Quickstart!

ChaosBench:

A Multi-Channel, Physics-Based Benchmark for Subseasonal-to-Seasonal Climate Prediction

Juan Nathaniel (*jn2808@columbia.edu*), Yongquan Qu, Tung Nguyen, Sungduk Yu, Julius Busecke, Aditya Grover, Pierre Gentine

What is Chaos?

Given a cup of coffee, you can see **different steam patterns** every time

Factors

Wind speed	Wind direction	Room temperature
Liquid density (milk?)	Cup geometry	...

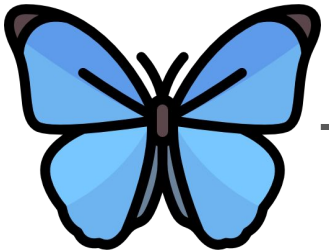
Chaos in real life

coffee steam

navier-stokes

Chaos (/ˈkāˌäs/): *stochastic patterns* of a **deterministic dynamics** due to **nonlinearity** and **initial condition sensitivities**

$F^N[x_0 + \varepsilon]$



butterfly flap in New York

hurricanes in Vancouver

Chaos in real life

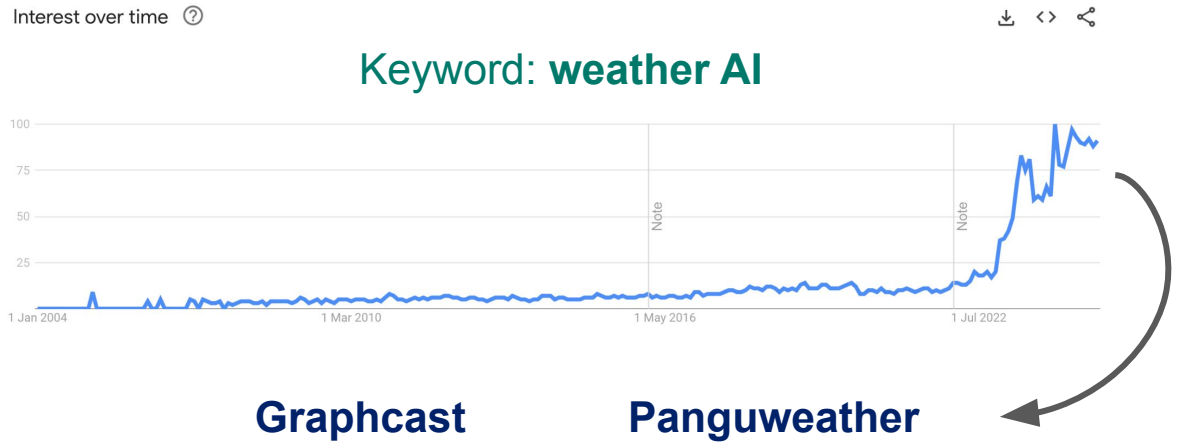
Chaotic systems are everywhere…

- Financial market → stock performance
- Virus spread → COVID-19 outbreaks
- **Subseasonal climate (2-6 weeks ahead)** → hurricanes, drought, flood, …

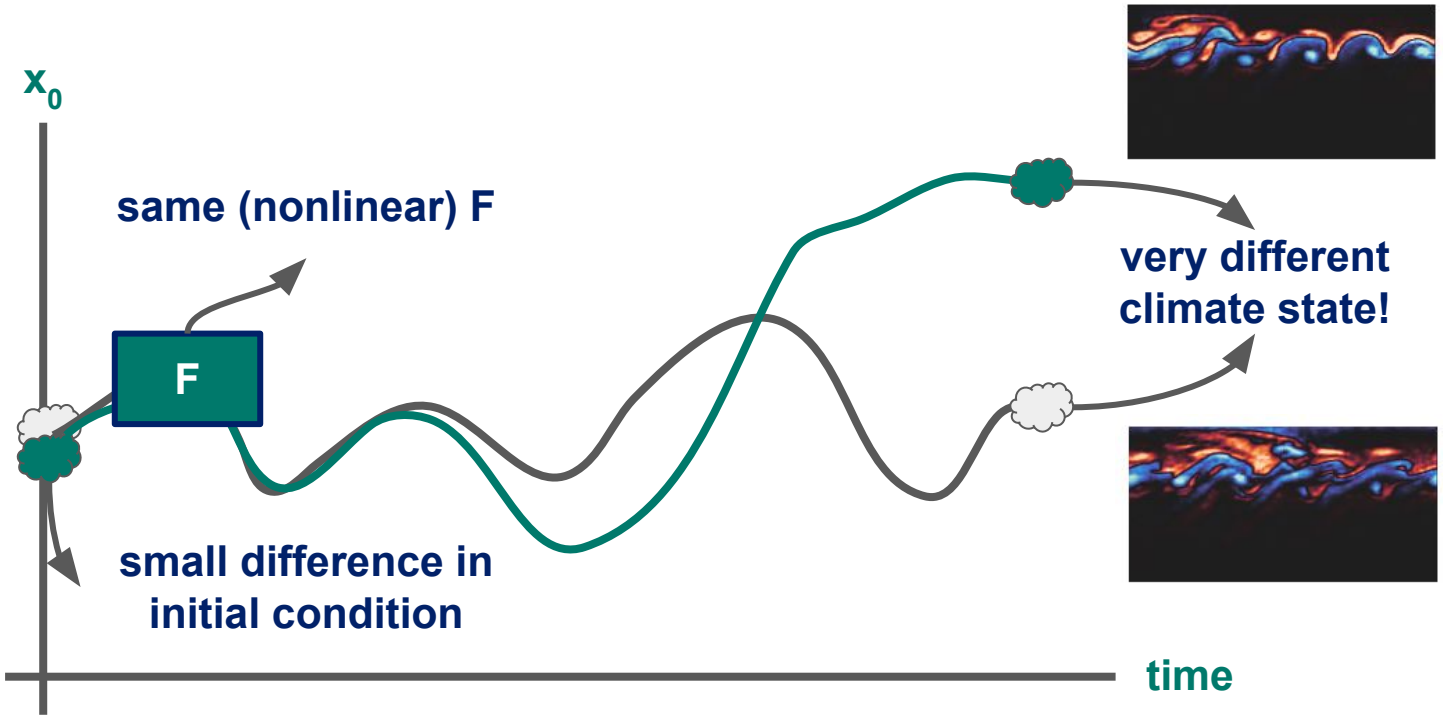
Q: Can we predict (and solve) Chaos?

The rise of weather ML

Weather (/ ˈwɛð ər /): short-term **(up to 2 weeks)** atmospheric states



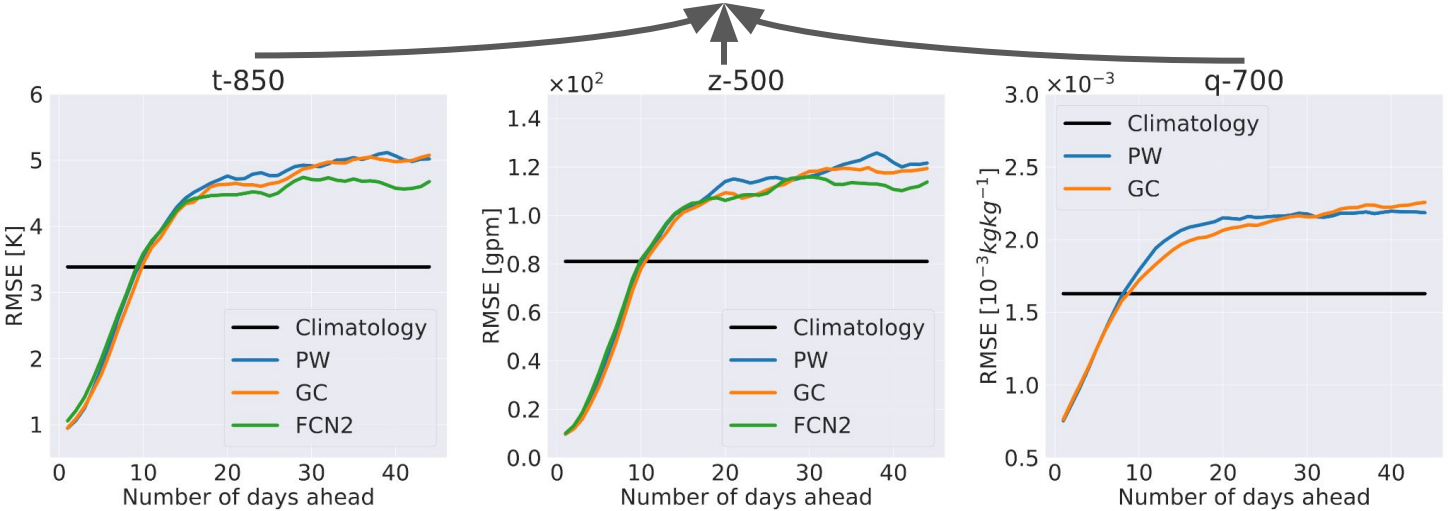
Chaos in planetary circulation



The collapse of deterministic weather ML

Variables at different vertical atmosphere level

Metrics



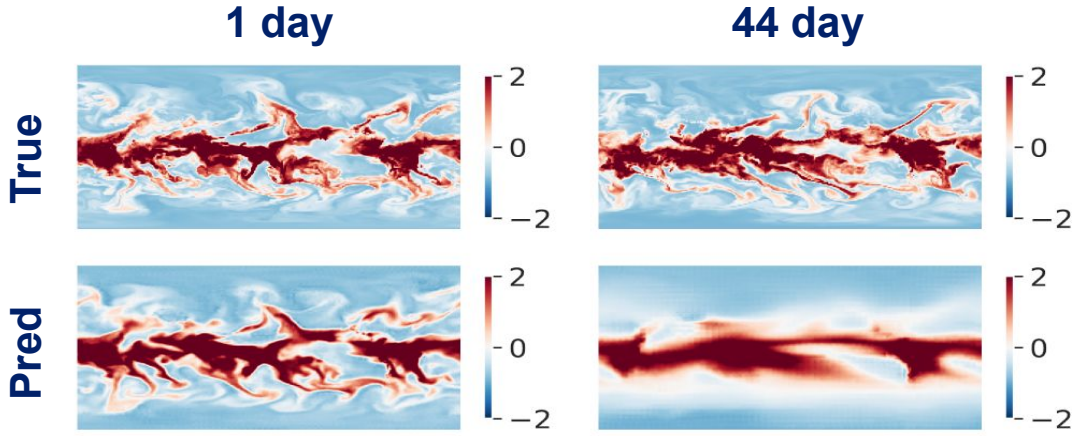
Climatology: long-term average (unskilled baseline)

PW: Panguweather
GC: Graphcast
FCN2: FourCastNetV2

~10 days predictability range for SoTA Weather ML

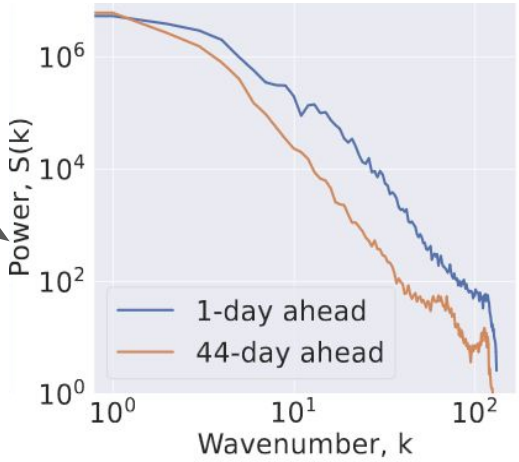
NSF LEAP

The collapse of deterministic weather ML



Blurrier as **forecast time** increases

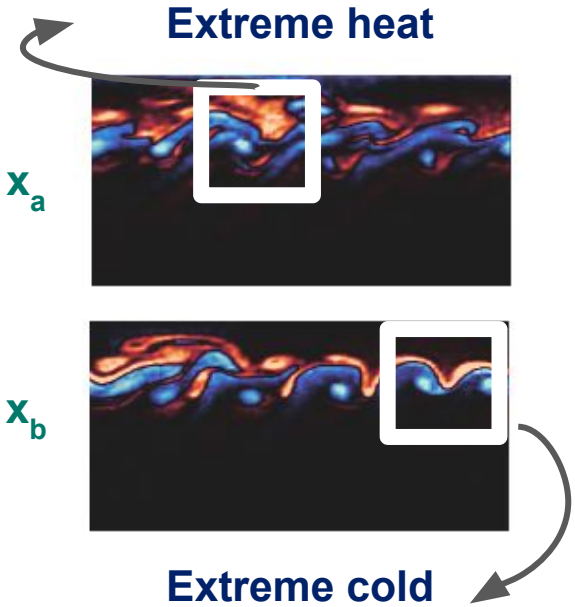
Failure to capture **high-frequency** details



*Analysis done with ViT/ClimaX: specific humidity (q) at pressure level 850 hPa;
** Insights are consistent across the board (models & variables)

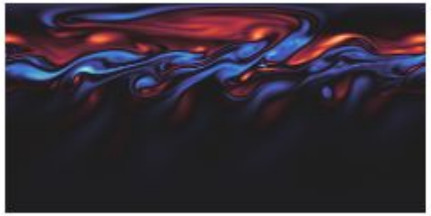
LEAP

The collapse of deterministic weather ML



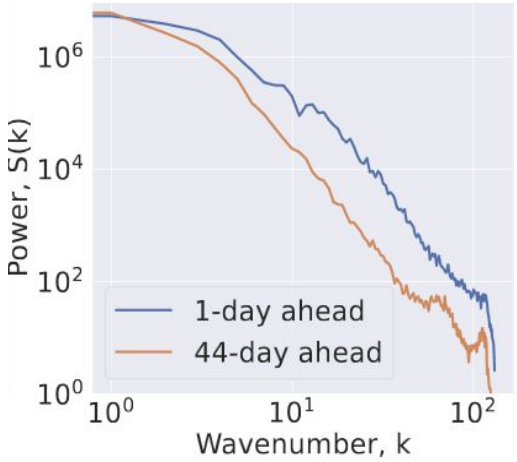
double penalty

distance loss (L1, L2)



Smoothed forecasts (no extremes, etc)

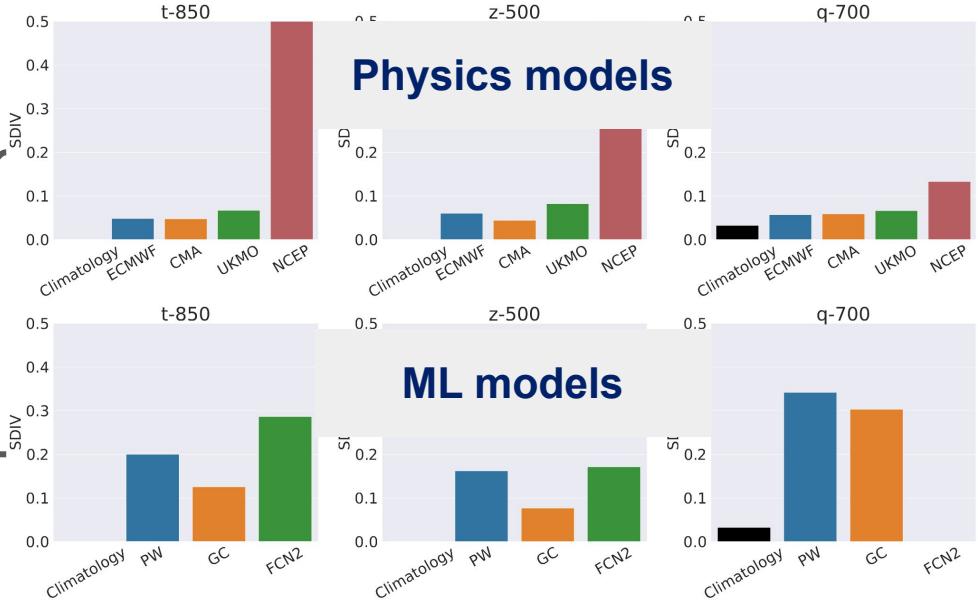
Closing the gap: physics-based metrics



Compute Spectra Discrepancy:

1. Kullback-Leibler Divergence
2. L1 Loss

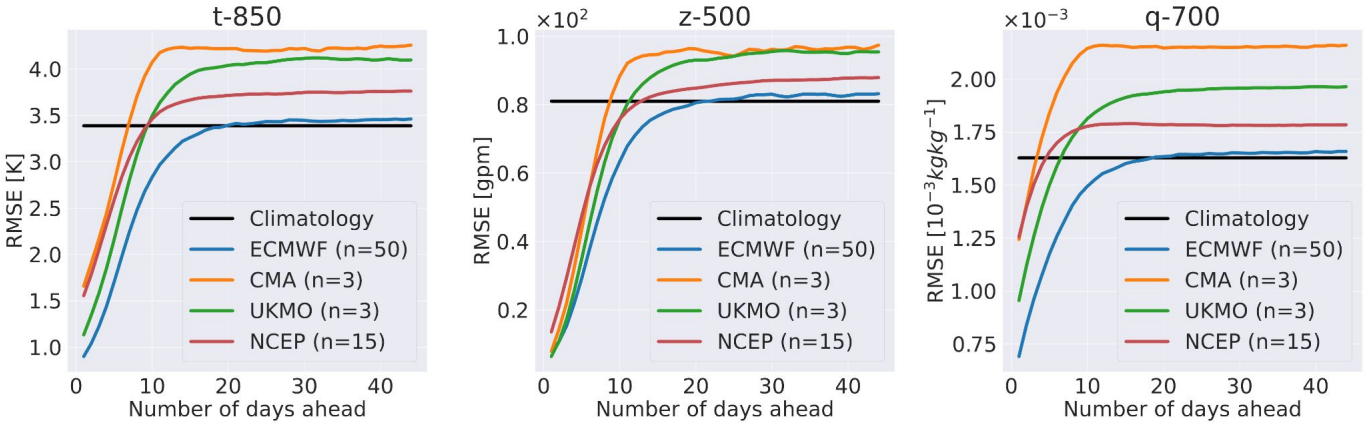
close the gap!



ML models have **higher spectra discrepancy** than **physics models**

What's next: consulting domain science

Physics models appear to be **better**, but still **plateauing early…**

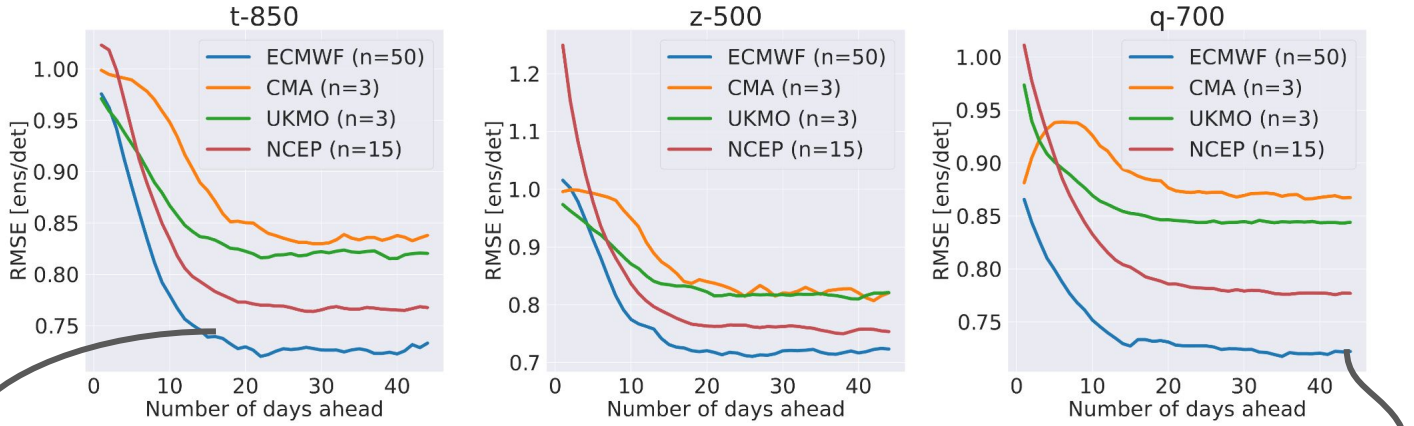


ECMWF: European Centre for Medium-Range Weather Forecasts
CMA: China Meteorological Administration
UKMO: UK Meteorology Office
NCEP: National Centers for Environmental Prediction

**(n = number of ensemble members)

What's next: ensemble scaling

Increasing the number of well-spread ensemble member (n) improves skillfulness



Improvement gain as number of members increases
Beneficial especially for long-range forecasts

ECMWF: European Centre for Medium-Range Weather Forecasts
CMA: China Meteorological Administration
UKMO: UK Meteorology Office
NCEP: National Centers for Environmental Prediction

**(n = number of ensemble members)

What's next: hybrid modeling

Hybrid physics-informed ML model* shows promise for long-range modeling**

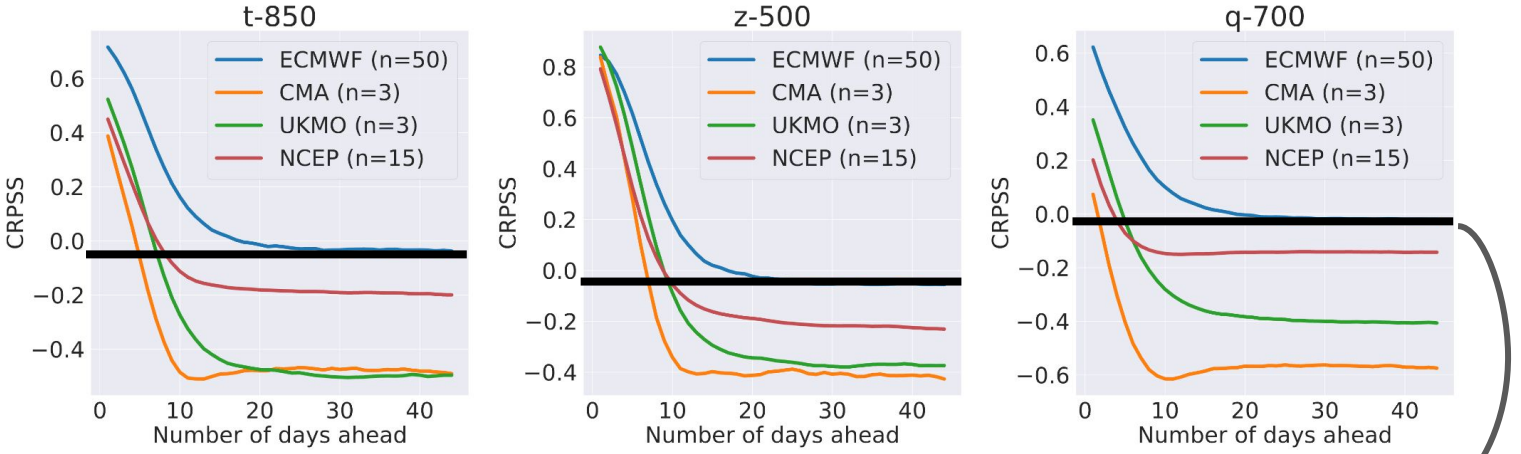
Models	RMSE ↓			MS-SSIM ↑			SpecDiv ↓		
	T850 (K)	Z500 (gpm)	Q700 ($\times 10^{-3}$)	T850	Z500	Q700	T850	Z500	Q700
Lagged AE	5.55	122.4	2.03	0.74	0.71	0.47	0.18	2.44	0.21
ResNet	5.67	125.3	2.07	0.73	0.70	0.47	0.21	0.37	0.26
UNet	5.47	121.5	2.13	0.73	0.71	0.45	0.30	1.16	2.20
FNO	**5.06**	**112.5**	**1.95**	**0.75**	**0.73**	**0.51**	**0.18**	**0.11**	**0.10**

FNO preserves some spectral physical information

*all experiments are performed with identical number of trainable parameters, hyperparameters
**results at final timestep T = 44 days ahead

On the limit of current predictability

Measuring Skillfulness: CRPSS* → 0 (Unskilled)



Even the best Physics model has 15-20 days limit on predictability

Challenge: Can we extend the predictability range with ML?

*CRPSS: Continuous Ranked Probability Skill Score

The path forward: ML + Physics synthesis

Predictability can be extended (more details in the paper):

- Well-spread ensemble → w/ Probabilistic ML
- Physics-based ML
- Robust control of error propagation

3-easy step Quickstart!

https://leap-stc.github.io/ChaosBench/