

Investigación en IA
frente al cambio climático y por la sostenibilidad ambiental

Claire Monteleoni

INRIA Paris

University of Colorado Boulder

Choose France™



Diciembre de 2021: condado de Boulder, Colorado

- La sequía de nieve en otoño e invierno de 2021 resecó la cobertura del suelo
- Vientos de 130-160 km/h, más una ignición, desataron una «tormenta de fuego» incontrolable
- 2 víctimas mortales. 1000 viviendas y 20 negocios destruidos, y más dañados





«La oportunidad que representa la IA para la Tierra es considerable. La explosión actual de la IA hará que la incorporemos a cada vez más cosas cada año.... Al pensar en los beneficios, las eficiencias y las nuevas soluciones que esto crea para las naciones, las empresas y la vida cotidiana, también debemos pensar en cómo maximizar los beneficios para la sociedad y para nuestro medio ambiente en general».

– Foro Económico Mundial (2018), «Harnessing Artificial Intelligence for the Earth»





Investigación en IA frente al cambio climático y por la sostenibilidad ambiental

CAMBIO CLIMÁTICO

ADAPTACIÓN

MITIGACIÓN

IMPACTOS





Tiempo extremo

Amenazas en cascada



Transición energética

Cambio de uso del suelo



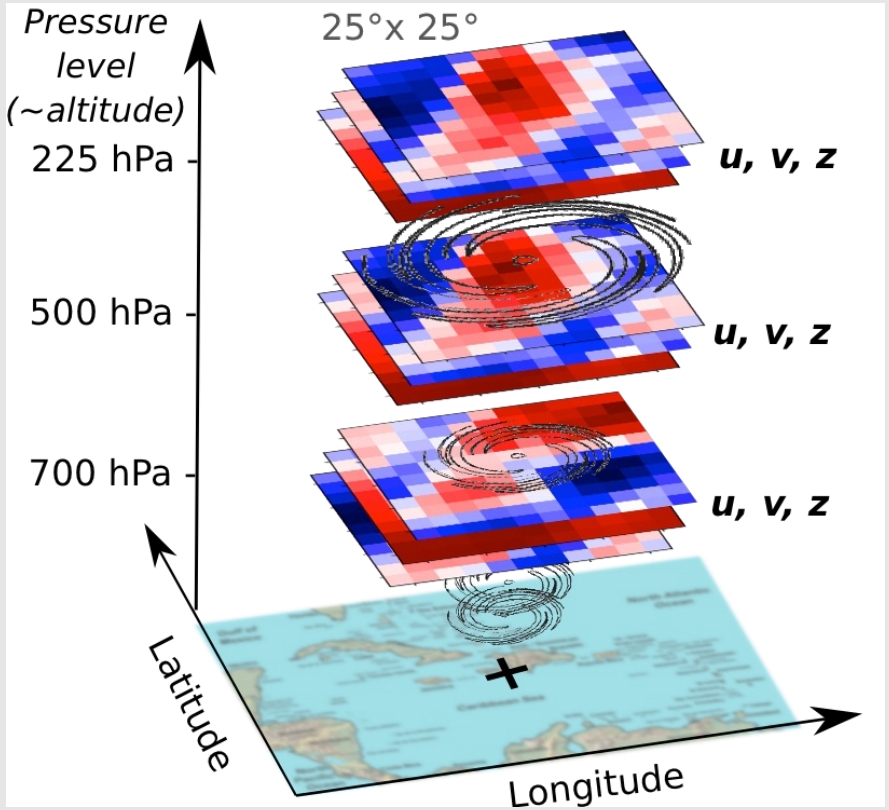
Emisiones de carbono

Subida del nivel del mar

Soluciones basadas en IA

ADAPTACIÓN

IA para el tiempo extremo y las amenazas en cascada



[Giffard-Roisin et al., Frontiers 2020]

Predicción de trayectorias de huracanes

Previsión de precipitaciones extremas del monzón indio de verano

Detección de avalanchas

IA generativa para la predicción meteorológica



MITIGACIÓN

Reducir las emisiones de carbono

Acelerar la transición a la energía verde

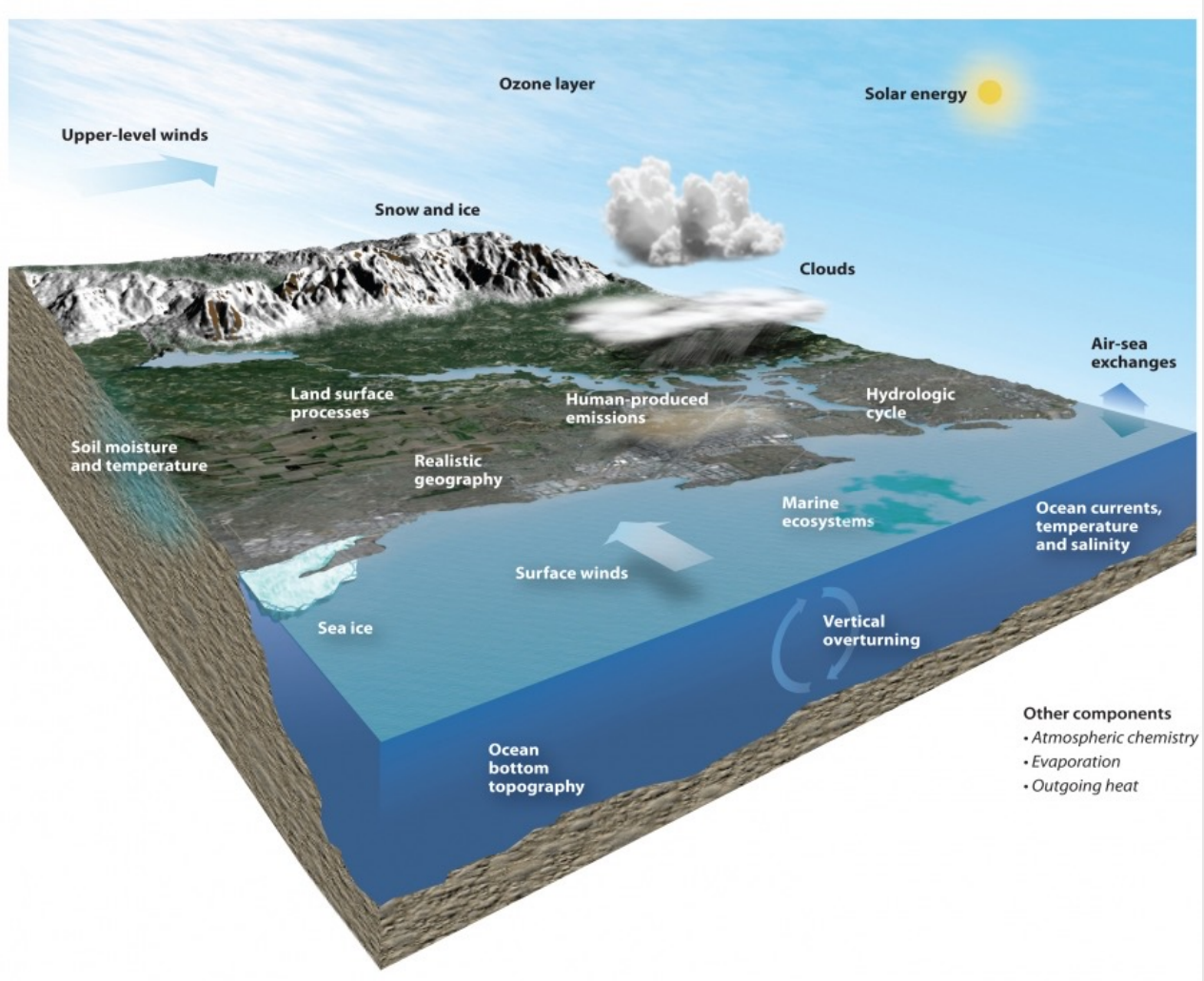
- Previsión solar y eólica con IA
- IA para la reducción de escala de datos solares y eólicos

Reducir el cómputo de la modelización meteorológica y climática

- Una vez entrenada, la IA predice mucho

IMPACTOS

IA para entender y prever el cambio climático



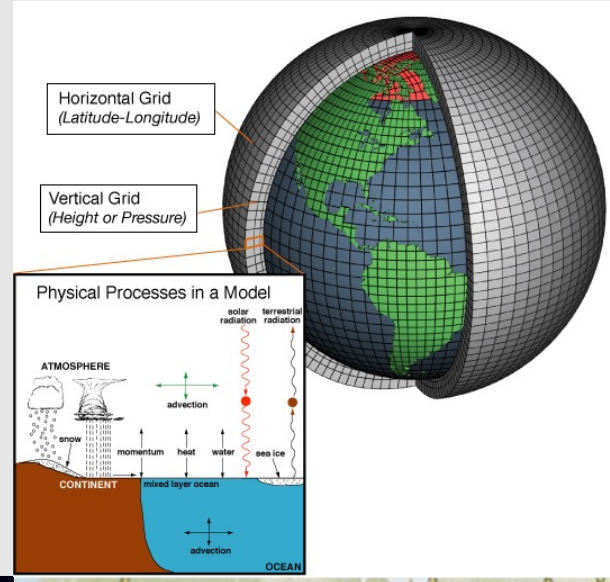
UCAR Science Education

Usar la IA para aprender las relaciones entre simulaciones del IPCC y observaciones

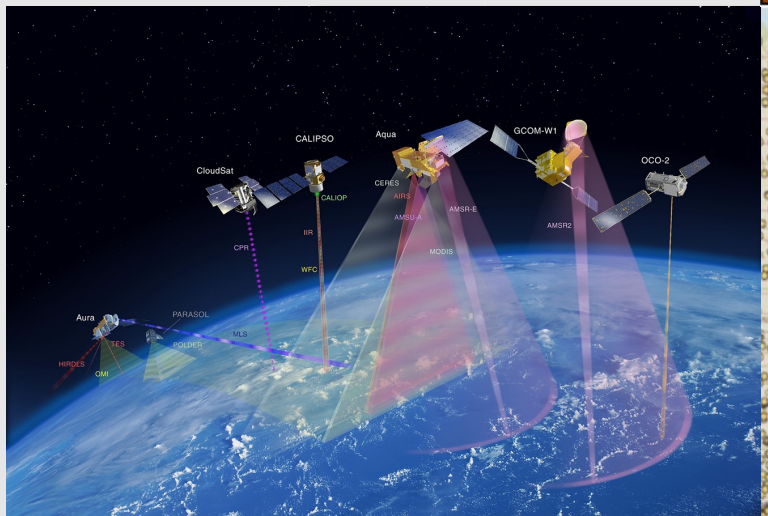
- Hacer más robustas las previsiones de conjuntos de modelos climáticos
- Proyectar la subida del nivel del mar a largo plazo
- Proyectar las emisiones de carbono a largo plazo

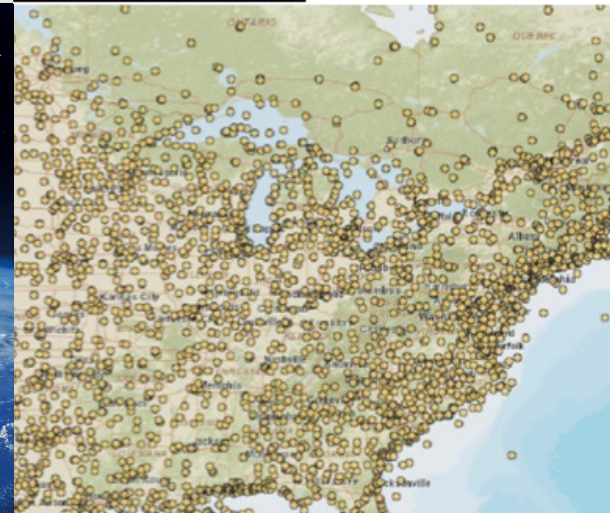
Enfoque: aprovechar todos los datos disponibles

- Datos simulados generados por modelos físicos
  - Modelos de predicción numérica del tiempo (PNT)
  - Modelos de circulación general (MCG)
  - Modelos climáticos regionales (MCR)



- Datos de reanálisis
  - Productos en malla a partir de la asimilación de datos: aplica leyes físicas a las observaciones





- Datos de observación
  - Datos de teledetección satelital
  - Datos in situ

Tareas de AA subyacentes a muchos de estos problemas

- ❑ Fusión de datos
- ❑ Previsión
- ❑ Aprendizaje no supervisado
- ❑ Alineación probabilística de dominios
- ❑ Reducción de escala e interpolación

Métodos de IA

- Aprendizaje no supervisado
- Aprendizaje profundo generativo

Esquema para el resto del día

I. **Contexto:** la revolución de la IA en la predicción meteorológica

Implicaciones para el estudio del cambio climático

II. **Métodos de IA:** ¿qué es el aprendizaje autosupervisado?

Aplicaciones en clima y tiempo

III. **Métodos de IA:** ¿qué es el aprendizaje profundo generativo?

Aplicaciones en clima y tiempo

Parte I: Contexto:
la revolución de la IA en la predicción meteorológica

La revolución silenciosa de la predicción meteorológica

- La predicción numérica del tiempo (PNT) ha progresado de forma constante en las últimas décadas
- Un gran logro científico
  - Un problema de simulación complejo resuelto **de forma rutinaria** en centros de predicción operativos de todo el mundo
- Gran valor socioeconómico
  - Mitigación de amenazas climáticas, logística, prevención de pérdidas materiales, …

Crédito de las primeras diapositivas: David Landry (INRIA Google DeepMind)

doi:10.1038/nature14956

REVIEW

doi:10.1038/nature14956

The quiet revolution of numerical weather prediction

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

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

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

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

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

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

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

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

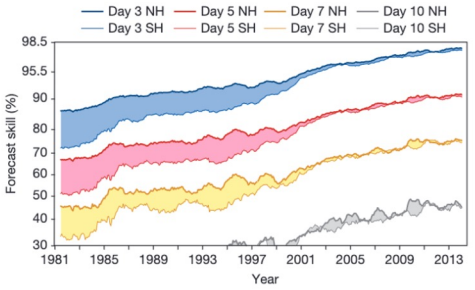
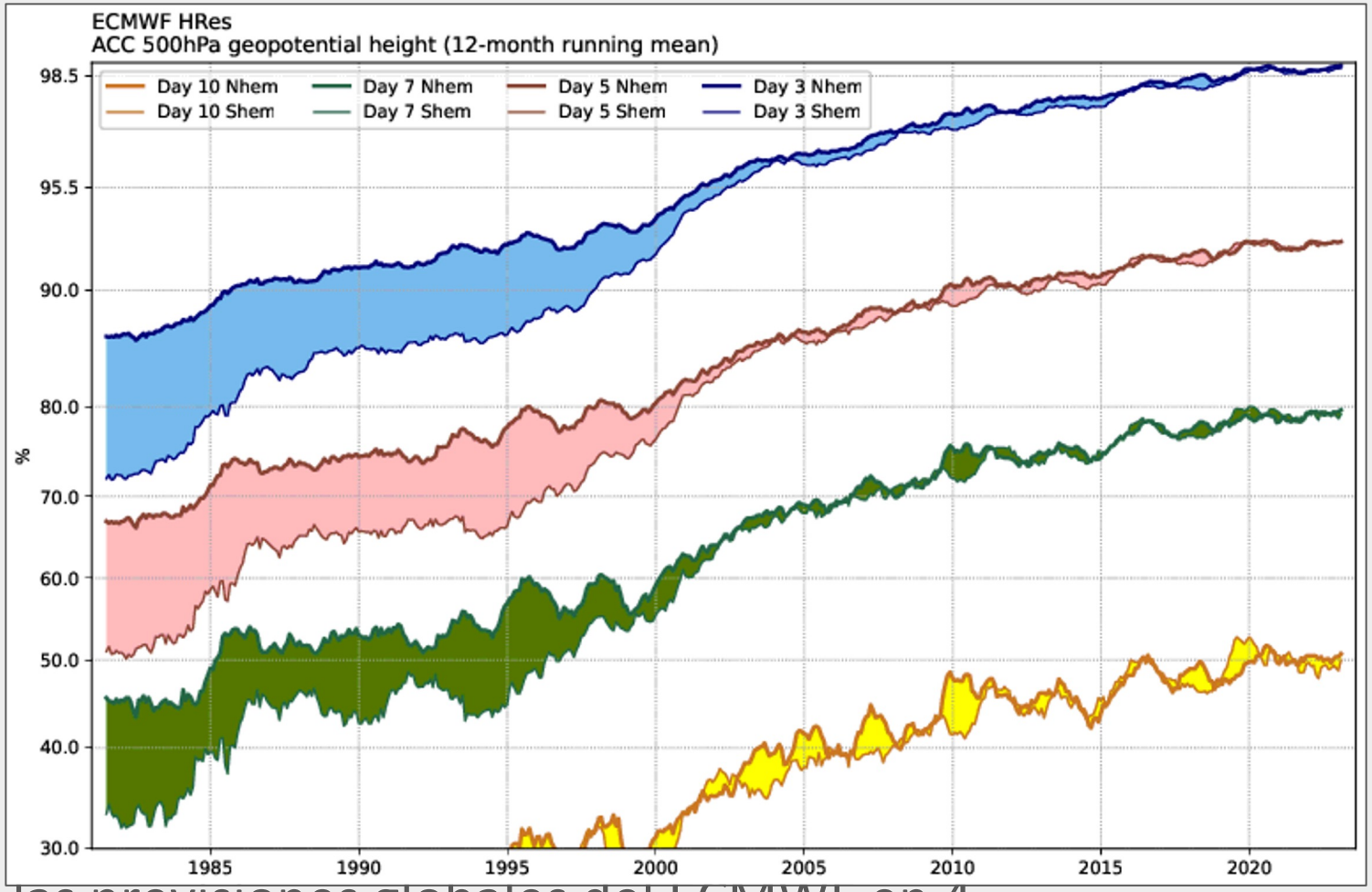


Figure 1 | A measure of forecast skill at three-, five-, seven- and ten-day ranges, computed over the extra-tropical northern and southern hemispheres. Forecast skill is the correlation between the forecasts and the verifying analysis of the height of the 500-hPa level, expressed as the anomaly with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

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

©2015 Macmillan Publishers Limited. All rights reserved

La revolución silenciosa de la predicción meteorológica

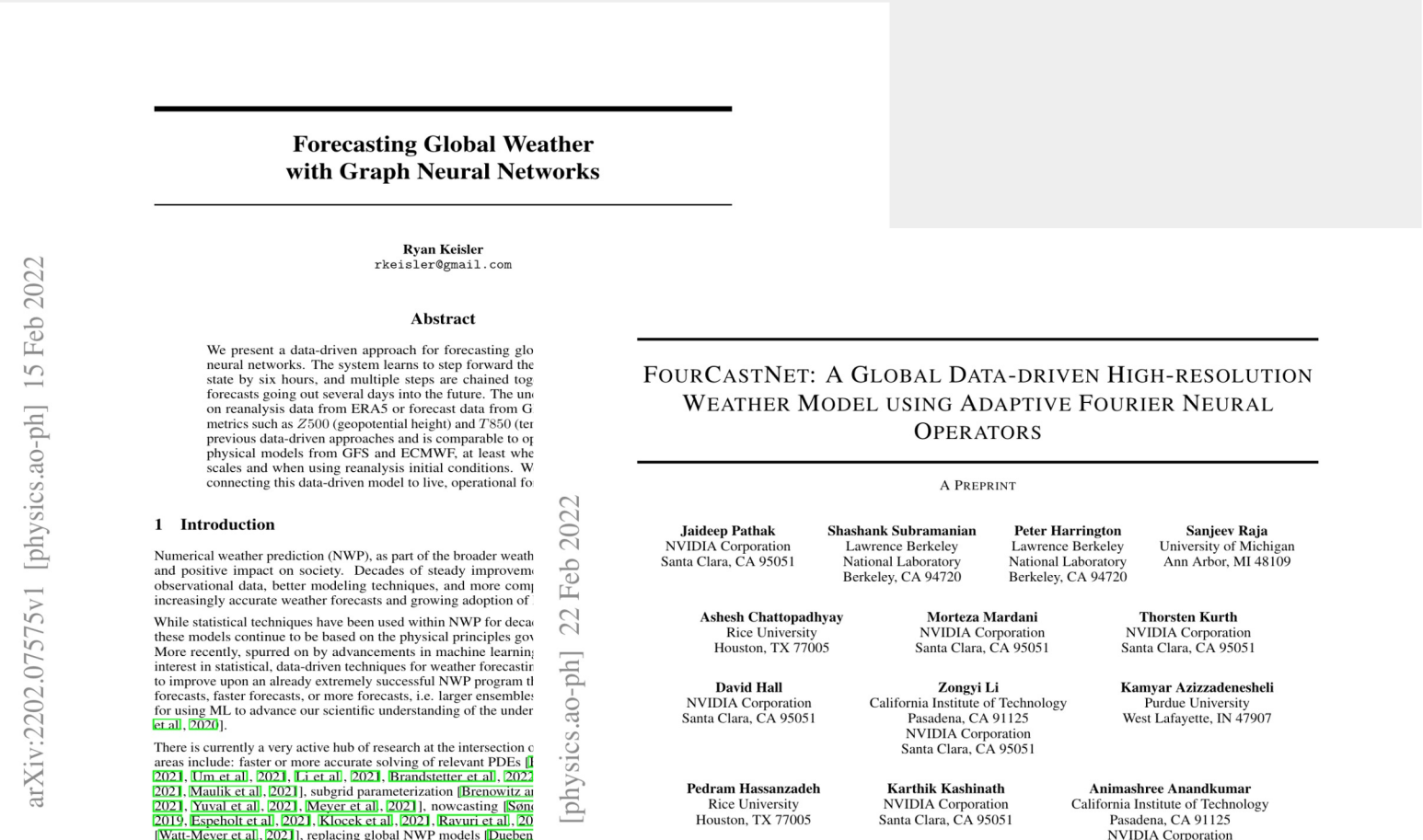


Rendimiento de las previsiones globales del ECMWF en 4 décadas

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

Primeros modelos IA-PNT

- Desde 2022, varios equipos de investigación han usado redes neuronales profundas para la predicción meteorológica a medio plazo
- La calidad de algunos de estos modelos llegó a ser ~equivalente a la de los mejores modelos de PNT



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

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 com 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

[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

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

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

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

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

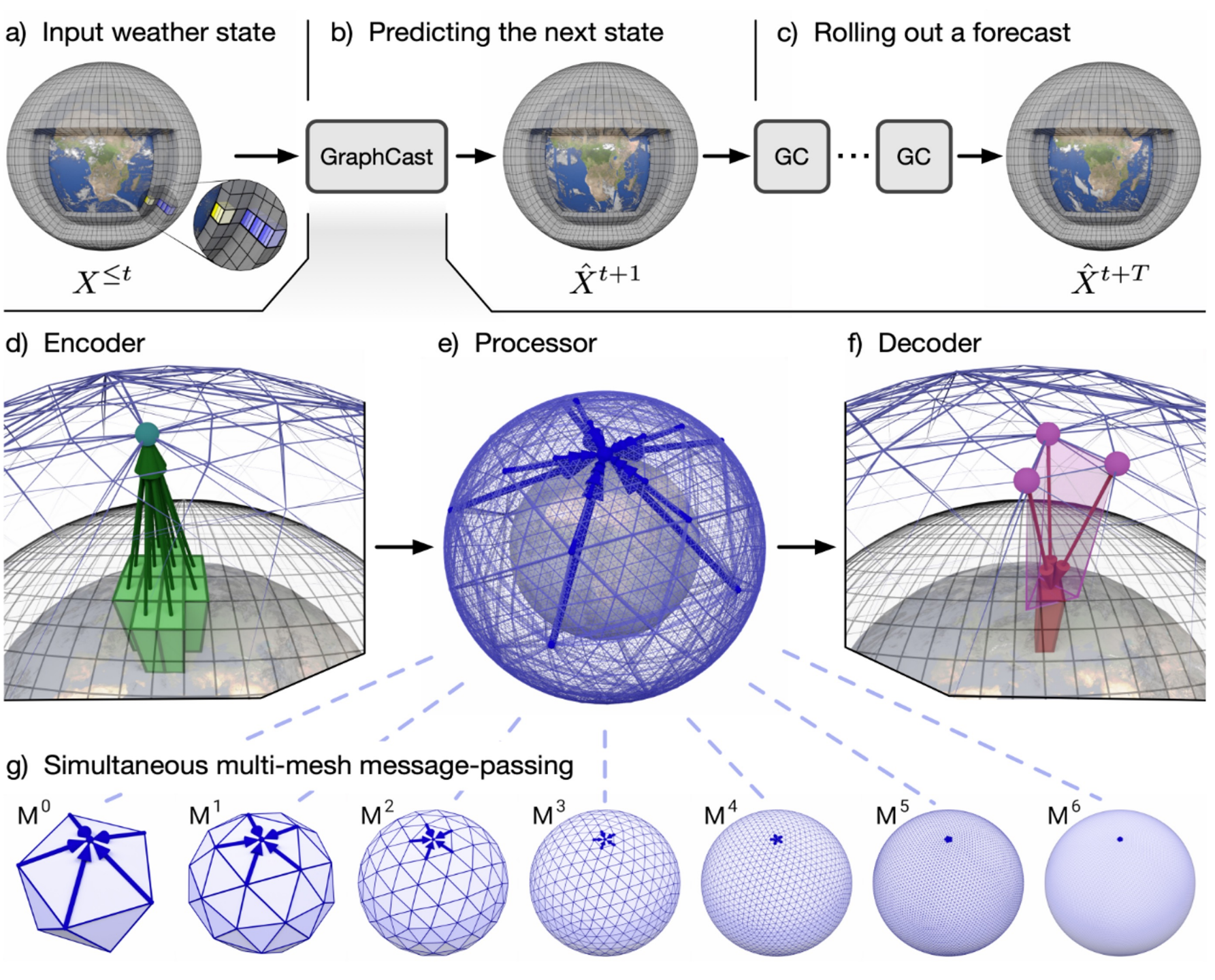
Introduction

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

Ahorro computacional de estos primeros modelos IA-PNT

Modelo	**Entrenamiento**	**Inferencia**
Pangu-Weather	16 días en 192 GPU V100	Previsión a 24 horas 1xV100 1,4 segundos en
GraphCast	21 días en 32 Cloud TPU V4 (~128 GPU)	Previsión a 10 días 1 Cloud TPU V4 < 60 segundos
PNT (ECMWF HRES) ***Mayor resolución**	-	Previsión a 10 días 1/3/6 horas Clúster HPC de 11 664 núcleos

Arquitectura de GraphCast



doi:10.48550/arXiv.2212.12794

IA para la predicción meteorológica: un nuevo paradigma

Asimilación de datos → Conjunto de datos ERA5

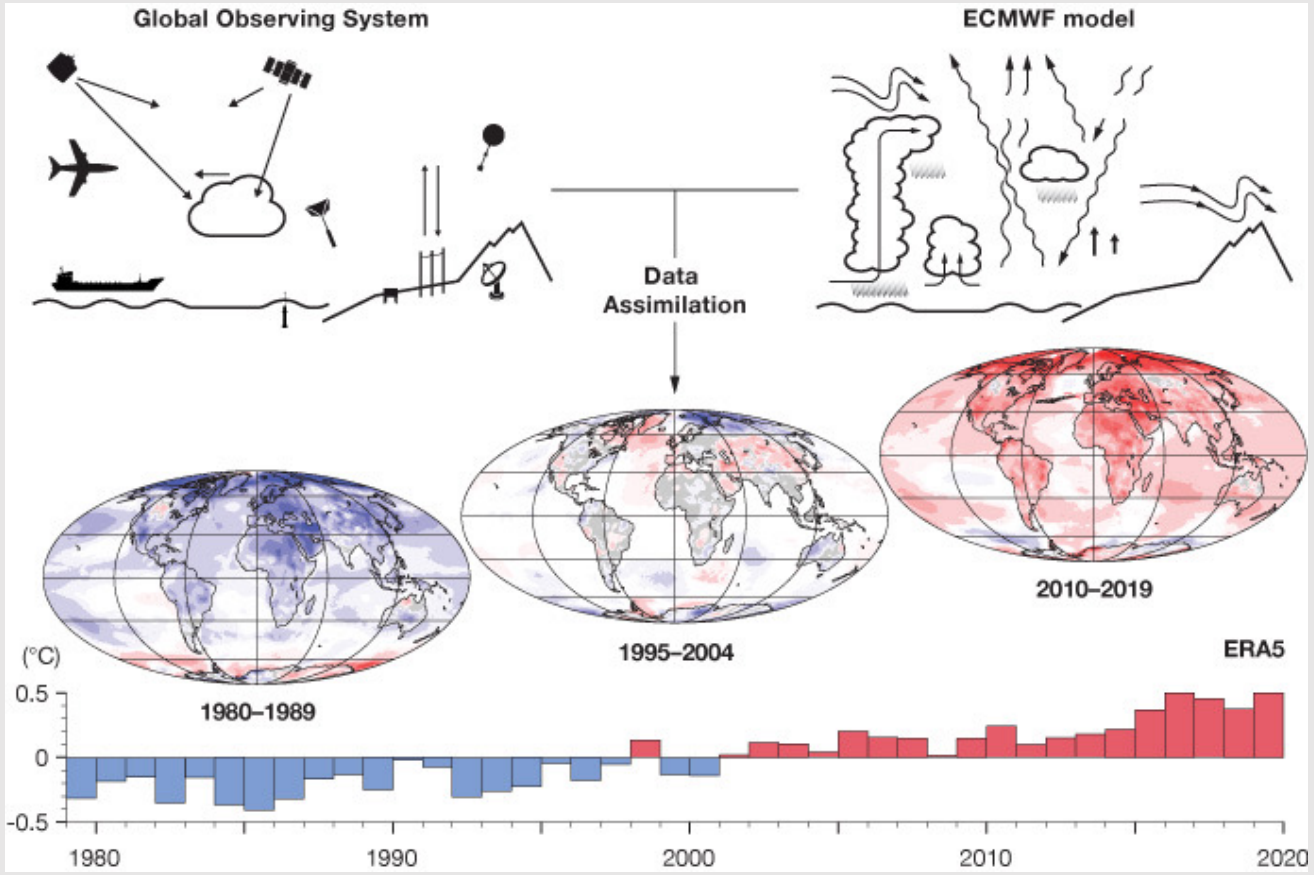
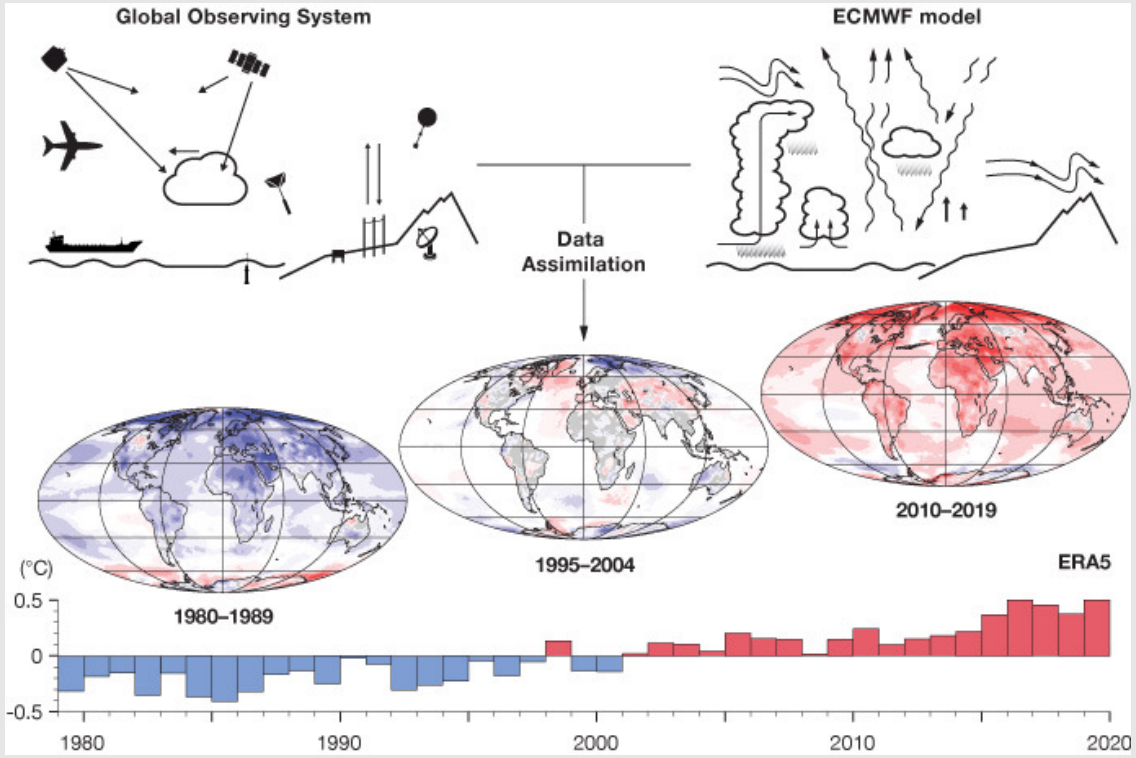


Figura: Centro Europeo de Previsiones Meteorológicas a Medio Plazo (ECMWF)

Conjunto de datos ERA5 → Simulación con red neuronal → $P(\text{Previsión})$

La revolución de la IA en la predicción meteorológica

Desde 2022, diversos modelos de aprendizaje profundo han mostrado un rendimiento de previsión comparable o **MEJOR** que la predicción numérica del tiempo (PNT), el estado del arte anterior.



Datos de entrenamiento: ERA5, un conjunto producido por asimilación de datos

Tarea de entrenamiento: autorregresión: pre a 6-24 horas

- Desplegado para prever a 7-10 días

[Sitio web del Centro Europeo de Previsiones Meteorológicas a Medio Plazo (ECMWF)]

Despliegues

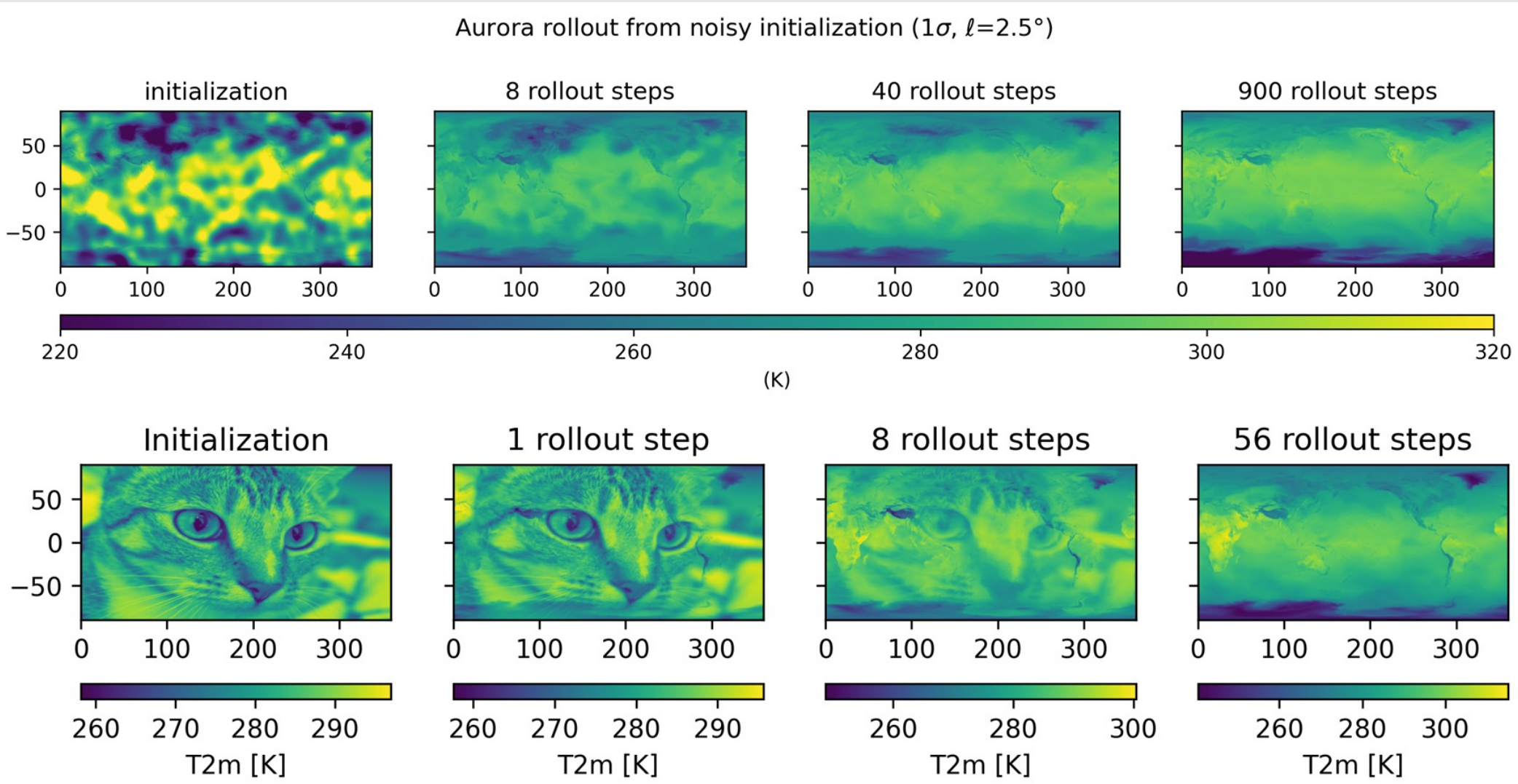


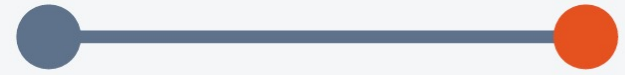
Figure A.9: 2-meter temperature predicted at different time stamps when (top) inputs are perturbed with an additive Gaussian random field with coefficient of variation σ^2 (where σ is the standard deviation of the physical variable) and correlation length of 10 pixels (equivalent to $2.5°$), (bottom)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

Time

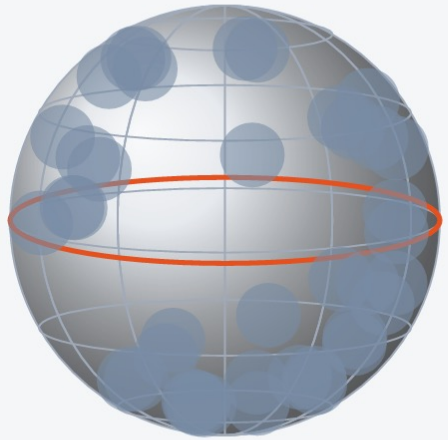
45+ years



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**

Reanálisis ERA5 del ECMWF

≈ 13 Petabytes

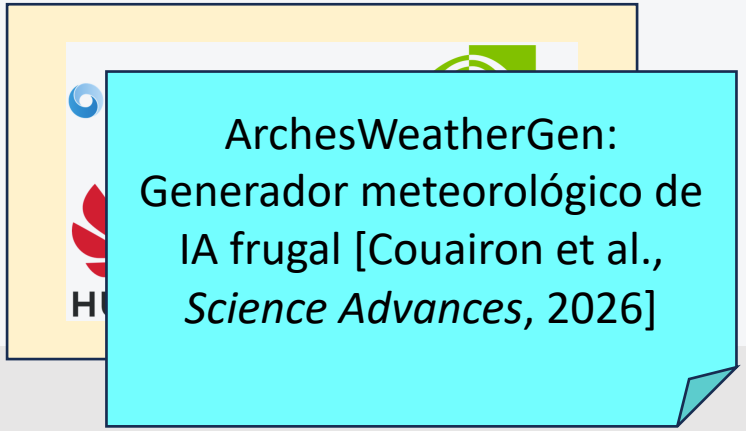


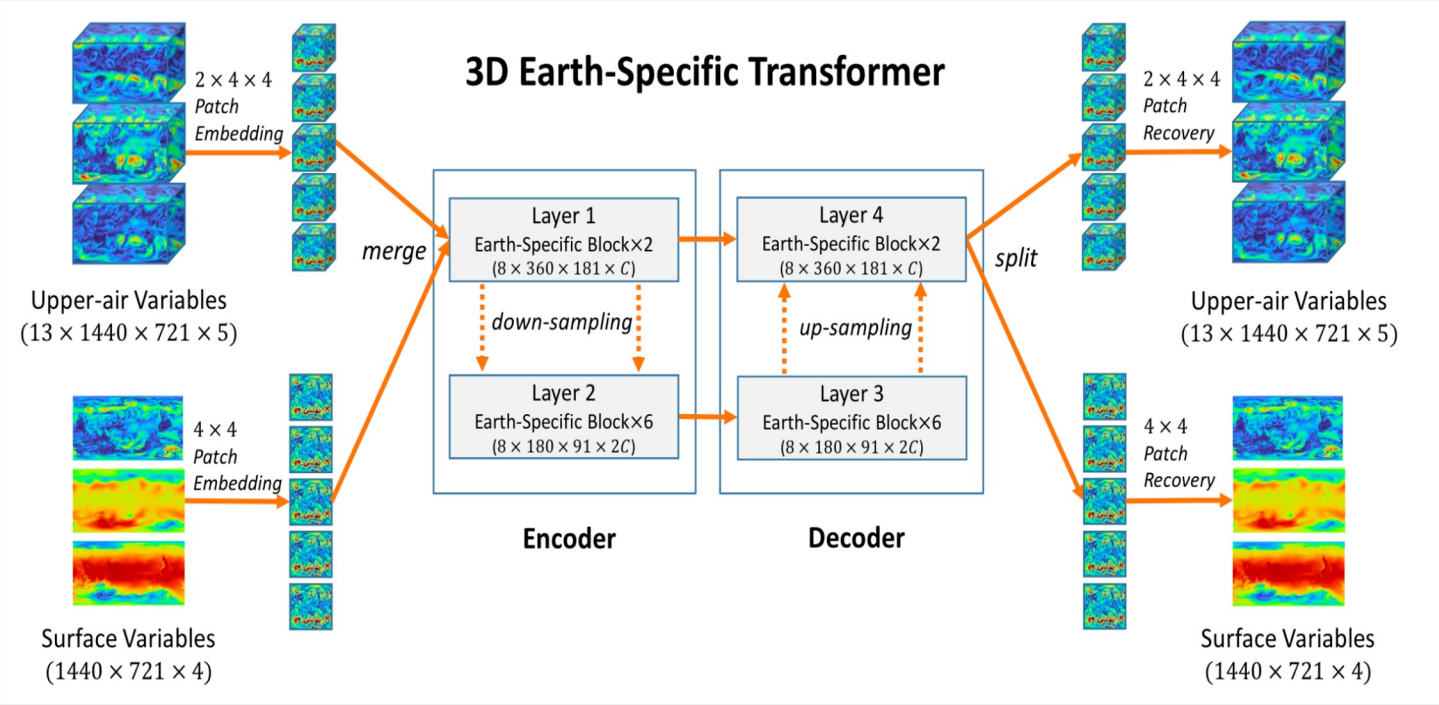
Figura: Emmanuel de Bézenac (INRIA)



ArchesWeather

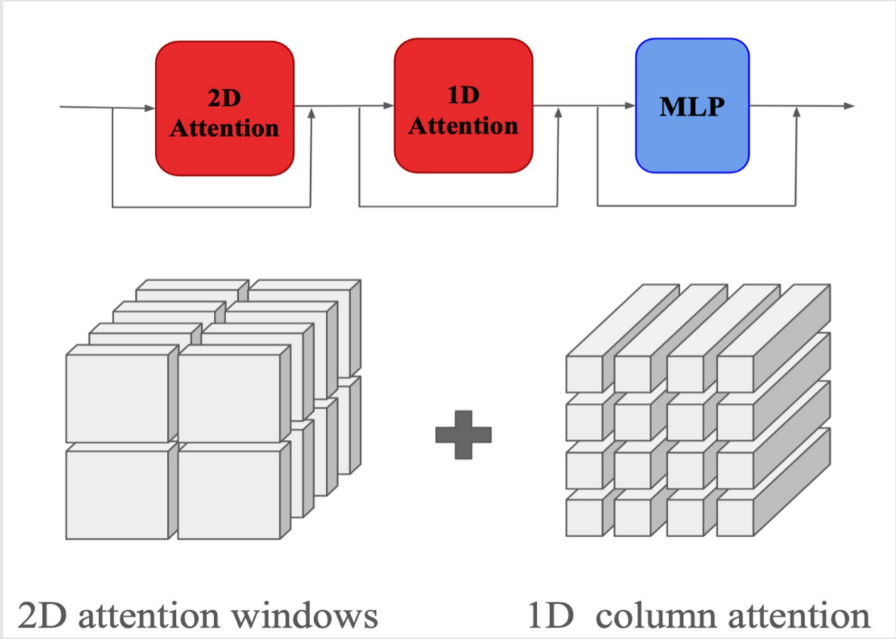
[Couairon et al., ArchesWeather: an efficient AI weather model at 1.5º resolution ICML 2024 workshop on Earth System Modeling]

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



Novedades de ArchesWeather:

- Entrenamiento con datos de resolución más gruesa
- Sustitución de la atención 3D por:

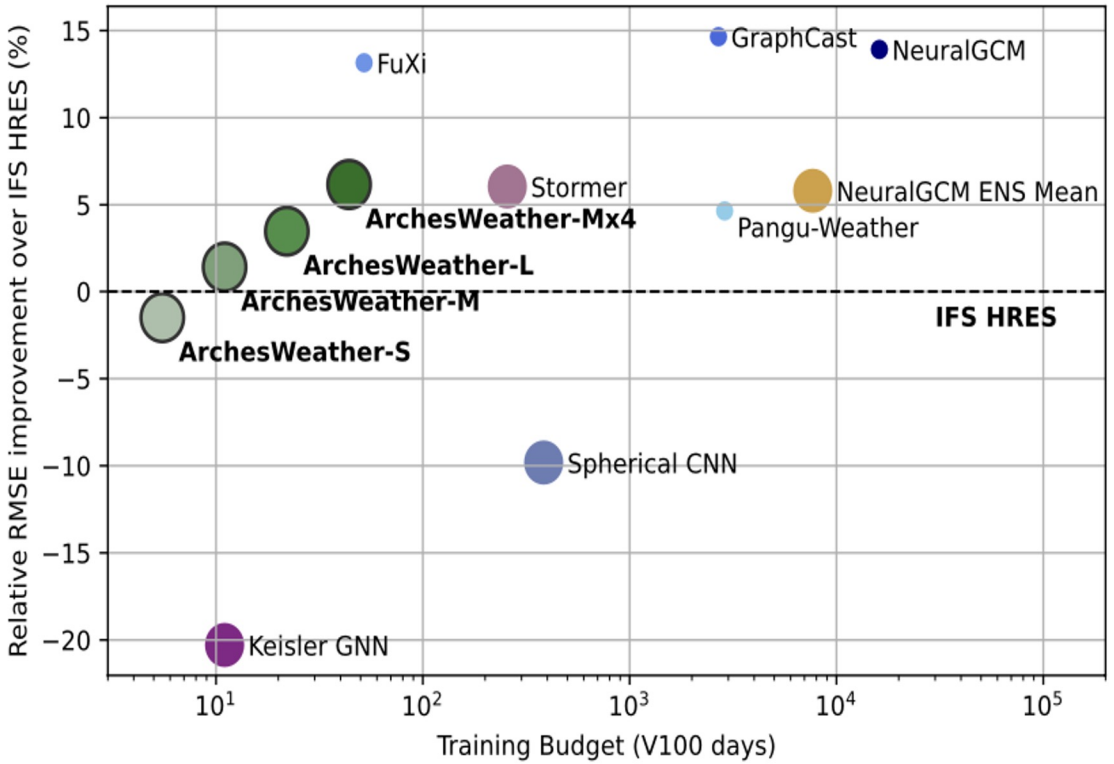


latitud + longitud

altitud

ArchesWeather

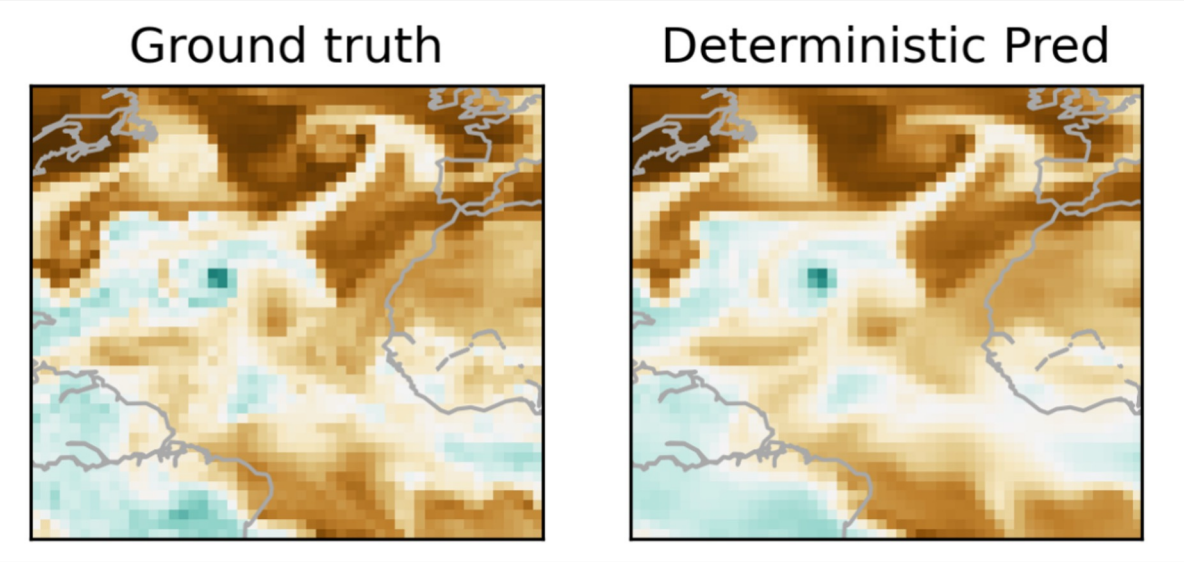
Compromiso de rendimiento



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

RMSE, plazo de 24 horas

Por qué pasar a la generación de conjuntos…



Previsión de Q700 a 24 h
fecha de inicialización: 28 sept. 2019

- La predicción determinista muestra un suavizado irreal
- Probar la generación de conjuntos mediante entrenamiento por difusión
- Objetivo: cada miembro muestreado debe ser más físico

ArchesWeatherGen

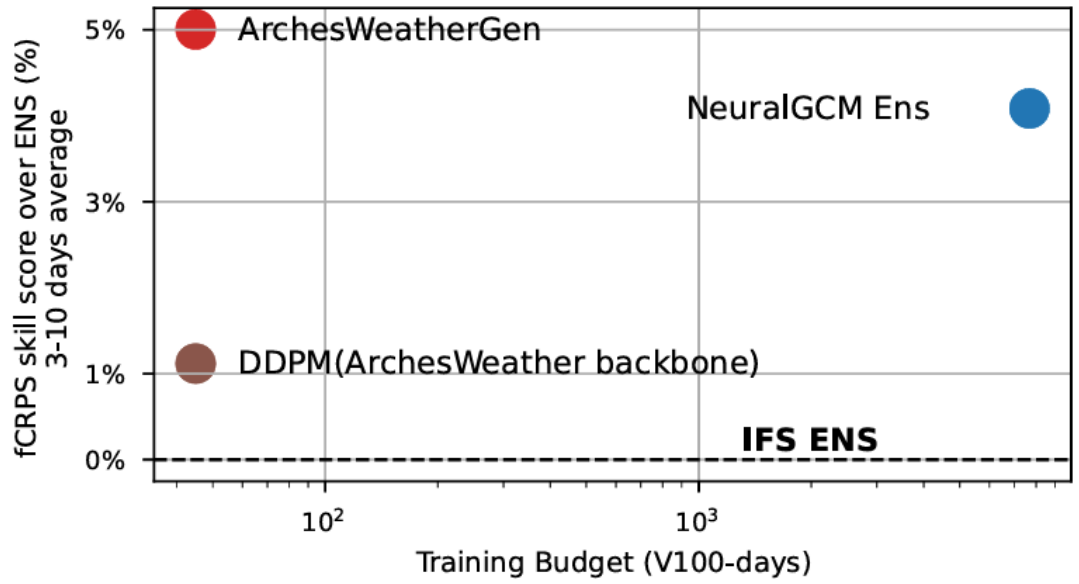
[Couairon et al., *Science Advances*, 2026]

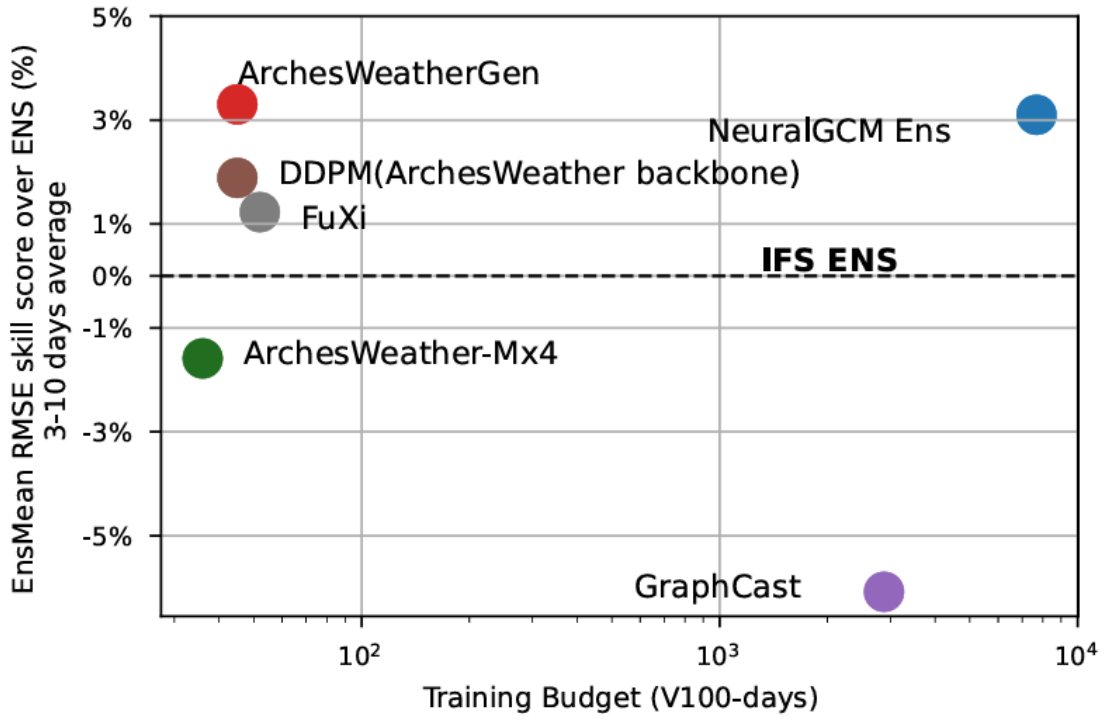
Red troncal: ArchesWeather (variante del Swin U-Transformer), 44 M de parámetros

Entrenado con Flow Matching

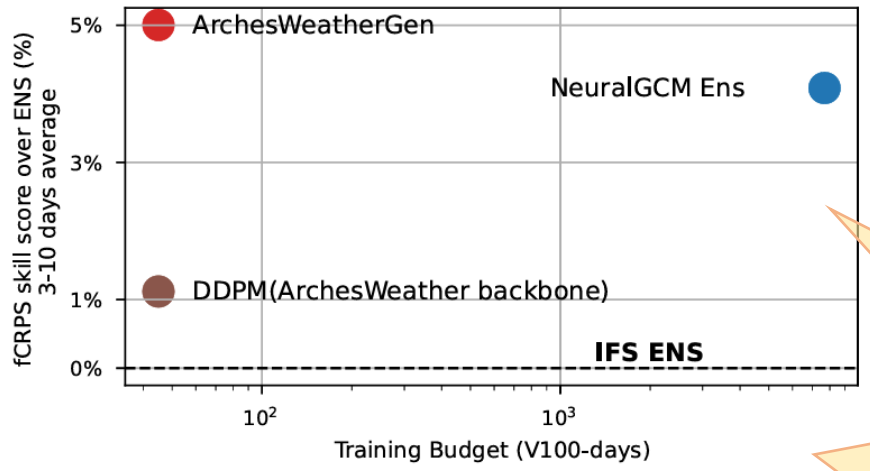
Inferencia con 25 pasos de muestreo cada 24 horas

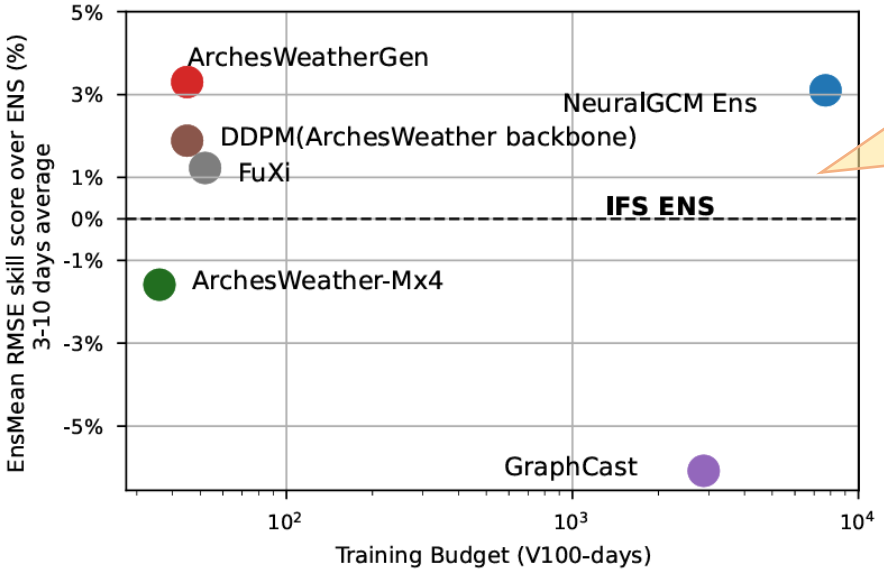
Despliegues con una muestra cada 24 horas. Esta salida se reinyecta en el modelo entrenado y se repite.

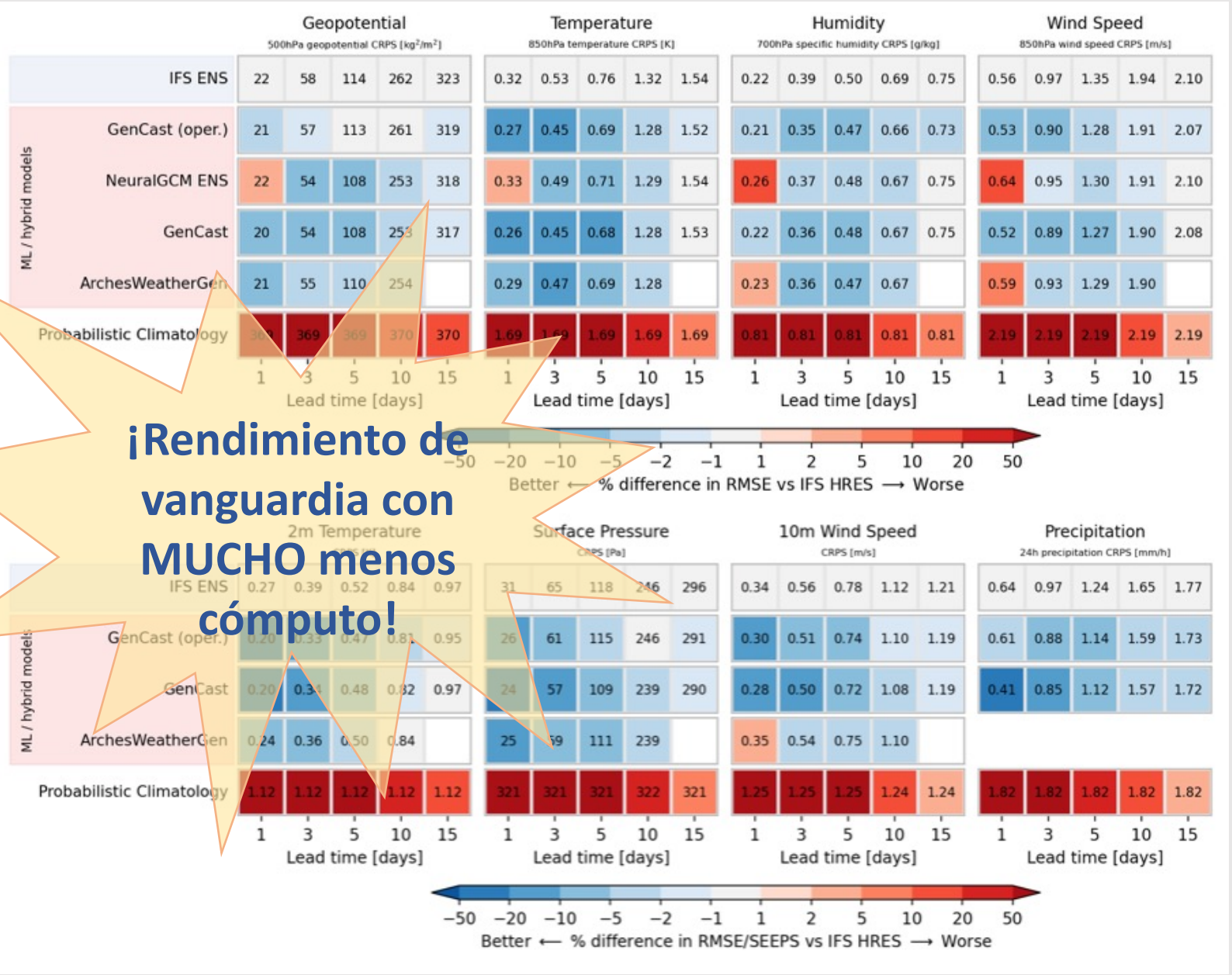




ArchesWeatherGen





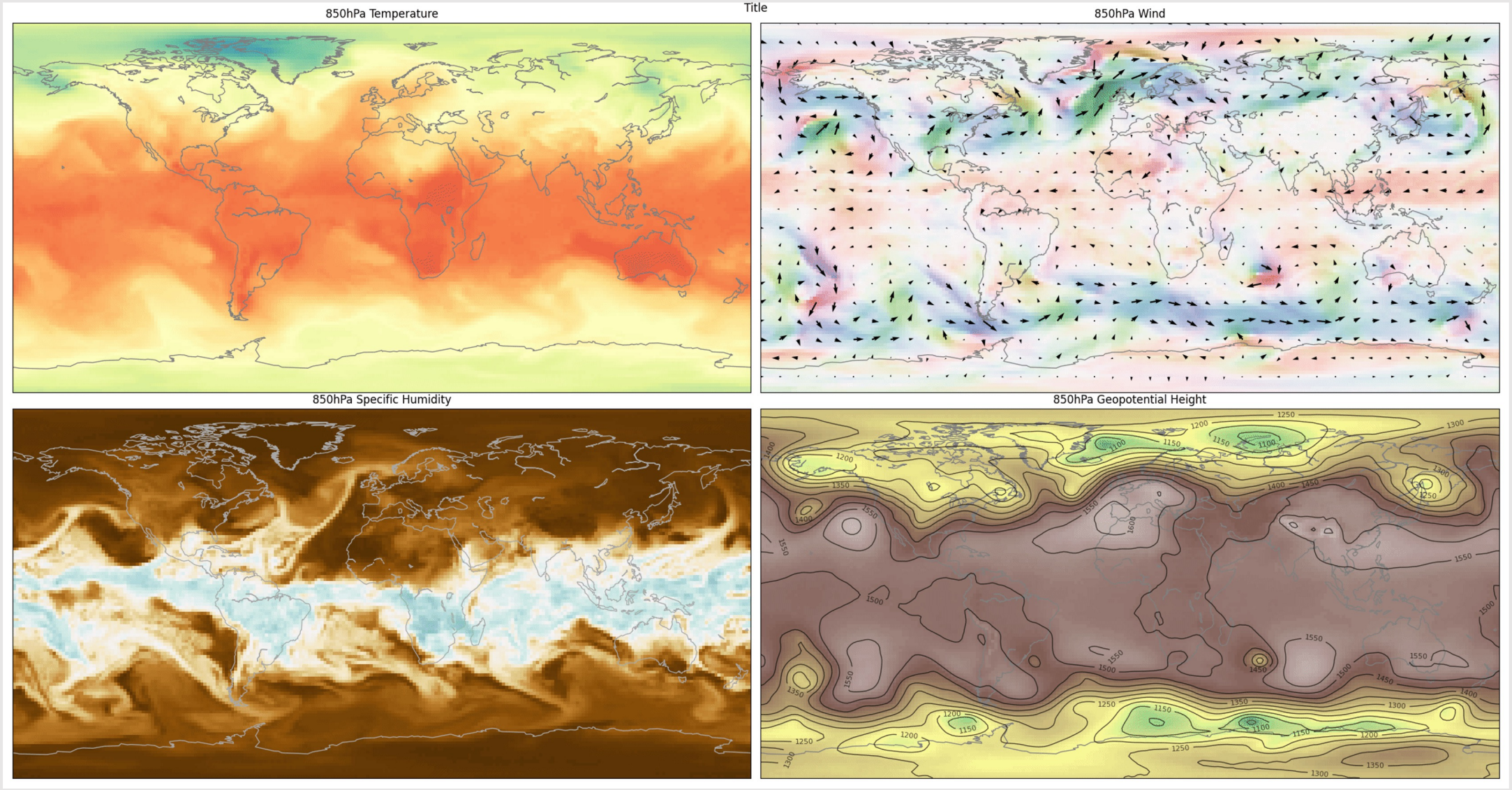


¡Rendimiento de vanguardia con MUCHO menos cómputo!

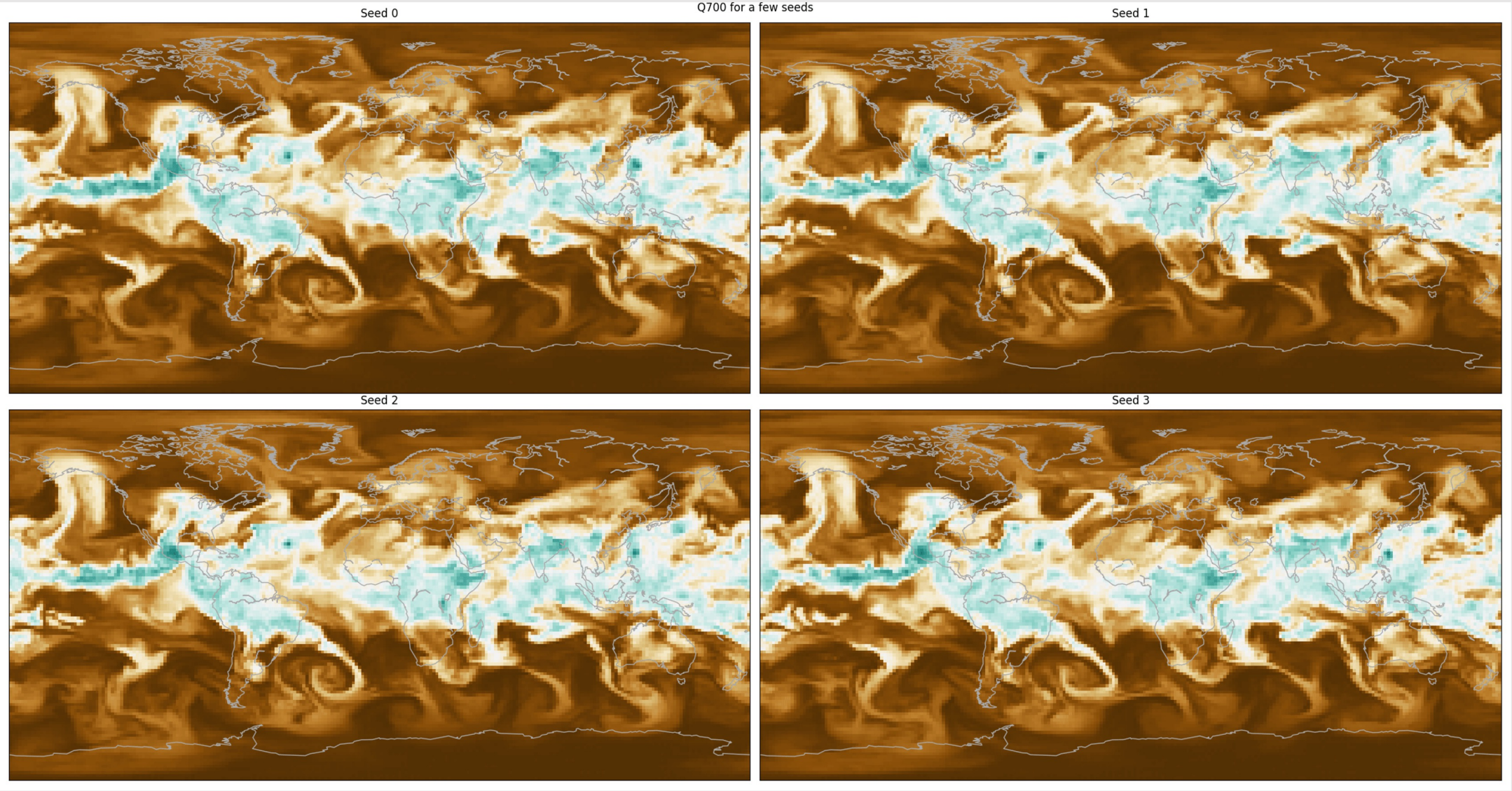
[Couairon, Singh, Charantonis, Lessig, Monteleoni, *Science Advances*, 2026]

Tablas de puntuación en WeatherBench 2 de Google

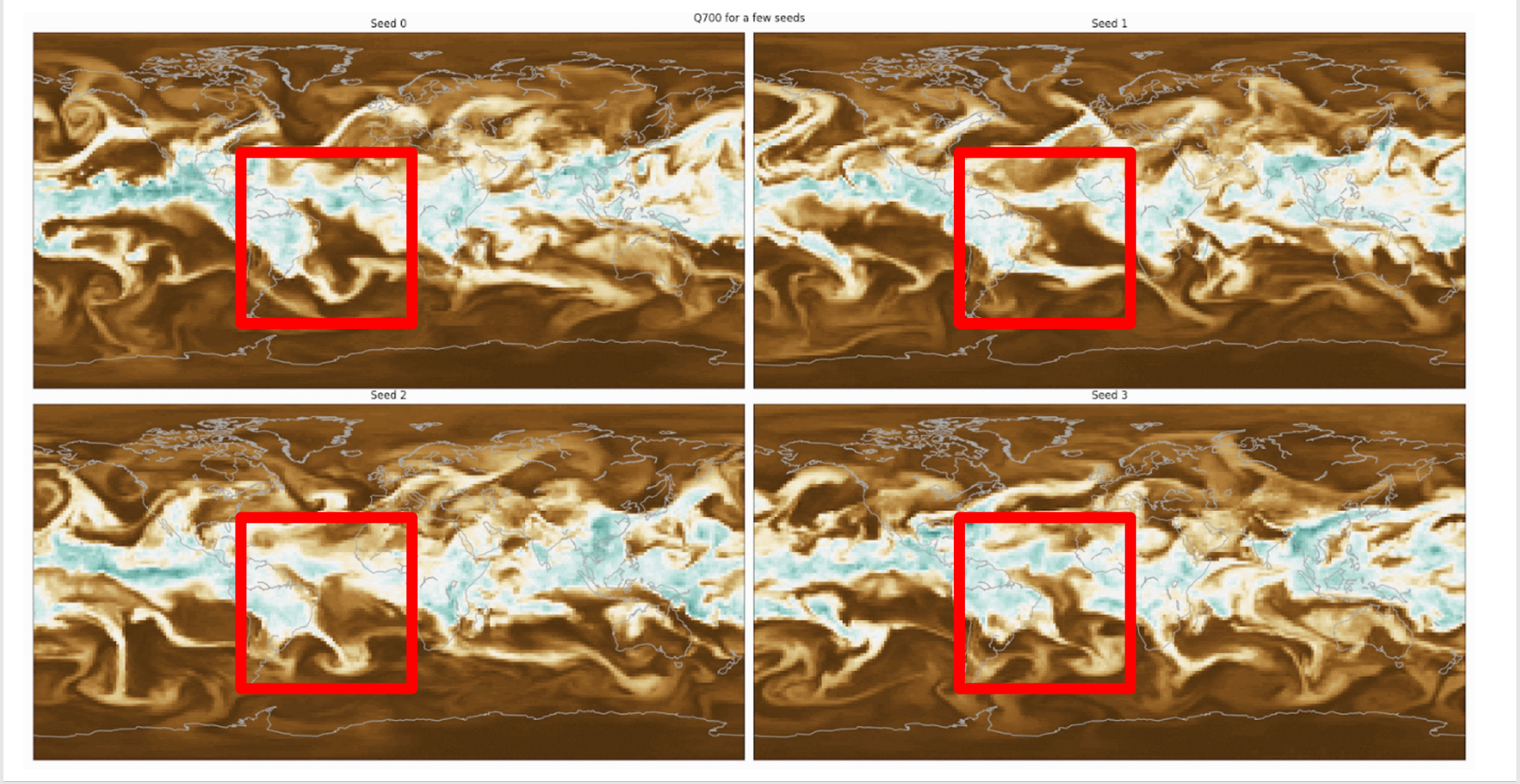
Una trayectoria de ejemplo



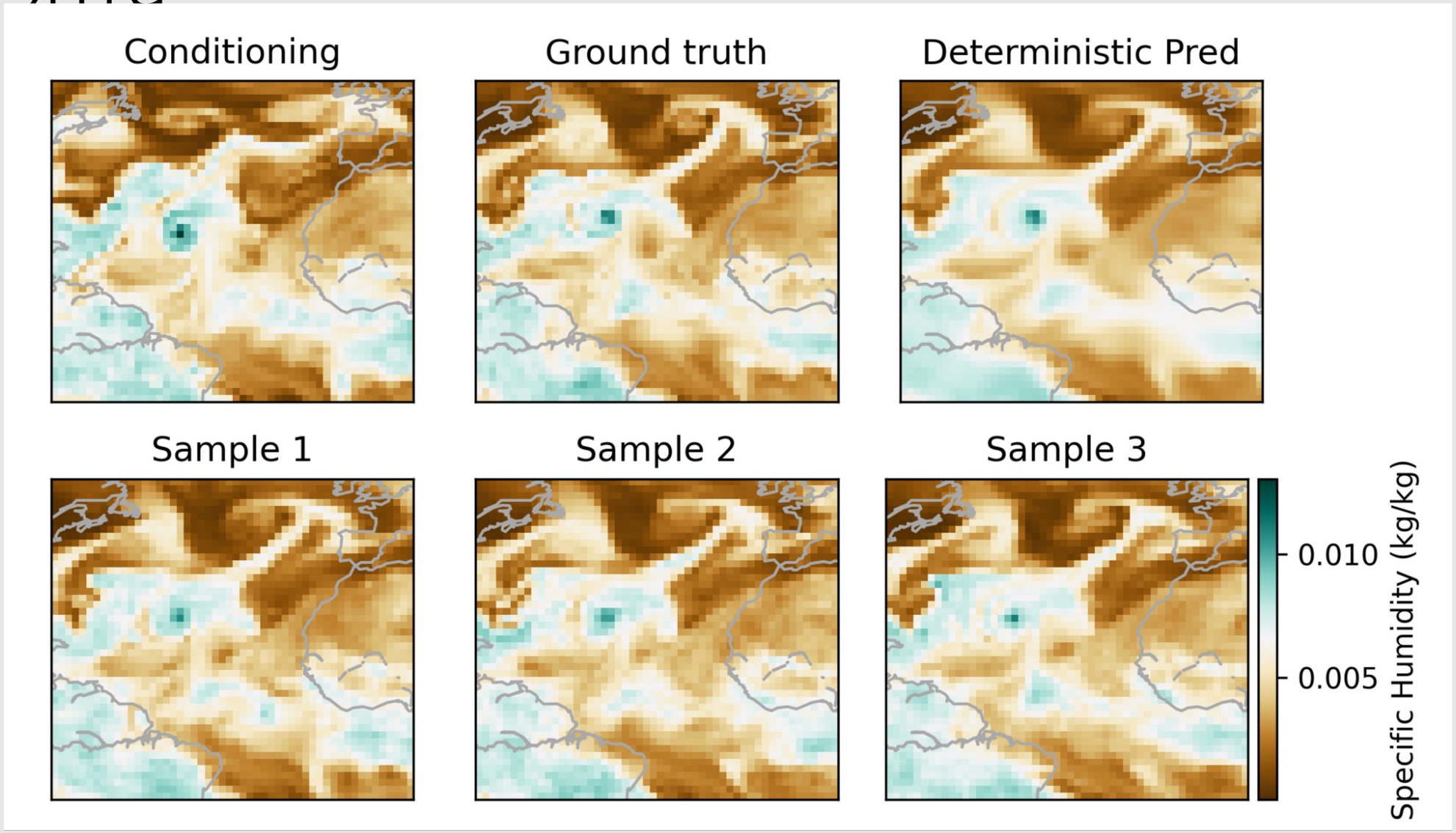
Distintas semillas - Q700



Diversidad de los miembros del conjunto; plazo de 10 días



Caso de estudio: Lorenzo - Q700 - inicialización 25 sept. 2019



ArchesClimate: emulación de modelos climáticos con IA



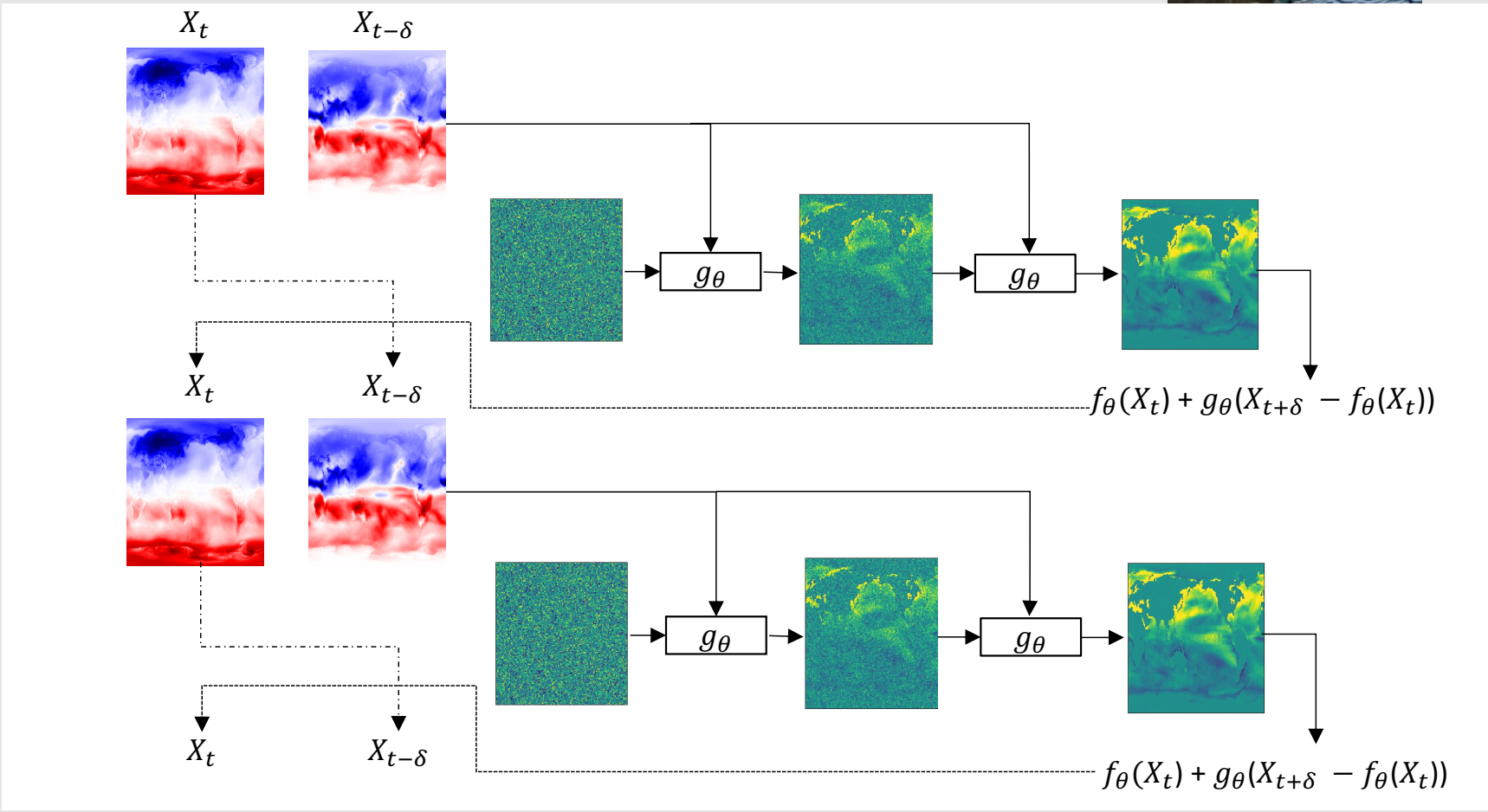
Tesis doctoral en curso de Graham Clyne
Director principal: Anastase Charantonis

Cadena: ArchesWeatherGen modificado:
- Sin atención 1D por columnas
- Inclusión de forzamientos

Datos de entrenamiento: simulaciones de modelos climáticos (IPSL DCPP)

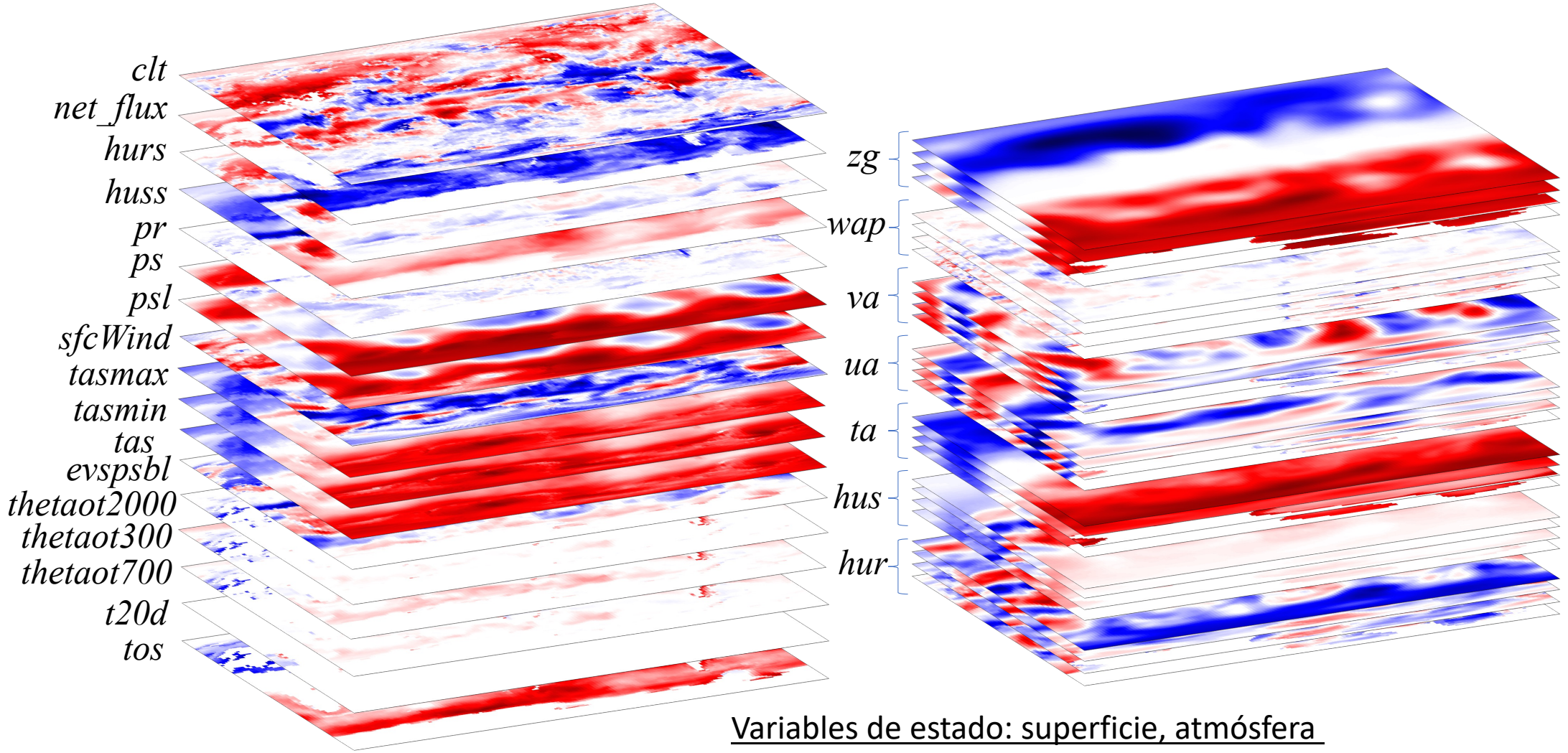
Entrenamiento: predecir un mes por delante

Despliegues con una muestra
cada mes. Esta salida se reinyecta en el modelo entrenado y se repite.



Clyne et al., *resultados preliminares presentados en EGU 2025 y Climate Informatics 2025*

Variables de entrada y forzamientos

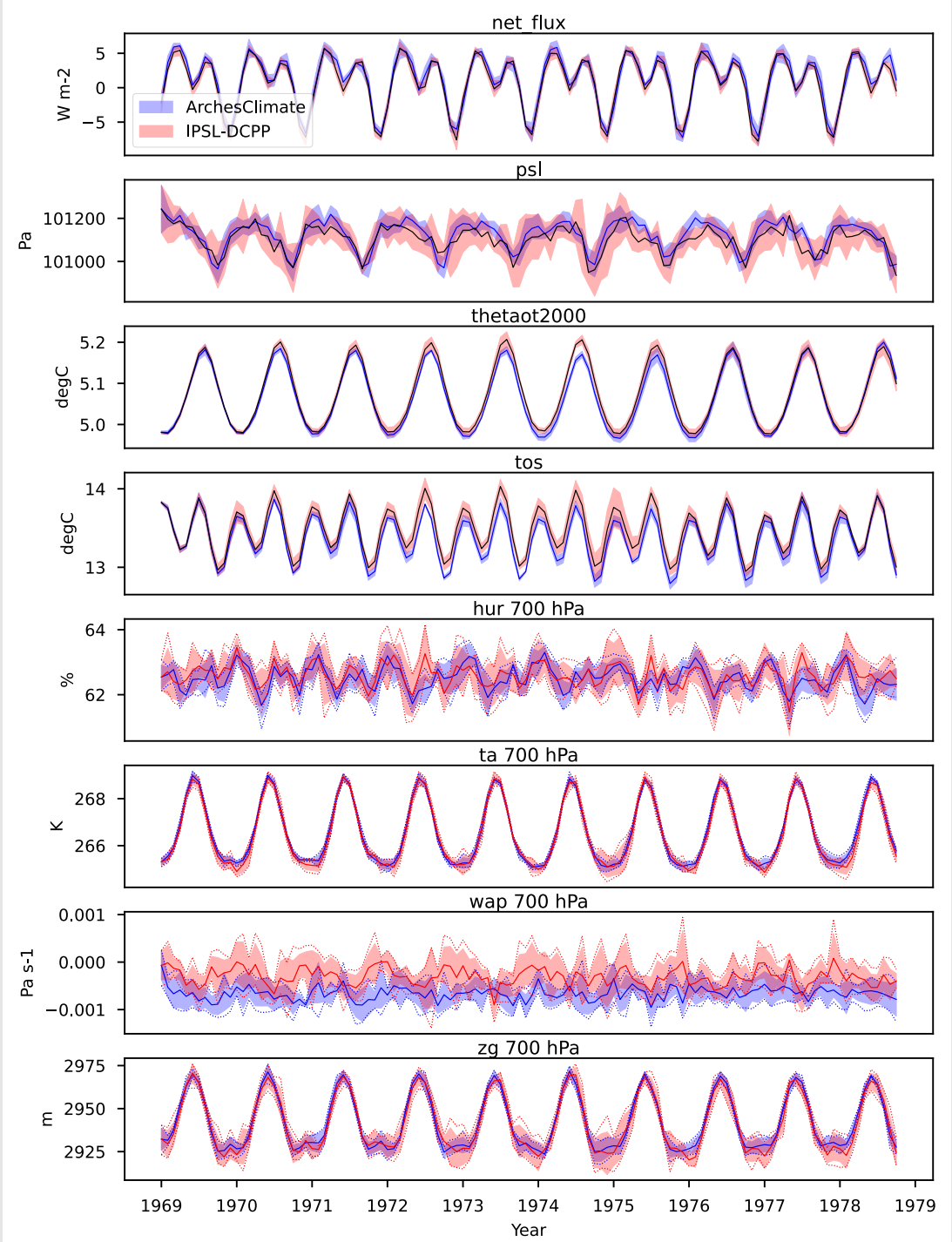


$$\begin{bmatrix} CO2 \\ CH4 \\ N2O \\ CFC11eq \\ SSI \end{bmatrix}$$

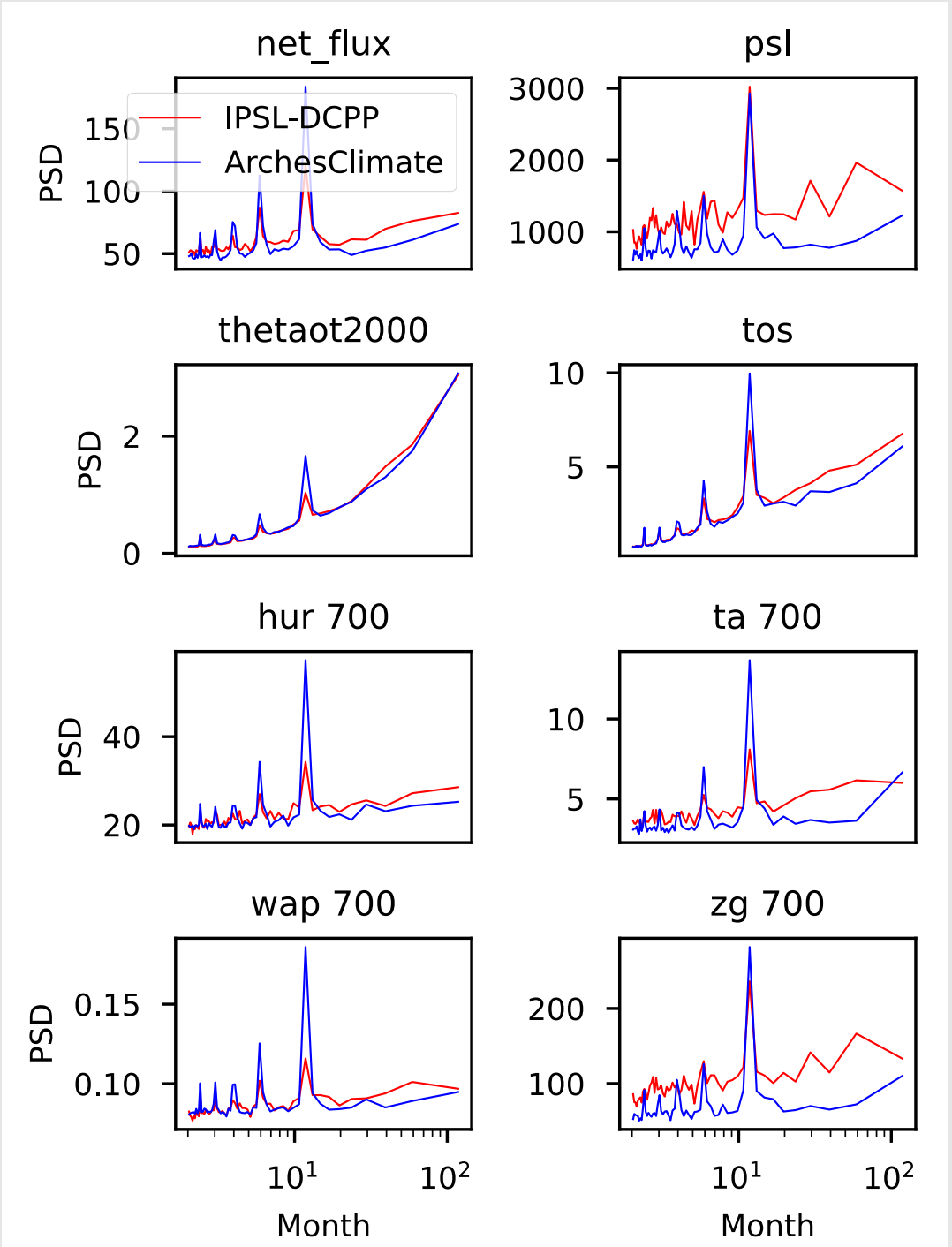
Variables de estado: superficie, atmósfera

Forzamientos

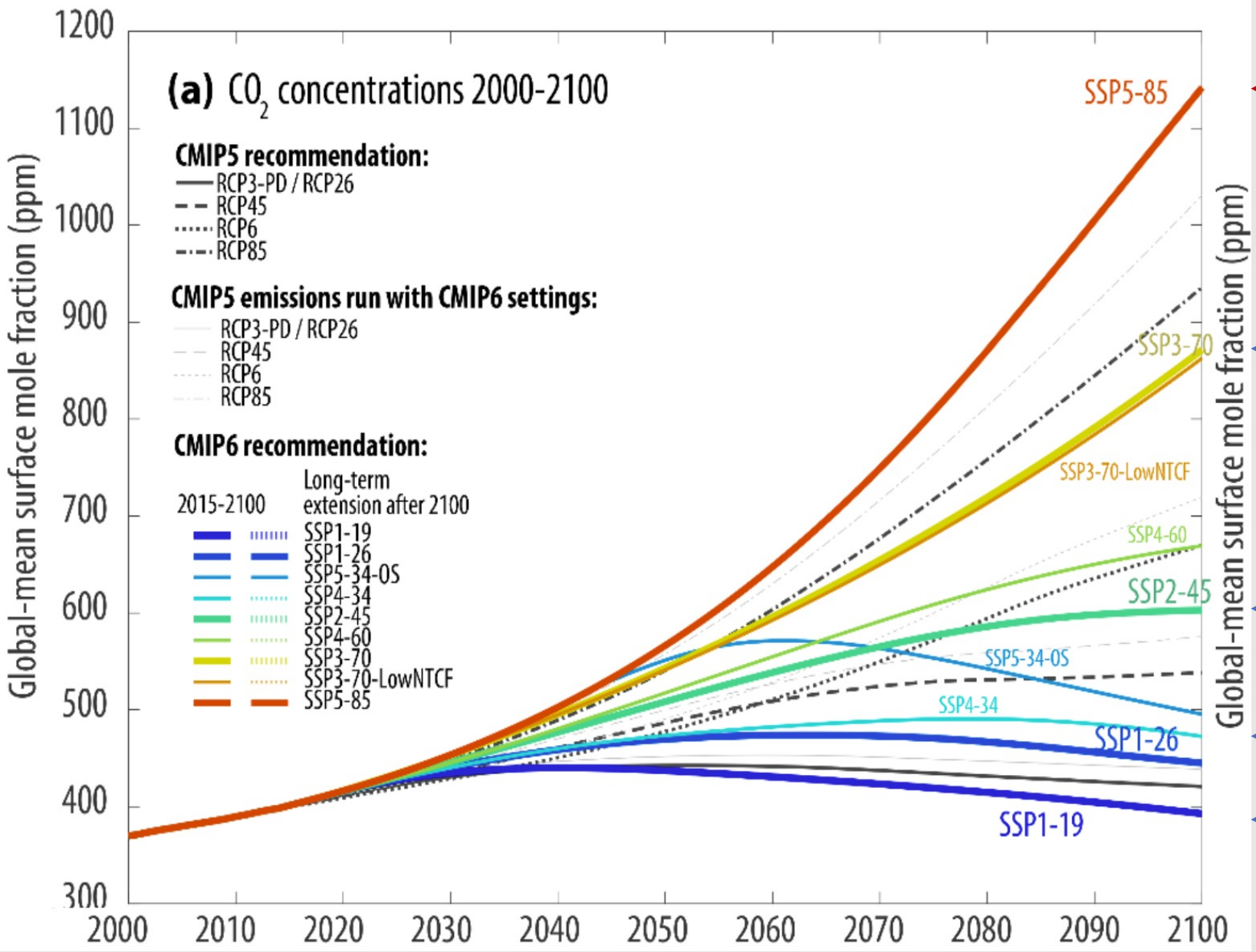
Despliegues decenales



Resultados preliminares



Extrapolar escenarios de modelos climáticos



Datos de prueba

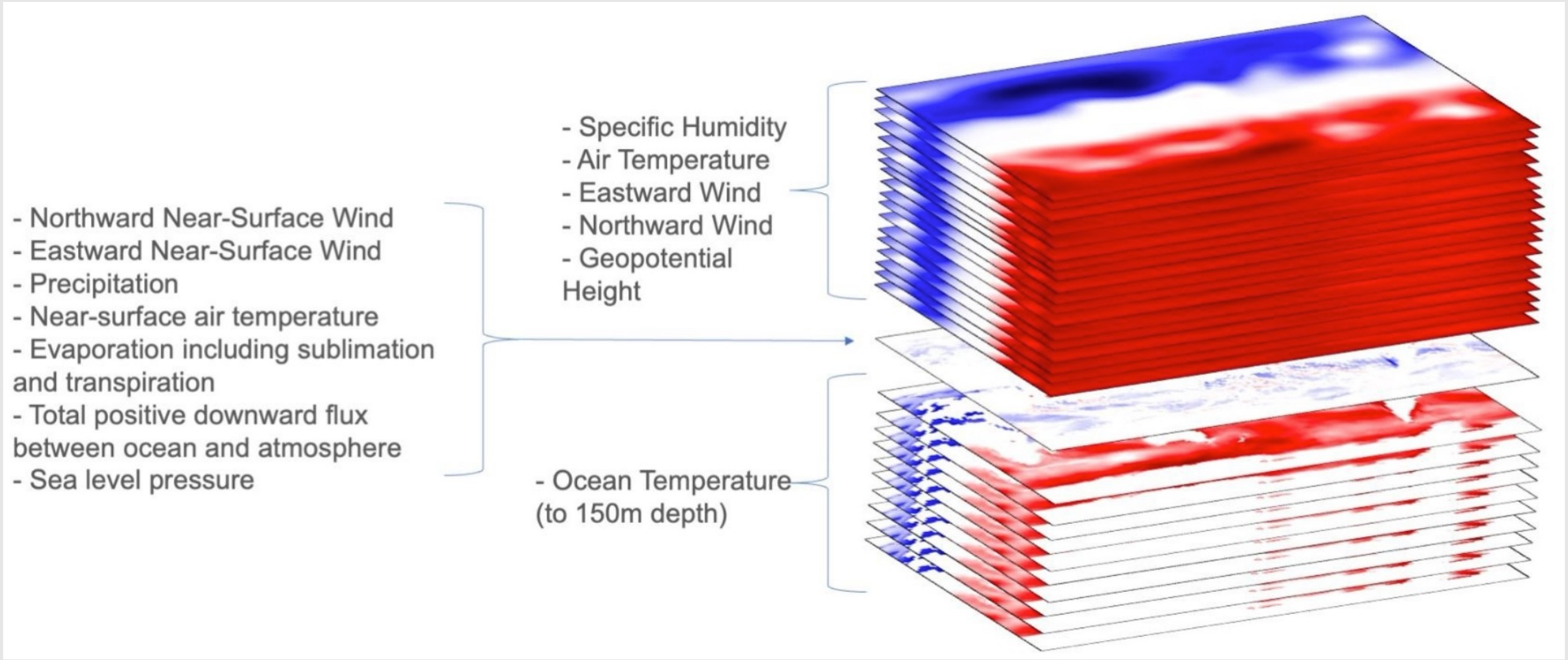
Tarea de extrapolación

Datos de entrenamiento

Estos 4 escenarios

Interpolación y extrapolación en el espacio de escenarios de modelos climáticos

Clyne et al., *en preparación*



Extrapolación de escenarios

Reproducir los patrones temporales de El Niño-Oscilación del Sur

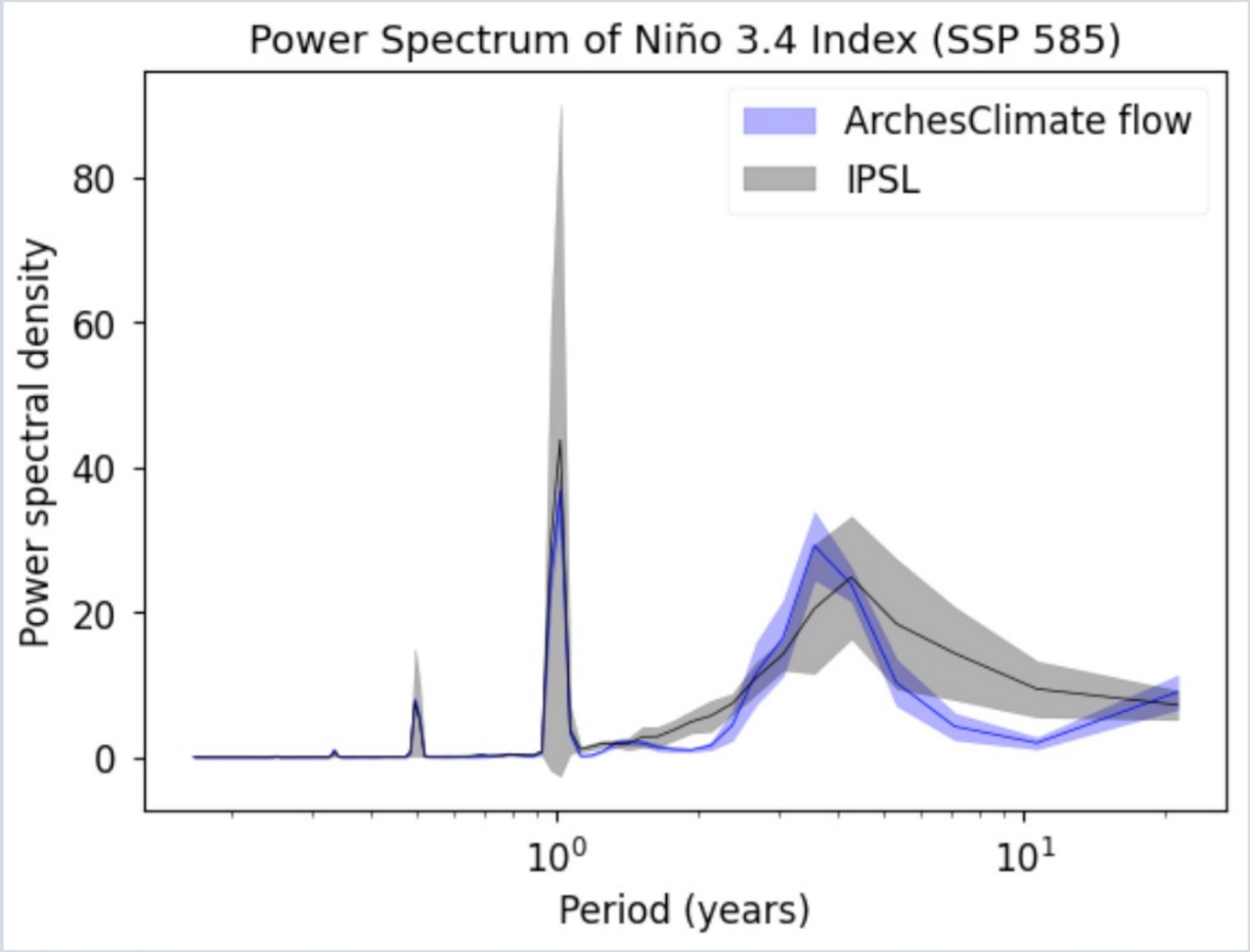


Figure 9. Power spectra for El Niño 3.4 index for SSP5-8.5.

Clyne et al., *en preparación*

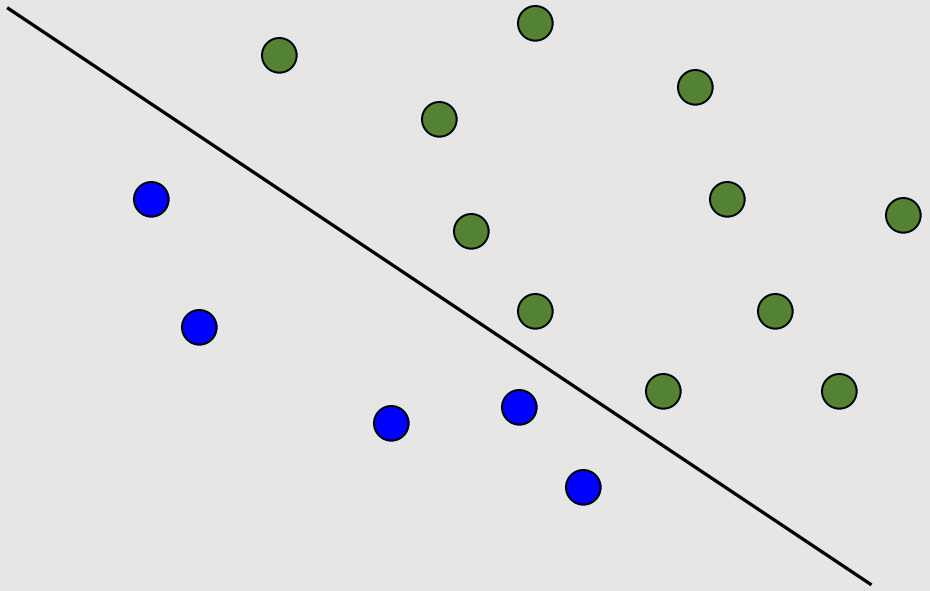
Parte II: Métodos de IA: aprendizaje autosupervisado

Aprendizaje supervisado

Dados puntos de datos etiquetados, encontrar una buena regla de clasificación.

Describe los datos

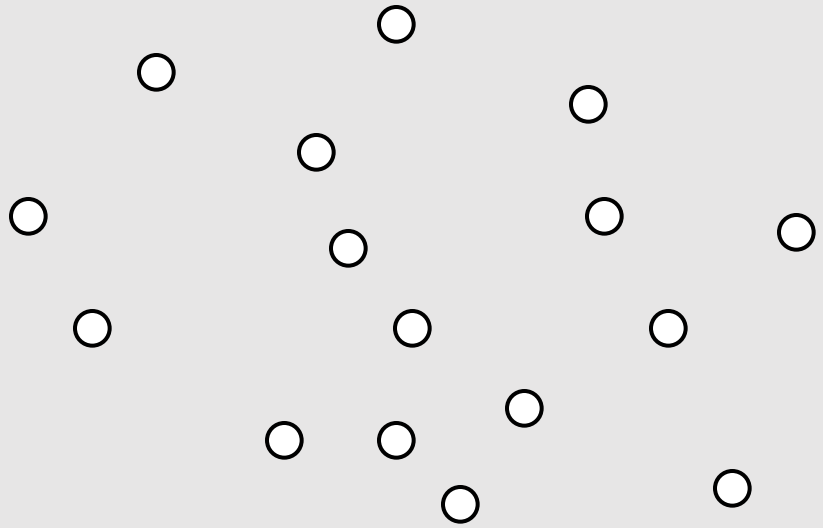
Generaliza bien



Aprendizaje no supervisado

¿Qué se puede hacer sin ninguna etiqueta?

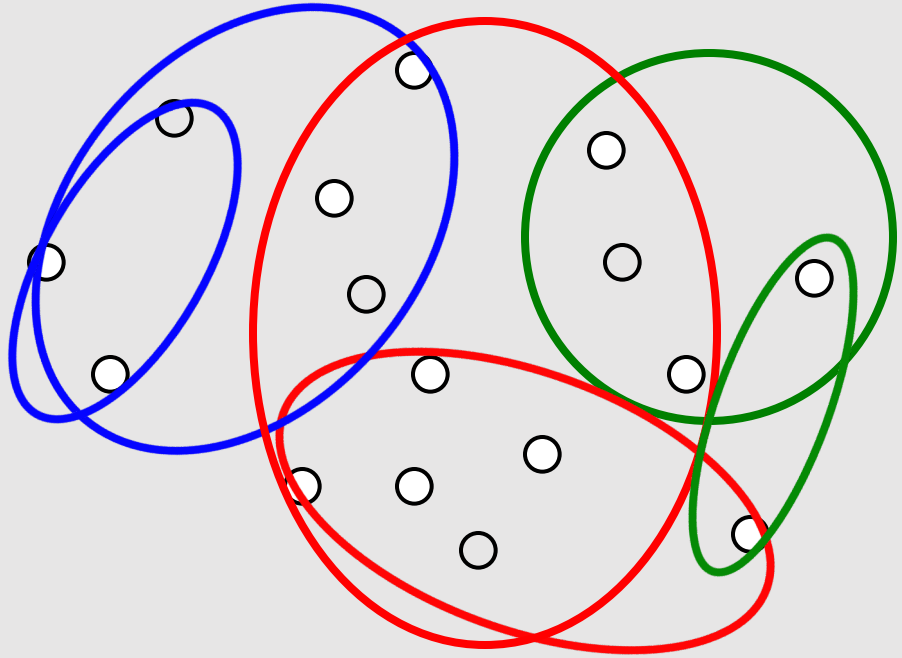
Aprendizaje no supervisado



Aprendizaje no supervisado

¿Qué se puede hacer sin ninguna etiqueta?

Aprendizaje no supervisado



Ejemplo: agrupamiento (clustering)

Aprender de datos «no etiquetados»

- Tareas de aprendizaje no supervisado
  1. Agrupamiento
  2. Incrustación (reducción de dimensionalidad, incluidos métodos no lineales)
  3. Interpolación
  4. Aprendizaje de un modelo generativo

- Hay enfoques de aprendizaje profundo (AP) para todas estas tareas

- Hay enfoques sin AP para todas estas tareas

- Hay enfoques de AP autosupervisados para todas estas tareas

Aprendizaje profundo no supervisado

- AP supervisado. La pérdida de predicción es función de la etiqueta y y de la salida de la red para la entrada x.

Salida de la red

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

Función de pérdida

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

- AP no supervisado. La pérdida de predicción solo depende de x y de la salida de la red para la entrada x. No hay etiqueta y.

Salida de la red

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

Función de pérdida

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

Enfoque autosupervisado del aprendizaje no supervisado

Aprendizaje autosupervisado

Un enfoque de vanguardia del aprendizaje (profundo) no supervisado

Diseñar una tarea pretexto:

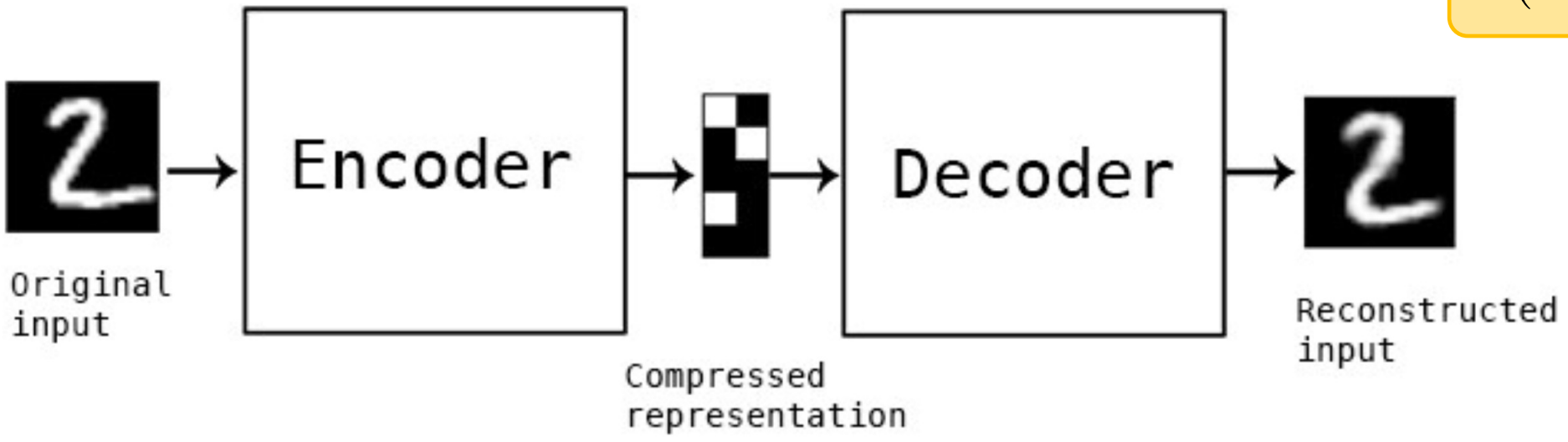
- Diseñar una tarea de aprendizaje supervisado usando solo los datos disponibles.
- Entrenar un modelo en esta tarea de modo que
- las características aprendidas (o la posterior aprendida sobre un espacio de características) sean útiles para otra tarea (posterior)

Tarea pretexto: ejemplo

Ejemplo clásico de tarea pretexto: el autoencoder

- Entrenar una red neuronal de forma no supervisada
  - Usar los datos no etiquetados tanto como entrada como para evaluar la salida
- Tras el entrenamiento, la capa cuello de botella será una representación compacta de la distribución de entrada

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



Aprendizaje autosupervisado

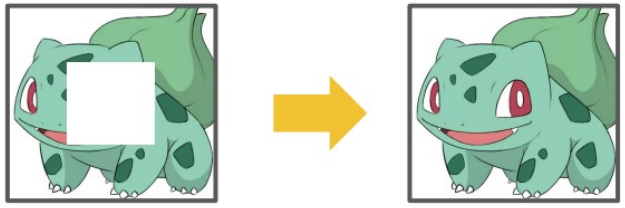
- Diseñar una tarea pretexto que haga aprender características útiles.

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



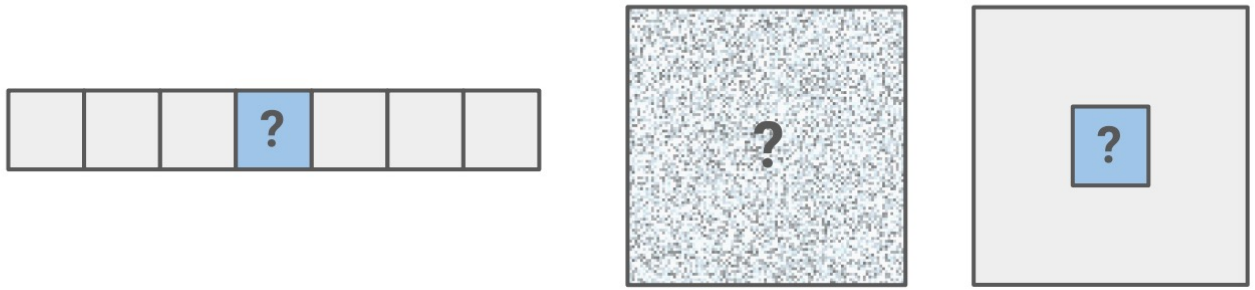
Lilian Weng, Jong Wook Kim, tutorial NeurIPS 2021

Aprendizaje autosupervisado

- Diseñar una tarea pretexto que haga aprender características útiles.

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)

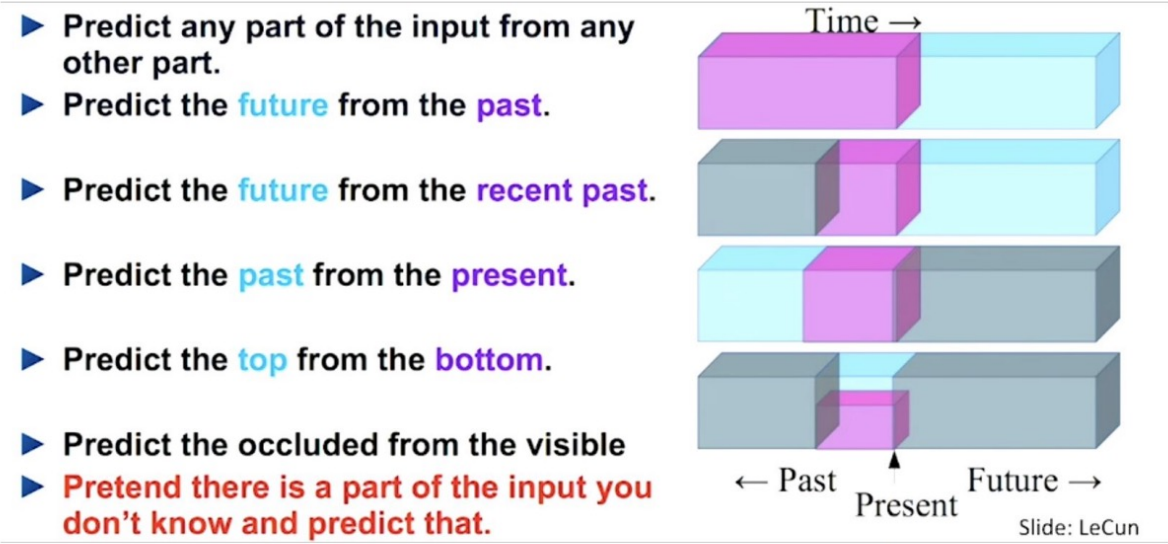
Lilian Weng, Jong Wook Kim, tutorial NeurIPS 2021

Aprendizaje autosupervisado

- Diseñar una tarea pretexto que haga aprender características útiles.

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

Aprendizaje autosupervisado: aprendizaje contrastivo

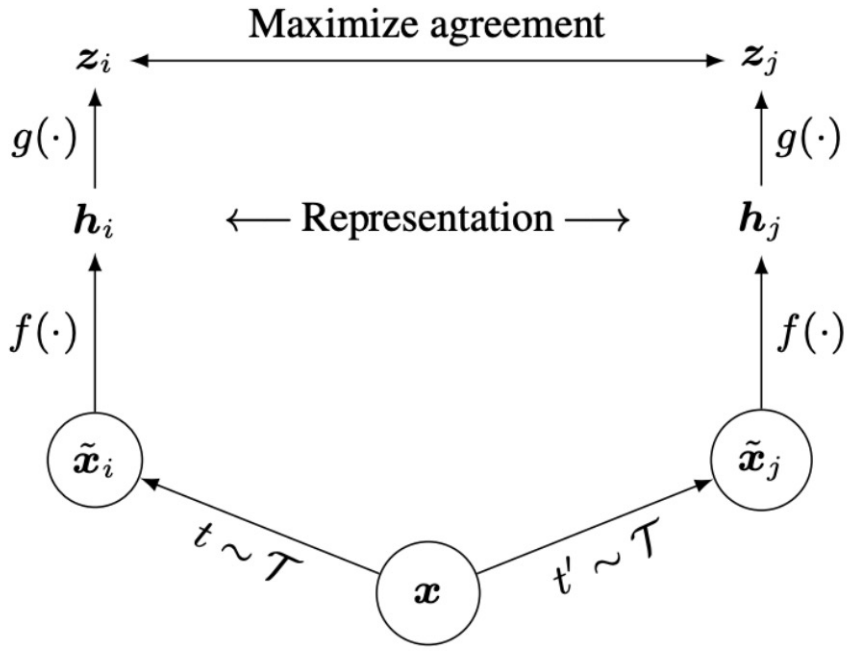


Figure 2. A simple framework for contrastive learning of visual representations. Two separate data augmentation operators are sampled from the same family of augmentations ($t \sim \mathcal{T}$ and $t' \sim \mathcal{T}$) and applied to each data example to obtain two correlated views. A base encoder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

Tarea pretexto:

1. Aumentar la entrada x dos veces: $\tilde{x}_i, \tilde{x}_j$
2. Aprender una correspondencia $g(f(\cdot))$, minimizando una función de pérdida para acercar $g(f(\tilde{x}_i)), g(f(\tilde{x}_j))$.

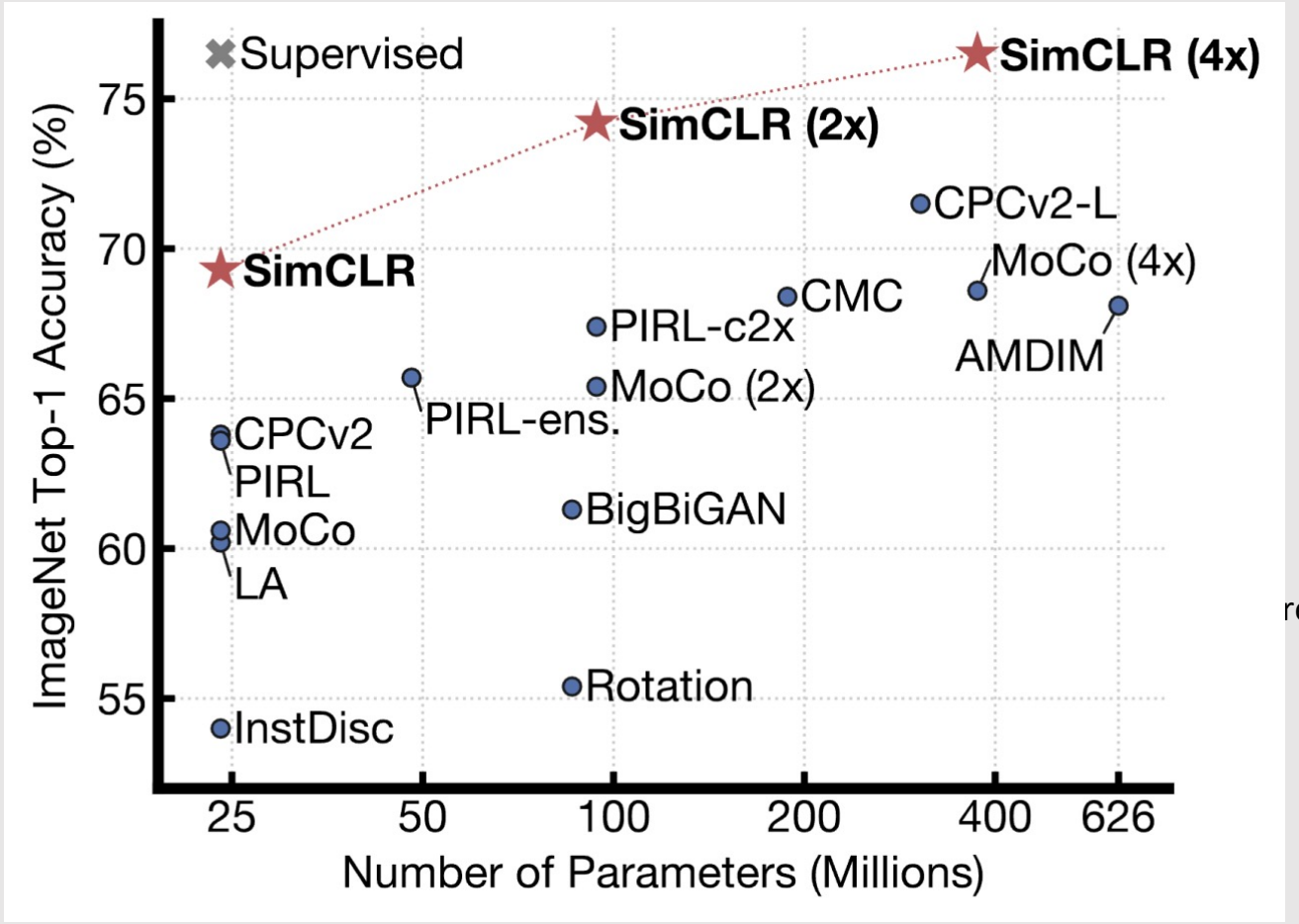
$$\mathcal{L}(\hat{x}, x)$$

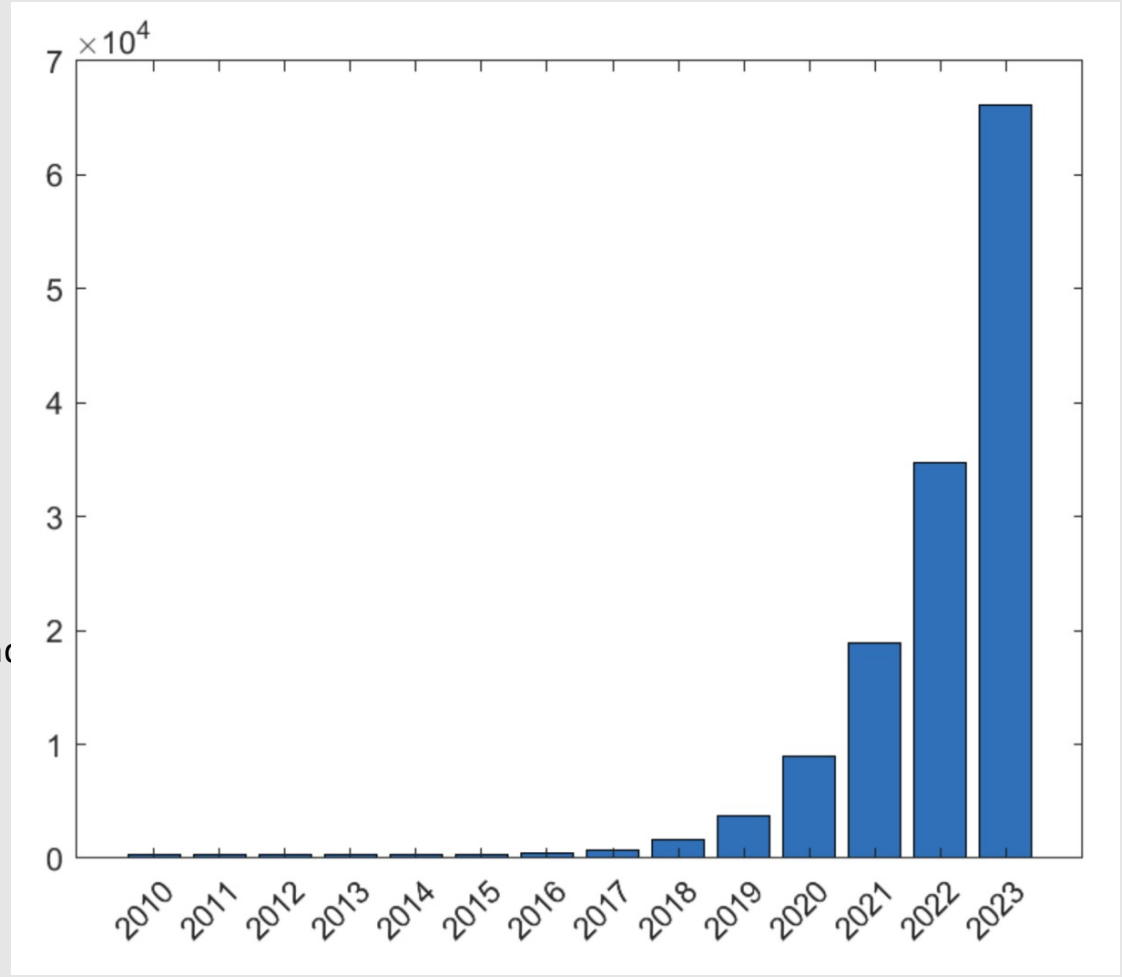
$$\hat{x} := \{g(f(\tilde{x}_i)), g(f(\tilde{x}_j))\}$$

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

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

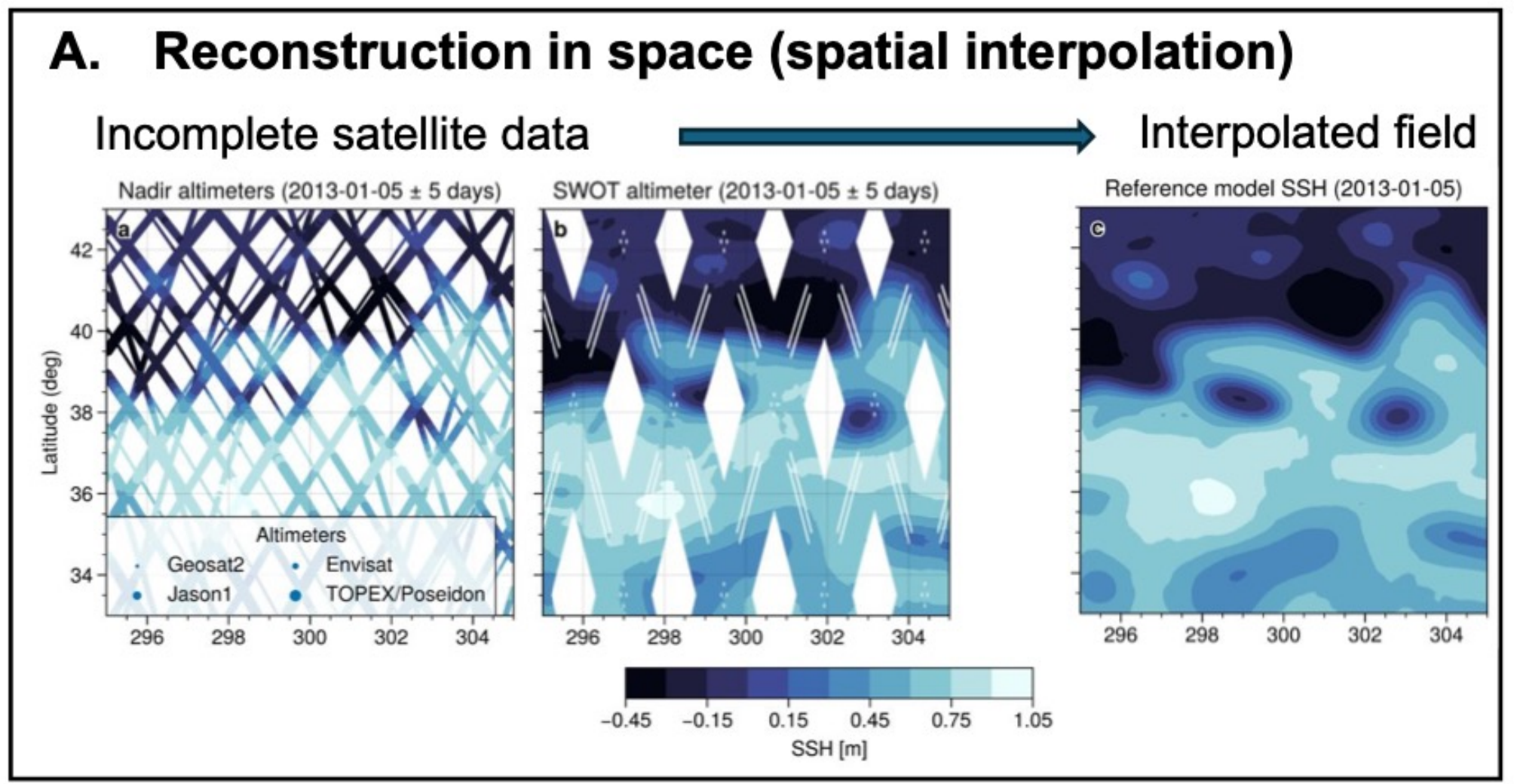
El aprendizaje autosupervisado puede superar al supervisado





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

Aprendizaje autosupervisado para vuestras aplicaciones



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

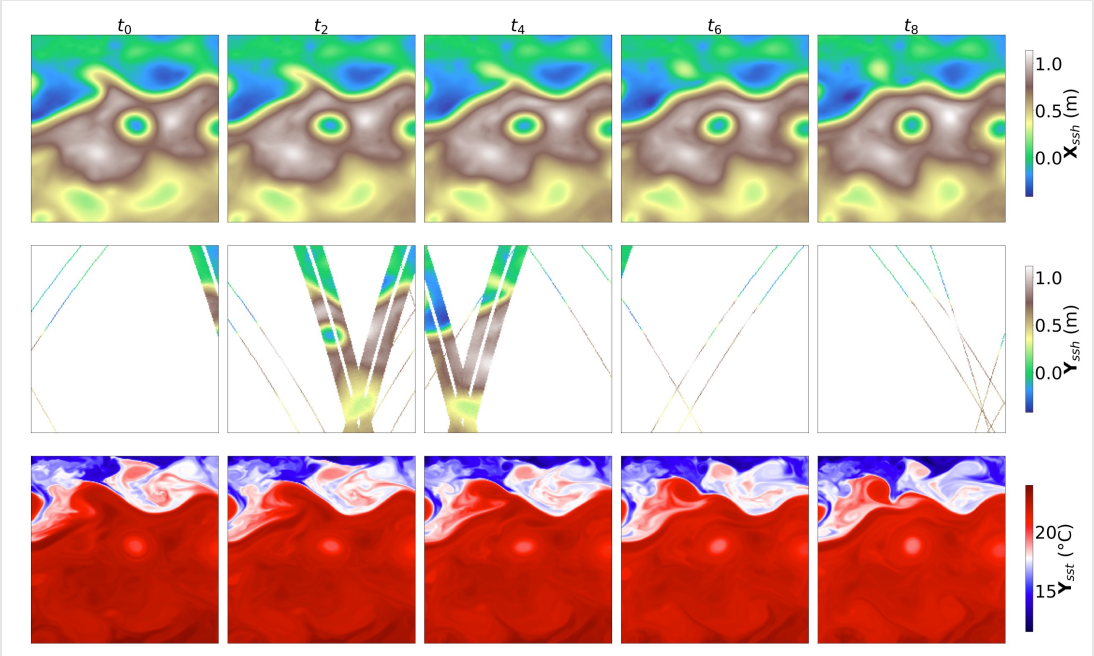
Interpolación espaciotemporal no supervisada

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

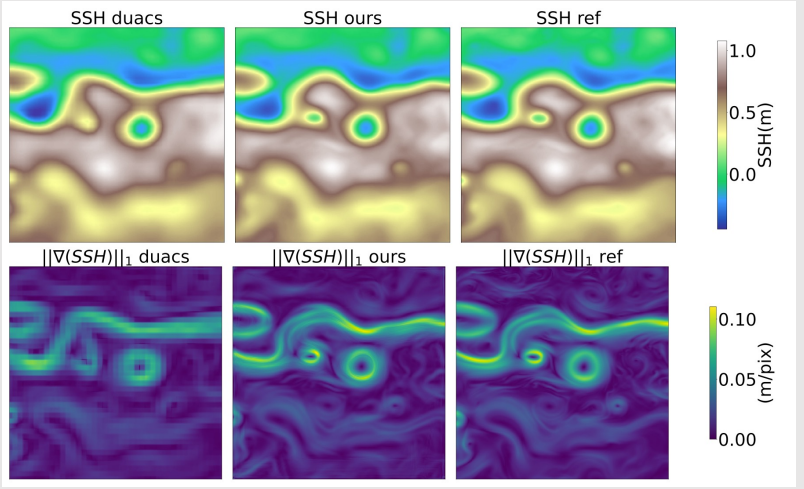
Campos SSH de referencia (no usados en el entrenamiento)

Datos SSH a lo largo de las trazas satelitales

Campos SST de un modelo físico



Datos de entrenamiento



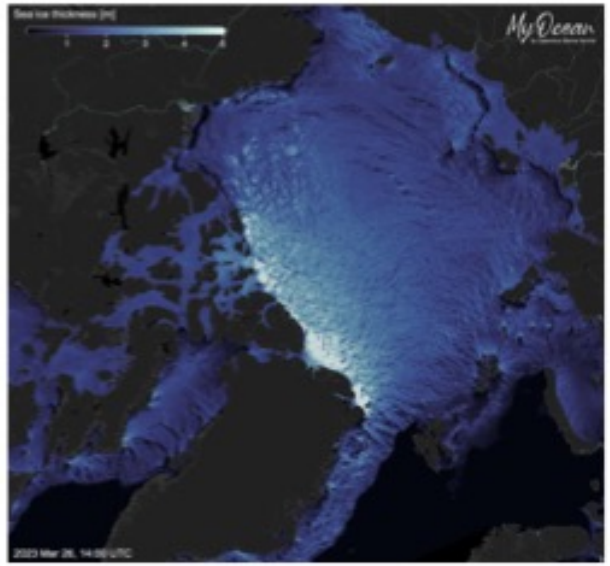
Resultados

Aprendizaje autosupervisado para vuestras aplicaciones

C. Reconstruction in scales (downscaling)

Low resolution field → High resolution field





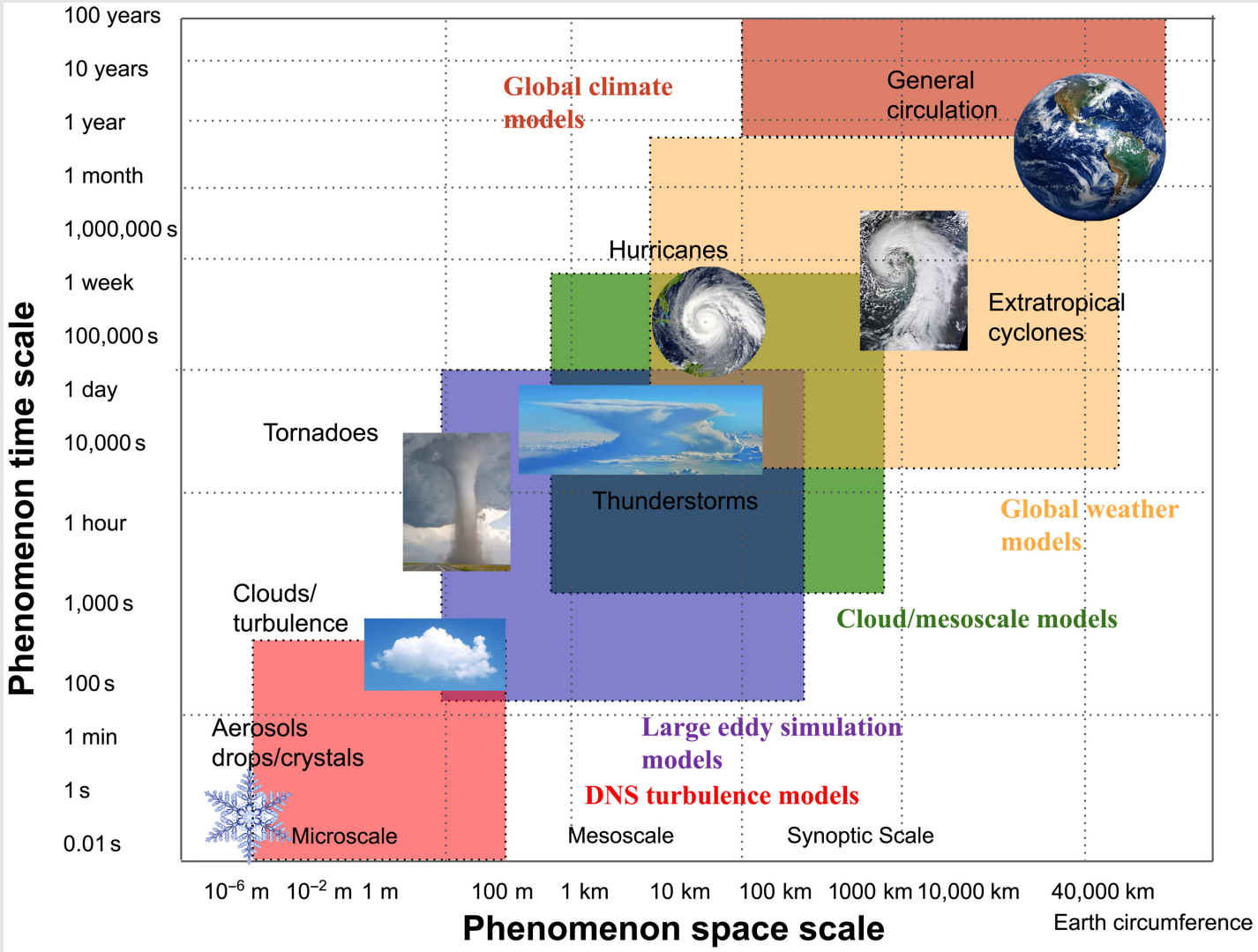
[Bracco et al., Machine Learning for the Physics of Climate. *Nature Reviews Physics*, 2024]

Reducción de escala de simulaciones de modelos climáticos

Las simulaciones de los modelos climáticos globales son más gruesas (en espacio y tiempo) de lo que requieren muchas tareas de:

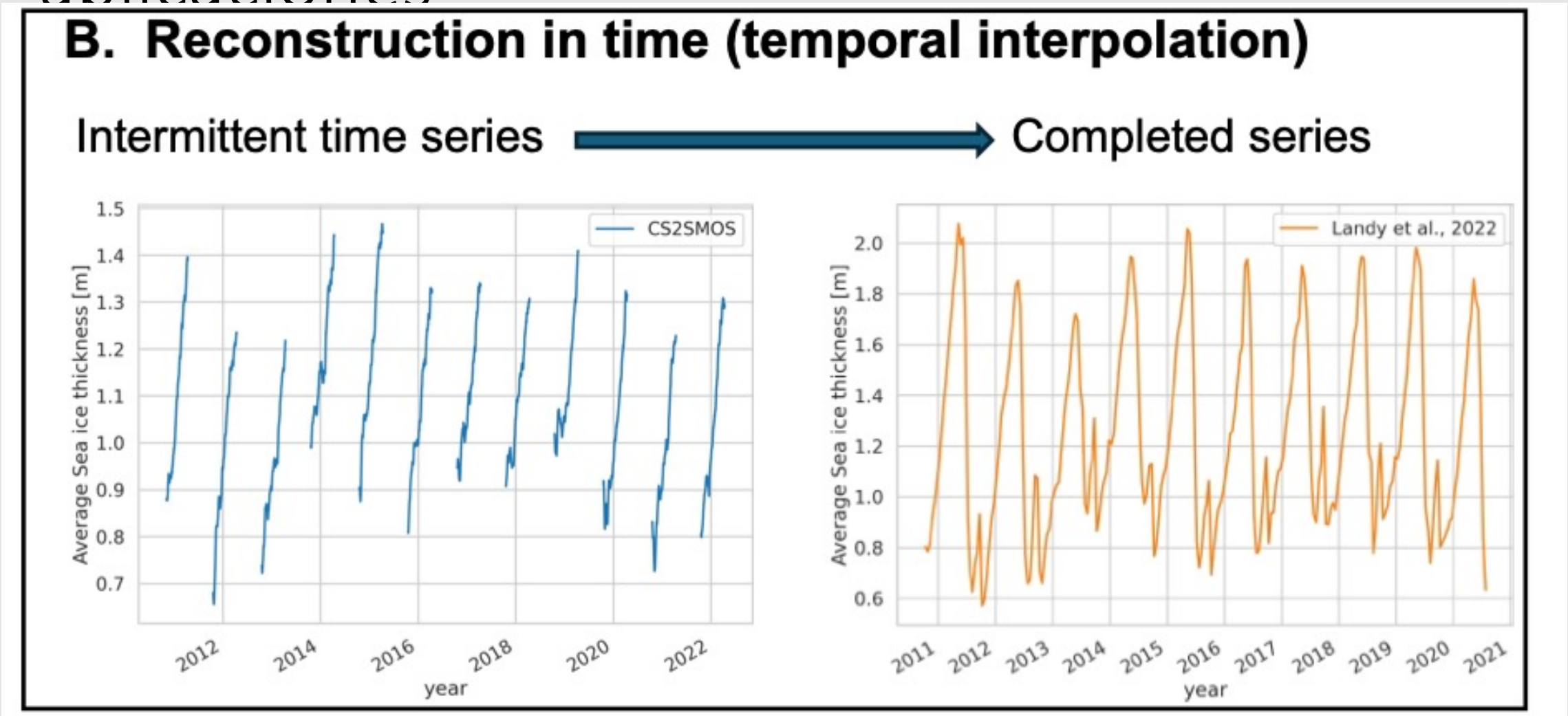
- Adaptación al cambio climático
- Mitigación del cambio climático
- Proyección de impactos a largo plazo

Enfoque: usar IA generativa para reducir la escala de los datos de modelos climáticos a escalas relevantes



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

Aprendizaje autosupervisado para vuestras aplicaciones



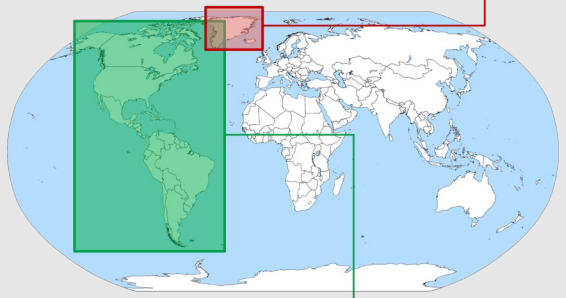
[Bracco et al., Machine Learning for the Physics of Climate. *Nature Reviews Physics*, 2024]

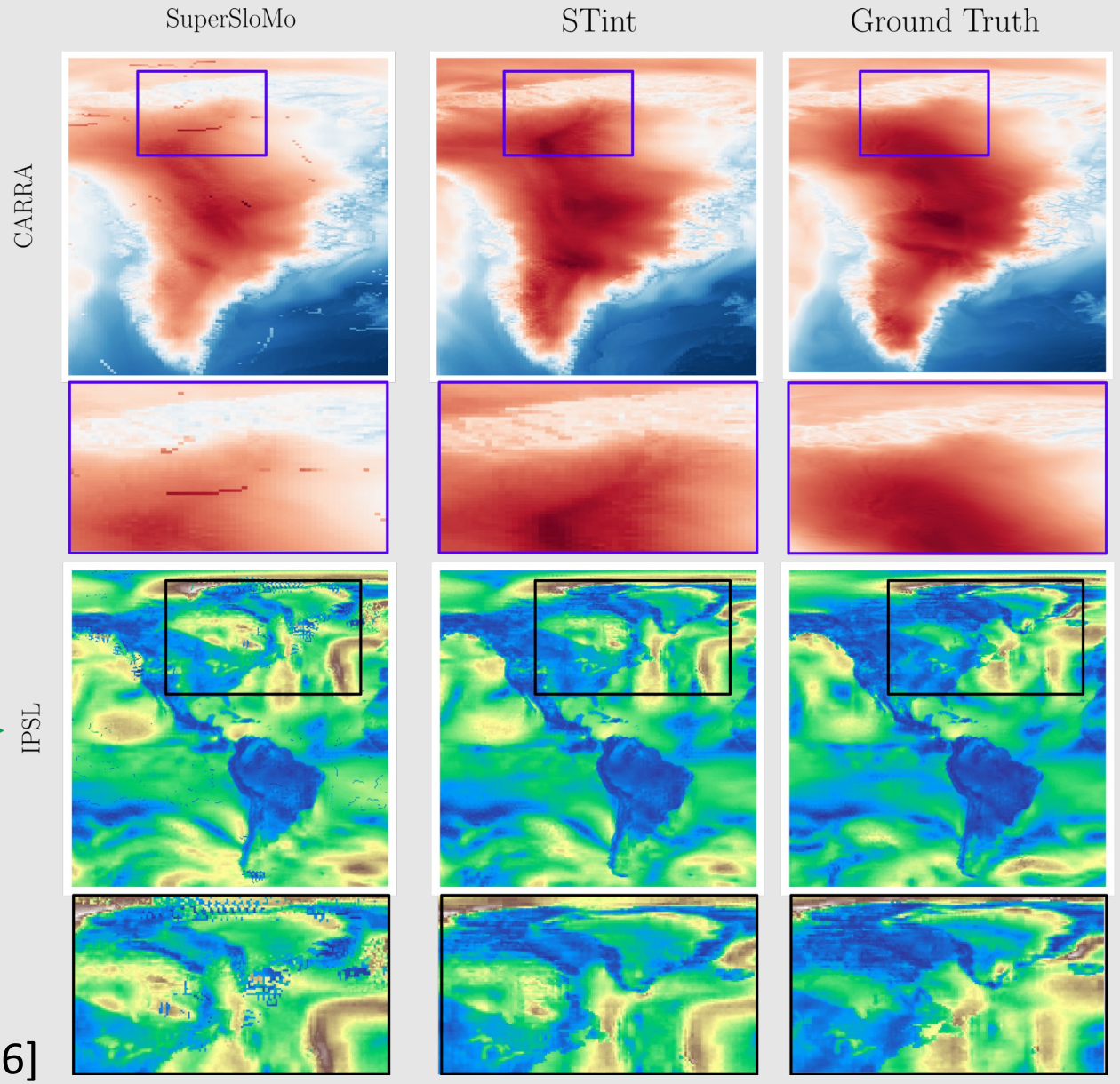
Ejemplo: reducción de escala temporal (interpolación) de datos geoespaciales

¿Por qué no usar simplemente los métodos de IA existentes?

- Las aplicaciones de cambio climático implican datos geoespaciales que evolucionan con el tiempo
  - Datos de observación en malla global mediante asimilación de datos
  - Simulaciones de modelos físicos (PNT, MCG, MCR)
- Son tensores de valores reales sobre latitud, longitud, tiempo,
  y posiblemente varias variables climatológicas
- Los algoritmos de visión por computador para «datos espaciotemporales» se basan en propiedades de los **datos de vídeo** que **no se generalizan bien a los datos geoespaciales**
  - p. ej. profundidad, bordes y «objetos»
  - frente a patrones efímeros en fluidos

STINT: interpolación temporal autosupervisada







El estado del arte en visión por computador para la interpolación temporal de vídeo usa el flujo óptico.
Esto es problemático con datos geoespaciales.

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

Limitaciones de trabajos previos: interpolación bicúbica

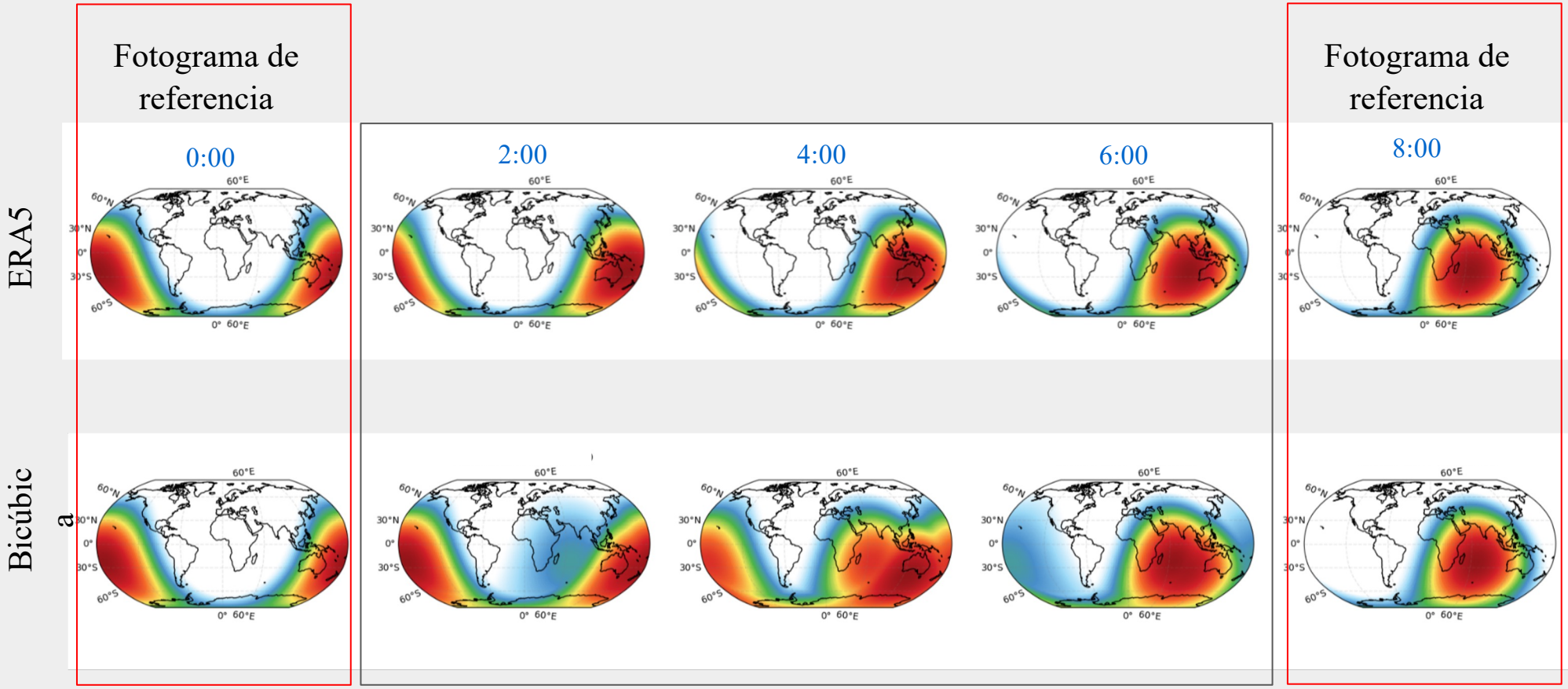
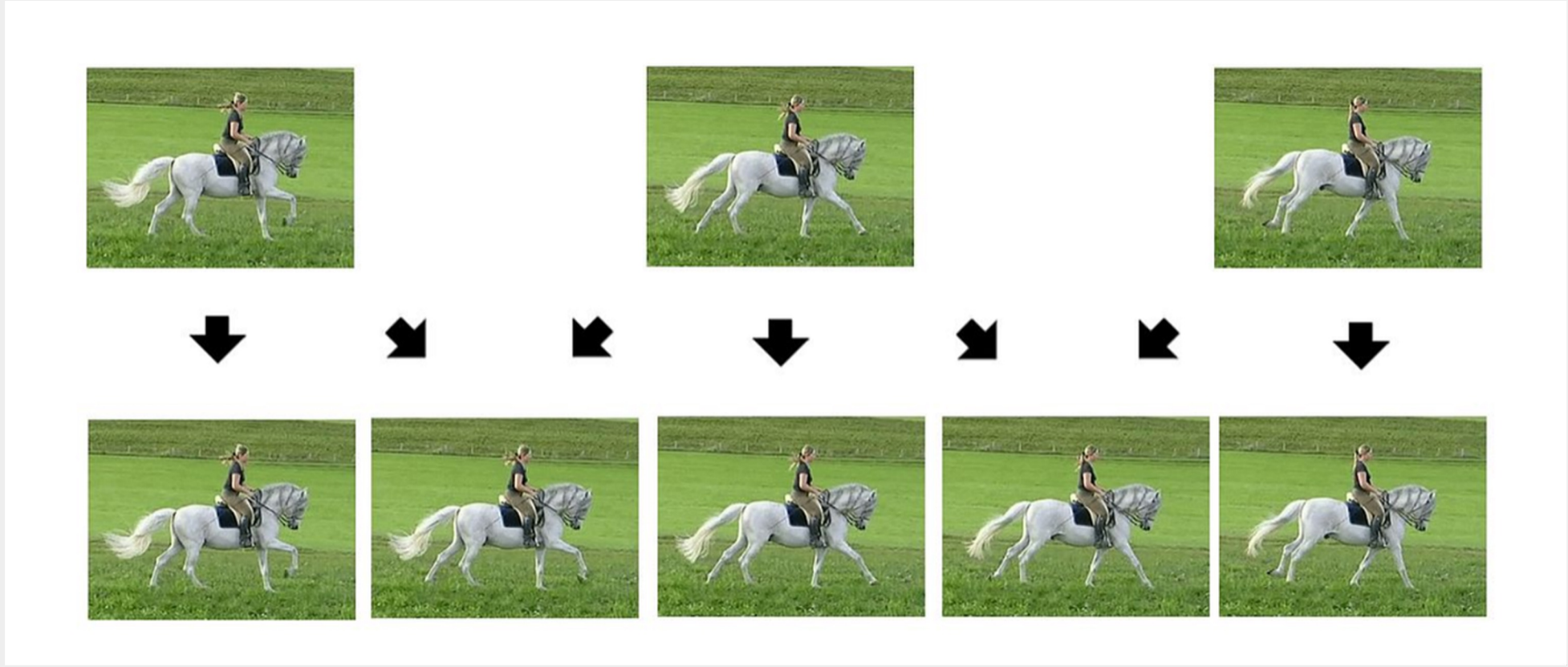


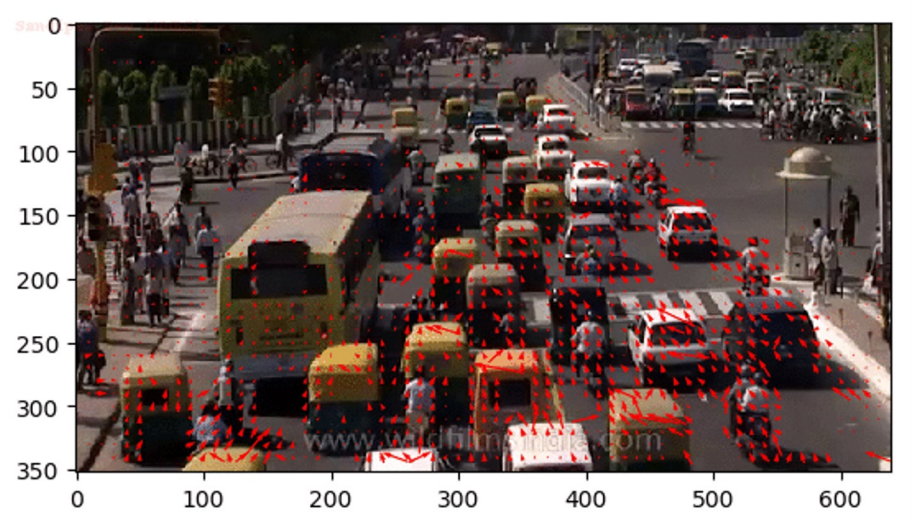
Figura. *Interpolación bicúbica temporal de campos de irradiancia solar ERA5*

Interpolación de vídeo en visión por computador

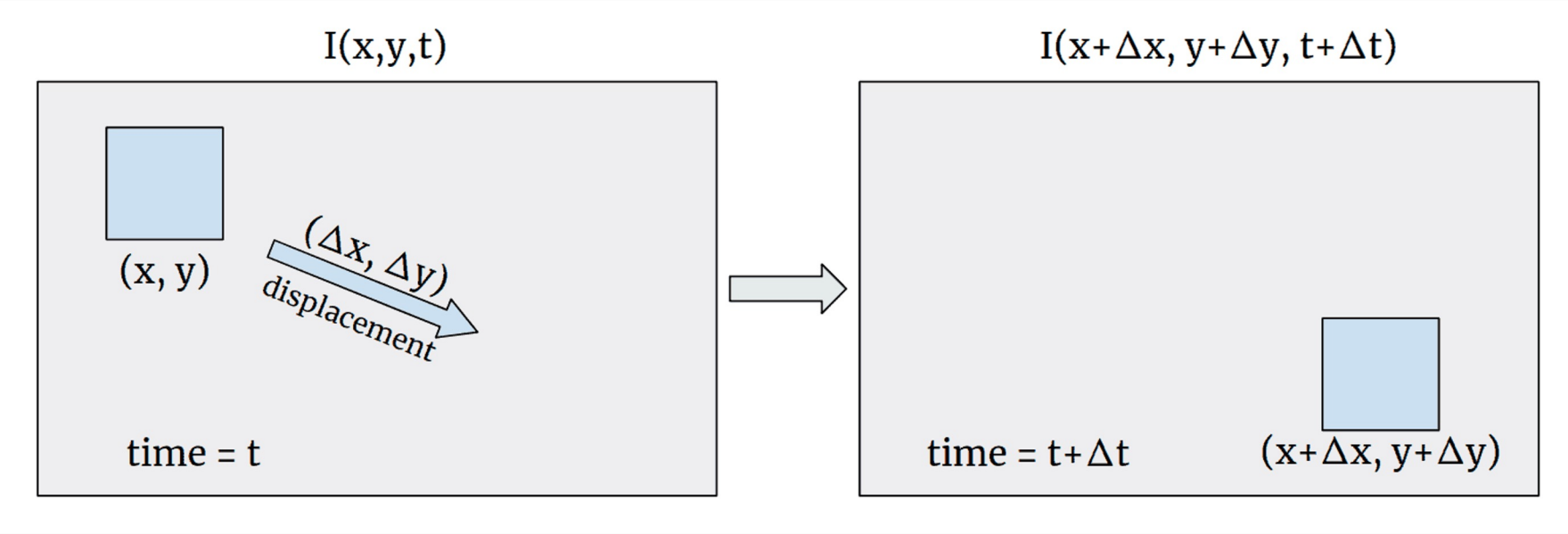


El flujo óptico

- Flujo óptico: campo de movimiento píxel a píxel entre dos imágenes (p. ej. fotogramas de vídeo).
- Codifica la proyección 2D «percibida» del movimiento físico de los puntos respecto al observador.



(fuente: Nvidia OF SDK)



El flujo óptico

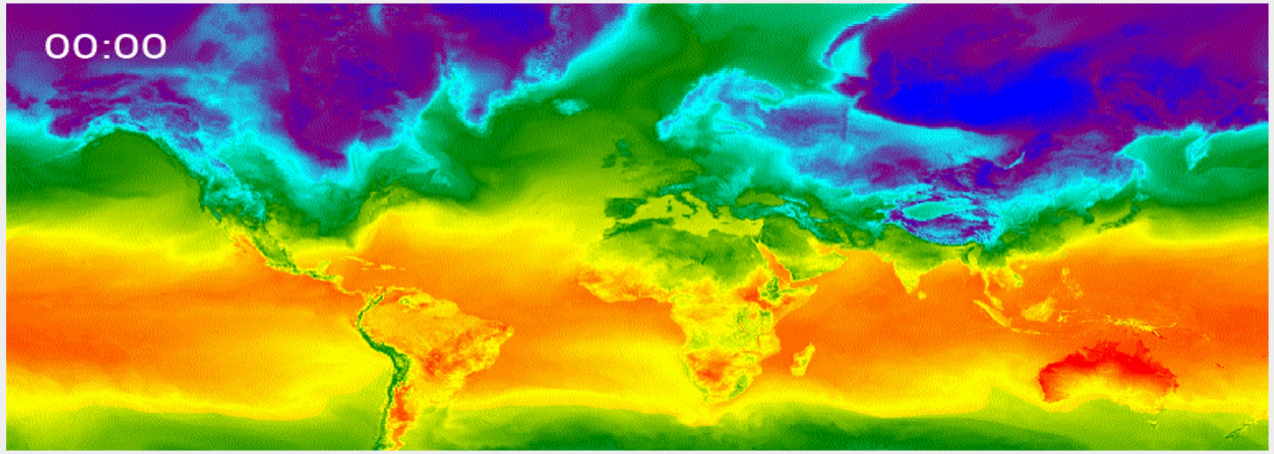
Supuestos:

- Desplazamiento pequeño (Δt).
- Ningún/poco cambio en la luz reflejada.
- Ningún/poco cambio de escala.
- Sin oclusión ni desoclusión.



¡No aplicable a datos geoespaciales!

- Sin noción de objetos ni de profundidad.
- Las entidades pueden formarse/deformarse arbitrariamente.
- Resolución temporal mucho más gruesa.



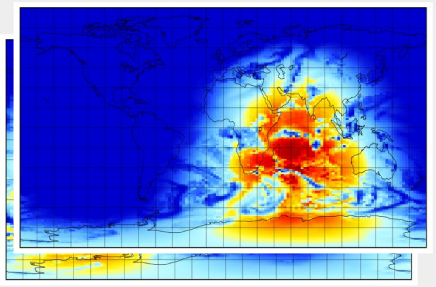
Temperatura horaria NOAA GFS

Objetivo: mejorar la resolución temporal

- La tarea es la interpolación temporal
- Pero no tenemos datos horarios para entrenar

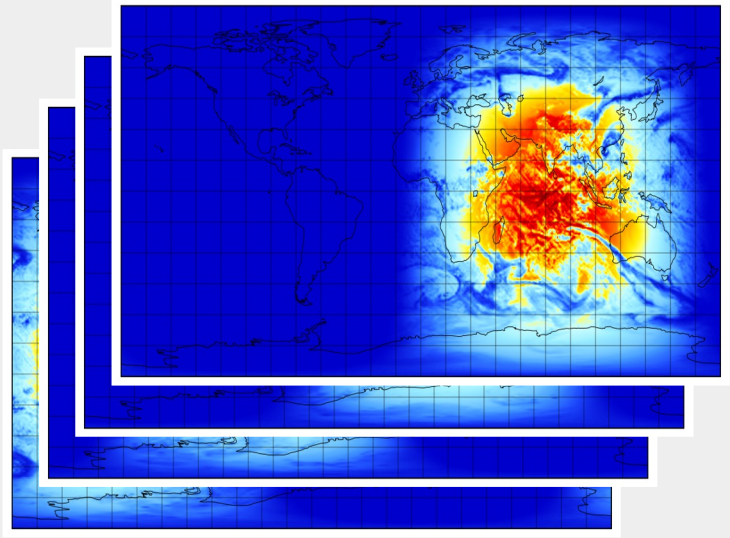
se necesitan técnicas no supervisadas (**autosupervisadas**).

CMIP6-IPSL



Trihoraria

?



Horaria

El autosupervisado en resumen

- Usar tareas pretexto/sustitutas de forma supervisada.
  - Etiquetas **generadas** automáticamente.

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

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

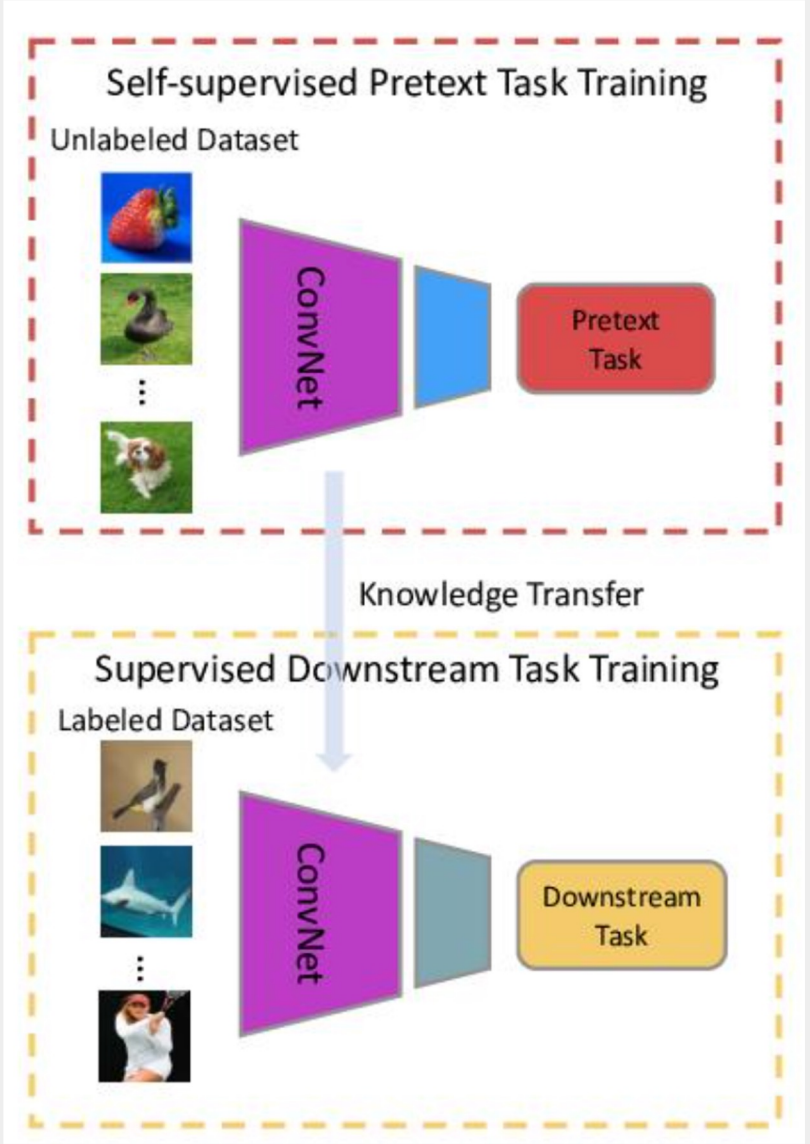
Aprendizaje supervisado

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

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

Aprendizaje autosupervisado

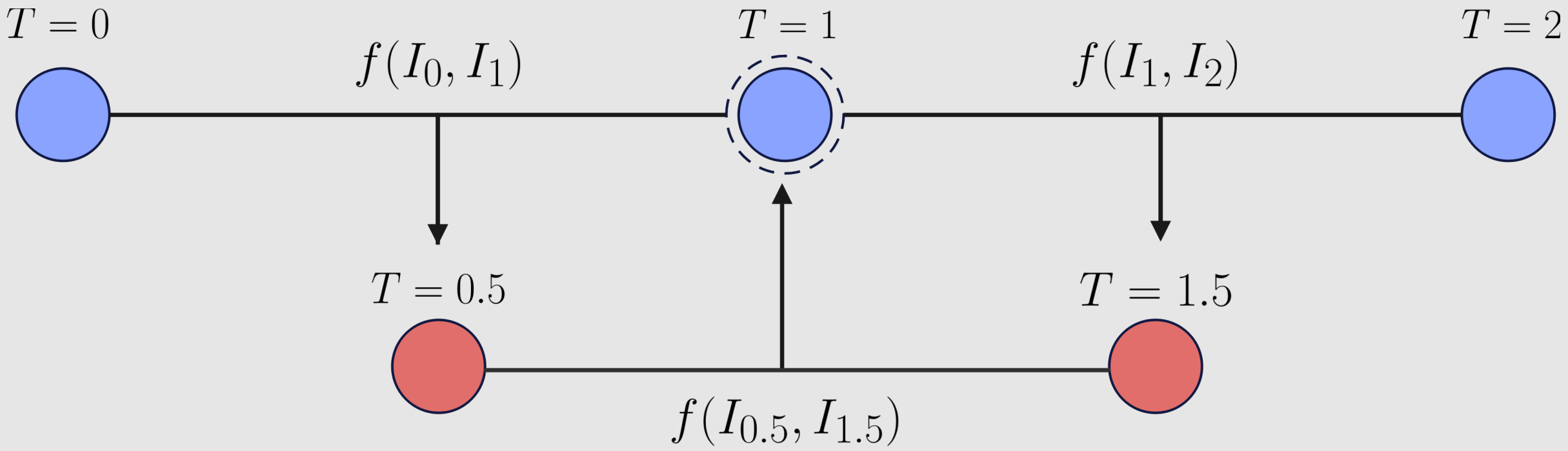
- Tarea posterior: recuperación, clasificación supervisada o semisupervisada, etc.



Una tarea pretexto para la interpolación temporal

STINT: interpolación temporal autosupervisada para datos geoespaciales

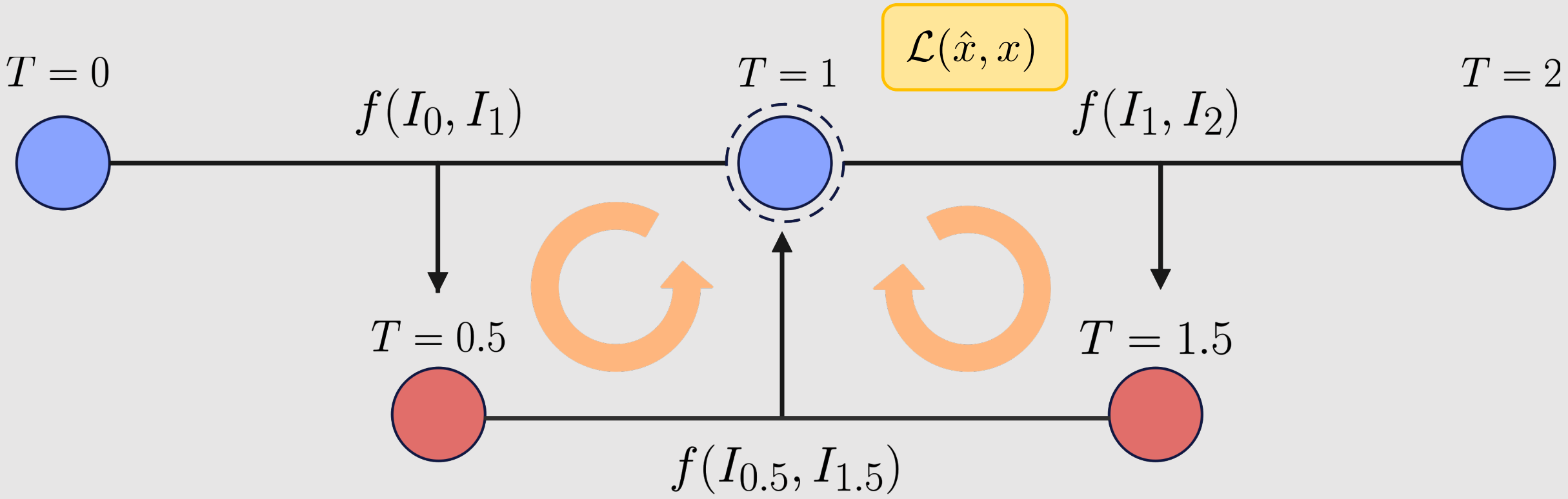
[Harilal, Hodge, Subramanian, & Monteleoni, Climate Informatics 2026]



Una tarea pretexto para la interpolación temporal

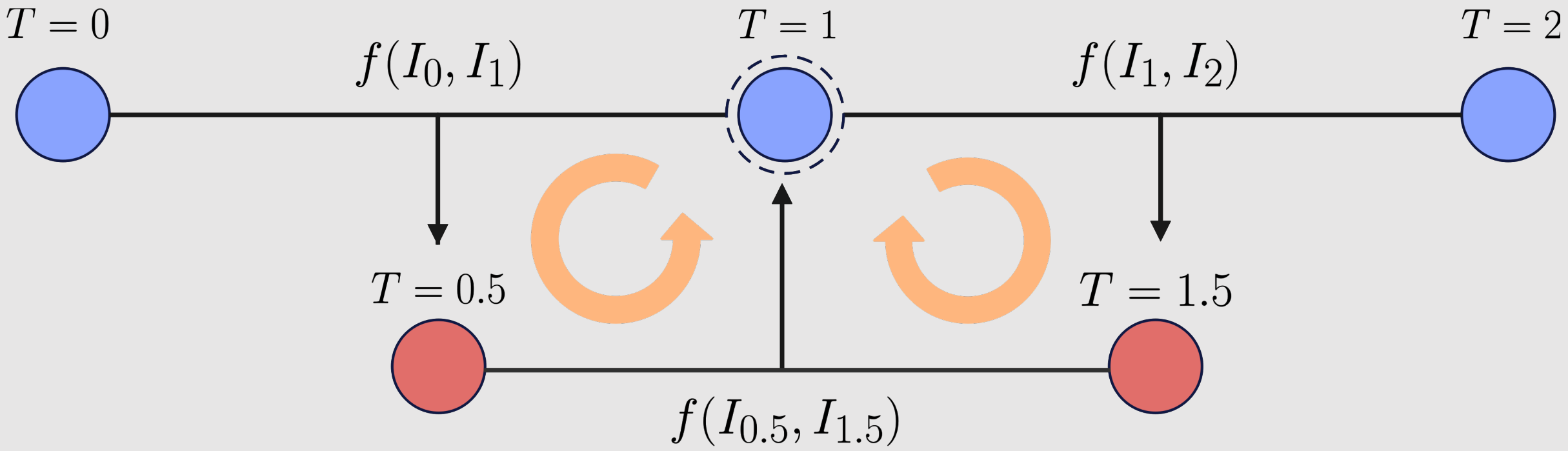
STINT: interpolación temporal autosupervisada para datos geoespaciales

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



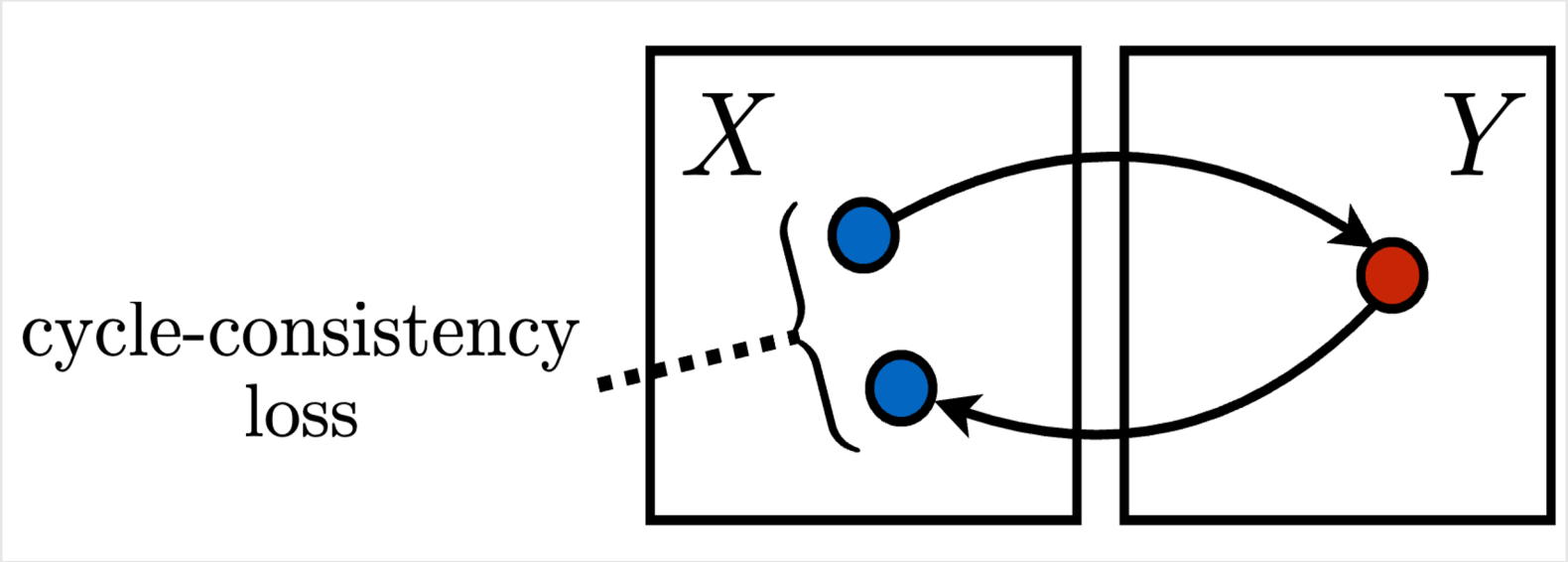
Una tarea pretexto para la interpolación temporal

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

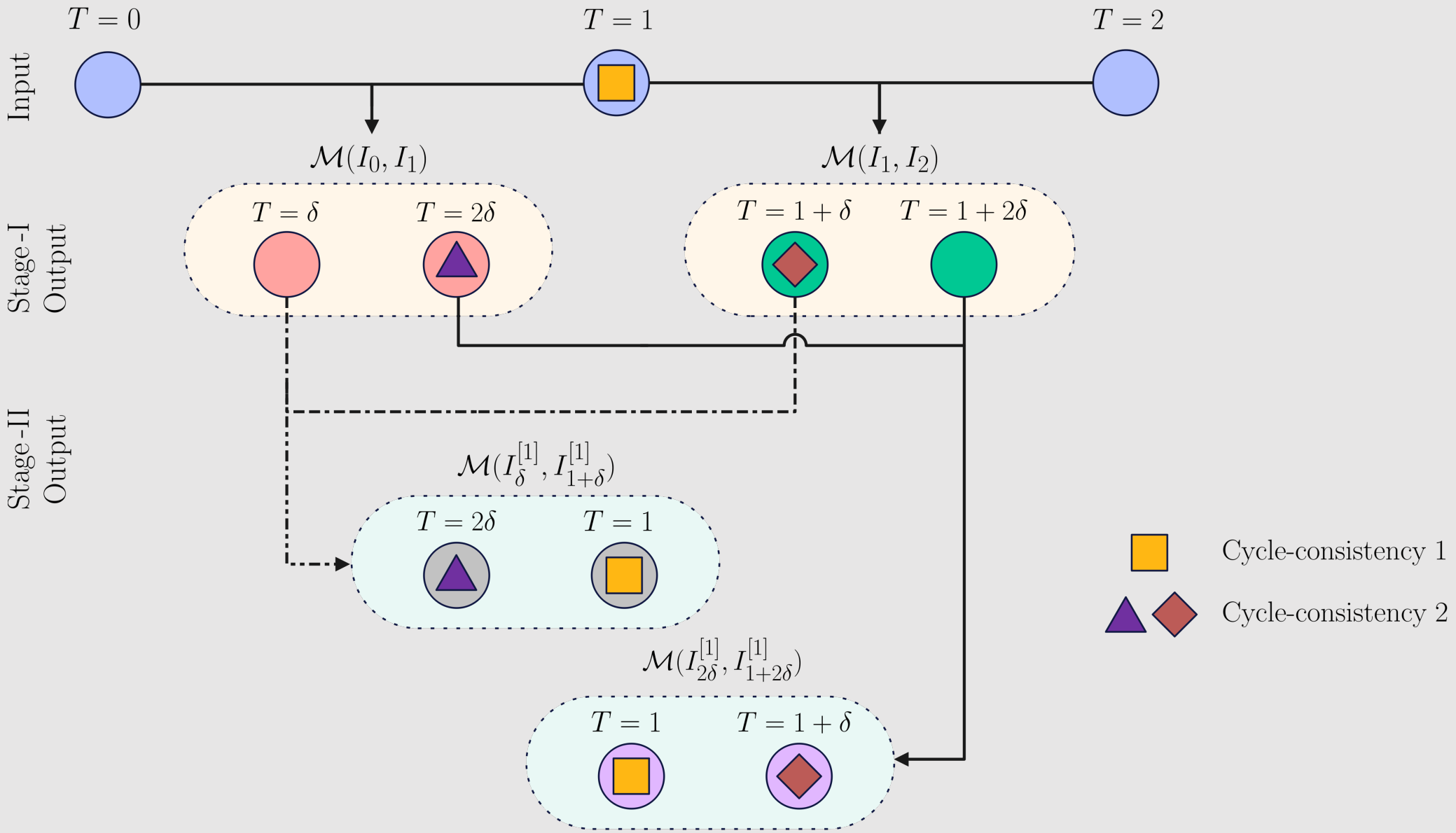


Pérdida de consistencia cíclica

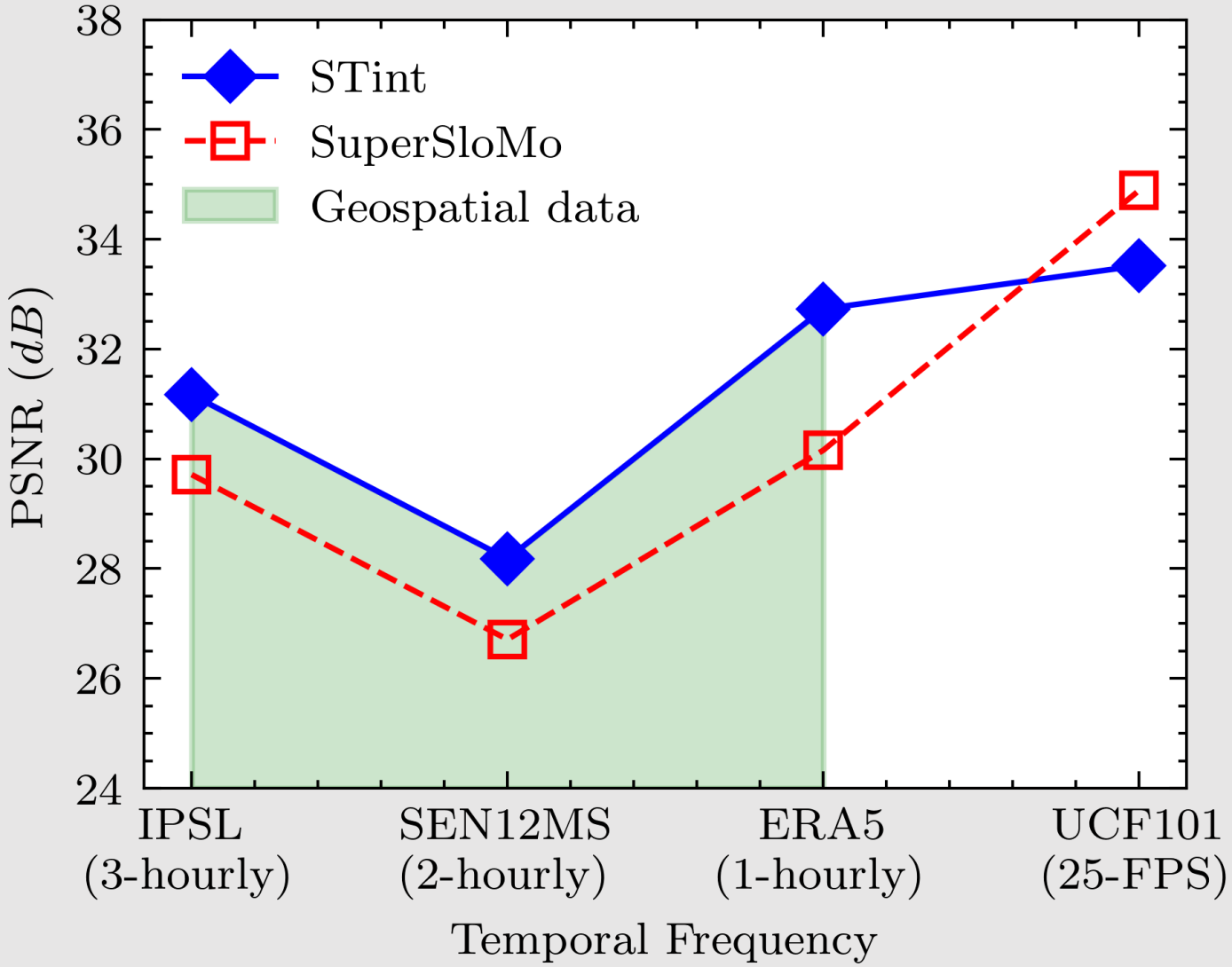
$$\mathcal{L}(\hat{x}, x) = \sum_{i=1}^{n} \mathcal{L}(T_i, \hat{T}_i) \qquad \hat{T}_i := f(f(T_{i-1}, T_i), f(T_i, T_{i+1}))$$



STINT: interpolación temporal autosupervisada



STINT: interpolación temporal autosupervisada



	$\frac{MSE}{Capacity}$ (↓)	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**

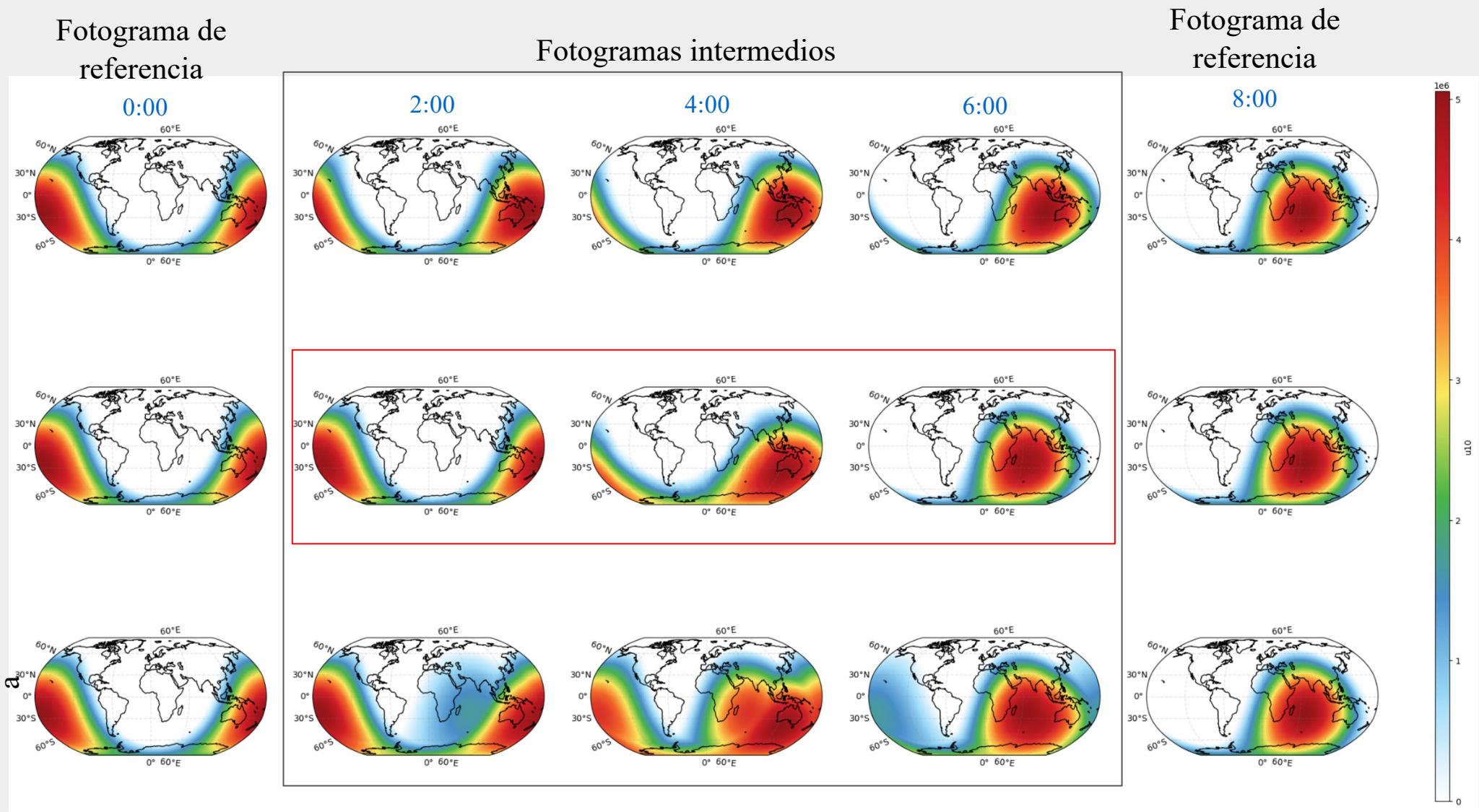


Figura. *Modelo entrenado con campos de irradiancia solar ERA5 e interpolación (3 fotogramas no vistos) entre las 0:00 y las 8:00.*

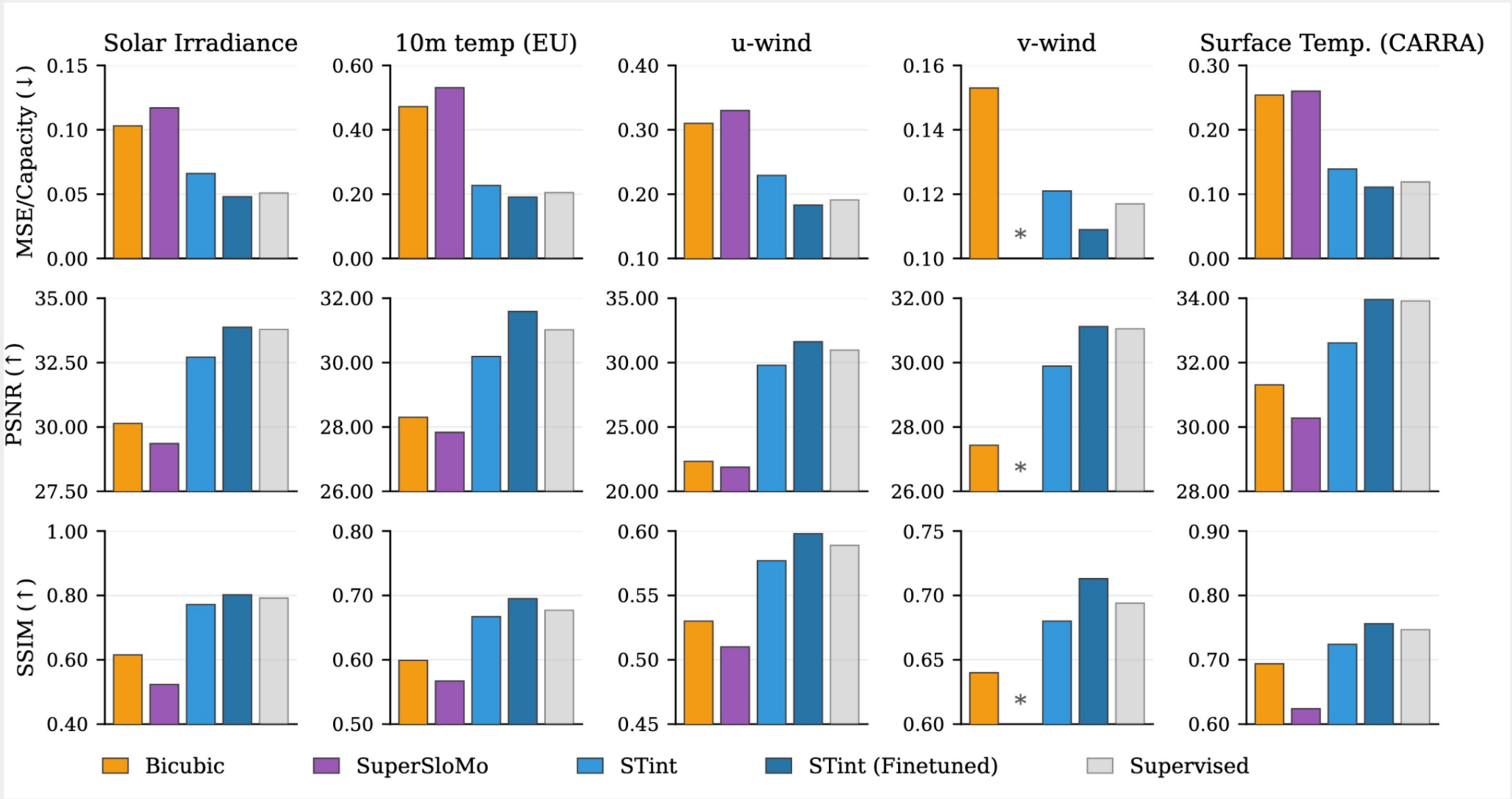
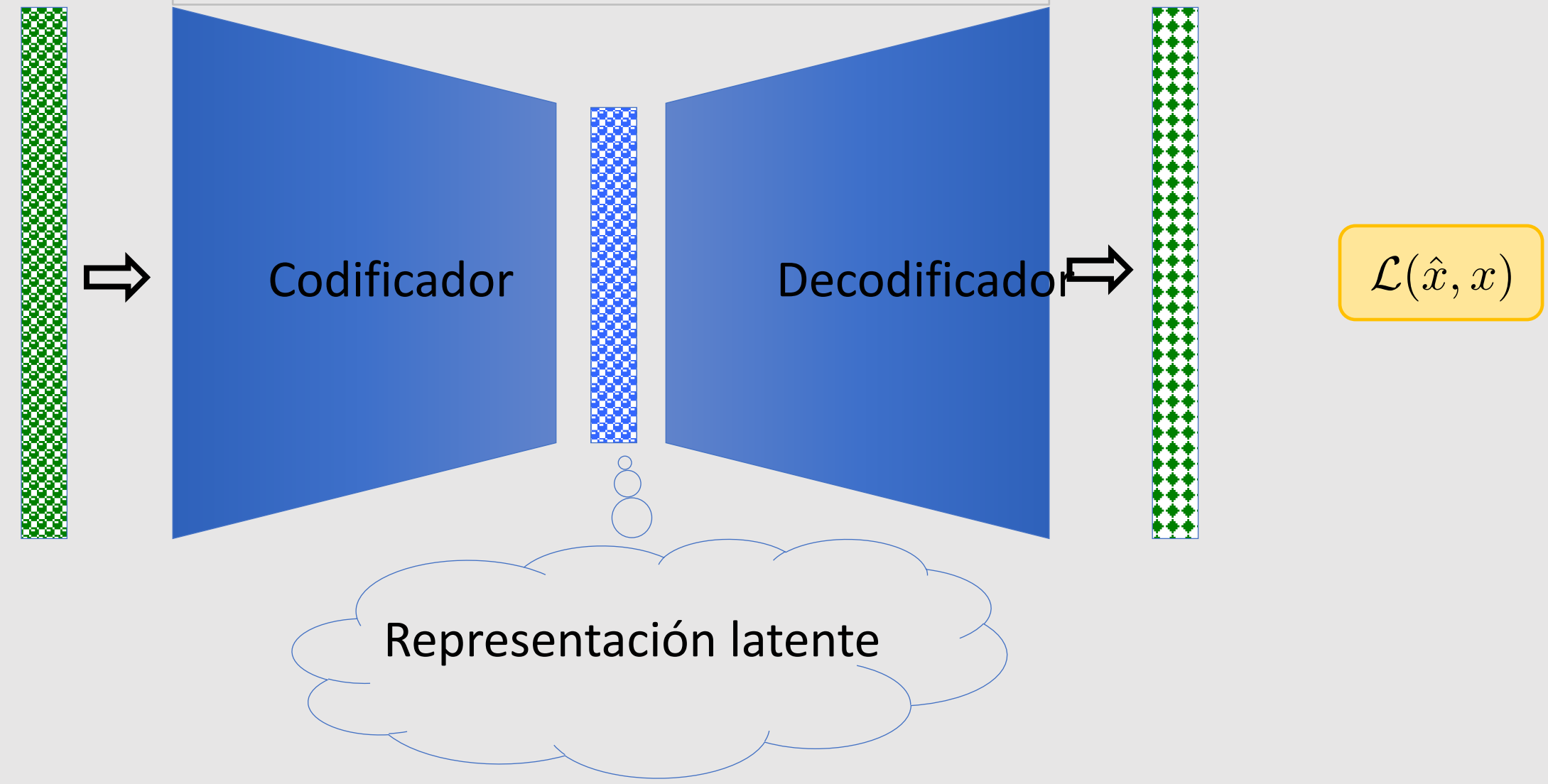


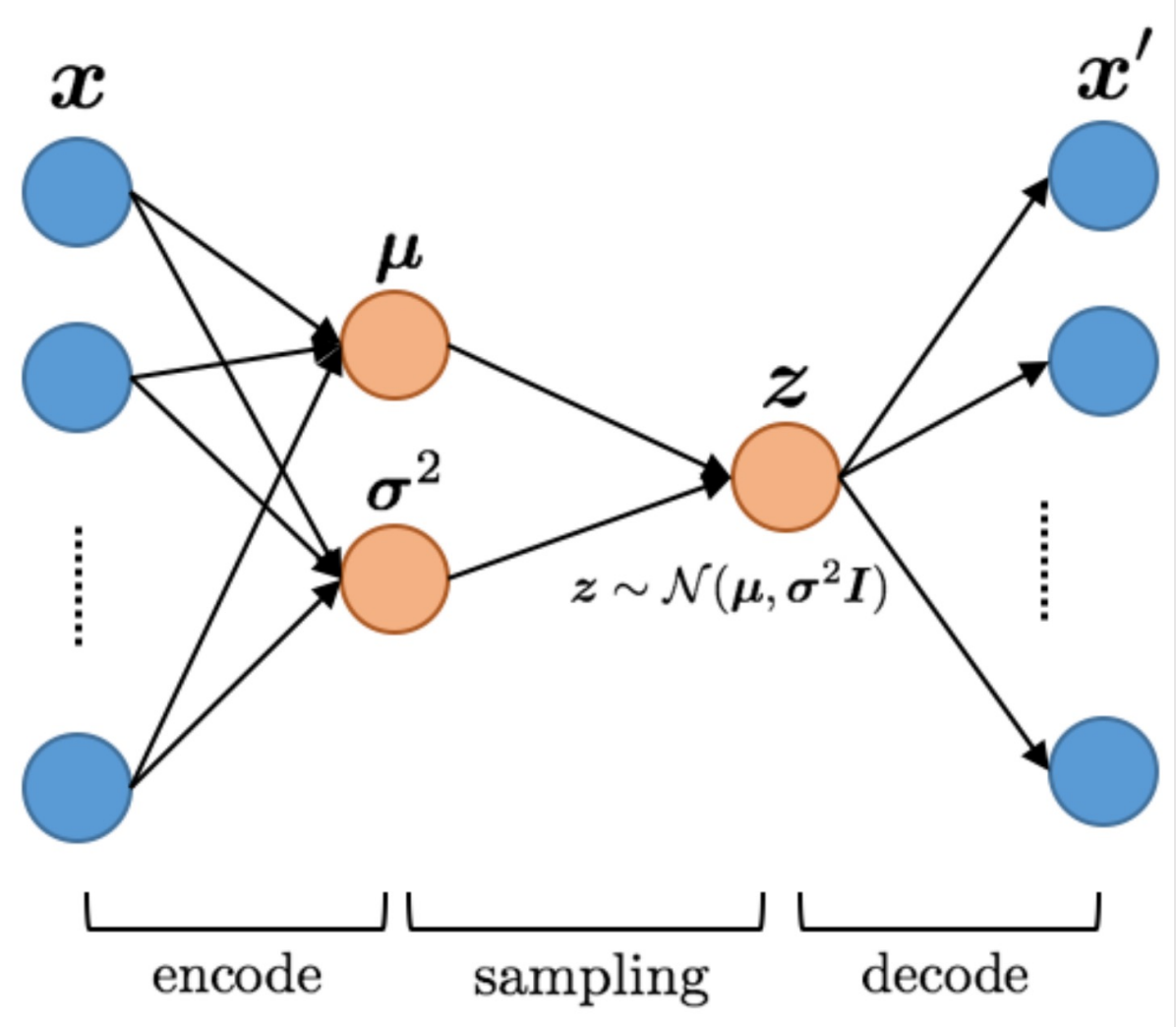
Figura. *Resultados de interpolación temporal en distintos conjuntos de datos geoespaciales.*

Autoencoder: los parámetros de las redes codificadora y decodificadora se entrenan para que la salida se aproxime a la entrada. Tras entrenar con muchos ejemplos, los parámetros de la capa cuello de botella forman una representación compacta de la distribución de entrada.



Autoencoder variacional (VAE)

Aprender una distribución sobre representaciones latentes, en lugar de una única codificación



Aprender un autoencoder variacional (VAE)

- El codificador es una distribución, q, que lleva el espacio de datos al espacio latente
  - Definido por los pesos de la red θ:

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

- El decodificador es una distribución, p, que lleva el espacio latente al espacio de datos
  - Definido por los pesos de la red ϕ:

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

- Ajustar los pesos θ y ϕ minimizando:

Error de reconstrucción en los datos de entrenamiento

Término de regularización: penaliza la distancia a un prior simple

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

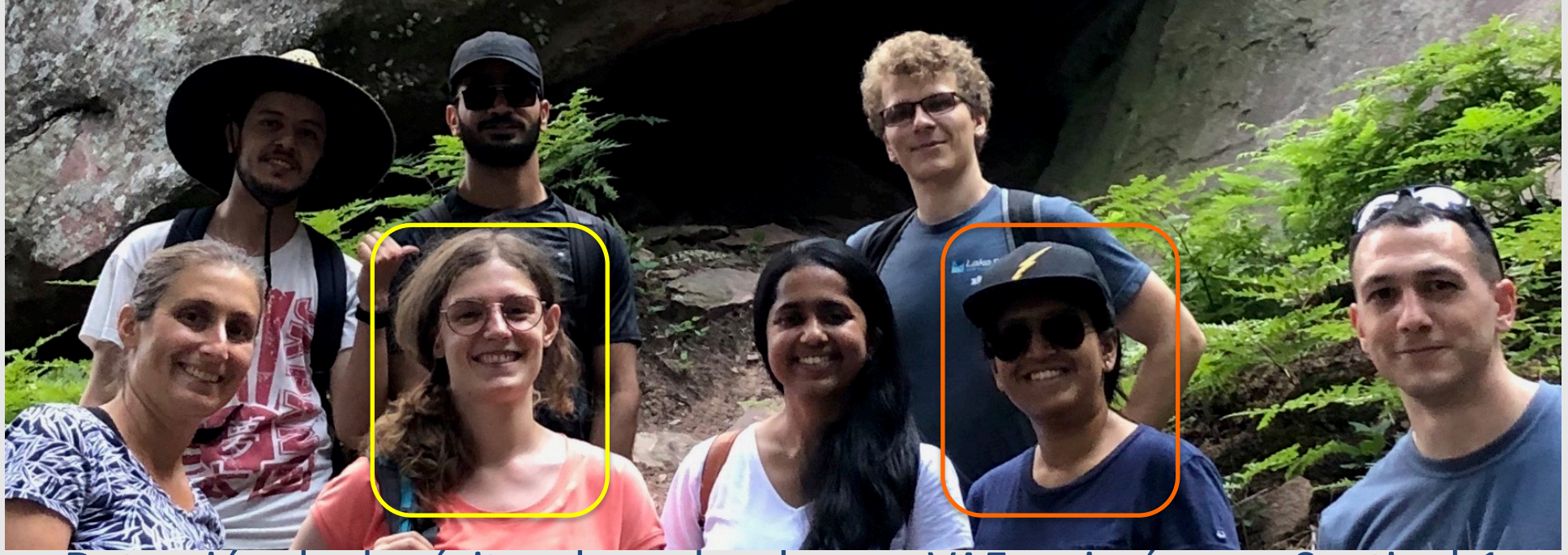
- Optimizado con una variante de esperanza-maximización (EM), con una cota inferior variacional

Ejemplo: VAE para la detección semisupervisada de anomalías aplicada a la detección de avalanchas

Detección de avalanchas



Fuente de la imagen: https://www.nytimes.com/2017/01/19/world/europe/italy-avalanche.html



Detección de depósitos de avalancha con VAE en imágenes Sentinel-1

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

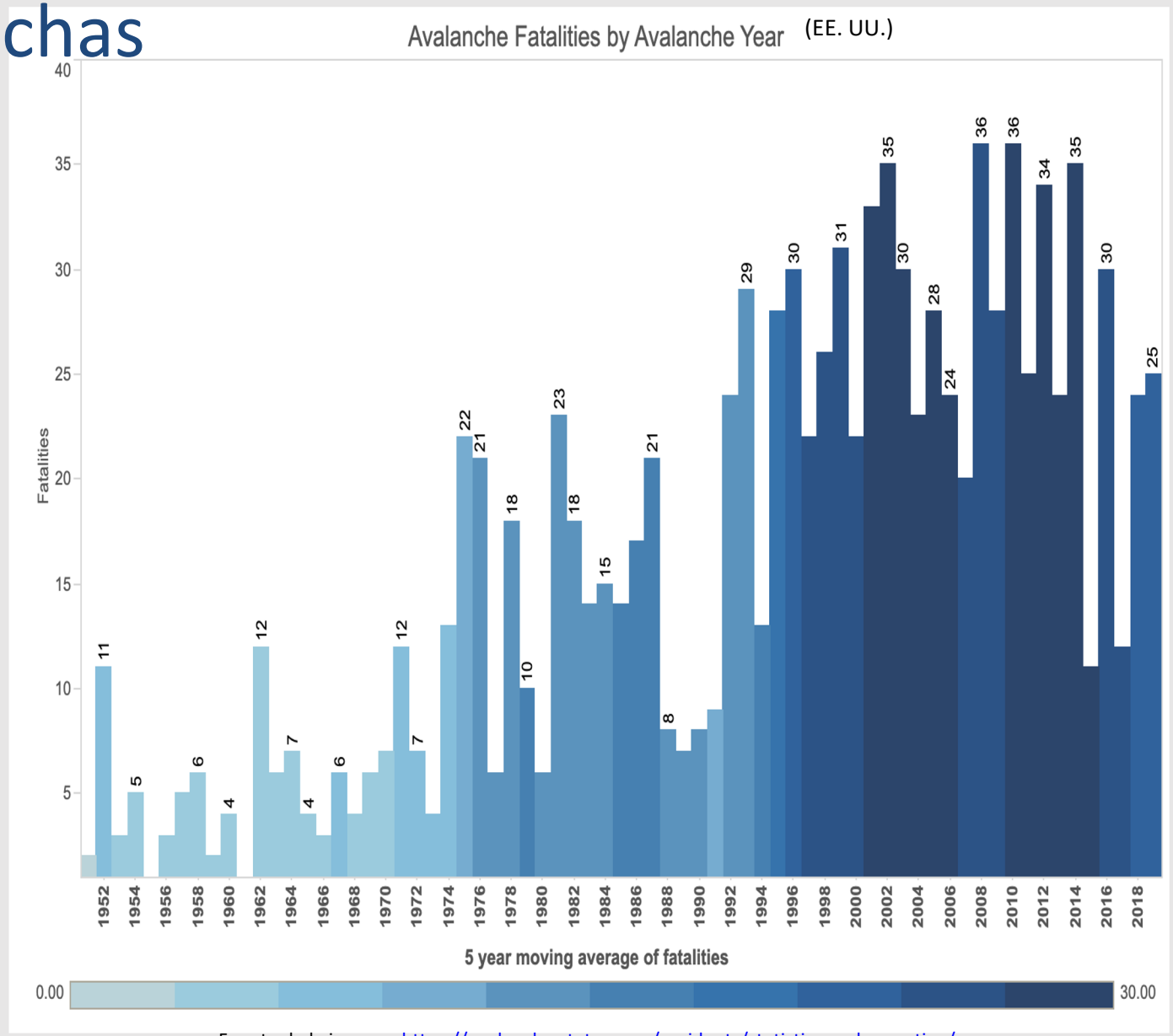
Variational autoencoder anomaly-detection of avalanche deposits in satellite SAR imagery. *Proc. 10th International Conference on Climate Informatics (CI) 2020*

Boulder

*Coautoría principal

METEO FRANCE

Muertes por avalancha según el año de avalanchas



Fuente de la imagen: https://avalanche.state.co.us/accidents/statistics-and-reporting/

Impactos del cambio climático



Fuente de la imagen: https://www.independent.co.uk/news/world/europe/mont-blanc-deaths-french-alps-avalanche-skiing-a9227361.html

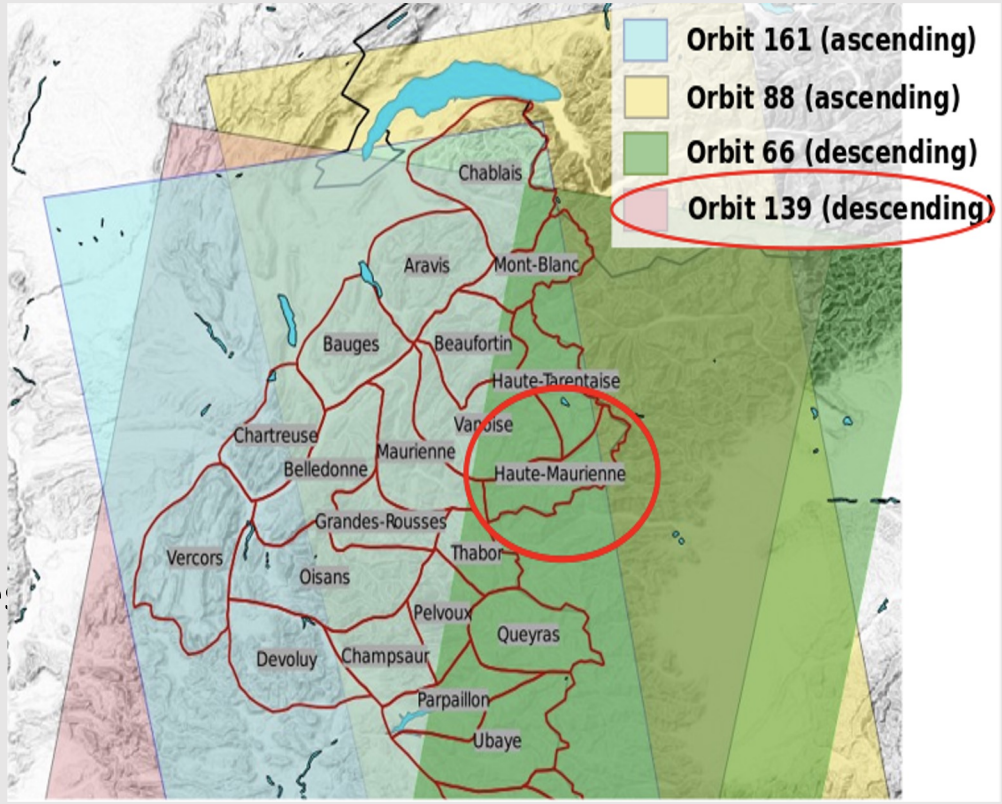
Retos para el aprendizaje automático

- Fuerte desequilibrio de clases
  - Las avalanchas son eventos raros
- Datos de referencia etiquetados difíciles de obtener
  - Accesibilidad del terreno
  - Condiciones meteorológicas
  - Peligro de avalanchas

El problema de la detección de avalanchas

- **Medidas de referencia** in situ limitadas
  - Météo-France

- **Imágenes SAR** no etiquetadas
  - Seguimiento de los Alpes franceses 2017-2018
  - Satélites Sentinel-1A y 1B
  - 4 características:
    - Coeficientes de retrodispersión en el momento actual y el anterior
    - Características topológicas: pendiente y orientación



METEO FRANCE

Nuestro enfoque

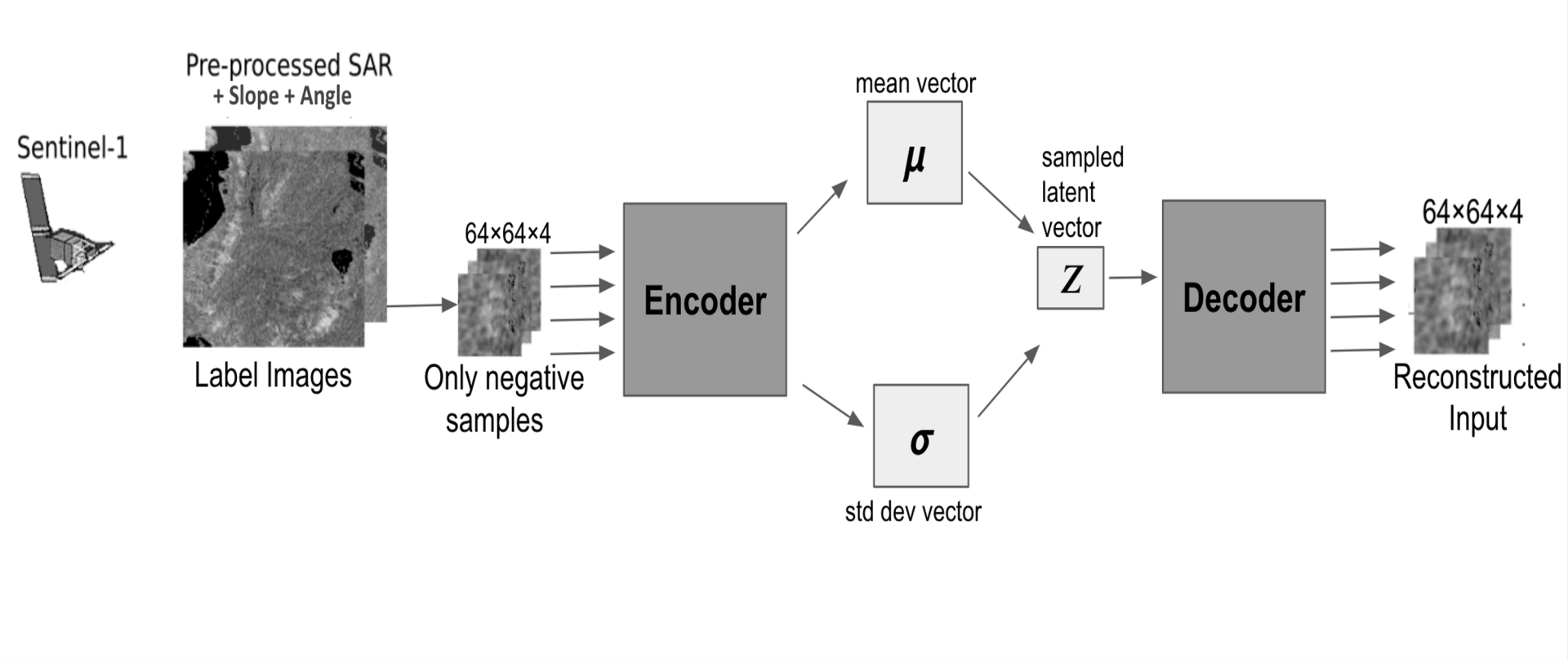
① Tratar una avalancha como un evento raro, o una anomalía

② Entrenar un autoencoder variacional (VAE) con los ejemplos negativos

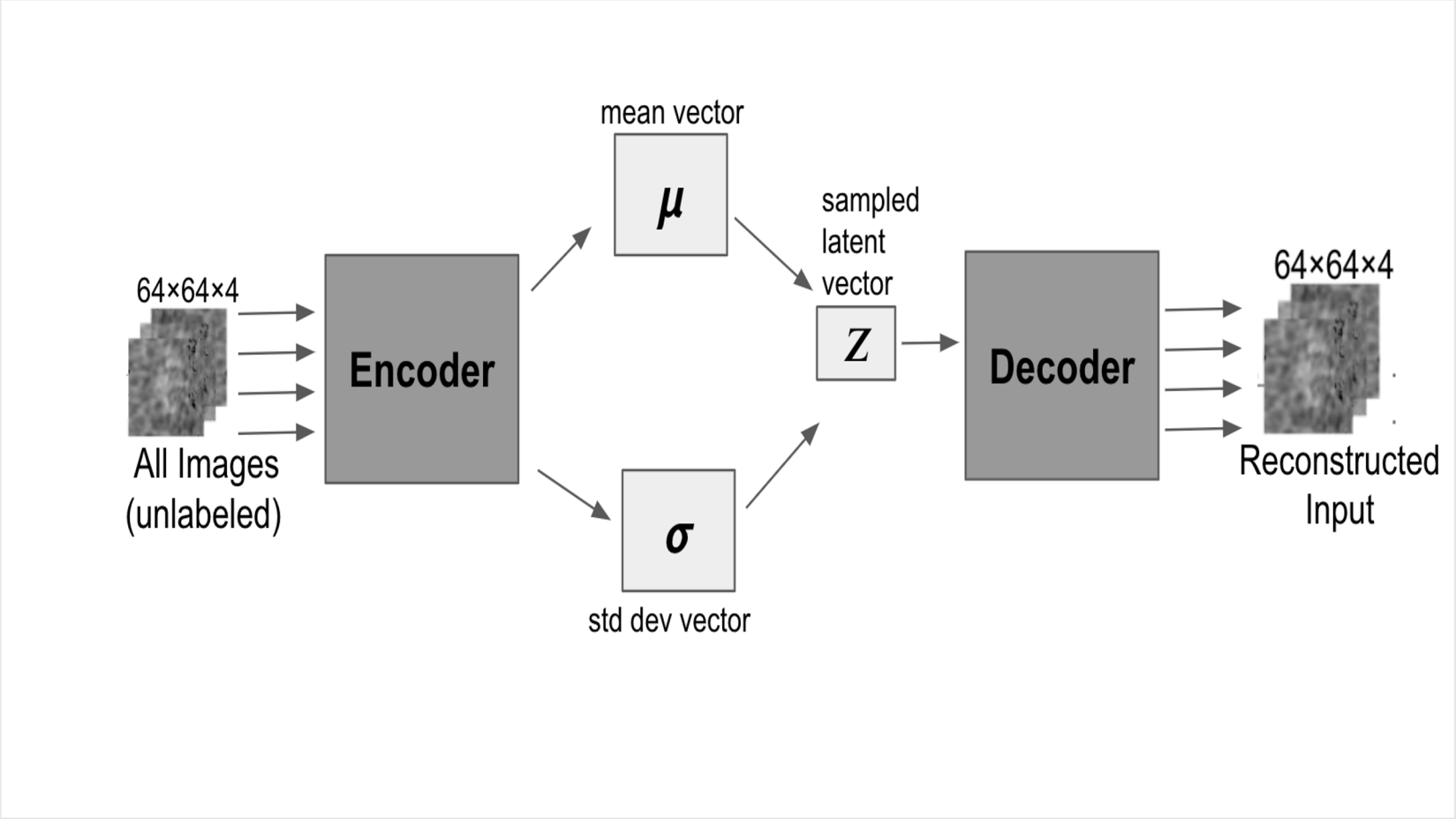
③ Umbralizar el error de reconstrucción del VAE para clasificar una nueva imagen

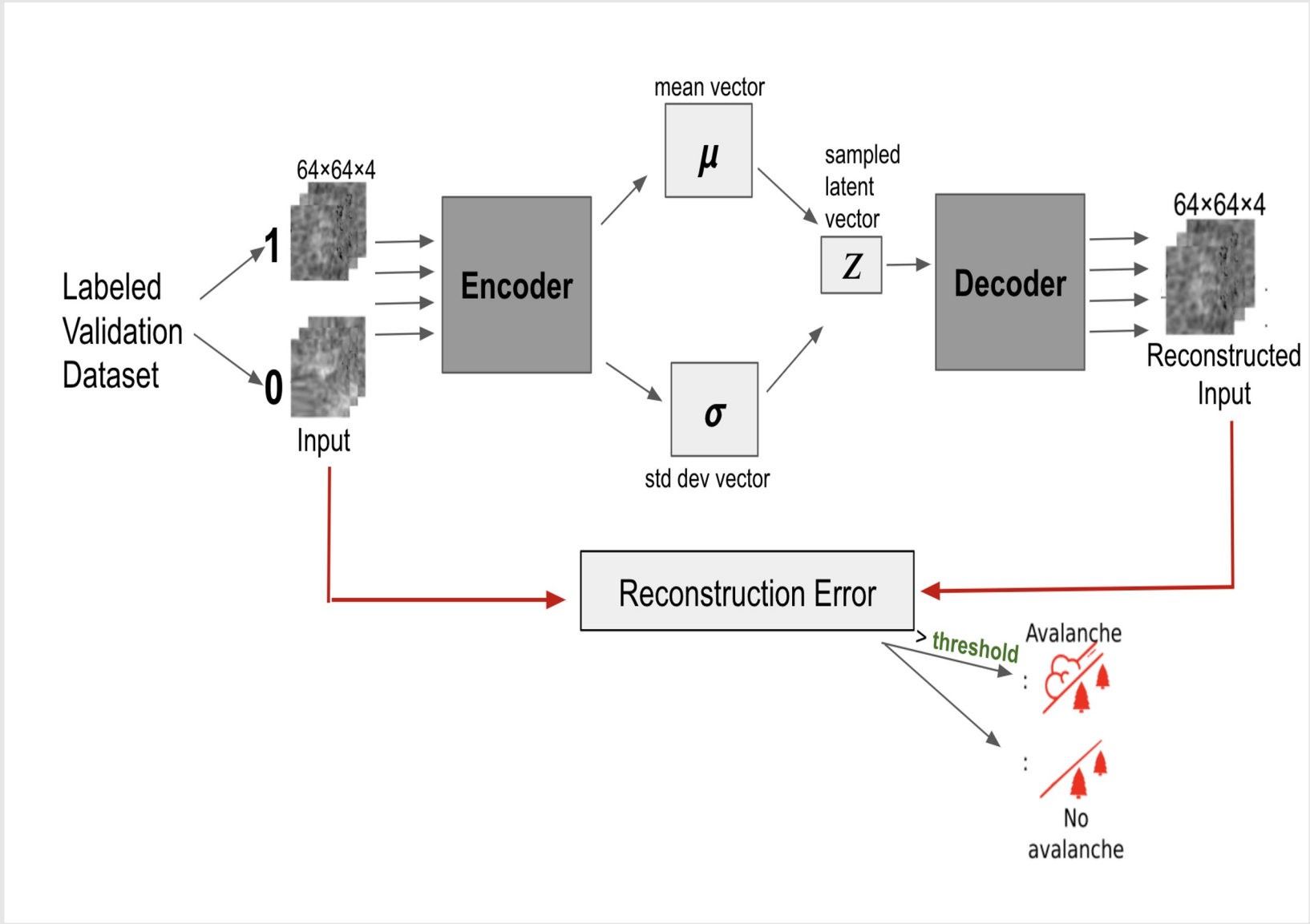
- Cuando los datos etiquetados escasean, el VAE puede entrenarse **sin** supervisión!

VAE entrenado con ejemplos negativos

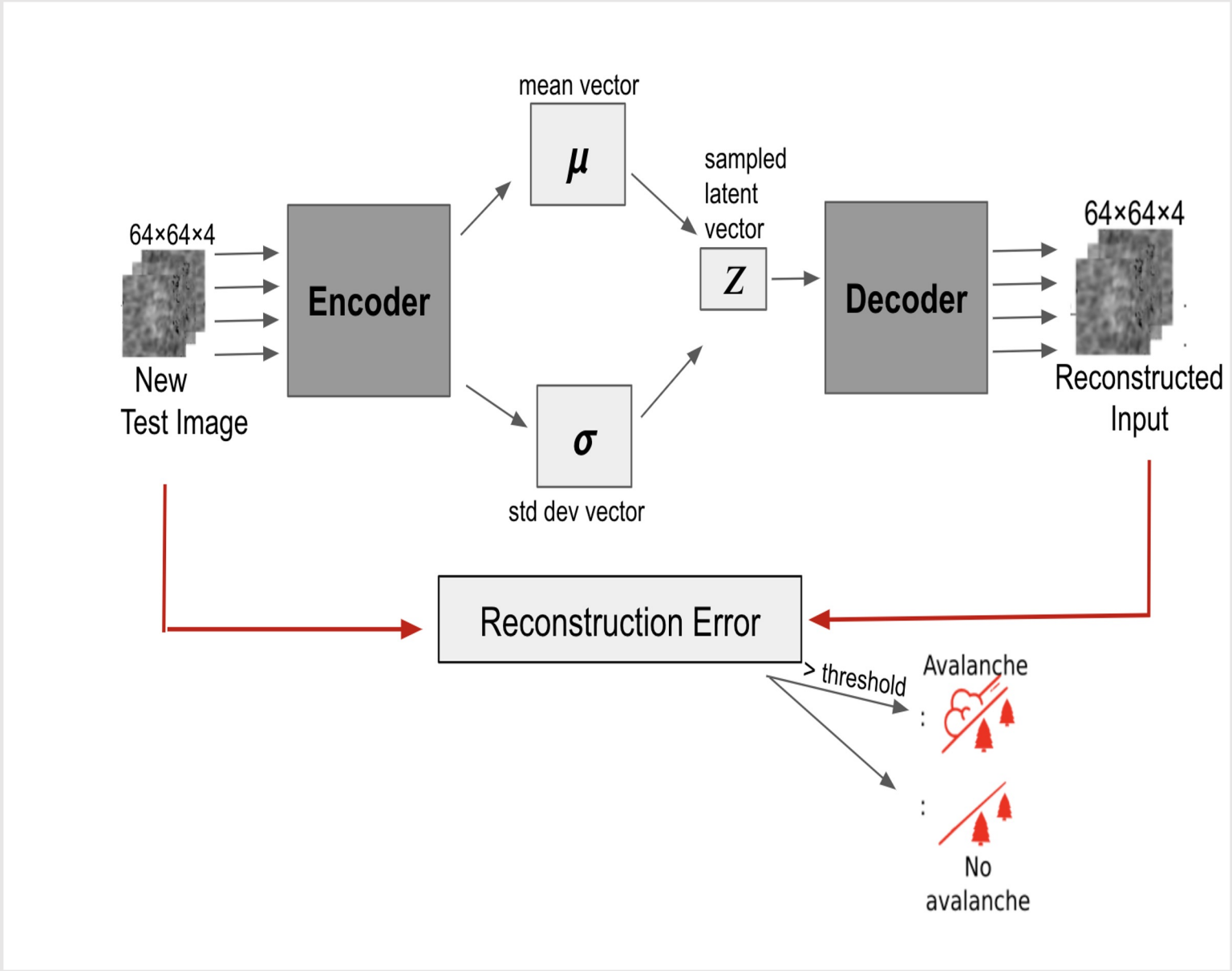


VAE entrenado con TODOS los ejemplos



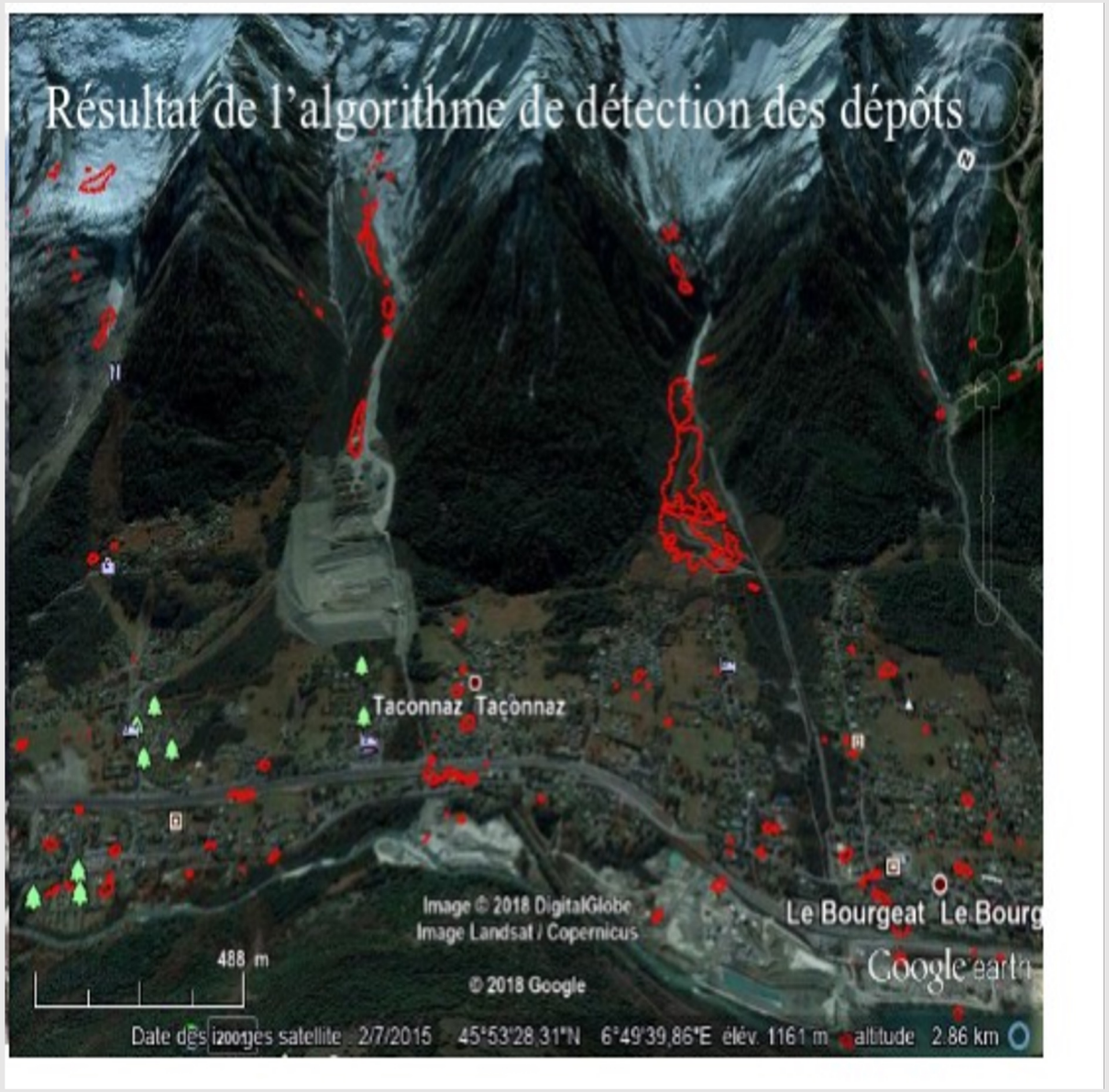


Ajuste del umbral para la detección de anomalías



Cadena de detección de anomalías

Método de referencia: umbralización



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

Evaluación

Uno de los macizos más expuestos (de los 18 de «Todos los Alpes»)

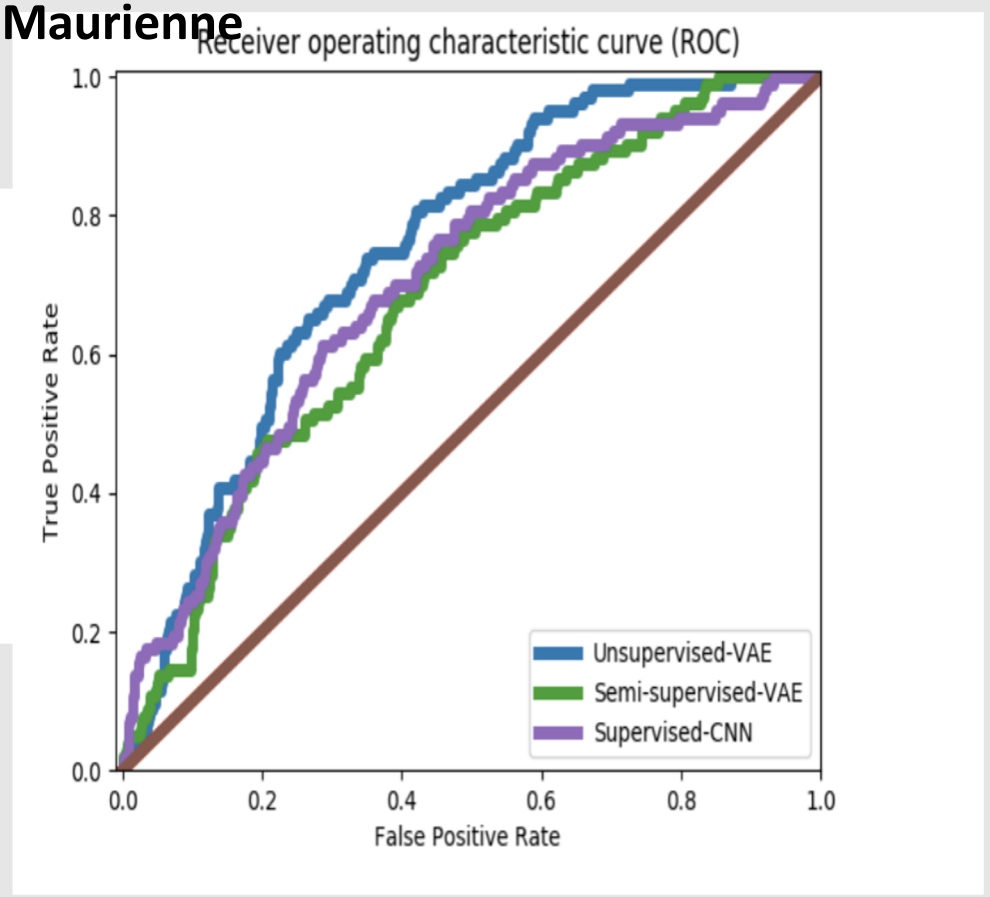
	All Alps		Haute Maurienne	
	Balanced Accuracy	F1-score	Balanced Accuracy	F1-score
Baseline	0.58	0.05	0.58	0.12
Supervised - CNN	0.53	0.10	0.53	0.12
Semi-supervised - VAE	0.59	0.11	0.6	0.23
Unsupervised - VAE	0.69	0.14	0.68	0.26

- Conjunto de prueba reservado: 6498 ejemplos etiquetados
- Método de referencia de la literatura sobre detección de avalanchas: umbralización [Karbou et al., ISSW 2018]
- Método de referencia de aprendizaje supervisado: red neuronal convolucional (CNN) entrenada con un conjunto equilibrado artificialmente [Sinha et al., Climate Informatics 2019]

Evaluación

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

Curvas ROC para la región de Haute-Maurienne



Resumen

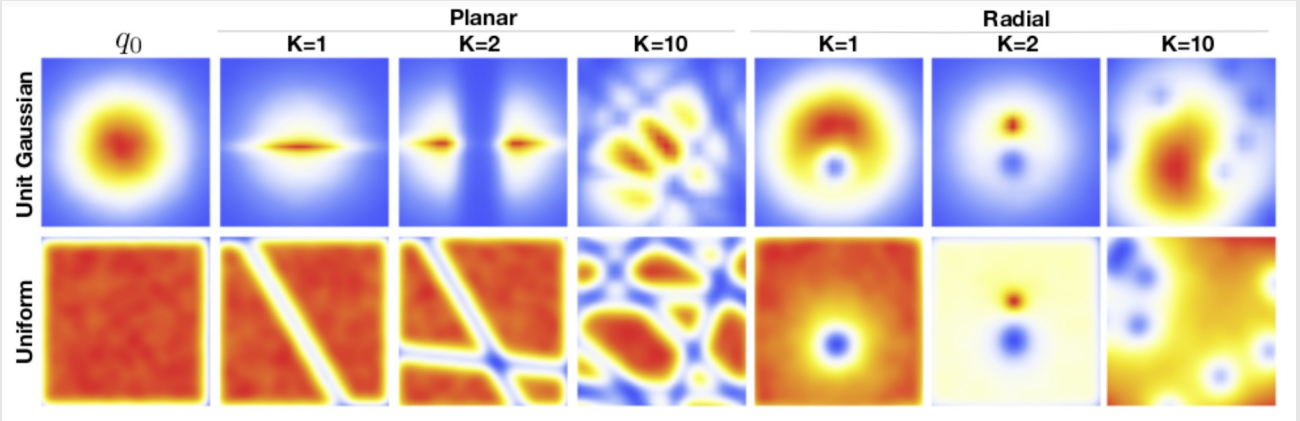
- Proponemos un enfoque semisupervisado para detectar eventos raros cuando los datos etiquetados son limitados
  - Esta cadena semisupervisada se basa en gran medida en el aprendizaje no supervisado
- Puede verse como una forma de sensor virtual

Parte III: Métodos de IA:
IA generativa

IA generativa: una distinción

- Probablemente estéis acostumbrados a que el término «IA generativa» se refiera a los grandes chatbots y agentes comerciales como ChatGPT, Gemini o Claude.
  - Para esos sistemas, los modelos de IA generativa se entrenaron con vídeo y texto extraídos de internet
  - ¡Pero esa **no es la ÚNICA definición** ni el único uso de la IA generativa!
- Nosotros podemos **entrenar modelos de IA generativa desde cero** para nuestros problemas
  - Para problemas de clima y tiempo NO entrenaremos con texto y vídeo, sino con los datos pertinentes, p. ej. campos de valores reales medidos en distintas latitudes, longitudes, momentos y quizá alturas en la atmósfera o profundidades en el océano, como:
    - Temperatura
    - Radiación solar
    - Humedad
- Hoy veremos los tipos de modelos generativos que luego podréis entrenar para vuestros propios problemas

Flujos normalizantes

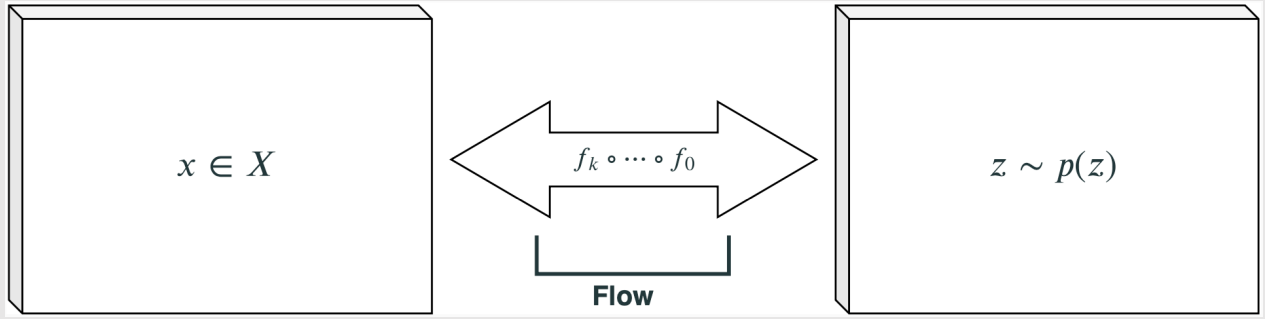


[Rezende & Mohamed, ICML 2015]

Pueden verse como una extensión del VAE más allá de la hipótesis gaussiana en el espacio latente

Aprender una serie de transformaciones invertibles, $\{f_i\}$, a partir de un prior simple en el espacio latente Z, para obtener distribuciones más informativas en el espacio latente:

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

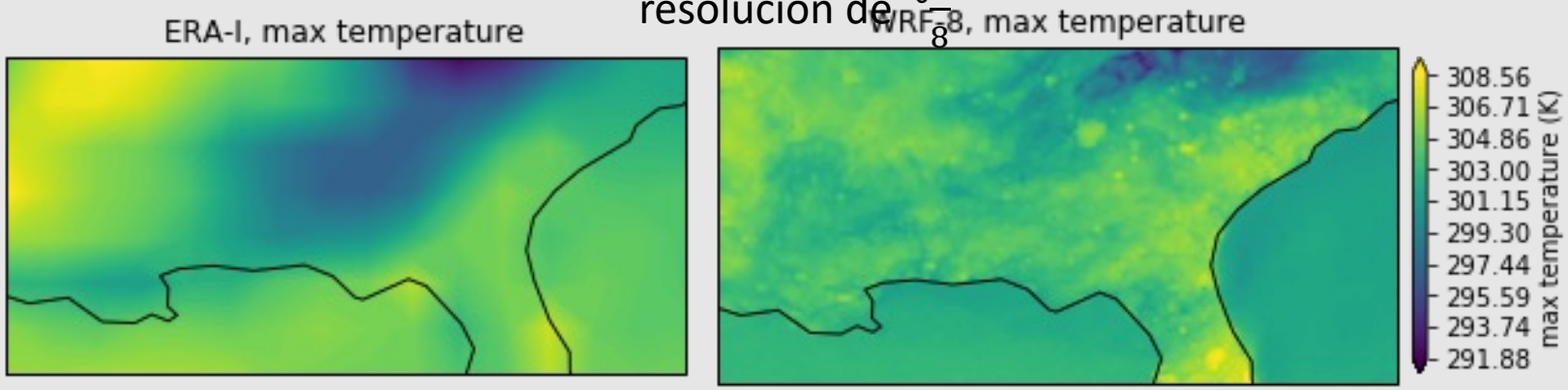


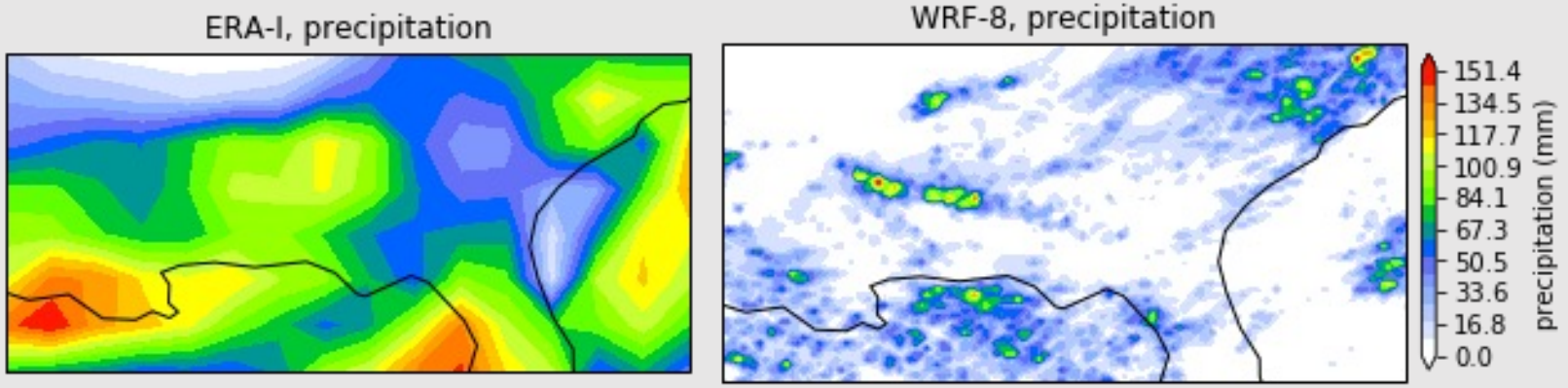
Flujos normalizantes: aplicación a la reducción de escala espacial

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

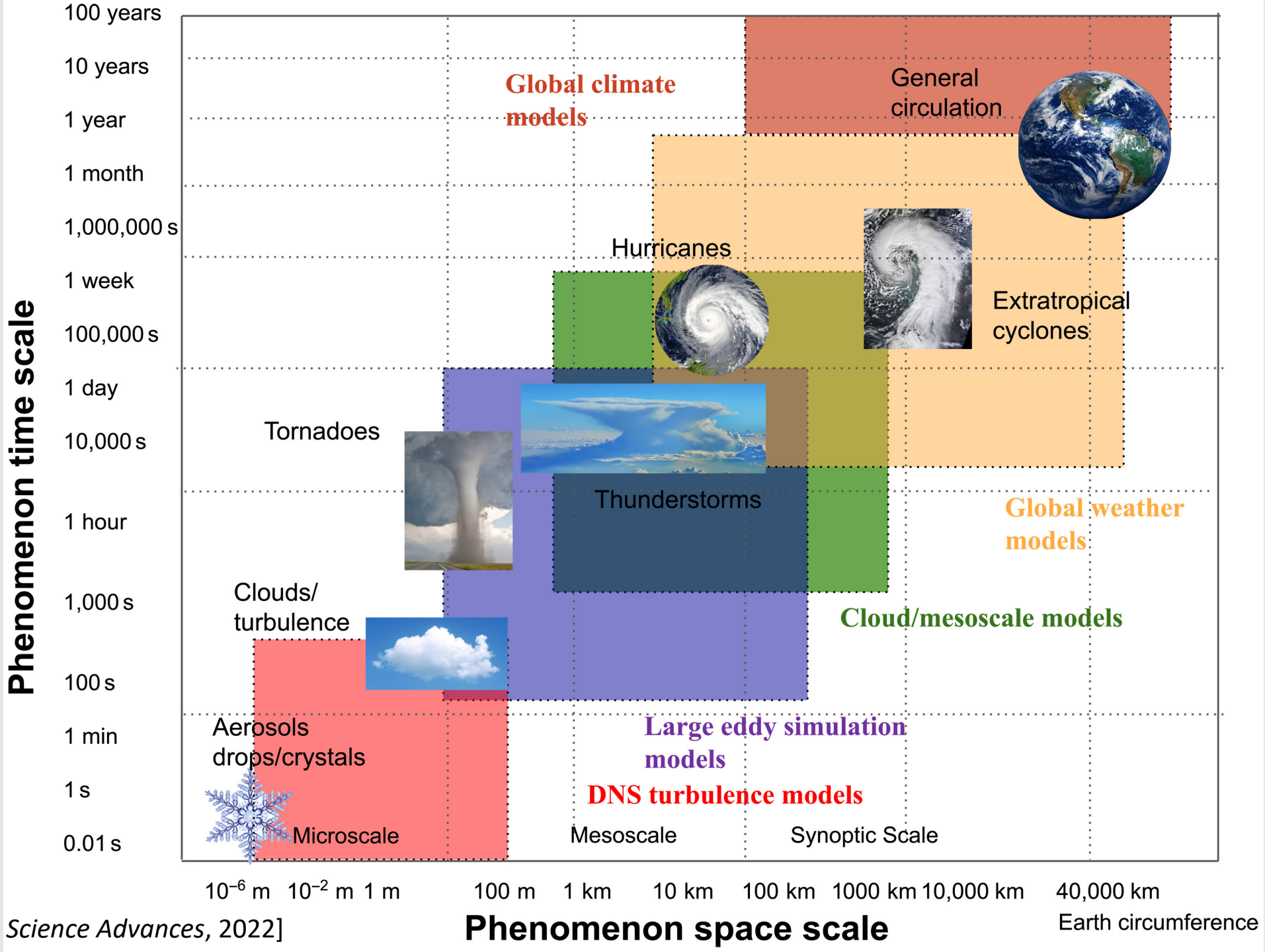
ERA: datos de reanálisis, resolución de 1°;

WRF: predicción de un modelo meteorológico numérico, resolución de $\frac{1}{8}°$









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

La reducción de escala como alineación de dominios

- Tarea de alineación de dominios: dadas variables aleatorias X, Y, aprender una aplicación f: X Y tal que, para todo x_i X e y_i Y,

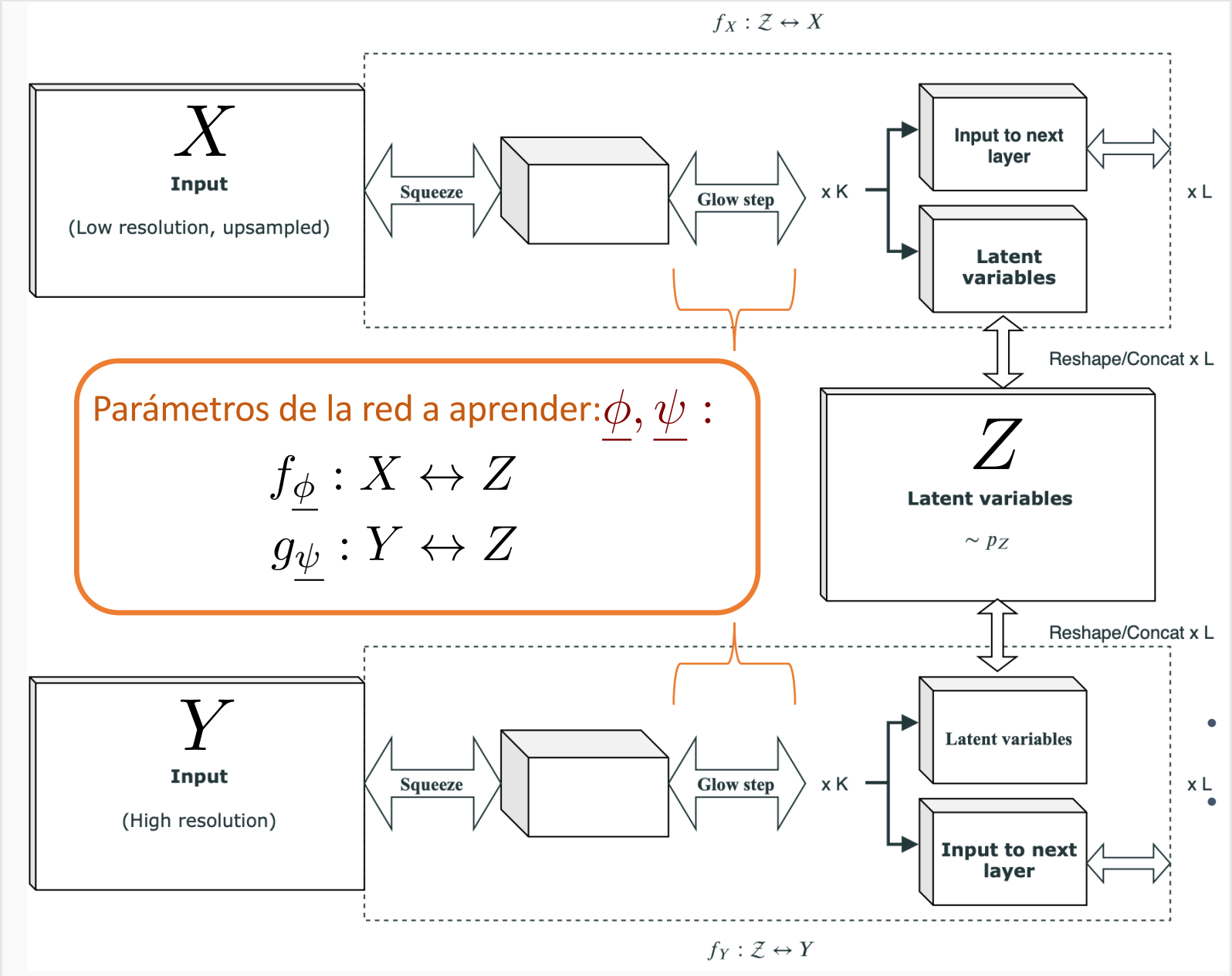
$$f(x_i) \sim P_Y \qquad f^{-1}(y_i) \sim P_X$$

- La reducción de escala como alineación de dominios
  - Dadas muestras i.i.d. a baja resolución (X) y alta resolución (Y)
  - Aprender la densidad conjunta de X e Y suponiendo independencia condicional a través de un espacio latente común 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 $P(x|z), P(y|z)$ con AlignFlow [Grover et al. 2020]
    - Partiendo de un prior simple sobre P_Z, aprender flujos normalizantes
    - ¡No hace falta emparejar ejemplos x e y!

Arquitectura de ClimAlign



- La arquitectura sigue a AlignFlow [Grover et al., 2020]
- Flujo normalizante: Glow [Kingma & Dhariwal, 2018]

Comparación con referencias supervisadas

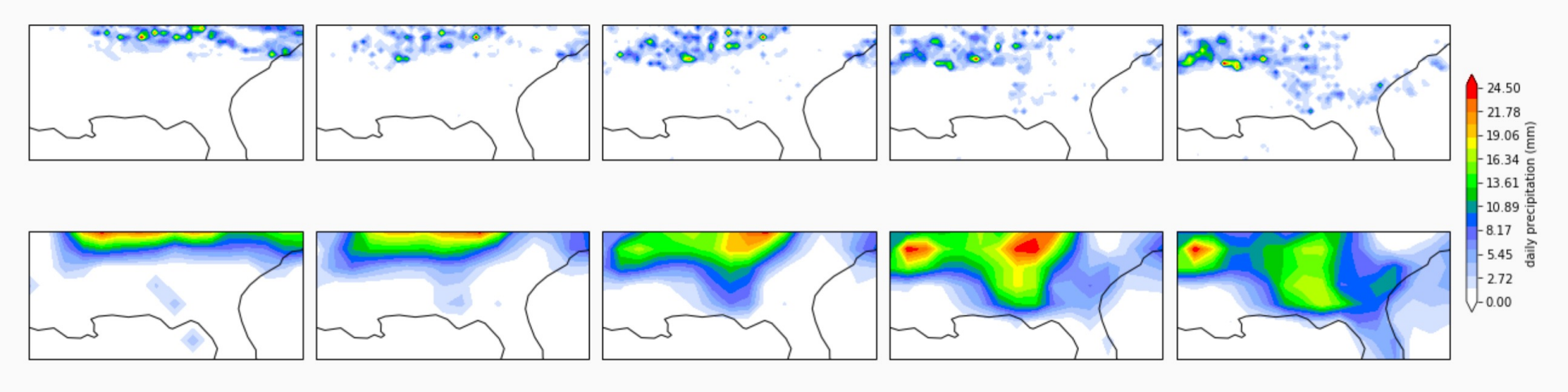
Temperatura máxima diaria

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

Precipitación diaria

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

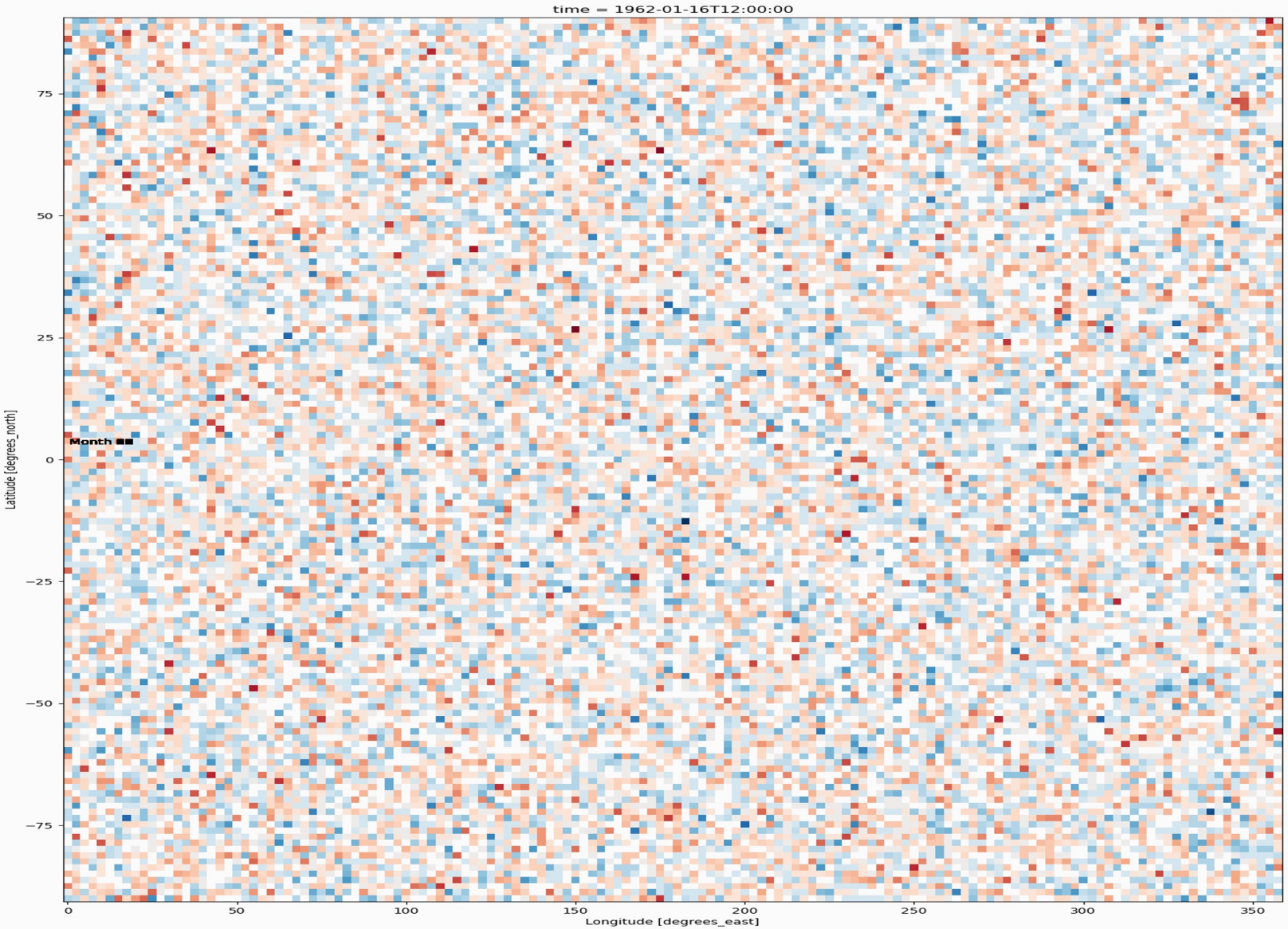
ClimAlign: reducción de escala generativa no supervisada

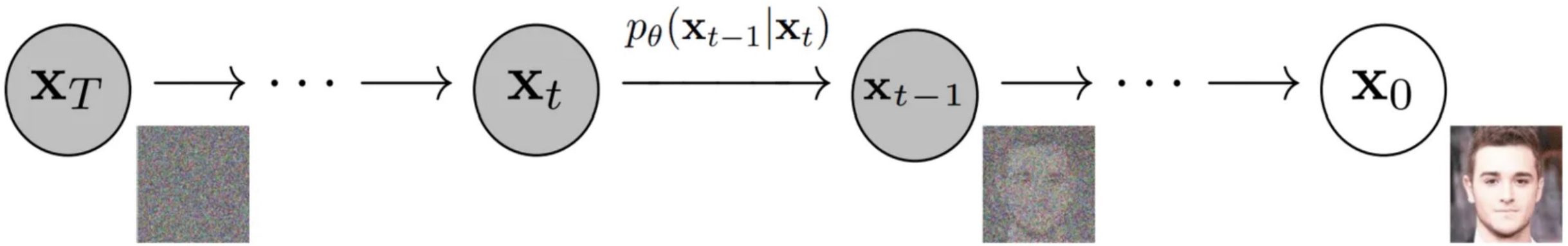


Técnica general de reducción de escala por alineación de dominios con flujos normalizantes [AlignFlow: Grover et al., AAAI 2020][Glow: Kingma & Dhariwal, NeurIPS 2018]

- **No supervisada**: no necesita mapas emparejados a baja y alta resolución
- **Generativa**: puede muestrear la posterior sobre la representación latente O muestrear condicionado a un mapa de baja (¡o alta!) resolución
- **Interpretable**, p. ej. mediante interpolación

Modelos generativos basados en flujos





Entrenamiento por difusión: aprender a generar imágenes realistas a partir de ruido aleatorio

Entrenamiento del modelo de difusión

1. Crear una tarea pretexto a partir de datos de entrenamiento no etiquetados: para cada punto de entrenamiento x_0, crear una secuencia de imágenes sintéticas añadiendo ruido gaussiano a la imagen anterior, hasta llegar a una imagen indistinguible del ruido gaussiano.

2. Hacer luego aprendizaje «supervisado» sobre los pares (x_{t-1}, x_t) para aprender $p_{\boldsymbol{\theta}}(x_{t-1} \mid x_t)$

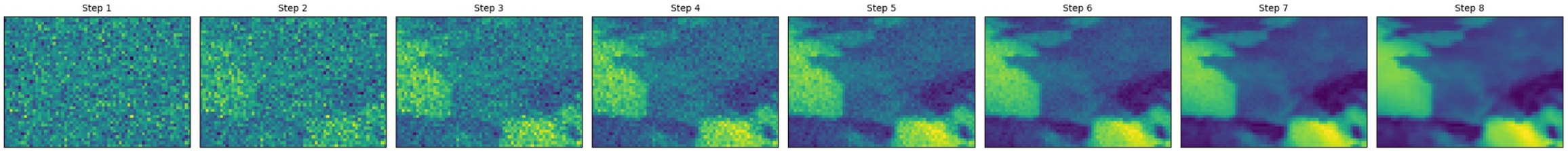
Inferencia, es decir, cómo generar muestras parecidas a los datos de entrenamiento

1. Muestrear x_T al azar (gaussiano).

2. Muestrear T veces de $p_{\boldsymbol{\theta}}$, condicionando cada vez en la muestra anterior

Modelos generativos basados en flujos

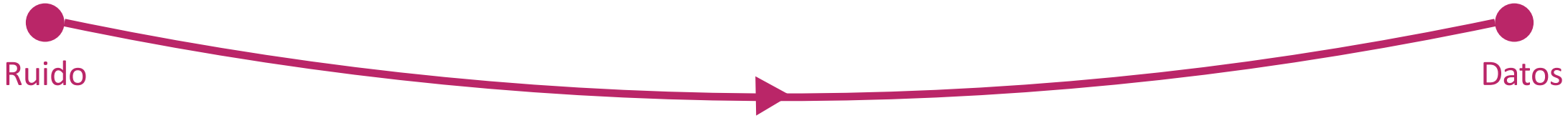
Modelo generativo basado en flujos: aprende un transporte (con una red neuronal) desde una distribución fácil de muestrear (p. ej. ruido gaussiano) hasta la distribución de datos.



Difusión: un «eliminador de ruido» invierte un proceso de ruido estocástico, quitando ruido gradualmente – trayectoria de EDE



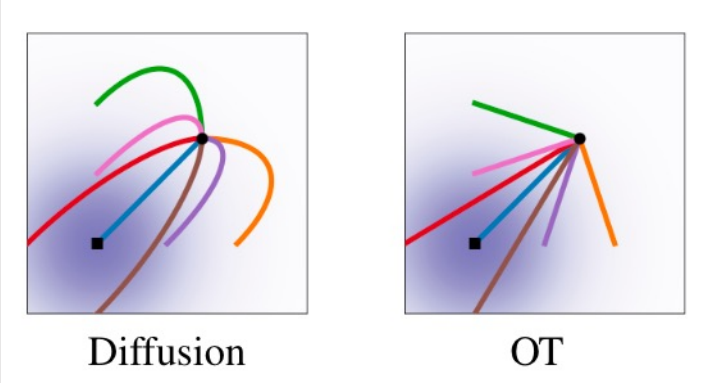
Flow matching [Lipman et al., 2023]: aprende un flujo determinista ruido → datos – trayectoria de EDO



Crédito de la diapositiva: Julie Keisler (INRIA)

Flow matching frente a difusión

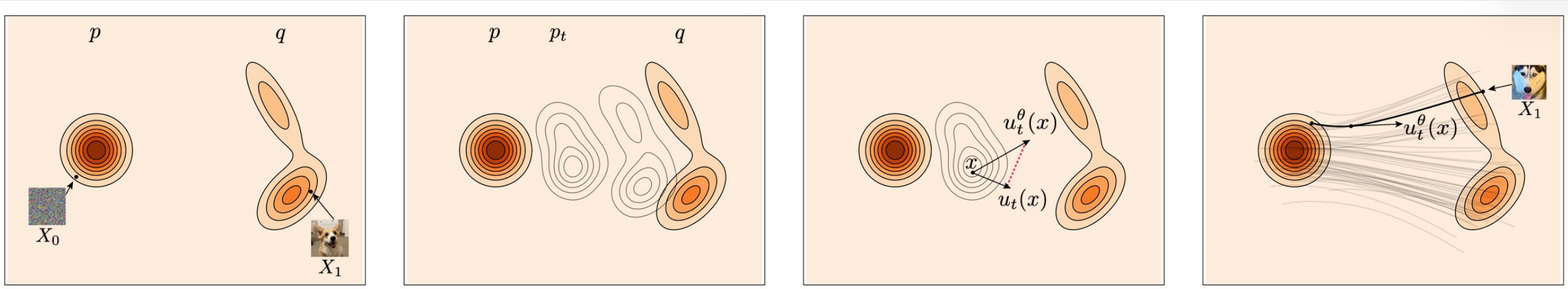
Intuición: crear caminos más rectos entre muestras de datos y muestras de ruido:



Resultados: caminos más rectos en los flujos de probabilidad, que permiten muestrear en menos pasos de difusión aceleraciones

Generalidad: el flow matching puede conectar dos distribuciones cualesquiera (o interpolantes estocásticos), a diferencia de la difusión

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.

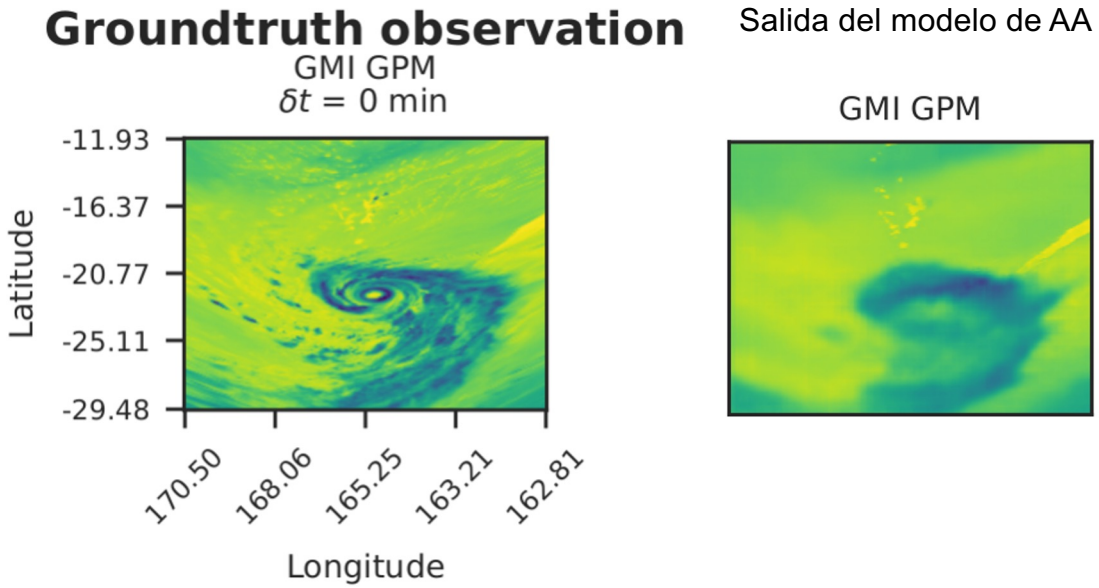
Lipman et al. (2024) Flow Matching Guide and Code

¿Por qué usar IA generativa en clima y tiempo?

Los modelos de AA deterministas suelen optimizar el rendimiento **medio**

Entrenados para minimizar una pérdida media, p. ej. la RMSE

➔ Salidas del modelo borrosas y no físicas (p. ej. imágenes, campos, trayectorias de tormentas)



Salida del modelo de AA

Salidas del modelo de AA, muestreadas tras el entrenamiento la misma red troncal, **en modo generativo**

¿Por qué usar IA generativa en clima y tiempo?

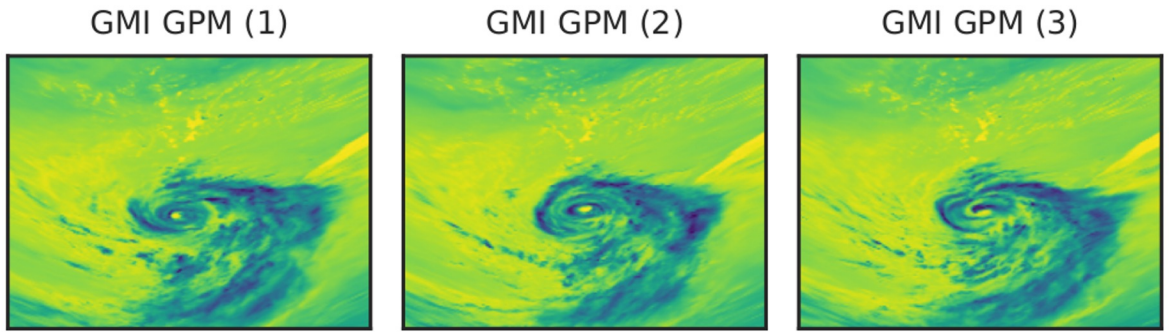
Para muchos problemas de clima, tiempo y geociencias

- Predicción meteorológica, emulación del clima, generación de escenarios
- Estimación de extremos o eventos raros, cuantiles
- Reducción de escala a escalas locales en tiempo o espacio
- Cuantificación del riesgo

La IA generativa aporta ventajas

- Mayor realismo físico
  - Mayor precisión de las altas frecuencias
- Conjuntos probabilísticos
  - Cuantificación de la incertidumbre
  - Información rica y realista para los decisores
- Generación de datos sintéticos

Muestras de un **modelo generativo**



IA generativa para:

✓ Previsión probabilística
- **ArchesWeatherGen**: predicción meteorológica eficiente entrenada a 1,5°
- **ArchesClimate**: emulación, interpolación y extrapolación del clima decenal

❑ Fusión de datos con alineación probabilística de dominios
- **MotifGen**: un modelo generativo, multifuente y autosupervisado para aprender de datos geoespaciales

❑ La alineación de dominios probabilística y la reducción de escala
- **ClimAlign**: reducción de escala generativa no emparejada con alineación de dominios
- **SerpentFlow**: alineación de dominios generativa no emparejada mediante descomposición en estructura compartida (con aplicación a la reducción de escala)

Retos del aprendizaje con datos multifuente

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

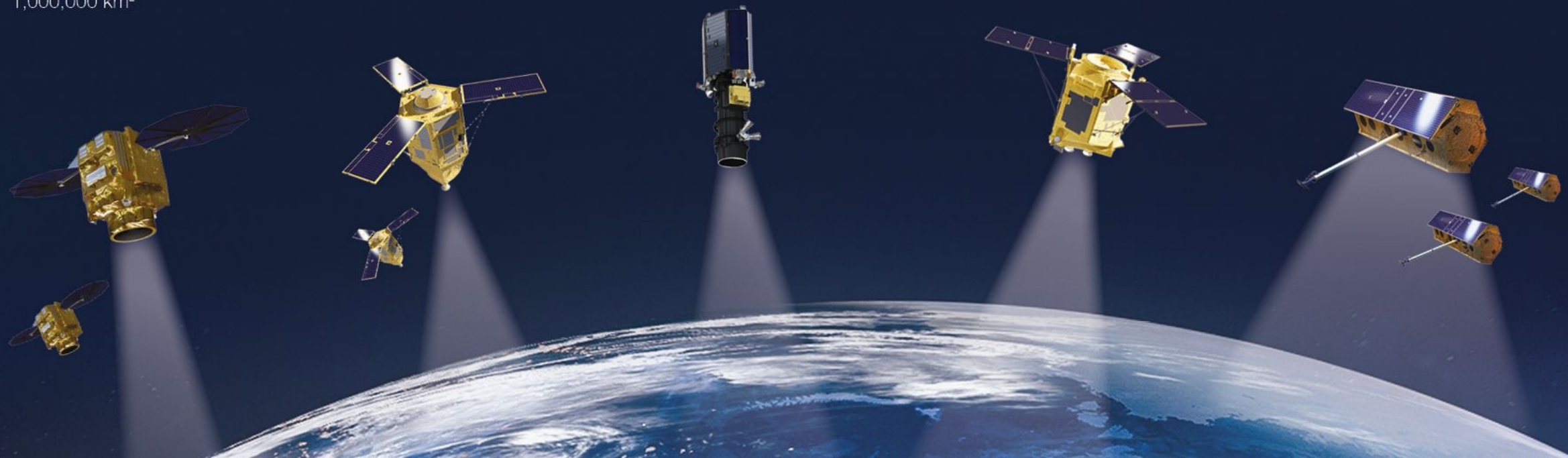
From 25 cm to 40 m

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)



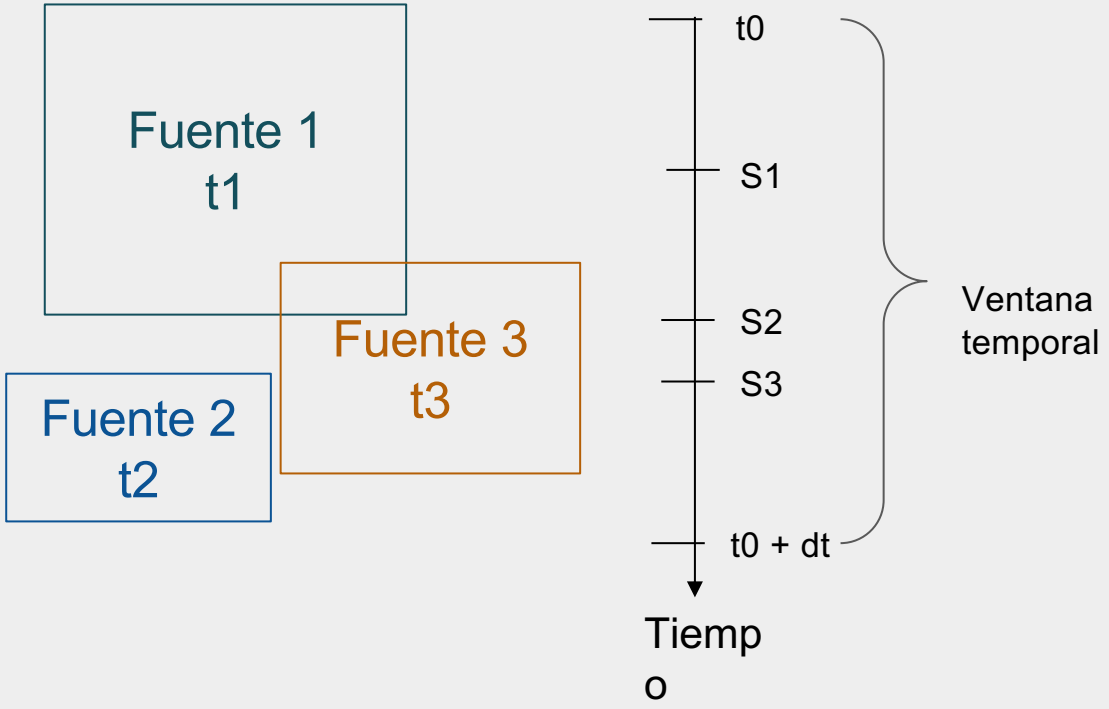
IA para la fusión de datos multifuente (desalineados)

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

Tradicionalmente: solo se entrenan modelos con datos colocalizados



Nuestro modelo de IA (MOTIF) puede aprender de **todas las fuentes de datos**, incluso de datos desalineados en tiempo y espacio



Una arquitectura multifuente para la fusión de datos geoespaciales

AA fundamental multimodal	**Aprendizaje de representaciones geoespaciales**
• Incrustar **valores** (p. ej. texto, imágenes) • Condicionamiento («perro o gato») Modelo DiT (2022)	• Incrustar **ubicaciones** • Entrenamiento autosupervisado a gran escala SatCLIP (2025), AlphaEarth (2025)

MOTIF: *Transformer multifuente con atención factorizada*

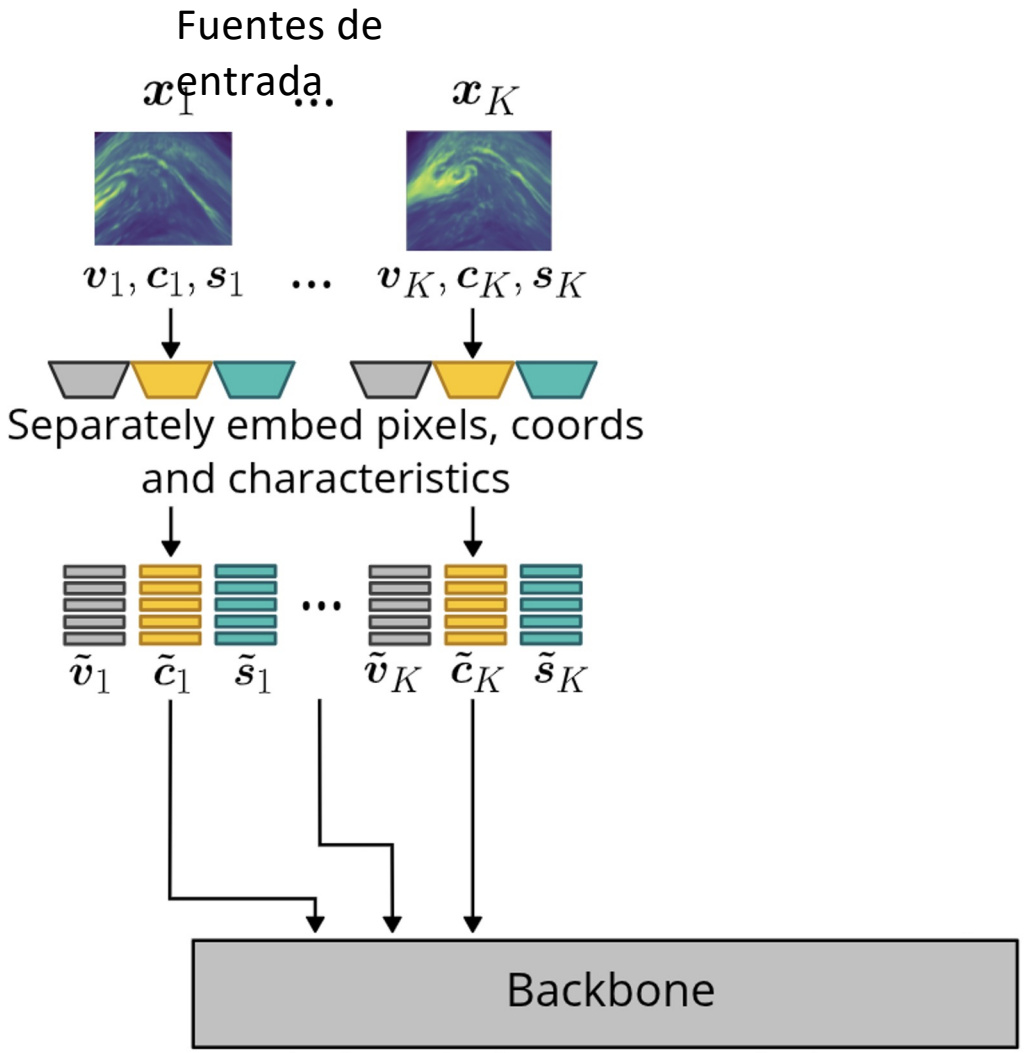
- Incrustar por separado **valores** y **coordenadas**
- Condicionar según las características de la fuente
- Preentrenamiento autosupervisado

[Dauvilliers & Monteleoni, Environmental Data Science, 2025]



Cadena de entrenamiento

Incrustar por separado:
- valores,
- coordenadas,
- características de la fuente



Incrustación posicional relativa

- Incrustación posicional relativa habitual en Transformers (Shaw et al., 2018)

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

- Aquí se usan las coordenadas espaciotemporales para la incrustación posicional:

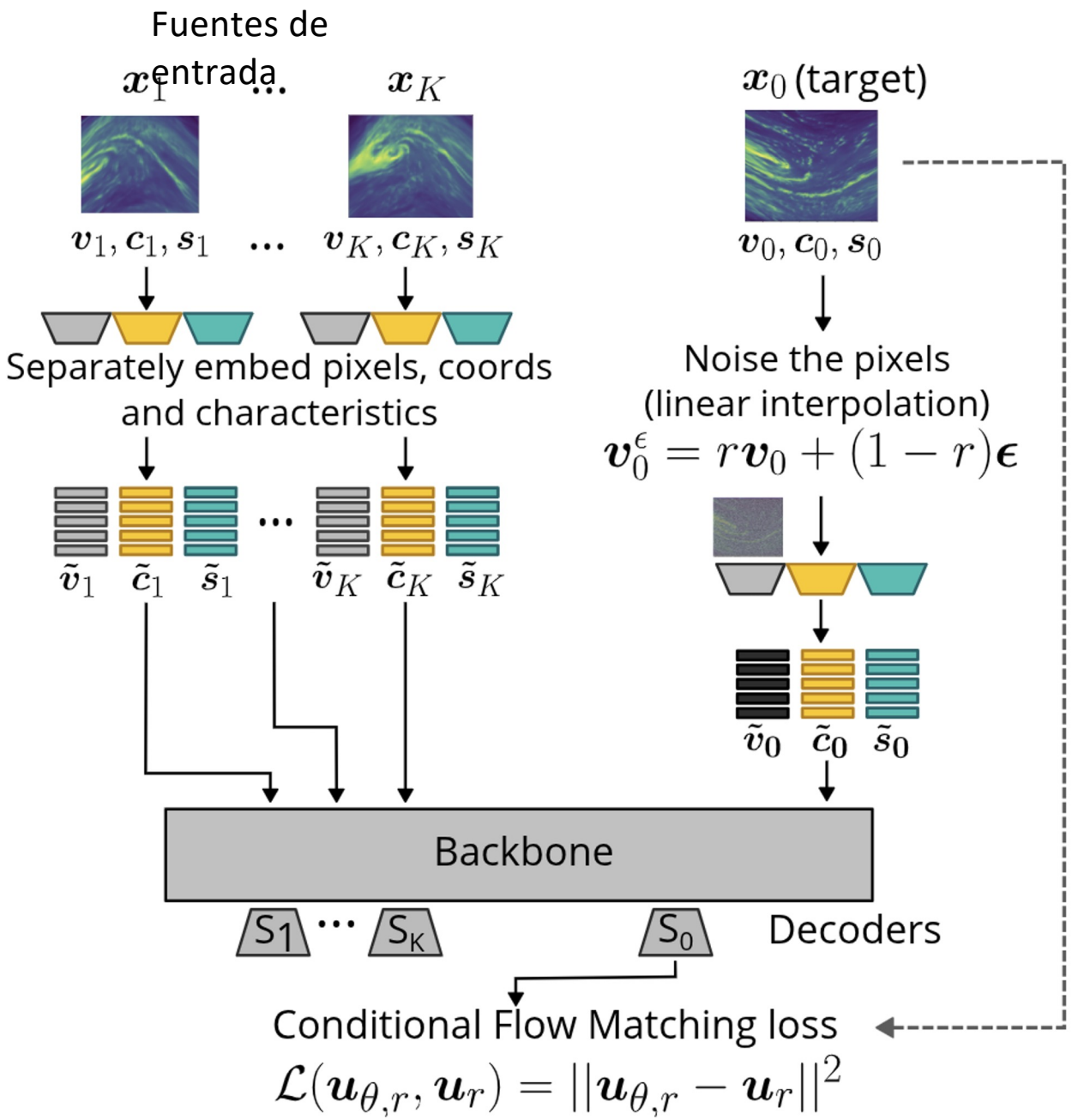
$$\tilde{v}_{out} = \frac{Softmax\left(Q_V K_V^T + Q_C K_C^T\right)}{\sqrt{d}} \tilde{v}_{in}$$

Valores (p. ej. píxeles)

Coordenadas espaciotemporales

Cadena de entrenamiento

Incrustar por separado:
- valores,
- coordenadas,
- características de la fuente



Tarea de entrenamiento autosupervisado: reconstruir una fuente enmascarada

Versión generativa entrenada con flow matching

[Dauvilliers & Monteleoni, Climate Informatics 2026]

Aplicación a ciclones tropicales

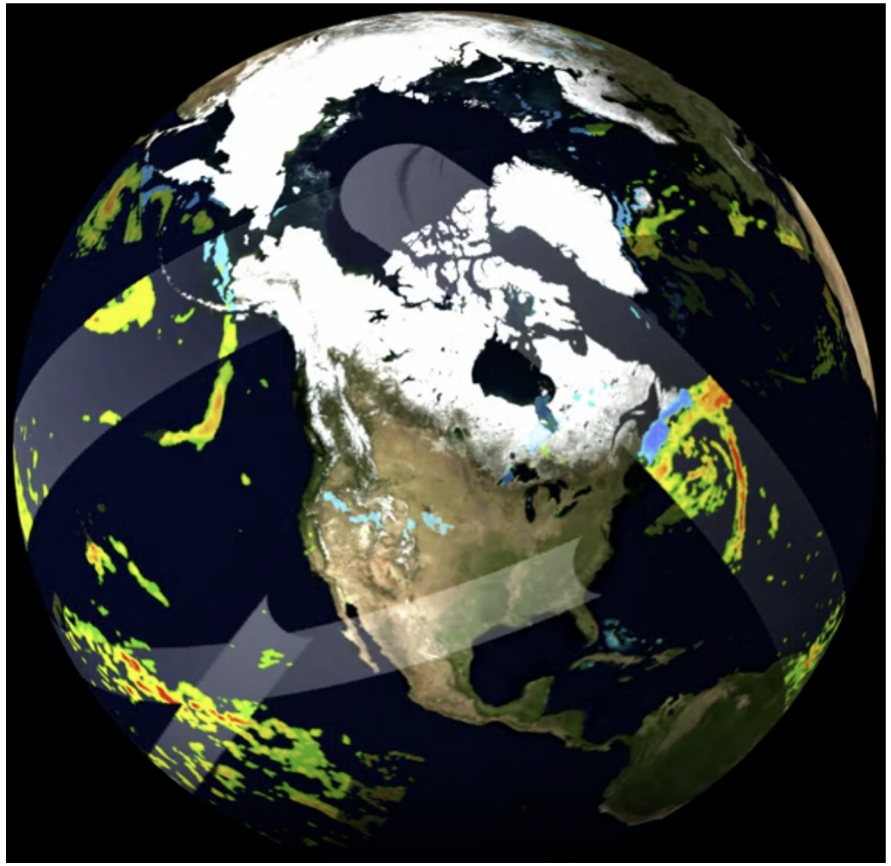
La constelación GPM

GMI (instrumento de referencia)



⇒ ¿Podemos hacer interpolación espaciotemporal de imágenes GMI?

La constelación GPM, NASA



La constelación GPM cubre la Tierra, NASA's Scientific Visualization Studio

156

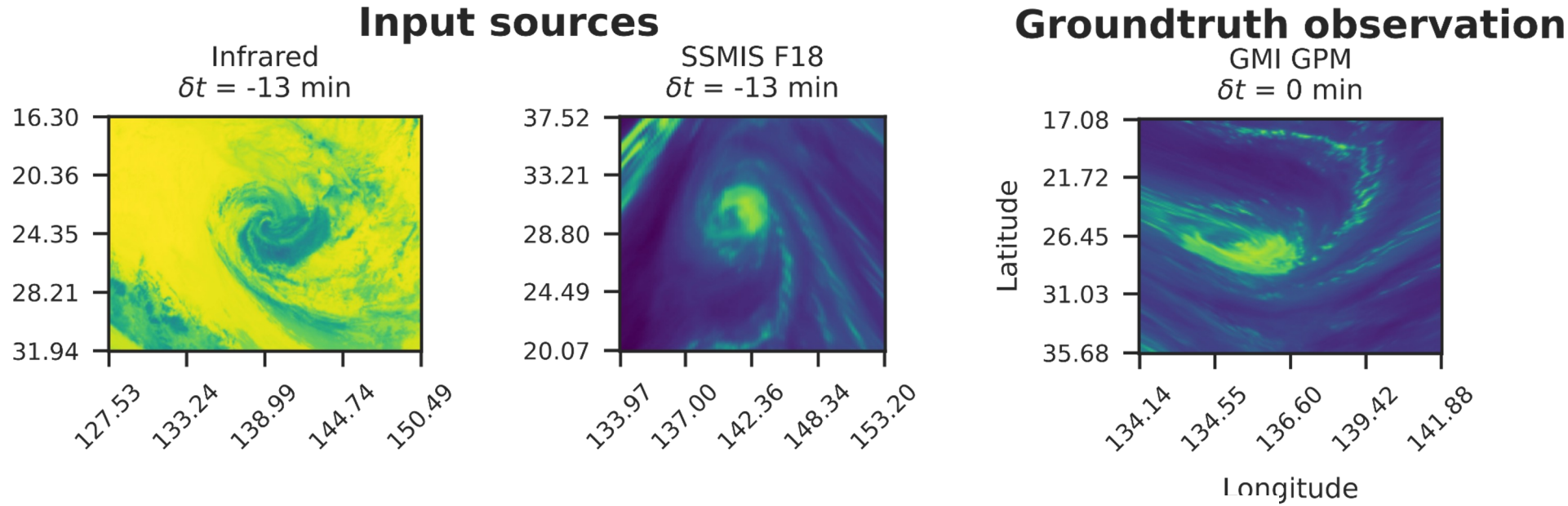
Conjunto de datos

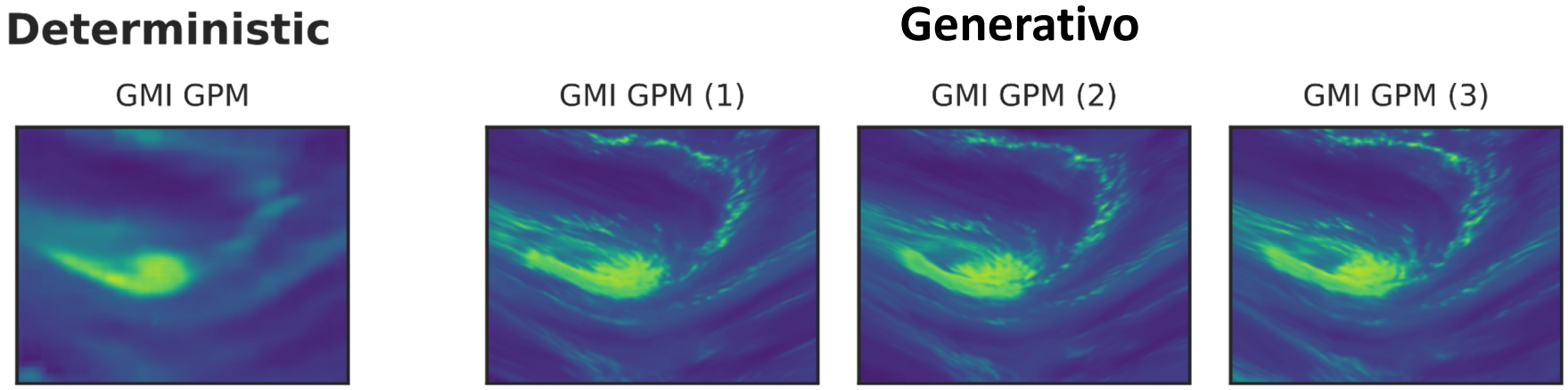
TC-PRIMED (Razin et al. 2023)

- ~3000 tormentas, de 1987 a 2024, todas las cuencas
- 2005-07-14-21-23 para validación, 2006-08-15-22-24 para prueba
- 11 sensores de microondas
- Observaciones infrarrojas

Ventanas de interpolación de 6 horas

Resultados: capta mejor la estructura fina que el modelo determinista





Cadenas de entrenamiento comparadas para la versión generativa

- Entrenamiento supervisado

Sensores no GMI → Modelo → GMI

⇒ ~6000 muestras

- Entrenamiento autosupervisado



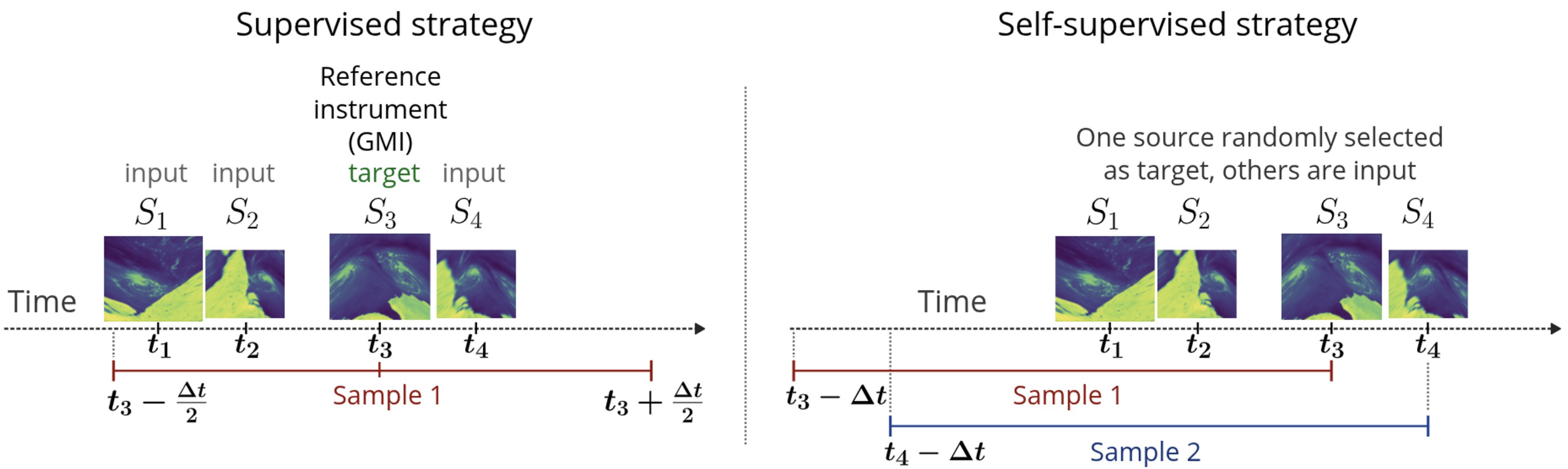
⇒ ~48 000 muestras

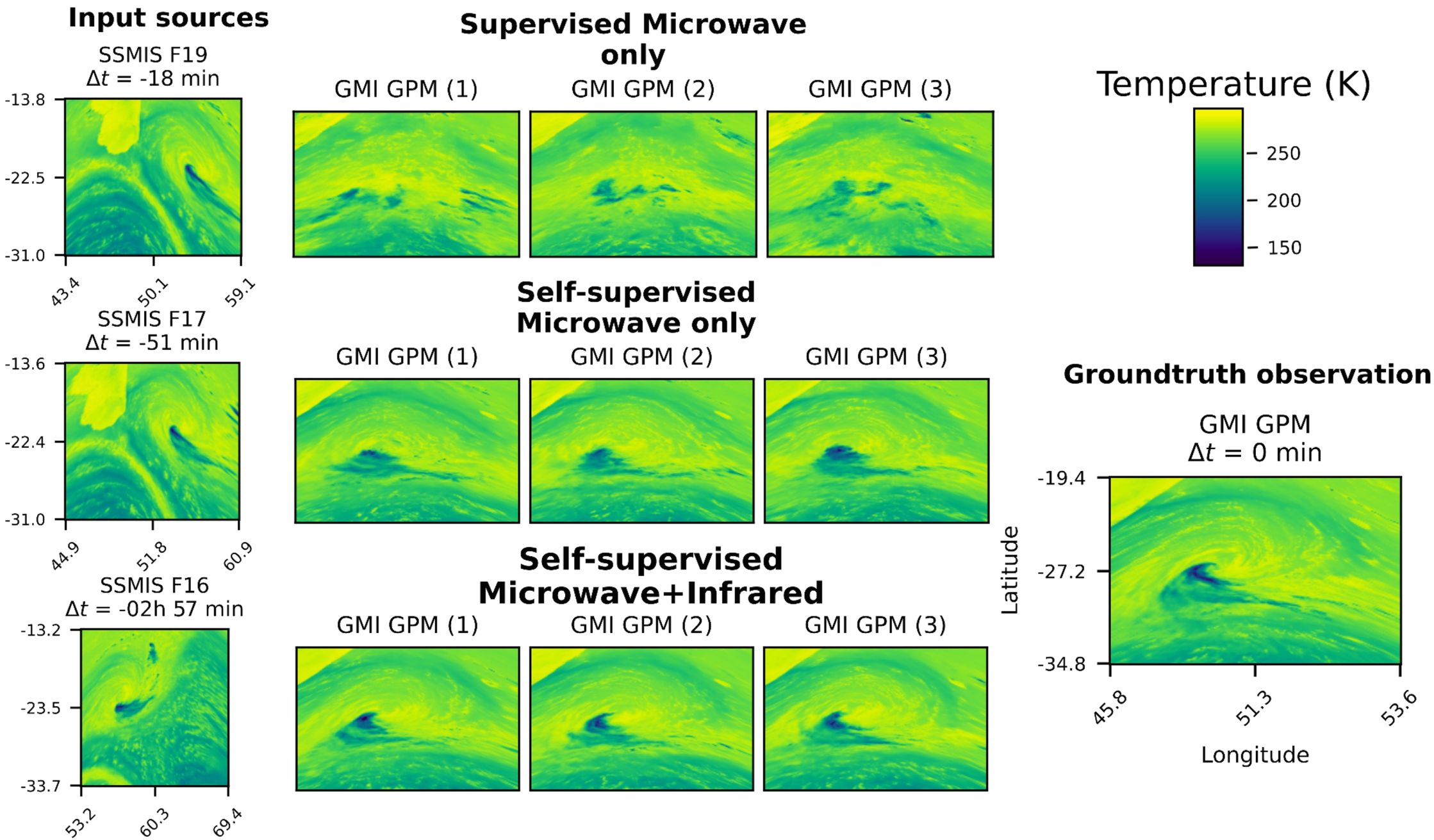
- Autosupervisado + infrarrojo

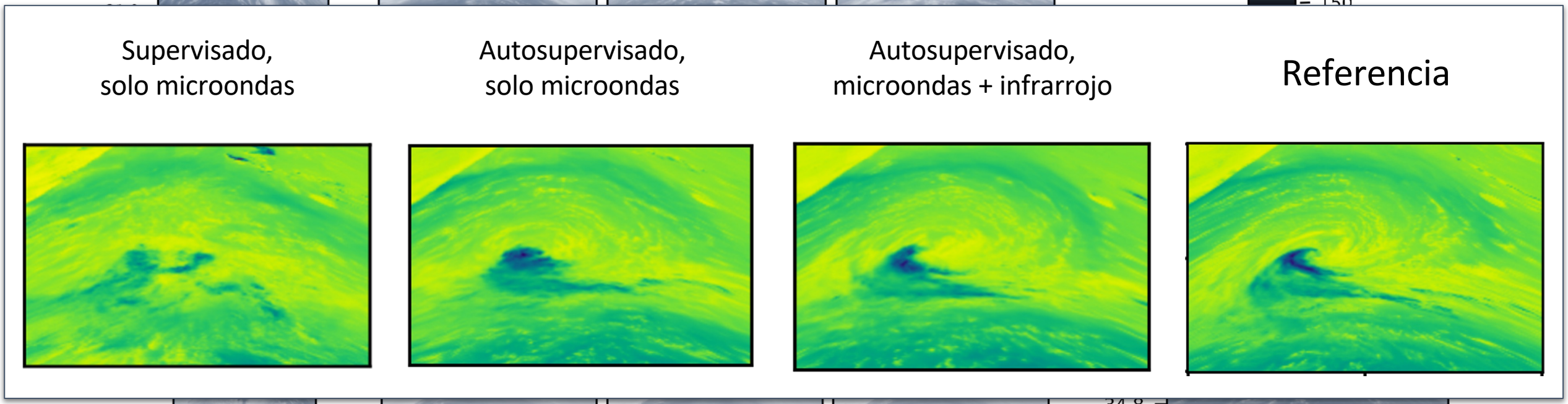


⇒ ~44 000 muestras

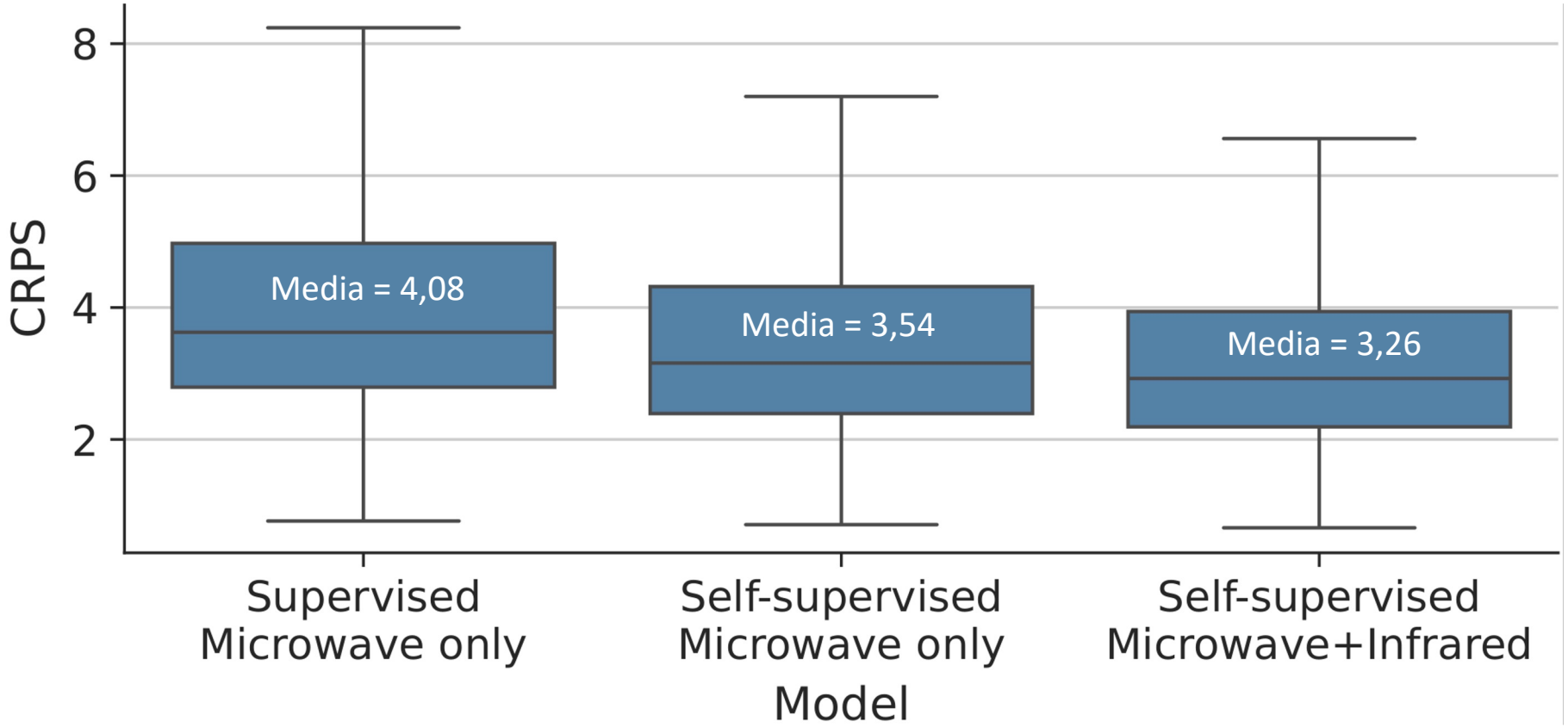
Tareas de entrenamiento




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


Supervisado,
solo microondas
Autosupervisado,
solo microondas
Autosupervisado,
microondas + infrarrojo
Referencia

Versión generativa: ablaciones



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 Referencia	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]

Los corchetes indican intervalos de confianza del 95 %

Puntuación de probabilidad clasificada continua (CRPS)

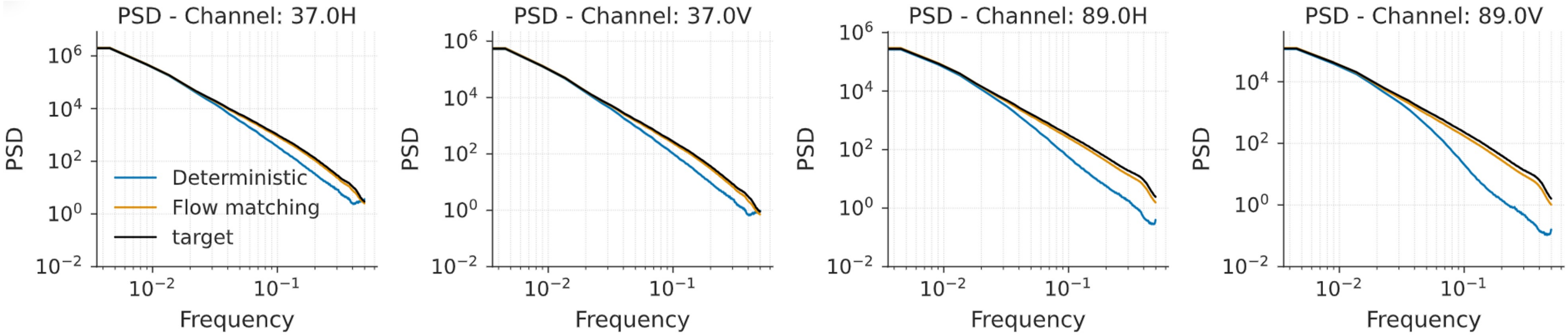
$$\mathrm{CRPS}(F, x) = -\int_{-\infty}^{\infty} (F(y) - \mathbb{1}\{y \geq x\})^2 \, \mathrm{d}y$$

- Otra formulación evita calcular esta integral
- El CRPS puede compararse directamente con el MAE de una previsión determinista
- El Energy Score es la extensión multivariante del CRPS

Lecturas complementarias sobre estas y otras medidas de evaluación de previsiones probabilísticas:

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

Evaluación frente al modelo determinista - espectro

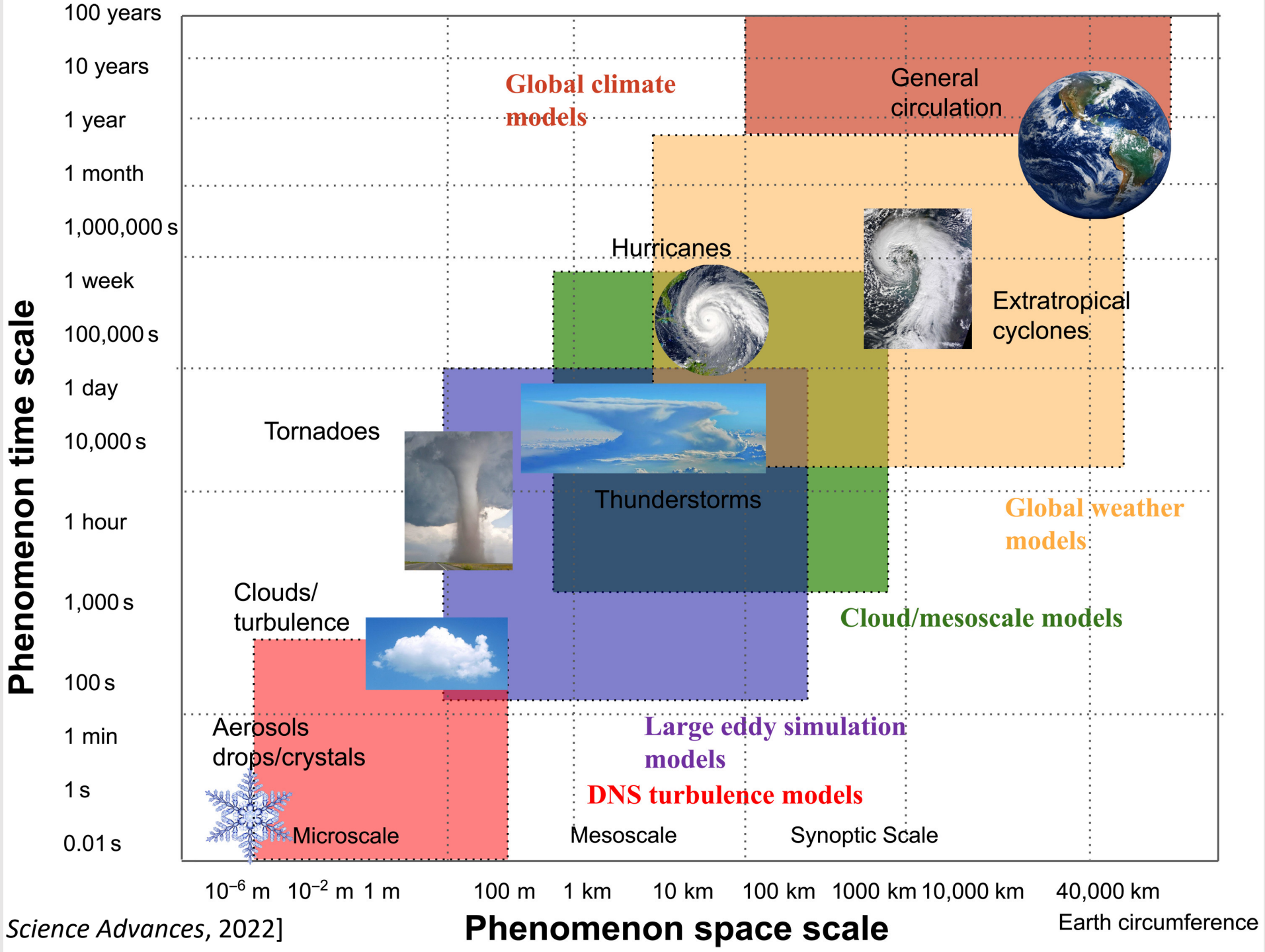


IA generativa para:

- ✓ Previsión probabilística
  - **ArchesWeatherGen**: predicción meteorológica eficiente entrenada a 1,5°
  - **ArchesClimate**: emulación, interpolación y extrapolación del clima decenal

- ✓ Fusión de datos con alineación probabilística de dominios
  - **MotifGen**: un modelo generativo, multifuente y autosupervisado para aprender de datos geoespaciales

- ❑ La alineación de dominios probabilística y la reducción de escala
  - **ClimAlign**: reducción de escala generativa no emparejada con alineación de dominios
  - **SerpentFlow**: alineación de dominios generativa no emparejada mediante descomposición en estructura compartida (con aplicación a la reducción de escala)



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



Reducción de escala no supervisada de modelos climáticos por alineación de dominios: *aplicación a campos de viento.*

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

Actas de Climate Informatics 2026

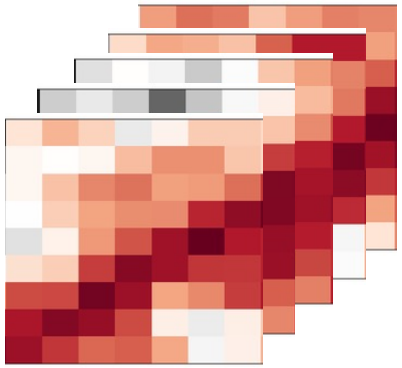
Reducción de escala de modelos climáticos

Resolución espacial gruesa: los MCG funcionan a ~100 km o más y omiten procesos locales clave

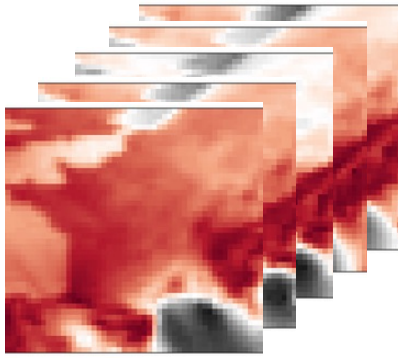
Representación limitada: los rasgos de escala fina, como los efectos topográficos y los contrastes tierra-mar, se captan mal

Sesgos sistemáticos: las salidas de los MCG se desvían mucho de los patrones climáticos observados

Estudios de impacto locales comprometidos: las salidas directas de los MCG no bastan para evaluar los patrones de viento futuros en la planificación energética

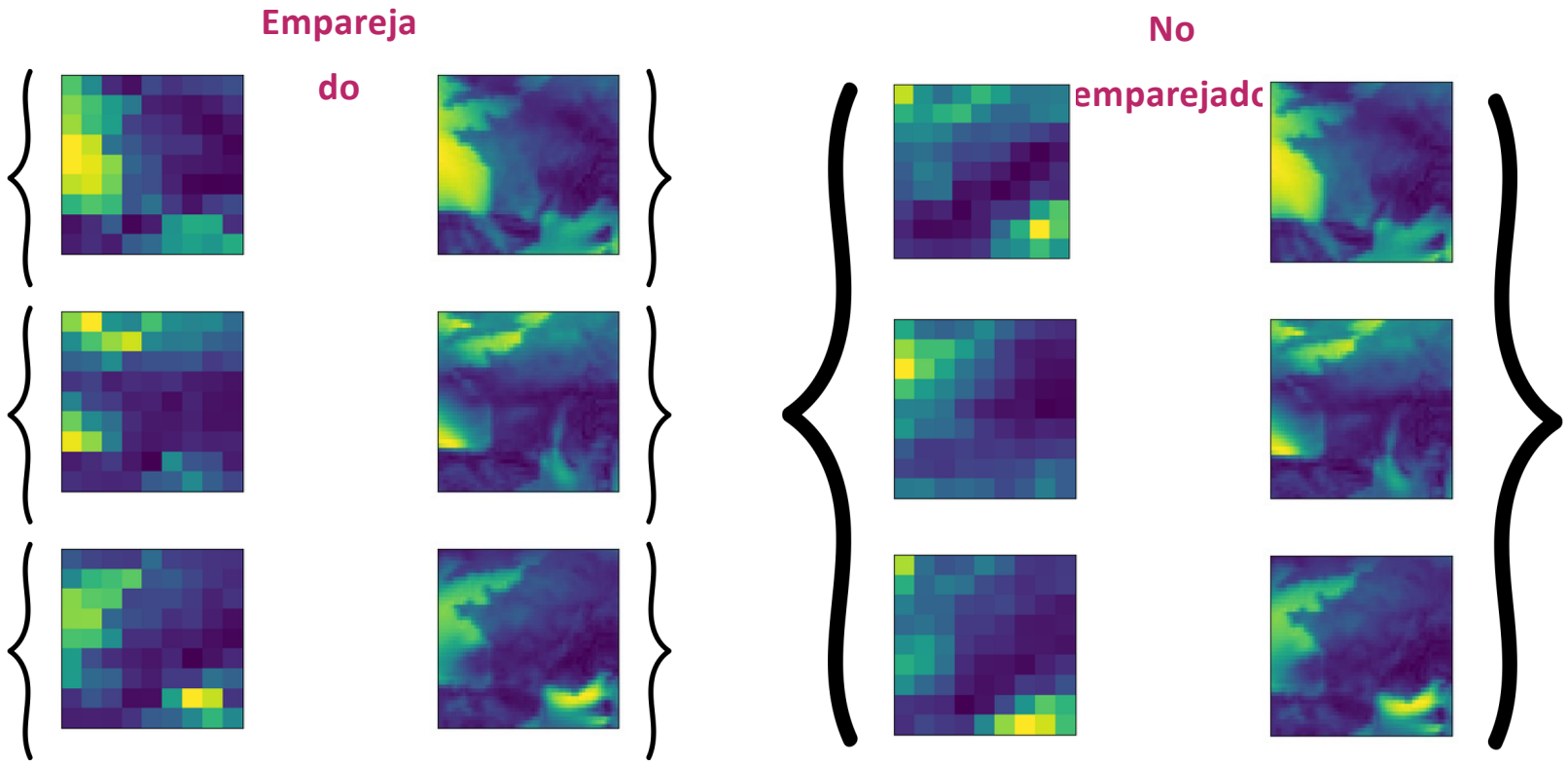


MCG ACCESS-ESM1 sobre Francia durante 5 días



ERA5 sobre Francia durante 5 días

La reducción de escala como alineación de dominios no emparejada



- MCG: **trayectorias físicamente plausibles**
- Reanálisis: la secuencia **meteorológica histórica** real → **MCG y reanálisis no están emparejados**
- Reducción de escala = **alineación de distribuciones**: proyectar las salidas del MCG sobre la distribución del reanálisis

Limitaciones de los métodos estadísticos tradicionales

Métodos basados en distribuciones

- Quantile mapping, CDF-t (Michelangeli et al. 2009), corrección de distribuciones marginales

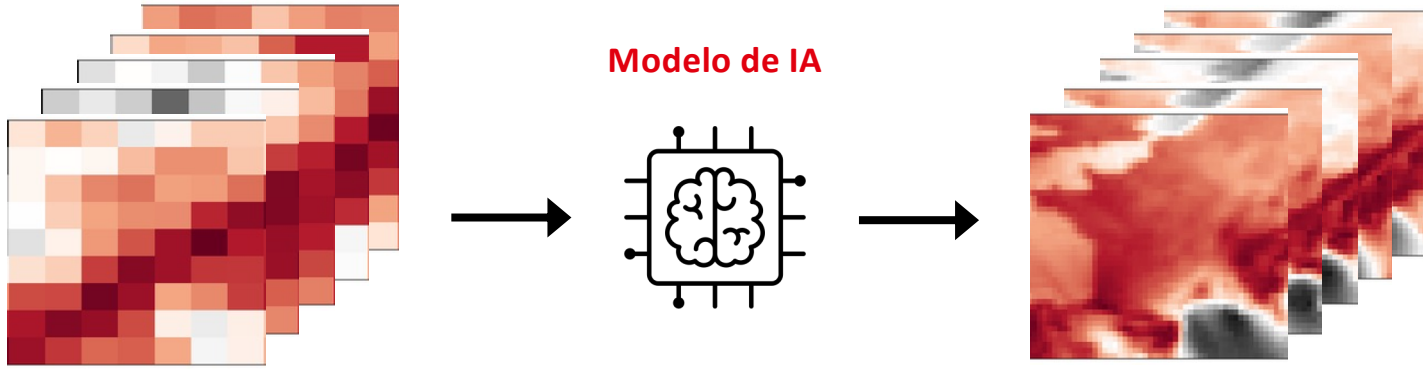
Métodos multivariantes (transporte óptimo)

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

Maldición de la dimensionalidad: los métodos tradicionales no logran modelar todas las interacciones en juego

⟶ Métodos basados en IA / aprendizaje profundo

IA generativa para la reducción de escala



Ventajas

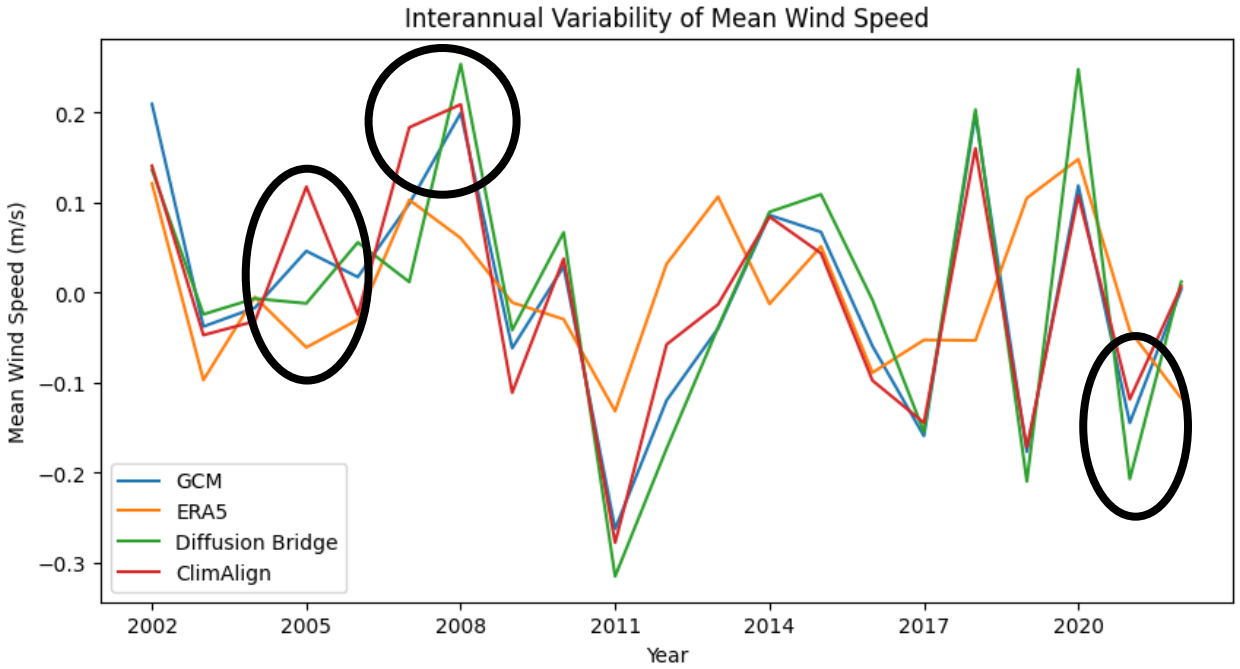
- Aprende la **distribución conjunta completa** de los campos de alta resolución
- Genera **múltiples realizaciones**

Retos

- Datos no emparejados ⟶ **sin verdad de referencia**
- Los métodos implícitos pueden desviarse de la seña del MCG

= proyecciones **poco fiables**

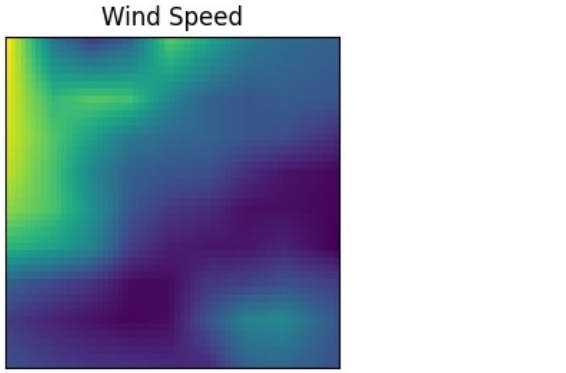
(ClimAlign/Groenke et al. 2020, puentes de difusión/Bischoff & Deck 2024)



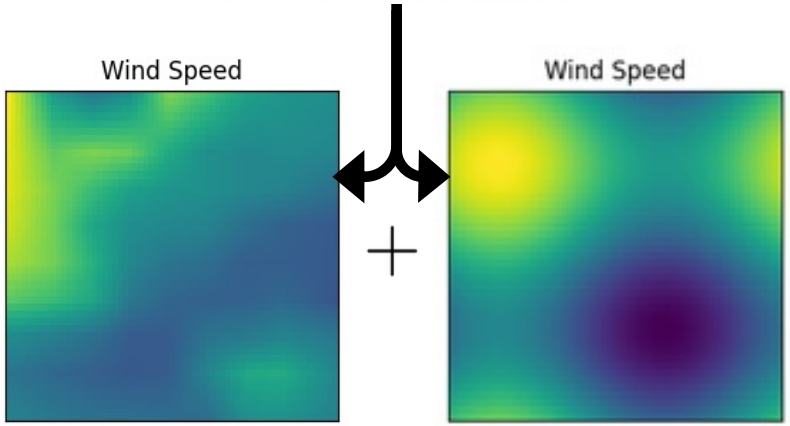
SerpentFlow: introducir pseudopares

SerpentFlow es un método generativo basado en IA para la **alineación de dominios no emparejada** mediante descomposición en estructura compartida

Velocidad media del viento del MCG



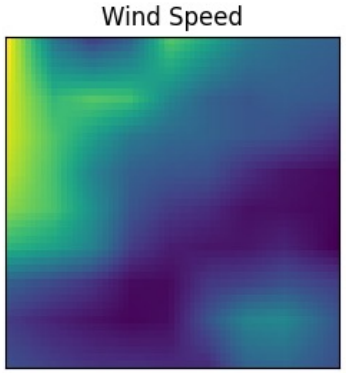
Separar las frecuencias espaciales altas de las bajas

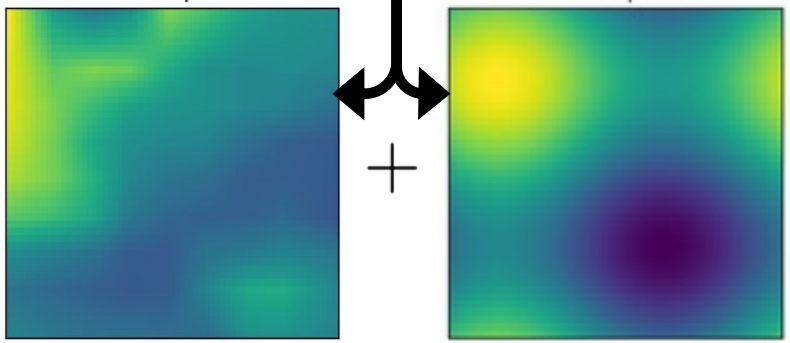


SerpentFlow: introducir pseudopares

SerpentFlow es un método generativo basado en IA para la **alineación de dominios no emparejada** mediante descomposición en estructura compartida

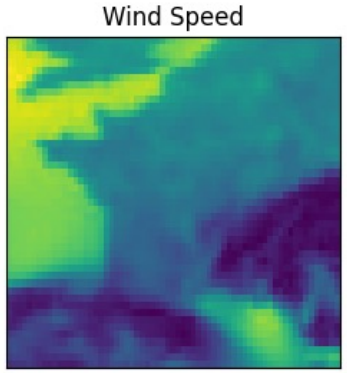
Velocidad media del viento del MCG

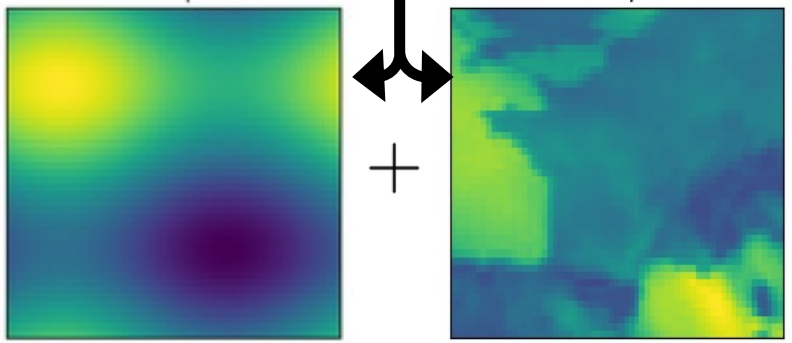




Separar las frecuencias espaciales altas de las bajas

Velocidad media del viento de ERA5

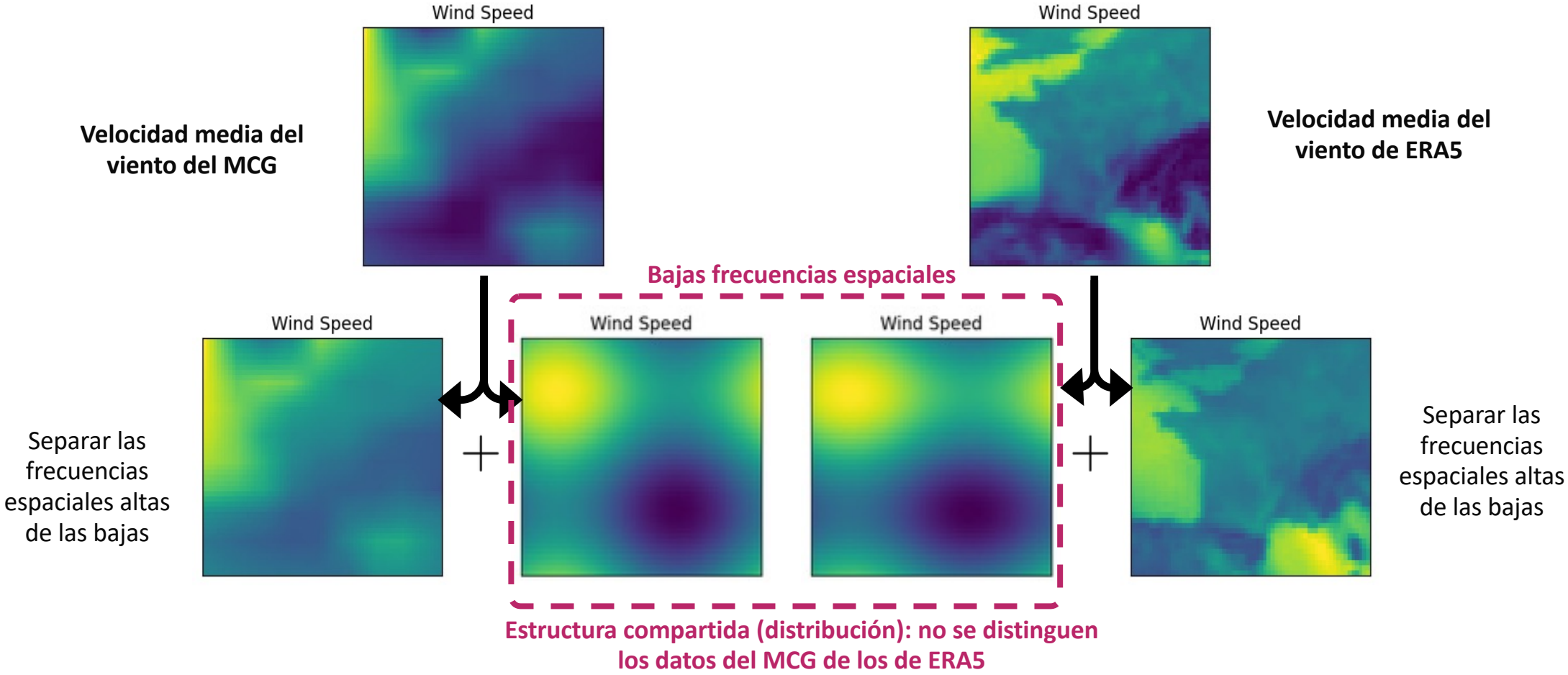




Separar las frecuencias espaciales altas de las bajas

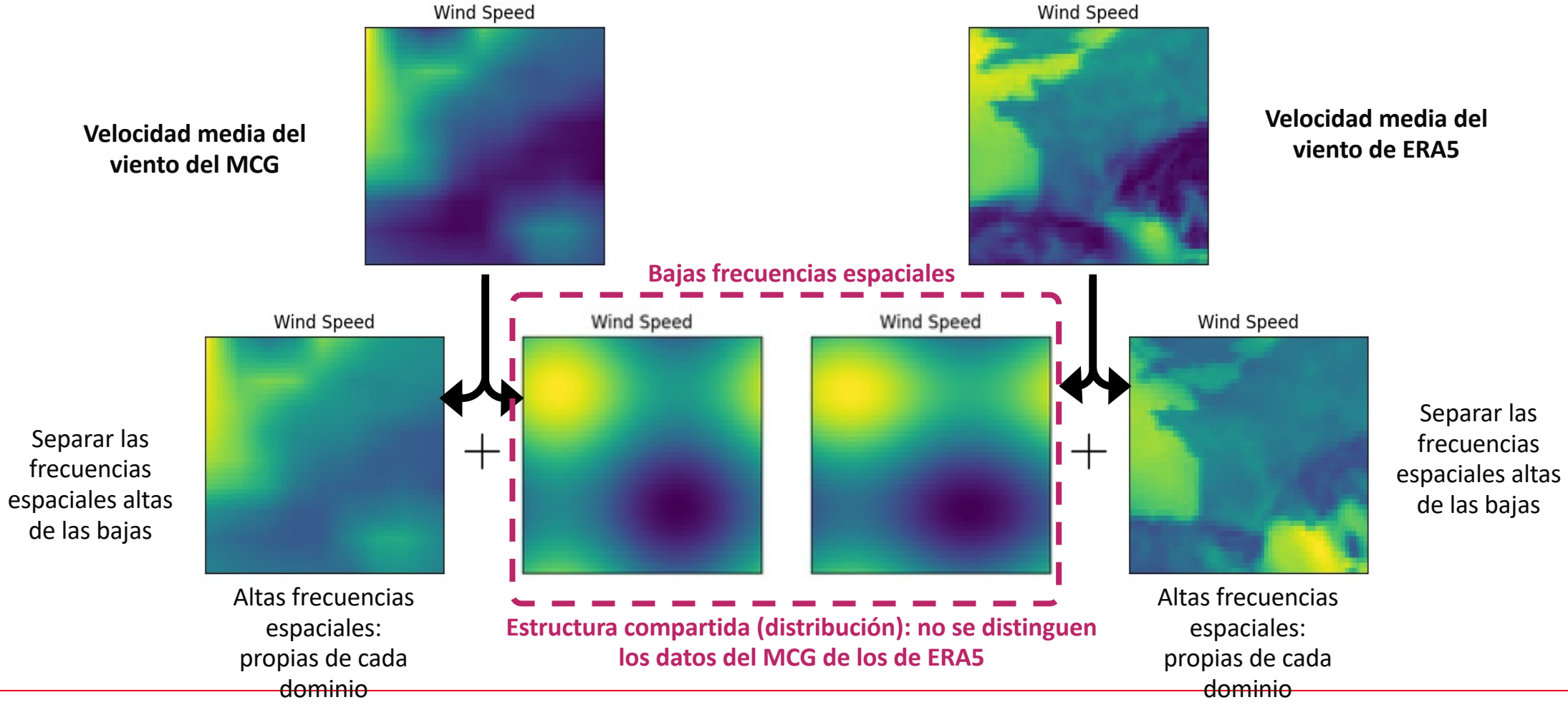
SerpentFlow: introducir pseudopares

SerpentFlow es un método generativo basado en IA para la **alineación de dominios no emparejada** mediante descomposición en estructura compartida

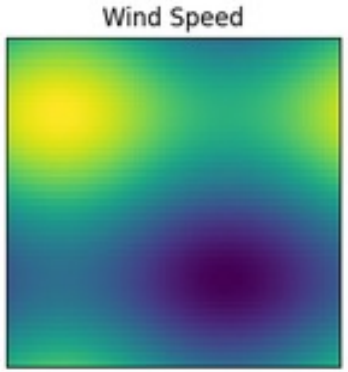


SerpentFlow: introducir pseudopares

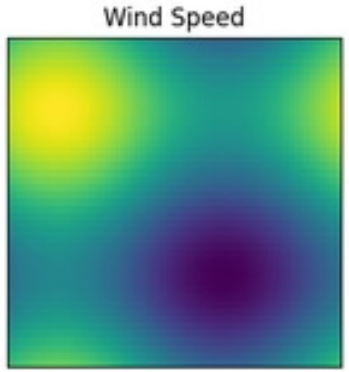
SerpentFlow es un método generativo basado en IA para la **alineación de dominios no emparejada** mediante descomposición en estructura compartida



Encontrar automáticamente la frecuencia de corte



Filtro paso bajo(AR) con frecuencia de corte w



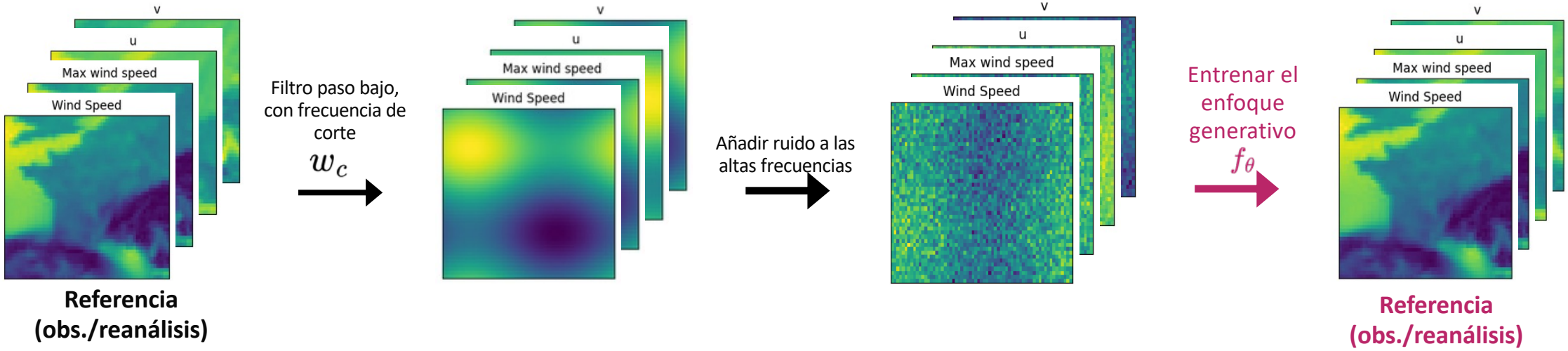
Filtro paso bajo(BR) con frecuencia de corte w

→ Clasificador → ¿Datos BR o AR?

- Dejar de reducir la frecuencia de corte cuando la exactitud <= 60 %
- Frecuencia a menudo hallada ~= **resolución efectiva** del MCG

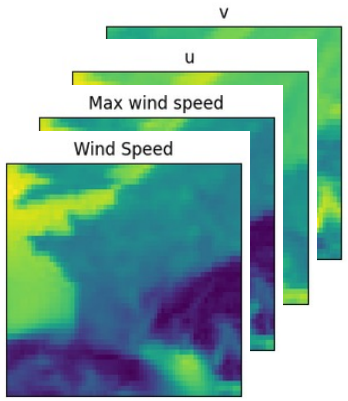
Cadena SerpentFlow

Entrenamiento



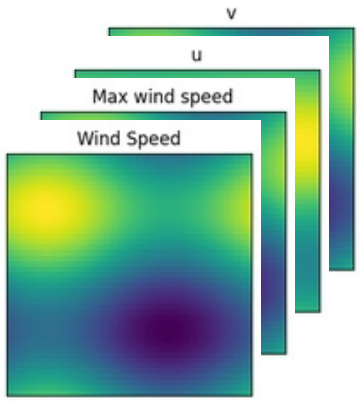
Cadena SerpentFlow

Entrenamiento

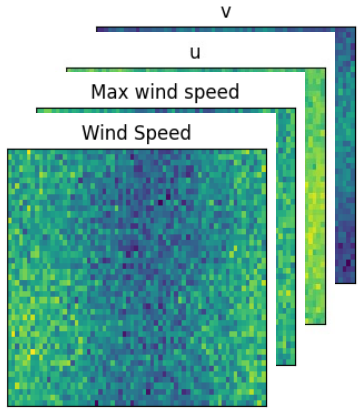


Referencia (obs./reanálisis)

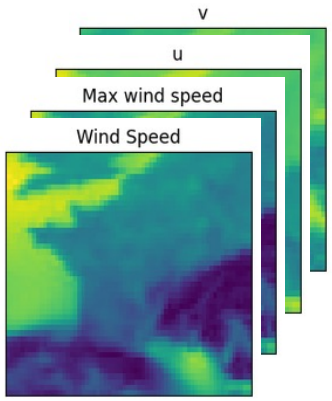
Filtro paso bajo, con frecuencia de corte w_c



Añadir ruido a las altas frecuencias

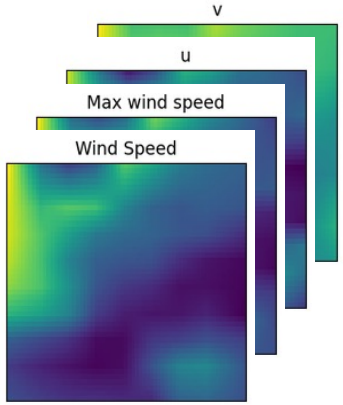


Entrenar el enfoque generativo f_θ



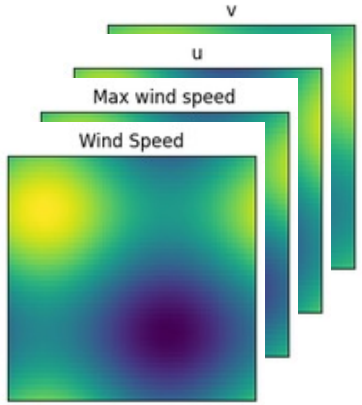
Referencia (obs./reanálisis)

Inferencia

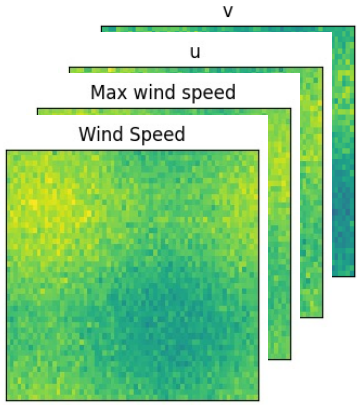


MCG interpolado

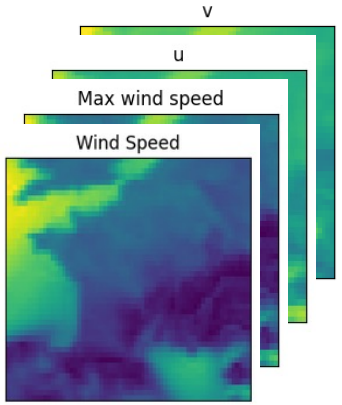
Filtro paso bajo, con frecuencia de corte w_c



Añadir ruido a las altas frecuencias



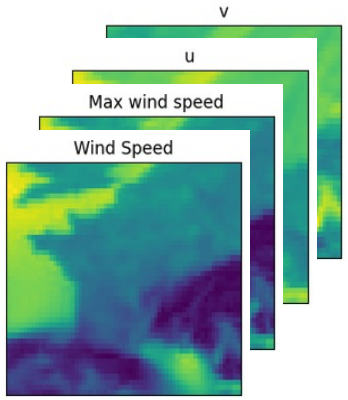
Inferencia con f_θ



MCG con escala reducida

Cadena SerpentFlow

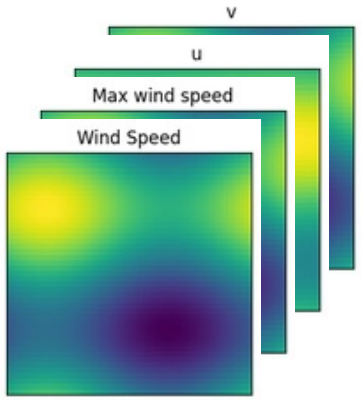
Entrenamiento



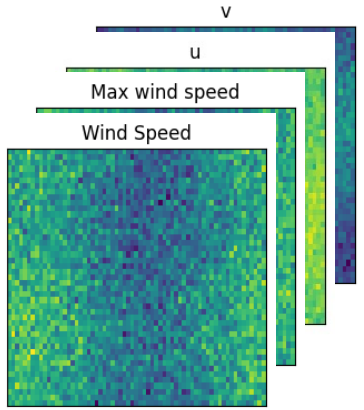
Referencia (obs./reanálisis)

Filtro paso bajo, con frecuencia de corte w_c

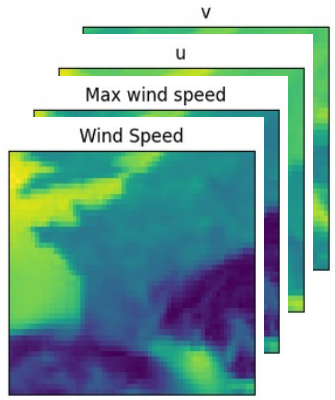
Distribución conjunta



Añadir ruido a las altas frecuencias

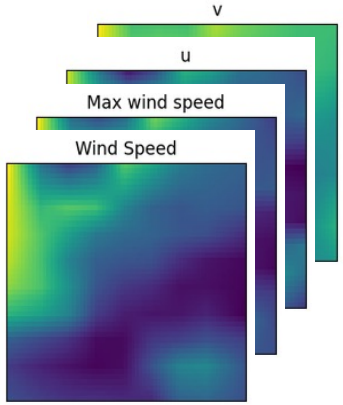


Entrenar el enfoque generativo f_θ



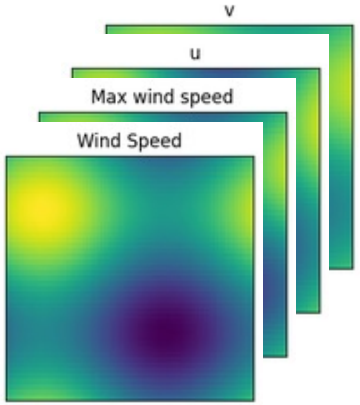
Referencia (obs./reanálisis)

Inferencia

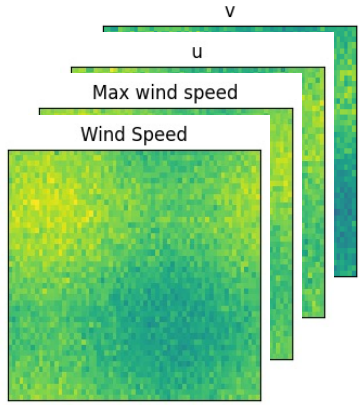


MCG interpolado

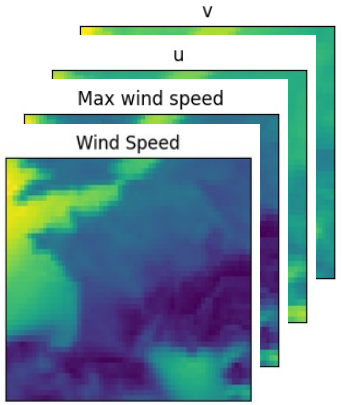
Filtro paso bajo, con frecuencia de corte w_c



Añadir ruido a las altas frecuencias

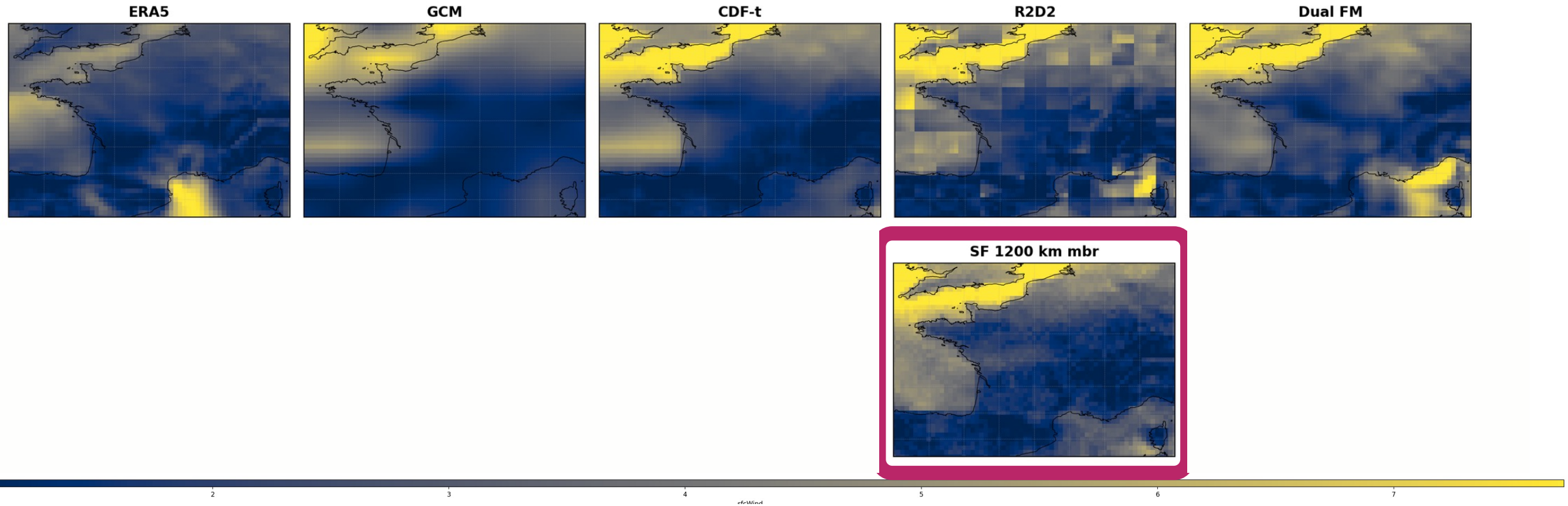


Inferencia con f_θ



MCG con escala reducida

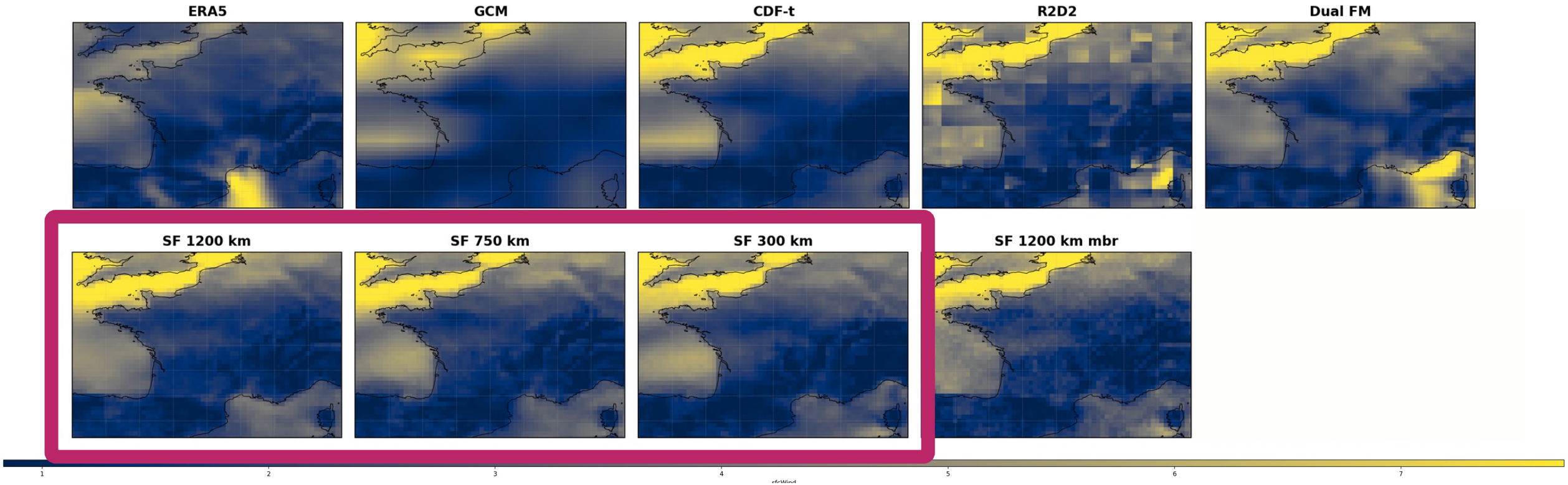
¿Hemos aproximado la distribución de ERA5?



Hallazgos clave

- Los métodos de AP superan a CDF-t y R2D2 en estructura espacial.
- **SF 1200 km** (1 miembro): **mejor aproximación** de la distribución de ERA5

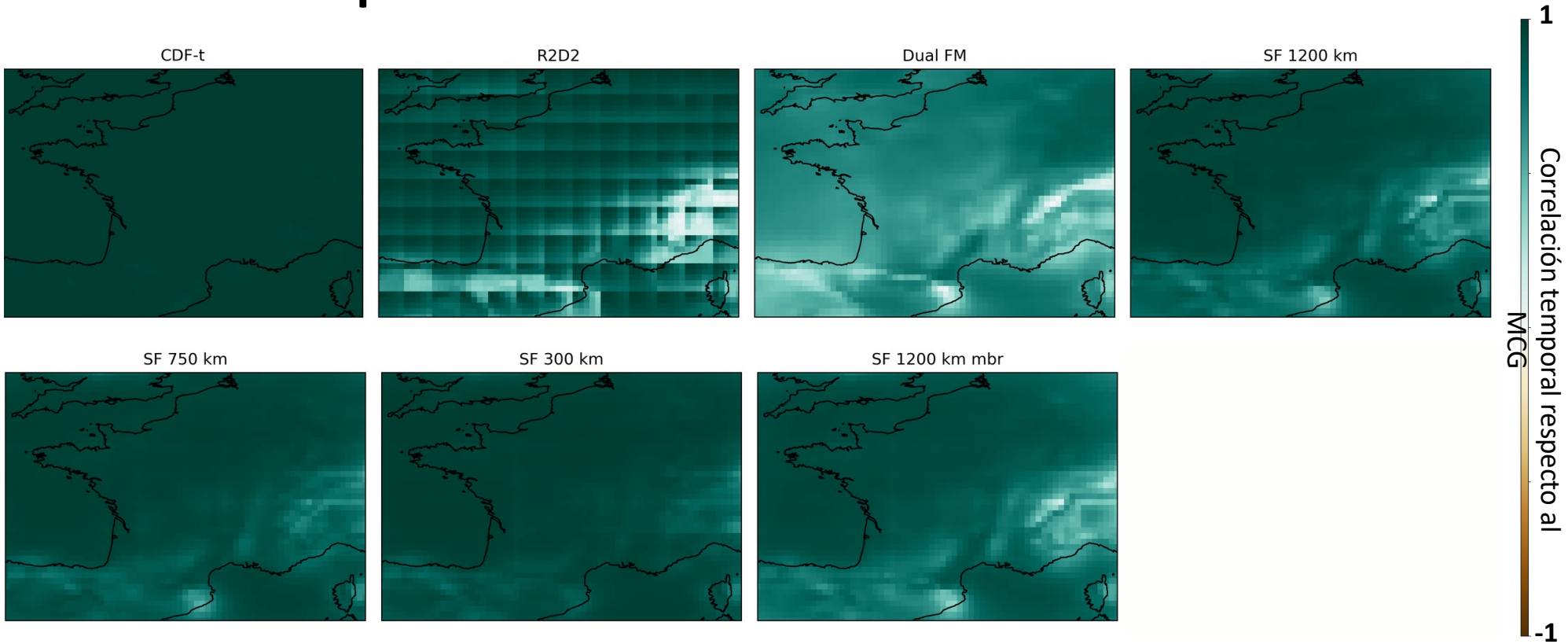
¿Hemos aproximado la distribución de ERA5?



Hallazgos clave

- Los métodos de AP superan a CDF-t y R2D2 en estructura espacial.
- **SF 1200 km** (1 miembro): **mejor aproximación** de la distribución de ERA5
- **frecuencia de corte más alta** ⟶ deriva respecto a la distribución de ERA5.

¿Seguimos la señal que da el MCG?



Hallazgos clave

- **CDF-t es el más cercano a la** dinámica del MCG; Dual FM no sigue la trayectoria del MCG
- La deriva de SF aparece en las regiones **montañosas**
- **Frecuencia de corte más baja** ⟶ más cerca de la dinámica del MCG

Perspectivas

- Funciona con distintos **MCG, resoluciones y dominios irregulares** — ERA5 (25 km) y SAFRAN (8 km)
- Supera a los métodos estadísticos al modelar la **distribución conjunta** de los datos de reanálisis
- Frente a otros enfoques generativos, se mantiene cerca de la dinámica del MCG
- **La frecuencia de corte es interpretable**
  - Vincula directamente la separación de escalas con la resolución efectiva del MCG
  - Gestiona **el compromiso** entre seguir la trayectoria del MCG y reproducir la distribución del reanálisis
- **Límites:** solo viento cerca de la superficie — próximos pasos: otras variables, extremos, reducción de escala temporal

IA generativa para:

- ✓ Previsión probabilística
  - **ArchesWeatherGen**: predicción meteorológica eficiente entrenada a 1,5°
  - **ArchesClimate**: emulación, interpolación y extrapolación del clima decenal

- ✓ Fusión de datos con alineación probabilística de dominios
  - **MotifGen**: un modelo generativo, multifuente y autosupervisado para aprender de datos geoespaciales

- ✓ La alineación de dominios probabilística y la reducción de escala
  - **ClimAlign**: reducción de escala generativa no emparejada con alineación de dominios
  - **SerpentFlow**: alineación de dominios generativa no emparejada mediante descomposición en estructura compartida (con aplicación a la reducción de escala)

Primeros pasos con la IA generativa basada en flujos

Flow matching

Tutoriales:

- 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

Lecturas complementarias sobre métodos basados en flujos

Interpolantes estocásticos

Marco general basado en flujos para transportar ruido a datos

- El flow matching y los rectified flows son ejemplos
- Generaliza el marco: se puede entrenar con una EDO y muestrear con una EDE, o al revés

Operadores neuronales

Los métodos basados en flujos operan en el espacio de operadores, lo que permite generalizar con independencia de la escala

Modelos de consistencia

Versión de un solo paso del flow matching. Área de investigación abierta, con algo de teoría y limitaciones prácticas.

Aprendizaje por cadena de Markov inversa

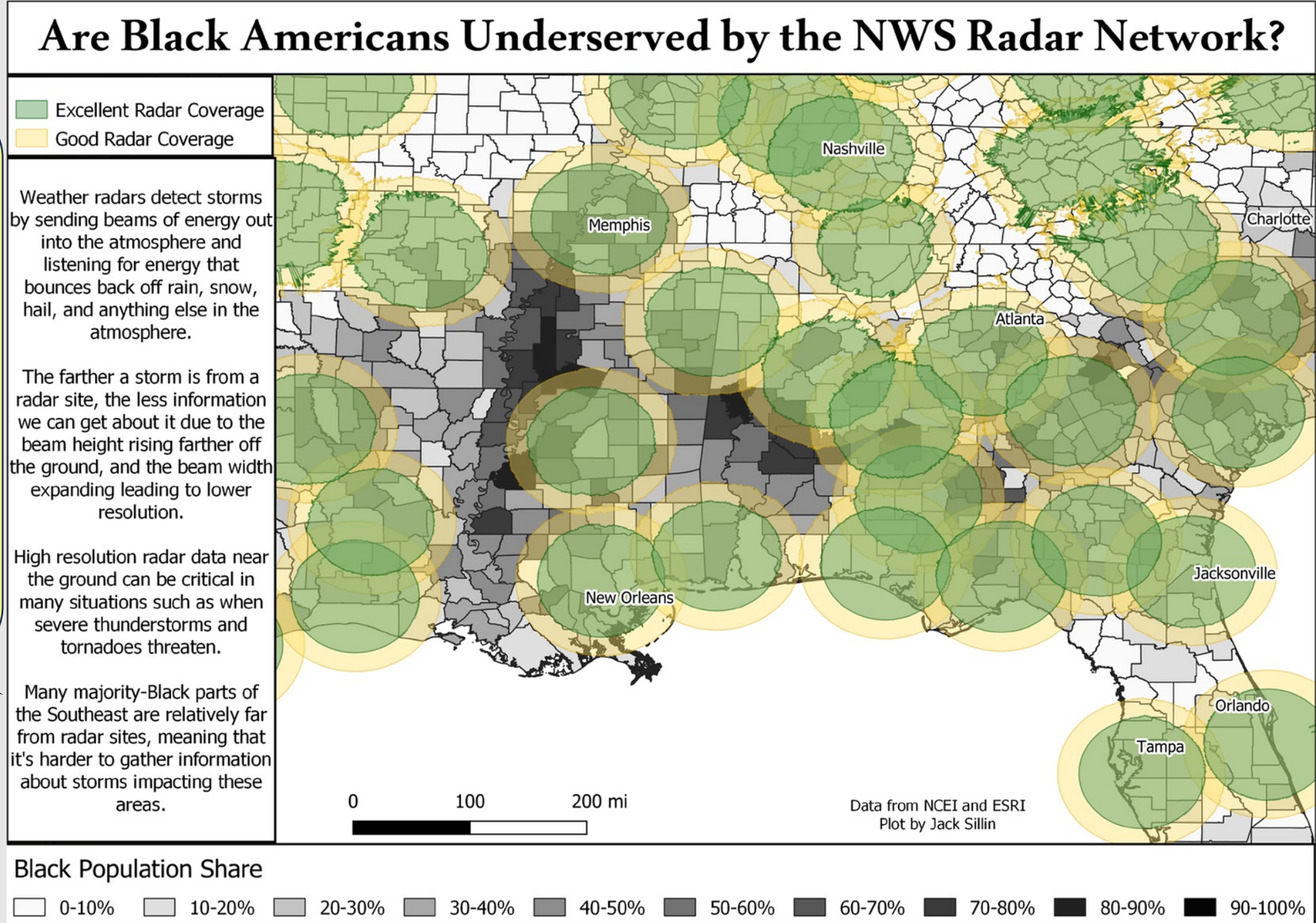
En esencia, flow matching de baja a alta resolución

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

Sección extra: IA para la equidad de los datos climáticos

«Muchas zonas de mayoría negra del Sudeste [de EE. UU.] están relativamente lejos de los radares, lo que dificulta recoger información sobre las tormentas que las afectan».

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



IA para la equidad de los datos climáticos

- Entrenar los modelos en regiones con muchos datos y aplicarlos en regiones con pocos datos
  - Entrenarlos y validarlos en regiones con muchos datos
  - Ajustarlos con los datos limitados de las regiones con pocos datos y usarlos para generar más datos.

- Contribución a la equidad de los datos climáticos
  - Escalas locales (p. ej. legado de injusticia ambiental en EE. UU.)
  - Escalas globales:
    - El Norte global históricamente ha emitido más carbono; además, allí suele haber más datos
    - El Sur global sufre los efectos más graves del calentamiento resultante