

La recherche en IA
face au changement climatique et pour la durabilité environnementale

Claire Monteleoni

INRIA Paris

University of Colorado Boulder

Choose France™



Décembre 2021 : comté de Boulder, Colorado

- La sécheresse neigeuse de l'automne et de l'hiver 2021 a asséché la couverture du sol
- Des vents de 130-160 km/h, plus une étincelle, ont déclenché une « tempête de feu » incontrôlable
- 2 morts. 1 000 maisons et 20 commerces détruits, et d'autres endommagés





« L'IA représente une occasion considérable pour la Terre. L'explosion actuelle de l'IA nous conduira à l'intégrer dans de plus en plus de choses chaque année.... En pensant aux gains, aux gains d'efficacité et aux nouvelles solutions qu'elle apporte aux nations, aux entreprises et à la vie quotidienne, nous devons aussi réfléchir à la manière de maximiser les bénéfices pour la société et pour notre environnement dans son ensemble. »

– Forum économique mondial (2018), « Harnessing Artificial Intelligence for the Earth »





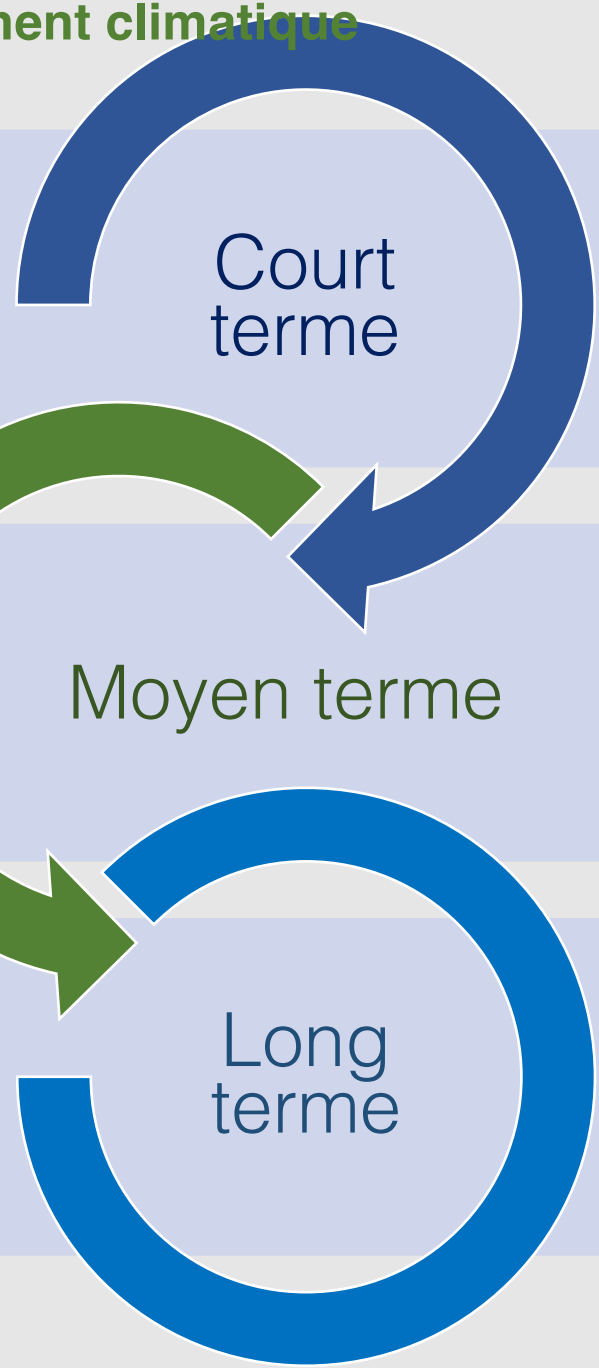
La recherche en IA face au changement climatique et pour la durabilité environnementale

CHANGEMENT CLIMATIQUE

ADAPTATION

ATTÉNUATION

IMPACTS



Météo extrême

Aléas en cascade

Transition énergétique

Changement d'usage des sols

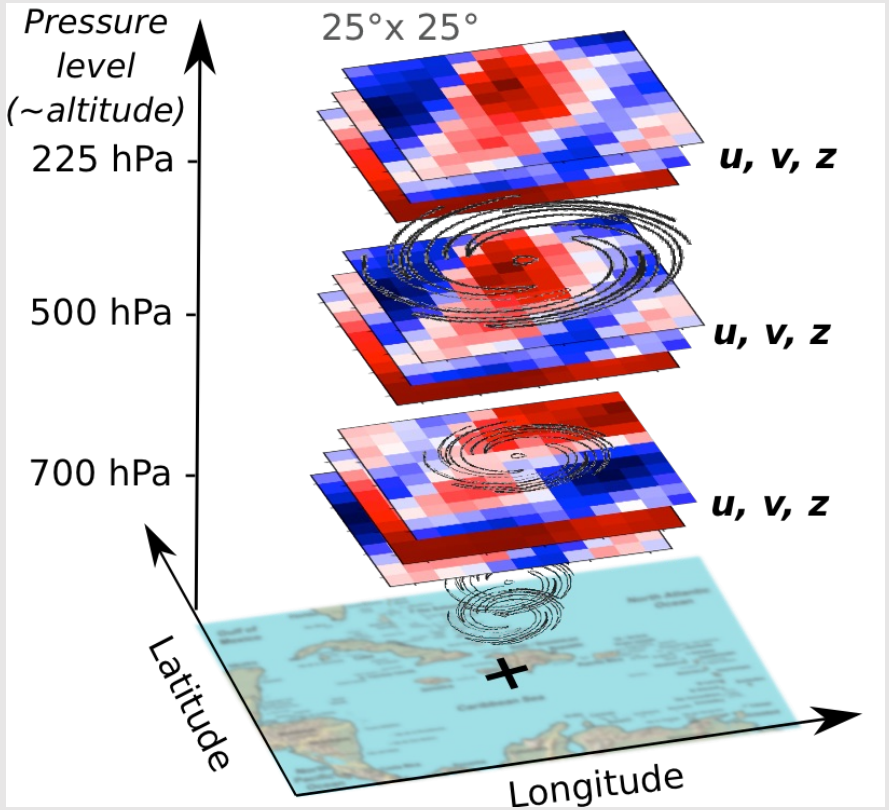
Émissions de carbone

Hausse du niveau de la mer

Solutions fondées sur l'IA

ADAPTATION

L'IA pour la météo extrême et les aléas en cascade



[Giffard-Roisin et al., Frontiers 2020]

Prévision des trajectoires d'ouragans

Prévision des précipitations extrêmes de la mousson indienne d'été

Détection des avalanches

IA générative pour la prévision météorologique



ATTÉNUATION

Réduire les émissions de carbone

Accélérer la transition vers l'énergie verte

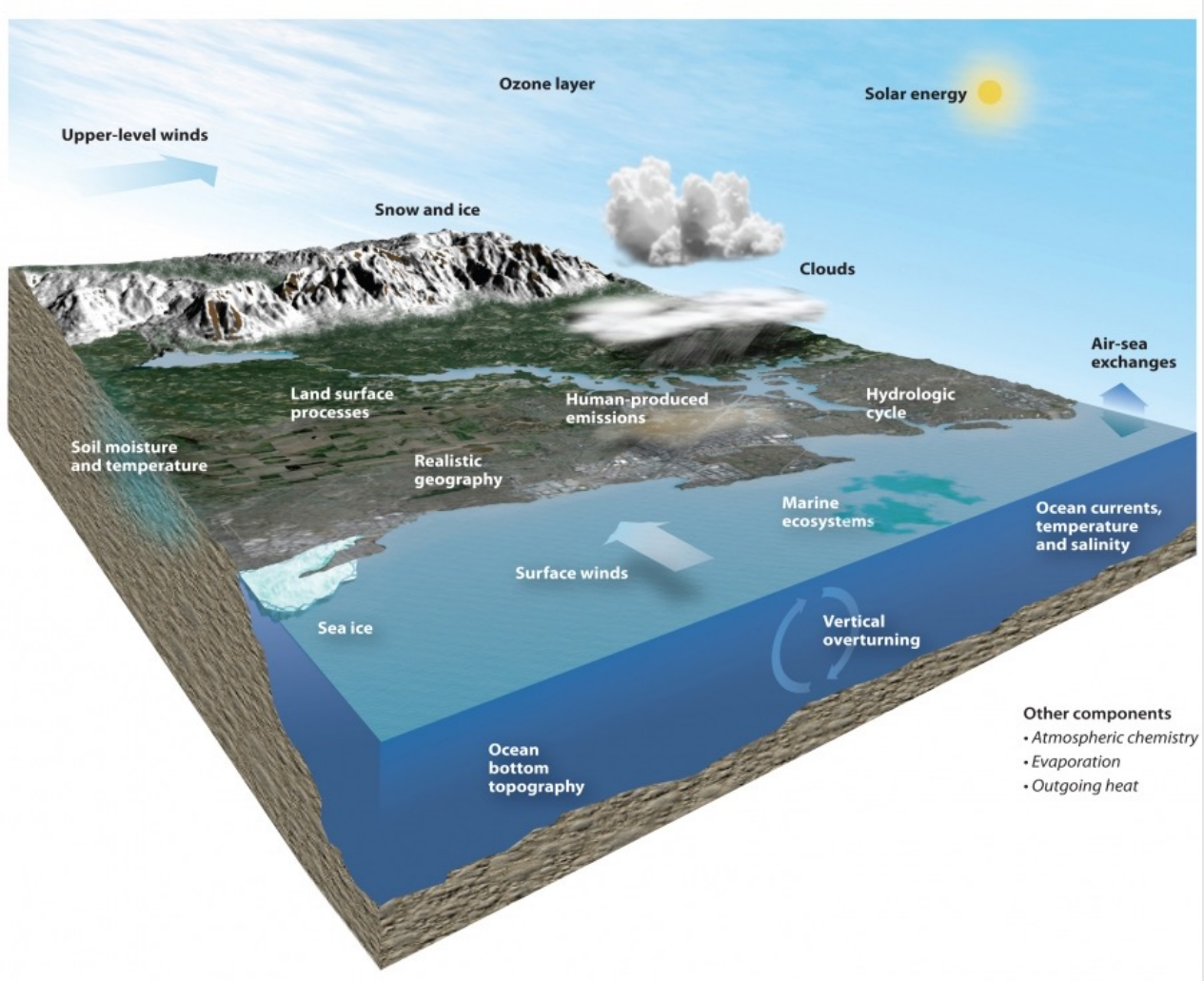
- Prévision du solaire et de l'éolien par l'IA
- L'IA pour la descente d'échelle des données solaires et éoliennes

Réduire le calcul de la modélisation météo et climat

- Une fois entraînée, l'IA prédit beaucoup plus vite que les modèles physiques

IMPACTS

L'IA pour comprendre et prévoir le changement climatique



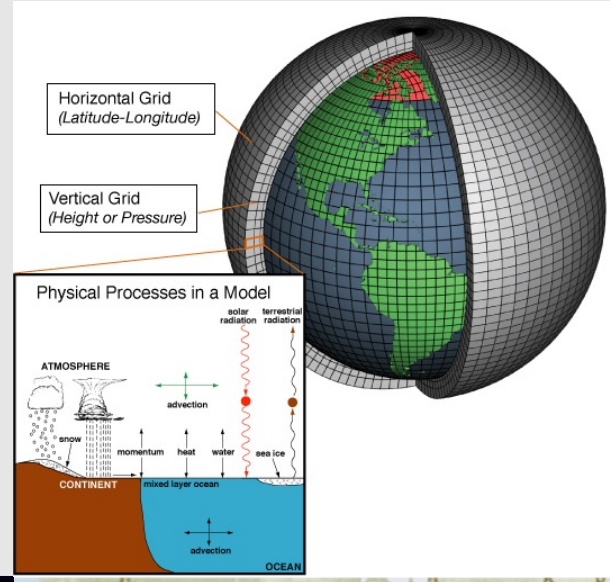
UCAR Science Education

Utiliser l'IA pour apprendre les relations entre simulations du GIEC et observations

- Rendre plus robustes les prévisions d'ensembles de modèles climatiques
- Projeter la hausse du niveau de la mer à long terme
- Projeter les émissions de carbone à long terme

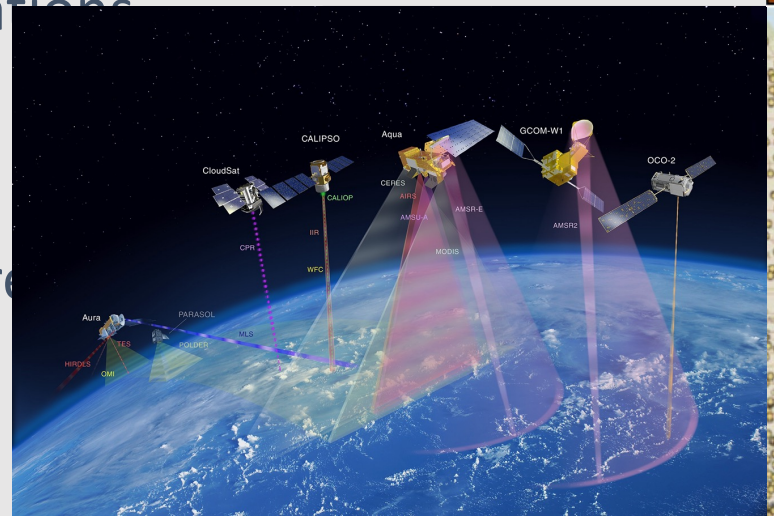
Approche : exploiter toutes les données disponibles

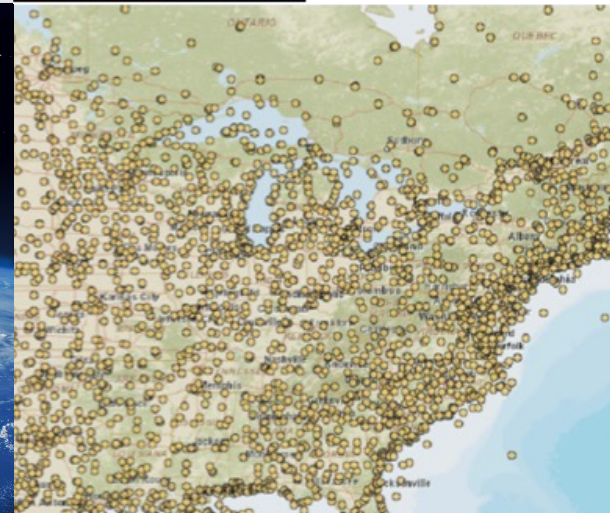
- Données simulées produites par des modèles physiques
  - Modèles de prévision numérique du temps (PNT)
  - Modèles de circulation générale (MCG)
  - Modèles climatiques régionaux (MCR)



- Données de réanalyse
  - Produits maillés issus de l'assimilation de données : applique les lois physiques aux observations

- Données d'observation
  - Données de télédétection satellitaire
  - Données in situ





Tâches d'AA sous-jacentes à beaucoup de ces problèmes

- ❑ Fusion de données
- ❑ Prévision
- ❑ Apprentissage non supervisé
- ❑ Alignement probabiliste de domaines
- ❑ Descente d'échelle et interpolation

Méthodes d’IA

- Apprentissage non supervisé
- Apprentissage profond génératif

Plan pour la suite de la journée

I. **Contexte :** la révolution de l'IA pour la prévision météorologique

Conséquences pour l'étude du changement climatique

II. **Méthodes d'IA :** qu'est-ce que l'apprentissage auto-supervisé ?

Applications au climat et à la météo

III. **Méthodes d'IA :** qu'est-ce que l'apprentissage profond génératif ?

Partie I : Contexte : la révolution de l'IA pour la prévision météorologique

La révolution silencieuse de la prévision météorologique

- La prévision numérique du temps (PNT) n'a cessé de progresser au fil des décennies
- Une grande réussite scientifique
  - Un problème de simulation complexe résolu **en routine** dans les centres de prévision opérationnels du monde entier
- Forte valeur socio-économique
  - Atténuation des aléas climatiques, logistique, prévention des pertes matérielles, …

Crédit des premières diapositives : David Landry (INRIA Google DeepMind)

doi:10.1038/nature14956

REVIEW

doi:10.1038/nature14956

The quiet revolution of numerical weather prediction

Peter Bauer[1], Alan Thorpe[1] & Gilbert Brunet[2]

Advances in numerical weather prediction represent a quiet revolution because they have resulted from a steady accumulation of scientific knowledge and technological advances over many years that, with only a few exceptions, have not been associated with the aura of fundamental physics breakthroughs. Nonetheless, the impact of numerical weather prediction is among the greatest of any area of physical science. As a computational problem, global weather prediction is comparable to the simulation of the human brain and of the evolution of the early Universe, and it is performed every day at major operational centres across the world.

At the turn of the twentieth century, Abbe[1] and Bjerknes[2] proposed that the laws of physics could be used to forecast the weather; they recognized that predicting the state of the atmosphere could be treated as an initial value problem of mathematical physics, wherein future weather is determined by integrating the governing partial differential equations, starting from the observed current weather. This proposition, even with the most optimistic interpretation of Newtonian determinism, is all the more audacious given that, at that time, there were few routine observations of the state of the atmosphere, no computers, and little understanding of whether the weather possesses any significant degree of predictability. But today, more than 100 years later, this paradigm translates into solving daily a system of nonlinear differential equations at about half a billion points per time step between the initial time and weeks to months ahead, and accounting for dynamic, thermodynamic, radiative and chemical processes working on scales from hundreds of metres to thousands of kilometres and from seconds to weeks.

A touchstone of scientific knowledge and understanding is the ability to predict accurately the outcome of an experiment. In meteorology, this translates into the accuracy of the weather forecast. In addition, today's numerical weather predictions also enable the forecaster to assess quantitatively the degree of confidence users should have in any particular forecast. This is a story of profound and fundamental scientific success built upon the application of the classical laws of physics. Clearly the success has required technological acumen as well as scientific advances and vision.

Accurate forecasts save lives, support emergency management and mitigation of impacts and prevent economic losses from high-impact weather, and they create substantial financial revenue—for example, in energy, agriculture, transport and recreational sectors. Their substantial benefits far outweigh the costs of investing in the essential scientific research, super-computing facilities and satellite and other observational programmes that are needed to produce such forecasts[3].

These scientific and technological developments have led to increasing weather forecast skill over the past 40 years. Importantly, this skill can be objectively and quantitatively assessed, as every day we compare the forecast with what actually occurs. For example, forecast skill in the range from 3 to 10 days ahead has been increasing by about one day per decade: today's 6-day forecast is as accurate as the 5-day forecast ten years ago, as shown in Fig. 1. Predictive skill in the Northern and Southern hemispheres is almost equal today, thanks to the effective use of observational information from satellite data providing global coverage.

More visible to society, however, are extreme events. The unusual path and intensification of hurricane Sandy in October 2012 was predicted 8 days ahead, the 2010 Russian heat-wave and the 2013 US cold spell were forecast with 1–2 weeks lead time, and tropical sea surface temperature variability following the El Niño/Southern Oscillation phenomenon can be predicted 3–4 months ahead. Weather and climate prediction skill are intimately linked, because accurate climate prediction needs a good representation of weather phenomena and their statistics, as the underlying physical laws apply to all prediction time ranges.

This Review explains the fundamental scientific basis of numerical weather prediction (NWP) before highlighting three areas from which the largest benefit in predictive skill has been obtained in the past—physical process representation, ensemble forecasting and model initialization. These are also the areas that present the most challenging science questions in the next decade, but the vision of running

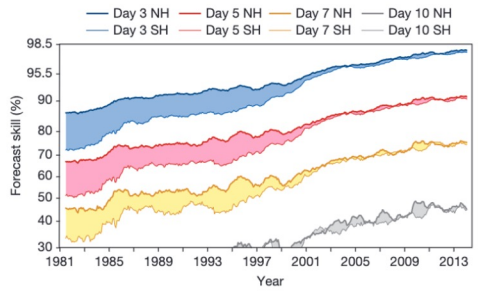
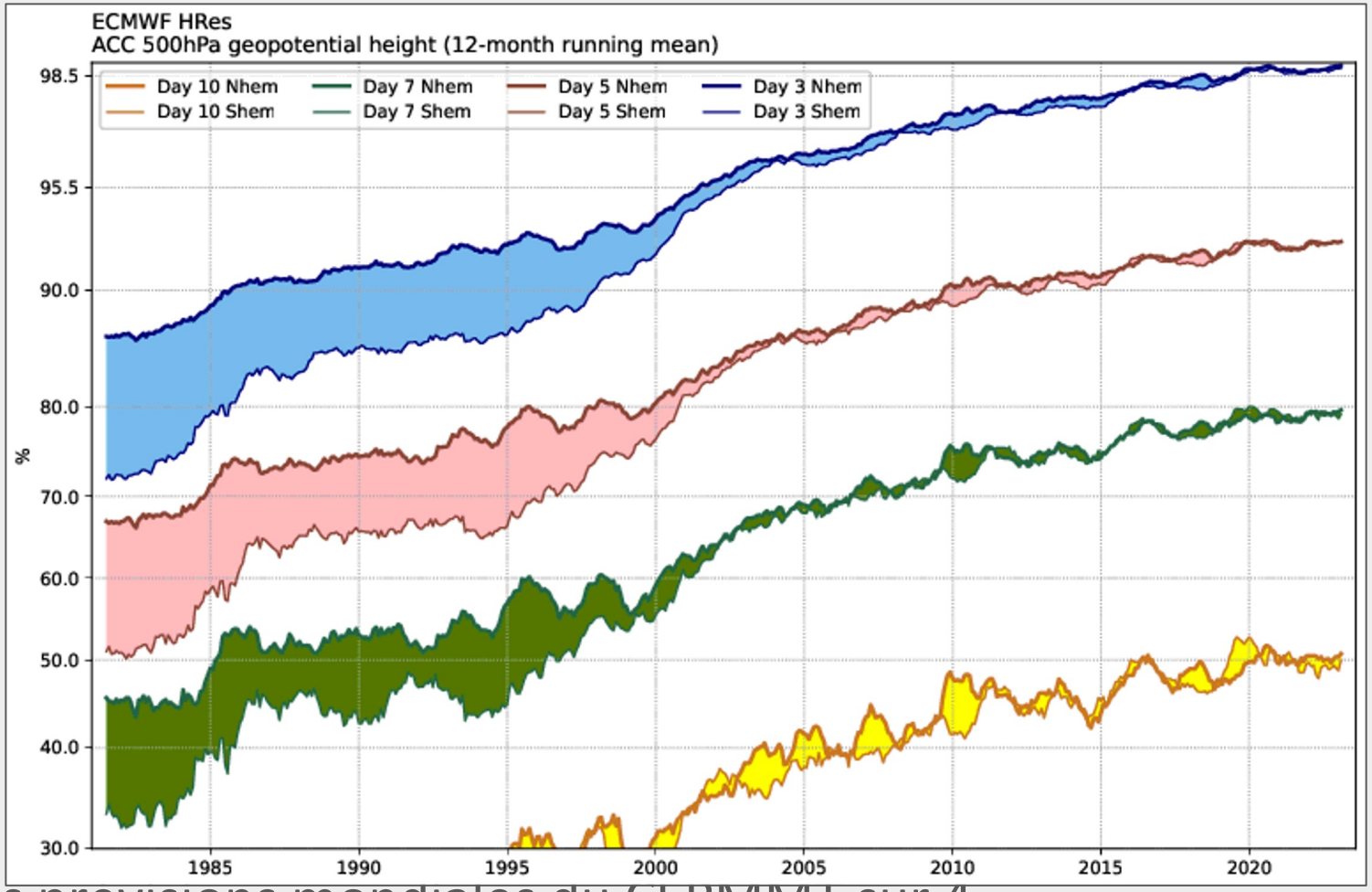


Figure 1 | A measure of forecast skill at three-, five-, seven- and ten-day ranges, computed over the extra-tropical northern and southern hemispheres. Forecast skill is the correlation between the forecasts and the verifying analysis of the height of the 500-hPa level, expressed as the anomaly with respect to the climatological height. Values greater than 60% indicate useful forecasts, while those greater than 80% represent a high degree of accuracy. The convergence of the curves for Northern Hemisphere (NH) and Southern Hemisphere (SH) after 1999 indicates the breakthrough in exploiting satellite data through the use of variational data[100].

[1]European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, UK. [2]Environment Canada, Trans-Canada Highway Dorval, Québec H9P 1J3, Canada.

©2015 Macmillan Publishers Limited. All rights reserved

La révolution silencieuse de la prévision météorologique

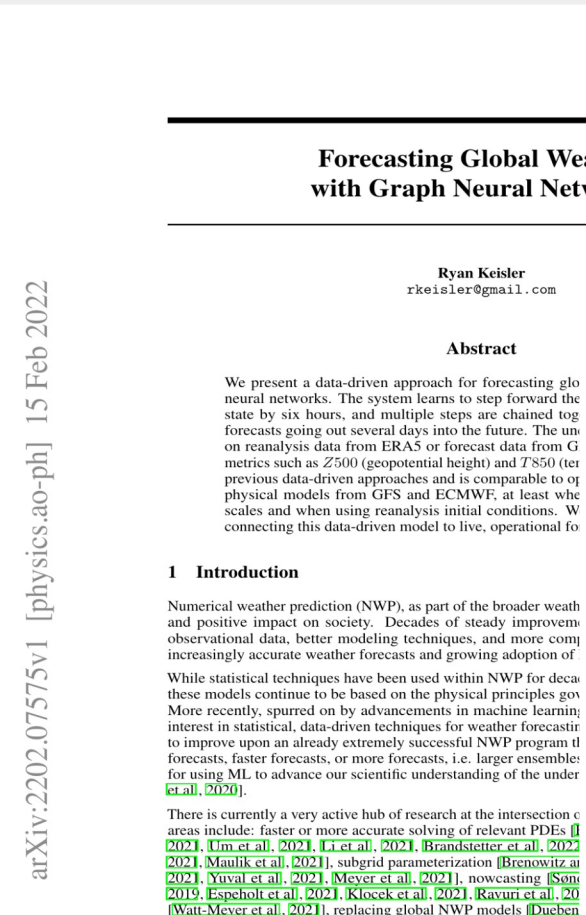


Performance des prévisions mondiales du CEPMMT sur 4 décennies

https://charts.ecmwf.int/products/plwww_m_hr_ccaf_adrian_ts?single_product=latest

Premiers modèles IA-PNT

- Depuis 2022, plusieurs équipes de recherche ont utilisé des réseaux de neurones profonds pour la prévision météorologique à moyen terme
- Certains de ces modèles ont atteint une qualité ~équivalente à celle des meilleurs modèles de PNT



Forecasting Global Weather with Graph Neural Networks

Ryan Keisler
rkeisler@gmail.com

Abstract

We present a data-driven approach for forecasting glo
neural networks. The system learns to step forward the
state by six hours, and multiple steps are chained tog
forecasts going out several days into the future. The un
on reanalysis data from ERA5 or forecast data from G
metrics such as $Z500$ (geopotential height) and $T850$ (ter
previous data-driven approaches and is comparable to op
physical models from GFS and ECMWF, at least whe
scales and when using reanalysis initial conditions. W
connecting this data-driven model to live, operational fo

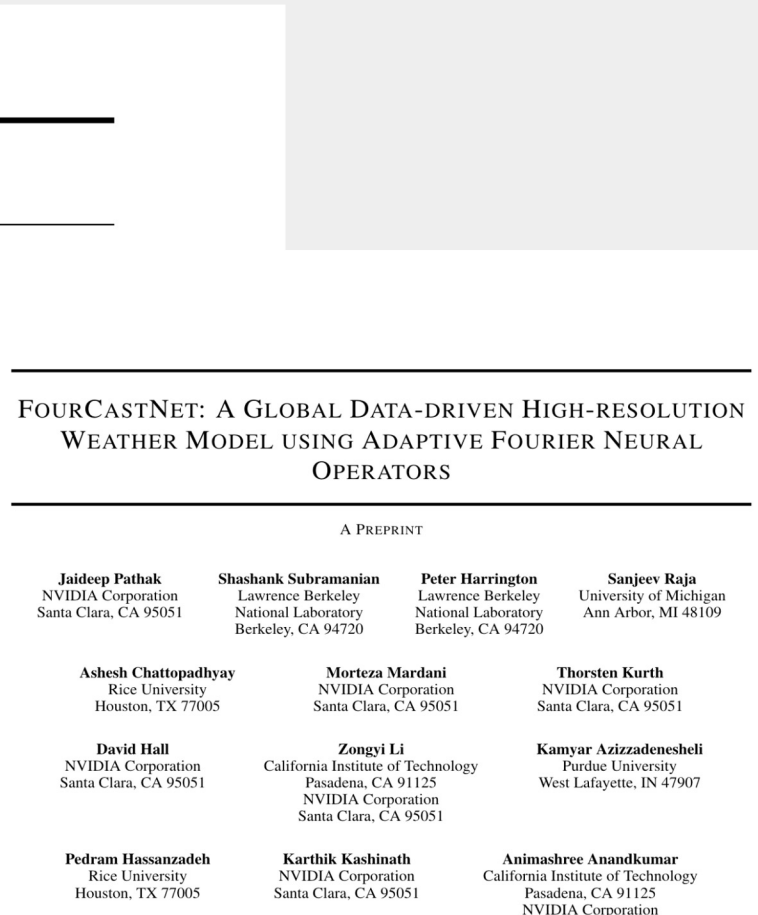
1 Introduction

Numerical weather prediction (NWP), as part of the broader weath
and positive impact on society. Decades of steady improveme
observational data, better modeling techniques, and more comp
increasingly accurate weather forecasts and growing adoption of

While statistical techniques have been used within NWP for deca
these models continue to be based on the physical principles gov
More recently, spurred on by advancements in machine learnin
interest in statistical, data-driven techniques for weather forecastin
to improve upon an already extremely successful NWP program th
forecasts, faster forecasts, or more forecasts, i.e. larger ensembles
for using ML to advance our scientific understanding of the under
et al., 2020].

There is currently a very active hub of research at the intersection o
areas include: faster or more accurate solving of relevant PDEs [
2021, Um et al., 2021, Li et al., 2021, Brandstetter et al., 2022
2021, Maulik et al., 2021], subgrid parameterization [Brenowitz a
2021, Yuval et al., 2021, Meyer et al., 2021], nowcasting [Søn
2019, Espeholt et al., 2021, Klocek et al., 2021, Ravuri et al., 20
[Watt-Meyer et al., 2021], replacing global NWP models [Dueben

arXiv:2202.07575v1 [physics.ao-ph] 15 Feb 2022



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

A PREPRINT

Jaideep Pathak
NVIDIA Corporation
Santa Clara, CA 95051

Shashank Subramanian
Lawrence Berkeley
National Laboratory
Berkeley, CA 94720

Peter Harrington
Lawrence Berkeley
National Laboratory
Berkeley, CA 94720

Sanjeev Raja
University of Michigan
Ann Arbor, MI 48109

Ashesh Chattopadhyay
Rice University
Houston, TX 77005

Morteza Mardani
NVIDIA Corporation
Santa Clara, CA 95051

Thorsten Kurth
NVIDIA Corporation
Santa Clara, CA 95051

David Hall
NVIDIA Corporation
Santa Clara, CA 95051

Zongyi Li
California Institute of Technology
Pasadena, CA 91125
NVIDIA Corporation
Santa Clara, CA 95051

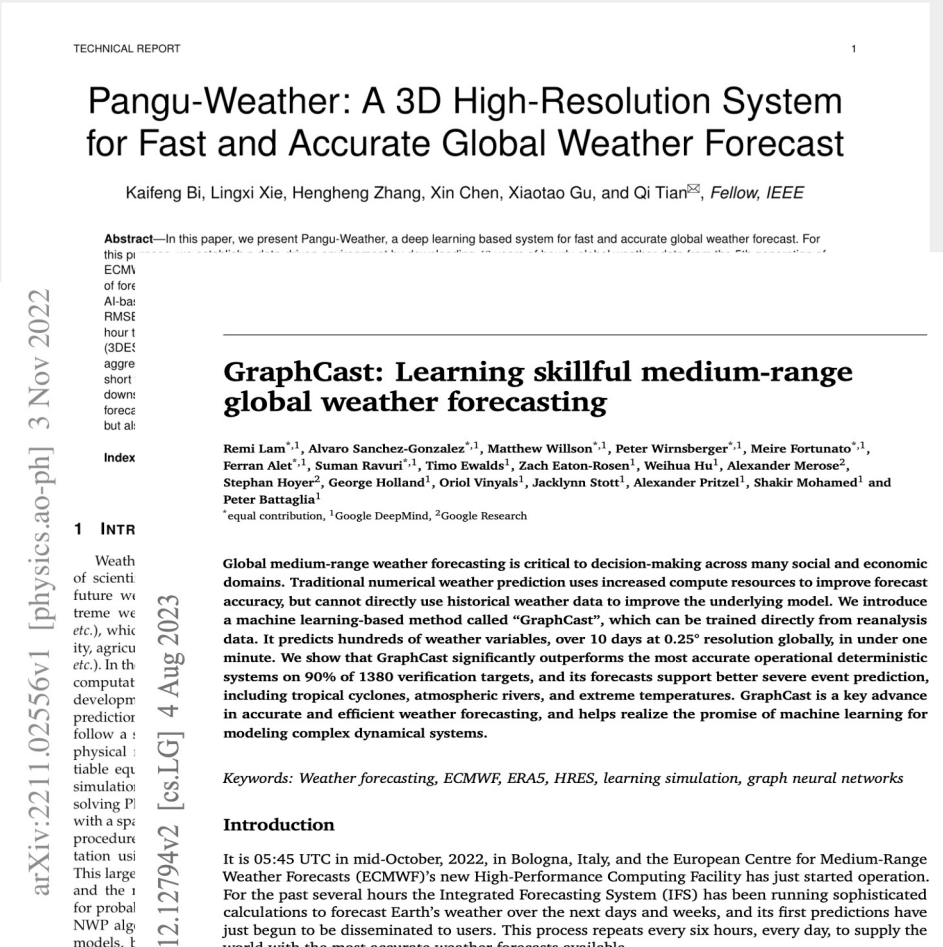
Kamyar Azizzadenesheli
Purdue University
West Lafayette, IN 47907

Pedram Hassanzadeh
Rice University
Houston, TX 77005

Karthik Kashinath
NVIDIA Corporation
Santa Clara, CA 95051

Animashree Anandkumar
California Institute of Technology
Pasadena, CA 91125
NVIDIA Corporation

[physics.ao-ph] 22 Feb 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*

Abstract—In this paper, we present Pangu-Weather, a deep learning based system for fast and accurate global weather forecast. For

arXiv:2211.02556v1 [physics.ao-ph] 3 Nov 2022

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

Global medium-range weather forecasting is critical to decision-making across many social and economic domains. Traditional numerical weather prediction uses increased compute resources to improve forecast accuracy, but cannot directly use historical weather data to improve the underlying model. We introduce a machine learning-based method called "GraphCast", which can be trained directly from reanalysis data. It predicts hundreds of weather variables, over 10 days at 0.25° resolution globally, in under one minute. We show that GraphCast significantly outperforms the most accurate operational deterministic systems on 90% of 1380 verification targets, and its forecasts support better severe event prediction, including tropical cyclones, atmospheric rivers, and extreme temperatures. GraphCast is a key advance in accurate and efficient weather forecasting, and helps realize the promise of machine learning for modeling complex dynamical systems.

Keywords: Weather forecasting, ECMWF, ERA5, HRES, learning simulation, graph neural networks

Introduction

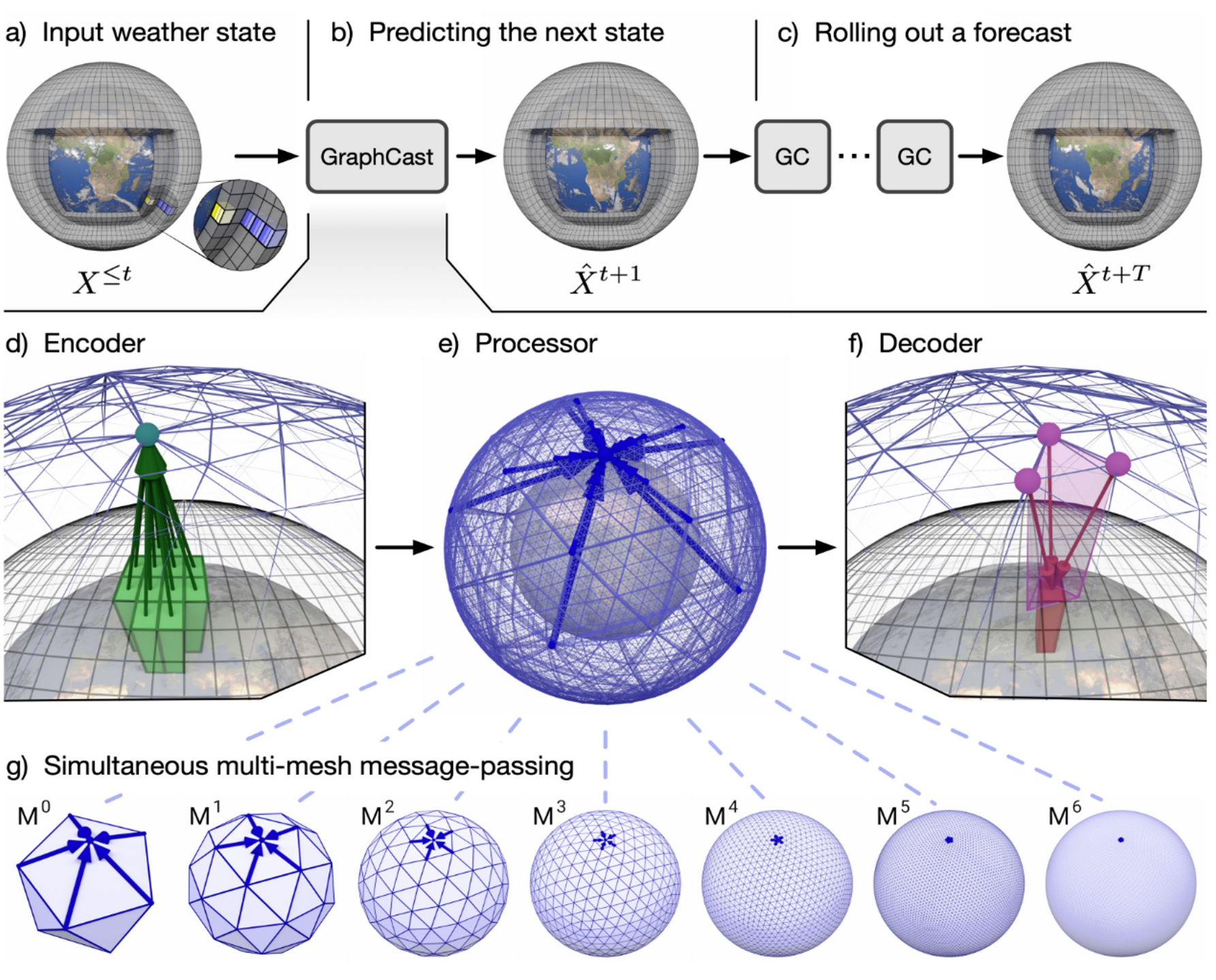
It is 05:45 UTC in mid-October, 2022, in Bologna, Italy, and the European Centre for Medium-Range Weather Forecasts (ECMWF)'s new High-Performance Computing Facility has just started operation. For the past several hours the Integrated Forecasting System (IFS) has been running sophisticated calculations to forecast Earth's weather over the next days and weeks, and its first predictions have just begun to be disseminated to users. This process repeats every six hours, every day, to supply the

2212.12794v2 [cs.LG] 4 Aug 2023

Économies de calcul de ces premiers modèles IA-PNT

Modèle	**Entraînement**	**Inférence**
Pangu-Weather	16 jours sur 192 GPU V100	Prévision à 24 heures 1xV100 1,4 seconde sur
GraphCast	21 jours sur 32 Cloud TPU V4 (~128 GPU)	Prévision à 10 jours 1 Cloud TPU V4 < 60 secondes
PNT (CEPMMT HRES) ***Résolution plus élevée**	-	Prévision à 10 jours 1/3/6 heures Grappe HPC de 11 664 cœurs

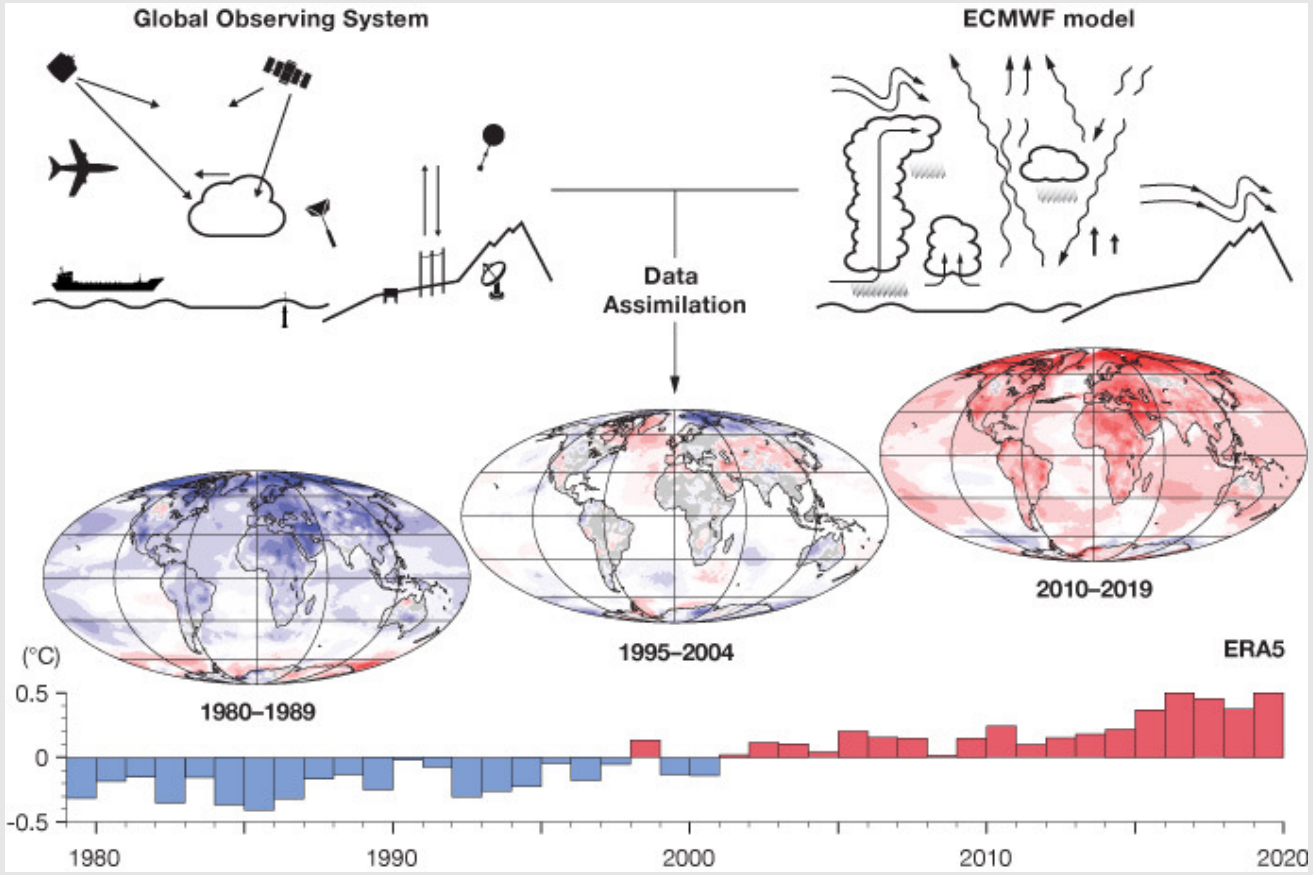
Architecture de GraphCast



doi:10.48550/arXiv.2212.12794

L'IA pour la prévision météorologique : un nouveau paradigme

Assimilation de données



Jeu de données ERA5

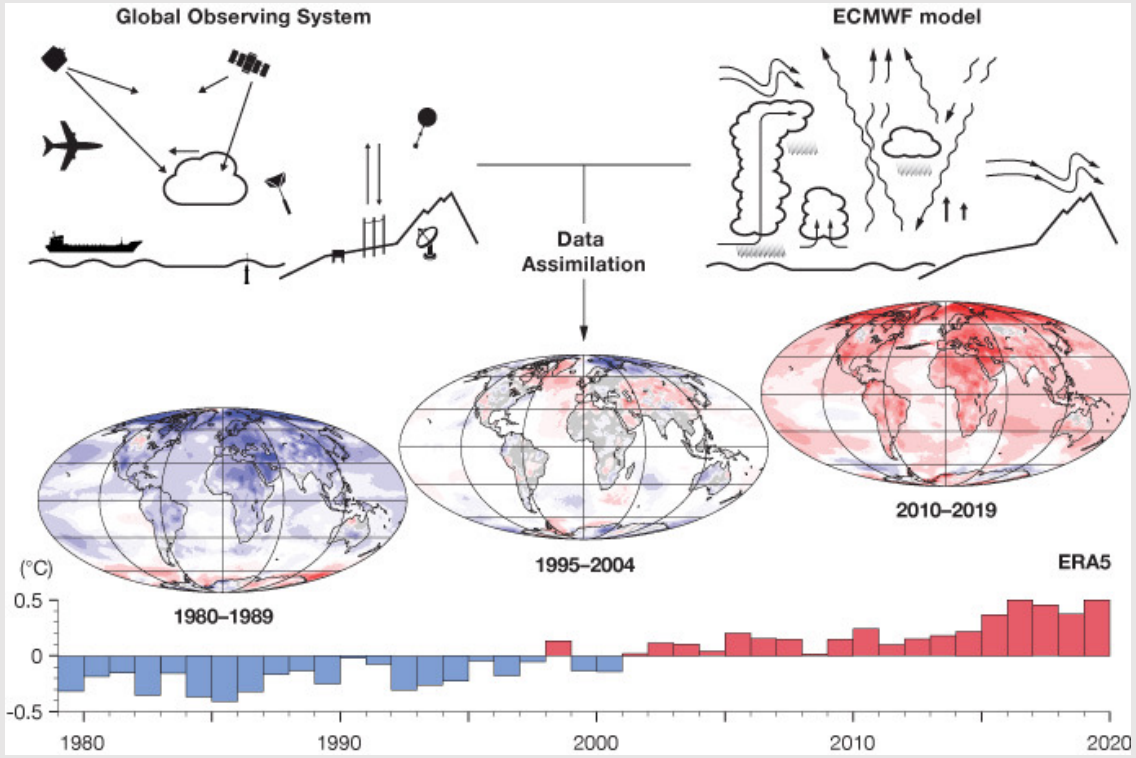
Simulation par réseau de neurones

P(**Prévision**)

Figure : Centre européen pour les prévisions météorologiques à moyen terme (CEPMMT)

La révolution de l'IA pour la prévision météorologique

Depuis 2022, divers modèles d'apprentissage profond ont atteint des performances de prévision comparables, voire **MEILLEURES**, que la prévision numérique du temps (PNT), l'état de l'art précédent.



Données d'entraînement : ERA5, un jeu de d
produit par assimilation de données

Tâche d'entraînement : autorégression : prév
à 6-24 heures

- Déroulée pour prévoir à 7-10 jours

[Site web du Centre européen pour les prévisions météorologiques à moyen terme (CEPMMT)]

Déroulés

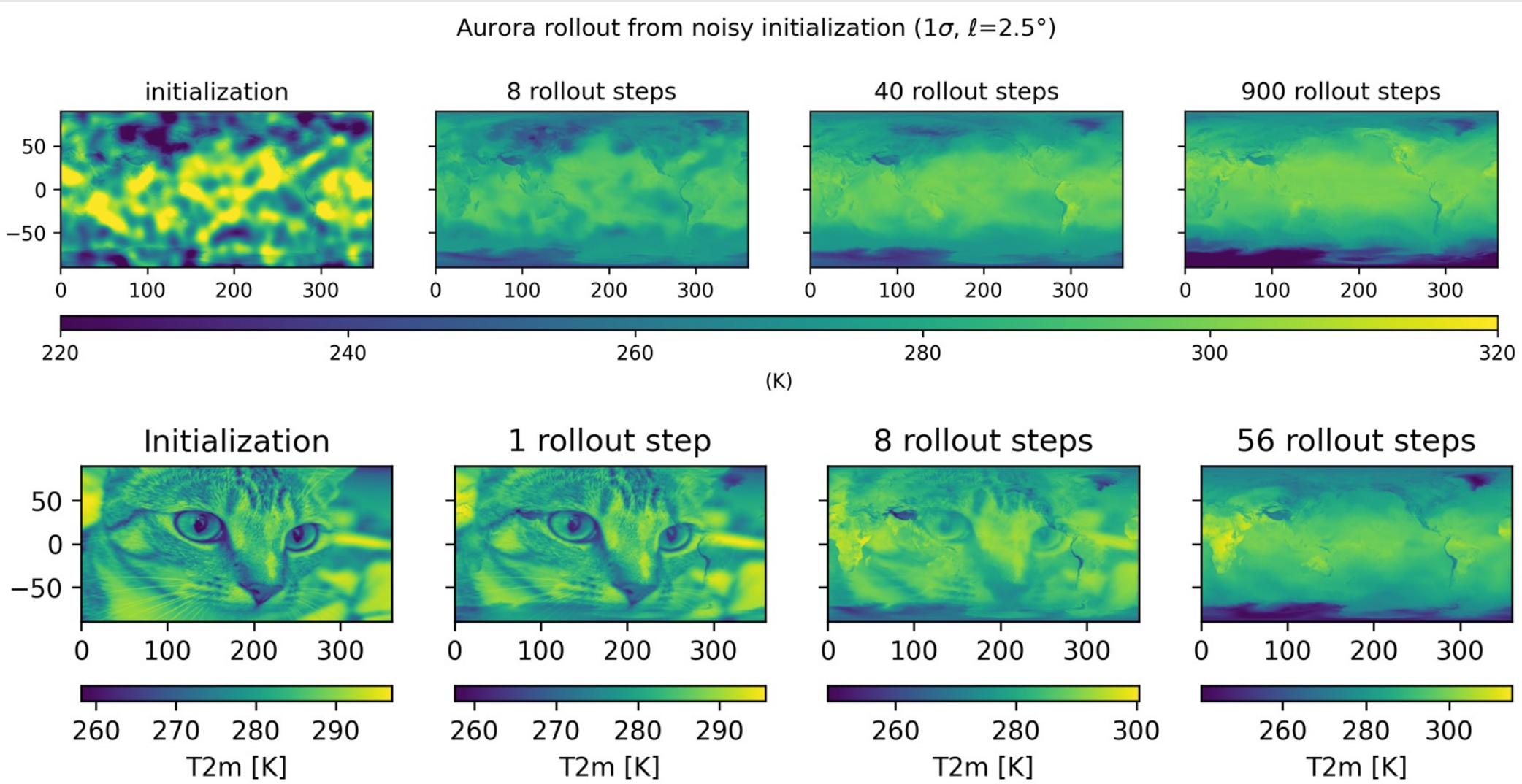


Figure A.9: 2-meter temperature predicted at different time stamps when (top) inputs are perturbed with an additive Gaussian random field with coefficient of variation σ^2 (where σ is the standard deviation of the physical variable) and correlation length of 10 pixels (equivalent to $2.5°$), (bottom) surface and atmospheric variables are initialized with a cat picture.

Lehmann et al., (2026) ”Can AI Weather Models Predict Beyond Two Weeks?” arXiv:2605.30184

L'IA pour la prévision météorologique : un nouveau paradigme

Time

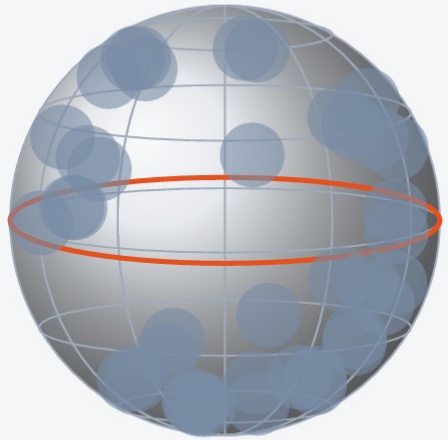
45+ years

1979 — 2024

Hourly

~400,000 timesteps

Space



721 × 1440 × 240

0.25° resolution

37 levels × 6 vars + 18 surface

Memory

per atmospheric state:

0.25° **0.94 GB**

×6.25

0.1° **5.9 GB**

Jeu de réanalyse ERA5 du CEPMMT ≈ **13 Petabytes**

ArchesWeatherGen: Générateur météo d'IA frugale [Couairon et al., *Science Advances*, 2026]

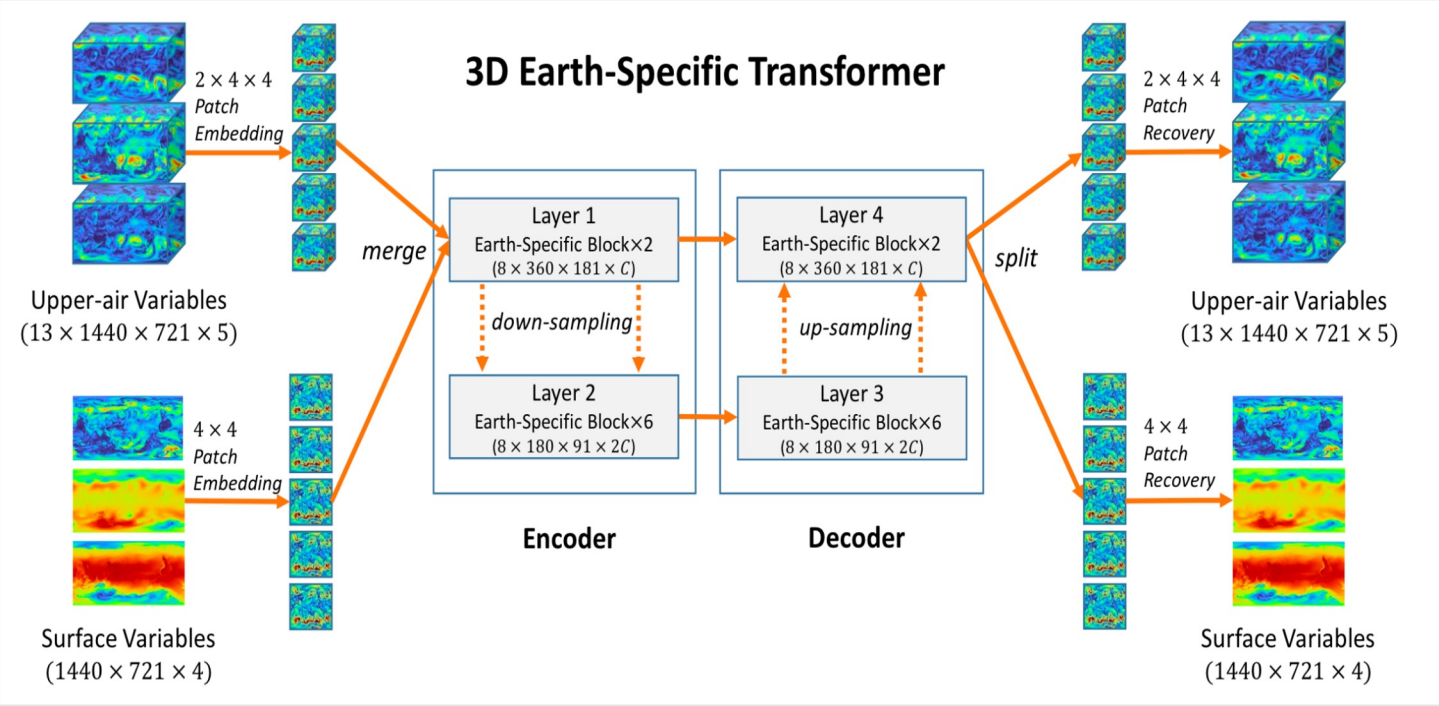
Figure : Emmanuel de Bézenac (INRIA)



ArchesWeather

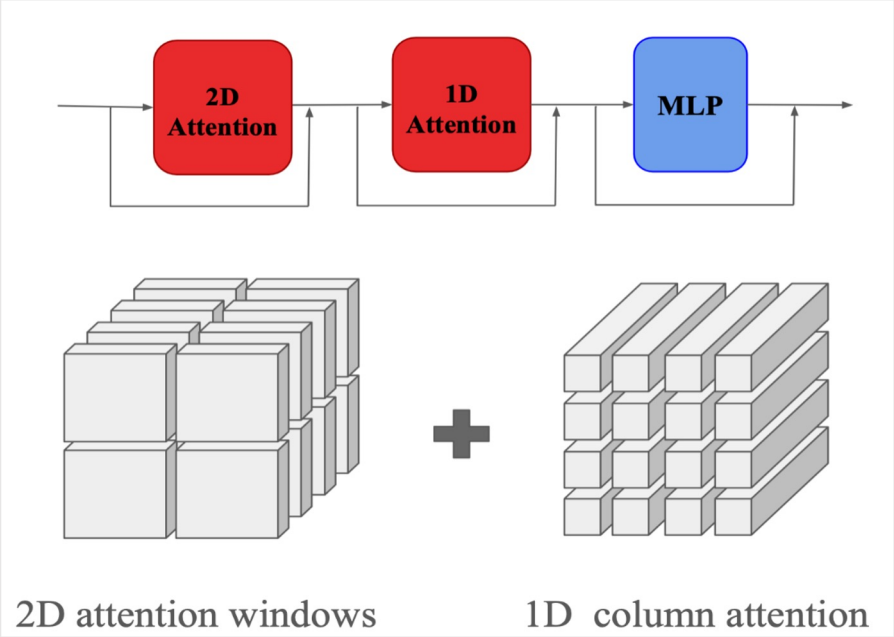
[Couairon et al., ArchesWeather: an efficient AI weather model at 1.5º resolution ICML 2024 workshop on Earth System Modeling]

Pangu-Weather: A 3D High-Resolution System for Fast and Accurate Global Weather Forecast [Bi et al., *Nature,* 2023]



Nouveautés d'ArchesWeather :

- Entraînement à une résolution de données plus grossière
- Remplacement de l'attention 3D par :

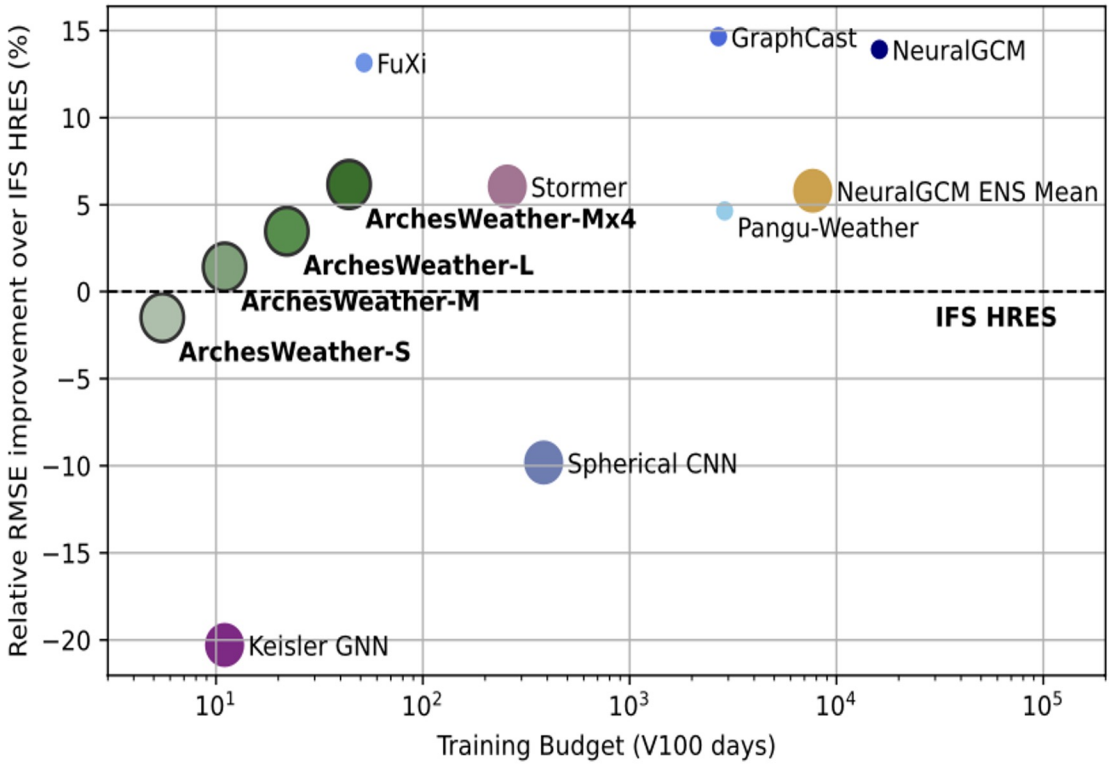


latitude + longitude

altitude

ArchesWeather

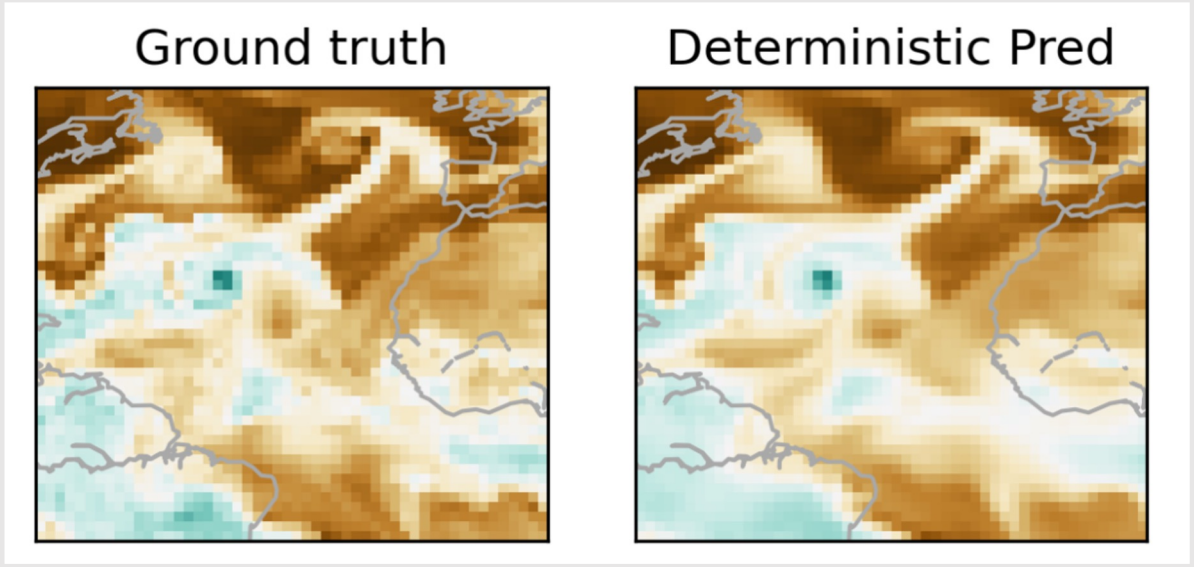
Compromis de performance



	Res.	Cost	Z500 m^2/s^2	T850 K	Q700 g/kg	U850 m/s	V850 m/s	T2m K	SP Pa	U10m m/s	V10m m/s
IFS HRES			42.30	0.625	0.556	1.186	1.206	0.513	60.16	0.833	0.872
Pangu-Weather	0.25°	2880	44.31	0.620	0.538	1.166	1.191	0.570	55.14	0.728	0.759
NeuralGCM	0.25°	16128	37.94	0.547	0.488	1.050	1.071	N/A	N/A	N/A	N/A
FuXi	0.25°	52	40.08	0.548	N/A	1.034	1.055	0.532	49.23	0.660	0.688
GraphCast	0.25°	2688	39.78	0.519	0.474	1.000	1.02	0.511	48.72	0.655	0.683
Keisler	1°	11	66.87	0.816	0.658	1.584	1.626	N/A	N/A	N/A	N/A
SphericalCNN	1.4°	384	54.43	0.738	0.591	1.439	1.471	N/A	N/A	N/A	N/A
Stormer	1.4°	256	45.12	**0.607**	0.527	**1.138**	**1.156**	0.570	**53.77**	**0.726**	**0.760**
NeuralGCM ENS (50)	1.4°	7680	**43.99**	0.658	0.540	1.239	1.256	N/A	N/A	N/A	N/A
ArchesWeather-S	1.5°	9	48.92	0.650	0.542	1.289	1.325	0.562	60.30	0.833	0.873
ArchesWeather-M	1.5°	18	44.96	0.615	**0.527**	1.219	1.251	**0.539**	56.22	0.783	0.819
ArchesWeather-Mx4	1.5°	36	**41.93**	**0.593**	**0.513**	**1.172**	**1.203**	**0.517**	**52.22**	**0.749**	**0.783**

RMSE, échéance 24 heures

Pourquoi passer à la génération d'ensembles…



Prévision Q700 à 24 h d'échéance
date d'initialisation : 28 sept. 2019

- La prédiction déterministe montre un lissage irréaliste
- Essayer la génération d'ensembles par entraînement de diffusion
- Objectif : chaque membre échantillonné doit être plus physique

ArchesWeatherGen

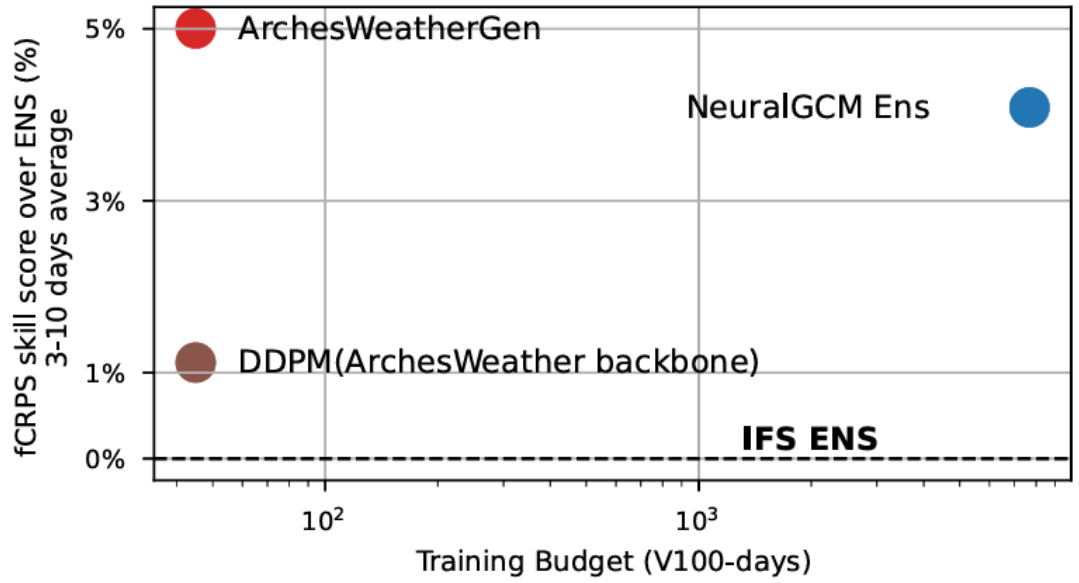
[Couairon et al., *Science Advances*, 2026]

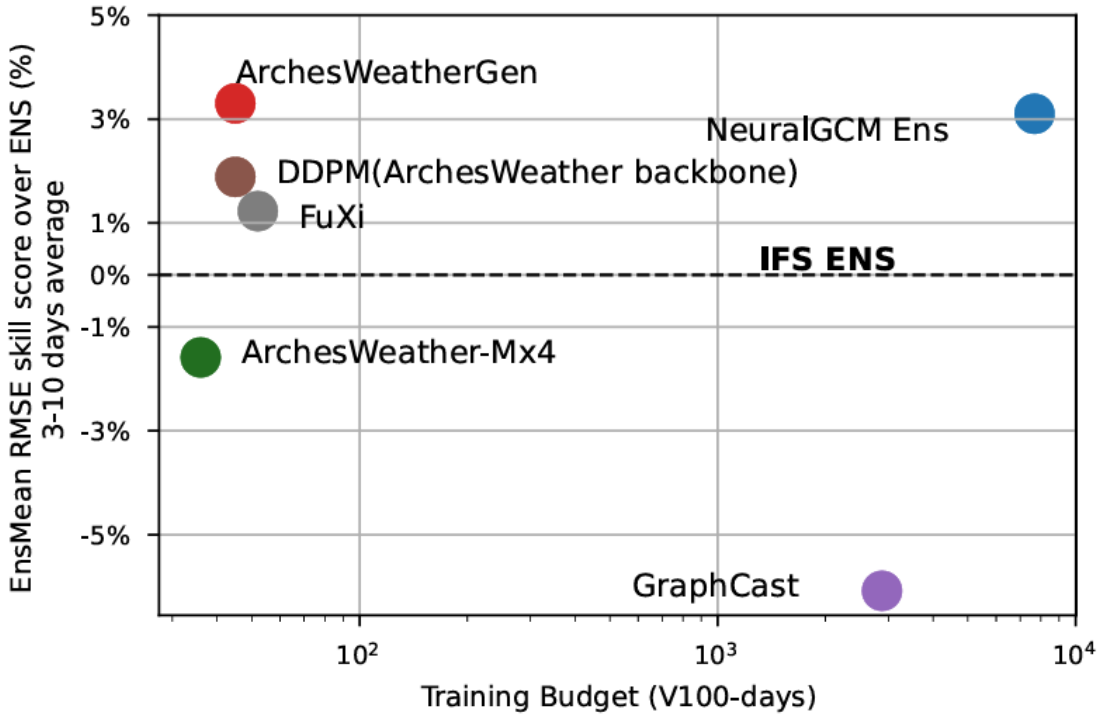
Ossature : ArchesWeather (variante du Swin U-Transformer), 44 M de paramètres

Entraîné par Flow Matching

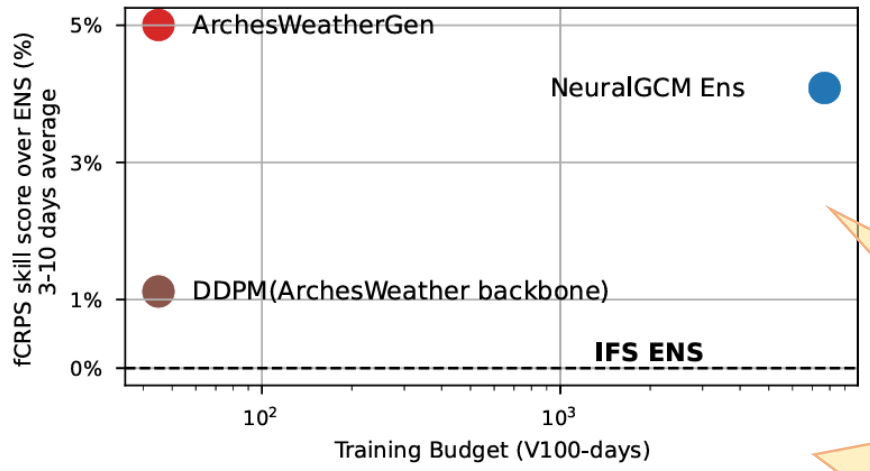
Inférence réalisée avec 25 pas d'échantillonnage par 24 heures

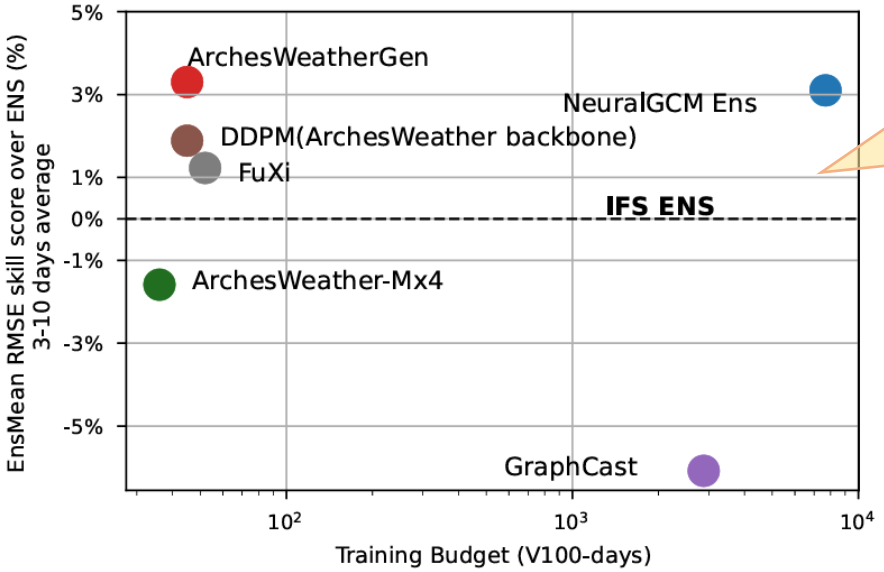
Déroulements réalisés avec un échantillon toutes les 24 heures. Cette sortie est réinjectée dans le modèle entraîné, et ainsi de suite.

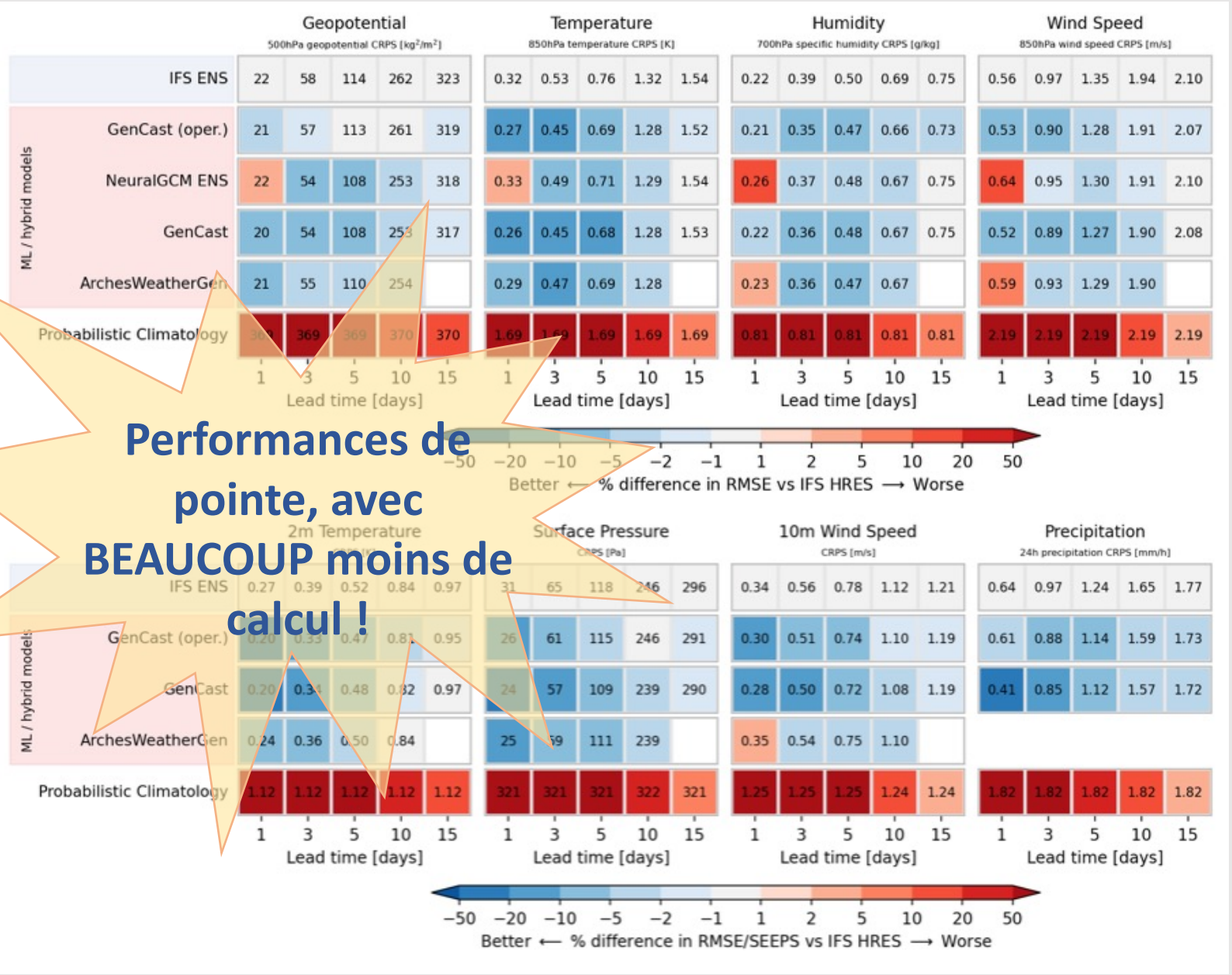




ArchesWeatherGen





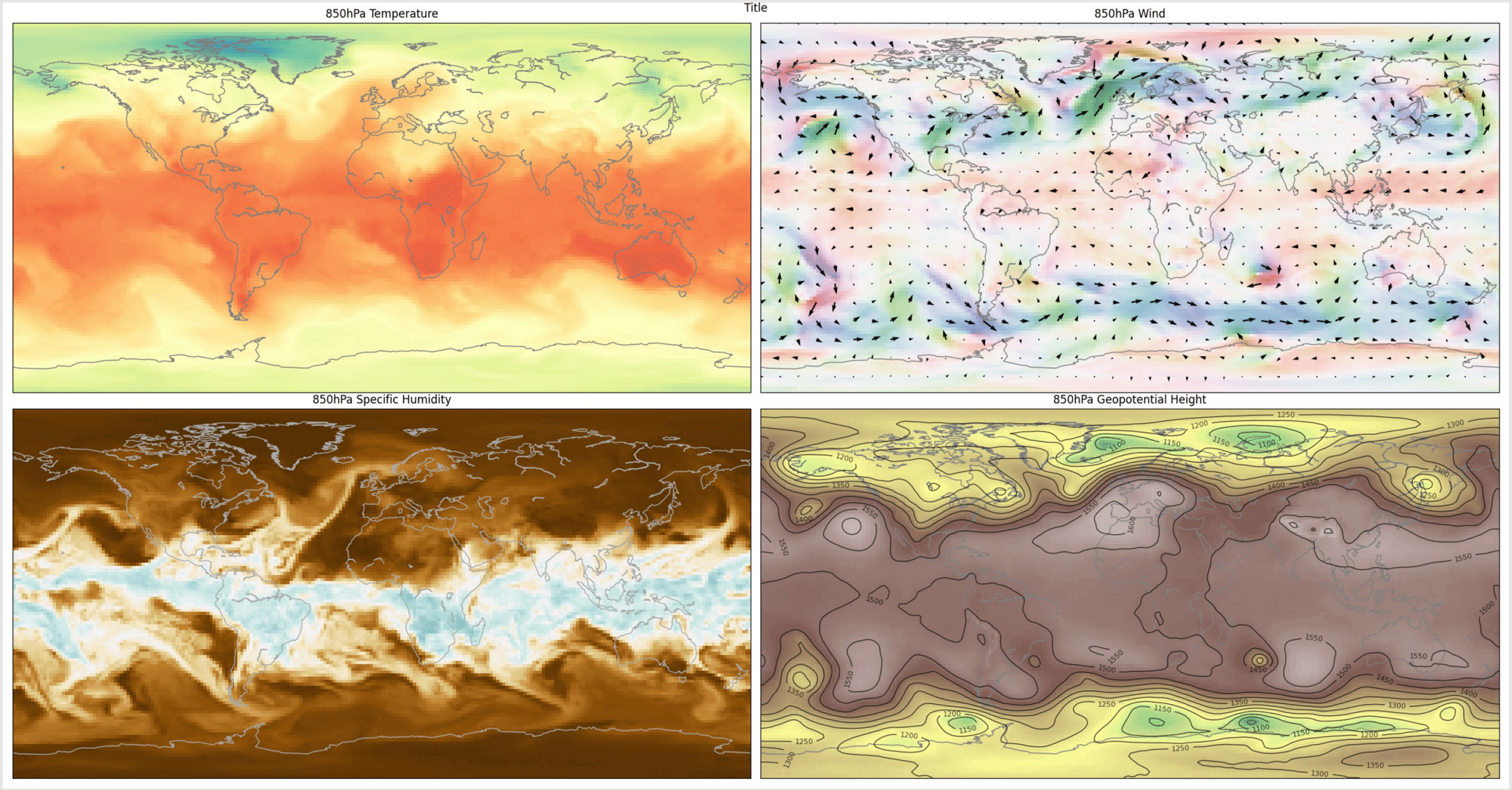


Performances de pointe, avec BEAUCOUP moins de calcul !

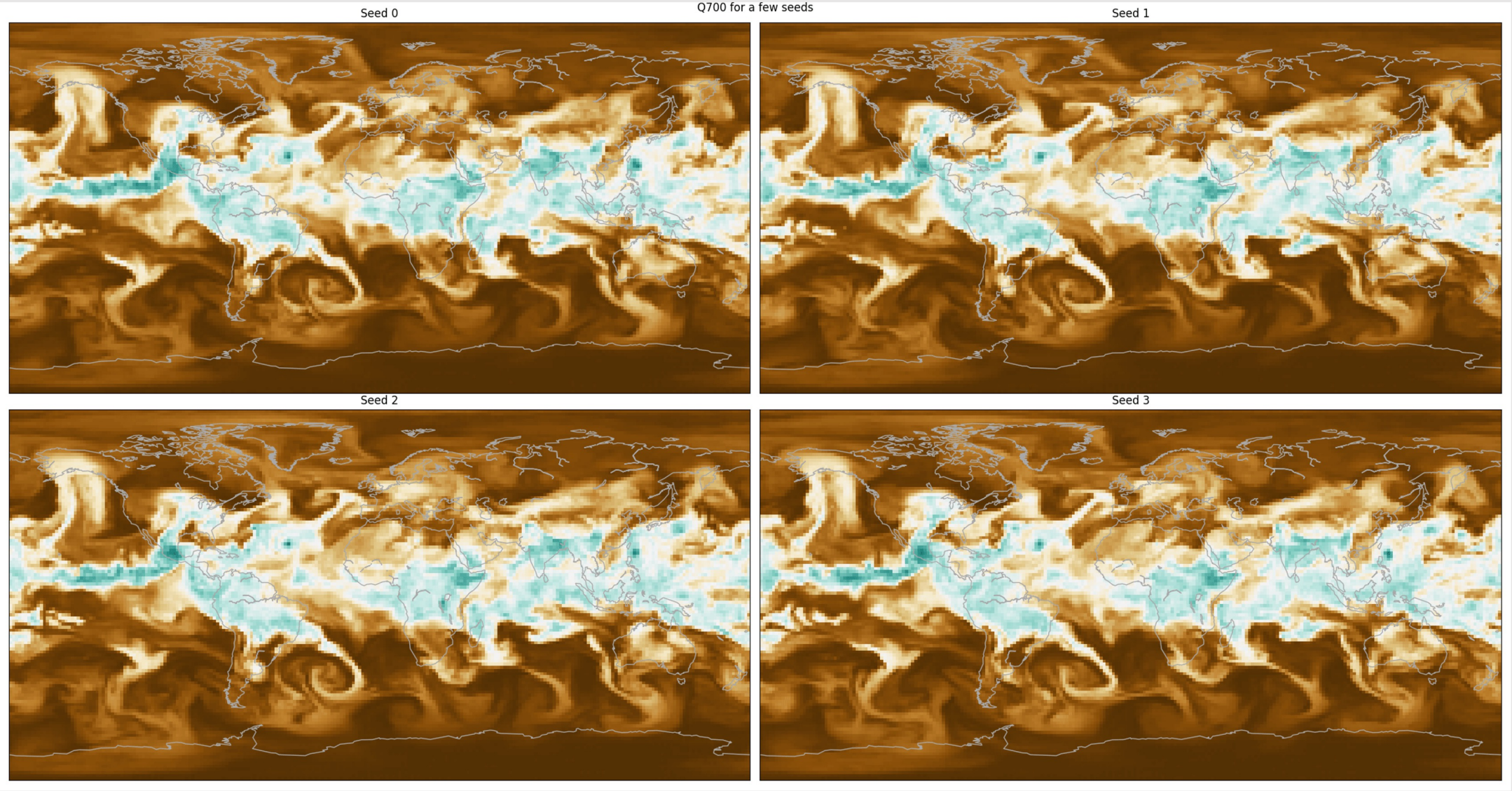
[Couairon, Singh, Charantonis, Lessig, Monteleoni, *Science Advances*, 2026]

Tableaux de scores sur WeatherBench 2 de Google

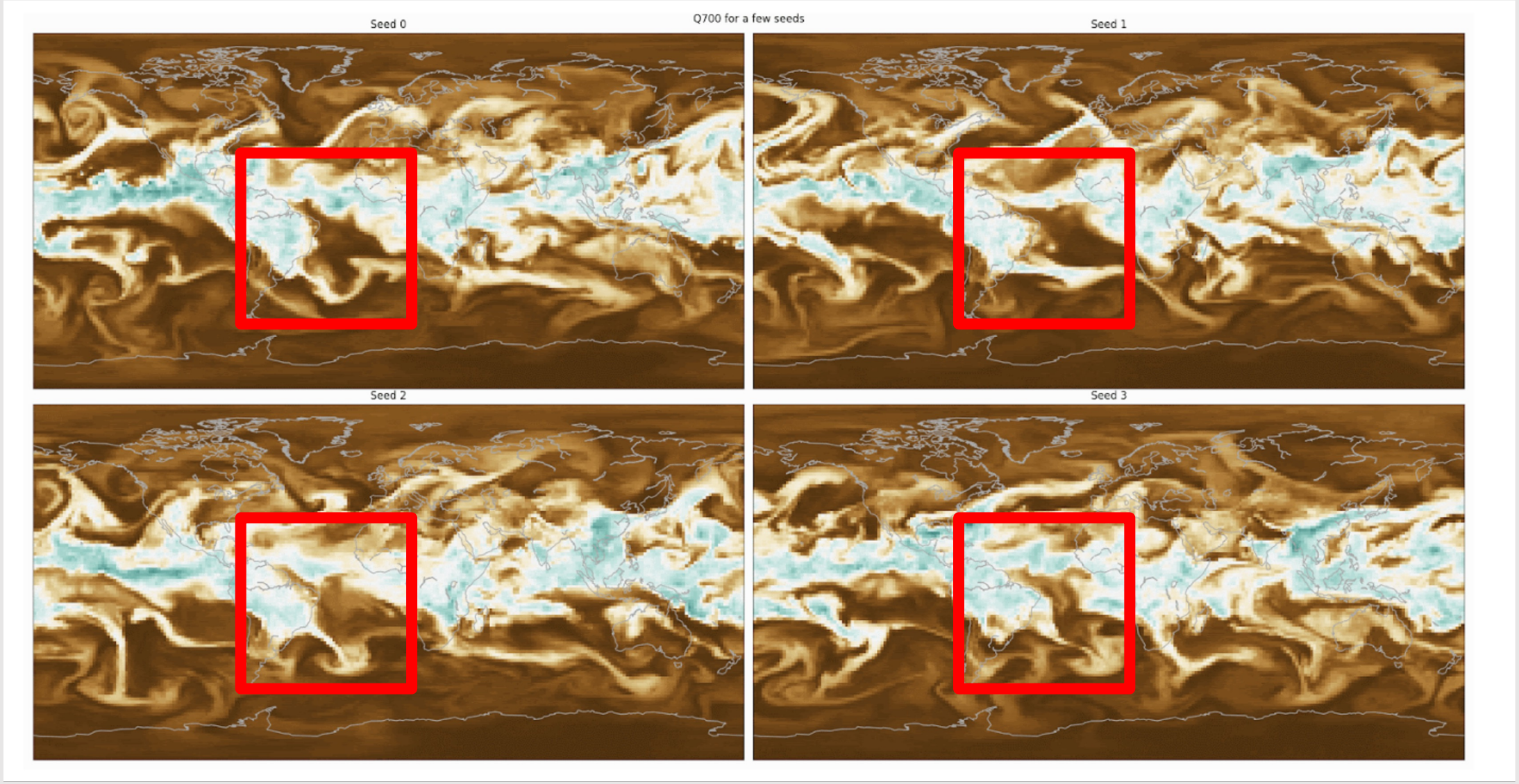
Un exemple de trajectoire



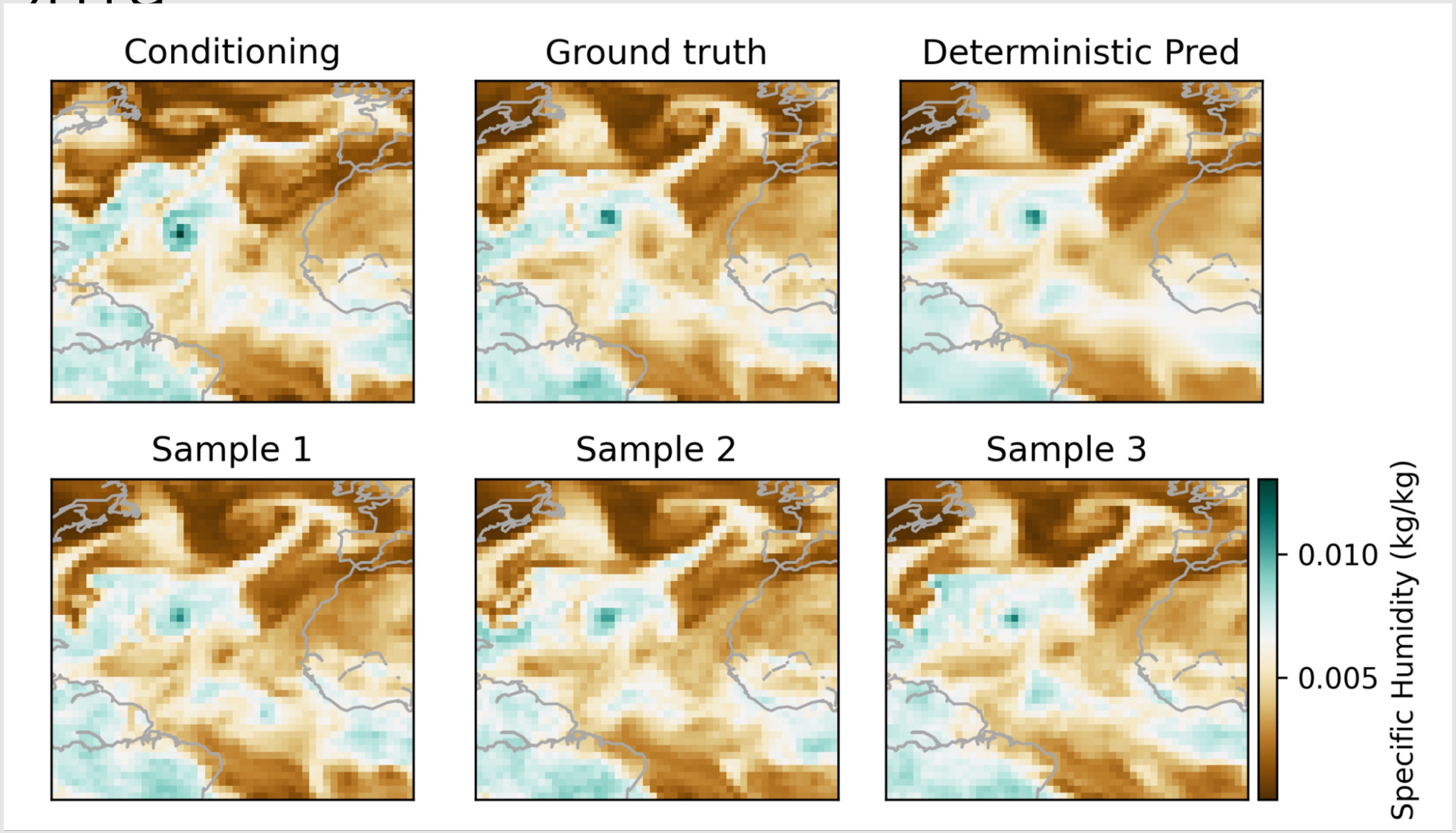
Différentes graines - Q700



Diversité des membres de l’ensemble ; échéance 10 jours



Étude de cas : Lorenzo - Q700 - initialisation le 25 sept. 2019



ArchesClimate : émulation de modèles climatiques par IA



Thèse en cours de Graham Clyne
Directeur principal : Anastase Charantonis

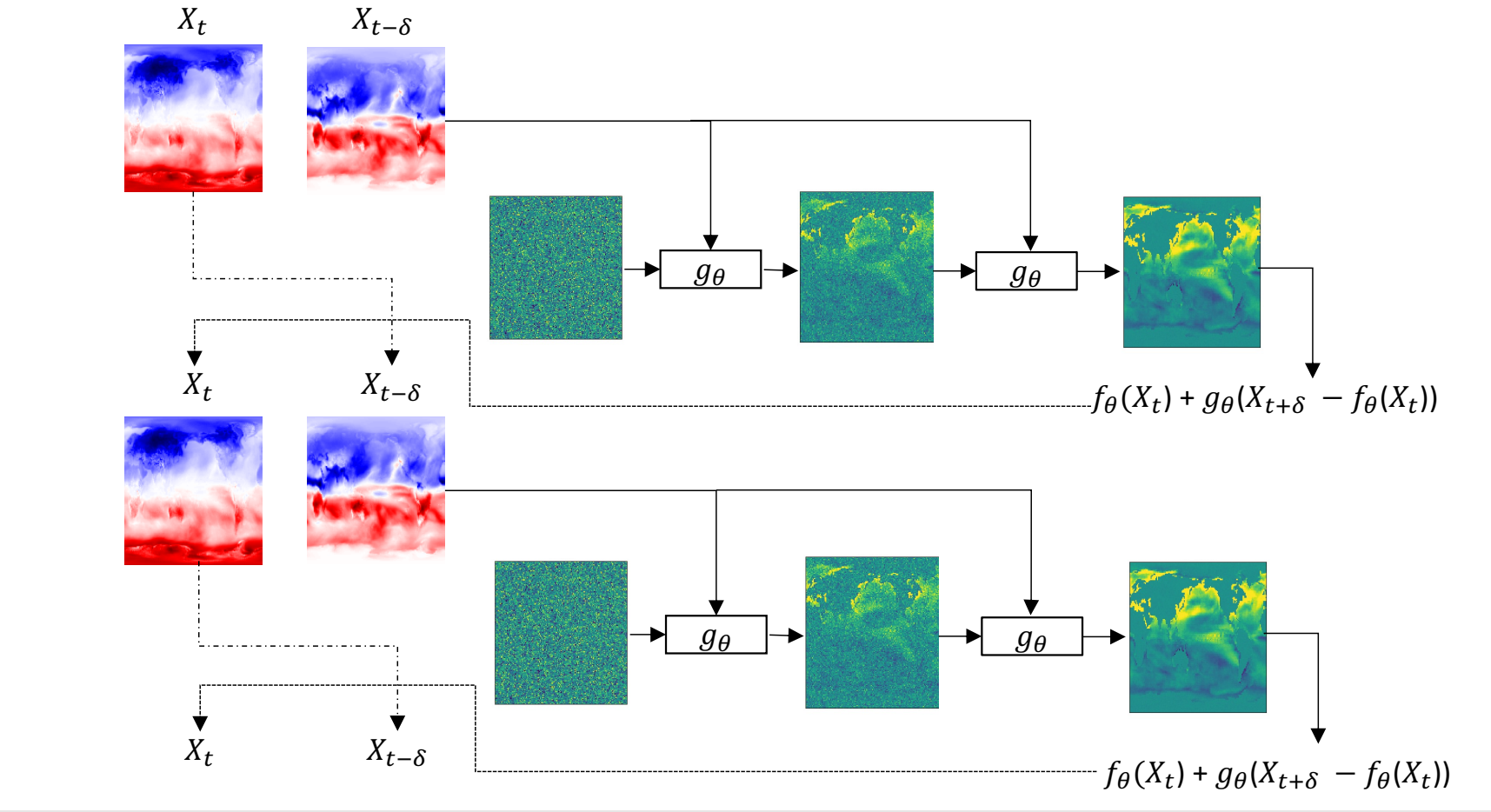
Chaîne : ArchesWeatherGen modifié :

- Pas d'attention 1D par colonne
- Inclusion des forçages

Données d'entraînement : simulations de modèles climatiques (IPSL DCPP)

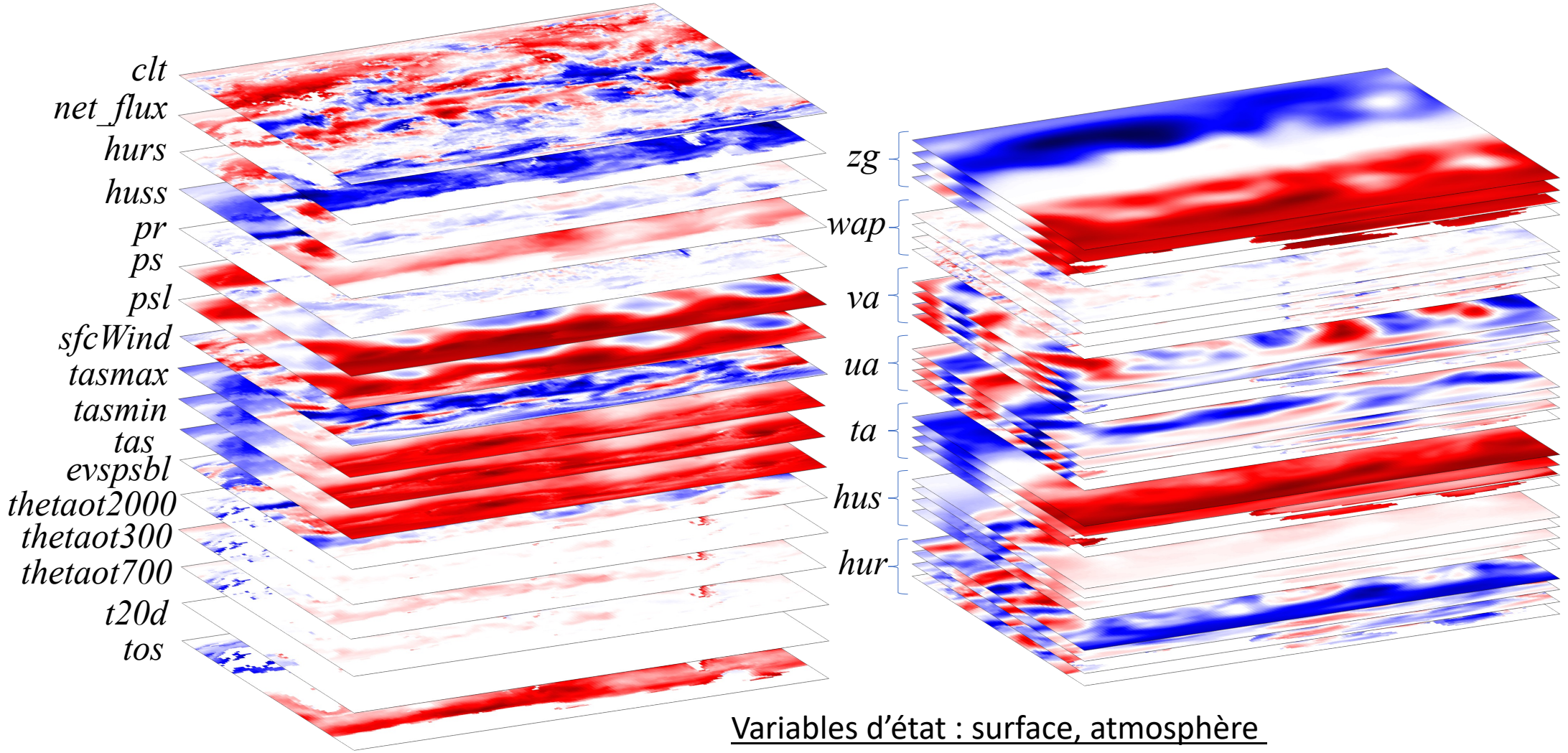
Entraînement : prédire un mois à l'avance

Déroulements réalisés avec un échantillon
par mois. Cette sortie est réinjectée dans le modèle entraîné, et ainsi de suite.



Clyne et al., *résultats préliminaires présentés à l'EGU 2025 et à Climate Informatics 2025*

Variables d’entrée et forçages

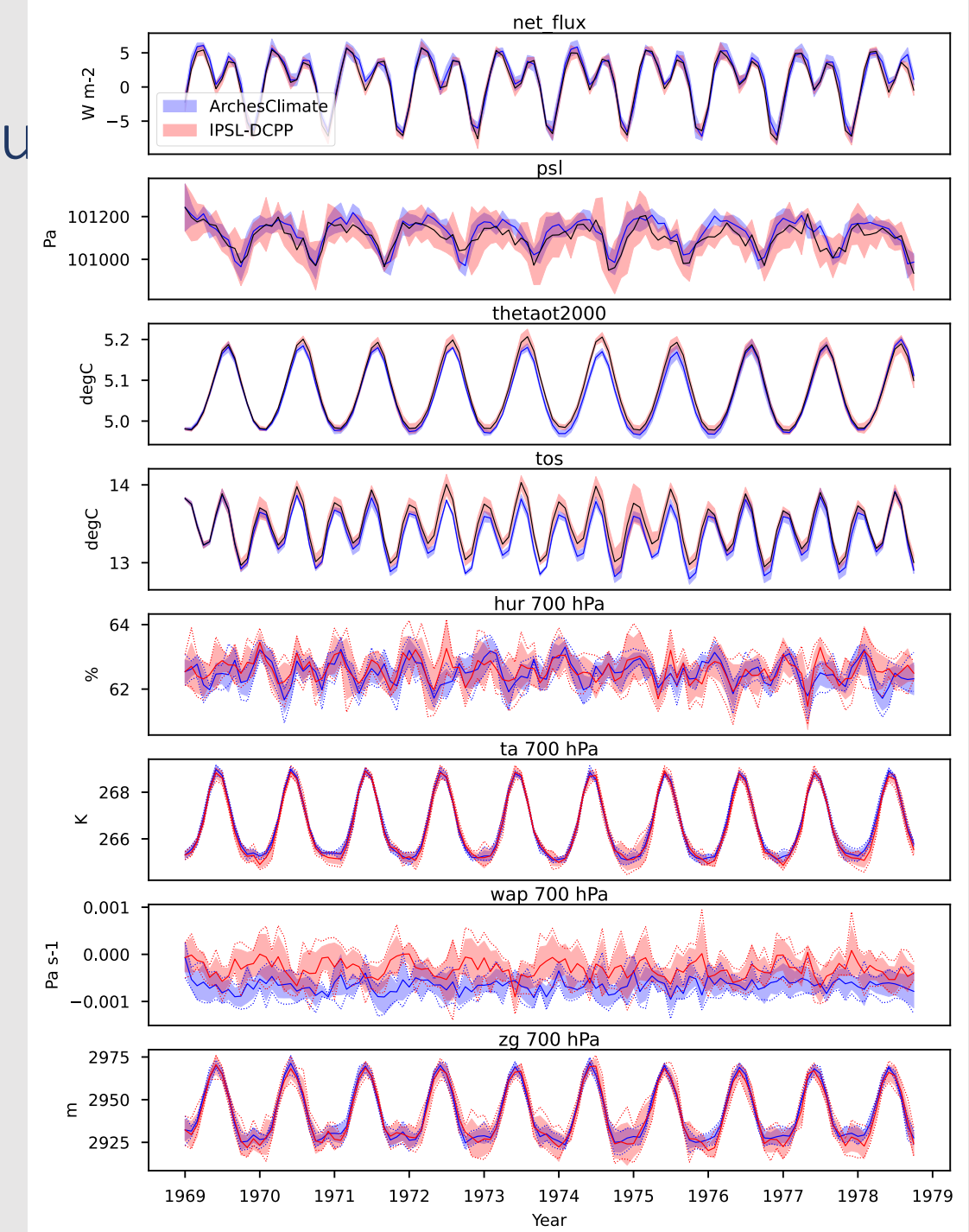


$$\begin{bmatrix} CO2 \\ CH4 \\ N2O \\ CFC11eq \\ SSI \end{bmatrix}$$

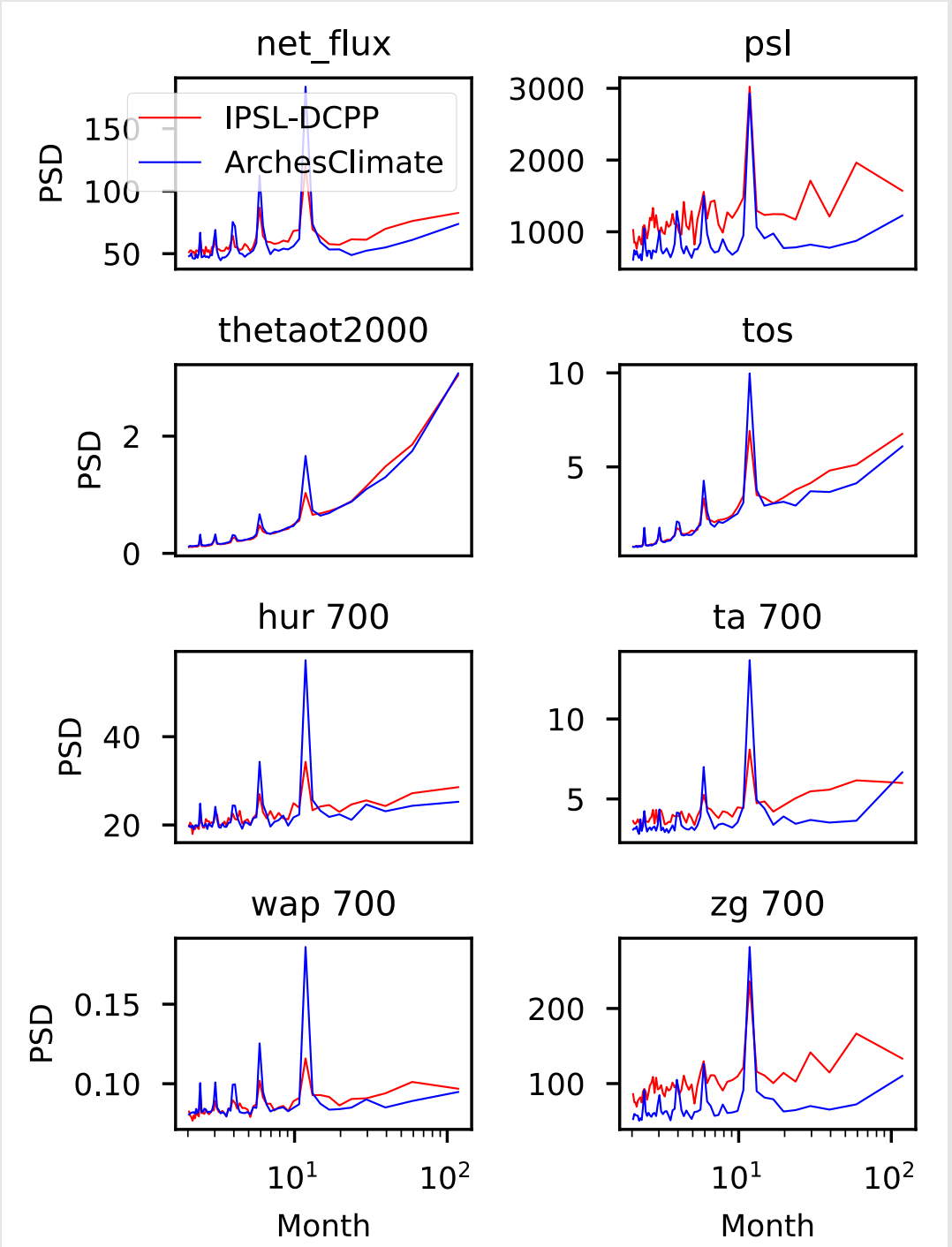
Variables d’état : surface, atmosphère

Forçages

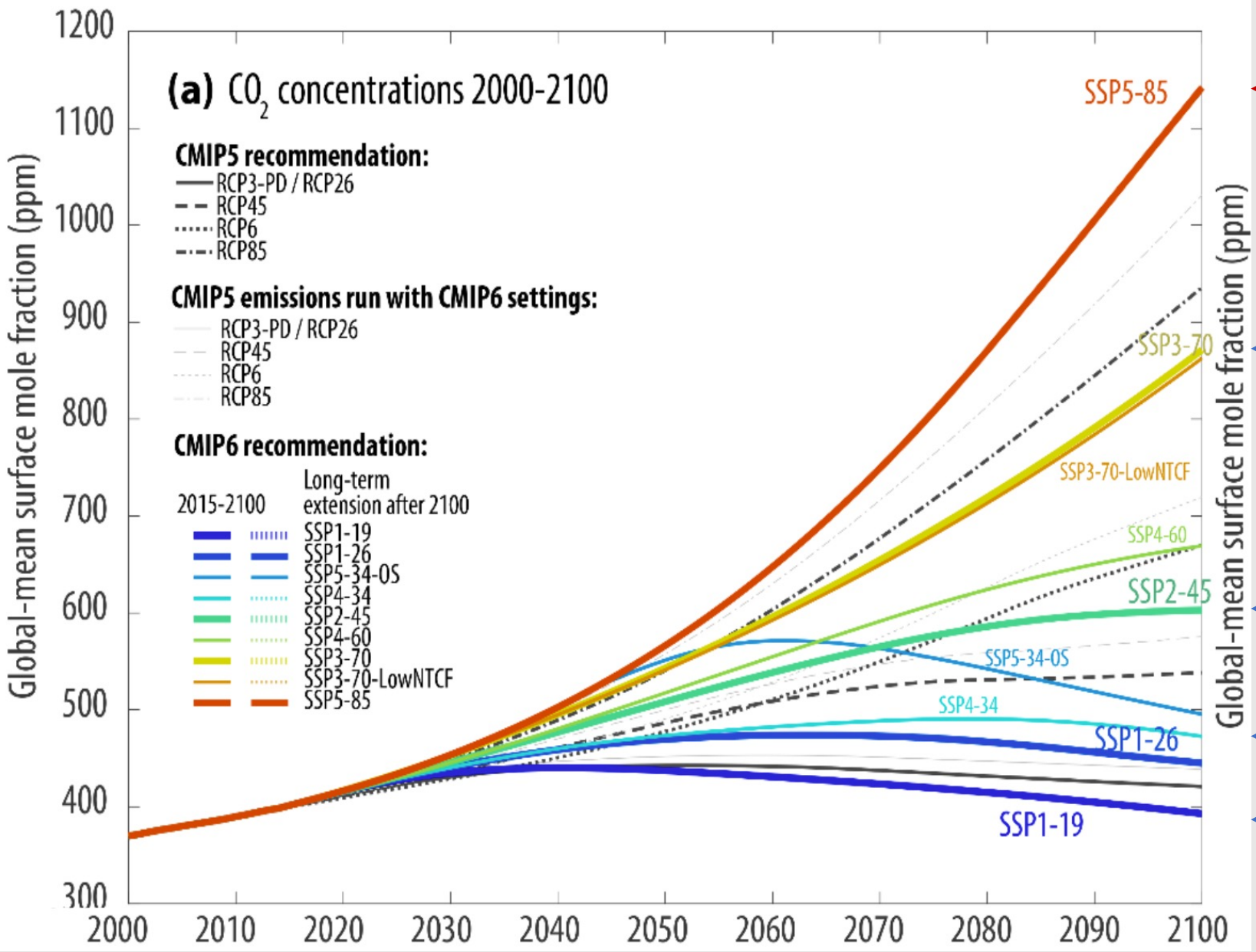
Déroulements décennau



Résultats préliminaires



Extrapoler les scénarios de modèles climatiques



Données de test

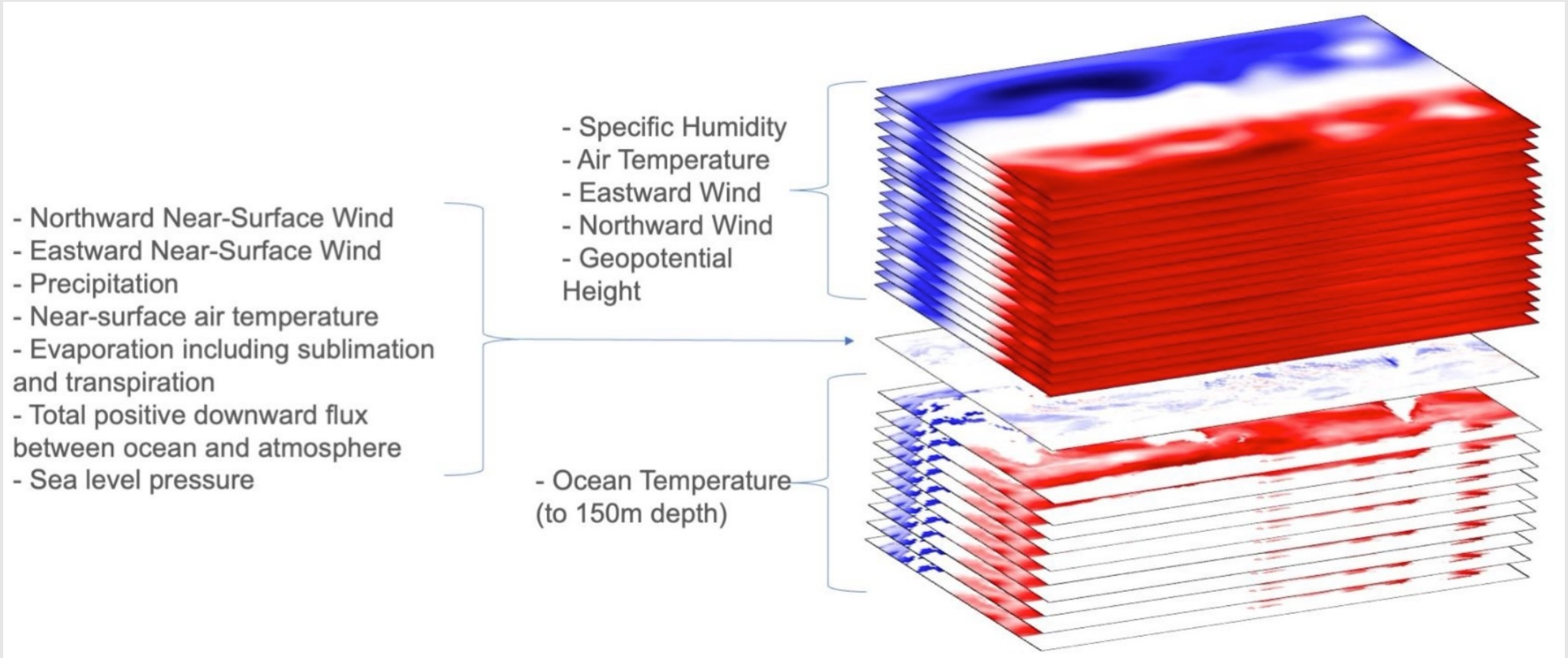
Tâche d'extrapolation

Données d'entraînement

Ces 4 scénarios

Interpolation et extrapolation dans l'espace des scénarios de modèles climatiques

Clyne et al., *en préparation*



Extrapolation de scénarios

Reproduire les schémas temporels d'El Niño-Oscillation australe

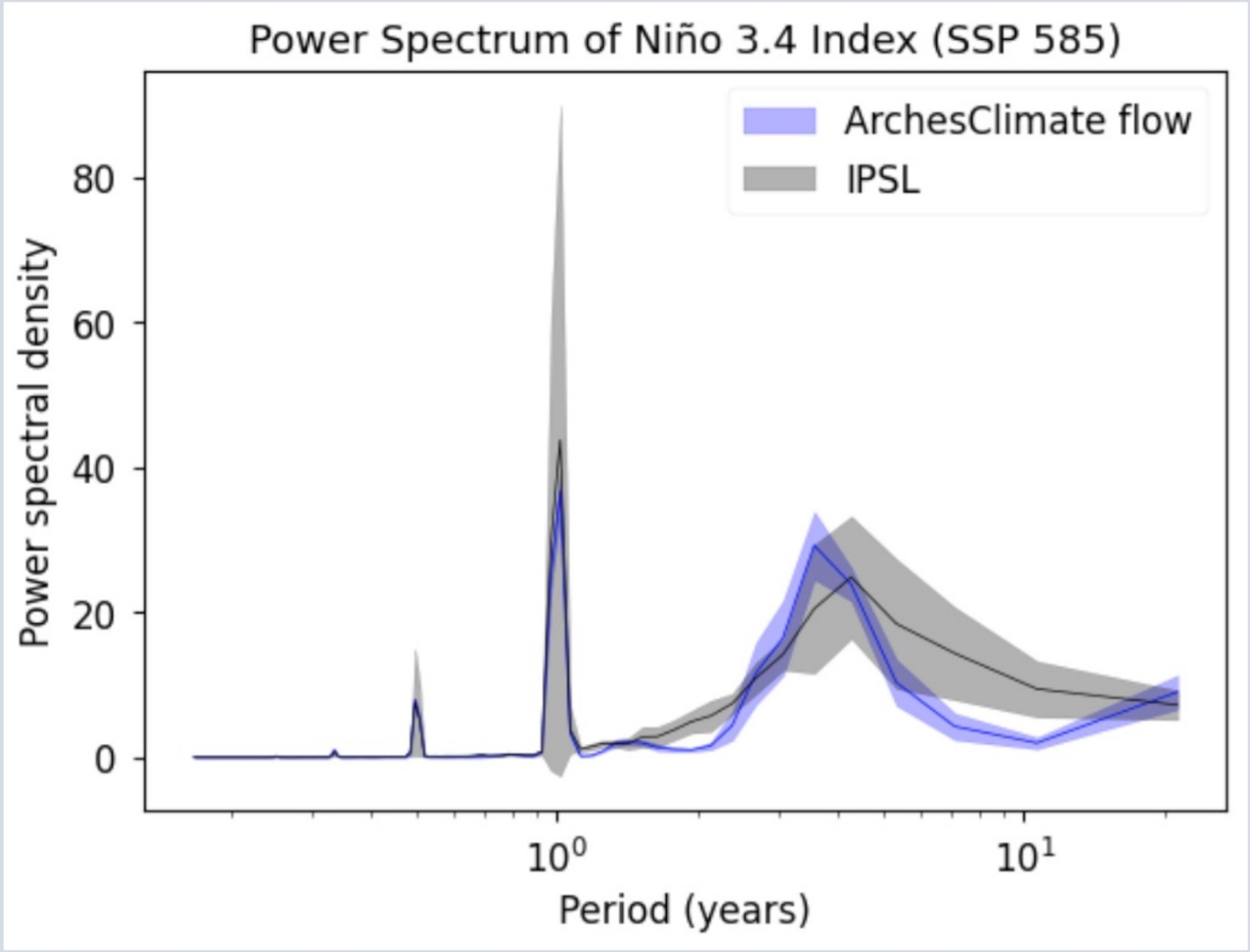


Figure 9. Power spectra for El Niño 3.4 index for SSP5-8.5.

Clyne et al., *en préparation*

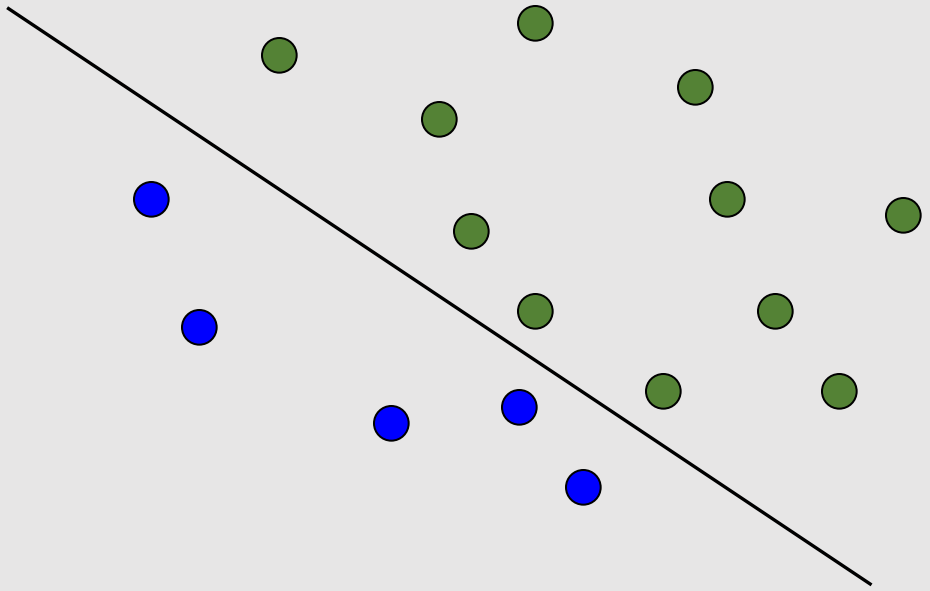
Partie II : Méthodes d'IA : l'apprentissage auto-supervisé

Apprentissage supervisé

À partir de points de données étiquetés, trouver une bonne règle de classification.

Décrit les données

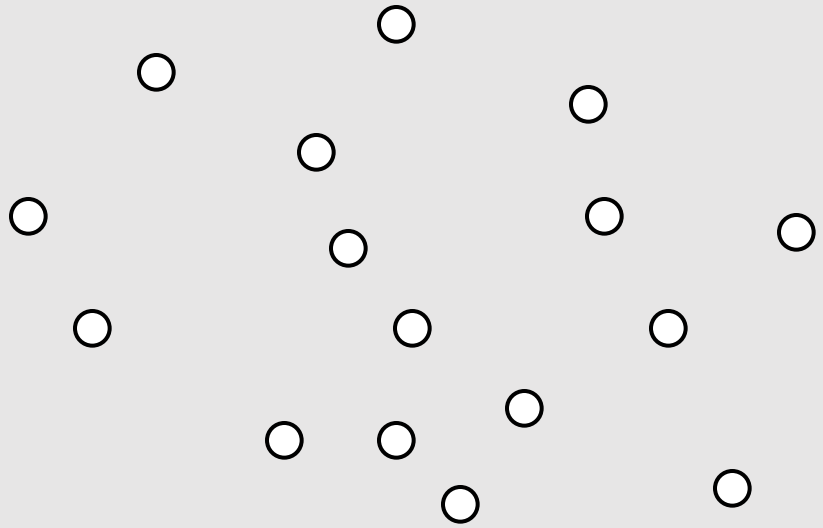
Généralise bien



Apprentissage non supervisé

Que peut-on faire sans aucune étiquette ?

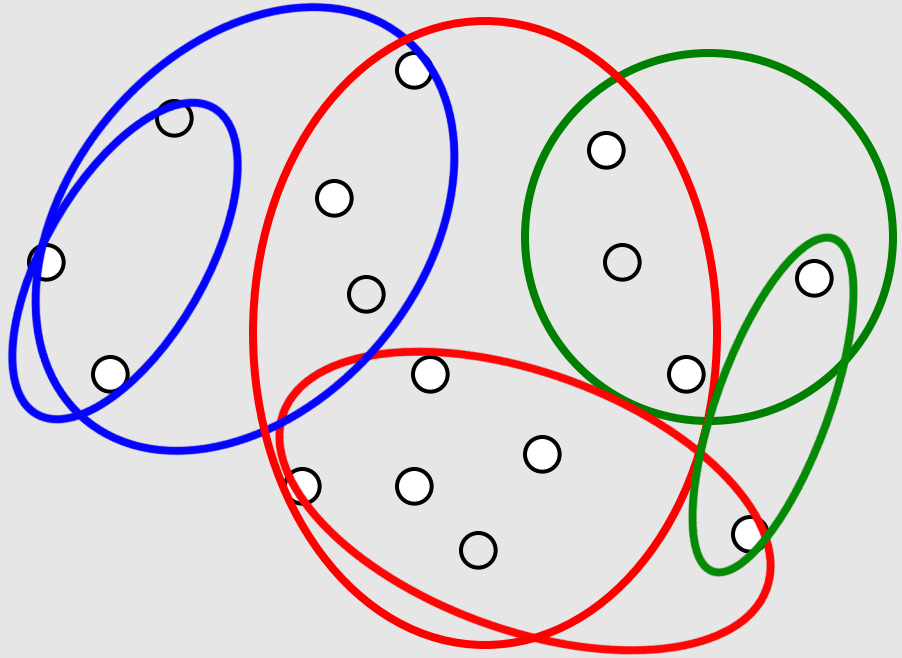
Apprentissage non supervisé



Apprentissage non supervisé

Que peut-on faire sans aucune étiquette ?

Apprentissage non supervisé



Exemple : le partitionnement (clustering)

Apprendre à partir de données « non étiquetées »

- Tâches d'apprentissage non supervisé
  1. Partitionnement
  2. Plongement (réduction de dimension, y compris par méthodes non linéaires)
  3. Interpolation
  4. Apprentissage d'un modèle génératif

- Il existe des approches d'apprentissage profond (AP) pour toutes ces tâches

- Il existe des approches hors AP pour toutes ces tâches

- Il existe des approches d'AP auto-supervisées pour toutes ces tâches

Apprentissage profond non supervisé

- AP supervisé. La perte de prédiction dépend de l'étiquette y et de la sortie du réseau pour l'entrée x.

Sortie du réseau

$$f_W(x) = \hat{y}$$

Fonction de perte

$$\mathcal{L}(\hat{y}, y)$$

- AP non supervisé. La perte de prédiction ne dépend que de x et de la sortie du réseau pour l'entrée x. Il n'y a pas d'étiquette y.

Sortie du réseau

$$f_W(x) = \hat{x}$$

Fonction de perte

$$\mathcal{L}(\hat{x}, x)$$

L’approche auto-supervisée de l’apprentissage non supervisé

Apprentissage auto-supervisé

Une approche de pointe de l’apprentissage (profond) non supervisé

Concevoir une tâche prétexte :

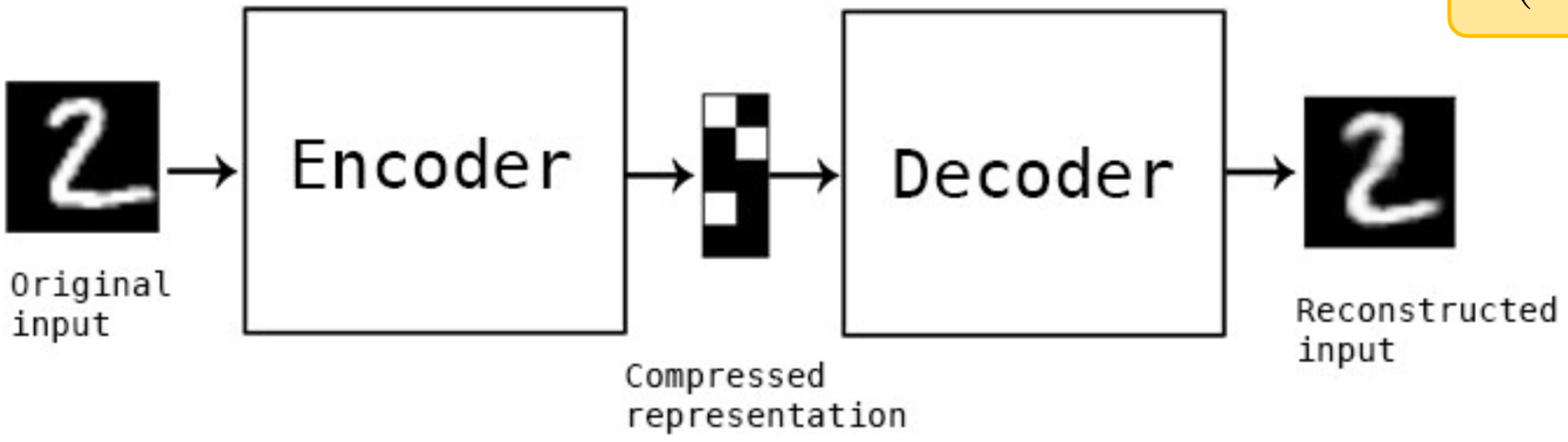
- Concevoir une tâche d’apprentissage supervisé à partir des seules données disponibles.
- Entraîner un modèle sur cette tâche de sorte que
- les caractéristiques apprises (ou la loi a posteriori apprise sur un espace de caractéristiques) soient utiles pour une autre tâche (en

Tâche prétexte : exemple

Exemple classique de tâche prétexte : l'autoencodeur

- Entraîner un réseau de neurones de façon non supervisée
  - Utiliser les données non étiquetées à la fois comme entrée et pour évaluer la sortie
- Après l'entraînement, la couche goulot sera une représentation compacte de la distribution d'entrée

$\mathcal{L}(\hat{x}, x)$



Apprentissage auto-supervisé

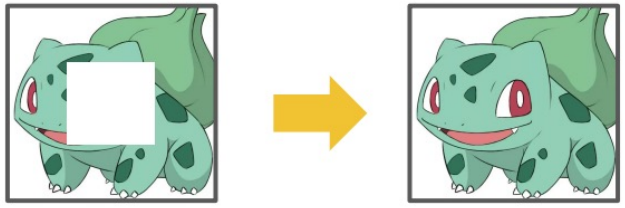
- Concevoir une tâche prétexte qui fera apprendre des caractéristiques utiles.

Vision Pretext Tasks: Masked Prediction

- **Denoising autoencoder** (Vincent et al. 2008)
  - Add noise = Randomly mask some pixels
  - Only reconstruction loss



- **Context autoencoder** (Pathak et al. 2016)
  - Mask a random region in the image
  - Reconstruction loss + adversarial loss



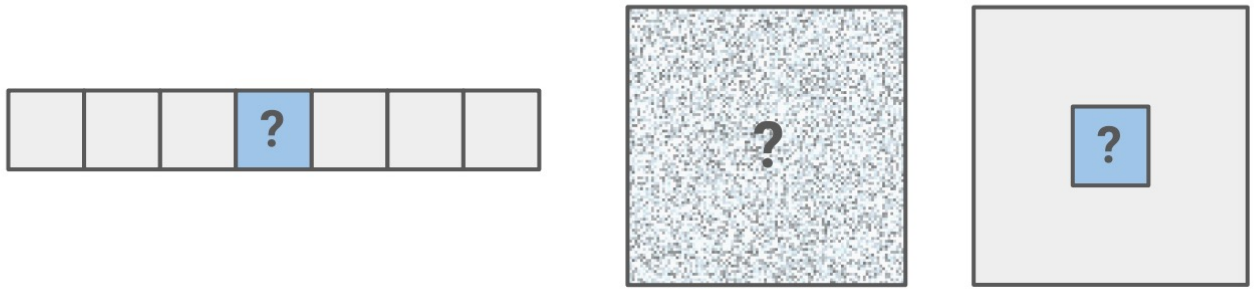
Lilian Weng, Jong Wook Kim, tutoriel NeurIPS 2021

Apprentissage auto-supervisé

- Concevoir une tâche prétexte qui fera apprendre des caractéristiques utiles.

Self-Prediction: Masked Generation

We mask a random portion of information and pretend it is missing, irrespective of the natural sequence. The model learns to predict the missing portion given other unmasked information.



Examples:

- Masked language modeling (BERT)
- Images with masked patch (denoising autoencoder, context autoencoder, colorization)

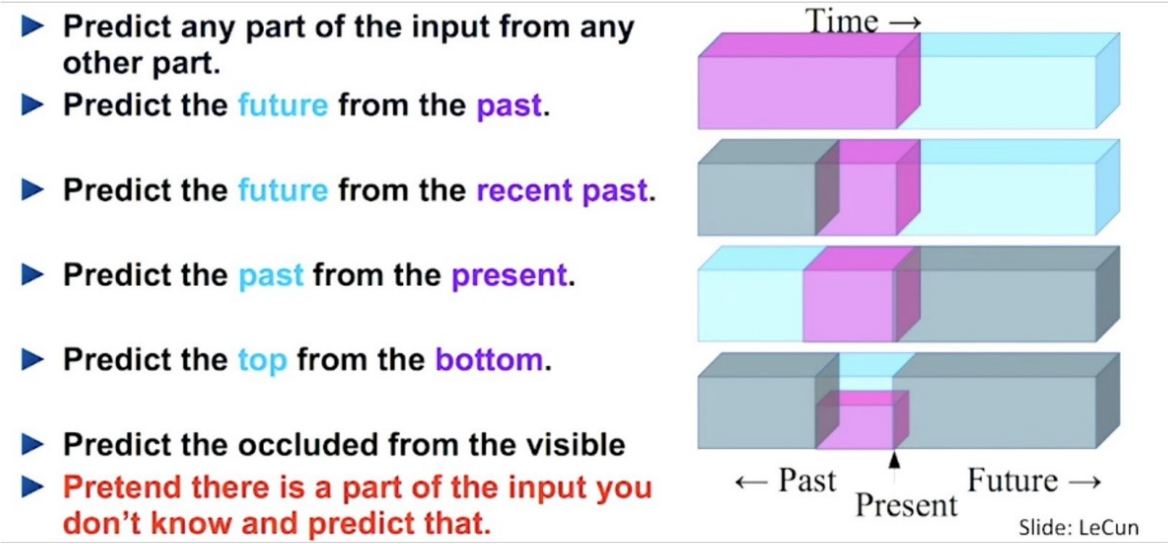
Lilian Weng, Jong Wook Kim, tutoriel NeurIPS 2021

Apprentissage auto-supervisé

- Concevoir une tâche prétexte qui fera apprendre des caractéristiques utiles.

Self-Prediction

Self-prediction construct prediction tasks within every individual data sample: to predict a part of the data from the rest while pretending we don't know that part.



(Famous illustration from Yann LeCun)

Lilian Weng, Jong Wook Kim, tutoriel NeurIPS 2021

Apprentissage auto-supervisé : apprentissage contrastif

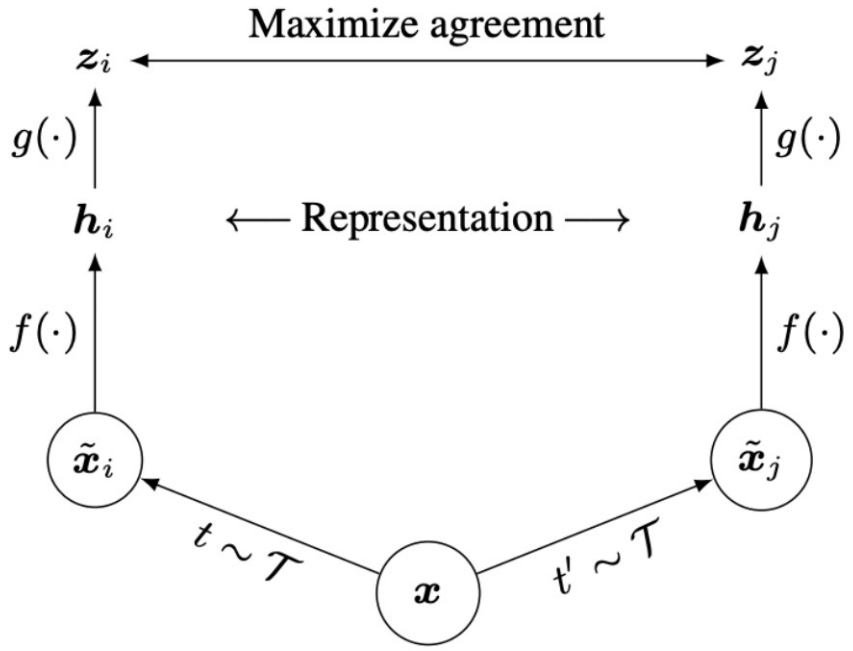


Figure 2. A simple framework for contrastive learning of visual representations. Two separate data augmentation operators are sampled from the same family of augmentations ($t \sim \mathcal{T}$ and $t' \sim \mathcal{T}$) and applied to each data example to obtain two correlated views. A base encoder network $f(\cdot)$ and a projection head $g(\cdot)$ are trained to maximize agreement using a contrastive loss. After training is completed, we throw away the projection head $g(\cdot)$ and use encoder $f(\cdot)$ and representation $\boldsymbol{h}$ for downstream tasks.

Tâche prétexte :

1. Augmenter l'entrée x deux fois : $\tilde{x}_i, \tilde{x}_j$
2. Apprendre une correspondance $g(f(\cdot))$, en minimisant une fonction de perte pour rapprocher $g(f(\tilde{x}_i)), g(f(\tilde{x}_j))$.

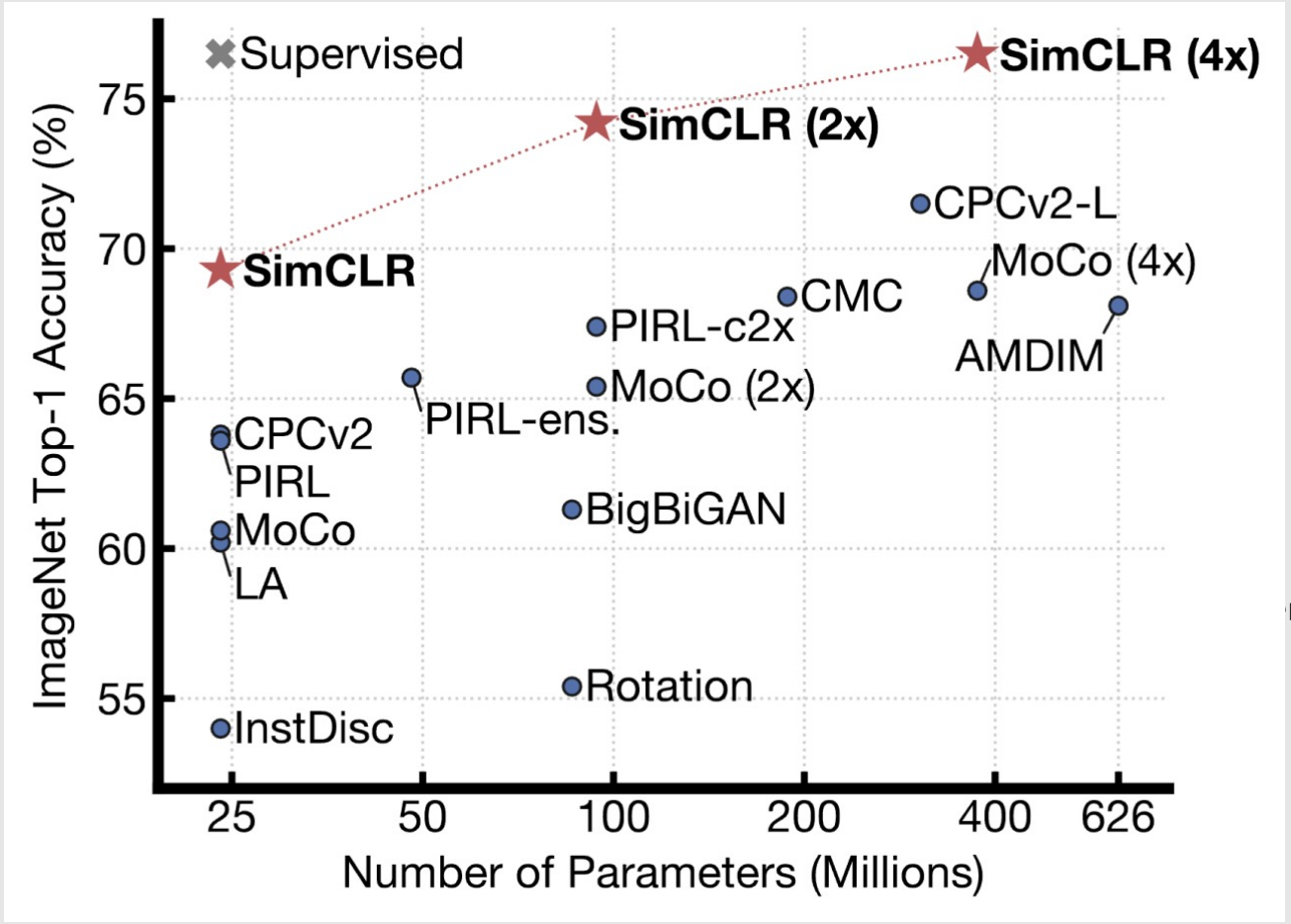
$$\mathcal{L}(\hat{x}, x)$$

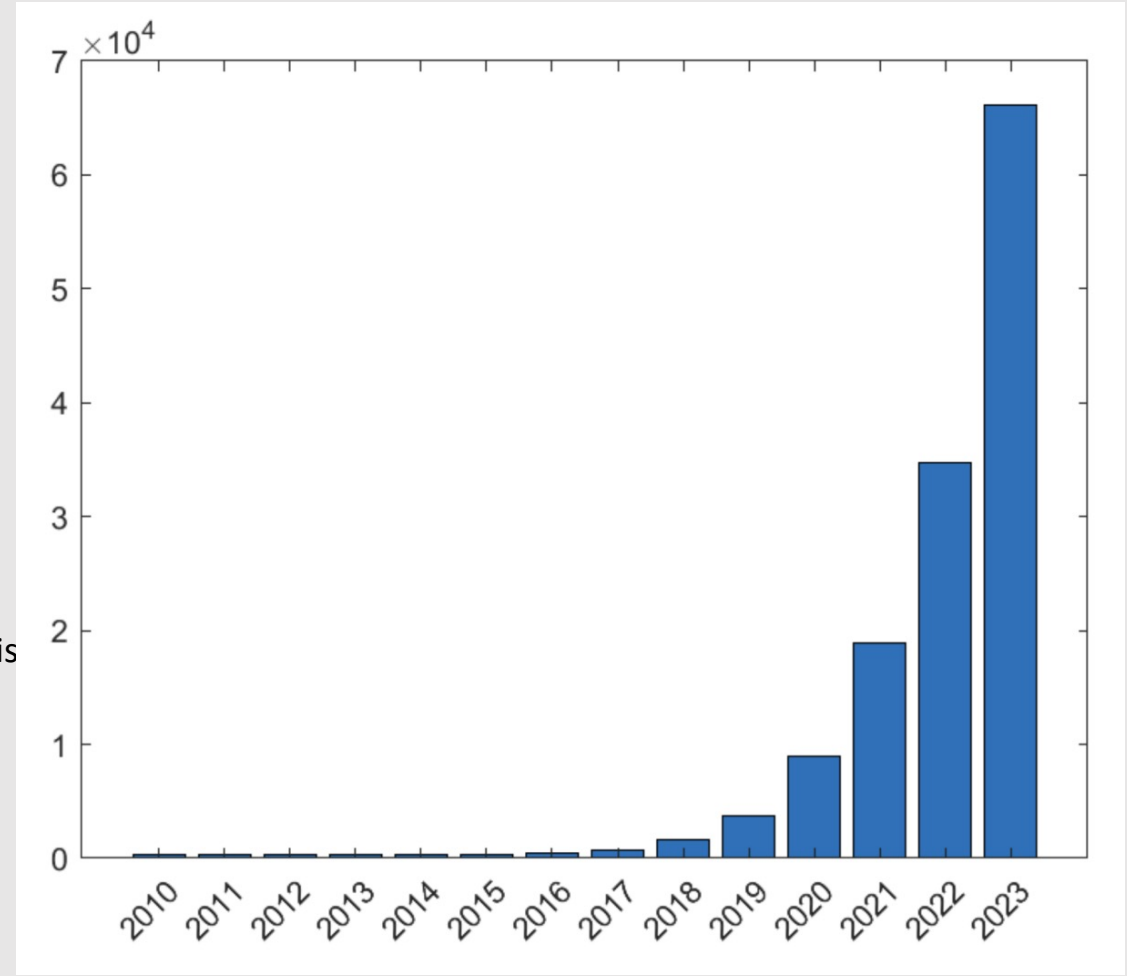
$$\hat{x} := \{g(f(\tilde{x}_i)), g(f(\tilde{x}_j))\}$$

$$\tilde{x} := aug_t(x), t \sim \mathcal{T}$$

SimCLR : [Chen et al., *ICML* 2020]

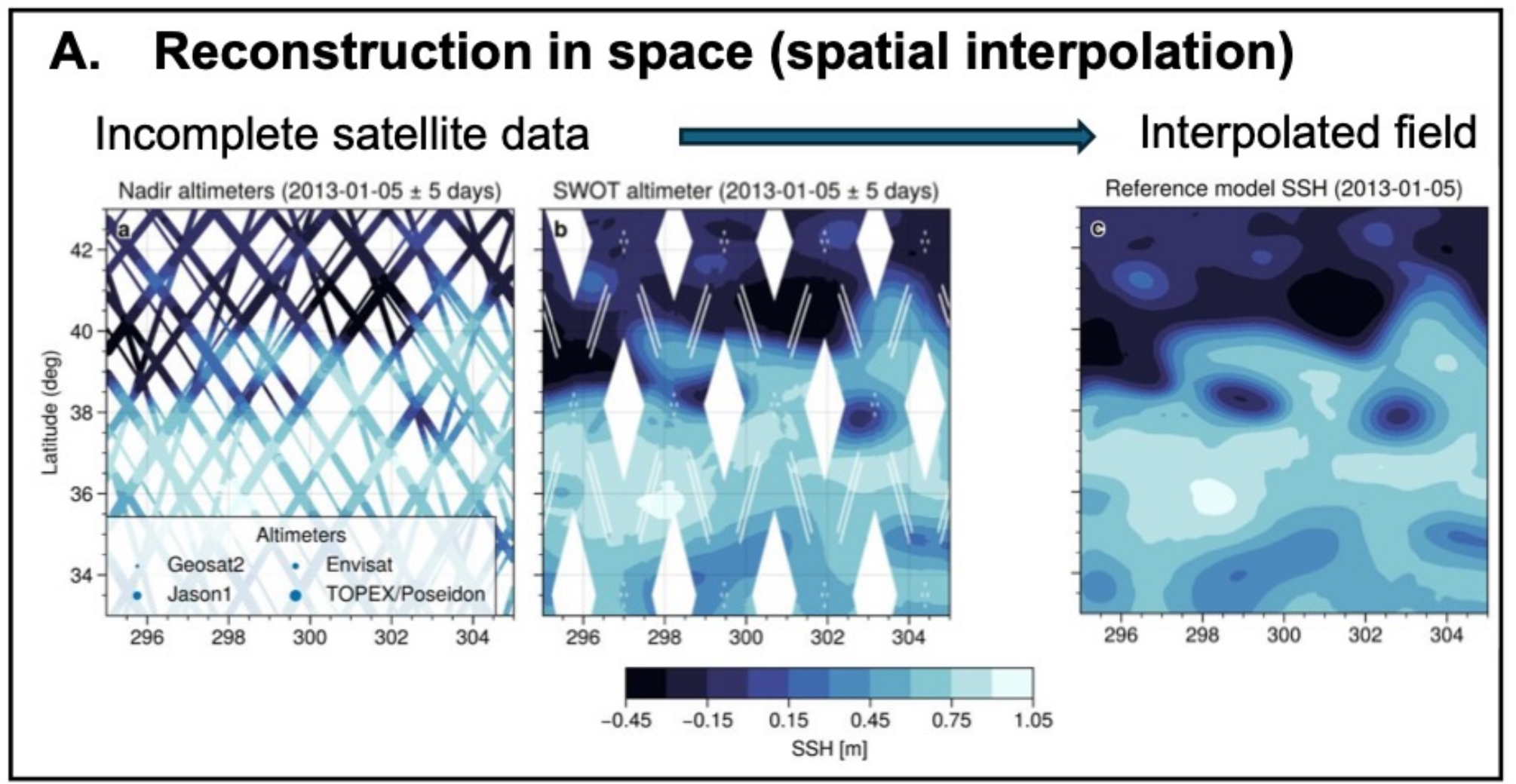
L’apprentissage auto-supervisé peut surpasser le supervisé





SimCLR : [Chen et al., *ICML* 2020]

L’apprentissage auto-supervisé pour vos applications



[Bracco et al., Machine Learning for the Physics of Climate. *Nature Reviews Physics*, 2024]

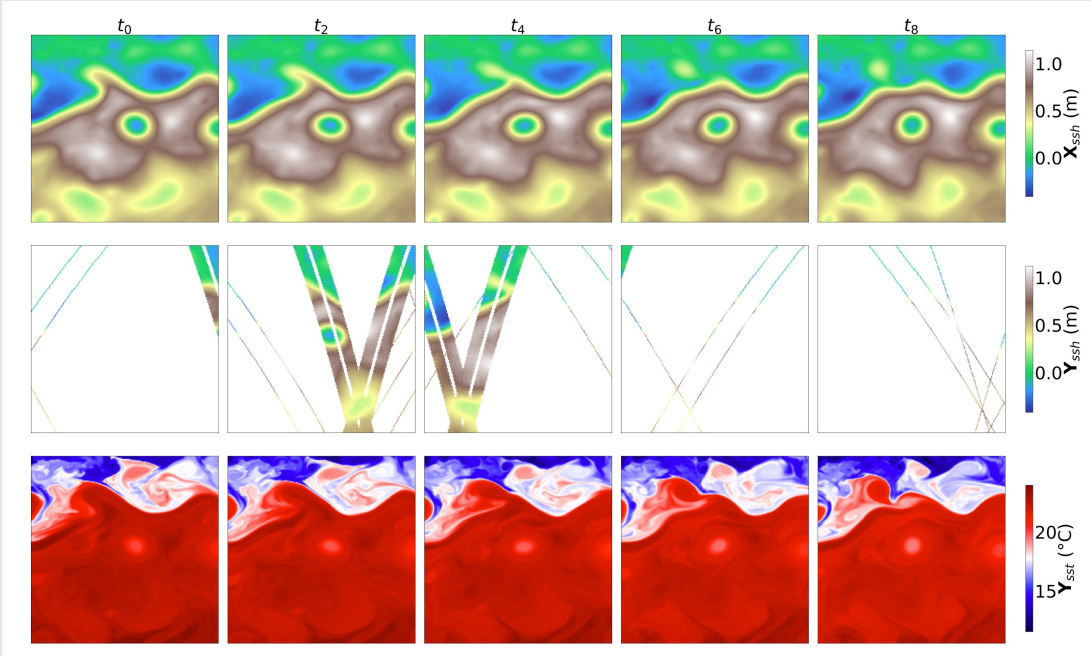
Interpolation spatio-temporelle non supervisée

MUSTI: Multi-modal Unsupervised Spatio-Temporal Interpolation of satellite ocean altimetry maps. [Archambault et al. VISAPP 2023]

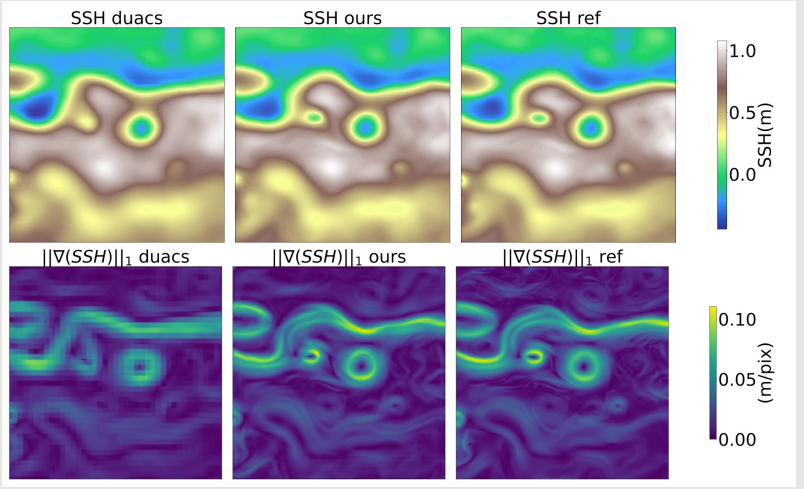
Champs SSH de référence (non utilisés à l'entraînement)

Données SSH le long des traces des satellites

Champs SST issus d'un modèle physique



Données d'entraînement



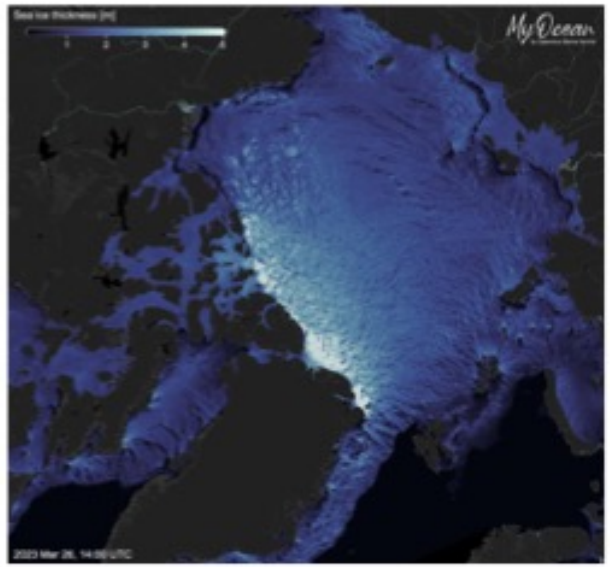
Résultats

L’apprentissage auto-supervisé pour vos applications

C. Reconstruction in scales (downscaling)

Low resolution field → High resolution field





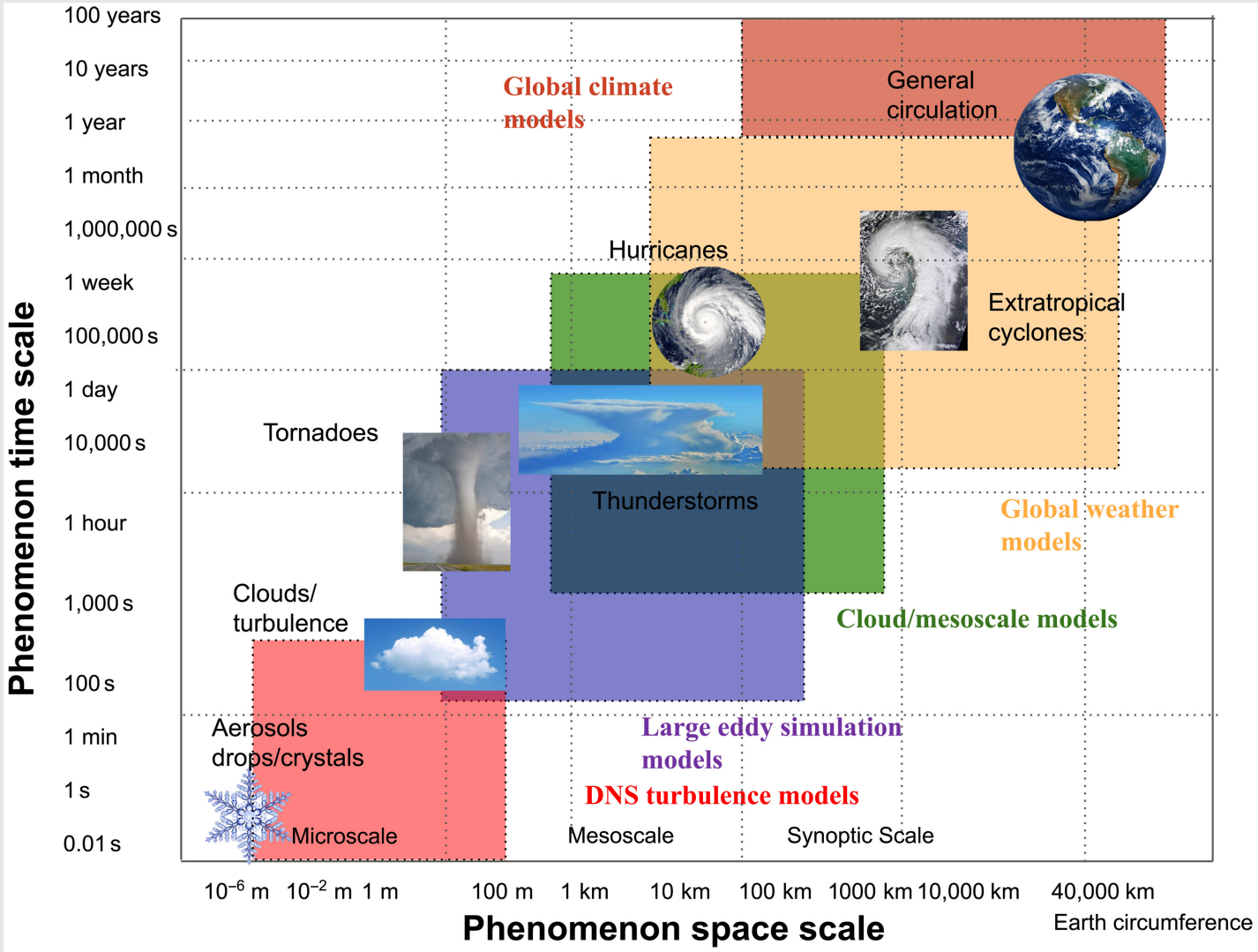
[Bracco et al., Machine Learning for the Physics of Climate. *Nature Reviews Physics*, 2024]

Descente d'échelle des simulations de modèles climatiques

Les simulations des modèles climatiques mondiaux sont plus grossières (en espace et en temps) que ne l'exigent de nombreuses tâches :

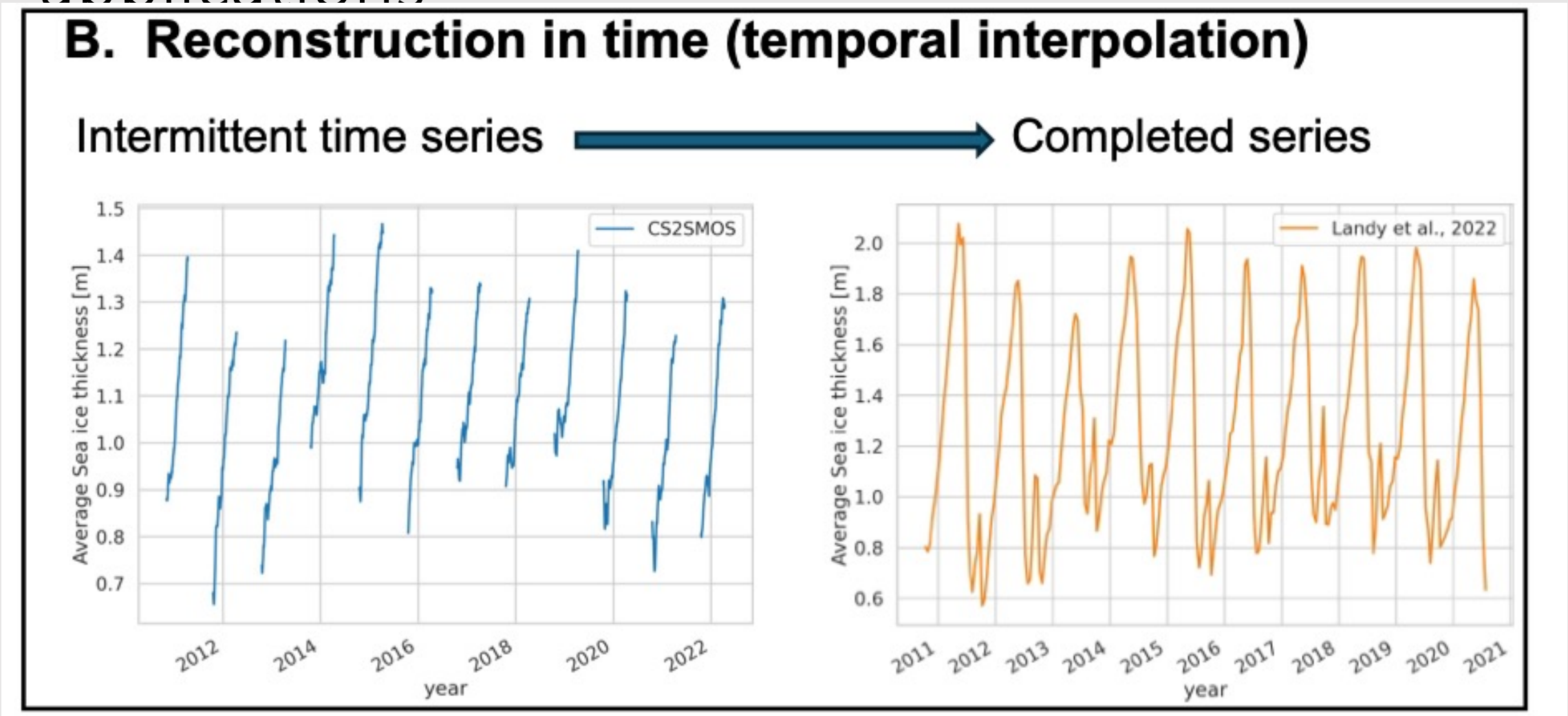
- Adaptation au changement climatique
- Atténuation du changement climatique
- Projection des impacts à long terme

Approche : utiliser l'IA générative pour la descente d'échelle des données de modèles climatiques vers les échelles pertinentes



[Gettelman, et al., Science Advances, 2022]

L’apprentissage auto-supervisé pour vos applications



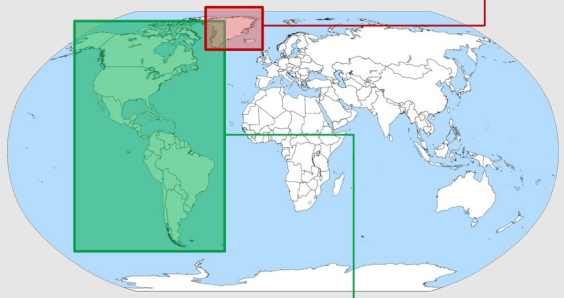
[Bracco et al., Machine Learning for the Physics of Climate. *Nature Reviews Physics*, 2024]

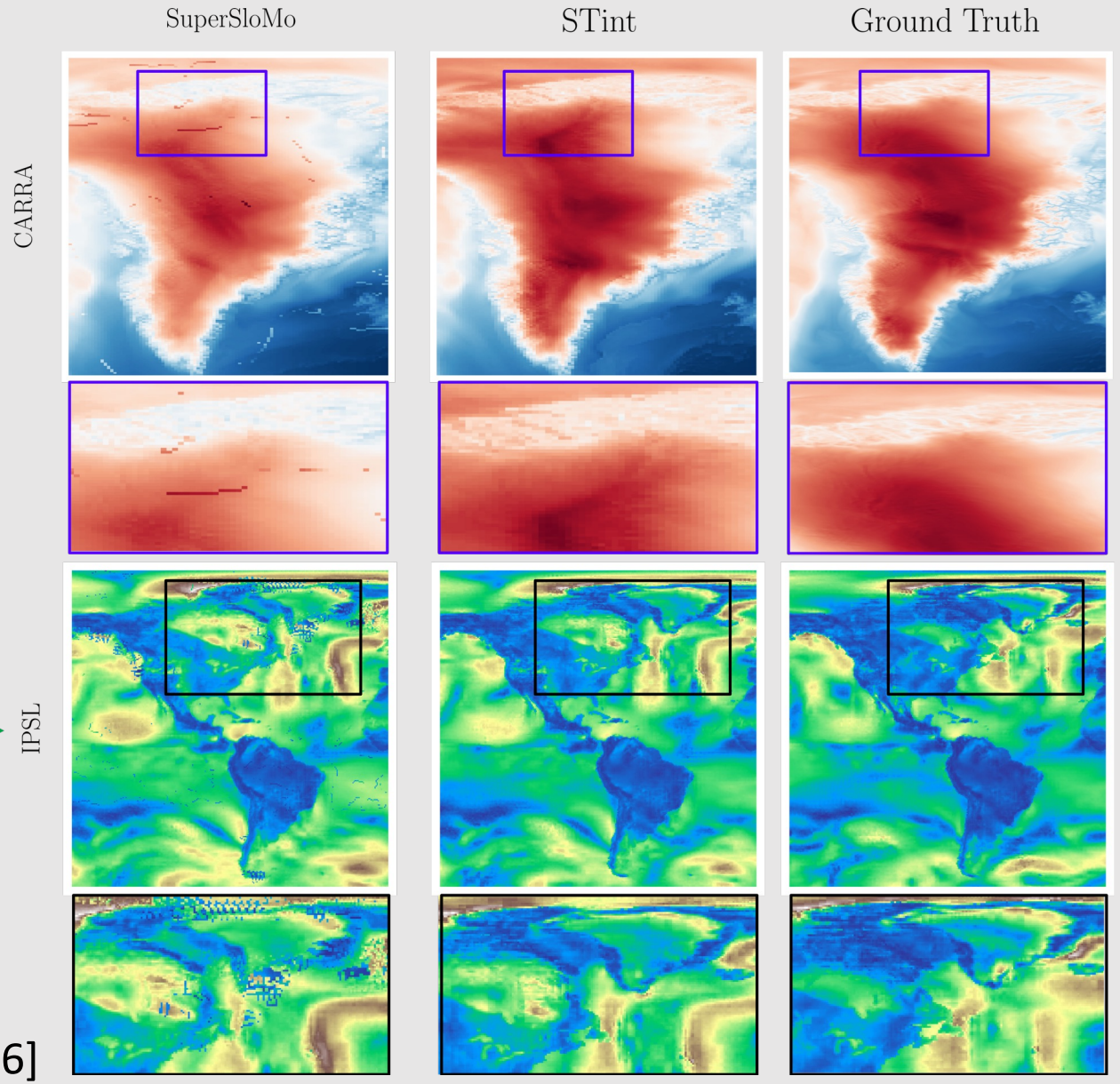
Exemple : descente d’échelle temporelle (interpolation) de données géospatiales

Pourquoi ne pas simplement utiliser les méthodes d'IA existantes ?

- Les applications climatiques portent sur des données géospatiales qui évoluent dans le temps
  - Données d'observation maillées sur le globe par assimilation de données
  - Simulations produites par des modèles physiques (PNT, MCG, MCR)
- Ce sont des tenseurs de valeurs réelles selon la latitude, la longitude, le temps,
  et éventuellement plusieurs variables climatologiques
- Les algorithmes de vision par ordinateur pour les « données spatio-temporelles » s'appuient sur des propriétés des **données vidéo** qui **se généralisent mal aux données géospatiales**
  - p. ex. profondeur, contours et « objets »
  - vs schémas éphémères dans les fluides

STINT : interpolation temporelle auto-supervisée







L'état de l'art en vision par ordinateur pour l'interpolation temporelle de vidéos utilise le flot optique.
C'est problématique pour les données géospatiales.

[Harilal, Hodge, Subramanian, & Monteleoni, 2026]

Limites des travaux antérieurs : interpolation bicubique

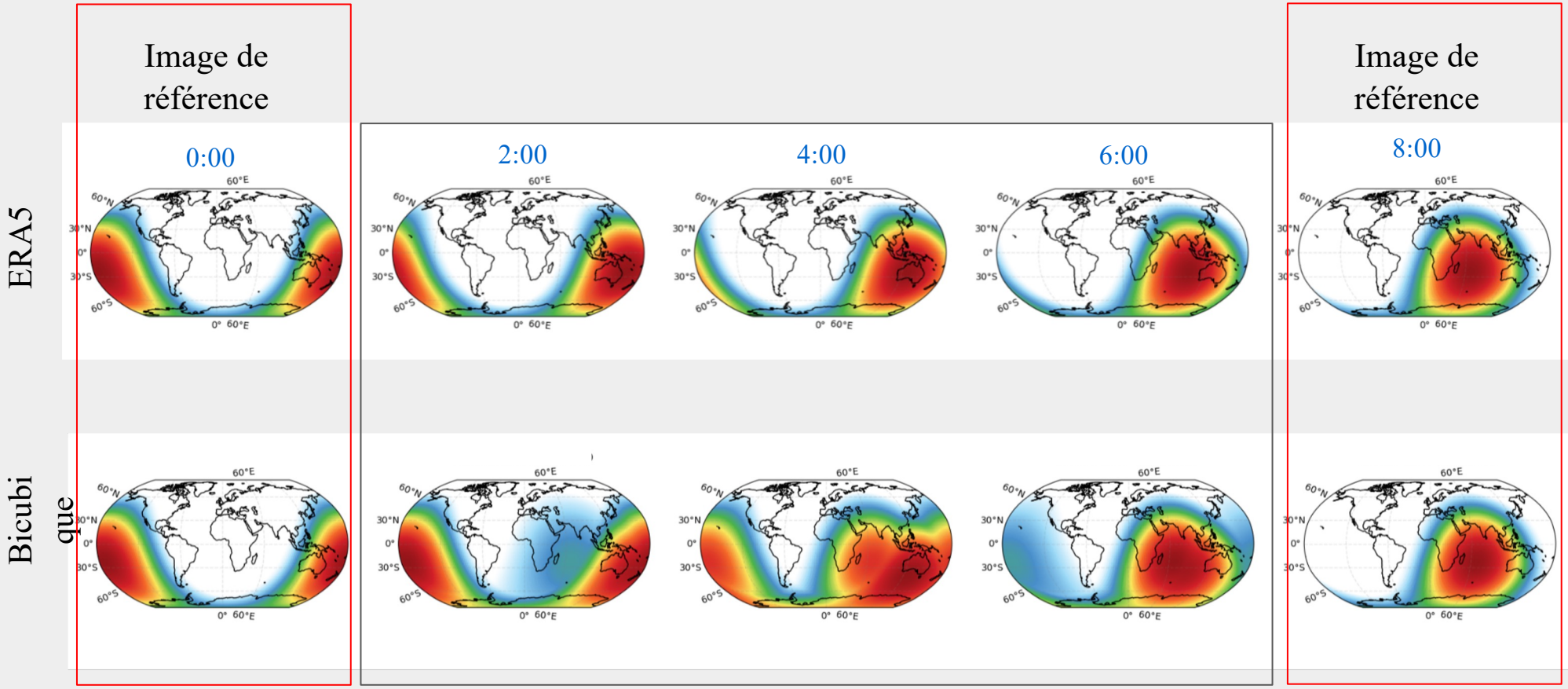
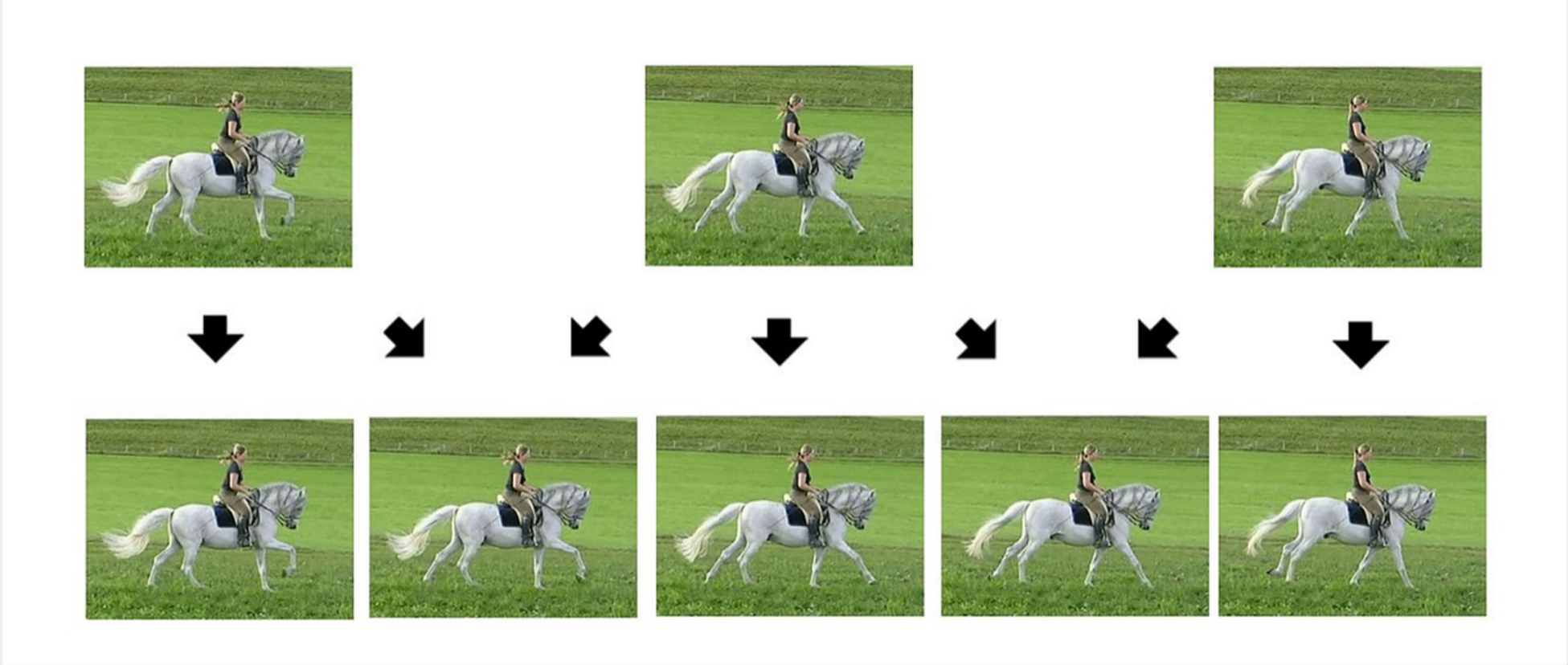


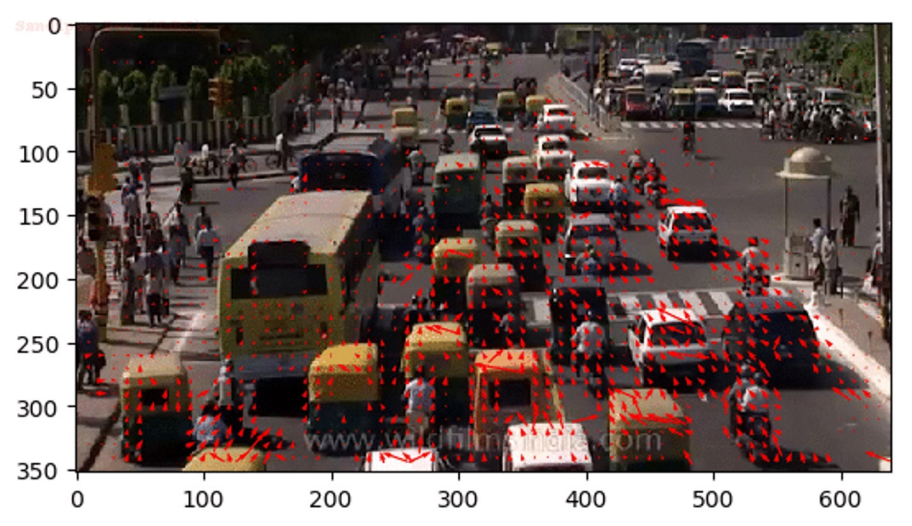
Figure. *Interpolation bicubique temporelle de champs d'irradiance solaire ERA5*

L’interpolation vidéo en vision par ordinateur

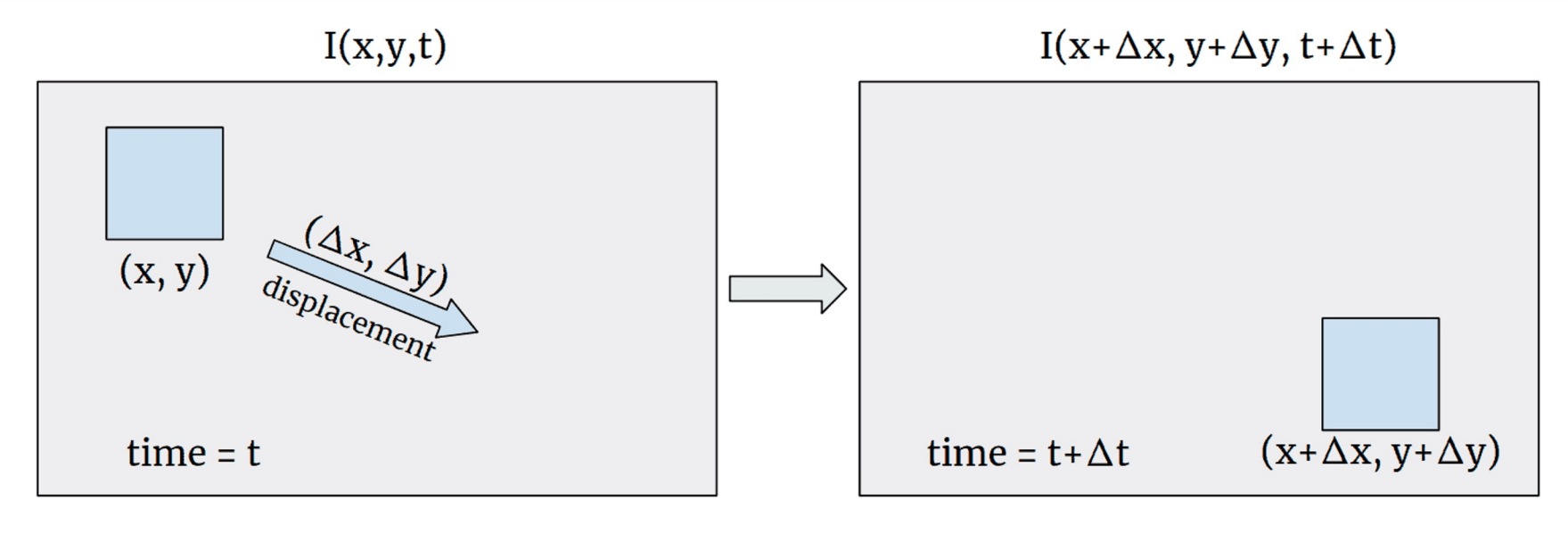


Le flot optique

- Flot optique : champ de mouvement pixel par pixel entre deux images (p. ex. des images vidéo).
- Il encode la projection 2D « perçue » du mouvement physique des points par rapport à l'observateur.



(source : Nvidia OF SDK)



Le flot optique

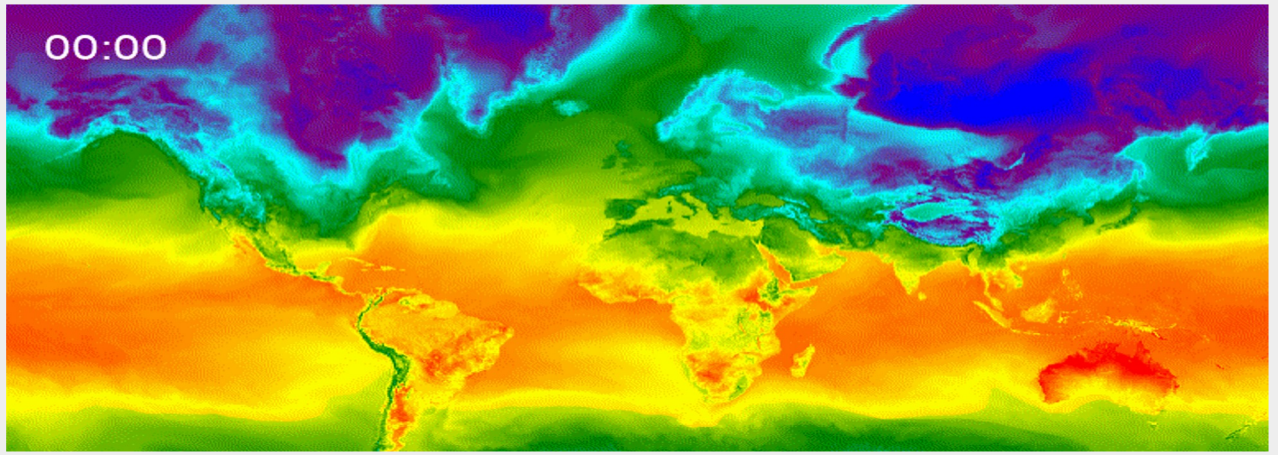
Hypothèses :

- Petit déplacement (Δt).
- Pas/peu de changement de la lumière réfléchie.
- Pas/peu de changement d'échelle.
- Pas d'occultation ni de désoccultation.



Inapplicable aux données géospatiales !

- Pas de notion d'objets ni de profondeur.
- Les entités peuvent se former/déformer arbitrairement.
- Résolution temporelle bien plus grossière.



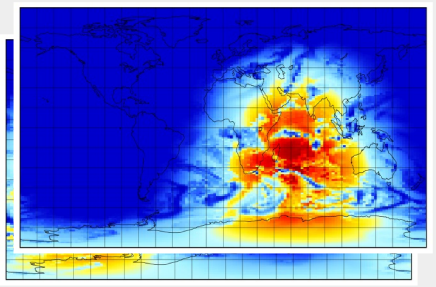
Température horaire NOAA GFS

Objectif : améliorer la résolution temporelle

- La tâche est l'interpolation temporelle
- Mais nous n'avons pas de données horaires pour l'entraînement

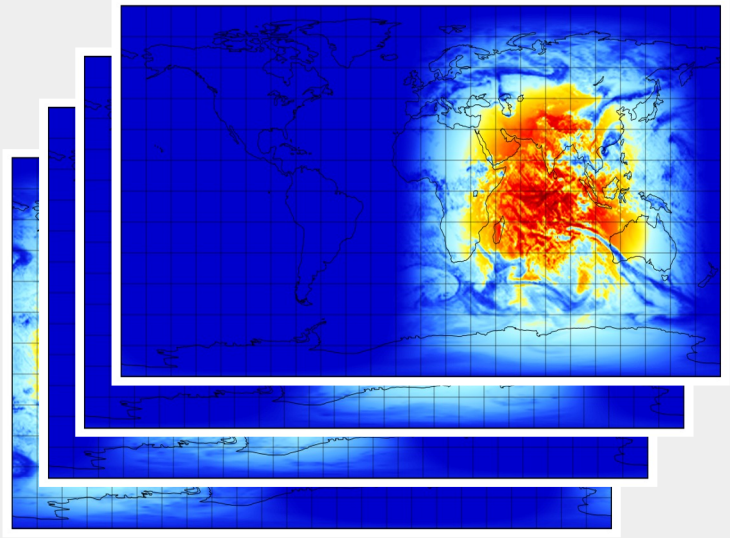
il faut des techniques non supervisées (**auto-supervisées**).

CMIP6-IPSL



Trihoraire

?



Horaire

L’auto-supervisé en bref

- Utiliser des tâches prétextes/substituts de façon supervisée.
  - Étiquettes **générées** automatiquement.

$$f_W(x) = \hat{y}$$

$$\mathcal{L}(\hat{y}, y)$$

Apprentissage supervisé

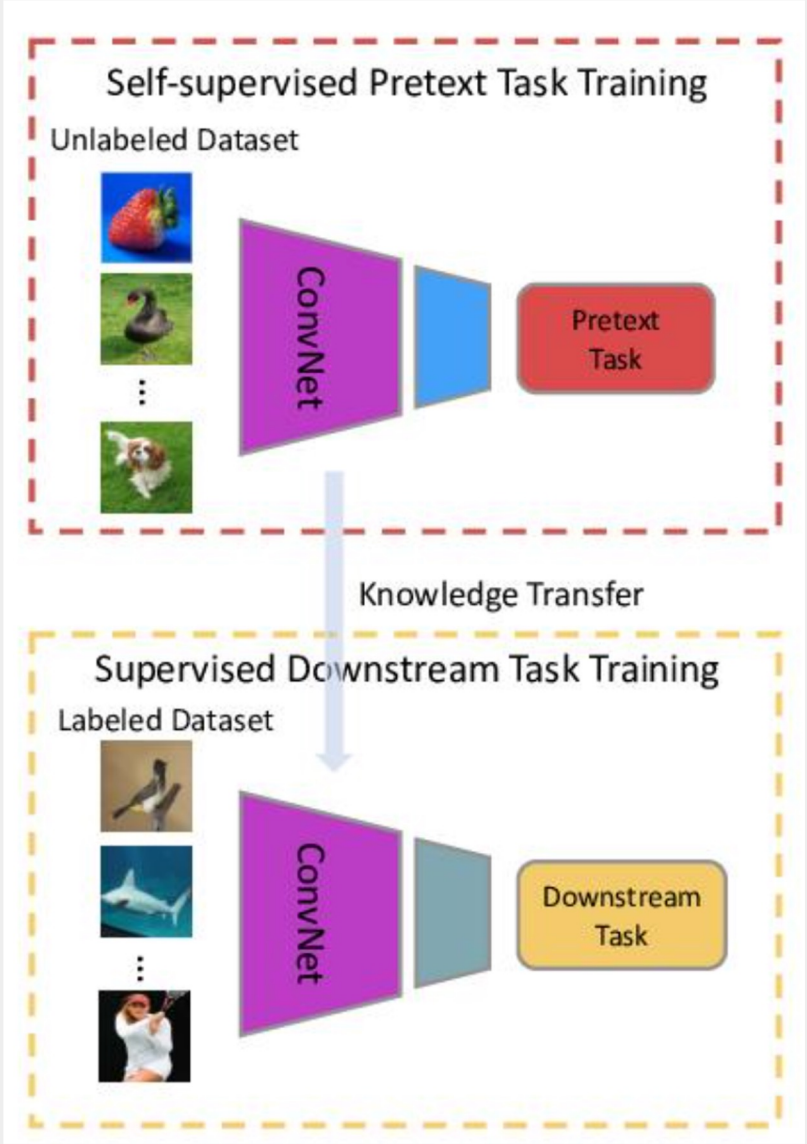
$$\longrightarrow$$

$$f_W(x) = \hat{x}$$

$$\mathcal{L}(\hat{x}, x)$$

Apprentissage auto-supervisé

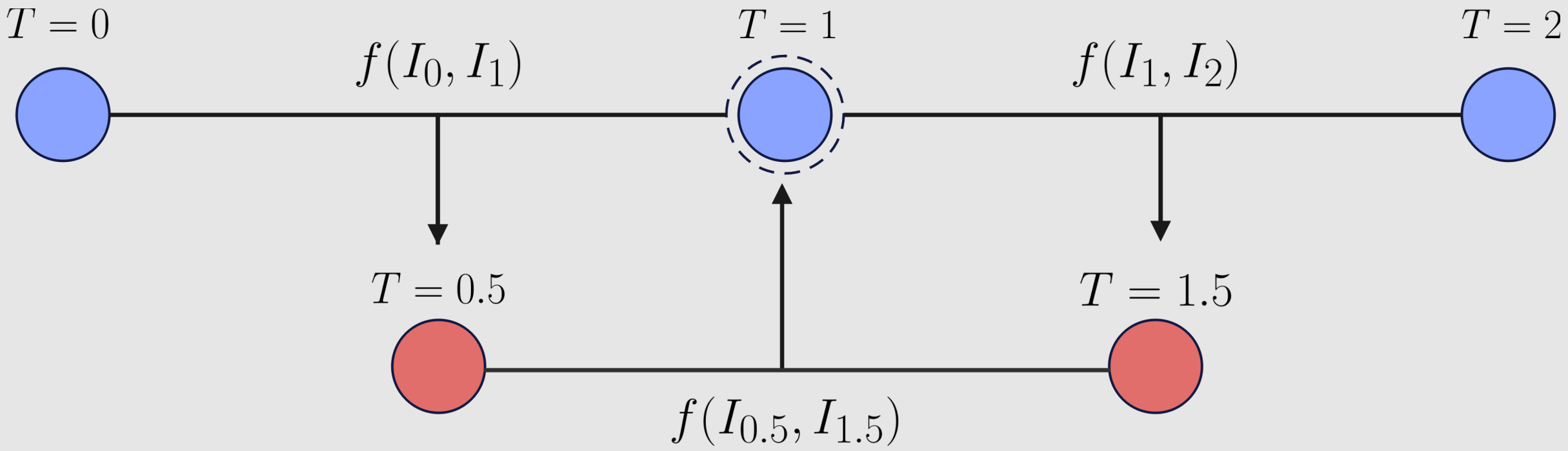
- Tâche en aval : recherche, classification supervisée ou semi-supervisée, etc.



Une tâche prétexte pour l'interpolation temporelle

STINT : interpolation temporelle auto-supervisée pour les données géospatiales

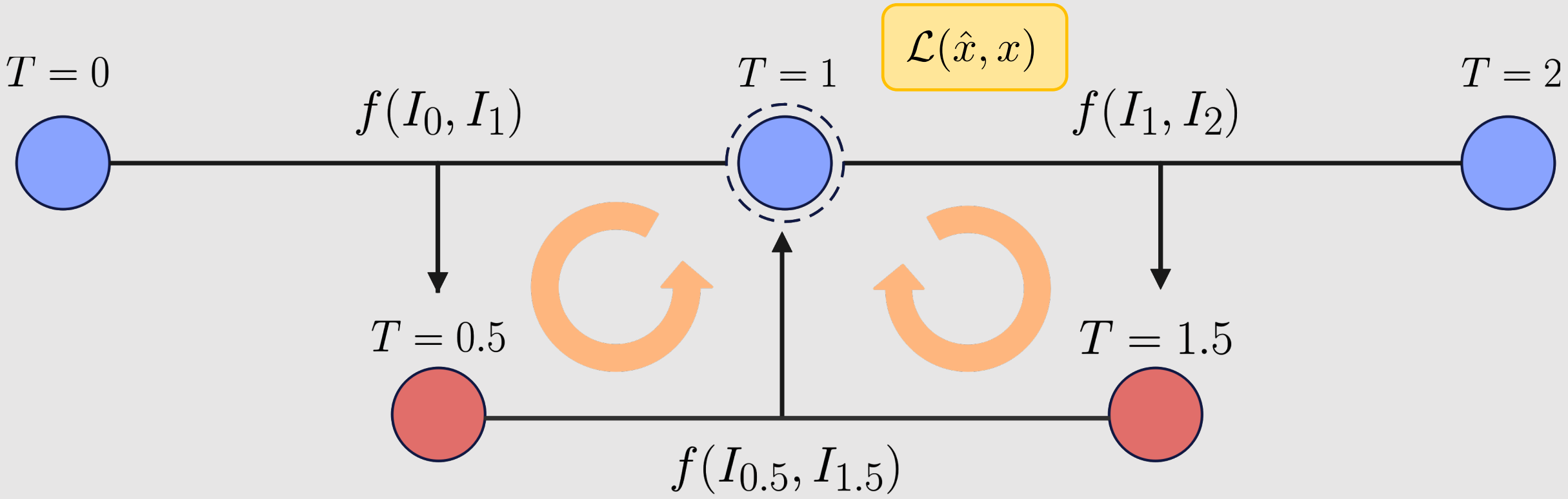
[Harilal, Hodge, Subramanian, & Monteleoni, Climate Informatics 2026]



Une tâche prétexte pour l'interpolation temporelle

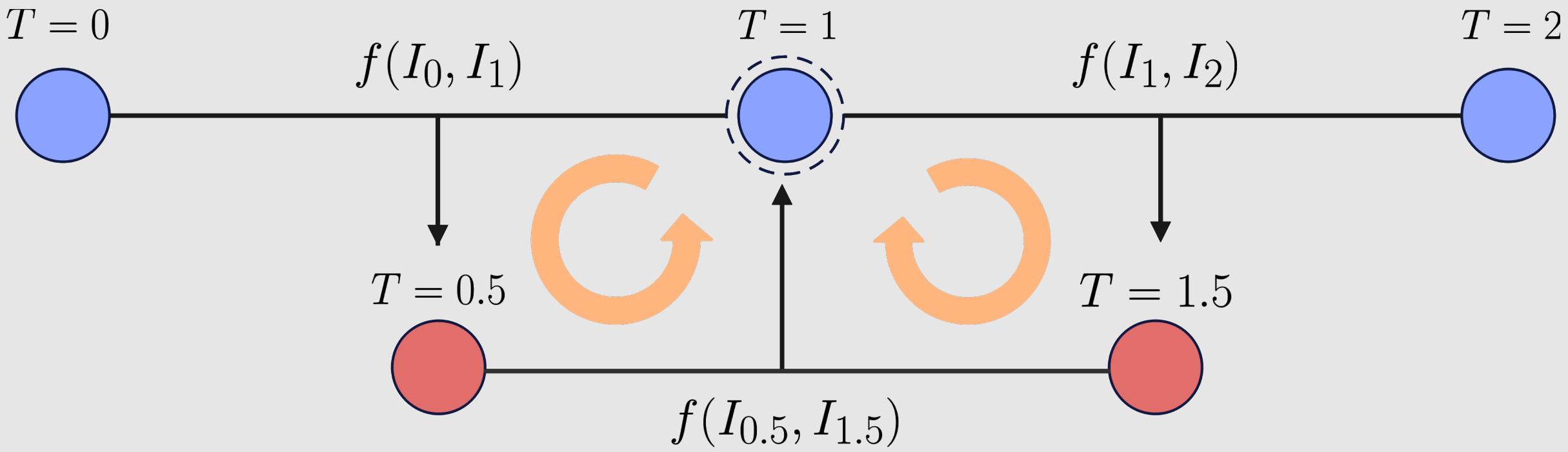
STINT : interpolation temporelle auto-supervisée pour les données géospatiales

[Harilal, Hodge, Subramanian, & Monteleoni, Climate Informatics 2026]



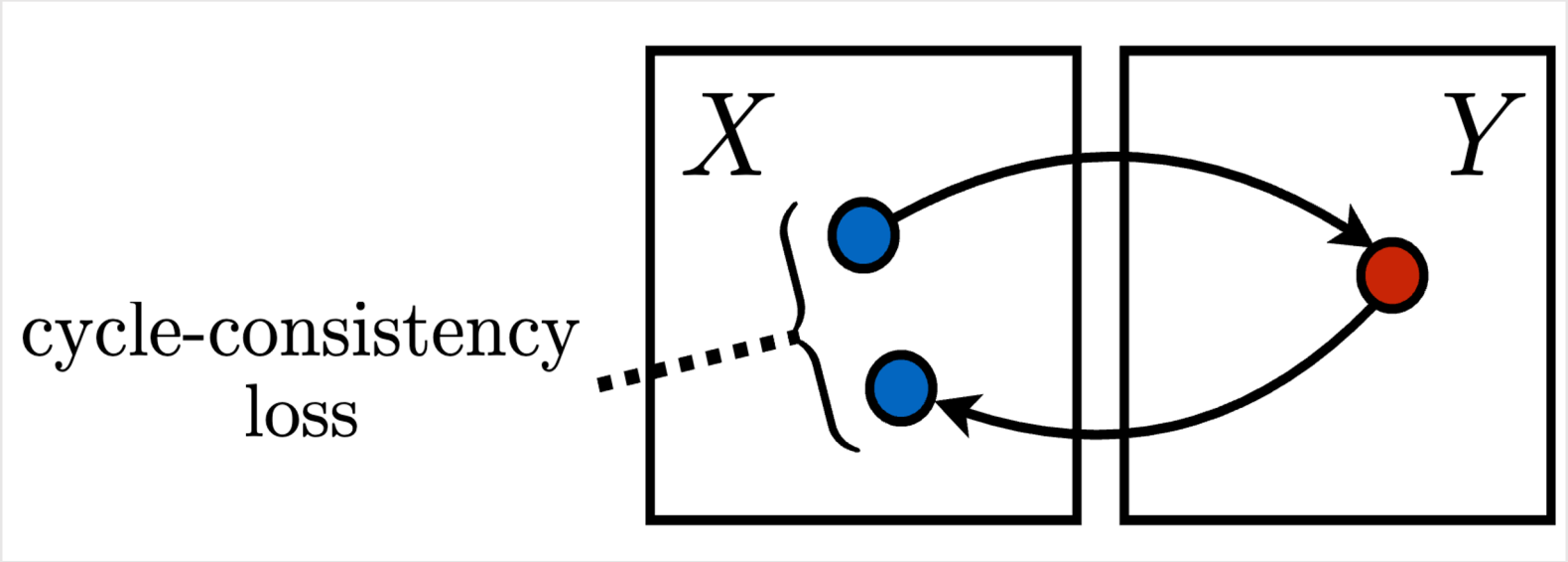
Une tâche prétexte pour l'interpolation temporelle

$$\mathcal{L}(\hat{x}, x) = \sum_{i=1}^{n} \mathcal{L}(T_i, \hat{T}_i) \qquad \hat{T}_i := f(f(T_{i-1}, T_i), f(T_i, T_{i+1}))$$

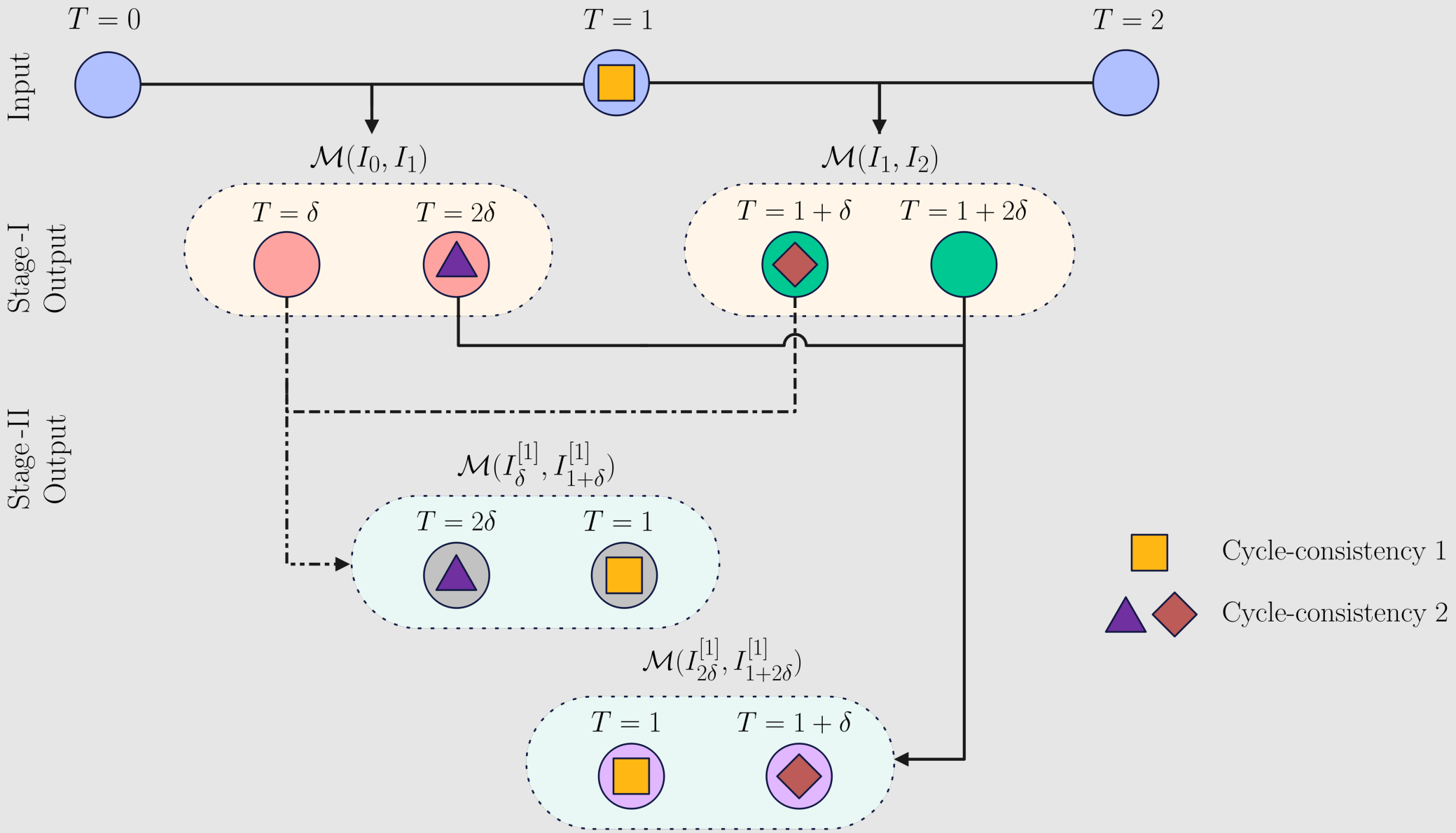


Perte de cohérence cyclique

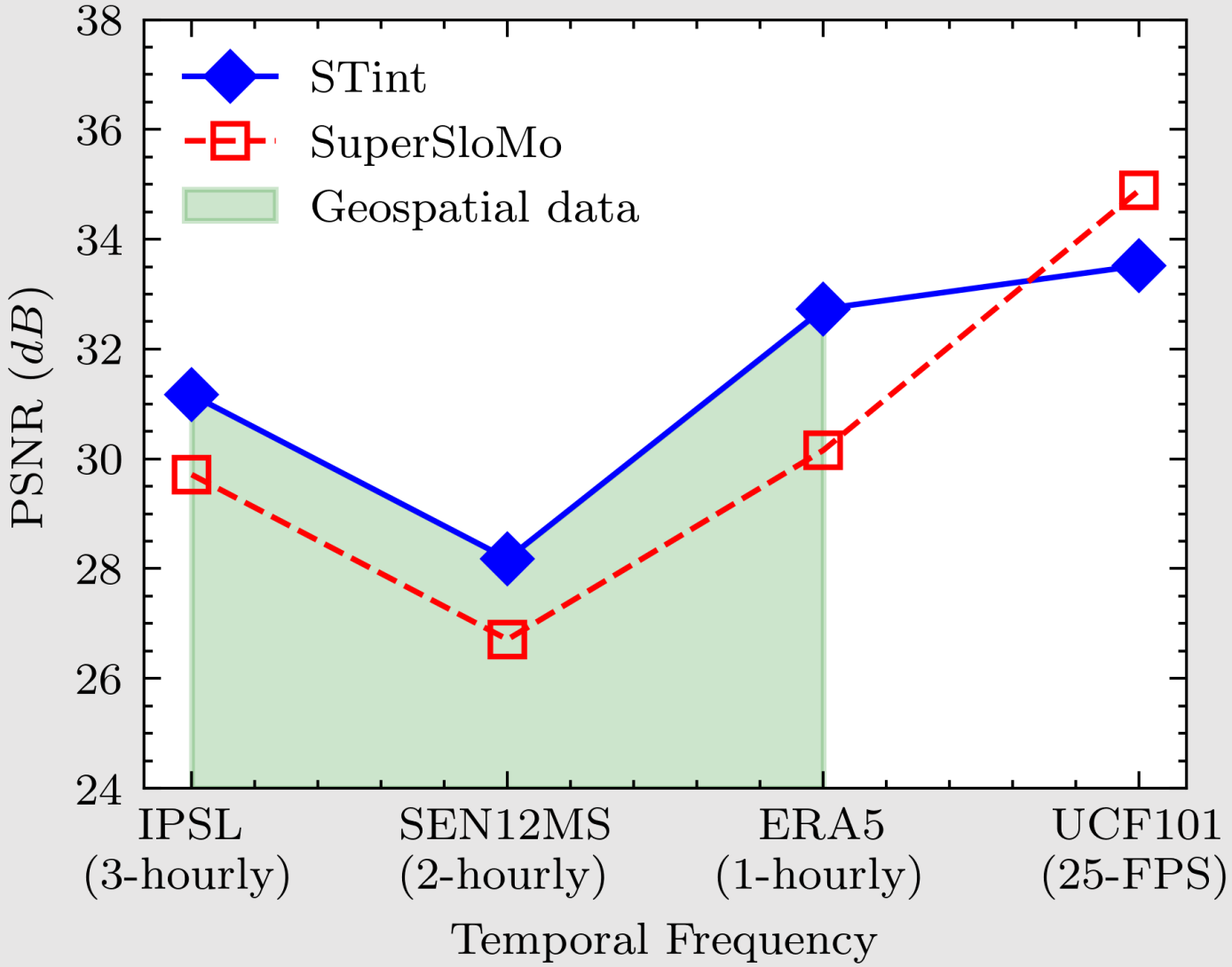
$$\mathcal{L}(\hat{x}, x) = \sum_{i=1}^{n} \mathcal{L}(T_i, \hat{T}_i) \qquad \hat{T}_i := f(f(T_{i-1}, T_i), f(T_i, T_{i+1}))$$



STINT : interpolation temporelle auto-supervisée



STINT : interpolation temporelle auto-supervisée



	$\frac{MSE}{Capacity}$ (↓)	PSNR (↑)	SSIM (↑)
ERA5 Solar			
Baseline	0.3086	25.238	0.623
SuperSloMo	0.0907	30.157	0.733
Proposed	**0.0561**	**32.731**	**0.792**
IPSL Wind			
Baseline	0.6206	24.097	0.619
SuperSloMo	0.4150	29.713	0.681
Proposed	**0.2904**	**31.167**	**0.713**
CARRA Temperature			
Baseline	0.5319	27.832	0.667
SuperSloMo	0.1604	30.276	0.724
Proposed	**0.0975**	**31.908**	**0.775**

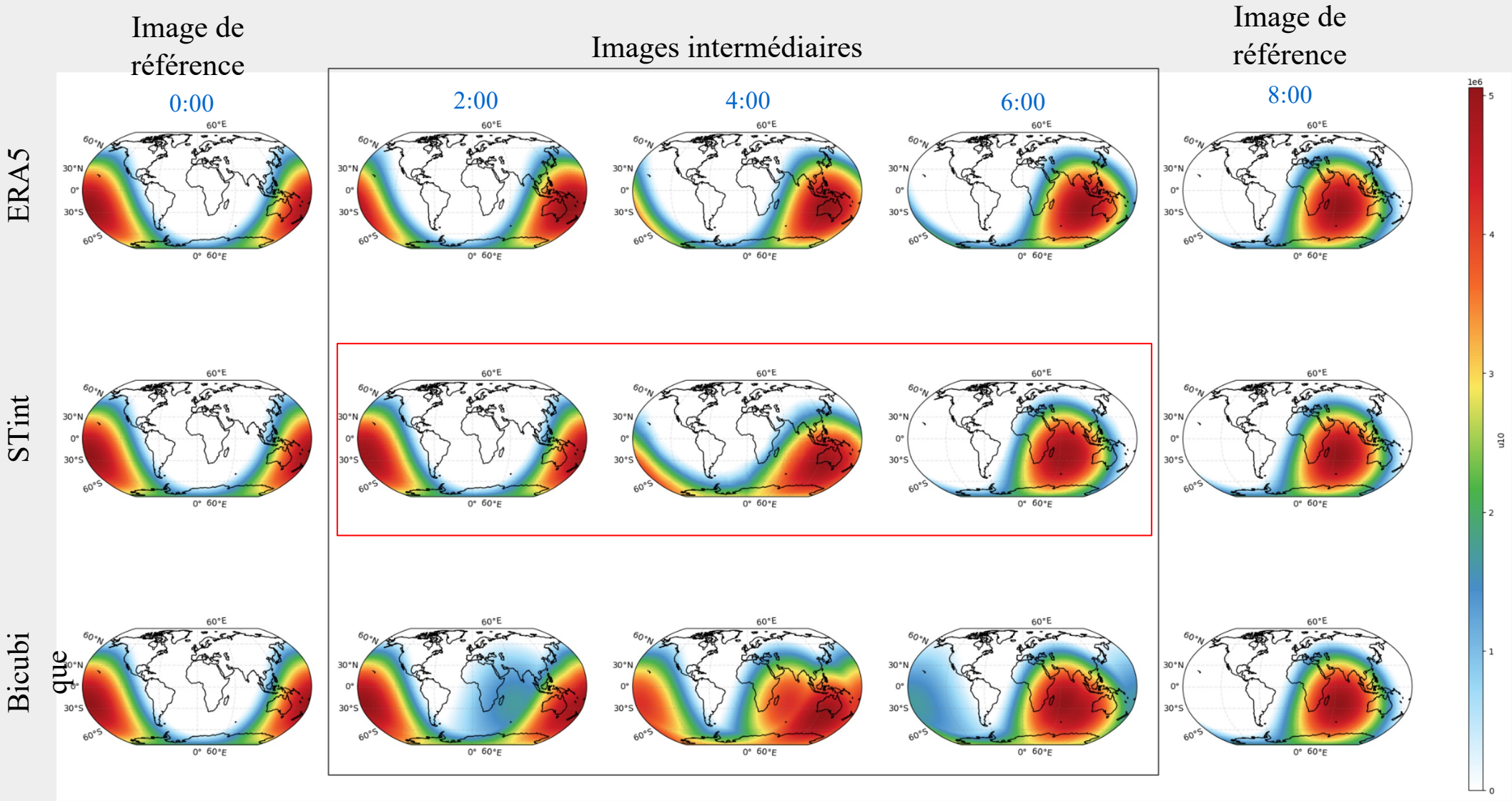


Figure. *Modèle entraîné sur des champs d'irradiance solaire ERA5 et interpolation (3 images non vues) entre 0 h et 8 h.*

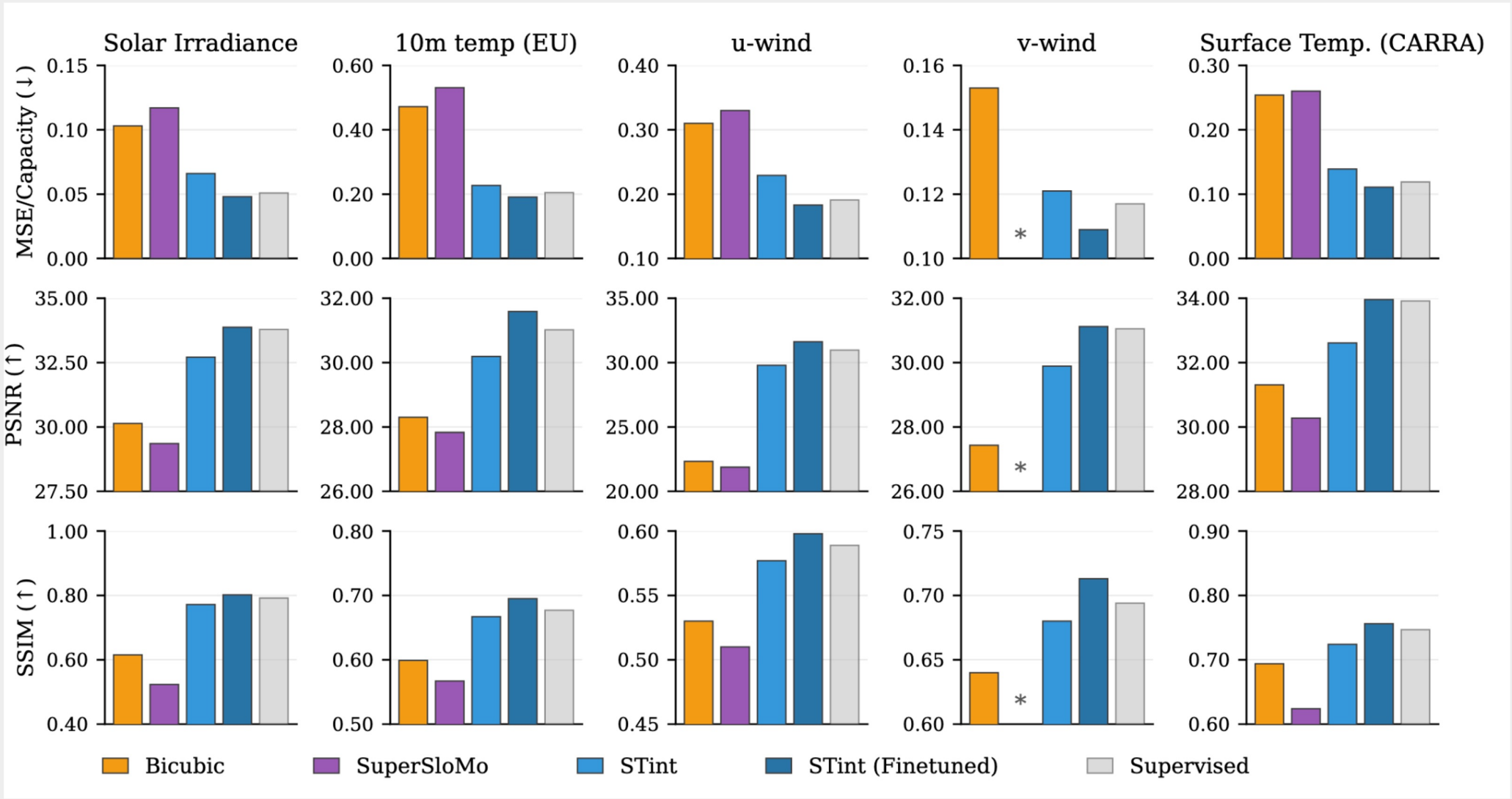
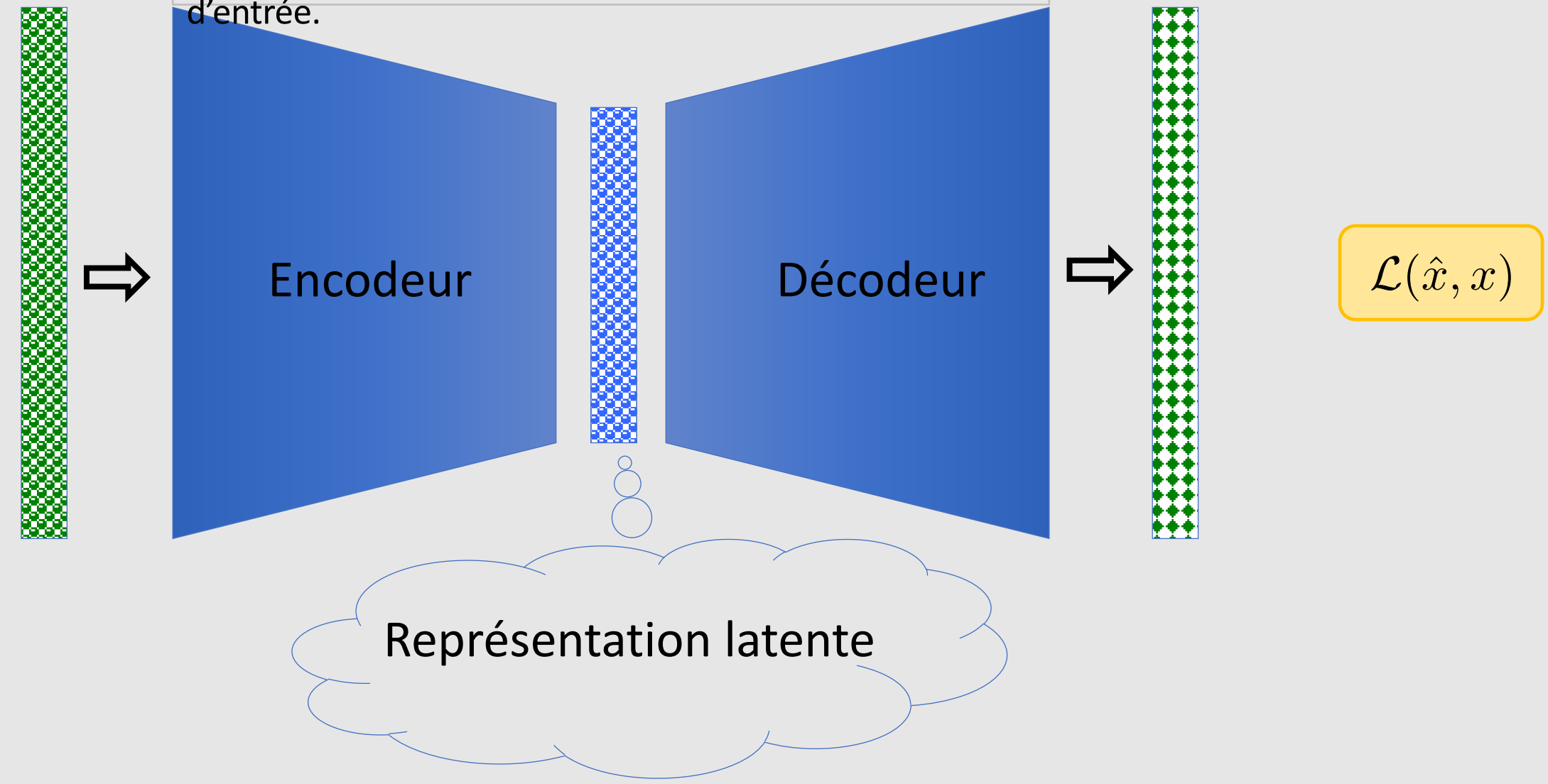


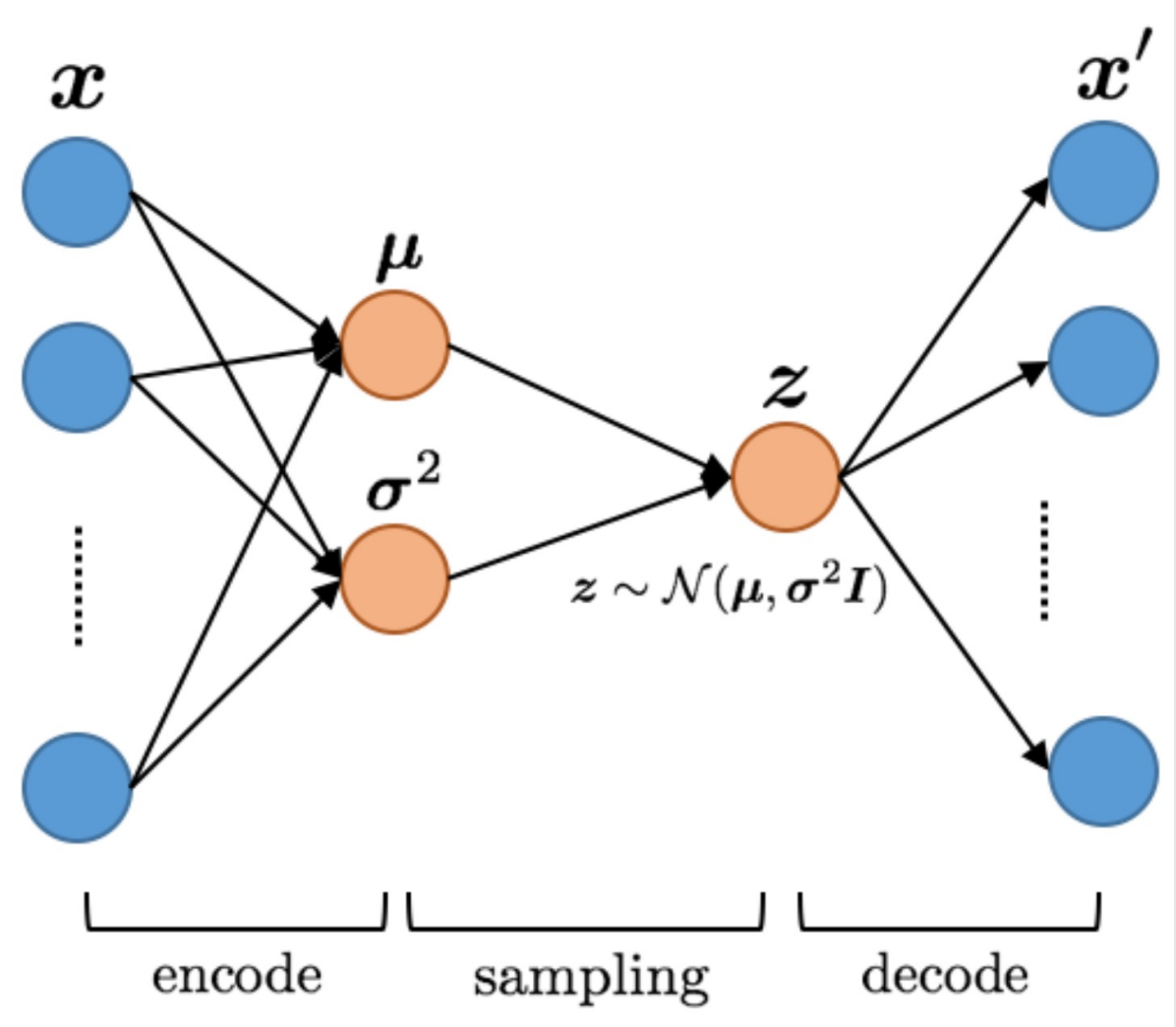
Figure. *Résultats d'interpolation temporelle sur différents jeux de données géospatiales.*

Autoencodeur : les paramètres des réseaux encodeur et décodeur sont entraînés pour que la sortie approche l'entrée. Après entraînement sur de nombreux exemples, les paramètres de la couche goulot forment une représentation compacte de la distribution d'entrée.



Autoencodeur variationnel (VAE)

Apprendre une distribution sur les représentations latentes, au lieu d'un encodage unique



Apprendre un autoencodeur variationnel (VAE)

- L'encodeur est une distribution, q, qui envoie l'espace des données vers l'espace latent
  - Défini par les poids du réseau θ :

$$q_\theta(z|x)$$

- Le décodeur est une distribution, p, qui envoie l'espace latent vers l'espace des données
  - Défini par les poids du réseau ϕ :

$$p_\phi(x|z)$$

- Ajuster les poids θ et ϕ en minimisant :

Erreur de reconstruction sur les données d'entraînement

Terme de régularisation : pénalise l'écart à un a priori simple

$$E_{q_\theta(z|x)}[\log p_\phi(x|z)] - \beta KL(q_\theta(z|x)||p(z))$$

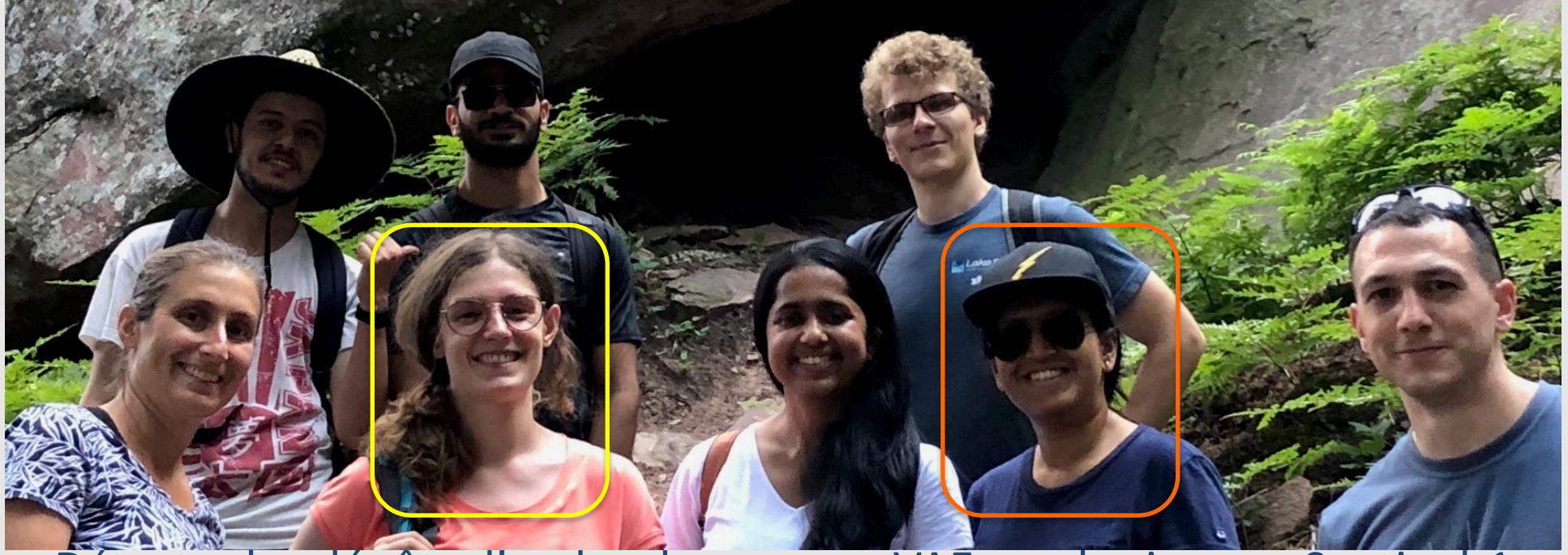
- Optimisé par une variante de l'espérance-maximisation (EM), avec une borne inférieure variationnelle

Exemple : VAE pour la détection semi-supervisée d’anomalies, appliqué à la détection d’avalanches

Détection d’avalanches



Source de l’image : https://www.nytimes.com/2017/01/19/world/europe/italy-avalanche.html



Détecter les dépôts d'avalanche avec un VAE sur des images Sentinel-1

Saumya Sinha*, Sophie Giffard-Roisin*, Fatima Karbou, Michael Deschatres, Anna Karas, Nicolas Eckert, Cécile Coléou, Claire Monteleoni

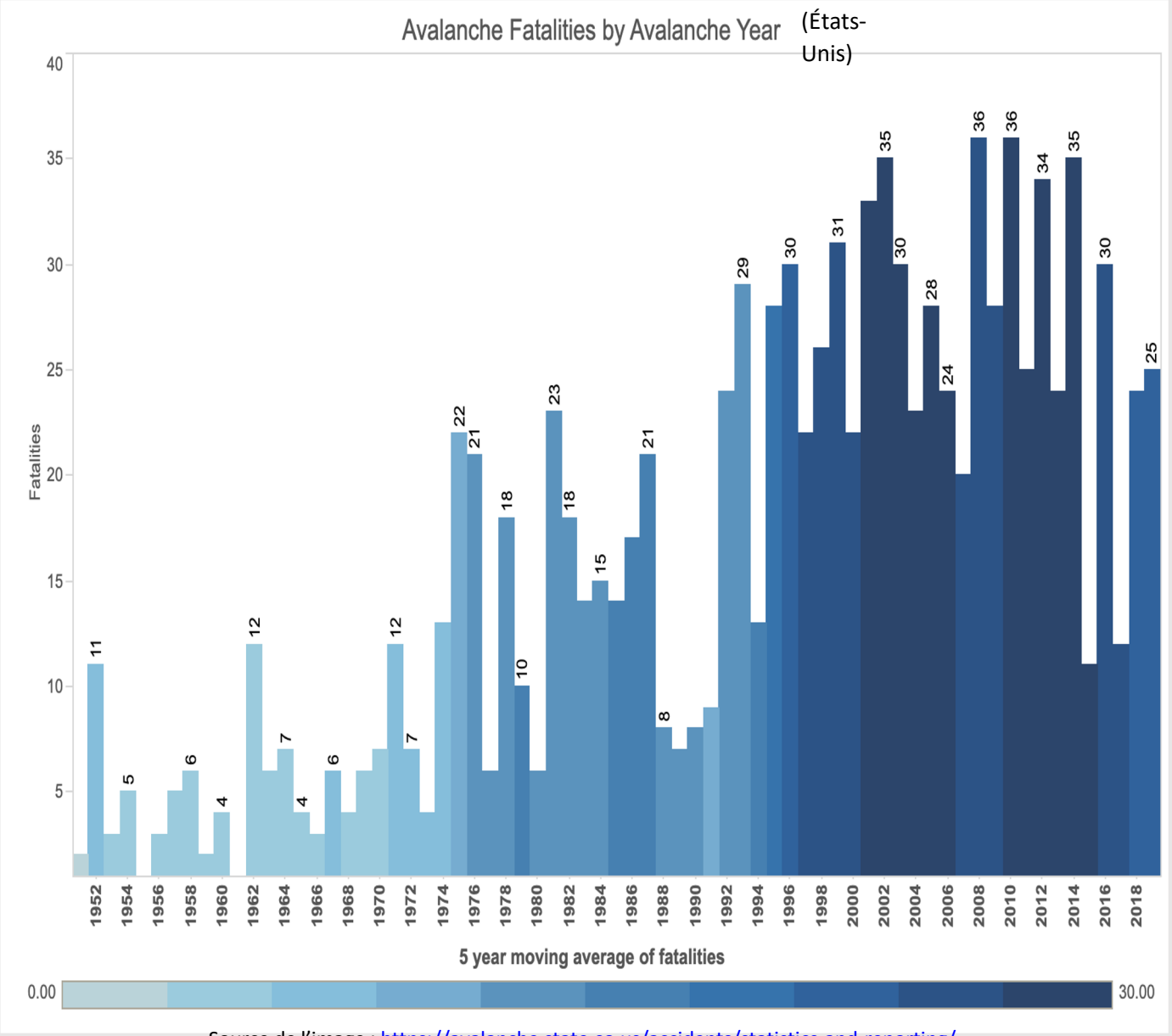
Variational autoencoder anomaly-detection of avalanche deposits in satellite SAR imagery. *Proc. 10th International Conference on Climate Informatics (CI) 2020*

Boulder

*Co-premiers auteurs

METEO FRANCE

Décès par avalanche selon l'année d'avalanche



Source de l'image : https://avalanche.state.co.us/accidents/statistics-and-reporting/

Impacts du changement climatique



Source de l'image : https://www.independent.co.uk/news/world/europe/mont-blanc-deaths-french-alps-avalanche-skiing-a9227461.html

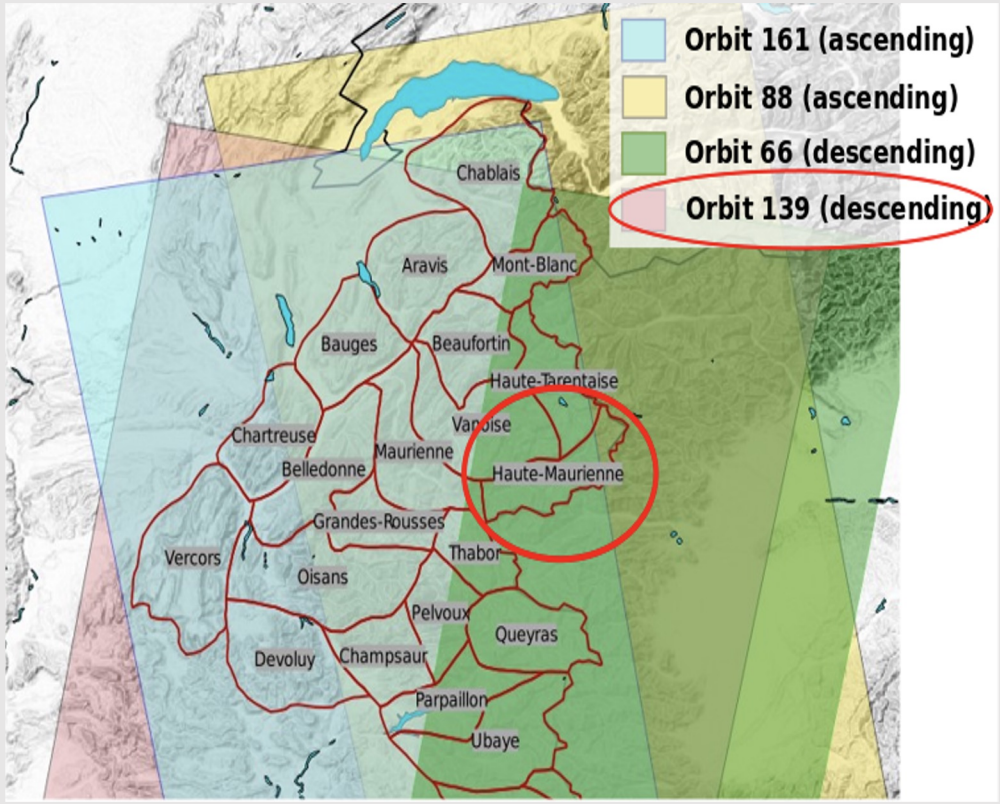
Les défis pour l'apprentissage automatique

- Fort déséquilibre des classes
  - Les avalanches sont des événements rares
- Données de référence étiquetées difficiles à obtenir
  - Accessibilité du terrain
  - Conditions météorologiques
  - Danger des avalanches

Le problème de la détection d'avalanches

- **Mesures de référence** in situ limitées
  - Météo-France

- **Images SAR** non étiquetées
  - Suivi des Alpes françaises en 2017-2018
  - Satellites Sentinel-1A et 1B
  - 4 caractéristiques :
    - Coefficients de rétrodiffusion à l'instant présent et précédent
    - Caractéristiques topologiques : pente et orientation



METEO FRANCE

Notre approche

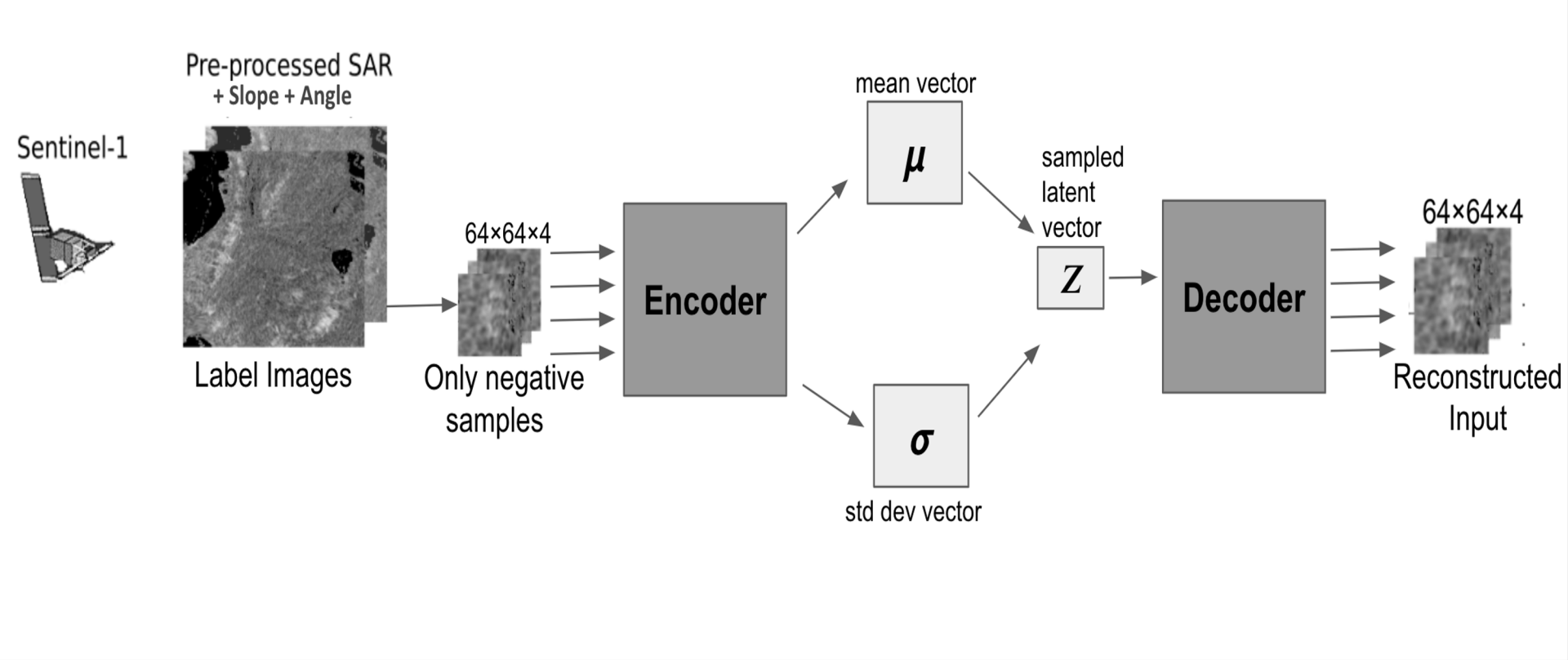
① Traiter une avalanche comme un événement rare, ou une anomalie

② Entraîner un autoencodeur variationnel (VAE) sur les exemples négatifs

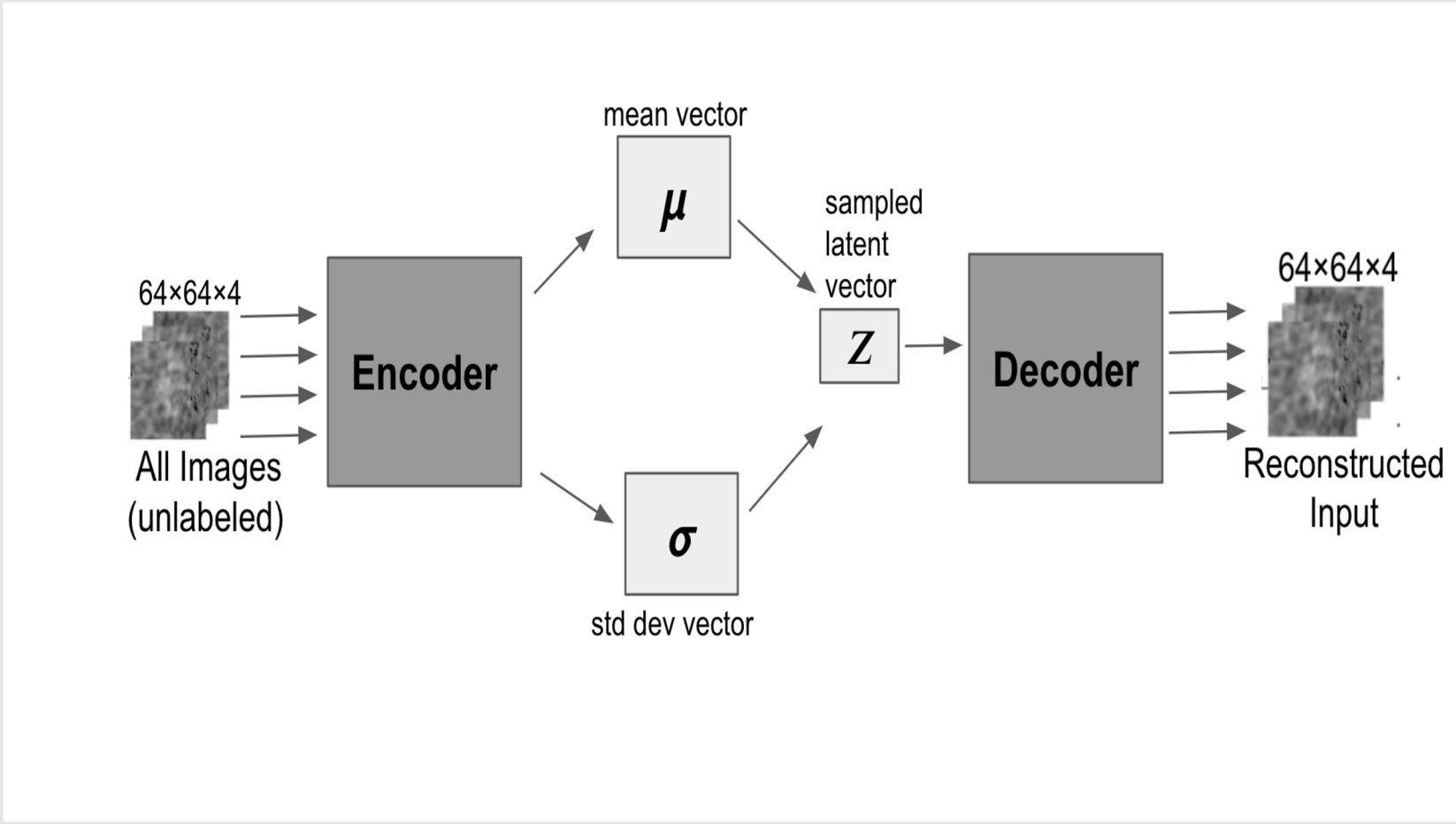
③ Seuiller l'erreur de reconstruction du VAE pour classer une nouvelle image

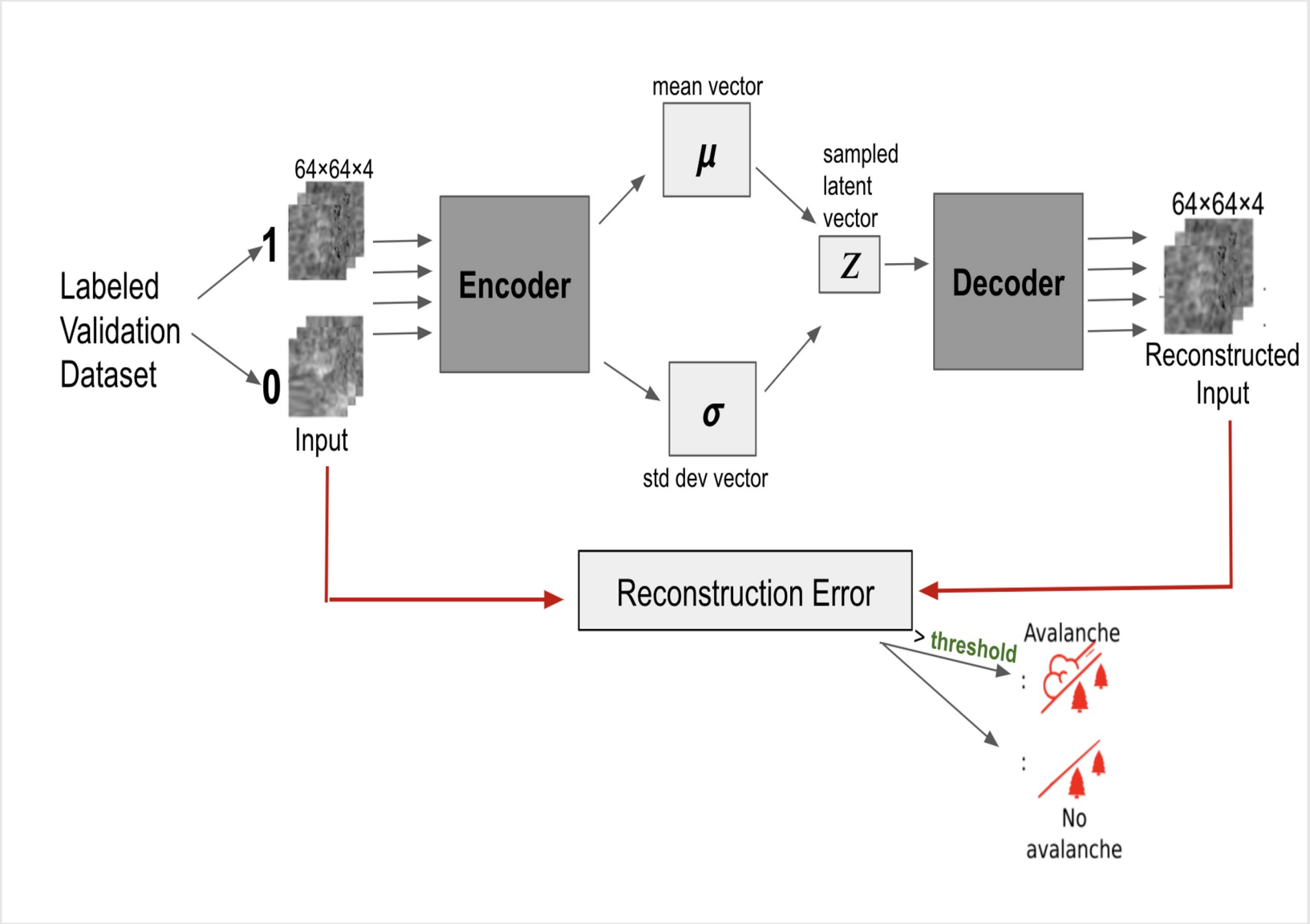
- Quand les données étiquetées sont rares, le VAE peut aussi être entraîné **sans** supervision !

VAE entraîné sur les exemples négatifs

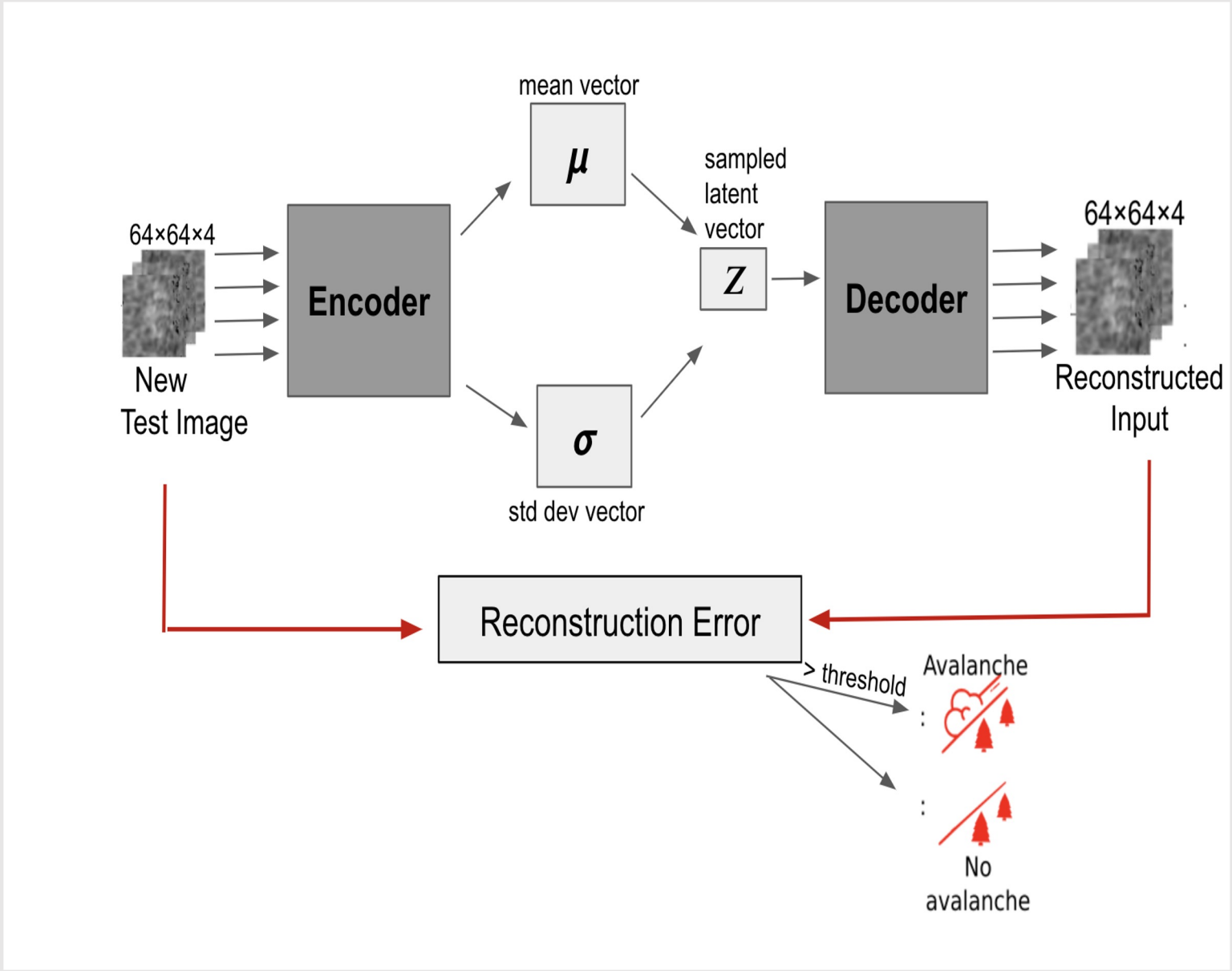


VAE entraîné sur TOUS les exemples



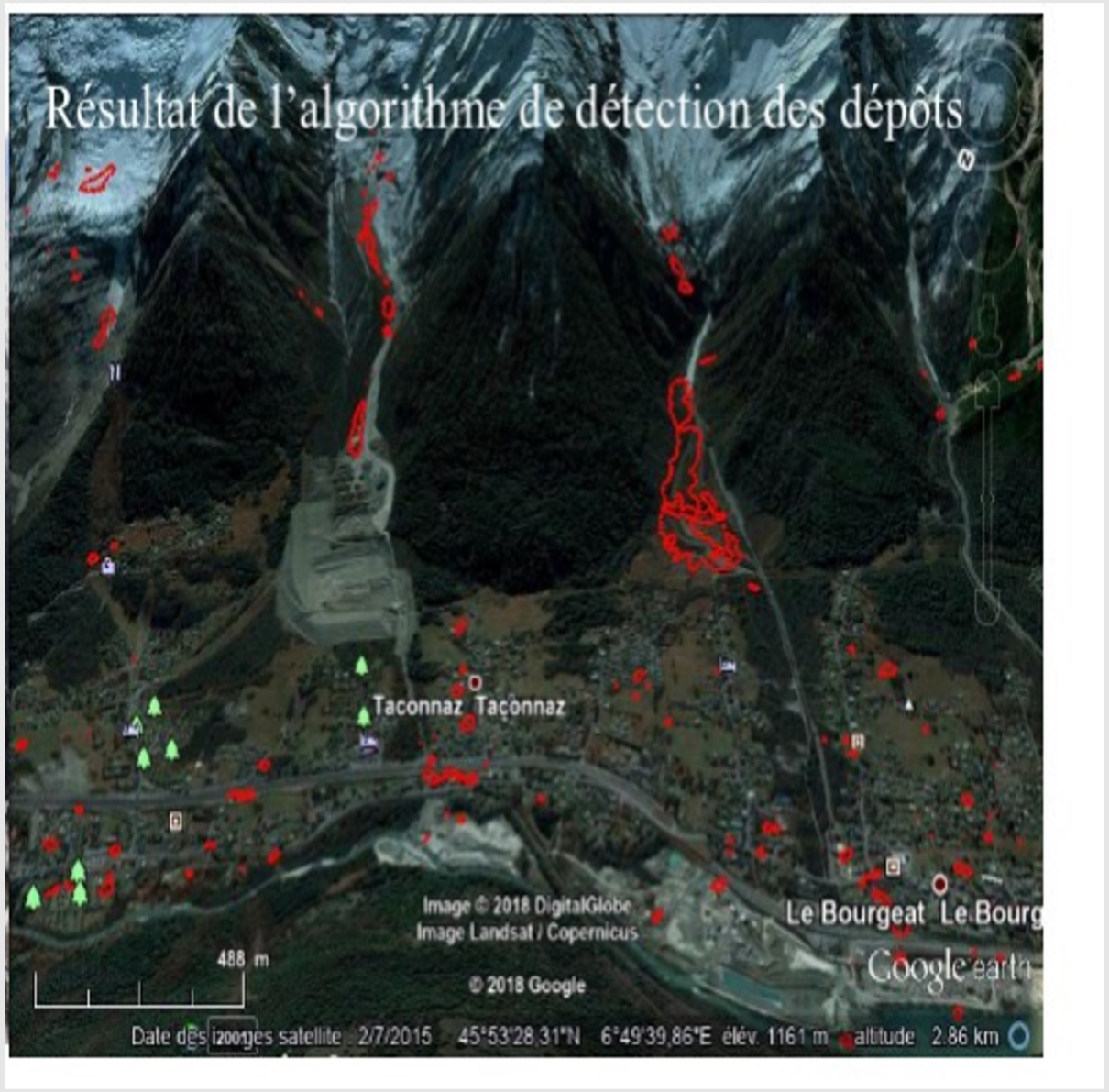


Régler le seuil de détection d'anomalies



Chaîne de détection d'anomalies

Méthode de référence : seuillage



[Karbou et al., International Snow Science Workshop 2018 & EGU 2018]

Évaluation

L'un des massifs les plus exposés (sur les 18 de « Toutes les Alpes »)

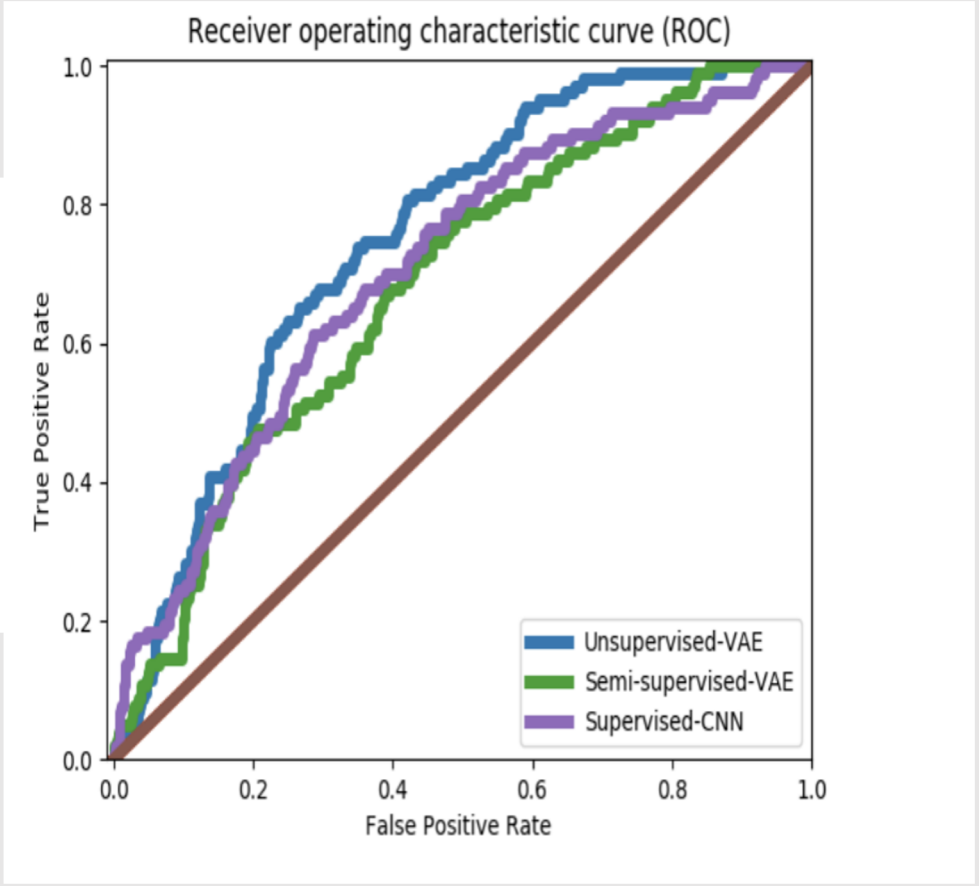
	All Alps		Haute Maurienne	
	Balanced Accuracy	F1-score	Balanced Accuracy	F1-score
Baseline	0.58	0.05	0.58	0.12
Supervised - CNN	0.53	0.10	0.53	0.12
Semi-supervised - VAE	0.59	0.11	0.6	0.23
Unsupervised - VAE	0.69	0.14	0.68	0.26

- Méthode de référence de la littérature sur la détection d'avalanches : seuillage [Karbou et al., ISSW 2018]
- Méthode de référence en apprentissage supervisé : réseau de neurones convolutif (CNN) entraîné sur un jeu artificiellement équilibré [Sinha et al., Climate Informatics 2019]

Évaluation

Method	AUC ROC
Supervised - CNN	70.7
Semi-supervised - VAE	68.3
Unsupervised - VAE	75

Courbes ROC pour la région de Haute-Maurienne



Résumé

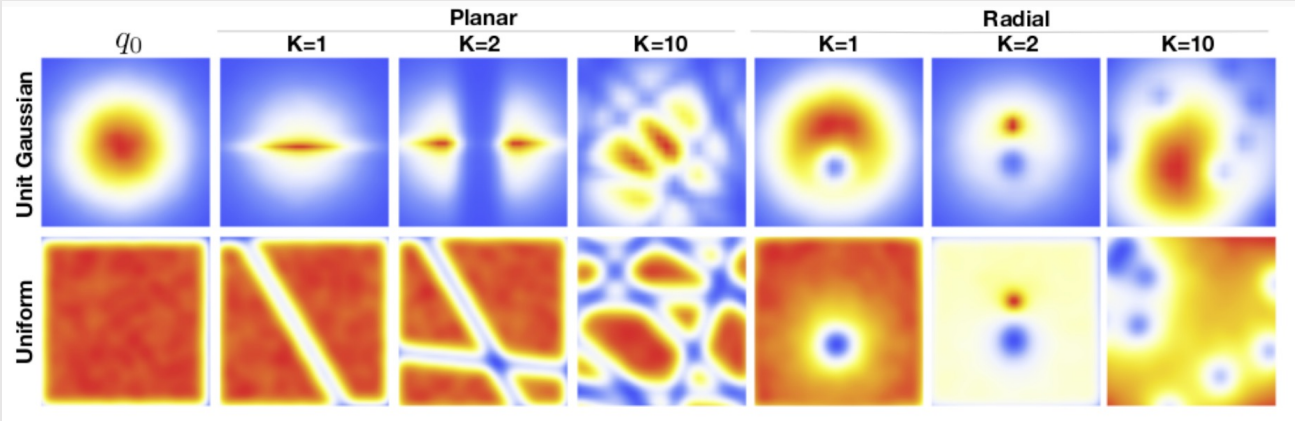
- Nous proposons une approche semi-supervisée pour détecter des événements rares quand les données étiquetées sont limitées
  - Cette chaîne semi-supervisée repose largement sur l'apprentissage non supervisé
- Elle peut être vue comme une forme de capteur virtuel

Partie III : Méthodes d'IA : l'IA générative

L'IA générative : une distinction

- Vous êtes sans doute habitués à ce que le terme « IA générative » désigne les grands chatbots et agents d'IA commerciaux comme ChatGPT, Gemini ou Claude.
  - Pour ces systèmes, les modèles d'IA générative ont été entraînés sur des vidéos et des textes collectés sur internet
  - Mais ce n'est **pas la SEULE définition** ni le seul usage de l'IA générative !
- Nous pouvons **entraîner des modèles d'IA générative à partir de zéro** pour nos problèmes
  - Pour les problèmes de climat et de météo, nous n'entraînerons PAS sur du texte et de la vidéo, mais sur les données pertinentes, p. ex. des champs de valeurs réelles mesurées à différentes latitudes, longitudes, dates, et éventuellement altitudes dans l'atmosphère ou profondeurs dans l'océan, dont :
    - Température
    - Rayonnement solaire
    - Humidité
- Aujourd'hui, nous allons découvrir les types de modèles génératifs que vous pourrez ensuite entraîner pour vos propres problèmes

Flots normalisants

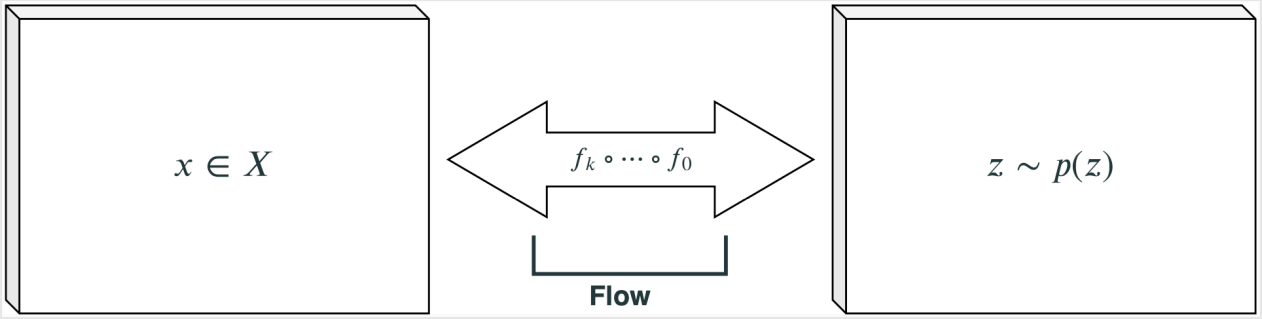


[Rezende & Mohamed, ICML 2015]

Peuvent être vus comme une extension du VAE au-delà de l'hypothèse gaussienne sur l'espace latent

Apprendre une série de transformations inversibles, $\{f_i\}$, à partir d'un a priori simple sur l'espace latent Z, pour obtenir des distributions plus informatives sur l'espace latent :

$$z_k = f_k \circ f_{k-1} \circ \cdots \circ f_1(z_0)$$

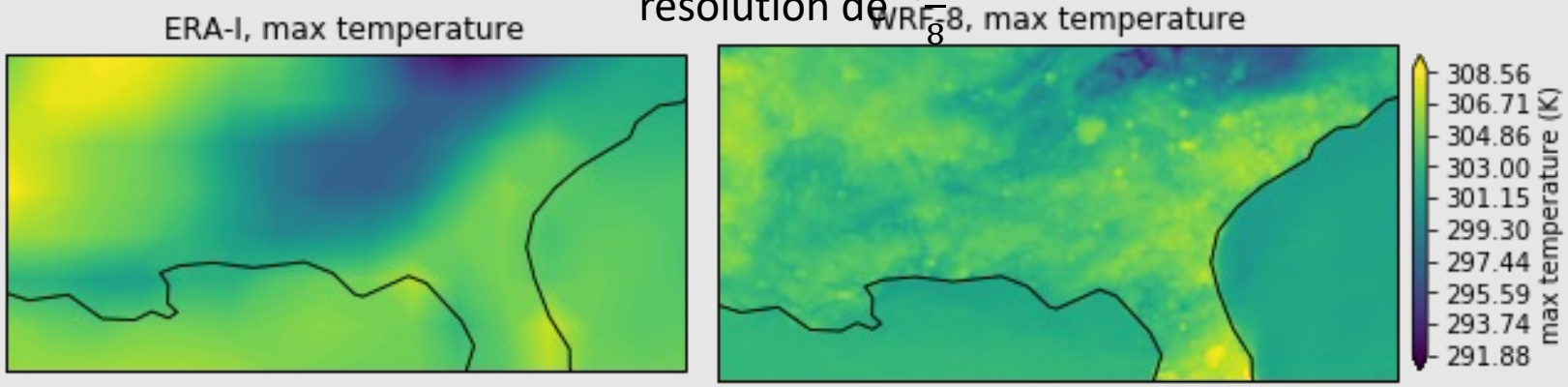


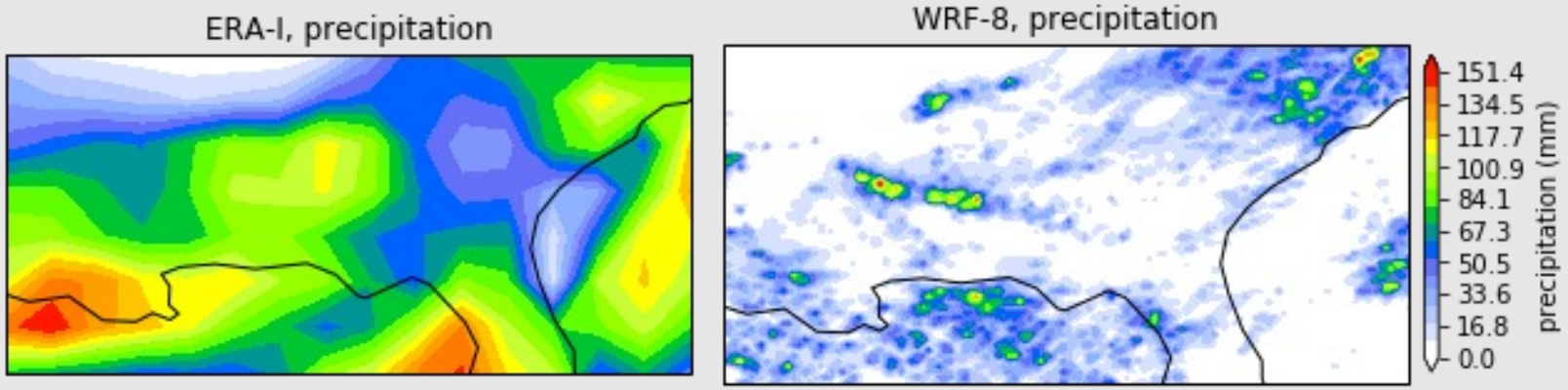
Flots normalisants : application à la descente d'échelle spatiale

[Groenke, Madaus, & Monteleoni, Climate Informatics 2020]

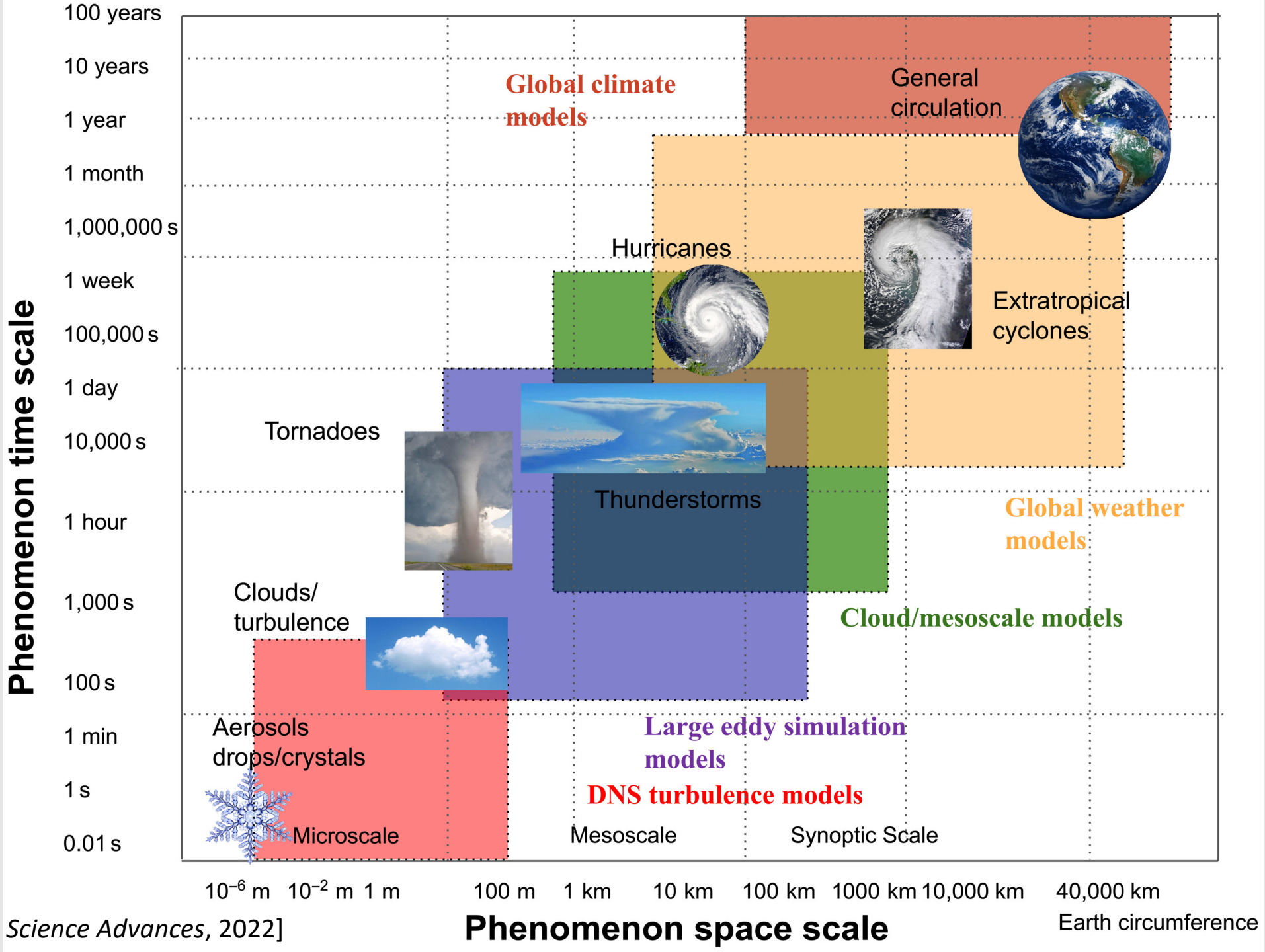
ERA : données de réanalyse, résolution de 1° ;

WRF : prévision d'un modèle météorologique numérique, résolution de $\frac{1}{8}°$









[Gettelman et al., *Science Advances*, 2022]

La descente d'échelle comme alignement de domaines

- Tâche d'alignement de domaines : étant données des variables aléatoires X, Y, apprendre une application f : X Y telle que, pour tout x_i X et y_i Y,

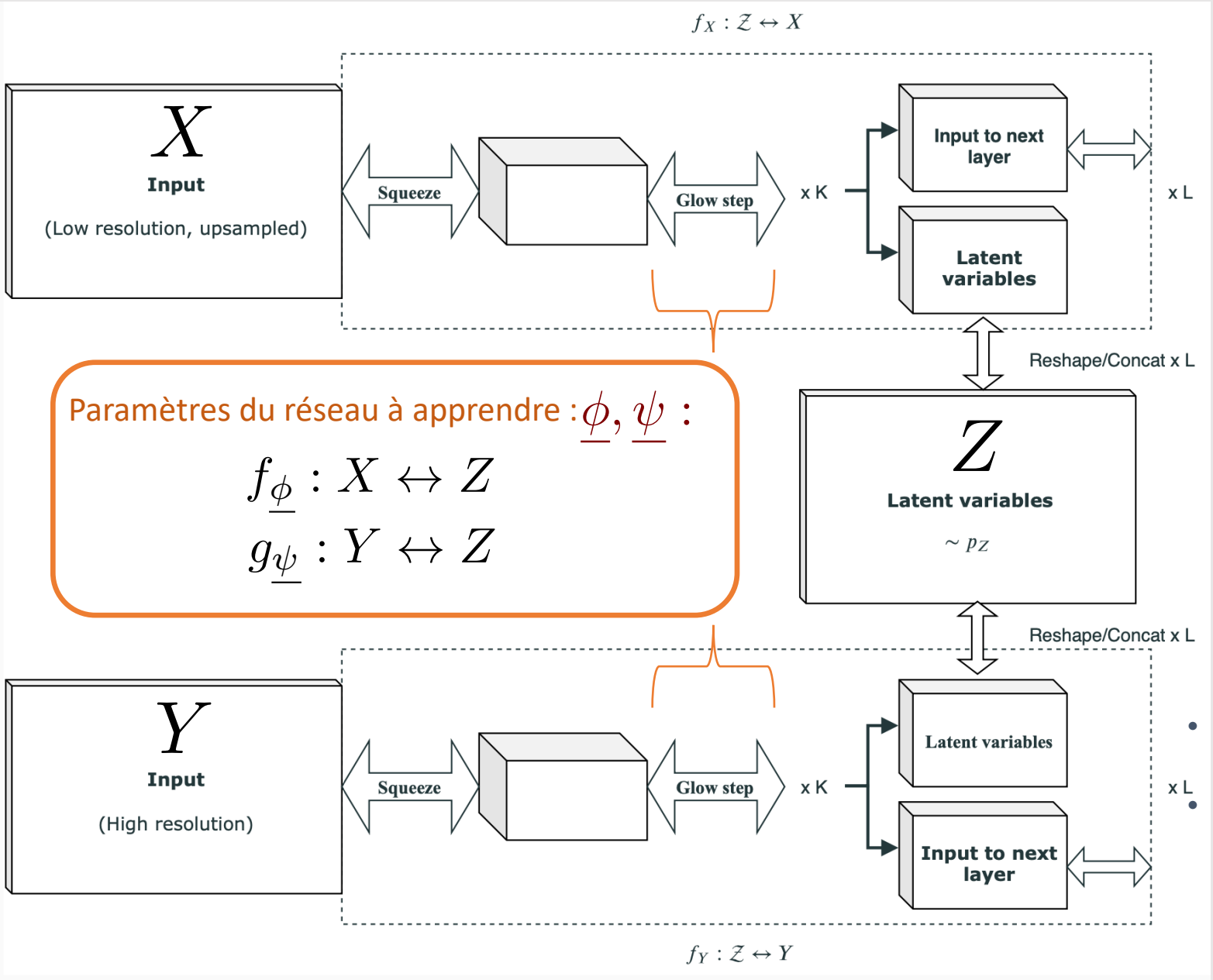
$$f(x_i) \sim P_Y \qquad\qquad f^{-1}(y_i) \sim P_X$$

- La descente d'échelle comme alignement de domaines
  - À partir d'échantillons i.i.d. à basse résolution (X) et à haute résolution (Y)
  - Apprendre la densité jointe de X et Y en supposant l'indépendance conditionnelle via un espace latent commun Z,

$$P_{XY}(x,y) = \int_{z \in Z} P_{XYZ}(x,y,z)dz = \int_{z \in Z} P(x|z)P(y|z)P_Z(z)dz$$

  - Modéliser $P(x|z), P(y|z)$ avec AlignFlow [Grover et al. 2020]
    - En partant d'un a priori simple sur P_Z, apprendre des flots normalisants
    - Aucun appariement entre exemples x et y n'est nécessaire !

Architecture de ClimAlign



- L'architecture suit AlignFlow [Grover et al., 2020]
- Flot normalisant : Glow [Kingma & Dhariwal, 2018]

Comparaison avec des références supervisées

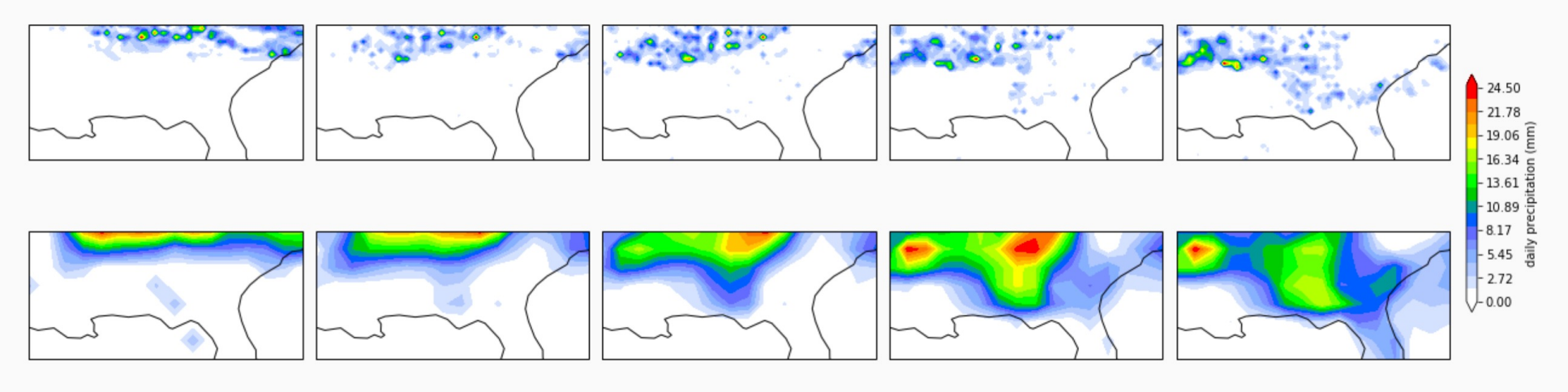
Température maximale journalière

Region	Method	RMSE	Bias	Corr
SE-US	BCSD	1.51 ± 0.15	-0.02 ± 0.21	**0.93** ± 0.05
	BMD-CNN	**1.30** ± 0.12	0.03 ± 0.13	0.90 ± 0.05
	ClimAlign (ours)	1.56 ± 0.13	**-0.005** ± 0.22	0.87 ± 0.06
P-NW	BCSD	1.54 ± 0.23	**0.01** ± 0.10	**0.95** ± 0.03
	BMD-CNN	**1.25** ± 0.14	-0.06 ± 0.05	0.93 ± 0.02
	ClimAlign (ours)	1.58 ± 0.18	0.03 ± 0.15	0.89 ± 0.04

Précipitations journalières

Region	Method	RMSE	Bias	Corr
SE-US	BCSD	27.32 ± 5.0	0.95 ± 1.4	0.39 ± 0.07
	BMD-CNN	**14.11** ± 2.18	-0.23 ± 0.47	**0.50** ± 0.10
	ClimAlign (ours)	18.40 ± 2.64	**0.08** ± 0.86	0.42 ± 0.07
P-NW	BCSD	8.90 ± 2.30	0.41 ± 0.26	0.61 ± 0.06
	BMD-CNN	**5.77** ± 0.72	**-0.18** ± 0.61	**0.70** ± 0.03
	ClimAlign (ours)	7.33 ± 0.69	0.54 ± 0.54	0.67 ± 0.03

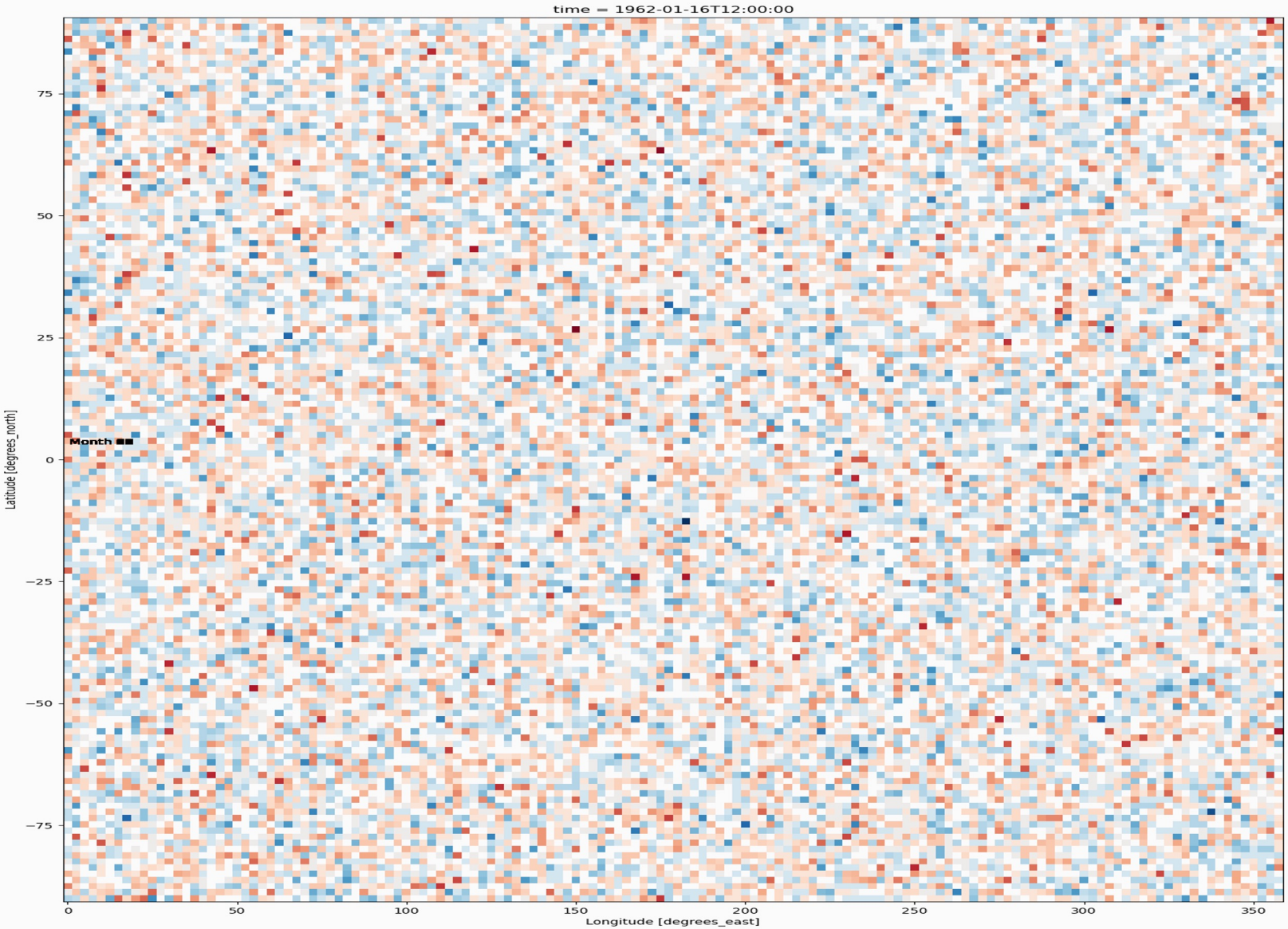
ClimAlign : descente d'échelle générative non supervisée

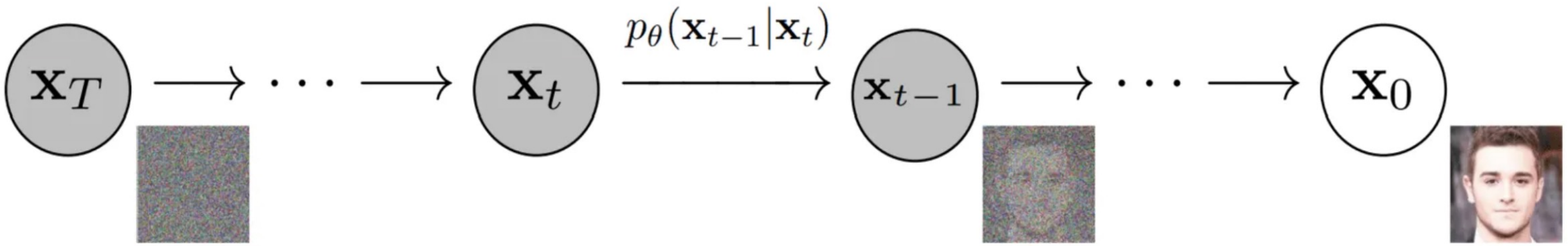


Technique générale de descente d'échelle par alignement de domaines avec flots normalisants [AlignFlow: Grover et al., AAAI 2020][Glow: Kingma & Dhariwal, NeurIPS 2018]

- **Non supervisée** : pas besoin de cartes appariées à basse et haute résolution
- **Générative** : on peut échantillonner la loi a posteriori sur la représentation latente OU échantillonner conditionnellement à une carte à basse (ou haute !) résolution
- **Interprétable**, p. ex. par interpolation

Modèles génératifs à base de flots





Entraînement par diffusion : apprendre à générer des images réalistes à partir de bruit aléatoire

Entraînement du modèle de diffusion

1. Créer une tâche prétexte à partir de données d'entraînement non étiquetées : pour chaque point d'entraînement x_0, créer une suite d'images synthétiques en ajoutant du bruit gaussien à l'image précédente, jusqu'à obtenir une image indiscernable du bruit gaussien.

2. Faire ensuite un apprentissage « supervisé » sur les paires (x_{t-1}, x_t) pour apprendre $p_\theta(x_{t-1} \mid x_t)$

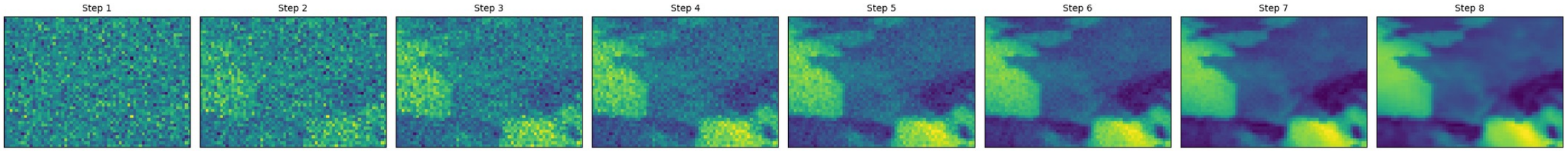
Inférence, c.-à-d. comment générer des échantillons proches des données d'entraînement

1. Tirer x_T au hasard (gaussien).

2. Tirer T fois selon p_θ, en conditionnant à chaque fois sur l'échantillon précédent

Modèles génératifs à base de flots

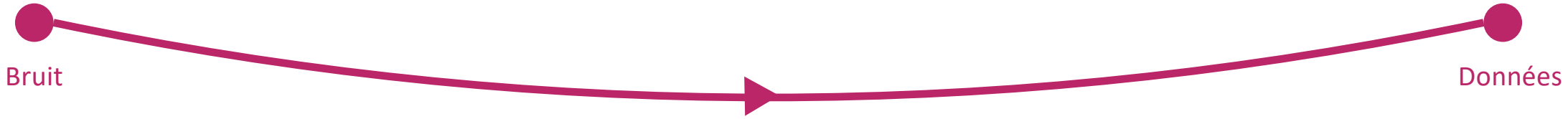
Modèle génératif à base de flots : apprend un transport (par un réseau de neurones) d'une distribution facile à échantillonner (p. ex. un bruit gaussien) vers la distribution des données.



Diffusion : un « débruiteur » inverse un processus de bruit stochastique, en débruitant progressivement – trajectoire d'EDS



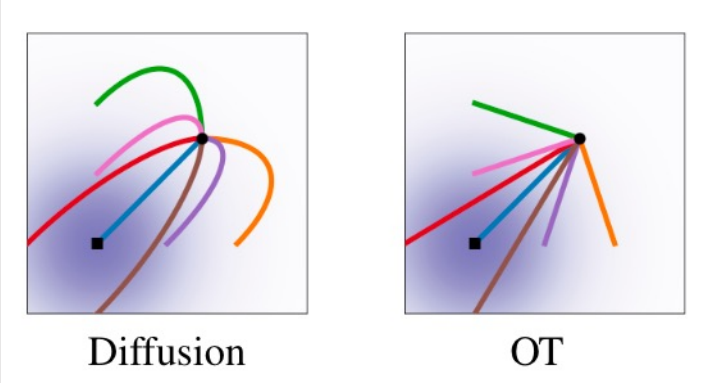
Flow matching [Lipman et al., 2023] : apprend un flot déterministe bruit → données – trajectoire d'EDO



Crédit de la diapositive : Julie Keisler (INRIA)

Flow matching vs diffusion

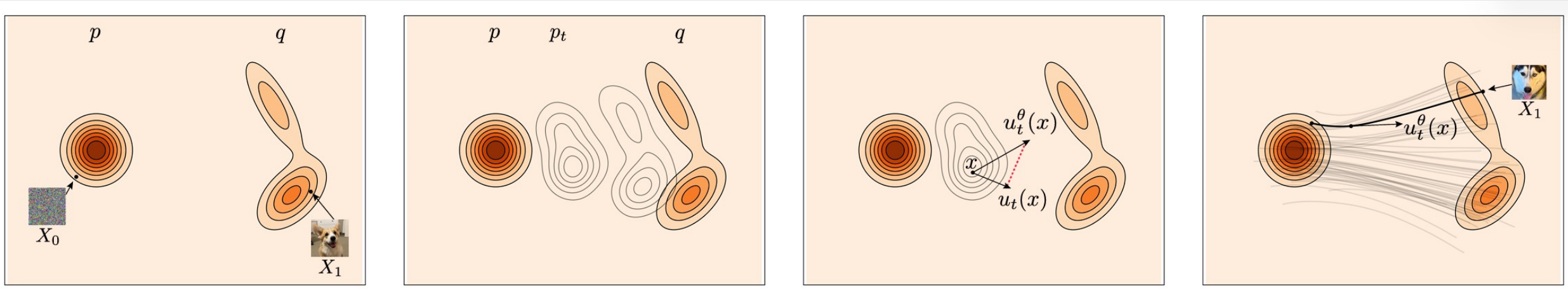
Intuition : créer des chemins plus droits entre échantillons de données et échantillons de bruit :



Résultats : des chemins plus droits dans les flots de probabilité, qui permettent d'échantillonner en moins de pas de diffusion accélérations

Généralité : le flow matching peut relier deux distributions quelconques (ou interpolants stochastiques), contrairement à la diffusion

Flow Matching



(a) Data. **(b)** Path design. **(c)** Training. **(d)** Sampling.

Figure 2 *The Flow Matching blueprint.* (a) The goal is to find a flow mapping samples X_0 from a known source or noise distribution q into samples X_1 from an unknown target or data distribution q. (b) To do so, design a time-continuous probability path $(p_t)_{0\leq t\leq 1}$ interpolating between $p := p_0$ and $q := p_1$. (c) During training, use regression to estimate the velocity field u_t known to generate p_t. (d) To draw a novel target sample $X_1 \sim q$, integrate the estimated velocity field $u_t^\theta(X_t)$ from $t = 0$ to $t = 1$, where $X_0 \sim p$ is a novel source sample.

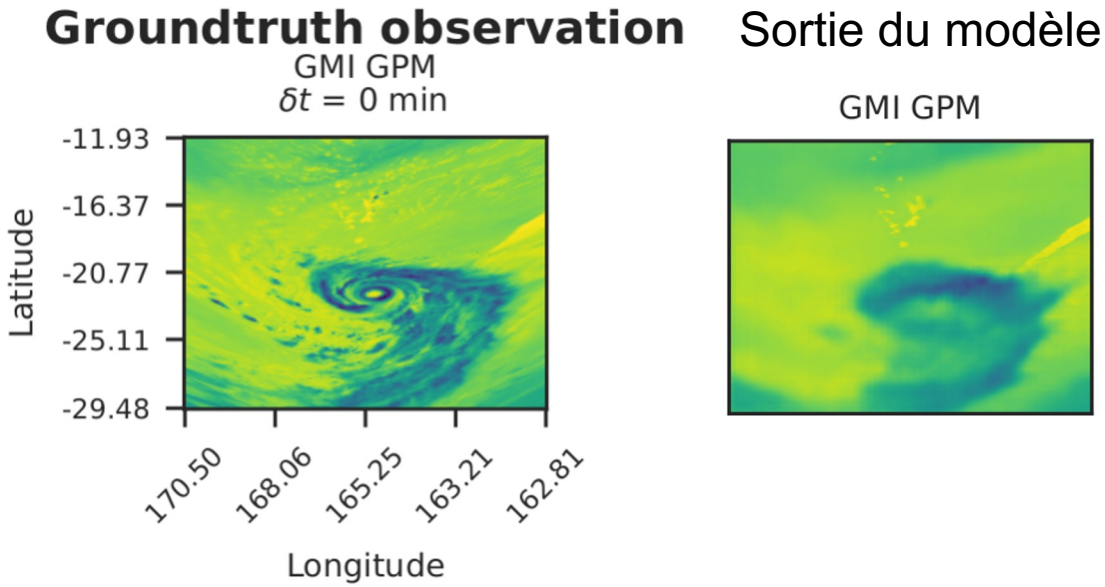
Lipman et al. (2024) Flow Matching Guide and Code

Pourquoi utiliser l'IA générative pour le climat et la météo ?

Les modèles d'AA déterministes optimisent souvent la performance **moyenne**

Entraînés à minimiser une perte moyenne, p. ex. la RMSE

➔ **Des** sorties de modèle floues et non physiques (p. ex. images, champs, trajectoires de tempêtes)



Sortie du modèle d'AA

Sorties du modèle d'AA, échantillonnées après entraînement la même ossature de modèle, **en mode génératif**

Pourquoi utiliser l'IA générative pour le climat et la météo ?

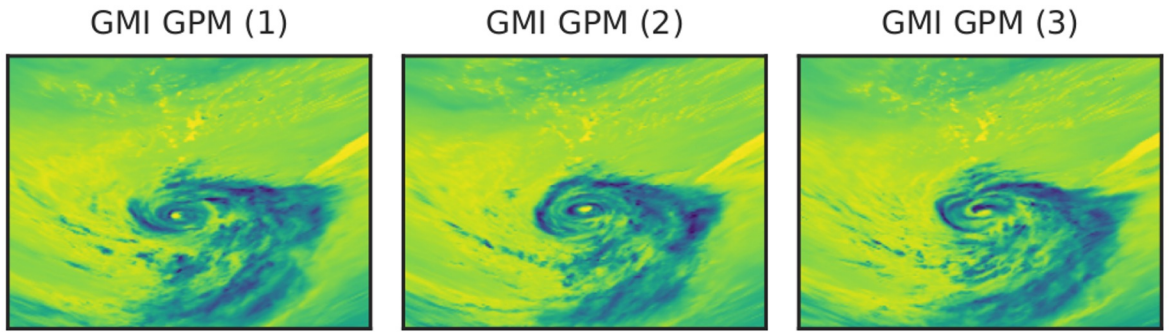
Pour de nombreux problèmes de climat, de météo et de géosciences

- Prévision météorologique, émulation du climat, génération de scénarios
- Estimation des extrêmes ou des événements rares, quantiles
- Descente d'échelle vers des échelles locales en temps ou en espace
- Quantification des risques

L'IA générative apporte des avantages

- Meilleur réalisme physique
  - Meilleure précision des hautes fréquences
- Ensembles probabilistes
  - Quantification de l'incertitude
  - Informations riches et réalistes pour les décideurs
- Génération de données synthétiques

Échantillons d'un **modèle génératif**



L’IA générative pour :

✓ Prévision probabiliste

- **ArchesWeatherGen** : prévision météo efficace entraînée à 1,5°
- **ArchesClimate** : émulation, interpolation et extrapolation du climat décennal

❑ Fusion de données avec alignement probabiliste de domaines

- **MotifGen** : un modèle génératif, multisource et auto-supervisé pour apprendre à partir de données géospatiales

❑ L’alignement de domaines probabiliste et la descente d’échelle

- **ClimAlign** : descente d’échelle générative non appariée avec alignement de domaines
- **SerpentFlow** : alignement de domaines génératif non apparié par décomposition en structure partagée (avec application à la descente d’échelle)

Les défis de l'apprentissage à partir de données multisources

Pléiades Neo

30 cm

Swath: 14 km

Spectral bands:
R, G, B, Deep Blue, Red Edge, NIR, Panchromatic

Revisit capacity:
Daily, anywhere

Daily acquisition capacity:
1,000,000 km²

Pléiades

50 cm

Swath: 20 km

Spectral bands:
R, G, B, NIR, Panchromatic

Revisit capacity:
Daily, anywhere

Daily acquisition capacity:
700,000 km²

Vision-1

87 cm

Swath: 20.8 km

Spectral bands:
R, G, B, NIR, Panchromatic

Revisit capacity:
Daily to 8 days, anywhere depending on latitude and partner satellites

Daily acquisition capacity:
20,000 km²

SPOT

1.5 m

Swath: 60 km

Spectral bands:
R, G, B, NIR, Panchromatic

Revisit capacity:
Up to daily, anywhere

Daily acquisition capacity:
3,000,000 km²

Radar constellation

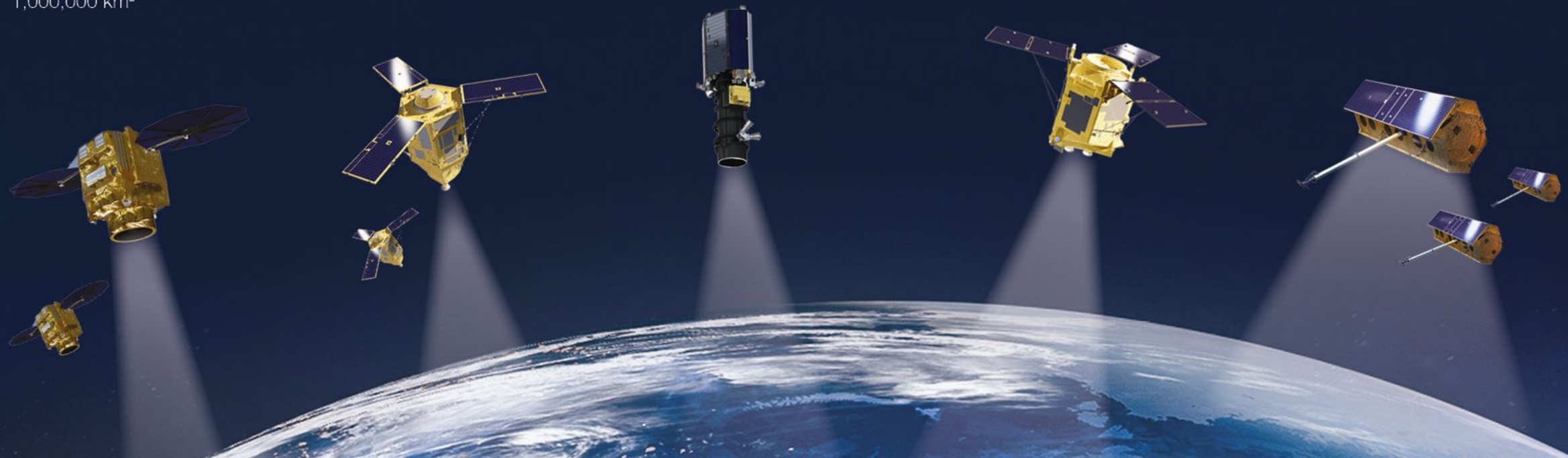
From 25 cm to 40 m

Swath: 270 km

Imaging modes:
SpotLight, StripMap, ScanSAR

Revisit capacity:
Daily for most latitudes

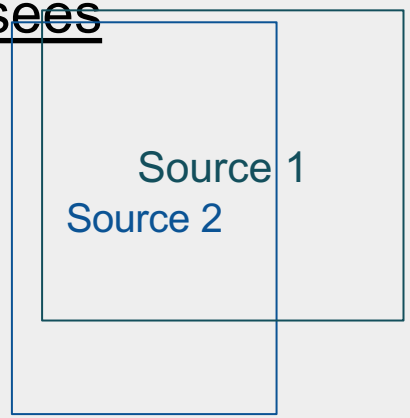
Daily acquisition capacity:
Up to 5,400,000 km²
(Wide ScanSAR Mode)



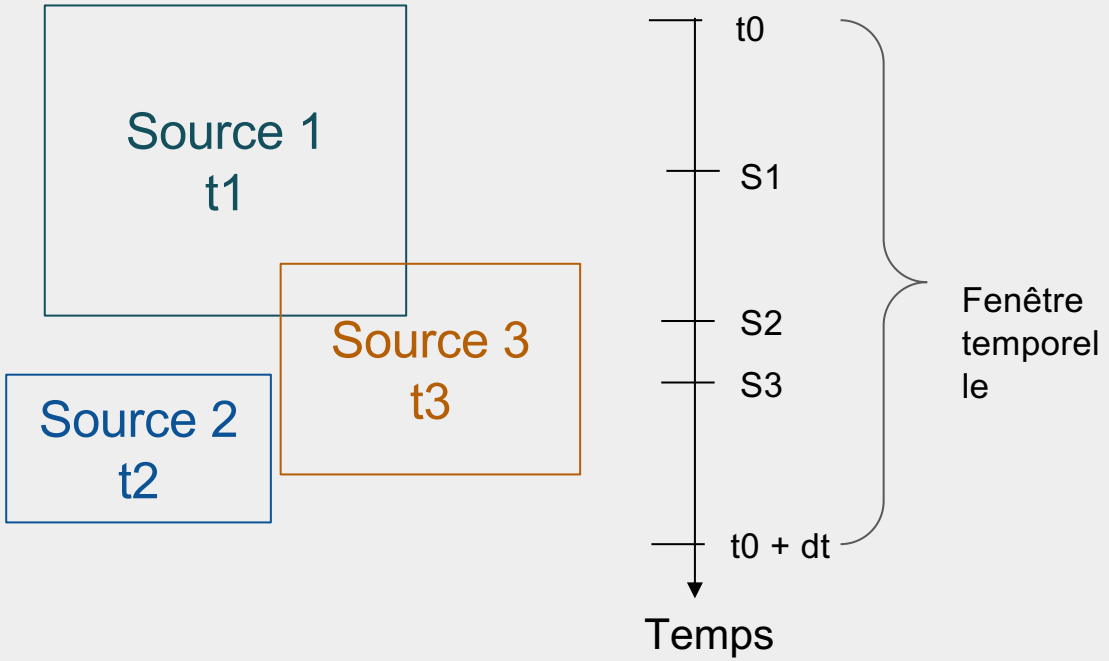
L'IA pour la **fusion** de données multisources (désalignées)

[Dauvilliers & Monteleoni, *Environmental Data Science*, 2025]

Traditionnellement : n'entraîner les modèles que sur des données colocalisées



Notre modèle d'IA (MOTIF) peut apprendre de **toutes les sources de données**, même de données désalignées dans le temps et l'espace



Une architecture multisource pour la fusion de données géospatiales

AA fondamental multimodal	**Apprentissage de représentations géospatiales**
• Plonger les **valeurs** (p. ex. texte, images) • Conditionnement (« chien ou chat »)	• Plonger les **positions** • Entraînement auto-supervisé à grande échelle
Modèle DiT (2022)	SatCLIP (2025), AlphaEarth (2025)

MOTIF : *Transformer multisource à attention factorisée*

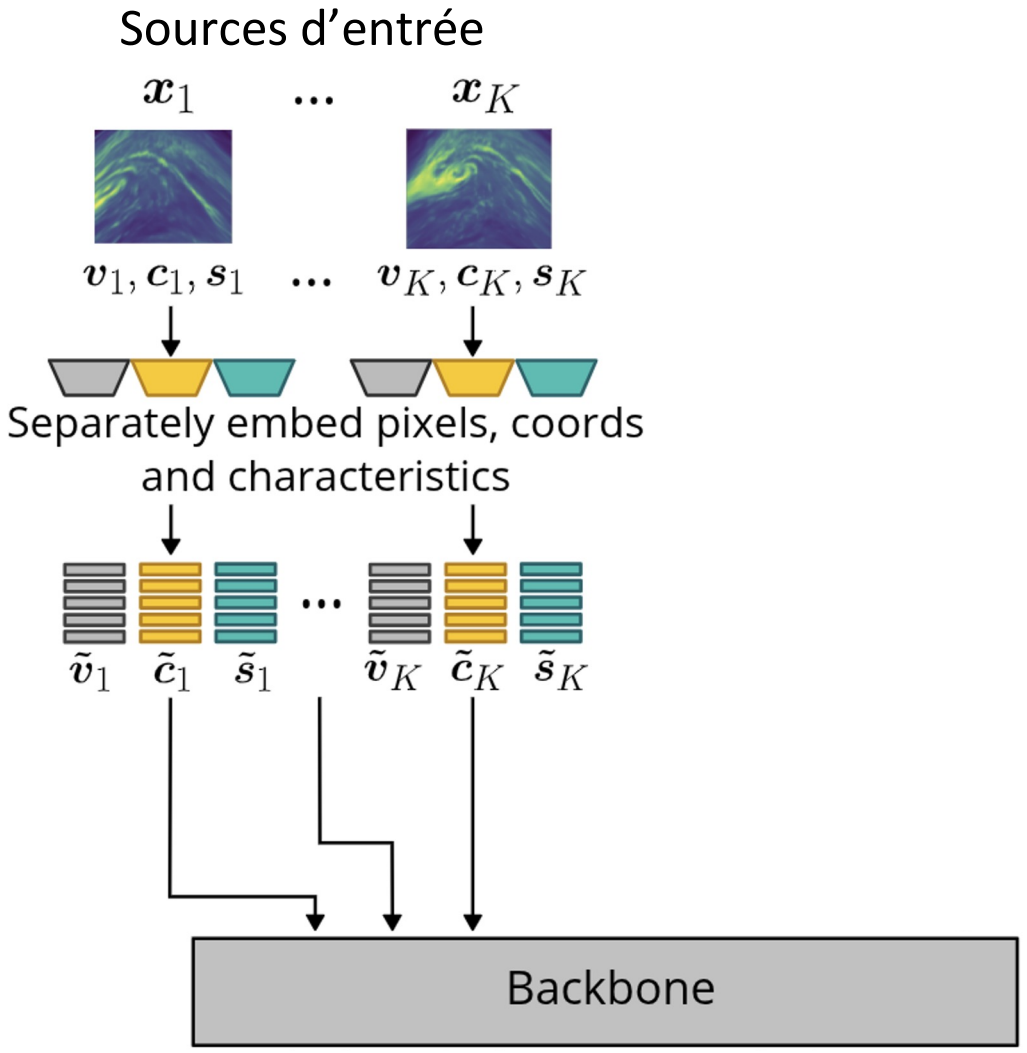
- Plonger séparément les **valeurs** et les **coordonnées**
- Conditionner sur les caractéristiques des sources
- Pré-entraînement auto-supervisé

[Dauvilliers & Monteleoni, Environmental Data Science, 2025]



Chaîne d'entraînement

Plonger séparément :
- les valeurs,
- les coordonnées,
- les caractéristiques des sources



Plongement positionnel relatif

- Plongement positionnel relatif habituel des Transformers (Shaw et al., 2018)

$$V_{out} = \frac{Softmax(QK^T + Q_{pos}K_{pos}^T)}{\sqrt{d}}$$

- Ici, les coordonnées spatio-temporelles servent au plongement positionnel :

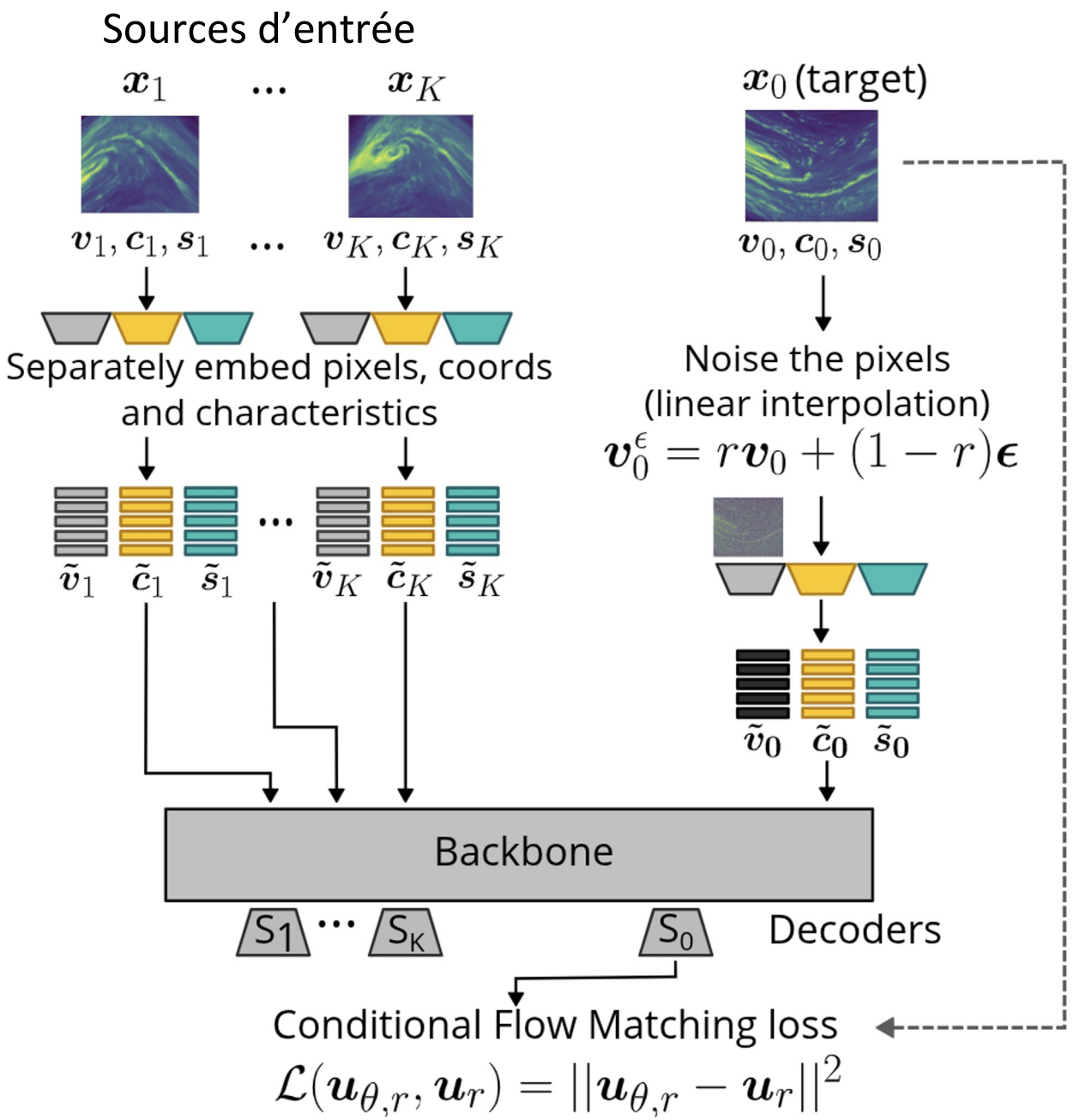
$$\tilde{v}_{out} = \frac{Softmax\left(Q_V K_V^T + Q_C K_C^T\right)}{\sqrt{d}} \tilde{v}_{in}$$

Valeurs (p. ex. pixels)

Coordonnées spatio-temporelles

Chaîne d'entraînement

Plonger séparément :
- les valeurs,
- les coordonnées,
- les caractéristiques des sources



Tâche d'entraînement auto-supervisé : reconstruire une source masquée

Version générative entraînée par flow matching

[Dauvilliers & Monteleoni, Climate Informatics 2026]

Application aux cyclones tropicaux

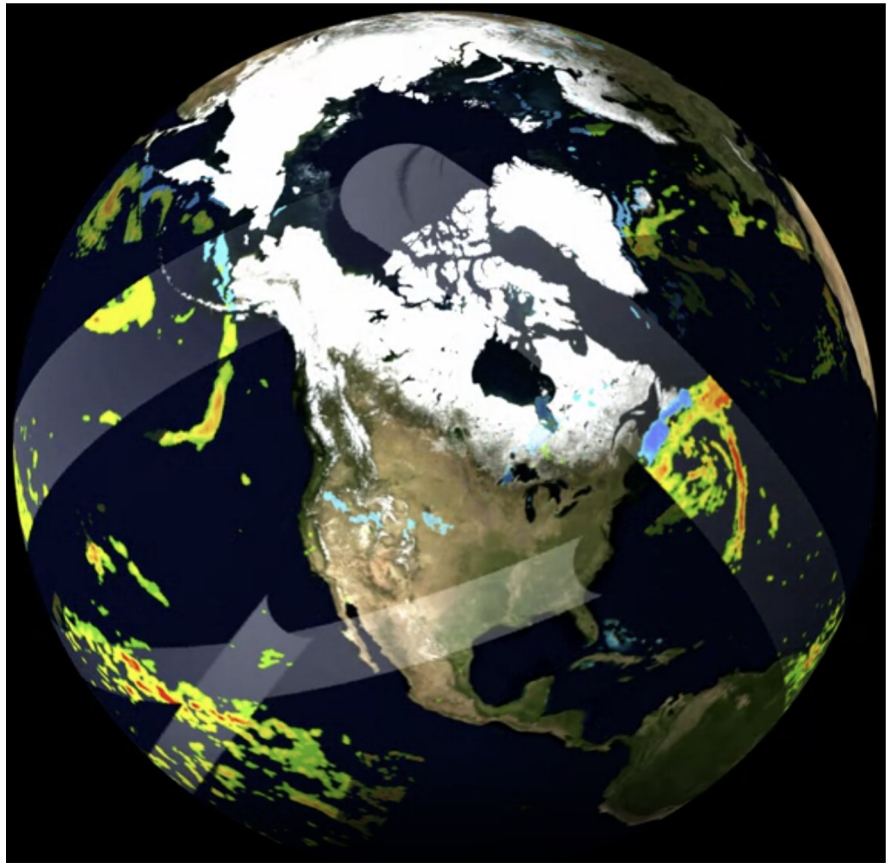
La constellation GPM

GMI (instrument de référence)



⇒ Peut-on faire de l'interpolation spatio-temporelle des images GMI ?

La constellation GPM, NASA



La constellation GPM couvre la Terre,
NASA's Scientific Visualization Studio

156

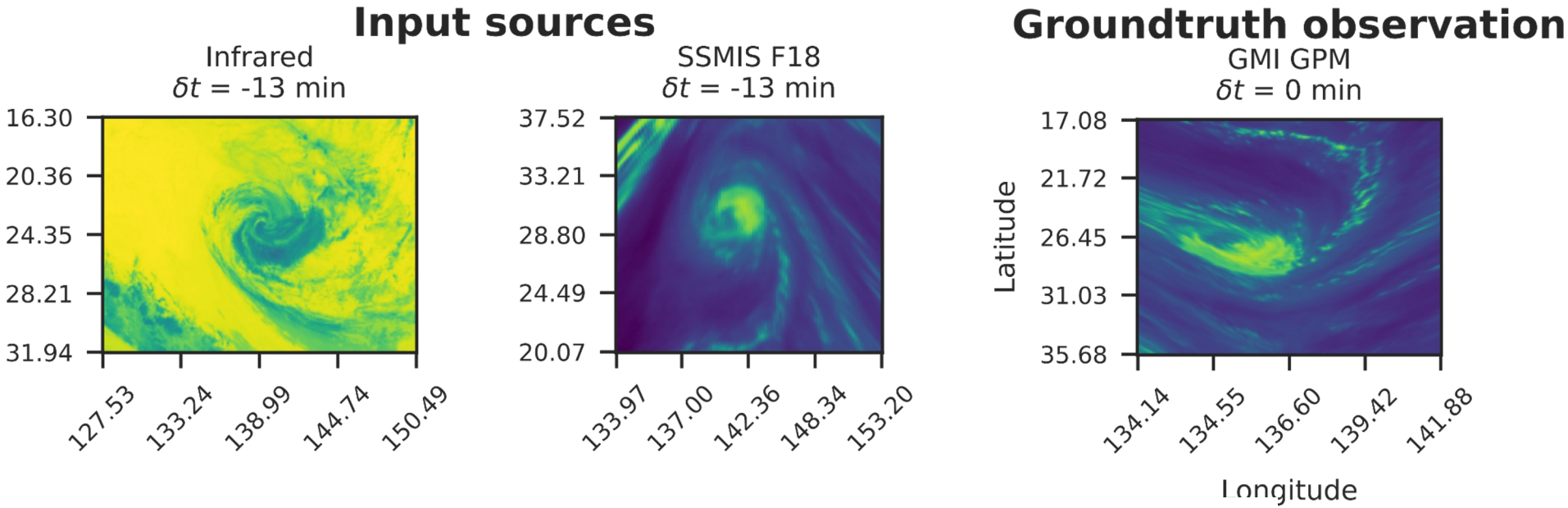
Jeu de données

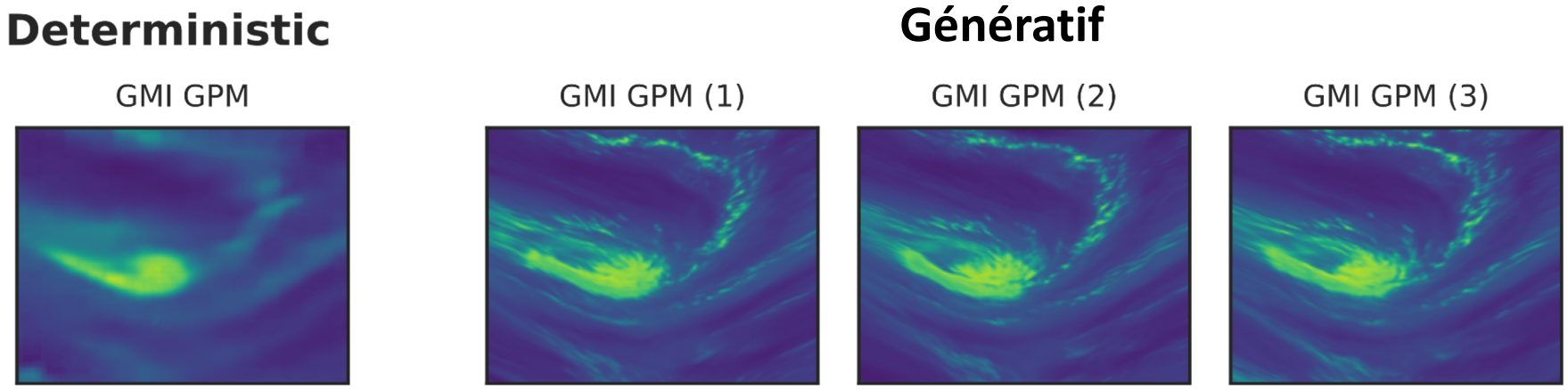
TC-PRIMED (Razin et al. 2023)

- ~3 000 tempêtes, de 1987 à 2024, tous bassins
- 2005-07-14-21-23 pour la validation, 2006-08-15-22-24 pour le test
- 11 capteurs micro-ondes
- Observations infrarouges

Fenêtres d’interpolation de 6 heures

Résultats : saisit mieux la structure fine que le modèle déterministe





Chaînes d'entraînement comparées pour la version générative

- Entraînement supervisé

Capteurs non GMI —Modèle→ GMI

⇒ ~6 000 échantillons

- Entraînement auto-supervisé

Tout capteur → Maximise les données d'entraînement disponibles → Tout capteur

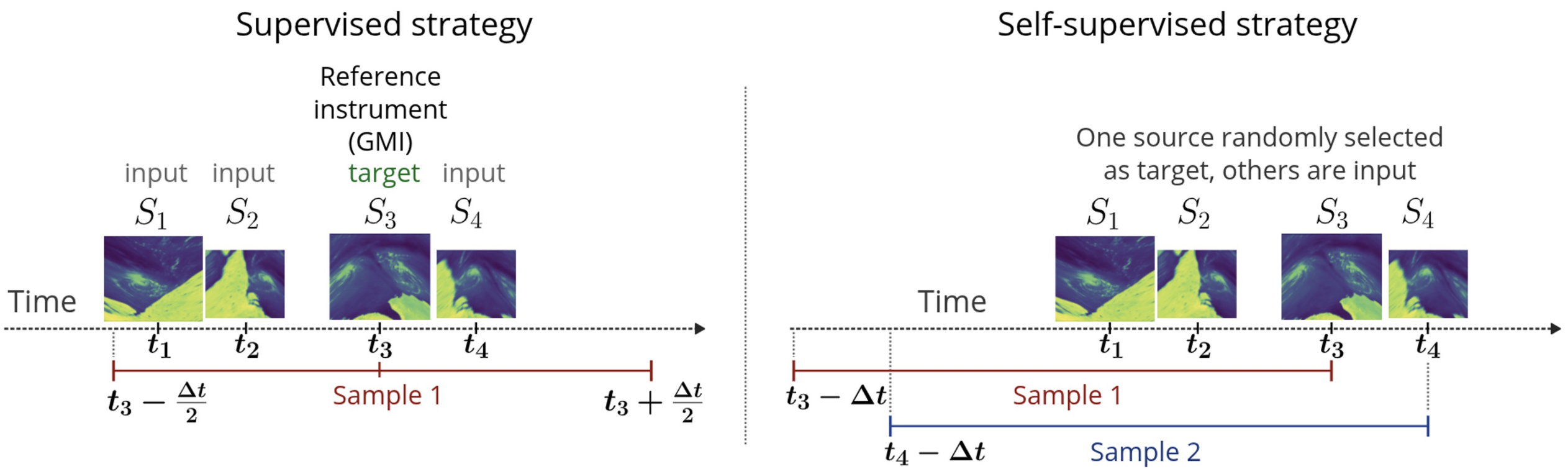
⇒ ~48 000 échantillons

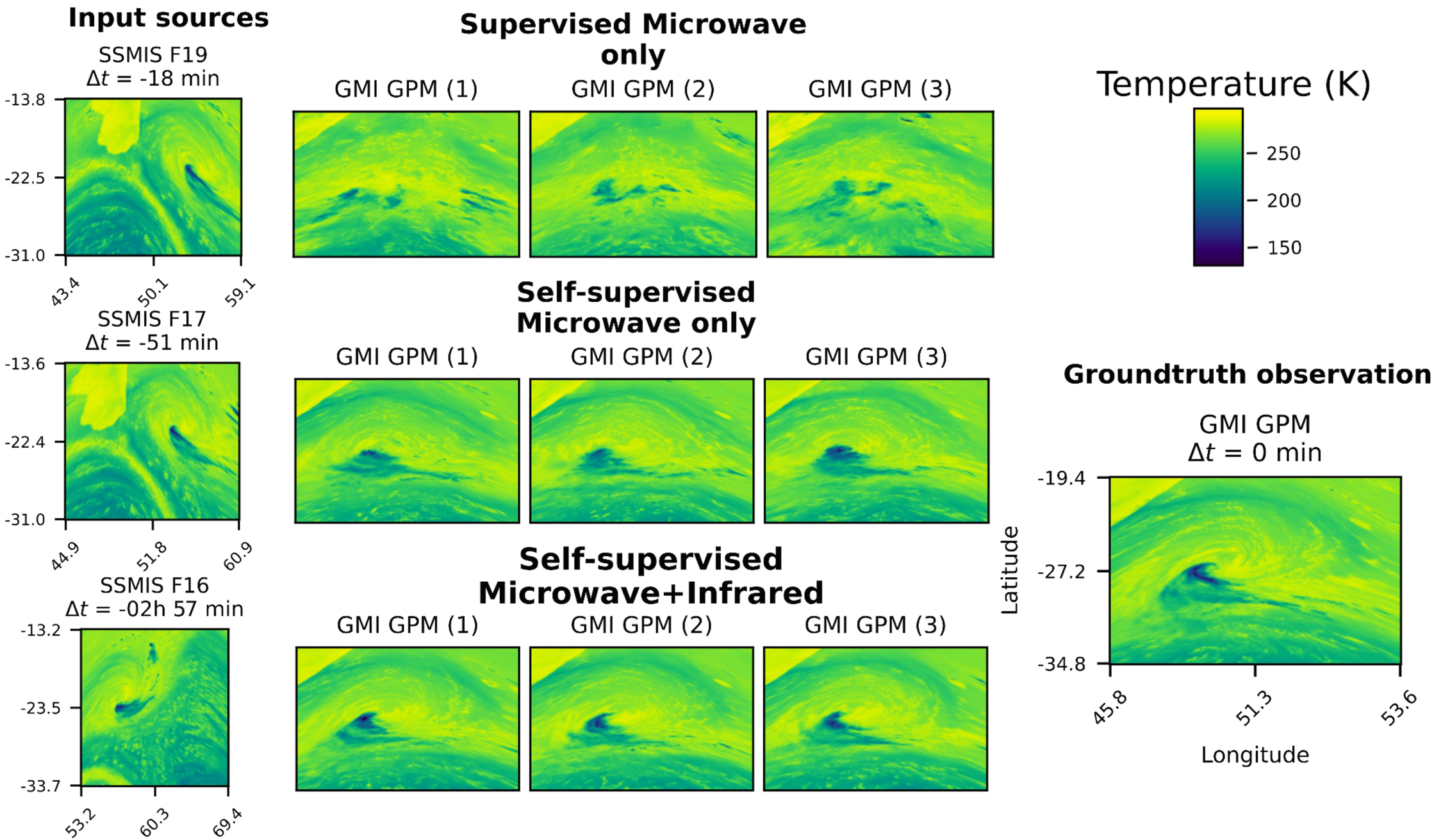
- Auto-supervisé + infrarouge

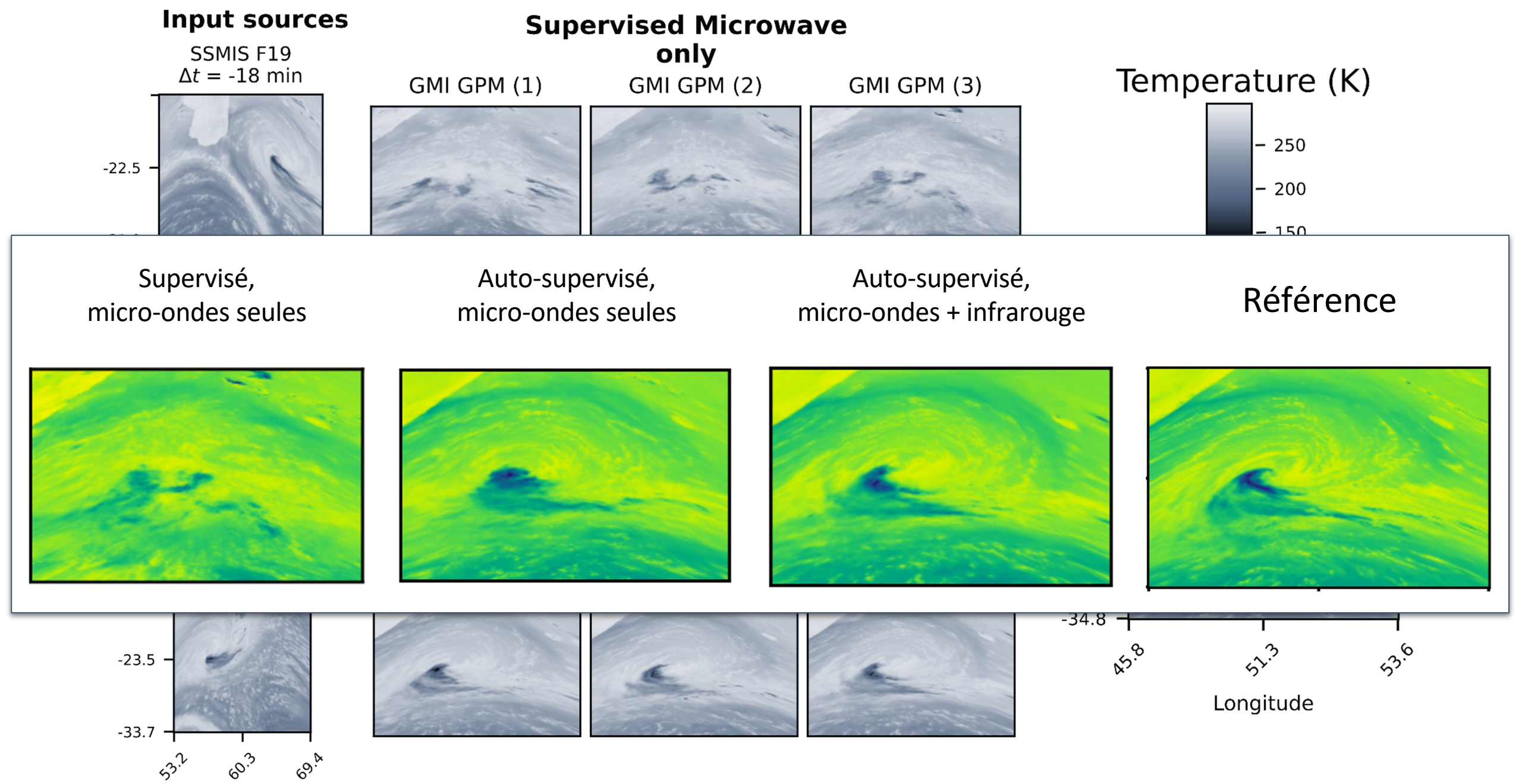
Tout capteur + Infrarouge → Utilise davantage de modalités d'entrée → Tout capteur

⇒ ~44 000 échantillons

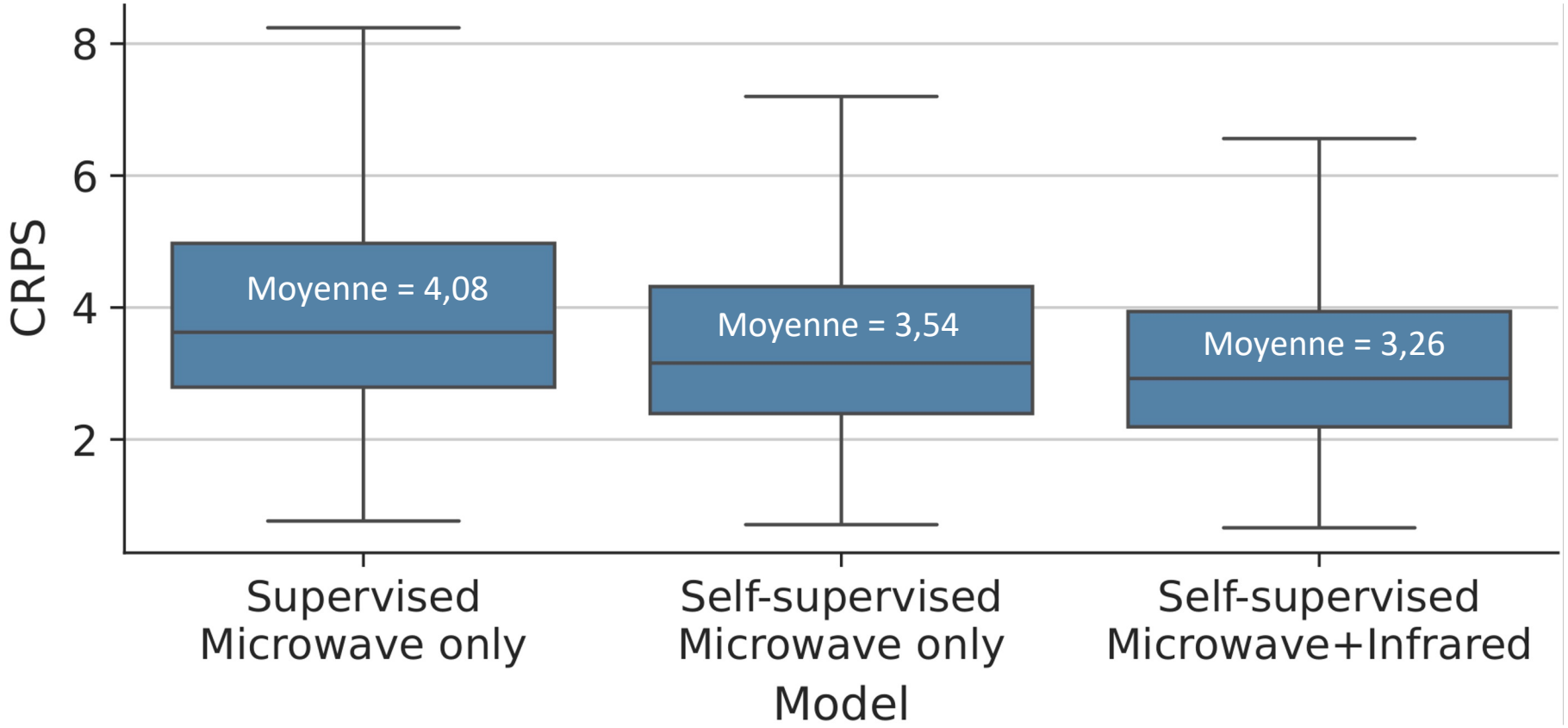
Tâches d'entraînement




Input sources
SSMIS F19
Δt = -18 min
-13.8
-22.5
-31.0
43.4
50.1
59.1
SSMIS F17
Δt = -51 min
-13.6
-22.4
-31.0
44.9
51.8
60.9
SSMIS F16
Δt = -02h 57 min
-13.2
-23.5
-33.7
53.2
60.3
69.4
Supervised Microwave only
GMI GPM (1)
GMI GPM (2)
GMI GPM (3)
Self-supervised Microwave only
GMI GPM (1)
GMI GPM (2)
GMI GPM (3)
Self-supervised Microwave+Infrared
GMI GPM (1)
GMI GPM (2)
GMI GPM (3)
Temperature (K)
250
200
150
Groundtruth observation
GMI GPM
Δt = 0 min
-19.4
-27.2
-34.8
45.8
51.3
53.6
Latitude
Longitude


Input sources
SSMIS F19
Δt = -18 min
Supervised Microwave only
GMI GPM (1)
GMI GPM (2)
GMI GPM (3)
Temperature (K)
250
200
150
-22.5
-23.5
-33.7
-34.8
53.2
60.3
69.4
45.8
51.3
53.6
Longitude
Supervisé, micro-ondes seules
Auto-supervisé, micro-ondes seules
Auto-supervisé, micro-ondes + infrarouge
Référence

Version générative : ablations



Résultats

Model type	Training strategy	Input sources	CRPS	RMSE	MAE	MAPE (%)
Deterministic	Self-supervised	MW+IR	-	9.32 [9.19, 9.45]	5.15 [5.07, 5.22]	2.26 [2.22, 2.30]
Flow matching	Supervised Référence	MW only	4.08 [4.02, 4.14]	10.92 [10.77, 11.07]	6.29 [6.19, 6.38]	2.79 [2.73, 2.84]
Flow matching	Self-supervised	MW only	3.54 [3.49, 3.60]	9.72 [9.58, 9.85]	5.31 [5.23, 5.39]	2.34 [2.30, 2.39]
Flow matching	Self-supervised	MW+IR	**3.26** [3.21, 3.31]	**9.01** [8.89, 9.13]	**4.76** [4.69, 4.83]	**2.09** [2.05, 2.13]

Les crochets indiquent les intervalles de confiance à 95 %

Score de probabilité classé continu (CRPS)

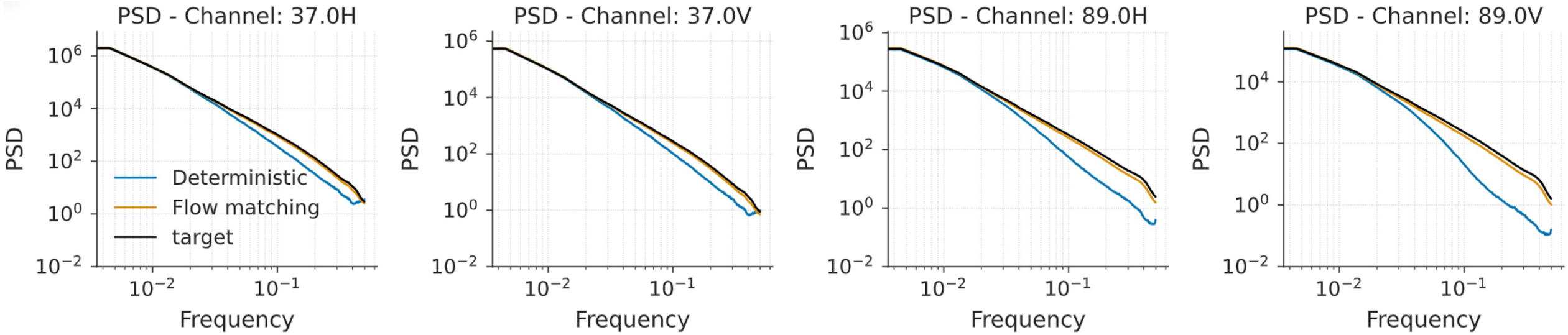
$$\mathrm{CRPS}(F, x) = -\int_{-\infty}^{\infty} (F(y) - \mathbb{1}\{y \geq x\})^2 \, \mathrm{d}y$$

- Une autre formulation évite de calculer cette intégrale
- Le CRPS peut être directement comparé à la MAE d'une prévision déterministe
- L'Energy Score est l'extension multivariée du CRPS

Lectures complémentaires sur ces mesures et d'autres pour évaluer les prévisions probabilistes :

Raftery & Gneiting (2007) Strictly Proper Scoring Rules, Prediction, and Estimation. JASA

Évaluation par rapport au modèle déterministe - spectre

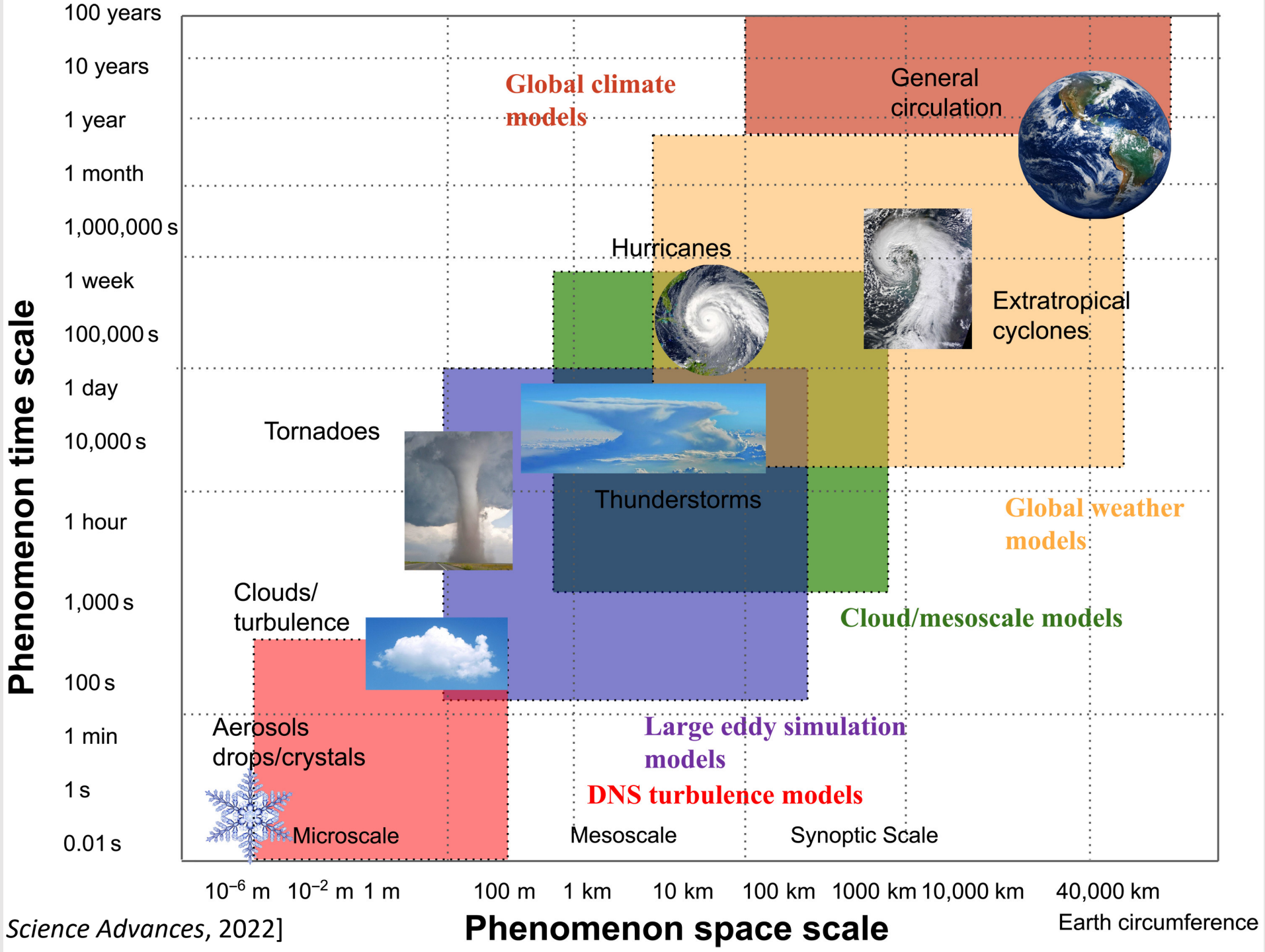


L’IA générative pour :

✓ Prévision probabiliste
- **ArchesWeatherGen** : prévision météo efficace entraînée à 1,5°
- **ArchesClimate** : émulation, interpolation et extrapolation du climat décennal

✓ Fusion de données avec alignement probabiliste de domaines
- **MotifGen** : un modèle génératif, multisource et auto-supervisé pour apprendre à partir de données géospatiales

❑ L’alignement de domaines probabiliste et la descente d’échelle
- **ClimAlign** : descente d’échelle générative non appariée avec alignement de domaines
- **SerpentFlow** : alignement de domaines génératif non apparié par décomposition en structure partagée (avec application à la descente d’échelle)



[Gettelman et al., *Science Advances*, 2022]



Descente d’échelle non supervisée des modèles climatiques par alignement de domaines : *application aux champs de vent.*

Julie Keisler, Boutheina Oueslati,
Anastase Charantonis, Yannig Goude,
Claire Monteleoni

Actes de Climate Informatics 2026

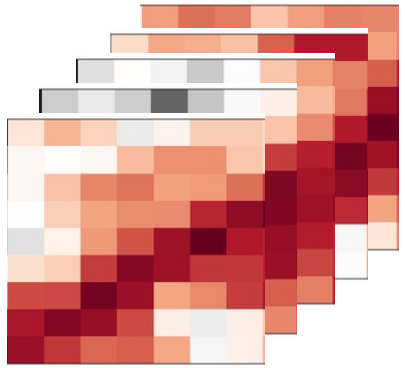
Descente d’échelle des modèles climatiques

Résolution spatiale grossière : les MCG fonctionnent à ~100 km ou plus et ratent des processus locaux essentiels

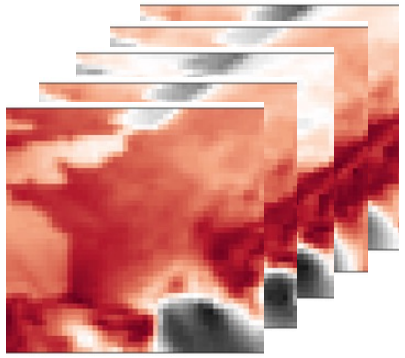
Représentation limitée : les structures fines comme les effets topographiques et les contrastes terre-mer sont mal captées

Biais systématiques : les sorties des MCG s’écartent nettement des régimes climatiques observés

Études d’impact locales compromises : les sorties brutes des MCG ne suffisent pas pour évaluer les régimes de vent futurs en vue de la planification énergétique

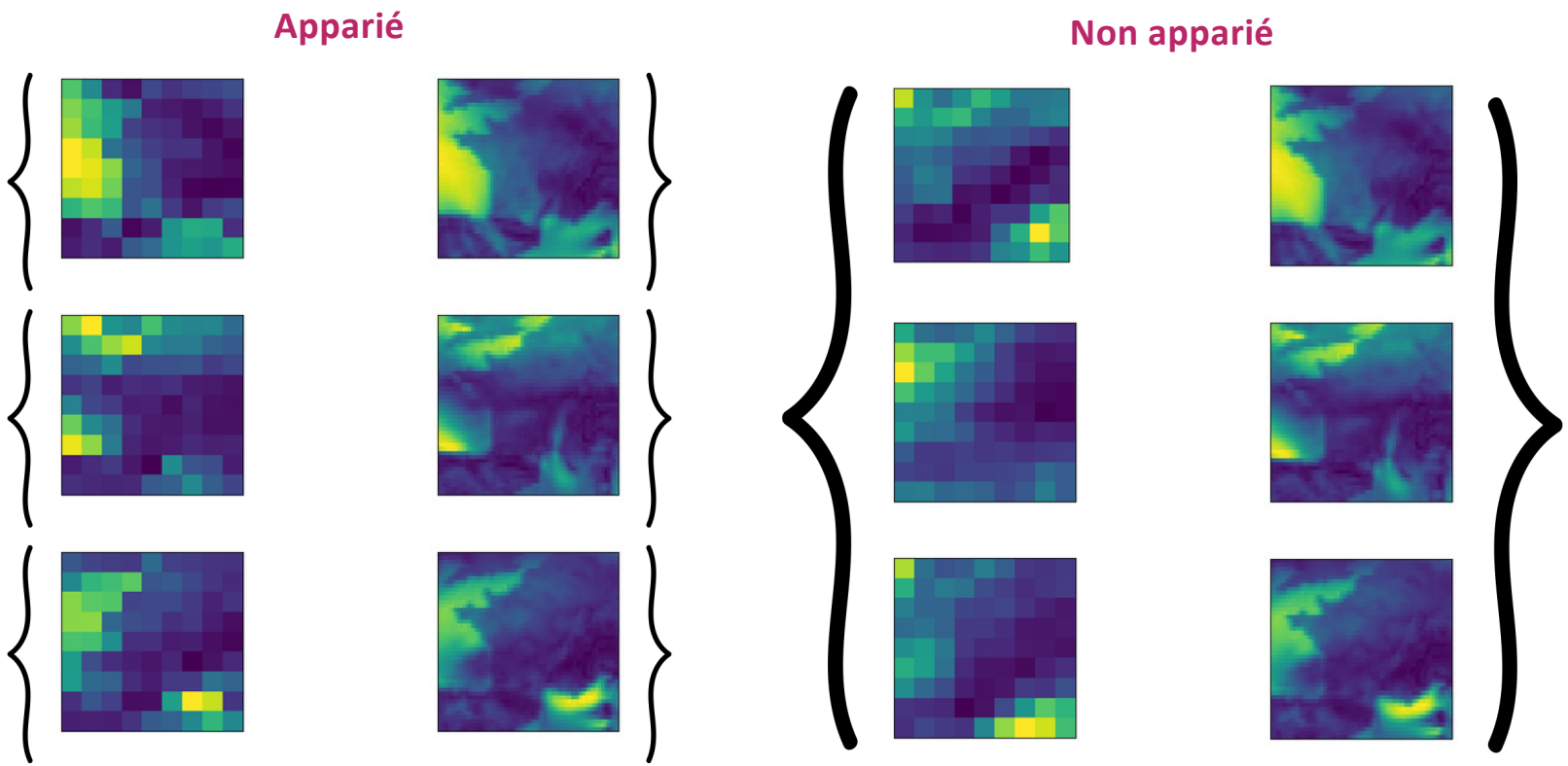


MCG ACCESS-ESM1 sur la France sur 5 jours



ERA5 sur la France sur 5 jours

La descente d’échelle comme alignement de domaines non apparié



- MCG : **trajectoires physiquement plausibles**
- Réanalyses : la séquence **météo historique** réelle → **MCG et réanalyse ne sont pas appariés**
- Descente d’échelle = **alignement de distributions** : projeter les sorties du MCG sur la distribution de la réanalyse

Limites des méthodes statistiques classiques

Méthodes fondées sur les distributions

- Quantile mapping, CDF-t (Michelangeli et al. 2009), correction des lois marginales

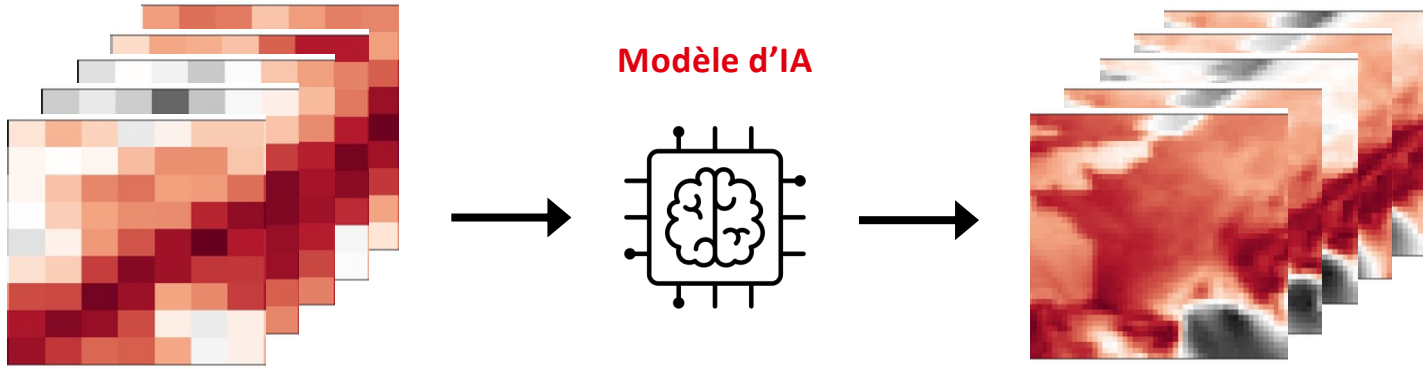
Méthodes multivariées (transport optimal)

- R2D2 (Mehrotra & Sharma 2020), dOTC (Robin et al. 2019)

Malédiction de la dimension : les méthodes classiques ne parviennent pas à modéliser toutes les interactions en jeu

⟶ Méthodes fondées sur l'IA / l'apprentissage profond

L'IA générative pour la descente d'échelle



Avantages

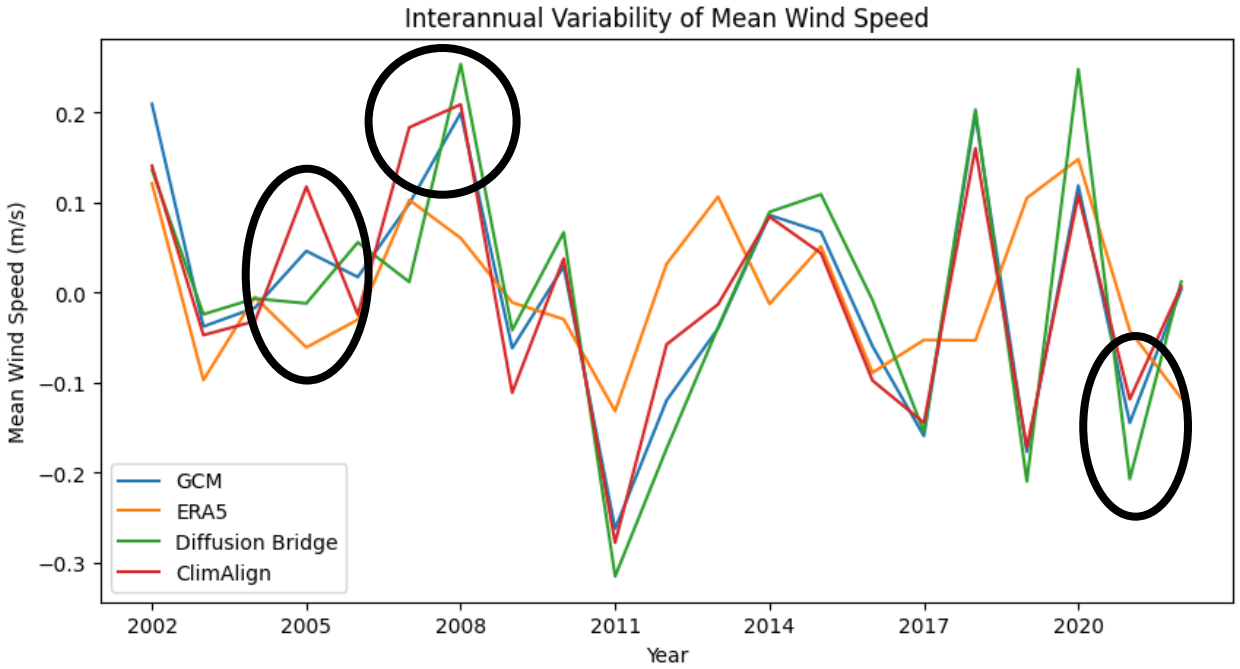
- Apprend la **distribution jointe complète** des champ haute résolution
- Génère **plusieurs réalisations**

Défis

- Données non appariées ⟶ **pas de vérité de terrain**
- Les méthodes implicites peuvent dériver par rappor au signal du MCG

= projections **peu fiables**

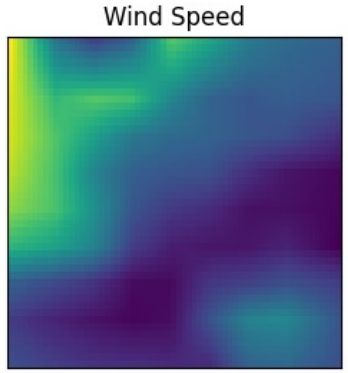
(ClimAlign/Groenke et al. 2020, ponts de diffusion/Bischoff & Deck 2024)



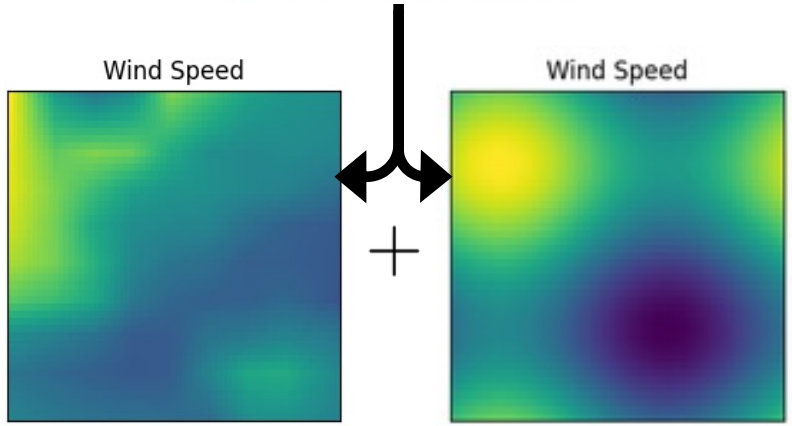
SerpentFlow : introduire des pseudo-paires

SerpentFlow est une méthode générative fondée sur l'IA pour l'**alignement de domaines non apparié** par décomposition en structure partagée

Vitesse moyenne du vent du MCG



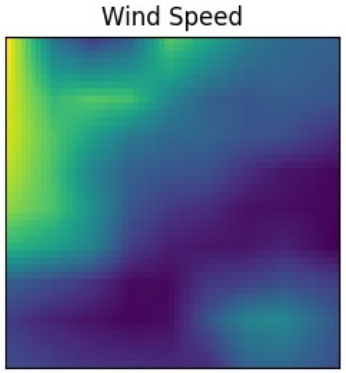
Séparer les hautes et les basses fréquences spatiales

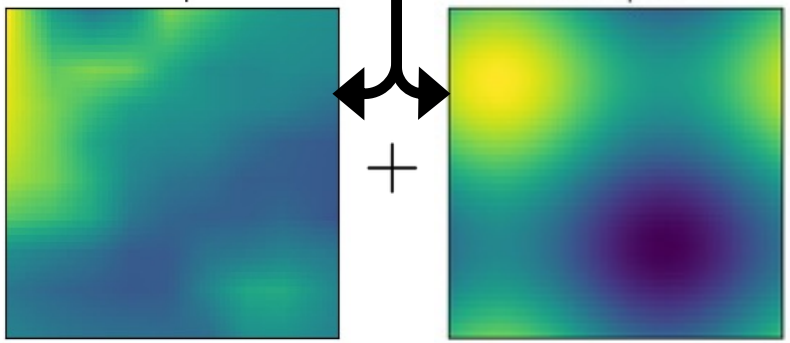


SerpentFlow : introduire des pseudo-paires

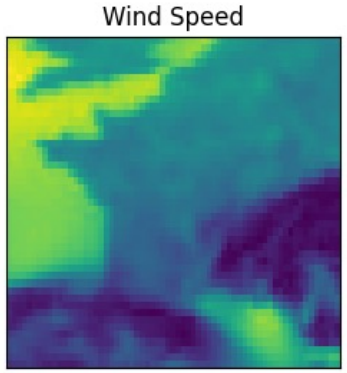
SerpentFlow est une méthode générative fondée sur l'IA pour l'**alignement de domaines non apparié** par décomposition en structure partagée

Vitesse moyenne du vent du MCG

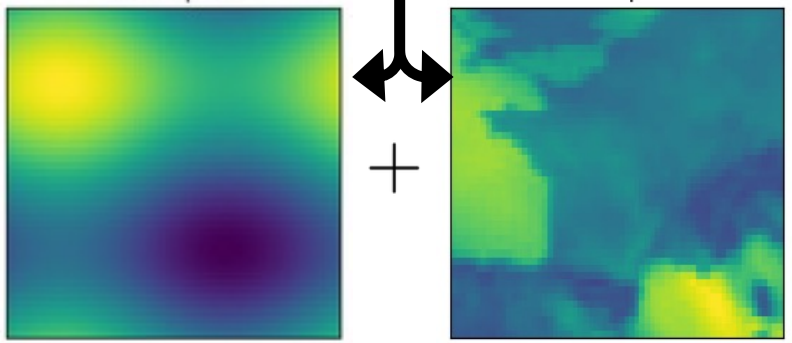




Séparer les hautes et les basses fréquences spatiales



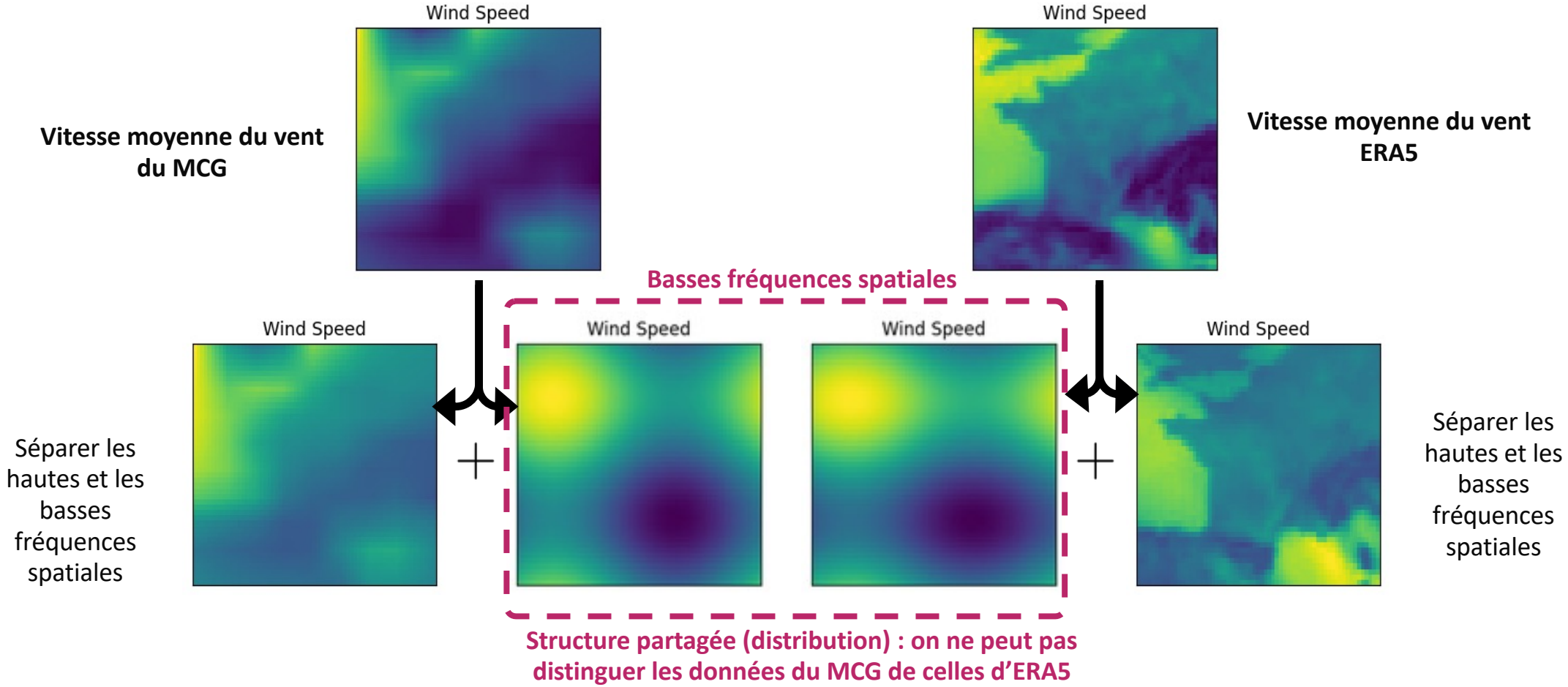
Vitesse moyenne du vent ERA5



Séparer les hautes et les basses fréquences spatiales

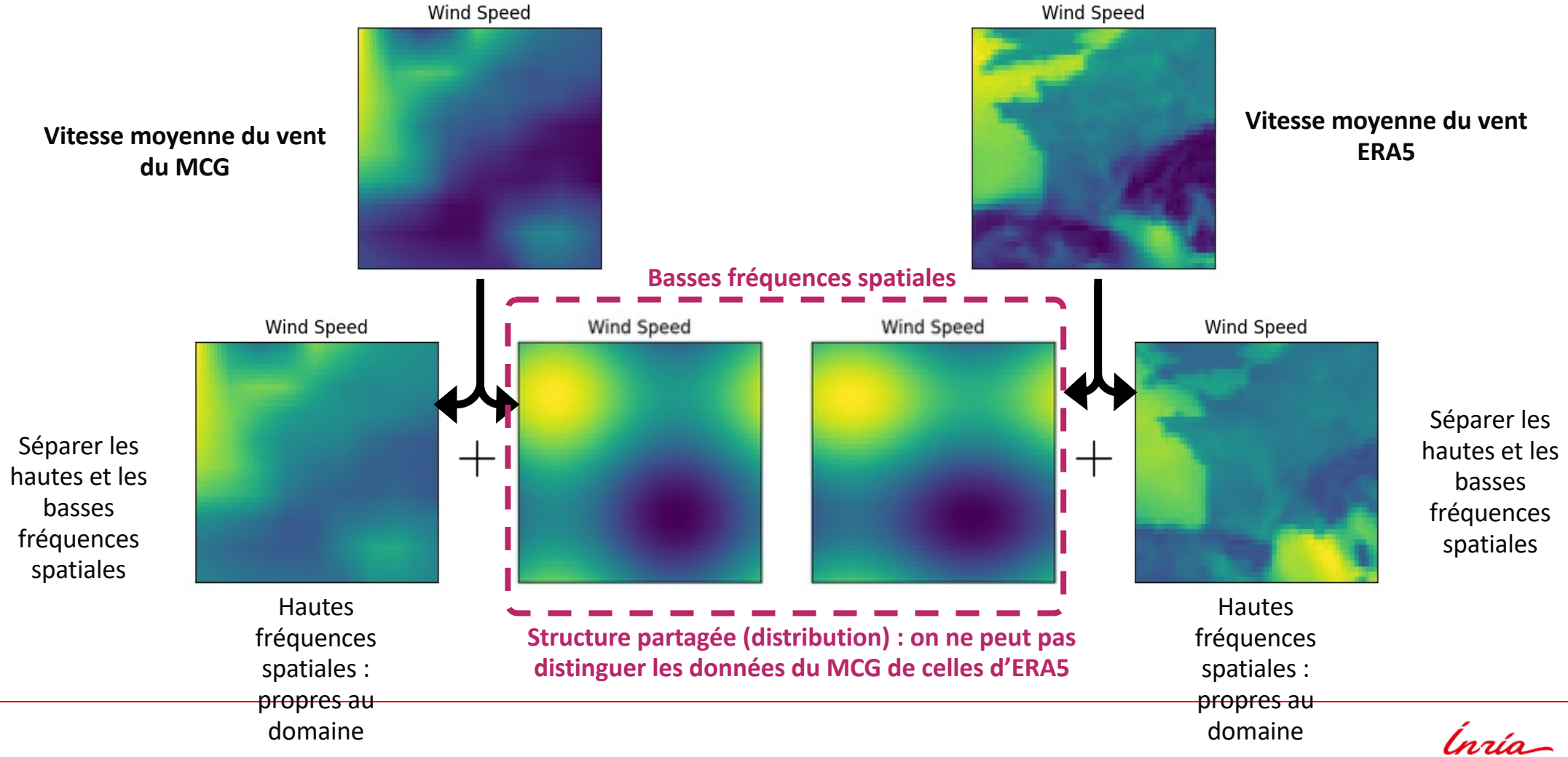
SerpentFlow : introduire des pseudo-paires

SerpentFlow est une méthode générative fondée sur l'IA pour l'**alignement de domaines non apparié** par décomposition en structure partagée

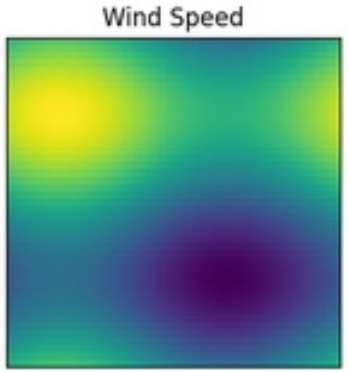


SerpentFlow : introduire des pseudo-paires

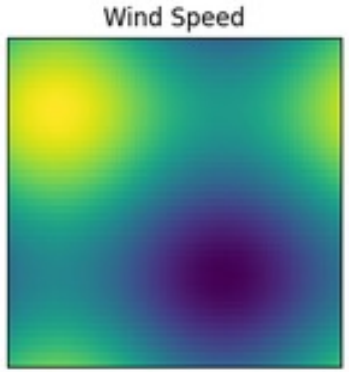
SerpentFlow est une méthode générative fondée sur l'IA pour l'**alignement de domaines non apparié** par décomposition en structure partagée



Trouver automatiquement la fréquence de coupure



Filtre passe-bas(HR) de fréquence de coupure w



Filtre passe-bas(BR) de fréquence de coupure w

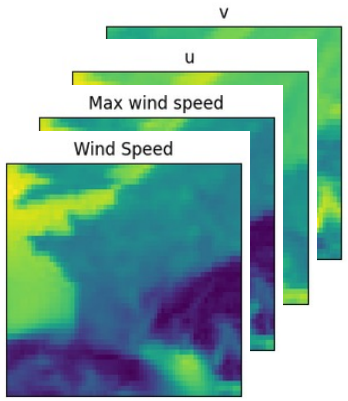
→ Classifieur →

Données BR ou HR ?

- Cesser de réduire la fréquence de coupure quand l'exactitude <= 60 %
- Fréquence souvent trouvée ~= **résolution effective** du MCG

Chaîne SerpentFlow

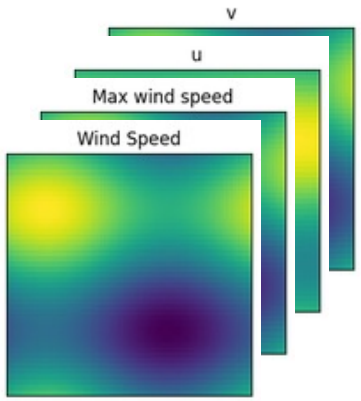
Entraînement



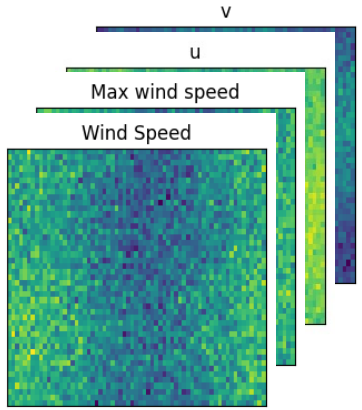
Référence (obs./réanalyse)

Filtre passe-bas, de fréquence de coupure

w_c

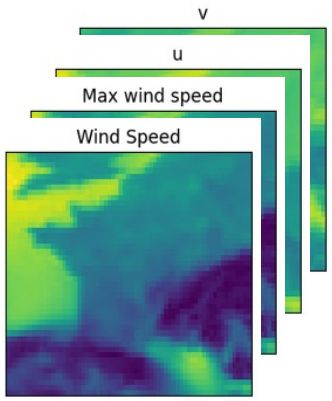


Ajouter du bruit sur les hautes fréquences



Entraîner l'approche générative

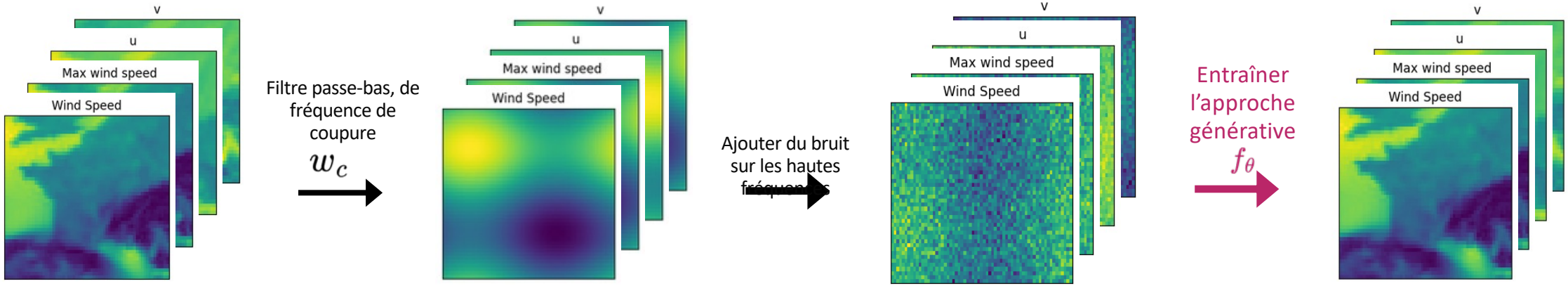
f_θ



Référence (obs./réanalyse)

Chaîne SerpentFlow

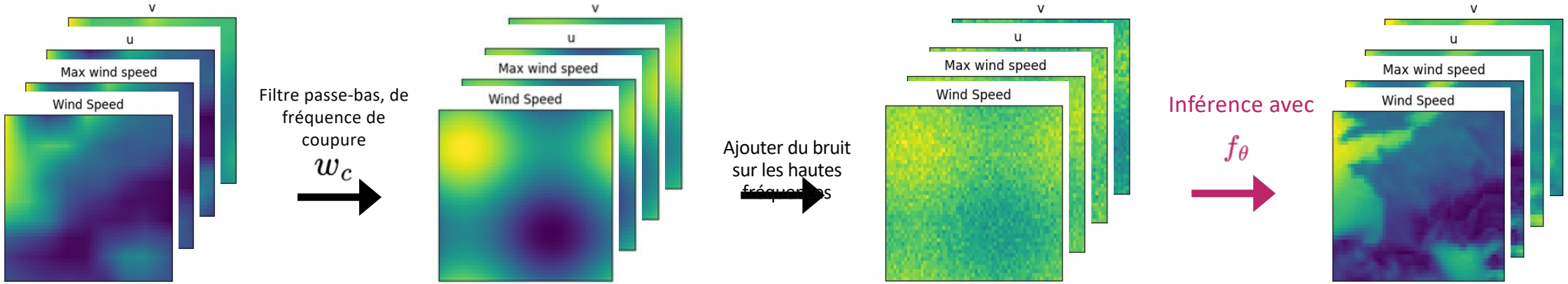
Entraînement



Référence (obs./réanalyse)

Référence (obs./réanalyse)

Inférence

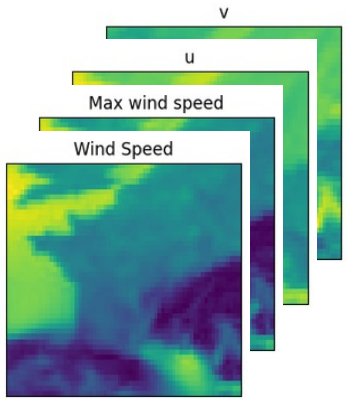


MCG interpolé

MCG après descente d'échelle

Chaîne SerpentFlow

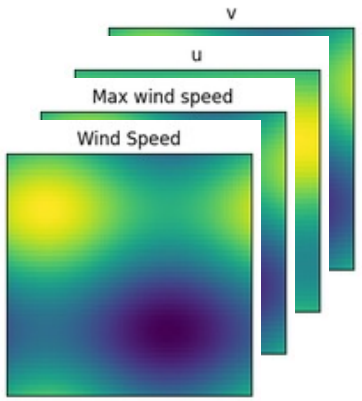
Entraînement



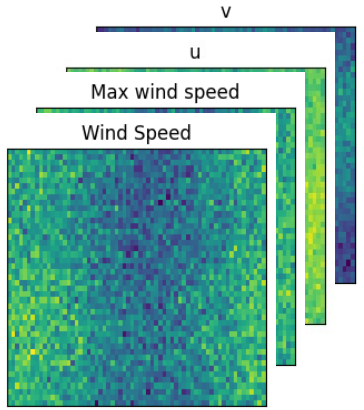
Référence (obs./réanalyse)

Filtre passe-bas, de fréquence de coupure w_c

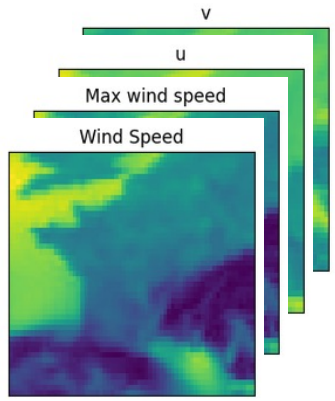
Distribution jointe



Ajouter du bruit sur les hautes fréquences

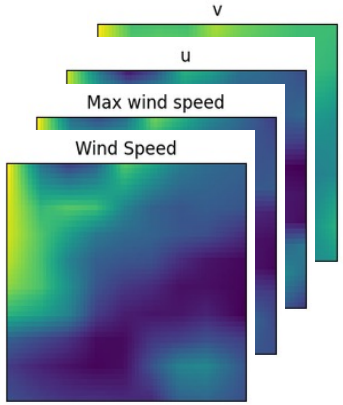


Entraîner l'approche générative f_θ



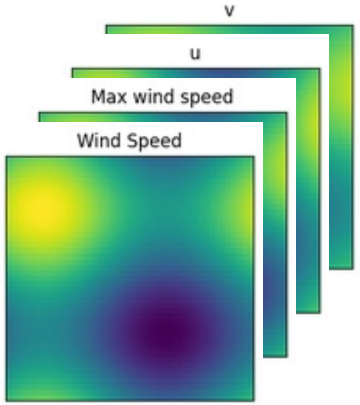
Référence (obs./réanalyse)

Inférence

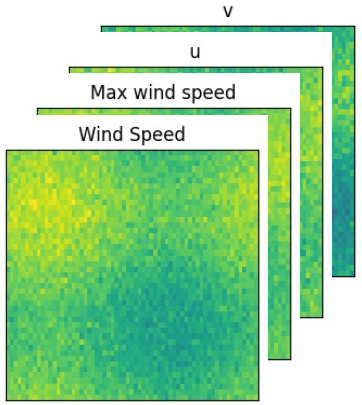


MCG interpolé

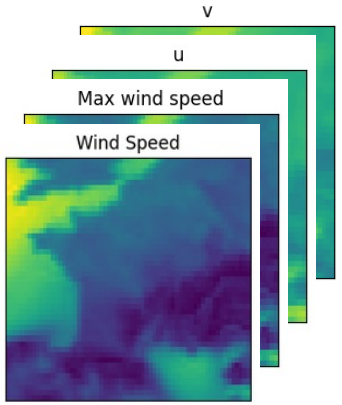
Filtre passe-bas, de fréquence de coupure w_c



Ajouter du bruit sur les hautes fréquences

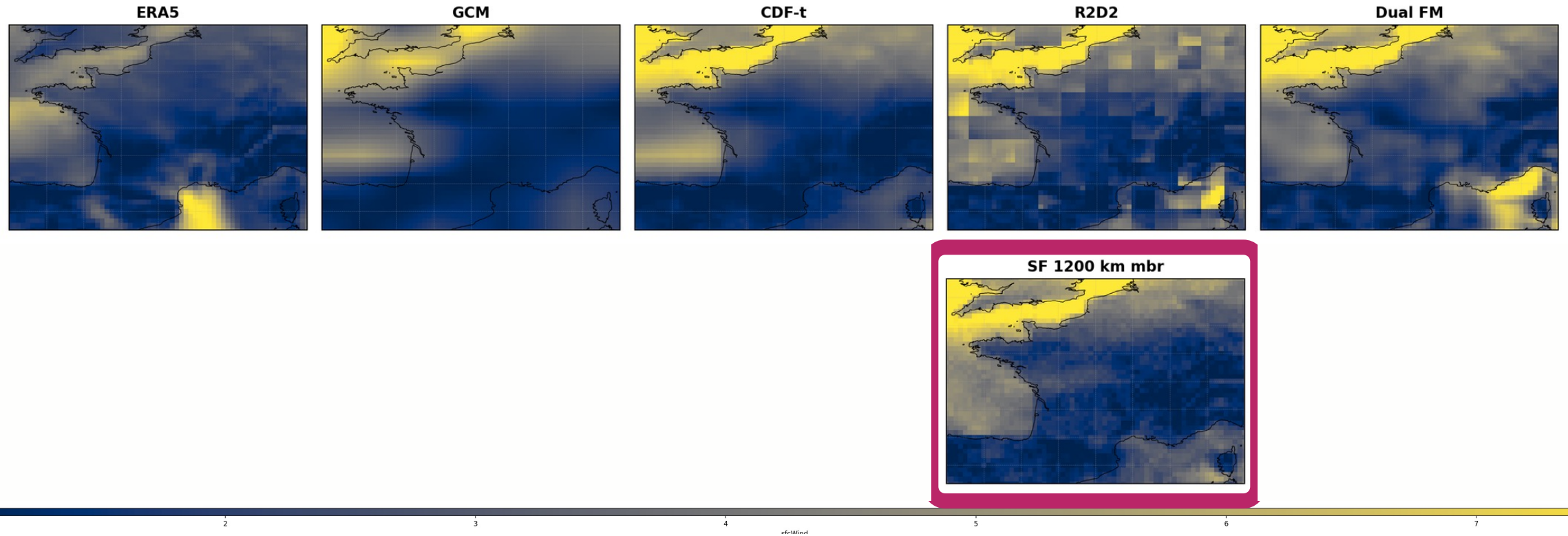


Inférence avec f_θ



MCG après descente d'échelle

Avons-nous approché la distribution d'ERA5 ?

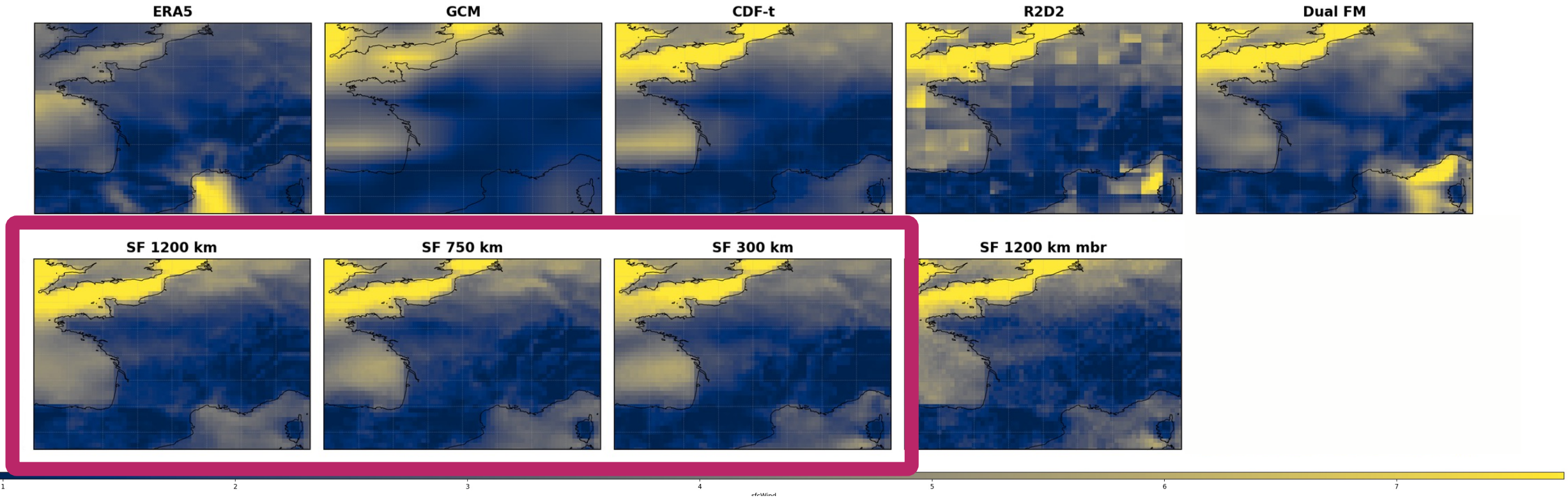


Principaux constats

- Les méthodes d'AP surpassent CDF-t et R2D2 sur la structure spatiale.
- **SF 1200 km** (1 membre) : **meilleure approximation** de la distribution d'ERA5

l'ERA5.

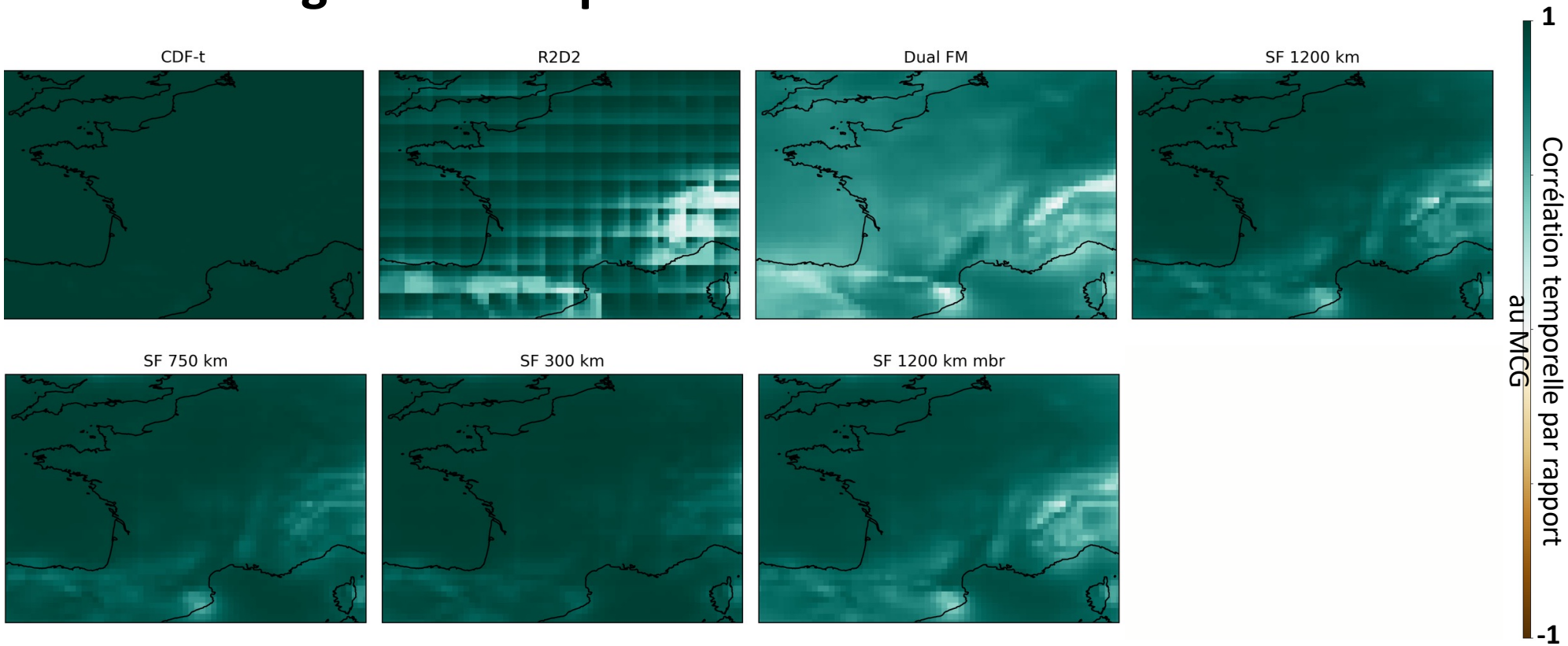
Avons-nous approché la distribution d'ERA5 ?



Principaux constats

- Les méthodes d'AP surpassent CDF-t et R2D2 sur la structure spatiale.
- **SF 1200 km** (1 membre) : **meilleure approximation** de la distribution d'ERA5
- **fréquence de coupure plus haute** ⟶ dérive par rapport à la distribution d'ERA5.

Suivons-nous le signal donné par le MCG ?



Principaux constats

- **CDF-t est le plus proche de la** dynamique du MCG ; Dual FM ne suit pas la trajectoire du MCG
- La dérive de SF apparaît dans les régions **montagneuses**
- **Fréquence de coupure plus basse** ⟶ plus proche de la dynamique du MCG

Perspectives

- Fonctionne avec différents **MCG, résolutions et domaines irréguliers** — ERA5 (25 km) et SAFRAN (8 km)
- Surpasse les méthodes statistiques pour modéliser la **distribution jointe** des données de réanalyse
- Par rapport aux autres approches génératives, reste proche de la dynamique du MCG
- **La fréquence de coupure est interprétable**
  - Relie directement la séparation des échelles à la résolution effective du MCG
  - Gère **le compromis** entre suivre la trajectoire du MCG et reproduire la distribution de la réanalyse
- **Limites :** vent près de la surface uniquement — prochaines étapes : autres variables, extrêmes, descente d'échelle temporelle

L’IA générative pour :

- ✓ Prévision probabiliste
  - **ArchesWeatherGen** : prévision météo efficace entraînée à 1,5°
  - **ArchesClimate** : émulation, interpolation et extrapolation du climat décennal

- ✓ Fusion de données avec alignement probabiliste de domaines
  - **MotifGen** : un modèle génératif, multisource et auto-supervisé pour apprendre à partir de données géospatiales

- ✓ L’alignement de domaines probabiliste et la descente d’échelle
  - **ClimAlign** : descente d’échelle générative non appariée avec alignement de domaines
  - **SerpentFlow** : alignement de domaines génératif non apparié par décomposition en structure partagée (avec application à la descente d’échelle)

Se lancer dans l'IA générative à base de flots

Flow matching

Tutoriels :

- Peter Holderrieth et al. (2026) "An Introduction to Flow Matching and Diffusion Models" Lecture notes and videos: `diffusion.csail.mit.edu`
- Yaron Lipman et al. (2024) "Flow Matching Guide and Code" arXiv:2412.06264
- Emile Pierret, Valentine Tosel, Julie Delon, Alasdair Newson (2026) "Flow Matching for Applied Mathematicians" hal-05538982v1

Lectures complémentaires sur les méthodes à base de flots

Interpolants stochastiques

Cadre général à base de flots pour transporter du bruit vers les données

- Le flow matching et les rectified flows en sont des exemples
- Généralise le cadre : on peut entraîner avec une EDO et échantillonner avec une EDS, ou inversement

Opérateurs neuronaux

Les méthodes à base de flots opèrent dans l'espace des opérateurs, ce qui permet une généralisation indépendante de l'échelle

Modèles de cohérence

Version en une seule étape du flow matching. Domaine de recherche ouvert, avec un peu de théorie et des limites pratiques.

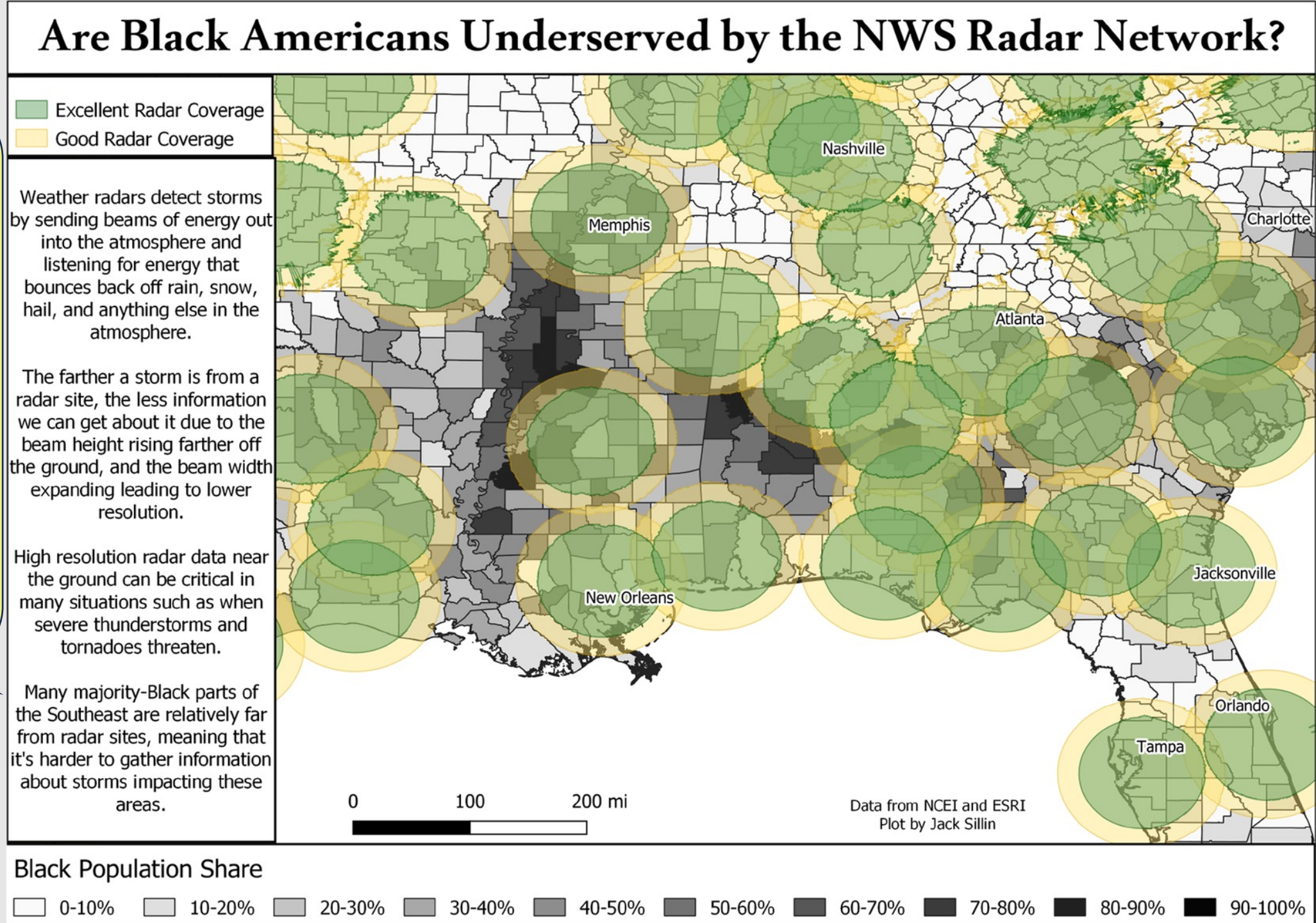
Apprentissage par chaîne de Markov inverse

En substance, du flow matching de la basse vers la haute résolution

Xinwei Shen, Nicolai Meinshausen, Tong Zhang (2025) "Reverse Markov Chain Learning: Multi-Step Generative Models for Complex Distributions" arXiv:2502.13747

Section bonus : l’IA pour l’équité des données climatiques

« De nombreuses zones à majorité noire du Sud-Est [des États-Unis] sont relativement éloignées des radars, si bien qu'il est plus difficile de recueillir des informations sur les tempêtes qui les frappent. »



Crédit : Jack Sillin

L’IA pour l’équité des données climatiques

- Entraîner les modèles dans les régions riches en données et les appliquer dans les régions pauvres en données
  - Les entraîner et les valider dans les régions riches en données
  - Les affiner avec les données limitées des régions pauvres en données et s’en servir pour générer davantage de données.

- Contribution à l’équité des données climatiques
  - Échelles locales (p. ex. héritage d’injustice environnementale aux États-Unis)
  - Échelles mondiales :
    - Le Nord global a historiquement émis plus de carbone ; c’est aussi là qu’il y a généralement plus de données