3097 ± 0.0007	$\mathbf{0.246 \pm 0.003}$	$\mathbf{0.314 \pm 0.005}$
CNN	$\mathbf{0.310 \pm 0.003}$	0.245 ± 0.007	0.311 ± 0.008

- Skill above linearity: GA is outperformed by more than 10%.
- We don’t need to learn everything: ScatNet has higher skill than CNN.
- Perfect candidate for low data regimes: ScatNet has fewer parameters than GA and is robust when trained with less data (e.g., 80 years; not shown).

Gaussian Approximation: What are the optimal projection patterns?

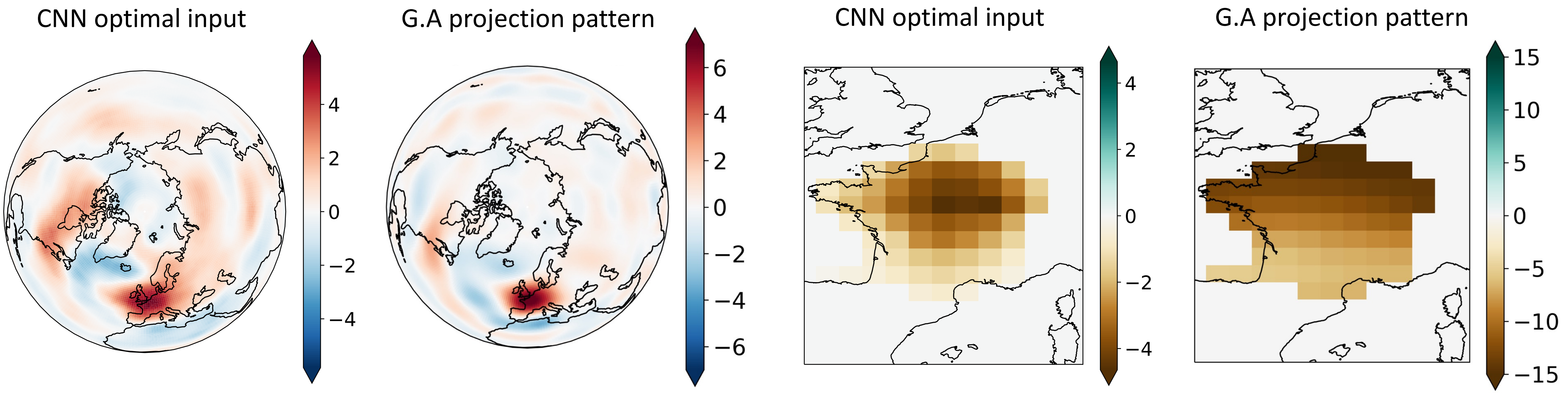




- Persistent anticyclonic anomaly over France
- Dry soil is important (more in North than South France)
- Coherent patterns over the North Atlantic linked with the path or strength of the Jet Stream.

Ad-hoc explainability tools reveal similar patterns for CNN and G.A

$$\min_{S} \mathcal{L}(S) = -\hat{\mu}_{\mathrm{CNN}}(S) + \lambda_2(\| S \| - n_0)^2 + \lambda_r\left(\sqrt{H_2(S)} - r_0\right)^2$$

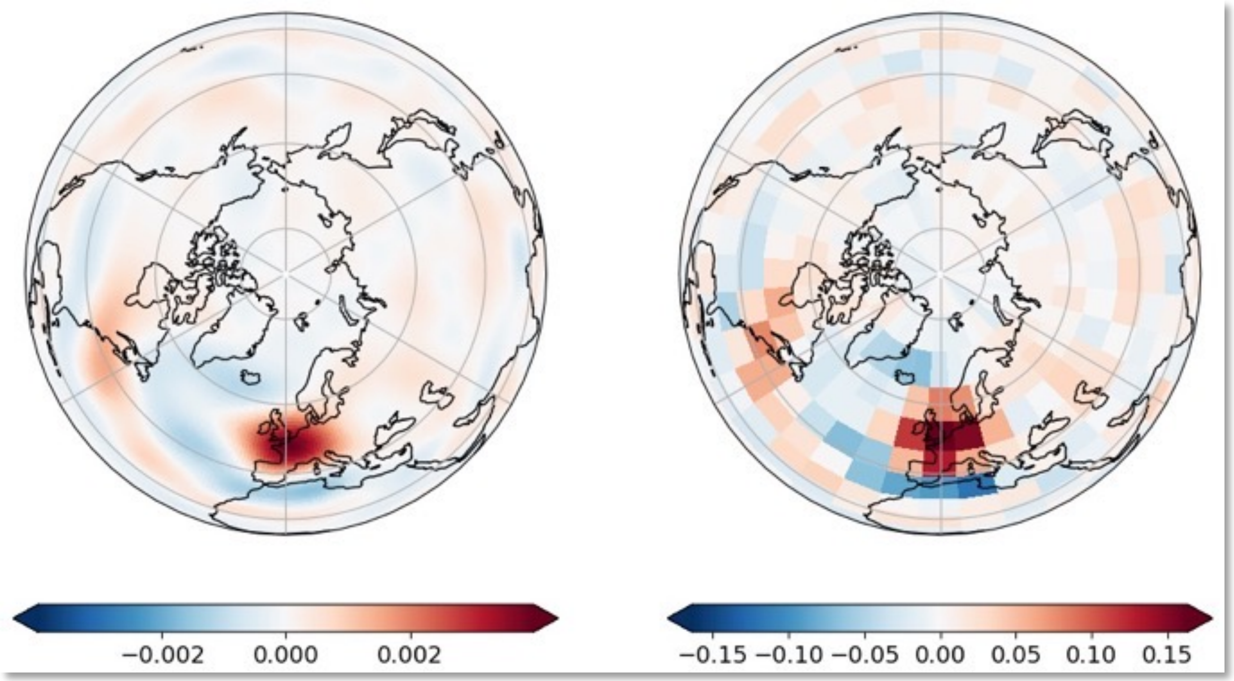


- Very similar patterns
- Smaller weight to the positive geopotential anomaly over Western Europe and extend further North over Scandinavia
- Hard to get more insights on how the CNN makes his predictions

At first order, ScatNet behaves as the Gaussian Approximation

G.A projection pattern

ScatNet coarse field projection pattern



The higher, the better

Model	Metric: CRPSS	NLLS	BCES
GA	0.2864 ± 0.0009	0.2169 ± 0.0009	0.293 ± 0.001
ScatNet$_{\text{coarse}}$	0.2862 ± 0.0005	0.203 ± 0.001	0.291 ± 0.001
CNN	0.310 ± 0.003	0.245 ± 0.007	0.311 ± 0.008

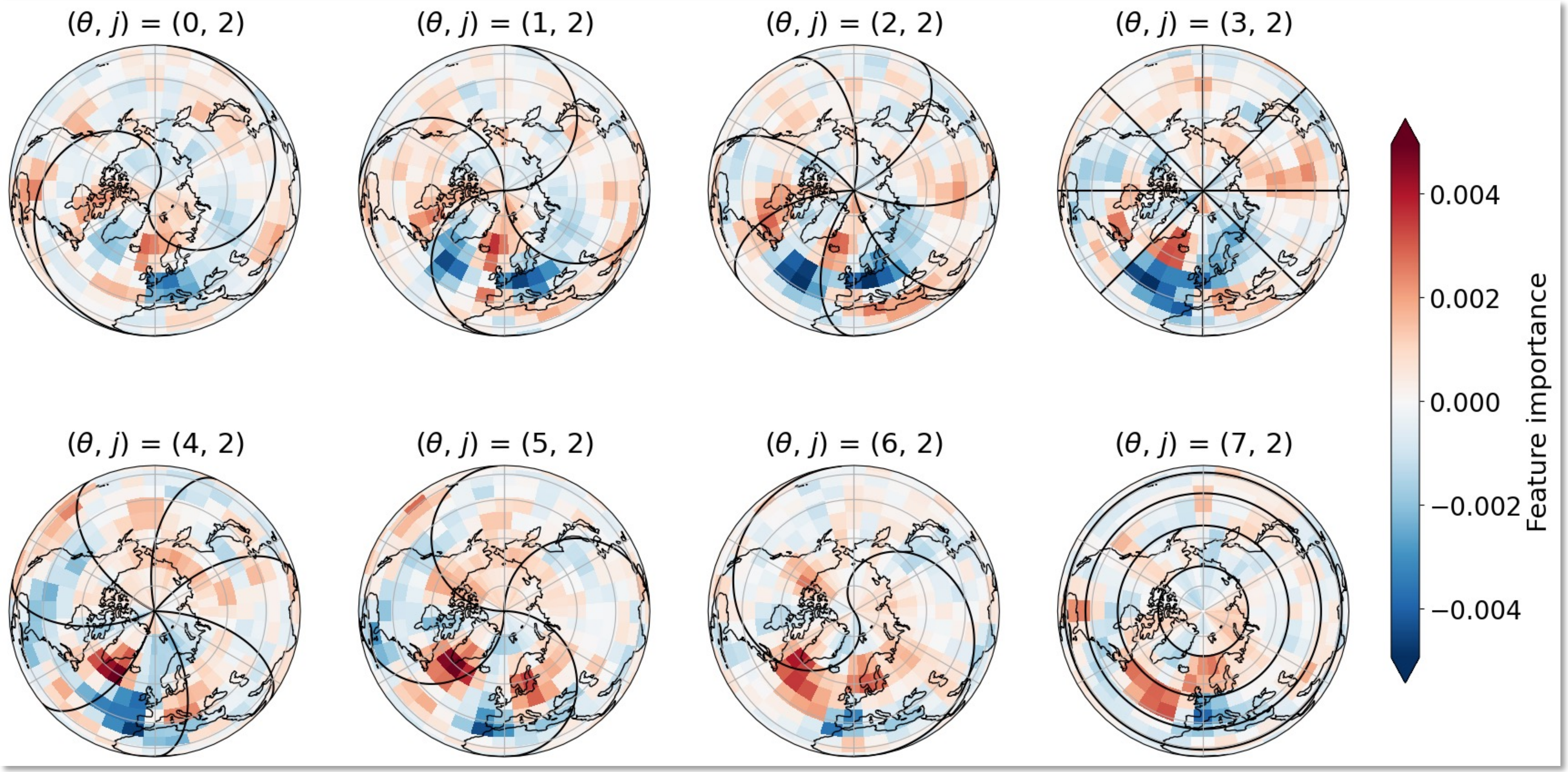
ScatNet can identifies the most relevant scales for prediction

Relative **feature importance** of various scales, expressed as percentages

	scale $j=0$	scale $j=1$	scale $j=2$	coarse field	soil moisture
Relative FI (in %)	5.2 ± 0.3	10.4 ± 0.5	20.0 ± 1.0	51.0 ± 2.4	13.5 ± 1.0

Correspond to oscillations that are 2^2 pixels large, i.e. ~ 400 km

Allow to inspect a whole new set of feature importance patterns for each part of the Fourier domain



Conclusion and perspectives

- Skill above linearity: GA is outperformed by more than 10%.
- CNN model does not outperform the interpretable ScatNet, even with enough data.
- CNN explanations are minor variations of known results from GA.
- ScatNet identifies relevant scales and orientations for prediction, linking performance gains from GA to 400 km wavelength oscillations in the geopotential height field over the North Atlantic.
- ScatNet is a promising tool for climate science applications.

2

Riding the (Large-Scale) Deep Learning wave: Using AI-Emulators to design better score functions

How traditional climate/weather models works?

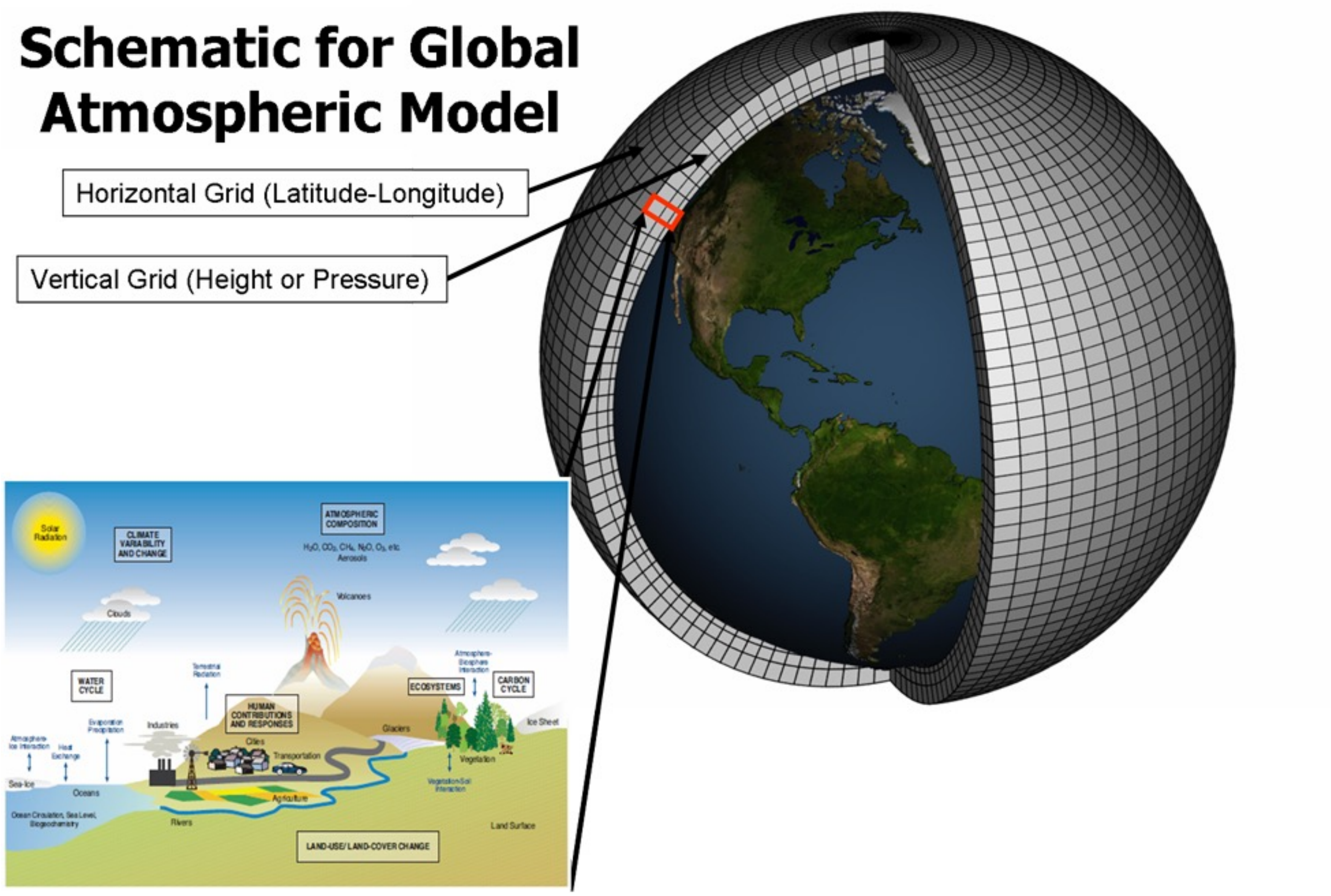
State variable X = [u(x,y,z), T(x,y,z), ...]

- Pressure p
- Free surface η
- salinity s (ocean), humidity H (atmos.)

Dynamics with partial differential equation (PDE)

$$\partial_t \mathbf{X} = F(\mathbf{X})$$

- Fluid-dynamics: Navier-stokes equation simplification
- These models are **very costly to run**



2022-today: The deep learning revolution

Big Tech is starting the race to pure data-driven weather forecast…

2022

FOURCASTNET: A GLOBAL DATA-DRIVEN HIGH-RESOLUTION WEATHER MODEL USING ADAPTIVE FOURIER NEURAL OPERATORS

A PREPRINT

NVIDIA®

2022

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*

HUAWEI

2023

GraphCast: Learning skillful medium-range global weather forecasting

Remi Lam[*,1], Alvaro Sanchez-Gonzalez[*,1], Matthew Willson[*,1], Peter Wirnsberger[*,1], Meire Fortunato[*,1], Ferran Alet[*,1], Suman Ravuri[*,1], Timo Ewalds[1], Zach Eaton-Rosen[1], Weihua Hu[1], Alexander Merose[2], Stephan Hoyer[2], George Holland[1], Oriol Vinyals[1], Jacklynn Stott[1], Alexander Pritzel[1], Shakir Mohamed[1] and Peter Battaglia[1]

[*]equal contribution, [1]Google DeepMind, [2]Google Research

Google DeepMind

Rte LMD

How these models work?

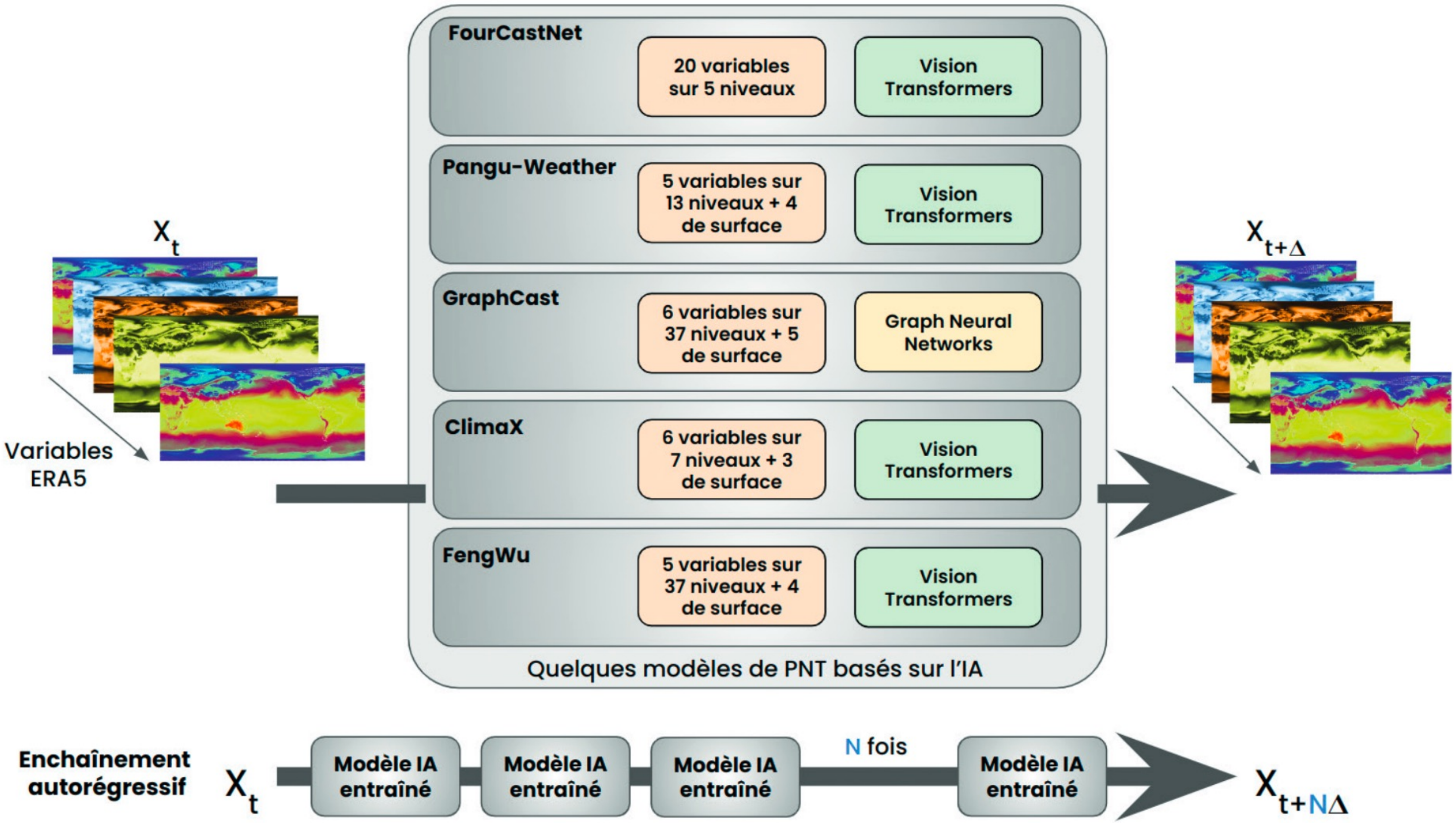
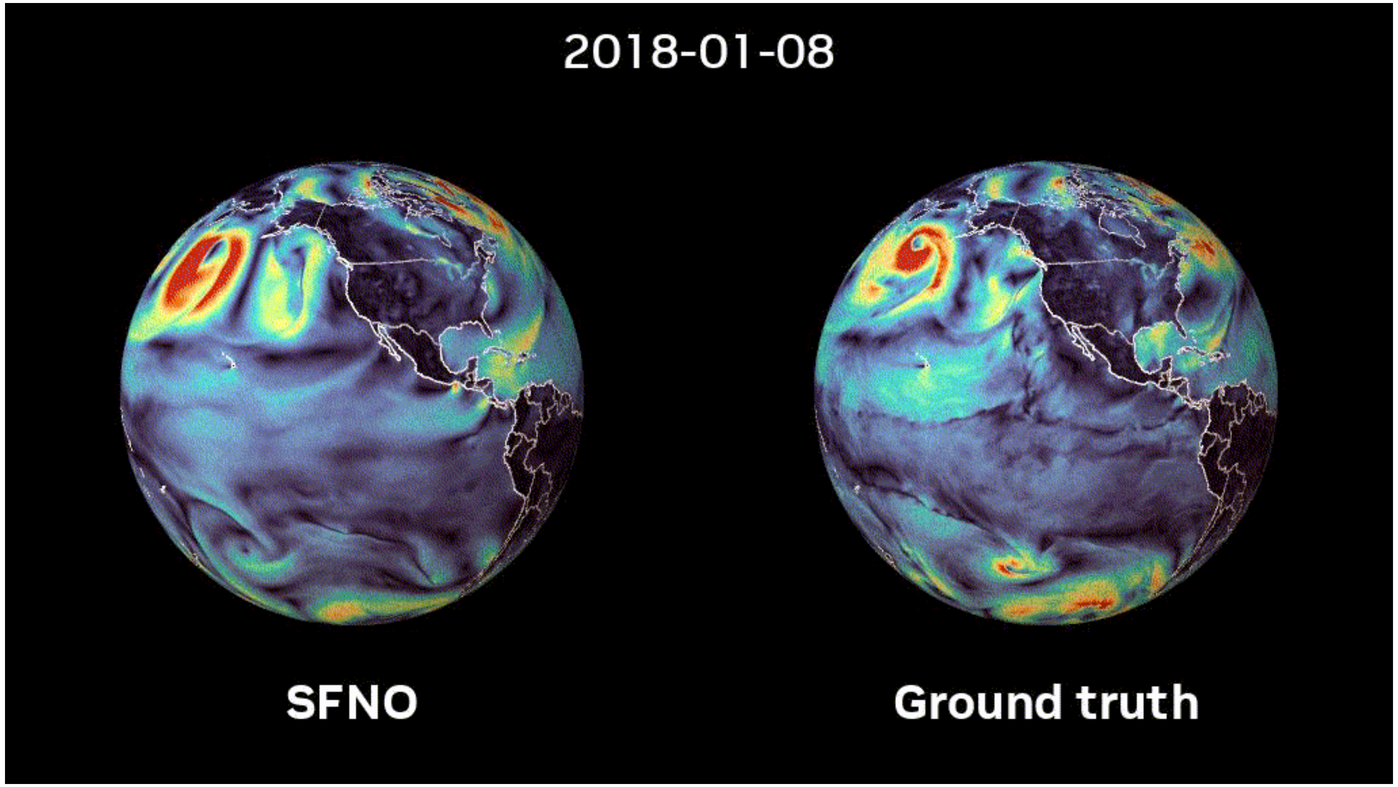


Figure 1. Schéma récapitulatif des modèles présentés ici qui partent d'un état de l'atmosphère pour le prévoir à une courte échéance Δ. Pour produire une prévision sur un horizon plus long *N*Δ, le modèle utilisé est enchaîné *N* fois. Il est à noter que GraphCast prend non seulement *X*(*t*), mais aussi *X*(*t* – Δ), mais cette spécificité est omise dans cette figure pour une meilleure lisibilité.

Source: R. Lguensat, *Les nouveaux modèles de prévision météorologique basés sur l'intelligence artificielle : opportunité ou menace ?*, La Métérologie, 2023.

Potential benefits of climate models emulators

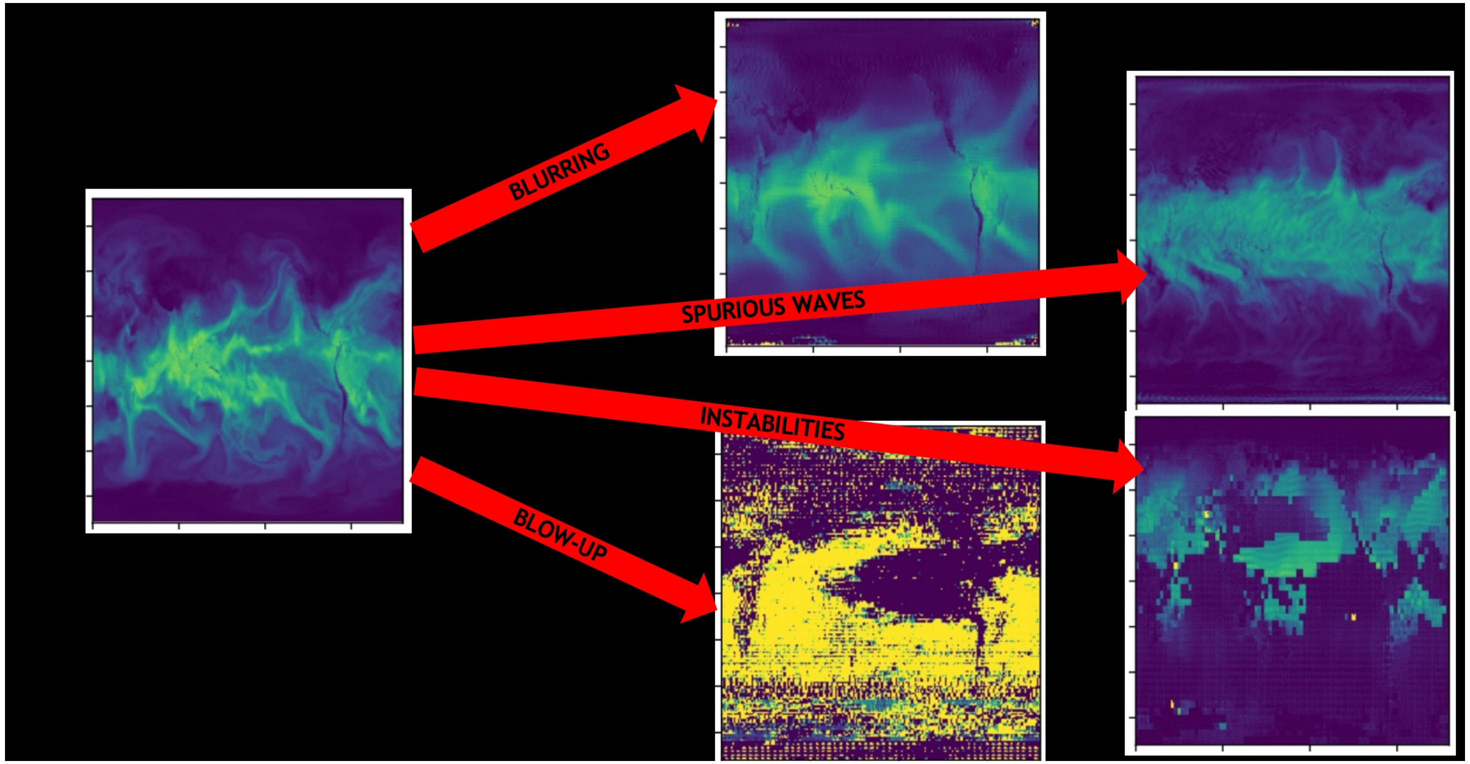
Source: ACE emulator, 2023



- **10,000 to 100,000 faster** than conventional GCMs allowing Huge Ensemble simulations. Particular interest to study **rare events** and climate projections.
- These are differentiable models. Particular interest for **data assimilation**.
- If trained directly on observations, could it help discover **new physics**?

Limitations of Deep Learning weather forecasts

- The lack of physical constraints in deep learning approaches can lead to **non-physical predictions**.
- These models make **deterministic predictions**, making it difficult to estimate the uncertainty surrounding their predictions.
- A less studied challenge is their limited ability to **extrapolate** beyond their training range.

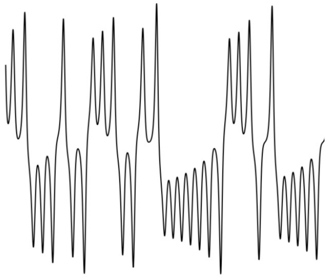


Limitations of Deep Learning weather forecasts

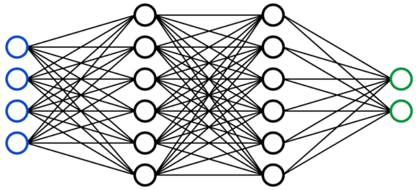
- The lack of physical constraints in deep learning approaches can lead to **non-physical predictions**.
- These models make **deterministic predictions**, making it difficult to estimate the uncertainty surrounding their predictions.
- A less studied challenge is their limited ability to **extrapolate** beyond their training range.

Our approach: AI-RES

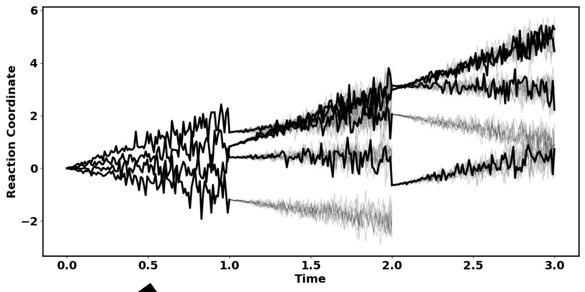
Data



Train emulator



Use an emulator in score function to drive rare event sampling in Climate Model



Repeat as needed using additional rare event instances

(1) Accurately estimate return period in Climate Model
(2) Improve emulation of rare events

We first want to test the procedure for a cheap-to-run Climate Model

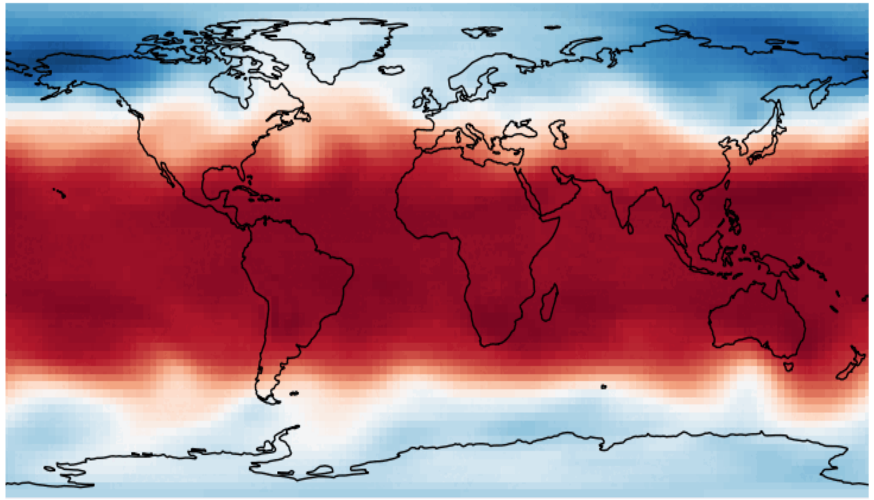
- **PlaSim**: an intermediate complexity GCM with simple parametrizations

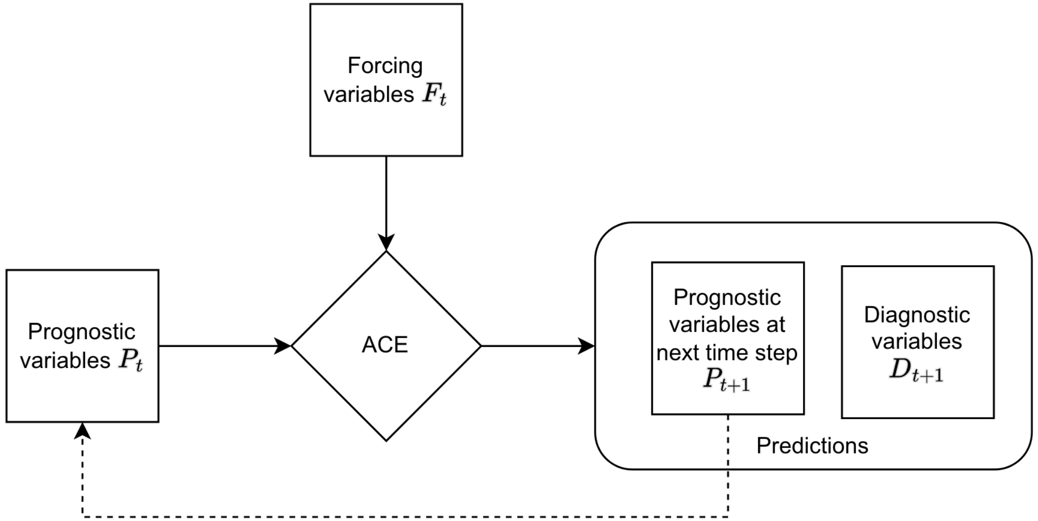
- We have a 100,000 years control run to study statistics of rare events

- We can impose forcings such as **soil moisture (or other slow drivers)** to study the impact on relevant extreme events (heatwaves, cold snaps, etc.)

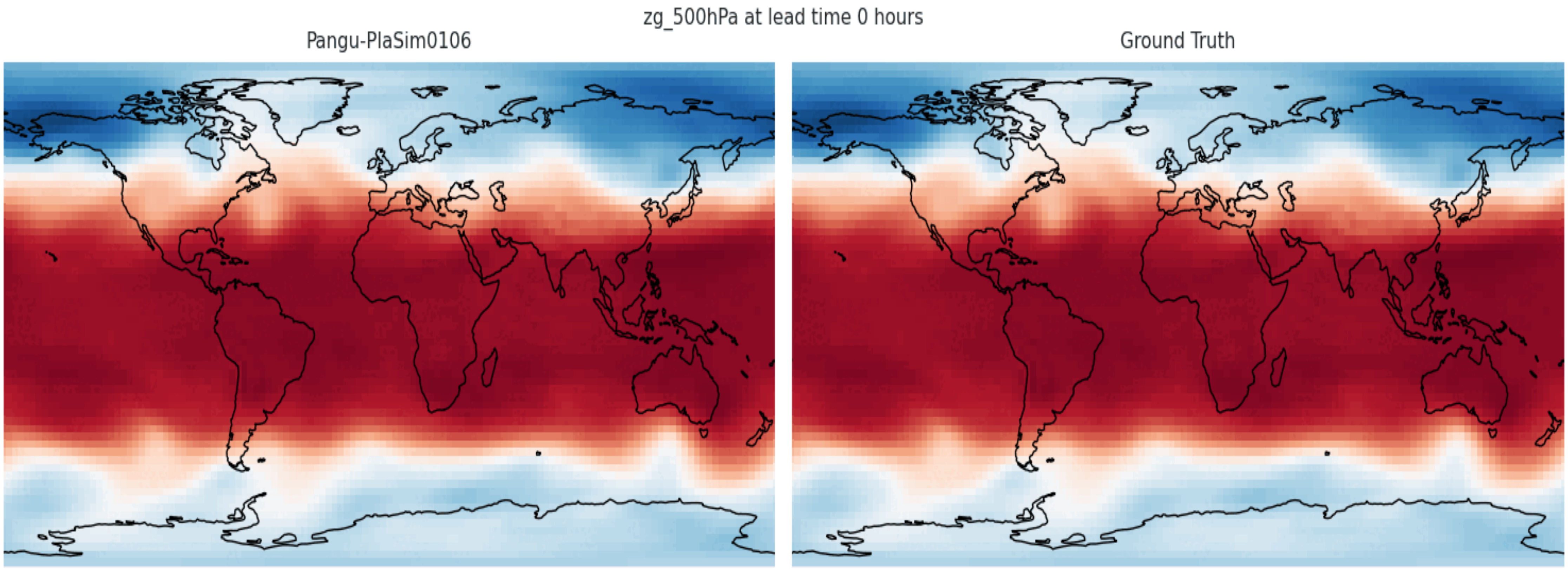
- The first step is to develop a (stable) **dynamical emulator** of this model.

Z500 from a PlaSim simulation





First step accomplished?



Next steps and questions to answer

- Finish the hyperparameter optimization of the emulator…
- Add other varying boundary variables such as soil moisture, snow cover, etc.
- Test the emulator on the prediction of extremes
- Test the emulator on probabilistic metrics
  - How to construct good ensembles with the emulator?
- Use emulator predictions as score function in Rare event algorithm
  - Which algo?
  - Which event of interest?
  - Which criteria of success for the algorithm?

Thank you for your attention!