

La ricerca sull'IA
per i cambiamenti climatici e la sostenibilità ambientale

Claire Monteleoni

INRIA Paris

University of Colorado Boulder

Choose France™



Dicembre 2021: contea di Boulder, Colorado

- La siccità nevosa dell’autunno e dell’inverno 2021 ha inaridito la copertura del suolo
- Venti a 130-160 km/h, più un innesco, hanno scatenato una «tempesta di fuoco» incontrollabile
- 2 vittime. 1.000 case e 20 attività distrutte, e altre danneggiate





«L’opportunità che l’IA rappresenta per la Terra è significativa. L’attuale esplosione dell’IA ci porterà a integrarla in sempre più cose ogni anno.... Pensando ai vantaggi, alle efficienze e alle nuove soluzioni che crea per nazioni, imprese e vita quotidiana, dobbiamo anche pensare a come massimizzare i benefici per la società e per l’ambiente nel suo complesso.»

– World Economic Forum (2018), «Harnessing Artificial Intelligence for the Earth»





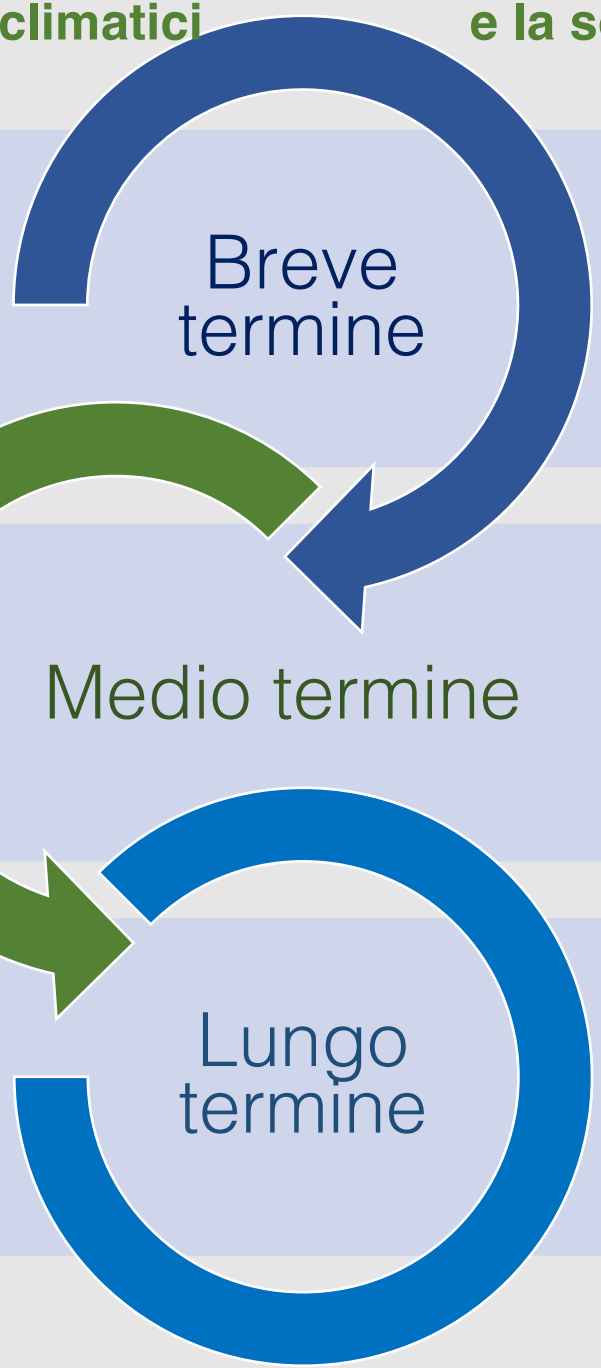
La ricerca sull'IA per i cambiamenti climatici e la sostenibilità ambientale

CAMBIAMENTI CLIMATICI

ADATTAMENTO

MITIGAZIONE

IMPATTI



Meteo estremo

Rischi a cascata

Transizione energetica

Cambio d'uso del suolo

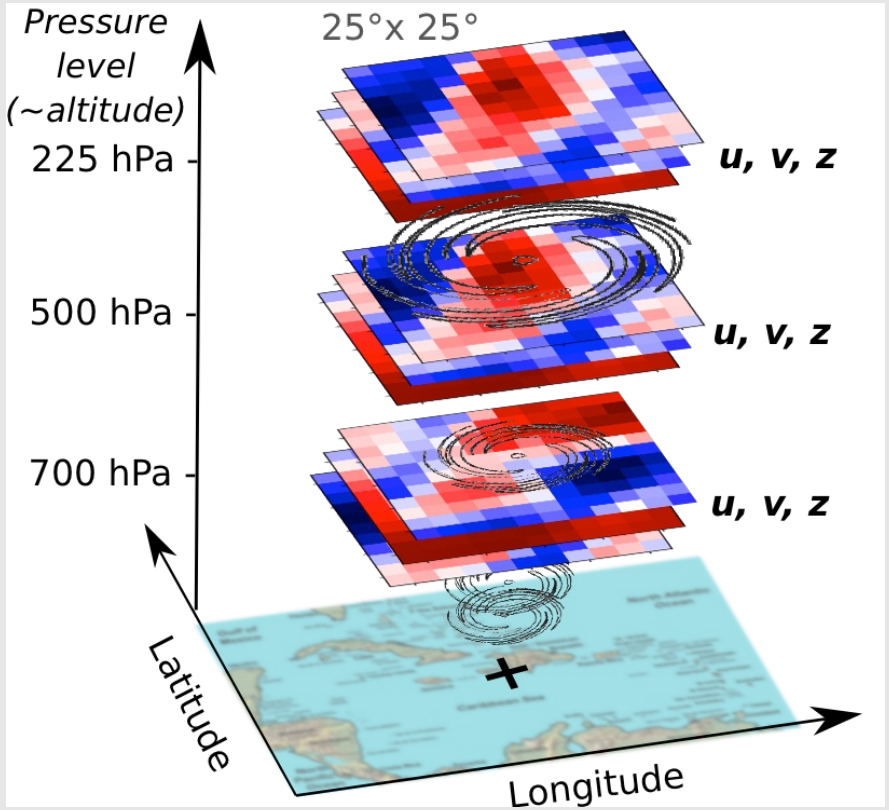
Emissioni di carbonio

Innalzamento del mare

Soluzioni basate sull'IA

ADATTAMENTO

IA per il meteo estremo e i rischi a cascata



[Giffard-Roisin et al., Frontiers 2020]

Previsione delle traiettorie degli uragani

Previsione delle precipitazioni estreme del monsone estivo indiano

Rilevamento delle valanghe

IA generativa per le previsioni meteo



MITIGAZIONE

Ridurre le emissioni di carbonio

Accelerare la transizione all'energia verde

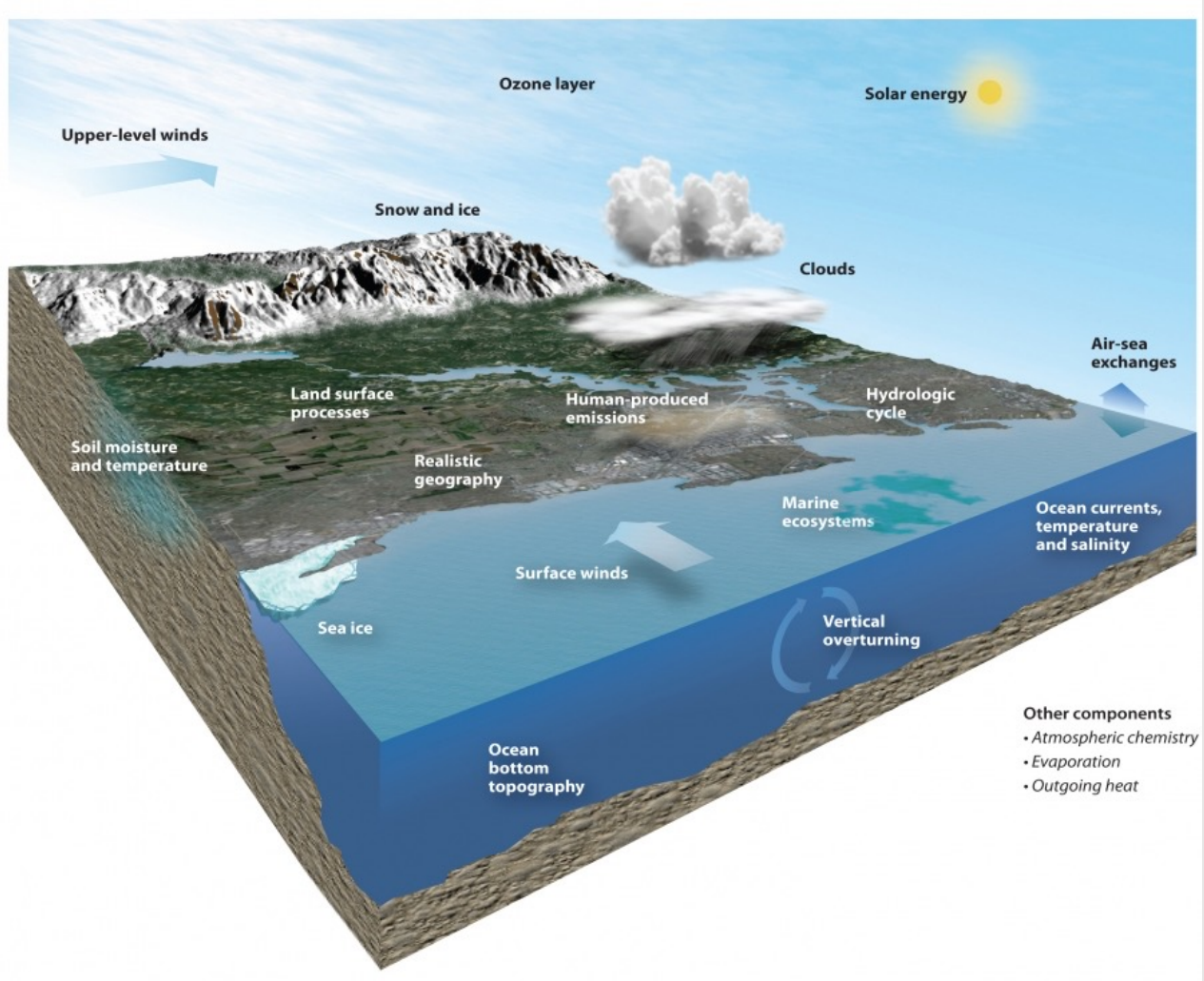
- Previsione di solare ed eolico con l'IA
- IA per il downscaling dei dati solari ed eolici

Ridurre il calcolo per la modellazione meteo e climatica

- Una volta addestrata, l'IA prevede molto più rapidamente dei modelli fisici

IMPATTI

IA per capire e prevedere i cambiamenti climatici



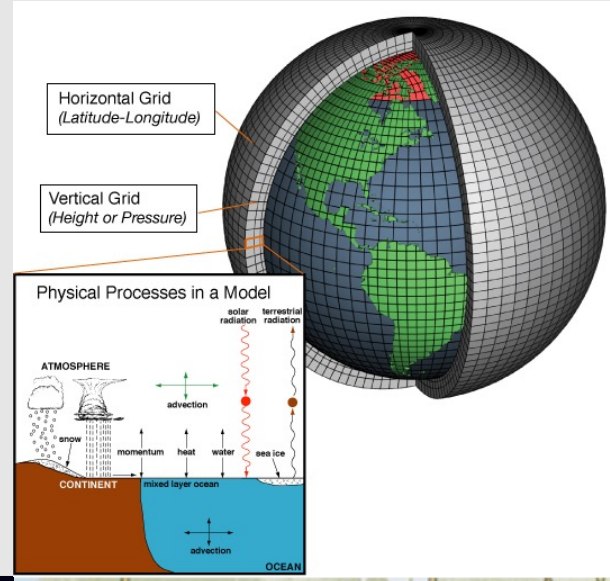
UCAR Science Education

Usare l'IA per apprendere le relazioni tra simulazioni IPCC e osservazioni

- Rendere più robuste le previsioni di ensemble di modelli climatici
- Proiettare l'innalzamento del livello del mare a lungo termine
- Proiettare le emissioni di carbonio a lungo termine

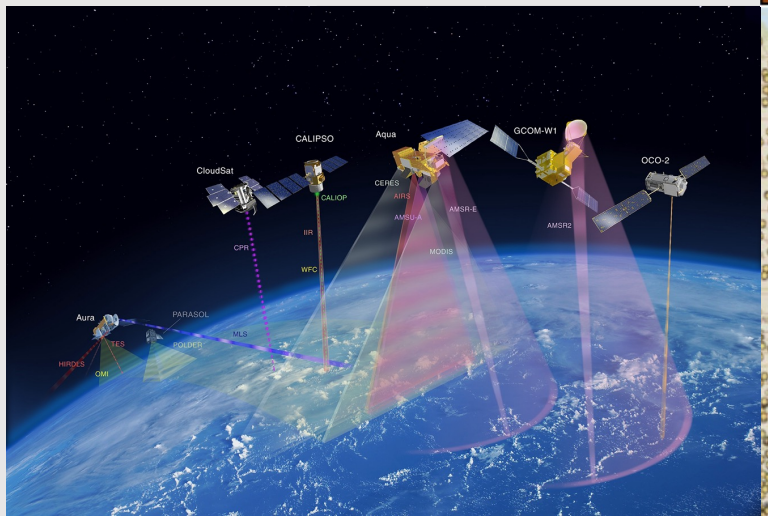
Approccio: sfruttare tutti i dati disponibili

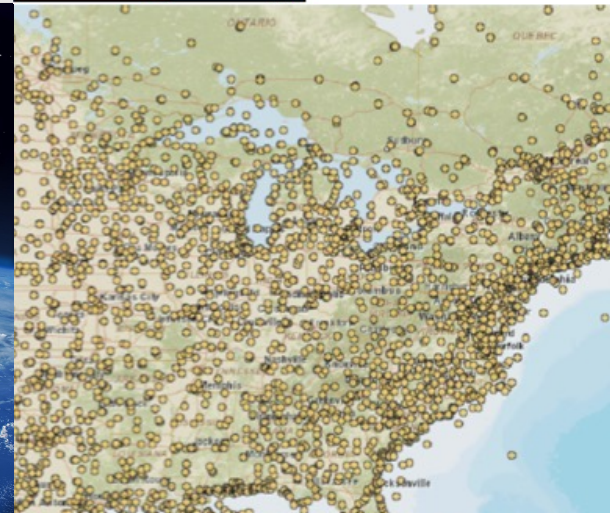
- Dati simulati generati da modelli fisici
  - Modelli di previsione numerica del tempo (NWP)
  - Modelli di circolazione generale (GCM)
  - Modelli climatici regionali (RCM)



- Dati di rianalisi
  - Prodotti su griglia dall'<u>assimilazione dei dati</u>:

applica leggi fisiche alle osservazioni





- Dati osservativi
  - Dati di telerilevamento satellitare
  - Dati in situ

Compiti di ML alla base di molti di questi problemi

- ❑ Fusione dei dati
- ❑ Previsione
- ❑ Apprendimento non supervisionato
- ❑ Allineamento probabilistico di domini
- ❑ Downscaling e interpolazione

Metodi di IA

- Apprendimento non supervisionato
- Deep learning generativo

Scaletta per il resto della giornata

I. **Contesto:** la rivoluzione dell'IA nelle previsioni meteo

Implicazioni per lo studio dei cambiamenti climatici

II. **Metodi di IA:** che cos'è l'apprendimento auto-supervisionato?

Applicazioni a clima e meteo

III. **Metodi di IA:** che cos'è il deep learning generativo?

Applicazioni a clima e meteo

Parte I: Contesto:
la rivoluzione dell'IA nelle previsioni meteo

La rivoluzione silenziosa delle previsioni meteo

- La previsione numerica del tempo (NWP) è progredita costantemente negli ultimi decenni
- Un grande risultato scientifico
  - Un problema di simulazione complesso risolto **di routine** nei centri di previsione operativi di tutto il mondo
- Forte valore socioeconomico
  - Mitigazione dei rischi climatici, logistica, prevenzione delle perdite materiali, …

Crediti delle prime slide: 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

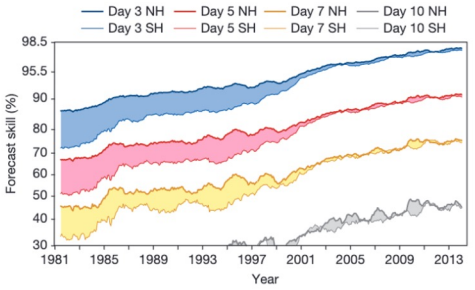
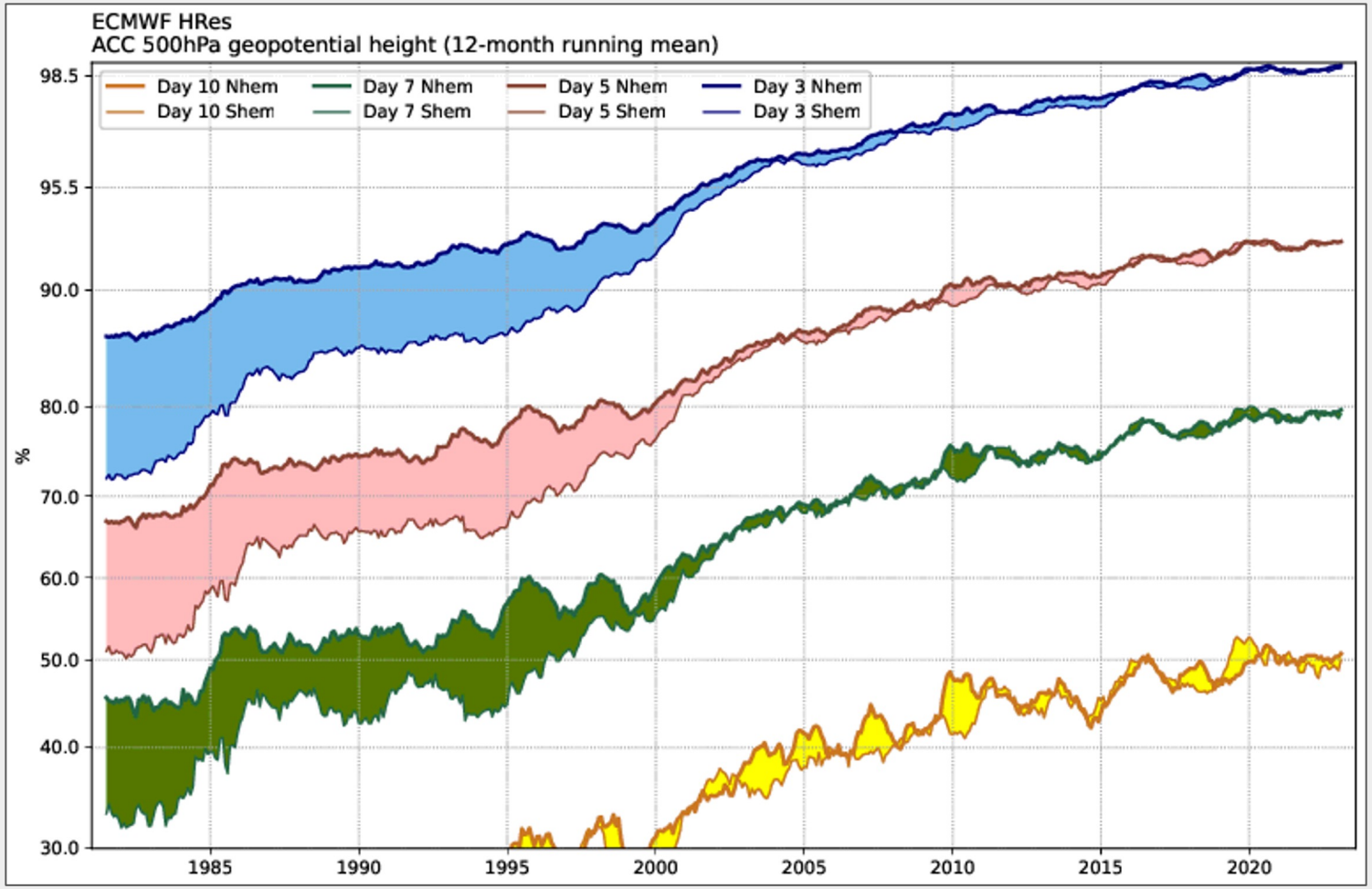


Figure 1 | A measure of forecast skill at three-, five-, seven- and ten-day ranges, computed over the extra-tropical northern and southern hemispheres. Forecast skill is the correlation between the forecasts and the verifying analysis of the height of the 500-hPa level, expressed as the anomaly with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

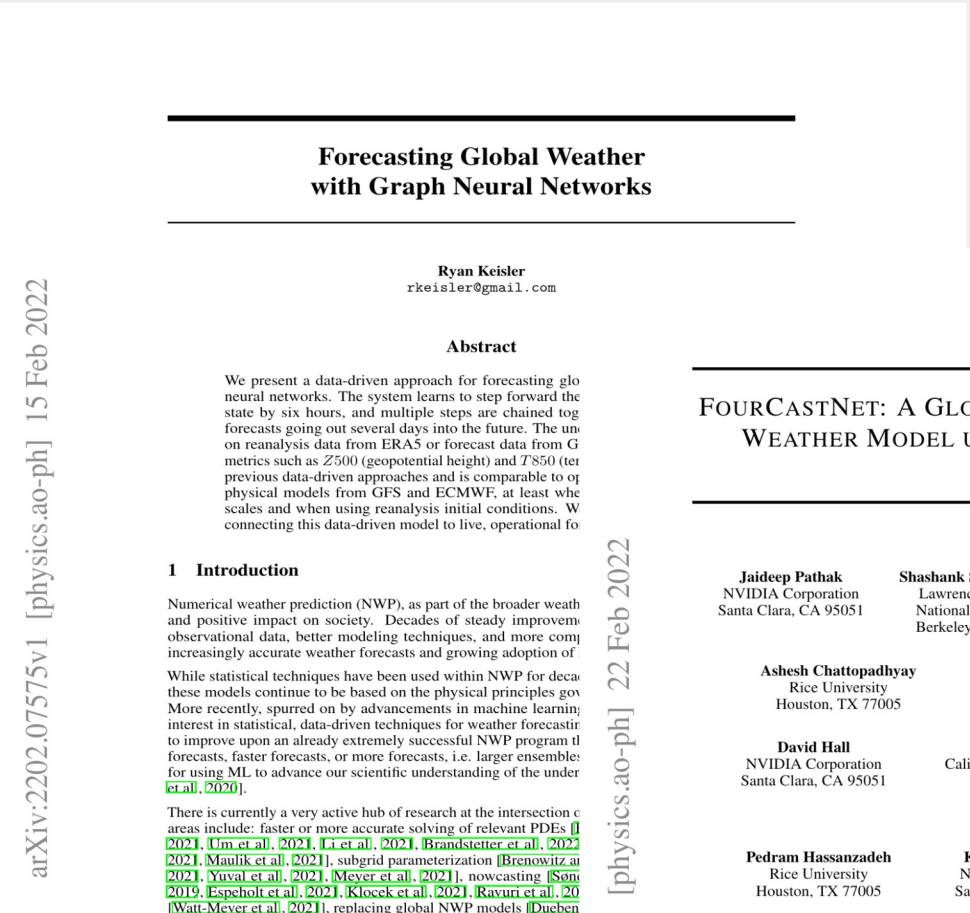


Prestazioni delle previsioni globali dell'ECMWF in 4 decenni

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

Primi modelli IA-NWP

- Dal 2022 diversi gruppi di ricerca hanno usato reti neurali profonde per le previsioni meteo a medio termine
- La qualità di alcuni di questi modelli è diventata ~equivalente a quella dei migliori modelli NWP



Forecasting Global Weather with Graph Neural Networks

Ryan Keisler
rkeisler@gmail.com

Abstract

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

1 Introduction

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

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

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

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

[physics.ao-ph] 22 Feb 2022

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

A PREPRINT

Jaideep Pathak
NVIDIA Corporation
Santa Clara, CA 95051

Shashank Subramanian
Lawrence Berkeley
National Laboratory
Berkeley, CA 94720

Peter Harrington
Lawrence Berkeley
National Laboratory
Berkeley, CA 94720

Sanjeev Raja
University of Michigan
Ann Arbor, MI 48109

Ashesh Chattopadhyay
Rice University
Houston, TX 77005

Morteza Mardani
NVIDIA Corporation
Santa Clara, CA 95051

Thorsten Kurth
NVIDIA Corporation
Santa Clara, CA 95051

David Hall
NVIDIA Corporation
Santa Clara, CA 95051

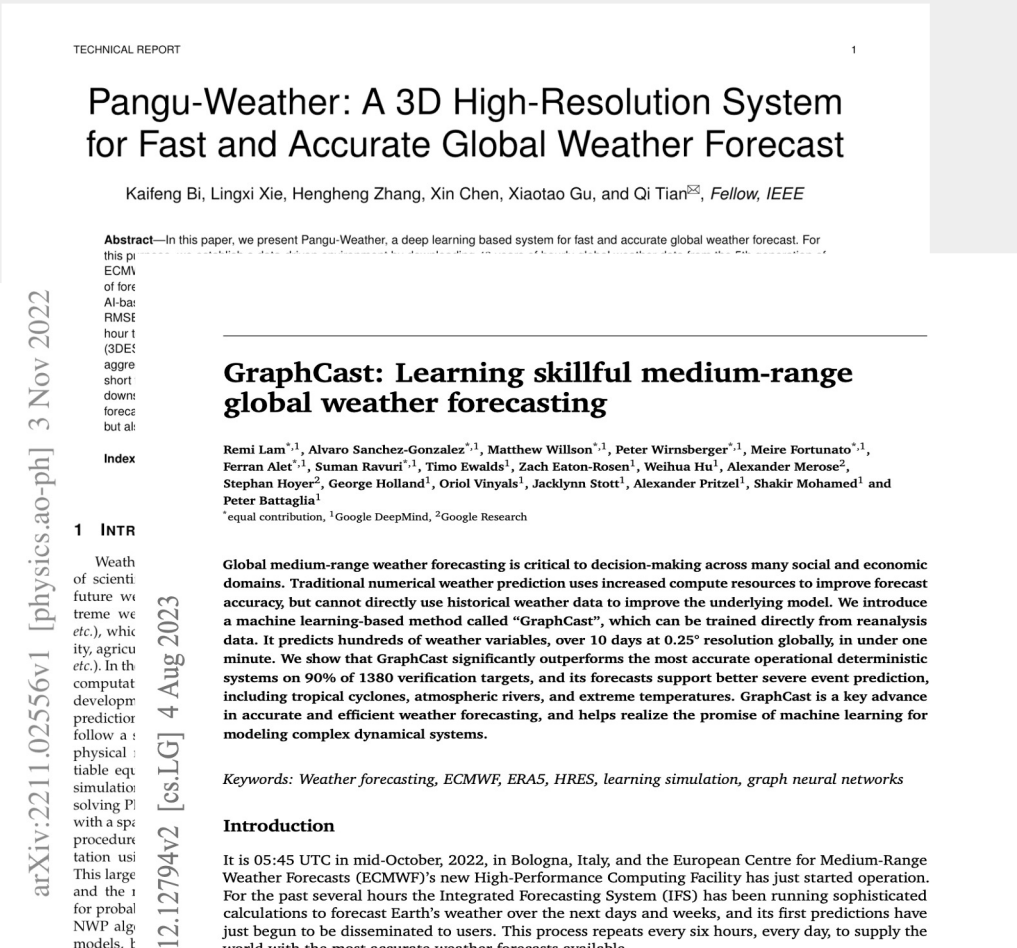
Zongyi Li
California Institute of Technology
Pasadena, CA 91125
NVIDIA Corporation
Santa Clara, CA 95051

Kamyar Azizzadenesheli
Purdue University
West Lafayette, IN 47907

Pedram Hassanzadeh
Rice University
Houston, TX 77005

Karthik Kashinath
NVIDIA Corporation
Santa Clara, CA 95051

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



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

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

Abstract—In this paper, we present Pangu-Weather, a deep learning based system for fast and accurate global weather forecast. For
this p
ECM
of fore
AI-ba
RMSE
hour t
(3DE
aggre
short
downs
foreca
but al

Index

1 INTR

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

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

12.12794v2 [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

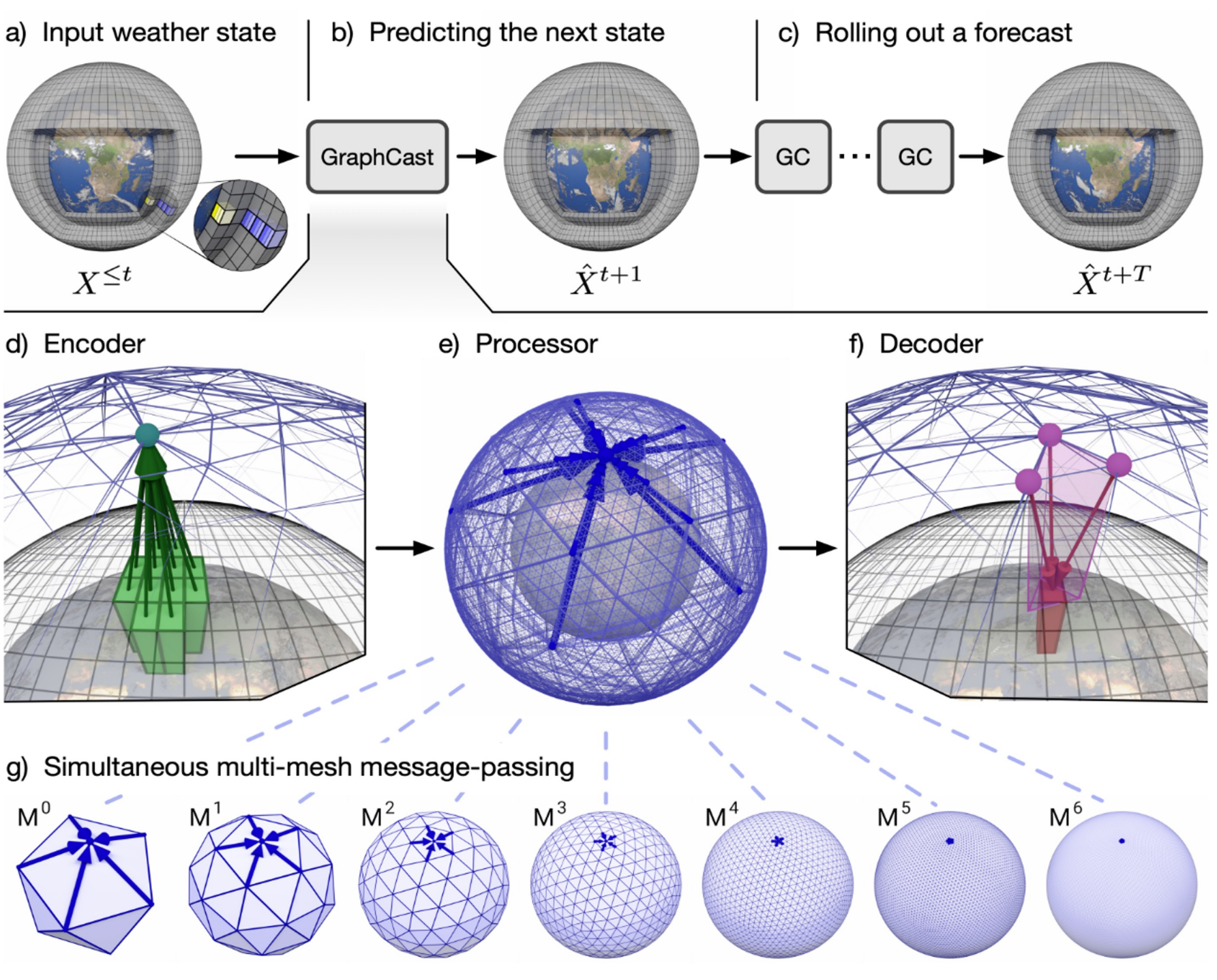
Introduction

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

Risparmi di calcolo di questi primi modelli IA-NWP

Modello	**Addestramento**	**Inferenza**
Pangu-Weather	16 giorni su 192 GPU V100	Previsione a 24 ore 1xV100 1,4 secondi su
GraphCast	21 giorni su 32 Cloud TPU V4 (~128 GPU)	Previsione a 10 giorni 1 Cloud TPU V4 < 60 secondi
NWP (ECMWF HRES) ***Risoluzione più alta**	-	Previsione a 10 giorni 1/3/6 ore Cluster HPC da

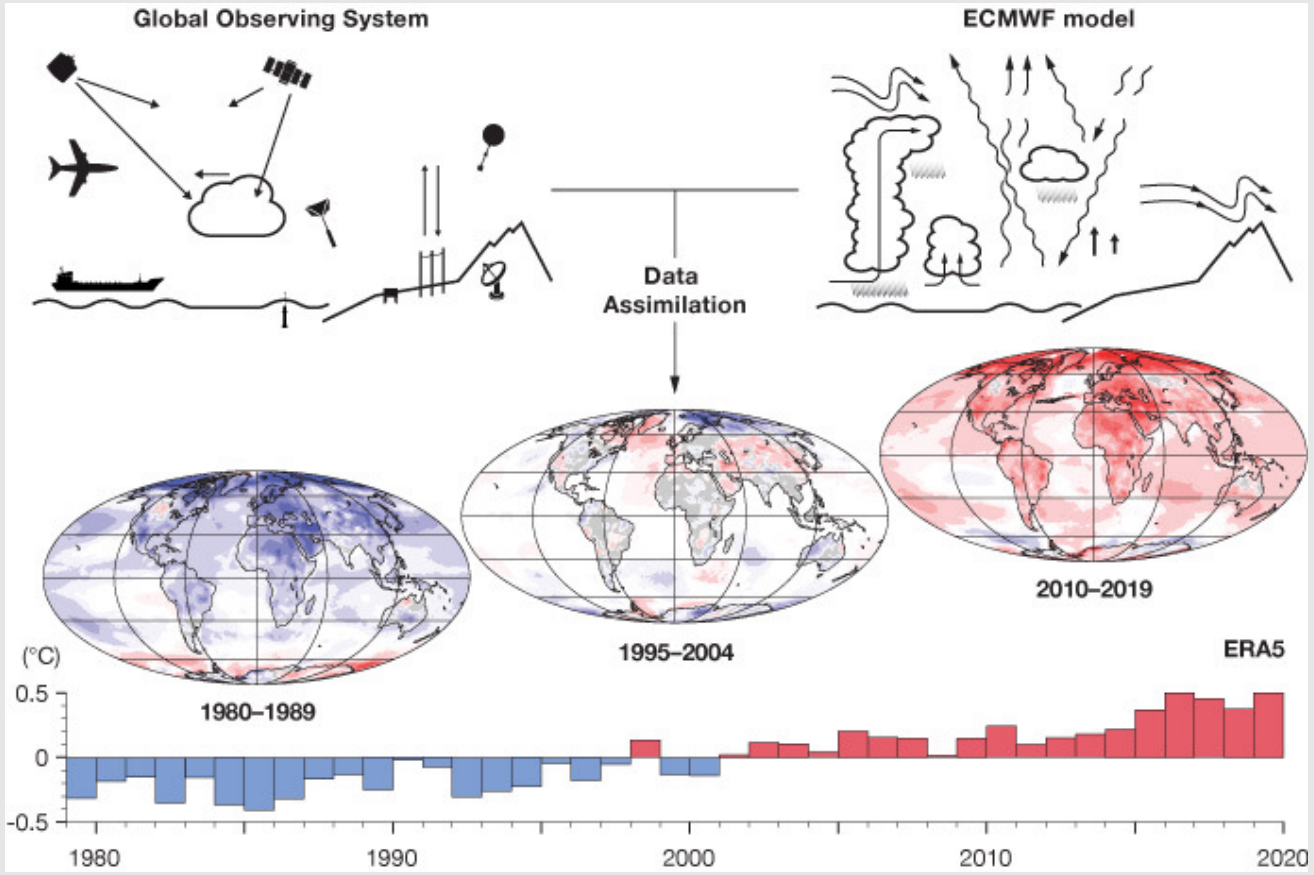
Architettura di GraphCast



doi:10.48550/arXiv.2212.12794

IA per le previsioni meteo: un nuovo paradigma

Assimilazione dei dati → Dataset ERA5

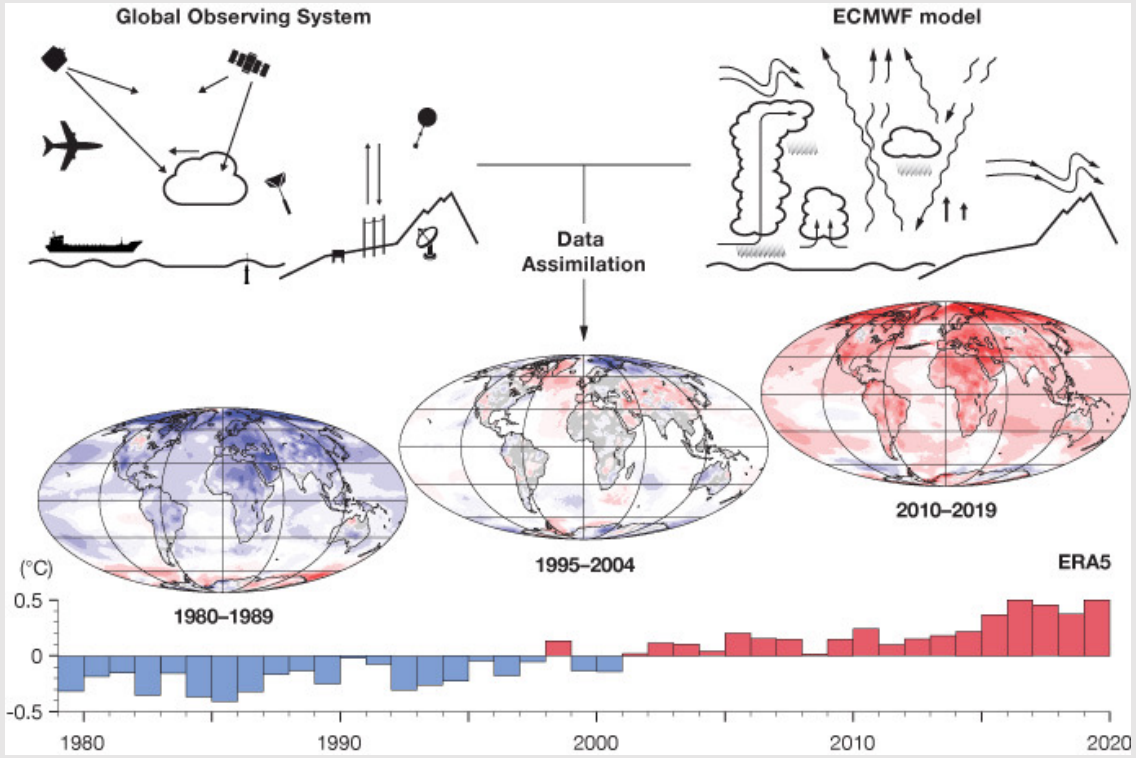


Dataset ERA5 → Simulazione con rete neurale → $P(\text{Previsione})$

Figura: Centro europeo per le previsioni meteorologiche a medio termine (ECMWF)

La rivoluzione dell'IA nelle previsioni meteo

Dal 2022 diversi modelli di deep learning hanno mostrato prestazioni di previsione paragonabili o **MIGLIORI** della previsione numerica del tempo (NWP), il precedente stato dell'arte.



Dati di addestramento: ERA5, un dataset di r
prodotto tramite assimilazione dei dati

Compito di addestramento: autoregressione
a 6-24 ore

- Esteso con rollout per prevedere a 7-10 giorni

[Sito web del Centro europeo per le previsioni meteorologiche a medio termine (ECMWF)]

Rollout

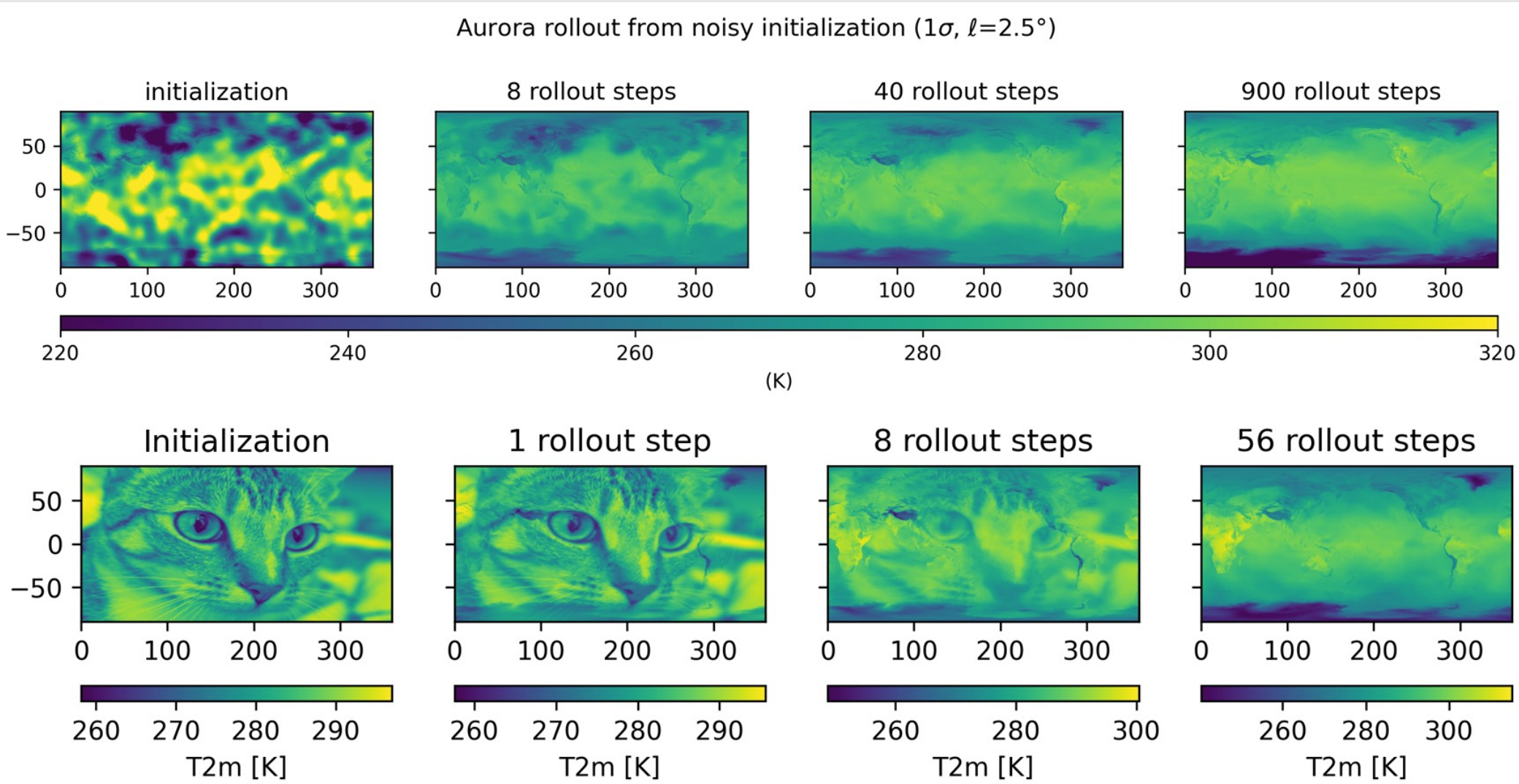


Figure A.9: 2-meter temperature predicted at different time stamps when (top) inputs are perturbed with an additive Gaussian random field with coefficient of variation σ^2 (where σ is the standard deviation of the physical variable) and correlation length of 10 pixels (equivalent to $2.5°$), (bottom)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

Time

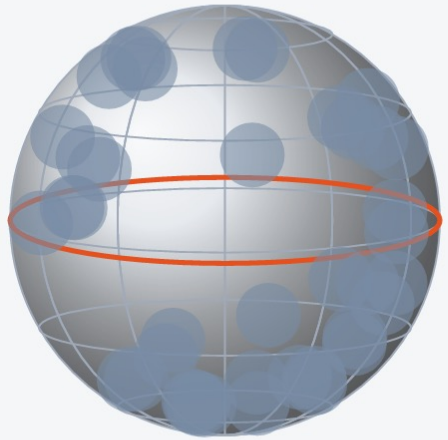
45+ years

1979 — 2024

Hourly

~400,000 timesteps

Space



721 × 1440 × 240

0.25° resolution

37 levels × 6 vars + 18 surface

Memory

per atmospheric state:

0.25° **0.94 GB**

×6.25

0.1° **5.9 GB**

Dataset di rianalisi ERA5 dell'ECMWF ≈ 13 Petabytes

ArchesWeatherGen: Generatore meteo di IA frugale [Couairon et al., *Science Advances*, 2026]

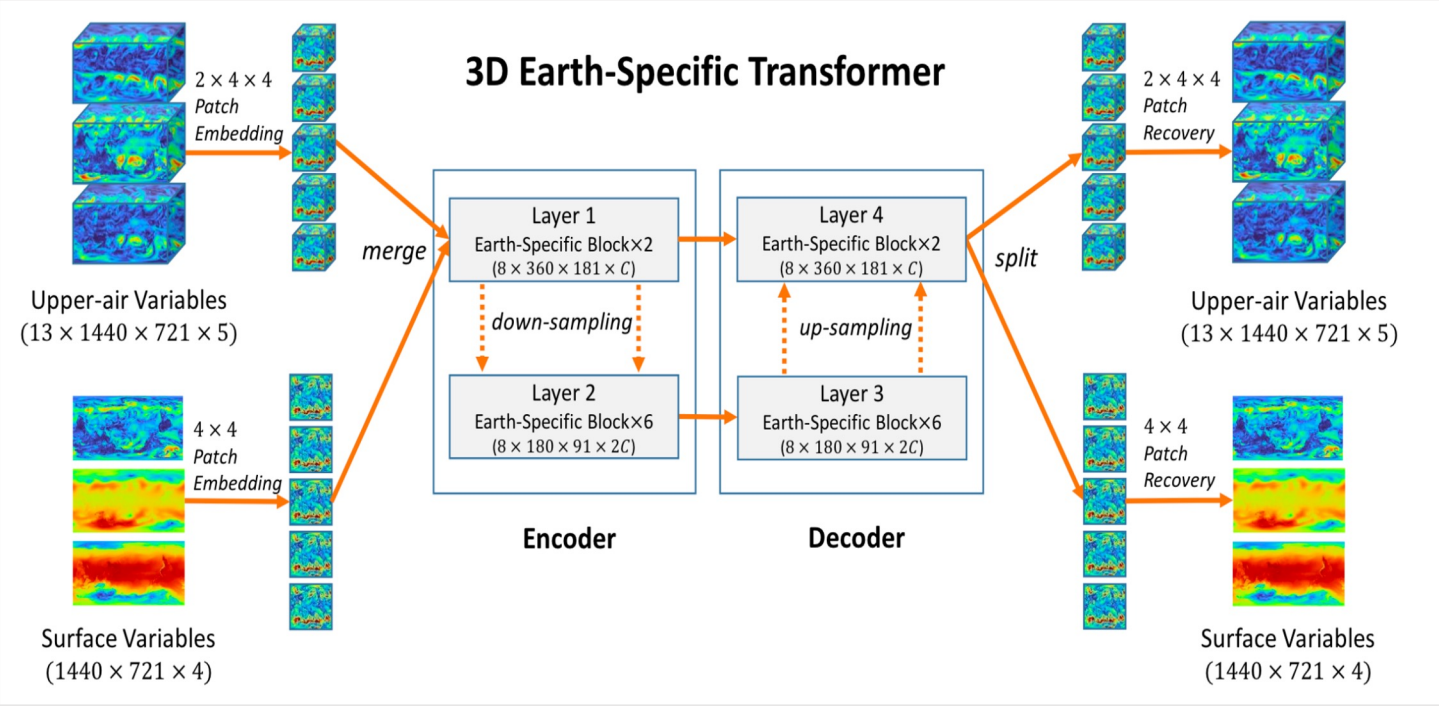
Figura: Emmanuel de Bézenac (INRIA)



ArchesWeather

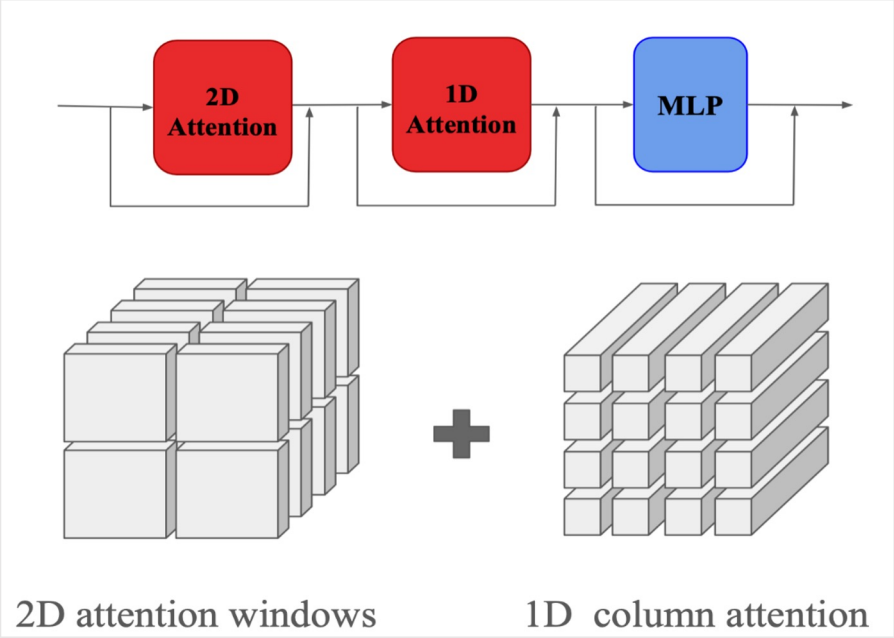
[Couairon et al., ArchesWeather: an efficient AI weather model at 1.5º resolution ICML 2024 workshop on Earth System Modeling]

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



Novità di ArchesWeather:

- Addestramento a una risoluzione dei dati più grossolana
- Sostituzione dell'attenzione 3D con:



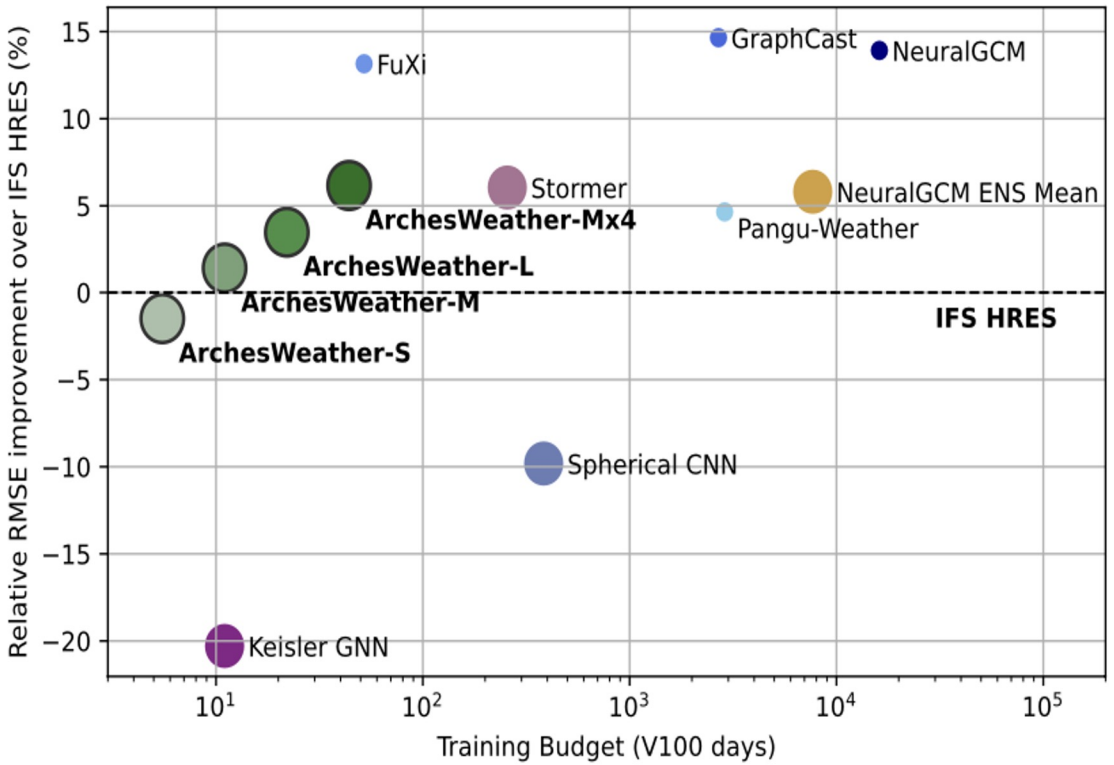
latitudine + longitudin

quota

ArchesWeather

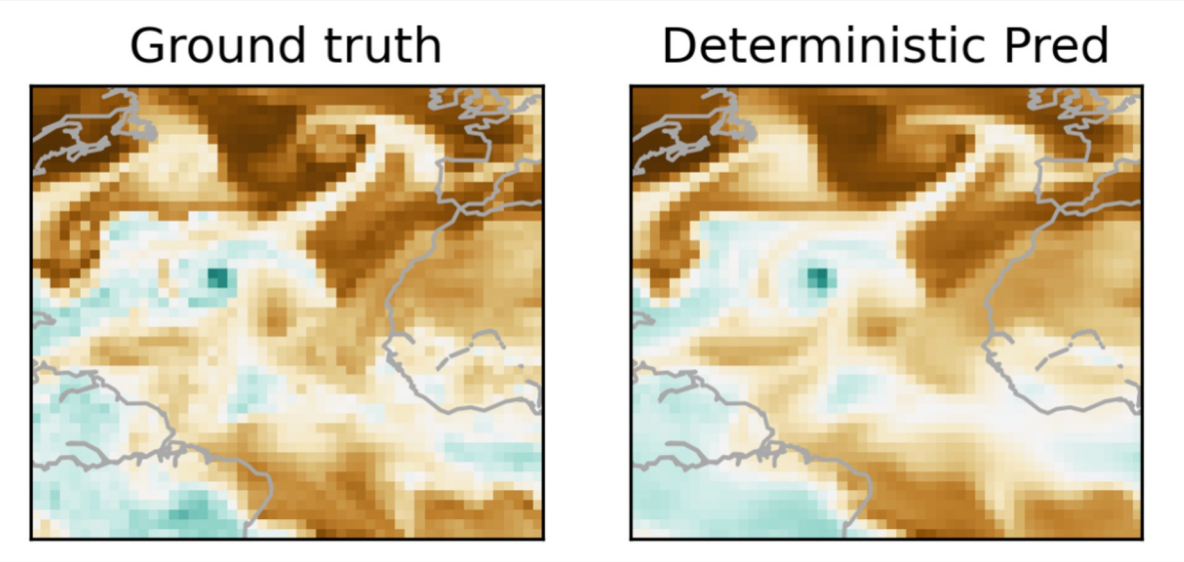
Compromesso sulle prestazioni

	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, lead time 24 ore

Perché passare alla generazione di ensemble…



Previsione Q700 a 24 ore
data di inizializzazione: 28 sett. 2019

- La previsione deterministica mostra un livellamento irrealistico
- Provare la generazione di ensemble con addestramento per diffusione
- Obiettivo: ogni membro campionato deve essere più fisico

ArchesWeatherGen

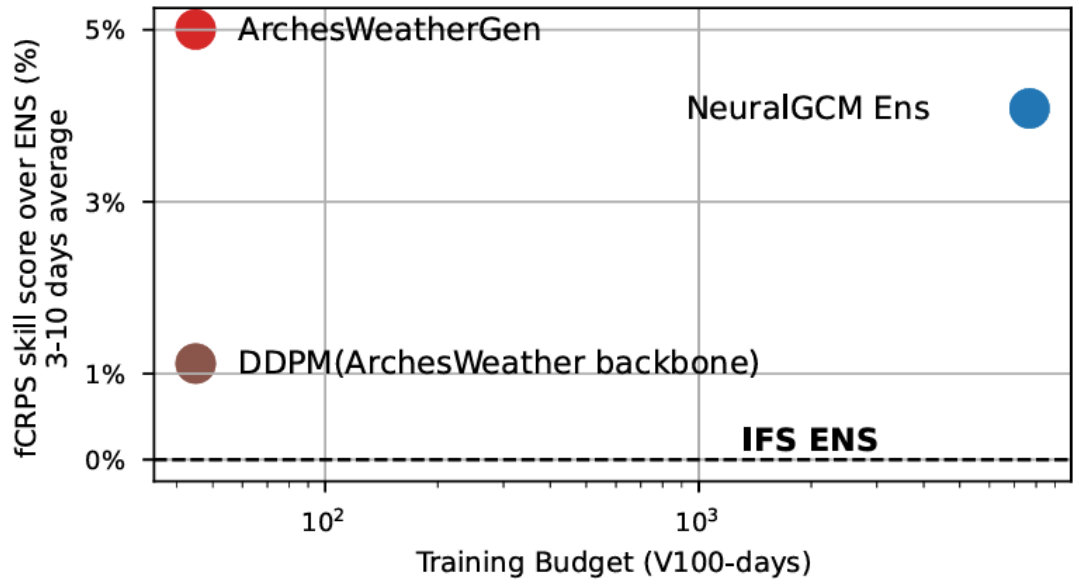
[Couairon et al., *Science Advances*, 2026]

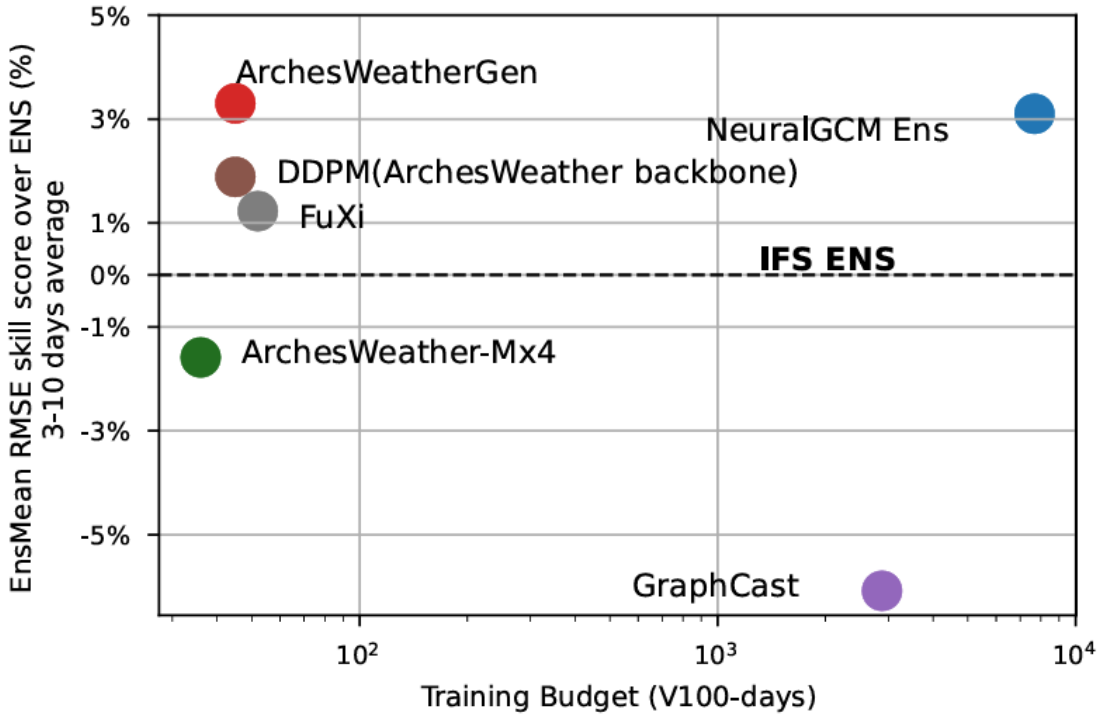
Backbone: ArchesWeather (variante dello Swin U-Transformer), 44 M di parametri

Addestrato con Flow Matching

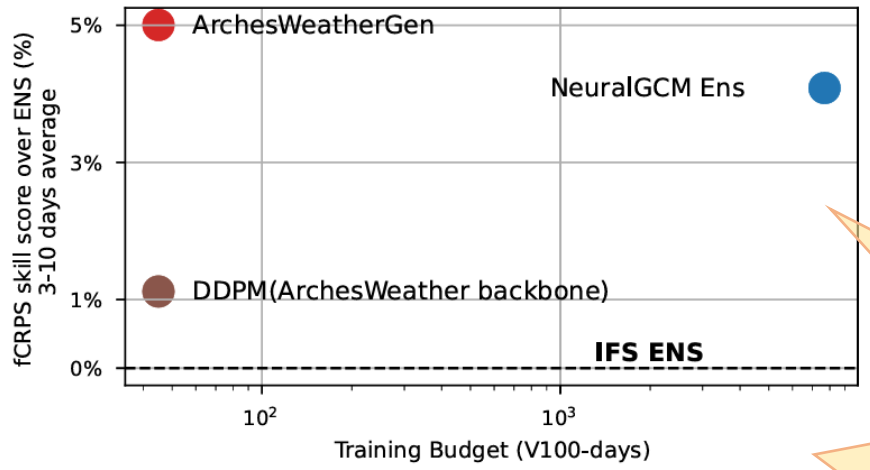
Inferenza con 25 passi di campionamento ogni 24 ore

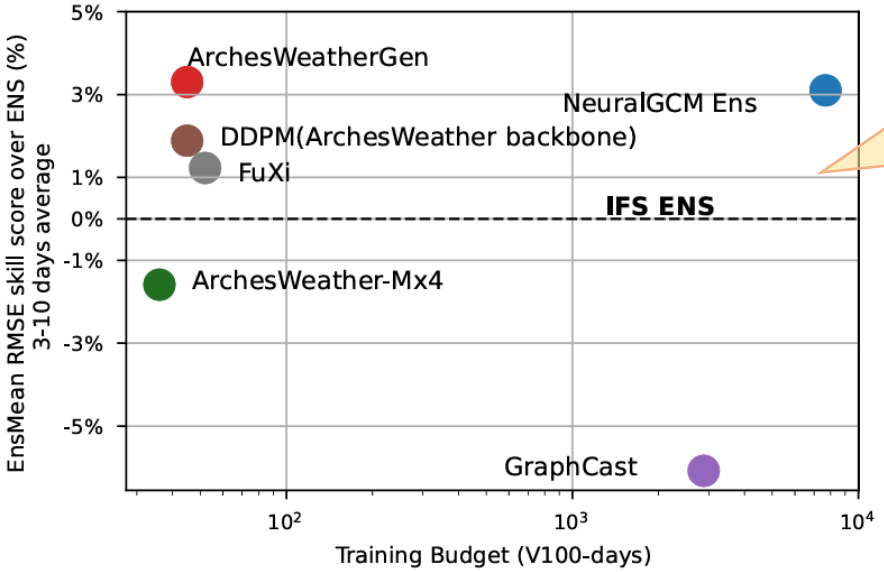
Rollout con un campione ogni 24 ore. L'output viene reinserito nel modello addestrato e si ripete.

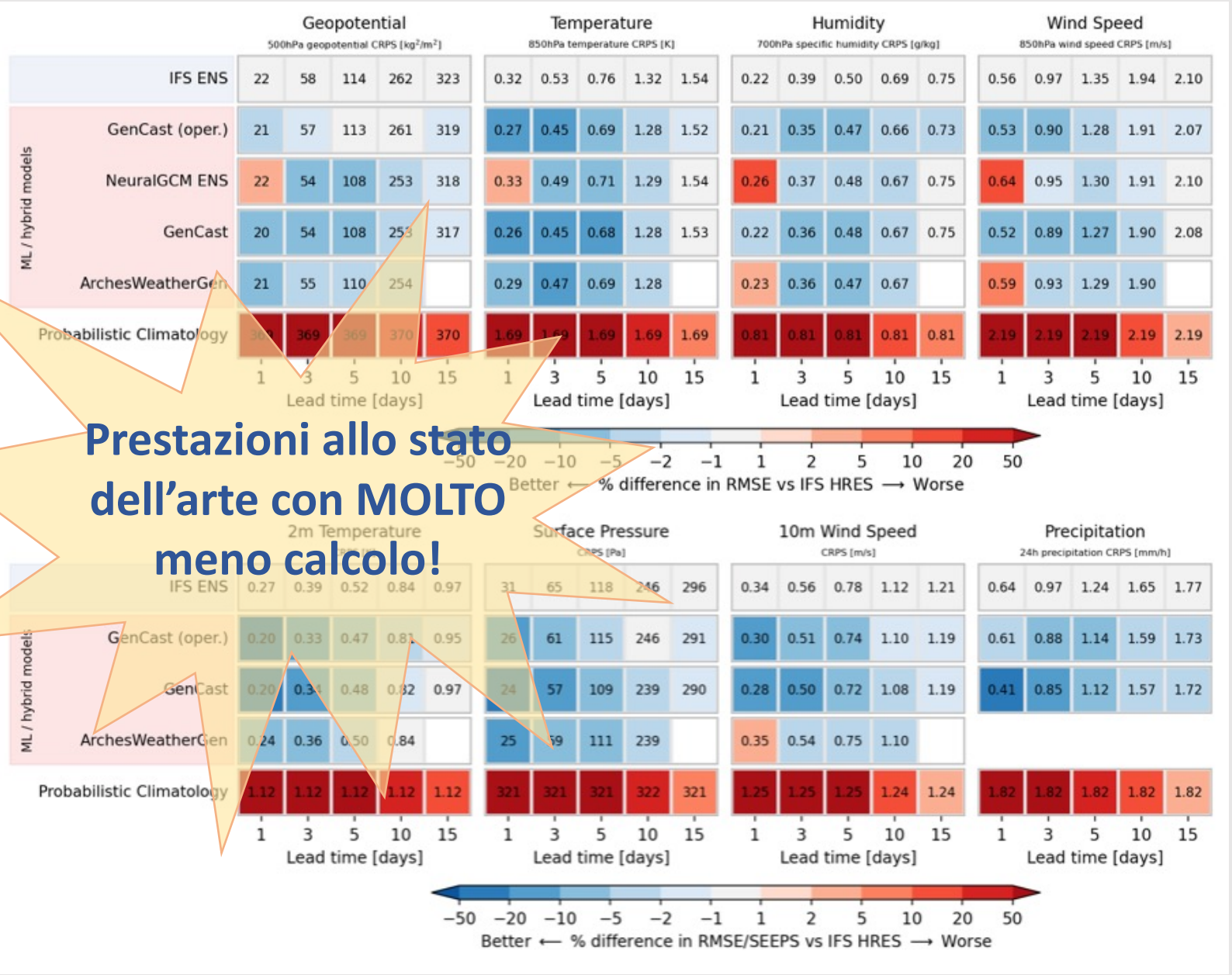




ArchesWeatherGen





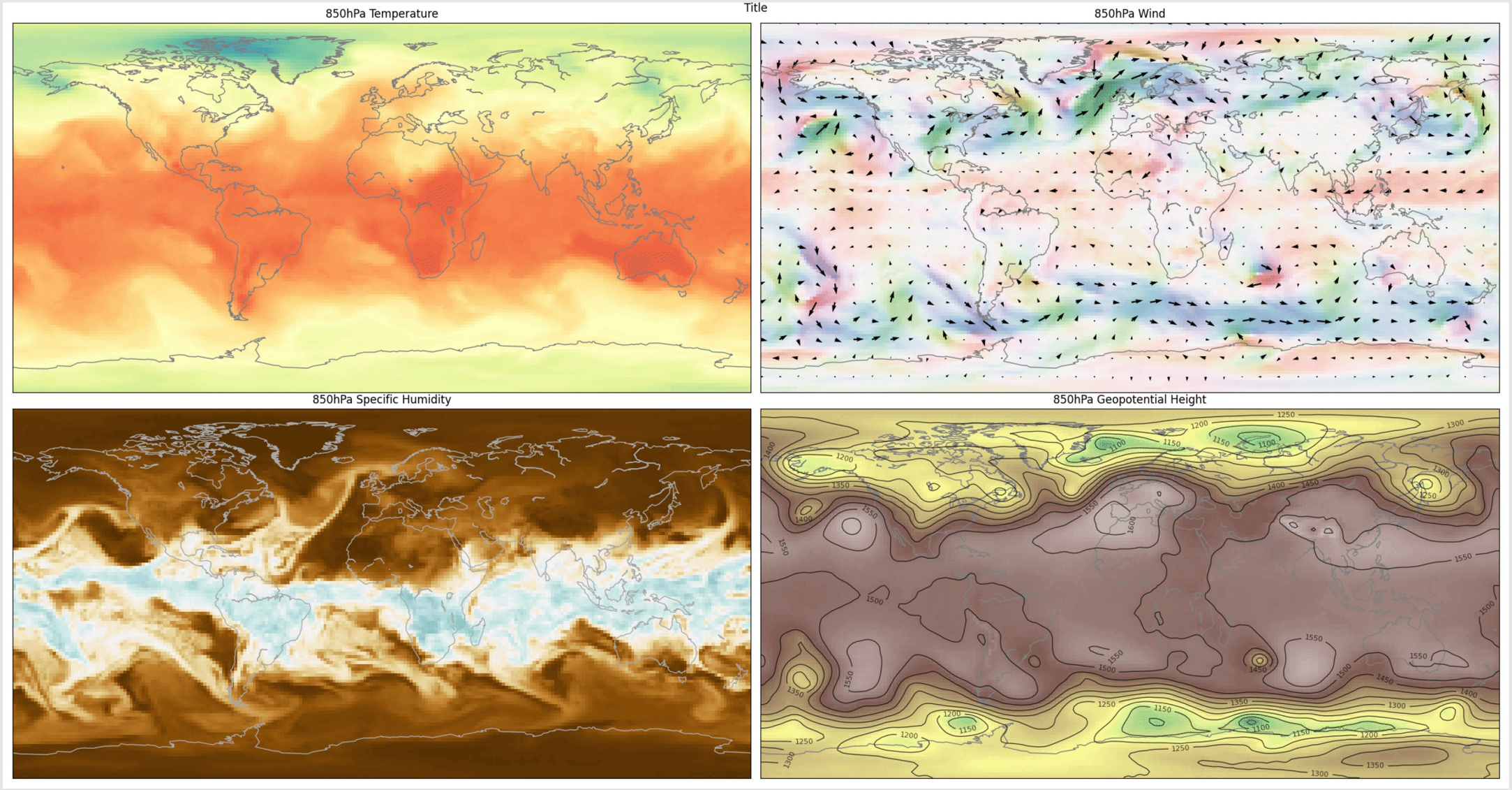


Prestazioni allo stato dell'arte con MOLTO meno calcolo!

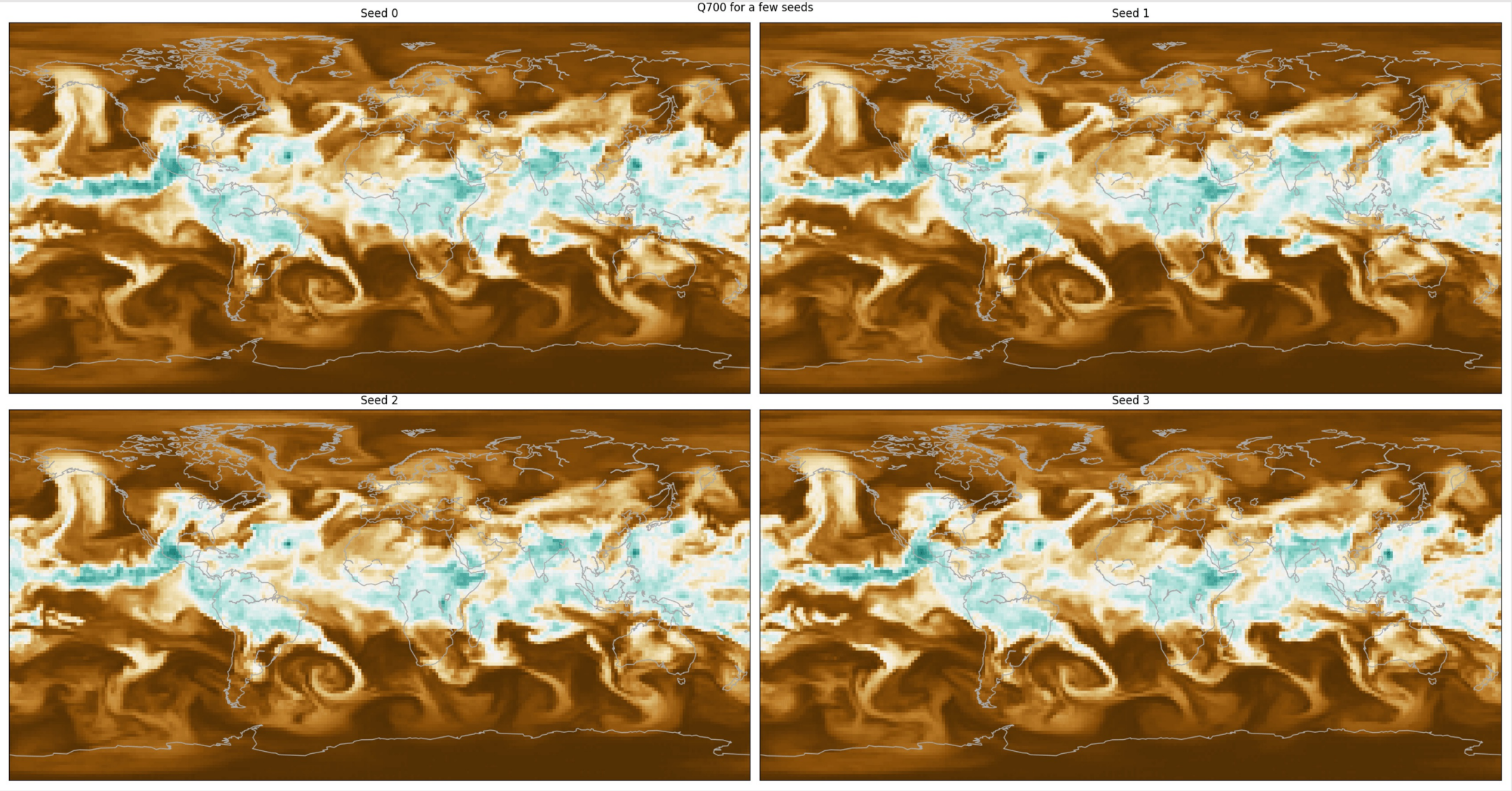
[Couairon, Singh, Charantonis, Lessig, Monteleoni, *Science Advances*, 2026]

Scorecard su WeatherBench 2 di Google

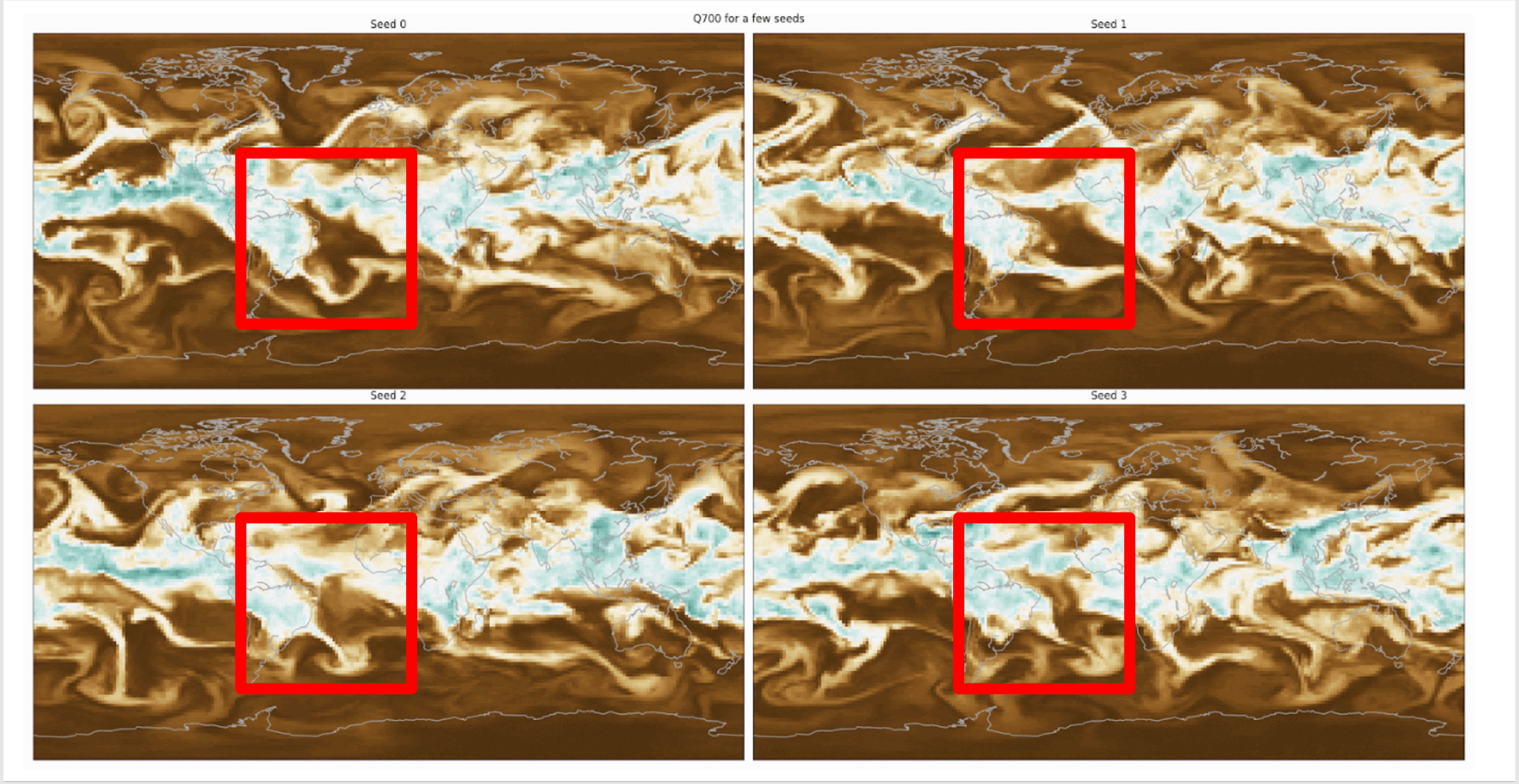
Una traiettoria di esempio



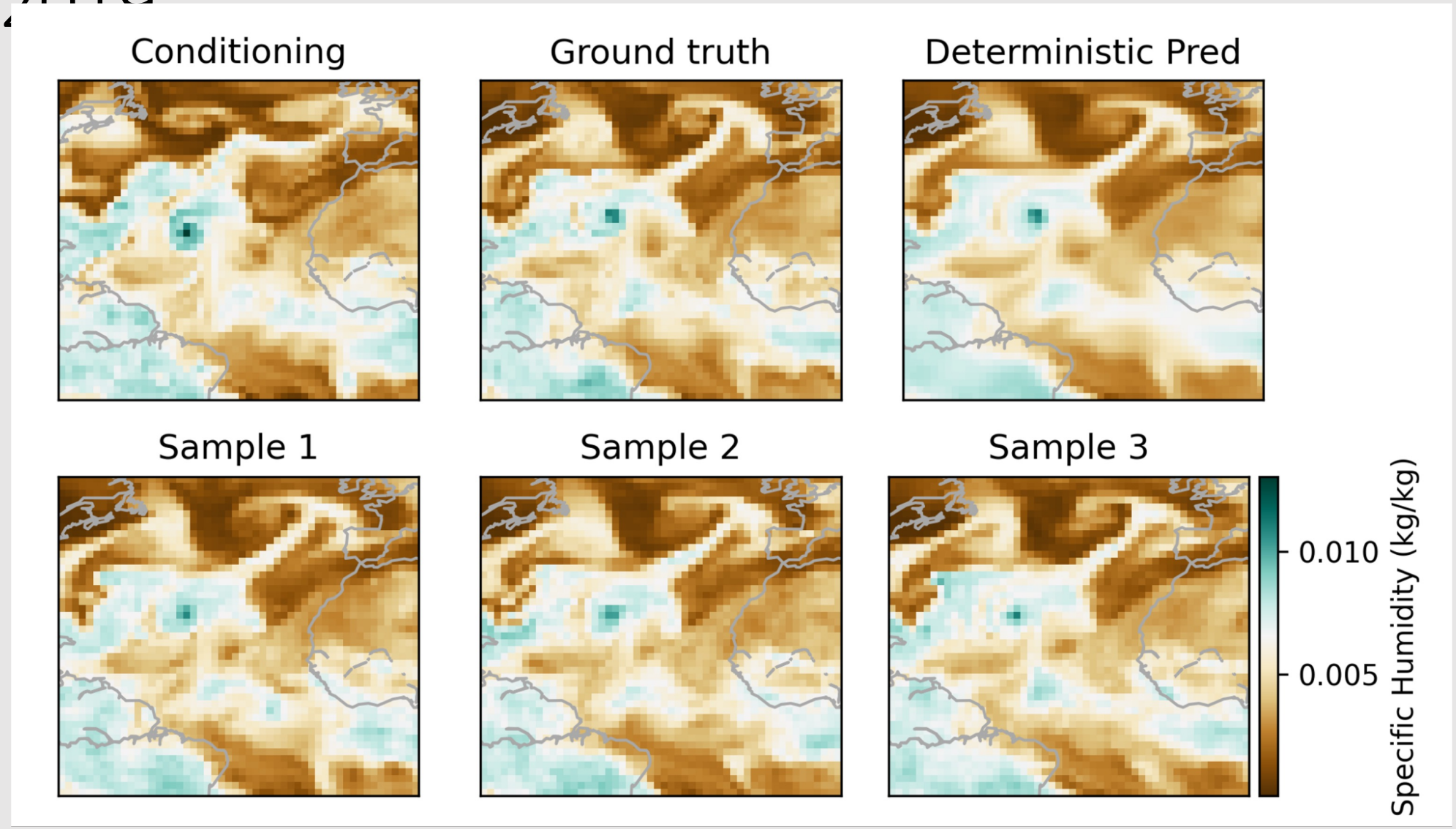
Semi diversi - Q700



Diversità dei membri dell’ensemble; lead time 10 giorni



Caso di studio: Lorenzo - Q700 - inizializzazione 25 sett. 2019



ArchesClimate: emulazione di modelli climatici con l'IA



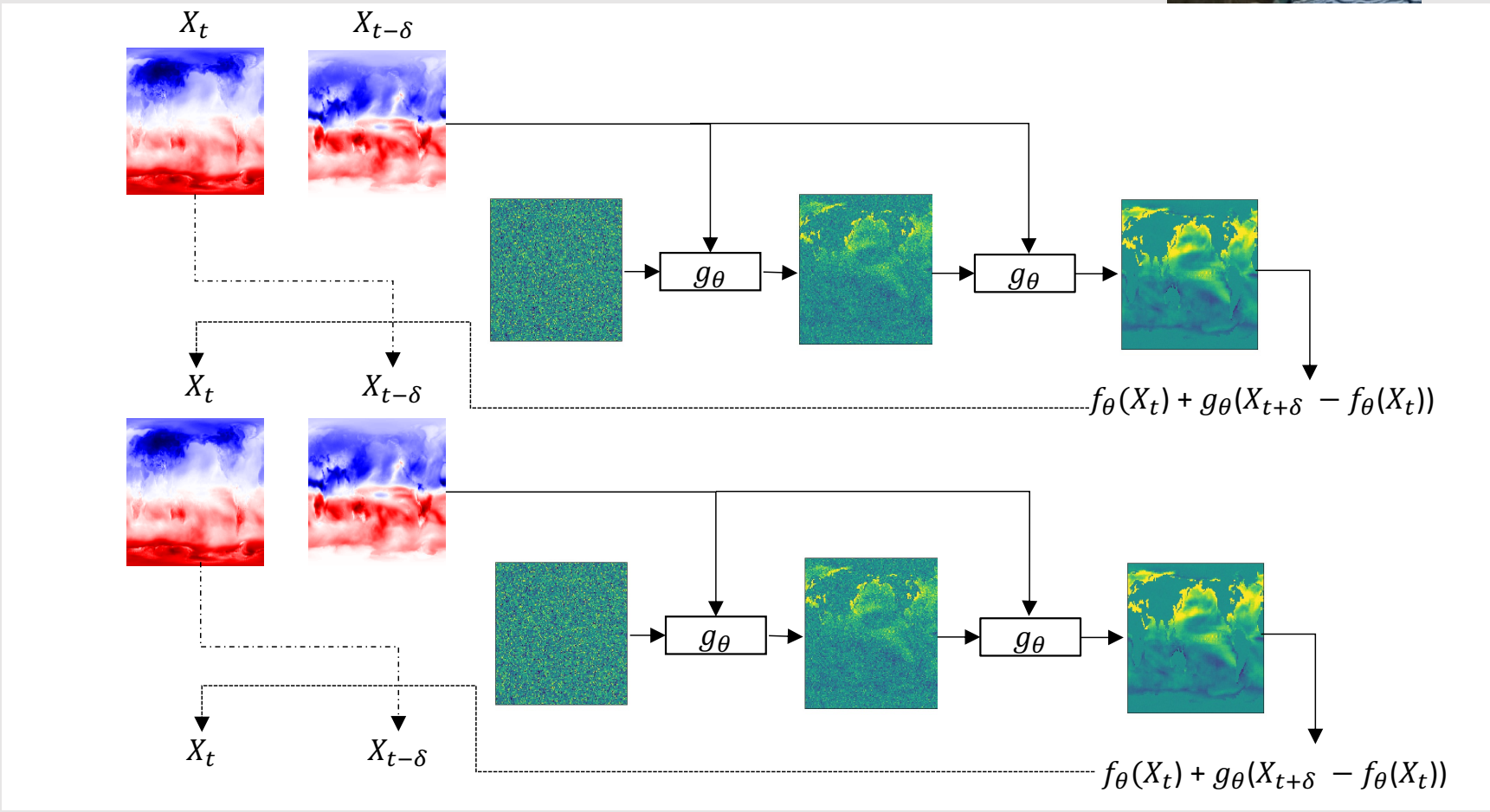
Dottorato in corso di Graham Clyne
Relatore principale: Anastase Charantonis

Pipeline: ArchesWeatherGen modificato:
- Nessuna attenzione 1D per colonna
- Inclusione dei forzanti

Dati di addestramento: simulazioni di modelli climatici (IPSL DCPP)

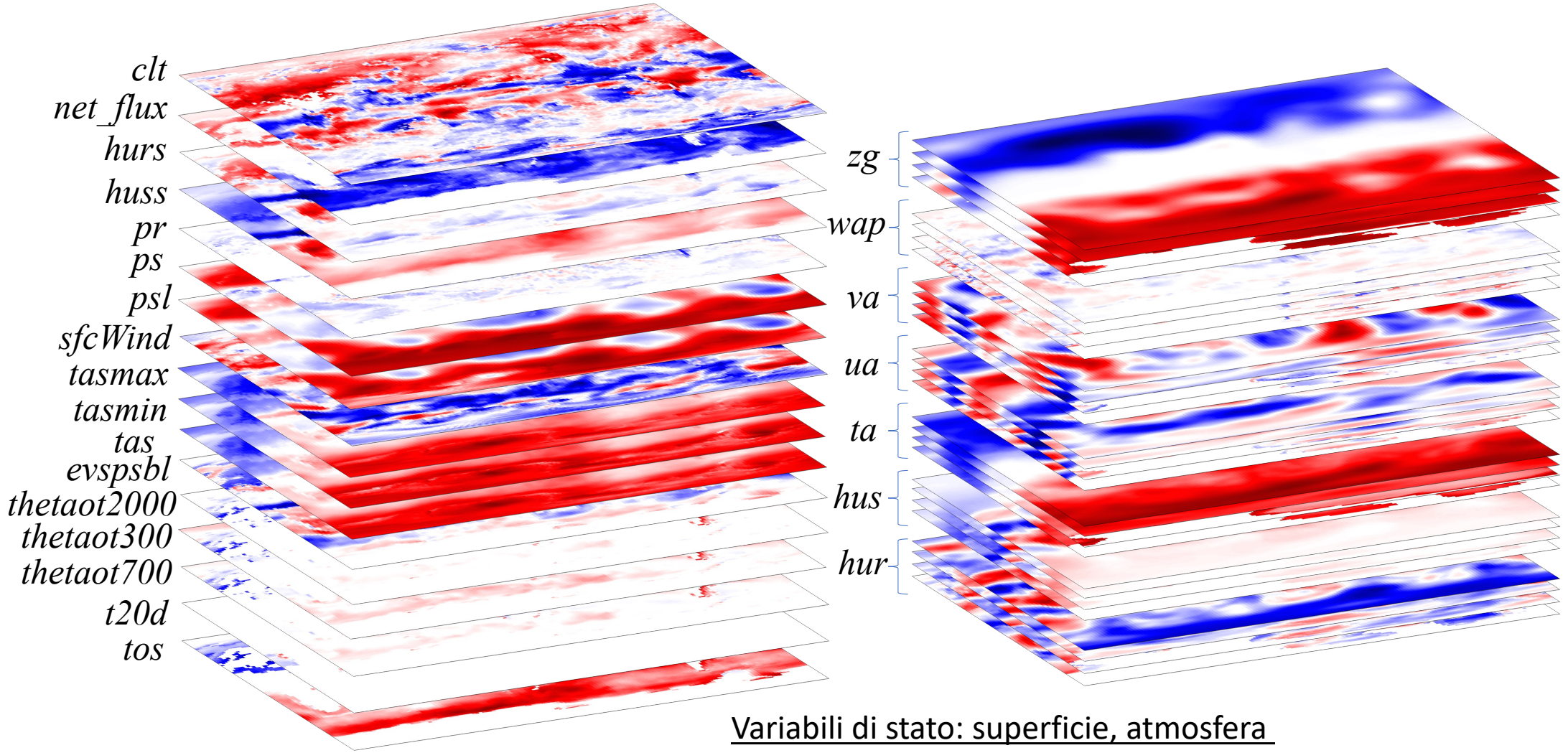
Addestramento: prevedere un mese avanti

Rollout con un campione al mese. L'output viene reinserito nel modello addestrato e si ripete.



Clyne et al., *risultati preliminari presentati a EGU 2025 e Climate Informatics 2025*

Variabili di input e forzanti

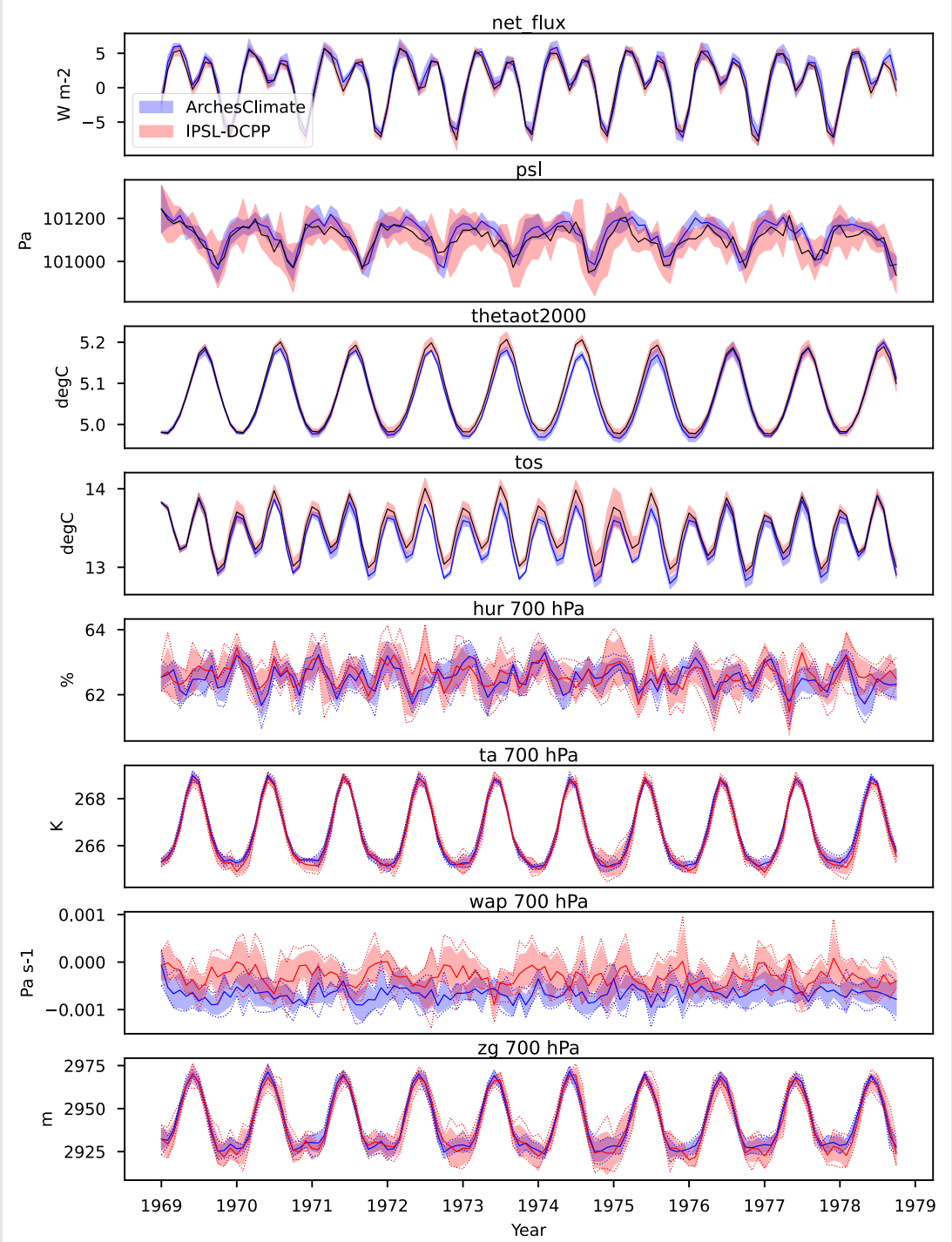


$$\begin{bmatrix} CO2 \\ CH4 \\ N2O \\ CFC11eq \\ SSI \end{bmatrix}$$

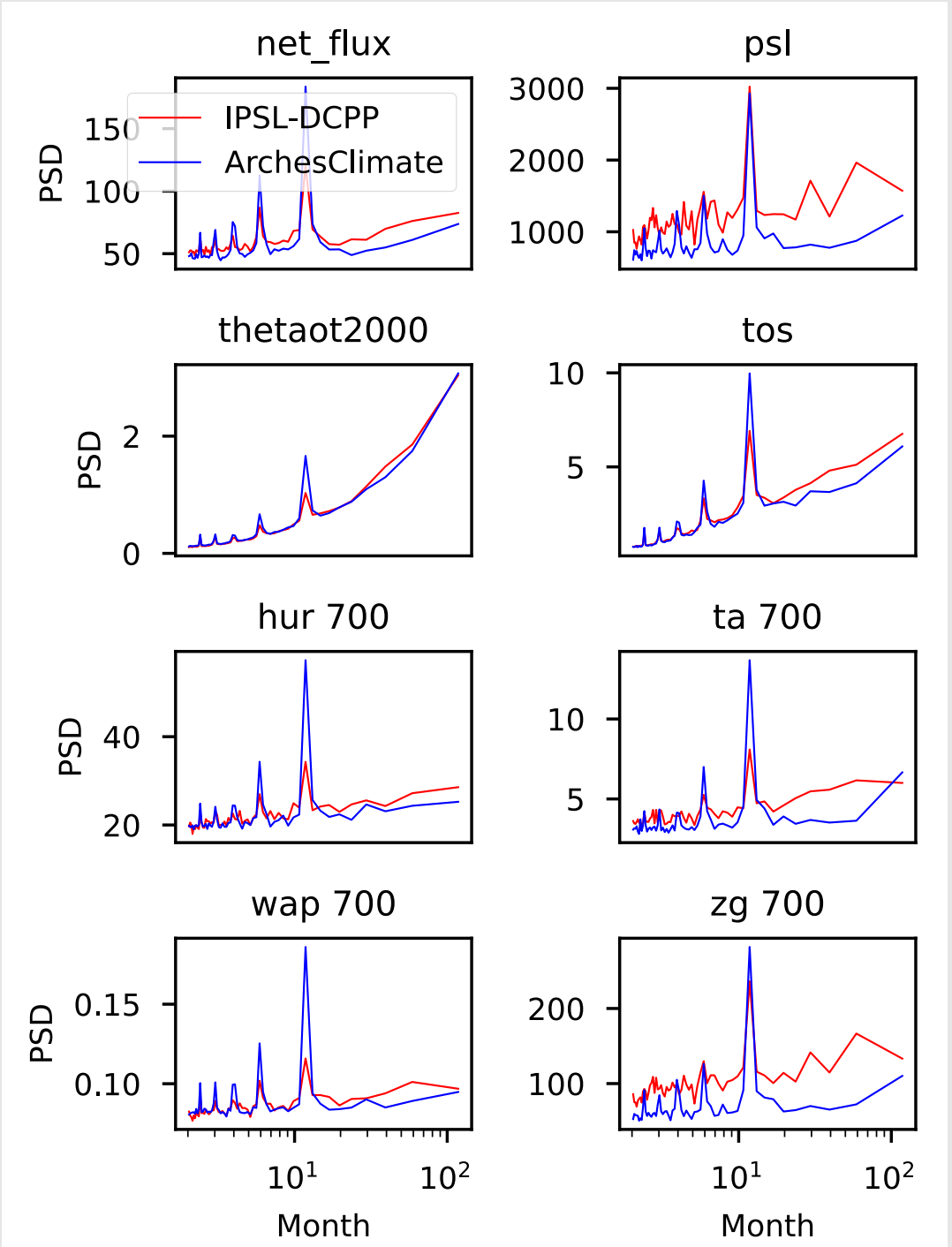
Variabili di stato: superficie, atmosfera

Forzanti

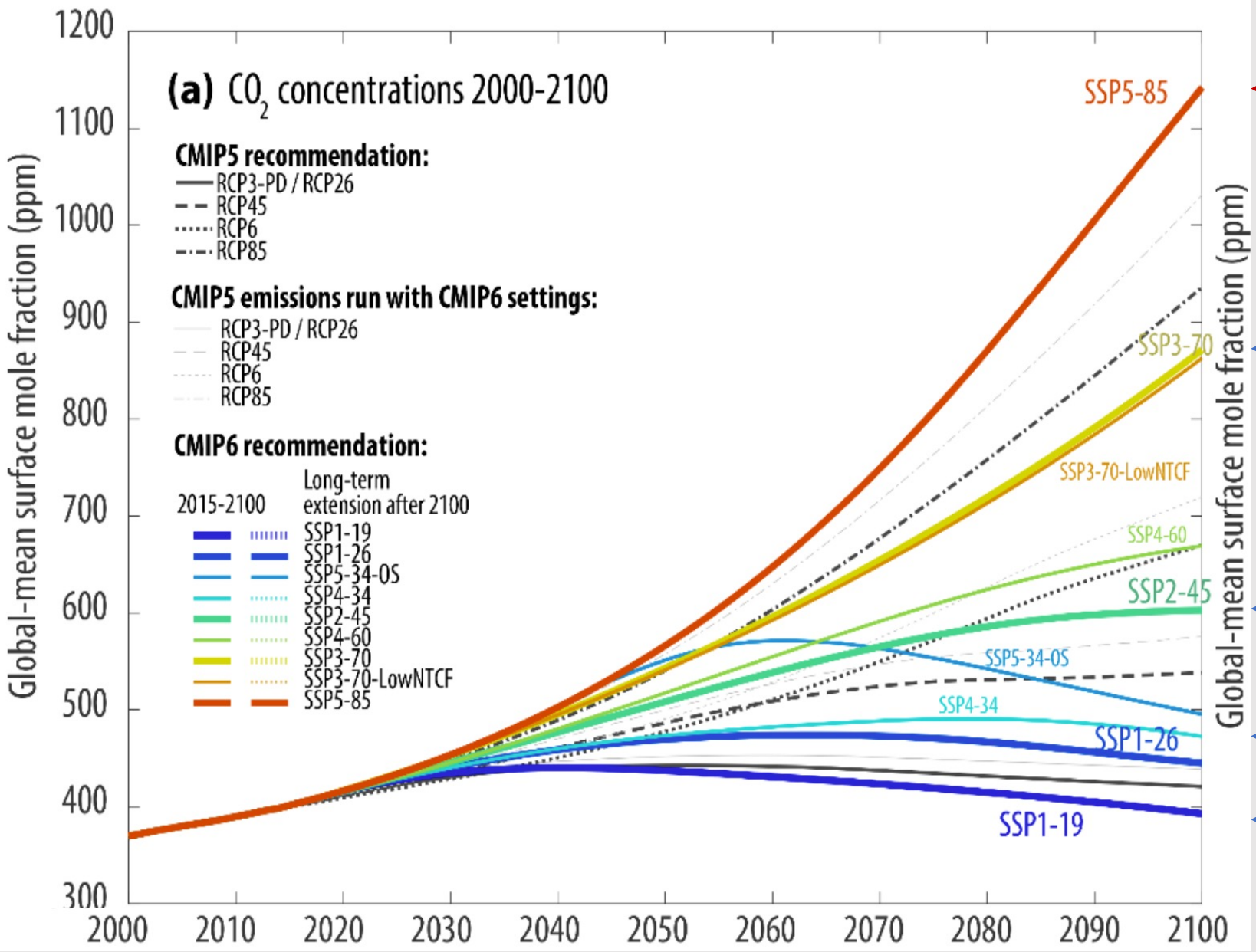
Rollout decennali



Risultati preliminari



Estrapolare gli scenari dei modelli climatici





Dati di test
Compito di estrapolazione

Dati di addestramento
Questi 4 scenari

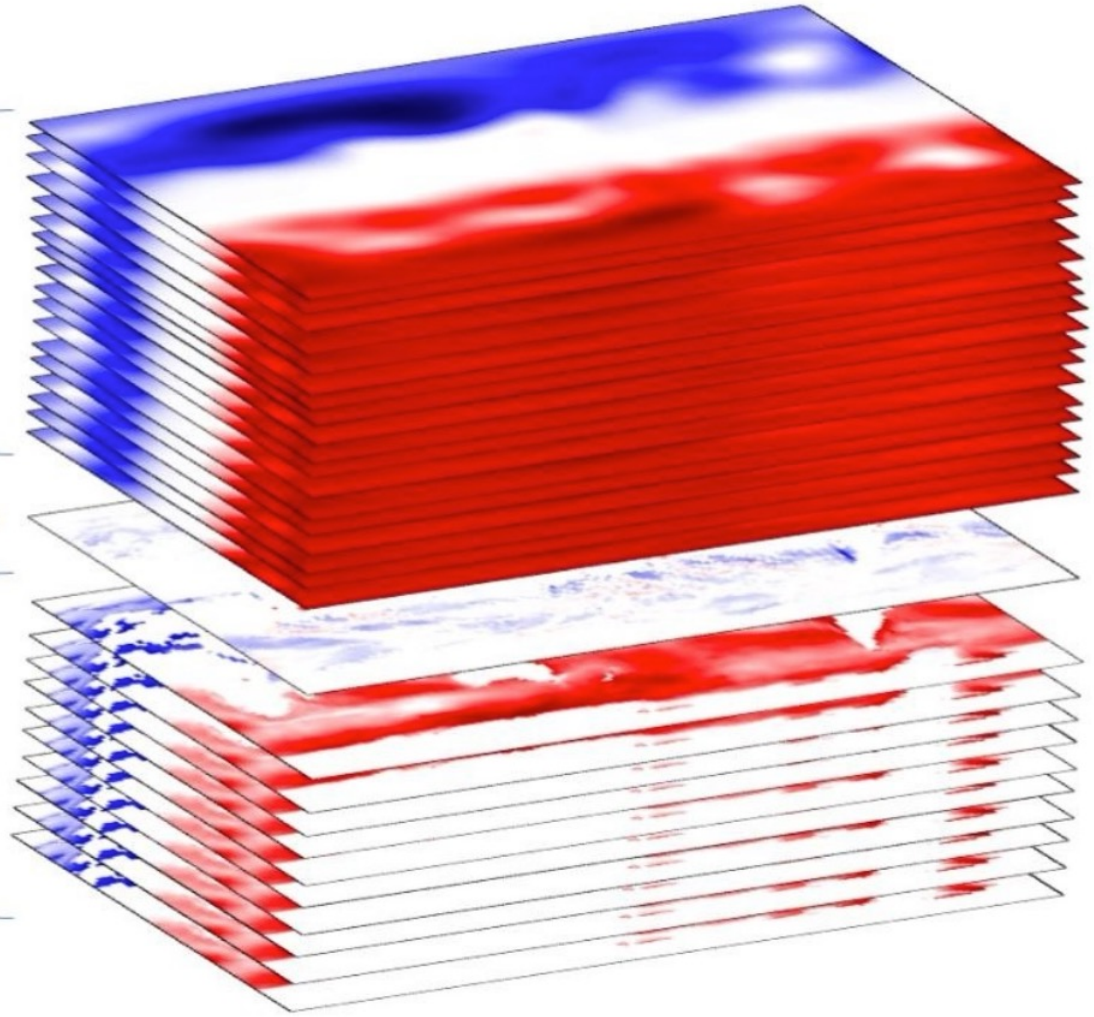
Interpolazione ed estrapolazione nello spazio degli scenari dei modelli climatici

Clyne et al., *in preparazione*

- Northward Near-Surface Wind
- Eastward Near-Surface Wind
- Precipitation
- Near-surface air temperature
- Evaporation including sublimation and transpiration
- Total positive downward flux between ocean and atmosphere
- Sea level pressure

- Specific Humidity
- Air Temperature
- Eastward Wind
- Northward Wind
- Geopotential Height

- Ocean Temperature (to 150m depth)



Estrapolazione di scenari

Riprodurre gli schemi temporali di El Niño-Oscillazione meridionale

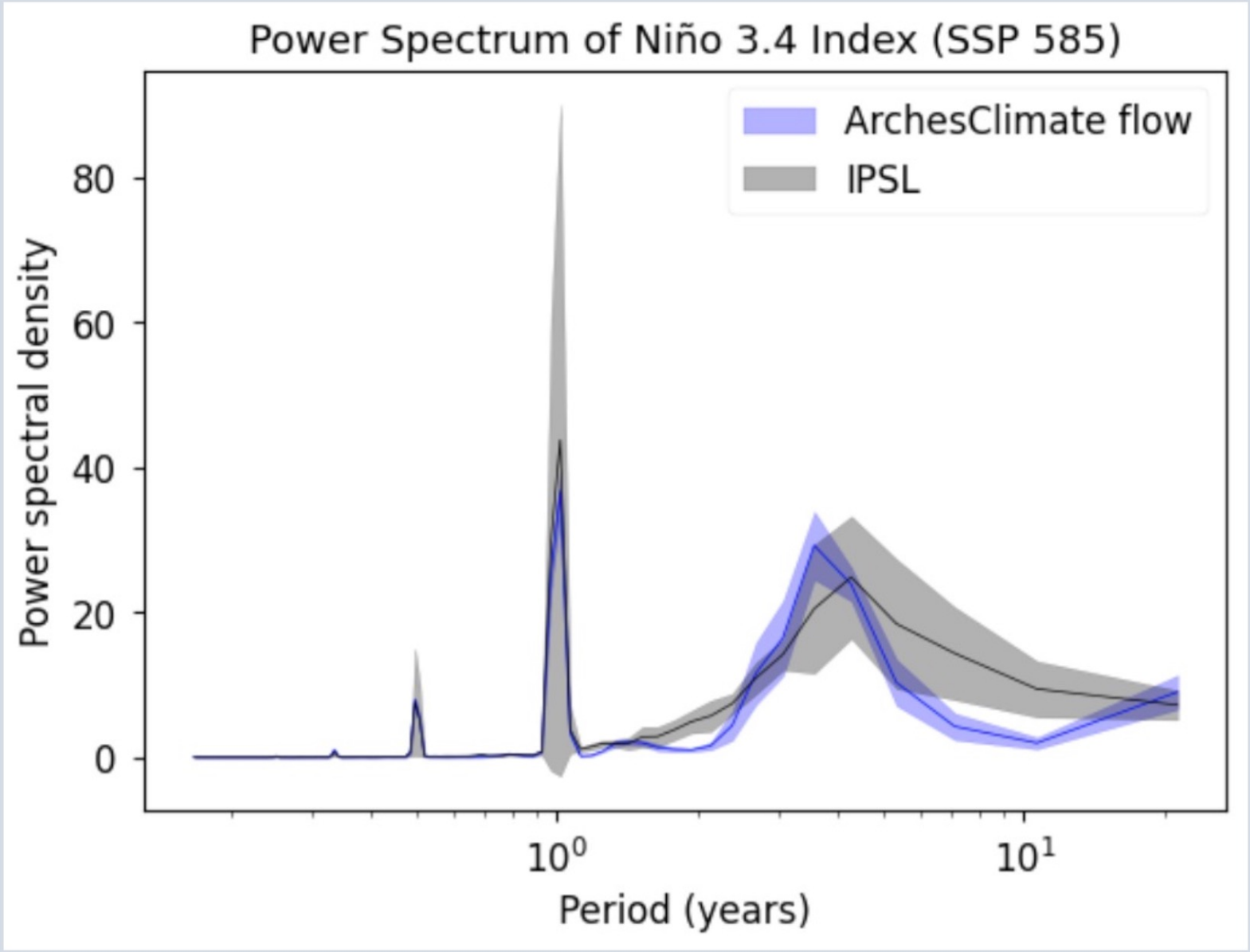


Figure 9. Power spectra for El Niño 3.4 index for SSP5-8.5.

Clyne et al., *in preparazione*

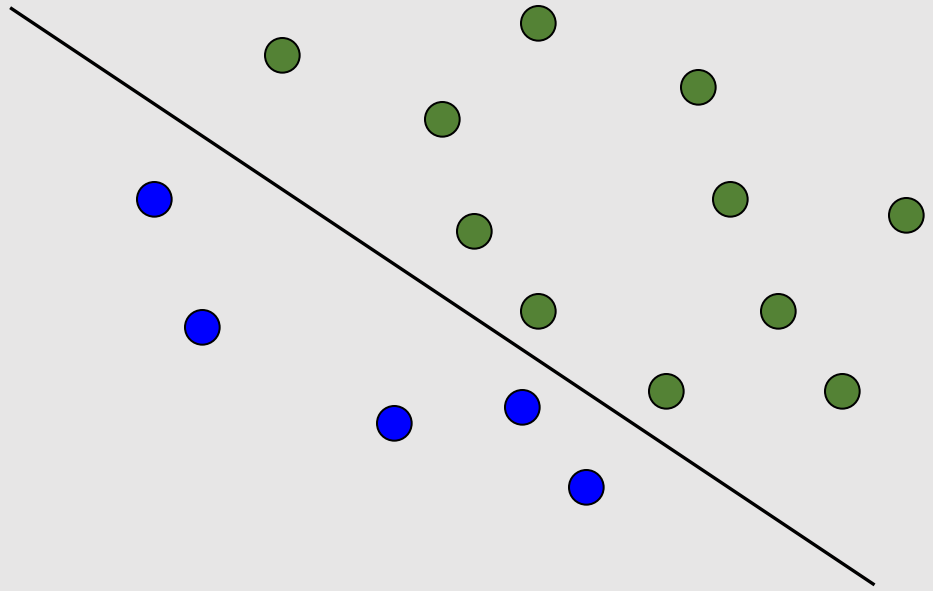
Parte II: Metodi di IA: apprendimento auto-supervisionato

Apprendimento supervisionato

Dati dei punti etichettati, trovare una buona regola di classificazione.

Descrive i dati

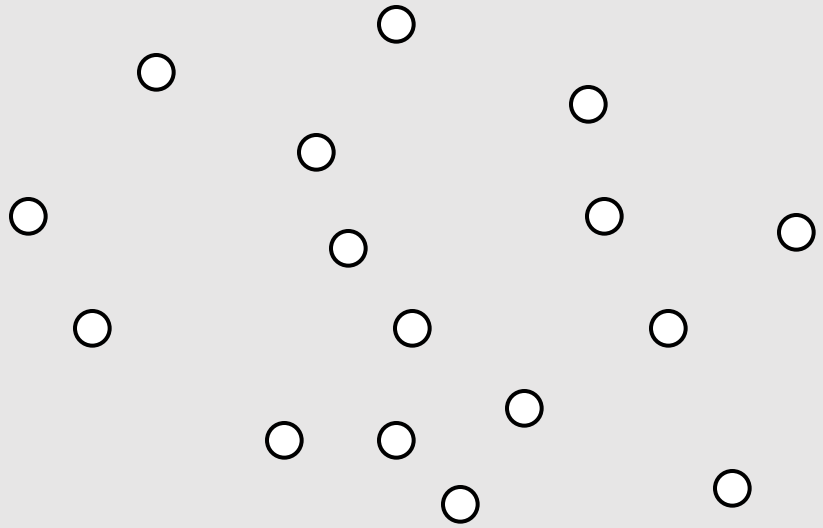
Generalizza bene



Apprendimento non supervisionato

Che cosa si può fare senza etichette?

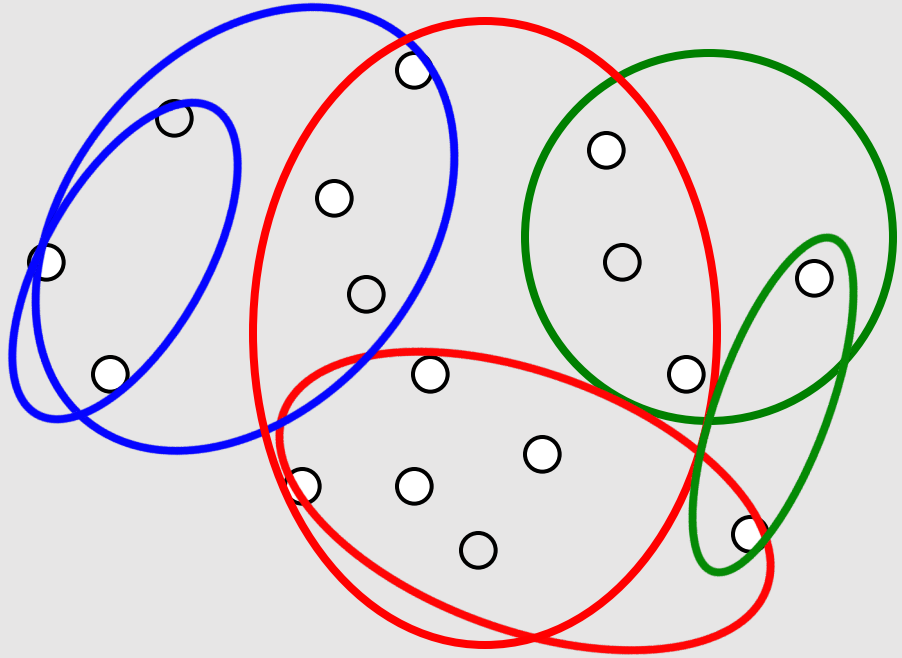
Apprendimento non supervisionato



Apprendimento non supervisionato

Che cosa si può fare senza etichette?

Apprendimento non supervisionato



Esempio: clustering

Apprendere da dati «non etichettati»

- Compiti di apprendimento non supervisionato
  1. Clustering
  2. Embedding (riduzione della dimensionalità, anche con metodi non lineari)
  3. Interpolazione
  4. Apprendimento di un modello generativo

- Esistono approcci di deep learning (DL) per tutti questi compiti

- Esistono approcci non DL per tutti questi compiti

- Esistono approcci DL auto-supervisionati per tutti questi compiti

Deep learning non supervisionato

- DL supervisionato. La loss di previsione è funzione dell'etichetta y e dell'output della rete sull'input x.

Output della rete

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

Funzione di loss

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

- DL non supervisionato. La loss di previsione dipende solo da x e dall'output della rete sull'input x. Non c'è etichetta y.

Output della rete

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

Funzione di loss

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

Approccio auto-supervisionato all'apprendimento non supervisionato

Apprendimento auto-supervisionato

Un approccio all'avanguardia all'apprendimento (profondo) non supervisionato

Progettare un compito pretesto:

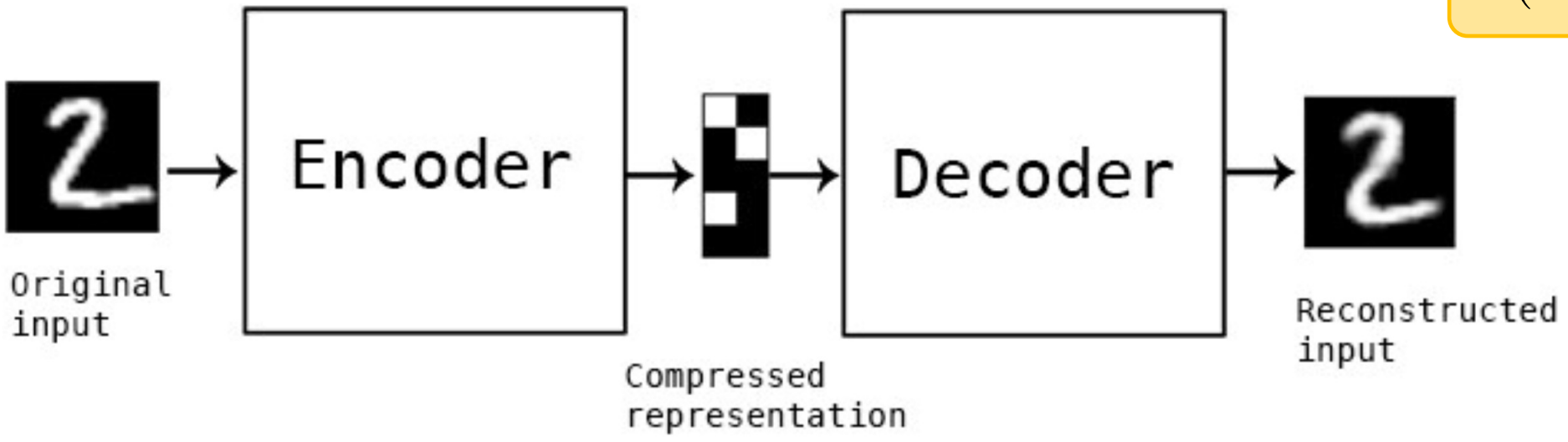
- Progettare un compito di apprendimento supervisionato usando solo i dati disponibili.
- Addestrare un modello su questo compito in modo che
- le caratteristiche apprese (o la posteriore appresa su uno spazio di caratteristiche) siano utili per un altro compito (a valle)

Compito pretesto: esempio

Esempio classico di compito pretesto: l'autoencoder

- Addestrare una rete neurale in modo non supervisionato
  - Usare i dati non etichettati sia come input sia per valutare l'output
- Dopo l'addestramento, lo strato collo di bottiglia sarà una rappresentazione compatta della distribuzione di input

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



Apprendimento auto-supervisionato

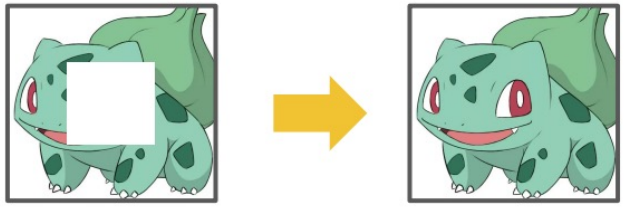
- Progettare un compito pretesto che porti ad apprendere caratteristiche utili.

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



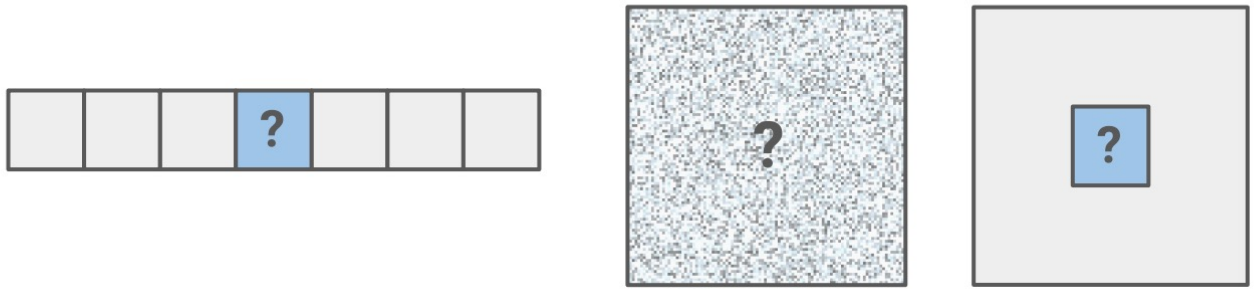
Lilian Weng, Jong Wook Kim, tutorial NeurIPS 2021

Apprendimento auto-supervisionato

- Progettare un compito pretesto che porti ad apprendere caratteristiche utili.

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)

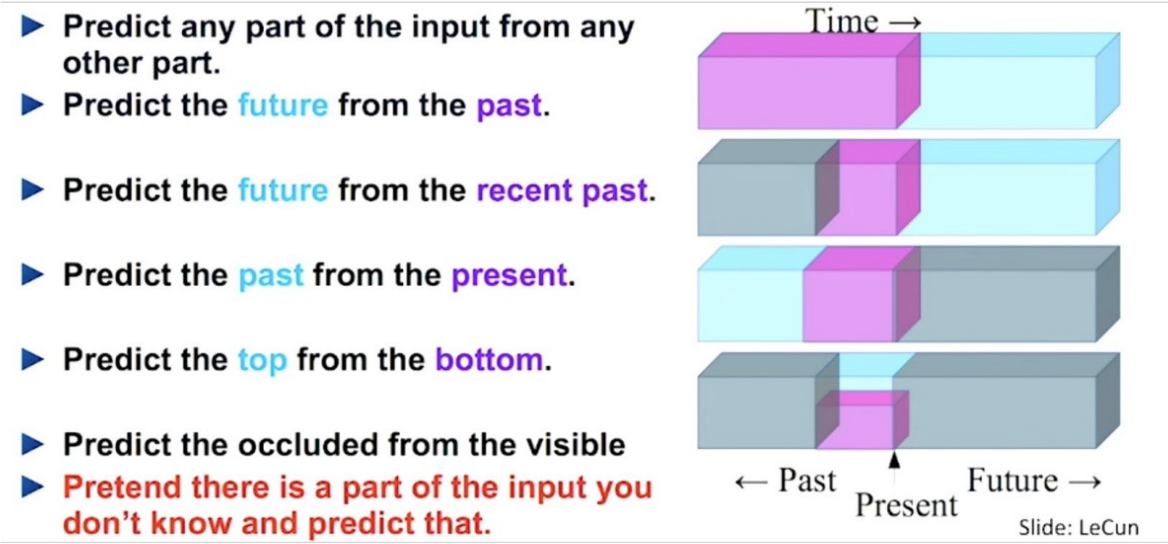
Lilian Weng, Jong Wook Kim, tutorial NeurIPS 2021

Apprendimento auto-supervisionato

- Progettare un compito pretesto che porti ad apprendere caratteristiche utili.

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

Apprendimento auto-supervisionato: apprendimento contrastivo

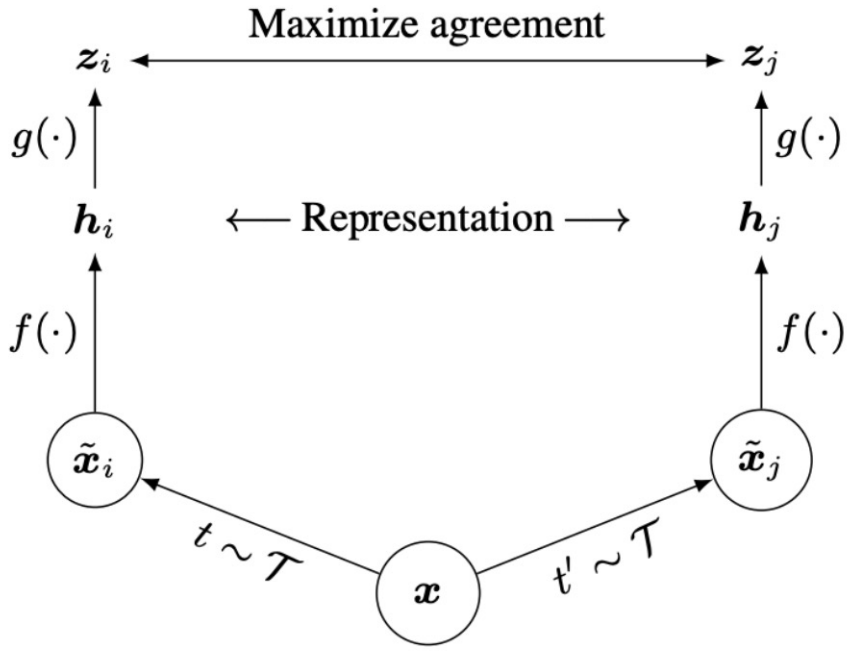


Figure 2. A simple framework for contrastive learning of visual representations. Two separate data augmentation operators are sampled from the same family of augmentations ($t \sim \mathcal{T}$ and $t' \sim \mathcal{T}$) and applied to each data example to obtain two correlated views. A base encoder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

Compito pretesto:

1. Aumentare l'input x due volte: $\tilde{x}_i, \tilde{x}_j$
2. Apprendere una mappatura $g(f(\cdot))$, minimizzando una funzione di loss per avvicinare $g(f(\tilde{x}_i)), g(f(\tilde{x}_j))$.

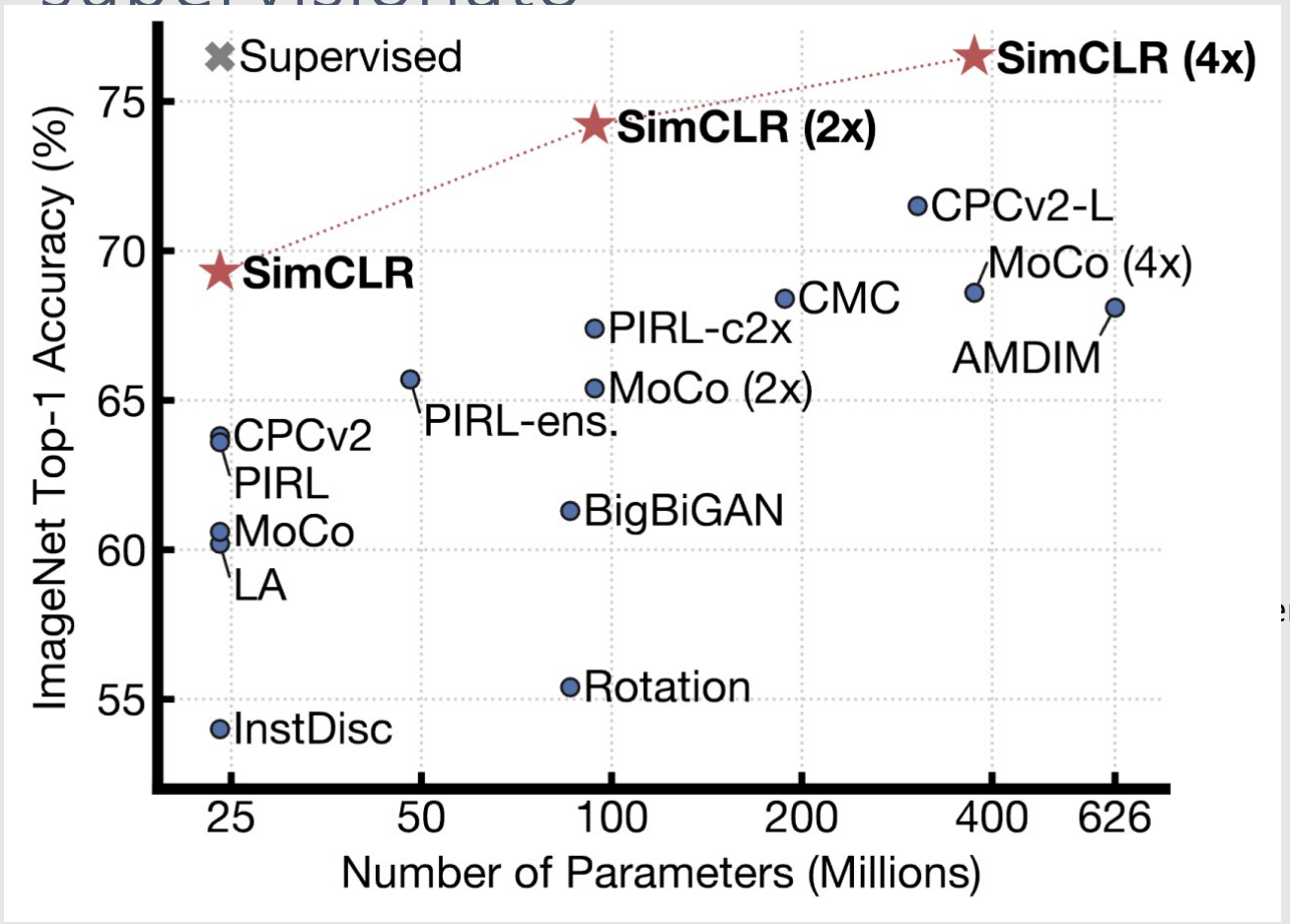
$$\mathcal{L}(\hat{x}, x)$$

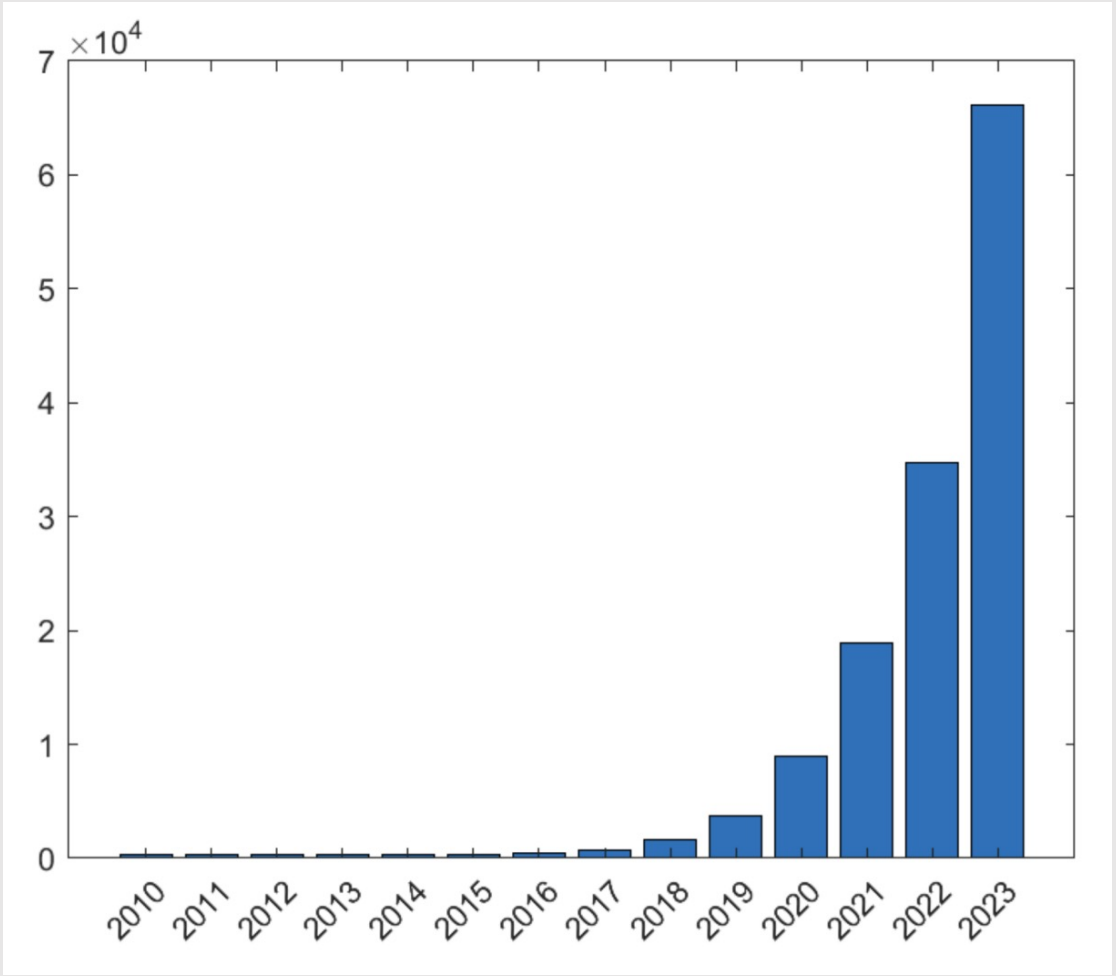
$$\hat{x} := \{g(f(\tilde{x}_i)), g(f(\tilde{x}_j))\}$$

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

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

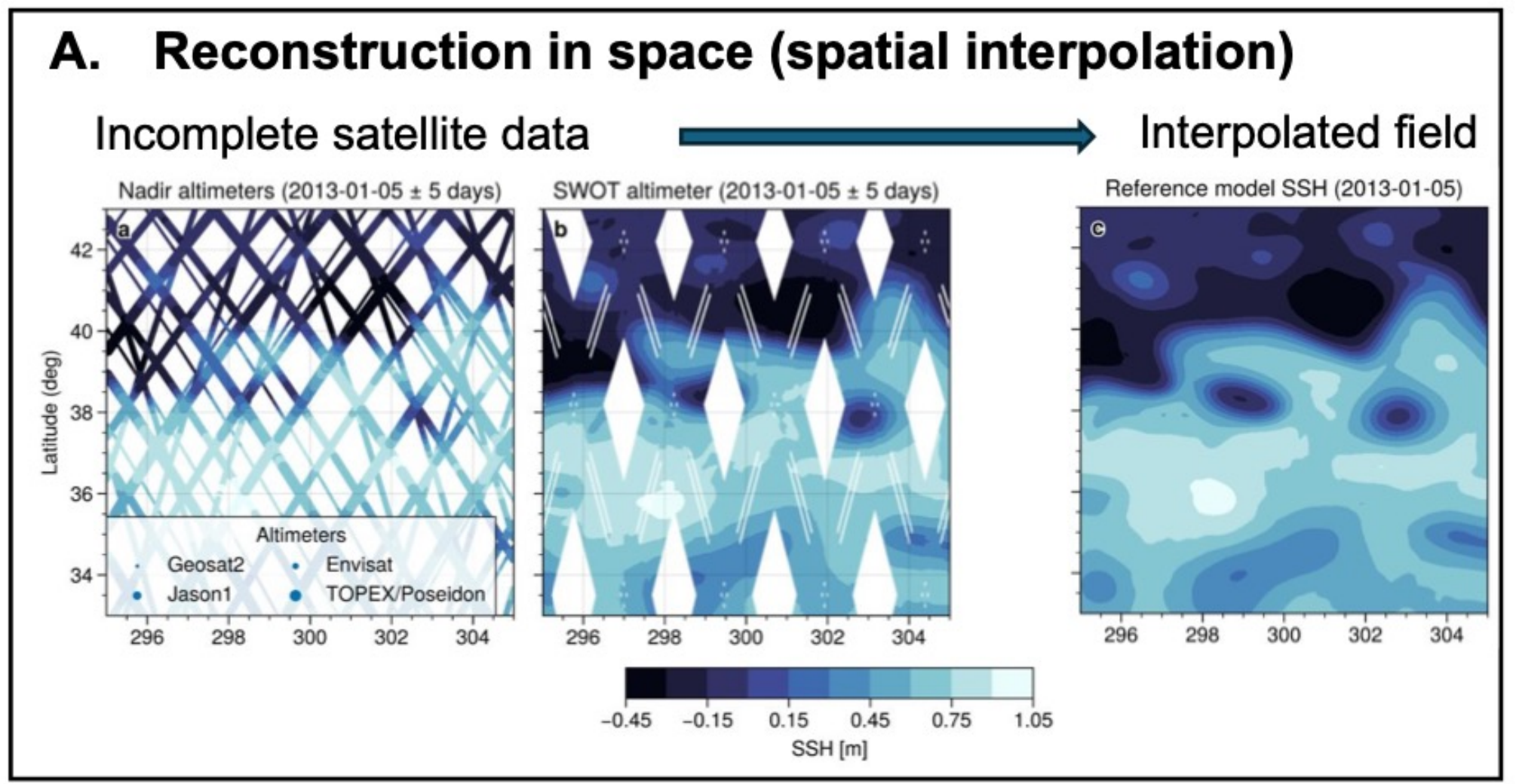
L’apprendimento auto-supervisionato può superare quello supervisionato





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

L’auto-supervisionato per le vostre applicazioni



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

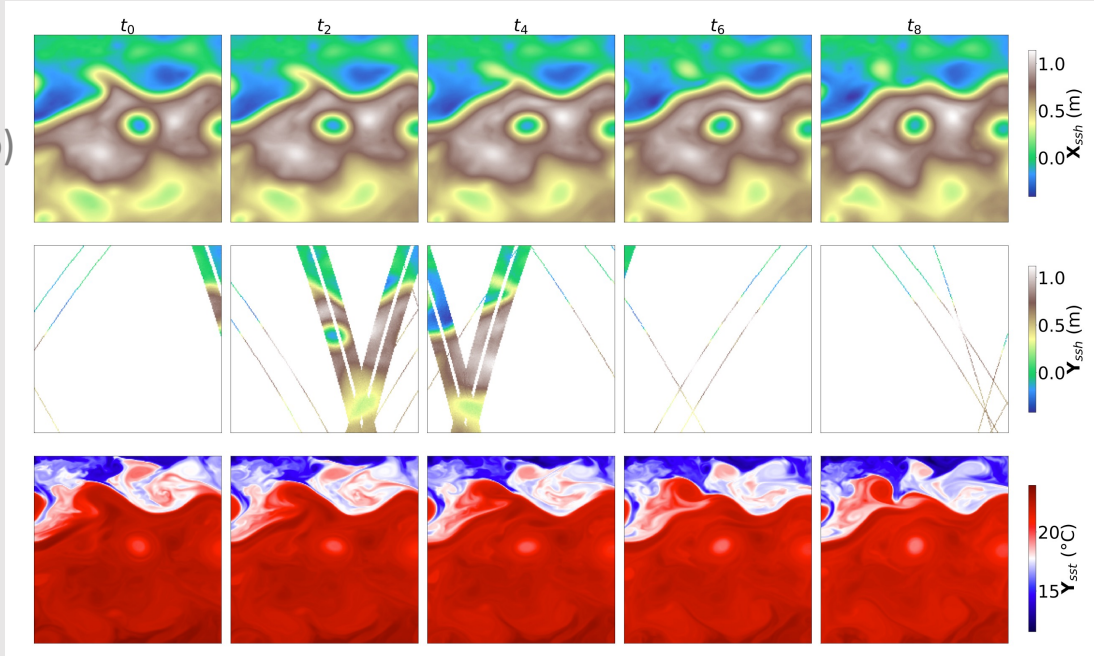
Interpolazione spazio-temporale non supervisionata

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

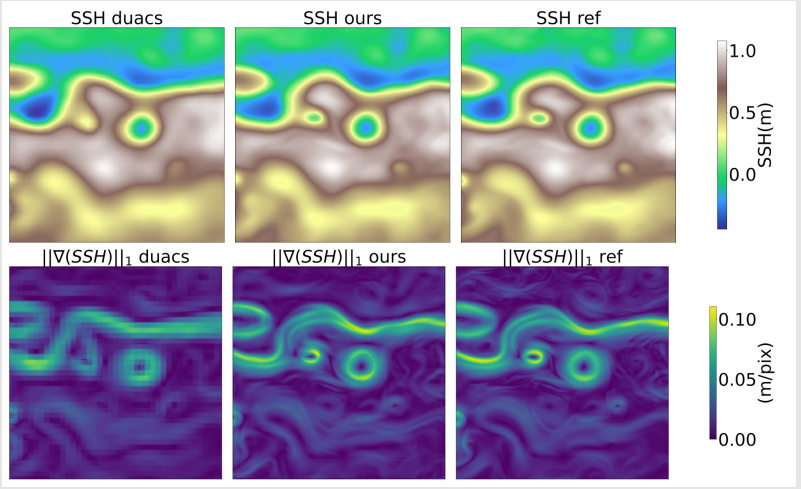
Campi SSH di riferimento
(non usati in addestramento)

Dati SSH lungo
le tracce dei satelliti

Campi SST da un
modello fisico



Dati di addestramento



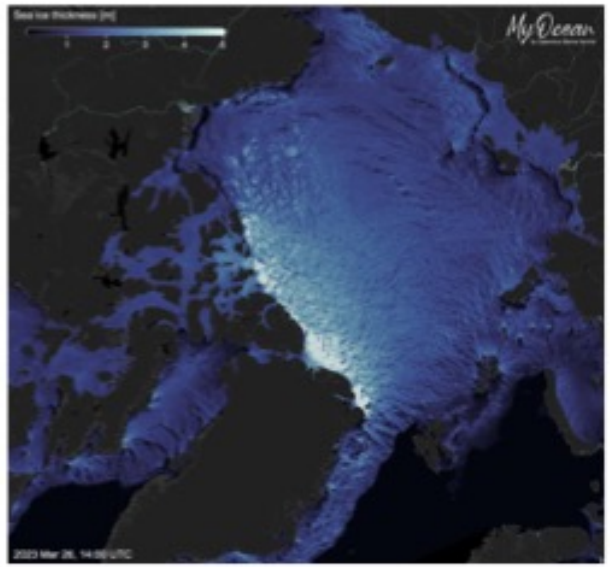
Risultati

L’auto-supervisionato per le vostre applicazioni

C. Reconstruction in scales (downscaling)

Low resolution field → High resolution field





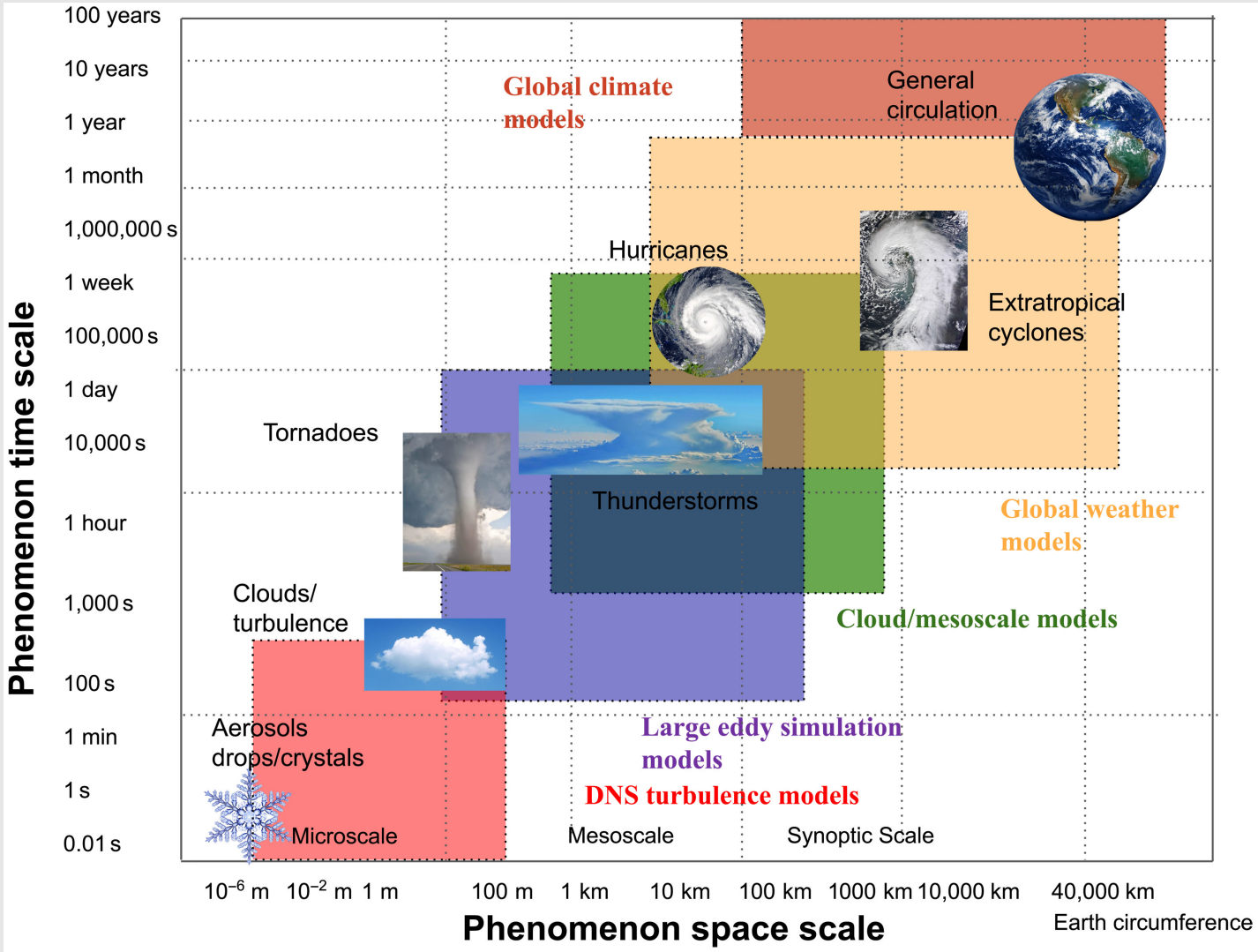
[Bracco et al., Machine Learning for the Physics of Climate. *Nature Reviews Physics*, 2024]

Downscaling delle simulazioni dei modelli climatici

Le simulazioni dei modelli climatici globali sono più grossolane (nello spazio e nel tempo) di quanto serva per molti compiti di:

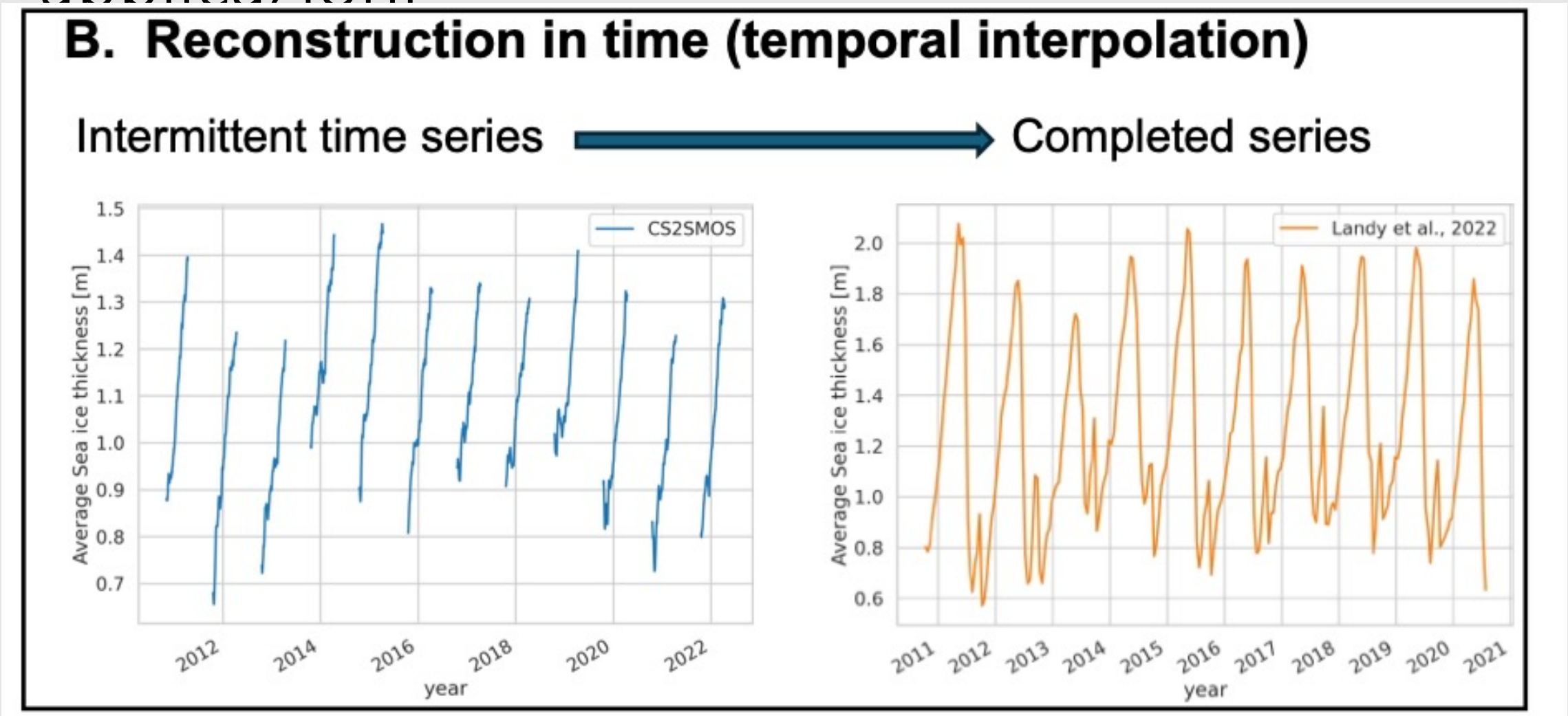
- Adattamento ai cambiamenti climatici
- Mitigazione dei cambiamenti climatici
- Proiezione degli impatti a lungo termine

Approccio: usare l'IA generativa per il downscaling dei dati dei modelli climatici alle scale rilevanti



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

L’auto-supervisionato per le vostre applicazioni



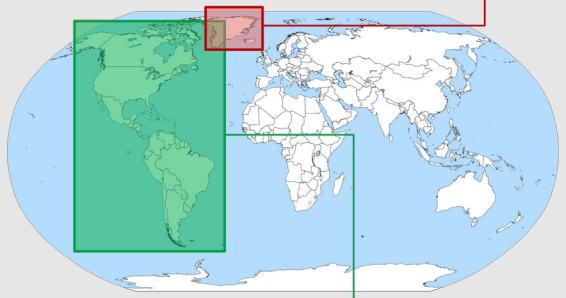
[Bracco et al., Machine Learning for the Physics of Climate. *Nature Reviews Physics*, 2024]

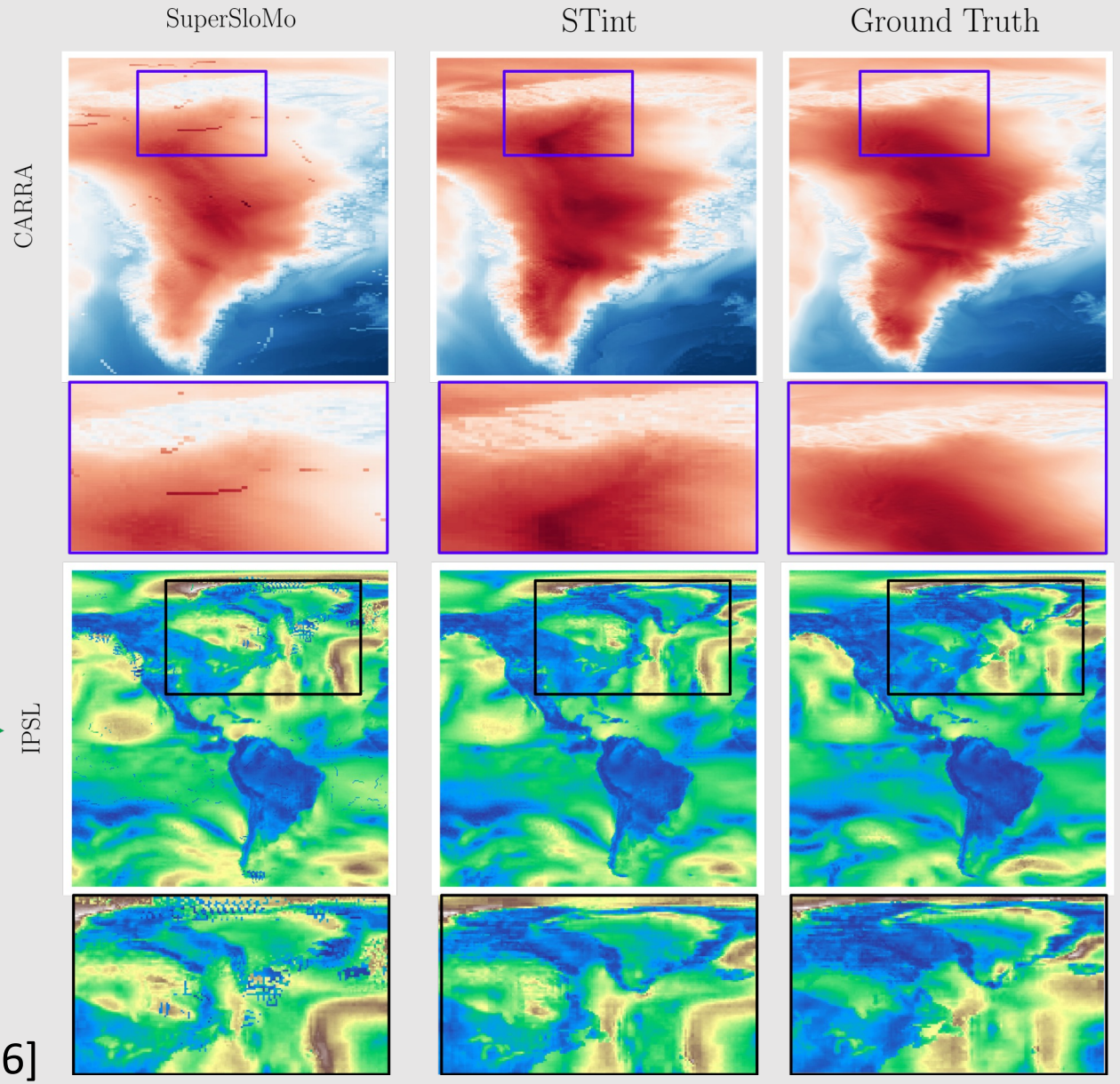
Esempio: downscaling temporale (interpolazione) di dati geospaziali

Perché non usare semplicemente i metodi di IA esistenti?

- Le applicazioni sui cambiamenti climatici riguardano dati geospaziali che evolvono nel tempo
  - Dati osservativi su griglia globale tramite assimilazione dei dati
  - Simulazioni da modelli fisici (NWP, GCM, RCM)
- Sono tensori di valori reali su latitudine, longitudine, tempo,
  ed eventualmente su più variabili climatologiche
- Gli algoritmi di computer vision per i «dati spazio-temporali» si basano su proprietà dei **dati video** che **non si generalizzano bene ai dati geospaziali**
  - ad es. profondità, contorni e «oggetti»
  - vs schemi effimeri nei fluidi

STINT: interpolazione temporale auto-supervisionata







Lo stato dell'arte della computer vision per l'interpolazione temporale dei video usa il flusso ottico.
Questo è problematico sui dati geospaziali.

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

Limiti dei lavori precedenti: interpolazione bicubica

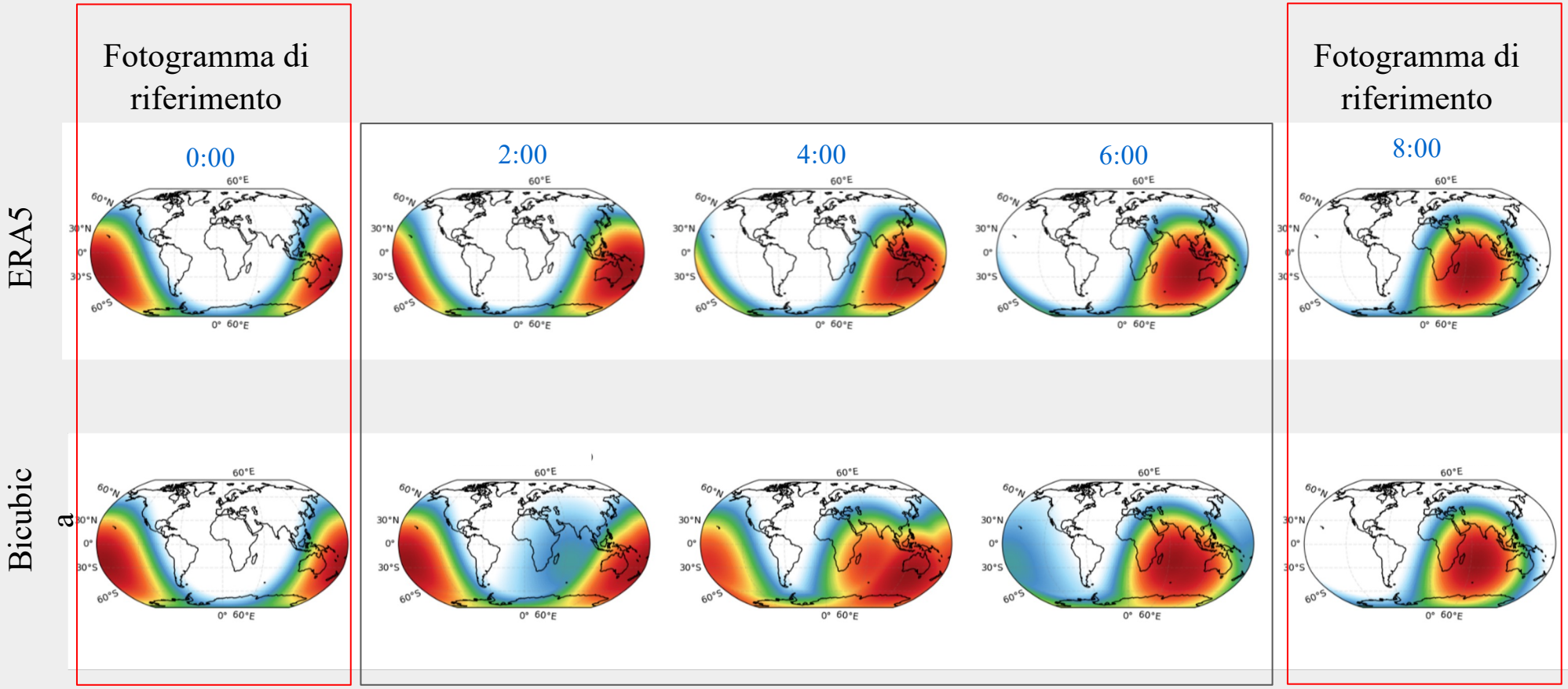
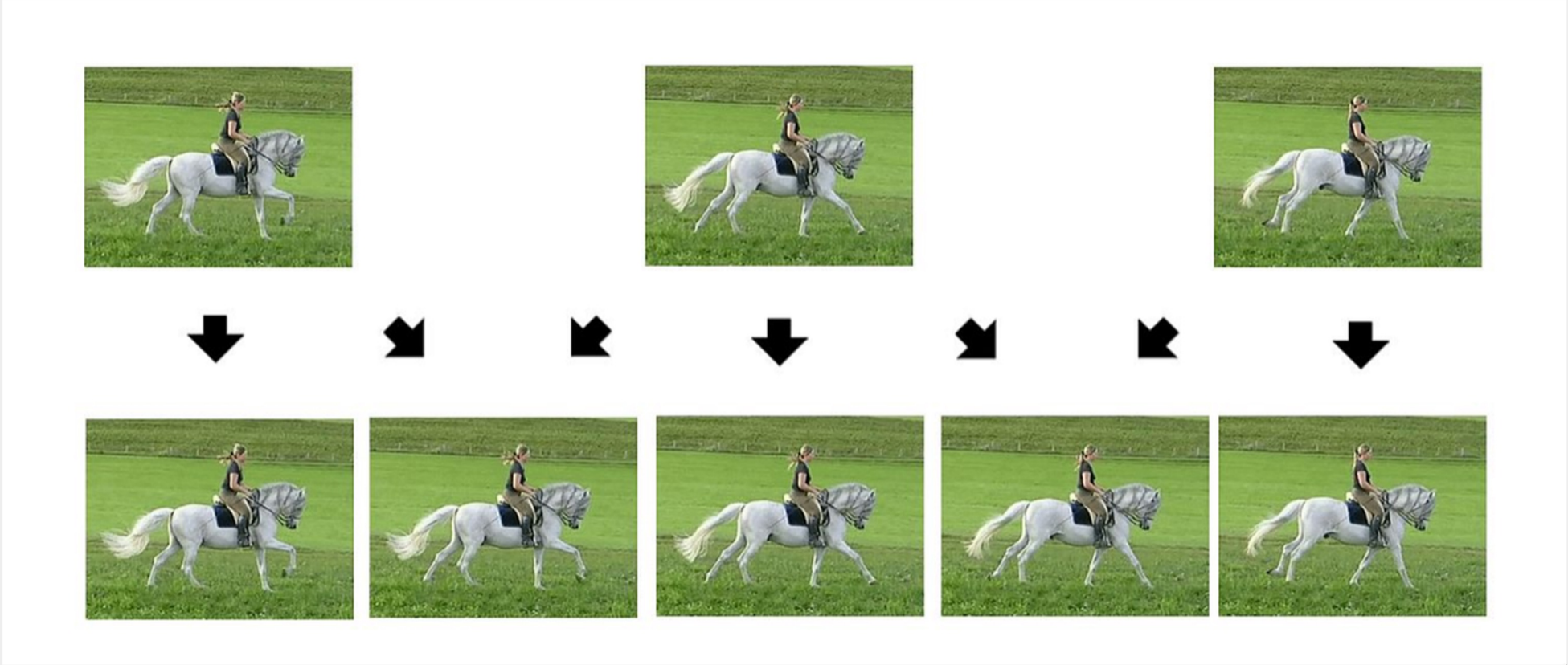


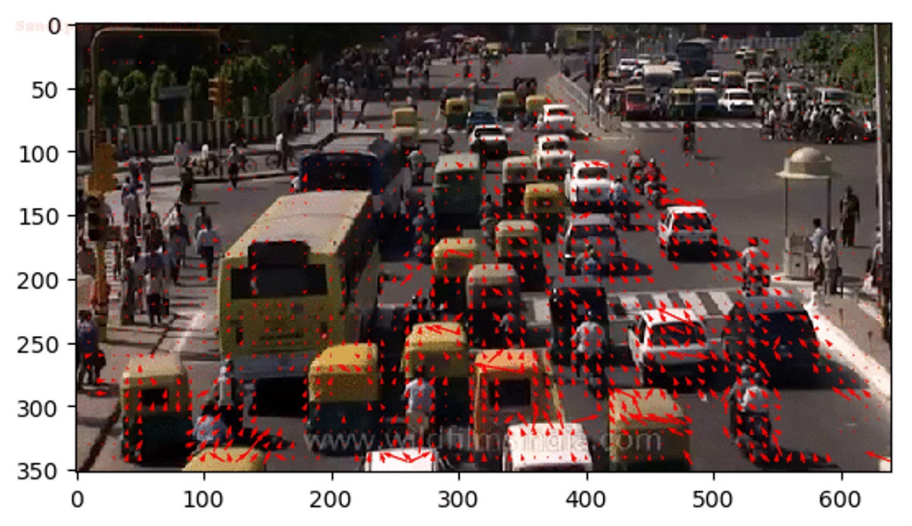
Figura. *Interpolazione bicubica temporale di campi di irradianza solare ERA5*

Interpolazione video nella computer vision

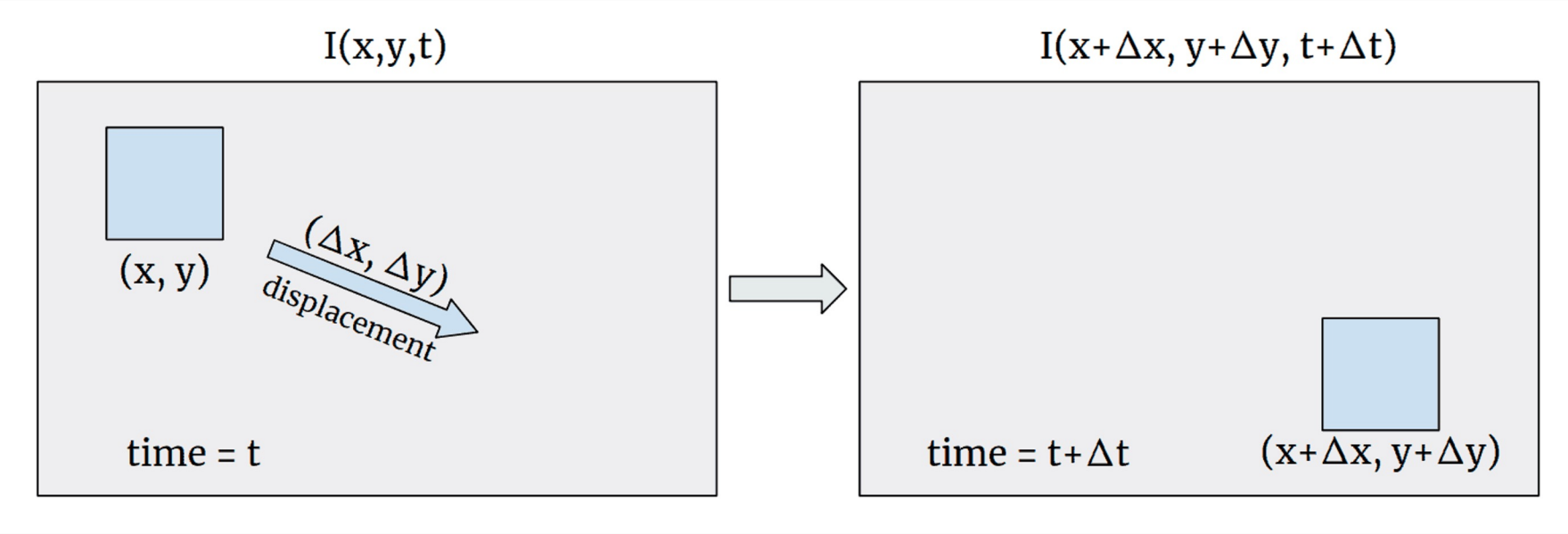


Il flusso ottico

- Flusso ottico: campo di moto pixel per pixel tra due immagini (ad es. fotogrammi video).
- Codifica la proiezione 2D «percepita» del moto fisico dei punti rispetto all'osservatore.



(fonte: Nvidia OF SDK)



Il flusso ottico

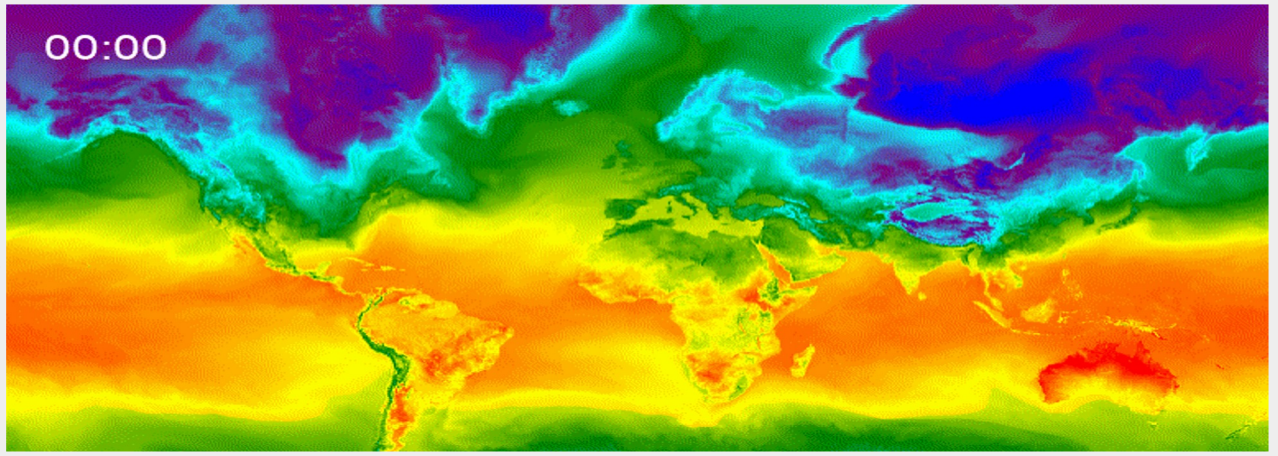
Ipotesi:

- Piccolo spostamento (Δt).
- Nessuna/poca variazione della luce riflessa.
- Nessuna/poca variazione di scala.
- Nessuna occlusione o disocclusione.



Non applicabile ai dati geospaziali!

- Nessun concetto di oggetti o profondità.
- Le entità possono formarsi/deformarsi arbitrariamente.
- Risoluzione temporale molto più grossolana



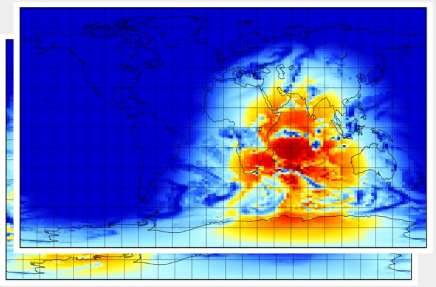
Temperatura oraria NOAA GFS

Obiettivo: migliorare la risoluzione temporale

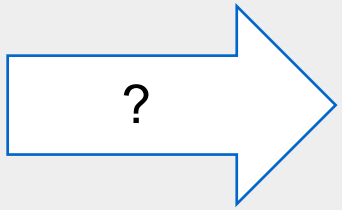
- Il compito è l'interpolazione temporale
- Ma <u>non abbiamo dati orari</u> per l'addestramento

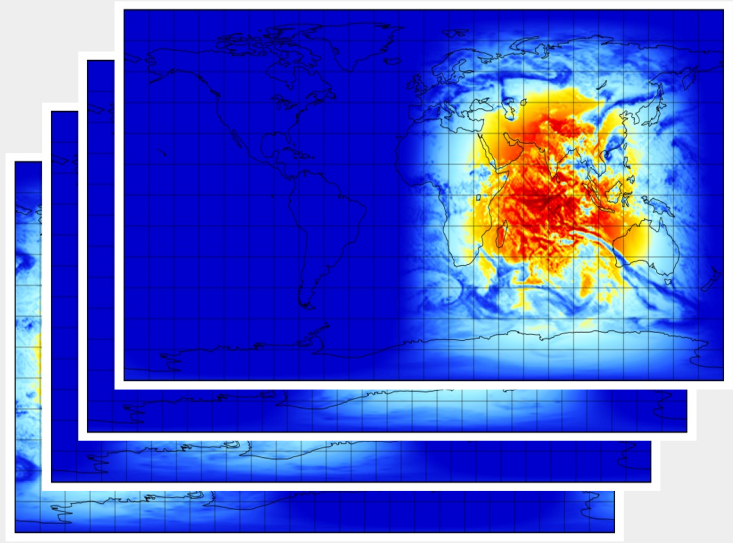
servono tecniche non supervisionate (**auto-supervisionate**).

CMIP6-IPSL



Triorario





Oraria

L’auto-supervisionato in breve

- Usare compiti pretesto/surrogati in modo supervisionato.
  - Etichette **generate** automaticamente.

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

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

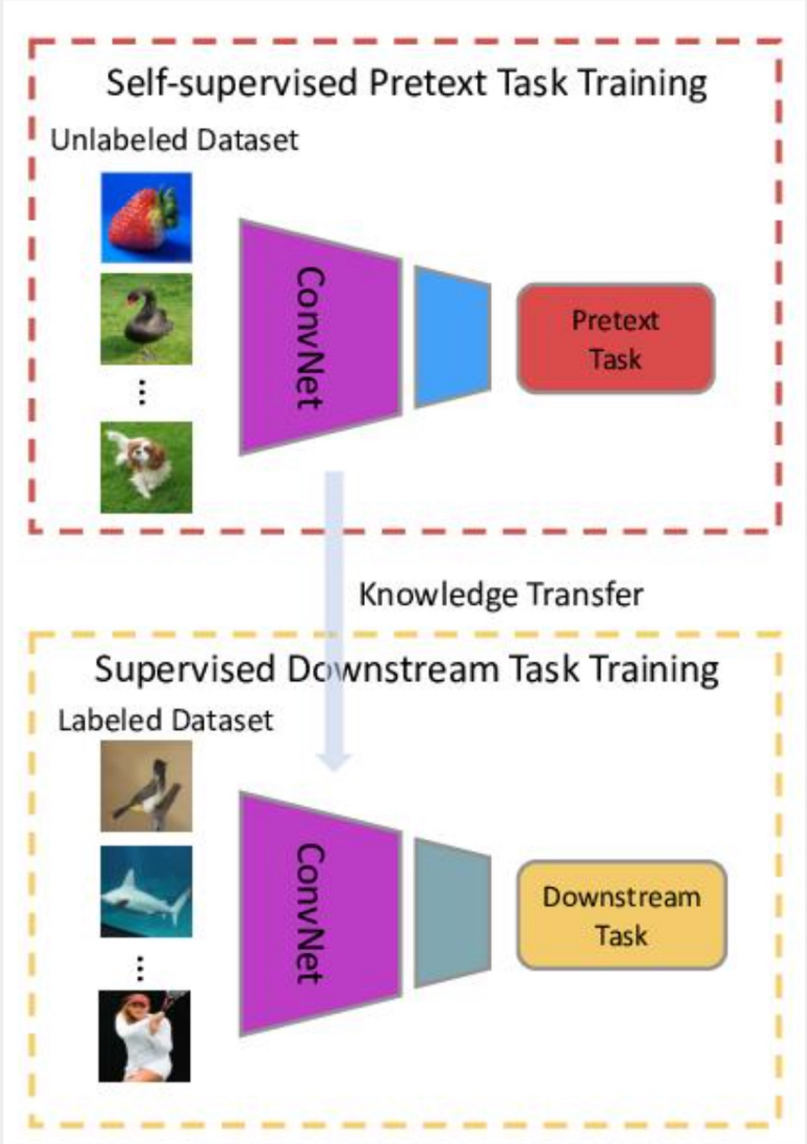
Apprendimento supervisionato

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

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

Apprendimento auto-supervisionato

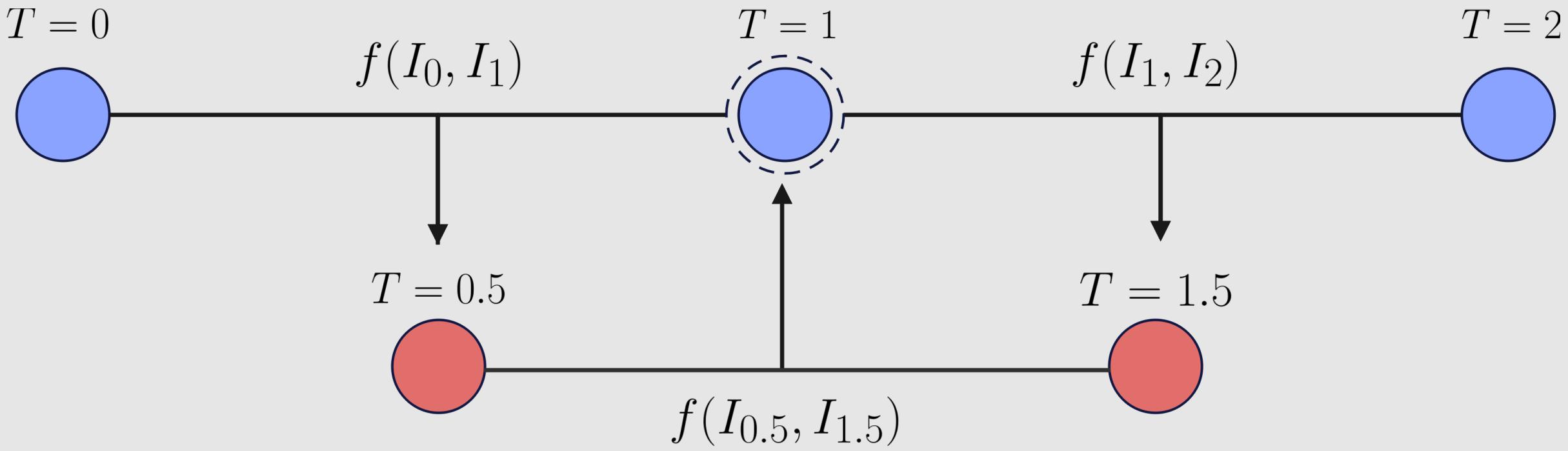
- Compito a valle: retrieval, classificazione supervisionata o semi-supervisionata, ecc.



Un compito pretesto per l'interpolazione temporale

STINT: interpolazione temporale auto-supervisionata per dati geospaziali

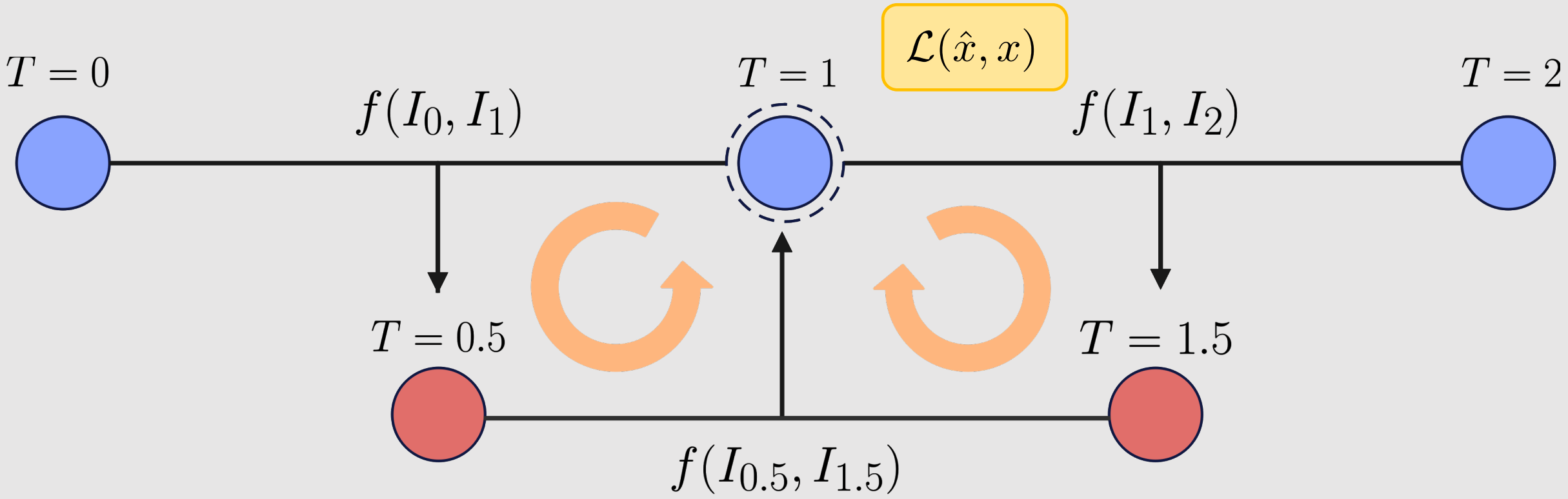
[Harilal, Hodge, Subramanian, & Monteleoni, Climate Informatics 2026]



Un compito pretesto per l'interpolazione temporale

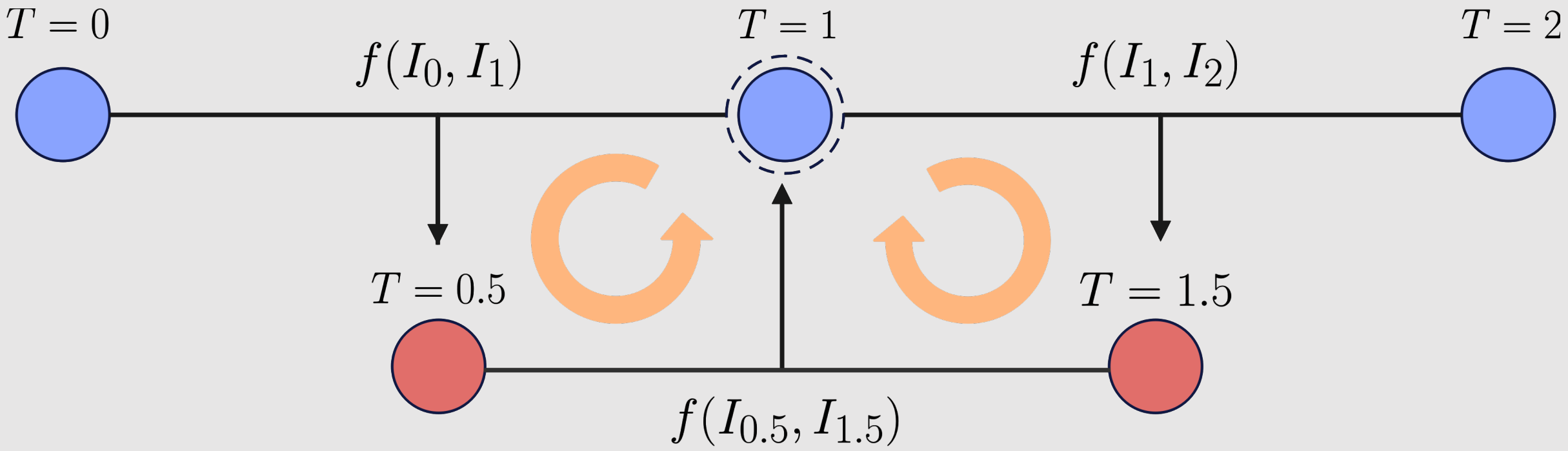
STINT: interpolazione temporale auto-supervisionata per dati geospaziali

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



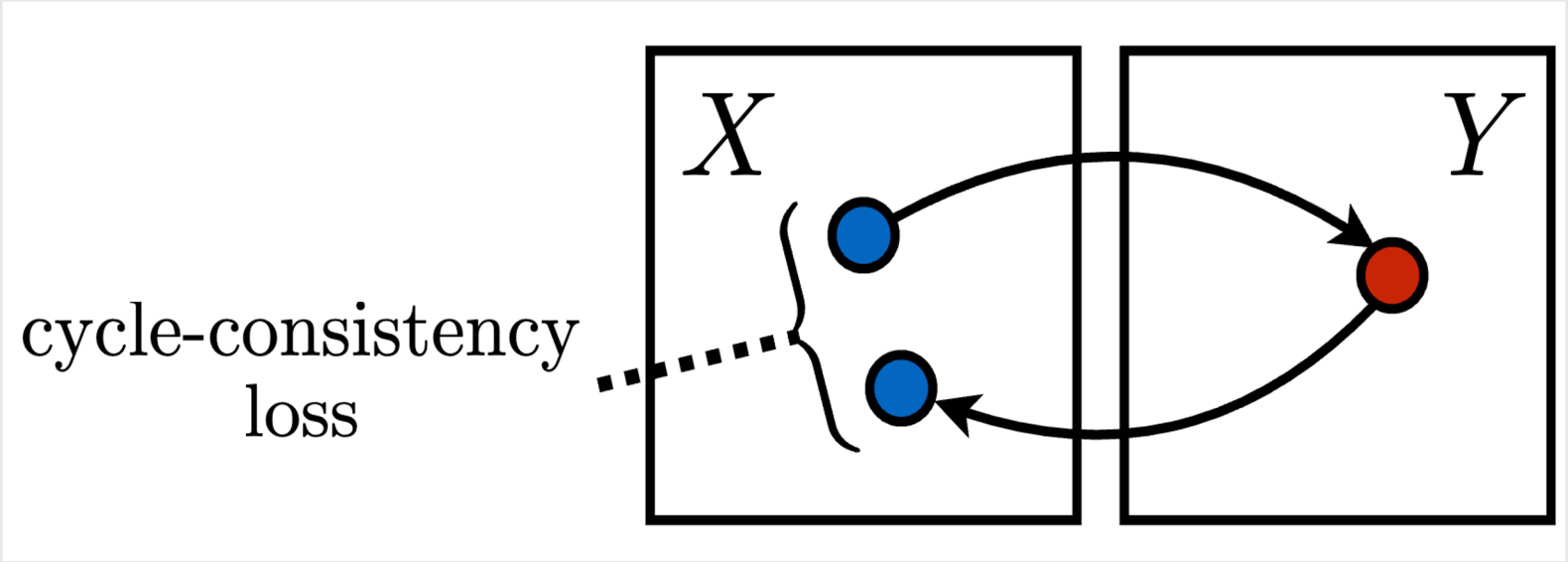
Un compito pretesto per l'interpolazione temporale

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

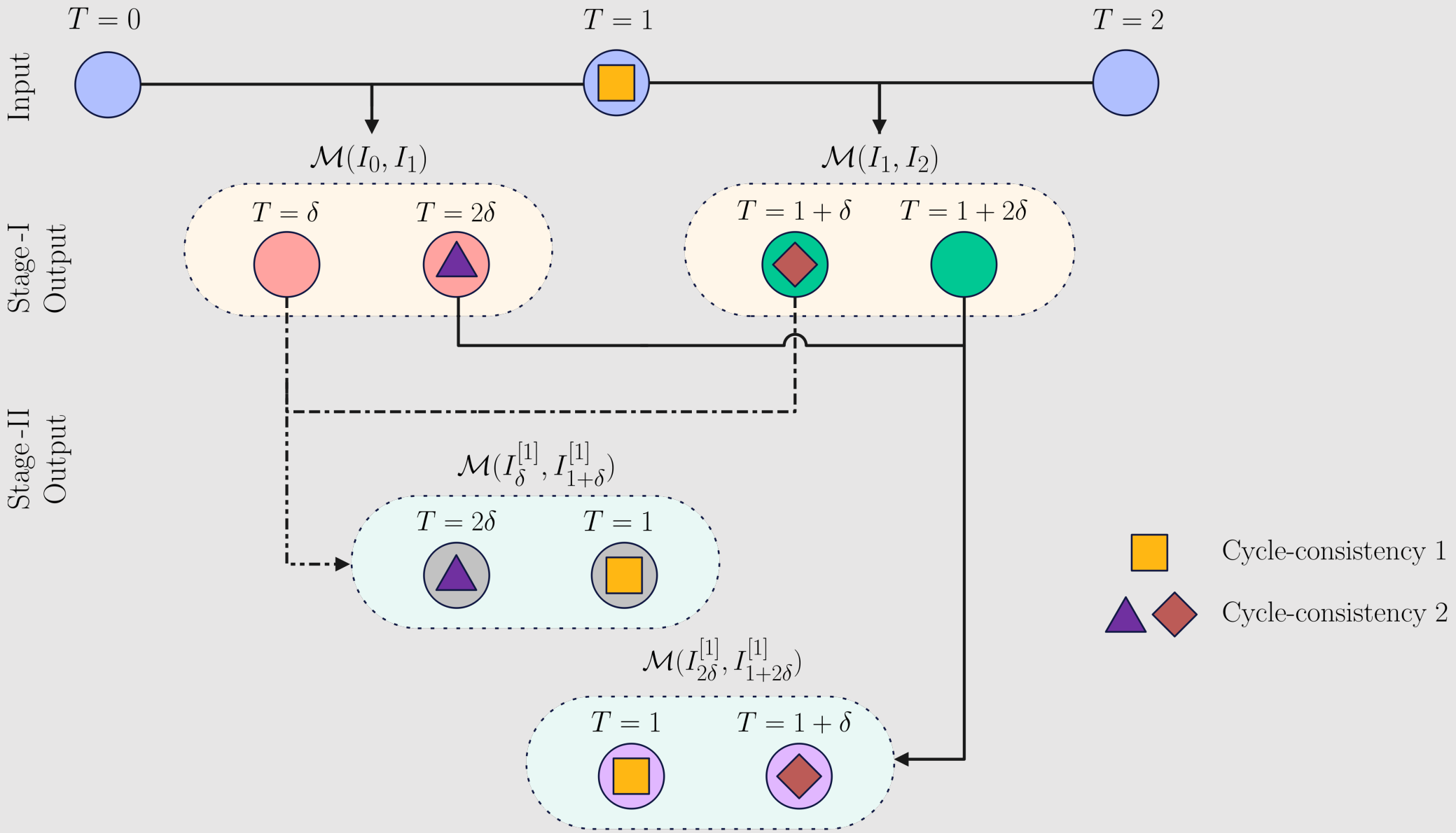


Loss di coerenza ciclica

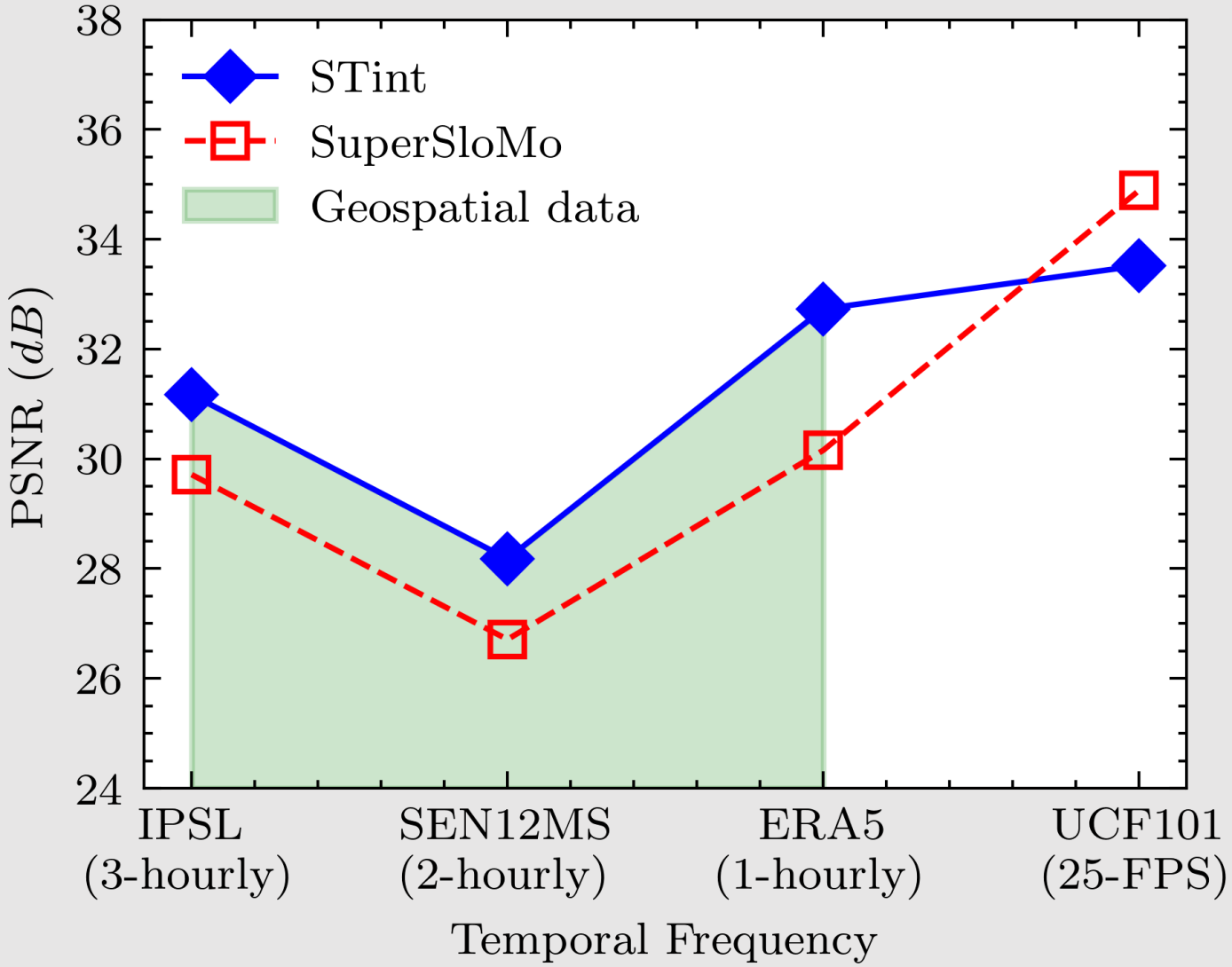
$$\mathcal{L}(\hat{x}, x) = \sum_{i=1}^{n} \mathcal{L}(T_i, \hat{T}_i) \qquad \hat{T}_i := f(f(T_{i-1}, T_i), f(T_i, T_{i+1}))$$



STINT: interpolazione temporale auto-supervisionata



STINT: interpolazione temporale auto-supervisionata



	$\frac{MSE}{Capacity}$ (↓)	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**

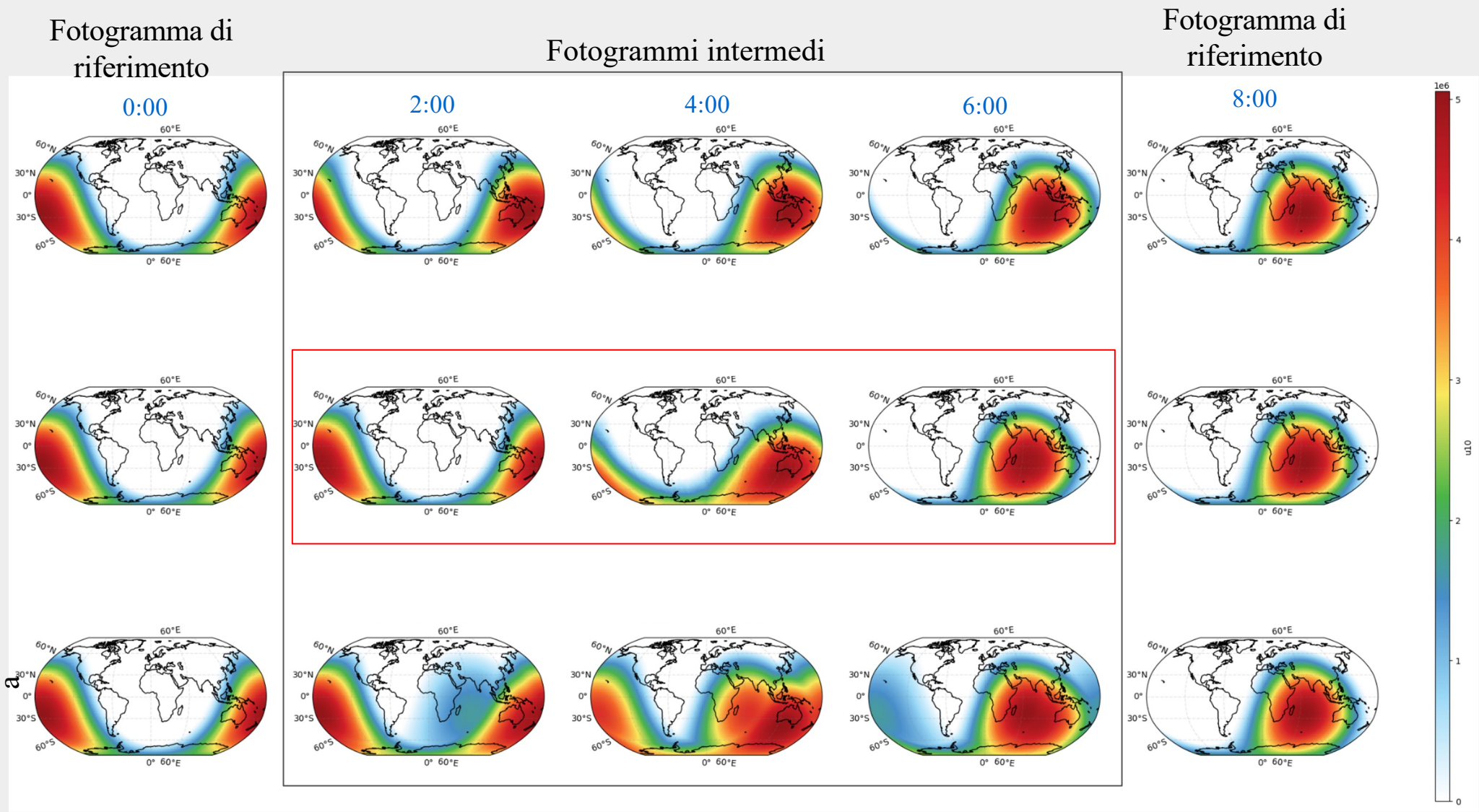


Figura. *Modello addestrato su campi di irradianza solare ERA5 e interpolazione (3 fotogrammi non visti) tra le 0:00 e le 8:00.*

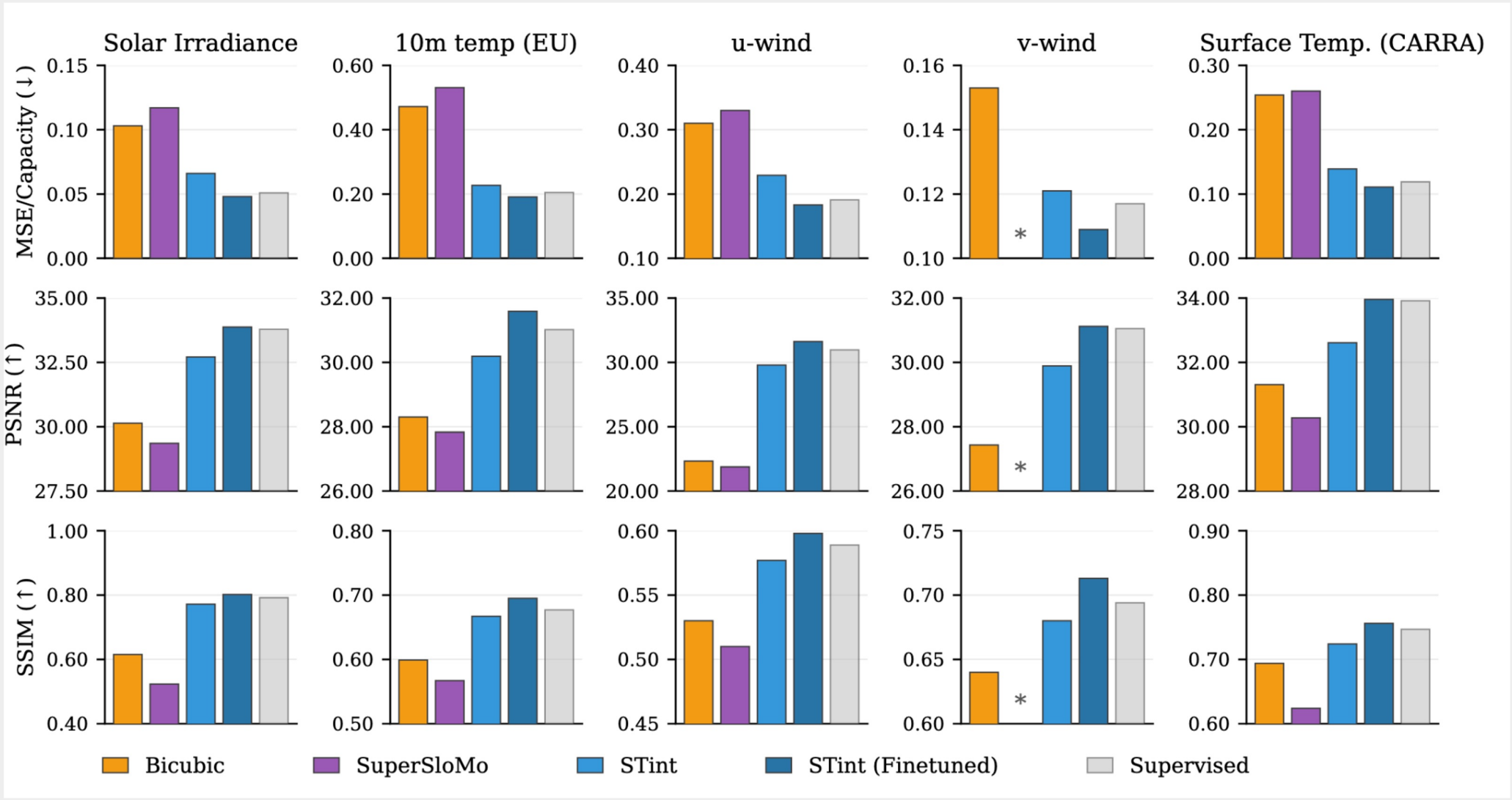
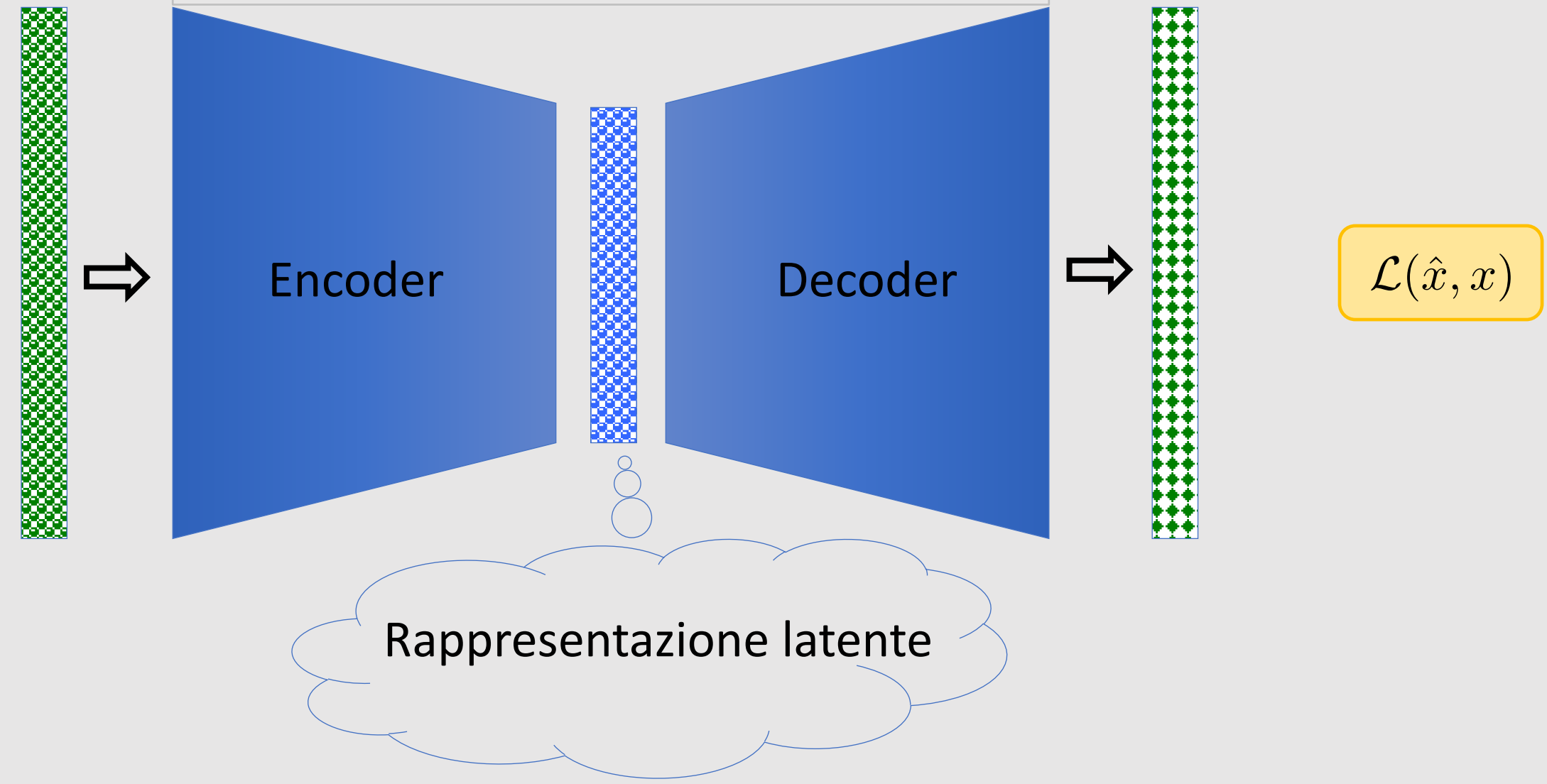


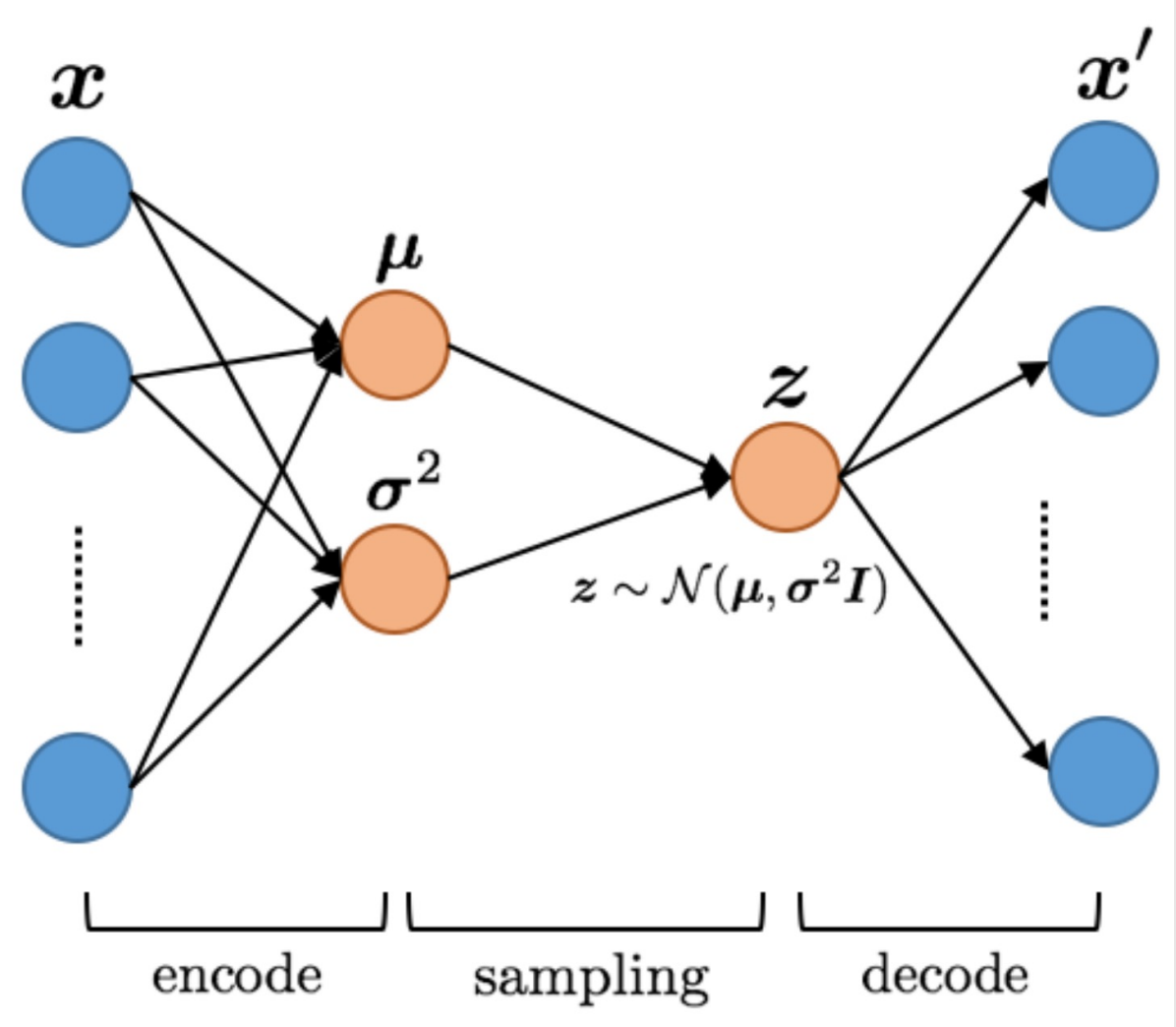
Figura. *Risultati di interpolazione temporale su diversi dataset geospaziali.*

Autoencoder: i parametri delle reti encoder e decoder sono addestrati perché l'output approssimi l'input. Dopo l'addestramento su molti esempi, i parametri dello strato collo di bottiglia formano una rappresentazione compatta della distribuzione di input.



Autoencoder variazionale (VAE)

Apprendere una distribuzione sulle rappresentazioni latenti, invece di una singola codifica



Apprendere un autoencoder variazionale (VAE)

- L'encoder è una distribuzione, q, che mappa lo spazio dei dati nello spazio latente
  - Definito dai pesi della rete θ:

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

- Il decoder è una distribuzione, p, che mappa lo spazio latente nello spazio dei dati
  - Definito dai pesi della rete ϕ:

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

- Adattare i pesi θ e ϕ minimizzando:

Errore di ricostruzione sui dati di addestramento

Termine di regolarizzazione: penalizza la distanza da una priori semplice

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

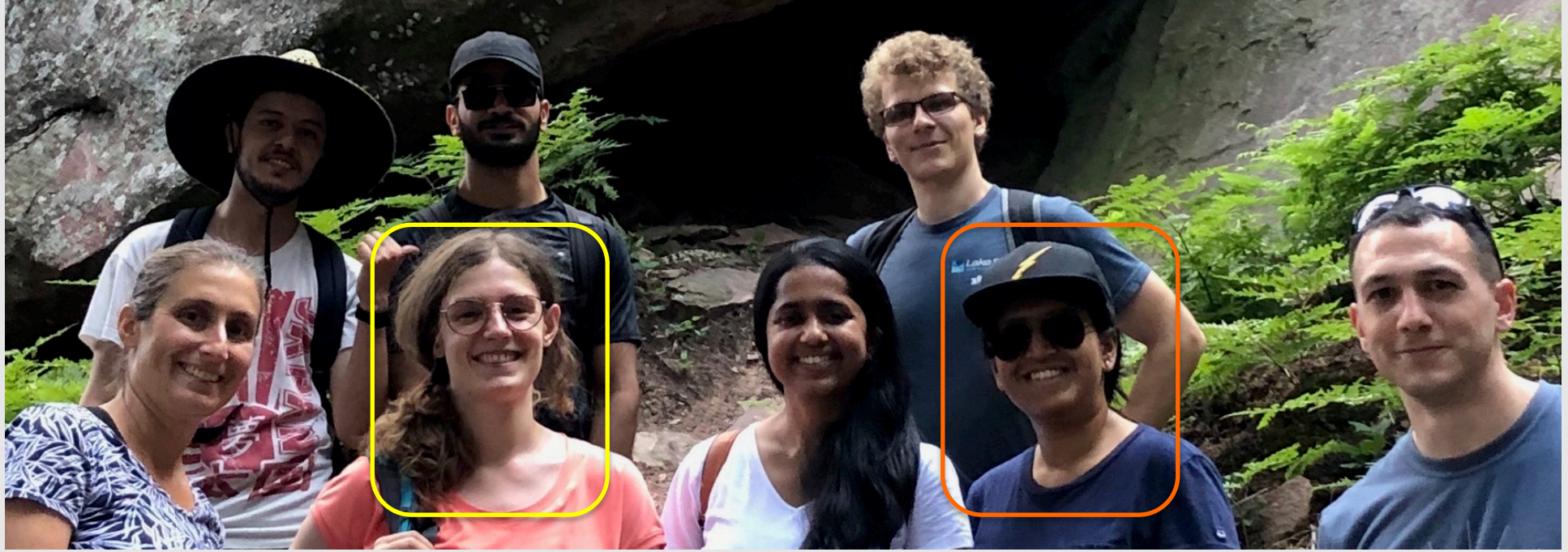
- Ottimizzato con una variante di expectation-maximization (EM), usando un limite inferiore variazionale

Esempio: VAE per il rilevamento semi-supervisionato di anomalie applicato al rilevamento delle valanghe

Rilevamento delle valanghe



Fonte dell'immagine: https://www.nytimes.com/2017/01/19/world/europe/italy-avalanche.html



Rilevamento di depositi di valanga con VAE su immagini Sentinel-1

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

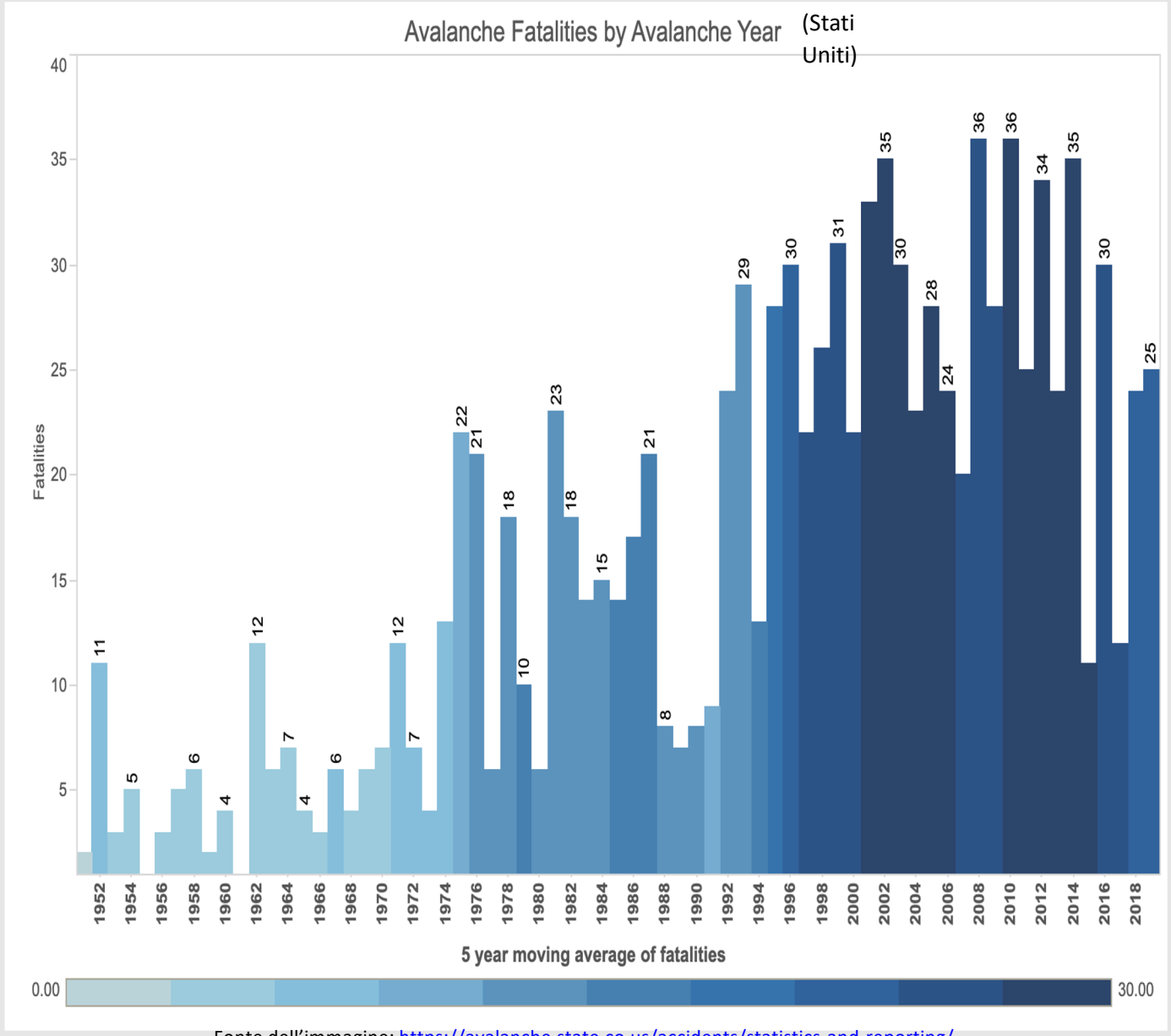
Variational autoencoder anomaly-detection of avalanche deposits in satellite SAR imagery.
Proc. 10th International Conference on Climate Informatics (CI) 2020

Boulder

*Co-primi autori

METEO FRANCE

Vittime di valanghe per anno di valanghe



Fonte dell'immagine: https://avalanche.state.co.us/accidents/statistics-and-reporting/

Impatti dei cambiamenti climatici



Fonte dell'immagine: https://www.independent.co.uk/news/world/europe/mont-blanc-deaths-french-alps-avalanche-skiing-a9227451.html

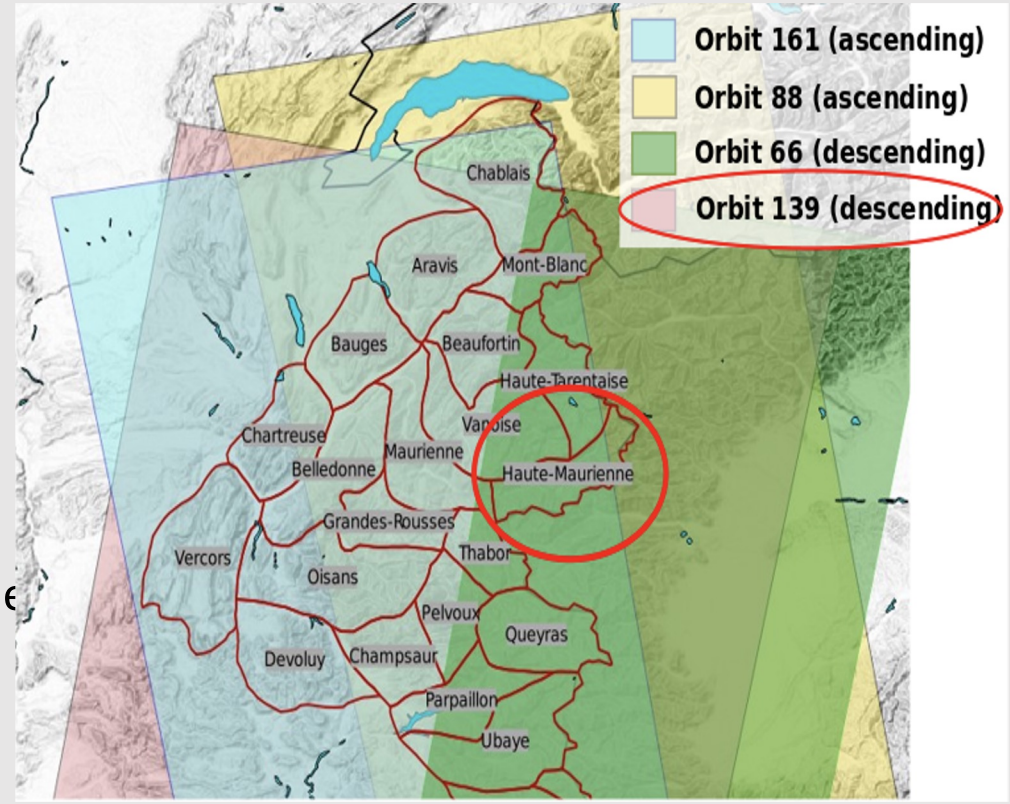
Sfide per il machine learning

- Forte squilibrio tra classi
  - Le valanghe sono eventi rari
- Dati di riferimento etichettati difficili da ottenere
  - Accessibilità del terreno
  - Condizioni meteorologiche
  - Pericolo delle valanghe

Il problema del rilevamento delle valanghe

- **Misure di riferimento** in situ limitate
  - Météo-France

- **Immagini SAR** non etichettate
  - Monitoraggio delle Alpi francesi ne 2017-2018
  - Satelliti Sentinel-1A e 1B
  - 4 caratteristiche:
    - Coefficienti di retrodiffusione all'istante attuale e precedente
    - Caratteristiche topologiche: pendenza ed esposizione



METEO FRANCE

Il nostro approccio

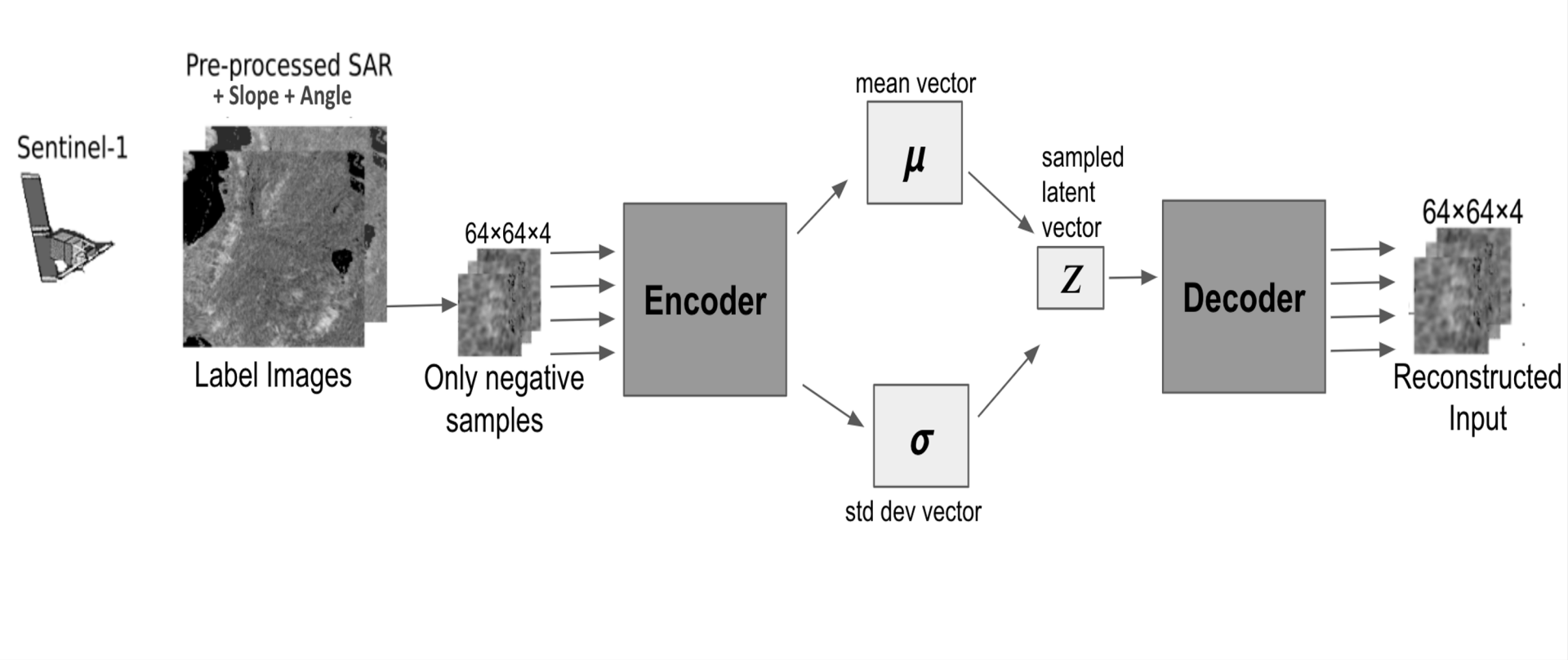
① Trattare una valanga come un evento raro, o un'anomalia

② Addestrare un autoencoder variazionale (VAE) sugli esempi negativi

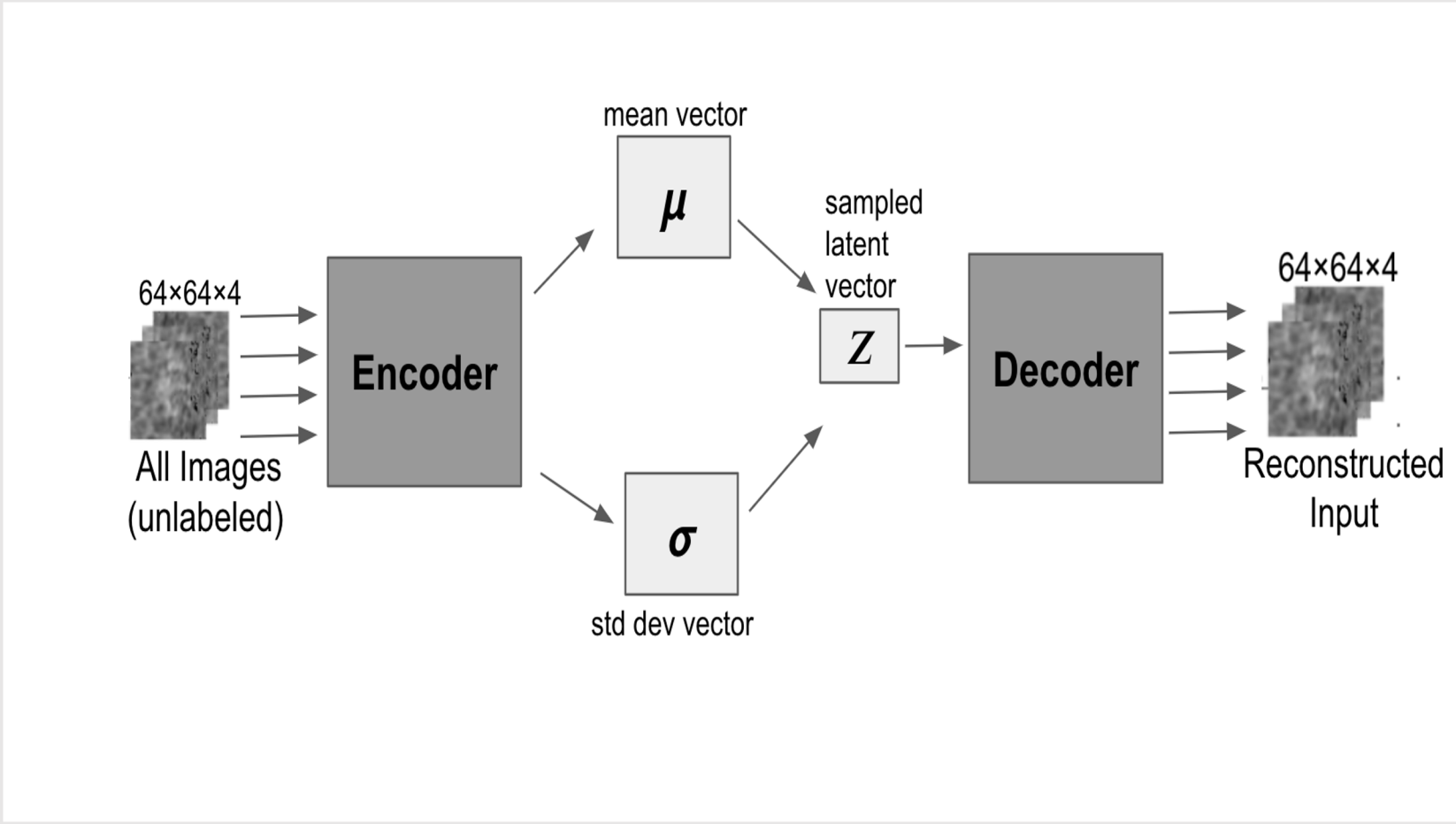
③ Applicare una soglia all'errore di ricostruzione del VAE per classificare una nuova immagine

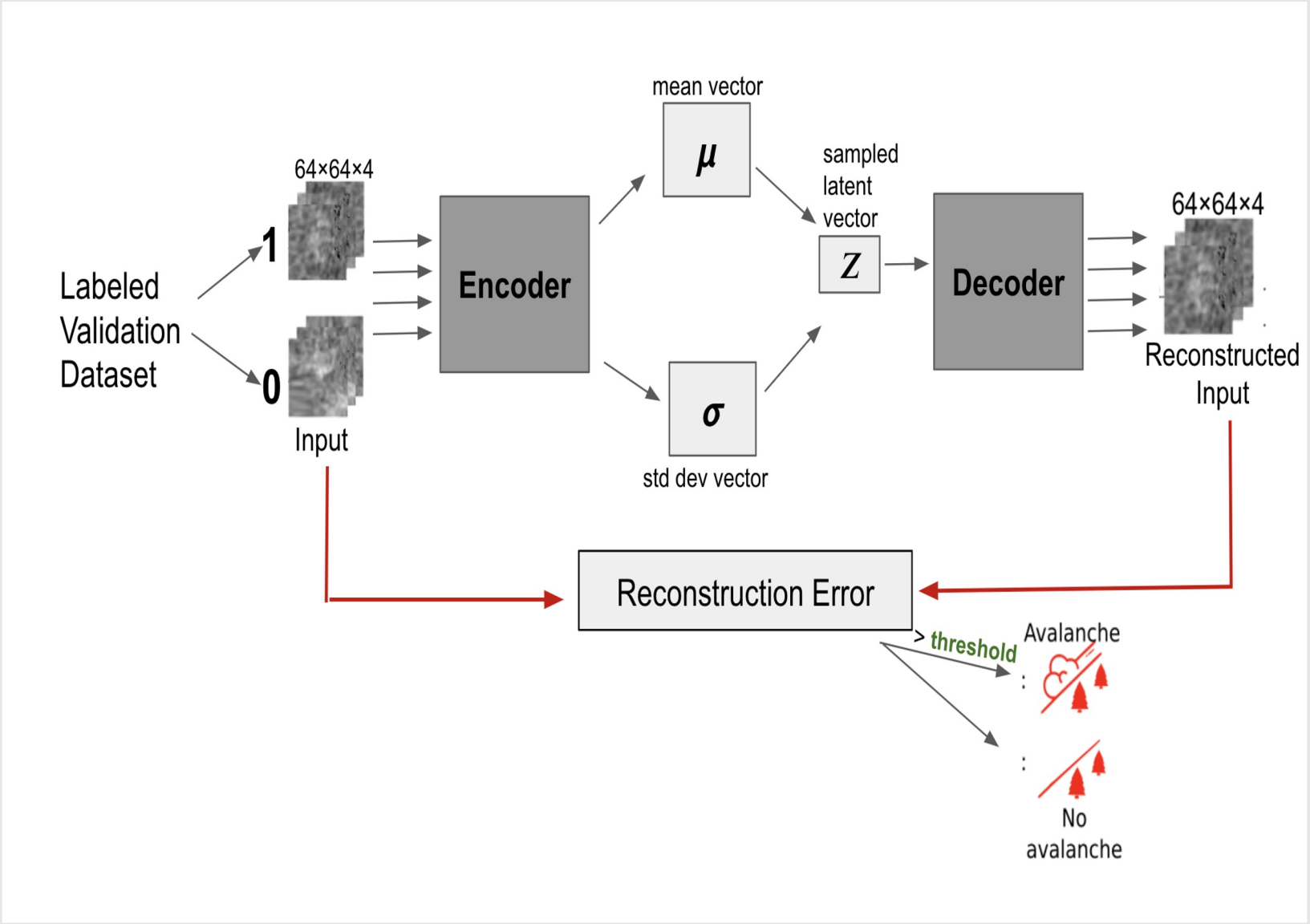
- Quando i dati etichettati scarseggiano, il VAE può essere addestrato **senza** supervisione!

VAE addestrato sugli esempi negativi

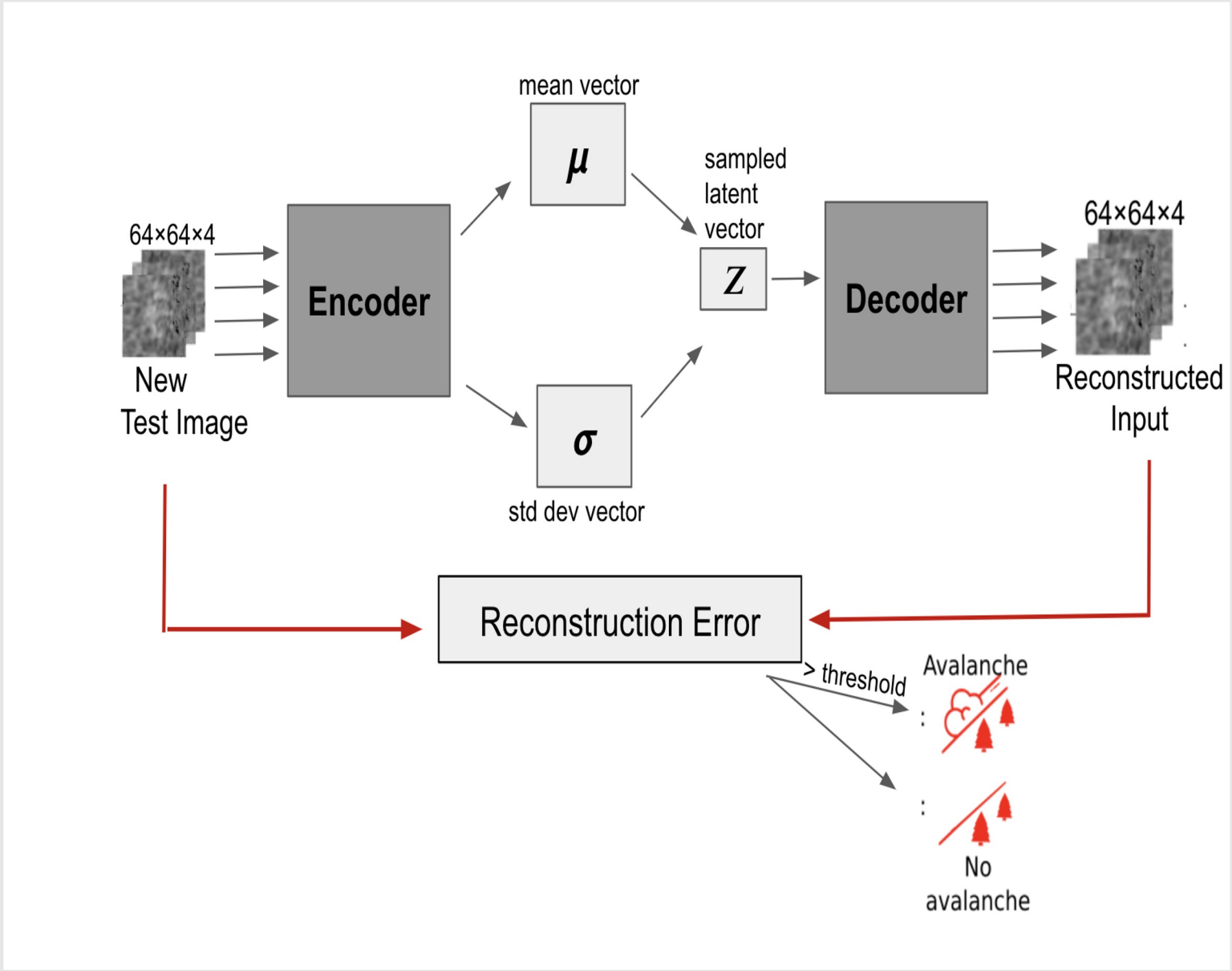


VAE addestrato su TUTTI gli esempi



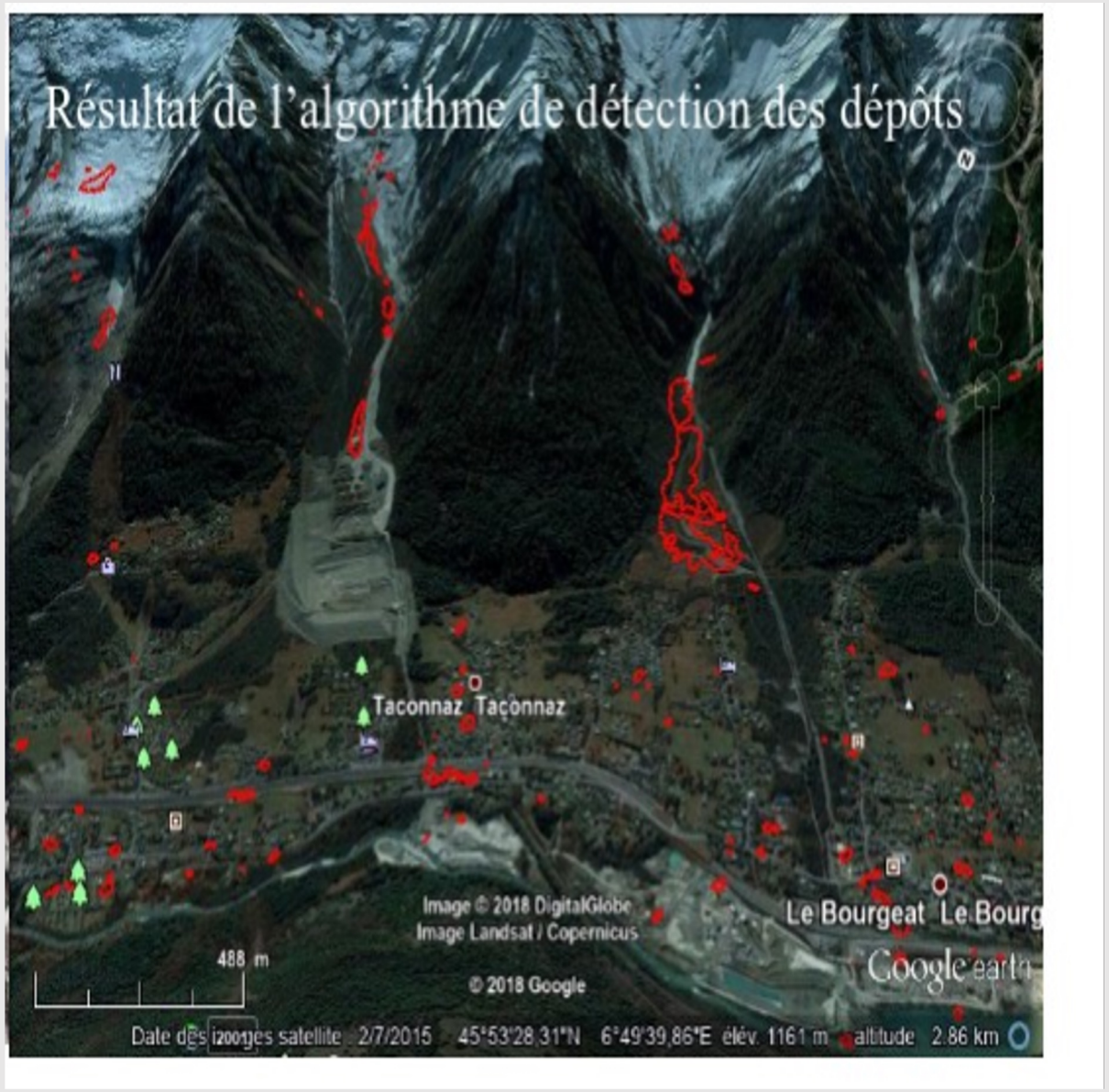


Regolare la soglia per il rilevamento di anomalie



Pipeline di rilevamento delle anomalie

Metodo di riferimento: sogliatura



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

Valutazione

Uno dei massicci più esposti (sui 18 di «Tutte le Alpi»)

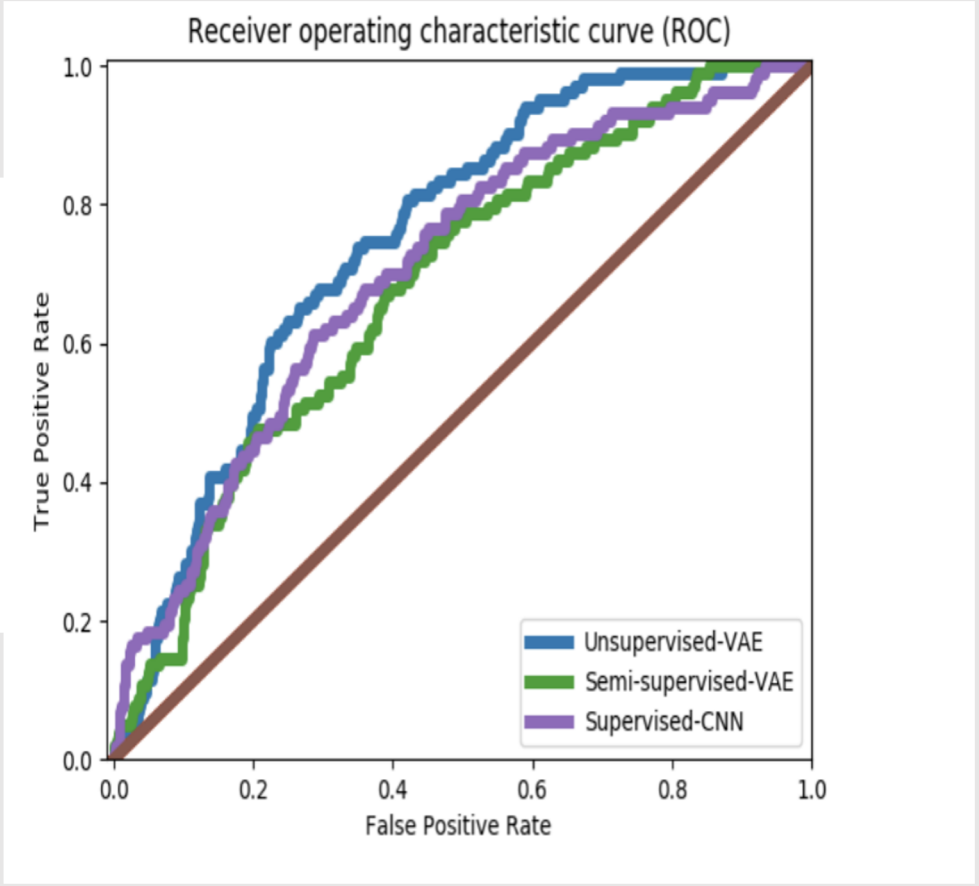
	All Alps		Haute Maurienne	
	Balanced Accuracy	F1-score	Balanced Accuracy	F1-score
Baseline	0.58	0.05	0.58	0.12
Supervised - CNN	0.53	0.10	0.53	0.12
Semi-supervised - VAE	0.59	0.11	0.6	0.23
Unsupervised - VAE	0.69	0.14	0.68	0.26

- Set di test riservato: 6.498 esempi etichettati
- Metodo di riferimento della letteratura sul rilevamento delle valanghe: sogliatura [Karbou et al., ISSW 2018]
- Metodo di riferimento di apprendimento supervisionato: rete neurale convoluzionale (CNN) addestrata su un dataset bilanciato artificialmente [Sinha et al., Climate Informatics 2019]

Valutazione

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

Curve ROC per la regione dell'Haute-Maurienne



Sintesi

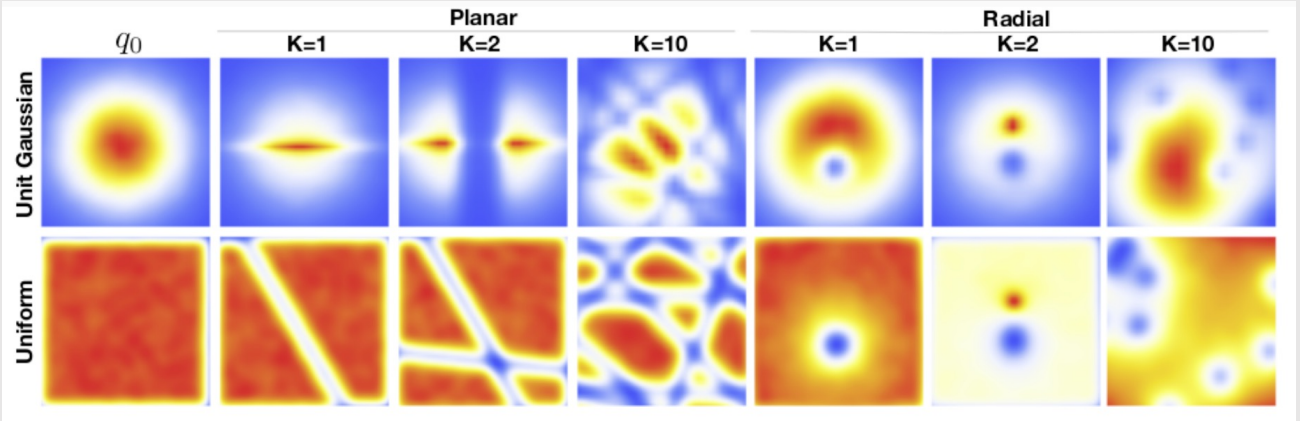
- Proposto un approccio semi-supervisionato per rilevare eventi rari quando i dati etichettati sono limitati
  - Questa pipeline semi-supervisionata si basa molto sull'apprendimento non supervisionato
- Può essere vista come una forma di sensore virtuale

Parte III: Metodi di IA:
IA generativa

IA generativa: una distinzione

- Probabilmente siete abituati al fatto che il termine «IA generativa» indichi grandi chatbot e agenti commerciali come ChatGPT, Gemini o Claude.
  - Per questi sistemi, i modelli di IA generativa sono stati addestrati su video e testi raccolti da internet
  - Ma questa **non è l'UNICA definizione** né l'unico uso dell'IA generativa!
- Noi possiamo **addestrare modelli di IA generativa da zero** per i nostri problemi
  - Per i problemi di clima e meteo NON addestreremo su testi e video, ma sui dati pertinenti, ad es. campi di valori reali misurati a diverse latitudini, longitudini, tempi ed eventualmente quote nell'atmosfera o profondità nell'oceano, tra cui:
    - Temperatura
    - Radiazione solare
    - Umidità
- Oggi vedremo i tipi di modelli generativi che potrete poi addestrare per i vostri problemi

Flussi normalizzanti

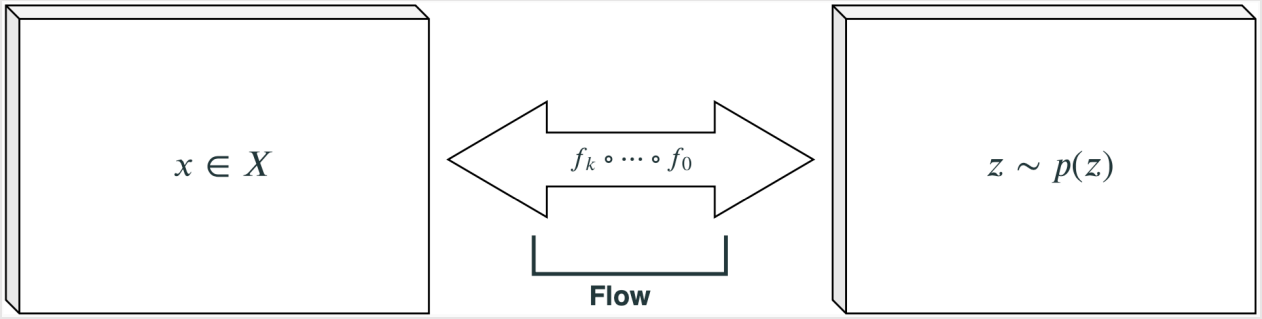


[Rezende & Mohamed, ICML 2015]

Si possono vedere come un'estensione del VAE oltre l'ipotesi gaussiana sullo spazio latente

Apprendere una serie di trasformazioni invertibili, $\{f_i\}$, a partire da una priori semplice sullo spazio latente Z, per ottenere distribuzioni più informative sullo spazio latente:

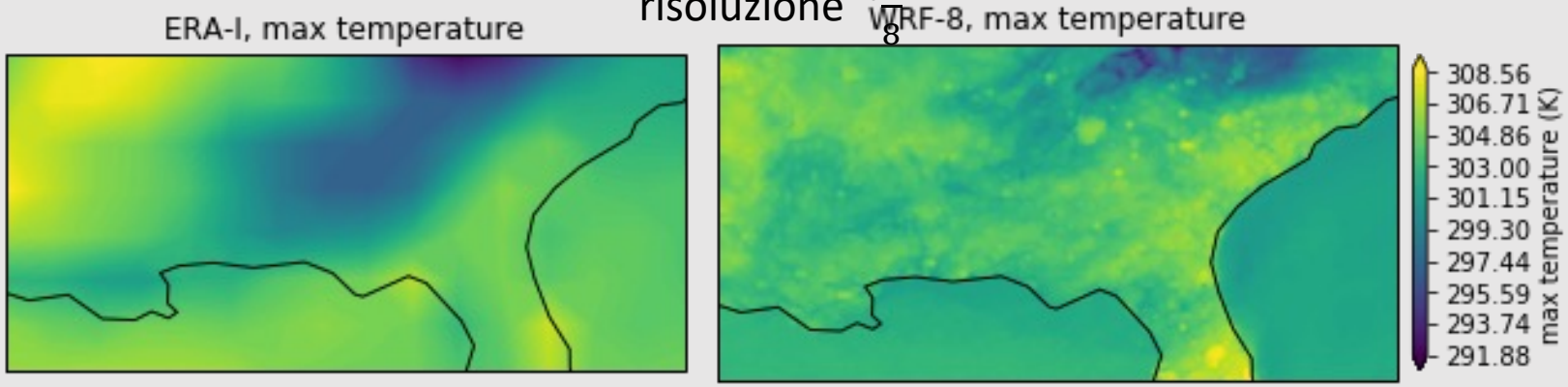
$$z_k = f_k \circ f_{k-1} \circ \cdots \circ f_1(z_0)$$

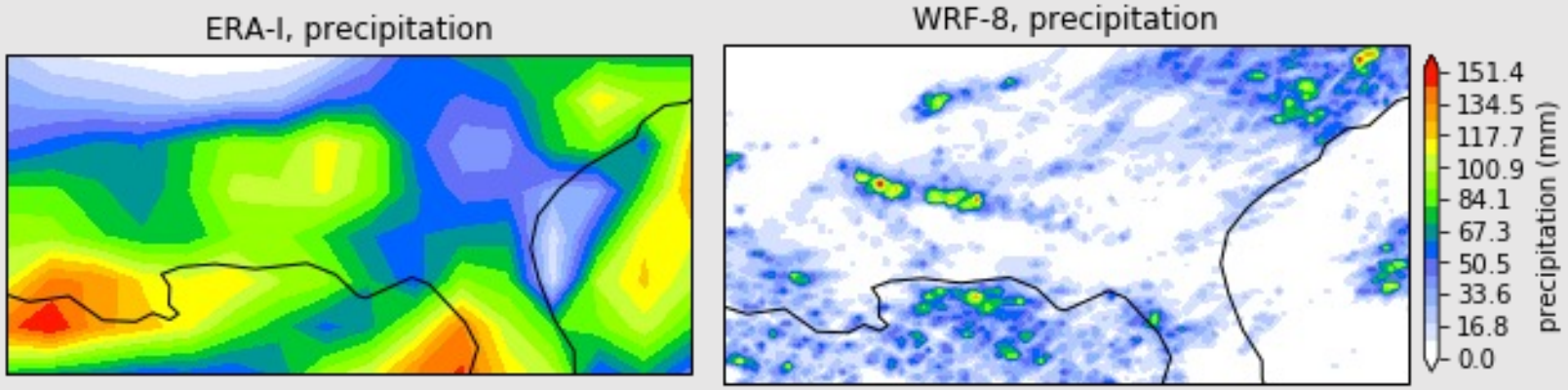


Flussi normalizzanti: applicazione al downscaling spaziale

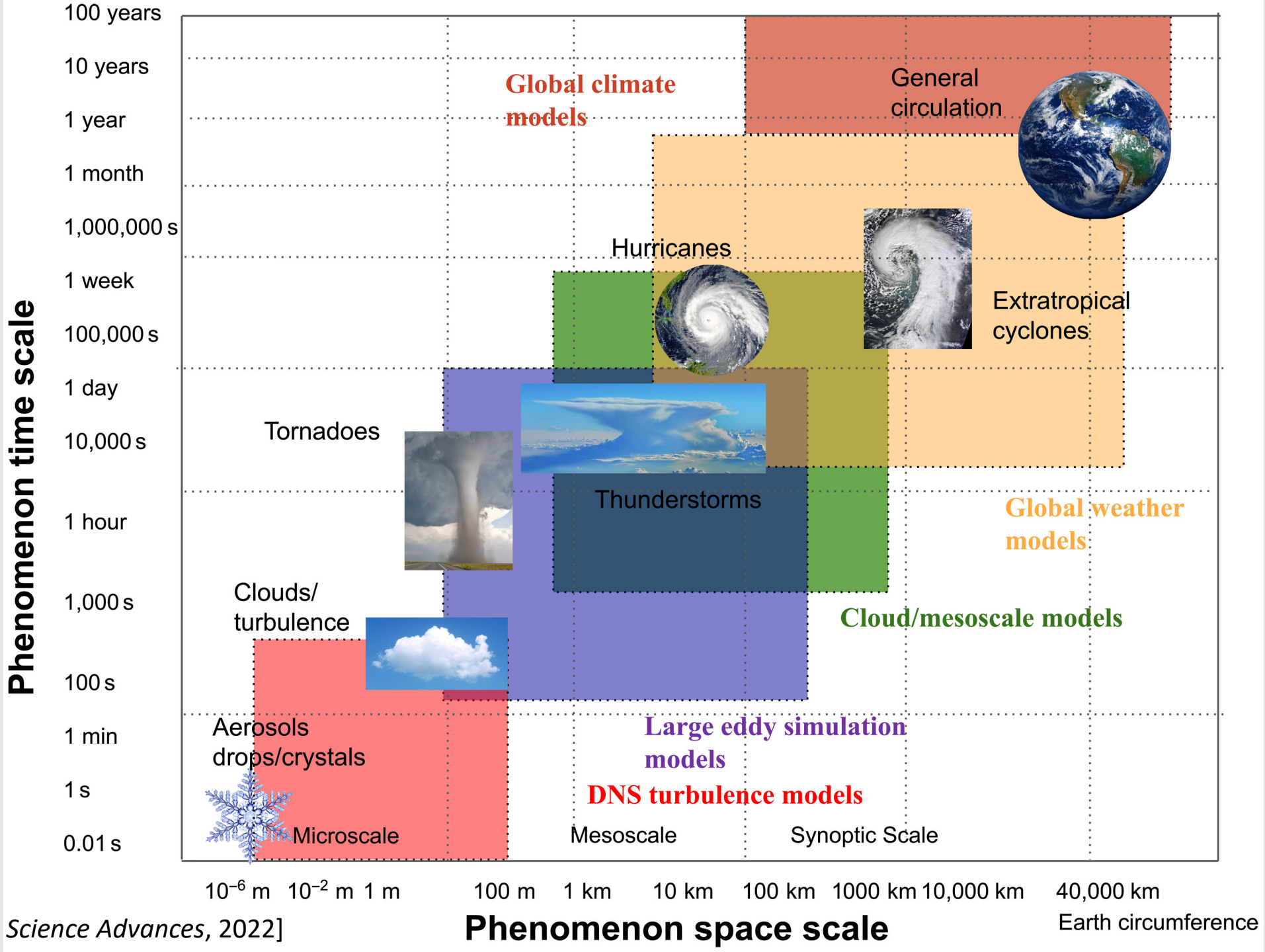
[Groenke, Madaus, & Monteleoni, Climate Informatics 2020]

ERA: dati di rianalisi, risoluzione 1°; WRF: previsione di modello meteorologico numerico, risoluzione $\frac{1}{8}°$









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

Il downscaling come allineamento di domini

- Compito di allineamento di domini: date variabili aleatorie X, Y, apprendere una mappatura f: X Y tale che, per ogni x_i X e y_i Y,

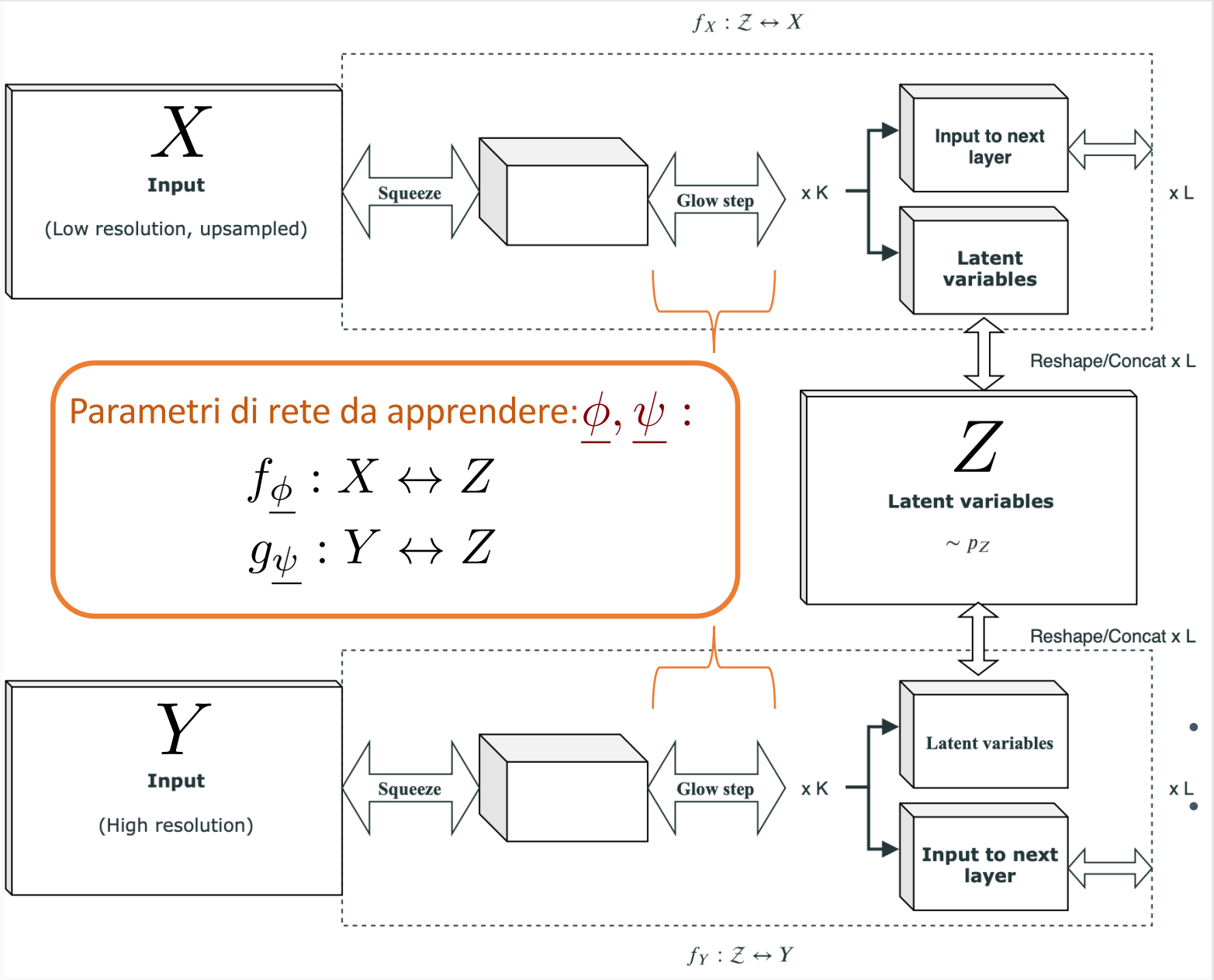
$$f(x_i) \sim P_Y \qquad f^{-1}(y_i) \sim P_X$$

- Il downscaling come allineamento di domini
  - Dati campioni i.i.d. a bassa risoluzione (X) e ad alta risoluzione (Y)
  - Apprendere la densità congiunta di X e Y assumendo indipendenza condizionale tramite uno spazio latente condiviso Z,

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

  - Modellare $P(x|z), P(y|z)$ con AlignFlow [Grover et al. 2020]
    - Partendo da una priori semplice su P_Z, apprendere flussi normalizzanti
    - Non serve accoppiare gli esempi x e y!

Architettura di ClimAlign



- L'architettura segue AlignFlow [Grover et al., 2020]
- Flusso normalizzante: Glow [Kingma & Dhariwal, 2018]

Confronto con baseline supervisionate

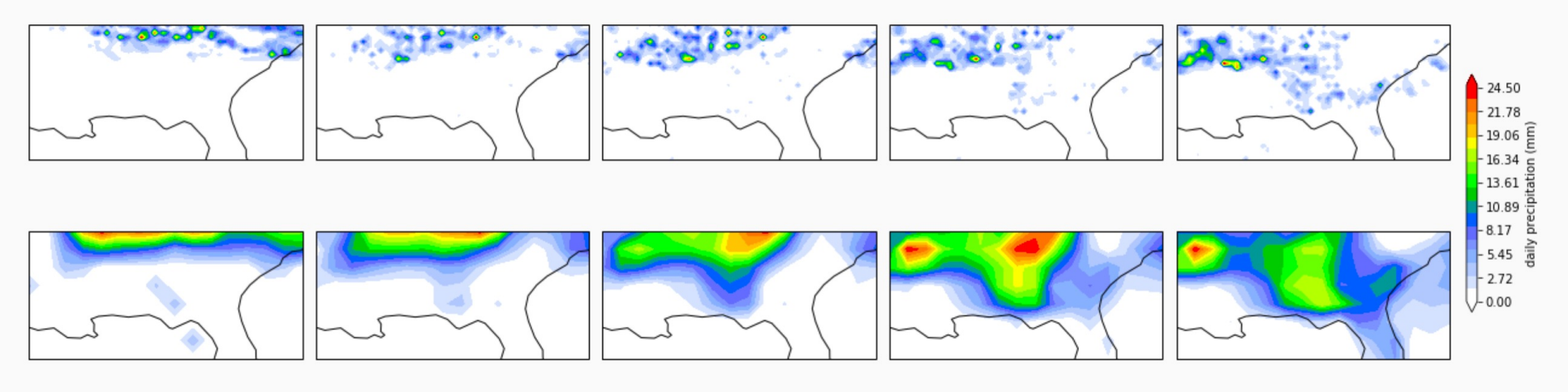
Temperatura massima giornaliera

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

Precipitazioni giornaliere

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

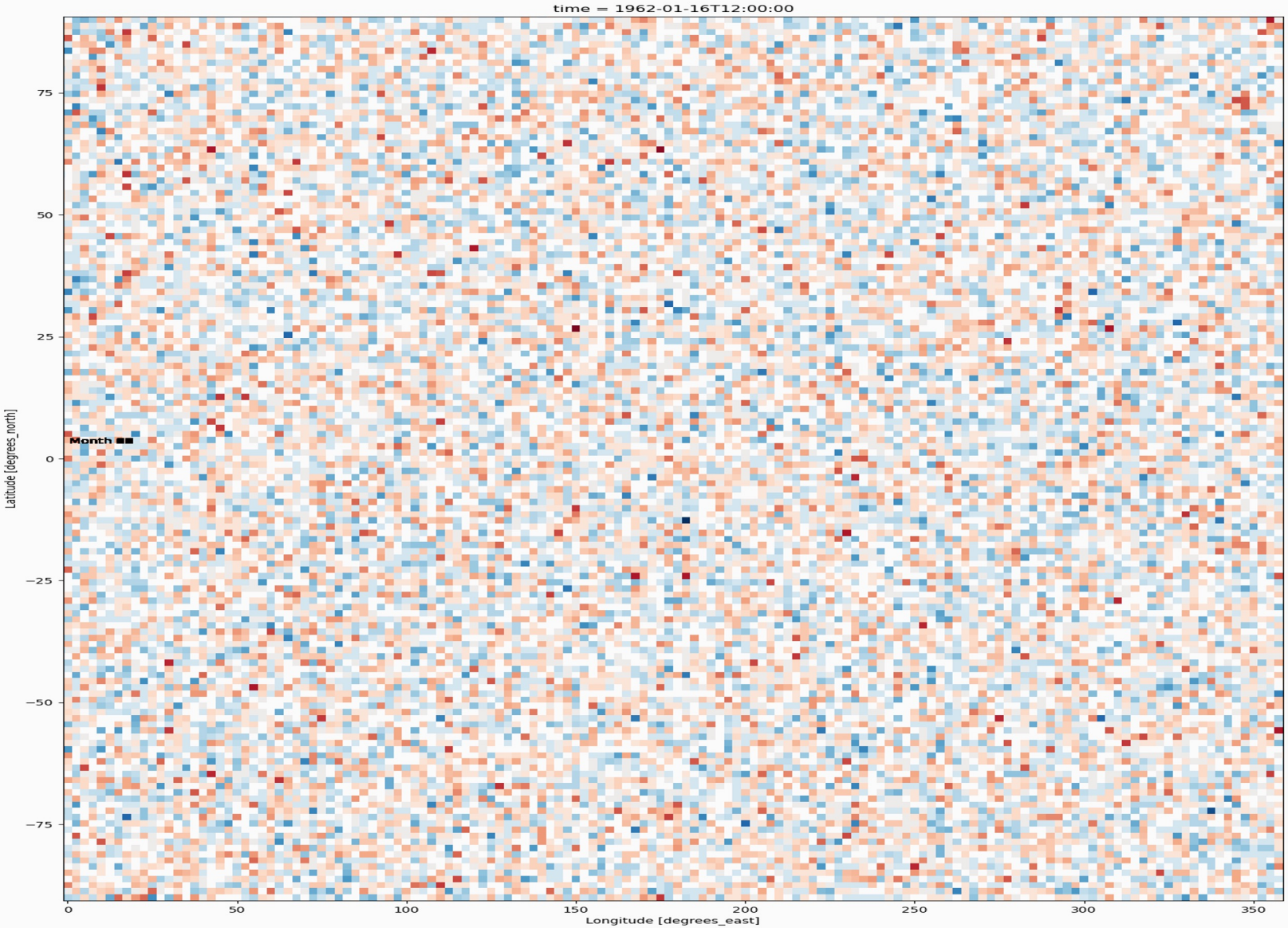
ClimAlign: downscaling generativo non supervisionato

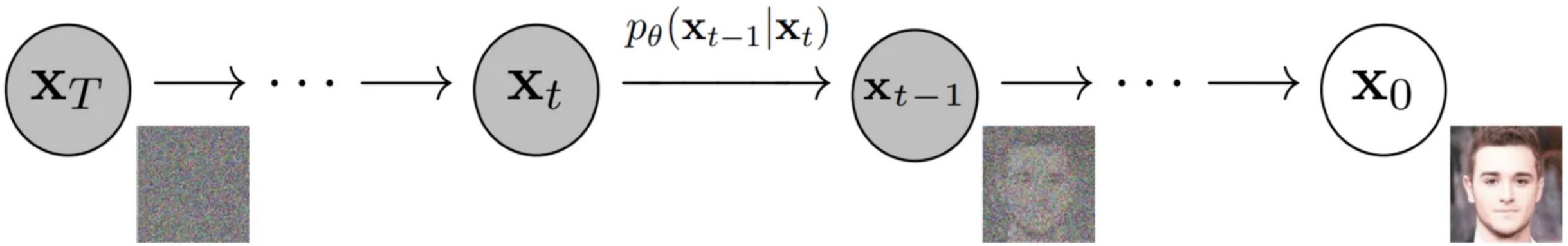


Tecnica generale di downscaling tramite allineamento di domini con flussi normalizzanti [AlignFlow: Grover et al., AAAI 2020][Glow: Kingma & Dhariwal, NeurIPS 2018]

- **Non supervisionato**: non servono mappe accoppiate a bassa e alta risoluzione
- **Generativo**: si può campionare dalla posteriore sulla rappresentazione latente O campionare condizionatamente a una mappa a bassa (o alta!) risoluzione
- **Interpretabile**, ad es. tramite interpolazione

Modelli generativi basati su flussi





Addestramento per diffusione: imparare a generare immagini realistiche da rumore casuale

Addestramento del modello di diffusione

1. Creare un compito pretesto da dati di addestramento non etichettati: per ogni punto di addestramento x_0, creare una sequenza di immagini sintetiche aggiungendo rumore gaussiano all'immagine precedente, fino a ottenere un'immagine indistinguibile dal rumore gaussiano.

2. Fare poi un apprendimento «supervisionato» sulle coppie (x_{t-1}, x_t) per apprendere $p_{\boldsymbol{\theta}}(x_{t-1} \mid x_t)$

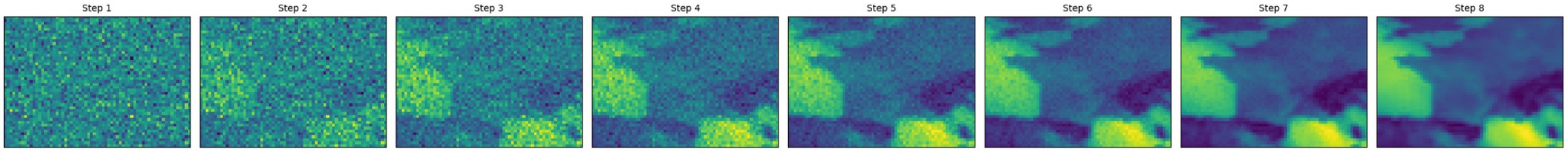
Inferenza, cioè come generare campioni simili ai dati di addestramento

1. Campionare x_T a caso (gaussiano).

2. Campionare T volte da $p_{\boldsymbol{\theta}}$, condizionando ogni volta sul campione precedente

Modelli generativi basati su flussi

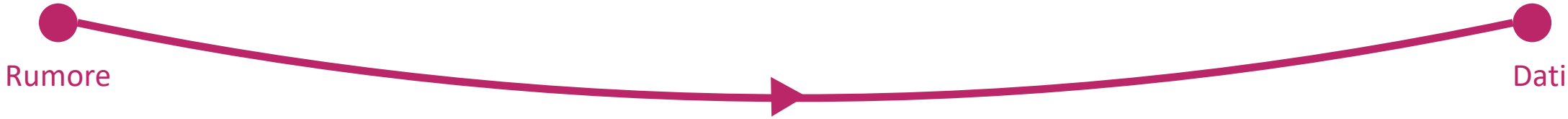
Modello generativo basato su flussi: apprende un trasporto (con una rete neurale) da una distribuzione facile da campionare (ad es. rumore gaussiano) alla distribuzione dei dati.



Diffusione: un «denoiser» inverte un processo di rumore stocastico, rimuovendo gradualmente il rumore – traiettoria di EDS



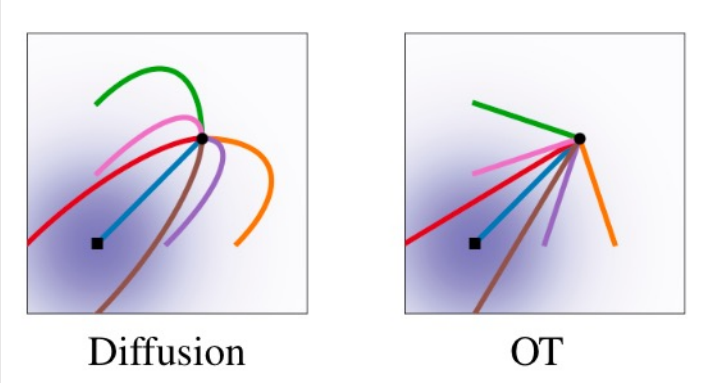
Flow matching [Lipman et al., 2023]: apprende un flusso deterministico rumore → dati – traiettoria di EDO



Crediti della slide: Julie Keisler (INRIA)

Flow matching vs diffusione

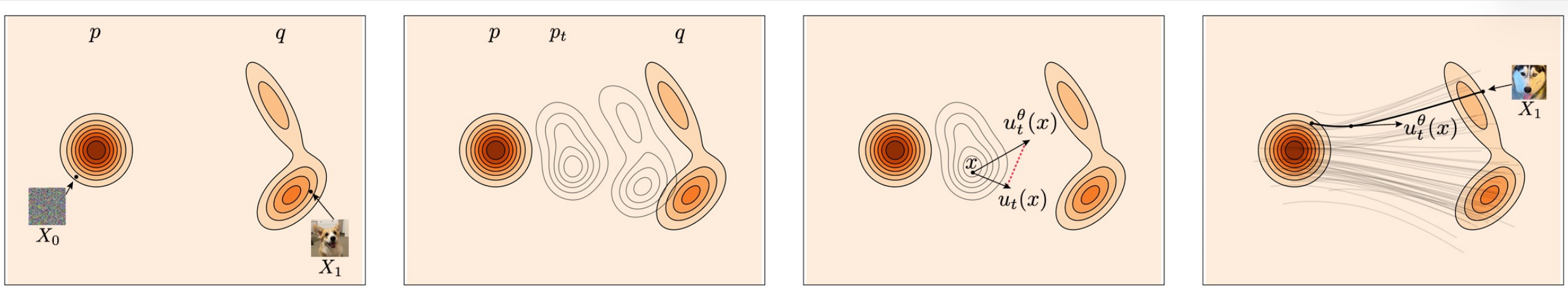
Intuizione: creare percorsi più rettilinei tra campioni di dati e campioni di rumore:



Risultati: percorsi più rettilinei nei flussi di probabilità, che consentono di campionare in meno passi di diffusione accelerazioni

Generalità: il flow matching può collegare due distribuzioni qualsiasi (o interpolanti stocastici), a differenza della diffusione

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.

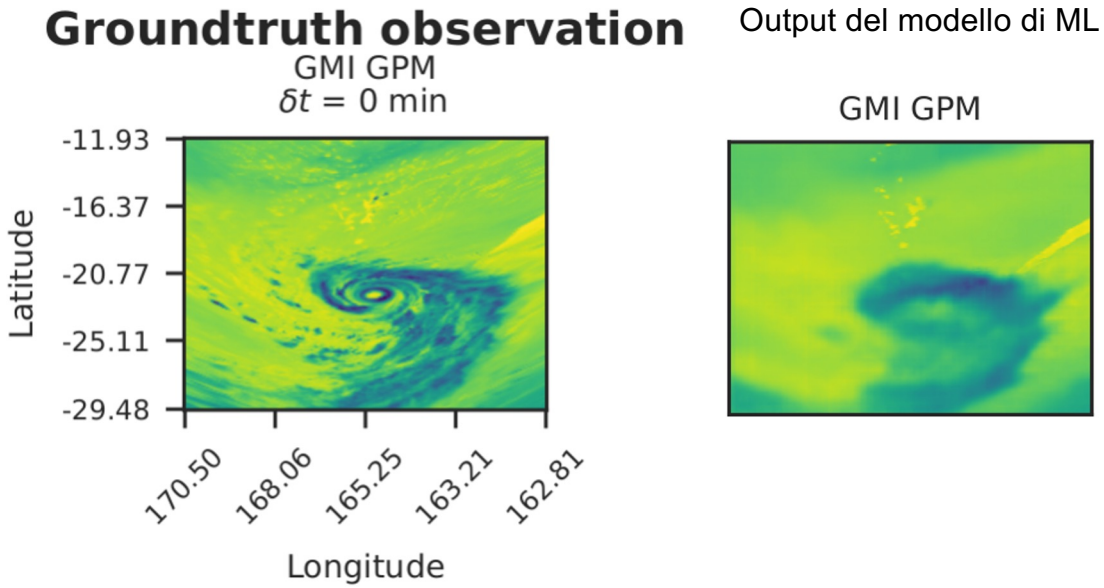
Lipman et al. (2024) Flow Matching Guide and Code

Perché usare l'IA generativa per clima e meteo?

I modelli di ML deterministici spesso ottimizzano le prestazioni **medie**

Addestrati a minimizzare una loss media, ad es. la RMSE

➔ Output di modello sfocati e non fisici (ad es. immagini, campi, traiettorie di tempeste)



Output del modello di ML

Output del modello di ML, campionati dopo l'addestramento la stessa backbone, **in modo generativo**

Perché usare l’IA generativa per clima e meteo?

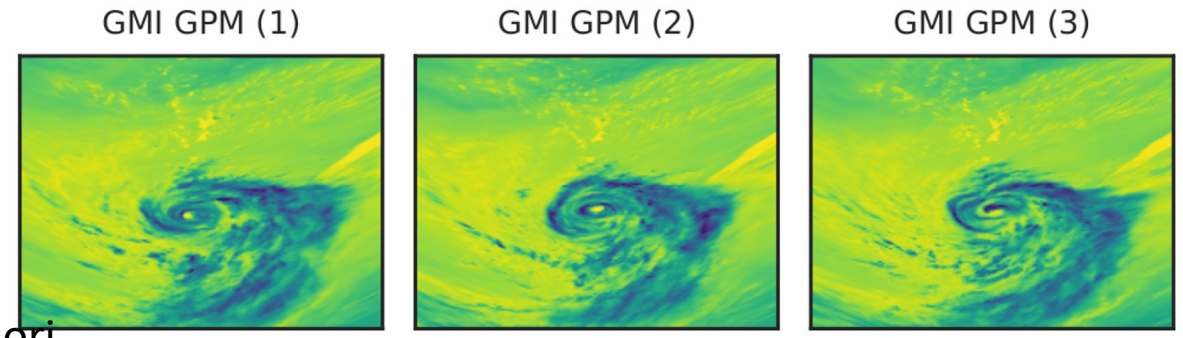
Per molti problemi di clima, meteo e geoscienze

- Previsioni meteo, emulazione del clima, generazione di scenari
- Stima di estremi o eventi rari, quantili
- Downscaling a scale locali nel tempo o nello spazio
- Quantificazione del rischio

L’IA generativa offre vantaggi

- Maggiore realismo fisico
  - Maggiore accuratezza delle alte frequenze
- Ensemble probabilistici
  - Quantificazione dell’incertezza
  - Informazioni ricche e realistiche per i decisori
- Generazione di dati sintetici

Campioni da un **modello generativo**



IA generativa per:

✓ Previsione probabilistica
- **ArchesWeatherGen**: previsioni meteo efficienti addestrate a 1,5°
- **ArchesClimate**: emulazione, interpolazione ed estrapolazione del clima decennale

❑ Fusione dei dati con allineamento probabilistico di domini
- **MotifGen**: un modello generativo, multi-fonte e auto-supervisionato per apprendere da dati geospaziali

❑ Allineamento di domini probabilistico e downscaling
- **ClimAlign**: downscaling generativo non accoppiato con allineamento di domini
- **SerpentFlow**: allineamento di domini generativo non accoppiato tramite scomposizione in struttura condivisa (con applicazione al downscaling)

Le sfide dell'apprendimento da dati multi-fonte

Pléiades Neo

30 cm

Swath: 14 km

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

Revisit capacity:
Daily, anywhere

Daily acquisition capacity:
1,000,000 km²

Pléiades

50 cm

Swath: 20 km

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

Revisit capacity:
Daily, anywhere

Daily acquisition capacity:
700,000 km²

Vision-1

87 cm

Swath: 20.8 km

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

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

Daily acquisition capacity:
20,000 km²

SPOT

1.5 m

Swath: 60 km

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

Revisit capacity:
Up to daily, anywhere

Daily acquisition capacity:
3,000,000 km²

Radar constellation

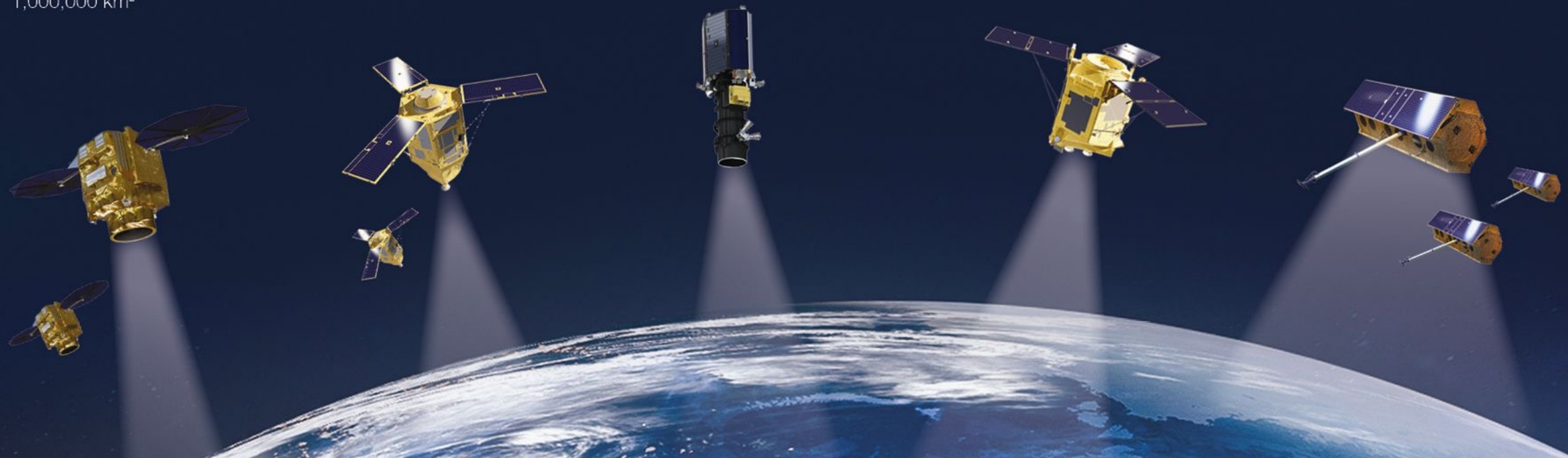
From 25 cm to 40 m

Swath: 270 km

Imaging modes:
SpotLight, StripMap, ScanSAR

Revisit capacity:
Daily for most latitudes

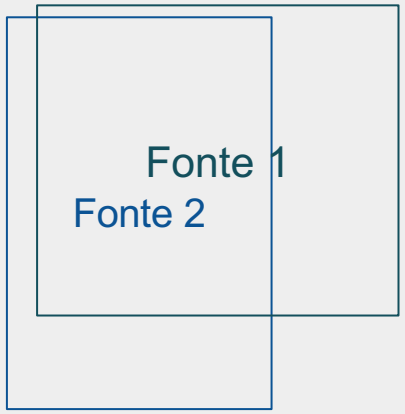
Daily acquisition capacity:
Up to 5,400,000 km²
(Wide ScanSAR Mode)



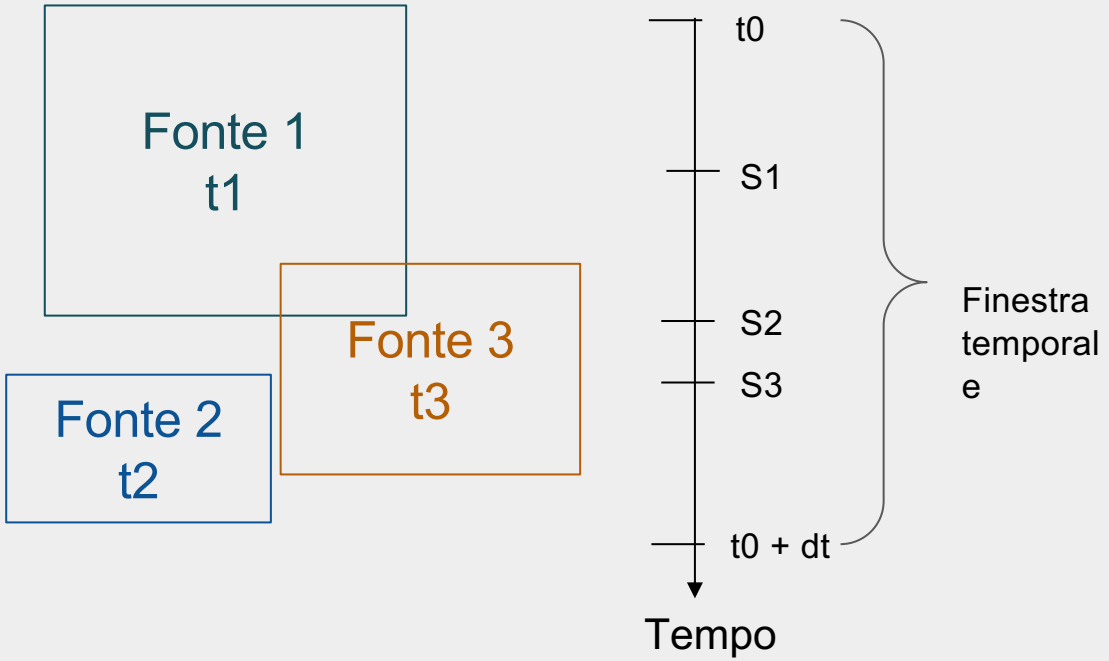
L’IA per la **fusione** di dati multi-fonte (disallineati)

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

Tradizionalmente: si addestrano modelli solo su dati co-localizzati



Il nostro modello di IA (MOTIF) può apprendere da **tutte le fonti di dati**, anche da dati disallineati nel tempo e nello spazio



Un’architettura multi-fonte per la fusione di dati geospaziali

ML di base multimodale	**Apprendimento di rappresentazioni geospaziali**
• Incorporare i **valori** (ad es. testo, immagini) • Condizionamento («cane o gatto»)	• Incorporare le **posizioni** • Addestramento auto-supervisionato su larga scala
Modello DiT (2022)	SatCLIP (2025), AlphaEarth (2025)

MOTIF: *Transformer multi-fonte con attenzione fattorizzata*

- Incorporare separatamente **valori** e **coordinate**
- Condizionare sulle caratteristiche della fonte
- Pre-addestramento auto-supervisionato

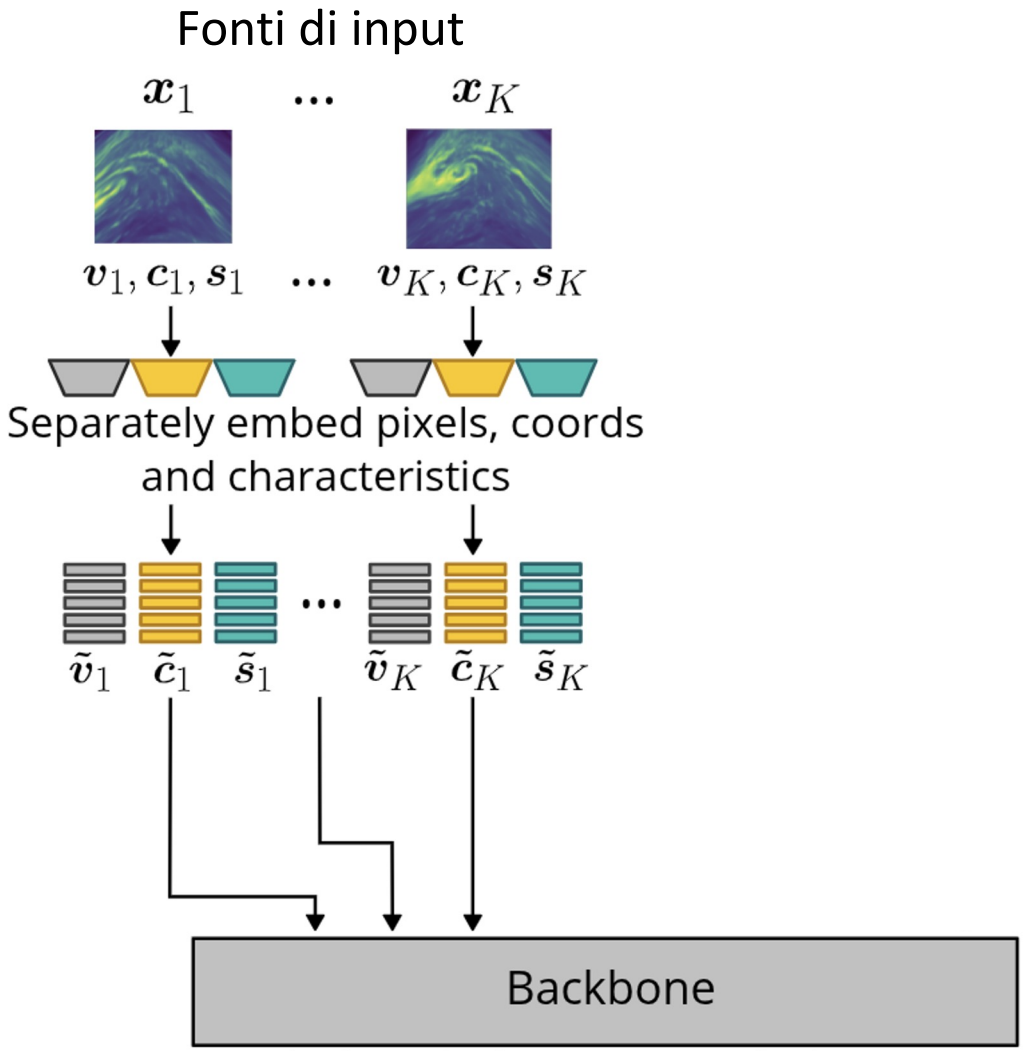
[Dauvilliers & Monteleoni, Environmental Data Science, 2025]



Pipeline di addestramento

Incorporare separatamente:

- valori,
- coordinate,
- caratteristiche della fonte



Embedding posizionale relativo

- Il consueto embedding posizionale relativo dei Transformer (Shaw et al., 2018)

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

- Qui le coordinate spazio-temporali sono usate per l'embedding posizionale:

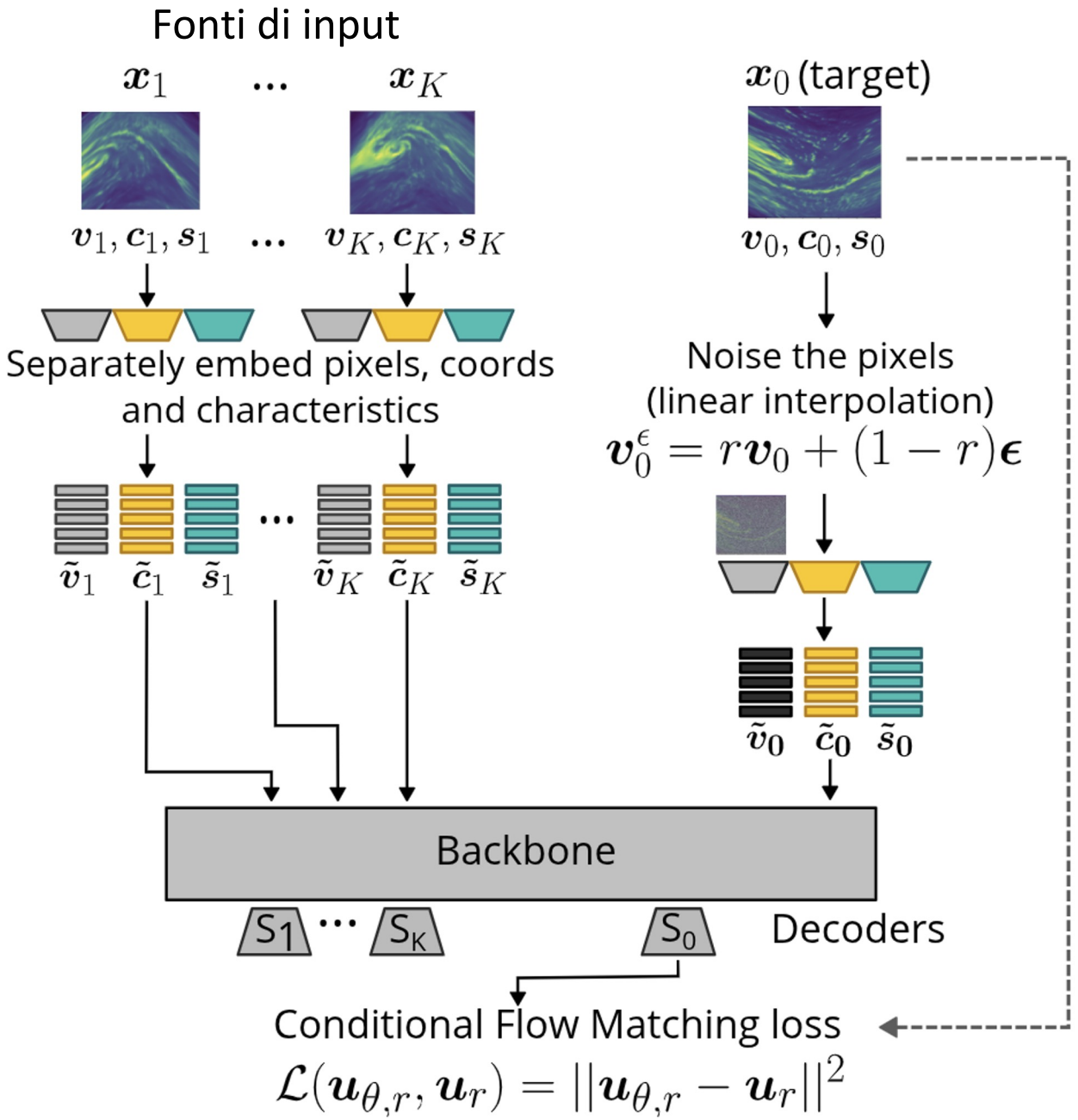
$$\tilde{v}_{out} = \frac{Softmax\left(Q_V K_V^T + Q_C K_C^T\right)}{\sqrt{d}} \tilde{v}_{in}$$

Valori (ad es. pixel)

Coordinate spazio-temporali

Pipeline di addestramento

Incorporare separatamente:
- valori,
- coordinate,
- caratteristiche della fonte



Compito di addestramento auto-supervisionato: ricostruire una fonte mascherata

Versione generativa addestrata con flow matching

[Dauvilliers & Monteleoni, Climate Informatics 2026]

Applicazione ai cicloni tropicali

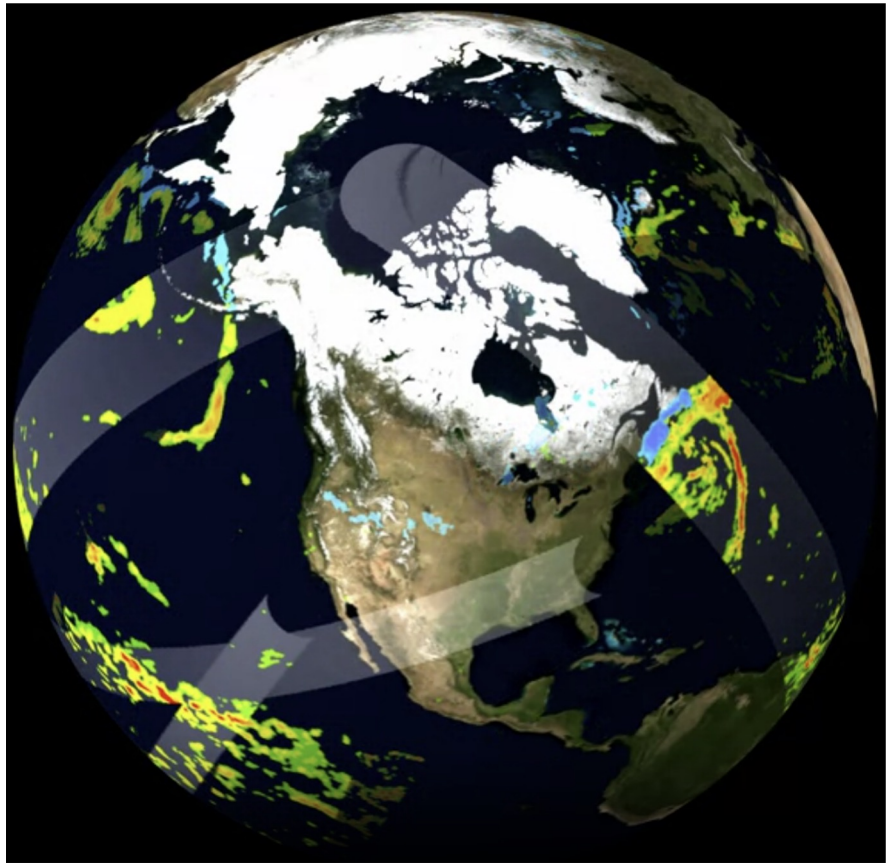
La costellazione GPM

GMI (strumento di riferimento)



⇒ Possiamo fare interpolazione spazio-temporale delle immagini GMI?

La costellazione GPM, NASA



La costellazione GPM copre la Terra, NASA's Scientific Visualization Studio

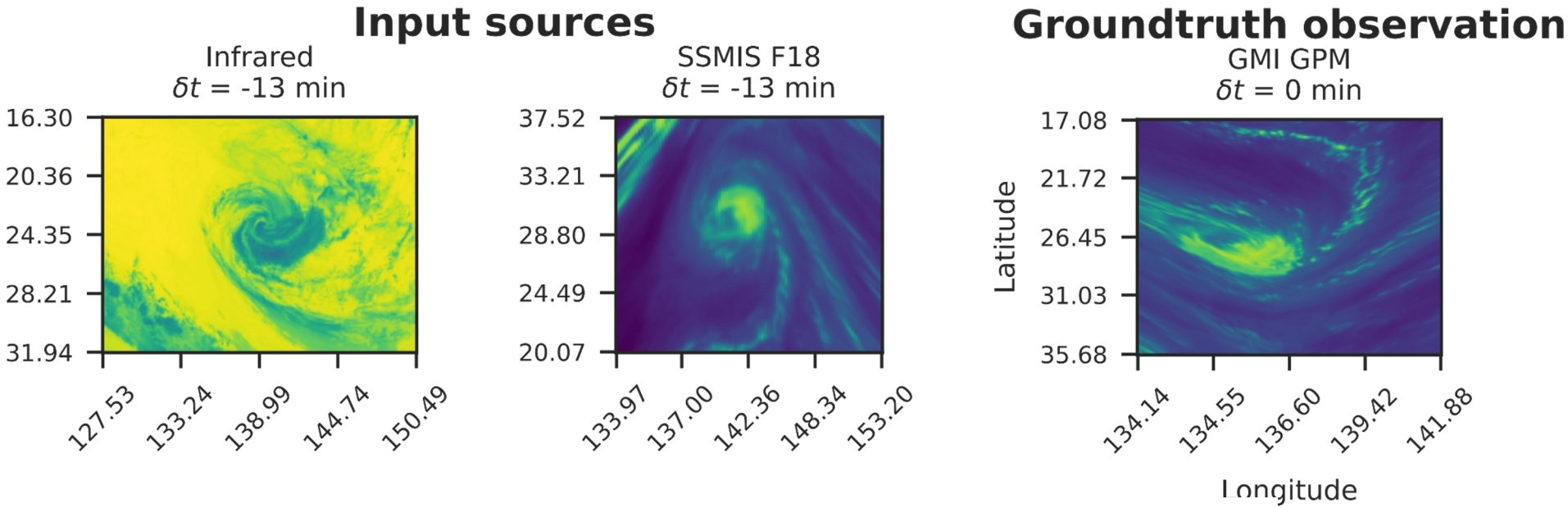
Dataset

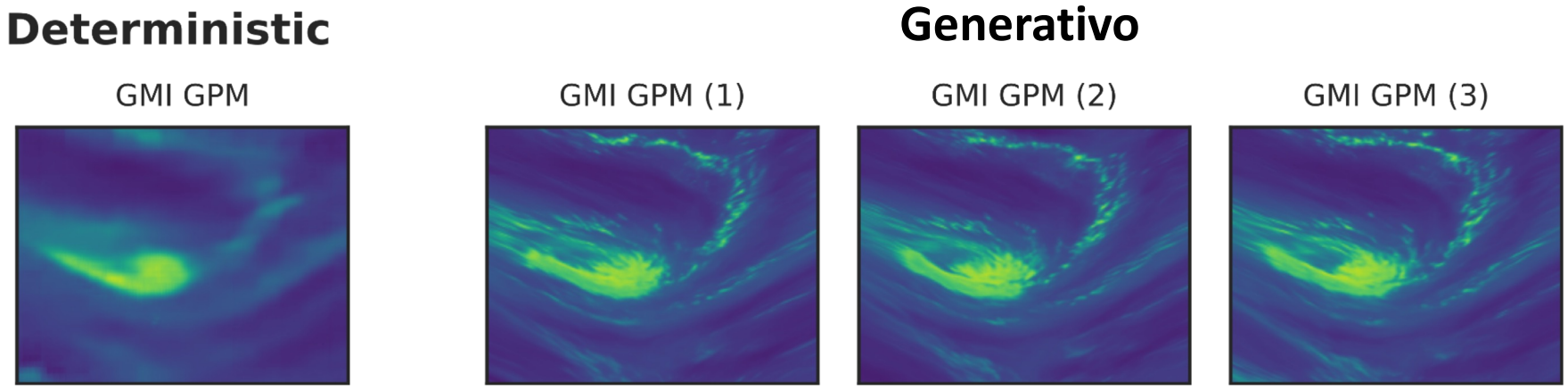
TC-PRIMED (Razin et al. 2023)

- ~3.000 tempeste, dal 1987 al 2024, tutti i bacini
- 2005-07-14-21-23 per la validazione, 2006-08-15-22-24 per il test
- 11 sensori a microonde
- Osservazioni nell'infrarosso

Finestre di interpolazione di 6 ore

Risultati: coglie meglio la struttura fine rispetto al modello deterministico





Pipeline di addestramento a confronto per la versione generativa

- Addestramento supervisionato



⇒ ~6.000 campioni

- Addestramento auto-supervisionato



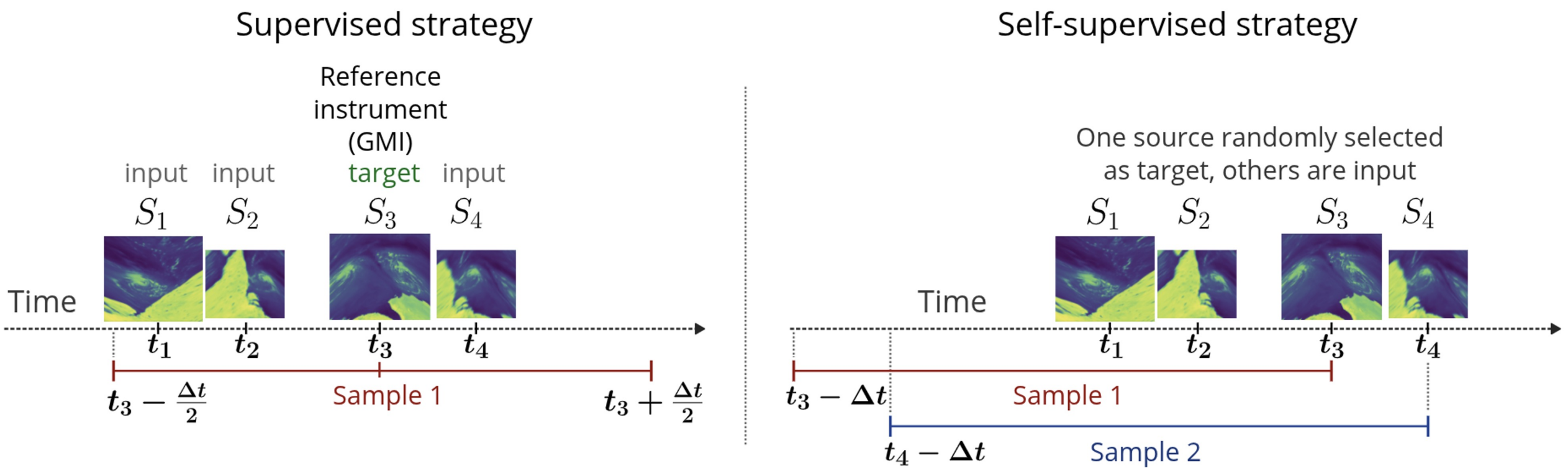
⇒ ~48.000 campioni

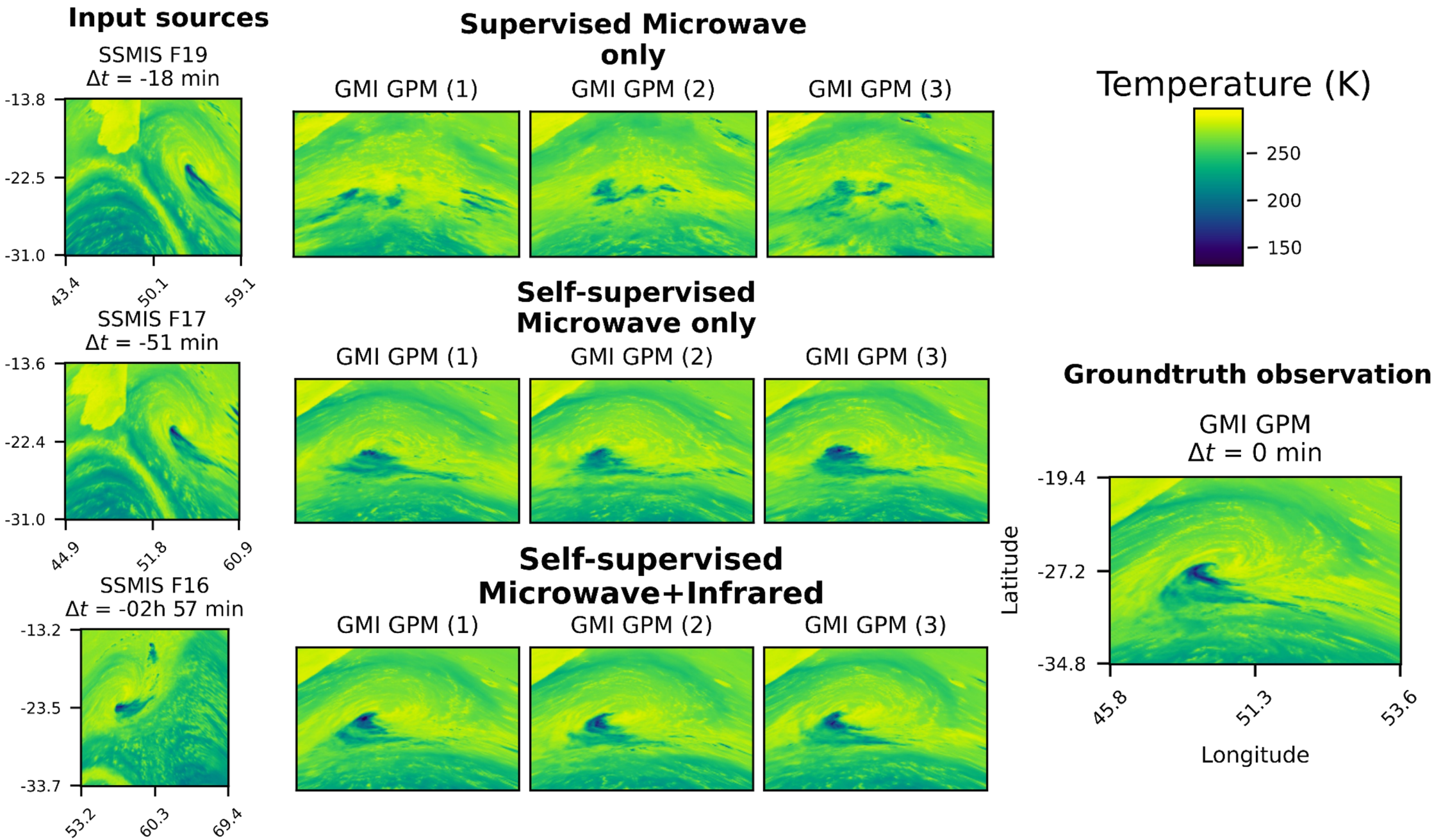
- Auto-supervisionato + infrarosso

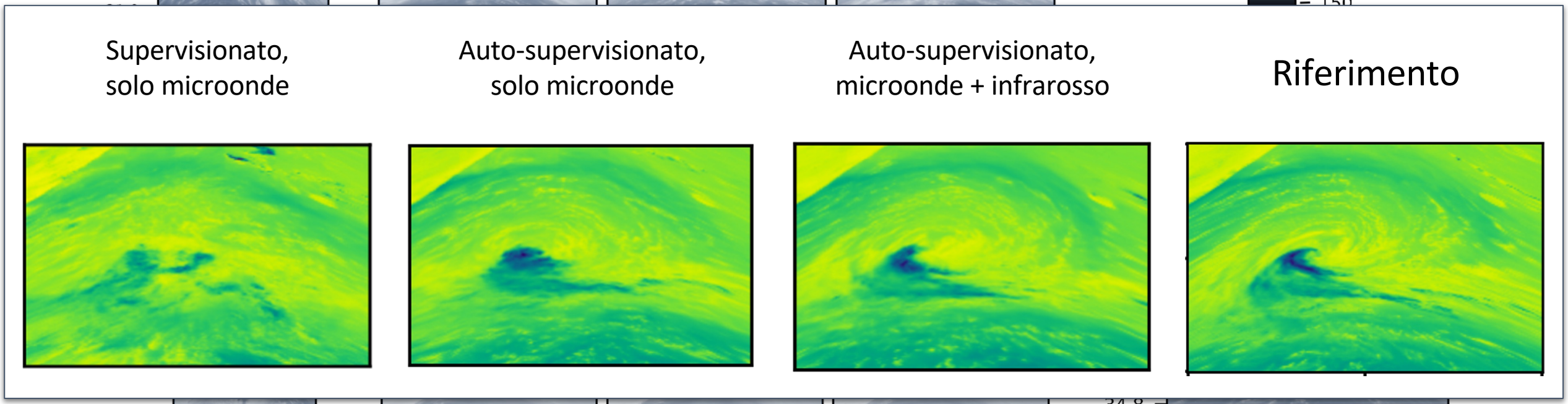


⇒ ~44.000 campioni

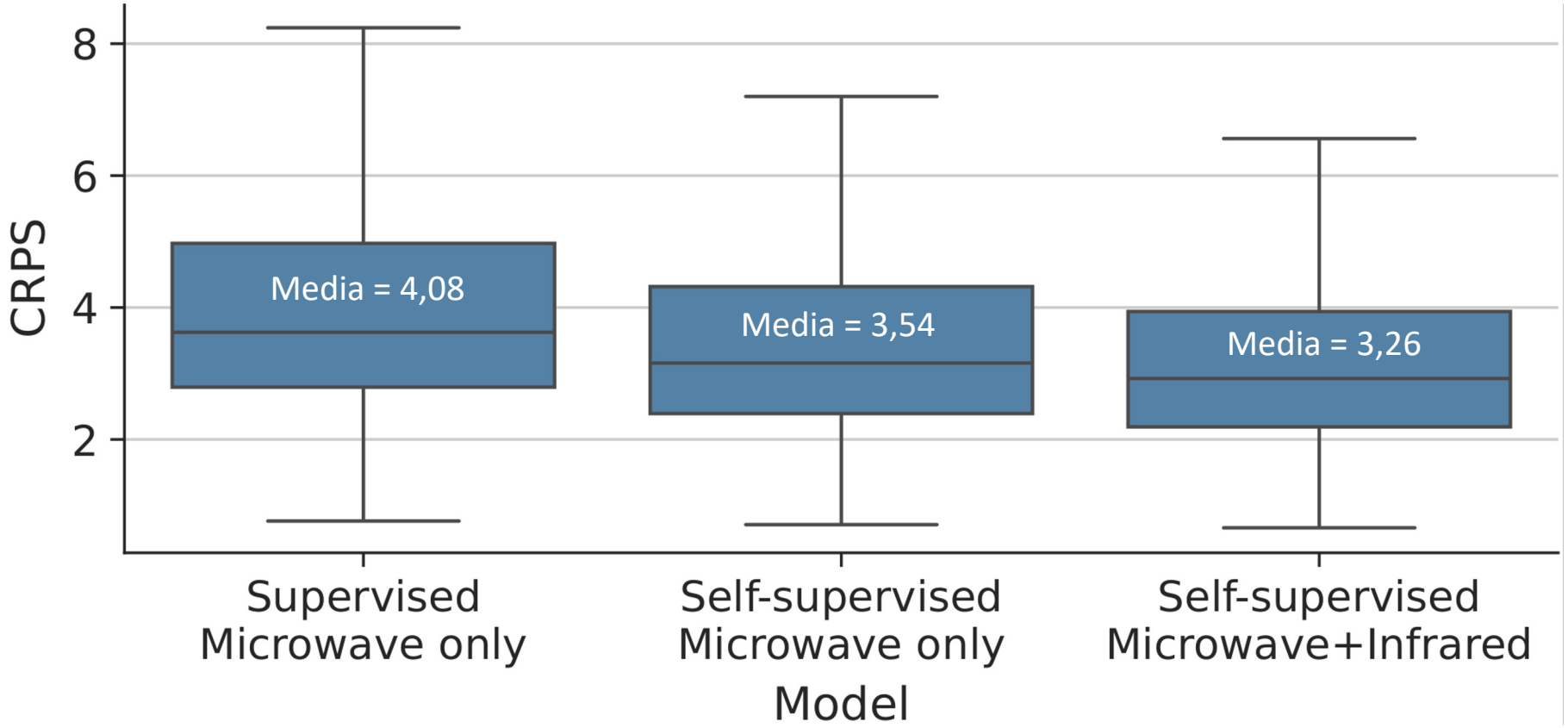
Compiti di addestramento




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


Supervisionato, solo microonde
Auto-supervisionato, solo microonde
Auto-supervisionato, microonde + infrarosso
Riferimento

Versione generativa: ablazioni



Risultati

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 Baseline	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]

Le parentesi indicano intervalli di confidenza al 95%

Continuous Ranked Probability Score (CRPS)

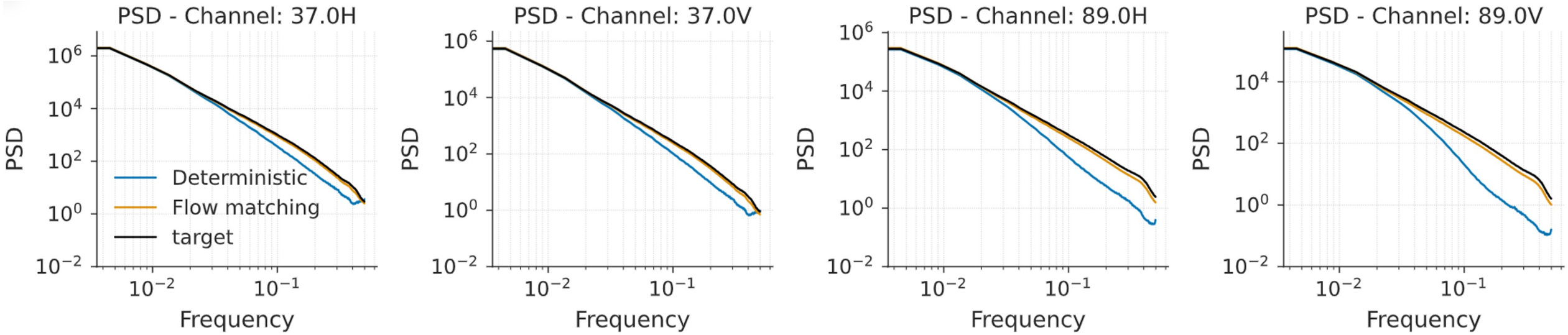
$$\mathrm{CRPS}(F, x) = -\int_{-\infty}^{\infty} (F(y) - \mathbb{1}\{y \geq x\})^2 \, \mathrm{d}y$$

- Una formulazione diversa evita il calcolo di questo integrale
- Il CRPS si può confrontare direttamente con il MAE di una previsione deterministica
- L'Energy Score è l'estensione multivariata del CRPS

Letture di approfondimento su queste e altre misure di valutazione delle previsioni probabilistiche:

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

Valutazione rispetto al modello deterministico - spettro

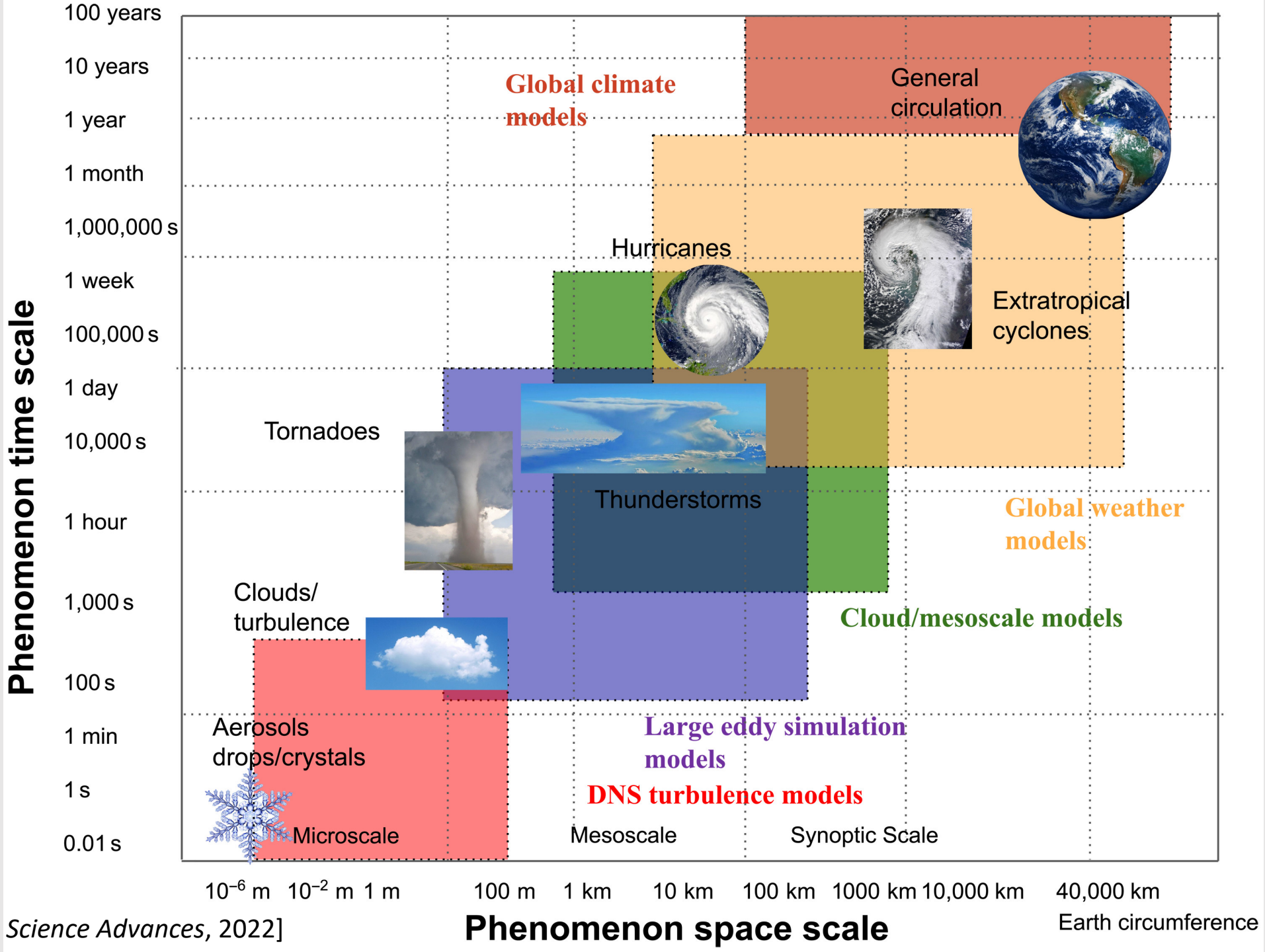


IA generativa per:

✓ Previsione probabilistica
- **ArchesWeatherGen**: previsioni meteo efficienti addestrate a 1,5°
- **ArchesClimate**: emulazione, interpolazione ed estrapolazione del clima decennale

✓ Fusione dei dati con allineamento probabilistico di domini
- **MotifGen**: un modello generativo, multi-fonte e auto-supervisionato per apprendere da dati geospaziali

❑ Allineamento di domini probabilistico e downscaling
- **ClimAlign**: downscaling generativo non accoppiato con allineamento di domini
- **SerpentFlow**: allineamento di domini generativo non accoppiato tramite scomposizione in struttura condivisa (con applicazione al downscaling)



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



Downscaling non supervisionato dei modelli climatici tramite allineamento di domini: *applicazione ai campi di vento.*

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

Atti di Climate Informatics 2026

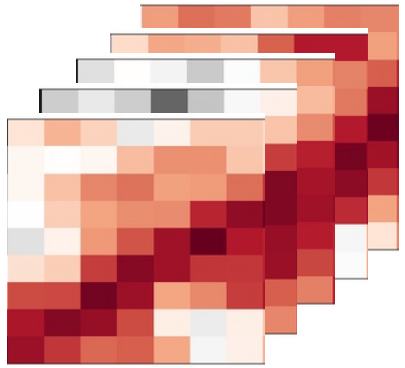
Downscaling dei modelli climatici

Risoluzione spaziale grossolana: i GCM operano a ~100+ km e trascurano processi locali cruciali

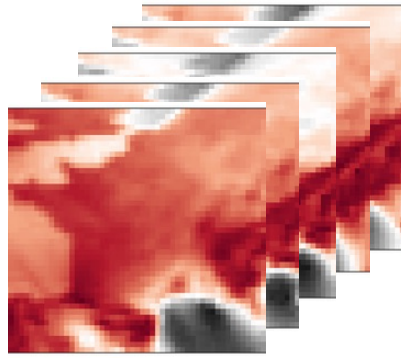
Rappresentazione limitata: le strutture fini, come gli effetti topografici e i contrasti terra-mare, sono colte male

Bias sistematici: gli output dei GCM si discostano molto dai regimi climatici osservati

Studi d'impatto locali compromessi: gli output diretti dei GCM non bastano per valutare i futuri regimi di vento per la pianificazione energetica

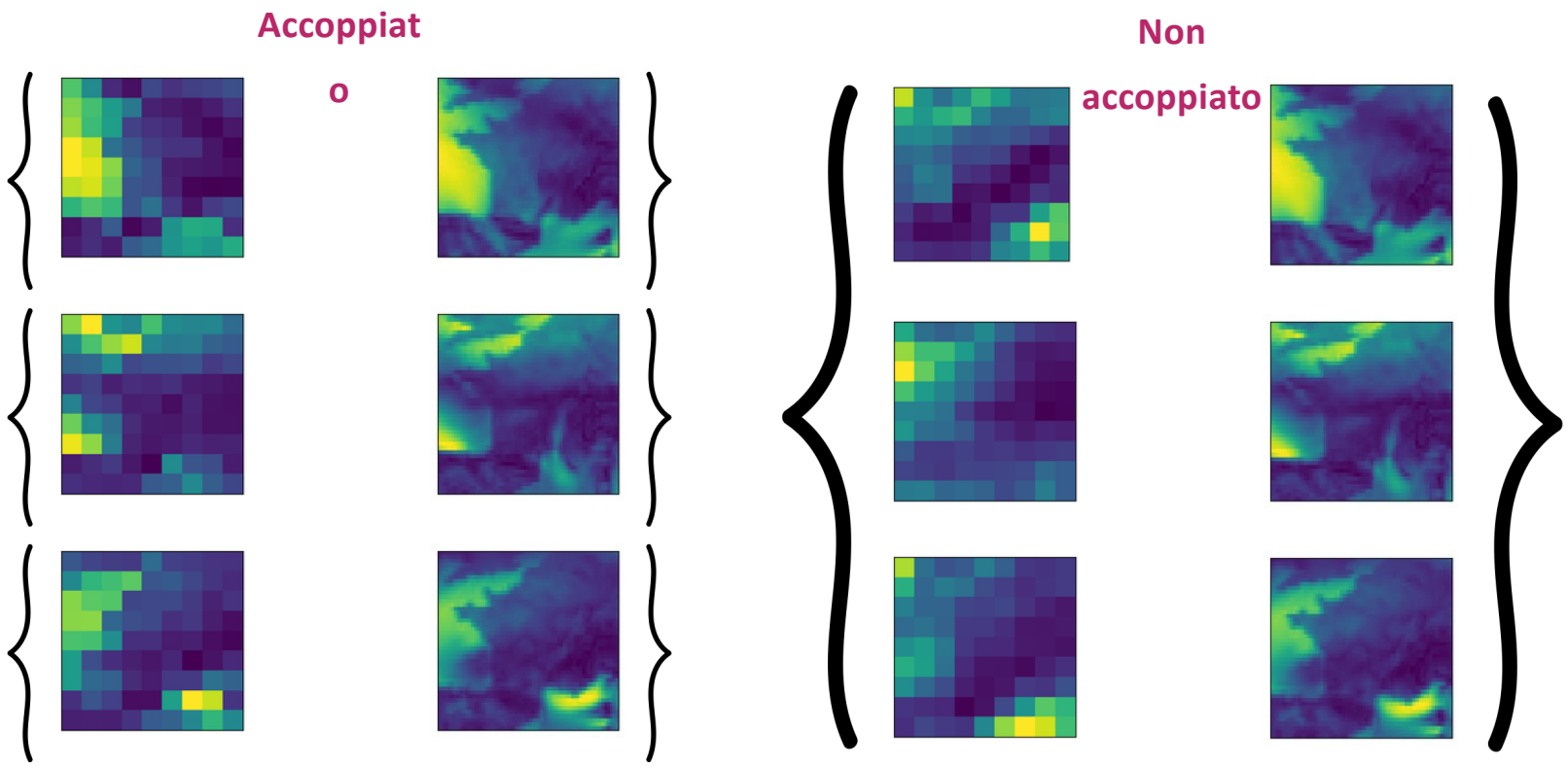


GCM ACCESS-ESM1 sulla Francia per 5 giorni



ERA5 sulla Francia per 5 giorni

Il downscaling come allineamento di domini non accoppiato



- GCM: **traiettorie fisicamente plausibili**
- Rianalisi: la sequenza **meteo storica** reale → **GCM e rianalisi non sono accoppiati**
- Downscaling = **allineamento di distribuzioni**: mappare gli output del GCM sulla distribuzione della rianalisi

Limiti dei metodi statistici tradizionali

Metodi basati sulle distribuzioni

- Quantile mapping, CDF-t (Michelangeli et al. 2009), correzione delle distribuzioni marginali

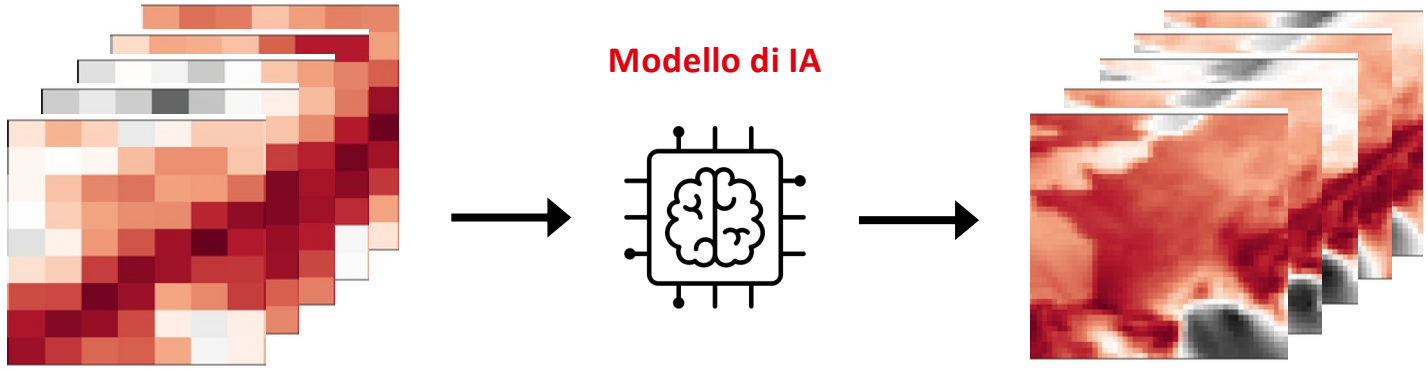
Metodi multivariati (trasporto ottimo)

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

Maledizione della dimensionalità: i metodi tradizionali non riescono a modellare tutte le interazioni in gioco

⟶ Metodi basati su IA / deep learning

L'IA generativa per il downscaling



Vantaggi

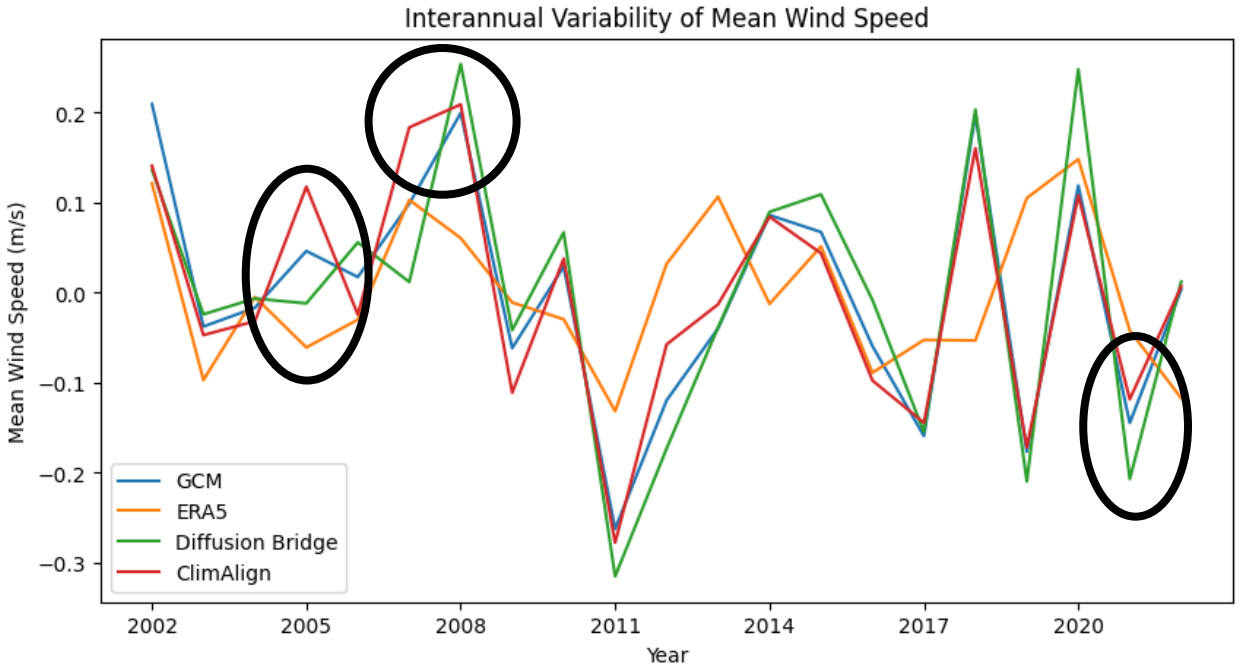
- Apprende la **distribuzione congiunta completa** dei campi ad alta risoluzione
- Genera **più realizzazioni**

Sfide

- Dati non accoppiati ⟶ **nessuna verità di riferiment**
- I metodi impliciti possono allontanarsi dal segnale d GCM

= proiezioni **poco affidabili**

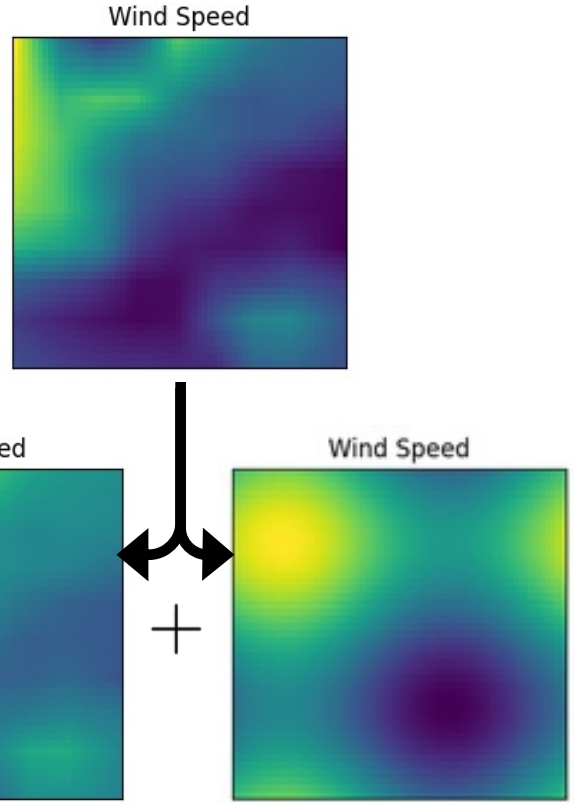
(ClimAlign/Groenke et al. 2020, ponti di diffusione/Bischoff & Deck 2024)



SerpentFlow: introdurre pseudo-coppie

SerpentFlow è un metodo generativo basato sull'IA per l'**allineamento di domini non accoppiato** tramite scomposizione in struttura condivisa

Velocità media del vento del GCM

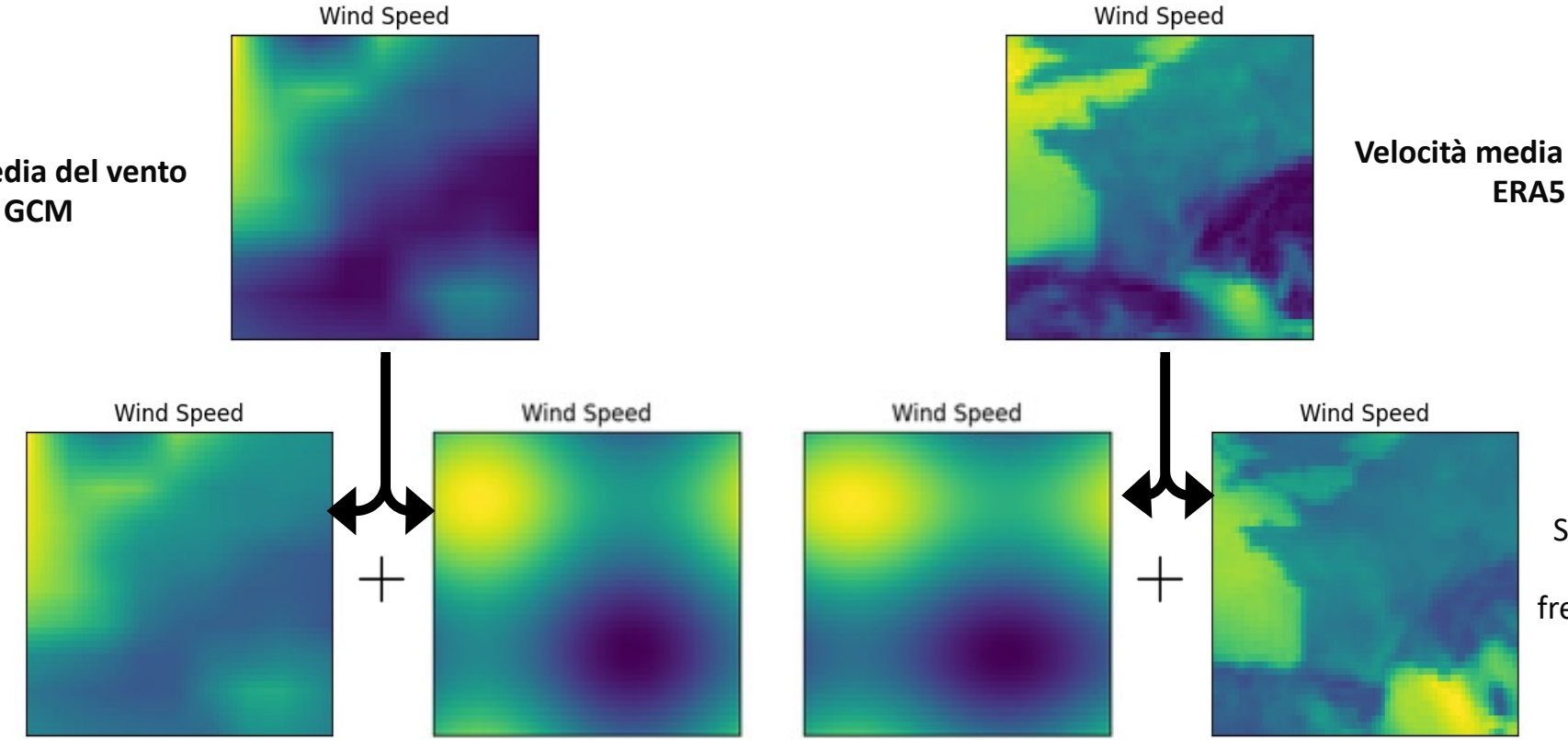


Separare le alte dalle basse frequenze spaziali

SerpentFlow: introdurre pseudo-coppie

SerpentFlow è un metodo generativo basato sull'IA per l'**allineamento di domini non accoppiato** tramite scomposizione in struttura condivisa

Velocità media del vento del GCM



Velocità media del vento ERA5

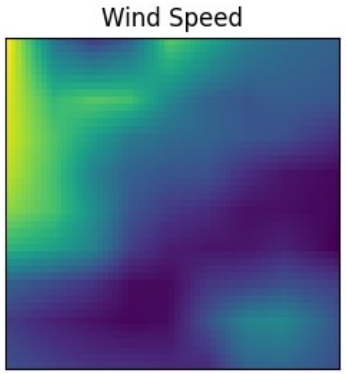
Separare le alte dalle basse frequenze spaziali

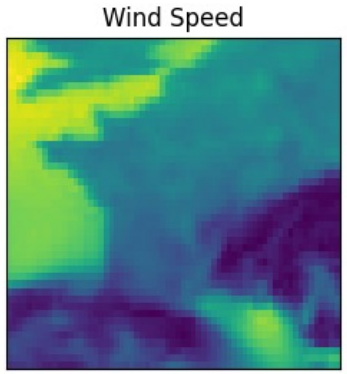
Separare le alte dalle basse frequenze spaziali

SerpentFlow: introdurre pseudo-coppie

SerpentFlow è un metodo generativo basato sull'IA per l'**allineamento di domini non accoppiato** tramite scomposizione in struttura condivisa

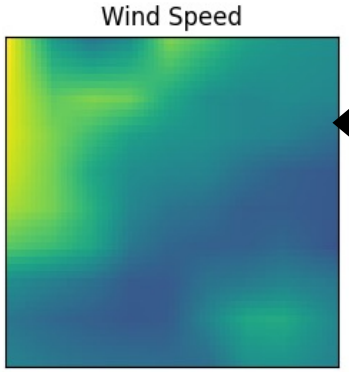
Velocità media del vento del GCM

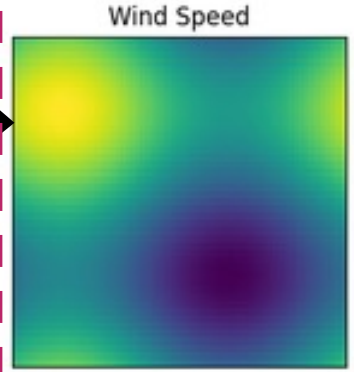


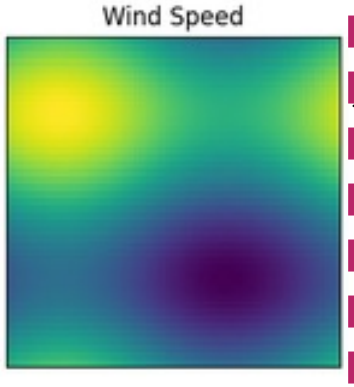


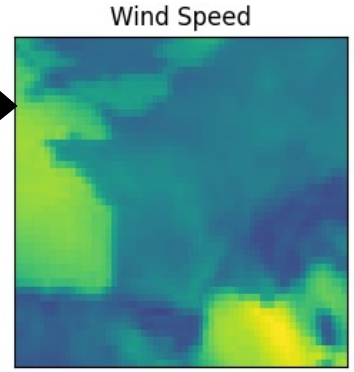
Velocità media del vento ERA5

Basse frequenze spaziali









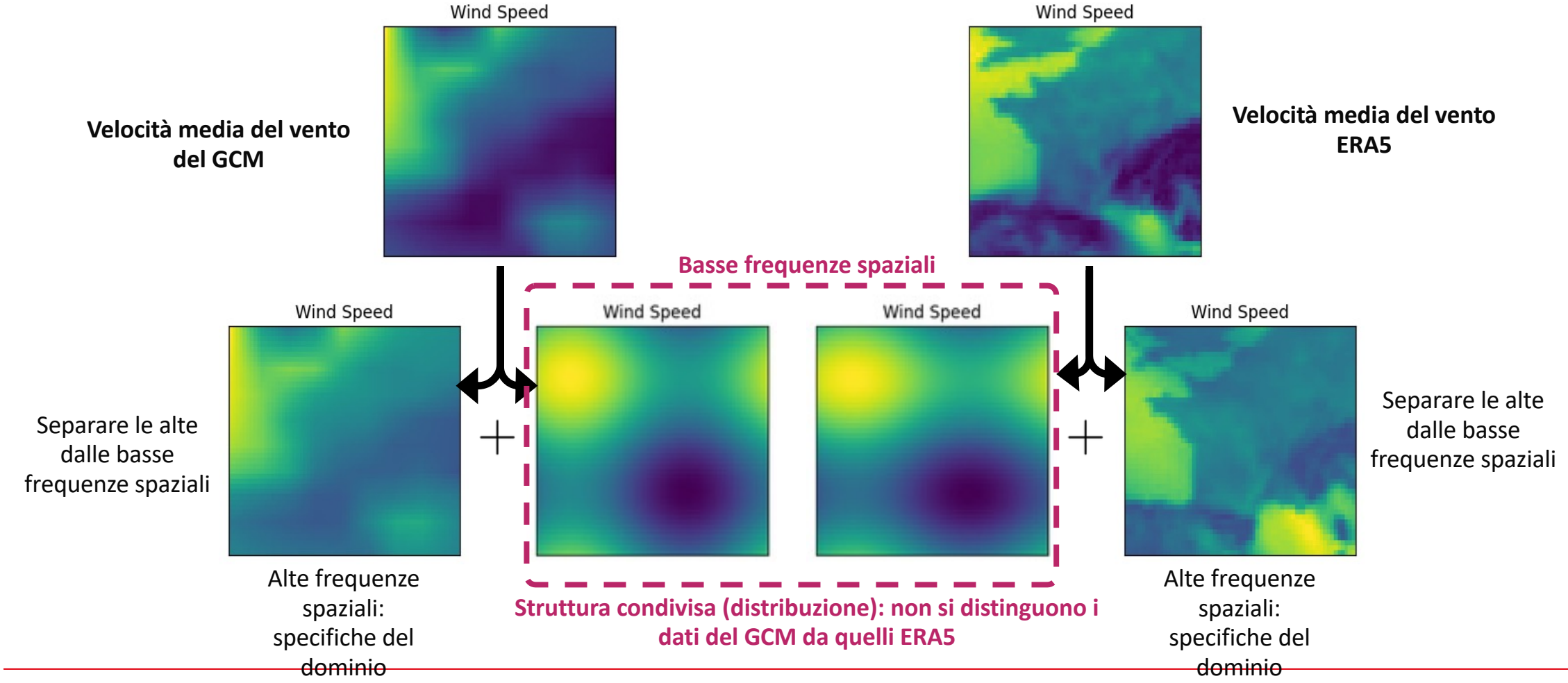
Separare le alte dalle basse frequenze spaziali

Separare le alte dalle basse frequenze spaziali

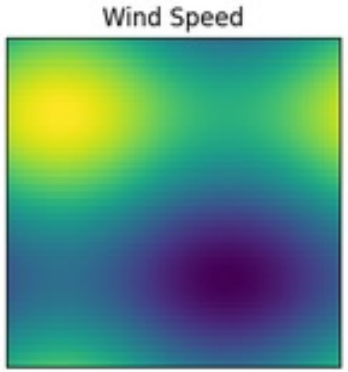
Struttura condivisa (distribuzione): non si distinguono i dati del GCM da quelli ERA5

SerpentFlow: introdurre pseudo-coppie

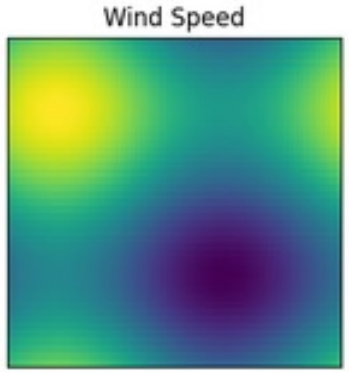
SerpentFlow è un metodo generativo basato sull'IA per l'**allineamento di domini non accoppiato** tramite scomposizione in struttura condivisa



Trovare automaticamente la frequenza di taglio



Filtro passa-basso(HR) con frequenza di taglio w



Filtro passa-basso(LR) con frequenza di taglio w

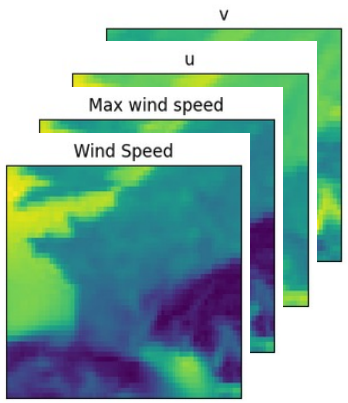
→ Classificatore →

Dati LR o HR?

- Smettere di ridurre la frequenza di taglio quando l'accuratezza <= 60%
- Frequenza trovata spesso ~= **risoluzione effettiva** del GCM

Pipeline di SerpentFlow

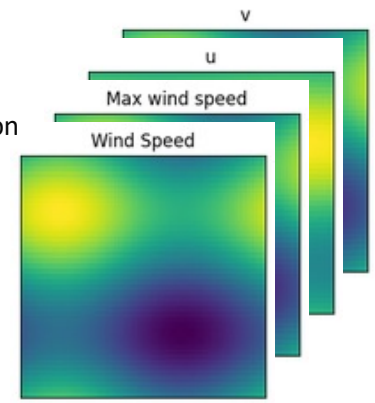
Addestramento



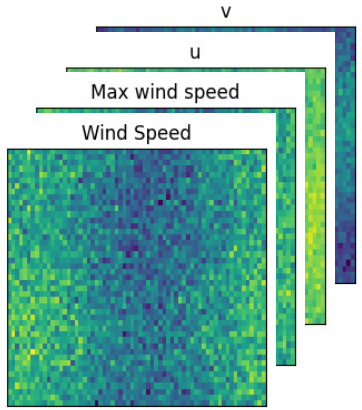
Riferimento (oss./rianalisi)

Filtro passa-basso, con frequenza di taglio

w_c

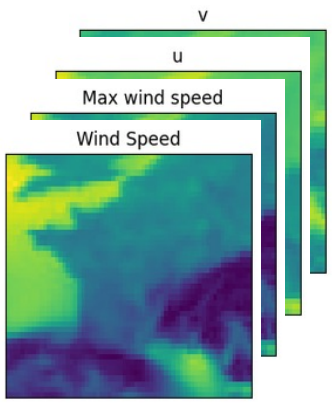


Aggiungere rumore alle alte frequenze



Addestrare l'approccio generativo

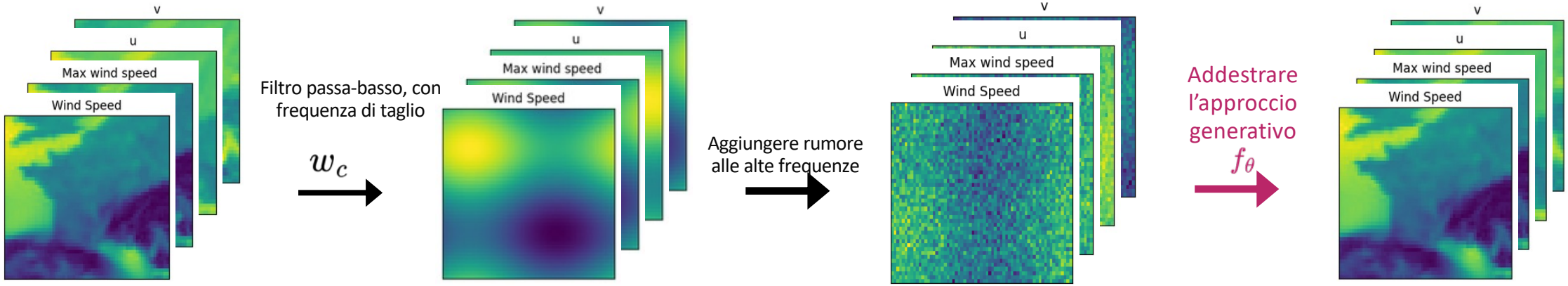
f_θ



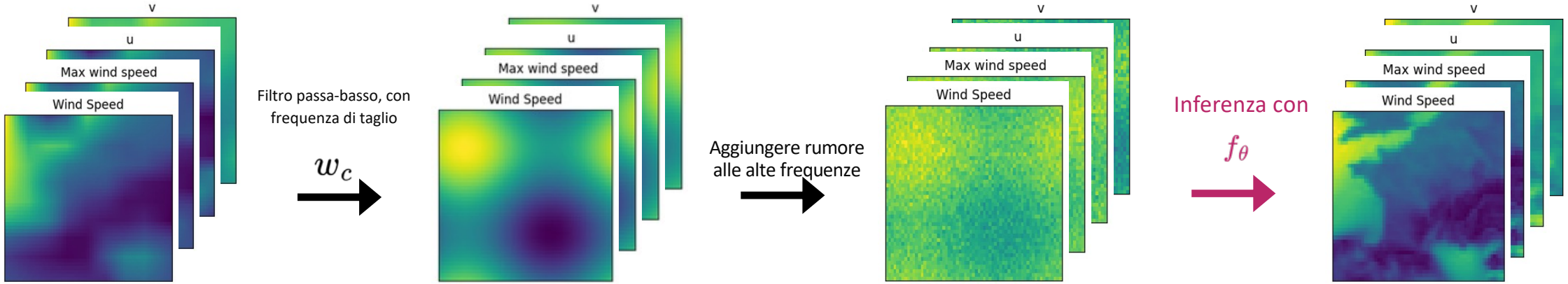
Riferimento (oss./rianalisi)

Pipeline di SerpentFlow

Addestramento

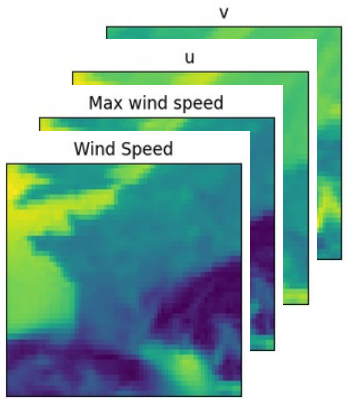


Inferenza



Pipeline di SerpentFlow

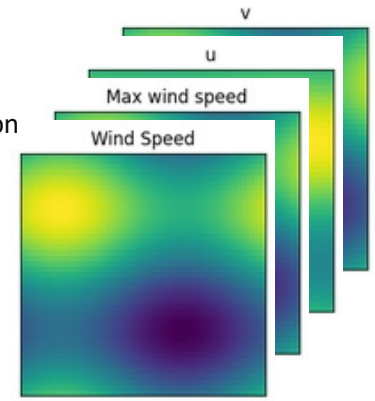
Addestramento



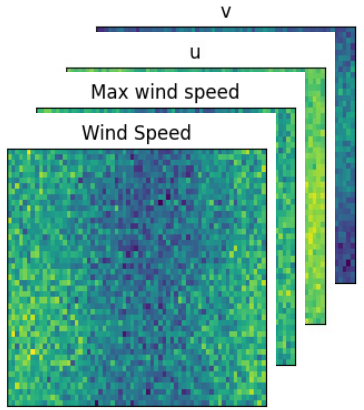
Riferimento (oss./rianalisi)

Filtro passa-basso, con frequenza di taglio w_c

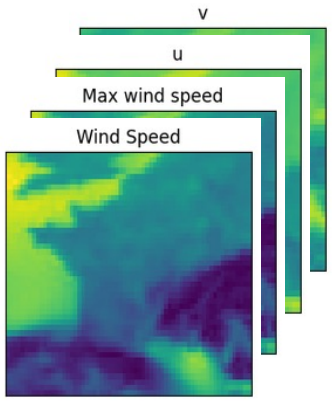
Distribuzione congiunta



Aggiungere rumore alle alte frequenze

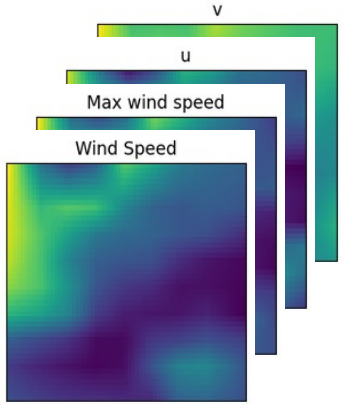


Addestrare l'approccio generativo f_θ



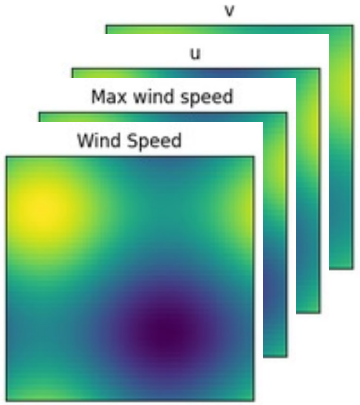
Riferimento (oss./rianalisi)

Inferenza

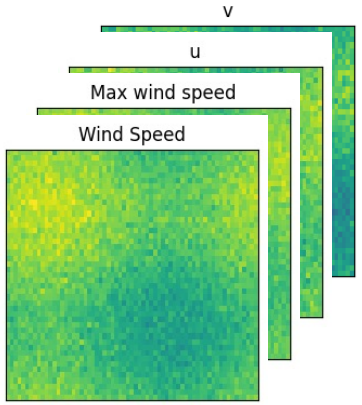


GCM interpolato

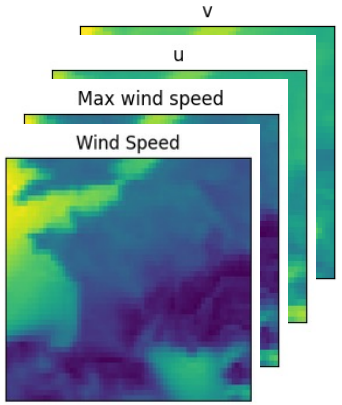
Filtro passa-basso, con frequenza di taglio w_c



Aggiungere rumore alle alte frequenze

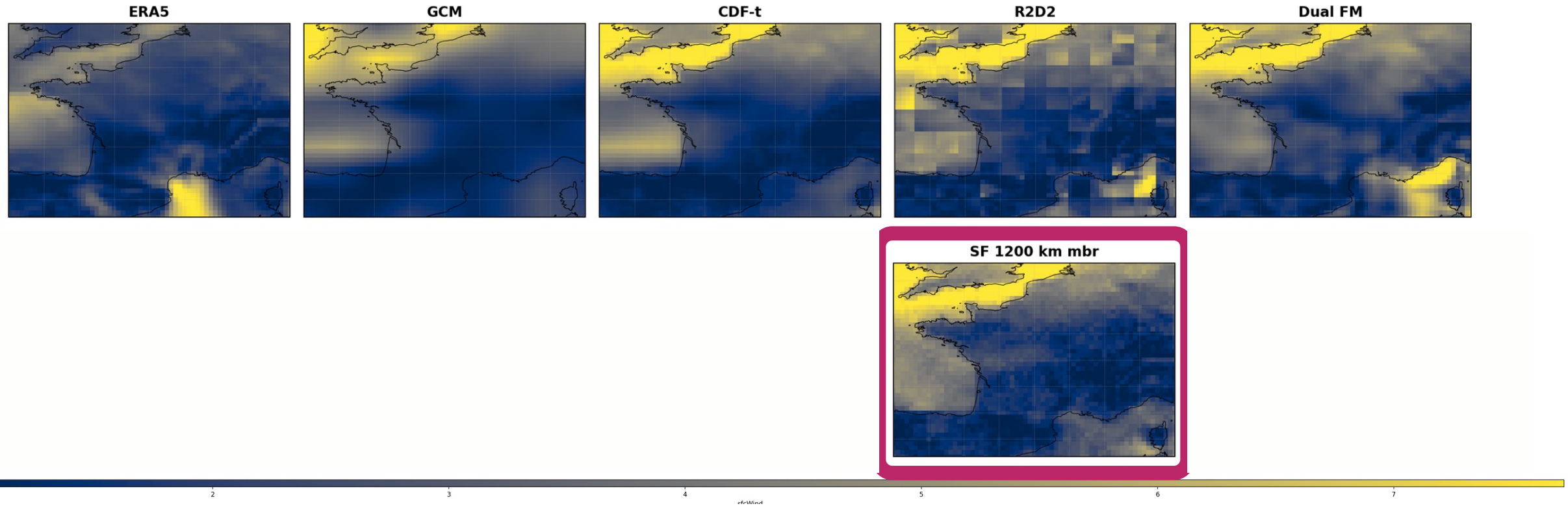


Inferenza con f_θ



GCM dopo downscaling

Abbiamo approssimato la distribuzione di ERA5?

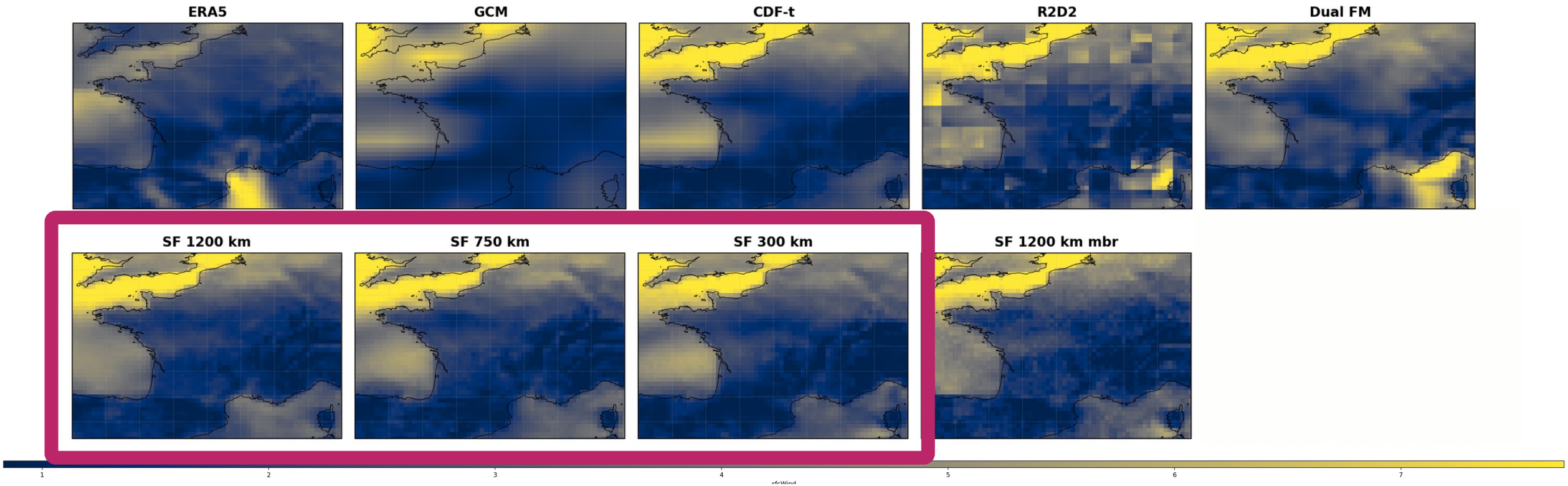


Risultati chiave

- I metodi DL superano CDF-t e R2D2 sulla struttura spaziale.
- **SF 1200 km** (1 membro): **migliore approssimazione** della distribuzione di ERA5

.

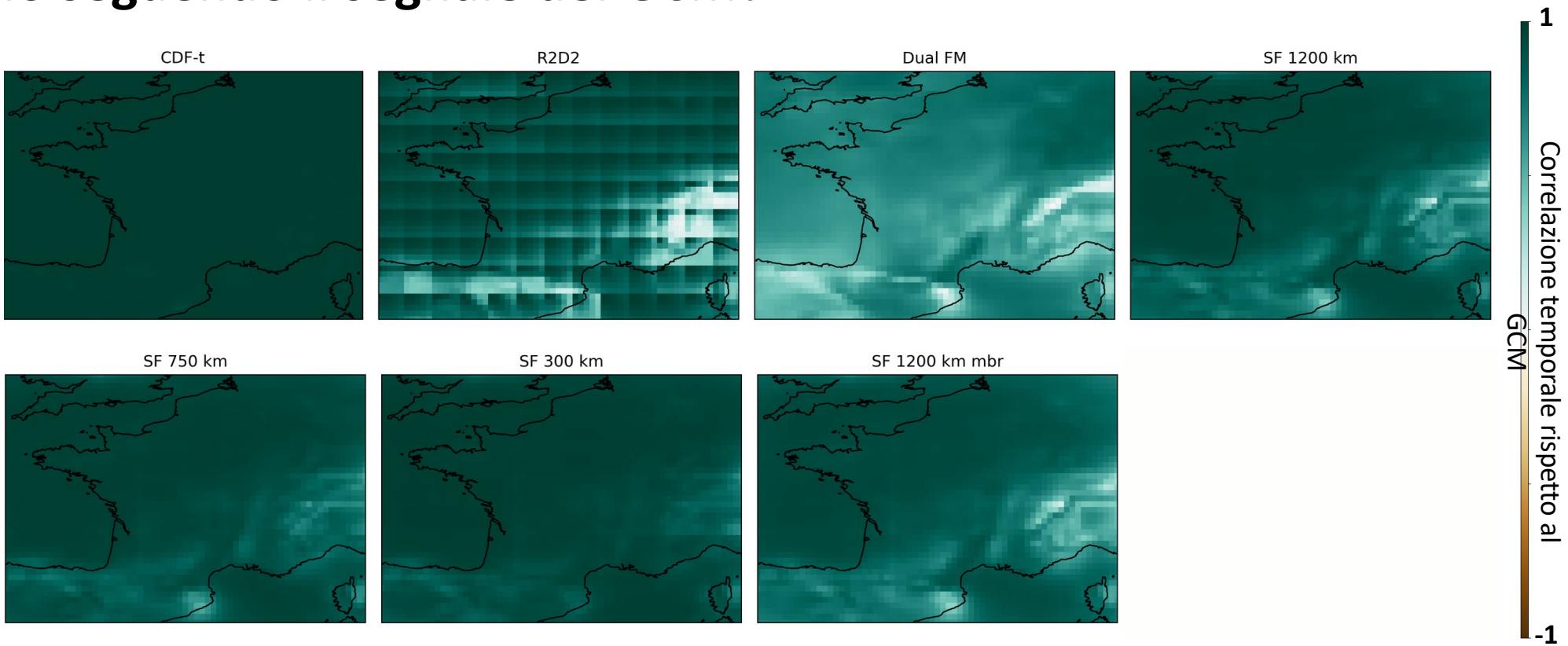
Abbiamo approssimato la distribuzione di ERA5?



Risultati chiave

- I metodi DL superano CDF-t e R2D2 sulla struttura spaziale.
- **SF 1200 km** (1 membro): **migliore approssimazione** della distribuzione di ERA5
- **frequenza di taglio più alta** ⟶ deriva rispetto alla distribuzione di ERA5.

Stiamo seguendo il segnale del GCM?



Risultati chiave

- **CDF-t è il più vicino alla** dinamica del GCM; Dual FM non segue la traiettoria del GCM
- La deriva di SF compare nelle regioni **montuose**
- **Frequenza di taglio più bassa** ⟶ più vicino alla dinamica del GCM

Prospettive

- Funziona con diversi **GCM, risoluzioni e domini irregolari** — ERA5 (25 km) e SAFRAN (8 km)
- Supera i metodi statistici nel modellare la **distribuzione congiunta** dei dati di rianalisi
- Rispetto ad altri approcci generativi, resta vicino alla dinamica del GCM
- **La frequenza di taglio è interpretabile**
  - Collega direttamente la separazione delle scale alla risoluzione effettiva del GCM
  - Gestisce **il compromesso** tra seguire la traiettoria del GCM e riprodurre la distribuzione della rianalisi
- **Limiti:** solo vento vicino alla superficie — prossimi passi: altre variabili, estremi, downscaling temporale

IA generativa per:

- ✓ Previsione probabilistica
  - **ArchesWeatherGen**: previsioni meteo efficienti addestrate a 1,5°
  - **ArchesClimate**: emulazione, interpolazione ed estrapolazione del clima decennale

- ✓ Fusione dei dati con allineamento probabilistico di domini
  - **MotifGen**: un modello generativo, multi-fonte e auto-supervisionato per apprendere da dati geospaziali

- ✓ Allineamento di domini probabilistico e downscaling
  - **ClimAlign**: downscaling generativo non accoppiato con allineamento di domini
  - **SerpentFlow**: allineamento di domini generativo non accoppiato tramite scomposizione in struttura condivisa (con applicazione al downscaling)

Primi passi con l’IA generativa basata su flussi

Flow matching

Tutorial:

- 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

Letture di approfondimento sui metodi basati su flussi

Interpolanti stocastici

Quadro generale basato su flussi per trasportare il rumore nei dati

- Flow matching e rectified flow ne sono esempi
- Generalizza il quadro: si può addestrare con un'EDO e campionare con un'EDS, o viceversa

Operatori neurali

I metodi basati su flussi operano nello spazio degli operatori, consentendo una generalizzazione indipendente dalla scala

Modelli di consistenza

Versione a passo singolo del flow matching. Area di ricerca aperta, con un po' di teoria e limiti pratici.

Apprendimento con catena di Markov inversa

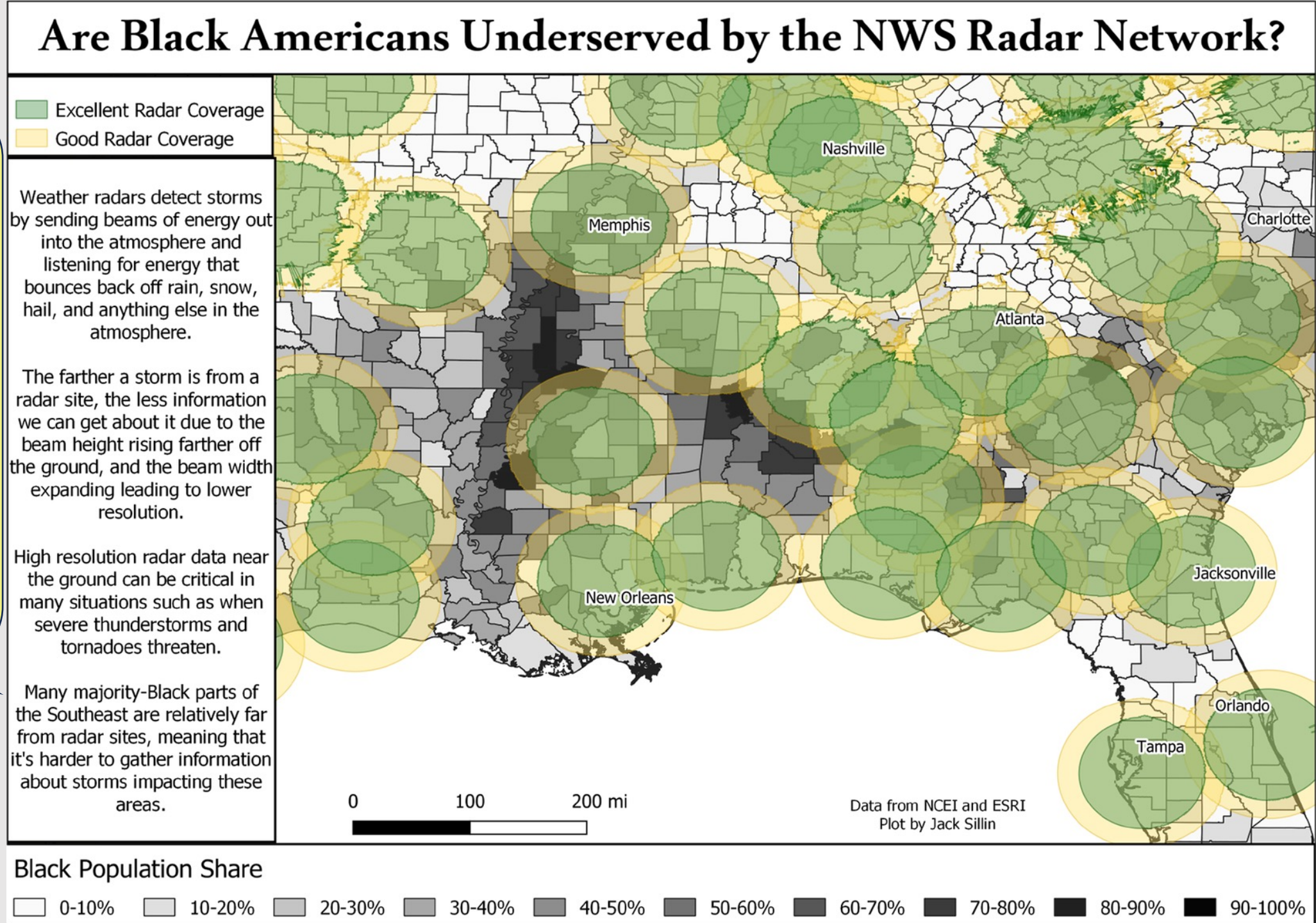
In sostanza, flow matching dalla bassa all'alta risoluzione

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

Sezione bonus: l’IA per l’equità dei dati climatici

«Molte aree a maggioranza nera del Sud-Est [degli Stati Uniti] sono relativamente lontane dai radar, il che rende più difficile raccogliere informazioni sulle tempeste che le colpiscono.»

Credito: Jack Sillin, in [McGovern et al., Environmental Data Science, 2022]



L’IA per l’equità dei dati climatici

- Addestrare i modelli in regioni ricche di dati e applicarli in regioni povere di dati
  - Addestrarli e validarli in regioni ricche di dati
  - Affinarli con i dati limitati delle regioni povere di dati e usarli per generare altri dati.

- Contributo all’equità dei dati climatici
  - Scale locali (ad es. eredità dell’ingiustizia ambientale negli Stati Uniti)
  - Scale globali:
    - Il Nord del mondo ha storicamente emesso più carbonio; e lì di solito ci sono più dati
    - Il Sud del mondo subisce gli effetti più gravi del riscaldamento che ne deriva