

Pesquisa em IA
para a mudança climática e a sustentabilidade ambiental

Claire Monteleoni

INRIA Paris

University of Colorado Boulder

Choose France™



Dezembro de 2021: condado de Boulder, Colorado

- A seca de neve no outono e no inverno de 2021 deixou a cobertura do solo seca
- Ventos de 130-160 km/h, mais uma ignição, desencadearam uma “tempestade de fogo” incontrolável
- 2 mortes. 1.000 casas e 20 estabelecimentos destruídos, e outros danificados





“A oportunidade da IA para a Terra é significativa. A explosão atual da IA fará com que a incorporemos a cada vez mais coisas a cada ano.... Ao pensar nos ganhos, nas eficiências e nas novas soluções que ela cria para países, empresas e o dia a dia, também precisamos pensar em como maximizar os ganhos para a sociedade e para o meio ambiente como um todo.”

– Fórum Econômico Mundial (2018), “Harnessing Artificial Intelligence for the Earth”





Pesquisa em IA para a mudança climática e a sustentabilidade ambiental

MUDANÇA CLIMÁTICA

ADAPTAÇÃO

MITIGAÇÃO

IMPACTOS

Curto prazo

Médio prazo

Longo prazo



Eventos extremos

Ameaças em cascata



Transição energética

Mudança no uso da terra



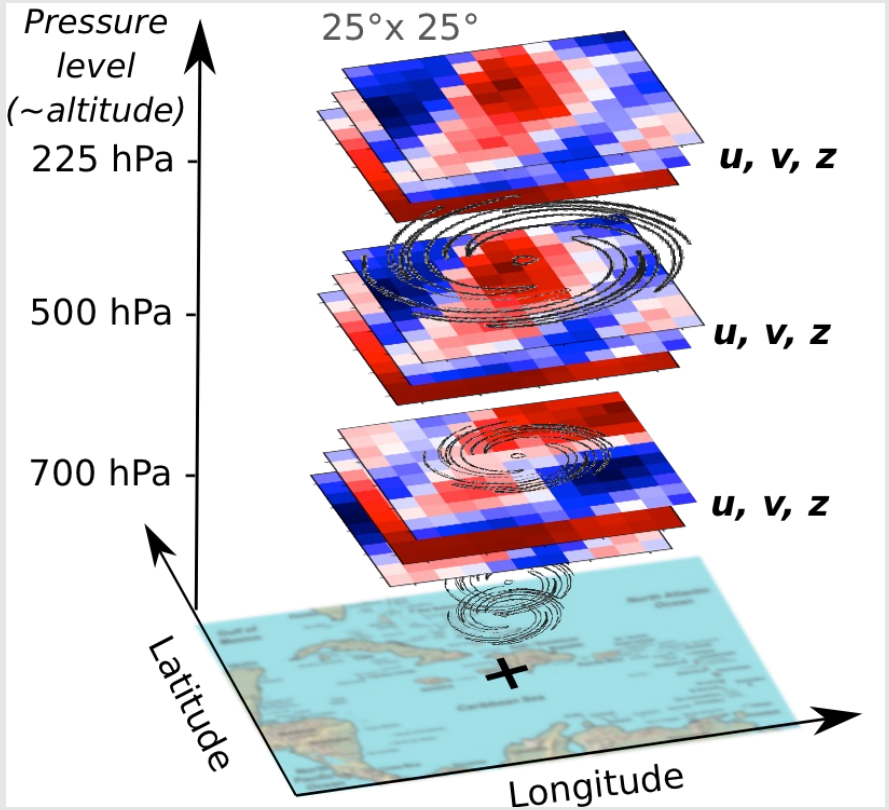
Emissões de carbono

Elevação do nível do mar

Soluções baseadas em IA

ADAPTAÇÃO

IA para eventos extremos e ameaças em cascata



[Giffard-Roisin et al., Frontiers 2020]

Previsão de trajetórias de furacões

Previsão de chuvas extremas da monção de verão indiana

Detecção de avalanches

IA generativa para a previsão do tempo



MITIGAÇÃO

Reduzir as emissões de carbono

Acelerar a transição para a energia verde

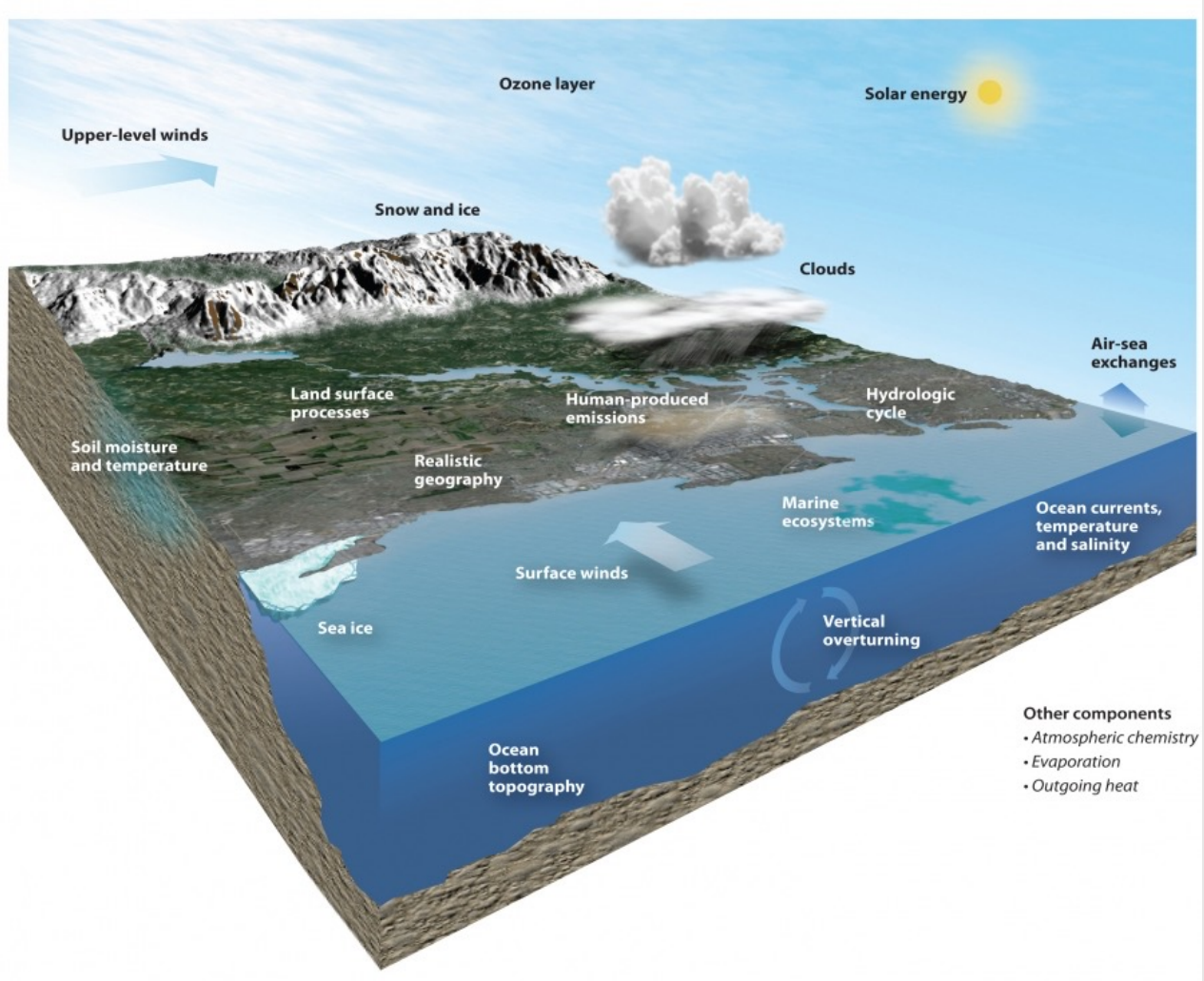
- Previsão solar e eólica com IA
- IA para o downscaling de dados solares e eólicos

Reduzir a computação da modelagem do tempo e do clima

- Depois de treinada, a IA prevê muito

IMPACTOS

IA para entender e prever a mudança climática



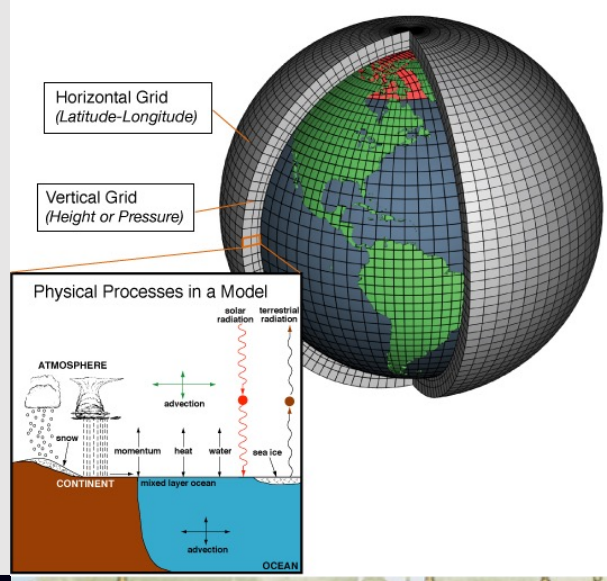
UCAR Science Education

Usar IA para aprender as relações entre simulações do IPCC e observações

- Tornar mais robustas as previsões de conjuntos de modelos climáticos
- Projetar a elevação do nível do mar a longo prazo
- Projetar as emissões de carbono a longo prazo

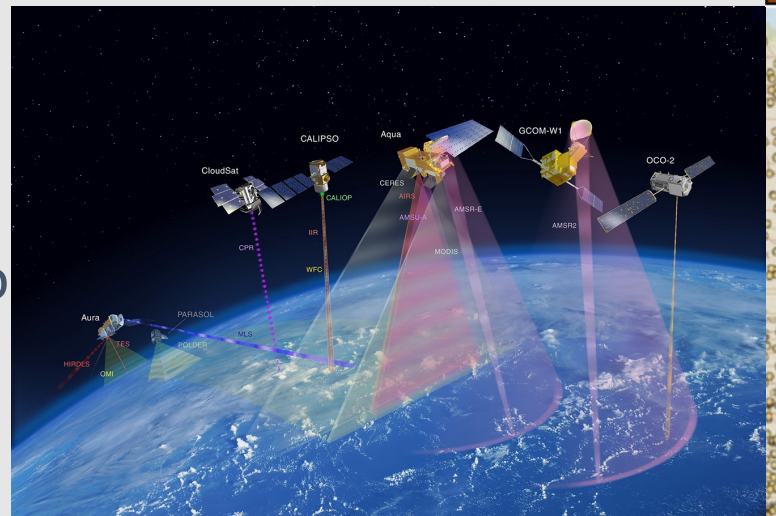
Abordagem: aproveitar todos os dados disponíveis

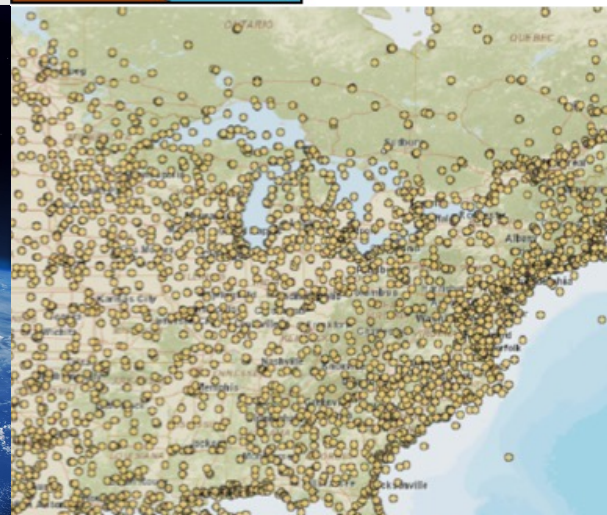
- Dados simulados gerados por modelos físicos
  - Modelos de previsão numérica do tempo (PNT)
  - Modelos de circulação geral (MCG)
  - Modelos climáticos regionais (MCR)



- Dados de reanálise
  - Produtos em grade a partir da assimilação de dados:

aplica leis físicas às observações





- Dados de observação
  - Dados de sensoriamento remoto po
  - Dados in situ

Tarefas de AM por trás de muitos desses problemas

- ❑ Fusão de dados
- ❑ Previsão
- ❑ Aprendizado não supervisionado
- ❑ Alinhamento probabilístico de domínios
- ❑ Downscaling e interpolação

Métodos de IA

- Aprendizado não supervisionado
- Aprendizado profundo generativo

Roteiro para o resto do dia

I. **Contexto:** a revolução da IA na previsão do tempo

Implicações para o estudo da mudança climática

II. **Métodos de IA:** o que é aprendizado autossupervisionado?

Aplicações em clima e tempo

III. **Métodos de IA:** o que é aprendizado profundo generativo?

Aplicações em clima e tempo

Parte I: Contexto:
a revolução da IA na previsão do tempo

A revolução silenciosa na previsão do tempo

- A previsão numérica do tempo (PNT) vem progredindo de forma constante nas últimas décadas
- Grande conquista científica
  - Um problema de simulação complexo resolvido **rotineiramente** em centros operacionais de previsão do mundo todo
- Alto valor socioeconômico
  - Mitigação de ameaças climáticas, logística, prevenção de perdas patrimoniais, …

Crédito dos primeiros slides: 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

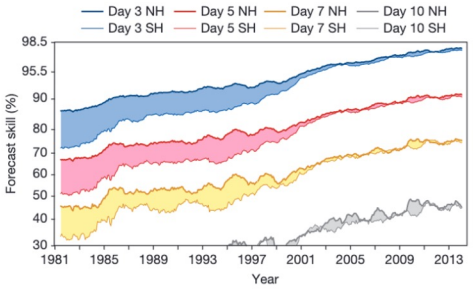
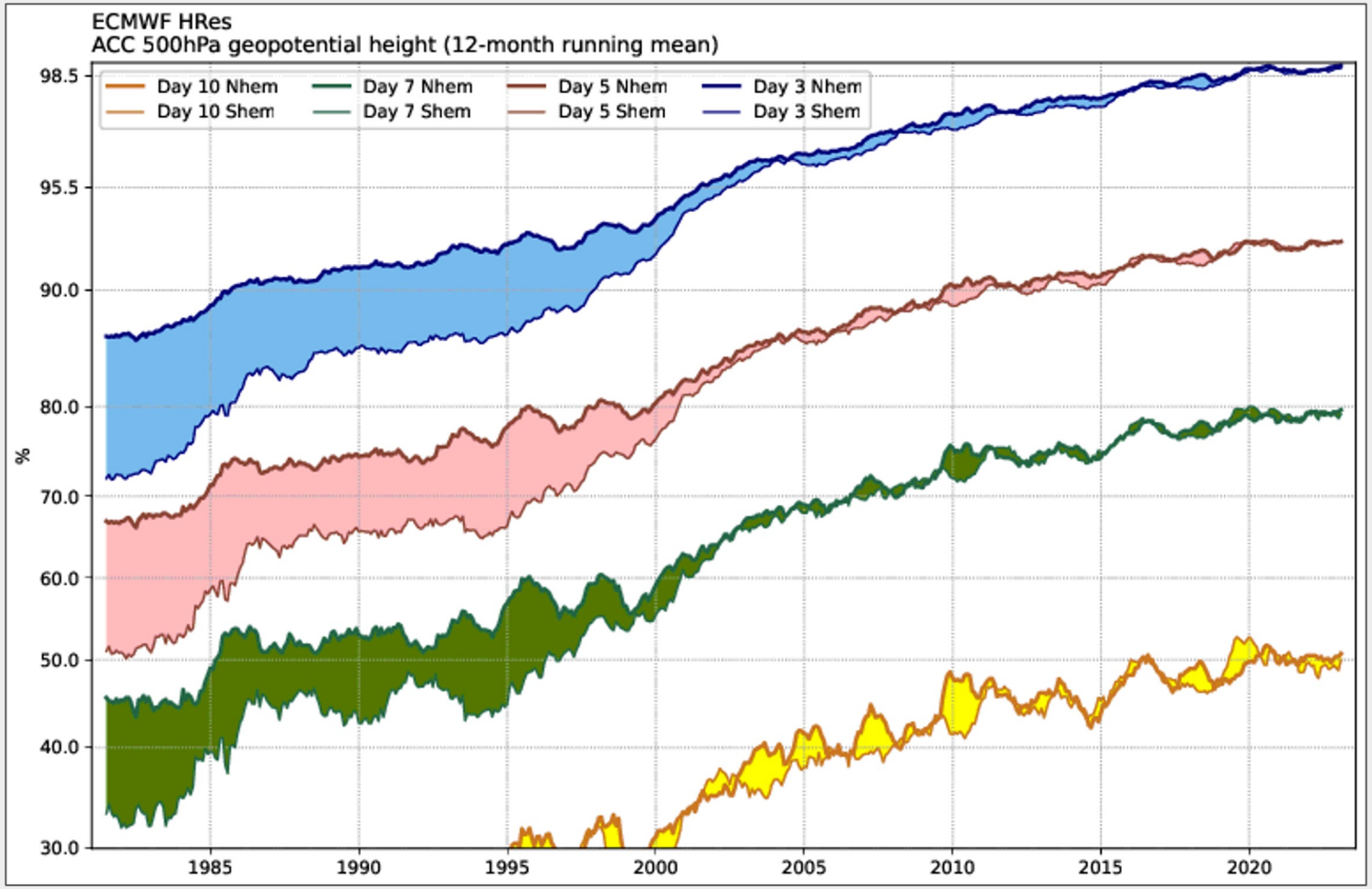


Figure 1 | A measure of forecast skill at three-, five-, seven- and ten-day ranges, computed over the extra-tropical northern and southern hemispheres. Forecast skill is the correlation between the forecasts and the verifying analysis of the height of the 500-hPa level, expressed as the anomaly with 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

A revolução silenciosa na previsão do tempo

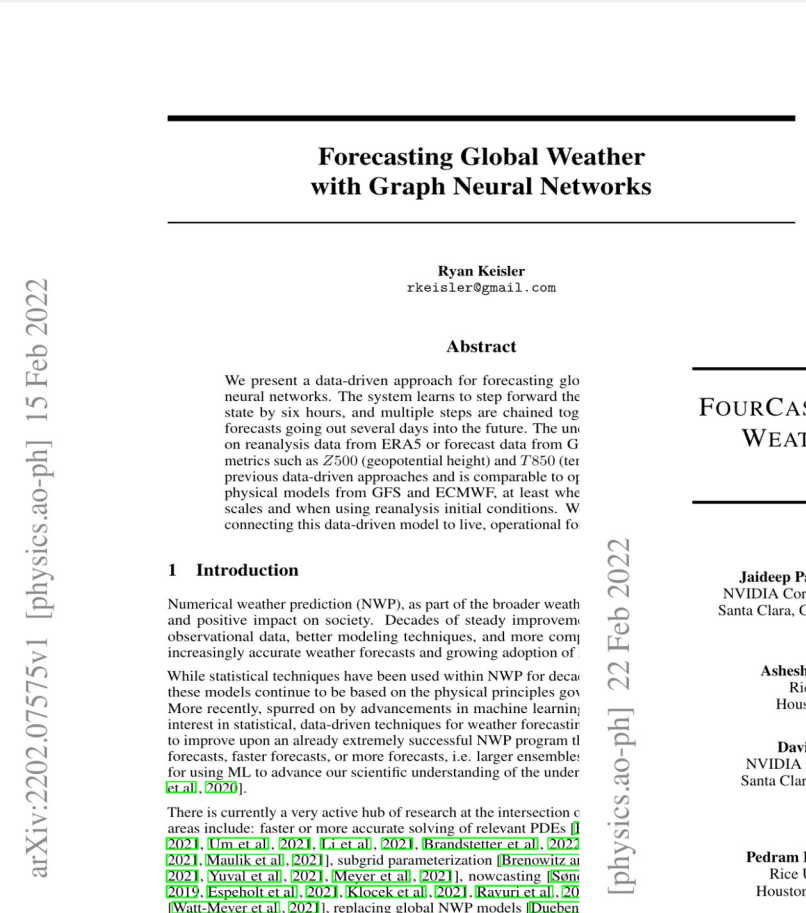


Desempenho das previsões globais do ECMWF em 4 décadas

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

Primeiros modelos IA-PNT

- Desde 2022, várias equipes de pesquisa usaram redes neurais profundas para a previsão do tempo de médio prazo
- A qualidade de alguns desses modelos ficou ~equivalente à dos principais modelos 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 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

[physics.ao-ph] 22 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

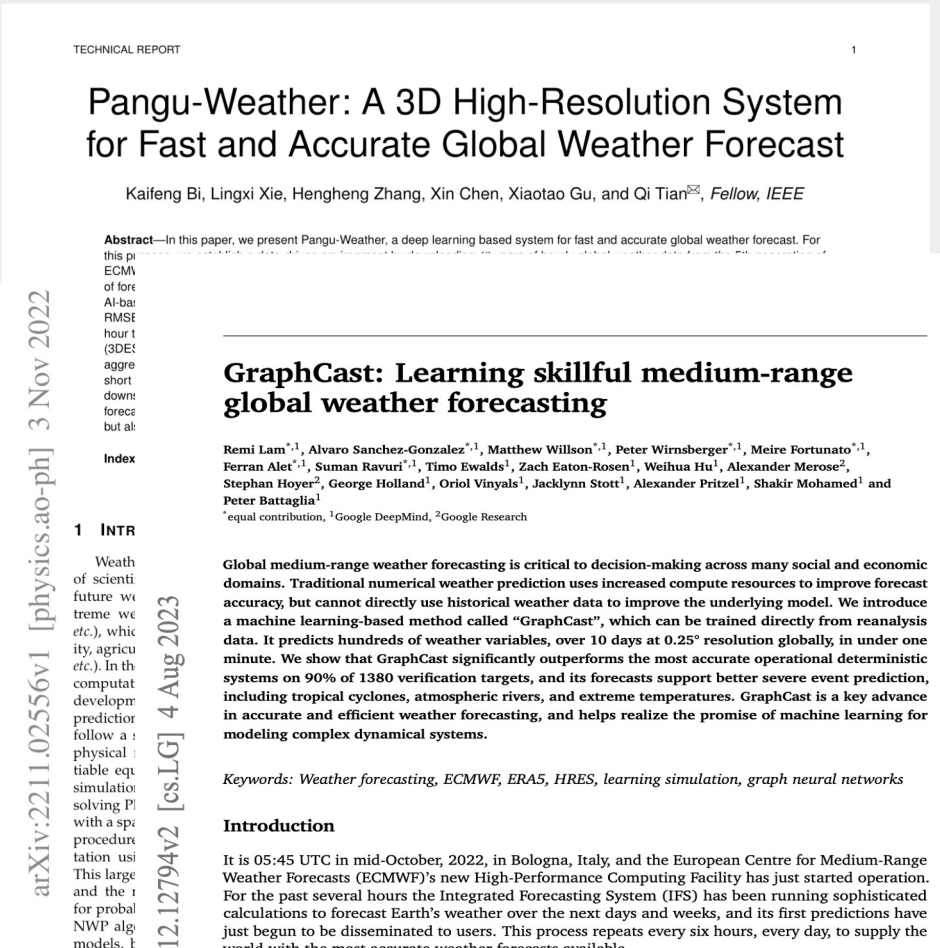
Zongyi Li
California Institute of Technology
Pasadena, CA 91125
NVIDIA Corporation
Santa Clara, CA 95051

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



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 this p ECM of fore AI-ba RMSE hour t (3DE aggre short down foreca but al

Index

1 INTR

Weath of scienti future we treme we *etc.*), whic ity, agricu *etc.*). In th computat developm predictio follow a physical tiable equ simulatio solving P with a spa procedure tation usi This large and the for probal NWP alg models, h

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

arXiv:2212.12794v2 [cs.LG] 4 Aug 2023

GraphCast: Learning skillful medium-range global weather forecasting

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

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

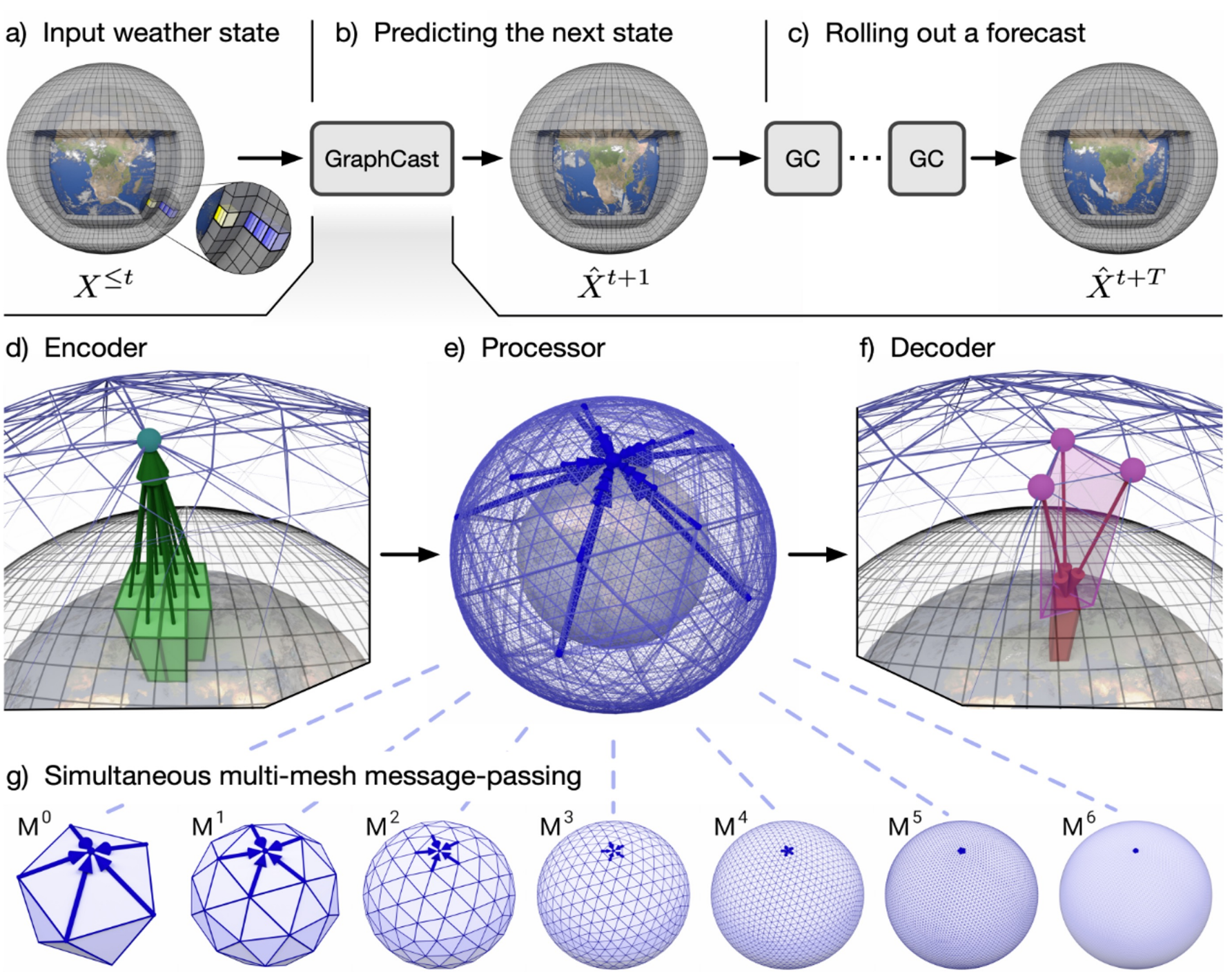
Introduction

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

Economia computacional desses primeiros modelos IA-PNT

Modelo	**Treinamento**	**Inferência**
Pangu-Weather	16 dias em 192 GPUs V100	Previsão de 24 horas 1xV100 1,4 segundo em
GraphCast	21 dias em 32 Cloud TPU V4 (~128 GPUs)	Previsão de 10 dias 1 Cloud TPU V4 < 60 segundos
PNT (ECMWF HRES) ***Resolução mais alta**	-	Previsão de 10 dias 1/3/6 horas Cluster HPC de 11.664 núcleos

Arquitetura do GraphCast



doi:10.48550/arXiv.2212.12794

IA para a previsão do tempo: um novo paradigma

Assimilação de dados → Conjunto de dados ERA5

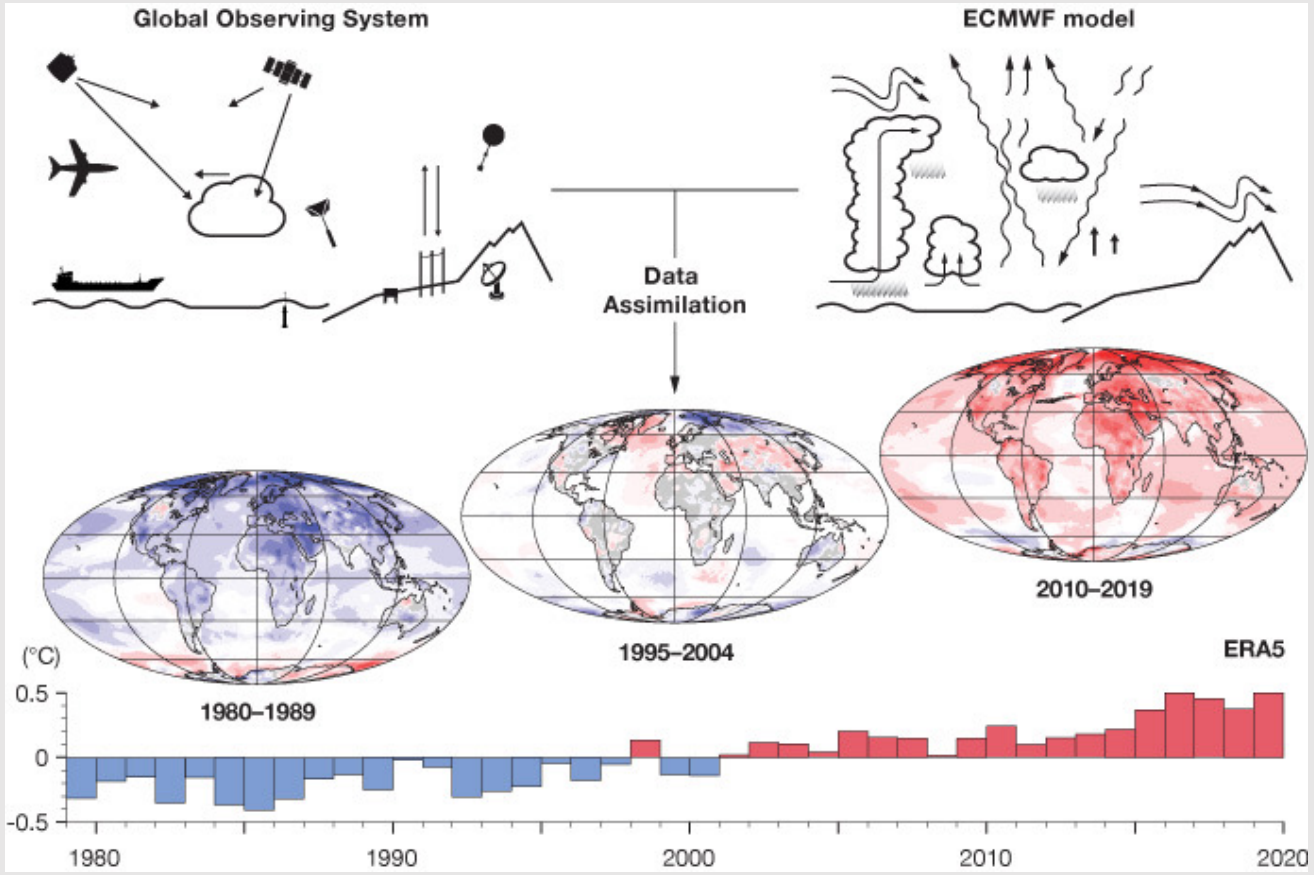
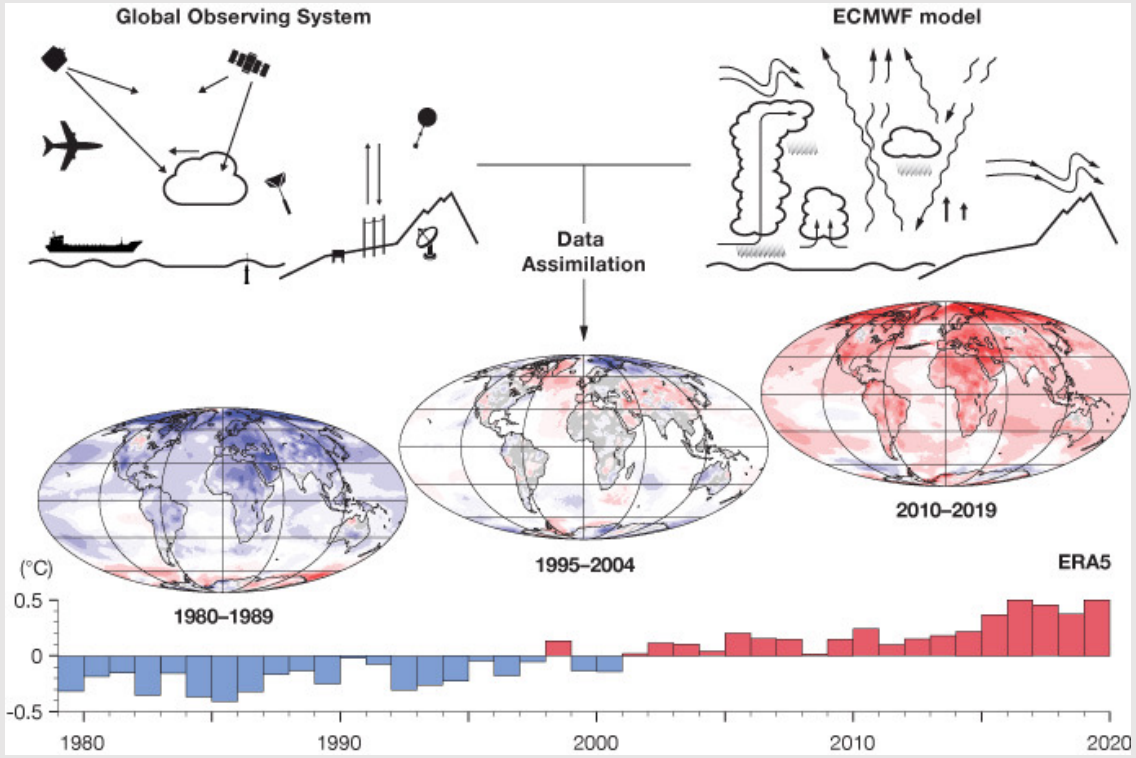


Figura: Centro Europeu de Previsão do Tempo a Médio Prazo (ECMWF)

Conjunto de dados ERA5 → Simulação com rede neural → $P(\textbf{Previsão})$

Revolução da IA na previsão do tempo

Desde 2022, vários modelos de aprendizado profundo mostraram desempenho de previsão comparável ou **MELHOR** que a previsão numérica do tempo (PNT), o estado da arte anterior.



Dados de treinamento: ERA5, um conjunto de dad produzido por assimilação de dados

Tarefa de treinamento: autorregressão: previsão de 6-24 horas

- Desdobrado para prever de 7 a 10 dias

[Site do Centro Europeu de Previsão do Tempo a Médio Prazo (ECMWF)]

Desdobramentos

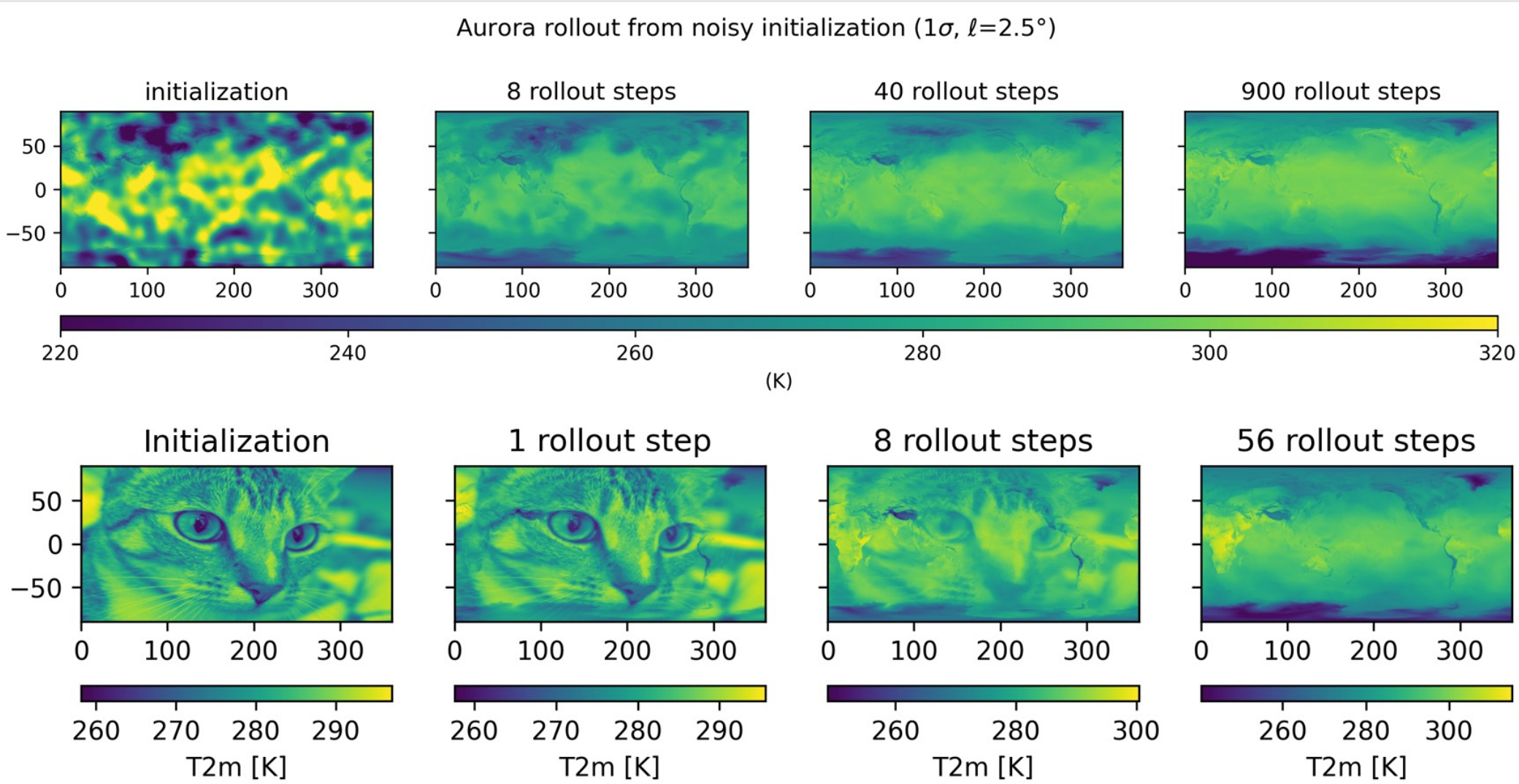


Figure A.9: 2-meter temperature predicted at different time stamps when (top) inputs are perturbed with an additive Gaussian random field with coefficient of variation σ^2 (where σ is the standard deviation of the physical variable) and correlation length of 10 pixels (equivalent to $2.5°$), (bottom)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

IA para a previsão do tempo: um novo paradigma

Time

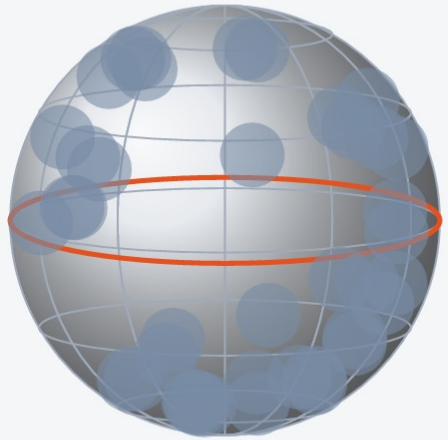
45+ years

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**

Reanálise ERA5 do ECMWF

≈ 13 Petabytes

ArchesWeatherGen: Gerador meteorológico de IA frugal [Couairon et al., *Science Advances*, 2026]

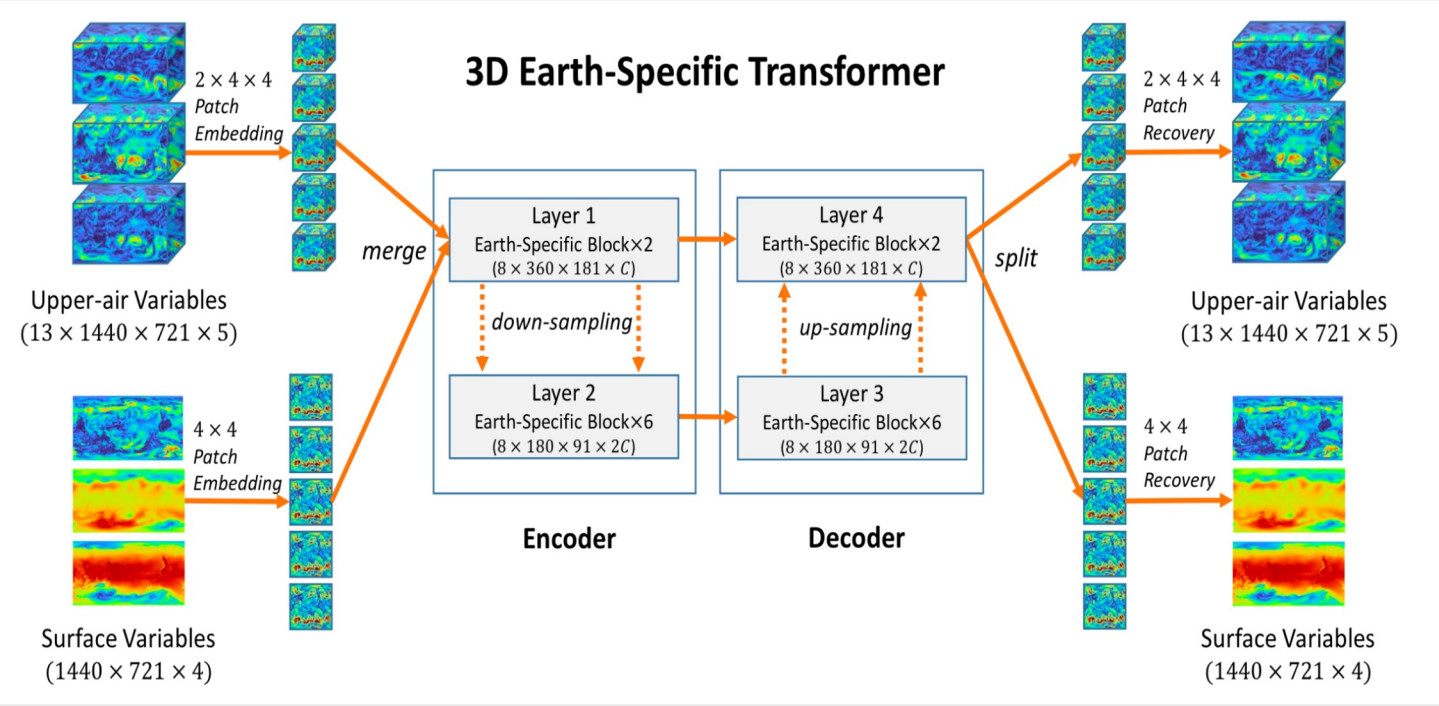
Figura: Emmanuel de Bézenac (INRIA)



ArchesWeather

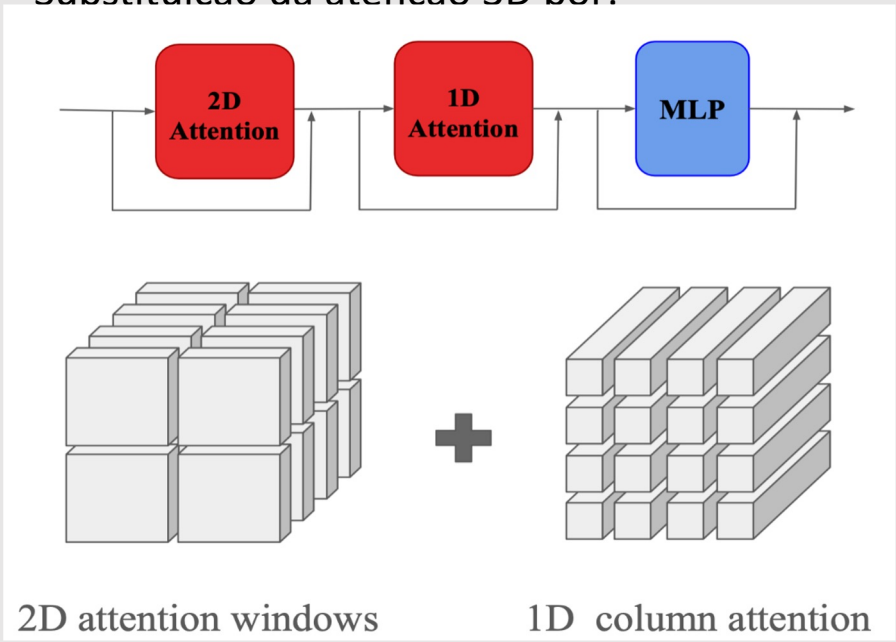
[Couairon et al., ArchesWeather: an efficient AI weather model at 1.5º resolution ICML 2024 workshop on Earth System Modeling]

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



Novidades do ArchesWeather:

- Treinamento com dados de resolução mais grossa
- Substituição da atenção 3D por:



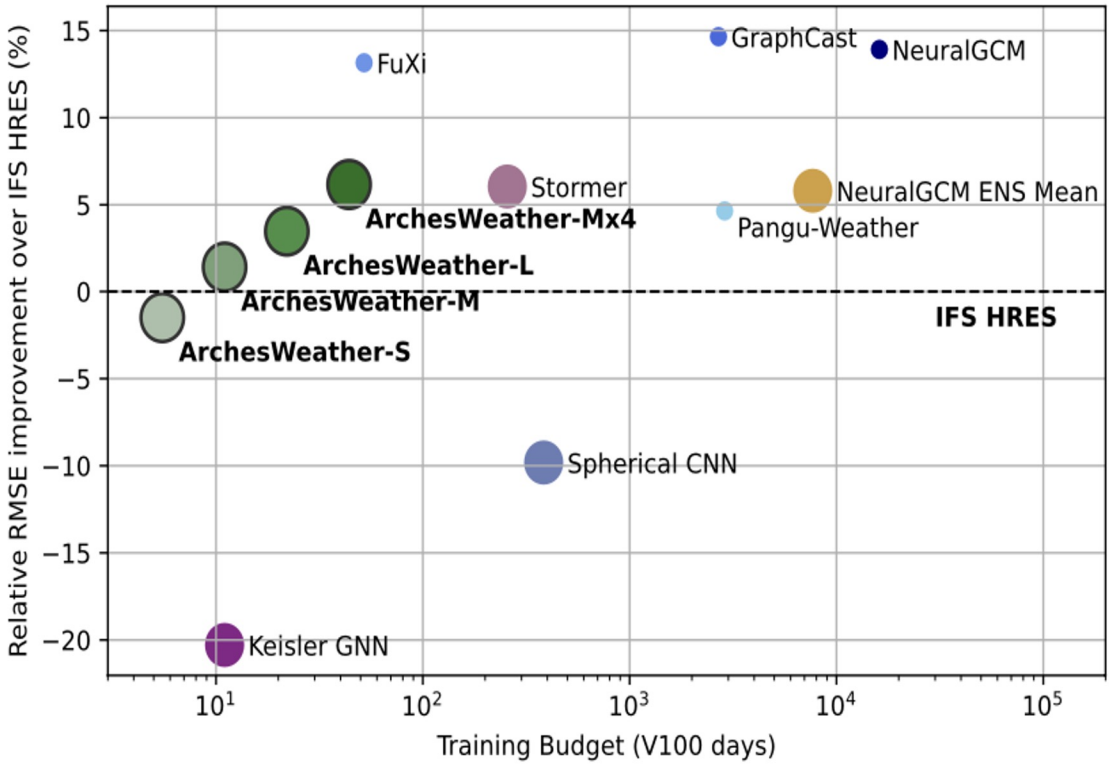
latitude + longitude

altitude

ArchesWeather

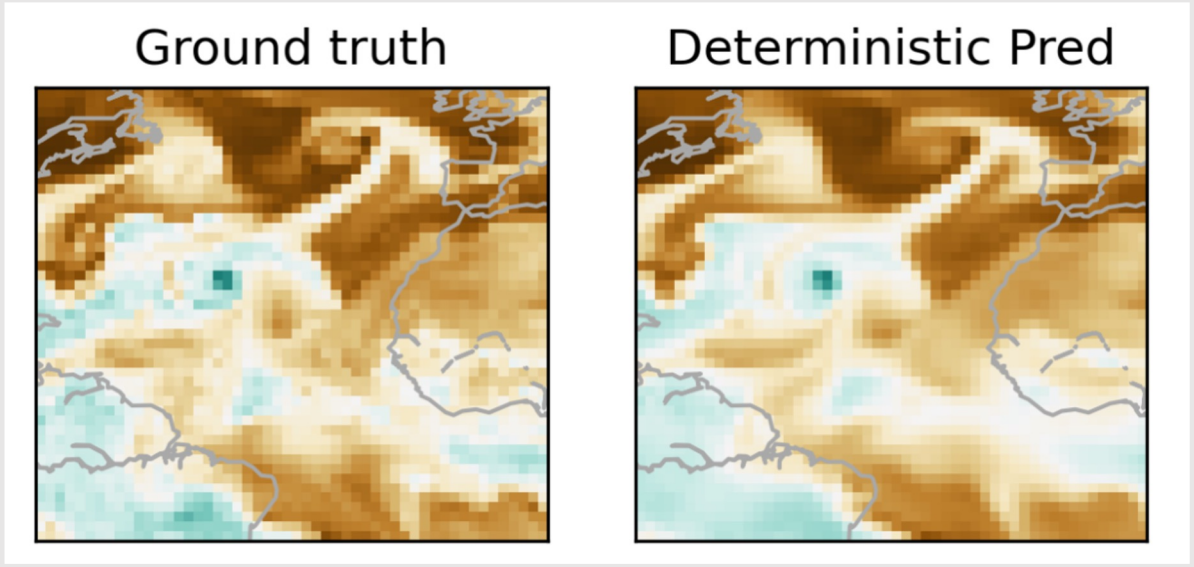
	Res.	Cost	Z500 m^2/s^2	T850 K	Q700 g/kg	U850 m/s	V850 m/s	T2m K	SP Pa	U10m m/s	V10m m/s
IFS HRES			42.30	0.625	0.556	1.186	1.206	0.513	60.16	0.833	0.872
Pangu-Weather	0.25°	2880	44.31	0.620	0.538	1.166	1.191	0.570	55.14	0.728	0.759
NeuralGCM	0.25°	16128	37.94	0.547	0.488	1.050	1.071	N/A	N/A	N/A	N/A
FuXi	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**

Trade-off de desempenho



RMSE, prazo de 24 horas

Por que passar à geração de conjuntos…



Previsão de Q700 com 24 h de antecedência
data de inicialização: 28 set. 2019

- A previsão determinística mostra suavização irreal
- Testar a geração de conjuntos com treinamento por difusão
- Objetivo: cada membro amostrado deve ser mais físico

ArchesWeatherGen

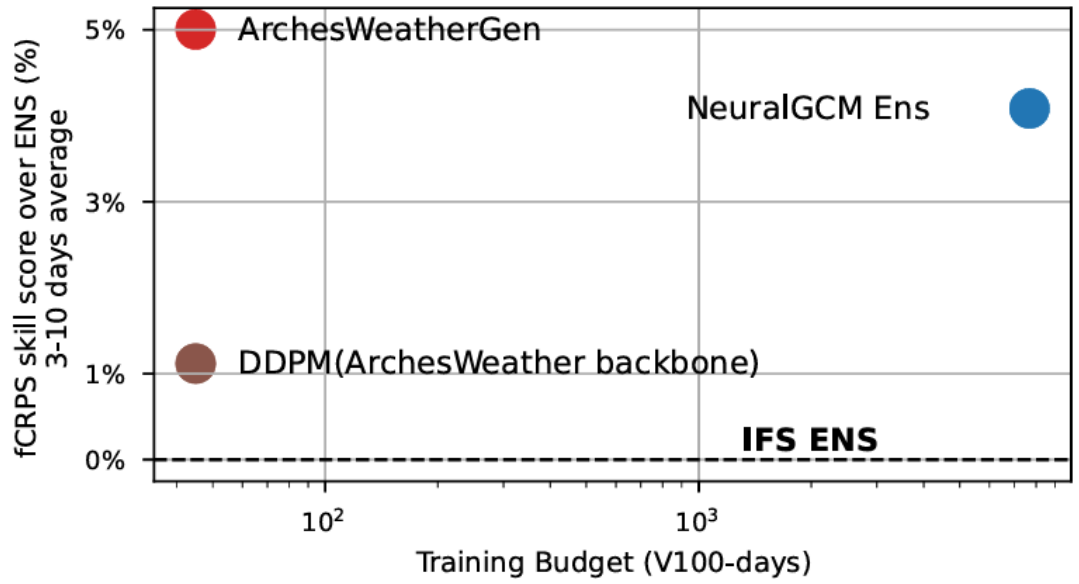
[Couairon et al., *Science Advances*, 2026]

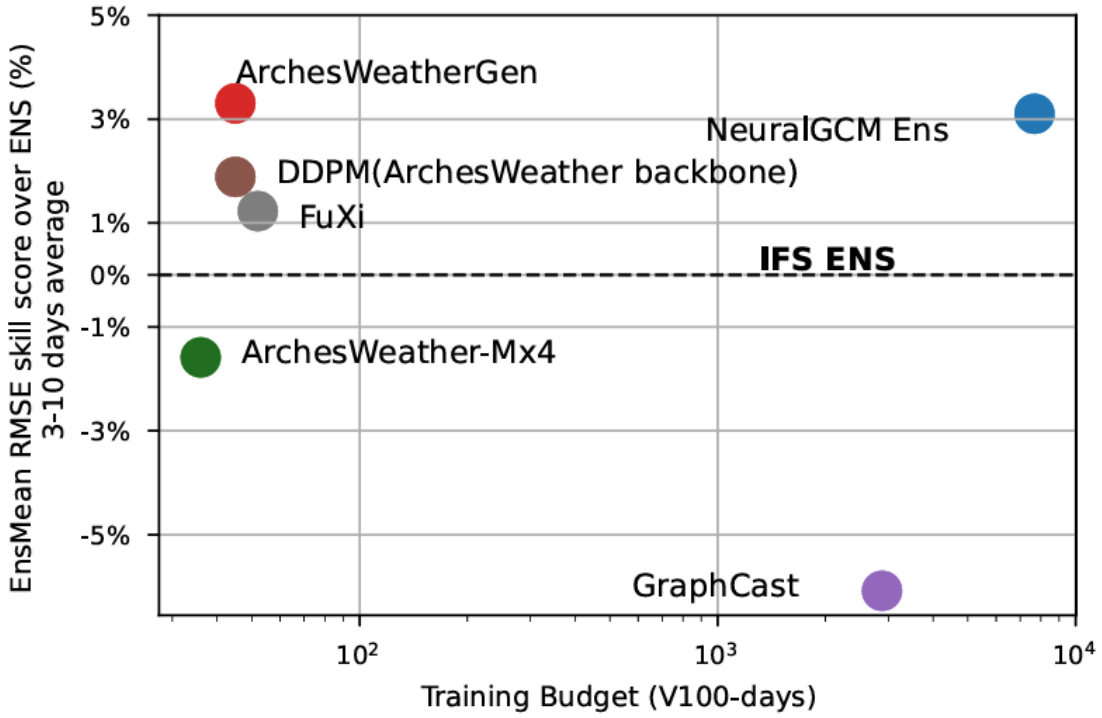
Backbone: ArchesWeather (variante do Swin U-Transformer), 44 M de parâmetros

Treinado com Flow Matching

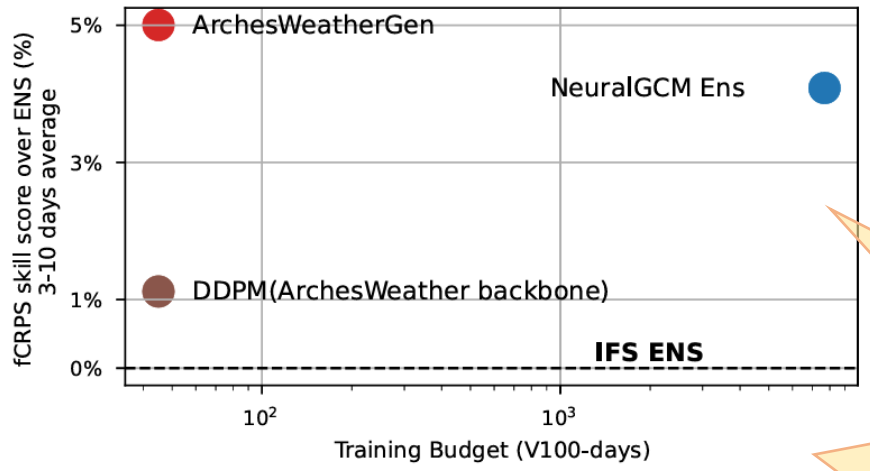
Inferência feita com 25 passos de amostragem a cada 24 horas

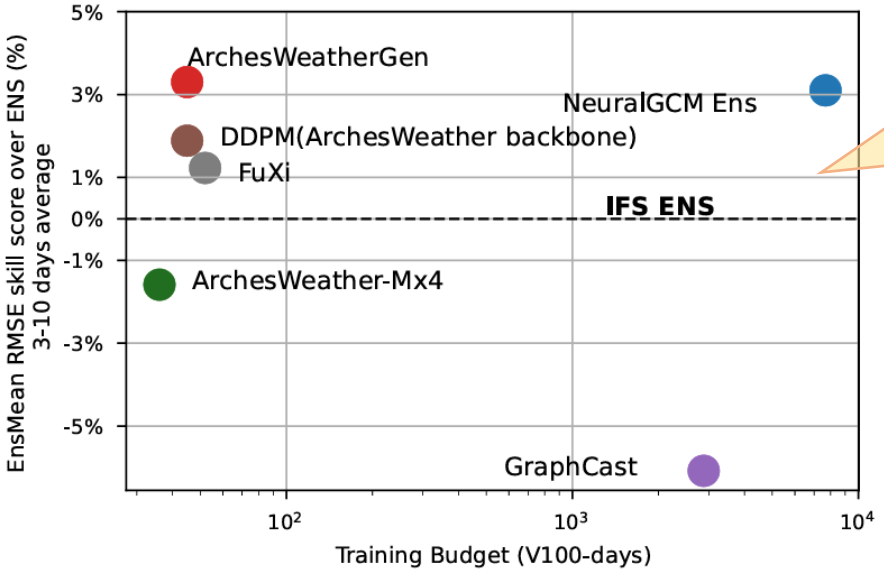
Desdobramentos feitos com uma amostra a cada 24 horas. Essa saída volta ao modelo treinado e o processo se repete.

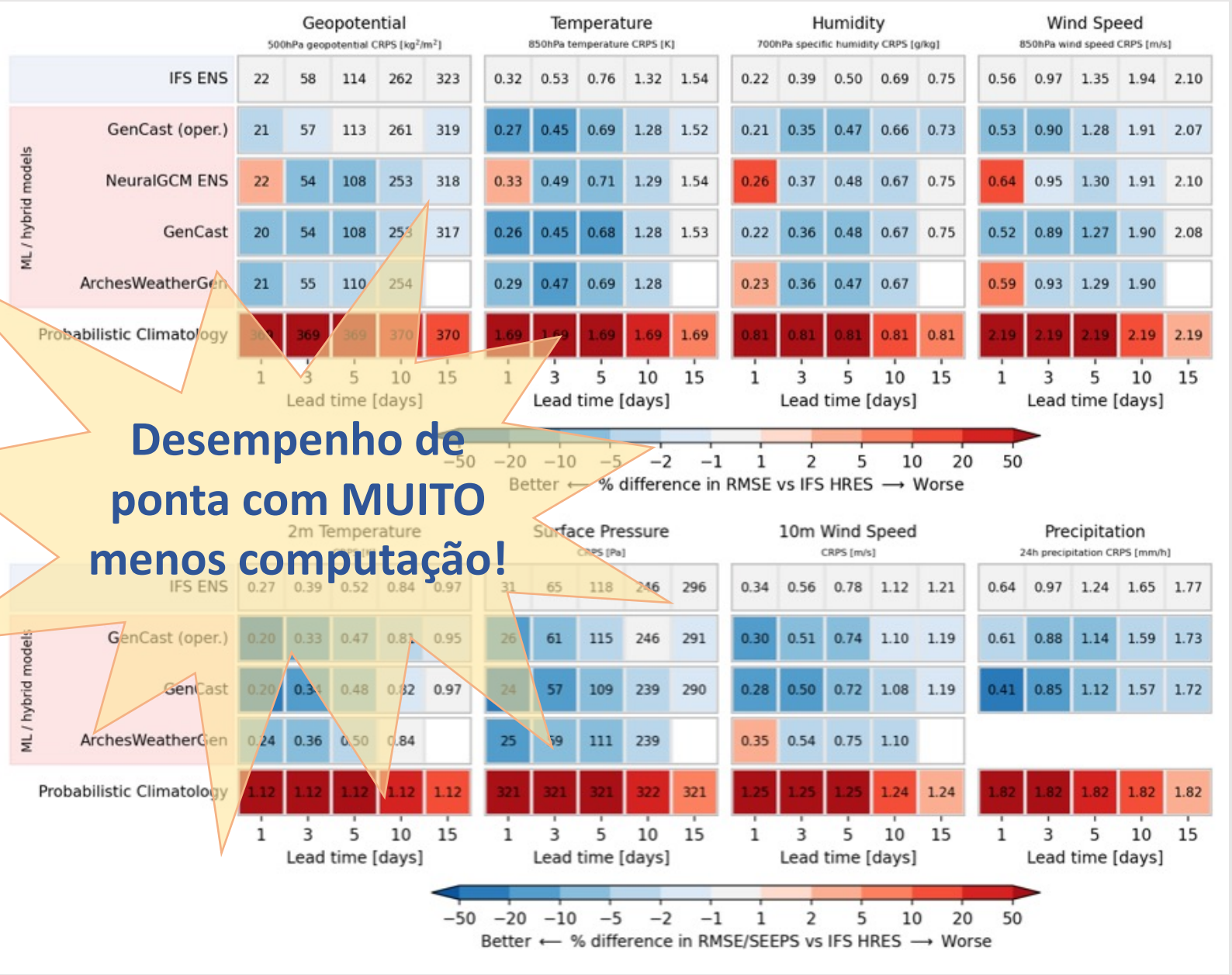




ArchesWeatherGen





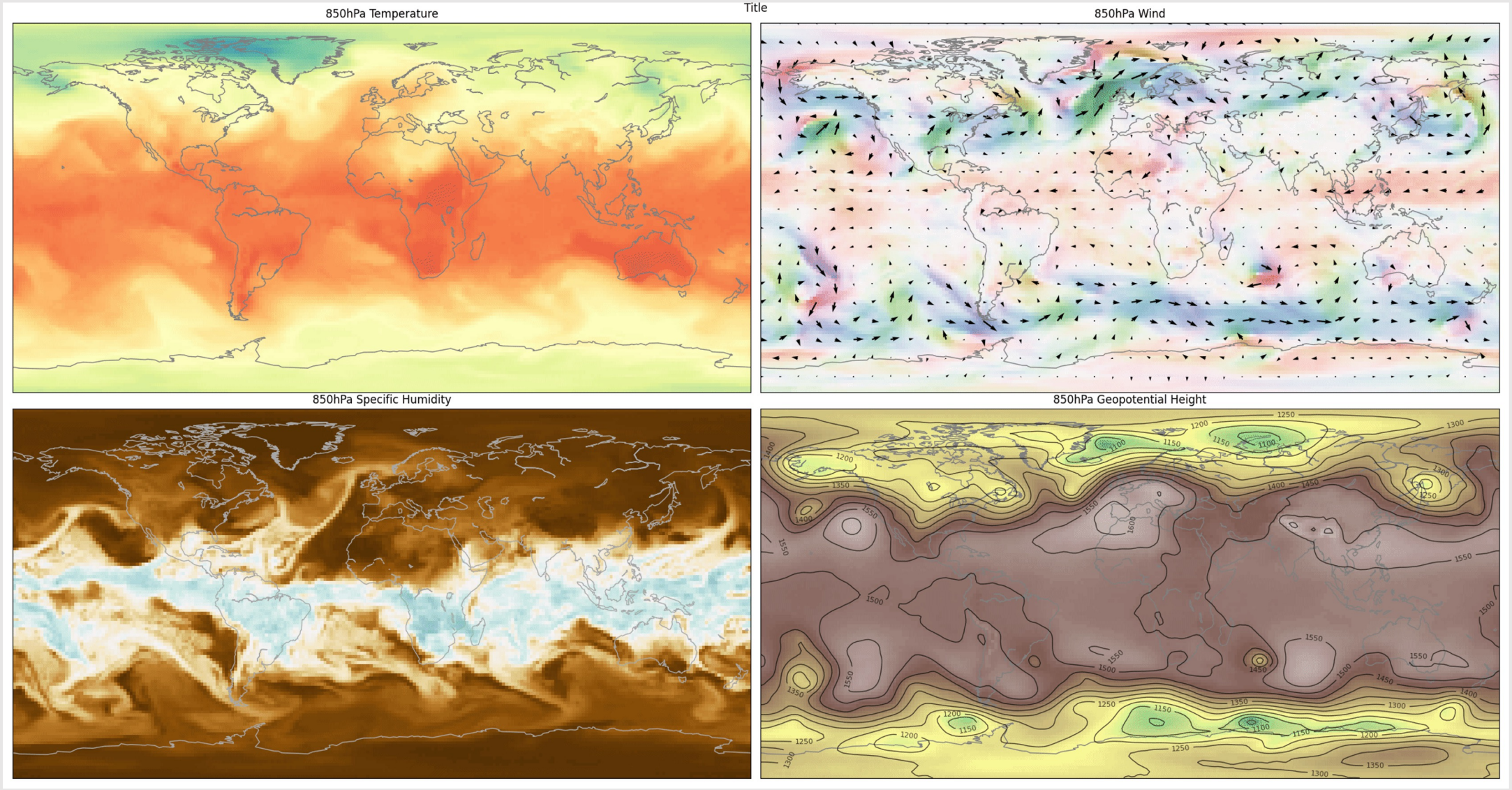


Desempenho de ponta com MUITO menos computação!

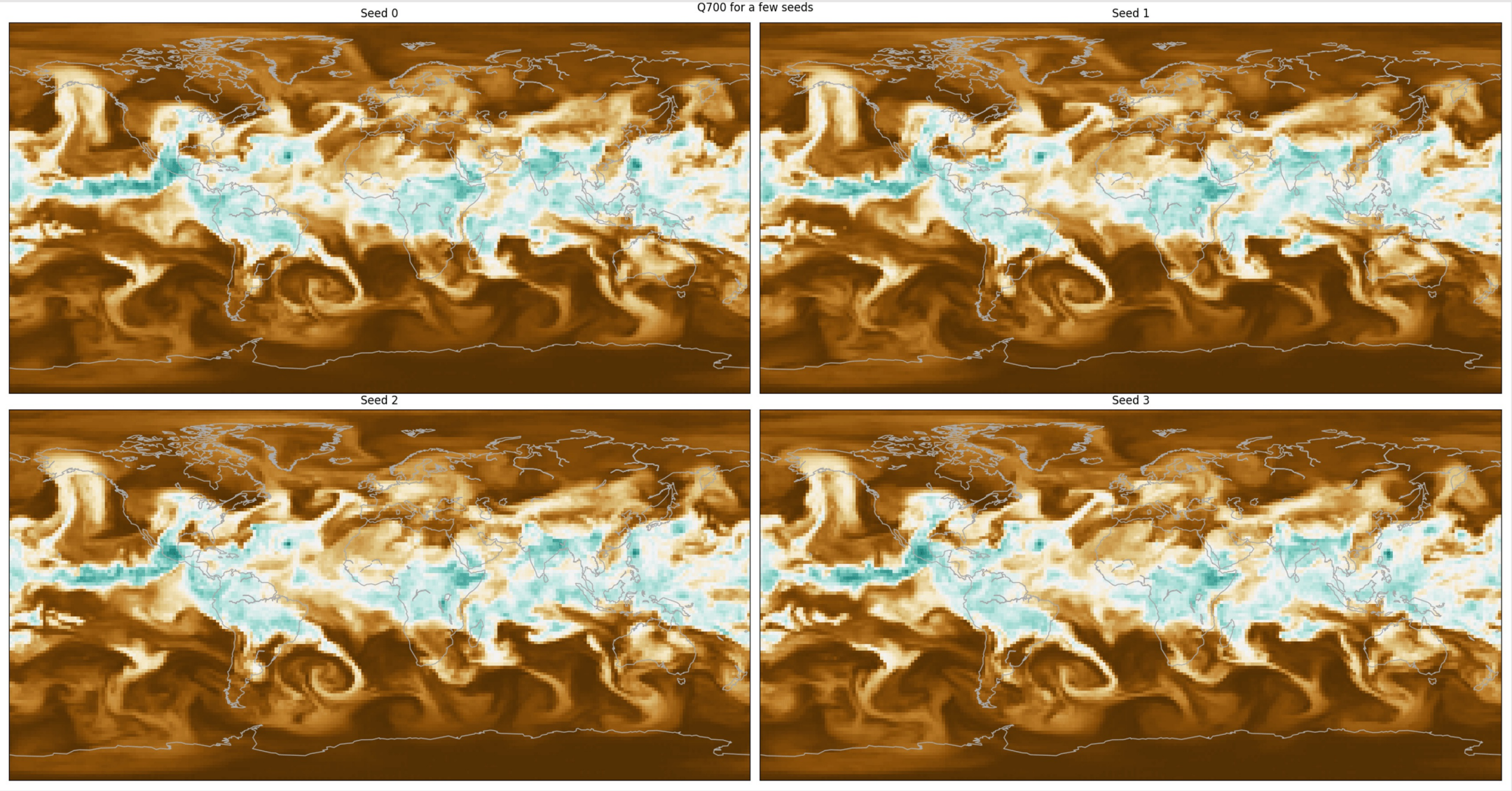
[Couairon, Singh, Charantonis, Lessig, Monteleoni, *Science Advances*, 2026]

Scorecards no WeatherBench 2 do Google

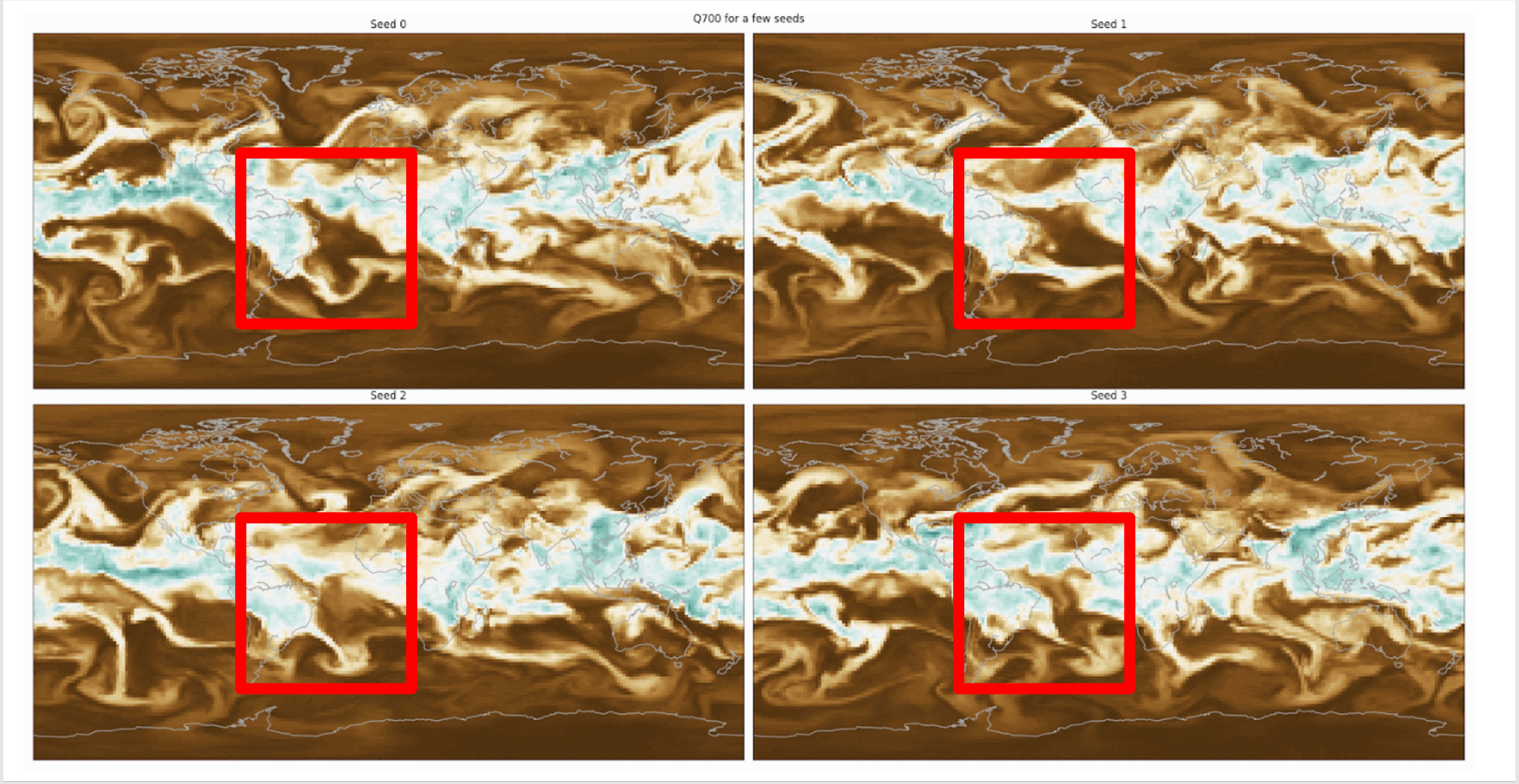
Uma trajetória de exemplo



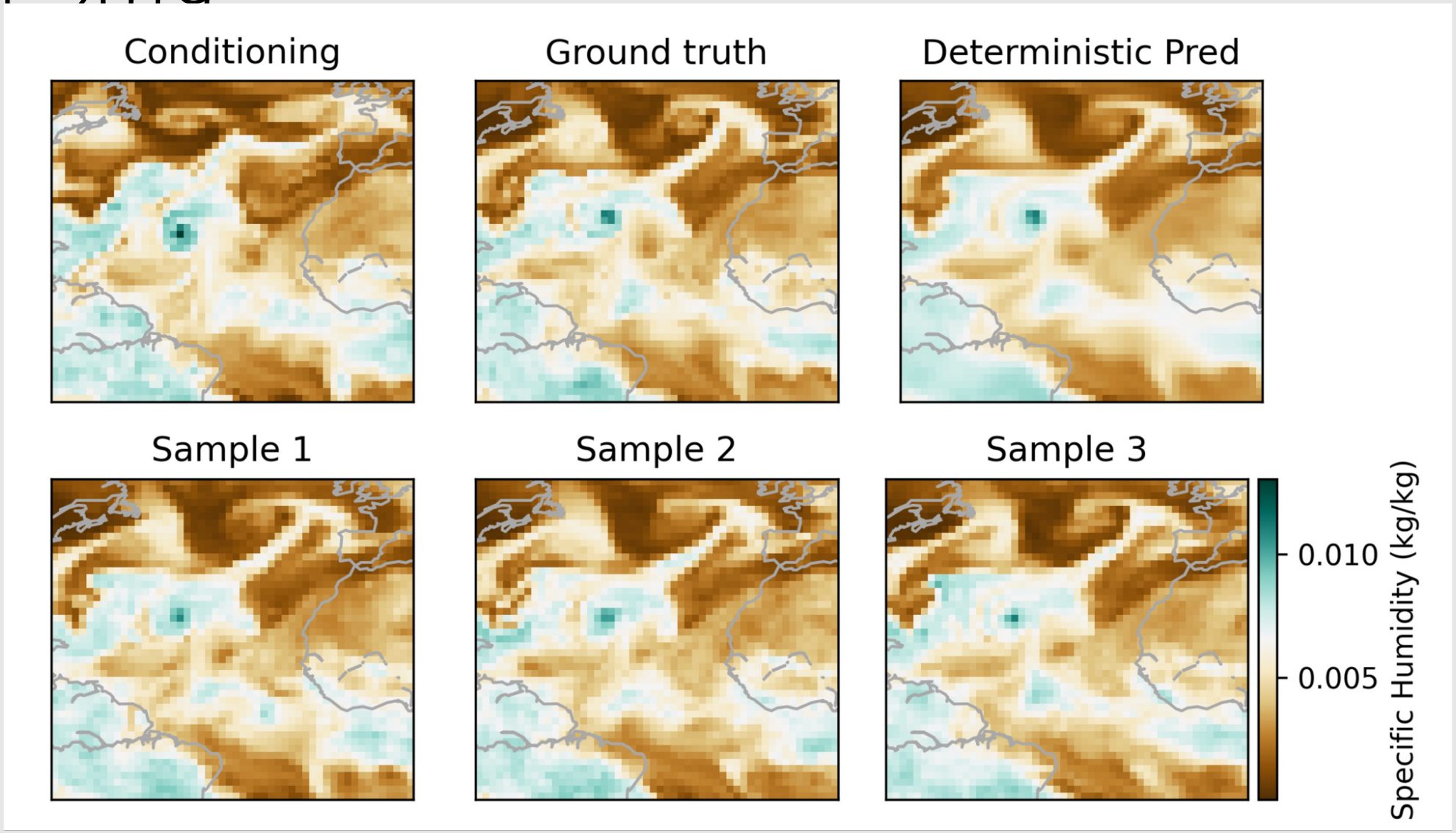
Sementes diferentes - Q700



Diversidade dos membros do conjunto; prazo de 10 dias



Estudo de caso: Lorenzo - Q700 - inicialização em 25 set. 2019



ArchesClimate: emulação de modelos climáticos com IA



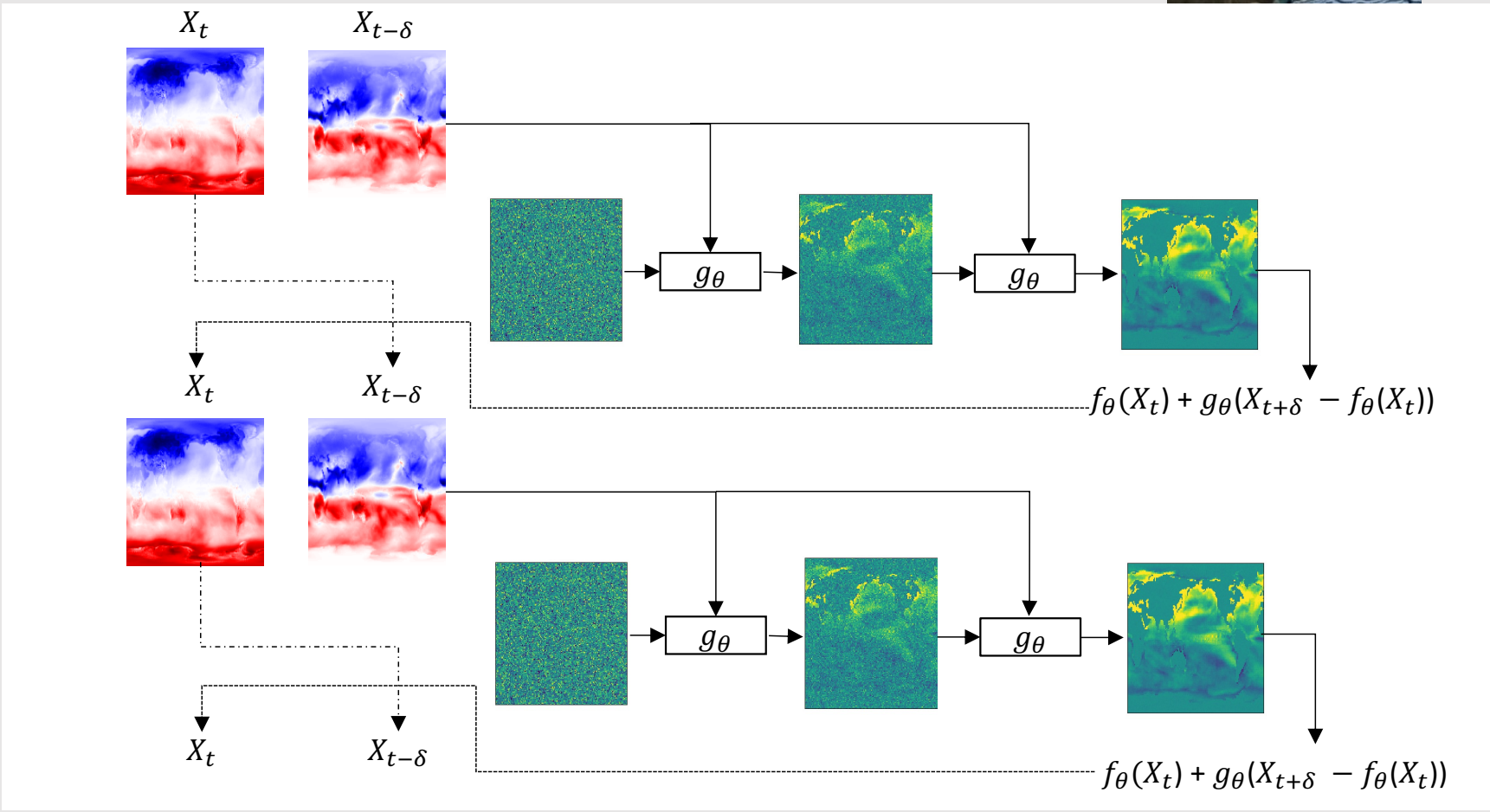
Doutorado em andamento de Graham Clyne
Orientador principal: Anastase Charantonis

Pipeline: ArchesWeatherGen modificado:
- Sem atenção 1D por coluna
- Inclusão das forçantes

Dados de treinamento: simulações de modelos climáticos (IPSL DCPP)

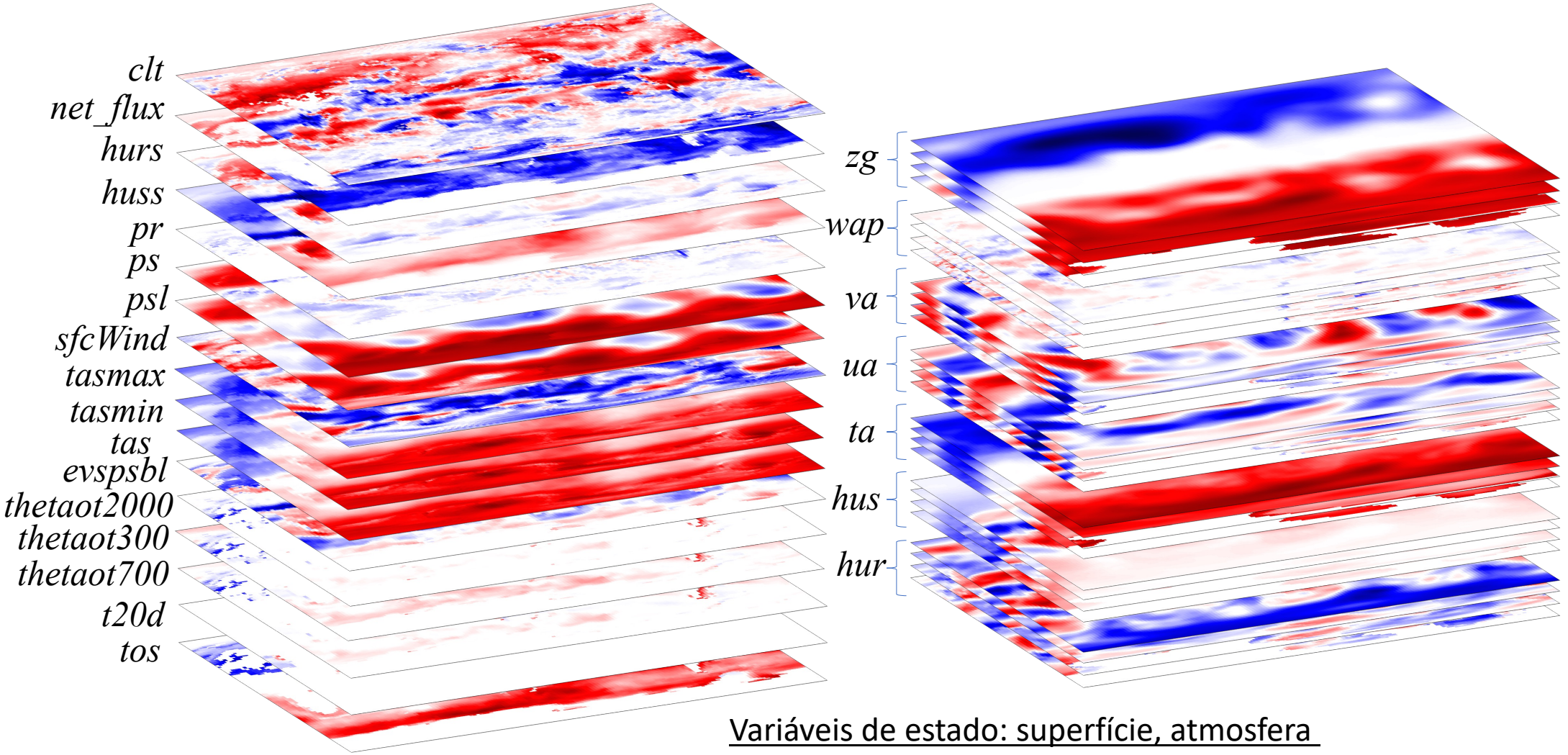
Treinamento: prever um mês à frente

Desdobramentos feitos com uma amostra por mês. Essa saída volta ao modelo treinado e o processo se repete.



Clyne et al., *resultados preliminares apresentados na EGU 2025 e no Climate Informatics 2025*

Variáveis de entrada e forçantes

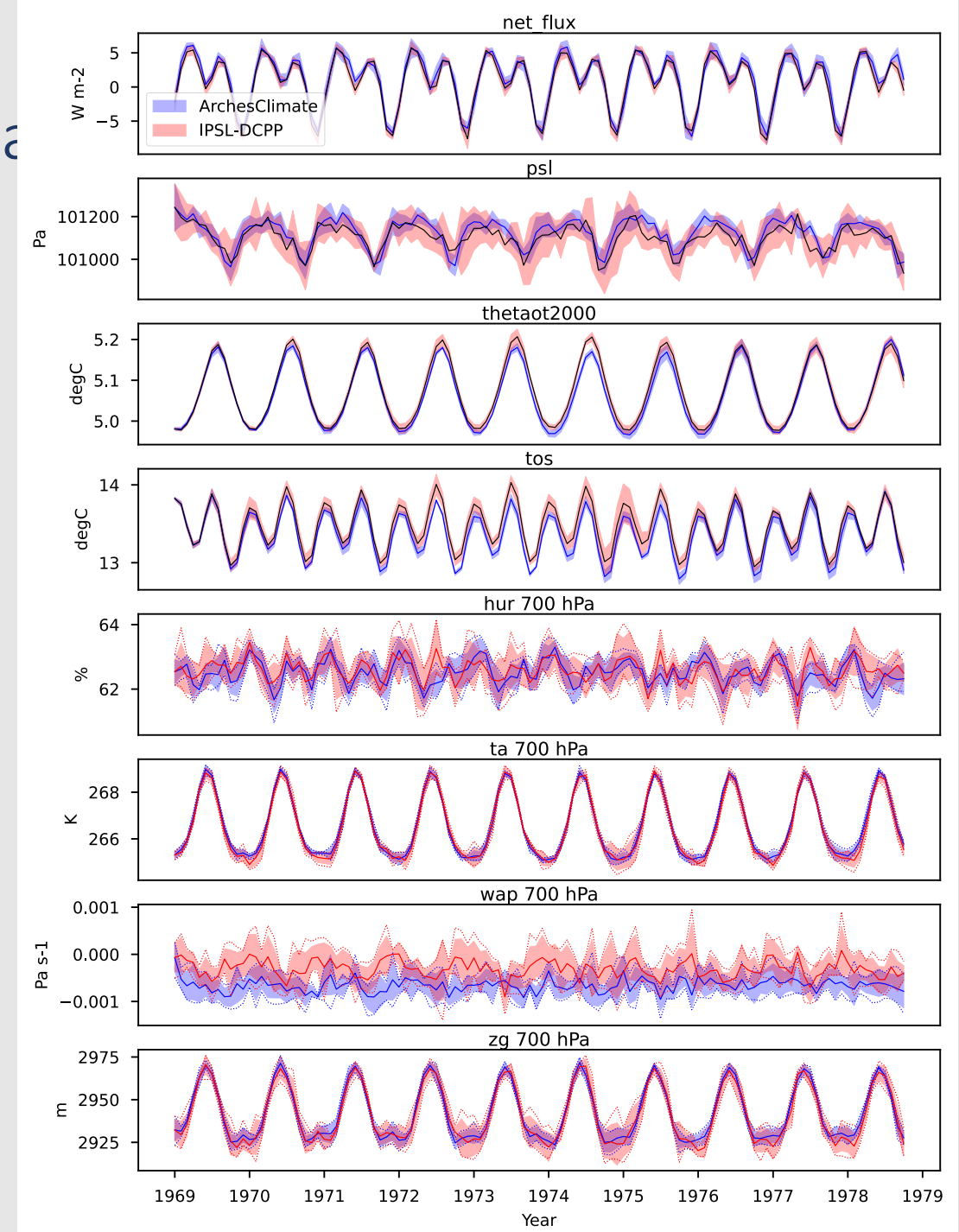


$$\begin{bmatrix} CO2 \\ CH4 \\ N2O \\ CFC11eq \\ SSI \end{bmatrix}$$

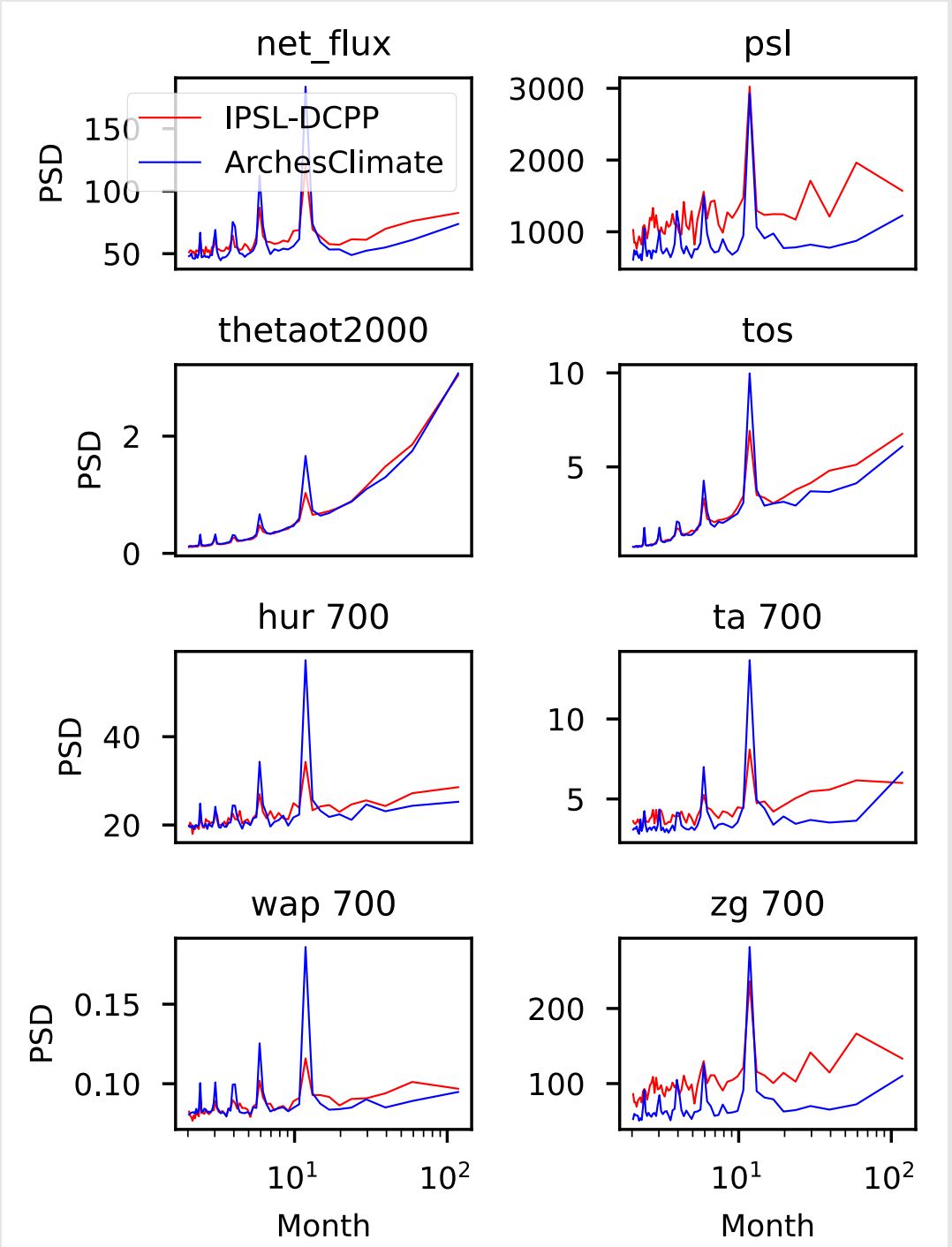
Variáveis de estado: superfície, atmosfera

Forçantes

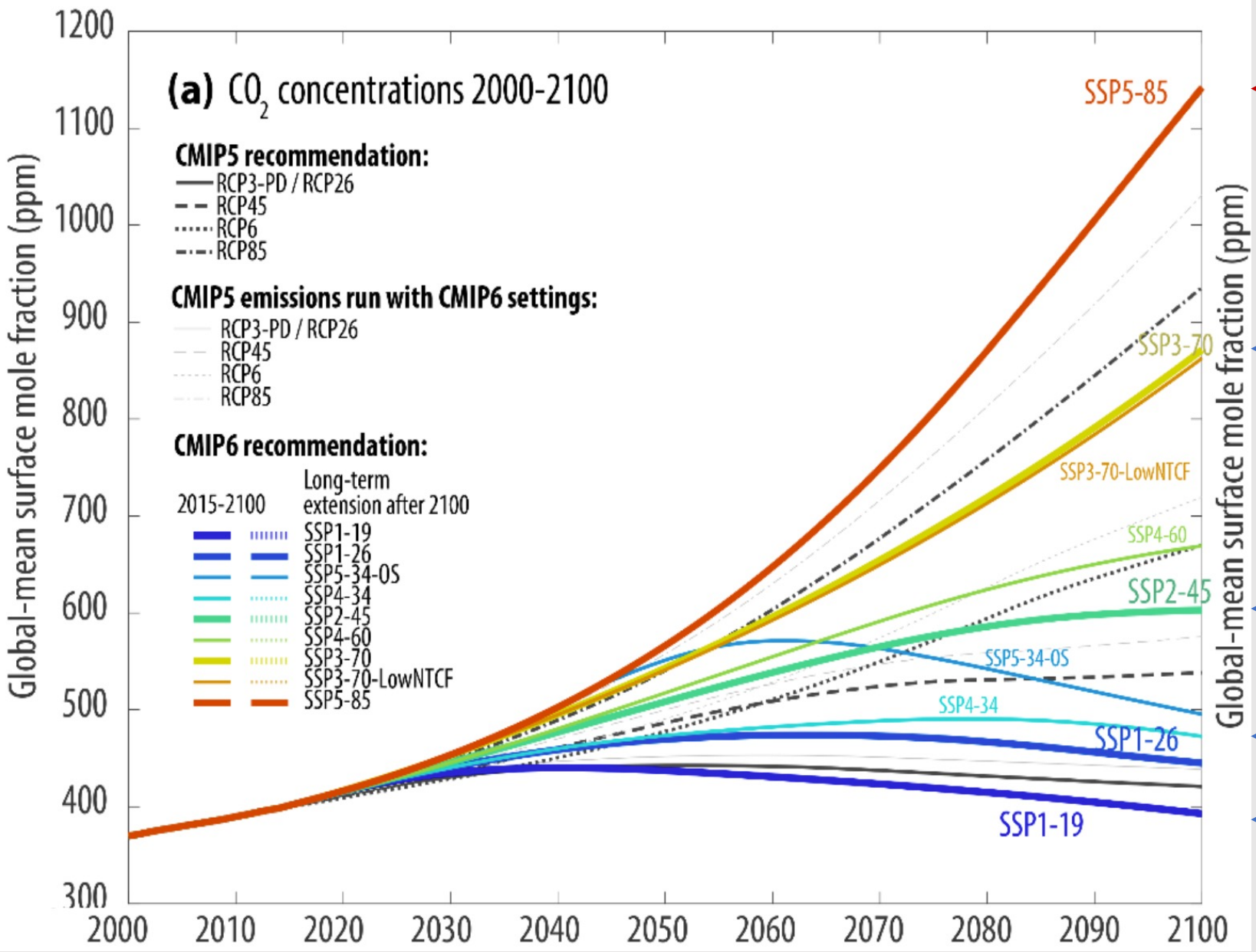
Desdobramentos decena



Resultados preliminares



Extrapolando cenários de modelos climáticos



Dados de teste

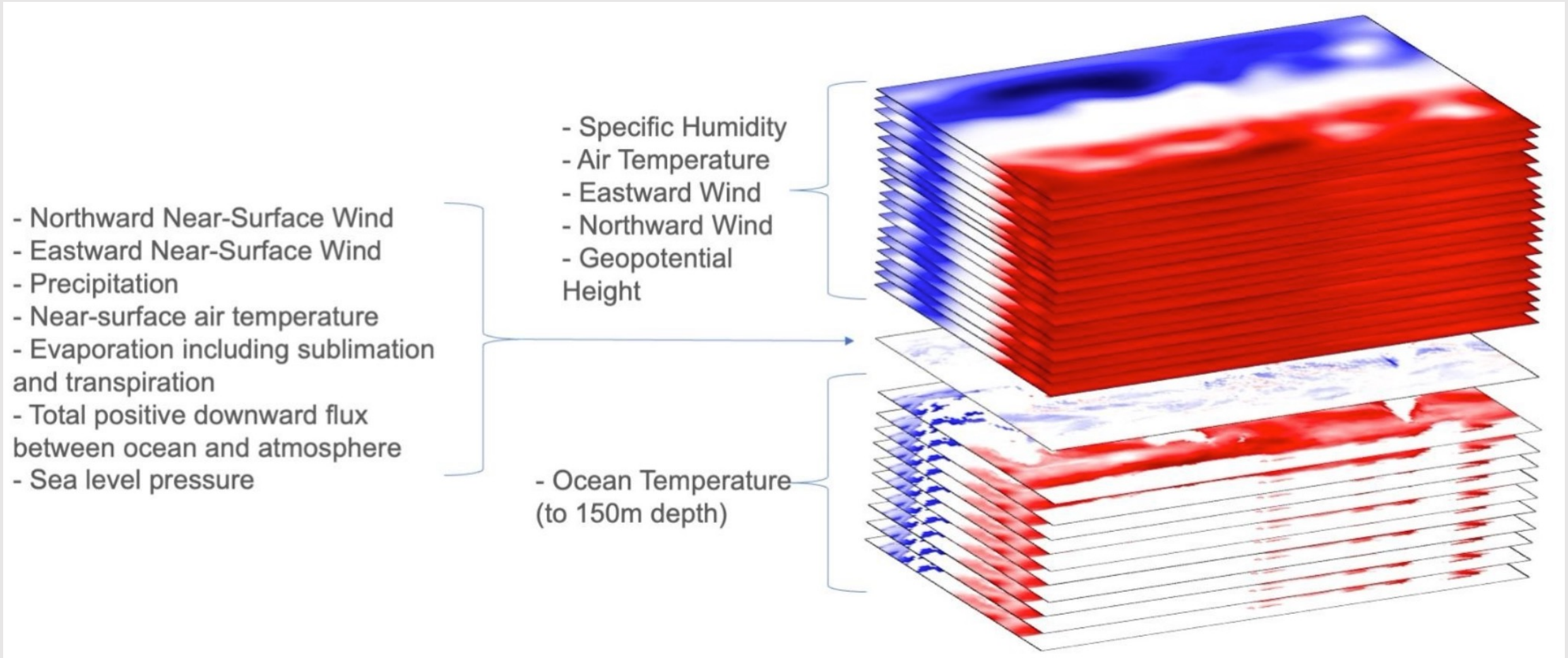
Tarefa de extrapolação

Dados de treinamento

Estes 4 cenários

Interpolação e extrapolação no espaço de cenários de modelos climáticos

Clyne et al., *em preparação*



Extrapolação de cenários

Reproduzir os padrões temporais do El Niño-Oscilação Sul

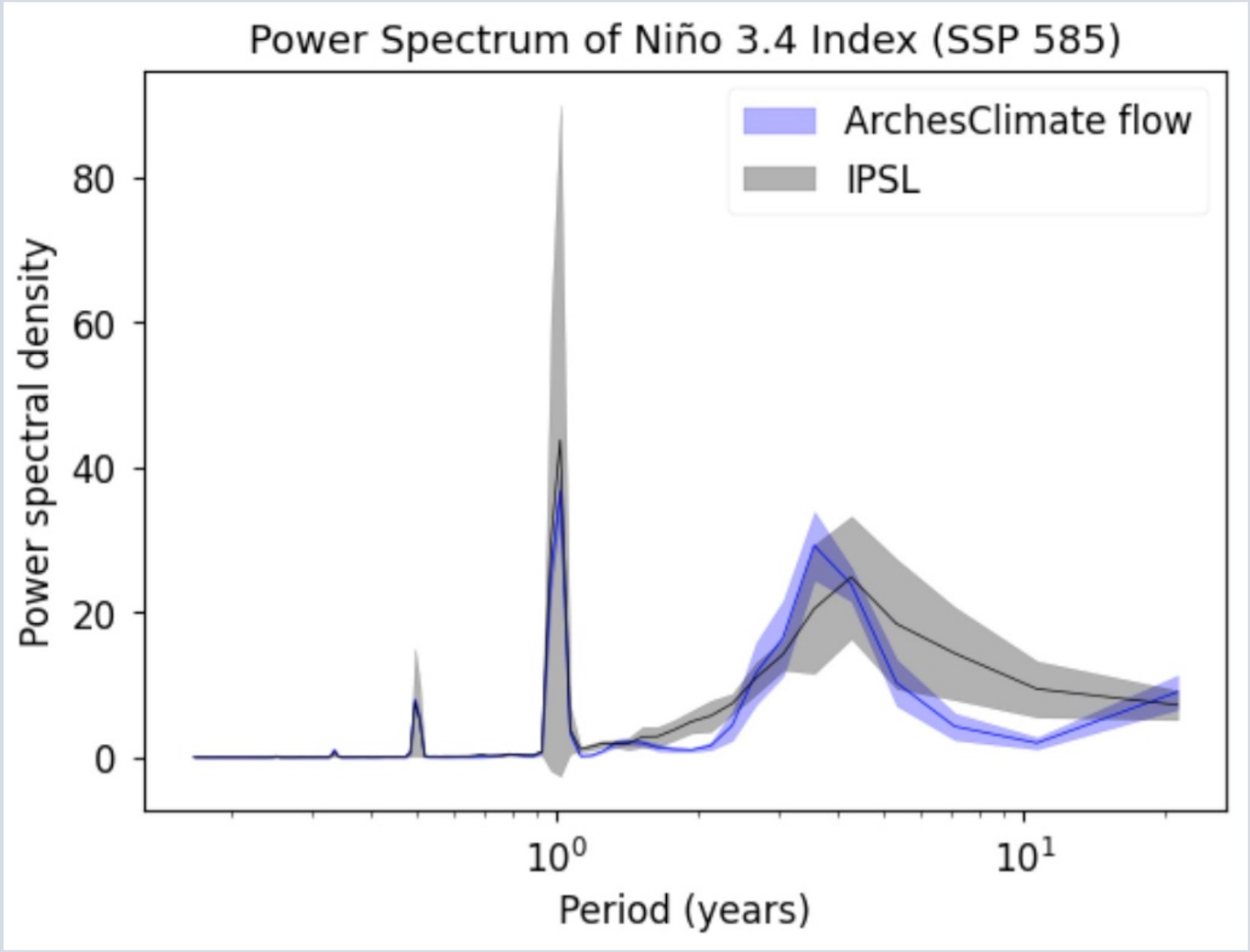


Figure 9. Power spectra for El Niño 3.4 index for SSP5-8.5.

Clyne et al., *em preparação*

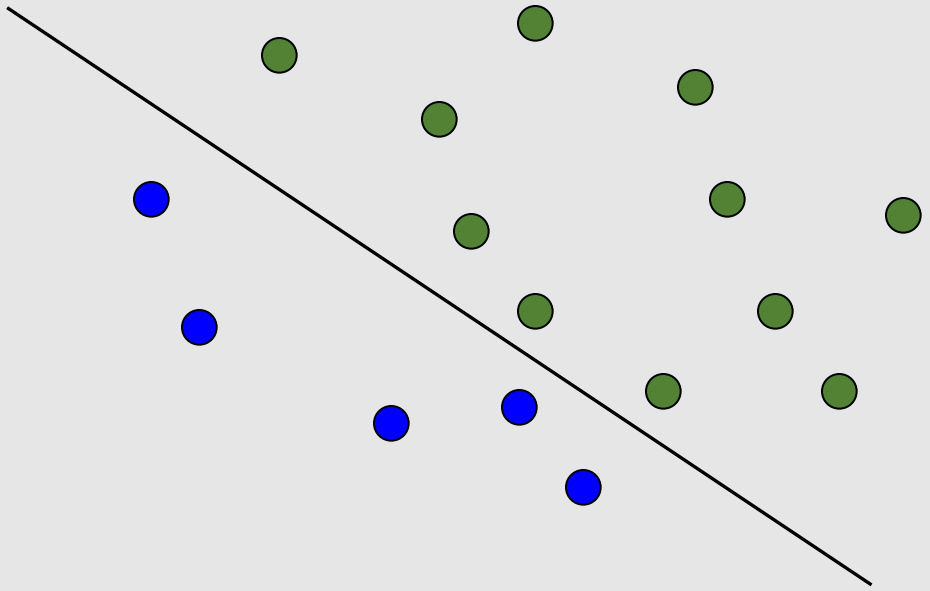
Parte II: Métodos de IA: aprendizado autossupervisionado

Aprendizado supervisionado

Dados pontos rotulados, encontrar uma boa regra de classificação.

Descreve os dados

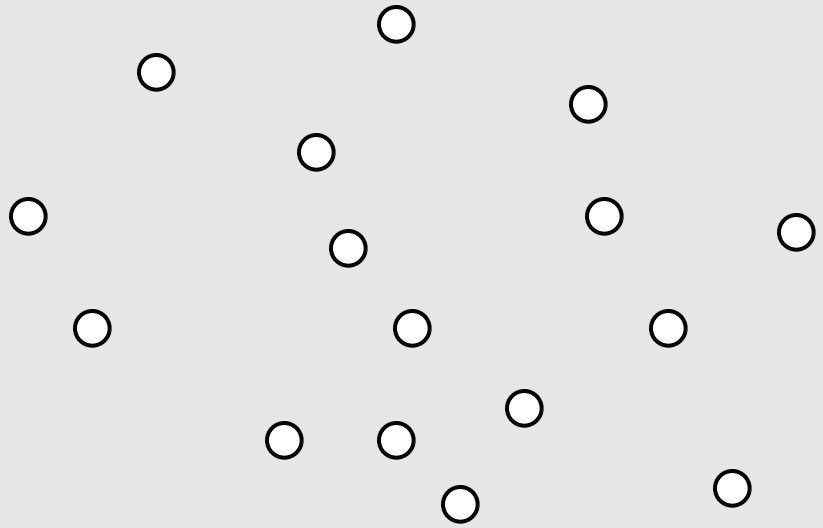
Generaliza bem



Aprendizado não supervisionado

O que dá para fazer sem nenhum rótulo?

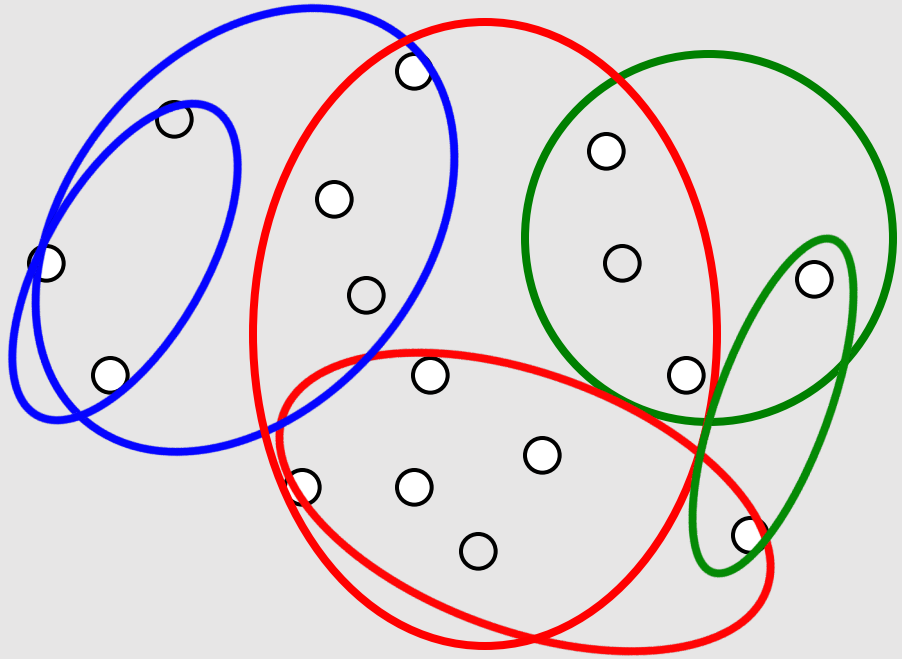
Aprendizado não supervisionado



Aprendizado não supervisionado

O que dá para fazer sem nenhum rótulo?

Aprendizado não supervisionado



Exemplo: agrupamento (clustering)

Aprender com dados “não rotulados”

- Tarefas de aprendizado não supervisionado
  1. Agrupamento
  2. Embedding (redução de dimensionalidade, incluindo métodos não lineares)
  3. Interpolação
  4. Aprendizado de um modelo generativo

- Há abordagens de aprendizado profundo (AP) para todas essas tarefas

- Há abordagens sem AP para todas essas tarefas

- Há abordagens de AP autossupervisionadas para todas essas tarefas

Aprendizado profundo não supervisionado

- AP supervisionado. A perda de previsão é função do rótulo y e da saída da rede para a entrada x.

Saída da rede

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

Função de perda

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

- AP não supervisionado. A perda de previsão só depende de x e da saída da rede para a entrada x. Não há rótulo y.

Saída da rede

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

Função de perda

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

Abordagem autossupervisionada do aprendizado não supervisionado

Aprendizado autossupervisionado

Uma abordagem de ponta para o aprendizado (profundo) não supervisionado

Criar uma tarefa pretexto:

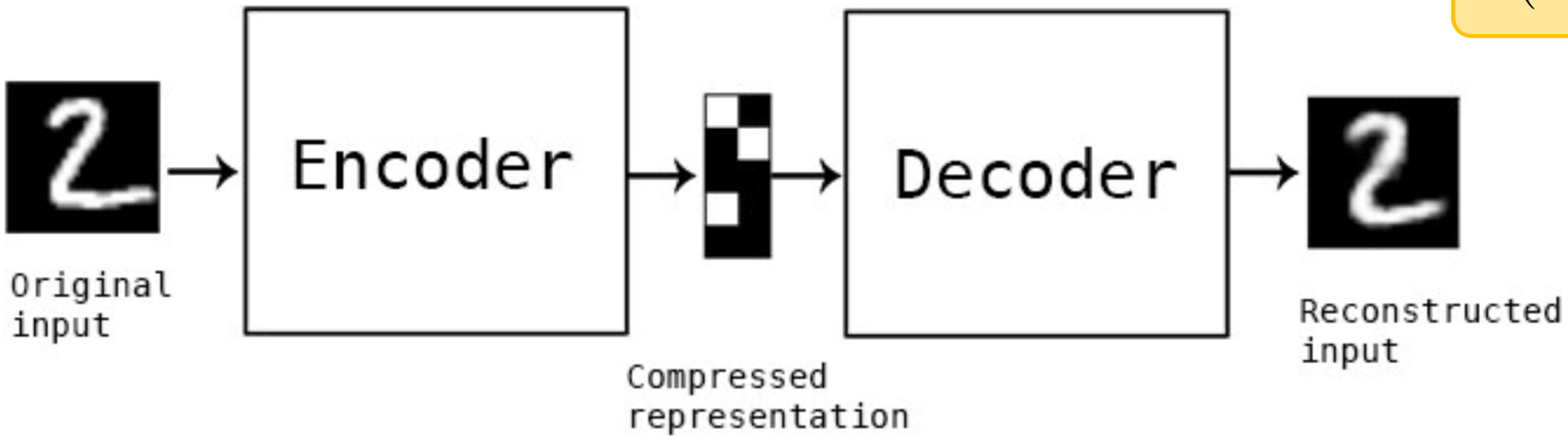
- Criar uma tarefa de aprendizado supervisionado usando só os dados disponíveis.
- Treinar um modelo nessa tarefa de modo que
- as características aprendidas (ou a posterior aprendida sobre um espaço de características) sejam úteis para outra tarefa (posterior)

Tarefa pretexto: exemplo

Exemplo clássico de tarefa pretexto: o autoencoder

- Treinar uma rede neural de forma não supervisionada
  - Usar os dados não rotulados como entrada e para avaliar a saída
- Após o treinamento, a camada de gargalo será uma representação compacta da distribuição de entrada

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



Aprendizado autossupervisionado

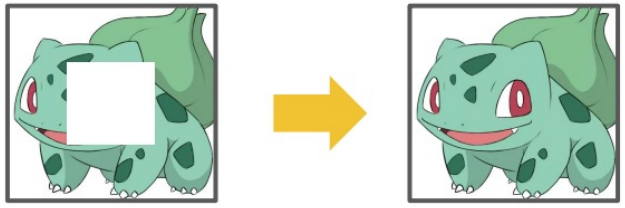
- Criar uma tarefa pretexto que faça aprender características úteis.

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



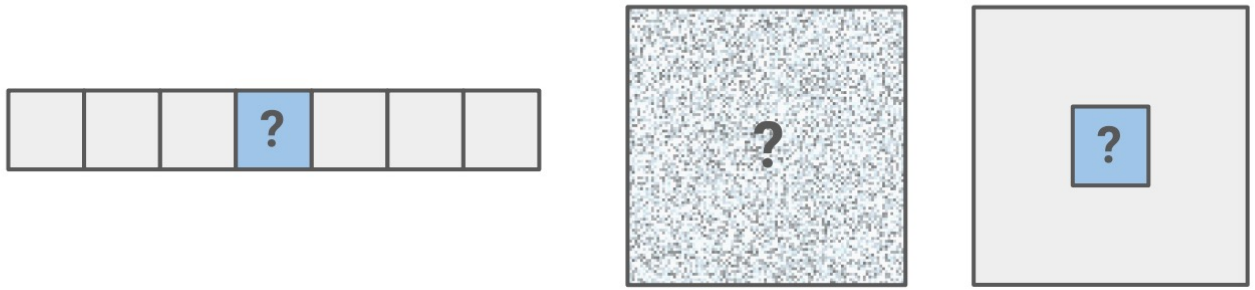
Lilian Weng, Jong Wook Kim, tutorial NeurIPS 2021

Aprendizado autossupervisionado

- Criar uma tarefa pretexto que faça aprender características úteis.

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)

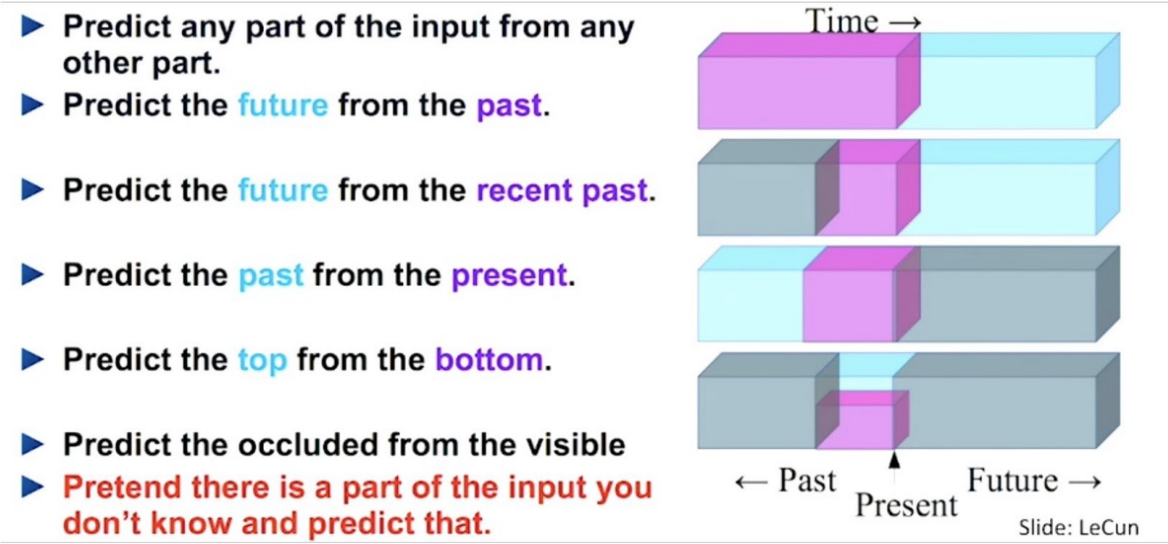
Lilian Weng, Jong Wook Kim, tutorial NeurIPS 2021

Aprendizado autossupervisionado

- Criar uma tarefa pretexto que faça aprender características úteis.

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, tutorial NeurIPS 2021

Aprendizado autossupervisionado: aprendizado contrastivo

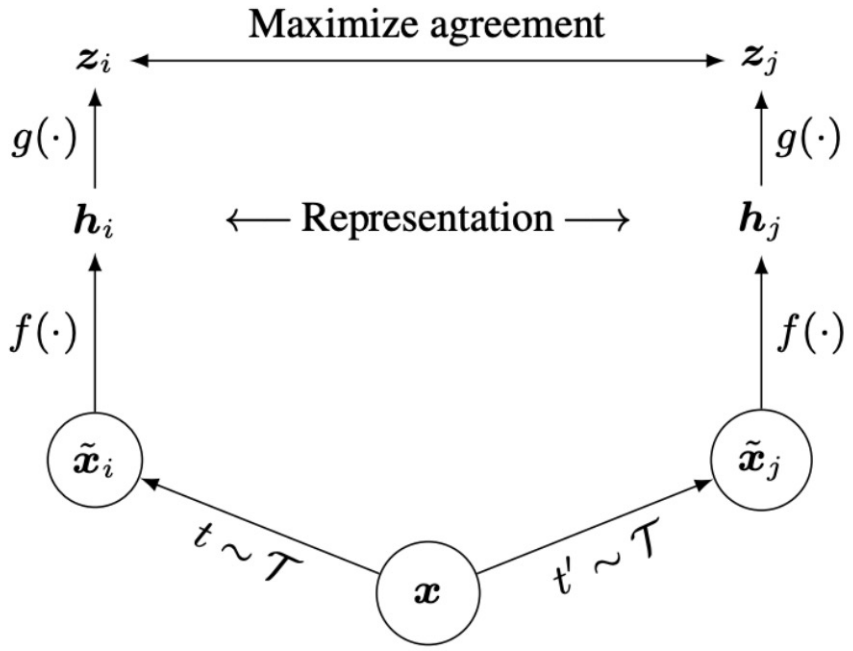


Figure 2. A simple framework for contrastive learning of visual representations. Two separate data augmentation operators are sampled from the same family of augmentations ($t \sim \mathcal{T}$ and $t' \sim \mathcal{T}$) and applied to each data example to obtain two correlated views. A base encoder 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.

Tarefa pretexto:

1. Aumentar a entrada x duas vezes: $\tilde{x}_i, \tilde{x}_j$
2. Aprender um mapeamento $g(f(\cdot))$, minimizando uma função de perda para aproximar $g(f(\tilde{x}_i)), g(f(\tilde{x}_j))$.

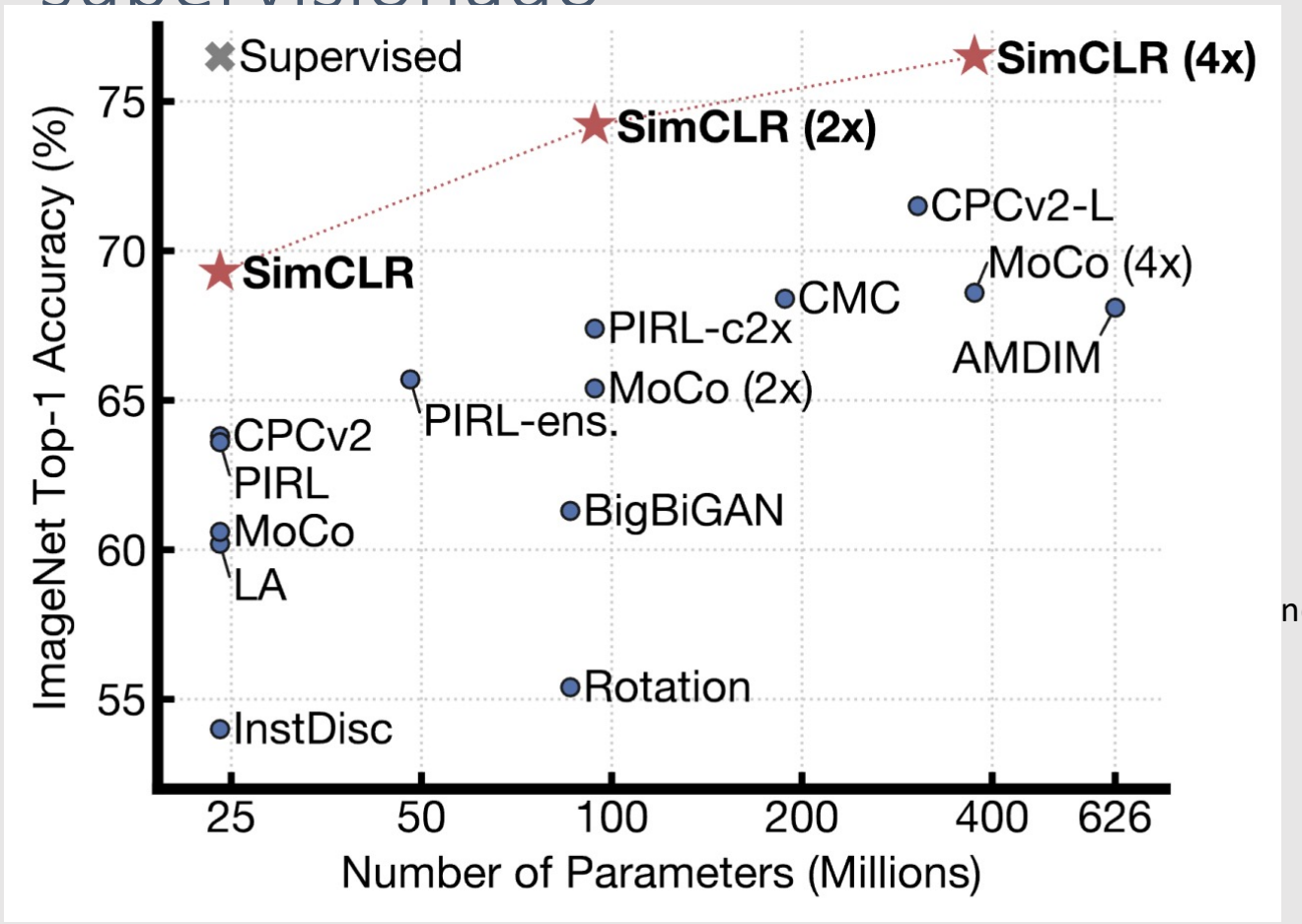
$$\mathcal{L}(\hat{x}, x)$$

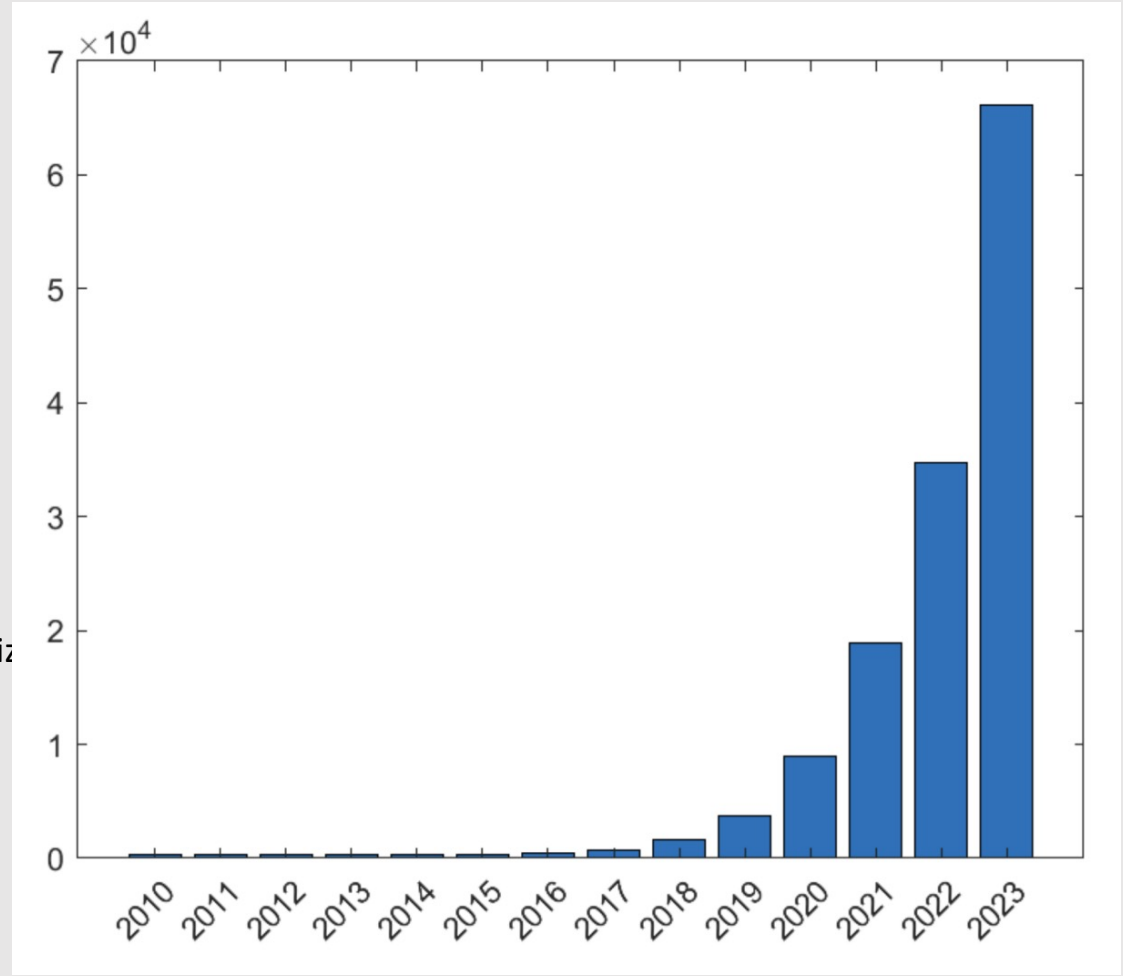
$$\hat{x} := \{g(f(\tilde{x}_i)), g(f(\tilde{x}_j))\}$$

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

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

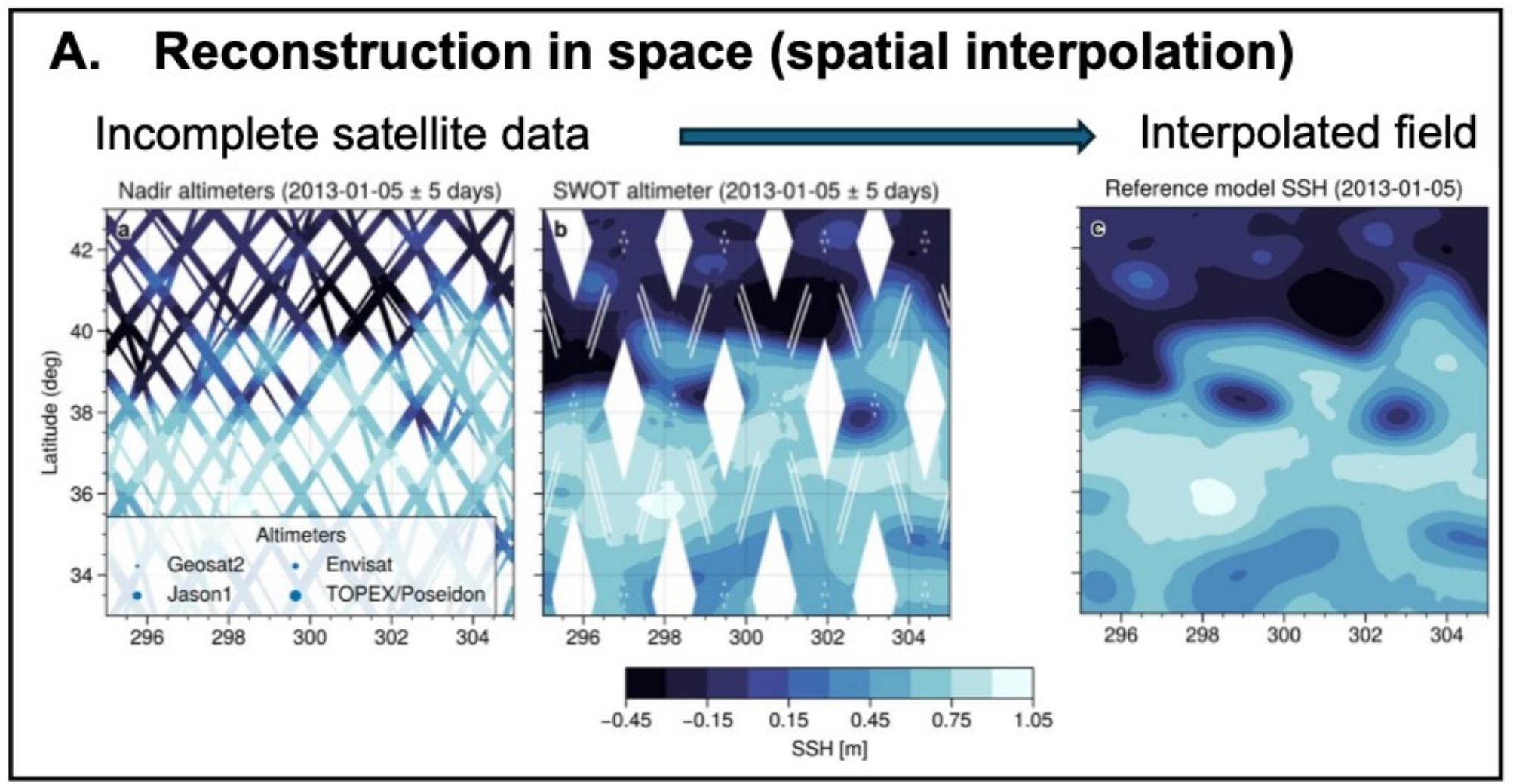
O aprendizado autossupervisionado pode superar o supervisionado





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

Autossupervisionado para suas aplicações



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

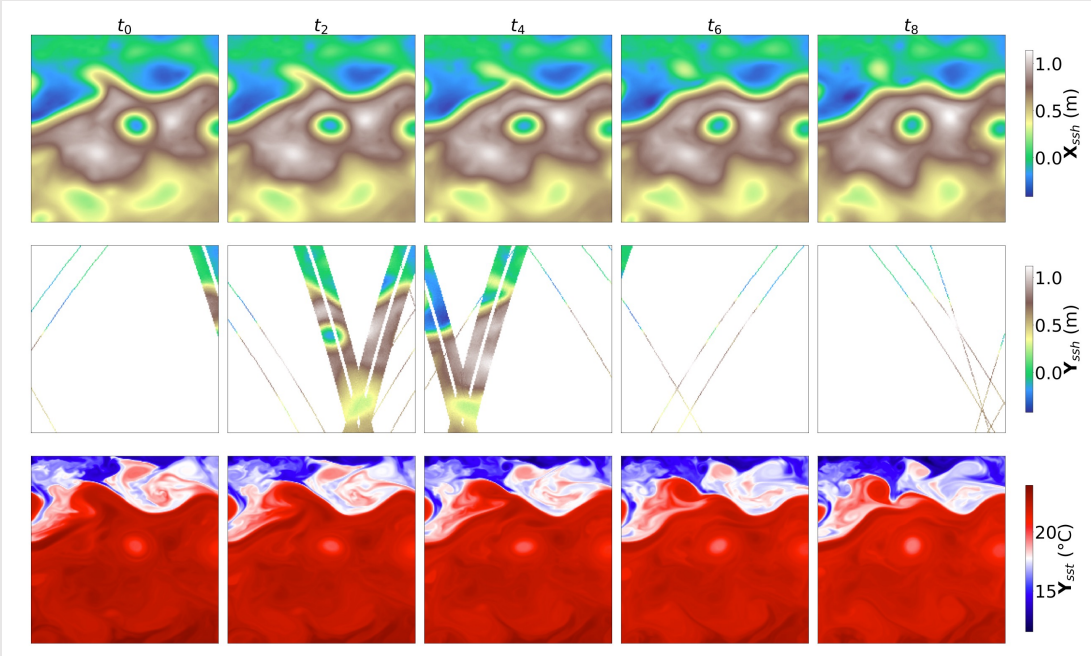
Interpolação espaço-temporal não supervisionada

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

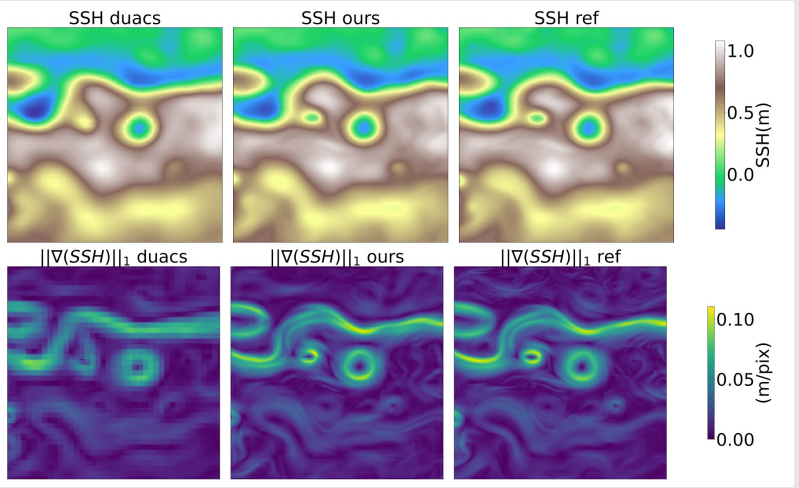
Campos SSH de referência (não usados no treinamento)

Dados SSH ao longo das trilhas dos satélites

Campos SST de um modelo físico



Dados de treinamento



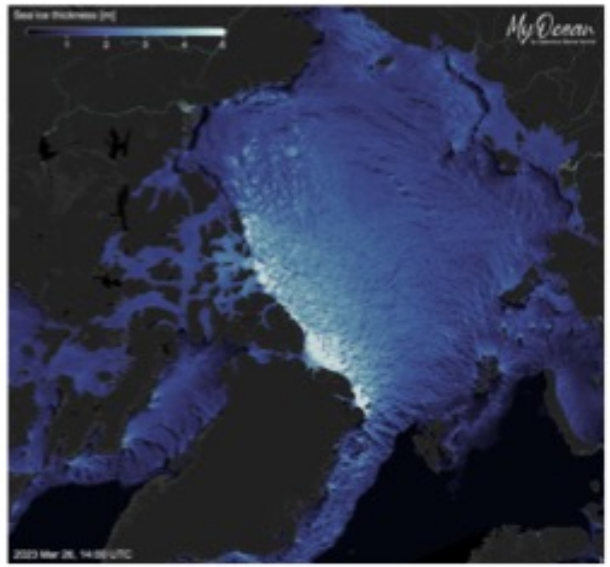
Resultados

Autossupervisionado para suas aplicações

C. Reconstruction in scales (downscaling)

Low resolution field → High resolution field





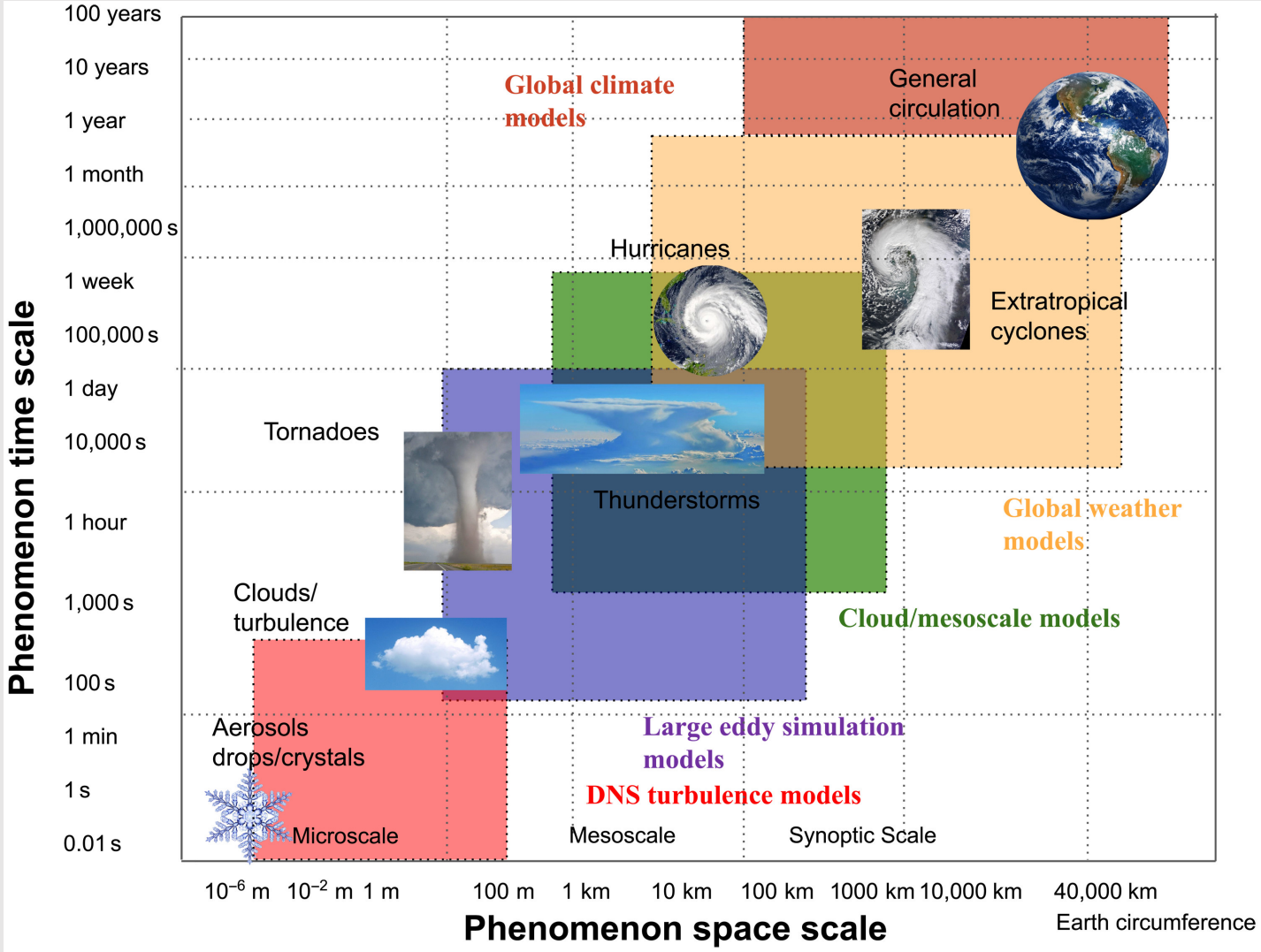
[Bracco et al., Machine Learning for the Physics of Climate. *Nature Reviews Physics*, 2024]

Downscaling de simulações de modelos climáticos

As simulações dos modelos climáticos globais são mais grossas (no espaço e no tempo) do que o necessário para várias tarefas de:

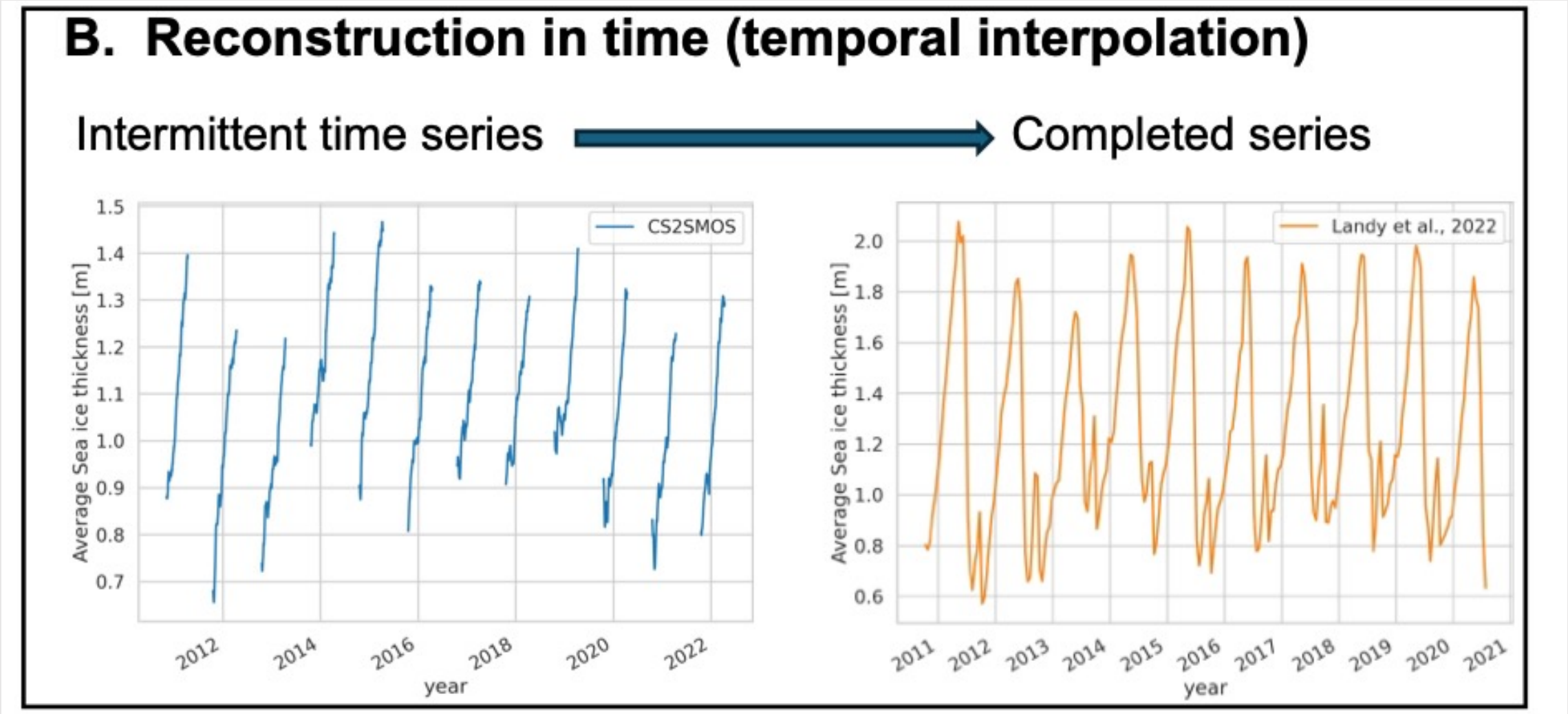
- Adaptação à mudança climática
- Mitigação da mudança climática
- Projeção de impactos de longo prazo

Abordagem: usar IA generativa para fazer o downscaling dos dados de modelos climáticos para escalas relevantes



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

Autossupervisionado para suas aplicações



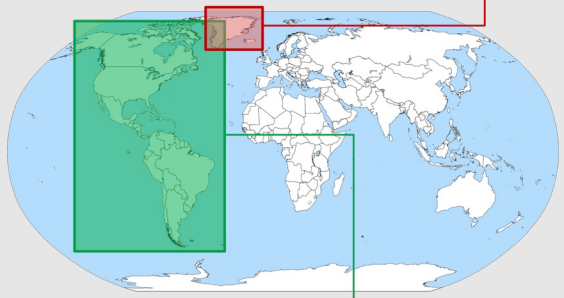
[Bracco et al., Machine Learning for the Physics of Climate. *Nature Reviews Physics*, 2024]

Exemplo: downscaling temporal (interpolação) de dados geoespaciais

Por que não usar simplesmente os métodos de IA existentes?

- Aplicações de mudança climática envolvem dados geoespaciais que evoluem no tempo
  - Dados de observação em grade global via assimilação de dados
  - Simulações de modelos físicos (PNT, MCG, MCR)
- São tensores de valores reais sobre latitude, longitude, tempo,
  e possivelmente várias variáveis climatológicas
- Os algoritmos de visão computacional para “dados espaço-temporais” se baseiam em propriedades de **dados de vídeo** que **não se generalizam bem para dados geoespaciais**
  - p. ex. profundidade, bordas e “objetos”
  - vs. padrões efêmeros em fluidos

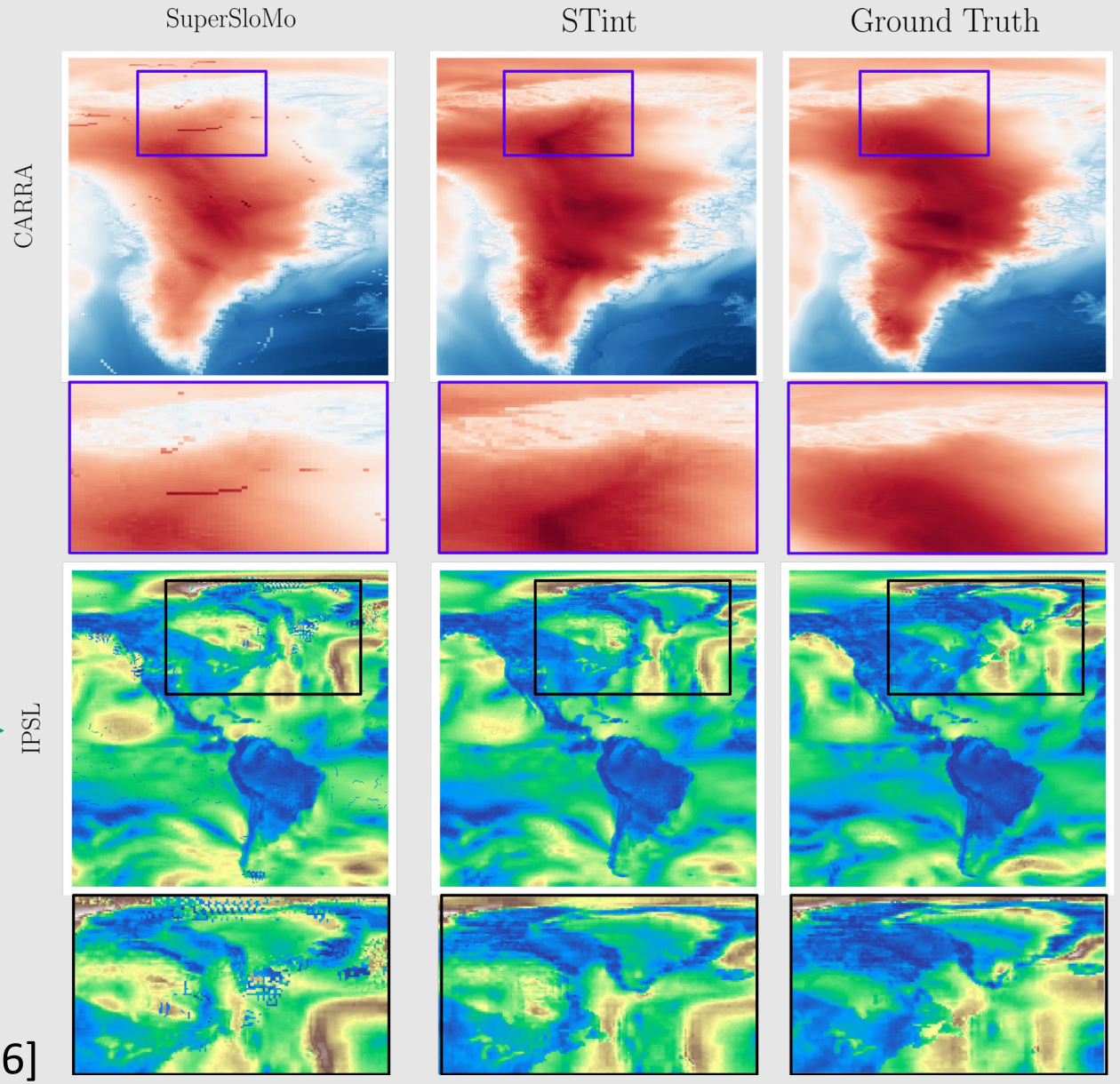
STINT: interpolação temporal autossupervisionada





O estado da arte em visão computacional para interpolação temporal de vídeo usa fluxo óptico.
Isso é problemático em dados geoespaciais.

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



Limitações de trabalhos anteriores: interpolação bicúbica

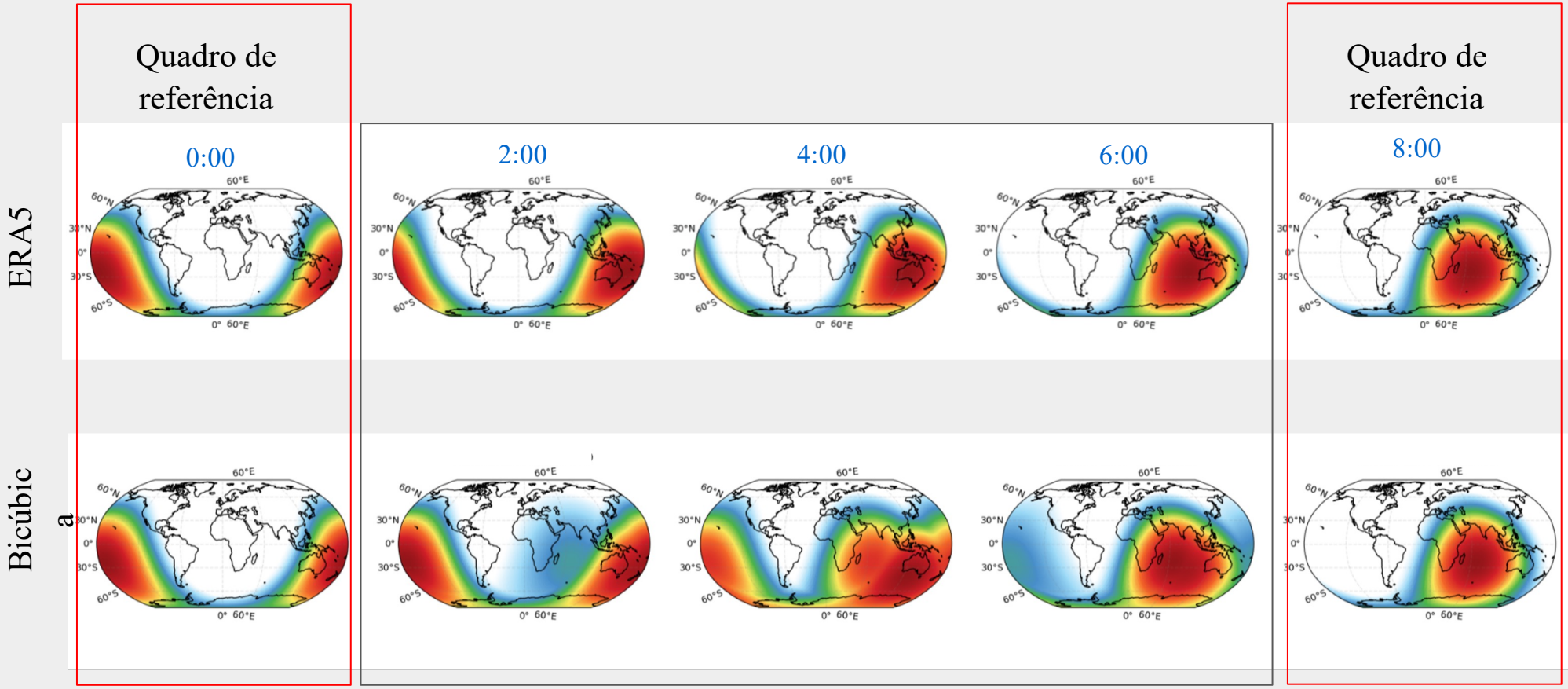
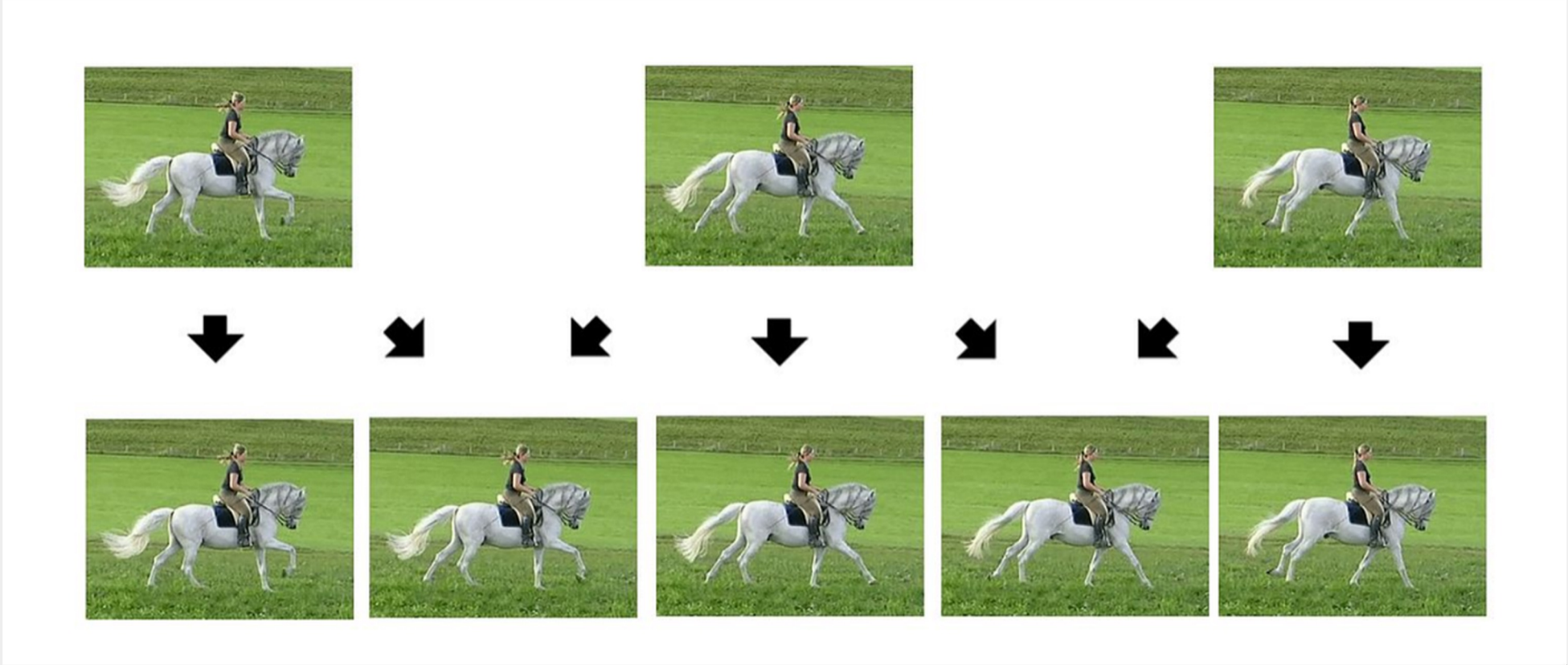


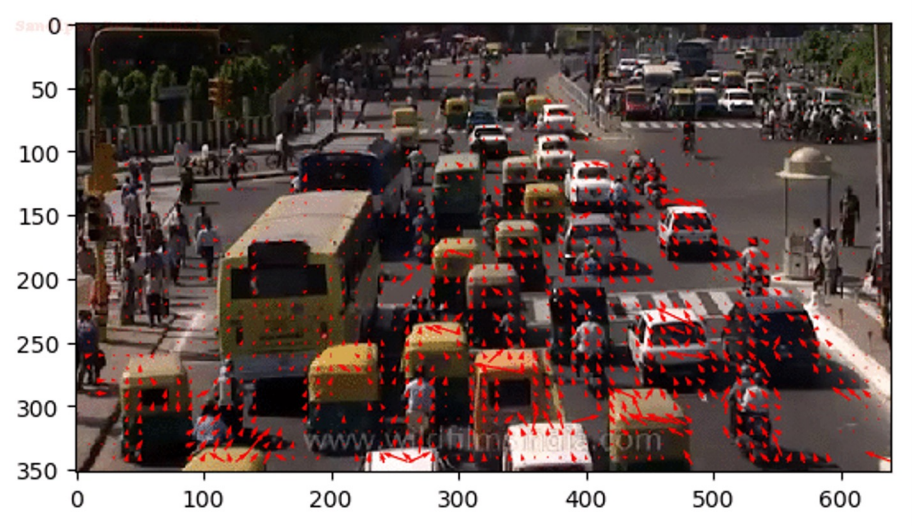
Figura. *Interpolação bicúbica temporal de campos de irradiância solar ERA5*

Interpolação de vídeo em visão computacional

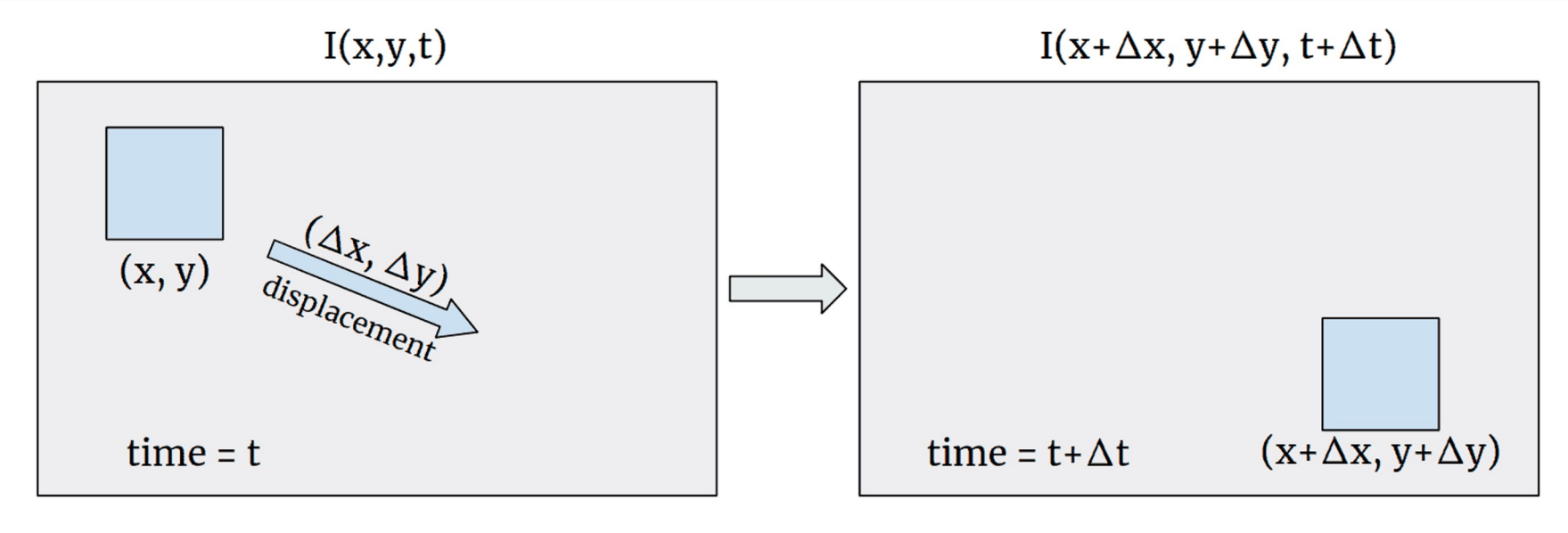


Fluxo óptico

- Fluxo óptico: campo de movimento pixel a pixel entre duas imagens (p. ex. quadros de vídeo).
- Codifica a projeção 2D “percebida” do movimento físico dos pontos em relação ao observador.



(fonte: Nvidia OF SDK)



Fluxo óptico

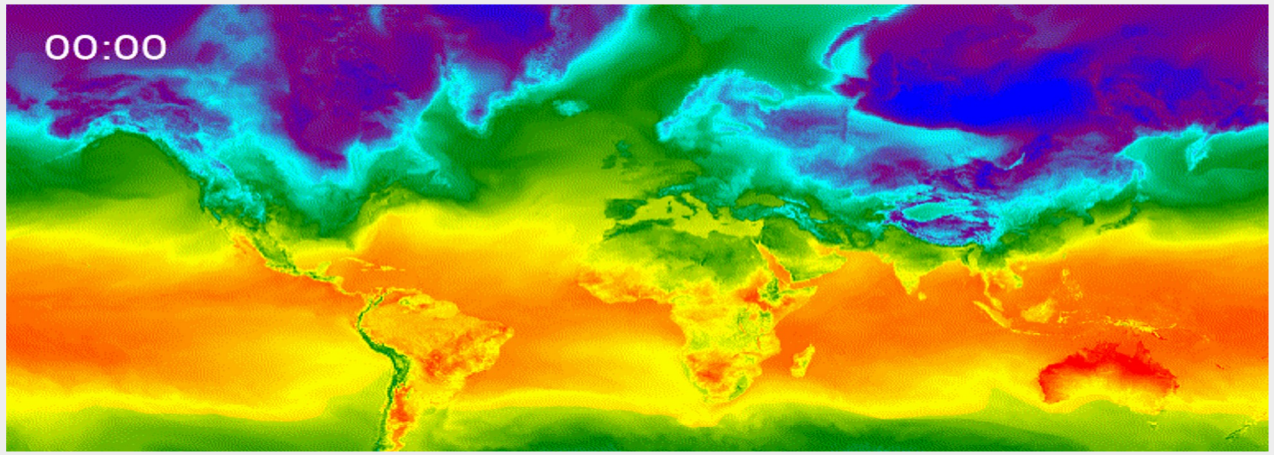
Premissas:

- Pequeno deslocamento (Δt).
- Nenhuma/pouca mudança na luz refletida.
- Nenhuma/pouca mudança de escala.
- Sem oclusão ou desoclusão.



Não se aplica a dados geoespaciais!

- Sem noção de objetos ou profundidade.
- Entidades podem se formar/deformar arbitrariamente.
- Resolução temporal muito mais grossa.



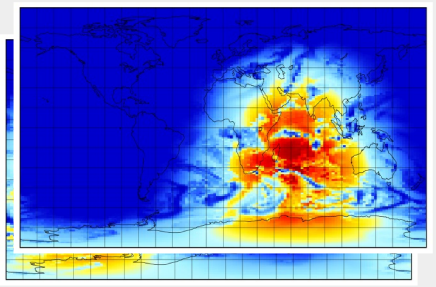
Temperatura horária NOAA GFS

Objetivo: melhorar a resolução temporal

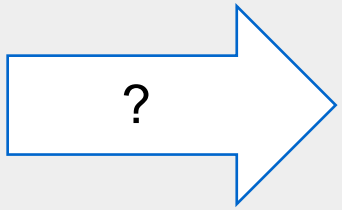
- A tarefa é a interpolação temporal
- Mas <u>não temos dados horários</u> para treinar

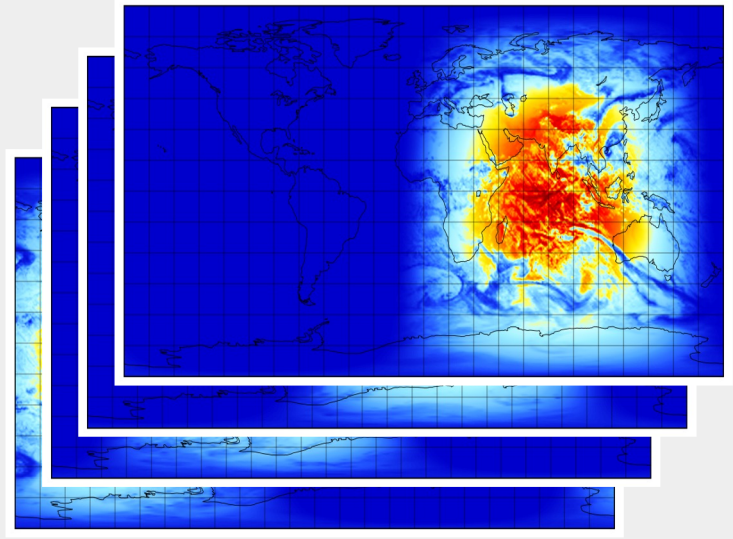
são necessárias técnicas não supervisionadas (**autossupervisionadas**).

CMIP6-IPSL



Trihorária





Horária

Autossupervisionado em resumo

- Usar tarefas pretexto/substitutas de forma supervisionada.
  - Rótulos **gerados** automaticamente.

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

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

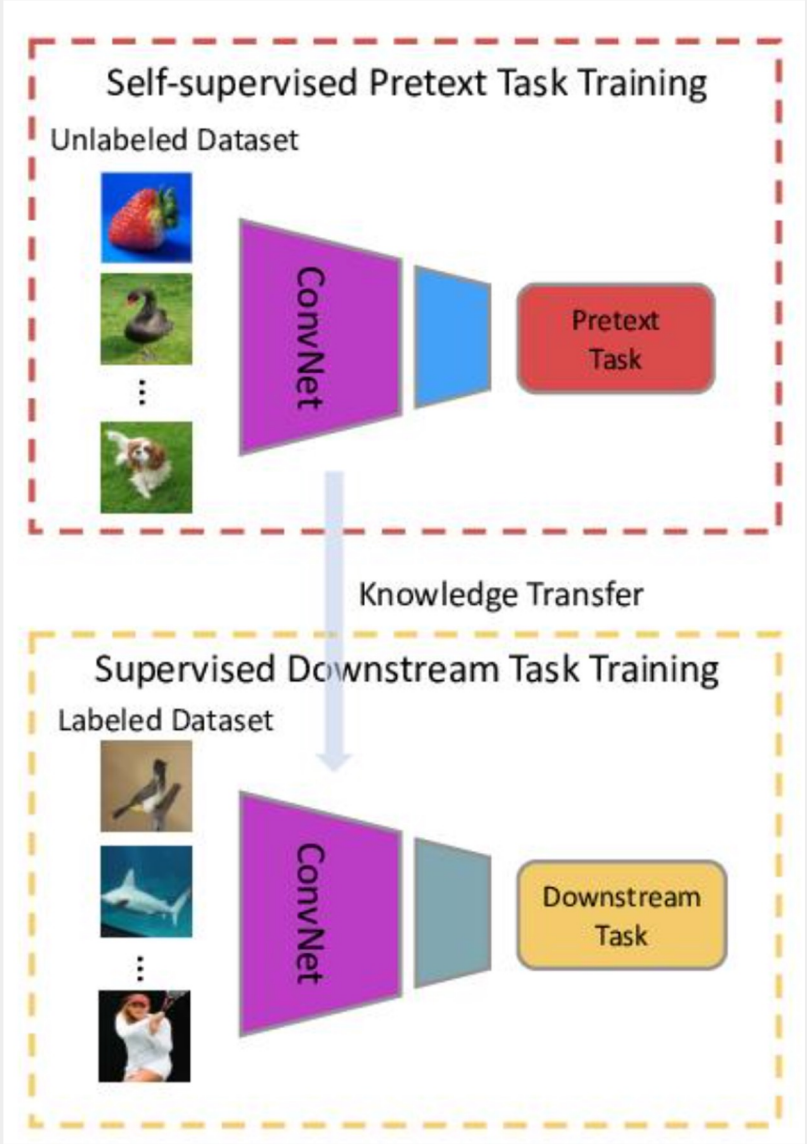
Aprendizado supervisionado

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

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

Aprendizado autossupervisionado

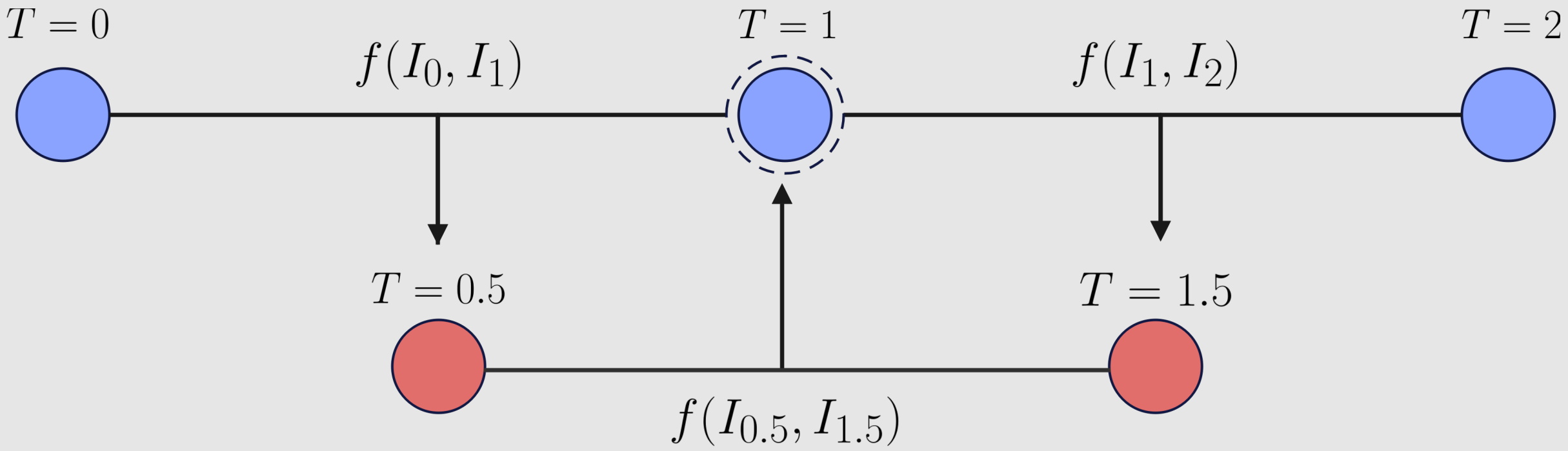
- Tarefa posterior: recuperação, classificação supervisionada ou semissupervisionada etc.



Uma tarefa pretexto para interpolação temporal

STINT: interpolação temporal autossupervisionada para dados geoespaciais

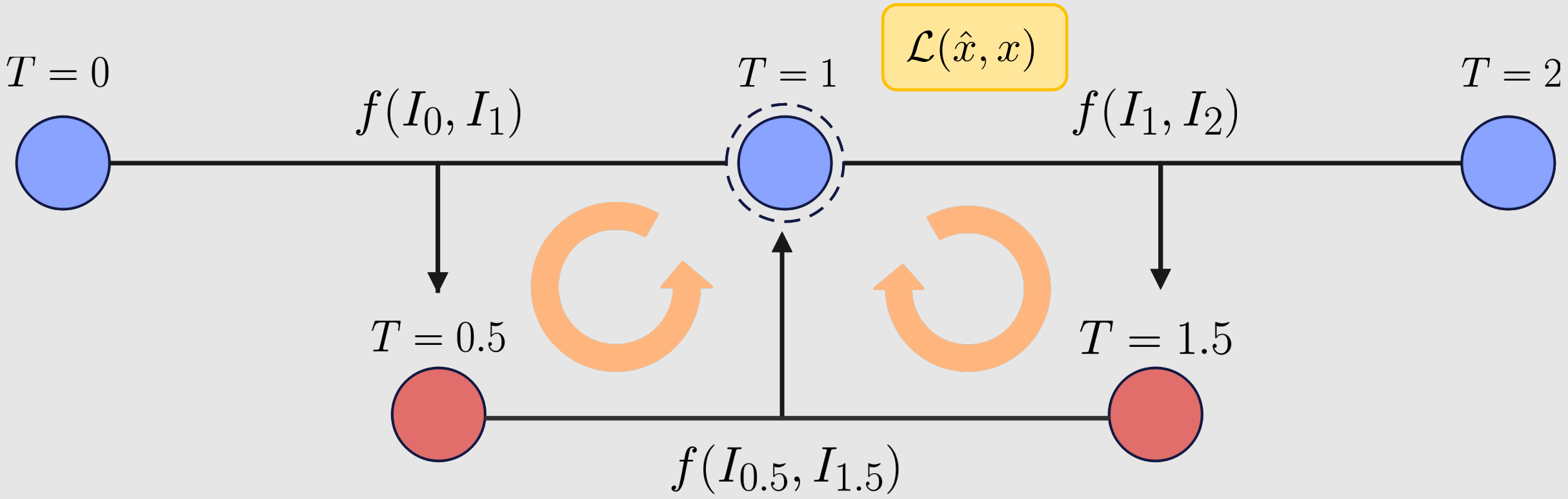
[Harilal, Hodge, Subramanian, & Monteleoni, Climate Informatics 2026]



Uma tarefa pretexto para interpolação temporal

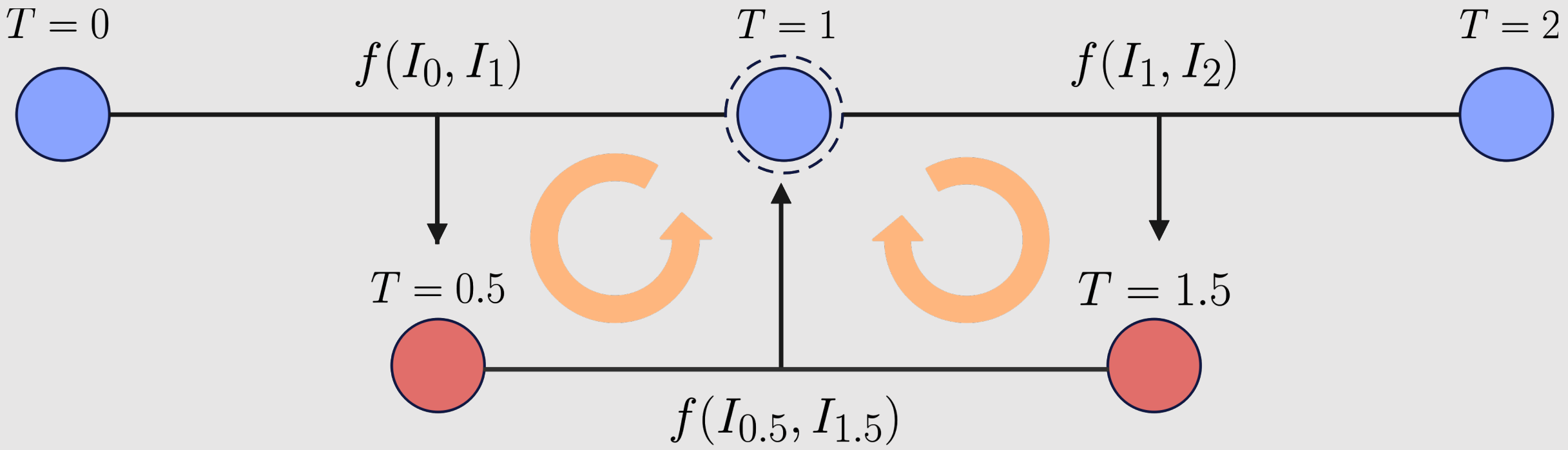
STINT: interpolação temporal autossupervisionada para dados geoespaciais

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



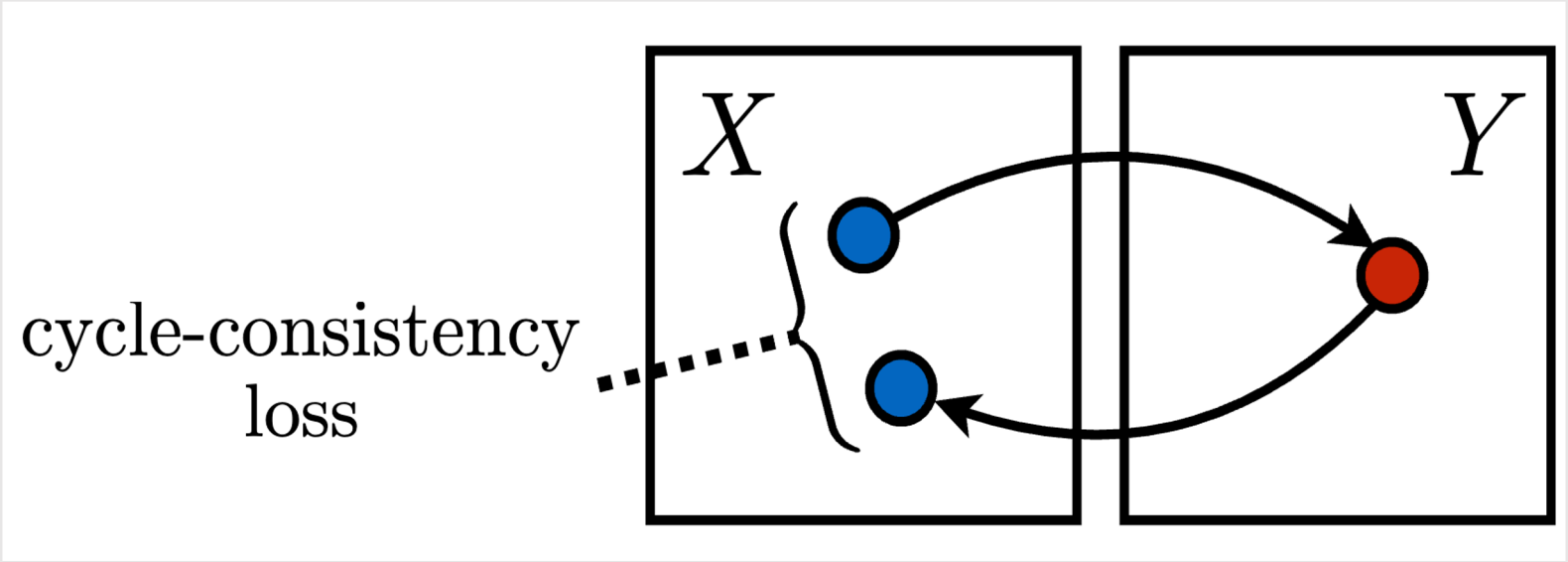
Uma tarefa pretexto para interpolação temporal

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

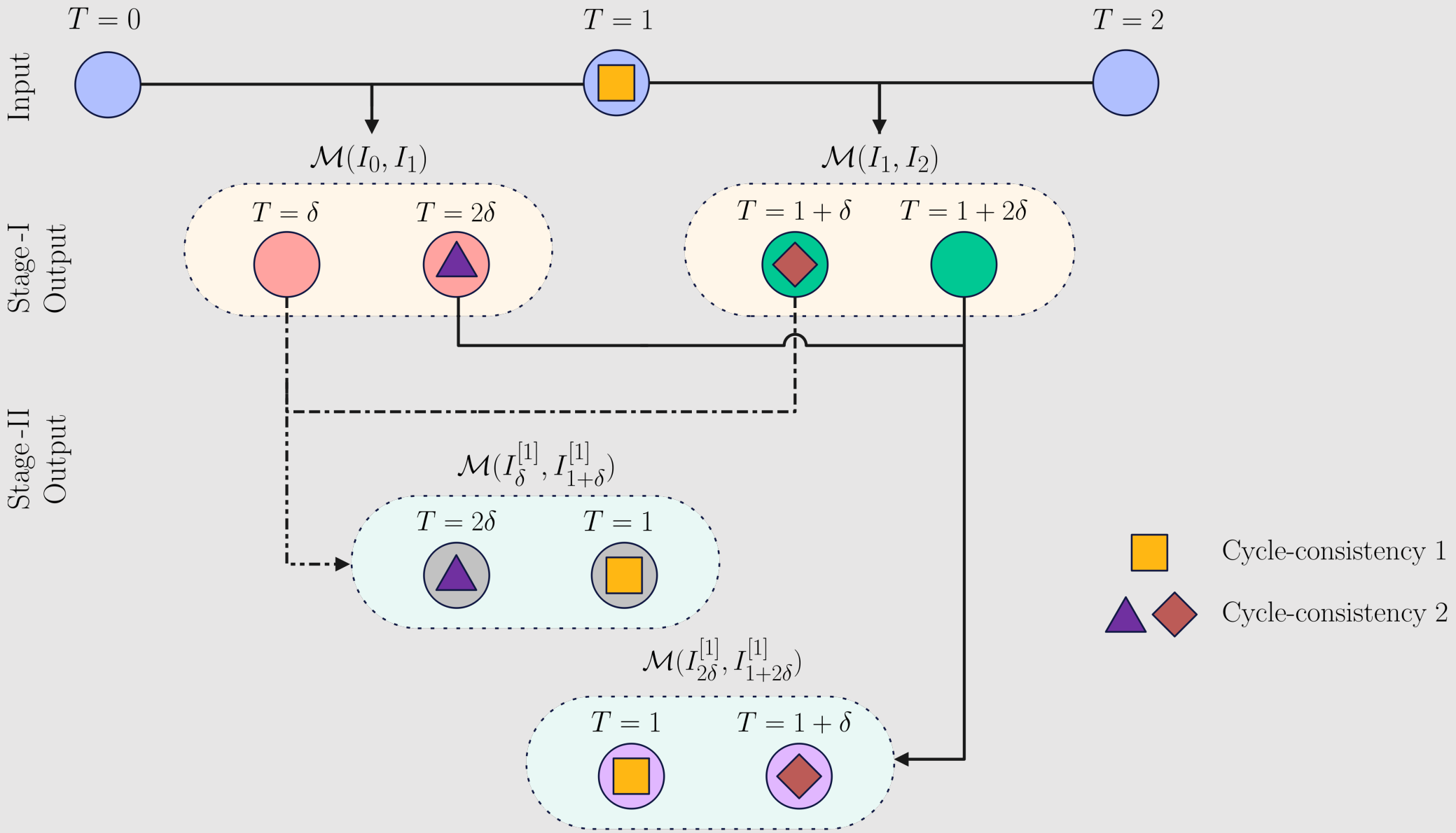


Perda de consistência cíclica

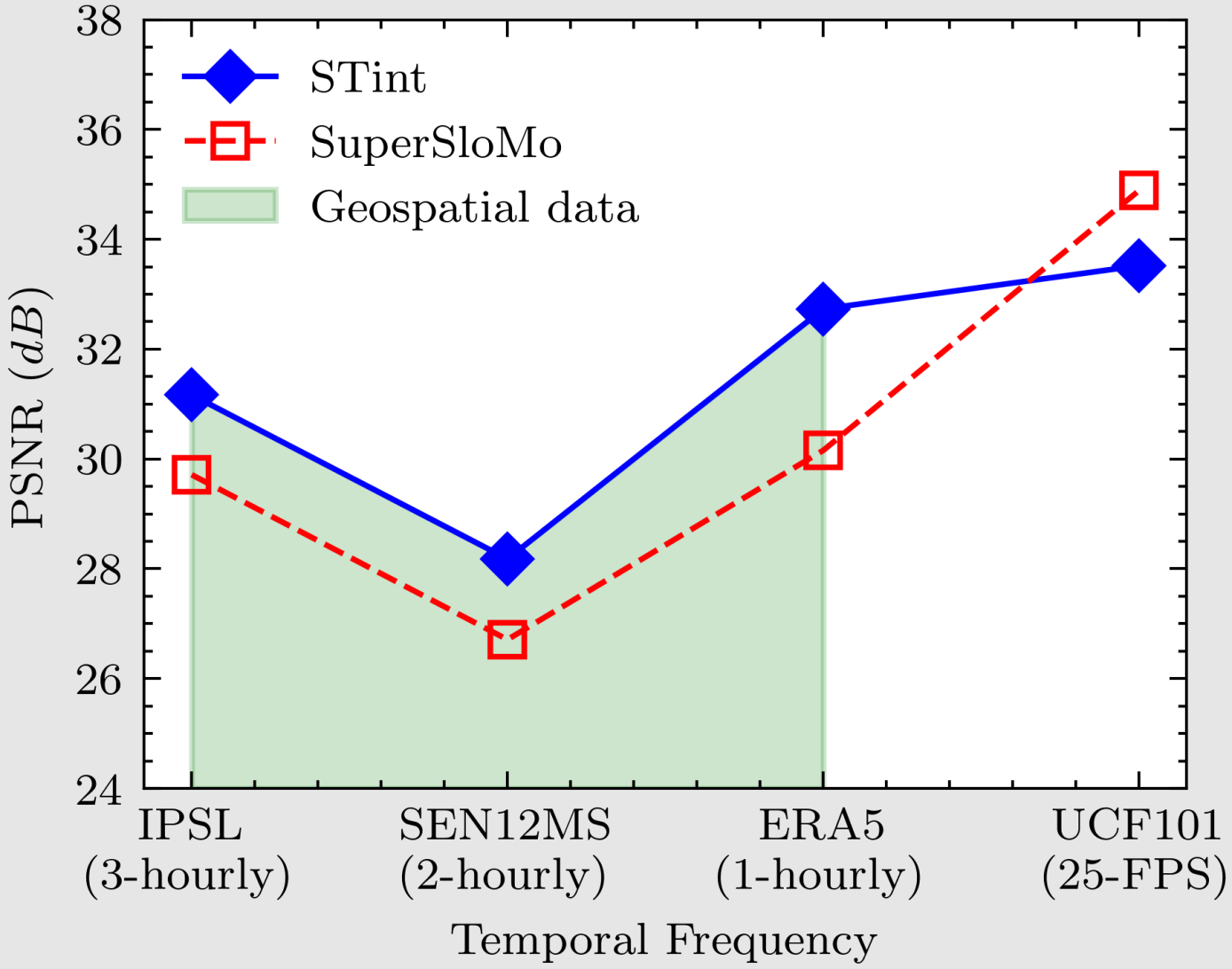
$$\mathcal{L}(\hat{x}, x) = \sum_{i=1}^{n} \mathcal{L}(T_i, \hat{T}_i) \qquad \hat{T}_i := f(f(T_{i-1}, T_i), f(T_i, T_{i+1}))$$



STINT: interpolação temporal autossupervisionada



STINT: interpolação temporal autossupervisionada



	$\frac{MSE}{Capacity}$ (↓)	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**

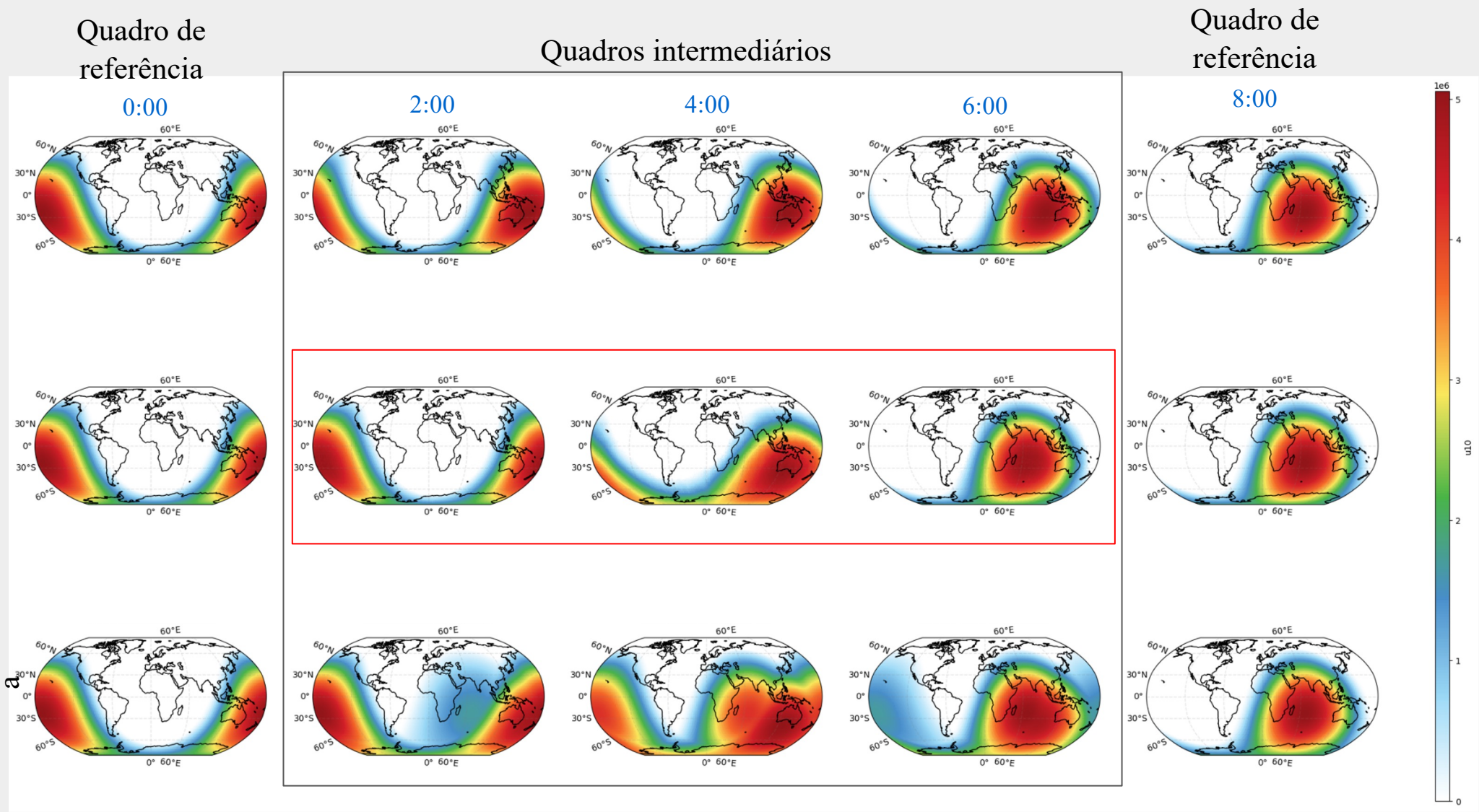


Figura. *Modelo treinado em campos de irradiância solar ERA5 e interpolação (3 quadros não vistos) entre 0h e 8h.*

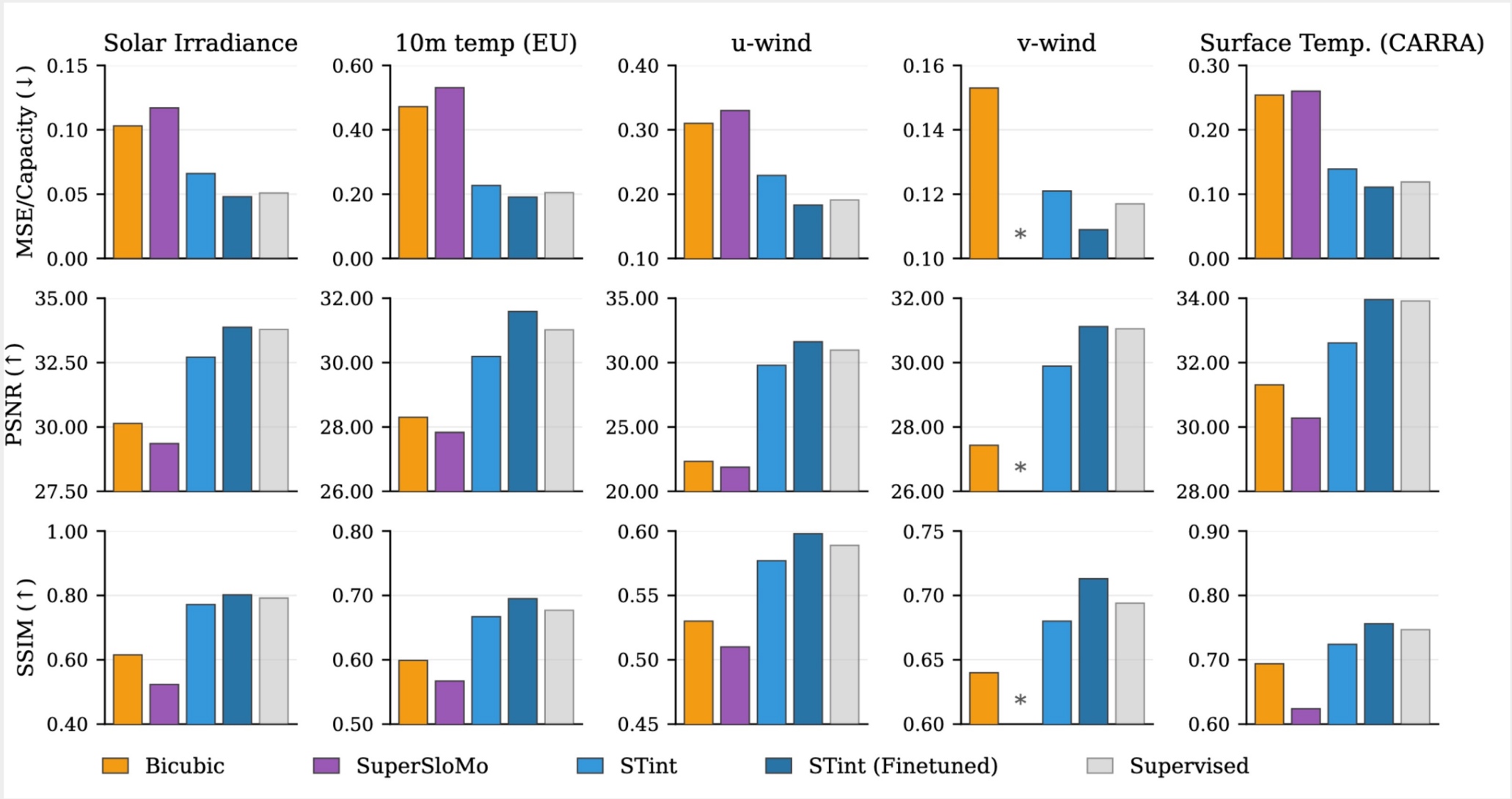
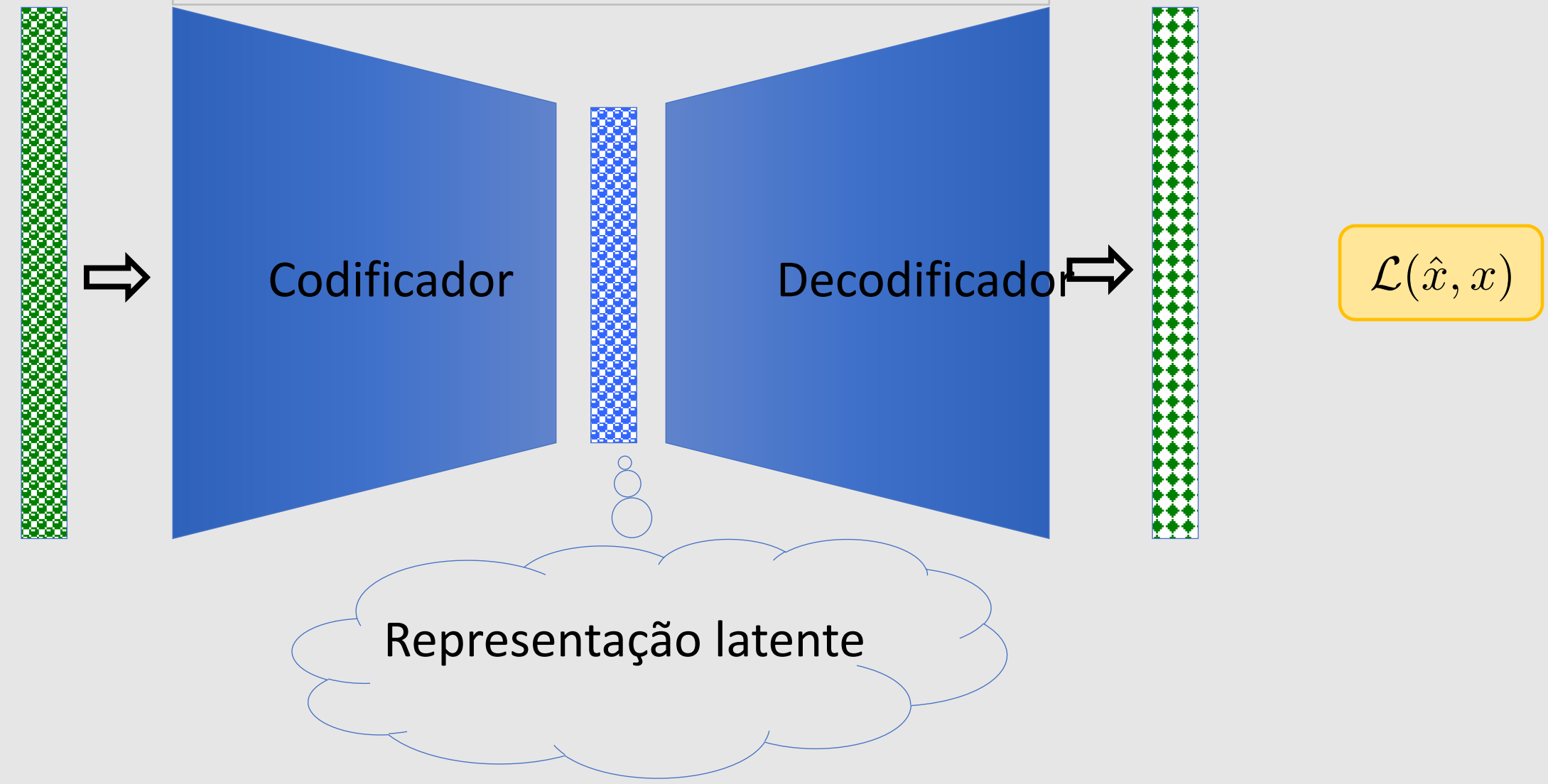


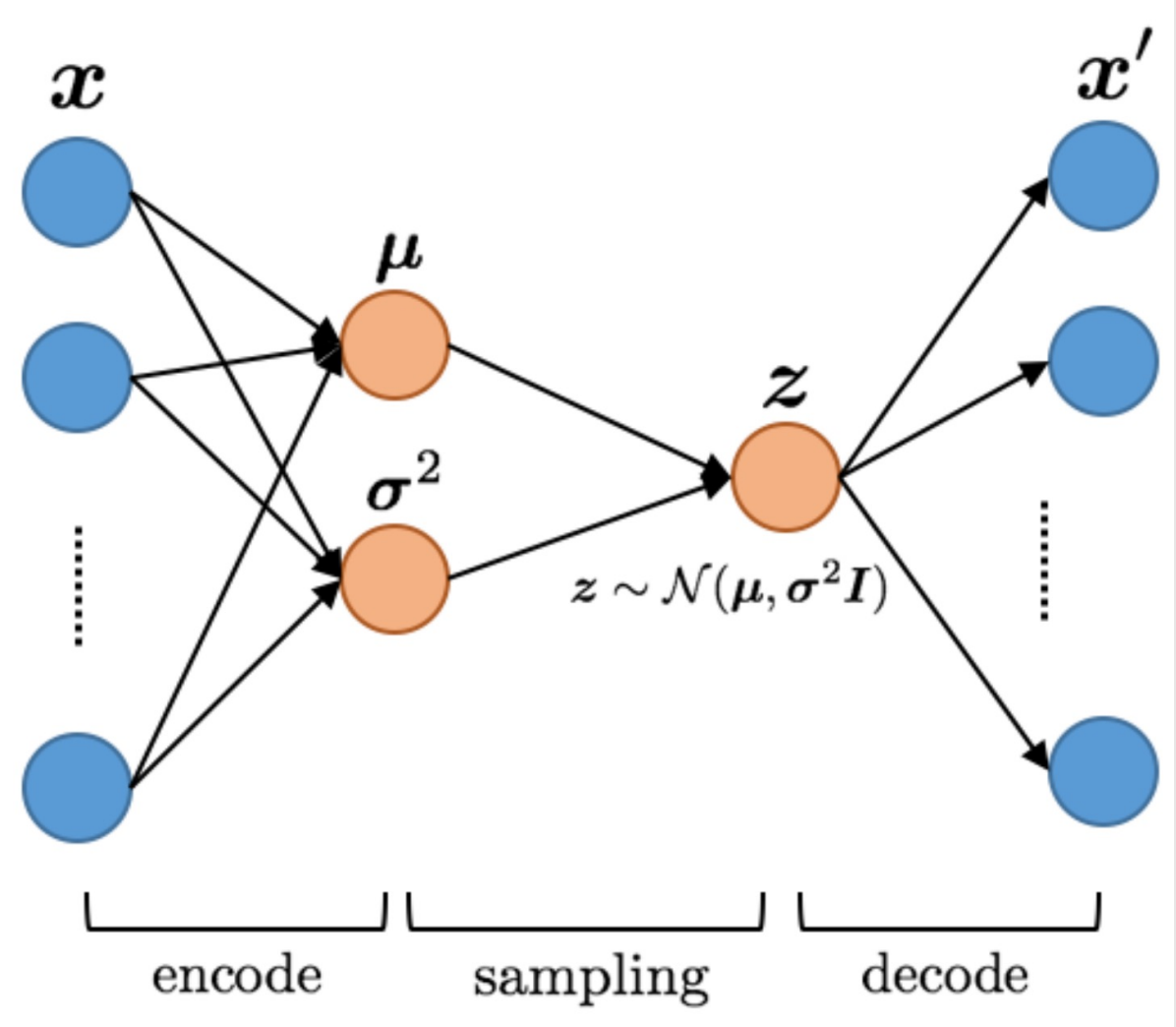
Figura. *Resultados de interpolação temporal em diferentes conjuntos de dados geoespaciais.*

Autoencoder: os parâmetros das redes codificadora e decodificadora são treinados para que a saída se aproxime da entrada. Após treinar com muitos exemplos, os parâmetros da camada de gargalo formam uma representação compacta da distribuição de entrada.



Autoencoder variacional (VAE)

Aprender uma distribuição sobre representações latentes, em vez de uma única codificação



Aprendendo um autoencoder variacional (VAE)

- O codificador é uma distribuição, q, que leva o espaço de dados ao espaço latente
  - Definido pelos pesos da rede θ:

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

- O decodificador é uma distribuição, p, que leva o espaço latente ao espaço de dados
  - Definido pelos pesos da rede ϕ:

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

- Ajustar os pesos θ e ϕ minimizando:

Erro de reconstrução nos dados de treinamento

Termo de regularização: penaliza a distância de uma priori simples

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

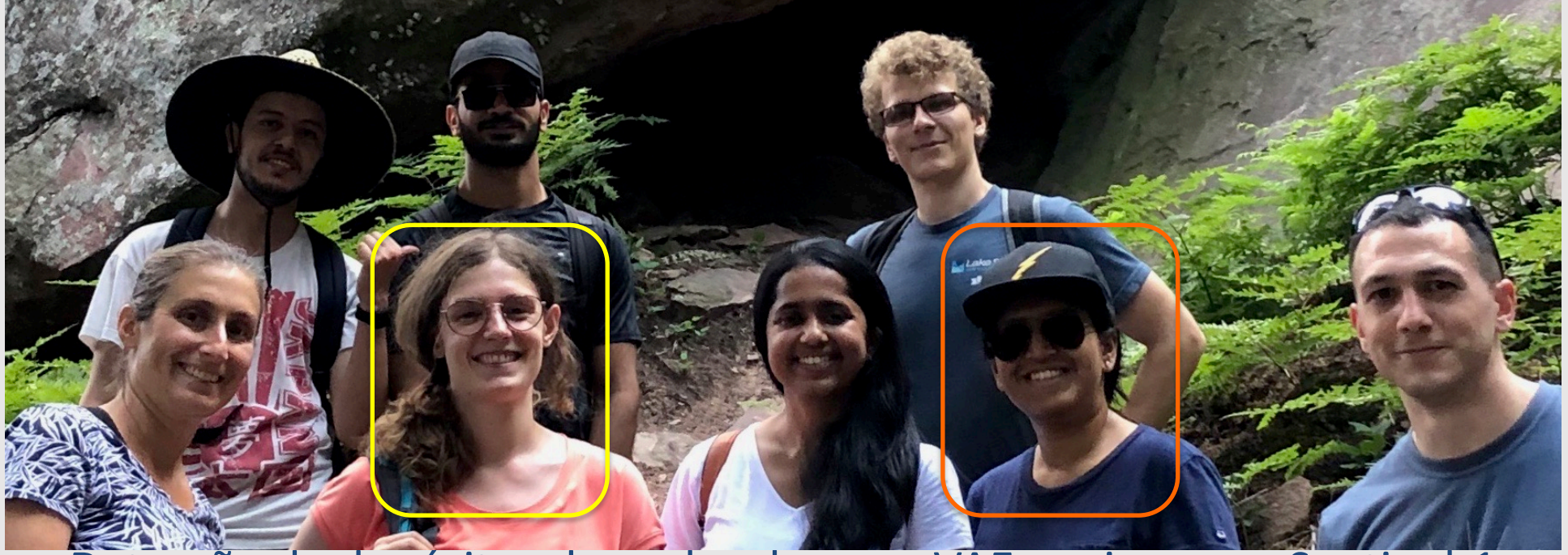
- Otimizado com uma variante de expectativa-maximização (EM), usando um limite inferior variacional

Exemplo: VAE para detecção semissupervisionada de anomalias aplicada à detecção de avalanches

Detecção de avalanches



Fonte da imagem: https://www.nytimes.com/2017/01/19/world/europe/italy-avalanche.html



Detecção de depósitos de avalanche com VAE em imagens Sentinel-1

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

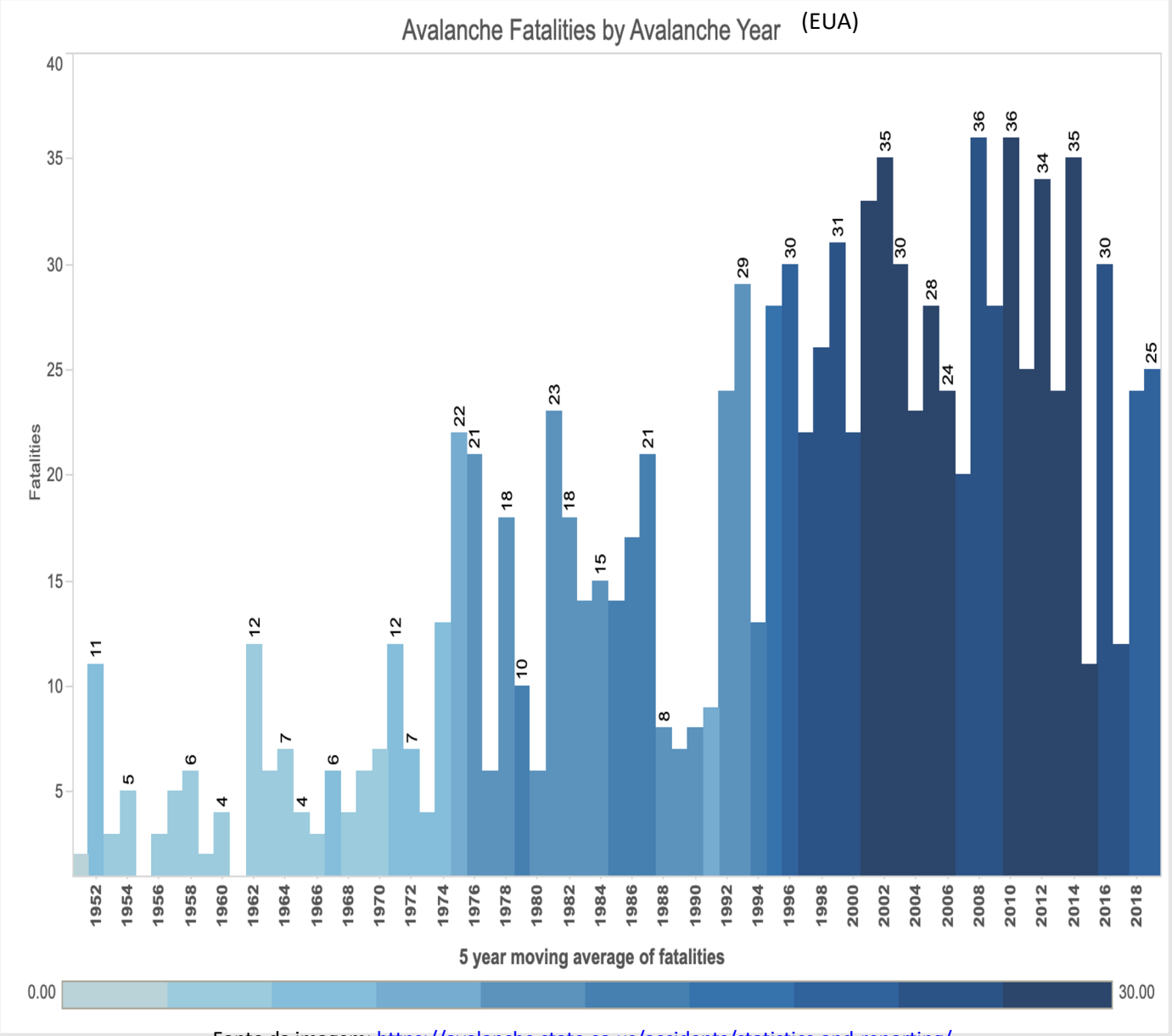
Variational autoencoder anomaly-detection of avalanche deposits in satellite SAR imagery. *Proc. 10th International Conference on Climate Informatics (CI) 2020*

Boulder

*Coautoria principal

METEO FRANCE

Mortes por avalanche por ano de avalanche



Fonte da imagem: https://avalanche.state.co.us/accidents/statistics-and-reporting/

Impactos da mudança climática



Fonte da imagem: https://www.independent.co.uk/news/world/europe/mont-blanc-deaths-french-alps-avalanche-skiing-a9223461.html

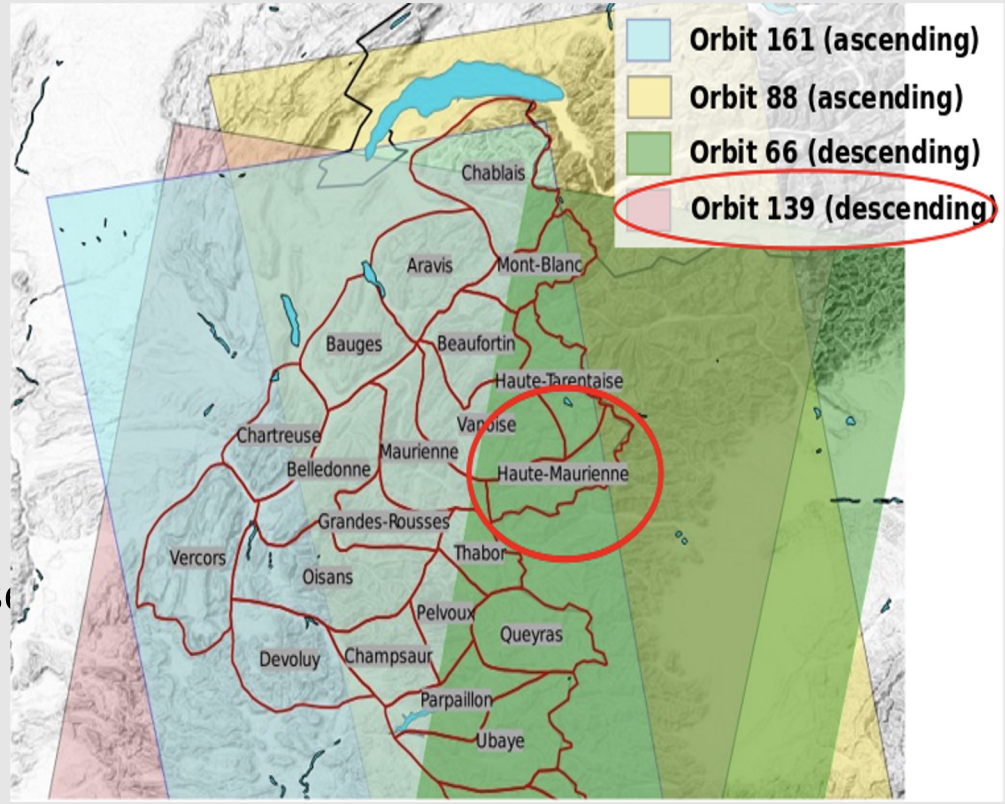
Desafios para o aprendizado de máquina

- Forte desbalanceamento de classes
  - Avalanches são eventos raros
- Dados de referência rotulados difíceis de obter
  - Acessibilidade do terreno
  - Condições meteorológicas
  - Perigo de avalanches

O problema da detecção de avalanches

- **Medições de referência** in situ limitadas
  - Météo-France

- **Imagens SAR** não rotuladas
  - Monitoramento dos Alpes franceses 2017-2018
  - Satélites Sentinel-1A e 1B
  - 4 características:
    - Coeficientes de retroespalhamento no momento atual e anterior
    - Características topológicas: declividade e orientação



METEO FRANCE

Nossa abordagem

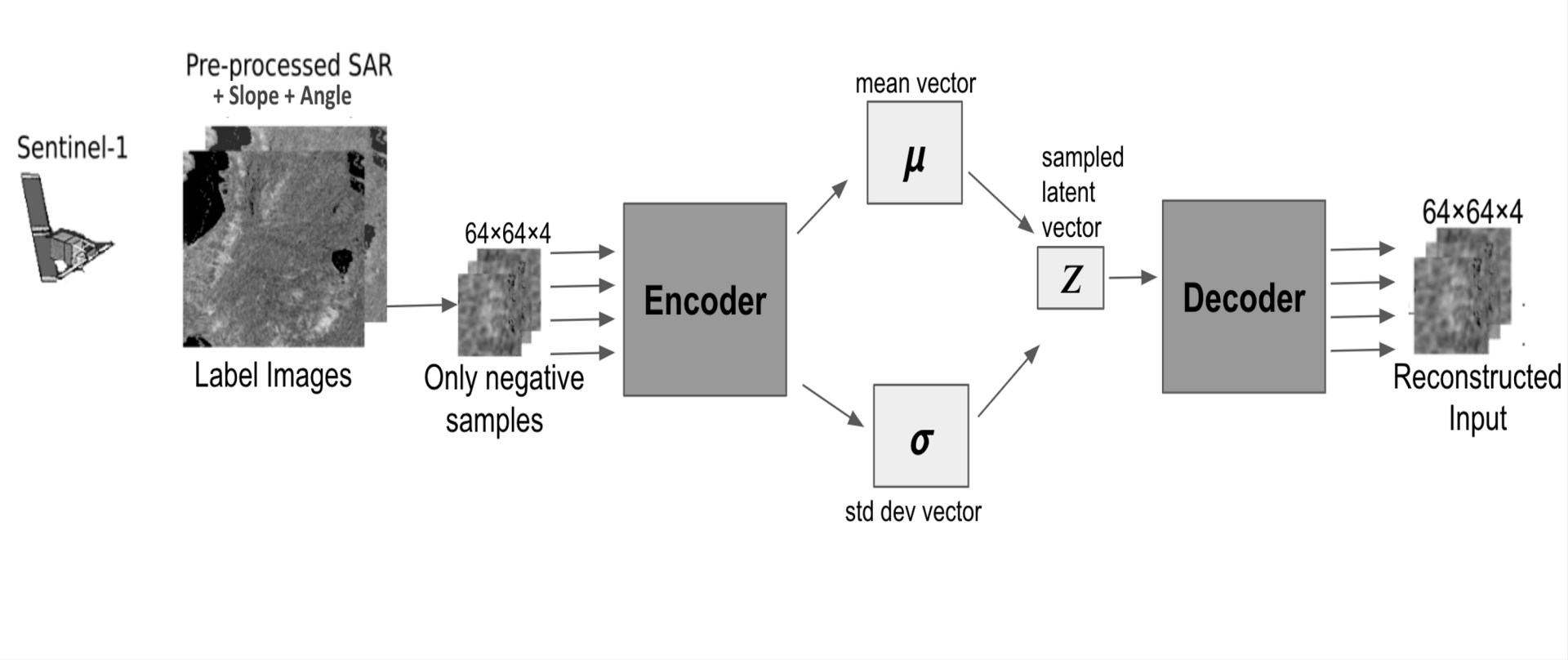
① Tratar uma avalanche como evento raro, ou anomalia

② Treinar um autoencoder variacional (VAE) com os exemplos negativos

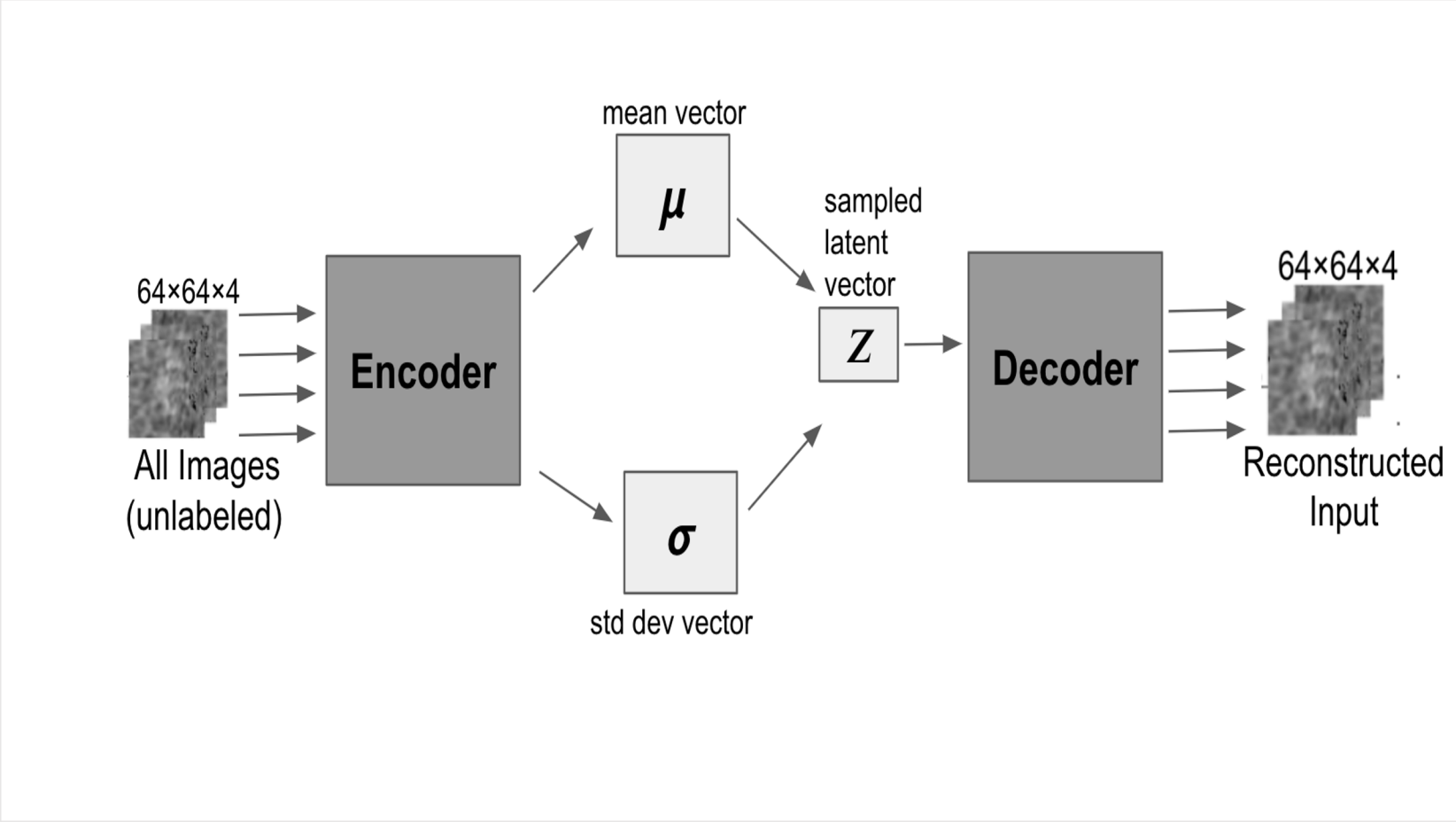
③ Aplicar um limiar ao erro de reconstrução do VAE para classificar uma nova imagem

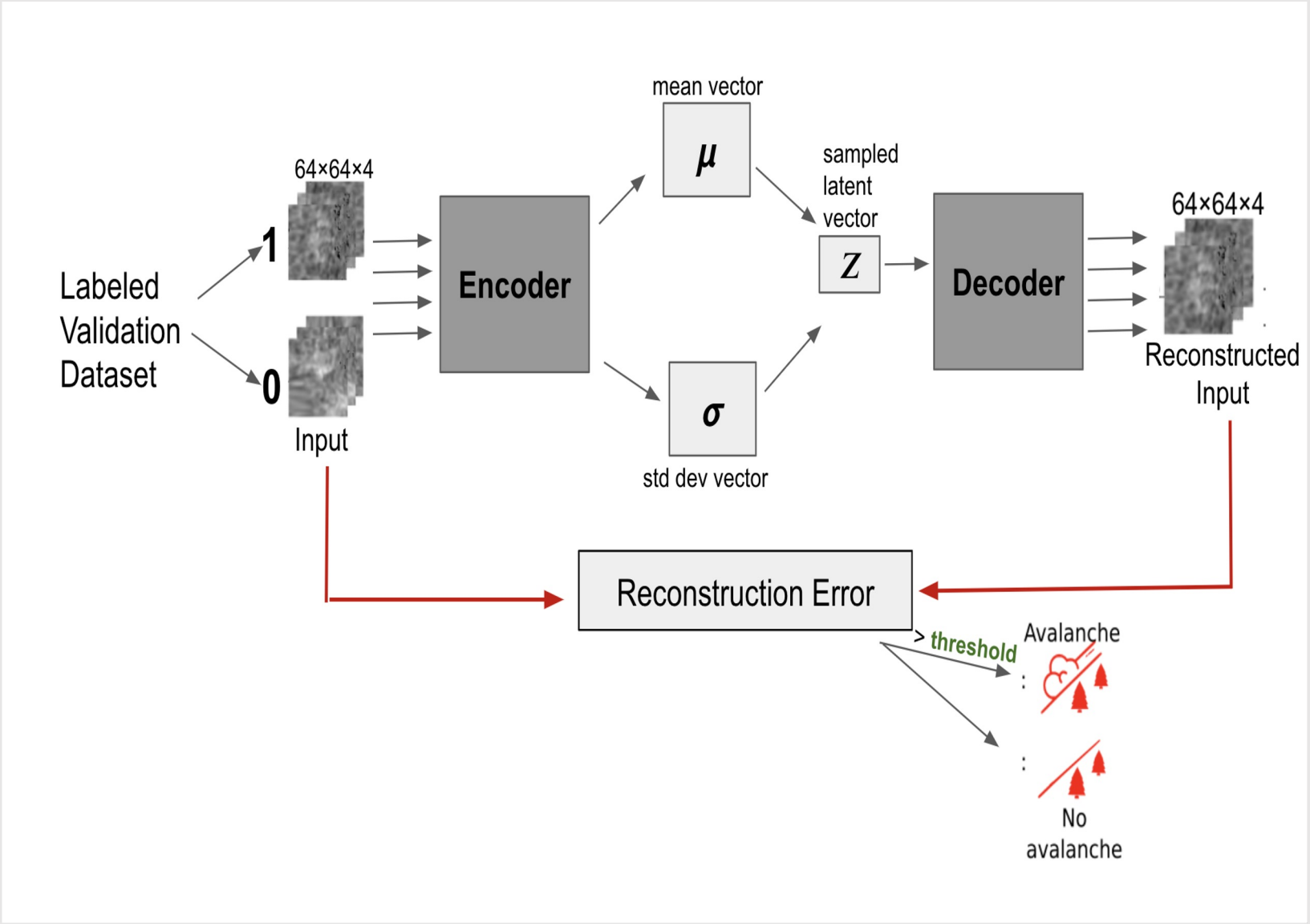
- Quando há poucos dados rotulados, o VAE pode ser treinado **sem** supervisão!

VAE treinado com exemplos negativos

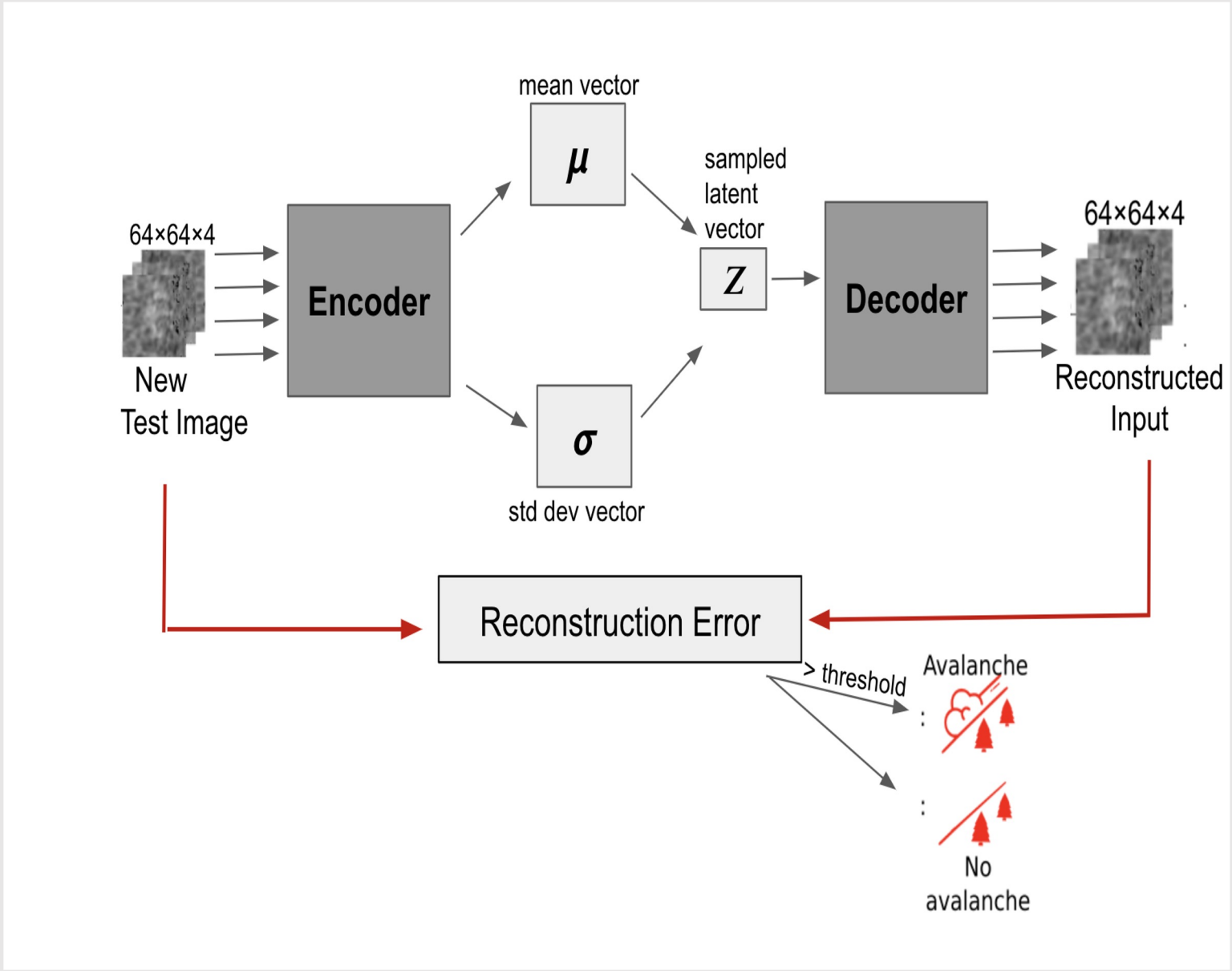


VAE treinado com TODOS os exemplos



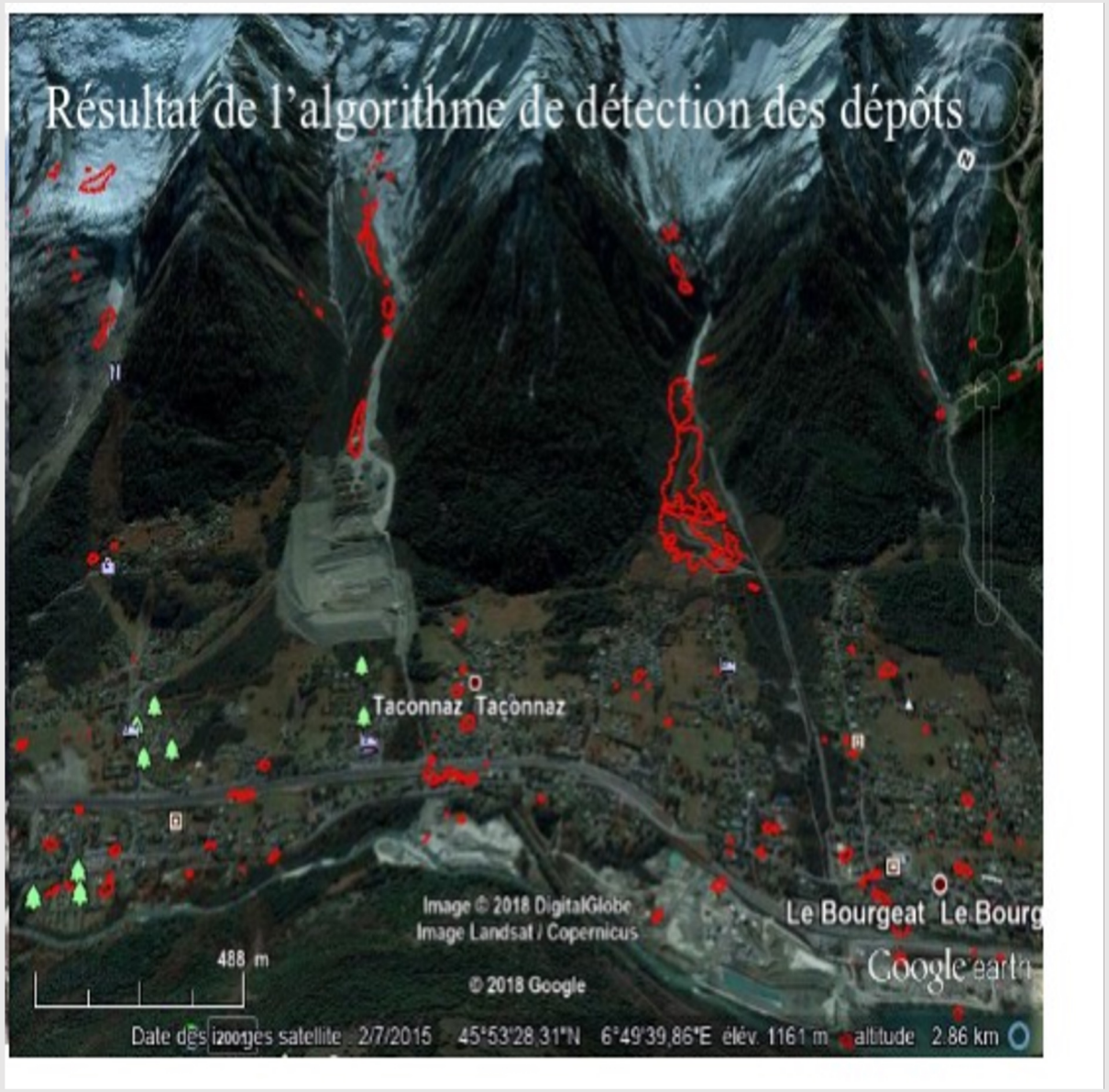


Ajuste do limiar para detecção de anomalias



Pipeline de detecção de anomalias

Método de referência: limiarização



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

Avaliação

Um dos maciços mais suscetíveis (dos 18 de “Todos os Alpes”)

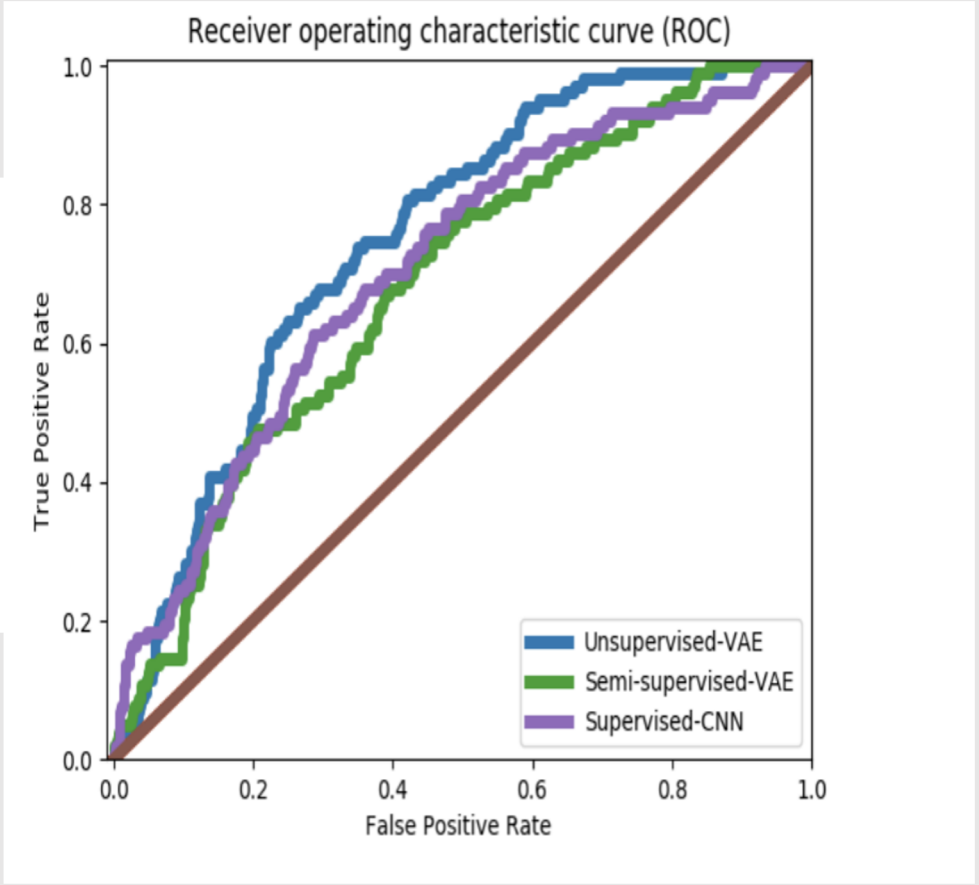
	All Alps		Haute Maurienne	
	Balanced Accuracy	F1-score	Balanced Accuracy	F1-score
Baseline	0.58	0.05	0.58	0.12
Supervised - CNN	0.53	0.10	0.53	0.12
Semi-supervised - VAE	0.59	0.11	0.6	0.23
Unsupervised - VAE	0.69	0.14	0.68	0.26

- Método de referência da literatura de detecção de avalanches: limiarização [Karbou et al., ISSW 2018]
- Método de referência em aprendizado supervisionado: rede neural convolucional (CNN) treinada com conjunto balanceado artificialmente [Sinha et al., Climate Informatics 2019]

Avaliação

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

Curvas ROC para a região de Haute-Maurienne



Resumo

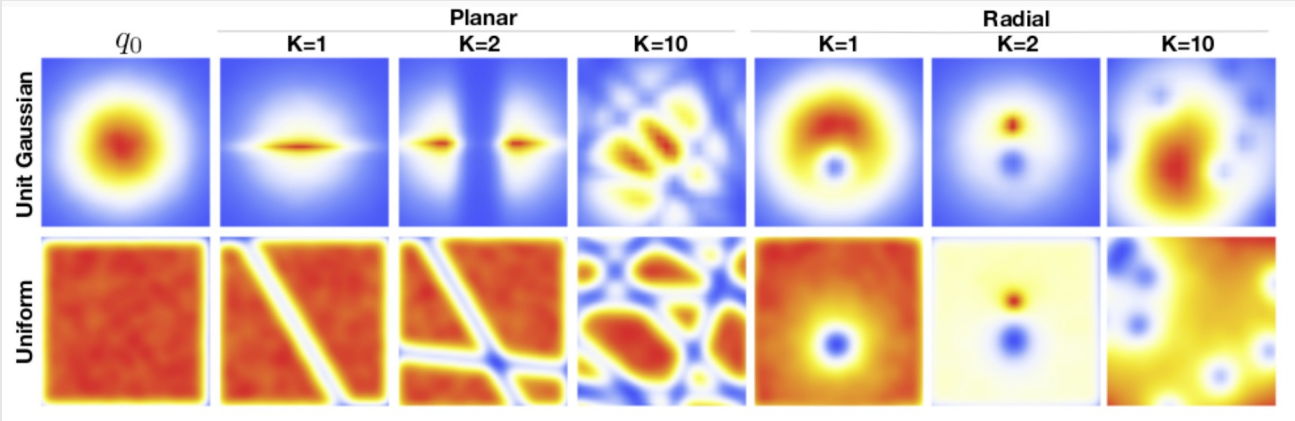
- Propusemos uma abordagem semissupervisionada para detectar eventos raros quando os dados rotulados são limitados
  - Este pipeline semissupervisionado depende muito do aprendizado não supervisionado
- Pode ser visto como uma forma de sensor virtual

Parte III: Métodos de IA:
IA generativa

IA generativa: uma distinção

- Vocês provavelmente estão acostumados com o termo "IA generativa" referindo-se a grandes chatbots e agentes comerciais como ChatGPT, Gemini ou Claude.
  - Nesses sistemas, os modelos de IA generativa foram treinados com vídeos e textos coletados da internet
  - Mas essa **não é a ÚNICA definição** nem o único uso da IA generativa!
- Nós podemos **treinar modelos de IA generativa do zero** para os nossos problemas
  - Para problemas de clima e tempo, NÃO vamos treinar com texto e vídeo, mas com os dados relevantes, p. ex. campos de valores reais medidos em diferentes latitudes, longitudes, momentos e talvez alturas na atmosfera ou profundidades no oceano, como:
    - Temperatura
    - Radiação solar
    - Umidade
- Hoje vamos conhecer os tipos de modelos generativos que vocês poderão treinar para seus próprios problemas

Fluxos normalizadores

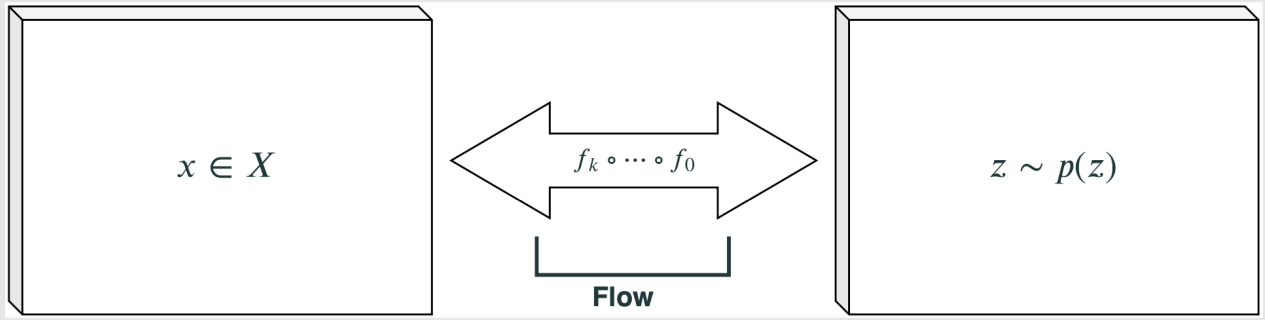


[Rezende & Mohamed, ICML 2015]

Podem ser vistos como extensão do VAE além da hipótese gaussiana no espaço latente

Aprender uma série de transformações inversíveis, $\{f_i\}$, a partir de uma priori simples no espaço latente Z, para obter distribuições mais informativas no espaço latente:

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

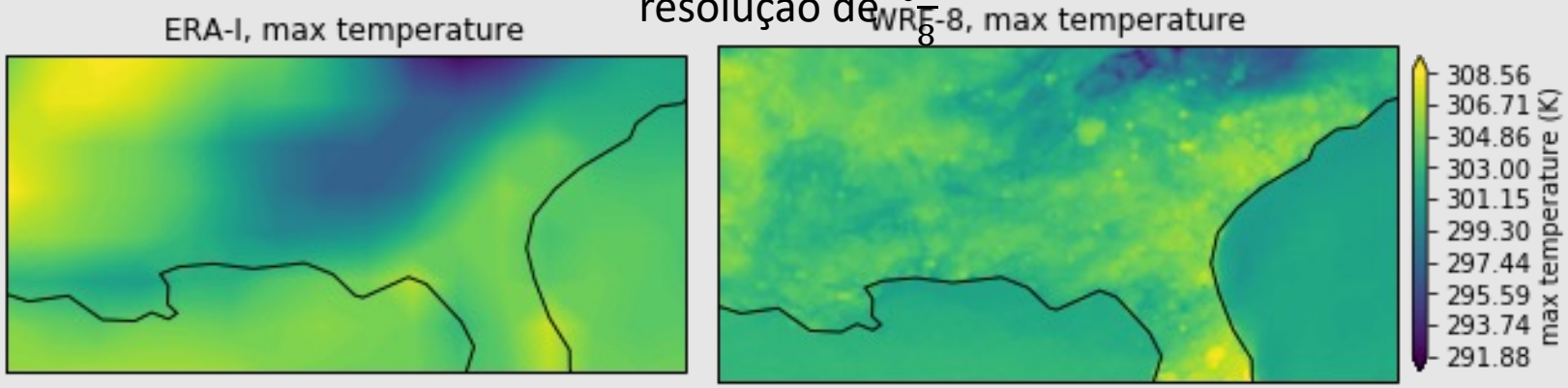


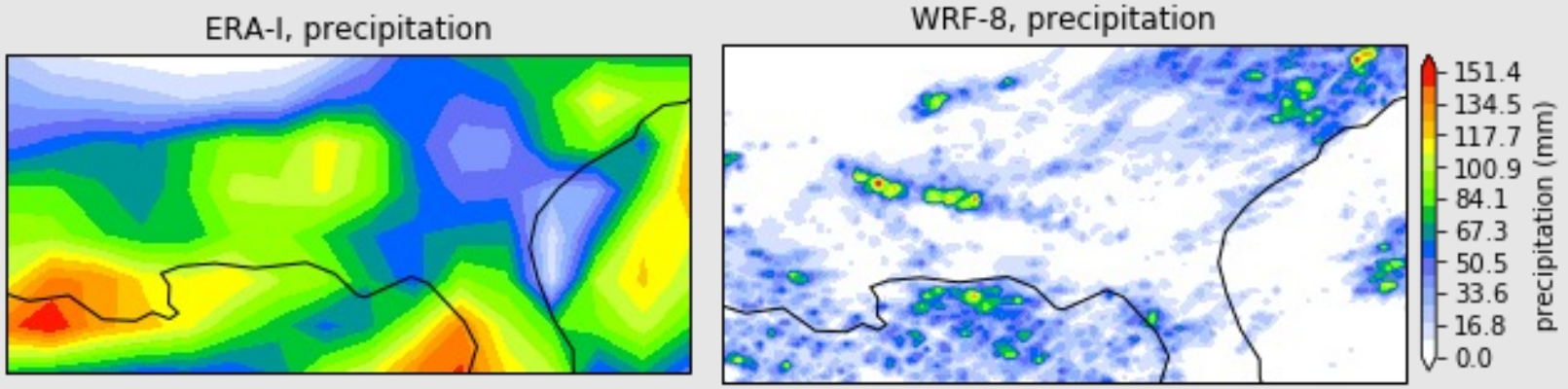
Fluxos normalizadores: aplicação ao downscaling espacial

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

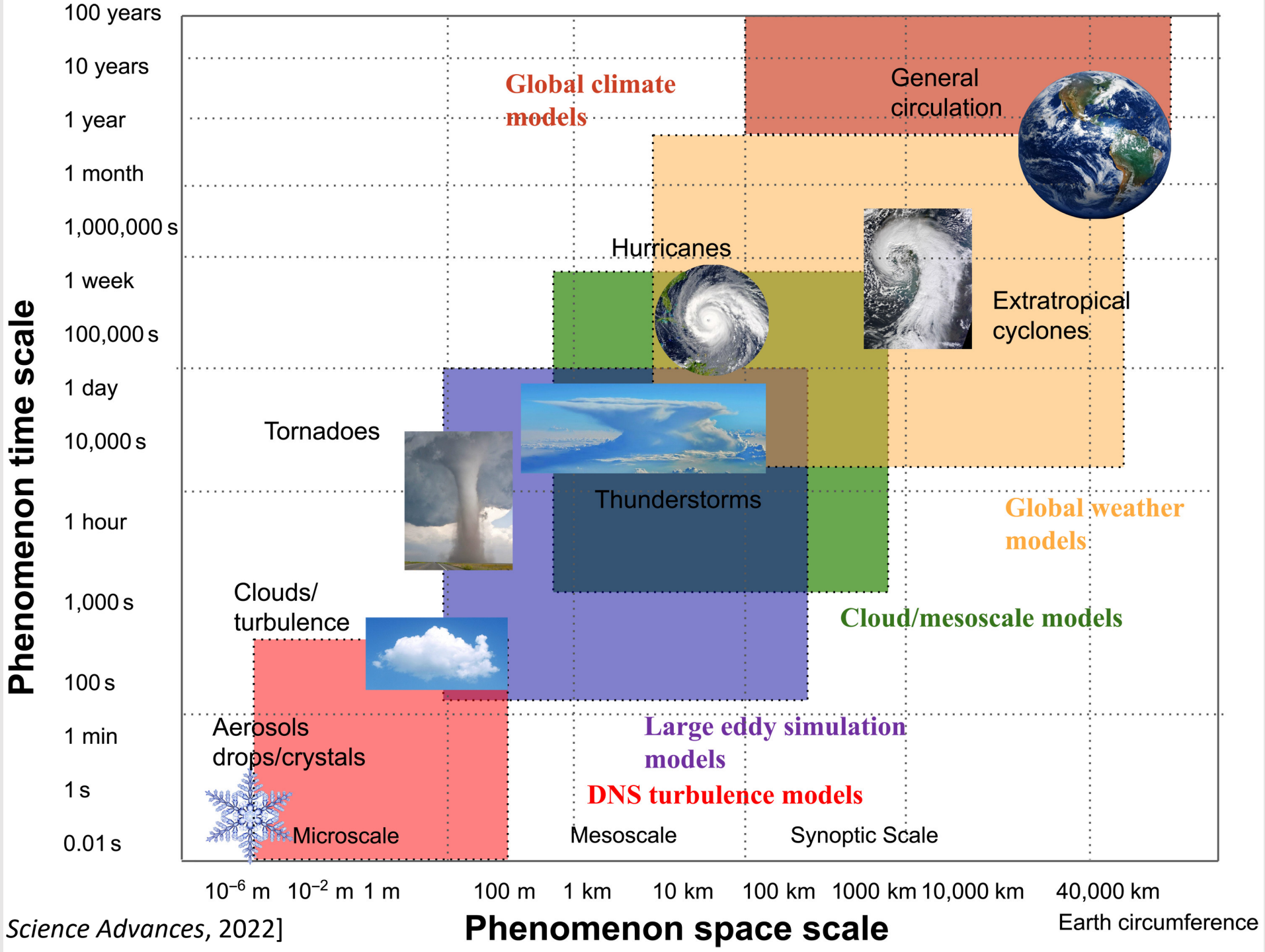
ERA: dados de reanálise, resolução de 1°;

WRF: previsão de modelo meteorológico numérico, resolução de $\frac{1}{8}°$









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

Downscaling como alinhamento de domínios

- Tarefa de alinhamento de domínios: dadas variáveis aleatórias X, Y, aprender um mapeamento f: X Y tal que, para todo x_i X e y_i Y,

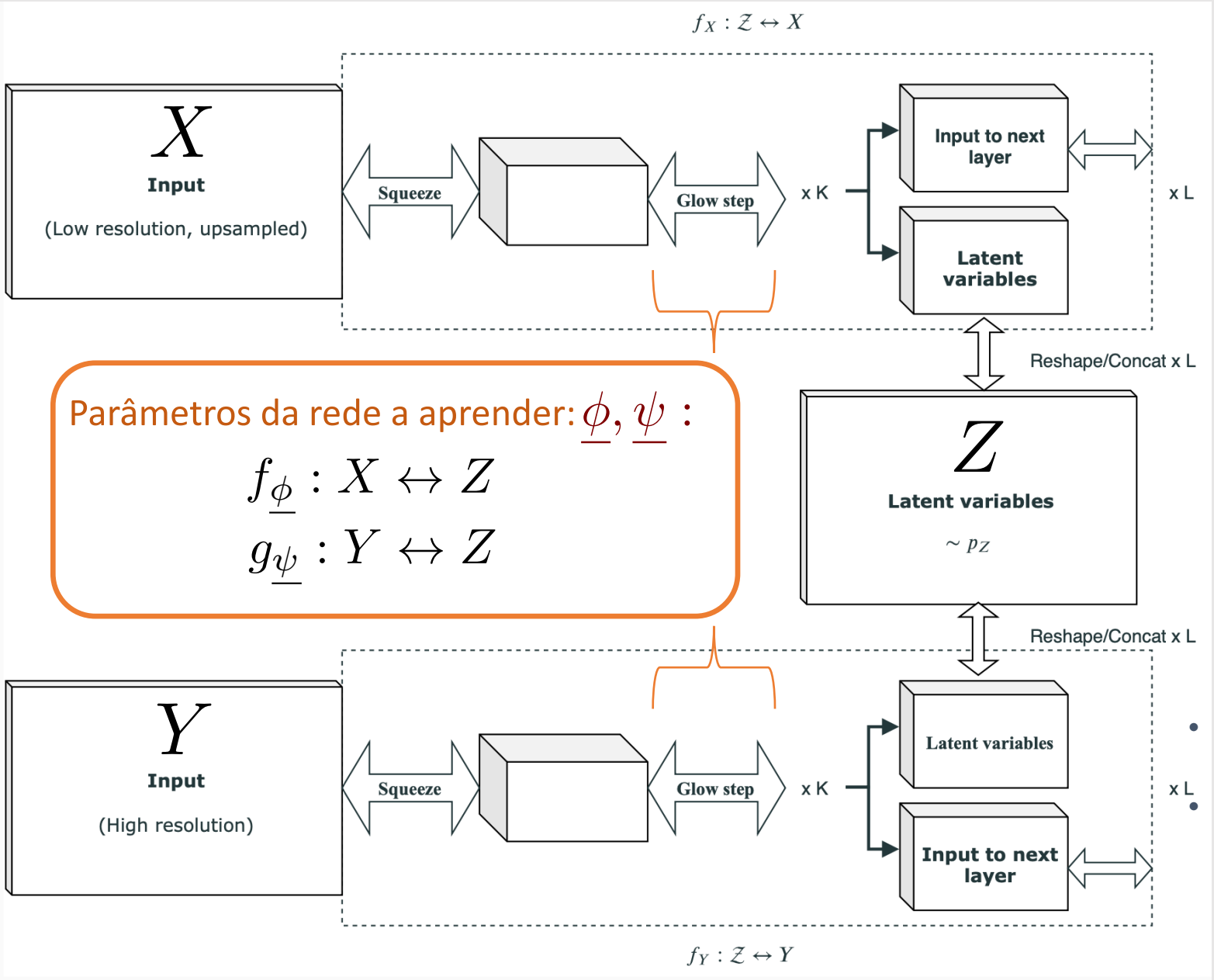
$$f(x_i) \sim P_Y \qquad f^{-1}(y_i) \sim P_X$$

- Downscaling como alinhamento de domínios
  - Dadas amostras i.i.d. em baixa resolução (X) e alta resolução (Y)
  - Aprender a densidade conjunta de X e Y supondo independência condicional por um espaço latente compartilhado Z,

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

  - Modelar $P(x|z), P(y|z)$ com AlignFlow [Grover et al. 2020]
    - Partindo de uma priori simples sobre P_Z, aprender fluxos normalizadores
    - Não é preciso parear exemplos x e y!

Arquitetura do ClimAlign



- A arquitetura segue o AlignFlow [Grover et al., 2020]
- Fluxo normalizador: Glow [Kingma & Dhariwal, 2018]

Comparação com referências supervisionadas

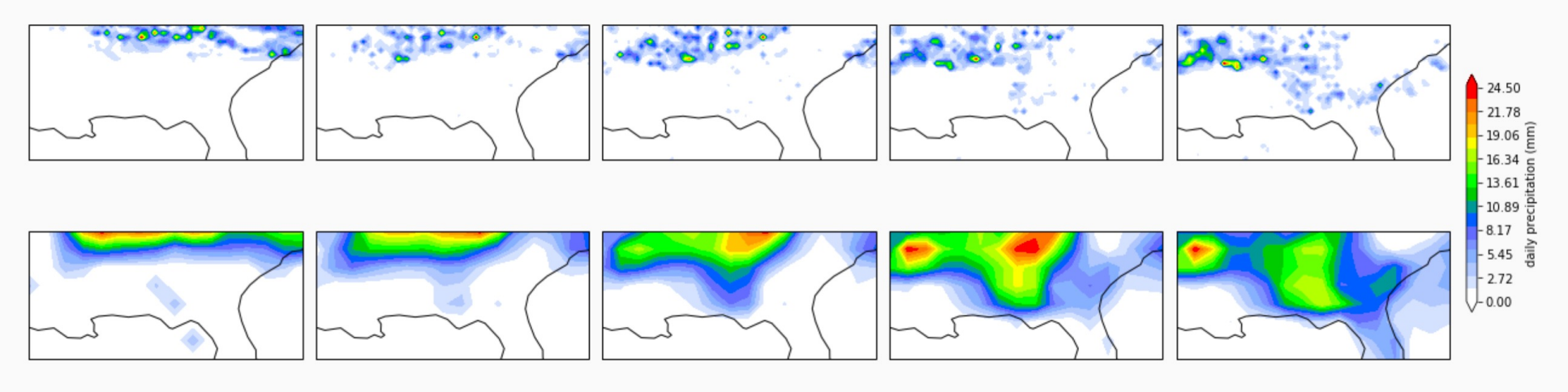
Temperatura máxima diária

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

Precipitação diária

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

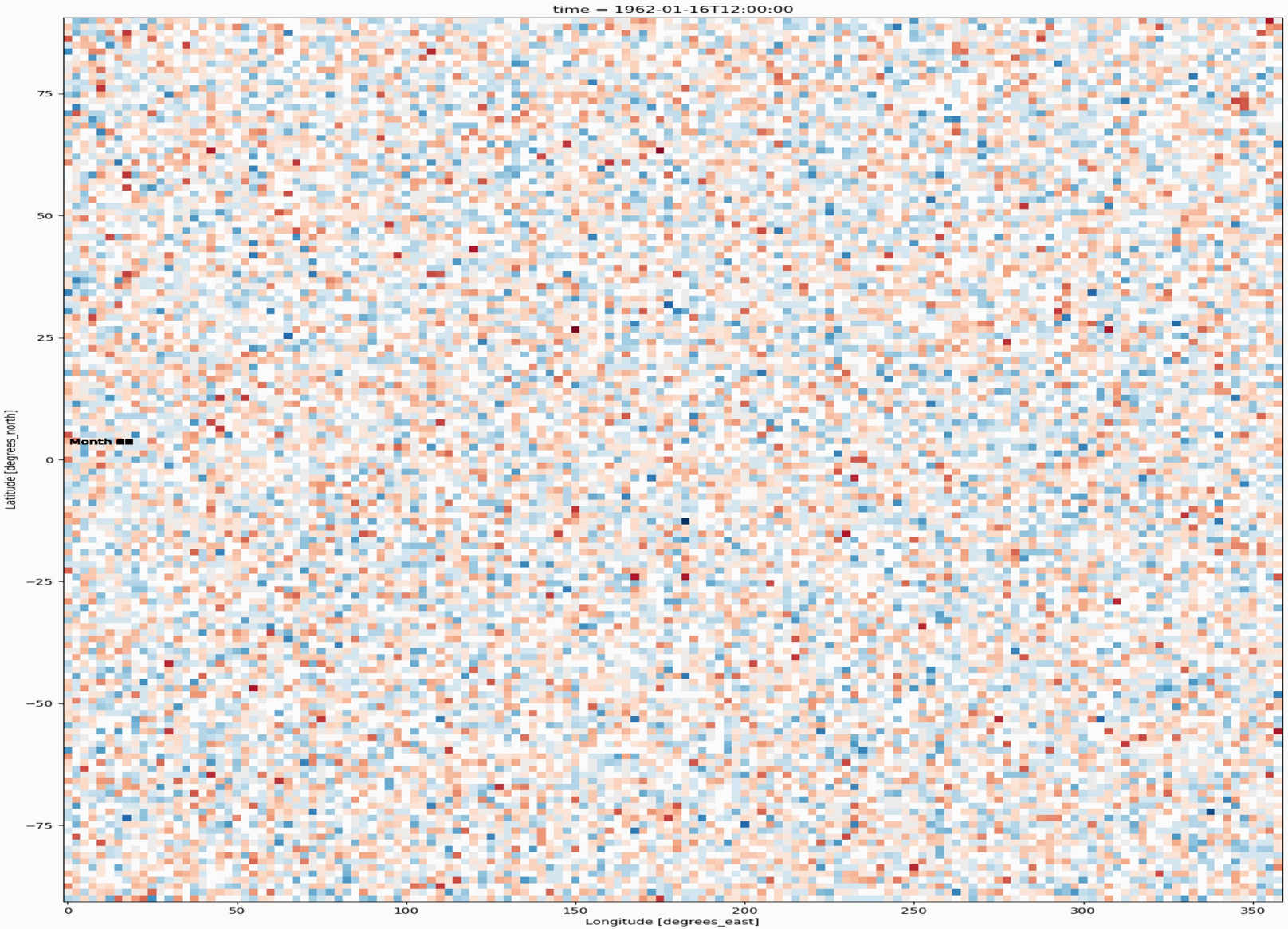
ClimAlign: downscaling generativo não supervisionado

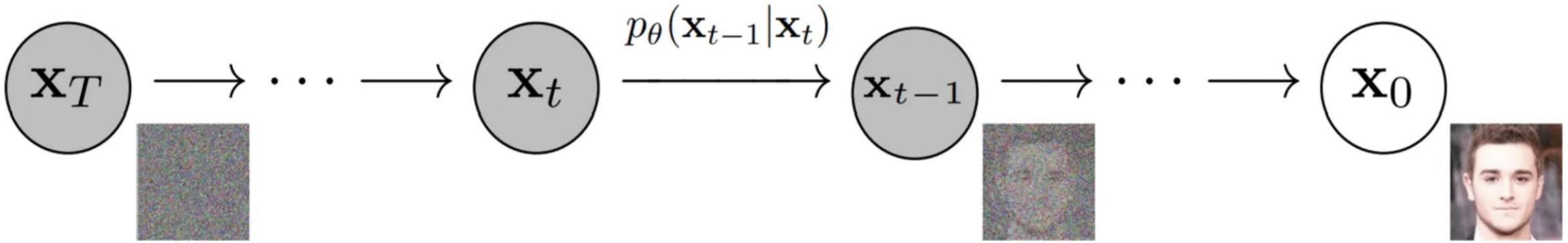


Técnica geral de downscaling por alinhamento de domínios com fluxos normalizadores [AlignFlow: Grover et al., AAAI 2020][Glow: Kingma & Dhariwal, NeurIPS 2018]

- **Não supervisionado**: não precisa de mapas pareados em baixa e alta resolução
- **Generativo**: pode amostrar a posterior sobre a representação latente OU amostrar condicionado a um mapa de baixa (ou alta!) resolução
- **Interpretável**, p. ex. por interpolação

Modelos generativos baseados em fluxos





Treinamento por difusão: aprender a gerar imagens realistas a partir de ruído aleatório

Treinamento do modelo de difusão

1. Criar uma tarefa pretexto a partir de dados de treinamento não rotulados: para cada ponto de treinamento x_0, criar uma sequência de imagens sintéticas adicionando ruído gaussiano à imagem anterior, até chegar a uma imagem indistinguível de ruído gaussiano.

2. Fazer então aprendizado "supervisionado" sobre os pares (x_{t-1}, x_t) para aprender $p_{\boldsymbol{\theta}}(x_{t-1} \mid x_t)$

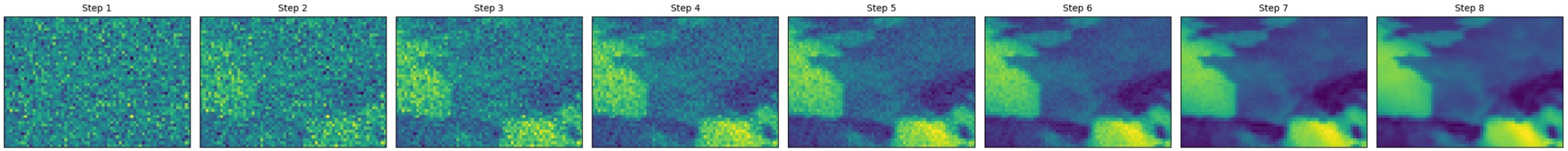
Inferência, ou seja, como gerar amostras parecidas com os dados de treinamento

1. Amostrar x_T aleatoriamente (gaussiano).

2. Amostrar T vezes de $p_{\boldsymbol{\theta}}$, condicionando a cada vez na amostra anterior

Modelos generativos baseados em fluxos

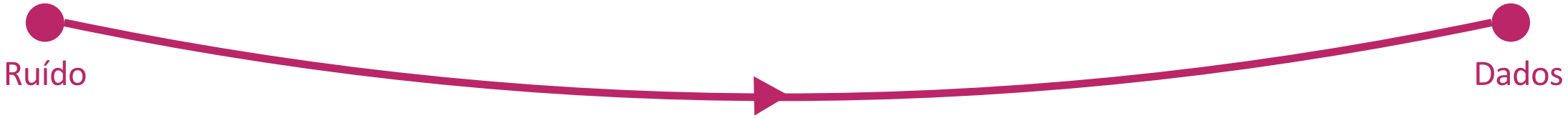
Modelo generativo baseado em fluxos: aprende um transporte (com uma rede neural) de uma distribuição fácil de amostrar (p. ex. ruído gaussiano) até a distribuição dos dados.



Difusão: um "removedor de ruído" inverte um processo de ruído estocástico, retirando ruído aos poucos – trajetória de EDE



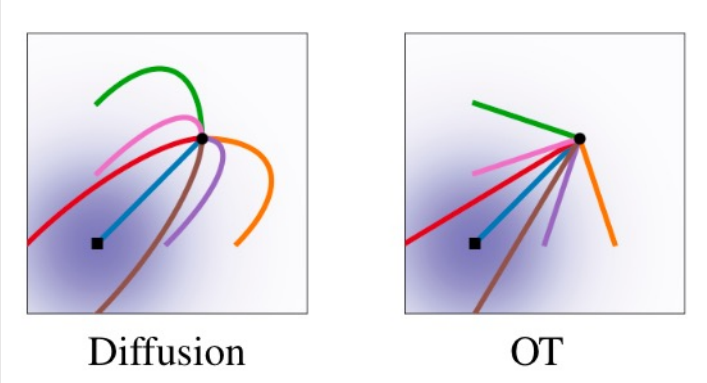
Flow matching [Lipman et al., 2023]: aprende um fluxo determinístico ruído → dados – trajetória de EDO



Crédito do slide: Julie Keisler (INRIA)

Flow matching vs. difusão

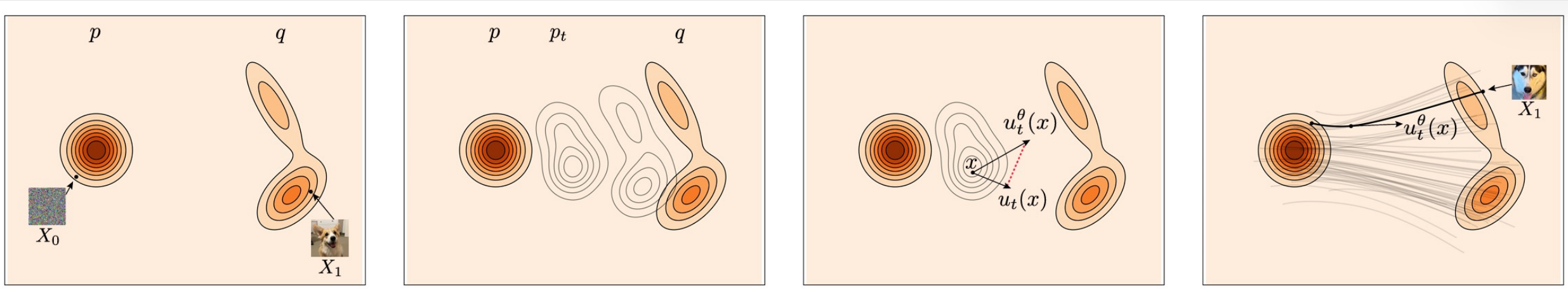
Intuição: criar caminhos mais retos entre amostras de dados e amostras de ruído:



Resultados: caminhos mais retos nos fluxos de probabilidade, que permitem amostrar em menos passos de difusão ganhos de velocidade

Generalidade: o flow matching pode ligar duas distribuições quaisquer (ou interpolantes estocásticos), ao contrário da difusão

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.

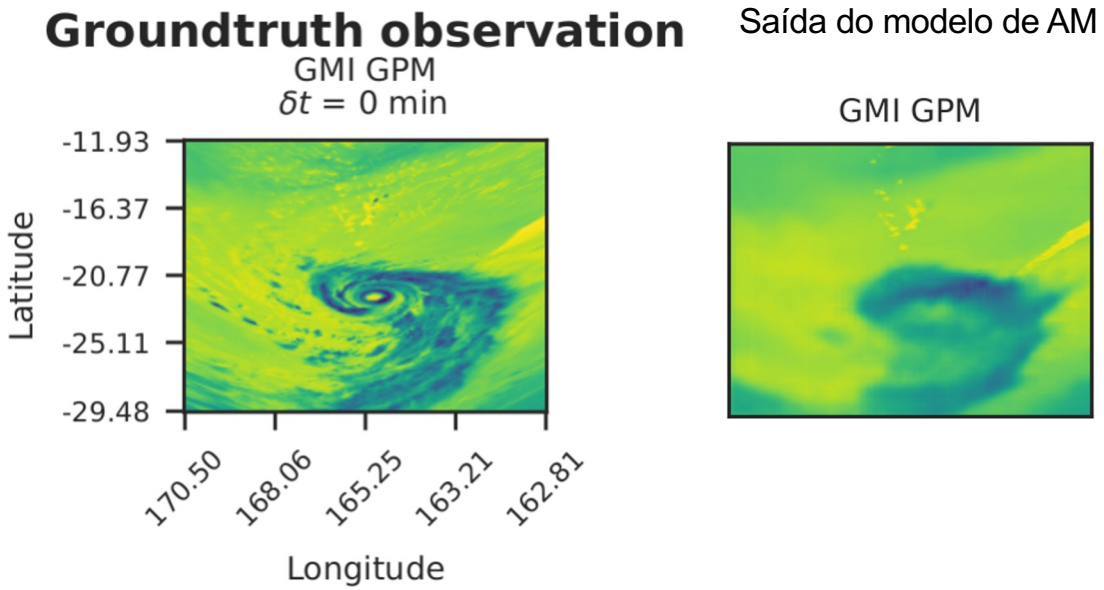
Lipman et al. (2024) Flow Matching Guide and Code

Por que usar IA generativa em clima e tempo?

Modelos de AM determinísticos costumam otimizar o desempenho **médio**

Treinados para minimizar uma perda média, p. ex. a RMSE

➔ Saídas de modelo borradas e não físicas (p. ex. imagens, campos, trajetórias de tempestades)



Saídas do modelo de AM, amostradas após o treinamento o mesmo backbone, **de forma generativa**

Por que usar IA generativa em clima e tempo?

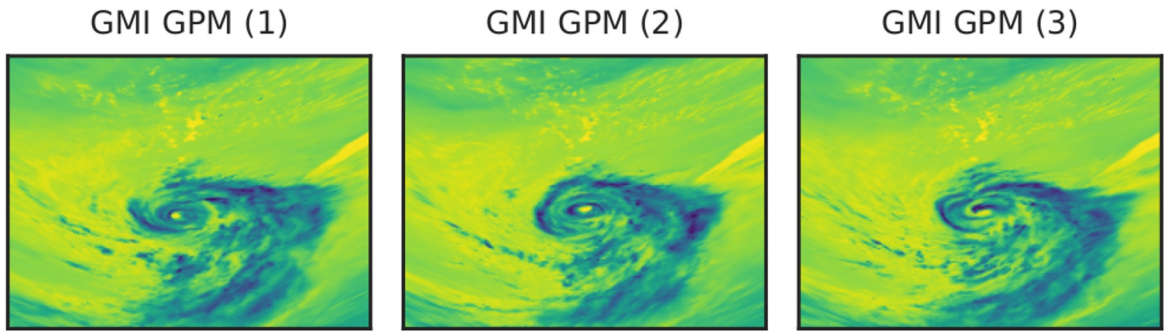
Para muitos problemas de clima, tempo e geociências

- Previsão do tempo, emulação do clima, geração de cenários
- Estimativa de extremos ou eventos raros, quantis
- Downscaling para escalas locais no tempo ou no espaço
- Quantificação de risco

A IA generativa traz benefícios

- Mais realismo físico
  - Mais precisão nas altas frequências
- Conjuntos probabilísticos
  - Quantificação da incerteza
  - Informação rica e realista para os gestores
- Geração de dados sintéticos

Amostras de um **modelo generativo**



IA generativa para:

✓ Previsão probabilística
- **ArchesWeatherGen**: previsão do tempo eficiente treinada a 1,5°
- **ArchesClimate**: emulação, interpolação e extrapolação do clima decenal

❑ Fusão de dados com alinhamento probabilístico de domínios
- **MotifGen**: um modelo generativo, multifonte e autossupervisionado para aprender com dados geoespaciais

❑ Alinhamento de domínios probabilístico e downscaling
- **ClimAlign**: downscaling generativo não pareado com alinhamento de domínios
- **SerpentFlow**: alinhamento de domínios generativo não pareado por decomposição em estrutura compartilhada (com aplicação ao downscaling)

Desafios de aprender com dados de várias fontes

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

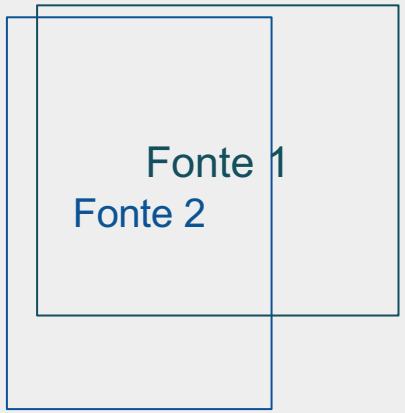
Daily acquisition capacity:
Up to 5,400,000 km²
(Wide ScanSAR Mode)



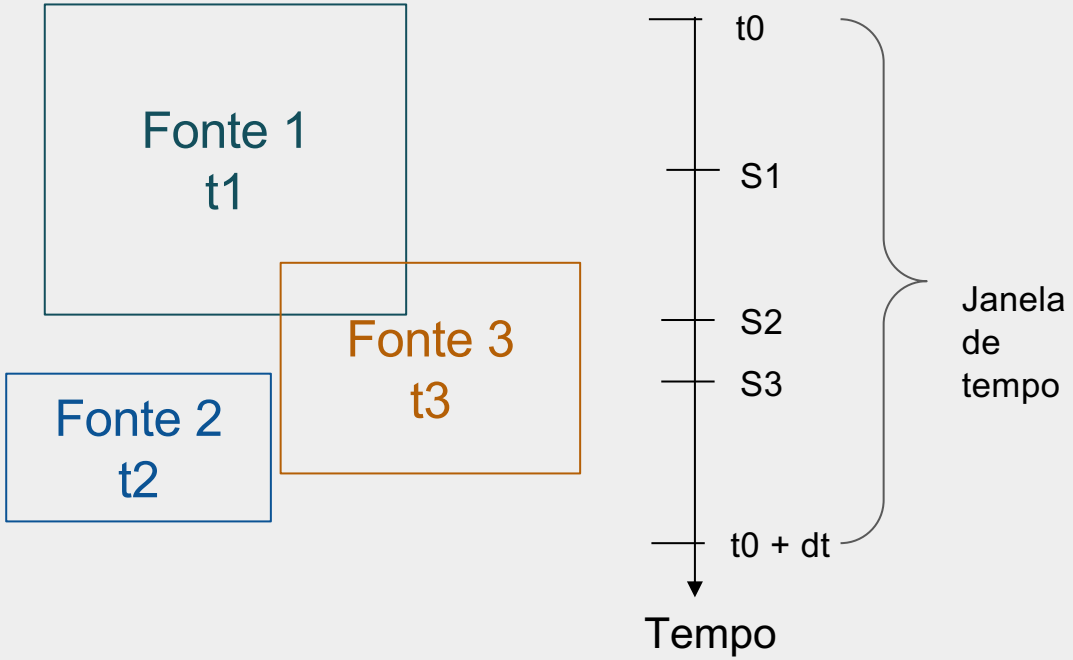
IA para fusão de dados de várias fontes (desalinhados)

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

Tradicionalmente: só treinar modelos com dados colocalizados



Nosso modelo de IA (MOTIF) consegue aprender com **todas as fontes de dados**, mesmo dados desalinhados no tempo e no espaço



Uma arquitetura multifonte para fusão de dados geoespaciais

AM fundamental multimodal

- Embutir **valores** (p. ex. texto, imagens)
- Condicionamento (“cachorro ou gato”)

Modelo DiT (2022)

Aprendizado de representações geoespaciais

- Embutir **localizações**
- Treinamento autossupervisionado em larga escala

SatCLIP (2025), AlphaEarth (2025)

MOTIF: *Transformer multifonte com atenção fatorada*

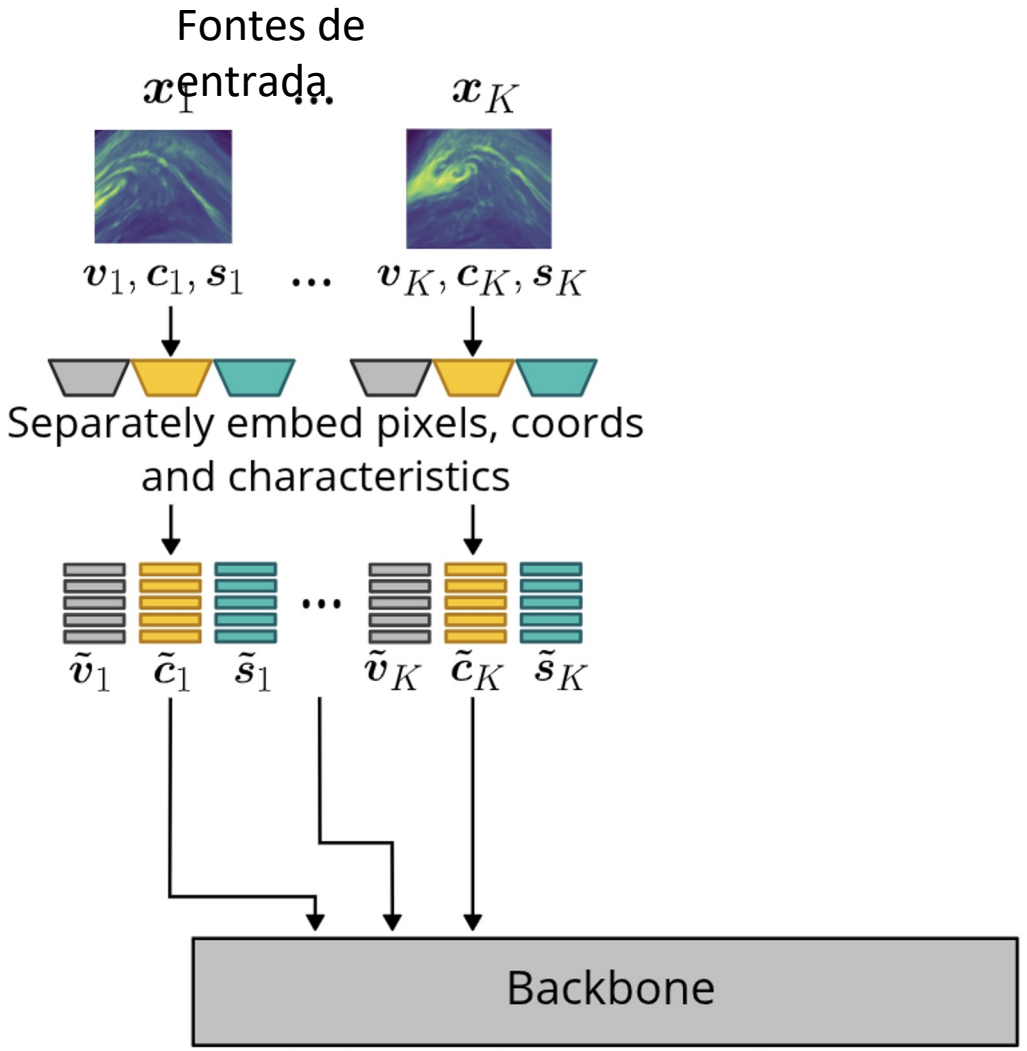
- Embutir separadamente **valores** e **coordenadas**
- Condicionar às características da fonte
- Pré-treinamento autossupervisionado

[Dauvilliers & Monteleoni, Environmental Data Science, 2025]



Pipeline de treinamento

Embutir separadamente:
- valores,
- coordenadas,
- características da fonte



Embedding posicional relativo

- Embedding posicional relativo usual nos Transformers (Shaw et al., 2018)

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

- Aqui, as coordenadas espaço-temporais são usadas no embedding posicional:

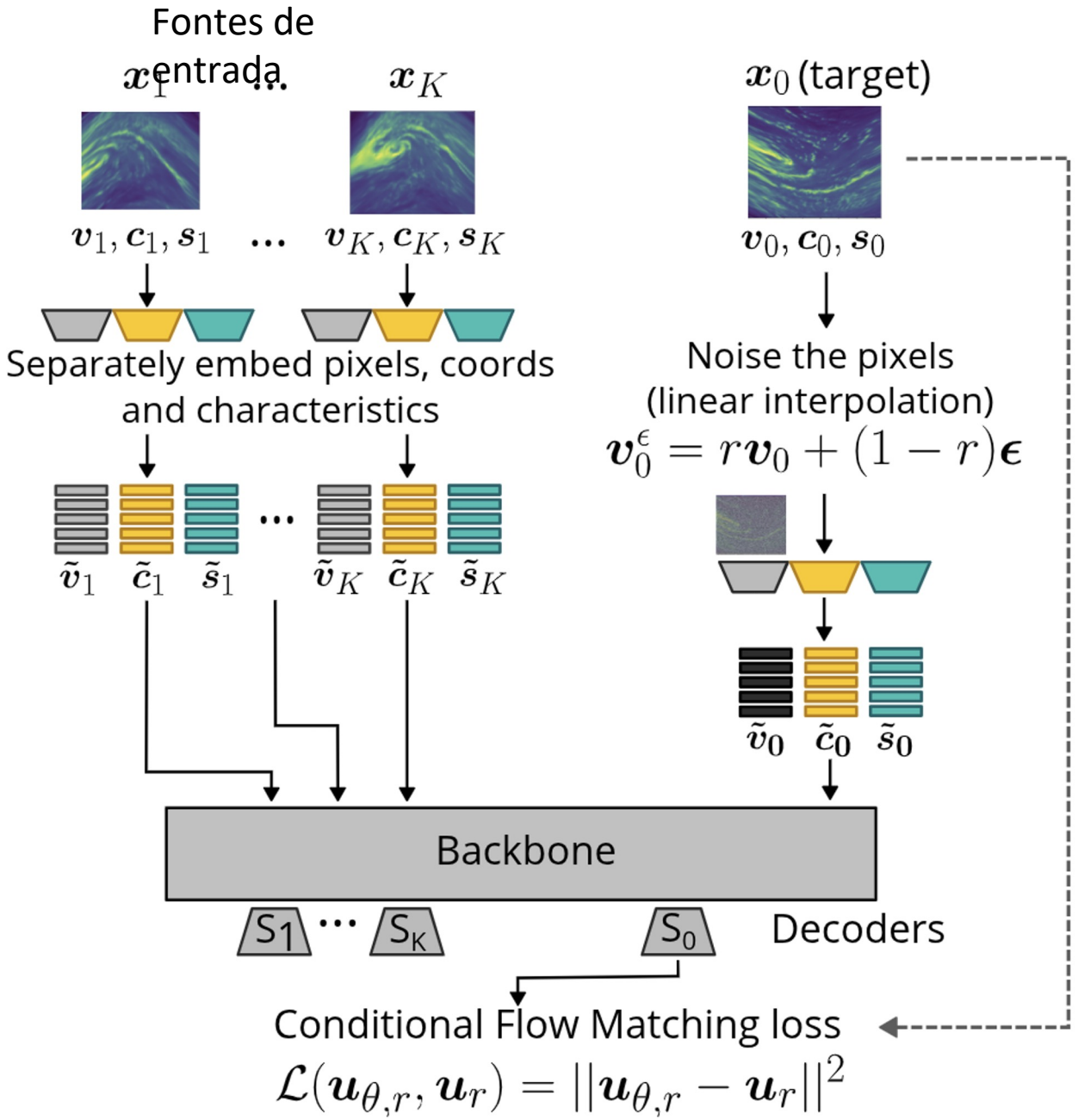
$$\tilde{v}_{out} = \frac{Softmax\left(Q_V K_V^T + Q_C K_C^T\right)}{\sqrt{d}} \tilde{v}_{in}$$

Valores (p. ex. pixels)

Coordenadas espaço-temporais

Pipeline de treinamento

Embutir separadamente:
- valores,
- coordenadas,
- características da fonte



Tarefa de treinamento autossupervisionado: reconstruir uma fonte mascarada

Versão generativa treinada com flow matching

[Dauvilliers & Monteleoni, Climate Informatics 2026]

Aplicação a ciclones tropicais

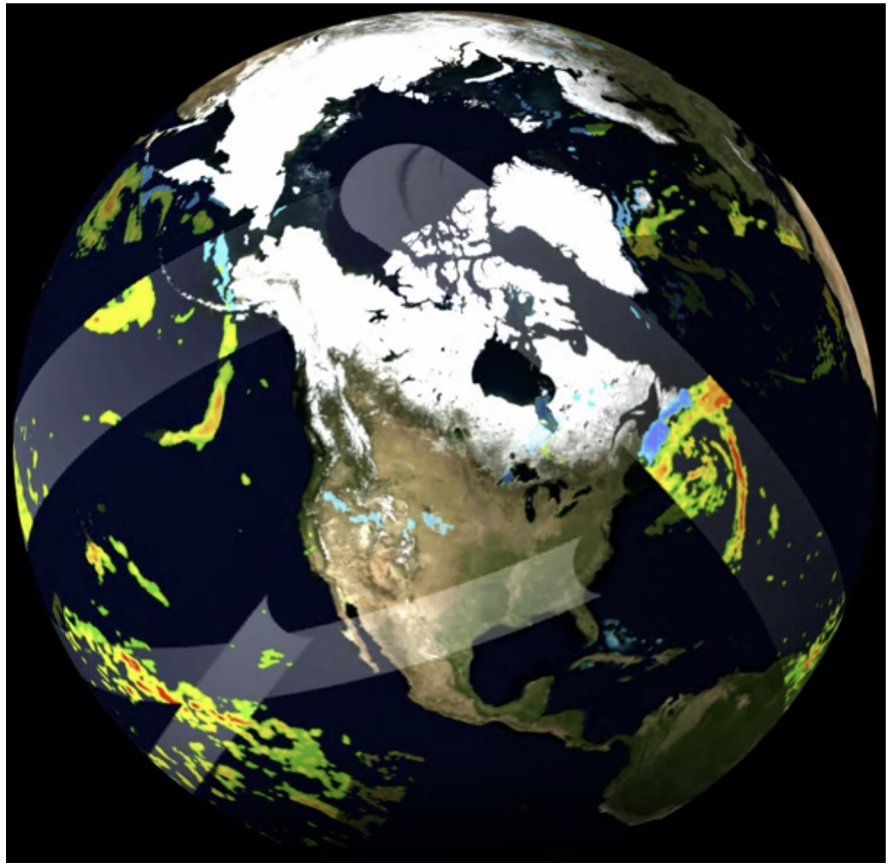
A constelação GPM

GMI (instrumento de referência)



⇒ Dá para fazer interpolação espaço-temporal de imagens GMI?

A constelação GPM, NASA



A constelação GPM cobre a Terra, NASA's Scientific Visualization Studio

156

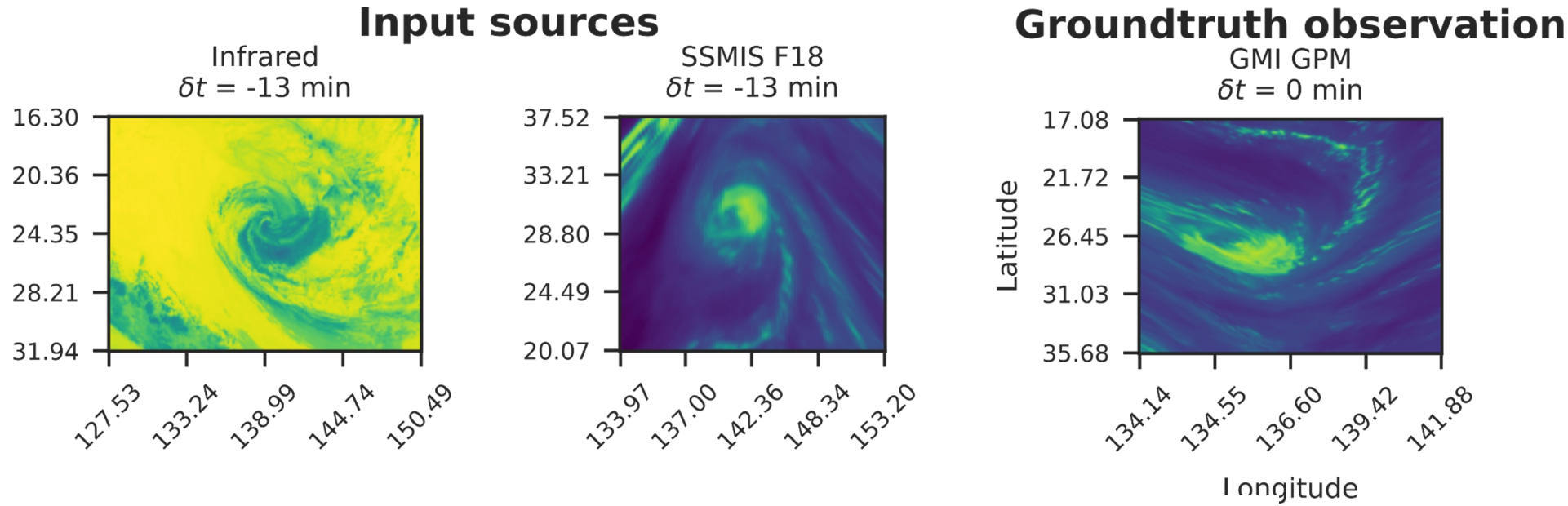
Conjunto de dados

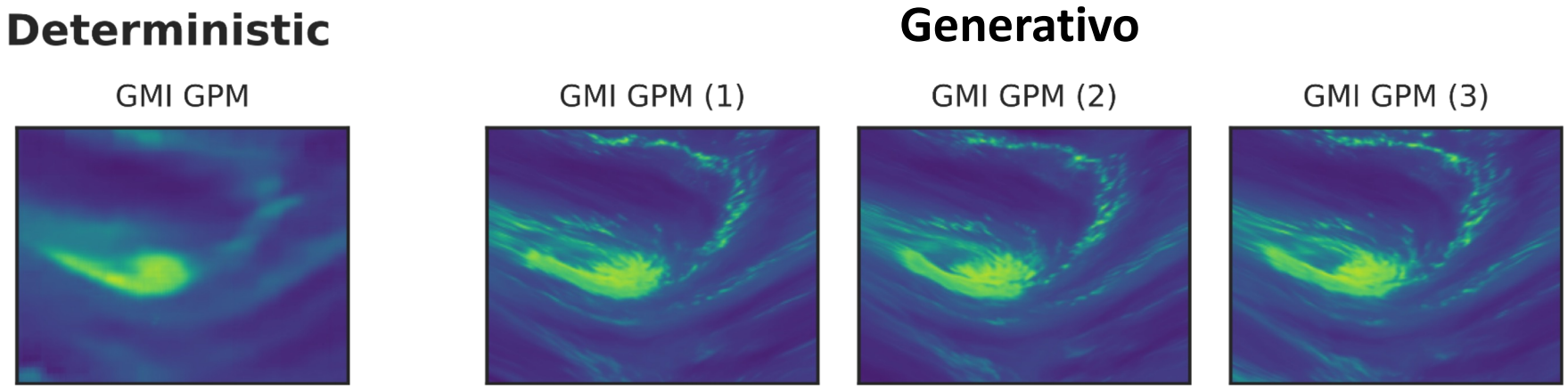
TC-PRIMED (Razin et al. 2023)

- ~3 mil tempestades, de 1987 a 2024, todas as bacias
- 2005-07-14-21-23 para validação, 2006-08-15-22-24 para teste
- 11 sensores de micro-ondas
- Observações no infravermelho

Janelas de interpolação de 6 horas

Resultados: capta melhor a estrutura fina que o modelo determinístico





Pipelines de treinamento comparados para a versão generativa

- Treinamento supervisionado

Sensores não GMI → Modelo → GMI

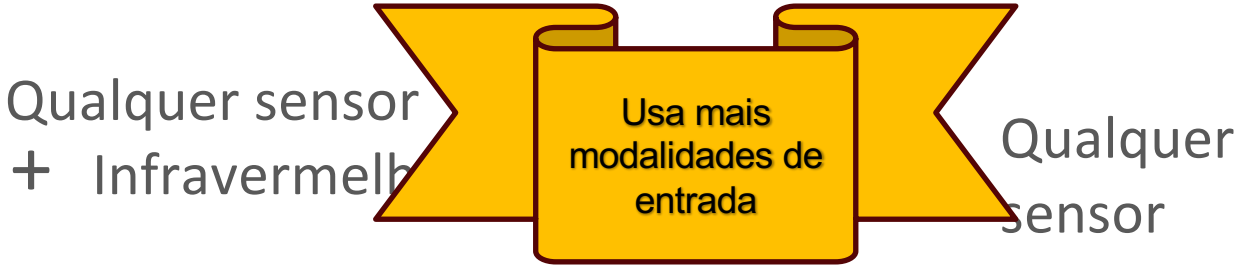
⇒ ~6 mil amostras

- Treinamento autossupervisionado



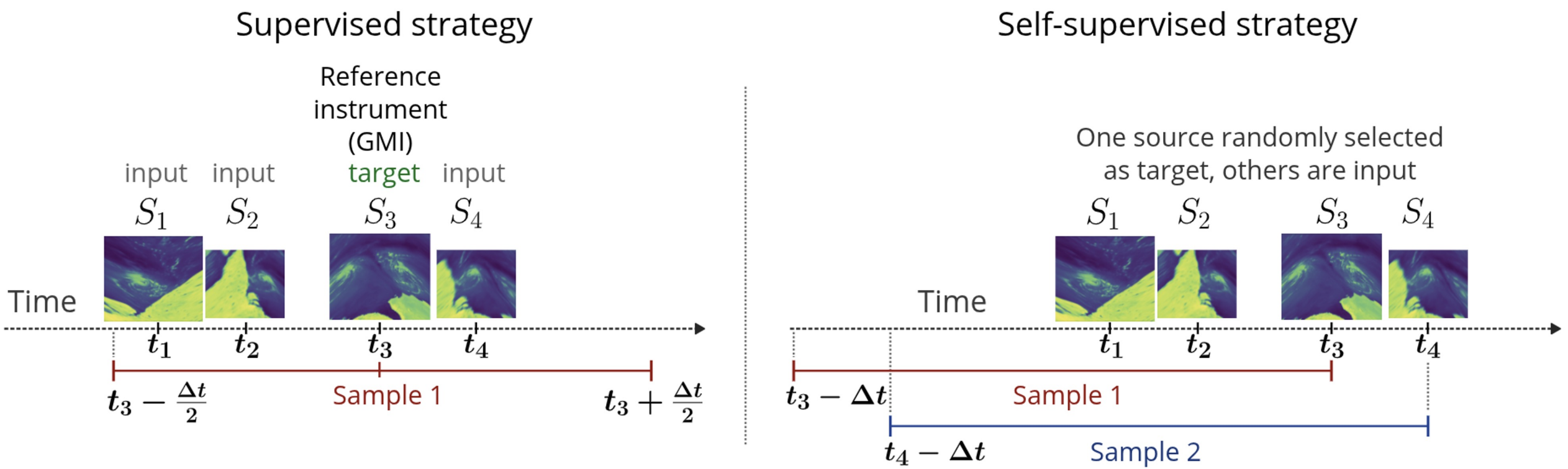
- Autossupervisionado + infravermelho

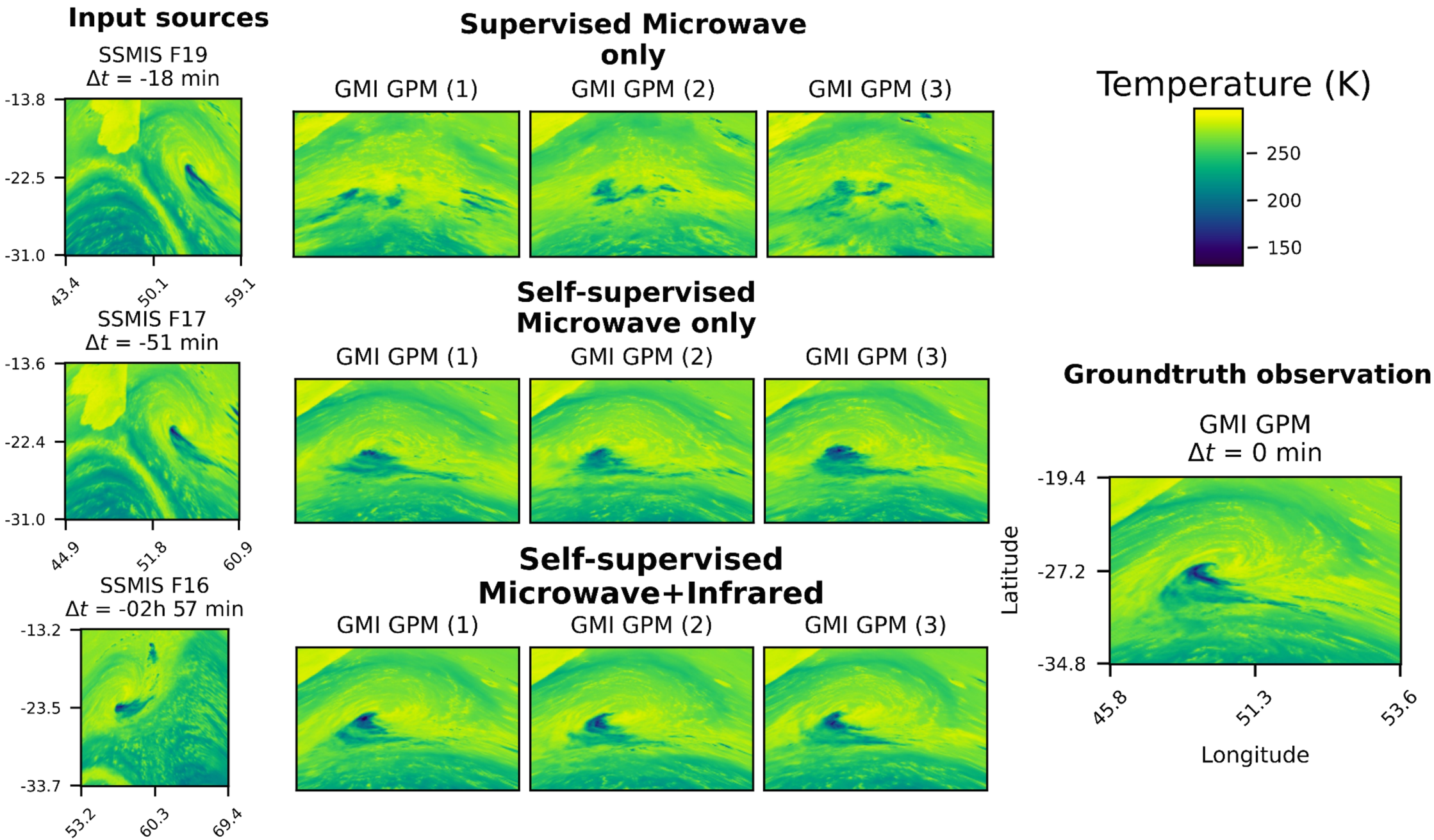
⇒ ~48 mil amostras

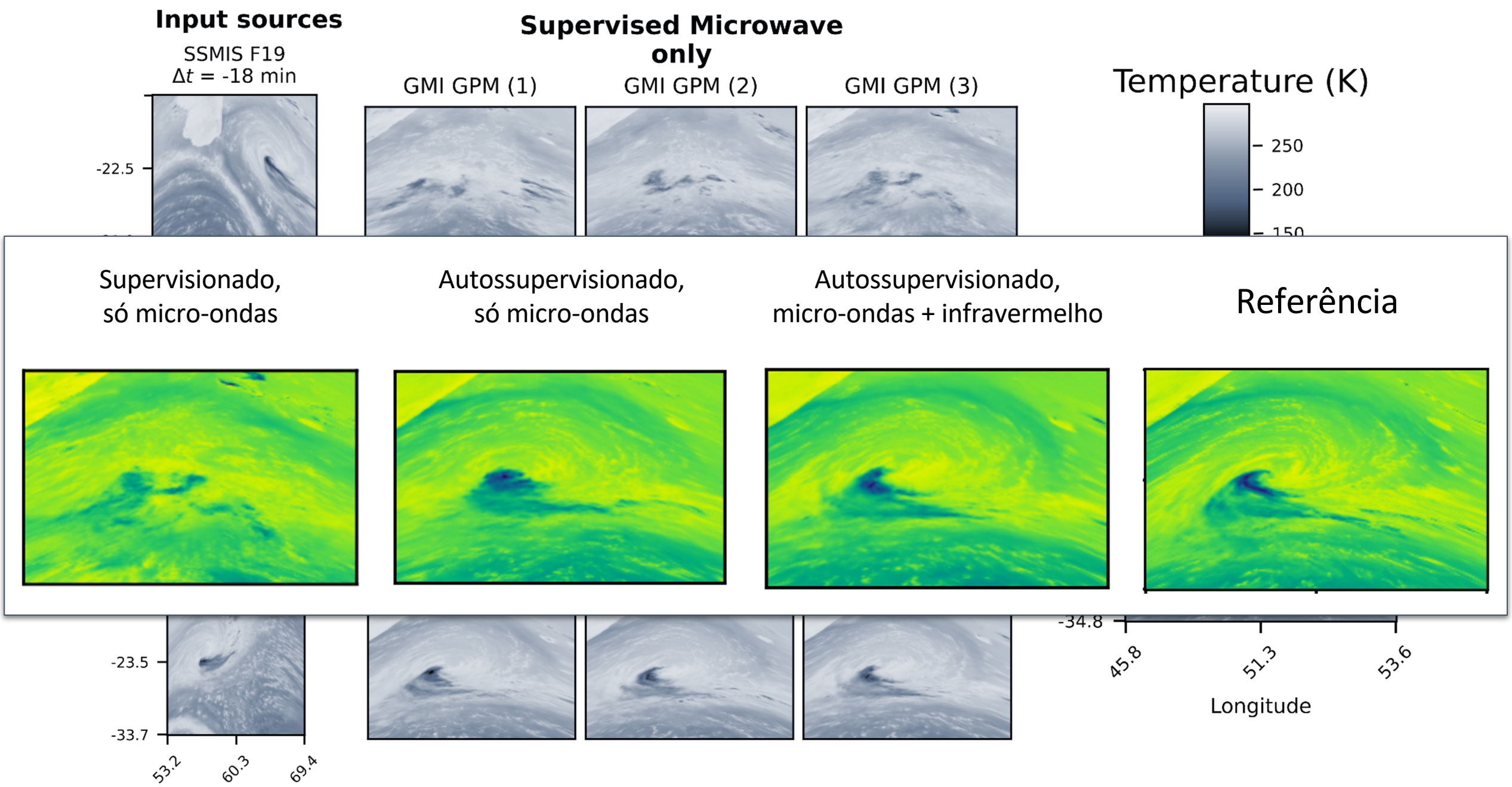


⇒ ~44 mil amostras

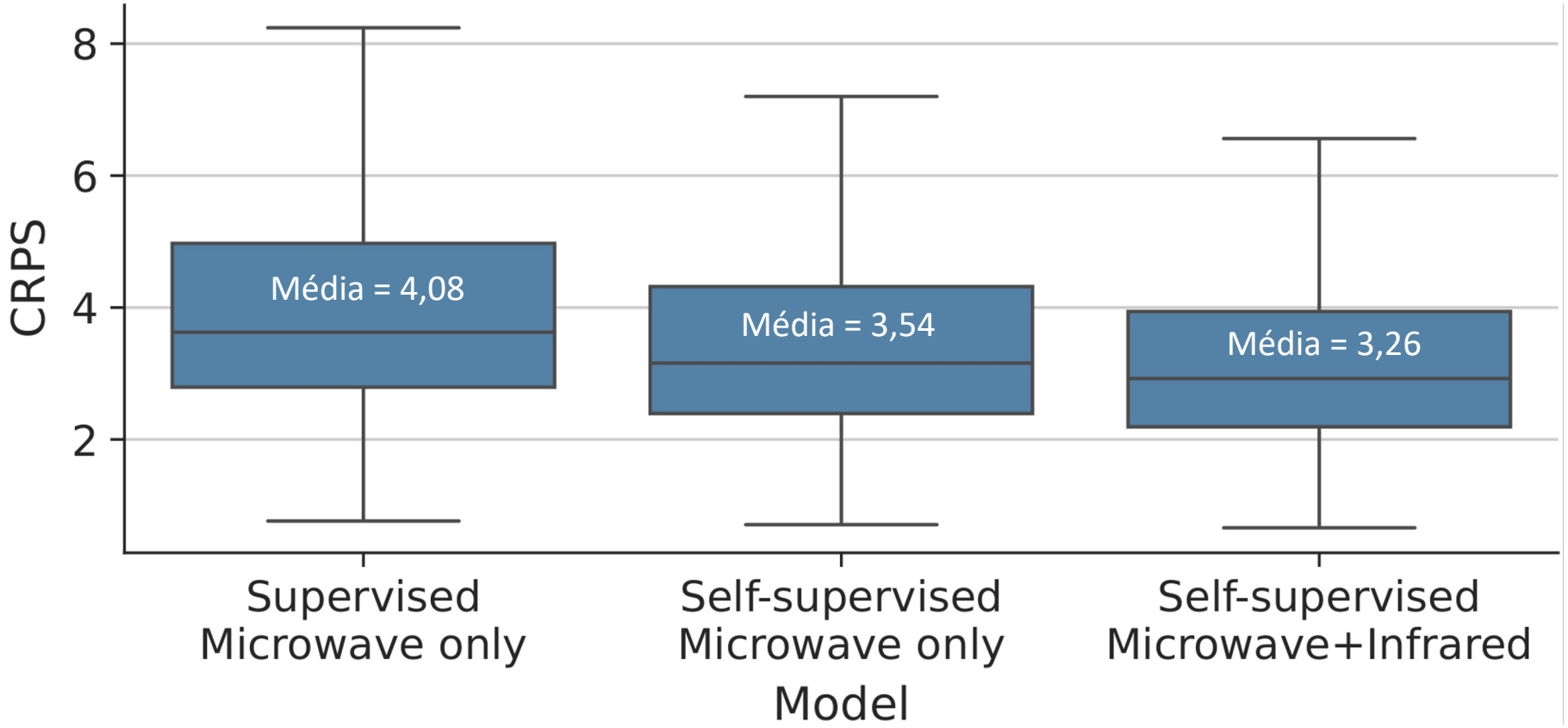
Tarefas de treinamento




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
Supervisionado, só micro-ondas
Autossupervisionado, só micro-ondas
Autossupervisionado, micro-ondas + infravermelho
Referência

Versão generativa: ablações



Resultados

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	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]

Linha de base

Os colchetes indicam intervalos de confiança de 95%

Escore de probabilidade ranqueada contínua (CRPS)

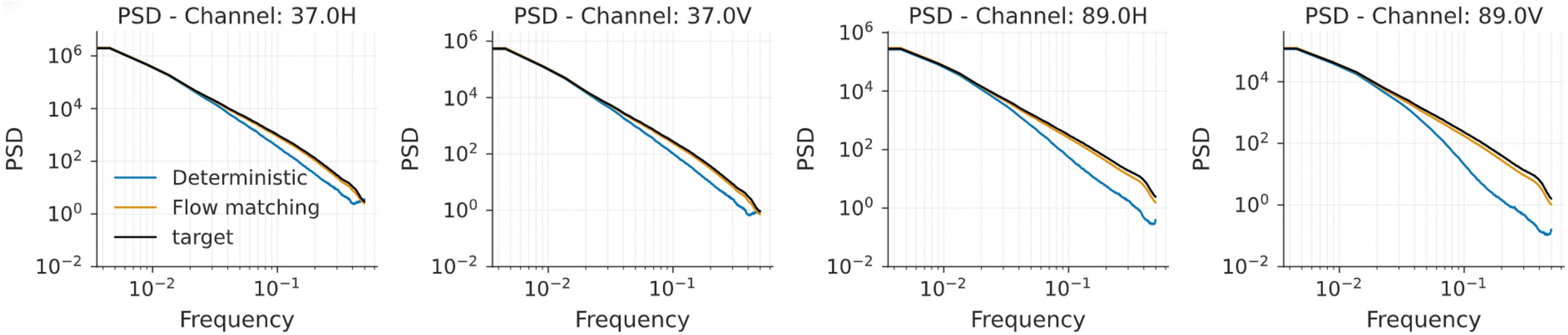
$$\mathrm{CRPS}(F, x) = -\int_{-\infty}^{\infty} (F(y) - \mathbb{1}\{y \geq x\})^2 \, \mathrm{d}y$$

- Outra formulação evita calcular esta integral
- O CRPS pode ser comparado diretamente ao MAE de uma previsão determinística
- O Energy Score é a extensão multivariada do CRPS

Leituras complementares sobre estas e outras medidas de avaliação de previsões probabilísticas:

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

Avaliação frente ao modelo determinístico - espectro

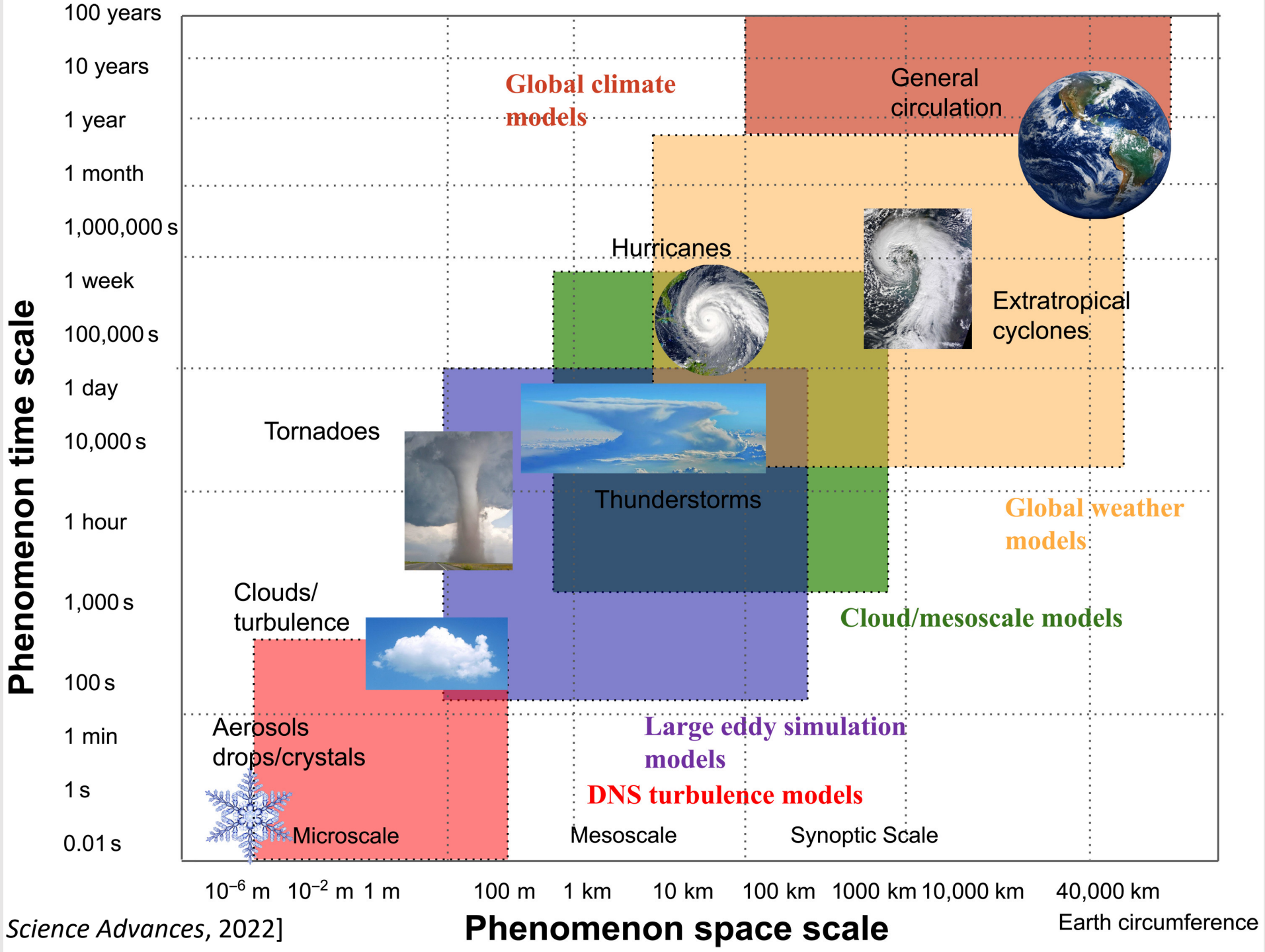


IA generativa para:

✓ Previsão probabilística
- **ArchesWeatherGen**: previsão do tempo eficiente treinada a 1,5°
- **ArchesClimate**: emulação, interpolação e extrapolação do clima decenal

✓ Fusão de dados com alinhamento probabilístico de domínios
- **MotifGen**: um modelo generativo, multifonte e autossupervisionado para aprender com dados geoespaciais

❑ Alinhamento de domínios probabilístico e downscaling
- **ClimAlign**: downscaling generativo não pareado com alinhamento de domínios
- **SerpentFlow**: alinhamento de domínios generativo não pareado por decomposição em estrutura compartilhada (com aplicação ao downscaling)



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



Downscaling não supervisionado de modelos climáticos por alinhamento de domínios: *aplicação a campos de vento.*

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

Anais do Climate Informatics 2026

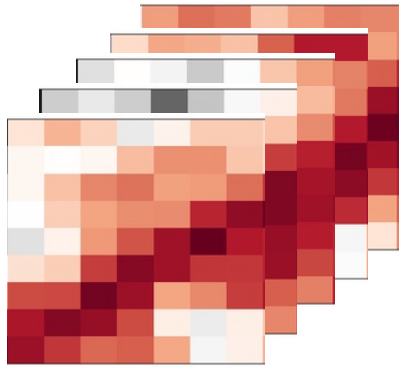
Downscaling de modelos climáticos

Resolução espacial grossa: os MCGs operam a ~100 km ou mais e deixam de fora processos locais essenciais

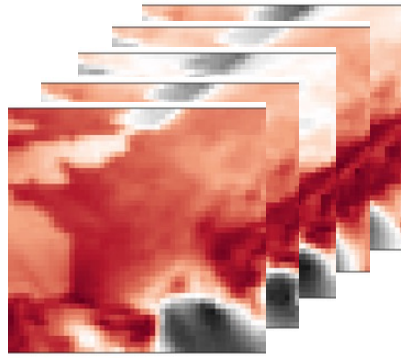
Representação limitada: feições de escala fina, como efeitos topográficos e contrastes terra-mar, são mal captadas

Vieses sistemáticos: as saídas dos MCGs se desviam bastante dos padrões climáticos observados

Estudos de impacto locais comprometidos: as saídas diretas dos MCGs não bastam para avaliar os padrões futuros de vento no planejamento energético

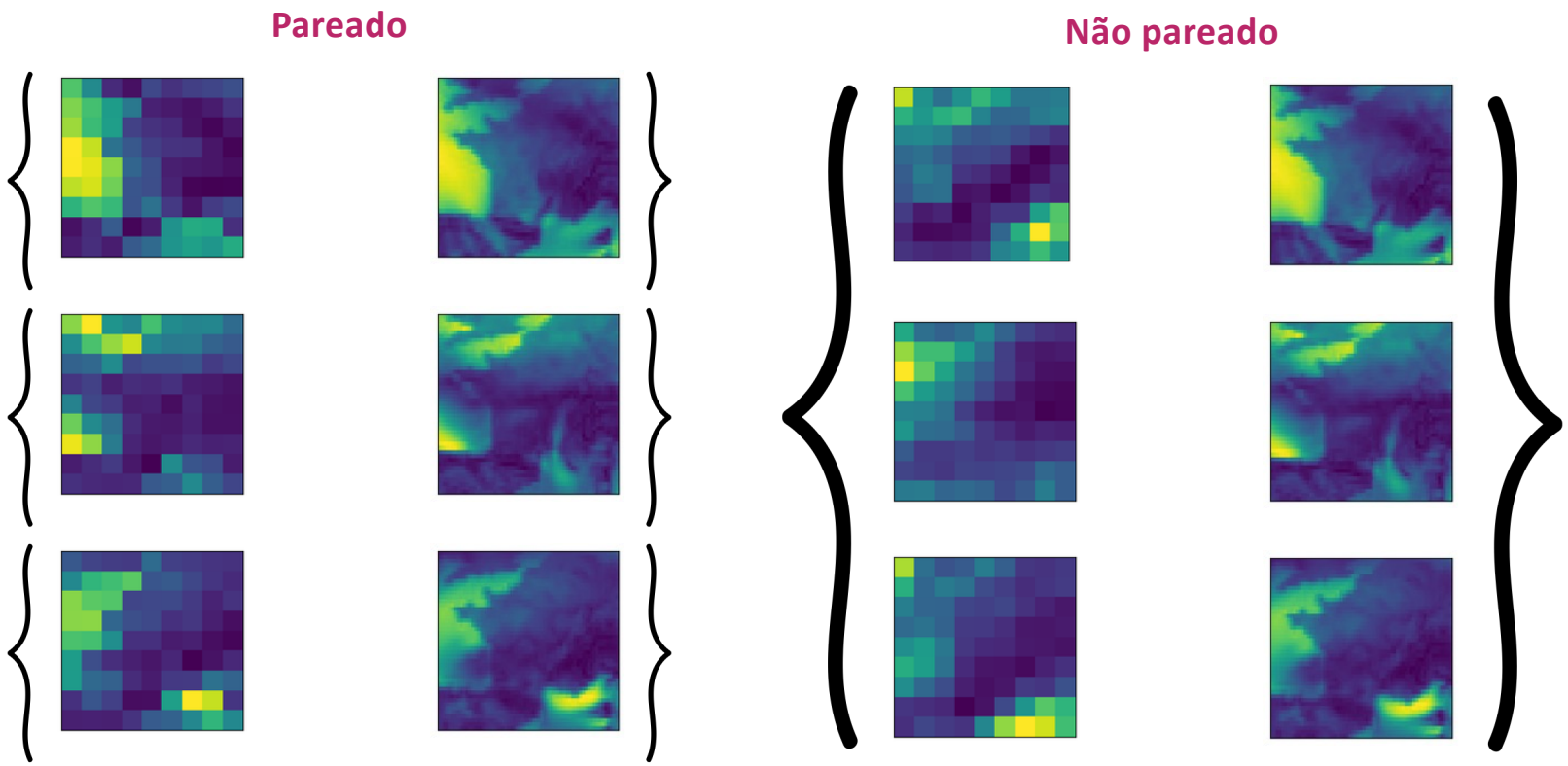


MCG ACCESS-ESM1 sobre a França em 5 dias



ERA5 sobre a França em 5 dias

Downscaling como alinhamento de domínios não pareado



- MCG: **trajetórias fisicamente plausíveis**
- Reanálises: a sequência **meteorológica histórica** real → **MCG e reanálise não são pareados**
- Downscaling = **alinhamento de distribuições**: mapear as saídas do MCG na distribuição da reanálise

Limitações dos métodos estatísticos tradicionais

Métodos baseados em distribuições

- Quantile mapping, CDF-t (Michelangeli et al. 2009), correção das distribuições marginais

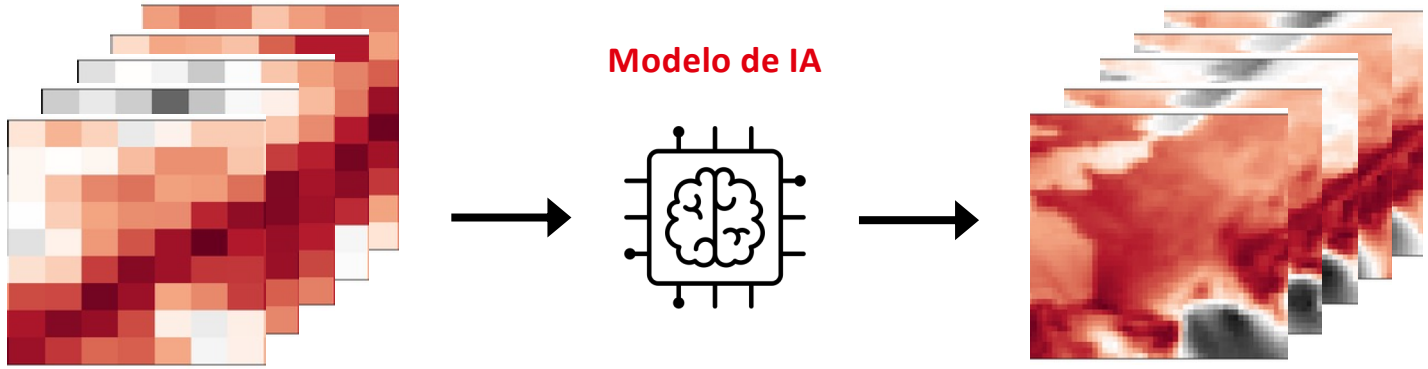
Métodos multivariados (transporte ótimo)

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

Maldição da dimensionalidade: os métodos tradicionais não conseguem modelar todas as interações em jogo

⟶ Métodos baseados em IA / aprendizado profundo

IA generativa para downscaling



Vantagens

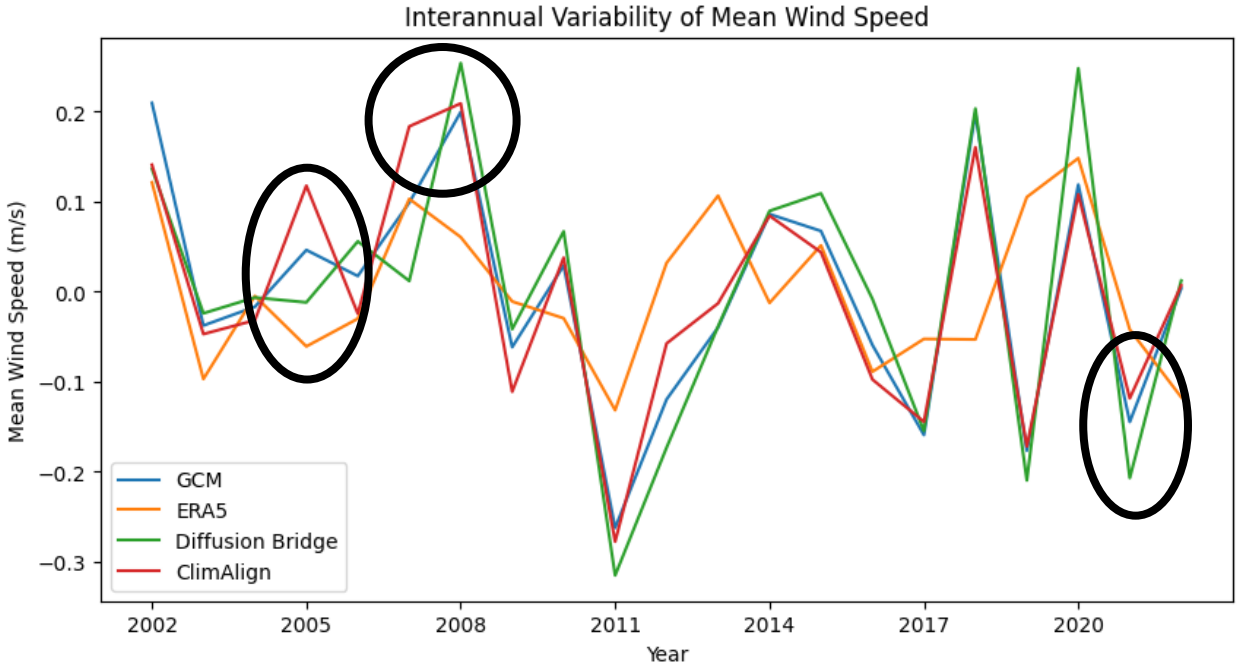
- Aprende a **distribuição conjunta completa** dos campos de alta resolução
- Gera **várias realizações**

Desafios

- Dados não pareados ⟶ **sem verdade de referência**
- Métodos implícitos podem se afastar do sinal do M

= projeções **pouco confiáveis**

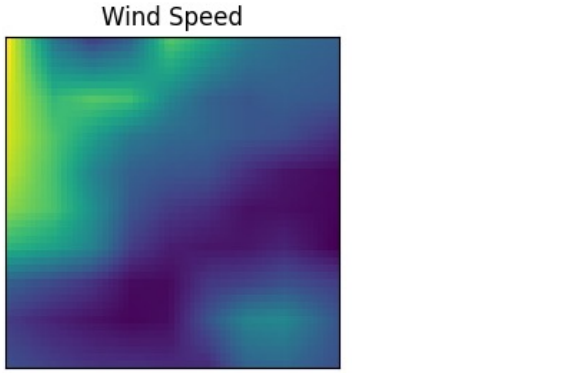
(ClimAlign/Groenke et al. 2020, pontes de difusão/Bischoff & Deck 2024)



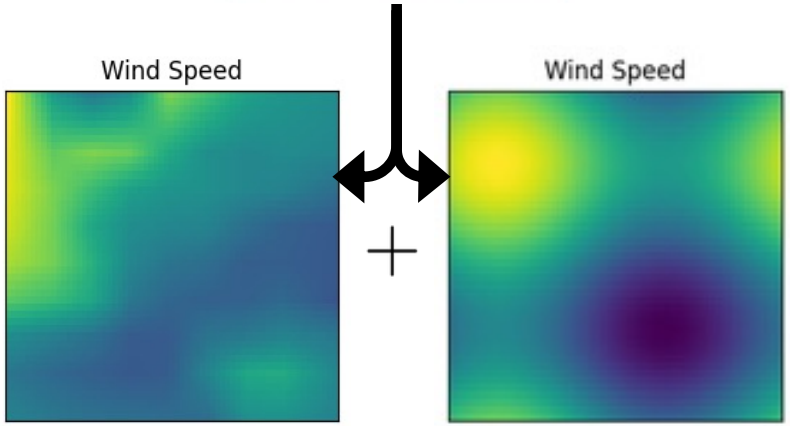
SerpentFlow: introduzindo pseudopares

SerpentFlow é um método generativo baseado em IA para **alinhamento de domínios não pareado** por decomposição em estrutura compartilhada

Velocidade média do vento do MCG



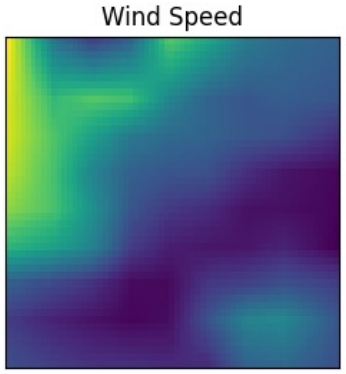
Separar as frequências espaciais altas das baixas

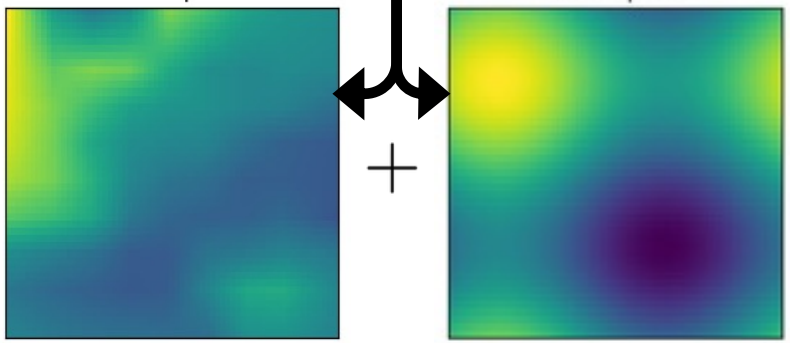


SerpentFlow: introduzindo pseudopares

SerpentFlow é um método generativo baseado em IA para **alinhamento de domínios não pareado** por decomposição em estrutura compartilhada

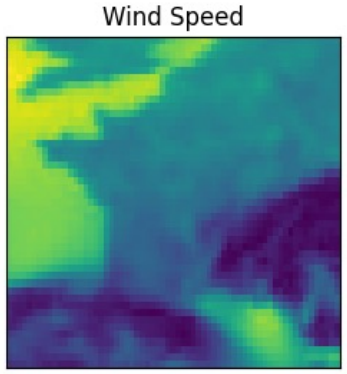
Velocidade média do vento do MCG

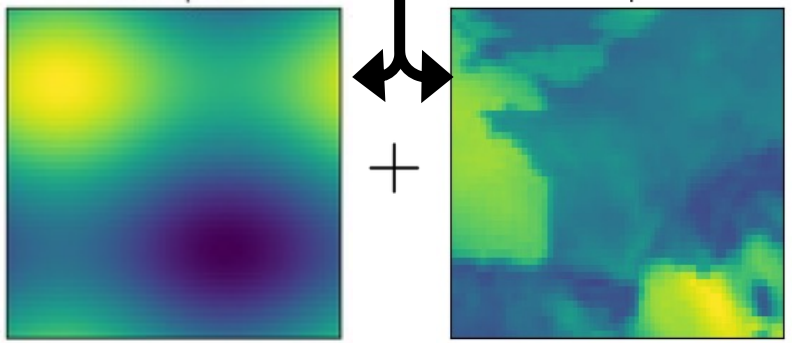




Separar as frequências espaciais altas das baixas

Velocidade média do vento do ERA5

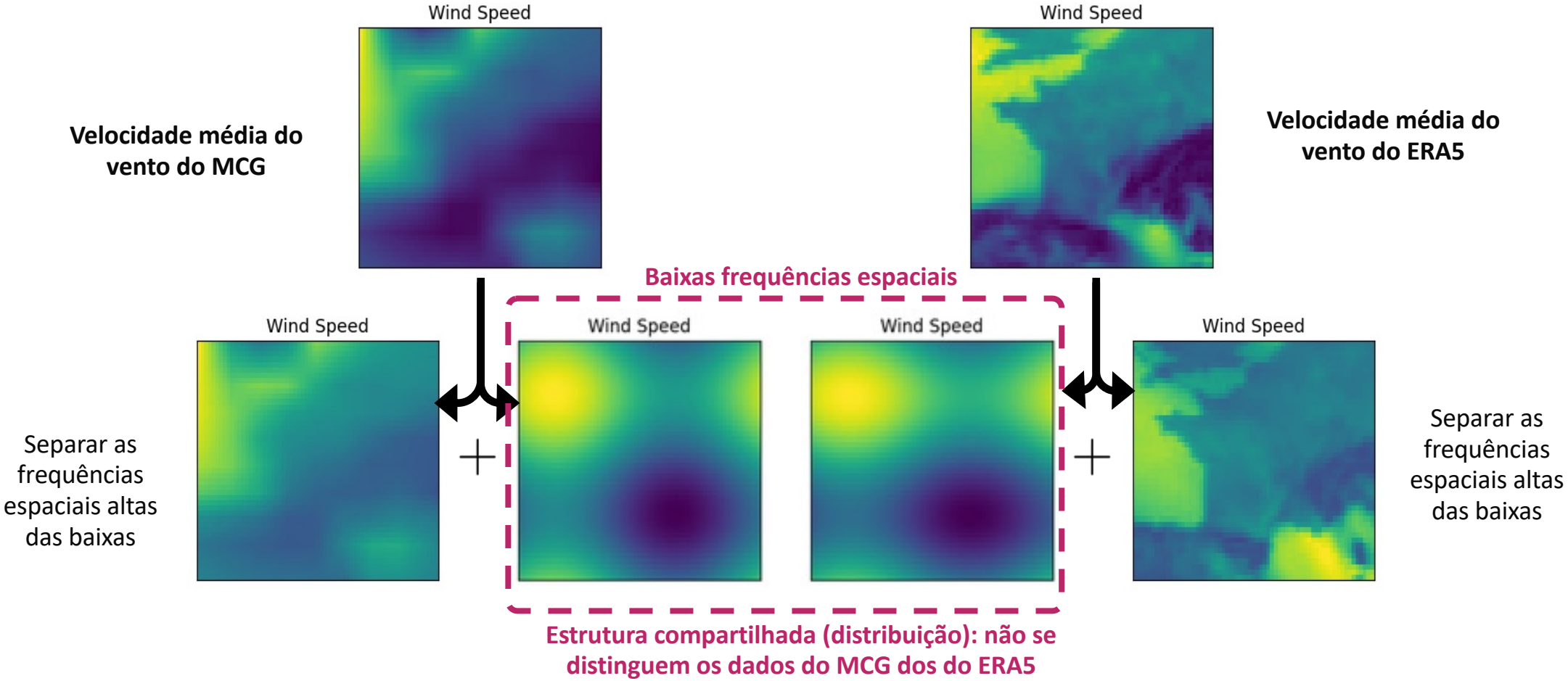




Separar as frequências espaciais altas das baixas

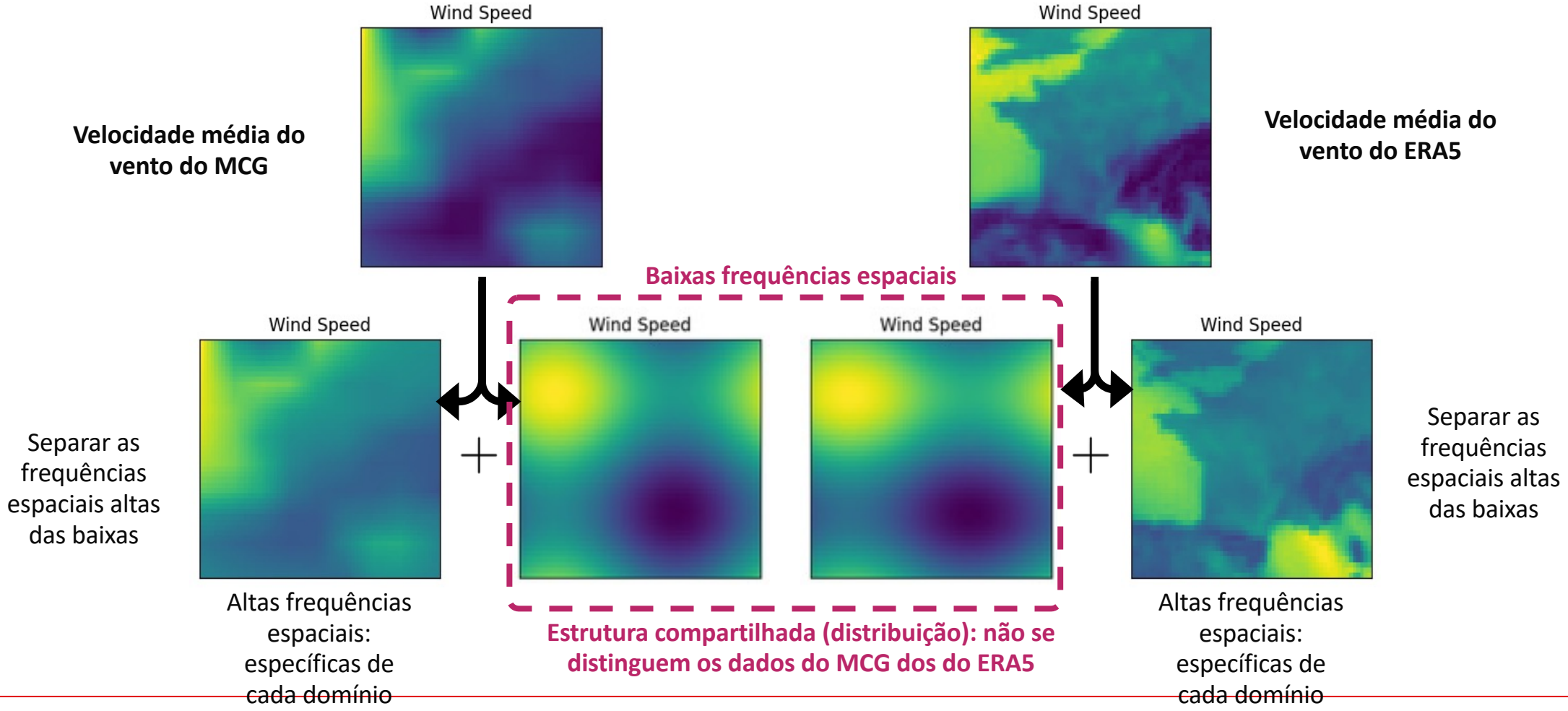
SerpentFlow: introduzindo pseudopares

SerpentFlow é um método generativo baseado em IA para **alinhamento de domínios não pareado** por decomposição em estrutura compartilhada

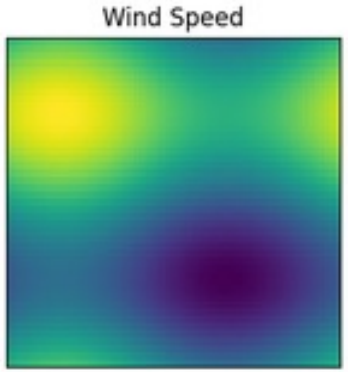


SerpentFlow: introduzindo pseudopares

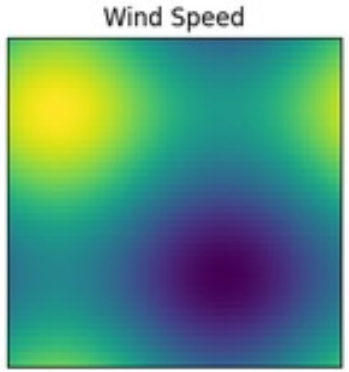
SerpentFlow é um método generativo baseado em IA para **alinhamento de domínios não pareado** por decomposição em estrutura compartilhada



Encontrar automaticamente a frequência de corte



Filtro passa-baixa(AR) com frequência de corte w



Filtro passa-baixa(BR) com frequência de corte w

→

Classificador

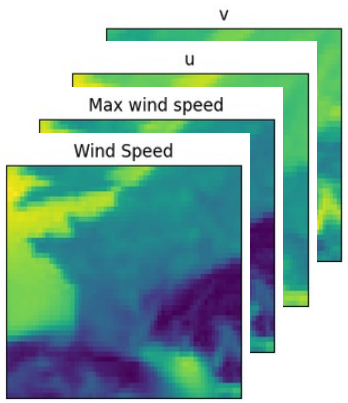
→

Dados BR ou AR?

- Parar de reduzir a frequência de corte quando a acurácia <= 60%
- Frequência muitas vezes encontrada ~= **resolução efetiva** do MCG

Pipeline do SerpentFlow

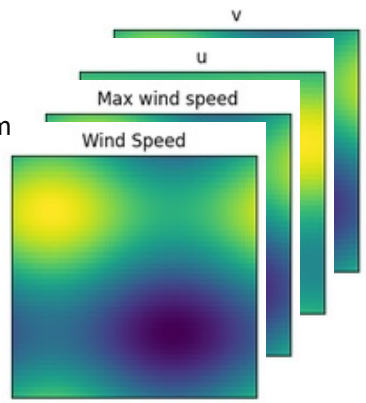
Treinamento



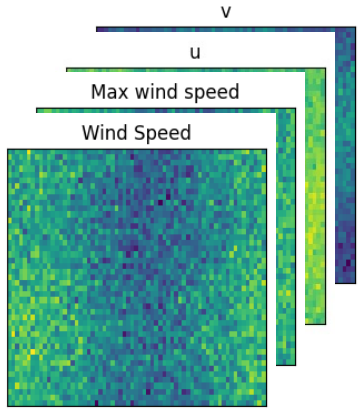
Referência (obs./reanálise)

Filtro passa-baixa, com frequência de corte

w_c

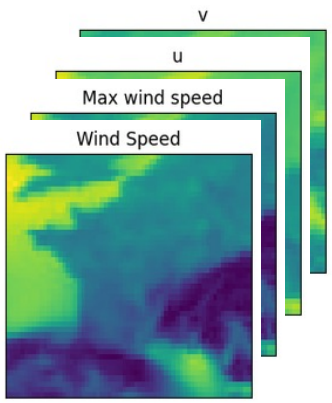


Adicionar ruído nas altas frequências



Treinar a abordagem generativa

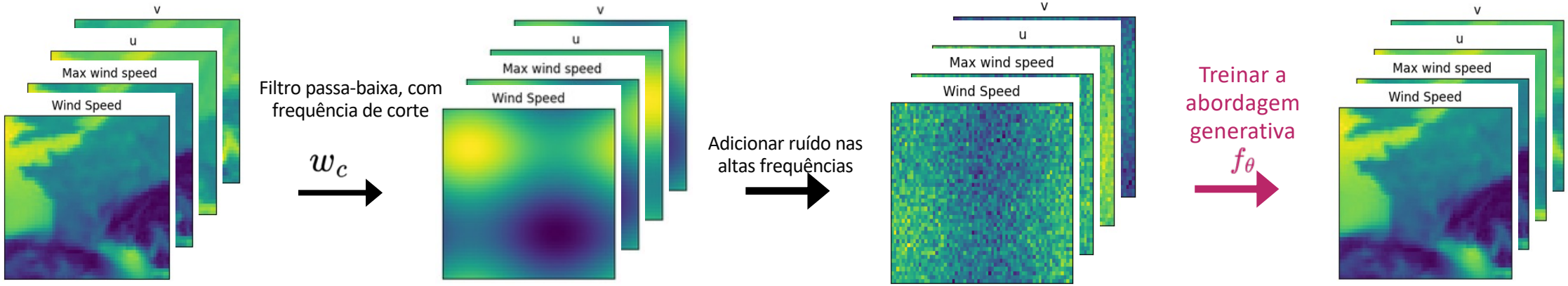
f_θ



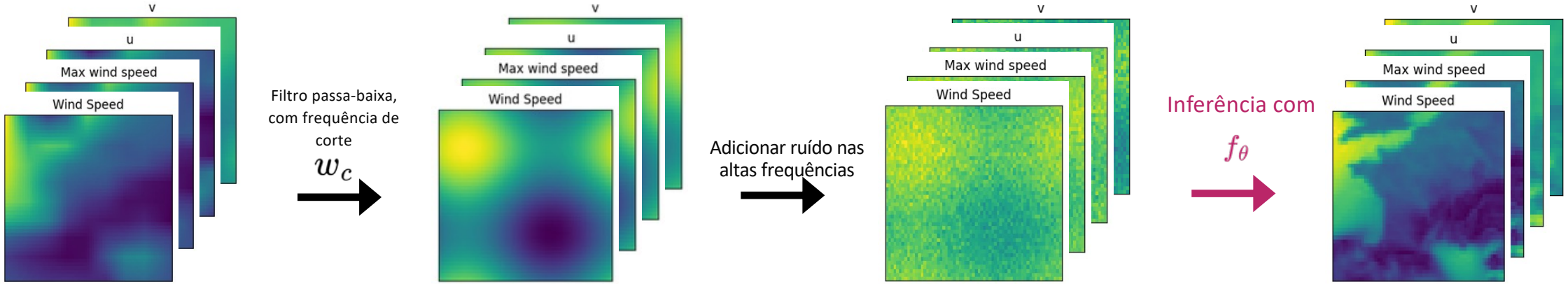
Referência (obs./reanálise)

Pipeline do SerpentFlow

Treinamento

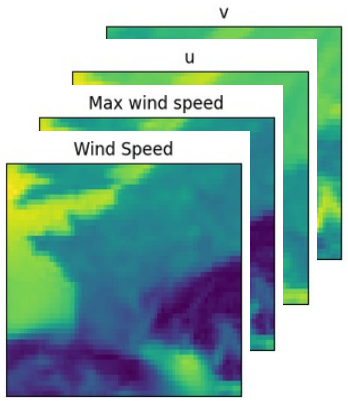


Inferência



Pipeline do SerpentFlow

Treinamento

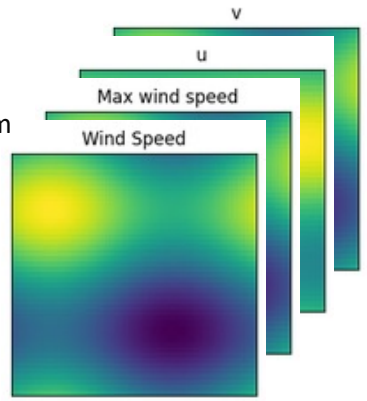


Referência (obs./reanálise)

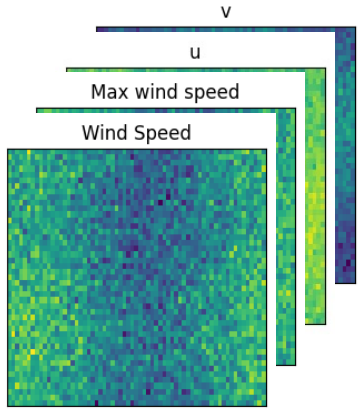
Filtro passa-baixa, com frequência de corte

w_c

Distribuição conjunta

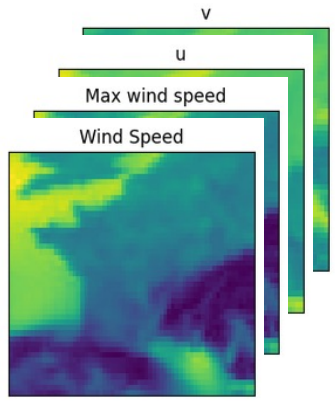


Adicionar ruído nas altas frequências



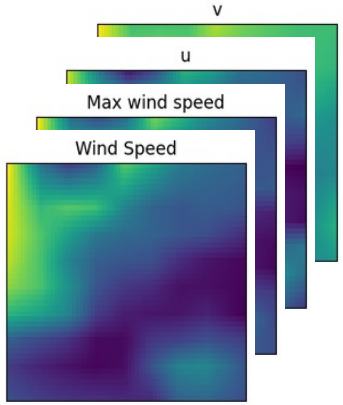
Treinar a abordagem generativa

f_θ



Referência (obs./reanálise)

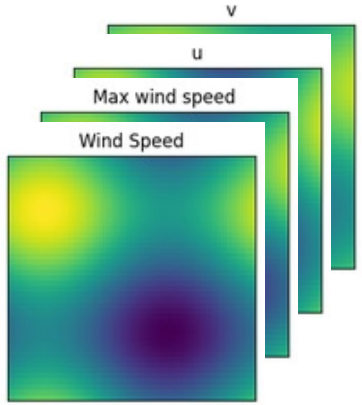
Inferência



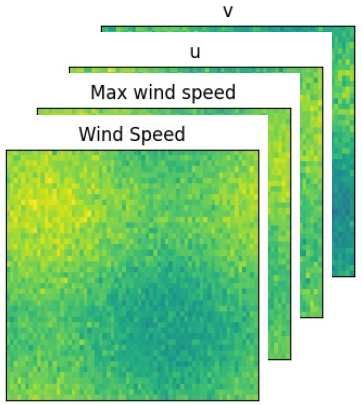
MCG interpolado

Filtro passa-baixa, com frequência de corte

w_c

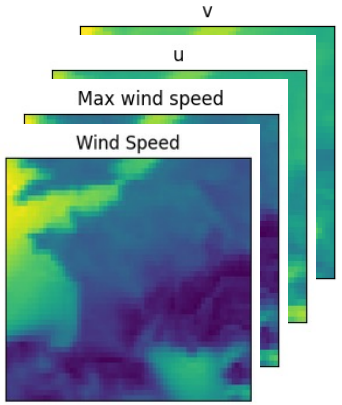


Adicionar ruído nas altas frequências



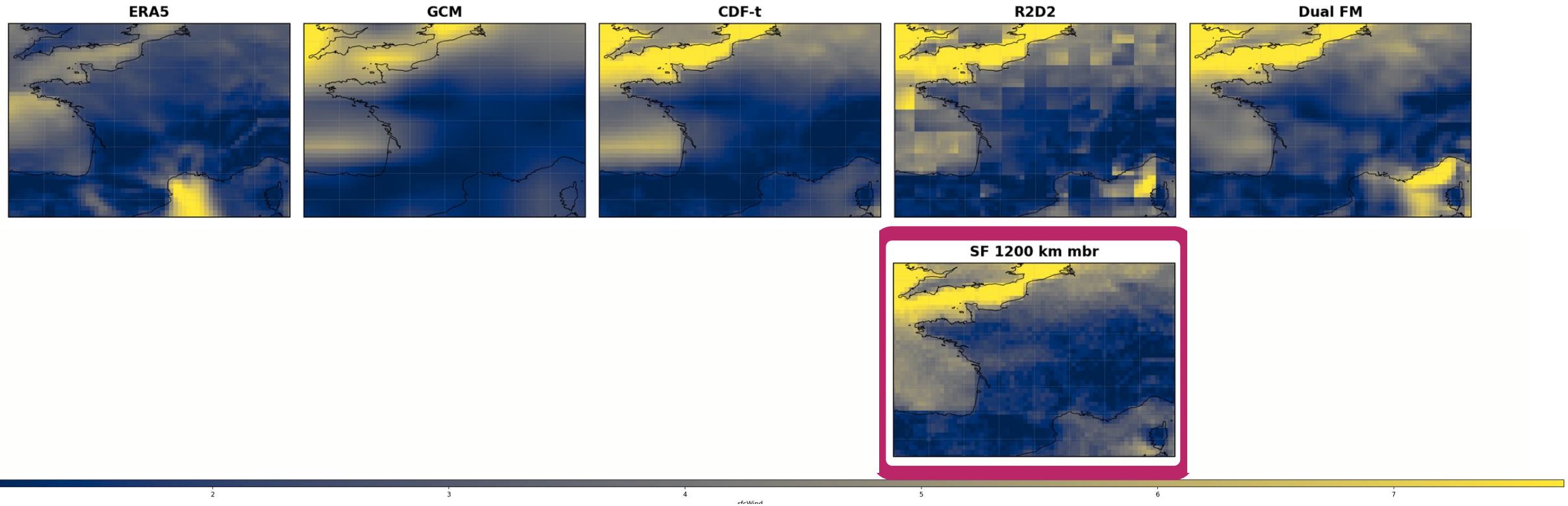
Inferência com

f_θ



MCG após downscaling

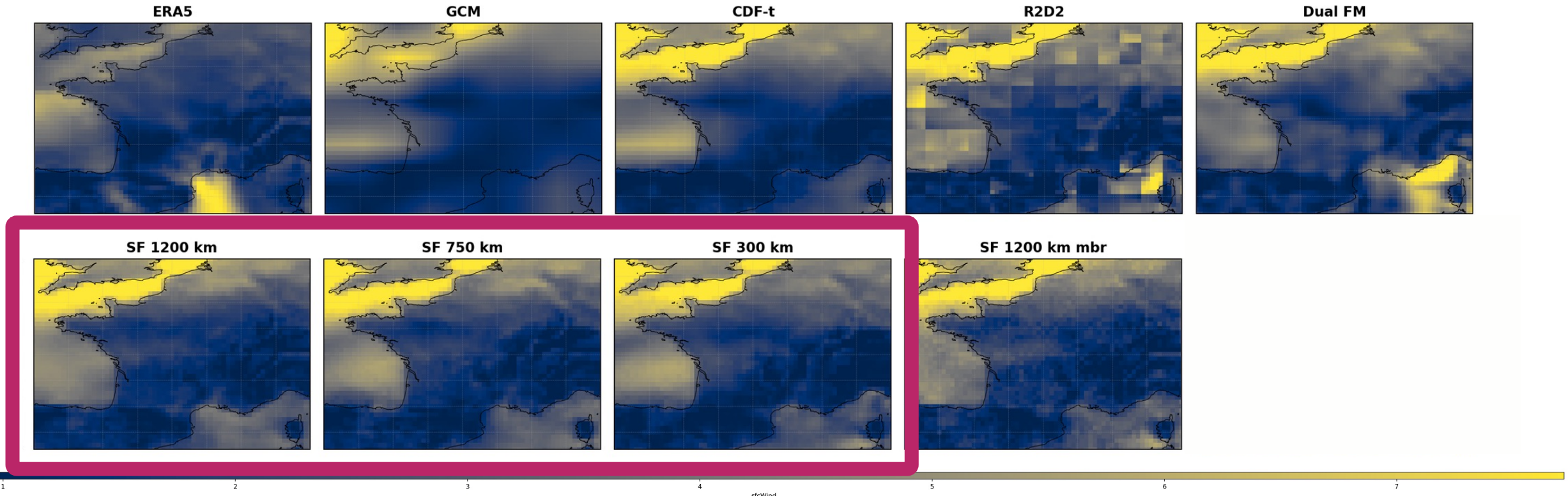
Aproximamos a distribuição do ERA5?



Principais achados

- Os métodos de AP superam CDF-t e R2D2 na estrutura espacial.
- **SF 1200 km** (1 membro): **melhor aproximação** da distribuição do ERA5

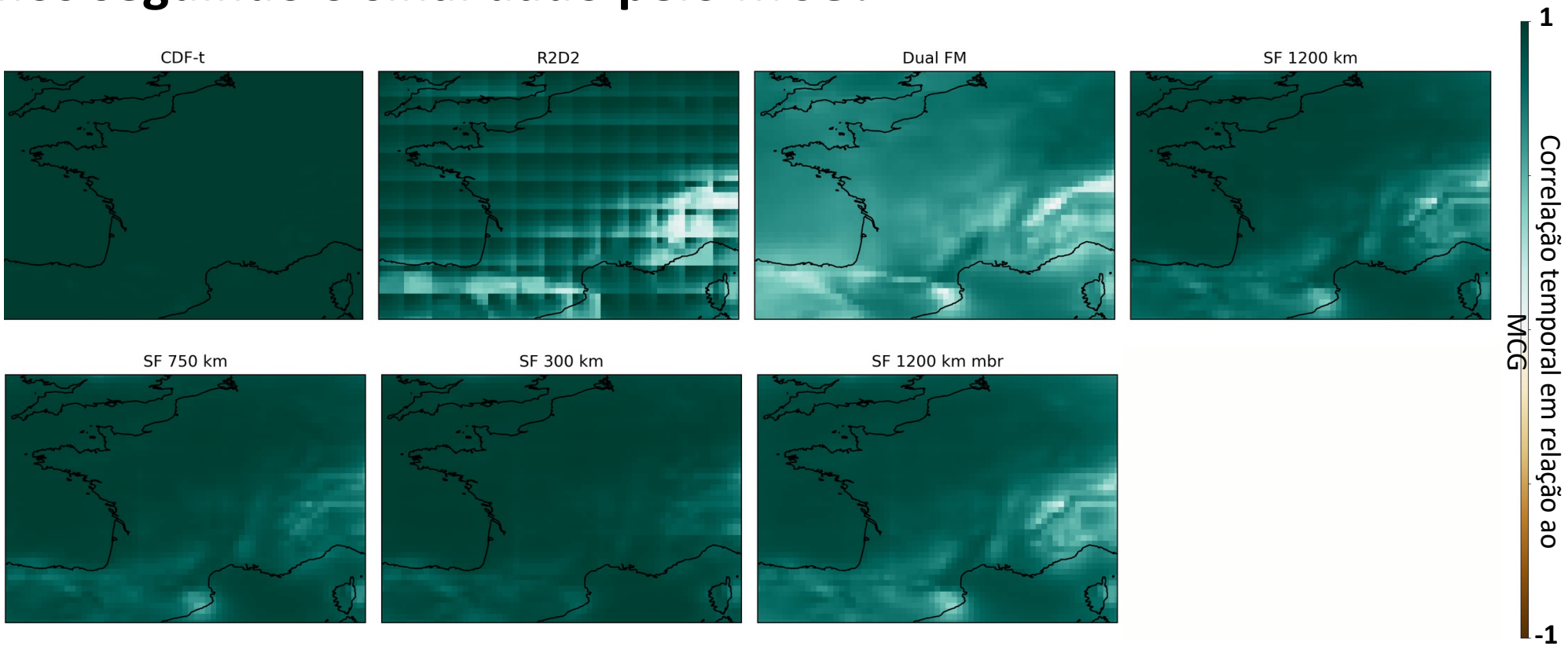
Aproximamos a distribuição do ERA5?



Principais achados

- Os métodos de AP superam CDF-t e R2D2 na estrutura espacial.
- **SF 1200 km** (1 membro): **melhor aproximação** da distribuição do ERA5
- **frequência de corte mais alta** ⟶ afastamento da distribuição do ERA5.

Estamos seguindo o sinal dado pelo MCG?



Principais achados

- **O CDF-t é o mais próximo da** dinâmica do MCG; o Dual FM não segue a trajetória do MCG
- O afastamento do SF aparece nas regiões **montanhosas**
- **Frequência de corte mais baixa** ⟶ mais perto da dinâmica do MCG

Perspectivas

- Funciona com diferentes **MCGs, resoluções e domínios irregulares** — ERA5 (25 km) e SAFRAN (8 km)
- Supera os métodos estatísticos ao modelar a **distribuição conjunta** dos dados de reanálise
- Em comparação com outras abordagens generativas, fica próximo da dinâmica do MCG
- **A frequência de corte é interpretável**
  - Liga diretamente a separação de escalas à resolução efetiva do MCG
  - Gerencia **o trade-off** entre seguir a trajetória do MCG e reproduzir a distribuição da reanálise
- **Limites:** só vento próximo à superfície — próximos passos: outras variáveis, extremos, downscaling temporal

IA generativa para:

- ✓ Previsão probabilística
  - **ArchesWeatherGen**: previsão do tempo eficiente treinada a 1,5°
  - **ArchesClimate**: emulação, interpolação e extrapolação do clima decenal

- ✓ Fusão de dados com alinhamento probabilístico de domínios
  - **MotifGen**: um modelo generativo, multifonte e autossupervisionado para aprender com dados geoespaciais

- ✓ Alinhamento de domínios probabilístico e downscaling
  - **ClimAlign**: downscaling generativo não pareado com alinhamento de domínios
  - **SerpentFlow**: alinhamento de domínios generativo não pareado por decomposição em estrutura compartilhada (com aplicação ao downscaling)

Primeiros passos com IA generativa baseada em fluxos

Flow matching

Tutoriais:

- 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

Leituras complementares sobre métodos baseados em fluxos

Interpolantes estocásticos

Arcabouço geral baseado em fluxos para transportar ruído em dados

- Flow matching e rectified flows são exemplos
- Generaliza o arcabouço: dá para treinar com EDO e amostrar com EDE, ou vice-versa

Operadores neurais

Métodos baseados em fluxos operam no espaço de operadores, permitindo generalização independente de escala

Modelos de consistência

Versão de passo único do flow matching. Área de pesquisa aberta, com alguma teoria e limitações práticas.

Aprendizado por cadeia de Markov reversa

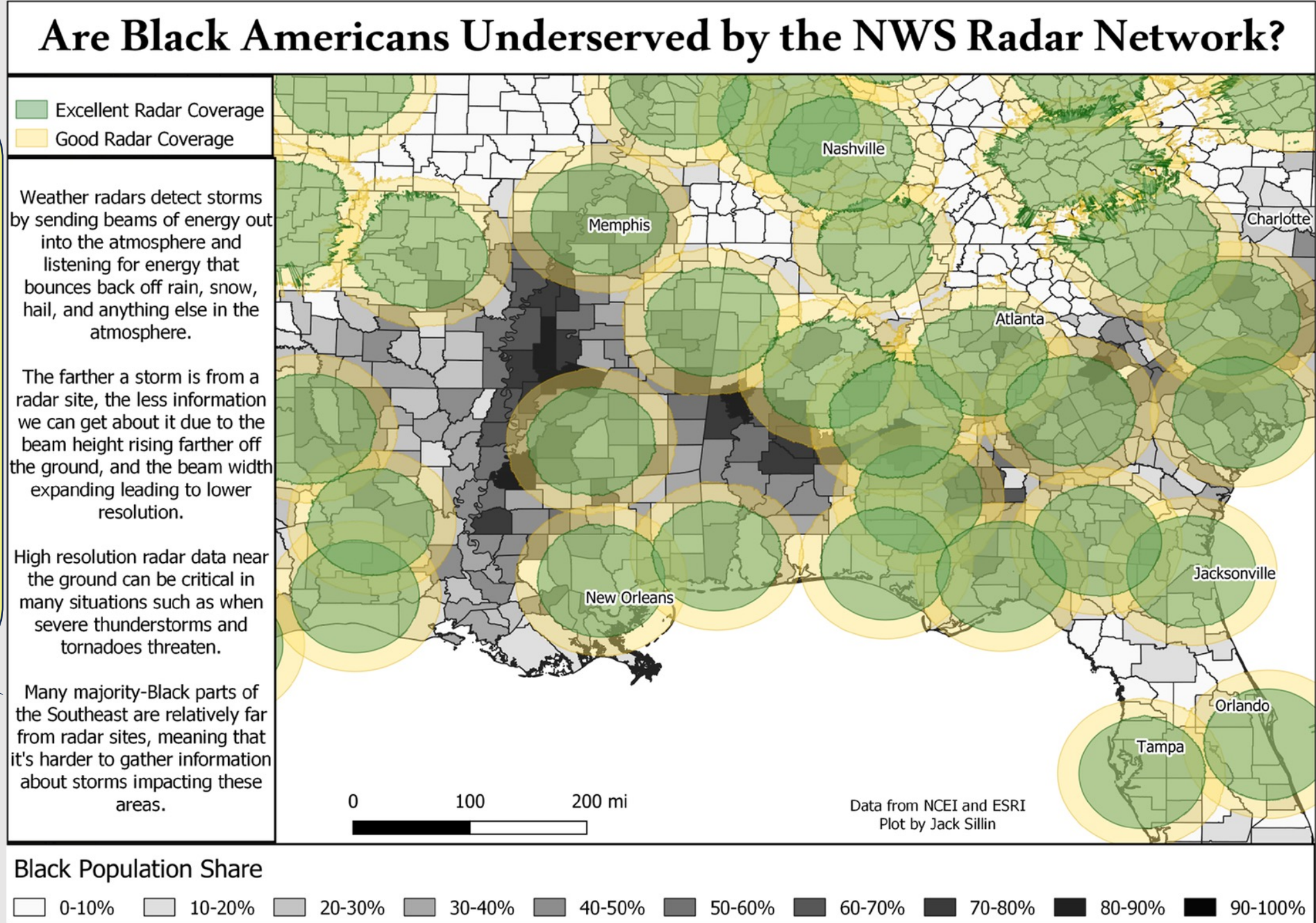
Basicamente, flow matching de baixa para alta resolução

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

Seção bônus: IA para a equidade dos dados climáticos

“Muitas áreas de maioria negra do Sudeste [dos EUA] ficam relativamente longe dos radares, o que dificulta coletar informações sobre as tempestades que as atingem.”

Crédito: Jack Sillin, em [McGovern et al., Environmental Data Science, 2022]



IA para a equidade dos dados climáticos

- Treinar modelos em regiões com muitos dados e aplicá-los em regiões com poucos dados
  - Treiná-los e validá-los em regiões com muitos dados
  - Fazer o ajuste fino com os dados limitados das regiões com poucos dados e usá-los para gerar mais dados.

- Contribuição para a equidade dos dados climáticos
  - Escalas locais (p. ex. herança de injustiça ambiental nos EUA)
  - Escalas globais:
    - O Norte Global historicamente emitiu mais carbono; além disso, costuma haver mais dados lá
    - O Sul Global sofre os efeitos mais graves do aquecimento resultante