

Weather Forecasting with Deep Learning

A paradigm shift

Lasse Espeholt

Google Research



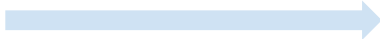
My Background



2009



2015



DeepMind

2018

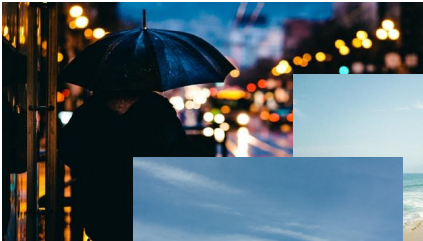


Google Brain



Angel

Purpose of Weather Forecasting



MetNet, a Neural Weather Model

“MetNet-2 outperforms the state-of-the-art physics-based ensemble model HREF for weather forecasts up to 12 hours ahead.”



Joint work within Google Research



Nal Kalchbrenner



Marcin Andrychowicz



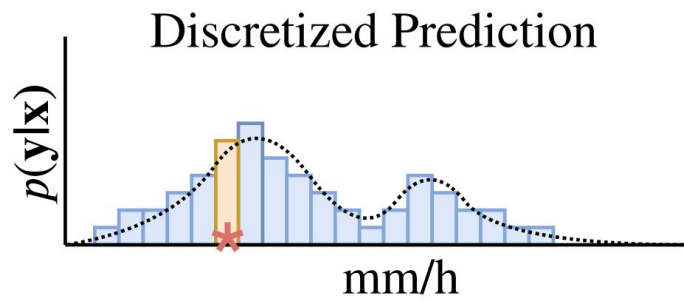
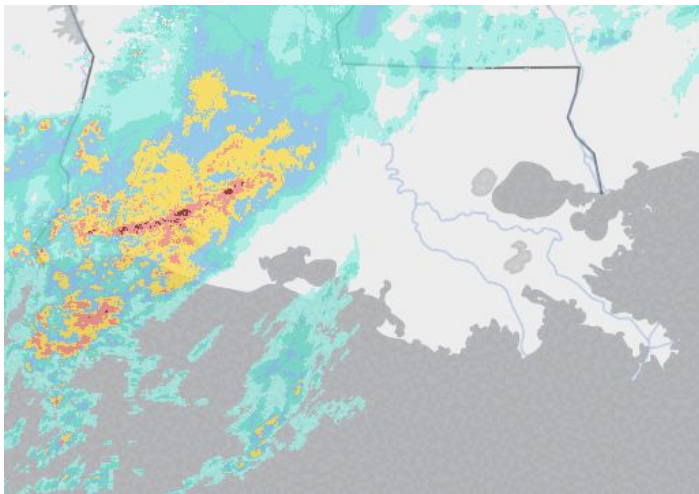
Shreya Agrawal



Casper Sønderby

Manoj Kumar, Jonathan Heek, Carla Bromberg, Cenk Gizen, Jason Hickey, Aaron Bell, Zack Ontiveros, Samier Merchant, Fred Zyda and others.

What is the probability of a given amount of precipitation (rain, snow, hail) occurring at a specific location and time?



Numerical Weather Prediction (NWP)

Majority of weather models are based on NWP. Other approaches include optical flow.

Simulation based.

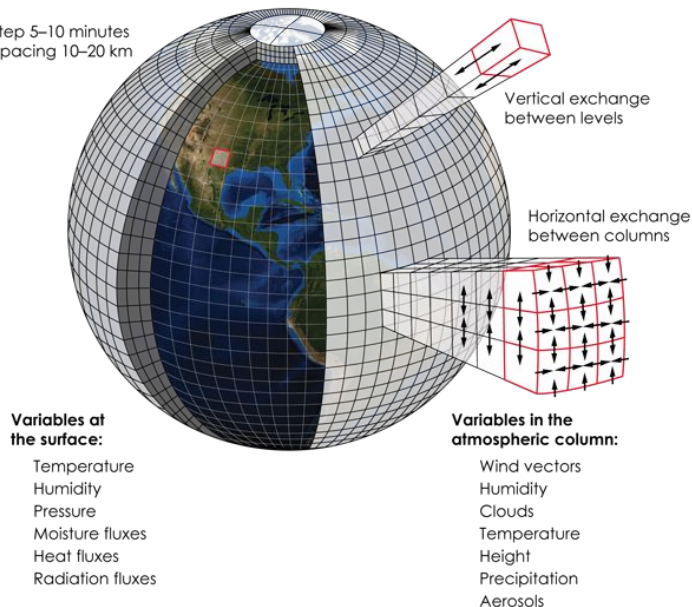
NOAA¹ HRRR² is primarily used for comparison in this work.
NWP for short-medium forecasts on continental USA.

¹ National Oceanic and Atmospheric Administration

² High Resolution Rapid Refresh (HRRR)

Weather forecast modeling

Timestep 5–10 minutes
Grid spacing 10–20 km



Credit: K. Cantner, AGI.

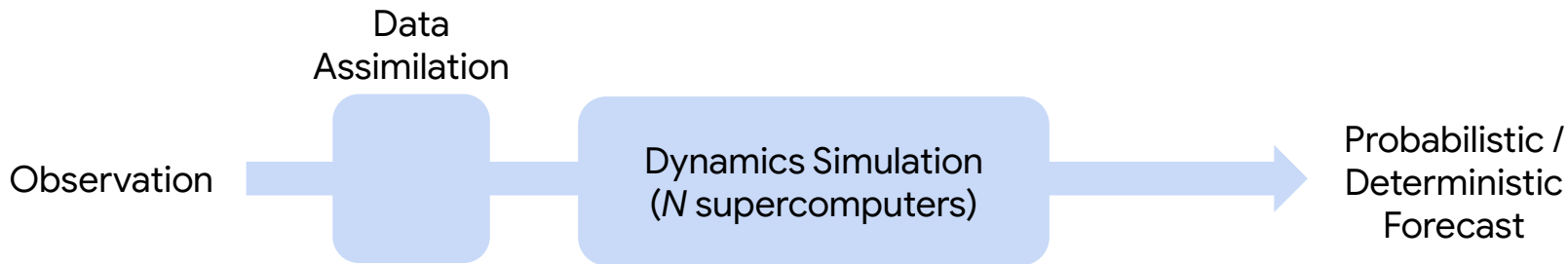
Numerical Weather Prediction (NWP) vs. MetNet-style



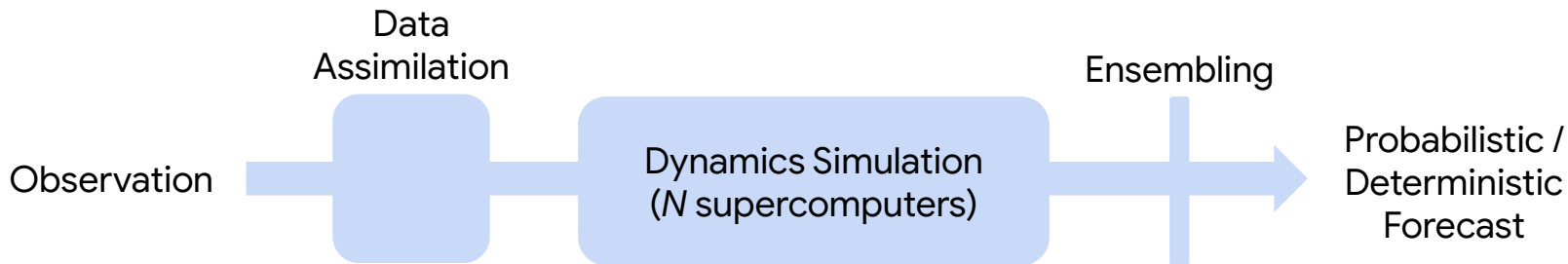
Numerical Weather Prediction (NWP) vs. MetNet-style



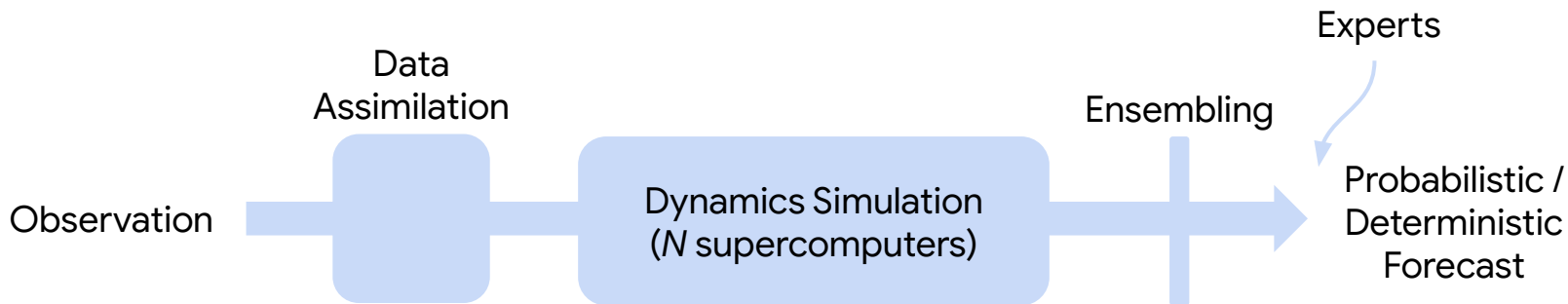
Numerical Weather Prediction (NWP) vs. MetNet-style



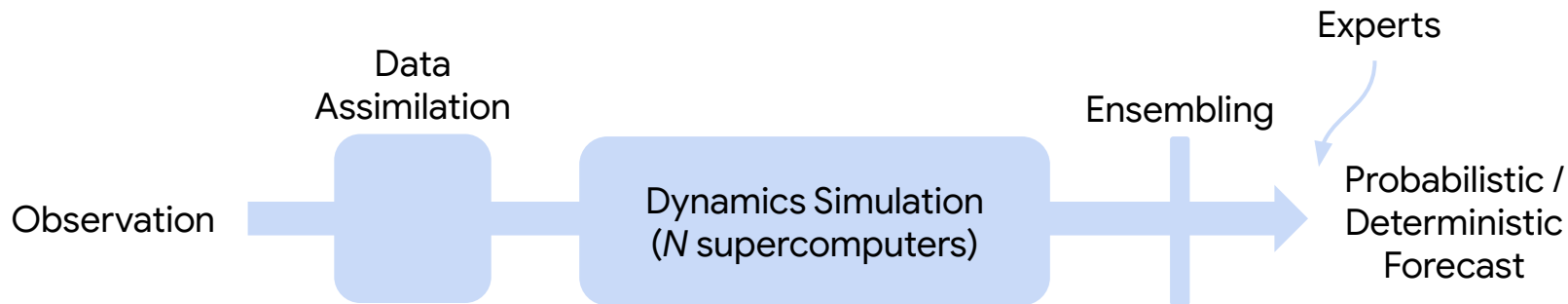
Numerical Weather Prediction (NWP) vs. MetNet-style



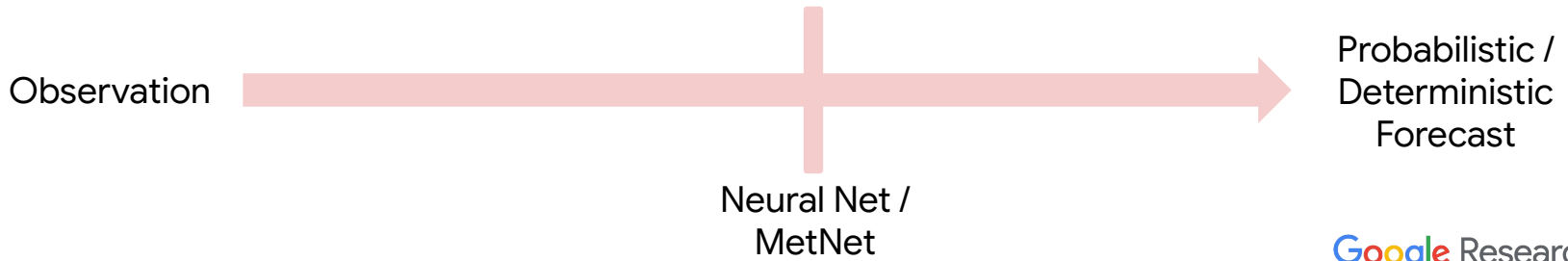
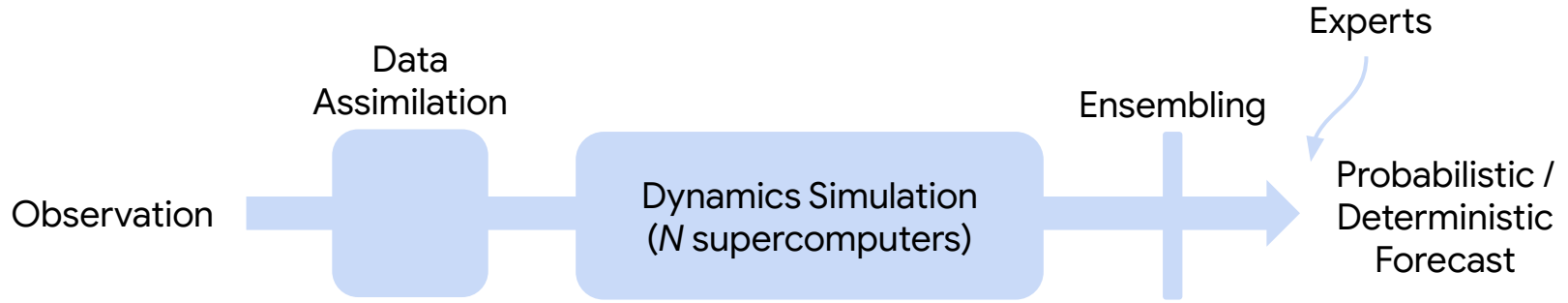
Numerical Weather Prediction (NWP) vs. MetNet-style



Numerical Weather Prediction (NWP) vs. MetNet-style



Numerical Weather Prediction (NWP) vs. MetNet-style



Features: Quick forecasts

NWPs, depending on model, takes approximately one hour to run. NWPs are generally not used for short-term forecasting.

Neural network approaches take a few minutes after data collection. They are highly efficient and highly parallelizable.

~1 hour



NWP

<5 min



MetNet

Features: High resolution

MetNet

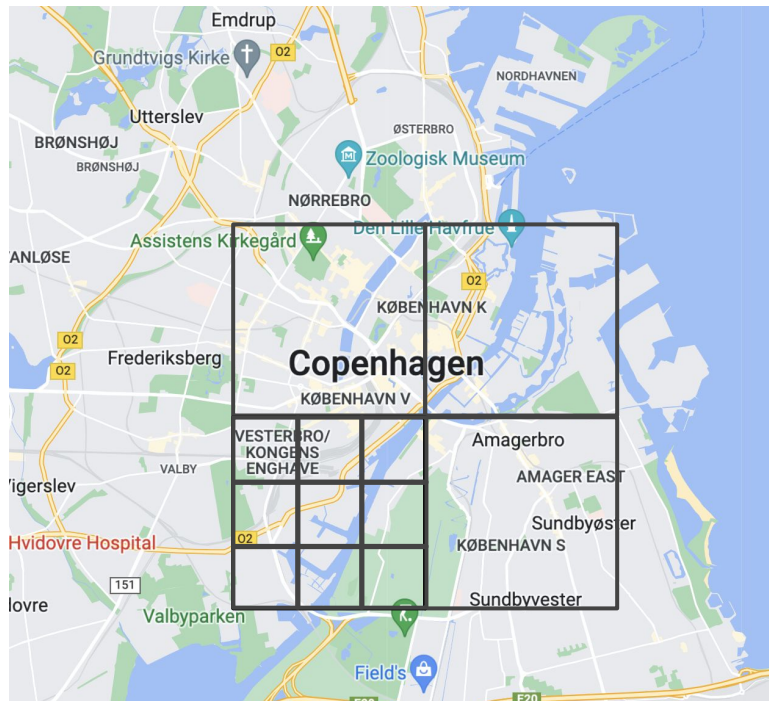
Temporal resolution: 2 min

Spatial resolution: 1 km

NWP (HRRR)

Temporal resolution: 1 hr

Spatial resolution: 3 km



Features: Modeling all phenomena

Precipitation is *one* variable to model, a hard one.

Predicting temperature, humidity, wind, etc. is straightforward.

Also, e.g. tornados, lightning which are annotated. Even ones that are not well understood.



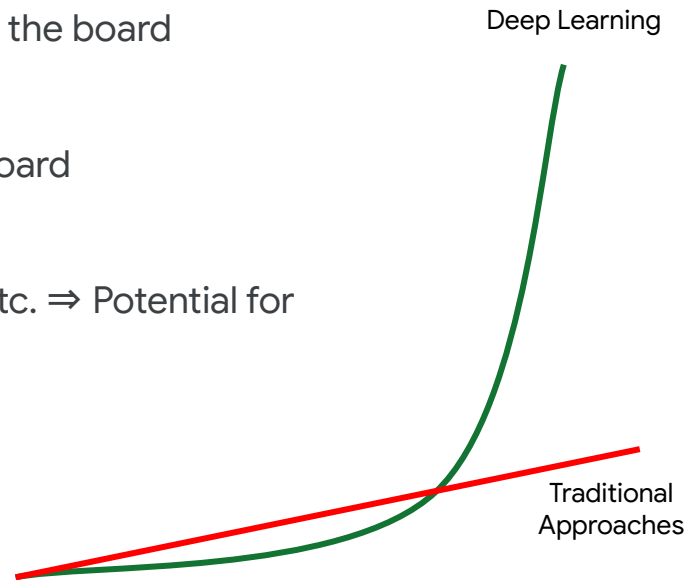
Features: Transfer learning

New variable predicted \Rightarrow Potential for improvement across the board

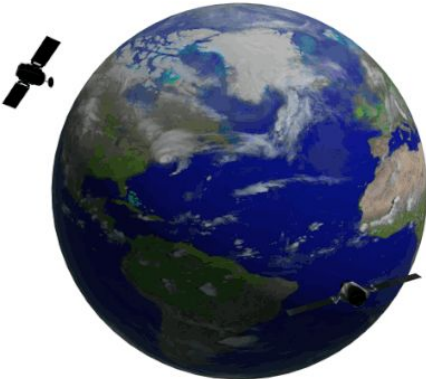
New input added \Rightarrow Potential for improvement across the board

Every new advancement in the community for NLP, vision, etc. \Rightarrow Potential for improvement across the board

The **essence** of modern machine learning success and dominance.



Goal of End to End Weather Forecasting (MetNet)



Results, comparing to HRRR

MetNet-2 is superior to HRRR on **12+ hour forecasts**, up from 8 hours with MetNet-1 (1 year prior).

High Resolution Rapid Refresh (HRRR)

NOAA's HRRR model, real-time 3 km² resolution, 1 hour time resolution, deterministic prediction.

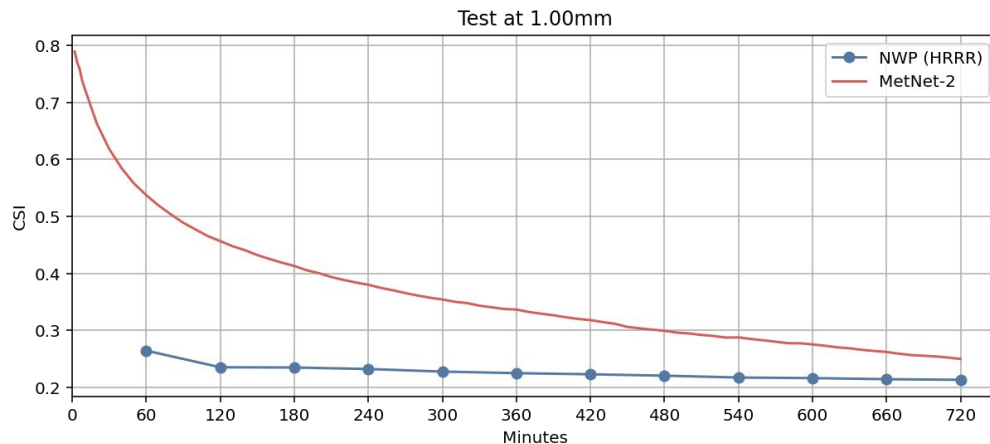
MetNet(-2)

Neural Network model.

Probabilistic output.

1 km² resolution,

2 min time resolution.



Critical Success Index (CSI). Higher is better.

Results, comparing to HREF

MetNet-2 is superior to HREF on **12+ hour forecasts**.

High Resolution Ensemble Forecast (HREF)

NOAA's HREF model produces ensemble products from 10 different models running at ~3 km horizontal grid spacing, and 1 hour time resolution.

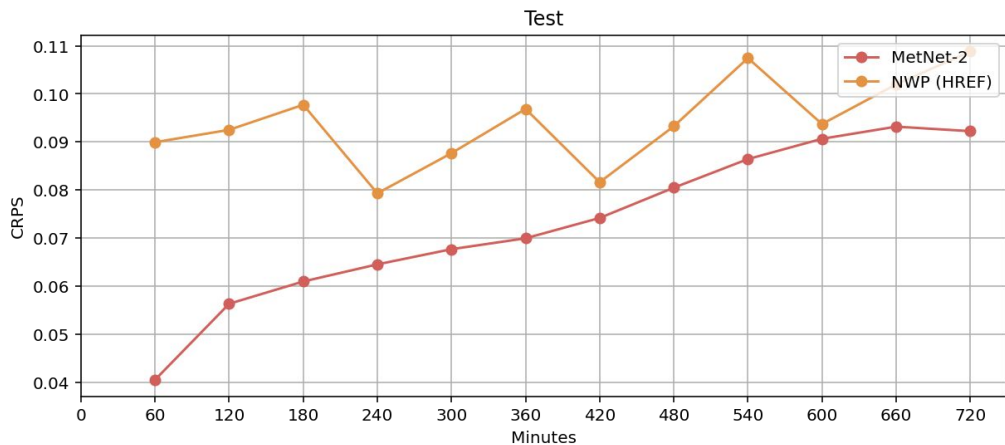
MetNet(-2)

Neural Network model.

Probabilistic output.

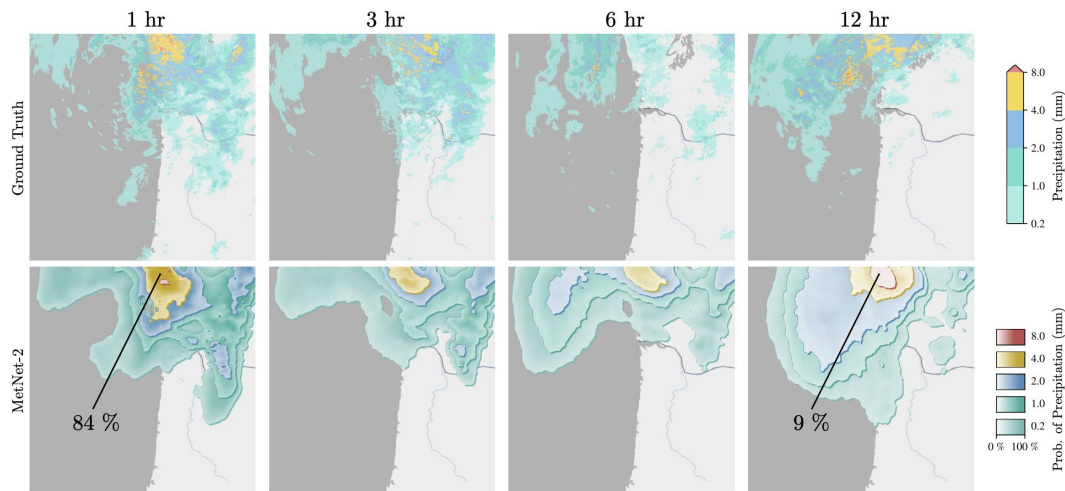
1 km² resolution,

2 min time resolution.



Continuous Ranked Probability Score (CRPS). Lower is better.

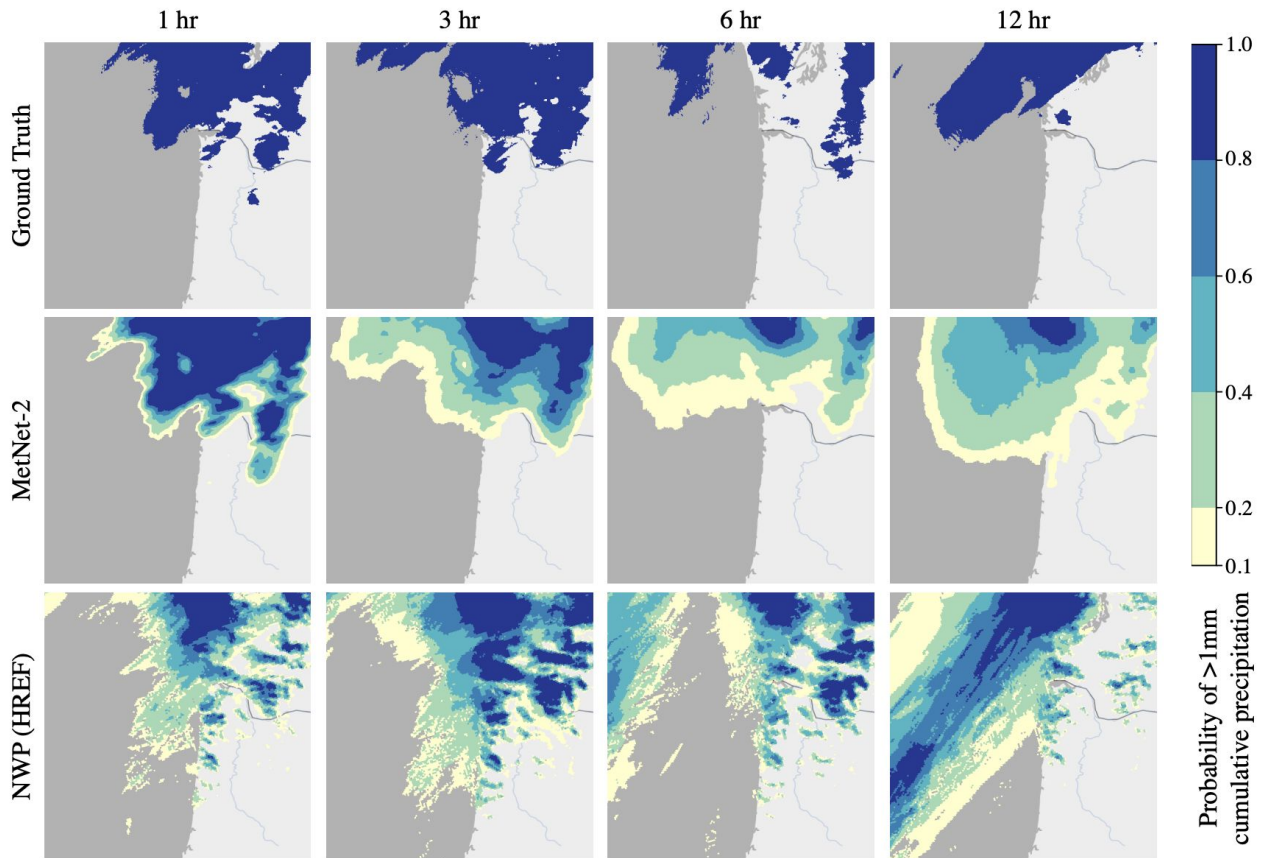
Probability Maps



MetNet-2 outputs probabilities, which we can optionally threshold to obtain a deterministic prediction.

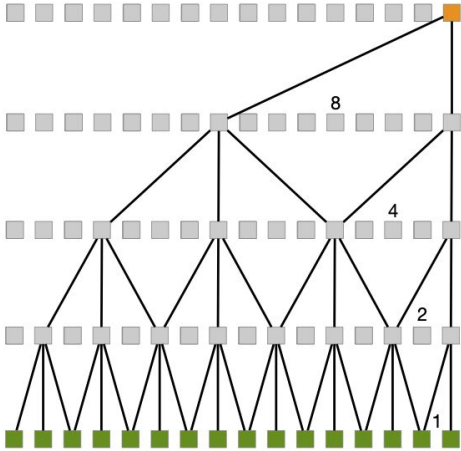
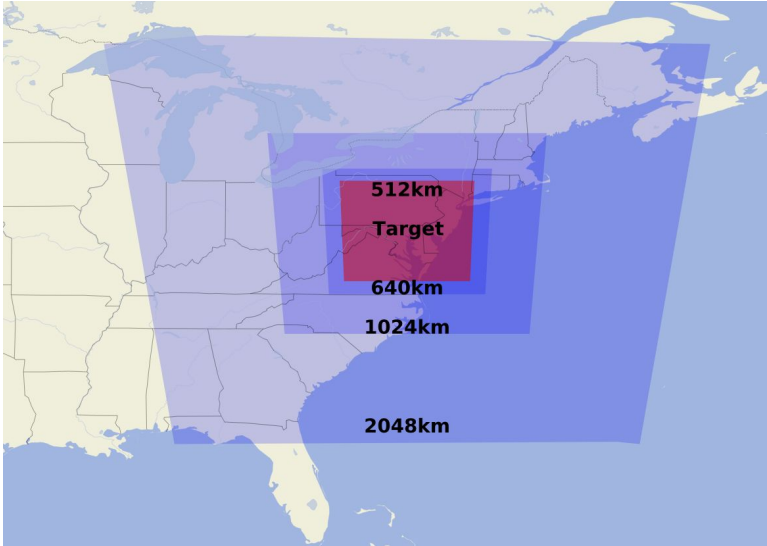
Here we show the probabilities for different amounts of precipitation over an optimized threshold.
Lighter color means less certain.

We see both additional structure and certainty, as well as declining certainty over time (as expected.)

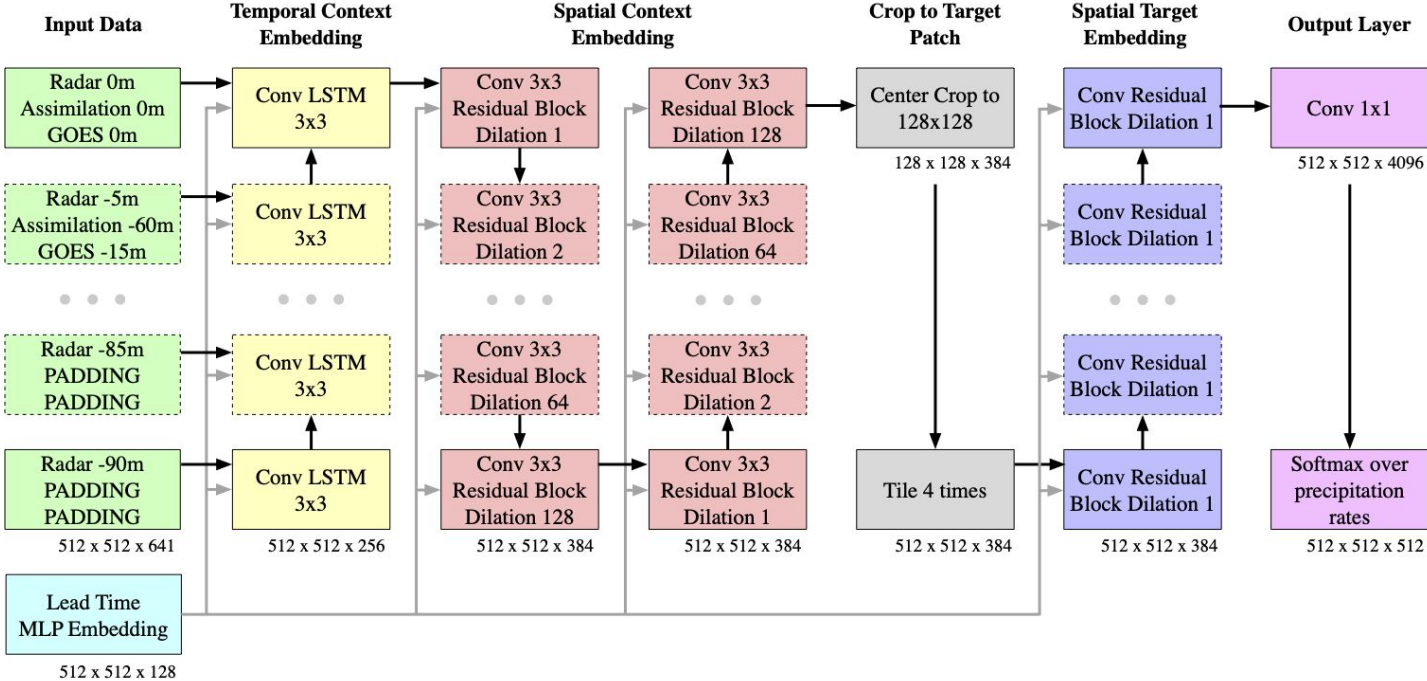


(a) Case study for Thu Jan 03 2019 12:00 UTC of the North West coast of the US.

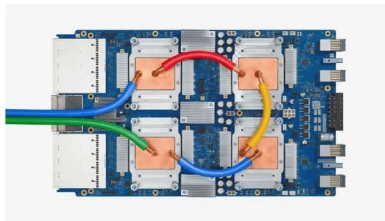
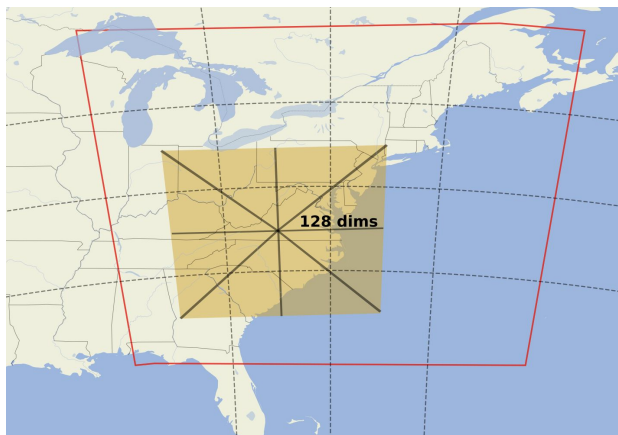
Architecture: Context and Dilation



Architecture: Inputs and Full View



Model Parallelism on TPUs



Cloud TPU v3

420 teraflops

128 GB HBM



Cloud TPU v3 Pod

100+ petaflops

32 TB HBM

2-D toroidal mesh network



JAX



Flax

Interpretation *by Integrated Gradients*

Consistent with Quasi-Geostrophic Theory.

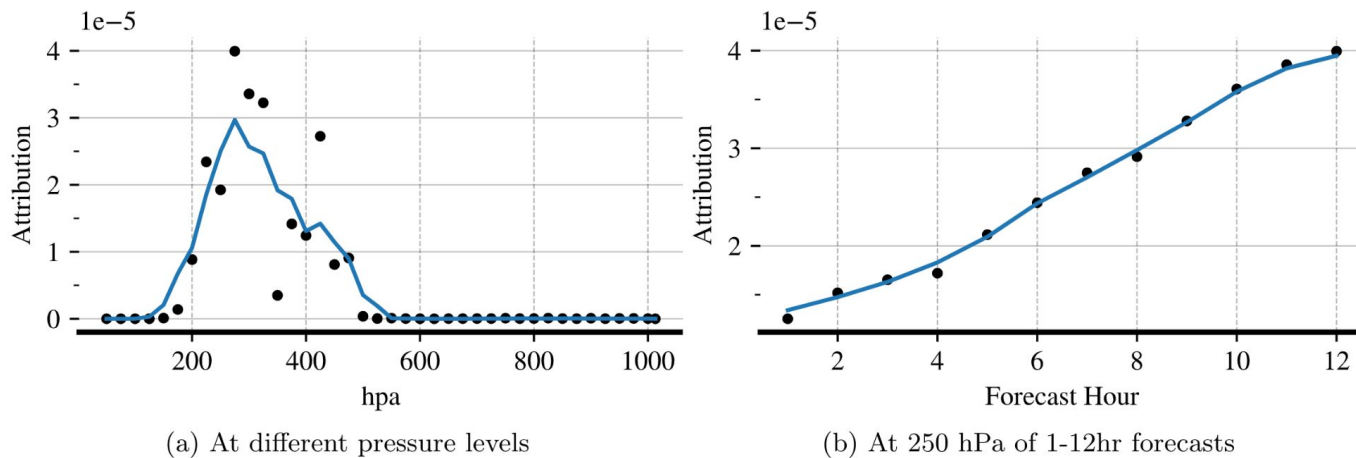


Figure 7: Attribution of Absolute Vorticity

Interpretation *by Integrated Gradients*

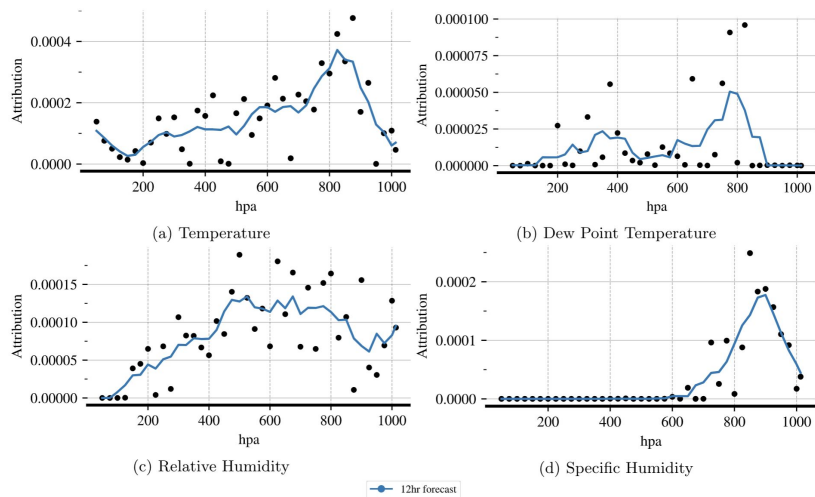
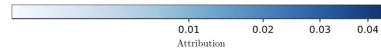
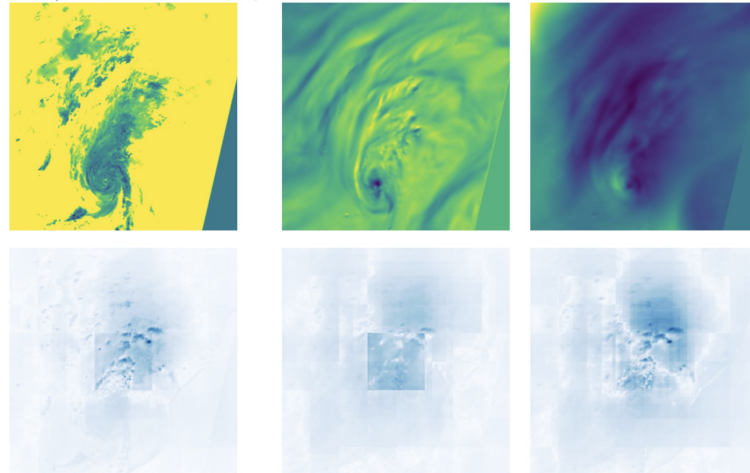
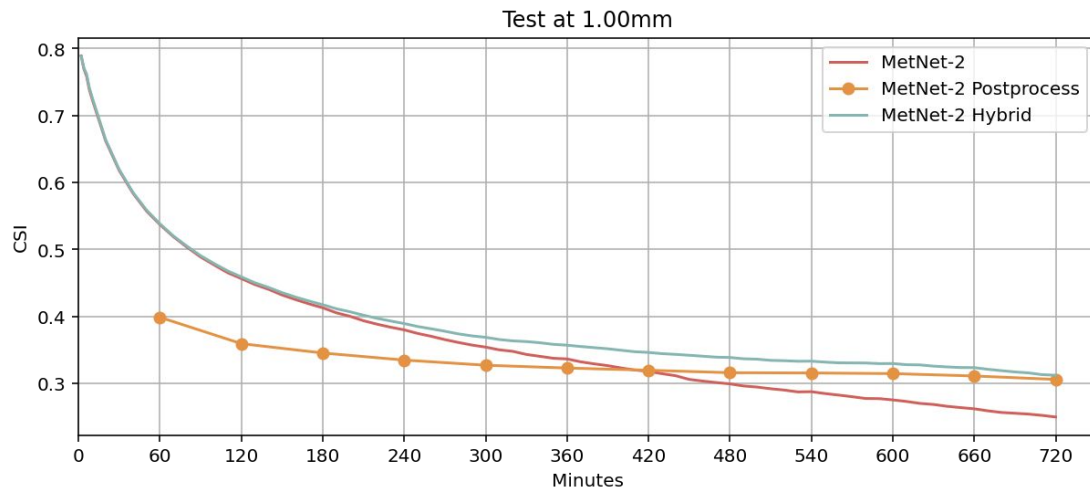


Figure 21: Attribution of different weather features at 12 hour forecast

Maximum/Composite radar reflectivity Absolute vorticity at 275hPa V component of wind at 375hPa



Hybrid Models



“We need to get rid of the dynamical core”

Prof. Dale Durran, University of Washington, '22

WE NEED TO GET RID OF THE DYNAMICAL CORE

- State-of-the-art NWP models require enormous computer resources for each forecast
- Completely replacing NWP with Deep Learning Weather Prediction (DLWP) may
 - *Reduce the time required for each forecast by orders of magnitude*
 - Address uncertainty by
 - Allowing a *large* number $O(1000)$ of simulations of likely future states (*ensembles*)
 - Giving better probabilistic forecasts
 - Capturing extreme events

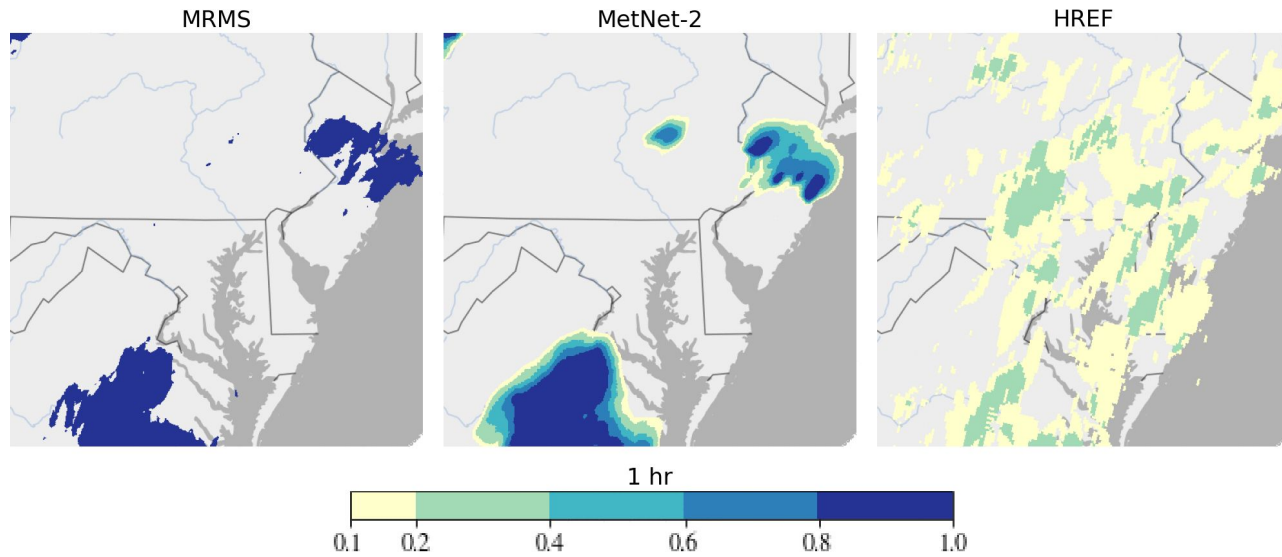


Thanks

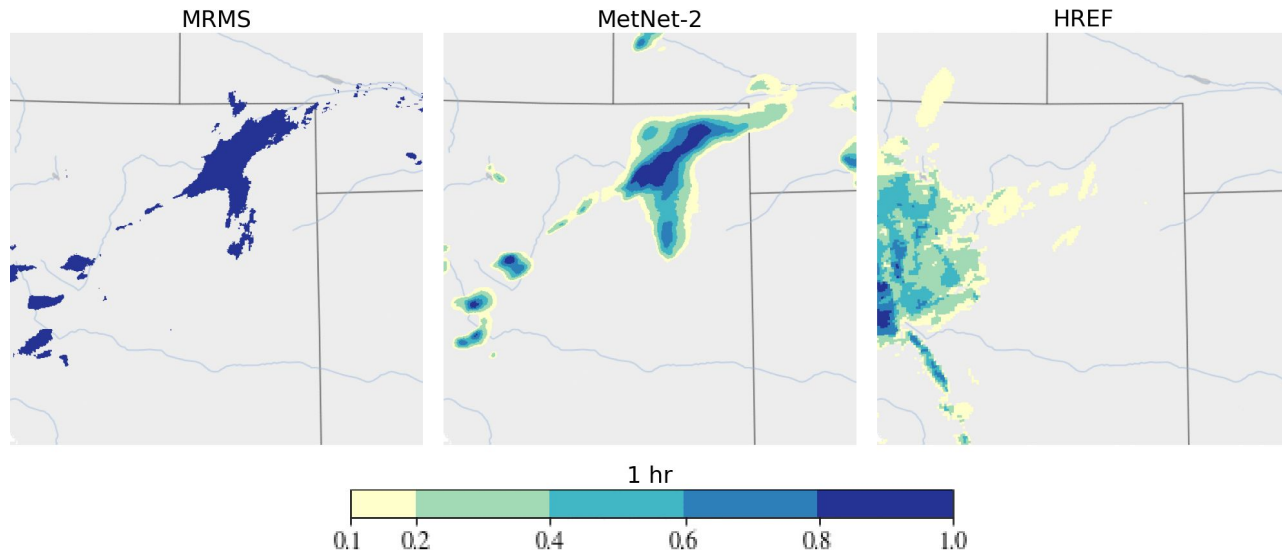
lespeholt@google.com / me@lasse.co

[linkedin.com/in/lasseespeholt](https://www.linkedin.com/in/lasseespeholt)

Visualization 1



Visualization 2



Visualization 3

