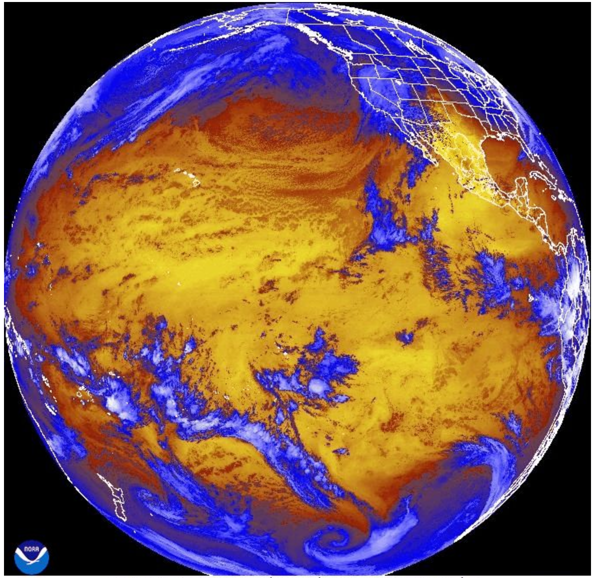
Google DeepMind



05 Mar 2025 20:10Z - NOAA/NESDIS/STAR - FULL DISK - Band 16

Data over Dogma

Ruthless Empiricism, Strange Ideas, and the Future of Weather Forecasting

Daniel Worrall
Research Scientist

3rd ERCOFTAC
"Machine learning for fluid dynamics" workshop

The Team


Aaron Bell


Peter Battaglia


Akib uddin


Aliyah Bond


Alvaro Sanchez


Andrew El-Kadi


Ben Gaiarin


Boris Babenko


Daniel Worrall


David Landry


Devaja Shah


Dominic Masters


Elinor Kruse


Emily Morris


Ferran Alet


Fred Zyda


Guy Shalev

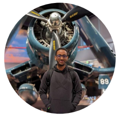

Ilan Price


Marc Deisenroth


Marcus Trail


Matt Willson


Megan Bela


Mohammed Alewi Hassen


Natalie Williams


Nofar Peled Levi

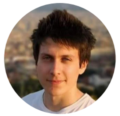

Rémi Lam


Samier Merchant


Shreya Agrawal


Stephan Rasp


Stephen Adams


Stratis Markou

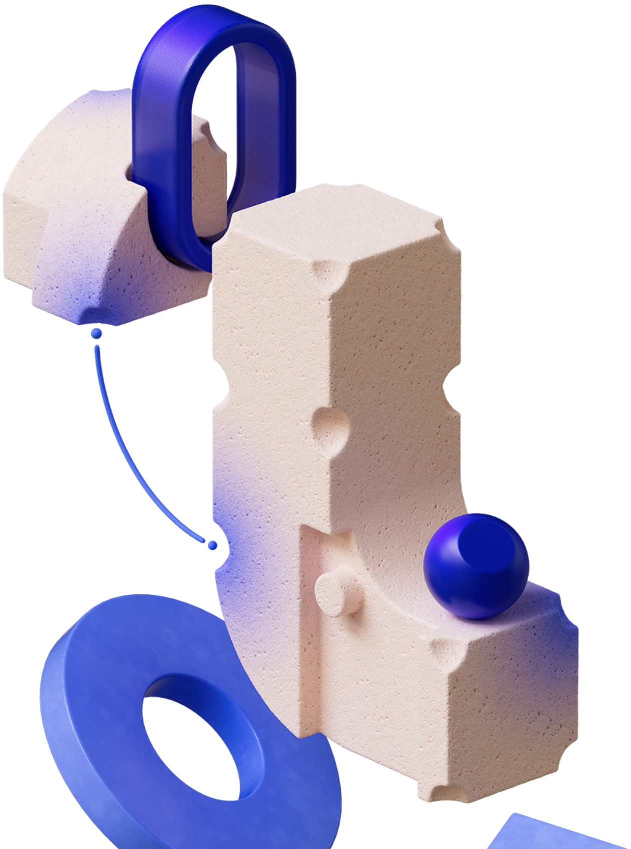

Suhani Vora


Sunny Mak


Thomas Turnbull


Tom Andersson

Google DeepMind



(Un)sustainability

(Un)sustainability

FIGURE C

Global risks ranked by severity over the short and long term

"Please estimate the likely impact (severity) of the following risks over a 2-year and 10-year period."

Risk categories

- Economic
- Environmental
- Geopolitical
- Societal
- Technological

	2 years		10 years
1st	Misinformation and disinformation	1st	Extreme weather events
2nd	Extreme weather events	2nd	Biodiversity loss and ecosystem collapse
3rd	State-based armed conflict	3rd	Critical change to Earth systems
4th	Societal polarization	4th	Natural resource shortages
5th	Cyber espionage and warfare	5th	Misinformation and disinformation
6th	Pollution	6th	Adverse outcomes of AI technologies
7th	Inequality	7th	Inequality
8th	Involuntary migration or displacement	8th	Societal polarization
9th	Geoeconomic confrontation	9th	Cyber espionage and warfare
10th	Erosion of human rights and/or civic freedoms	10th	Pollution

from World Economic Forum - The Global Risks Report 2025 (sourced from WEF: Global Risks Perception Survey 2025)

(Un)sustainability

FIGURE C | **Global risks ranked by severity over the short and long term**

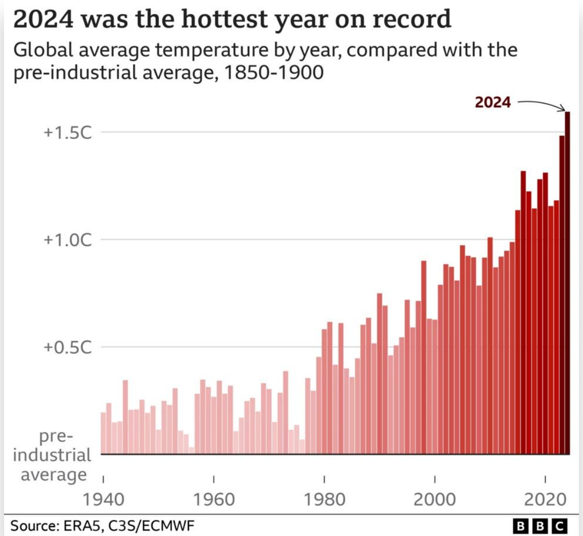
"Please estimate the likely impact (severity) of the following risks over a 2-year and 10-year period."

Risk categories
- Economic
- Environmental
- Geopolitical
- Societal
- Technological

	2 years	10 years
1st	Misinformation and disinformation	Extreme weather events
2nd	Extreme weather events	Biodiversity loss and ecosystem collapse
3rd	State-based armed conflict	Critical change to Earth systems
4th	Societal polarization	Natural resource shortages
5th	Cyber espionage and warfare	Misinformation and disinformation
6th	Pollution	Adverse outcomes of AI technologies
7th	Inequality	Inequality
8th	Involuntary migration or displacement	Societal polarization
9th	Geoeconomic confrontation	Cyber espionage and warfare
10th	Erosion of human rights and/or civic freedoms	Pollution

from [World Economic Forum - The Global Risks Report 2025] (sourced from [WEF: Global Risks Perception Survey 2025])

(Un)sustainability



Trump signs executive order directing US withdrawal from the Paris climate agreement — again

President Donald Trump has signed an executive order directing the United States to again withdraw the United States from the landmark Paris climate agreement

January 20, 2025

News

Climate Change Is Intensifying And Affecting Every Region On Earth: IPCC

Backlash Erupts Over Europe's Anti-Deforestation Law

Leaders around the world are asking the European Union to delay rules that would require companies to police their global supply chains.

Sept. 19, 2024

AI's hunger for electric power is threatening U.S. climate goals

Electric utility companies are building more power plants that will burn natural gas to meet demands of a data center construction boom.

Updated November 21, 2024

About Weather

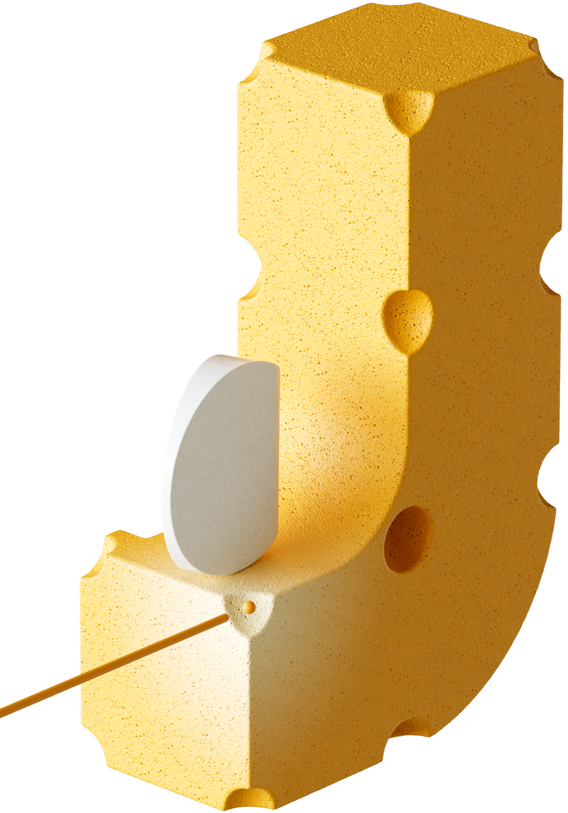
Weather touches all our lives, every day

- Should I carry an umbrella?
- Should I flee an approaching hurricane?
- How should I price wind power on an energy market?
- What should I plant to be resilient to drought?

Google Search: **Significant % of daily users** check the weather

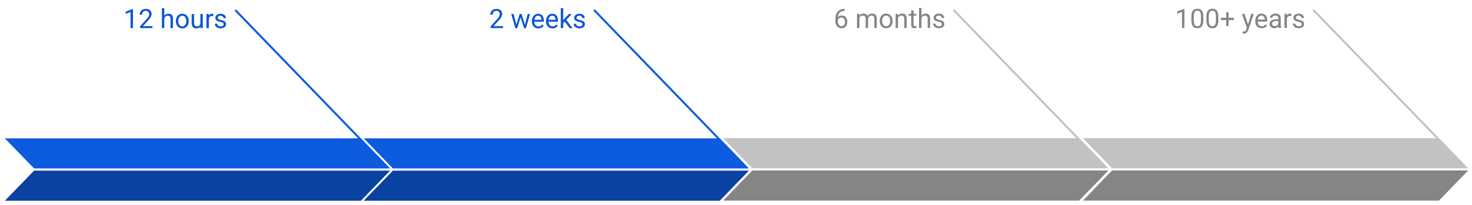
World Bank estimates

- **Global: $160 Billion** - Current annual benefits from weather and climate prediction
    - **Potential: $5 Billion** - Additional annual benefits from better models



The Weather Problem & AI Disruption

Weather forecasting context



Nowcasting

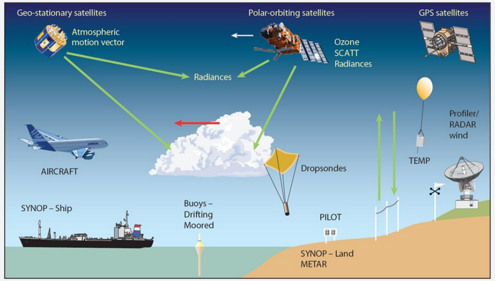
Medium-range

Subseasonal-to-seasonal

Climate

Weather forecasting context

Observations



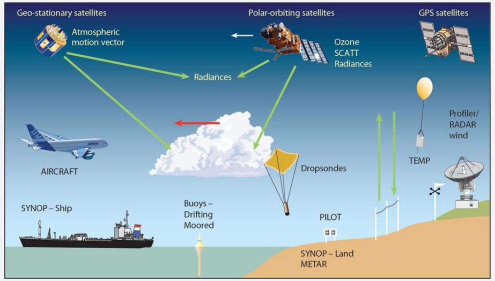
800 million observations assimilated daily

~20 satellites from 5 agencies

Near-real time exchange of observations across political boundaries

Weather forecasting context

Observations

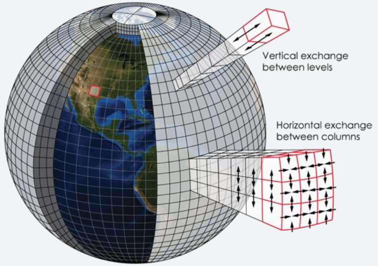


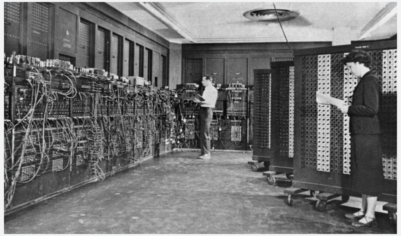
800 million observations assimilated daily

~20 satellites from 5 agencies

Near-real time exchange of observations across political boundaries

Model





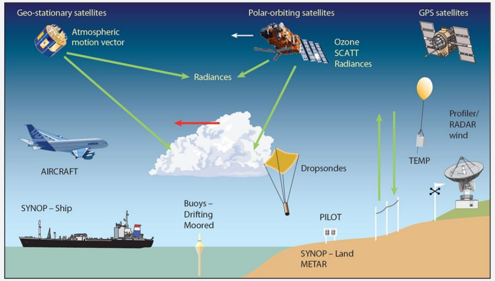
ENIAC supercomputer ~1950

10 x 10 km resolution

>100 vertical levels ≈ 500M grid points

Weather forecasting context

Observations

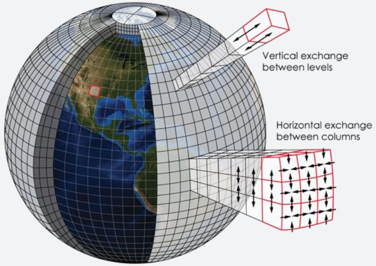


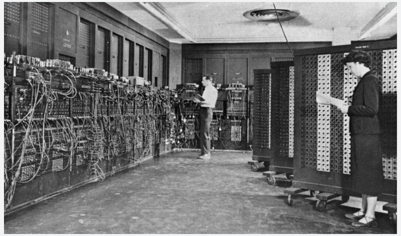
800 million observations assimilated daily

~20 satellites from 5 agencies

Near-real time exchange of observations across political boundaries

Model

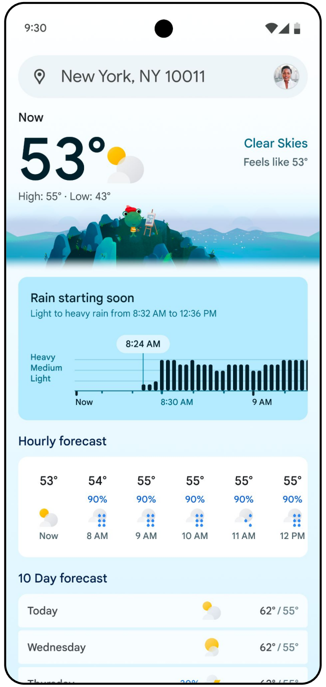




ENIAC supercomputer ~1950

10 x 10 km resolution

>100 vertical levels ≈ 500M grid points



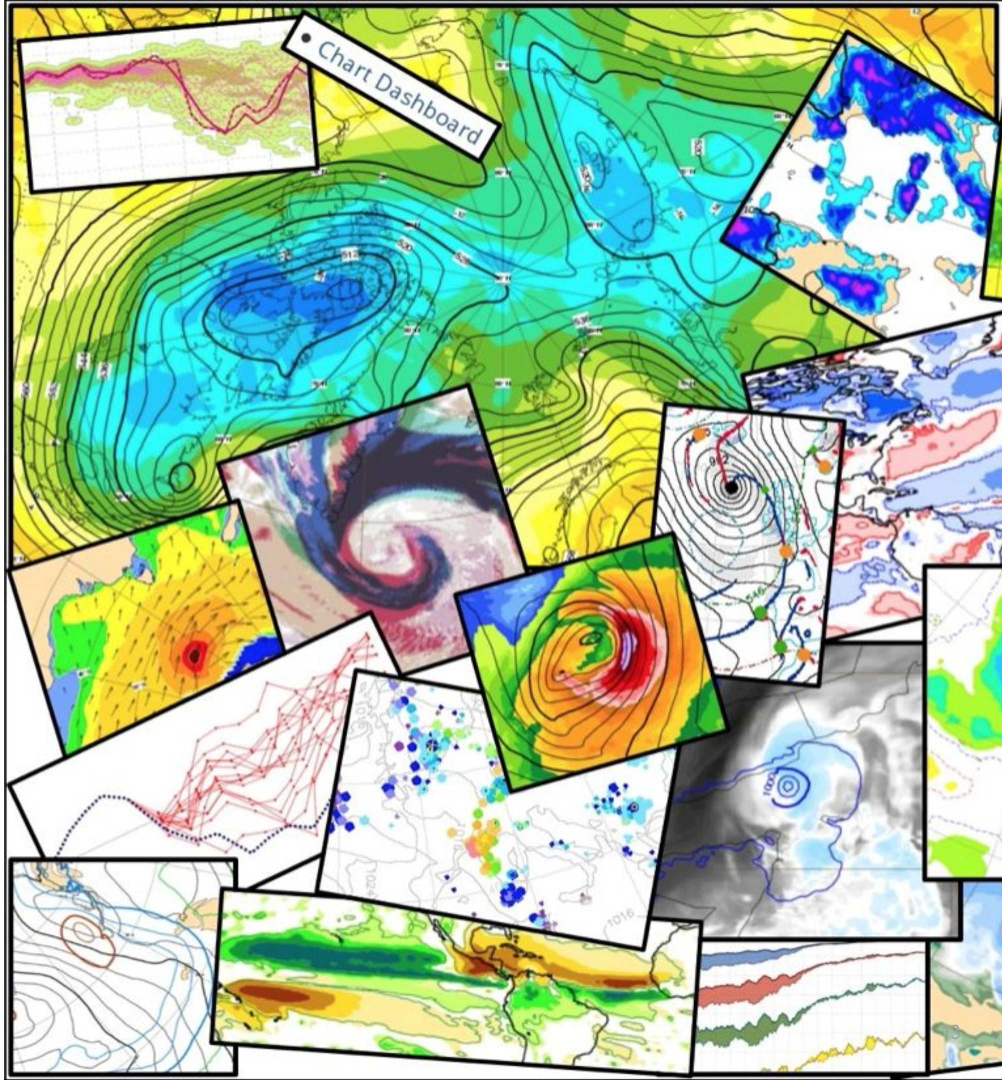
Physics-based model

ECMWF Integrated Forecast System (IFS)

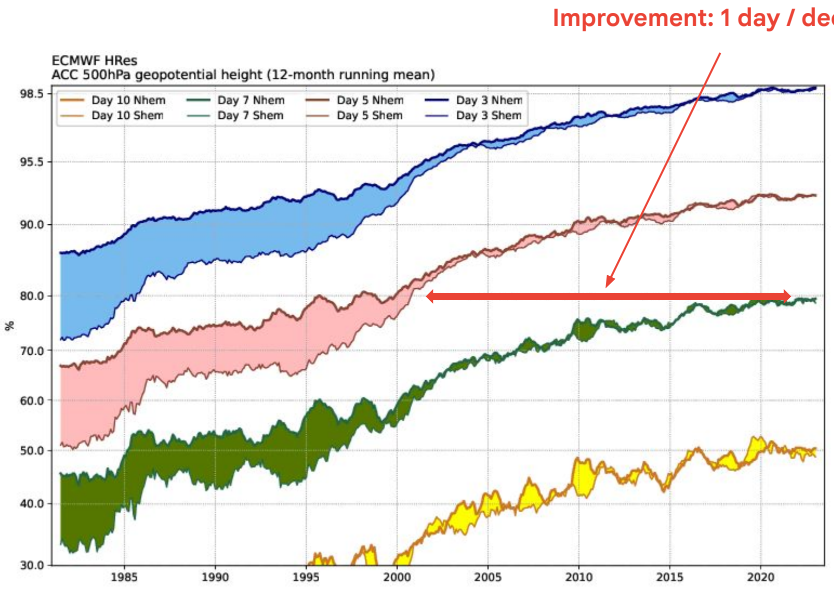
- Global, hydrostatic, primitive-equation model
- Horizontal velocity, temperature, specific humidity, log surface pressure

- Hydrostatic Equation
- Conservation of Mass, Momentum, Energy, & Water Species
- The Gas Law

- 9 km, 200 million grid points, 7.5 min timesteps
- 10^6 lines of code
- 1 forecast = 6-7 hours compute



Progress over the decades, and recent years



Source: European Center for Medium-Range Weather Forecasting (ECMWF)
ACC = Anomaly Correlation Coefficient
Nhem = Northern Hemisphere, Shem = Southern Hemisphere

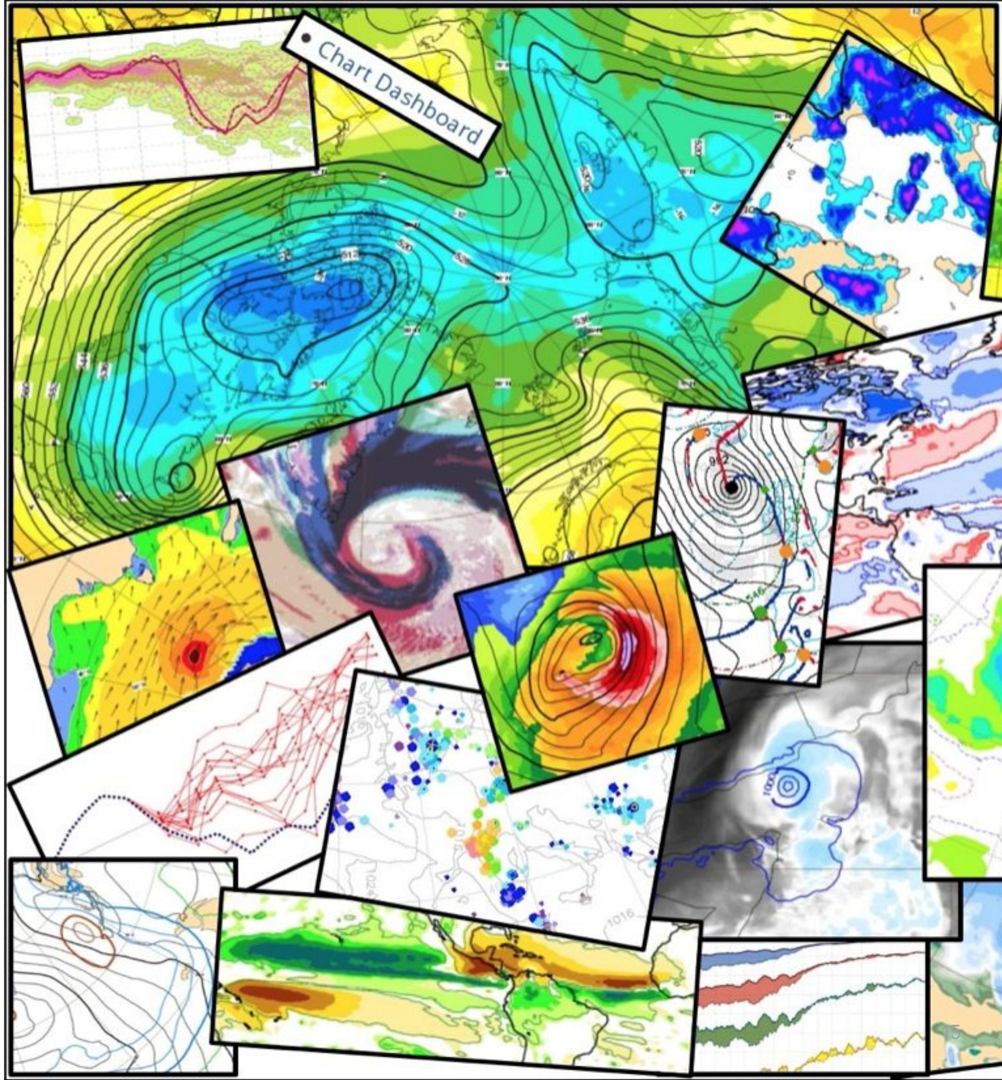
Physics-based model

ECMWF Integrated Forecast System (IFS)

- Global, hydrostatic, primitive-equation model
- Horizontal velocity, temperature, specific humidity, log surface pressure

- Hydrostatic Equation
- Conservation of Mass, Momentum, Energy, & Water Species
- The Gas Law

- 9 km, 200 million grid points, 7.5 min timesteps
- 10^6 lines of code
- 1 forecast = 6-7 hours compute



Physics-based model

ECMWF Integrated Forecast System (IFS)

Best representation?

- Global, hydrostatic, primitive-equation model
- Horizontal velocity, temperature, specific humidity, log surface pressure

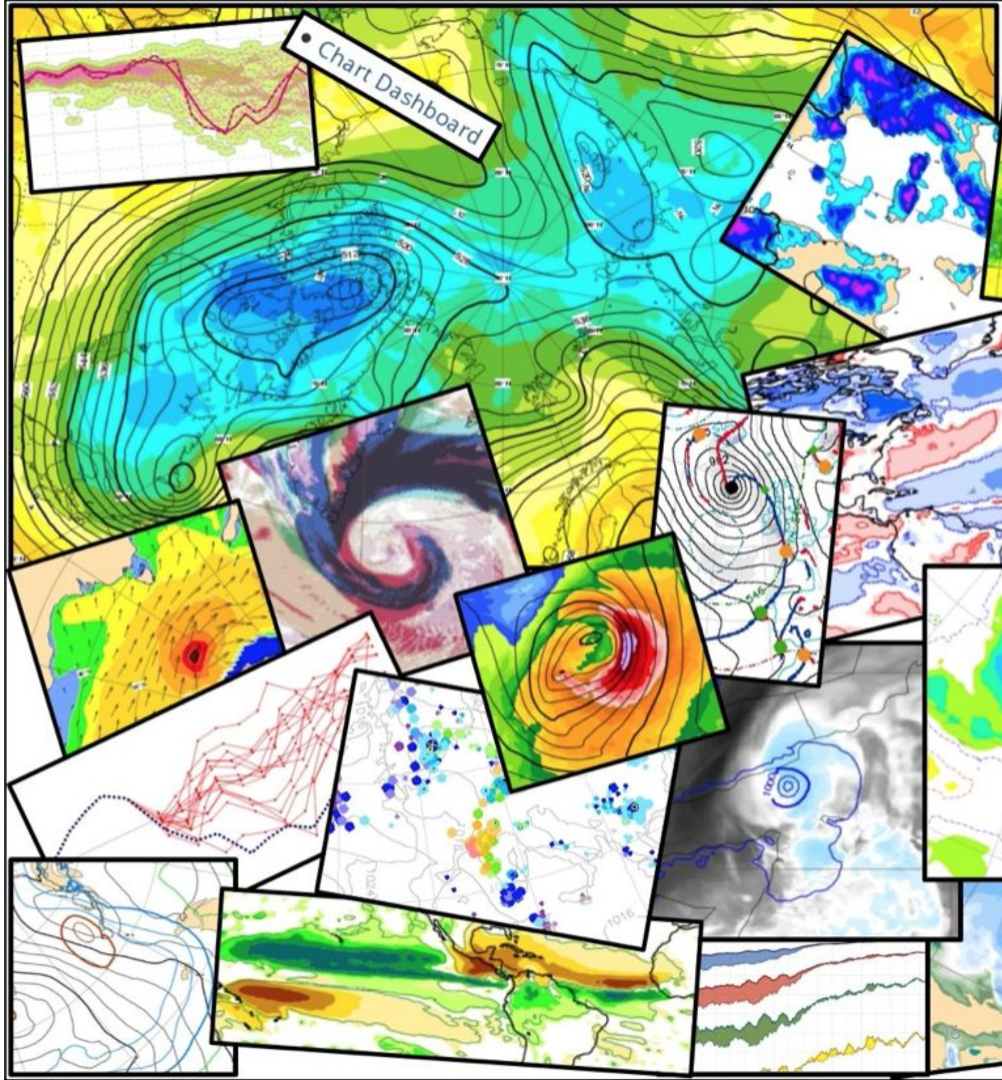
Are these right?

- Hydrostatic Equation
- Conservation of Mass, Momentum, Energy, & Water Species
- The Gas Law

What about clouds?

- 9 km, 200 million grid points, 7.5 min timesteps
- 10^6 lines of code
- 1 forecast = 6-7 hours compute

Why so high?



ML-based models

ECMWF Integrated Forecast System (IFS)

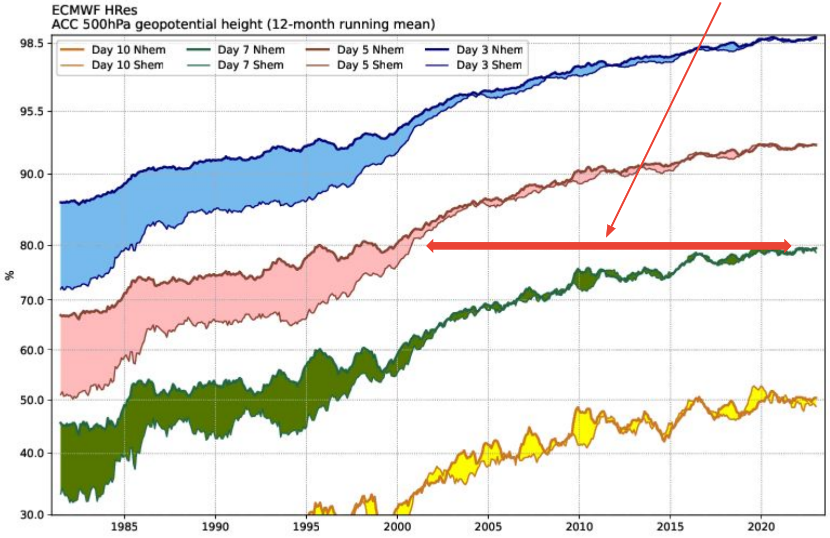
- Horizontal velocity, temperature, specific humidity, log surface pressure

- No equations!

- 28 km, 1 million grid points, 6h timesteps
- 10^5 lines of code
- 1 forecast = ~ 1min

Progress over the decades, and recent years

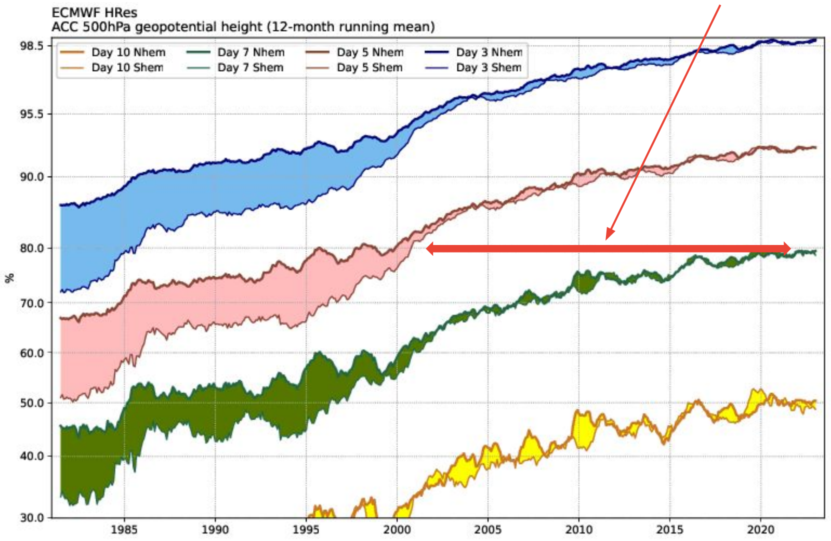
Improvement: 1 day / decade



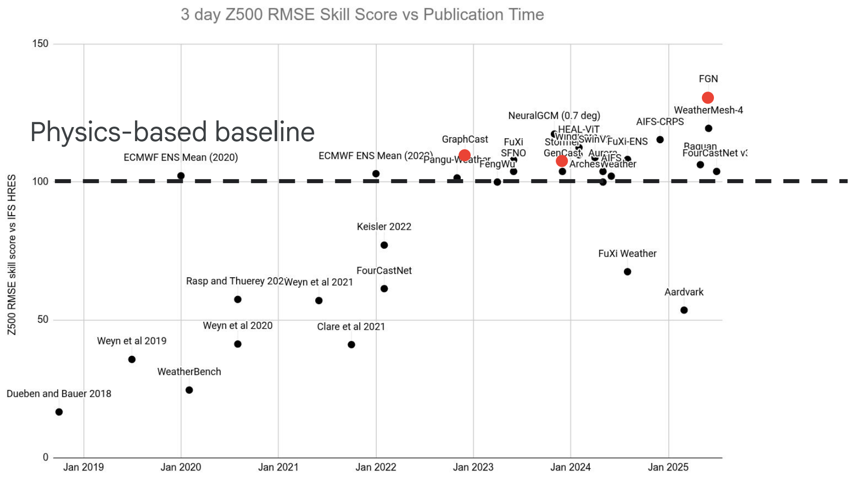
Source: European Center for Medium-Range Weather Forecasting (ECMWF)
ACC = Anomaly Correlation Coefficient
Nhem = Northern Hemisphere, Shem = Southern Hemisphere

Progress over the decades, and recent years

Improvement: 1 day / decade



WeatherBench 2



Improvement: 1 day / 1 year

Source: European Center for Medium-Range Weather Forecasting (ECMWF)
ACC = Anomaly Correlation Coefficient
Nhem = Northern Hemisphere, Shem = Southern Hemisphere

Google’s role

Precipitation Nowcasting

Regional
2 h
Probabilistic
(GAN)

Sep 2021

MetNet 3

Regional
Precipitation
12hr
Launched on
Search!

Nov 2023

WeatherNext

Launching our
real time
forecasts to cloud
customers

Jan 2025

Dec 2022 / Nov 2023

GraphCast

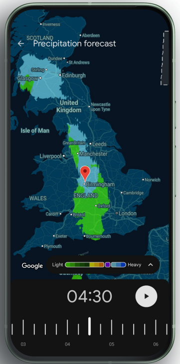
Global
10-day
Deterministic
GNN

Dec 2023 / Dec 2024

GenCast

Global
15-day
Probabilistic
(diffusion)
GNN +
Transformer

Dec 2024



Global Nowcasting

Global
Precipitation
12hr
Launched on Search and
Pixel Weather

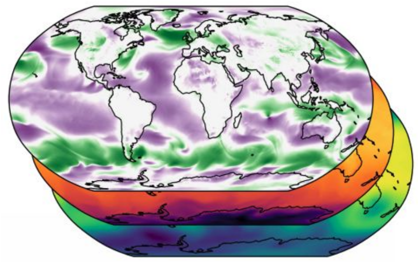
Google DeepMind

Weather forecasting context

Data

ERA5 reanalysis

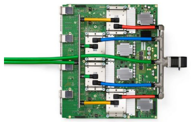
40+ yrs NWP-assimilated data



Compute

TPUs/GPUs

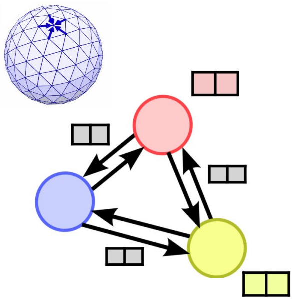
Powerful parallel computing and specialized hardware for large-scale ML models



Methods

Transformers, diffusion, etc

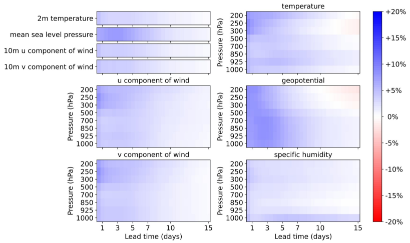
Good inductive biases capturing structure and aiding optimization



Evals & Competitions

Scorecards

Measurable, repeatable, comparative, external validation

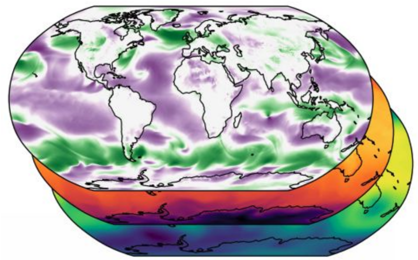


Weather forecasting context

Data

ERA5 reanalysis

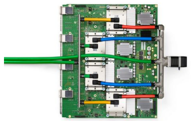
40+ yrs NWP-assimilated data



Compute

TPUs/GPUs

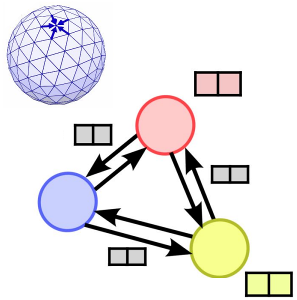
Powerful parallel computing and specialized hardware for large-scale ML models



Methods

Transformers, diffusion, etc

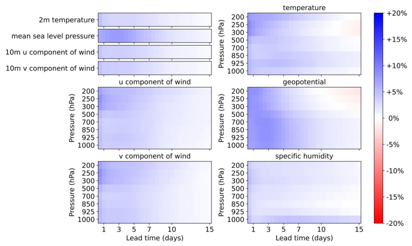
Good inductive biases capturing structure and aiding optimization



Evals & Competitions

Scorecards

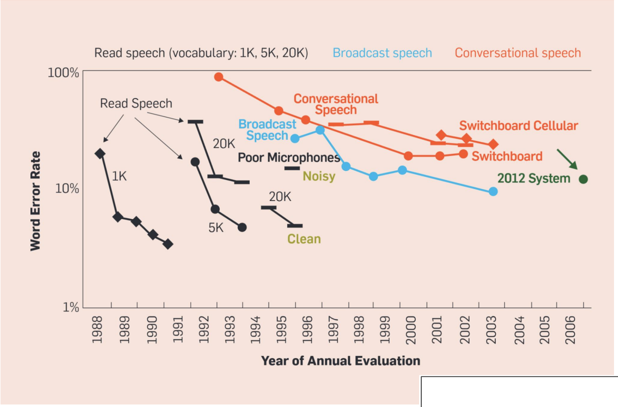
Measurable, repeatable, comparative, external validation



Common task framework

1. Shared dataset (training and testing)
2. Set of competitors (researchers/developers)
3. Automated scoring system (leaderboard)

Figure 1. Historical progress of speech recognition word error rate on more and more difficult tasks.[10] The latest system for the switchboard task is marked with the green dot.



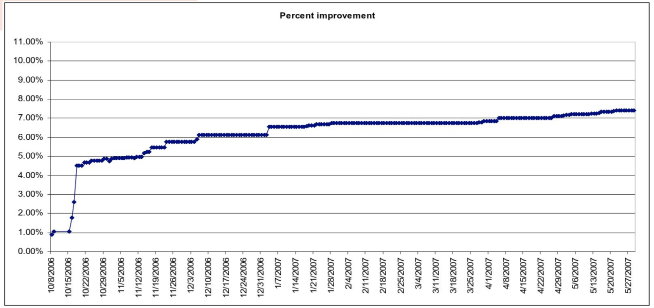
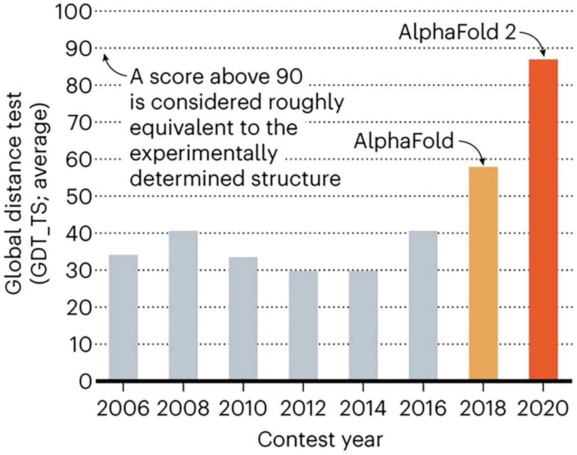
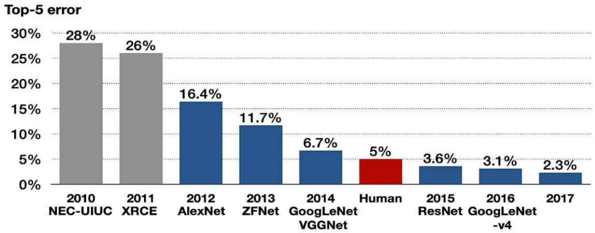


Figure 3: Aggregate improvement over Cinematch by time



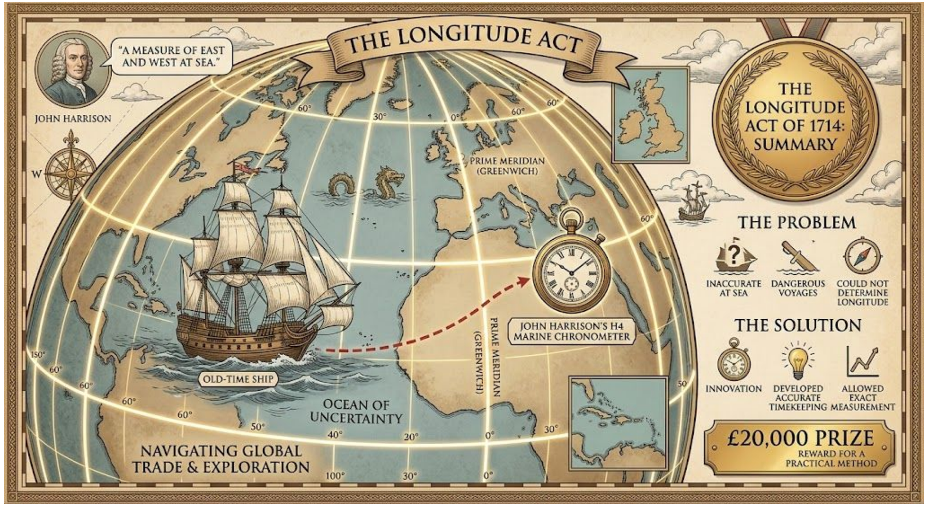


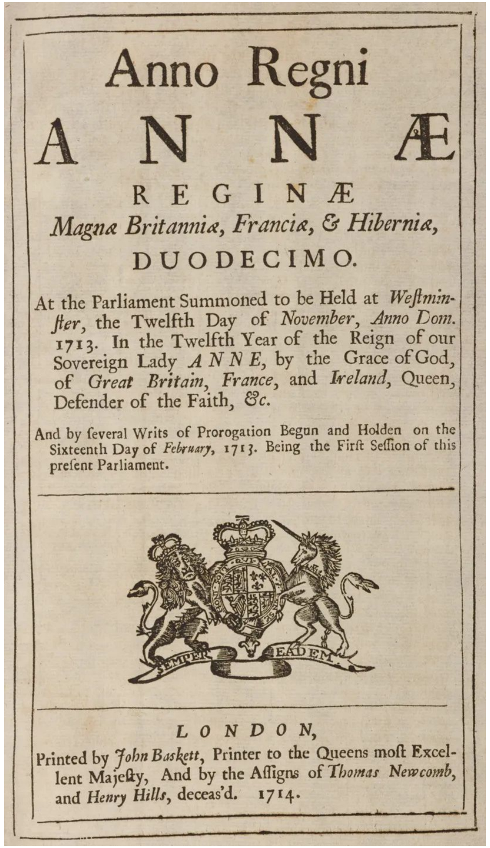
- https://viso.ai/deep-learning/imagenet/
- Xuedong Huang, James Baker, and Raj Reddy. 2014. A historical perspective of speech recognition. Commun. ACM 57, 1 (January 2014), 94–103. https://doi.org/10.1145/2500887
- Strodel, B. (2021). Energy landscapes of protein aggregation and conformation switching in intrinsically disordered proteins. Journal of Molecular Biology, 433(20), 167182. https://doi.org/10.1016/j.jmb.2021.167182
- Bennett, James and Stan Lanning. “The Netflix Prize.” (2007).

The Longitude Act 1714

First government funding and rewards for a specific ‘scientific’ problem

- £10,000 (€2.5 M today): find the longitude to 1° (60 NM)
- £15,000 (€3.8 M today): 40 NM
- £20,000 (€5.1 M today): <30 NM




Anno Regni

A N N Æ

R E G I N Æ

Magnæ Britanniæ, Franciæ, & Hiberniæ,

DUODECIMO.

At the Parliament Summoned to be Held at *Weſtminſter*, the Twelfth Day of *November*, *Anno Dom.* 1713. In the Twelfth Year of the Reign of our Sovereign Lady *A N N E*, by the Grace of God, of *Great Britain*, *France*, and *Ireland*, Queen, Defender of the Faith, &c.

And by ſeveral Writs of Prorogation Begun and Holden on the Sixteenth Day of *February*, 1713. Being the Firſt Seſſion of this preſent Parliament.

L O N D O N,

Printed by *John Baskett*, Printer to the Queens moſt Excellent Majeſty, And by the Aſſigns of *Thomas Newcomb*, and *Henry Hills*, deceas'd. 1714.

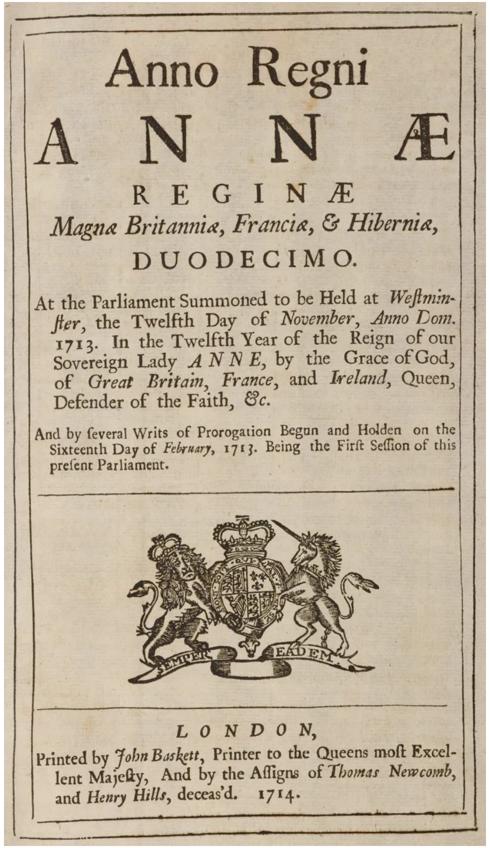
- https://historyofparliament.com/2014/07/03/finding-latitude-in-longitude-parliamentary-funding-of-early-modern-science-and-technology/
- https://watchesbysjx.com/2019/09/john-harrison-marine-chronometer-h4-diamond-pallets.html

The Longitude Act 1714

First government funding and rewards for a specific ‘scientific’ problem

- £10,000 (€2.5 M today): find the longitude to 1° (60 NM)
- £15,000 (€3.8 M today): 40 NM
- £20,000 (€5.1 M today): <30 NM

John Harrison’s H4 Chronometer



Anno Regni
ANNÆ
REGINÆ
Magnæ Britanniæ, Franciæ, & Hiberniæ,
DUODECIMO.

At the Parliament Summoned to be Held at *Weſtminſter*, the Twelfth Day of *November*, *Anno Dom.* 1713. In the Twelfth Year of the Reign of our Sovereign Lady *ANNE*, by the Grace of God, of *Great Britain*, *France*, and *Ireland*, Queen, Defender of the Faith, &c.

And by ſeveral Writs of Prorogation Begun and Holden on the Sixteenth Day of *February*, 1713. Being the Firſt Seſſion of this preſent Parliament.

SEMPER EADEM

LONDON,
Printed by *John Baskett*, Printer to the Queens moſt Excellent Majeſty, And by the Aſſigns of *Thomas Newcomb*, and *Henry Hills*, deceas’d. 1714.

- https://historyofparliament.com/2014/07/03/finding-latitude-in-longitude-parliamentary-funding-of-early-modern-science-and-technology/
- https://watchesbysjx.com/2019/09/john-harrison-marine-chronometer-h4-diamond-pallets.html

The scorecard

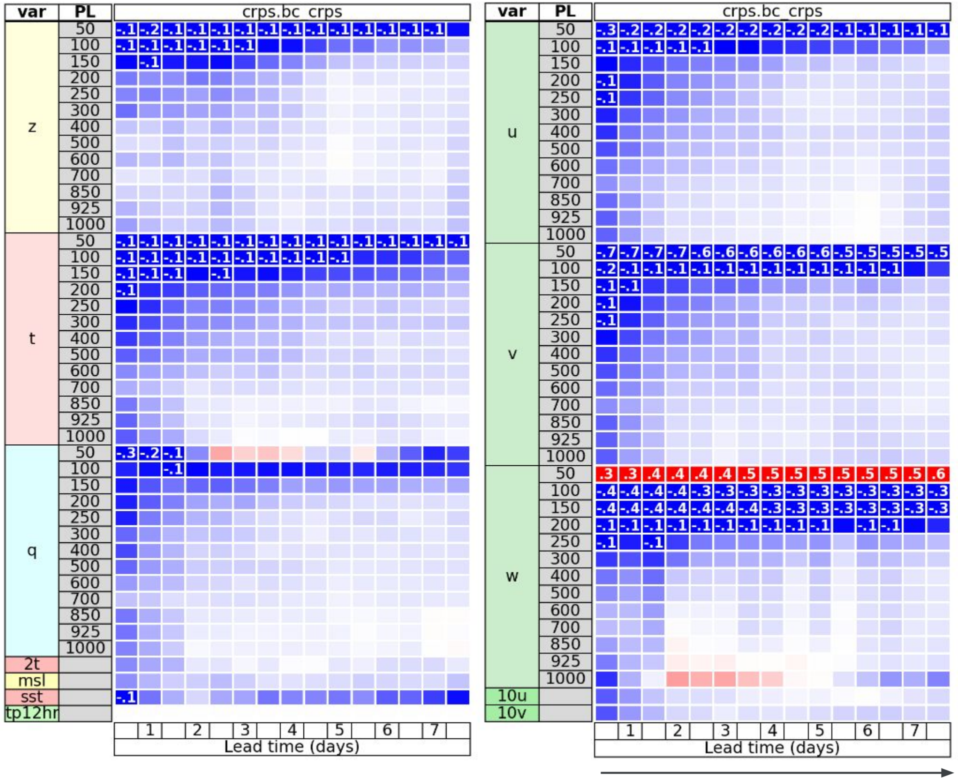
Make it blue!

geopotential

temperature

specific humidity

2m temperature
mean sea level
10m sea surface temp.
total precip. over 12 hr



zonal wind

meridional wind

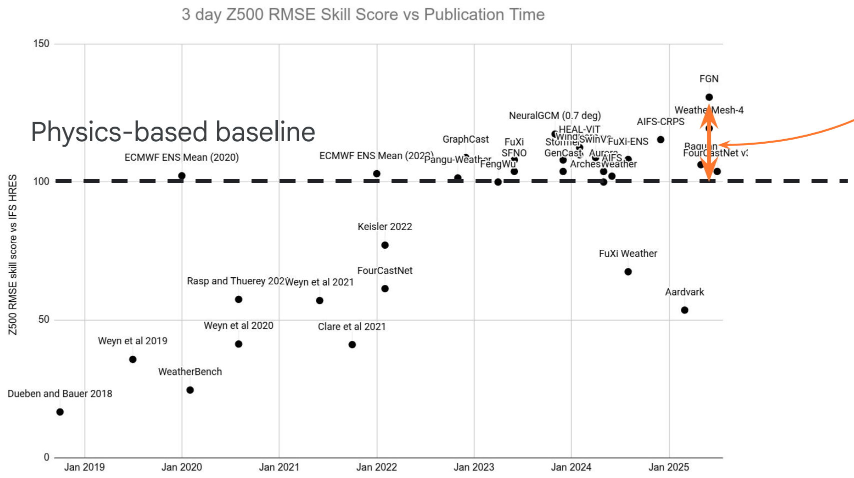
vertical wind

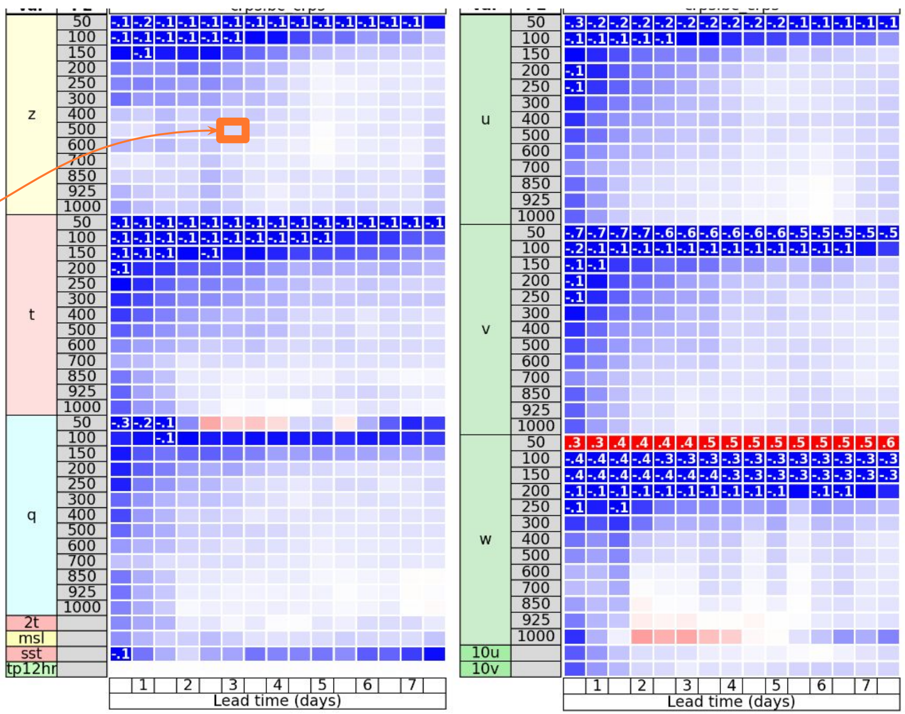
10m zonal wind
10m meridional wind

lead time

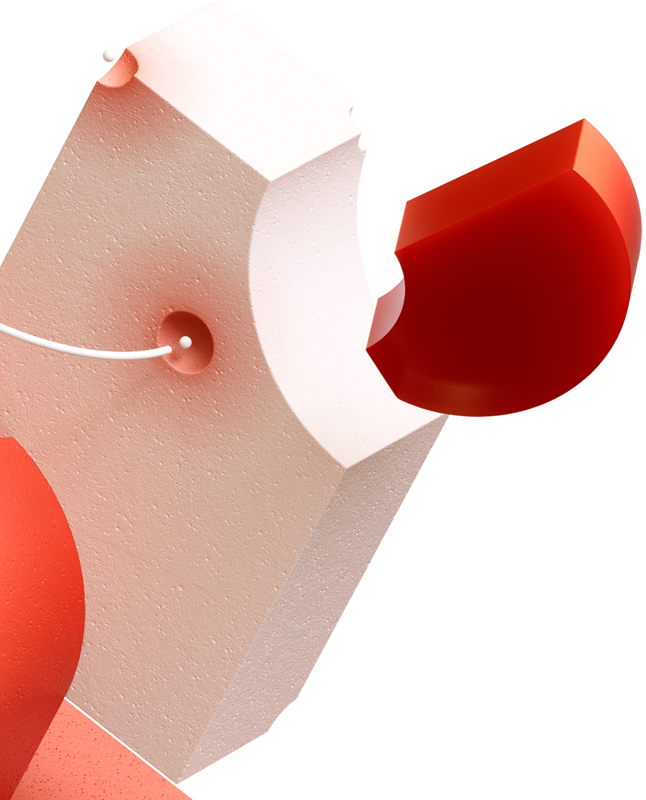
A scorecard is a 1000 words

One box = diff to baseline



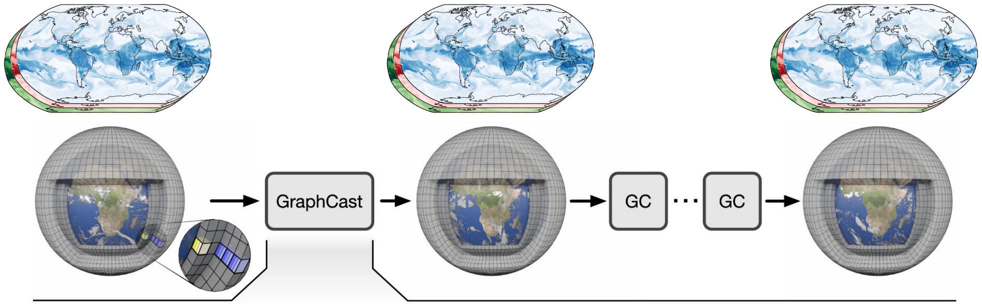


Google DeepMind

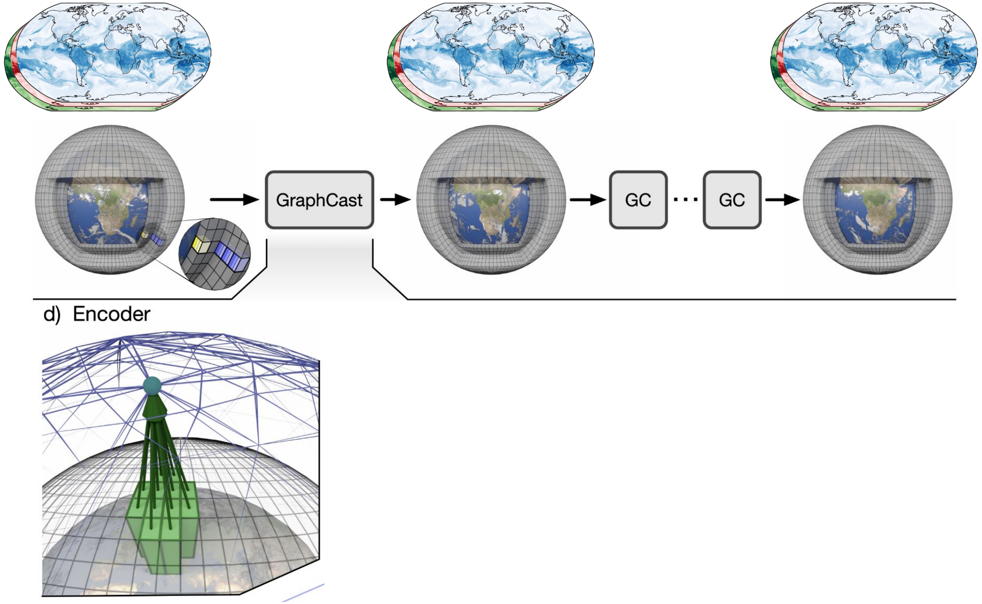


WeatherNext Series
GraphCast

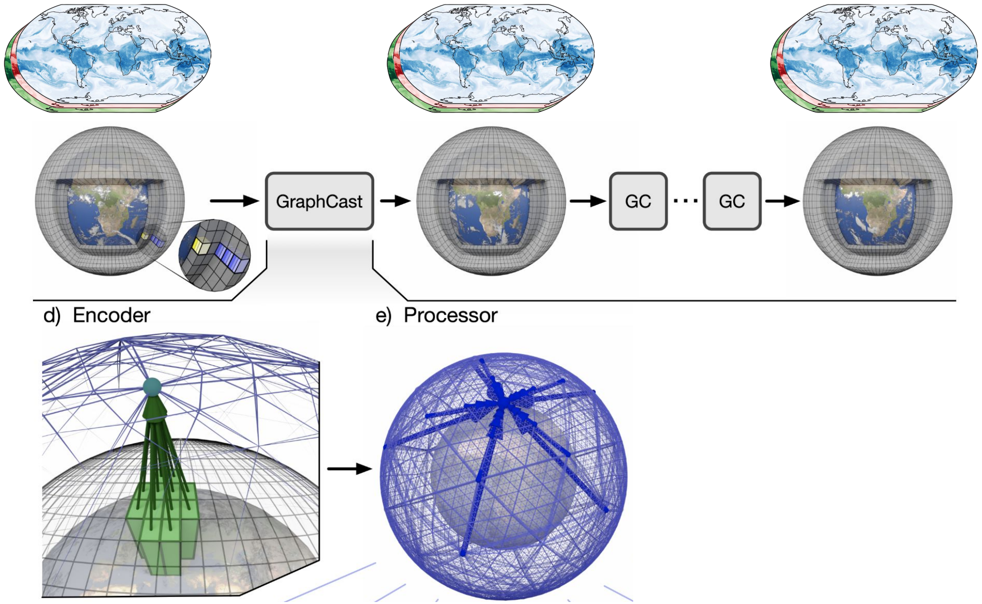
Graphcast in a slide



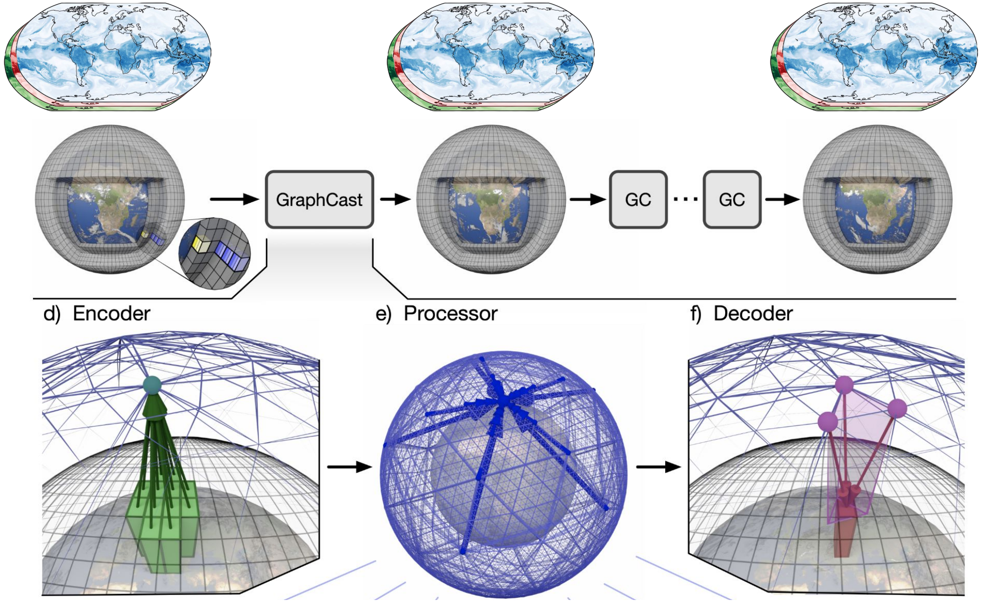
Graphcast in a slide



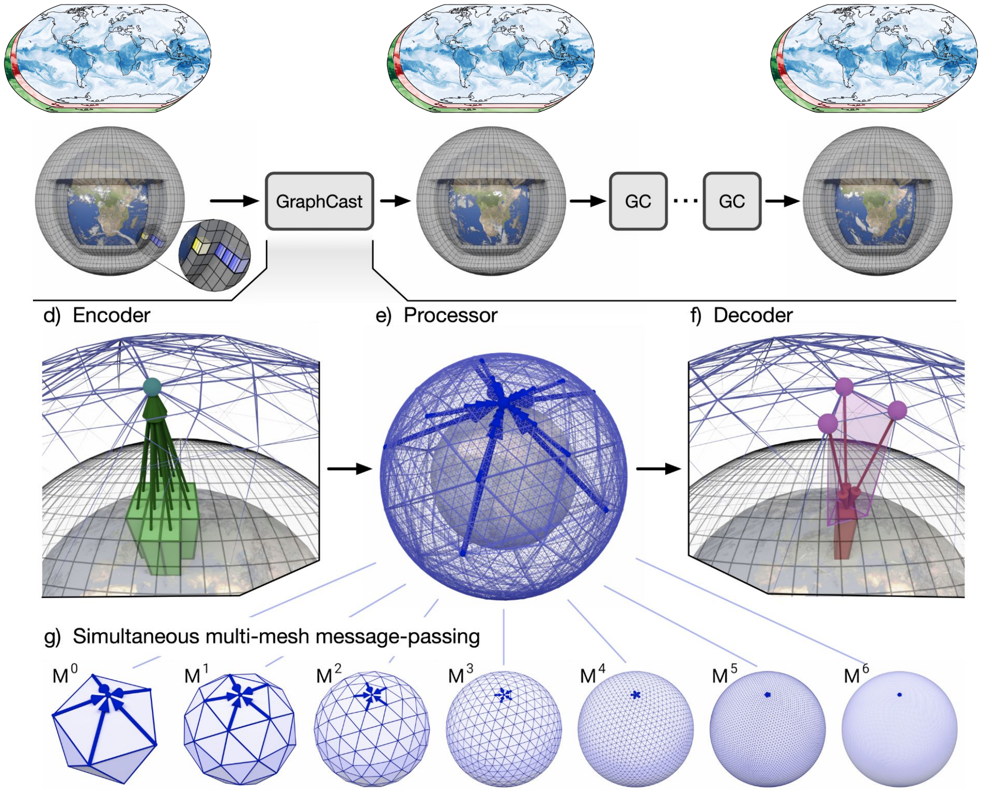
Graphcast in a slide



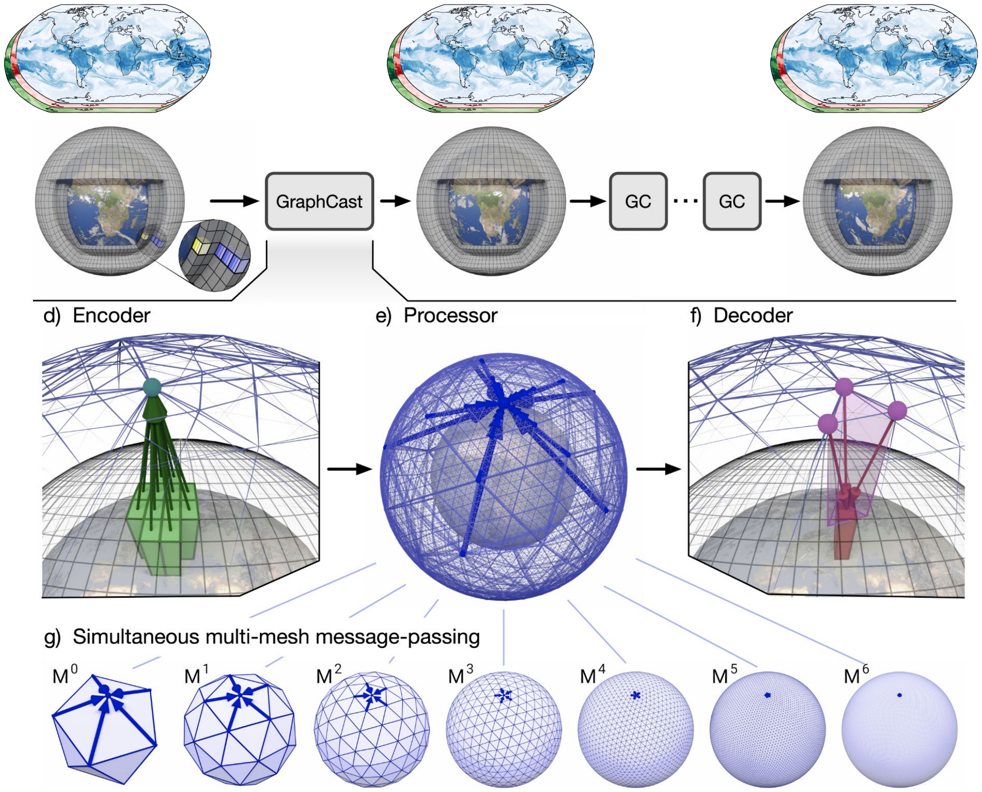
Graphcast in a slide

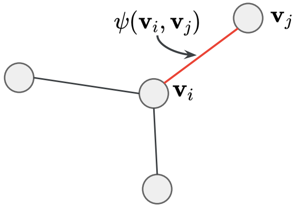


Graphcast in a slide



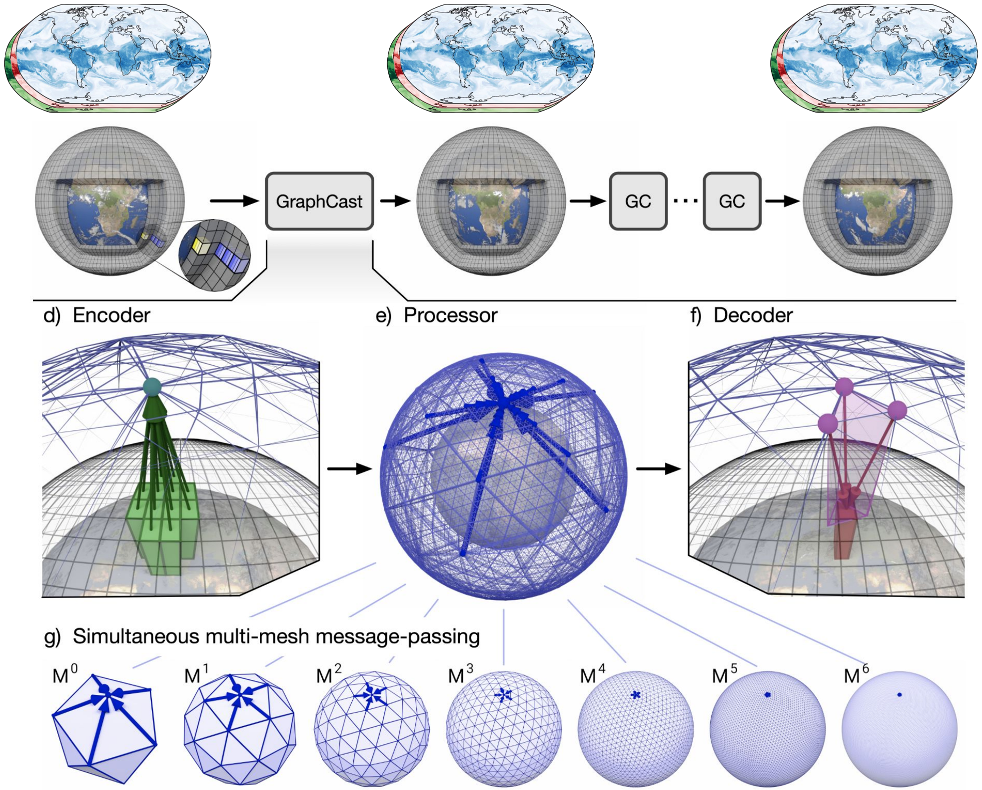
Graphcast in a slide

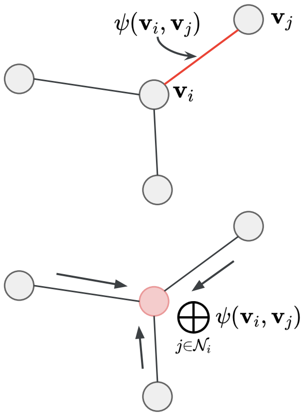




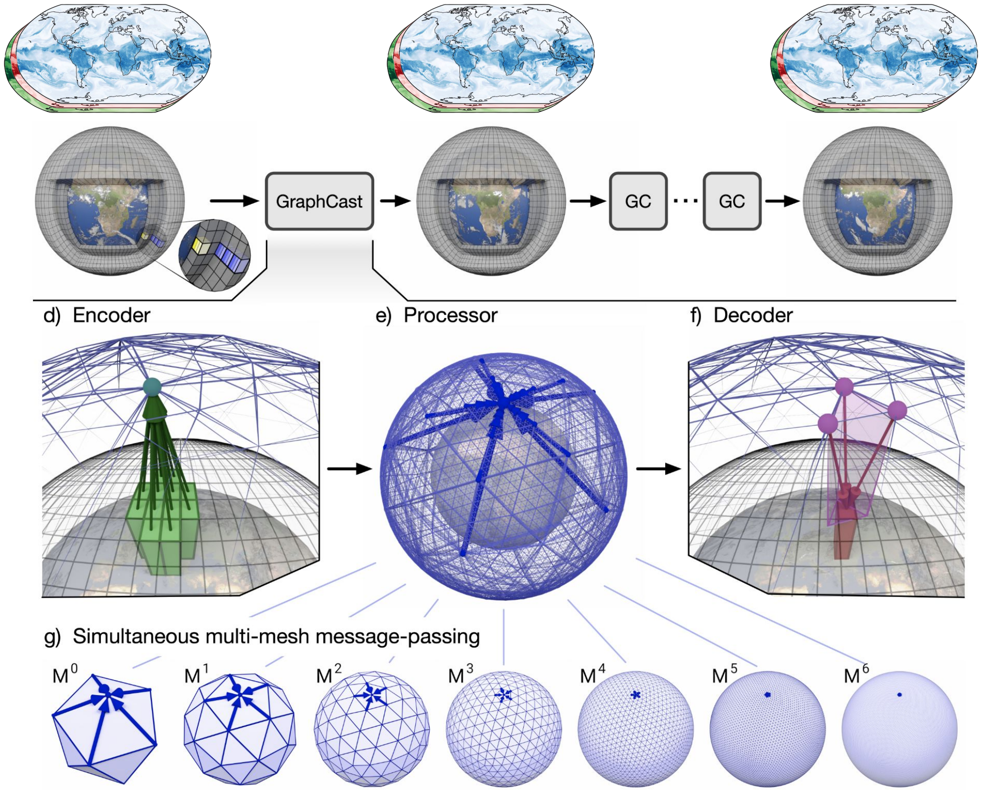
Edge updates

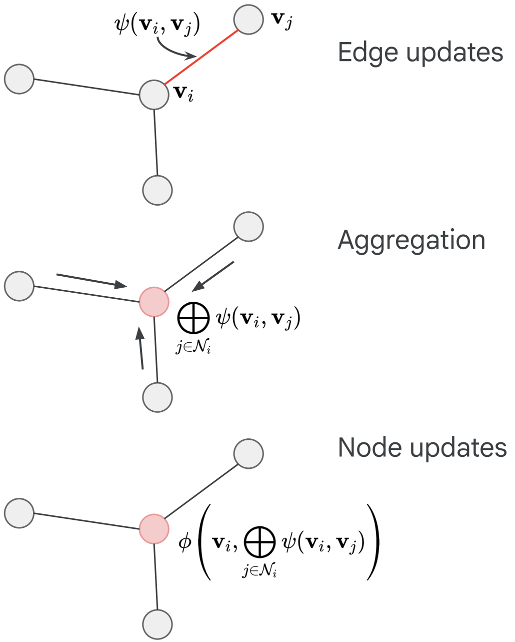
Graphcast in a slide



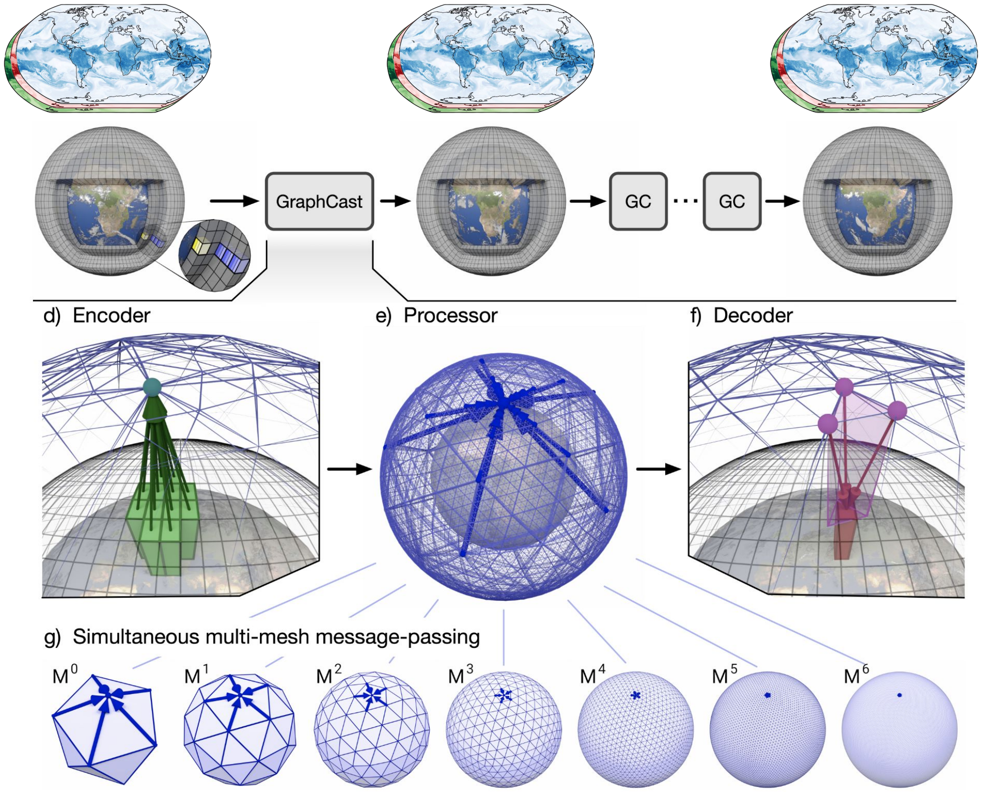


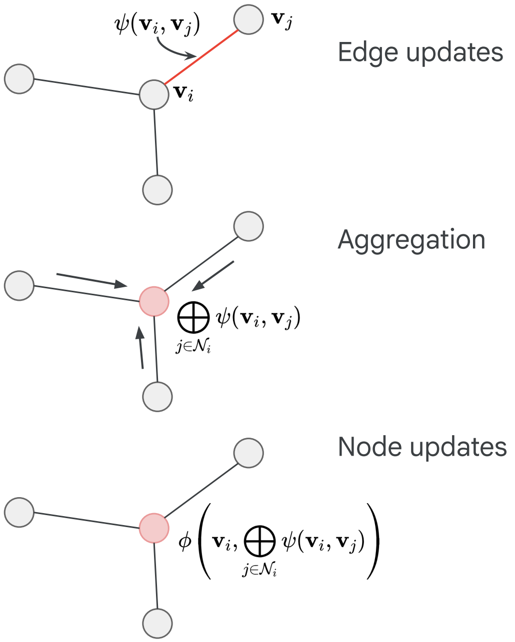
Graphcast in a slide





Graphcast in a slide

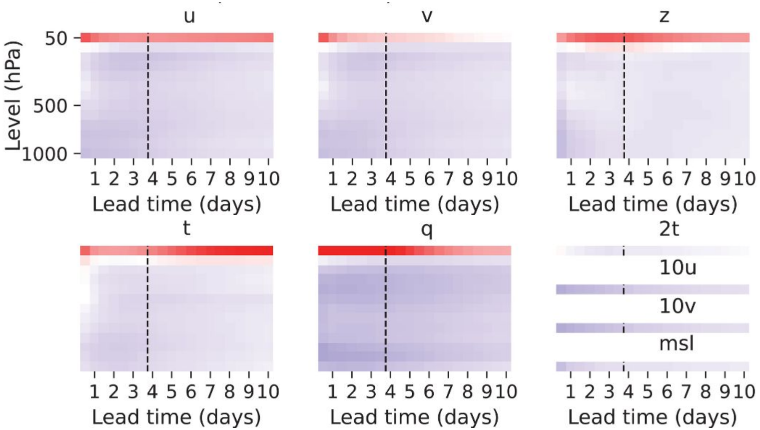




- Train on MSE
- 36.7 million weights, 40962 mesh nodes, 1038240 lat-lon points, 6 hour steps, 10 days
- No guarantees! Strange idea alert!

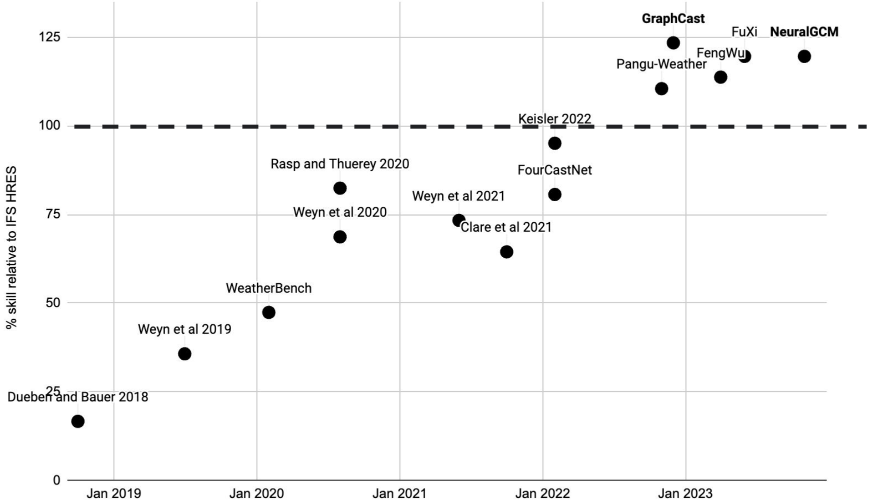
Results

Outperformed the industry-leading physics-based models (HRES) in **accuracy** and **speed (**10^5x acceleration**)**



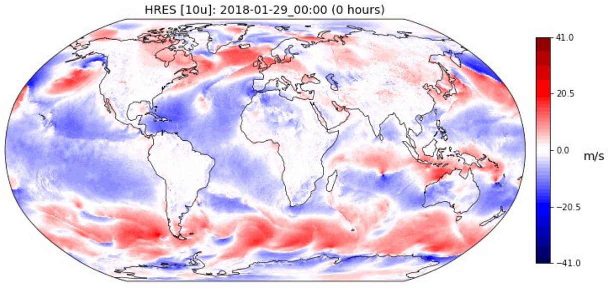
AI wins

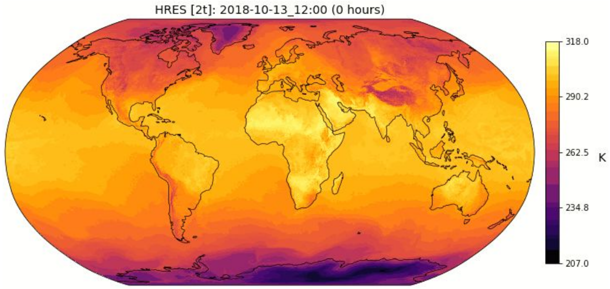
Among 1st AI systems to beat SoTA physics-based simulations

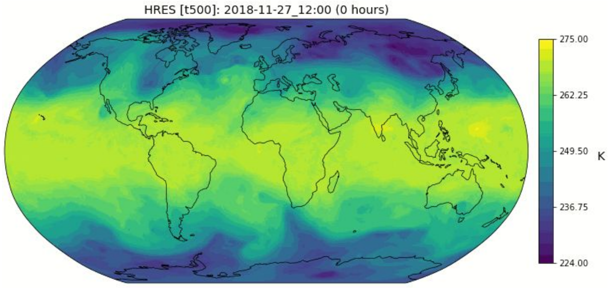


Example forecasts

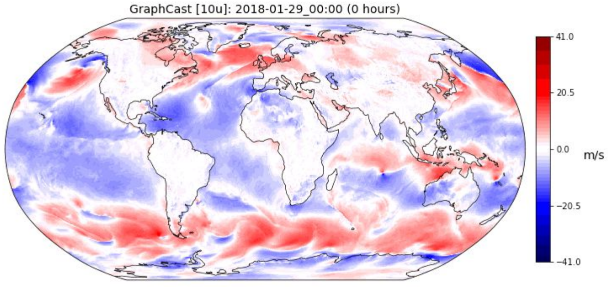
HRES

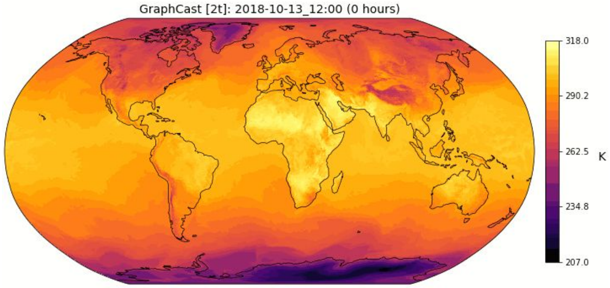


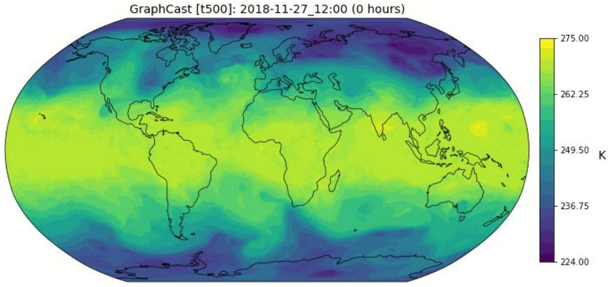




GraphCast







Surface E-W wind

Surface temperature

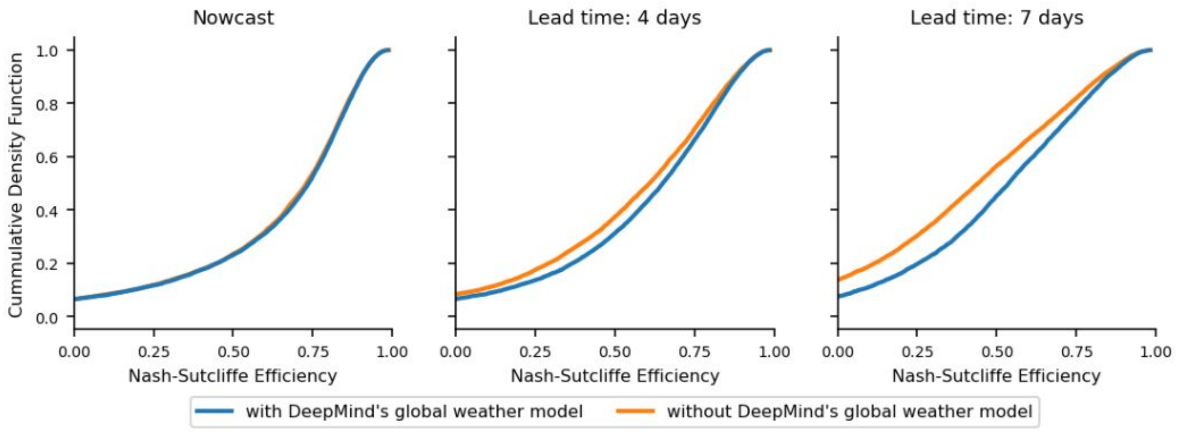
Temperature @ 500 hPa

Improved flood prediction with an improved weather forecasting model

Past experiments showed that LSTM models learn to combine and leverage multiple weather products. Thus, by adding more sources of weather data we expected to improve the skill of the flood forecasting model. So, we incorporated Google DeepMind's medium-range global weather forecasting model as input to the existing hydrologic model.

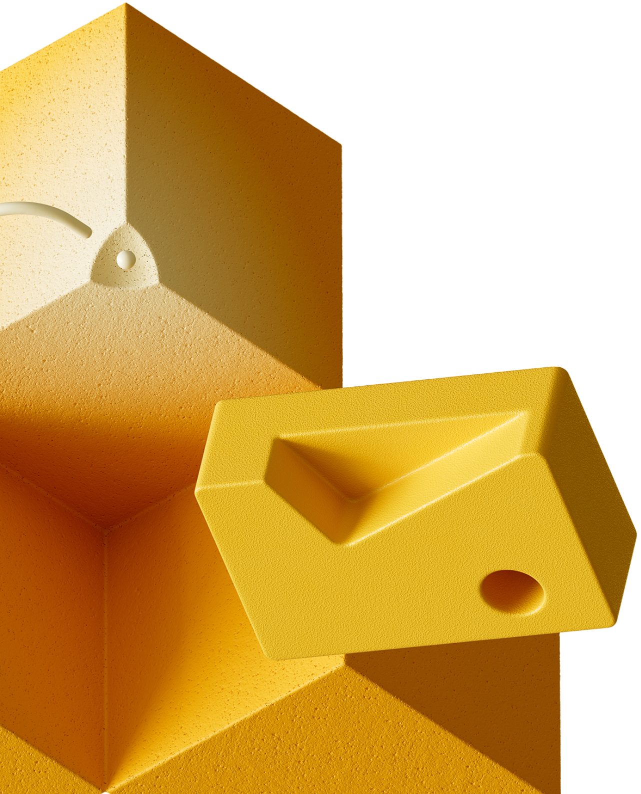
This model is able to make medium-range weather forecasts with higher accuracy than the traditional deterministic gold standard in weather forecasting. It predicts five Earth-surface variables and six atmospheric variables. We integrate two of these surface variables, precipitation and temperature, into the flood forecasting hydrologic model.

To measure the impact of this global forecasting weather model on the reliability of our global hydrologic model, we compare our model performance with and without the addition of the weather forecasting model as input. As shown in the graphs below, the impact grows as lead time increases.

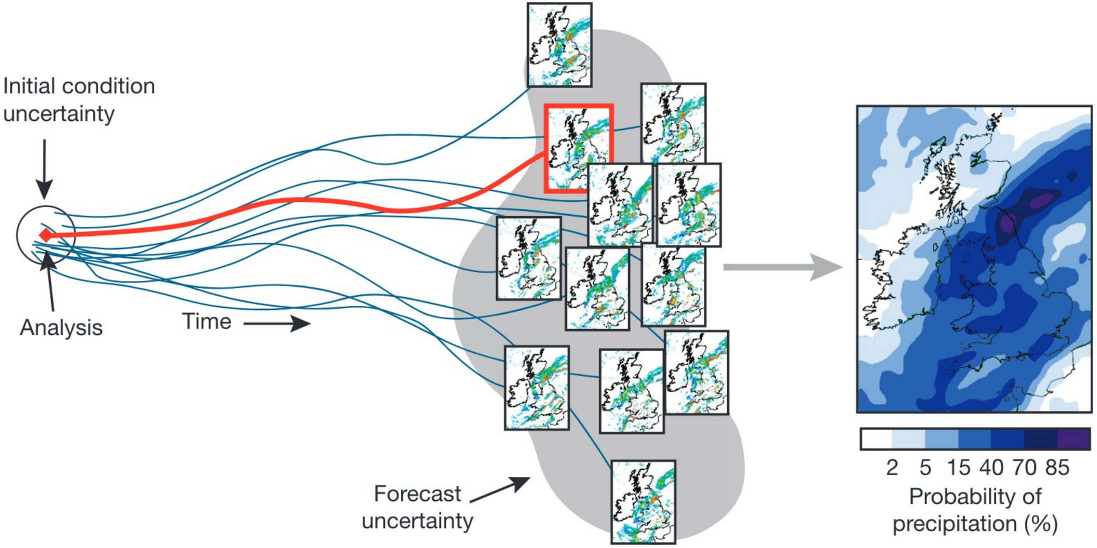


Google DeepMind

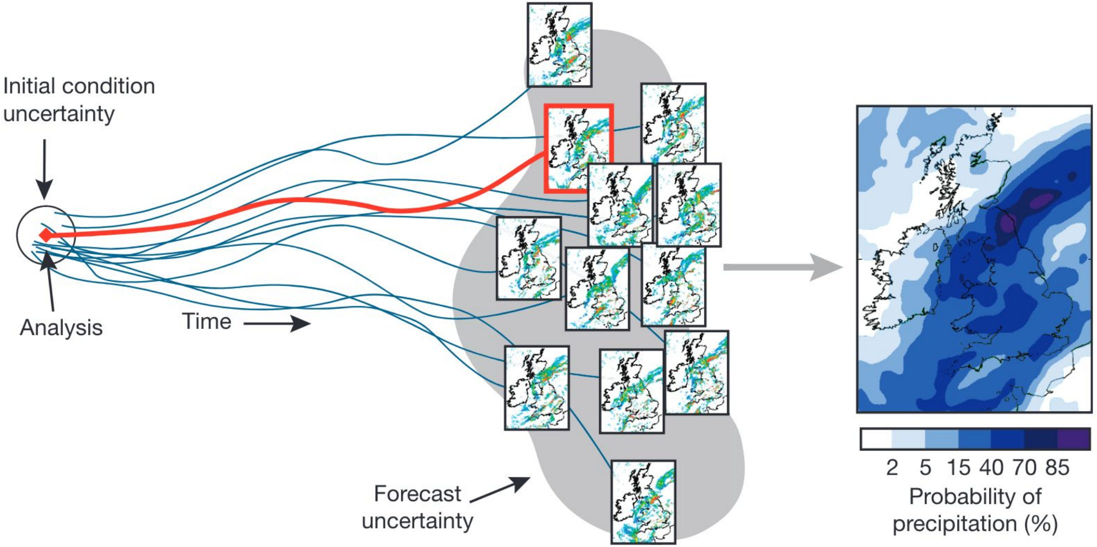
WeatherNext Series
GenCast

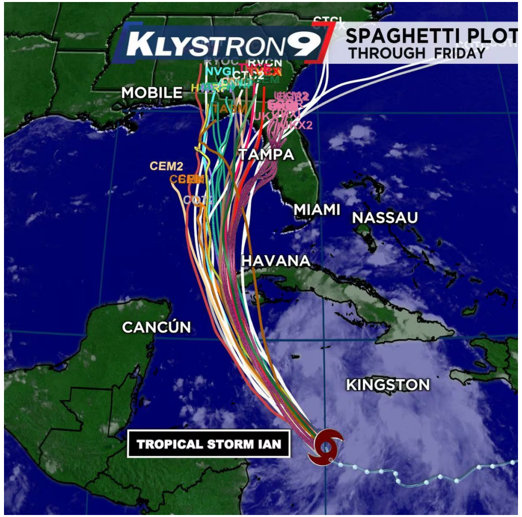


Weather is chaotic



Weather is chaotic



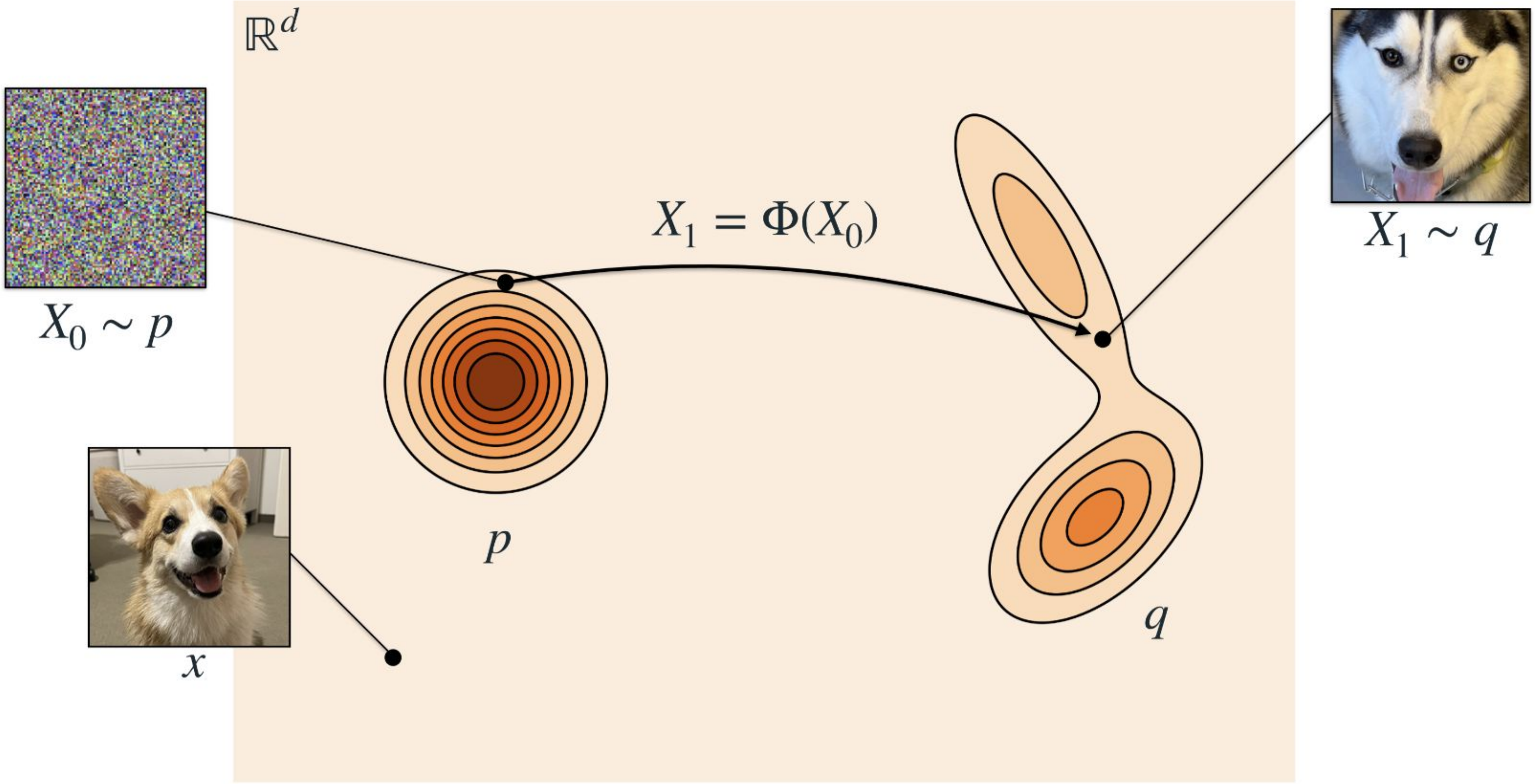


Given the present/past weather states, sample possible futures

$$p(\text{future weather}|\text{past weather})$$

- Models confidence
- Fixes blurring (more on this later)
- Easier modeling of extremes
- Strictly more general than deterministic
- Graceful convergence to climatology

Most generative modeling



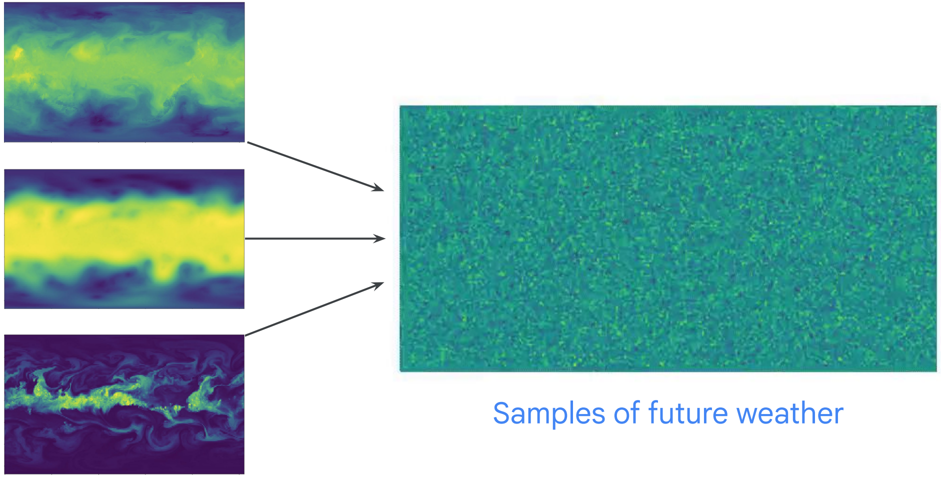
Diffusion recap

Diffusion (et al.) models are driving wider advances in generative AI

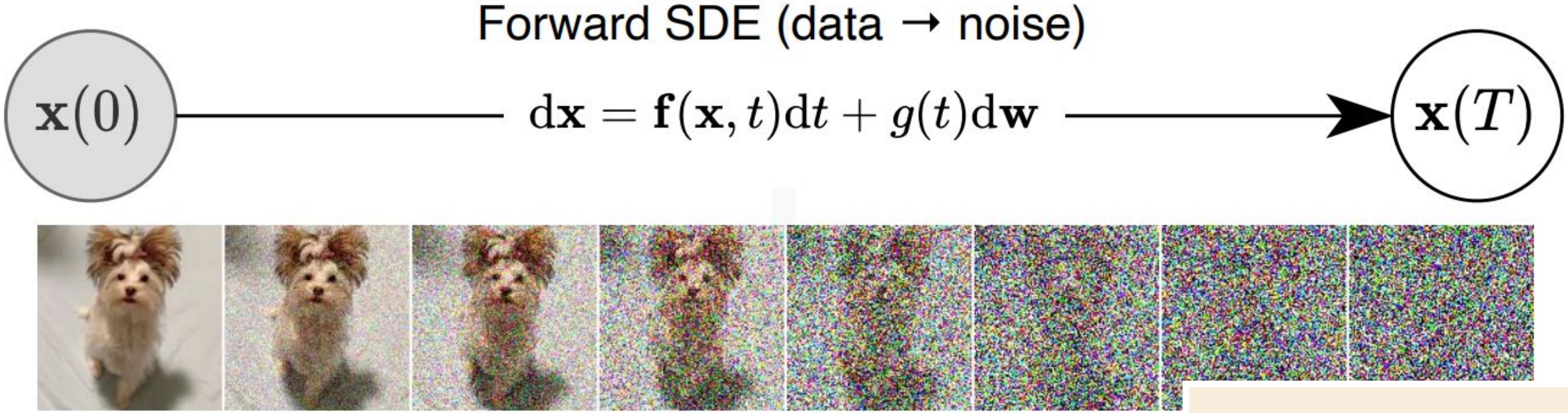
- Nano banana, SORA, Midjourney, etc

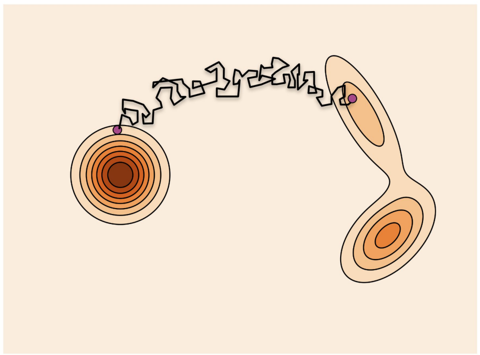
Figure from Saharia et al. 2022 “Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding”

$$x \sim p\left(\mathrm{weather}_{\mathrm{t+12hr}} \,\middle|\, \mathrm{weather}_{\mathrm{t}}, \mathrm{weather}_{\mathrm{t\text{-}12hr}}\right)$$



Diffusion recap

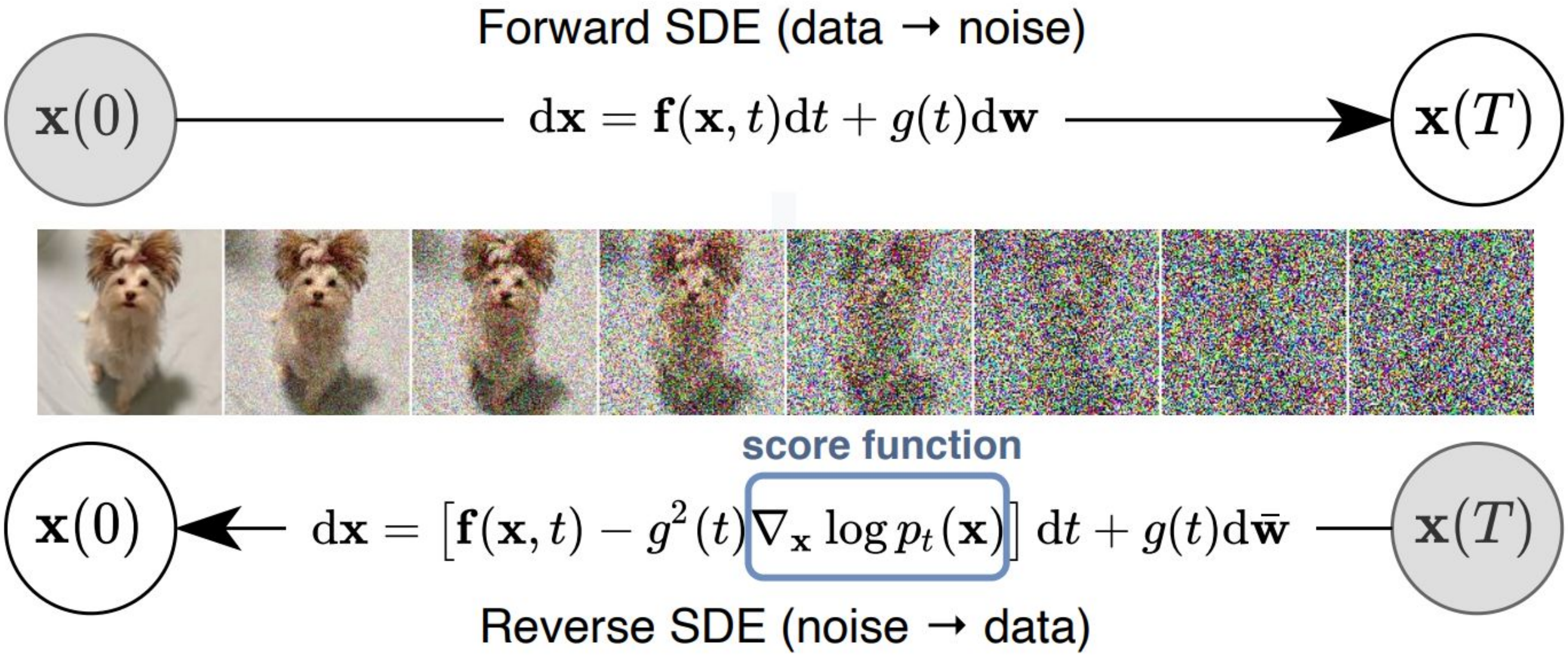




Diffusion

Song et al. 2021 “Score-Based Generative Modeling through Stochastic Differential Equations”

Diffusion recap



Song et al. 2021 “Score-Based Generative Modeling through Stochastic Differential Equations”

Denoising score matching

Let $D(\boldsymbol{x};\sigma)$ be a denoiser minimising $\mathbb{E}_{\boldsymbol{y}\sim p_{\text{data}}}\mathbb{E}_{\boldsymbol{n}\sim\mathcal{N}(\mathbf{0},\sigma^2\mathbf{I})}\|D(\boldsymbol{y}+\boldsymbol{n};\sigma)-\boldsymbol{y}\|_2^2$

Then $\boxed{\nabla_{\boldsymbol{x}}\log p(\boldsymbol{x};\sigma)=\big(D(\boldsymbol{x};\sigma)-\boldsymbol{x}\big)/\sigma^2}$

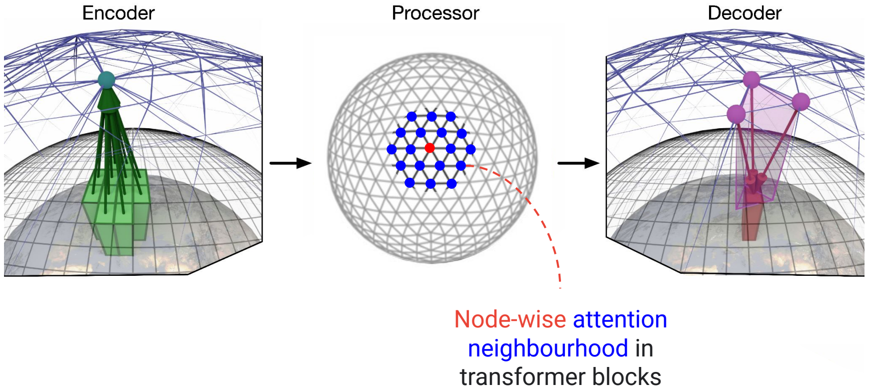
Approximate score function by **training denoiser**

To sample numerically solve reverse diffusion SDE*, using score-function approximation

* or its associated probability flow ODE

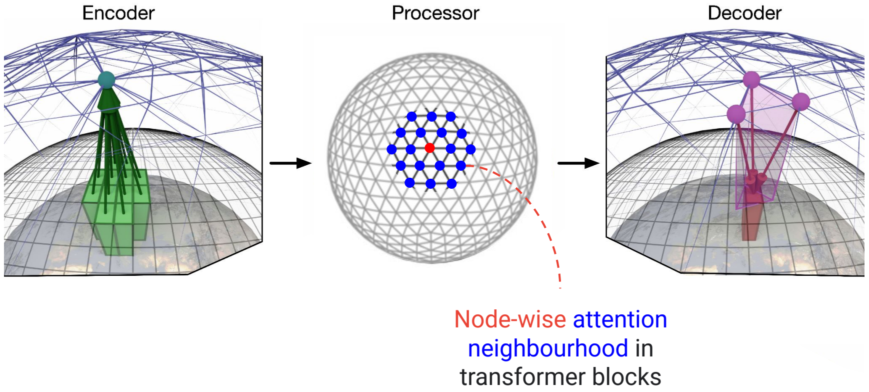
Denoiser Architecture

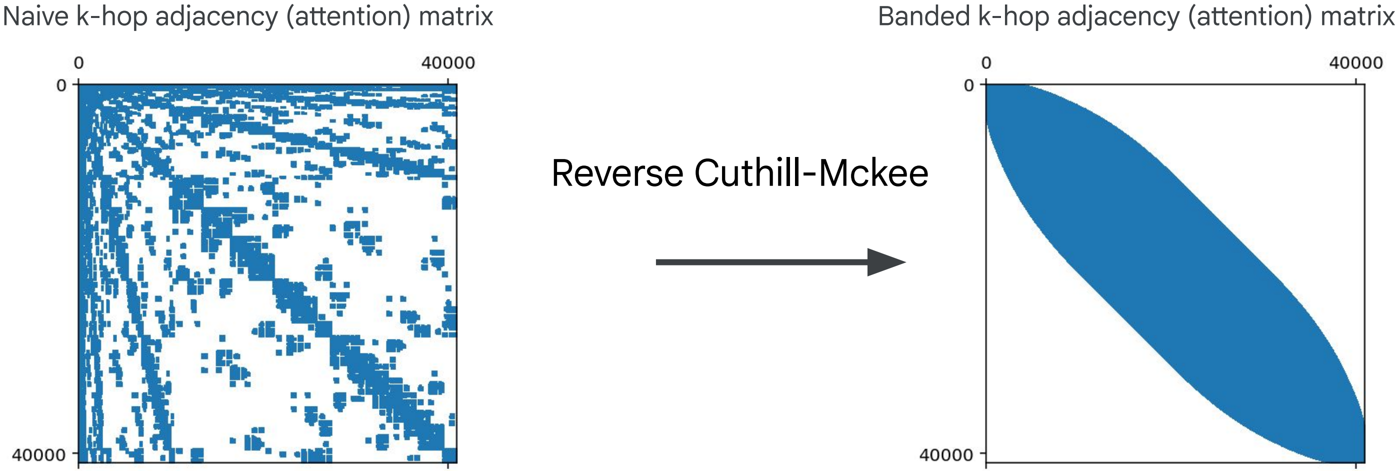
- Processor GNN ⟶ Processor transformer
- k-hop neighborhoods
  - Loosening of geometric assumptions.
  - Strange idea alert!
- ~40000 mesh nodes

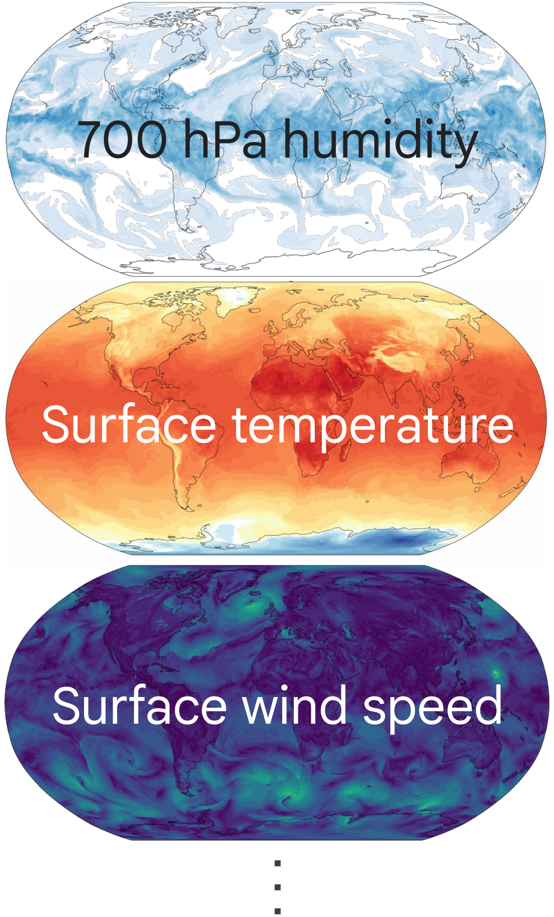


Denoiser Architecture

- Processor GNN ⟶ Processor transformer
- k-hop neighborhoods
  - Loosening of geometric assumptions
  - Strange idea alert!
- ~40000 mesh nodes







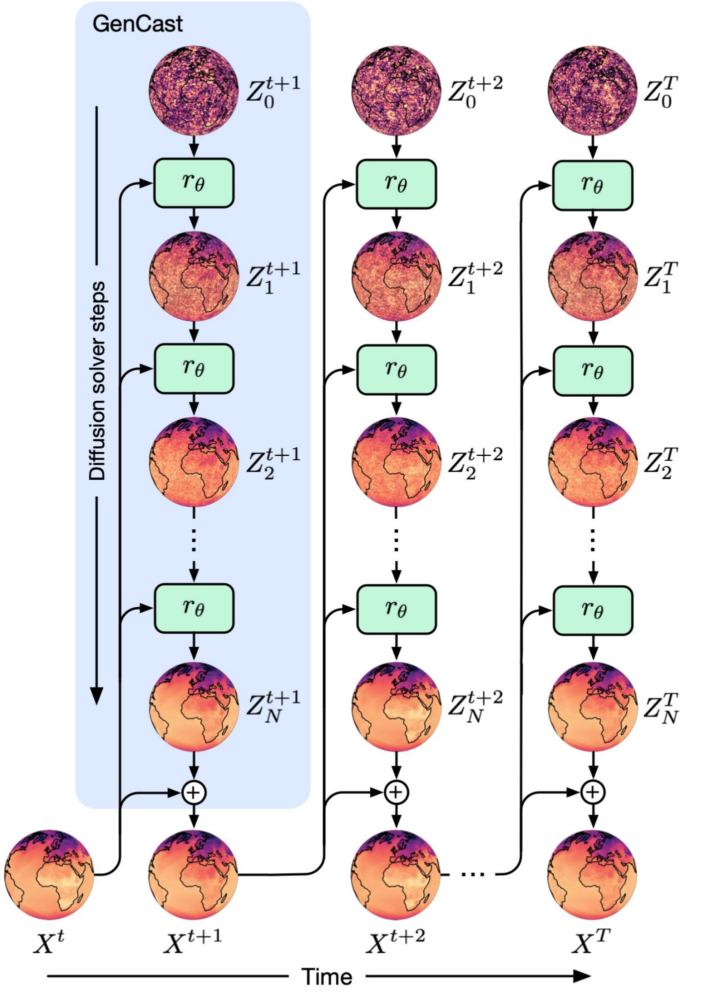
Variables
83 weather variables
(i.e. "channels")

Spatial
0.25° resolution
1M+ pixels

Temporal
12h steps
15 day predictions
8 mins/member on a TPU

Uncertainty
50 (or more) ensemble
members

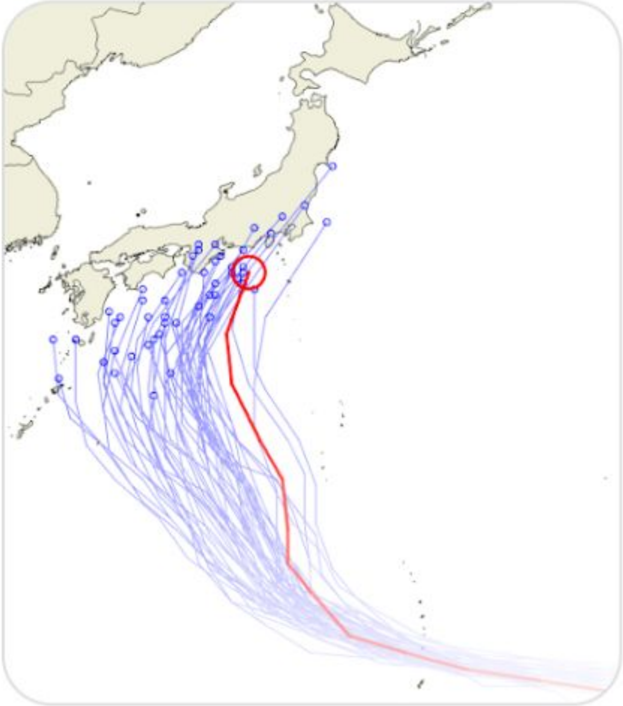
1 ensemble forecast =
1.3×10^{11} values

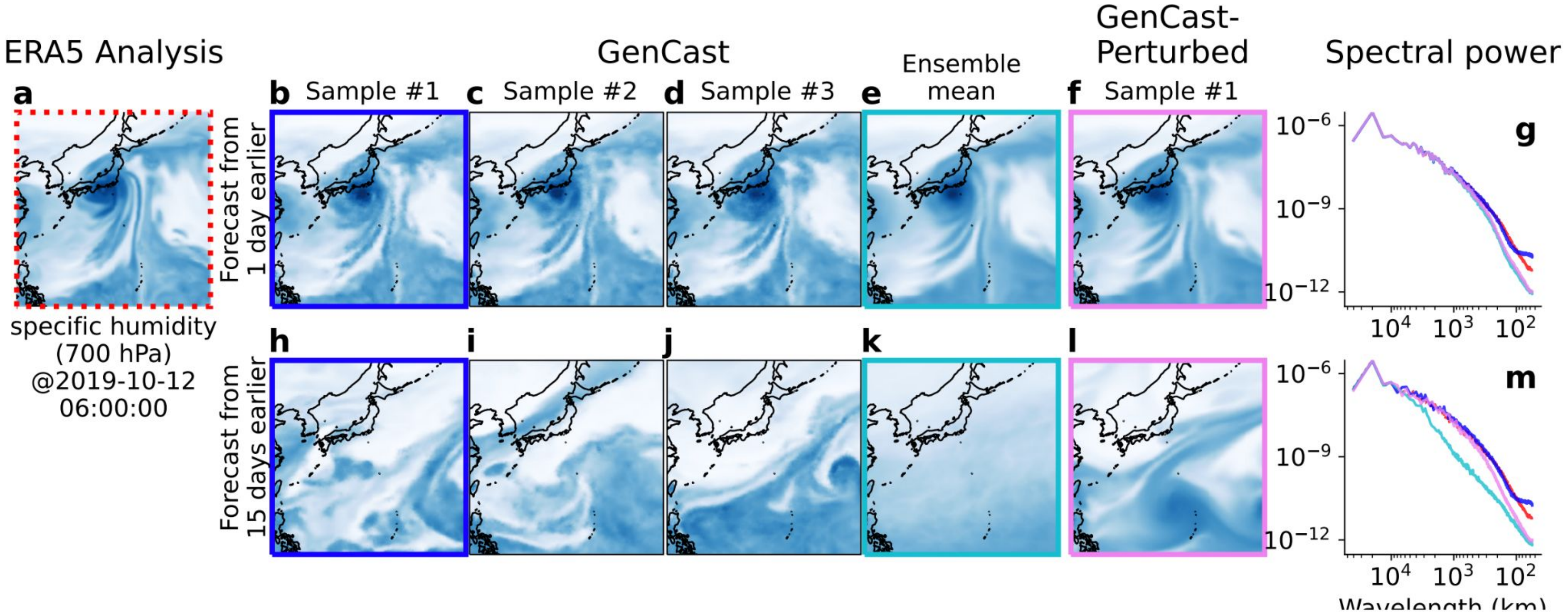


Probabilistic Forecasting

Standard conditional diffusion methods apply to the weather

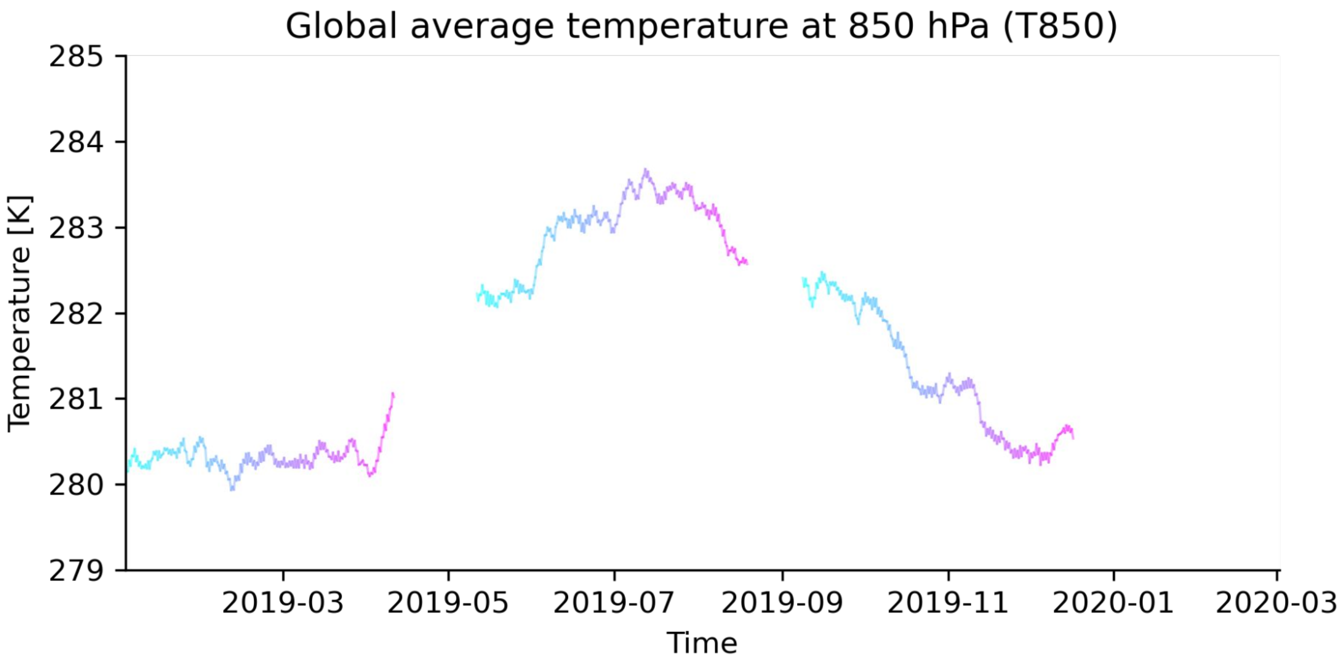
- ✚ This allows to implicitly sample from the full joint distribution of outputs
- ✚ Simply produce multiple ensemble members
- ▬ Takes many forward passes to produce a forecast
- ▬ Cannot back propagate through the diffusion sampling process





ERA5 Analysis
a
specific humidity
(700 hPa)
@2019-10-12
06:00:00
GenCast
b
Sample #1
c
Sample #2
d
Sample #3
Ensemble
mean
e
GenCast-
Perturbed
f
Sample #1
Spectral power
Forecast from
1 day earlier
g
h
i
j
k
l
Forecast from
15 days earlier
m
10^{-6}
10^{-9}
10^{-12}
10^4
10^3
10^2
Wavelength (km)

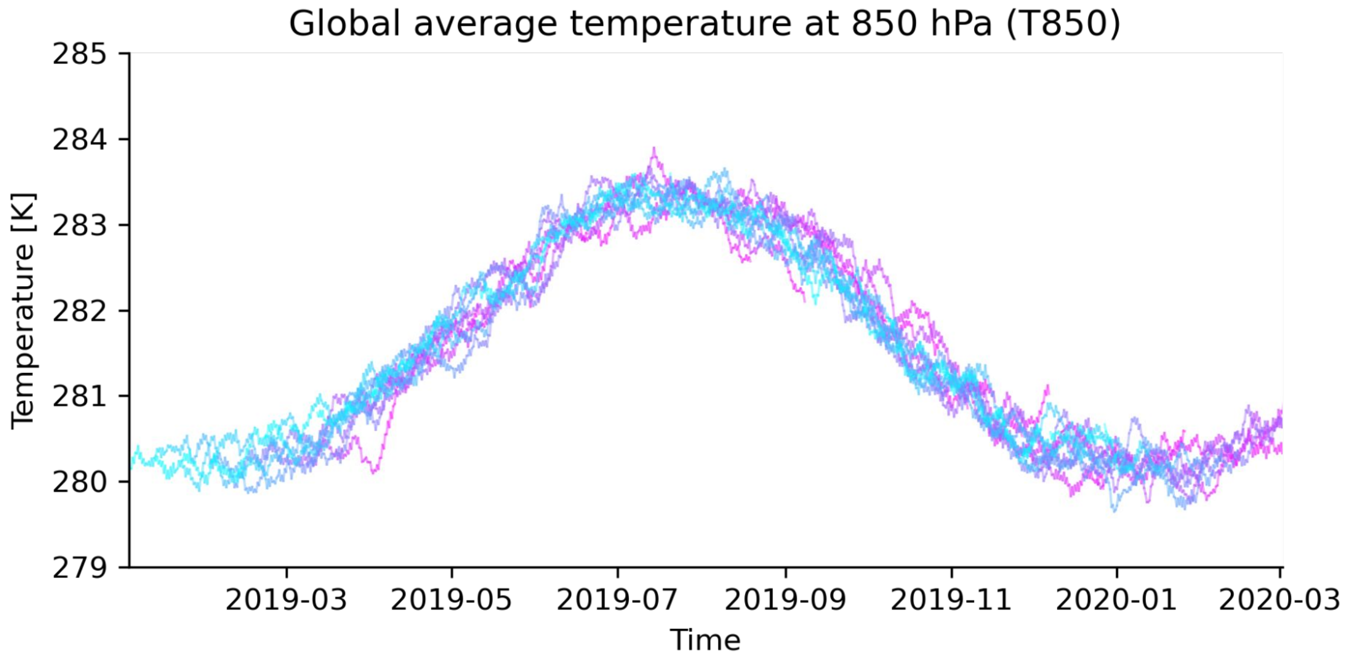
Longer rollouts with 1° version of GenCast



100-day rollouts from different start dates

Work by Dmitrii Kochkov

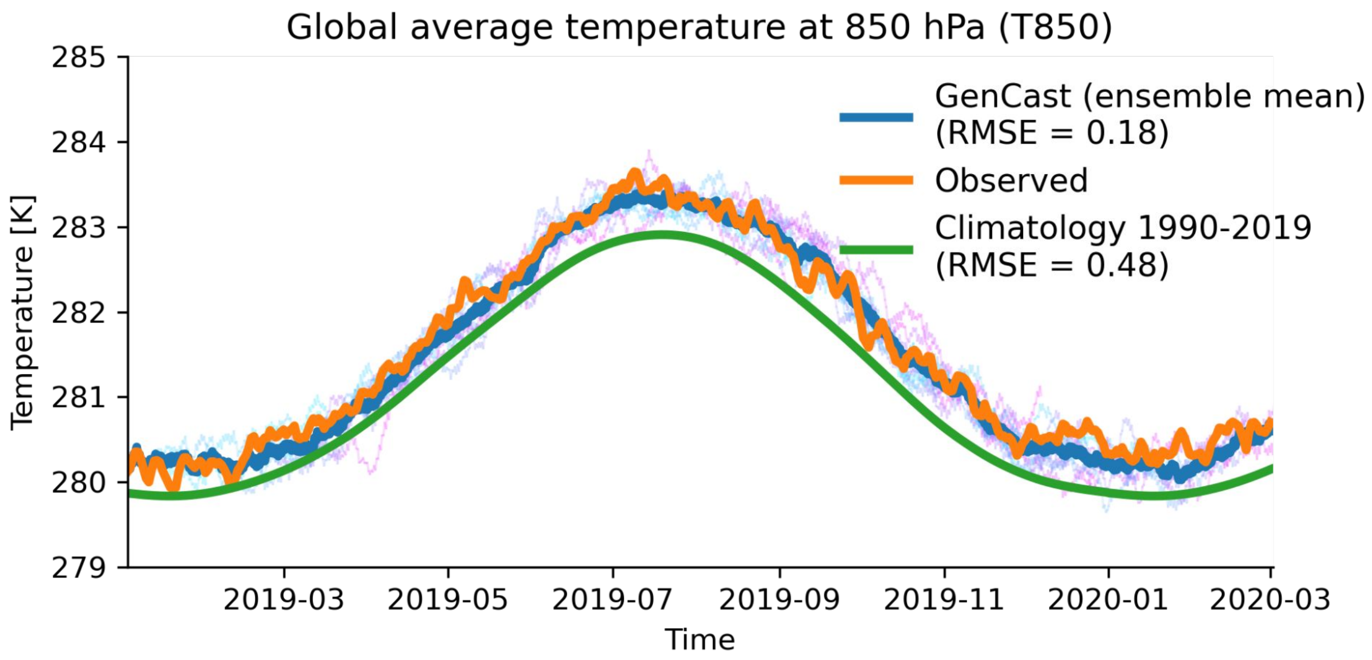
Longer rollouts with 1° version of GenCast



36 100-day rollouts from different start dates

Work by Dmitrii Kochkov

Longer rollouts with 1° version of GenCast:



Stable and track the global temperature cycle

Work by Dmitrii Kochkov

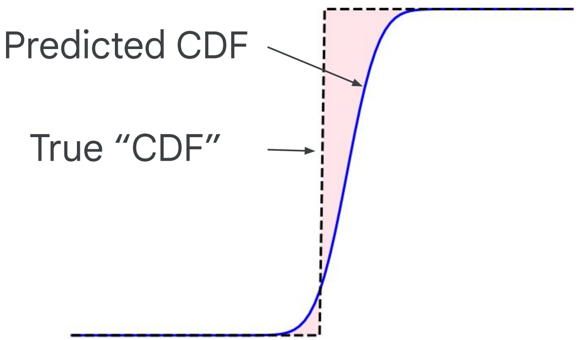
Google DeepMind

WeatherNext Series

Functional Generative Networks

How probabilistic forecasts are evaluated

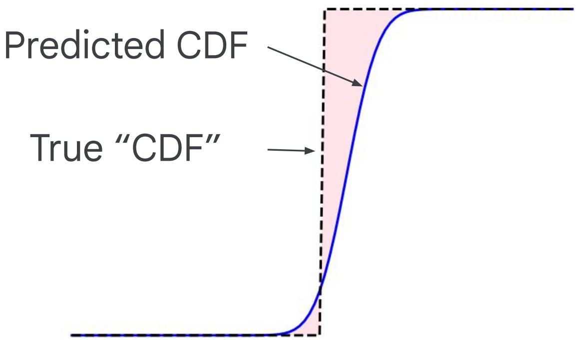
- Continuous ranked probability score (CRPS)
- Error between the predicted and true CDF
- Balances GT alignment and uncertainty



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

How probabilistic forecasts are evaluated

- Continuous ranked probability score (CRPS)
- Error between the predicted and true CDF
- Balances GT alignment and uncertainty



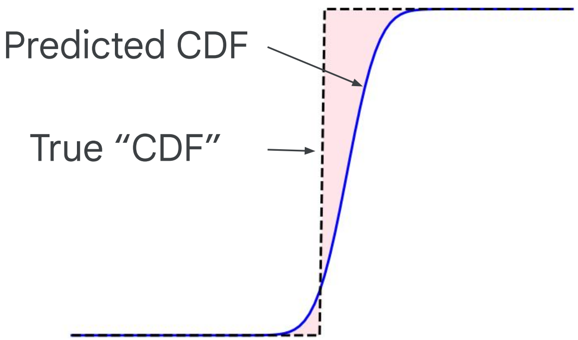
Ground truth

Ensemble members

$$\mathrm{CRPS}(F; y) = \int_{-\infty}^{\infty} \left(F(x) - \mathbb{1}[x \geq y]\right) \mathrm{d}x = \underbrace{\mathbb{E}_{X \sim F}[|X - y|]}_{\text{Errors}} - \frac{1}{2} \underbrace{\mathbb{E}_{X,X' \sim F}\left[|X - X'|\right]}_{\text{Predictive uncertainty}}$$

How probabilistic forecasts are evaluated

- Continuous ranked probability score (CRPS)
- Error between the predicted and true CDF
- Balances GT alignment and uncertainty



Idea: Just train on CRPS

Strange idea alert!

Ground truth

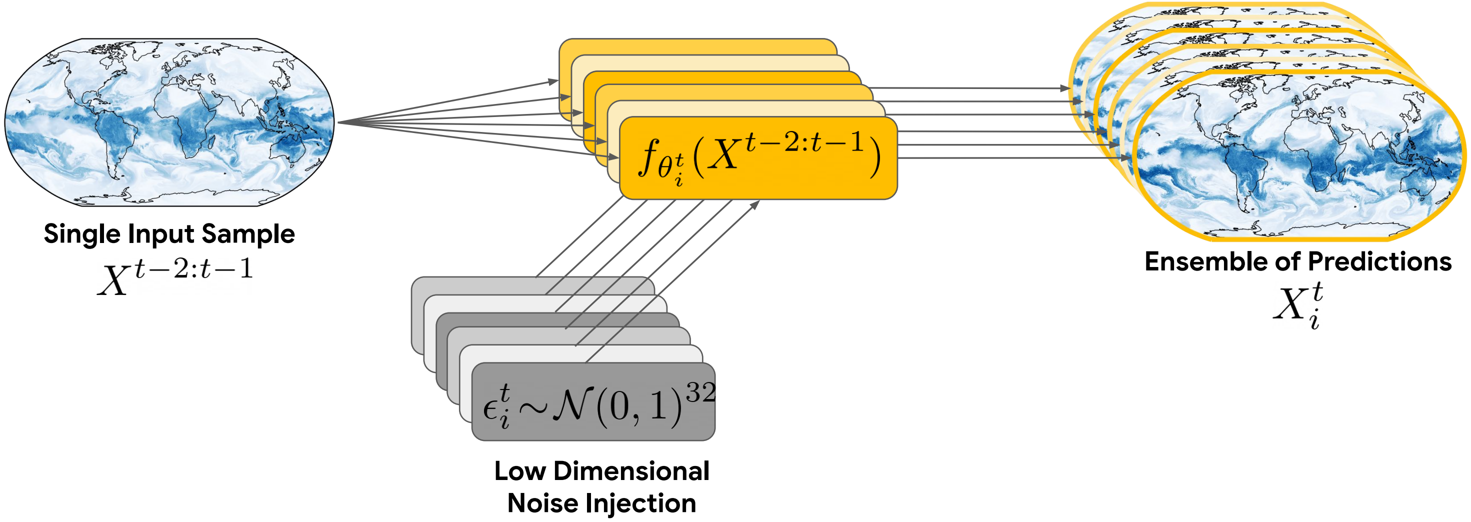
Ensemble members

$$\mathrm{CRPS}(F; y) = \int_{-\infty}^{\infty} (F(x) - \mathbb{1}[x \geq y])\,\mathrm{d}x = \underbrace{\mathbb{E}_{X\sim F}[|X - y|]}_{\text{Errors}} - \frac{1}{2}\underbrace{\mathbb{E}_{X,X'\sim F}[|X - X'|]}_{\text{Predictive uncertainty}}$$

Functional Generative Networks (FGN)

FGNs simpler than GenCast:

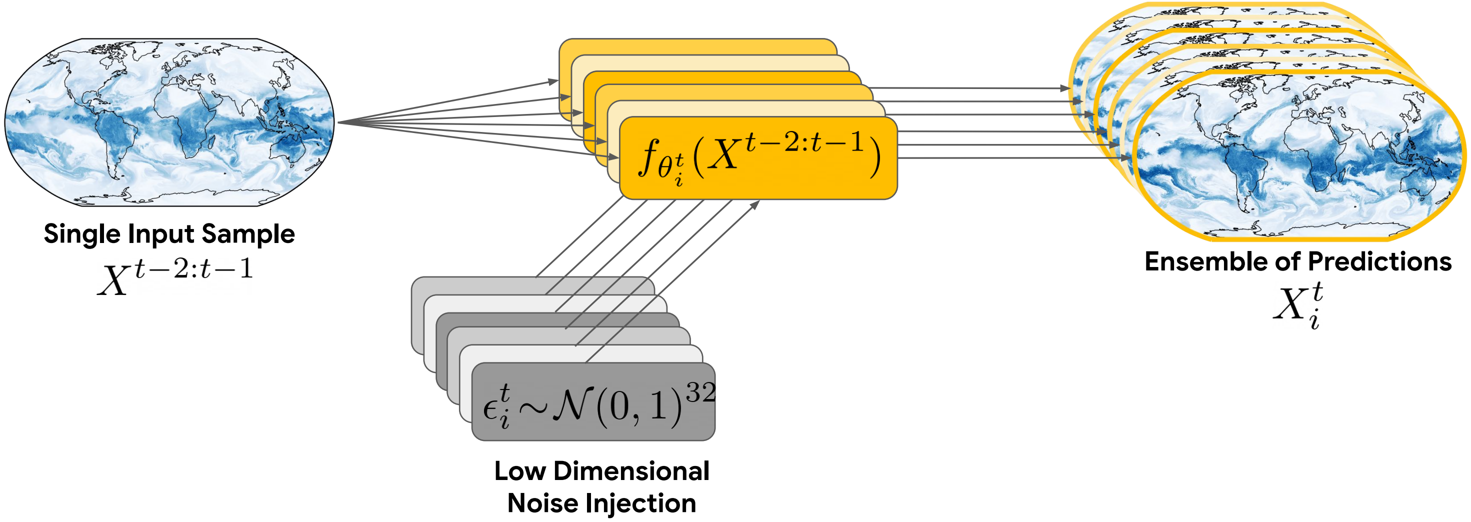
- Single pass 1AR prediction



Functional Generative Networks (FGN)

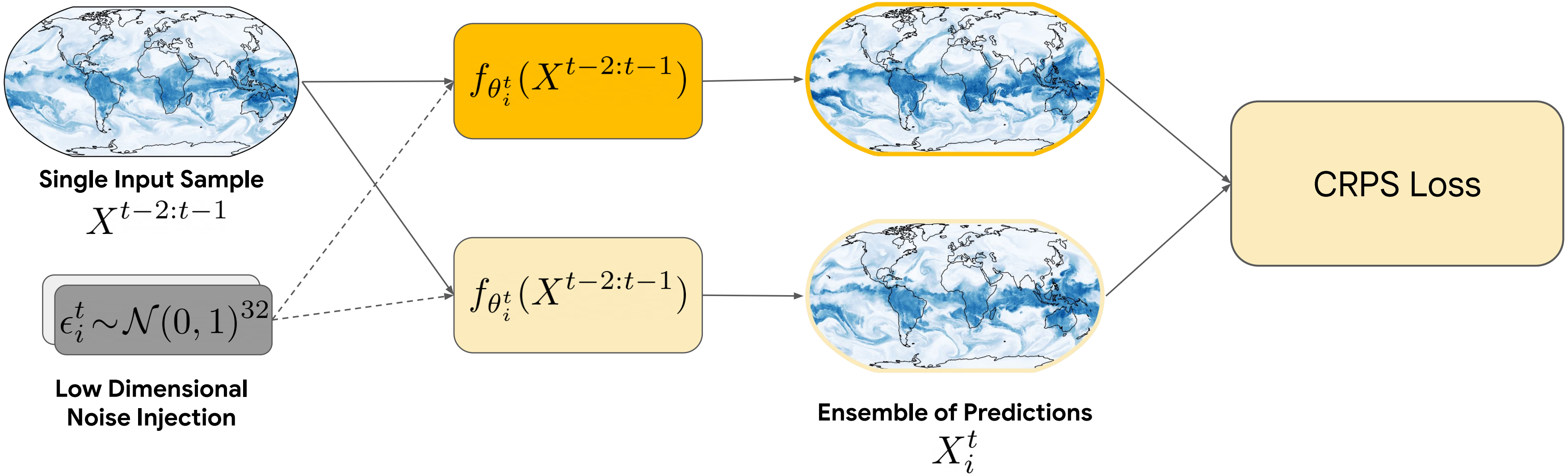
FGNs simpler than GenCast:

- Single pass 1AR prediction
- Low dimensional noise conditioning
    - All weather variability in 32 dimensions. Strange idea alert!



FGN Training

Training: Draw just **2** noise samples, same inputs, feed into CRPS [1,2,3].



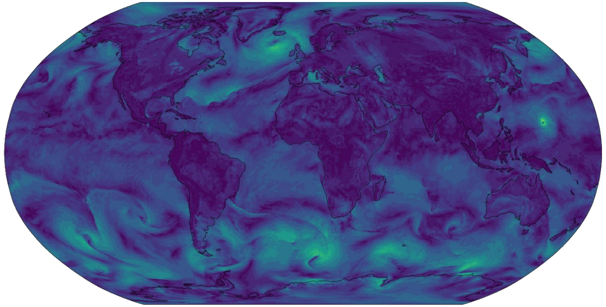
[1] D. Bouchacourt et al., **Disco nets: Dissimilarity coefficients networks**. Advances in Neural Information Processing Systems, (2016)
[2] D. Kochkov *et al.*, **Neural general circulation models for weather and climate**. *Nature* 632, (2024)
[3] *S. Lang e al.*, **AIFS-CRPS: Ensemble forecasting using a model trained with a loss function based on the Continuous Ranked Probability Score**. Arxiv, (2024)

Research surprise

Model is trained only on marginals, but generalizes to skillful joints

Extensive testing over the last months to corroborate finding

Marginals (pixels)
Humidity in London
Temperature in New York
Wind in Boston



Joints (global structure)
Wind power over all of the UK
Cold snap over the US
Cyclones; shared on Google Weather Lab

Marginals are enough!

- Same architecture and size as GenCast
- Marginal loss
  - Fair CRPS
  - 2 samples IID Gaussian noise
- No AR training

It would have already been SotA

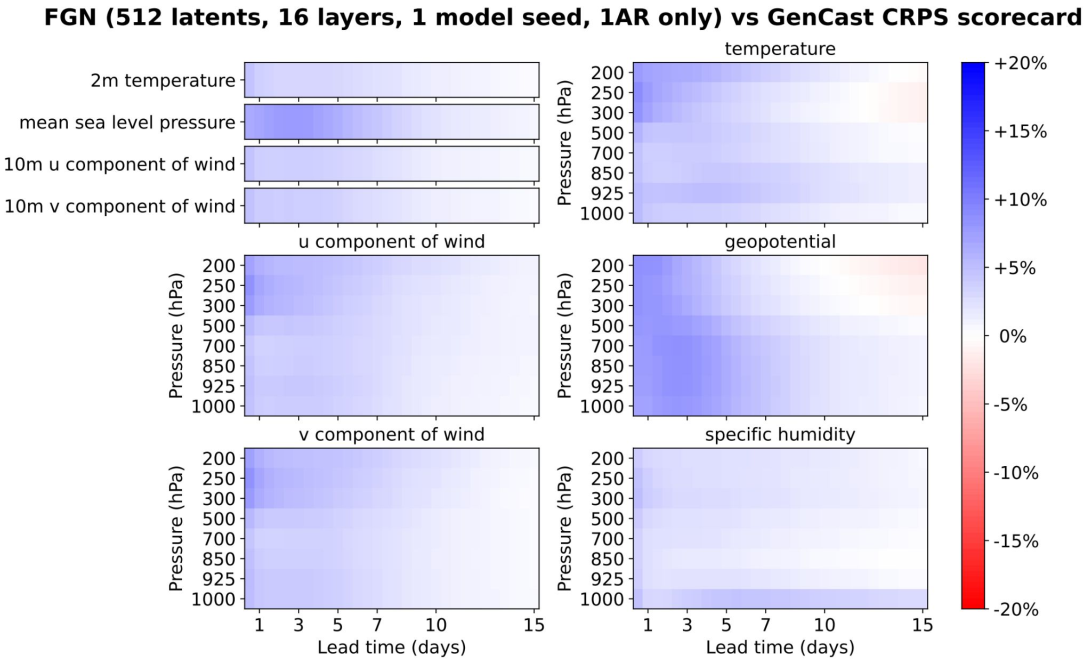


Figure A.1

Marginals are enough!

- Same architecture and size as GenCast
- Marginal loss
  - Fair CRPS
  - 2 samples IID Gaussian noise
- No AR training

It would have already been SotA

Skillful model with just marginals

Can't get much simpler than this!

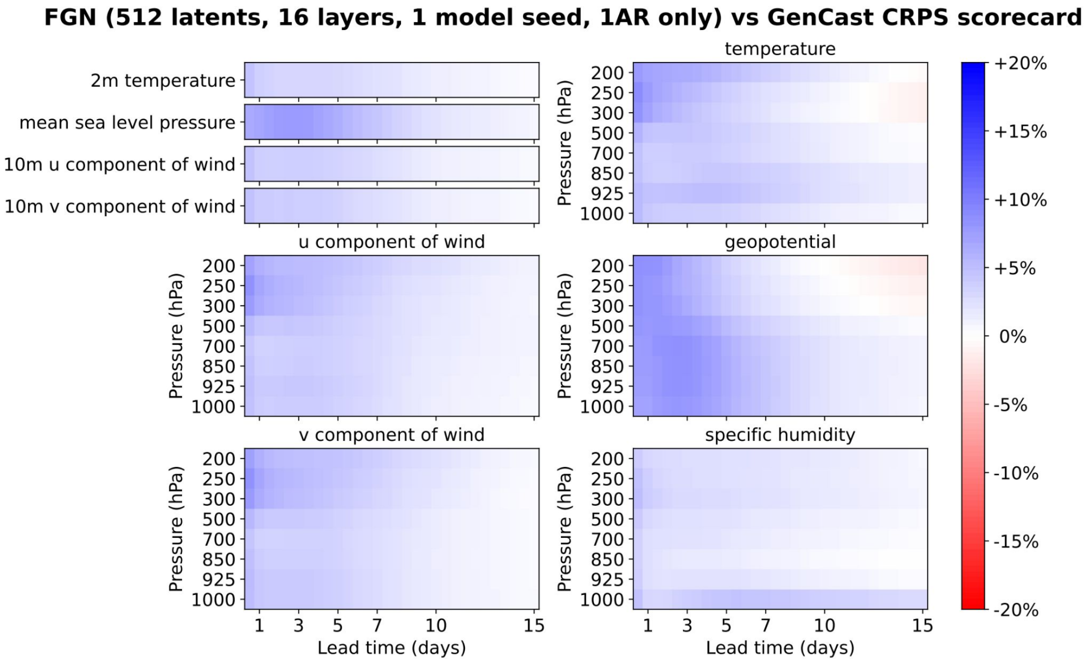
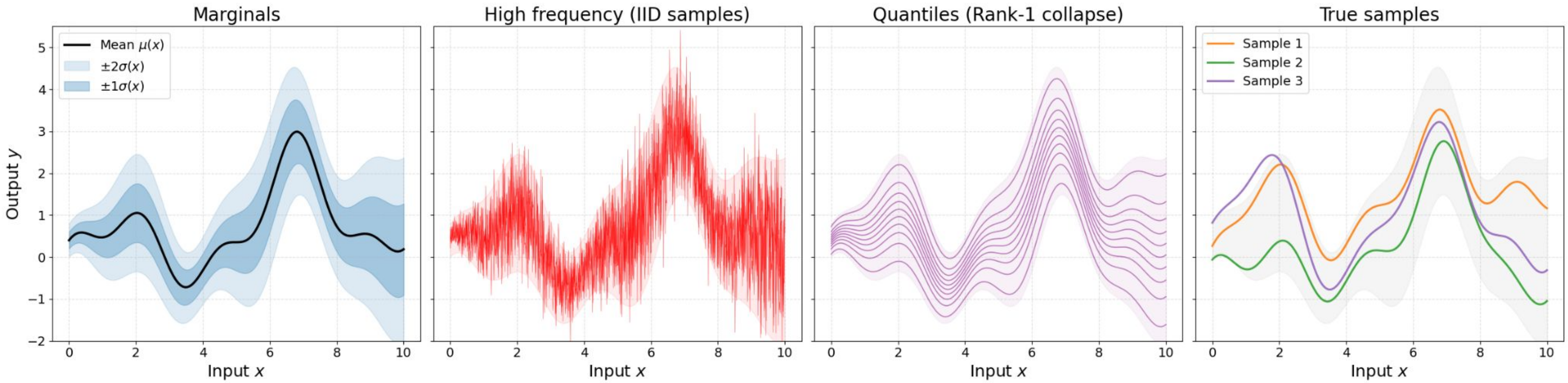


Figure A.1

This is *really* surprising: marginals should not be enough



Similar methods in ML literature provide[1,2] value, but not SotA high-dimensional joints

What is special about weather that makes CRPS variants[3,4,5] so effective?

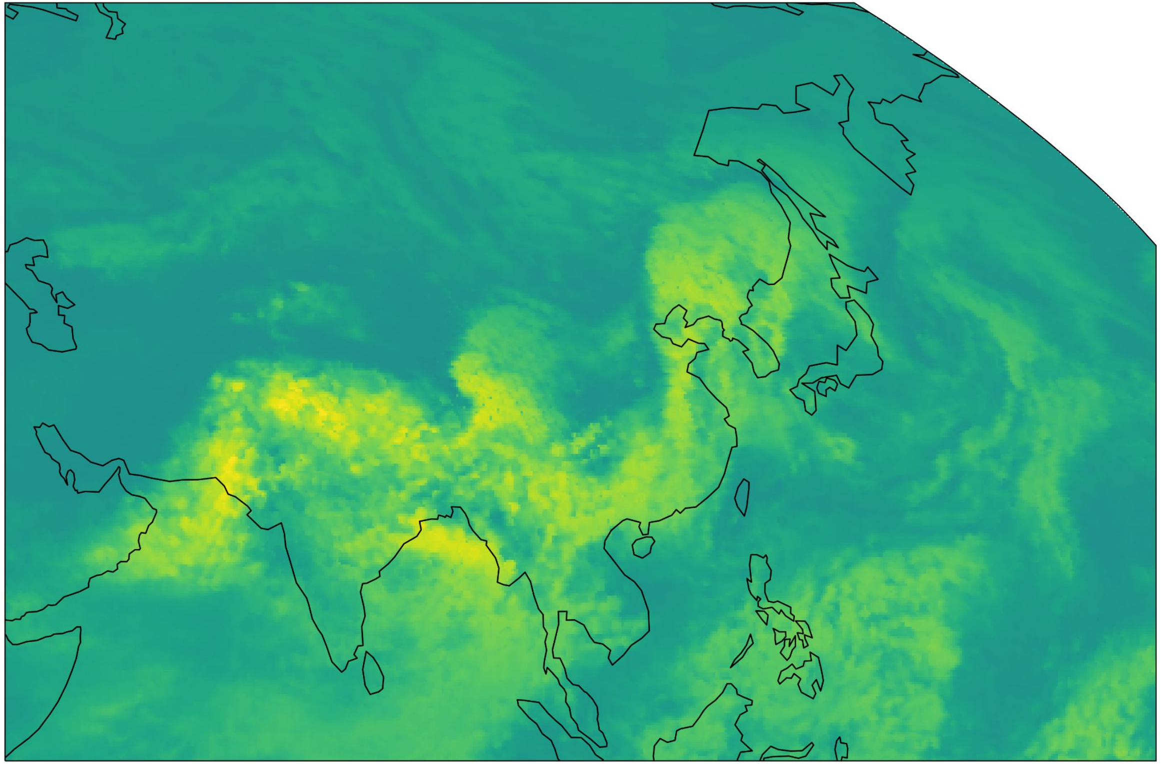
Are there generalizable insights to other domains?

[1] Gretton et al. A Kernel method for two-sample-problem
[2] Bouchacourt et al. DISCO Nets

[3] Kochkov et al. NeuralGCM
[4] Lang et al. AIFS-CRPS
[5] Bonev et al. FourCastNet 3 …

Marginals are *almost* enough

Sample 0, median-CRPS forecast of q300 at 15 day lead time

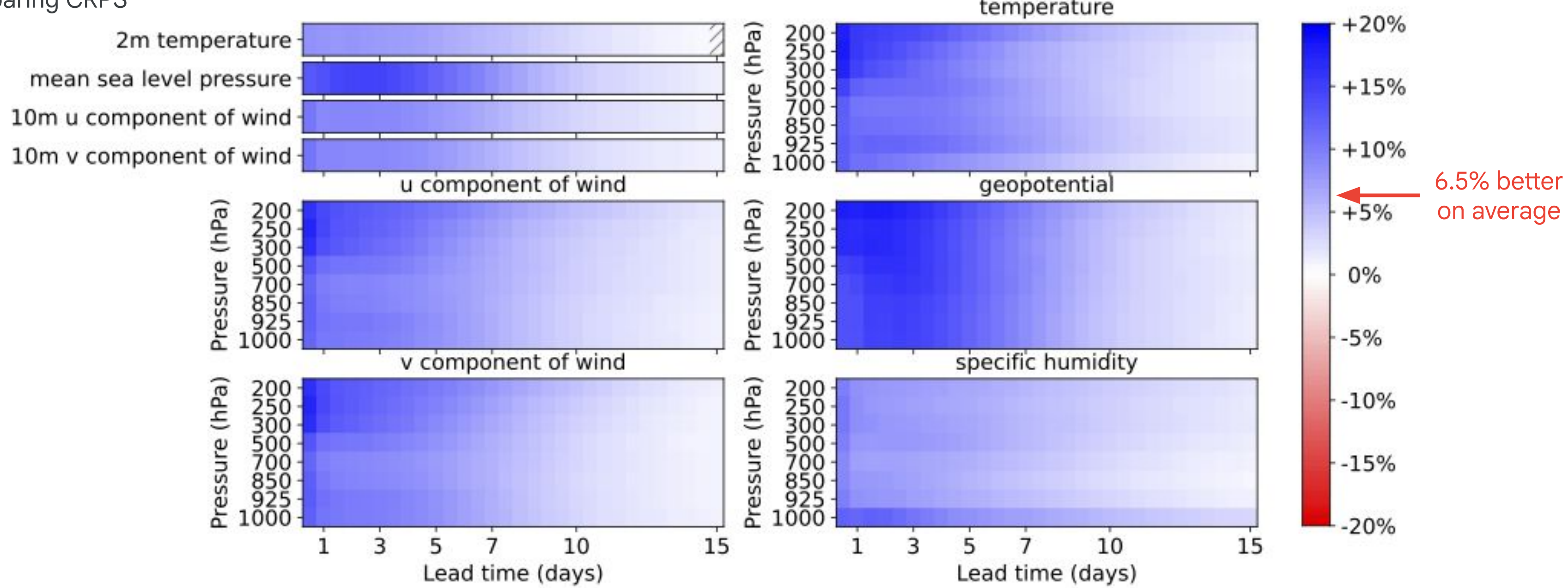


- Artifacts in different variables
- **Goodhart's Law?**

- Often *improved* with skill
- More analysis needed

FGN vs GenCast

Comparing CRPS



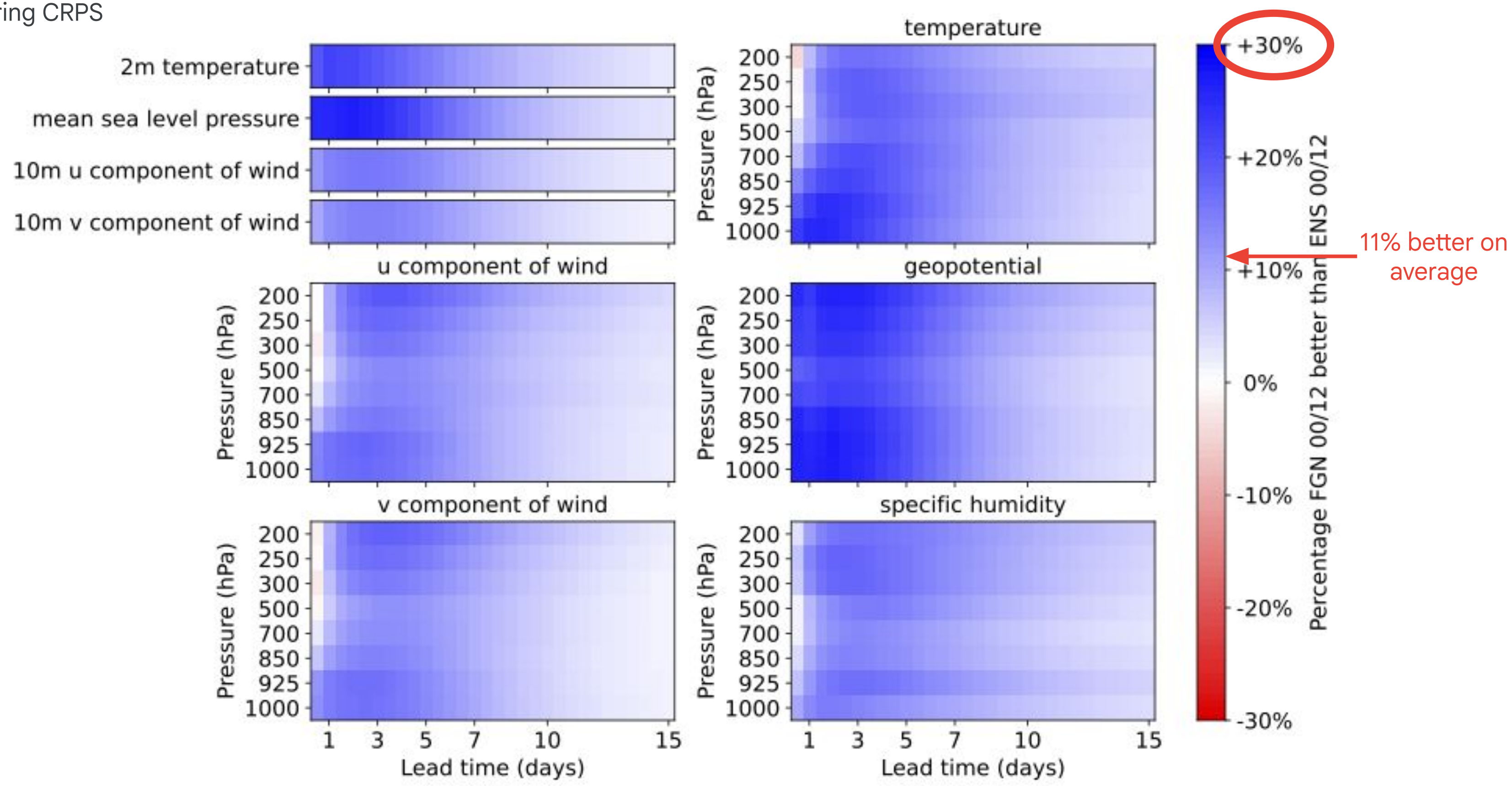
Note: The FGN model has a number of other changes to GenCast:

- 3.2x the number of parameters
- Ensembles 4 independently trained models
- Does 8 step autoregessive training
- 6hr rather than 12hr timesteps

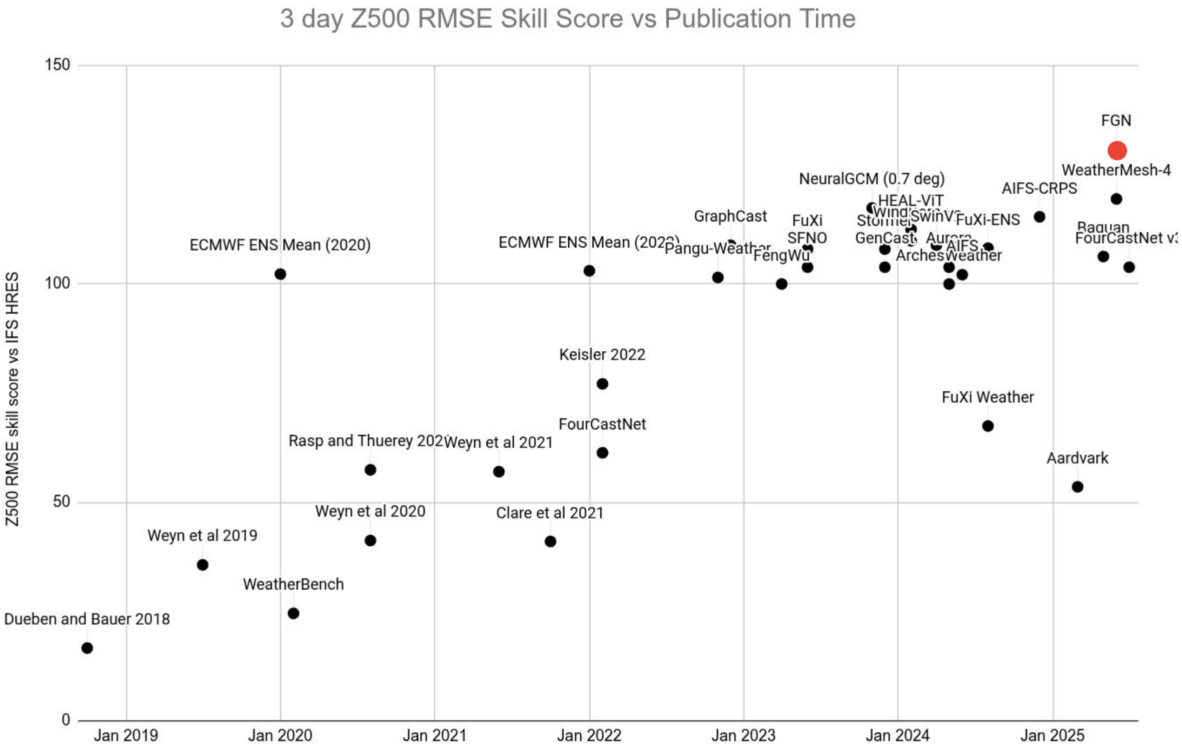
However a clean ablation of the methods is included in the paper which shows that even without these changes FGN is performs better than GenCast.

FGN vs ENS (Best NWP baseline)

Comparing CRPS



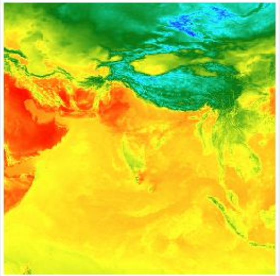
Current SoTA worldwide



FGN is running live

Forecasts freely available as “WeatherNext 2” on Earth Engine and BigQuery

WeatherNext 2



This dataset is part of a Publisher Catalog, and not managed by Google Earth Engine. Contact weathernext@google.com for bugs or view more datasets from the WeatherNext Catalog. Learn more about Publisher datasets.

Catalog Owner
WeatherNext

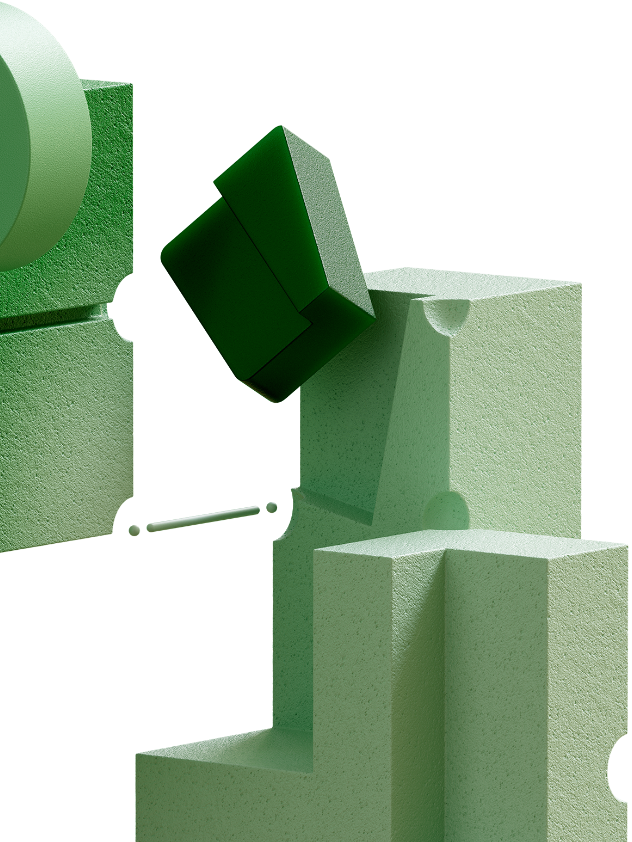
Dataset Availability
2022-01-01T00:00:00Z–2025-12-06T06:00:00Z

Dataset Provider
Google

Tags
climate forecast gcp-public-data-weathernext precipitation publisher-dataset temperature weather weathernext wind cyclones fgn

Google DeepMind

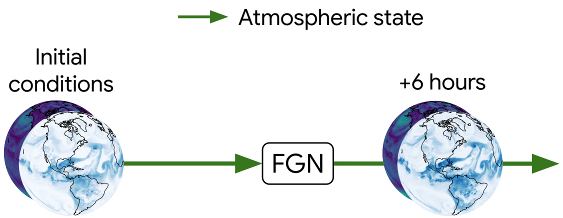
Cyclones



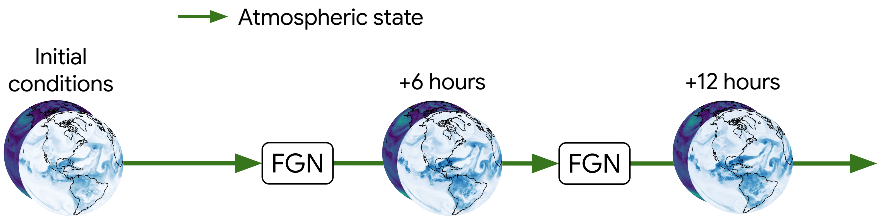
Tropical Cyclones

- Cyclones are particularly devastating
- Significant cyclones > $10 billion damage
- Hurricane Helene (2024)
  - $79 billion damage
  - Killed 150 people
- Predict cyclone trajectory and intensity.
  - Significantly improve disaster preparedness
  - Mitigate some impacts

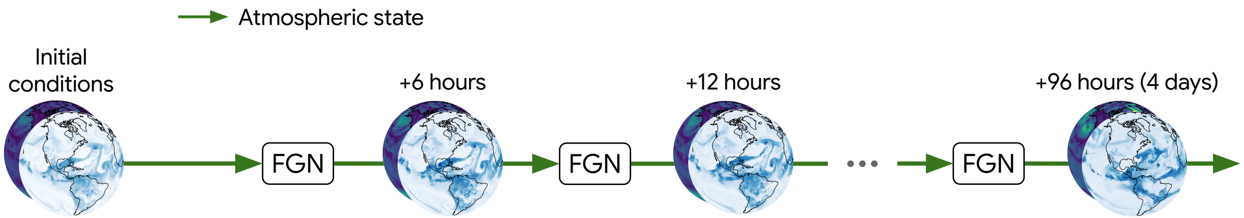
Issues with analysis-based cyclone predictions



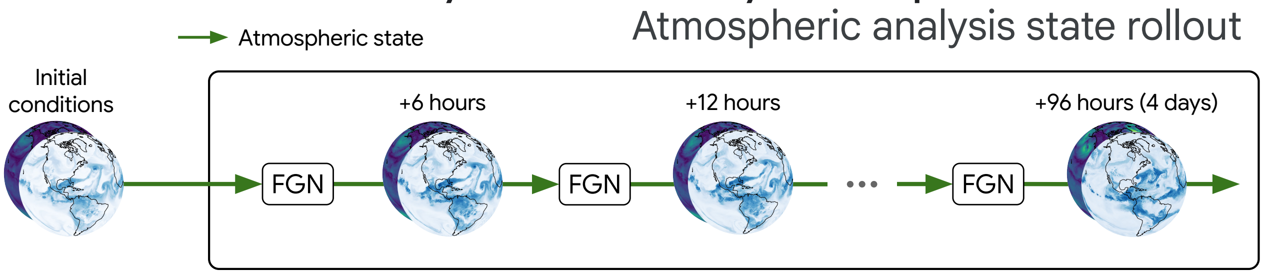
Issues with analysis-based cyclone predictions



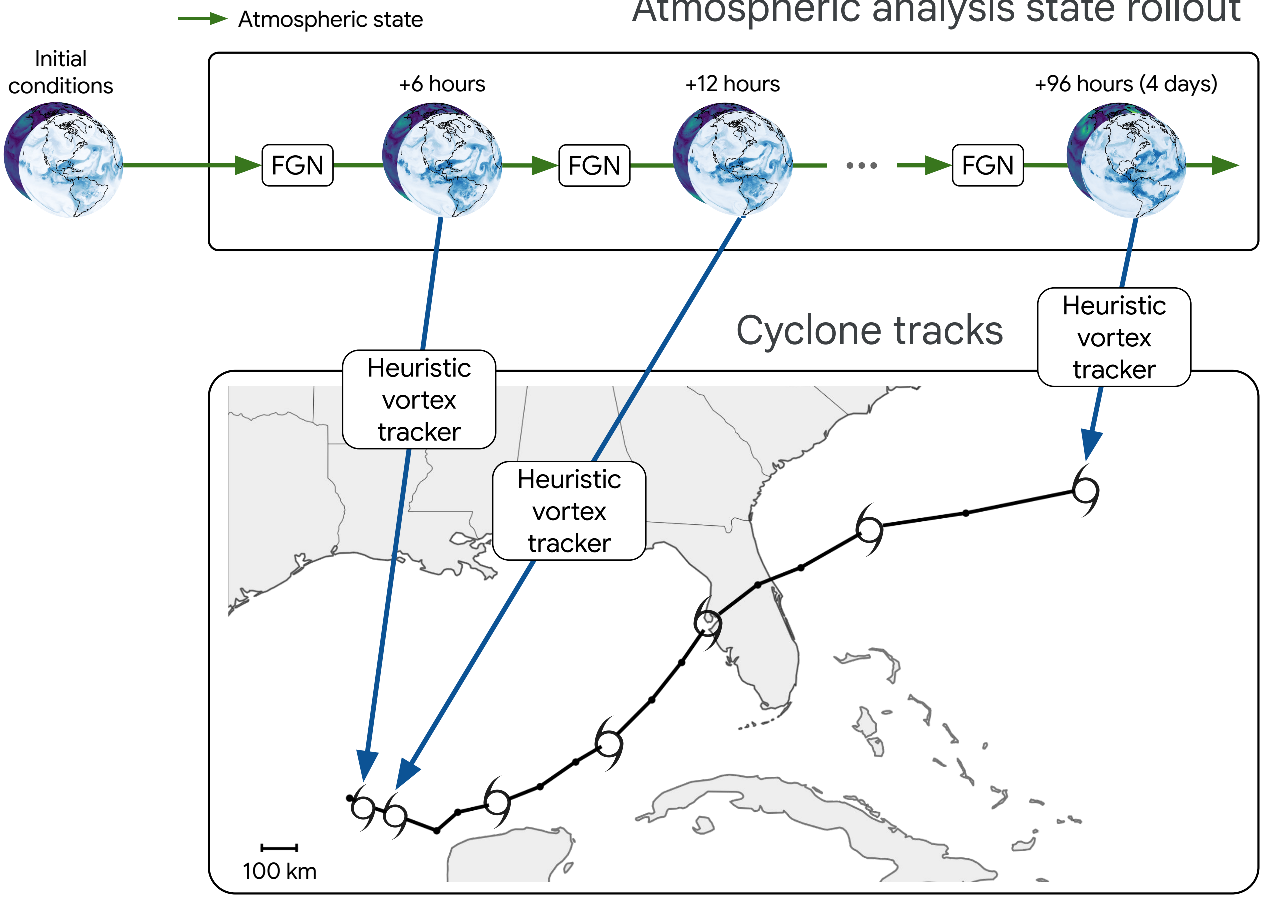
Issues with analysis-based cyclone predictions



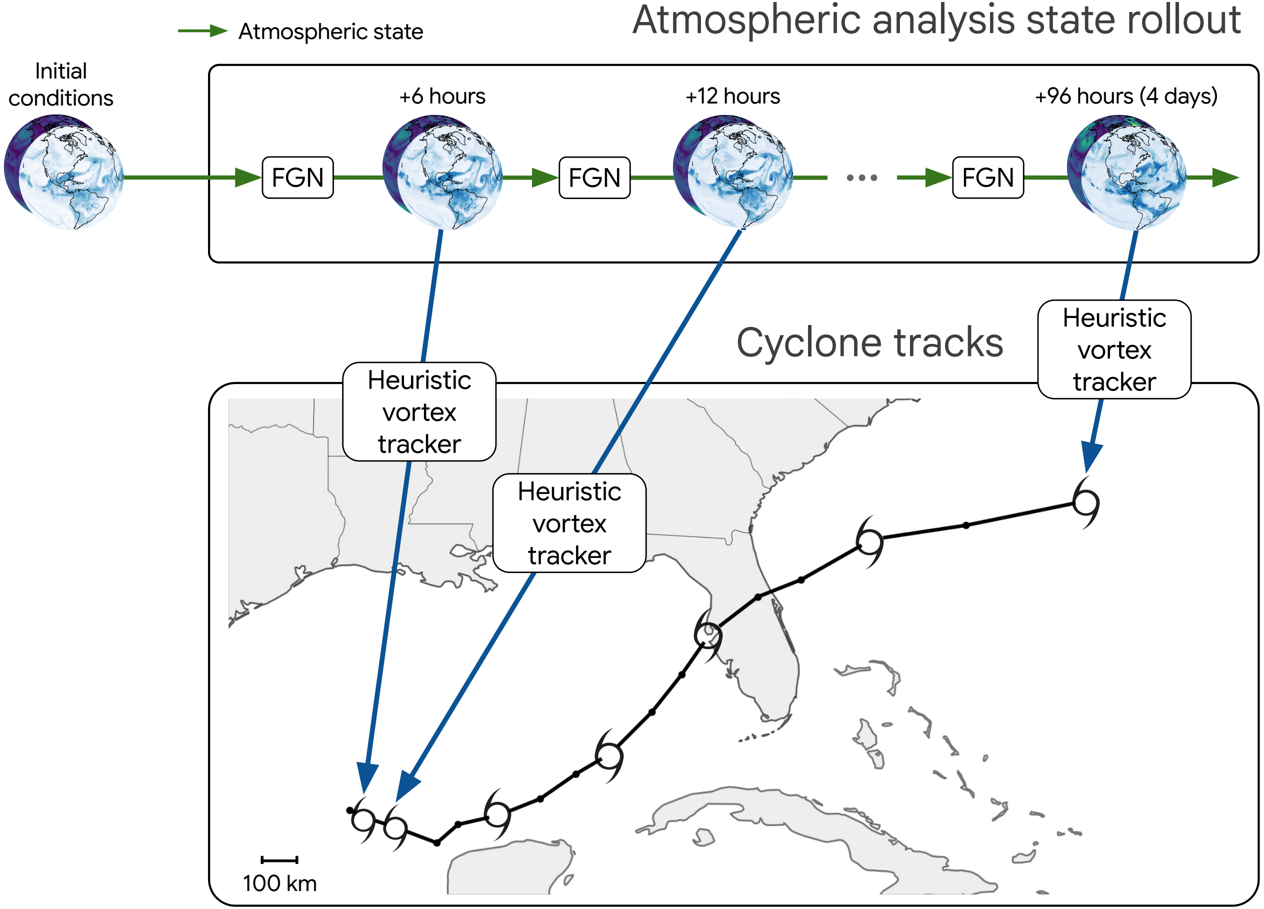
Issues with analysis-based cyclone predictions



Issues with analysis-based cyclone predictions



Issues with analysis-based cyclone predictions



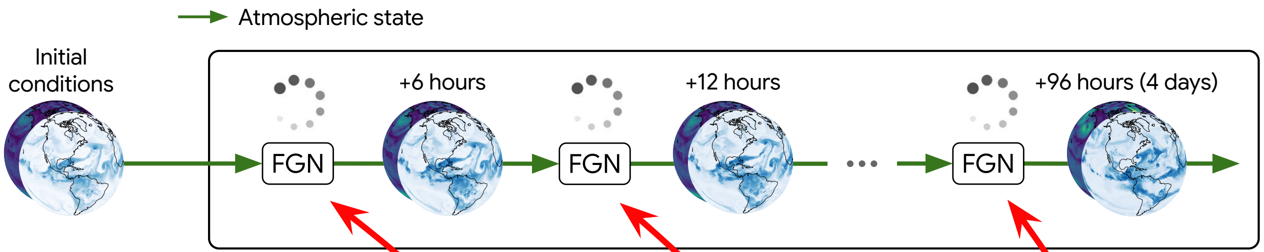
Issues with traditional vortex trackers:

❌ Surface winds negatively biased (at 0.25°)

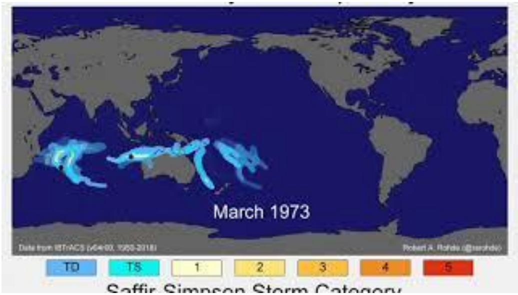
❌ Cyclone size (wind radii) negatively biased

❌ Poor rapid intensification recall (true positive rate)

End-to-end AI cyclone predictions: 1) Training

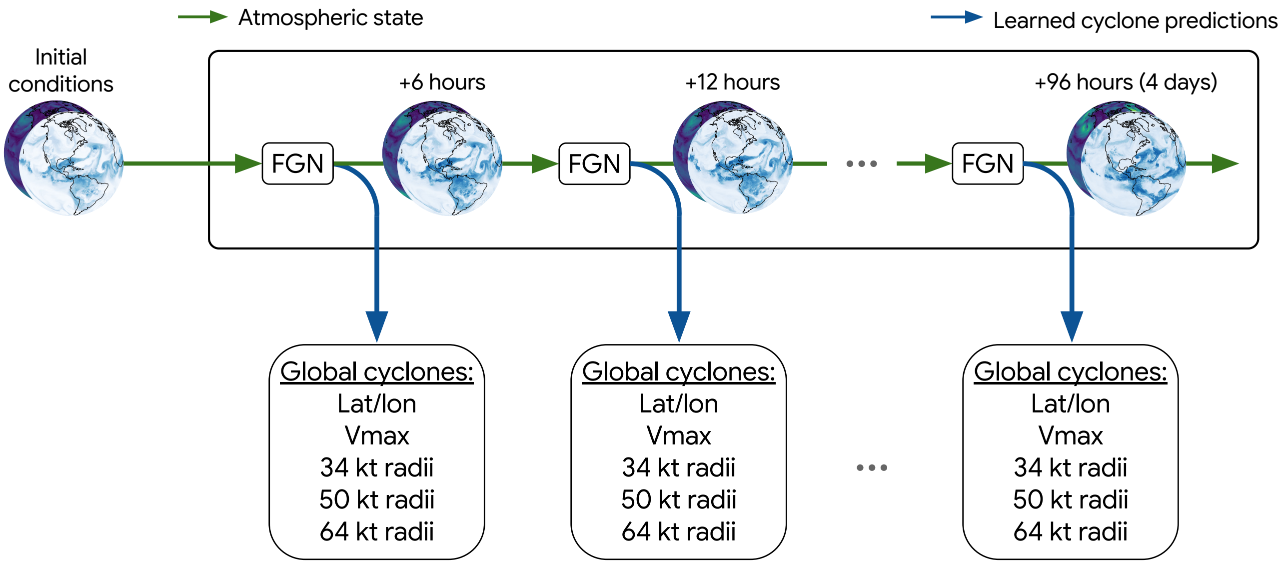


IBTrACS (1979-2021)

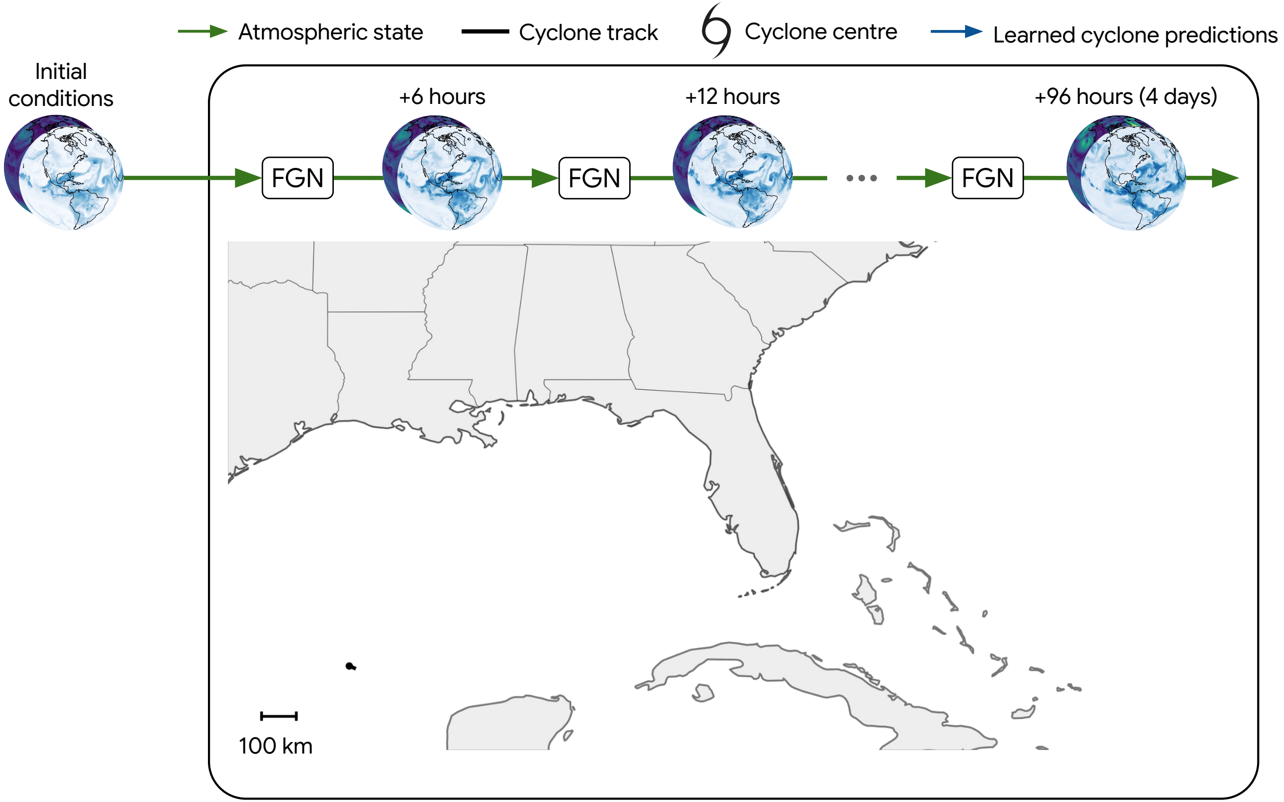


Solution: Train FGN directly on cyclone observations

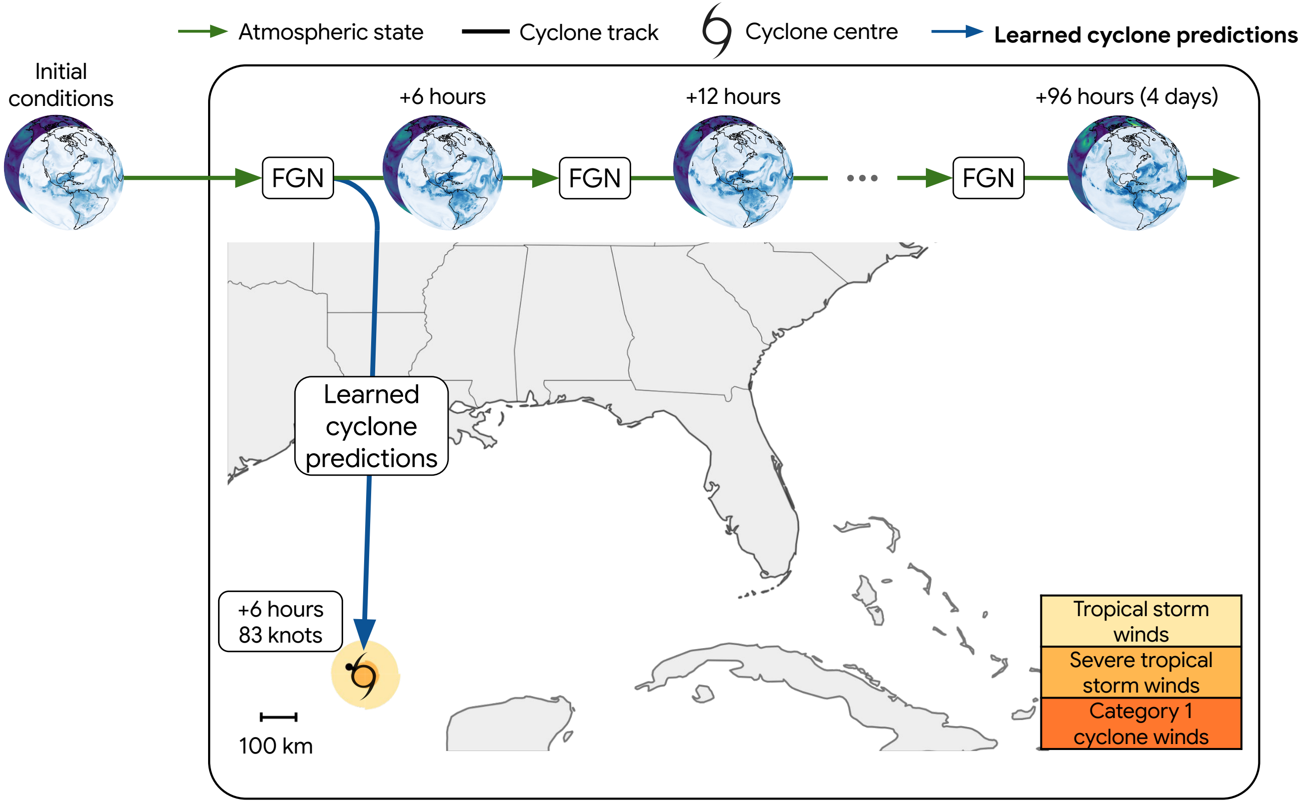
End-to-end AI cyclone predictions: 2) Inference



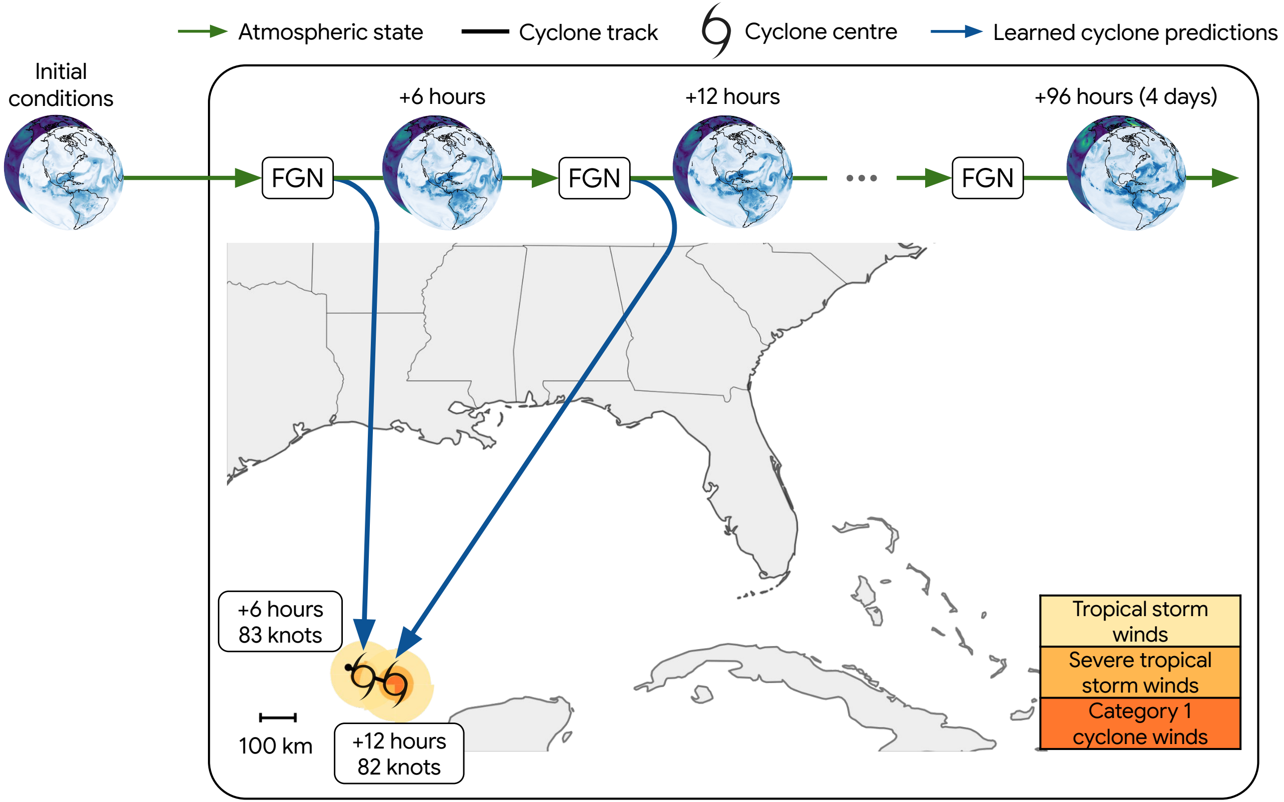
End-to-end AI cyclone predictions: 3) Tracking



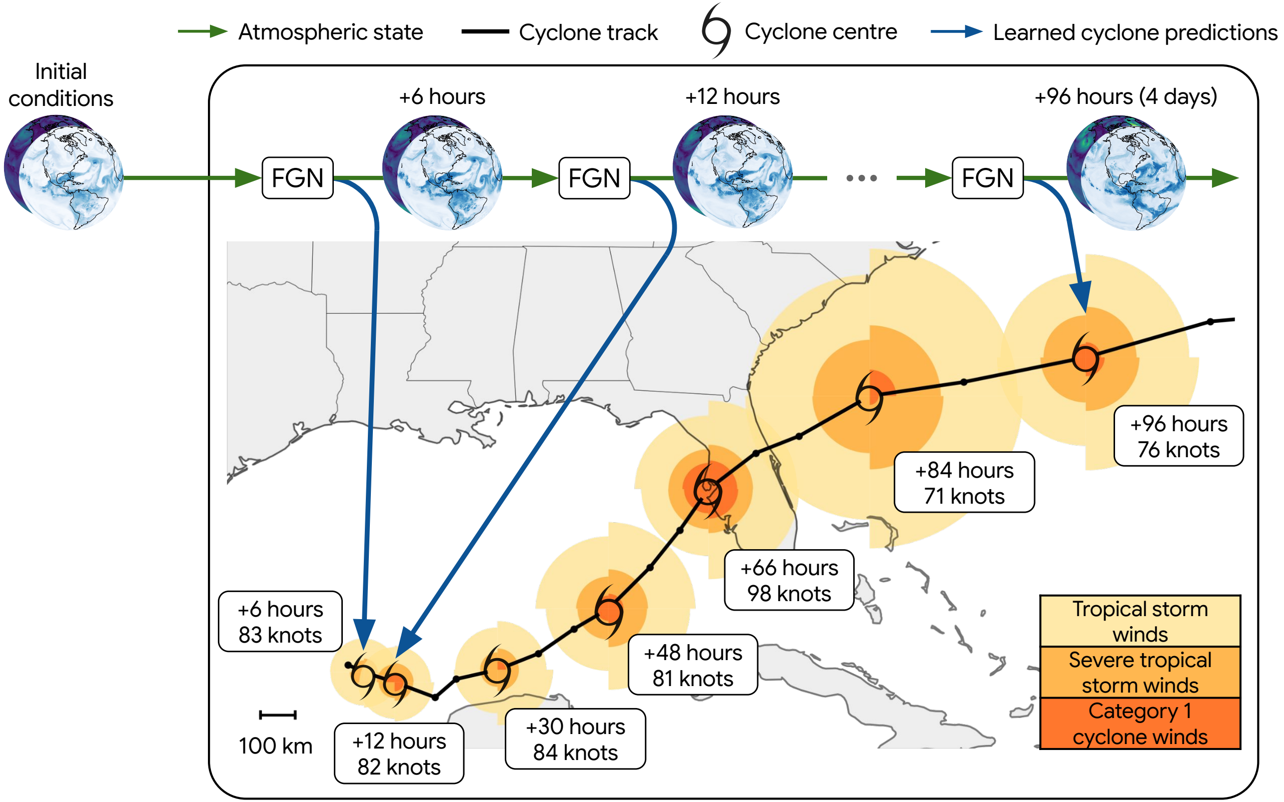
End-to-end AI cyclone predictions: 3) Tracking



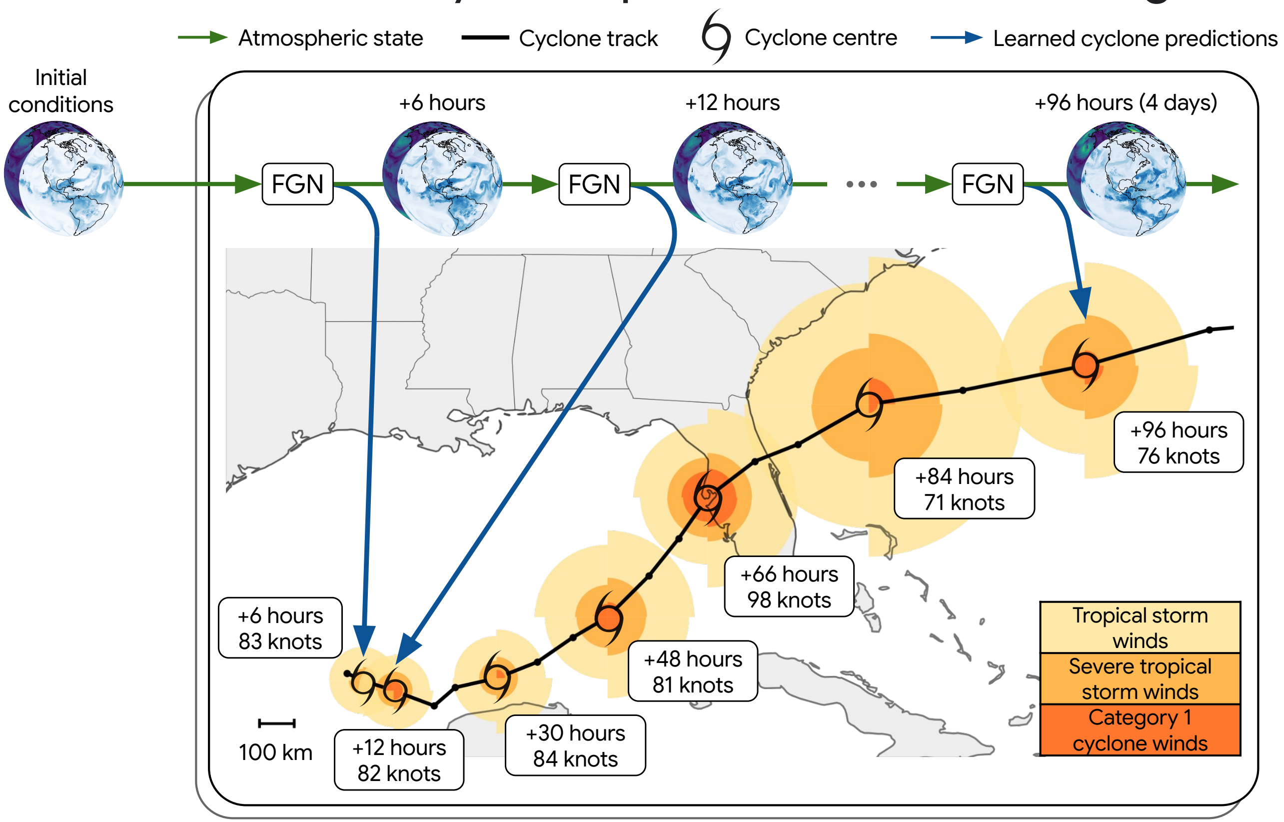
End-to-end AI cyclone predictions: 3) Tracking



End-to-end AI cyclone predictions: 3) Tracking

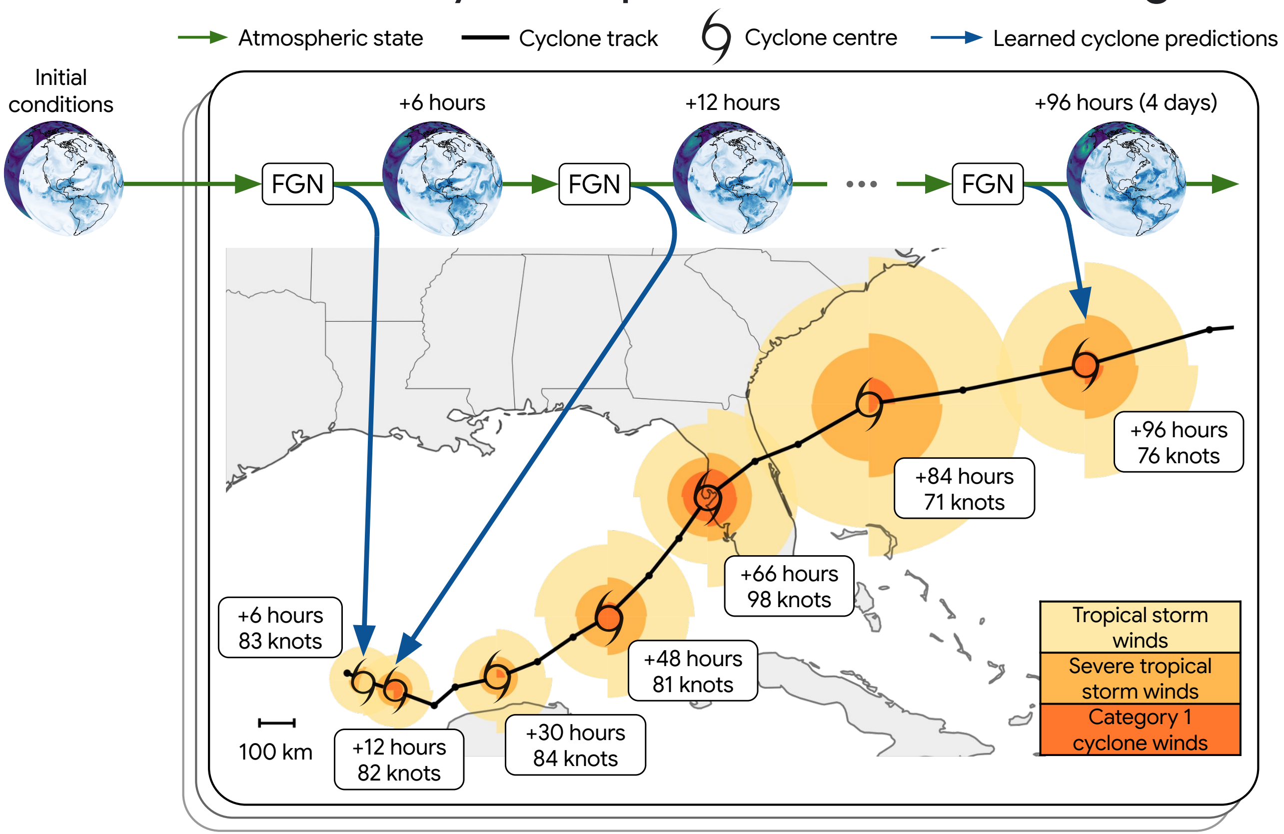


End-to-end AI cyclone predictions: 3) Tracking



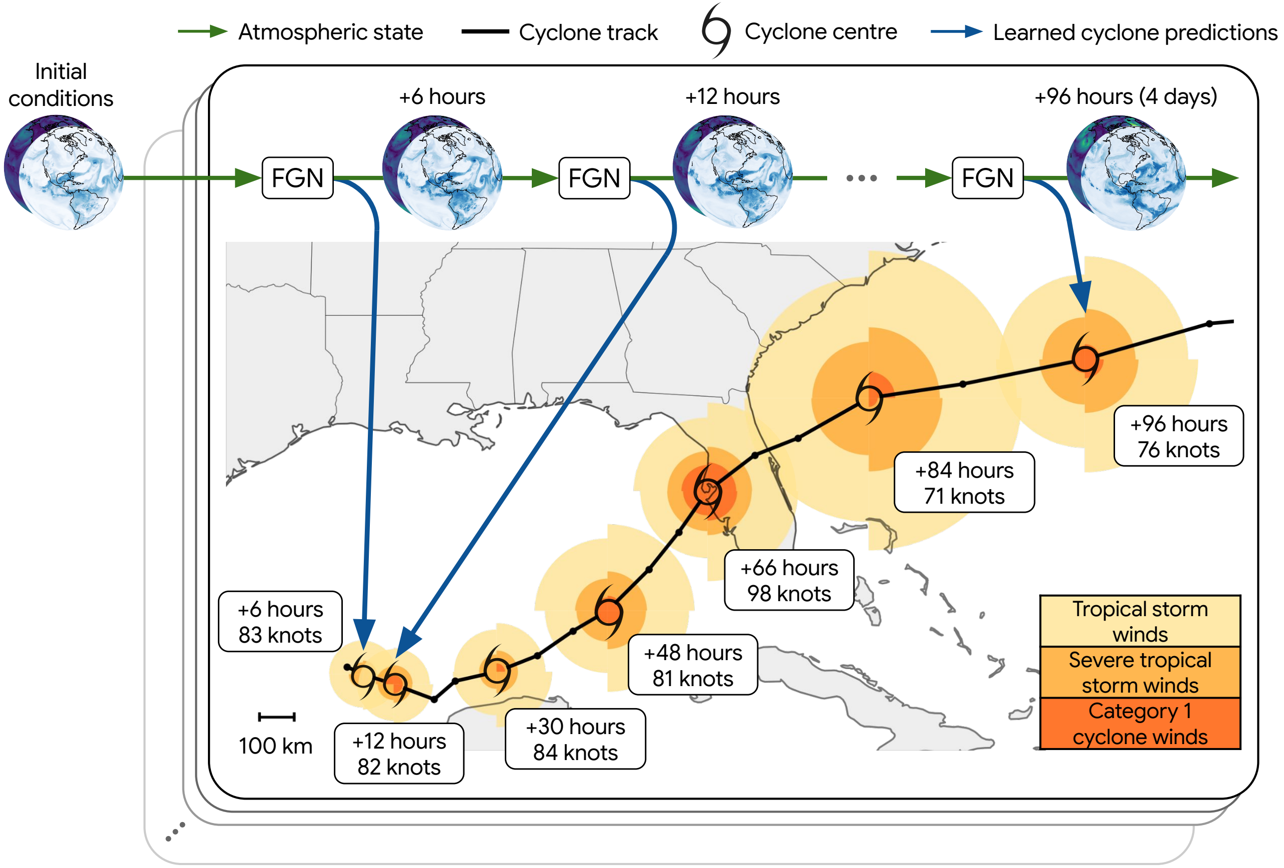
N=2 ensemble members

End-to-end AI cyclone predictions: 3) Tracking



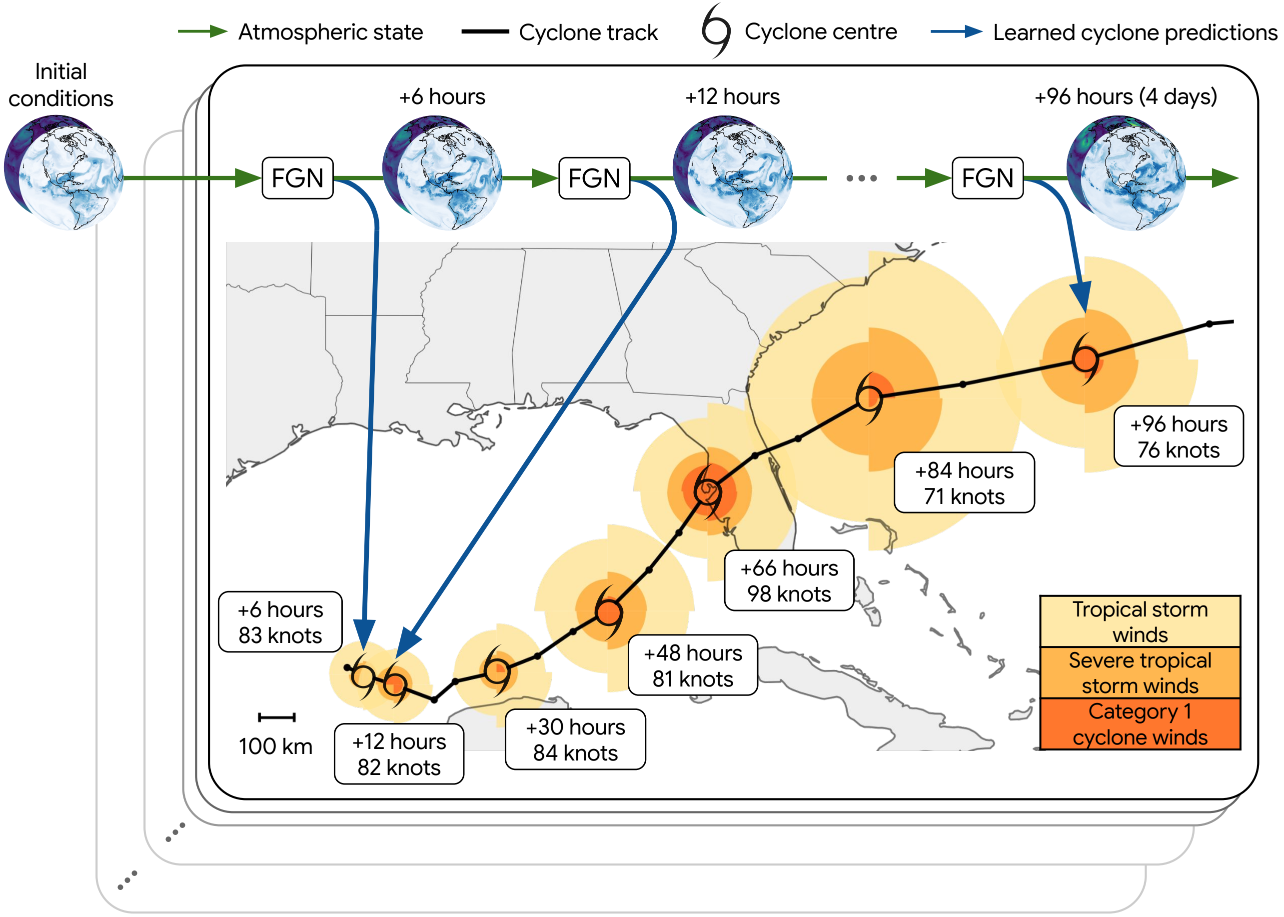
N=3 ensemble members

End-to-end AI cyclone predictions: 3) Tracking



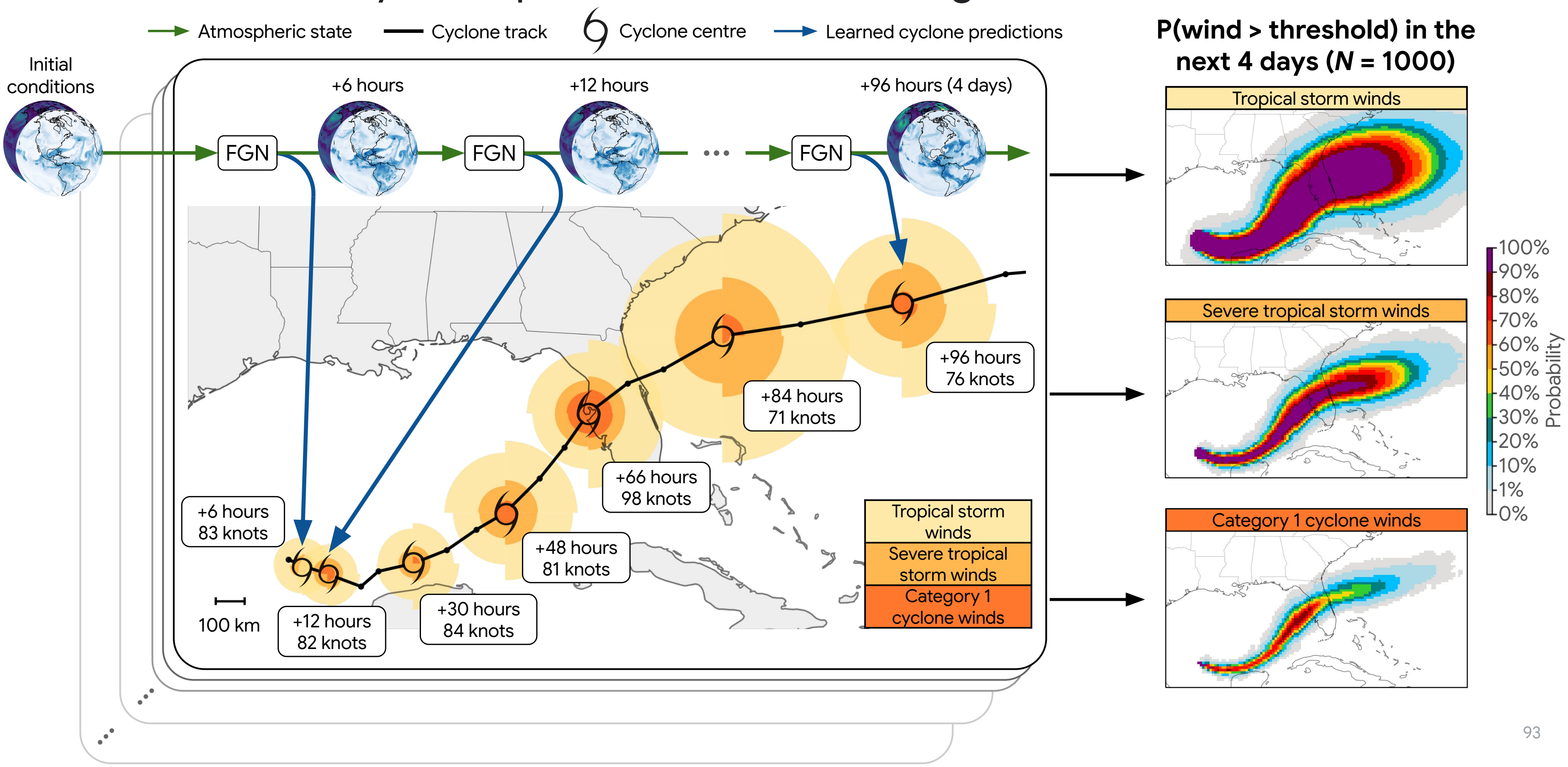
N=50 ensemble members

End-to-end AI cyclone predictions: 3) Tracking

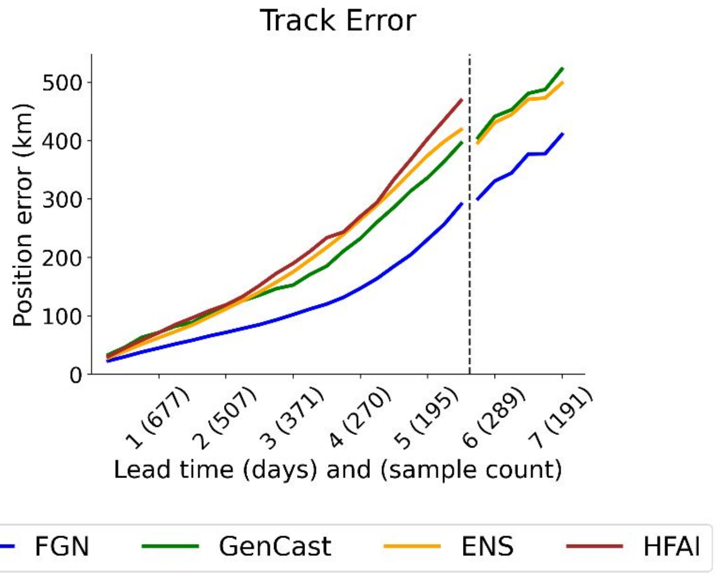


N=1000 ensemble members

End-to-end AI cyclone predictions: 3) Tracking



State-of-the-art cyclone forecasting (2023-2024)

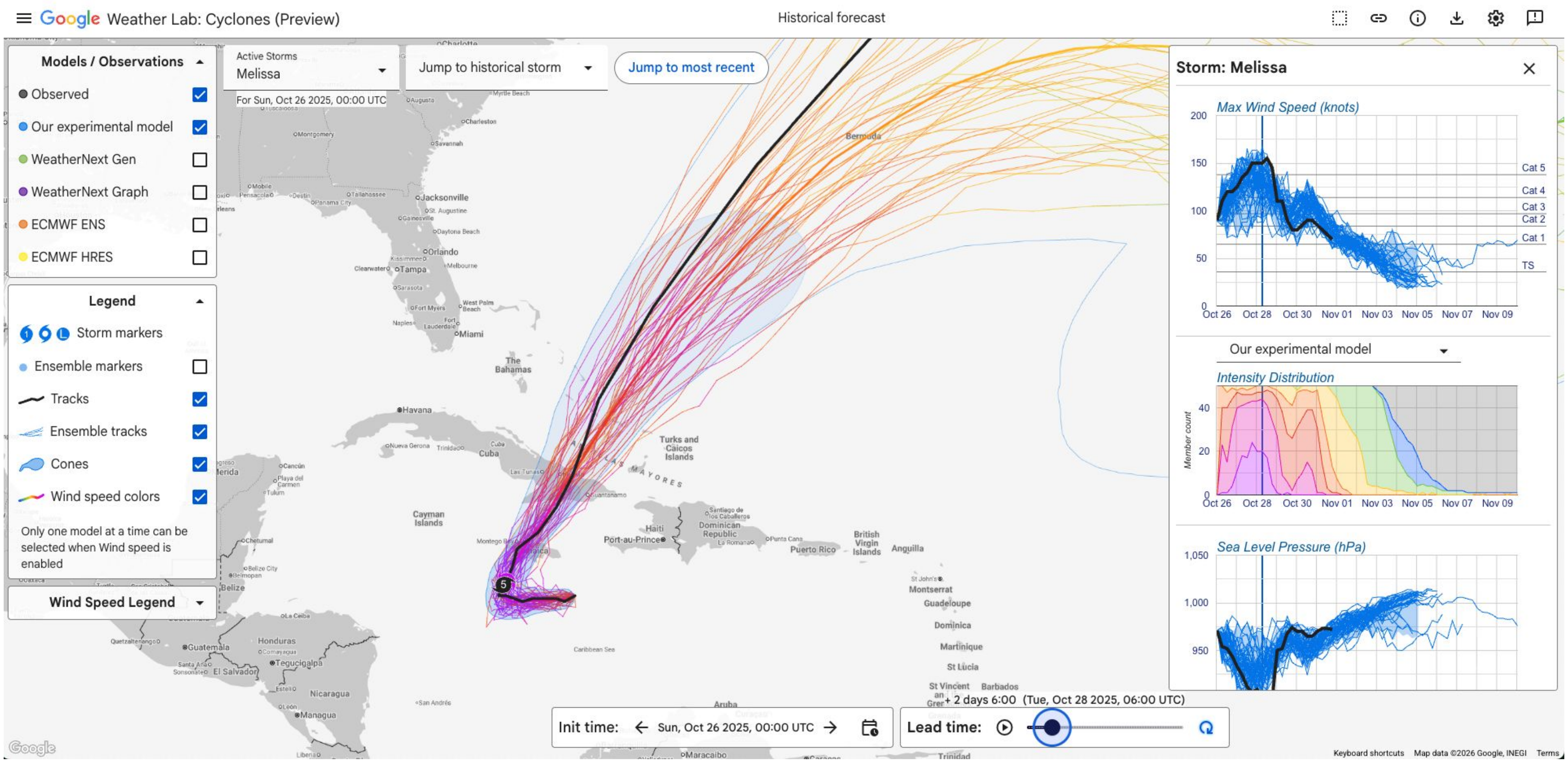


- First AI model to be good (great!) at intensities
- Best single model at tracks, intensities, and structure; ~1 decade of progress


Google Weather Lab: Cyclones (Preview)
Models / Observations
Observed
Our experimental model
WeatherNext Gen
WeatherNext Graph
ECMWF ENS
ECMWF HRES
Legend
Storm markers
Ensemble markers
Tracks
Ensemble tracks
Cones
Wind speed colors
Only one model at a time can be selected when Wind speed is enabled
Wind Speed Legend
Active Storms
Jump to historical storm
For Thu, Mar 5 2026, 00:00 UTC
Indian Ocean
North Pacific Ocean
South Pacific Ocean
Australia
+ 2 days 18:00 (Sat, Mar 7 2026, 18:00 UTC)
Init time:
Thu, Mar 5 2026, 00:00 UTC
Lead time:

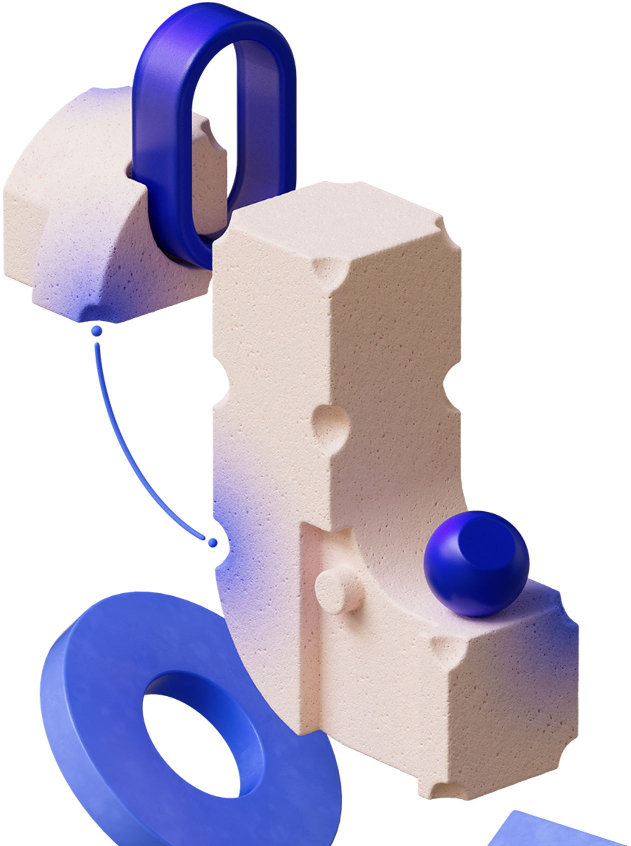
Live predictions on Weather Lab during 2025

https://deepmind.google.com/science/weatherlab



Google DeepMind

Broader Implications



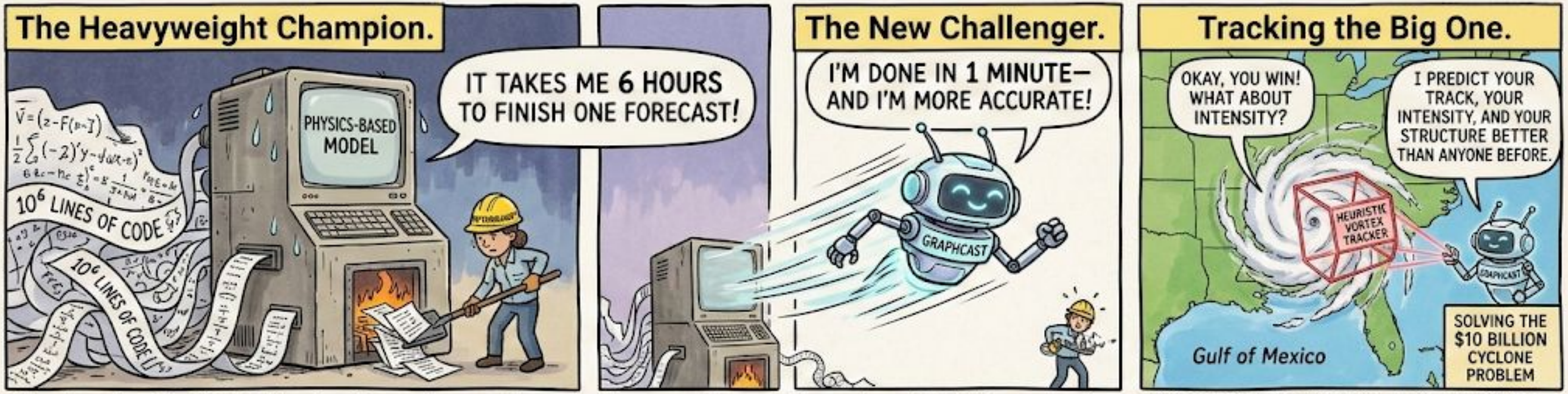
Our learnings

- We are *not*
  - Meteorologists
  - CFD folks
  - Numerical methods analysts
- Some of us never studied machine learning!

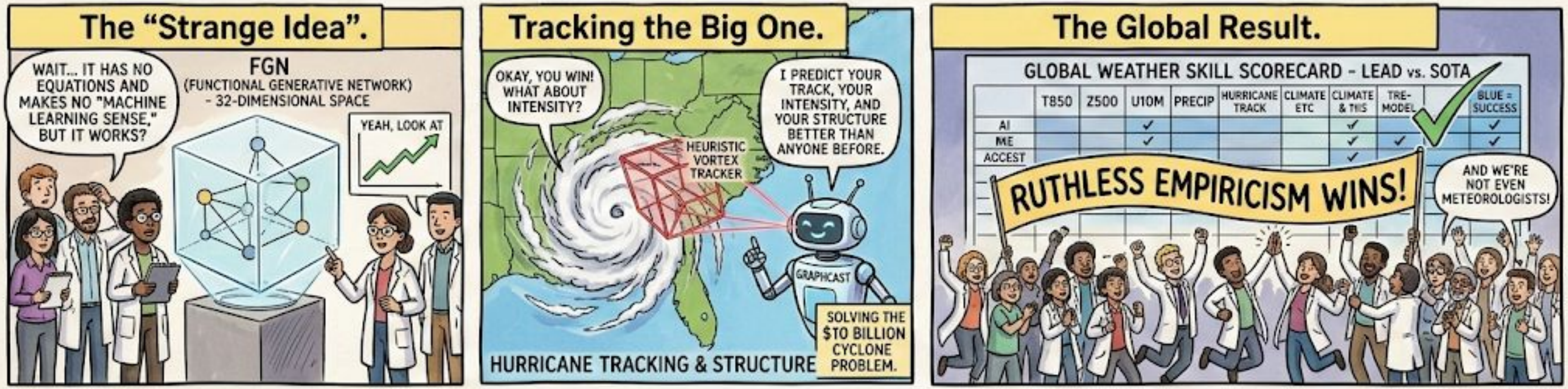
- Our best model makes little (machine learning) sense!

- CTF principles
  - Externally available datasets
  - Competition (other labs)
  - External validation (forecasts speak for themselves)

- Data over dogma
  - Nowadays theory is cheap, experiments are cheaper
  - Neural scaling laws ⟶ “Age of architectures is over”
  - Guarantees necessary? (Cyclones model deployed!)
  - Evals are king 👑



For decades, the "Gold Standard" was slow and heavy. AI-powered forecasting arrives with 10^5x acceleration. HURRICANE TRACKING & STRUCTURE.



Data over Dogma: If the evals are king, the model stays. 'Ruthless Empiricism' leads to a new era of weather science.

Google DeepMind