Microsoft Start

Machine learning in weather forecasting: An operational perspective

(successes and challenges)

JONATHAN WEYN (MICROSOFT, ECMWF)

ML + Weather: Two strategies?

Augmenting NWP

- Post-processing and calibration
- Super-resolution
- …

New models or replacement

- Physics-free deep learning models
- Precipitation nowcasting
- …

- Physics-constrained deep learning models
- ML-based parameterizations

ML + Weather: Two strategies?

Augmenting NWP

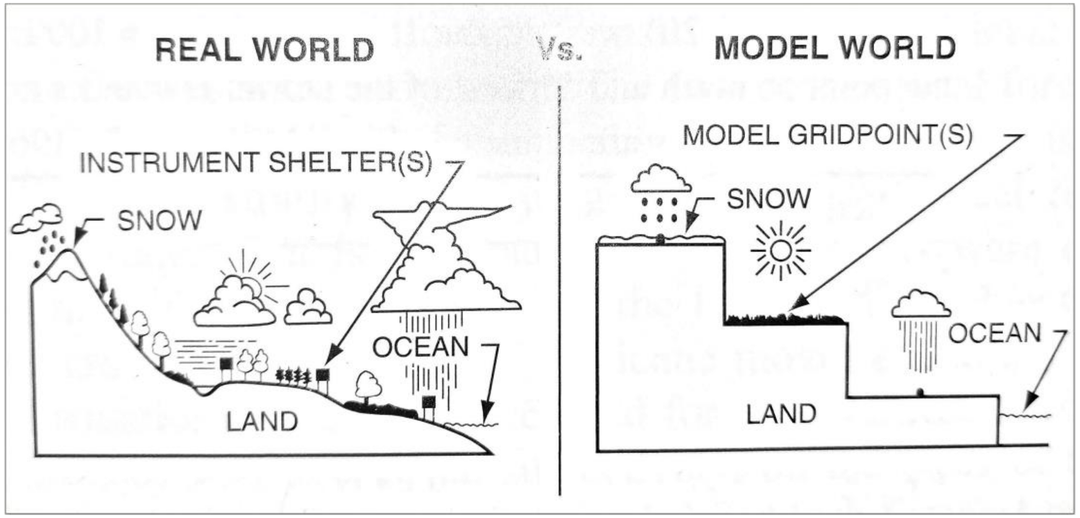
- Post-processing and calibration
- Super-resolution
- …

New models or replacement

- Physics-free deep learning models
- Precipitation nowcasting
- …

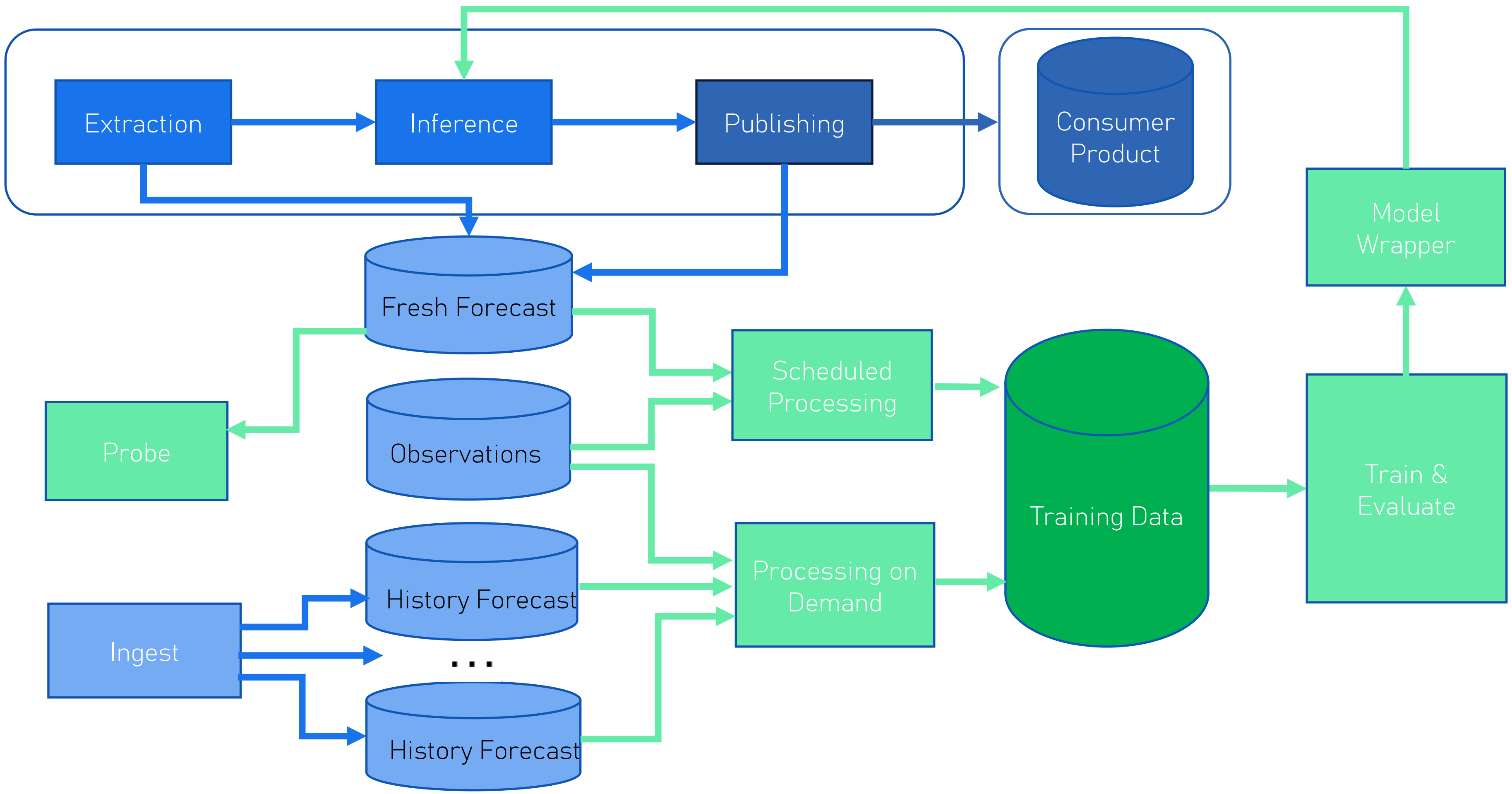
NWP post-processing: success

- Statistical post-processing is almost as old as NWP itself!
- Model Output Statistics (MOS) – 1970s
- Successful adaptation of modern ML methods, employed across the industry



Karl et al. 1972

Operations



Extraction
Inference
Publishing
Consumer Product
Model Wrapper
Fresh Forecast
Scheduled Processing
Probe
Observations
Training Data
Train & Evaluate
History Forecast
Processing on Demand
Ingest
...
History Forecast

Operations design

- All about automation: not only inference, but also data extraction, quality control, model training, model validation
  - Budget for online model training
- Avoid common pitfalls
  - Real-time data quality control (garbage in, garbage out)
  - Online model training (poor model generalization; performance decline)

Next in post-processing: Extreme events

- To protect life and property, key mission of any weather service, we must be able to forecast extreme events
- Key revolutions in NWP: ensemble forecasting and data assimilation
- Can machine learning utilize ensemble NWP to improve forecasts of extreme events? Can such a model be operationalized?

[Submitted on 15 Jul 2022 (v1), last revised 20 Sep 2022 (this version, v2)]

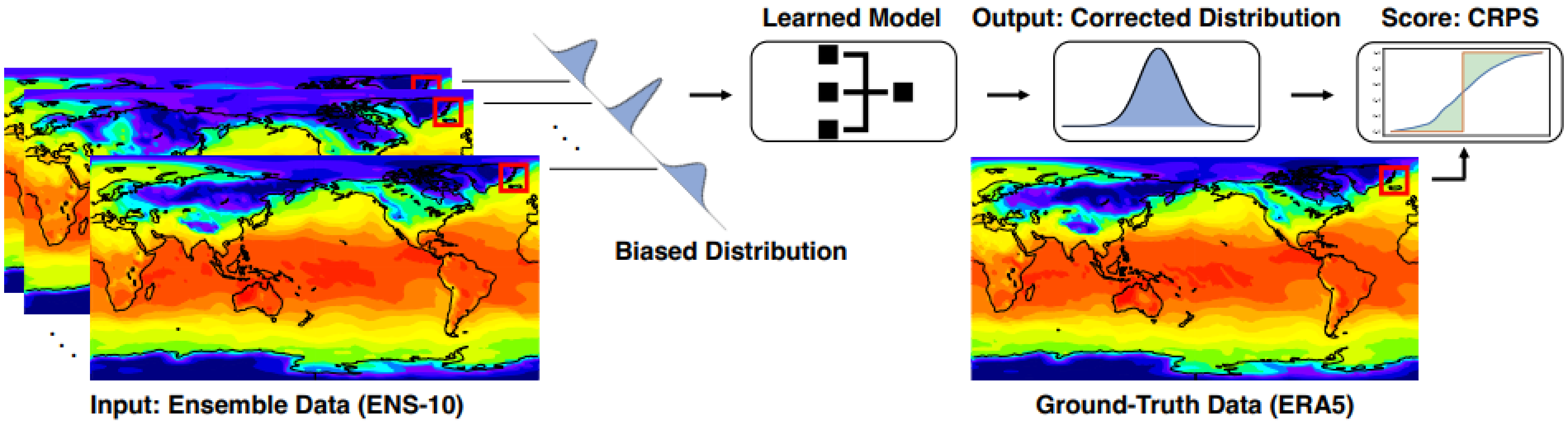
Machine learning applications for weather and climate need greater focus on extremes

Peter AG Watson

The ENS-10 benchmark

Ashkboos et al. (2022)

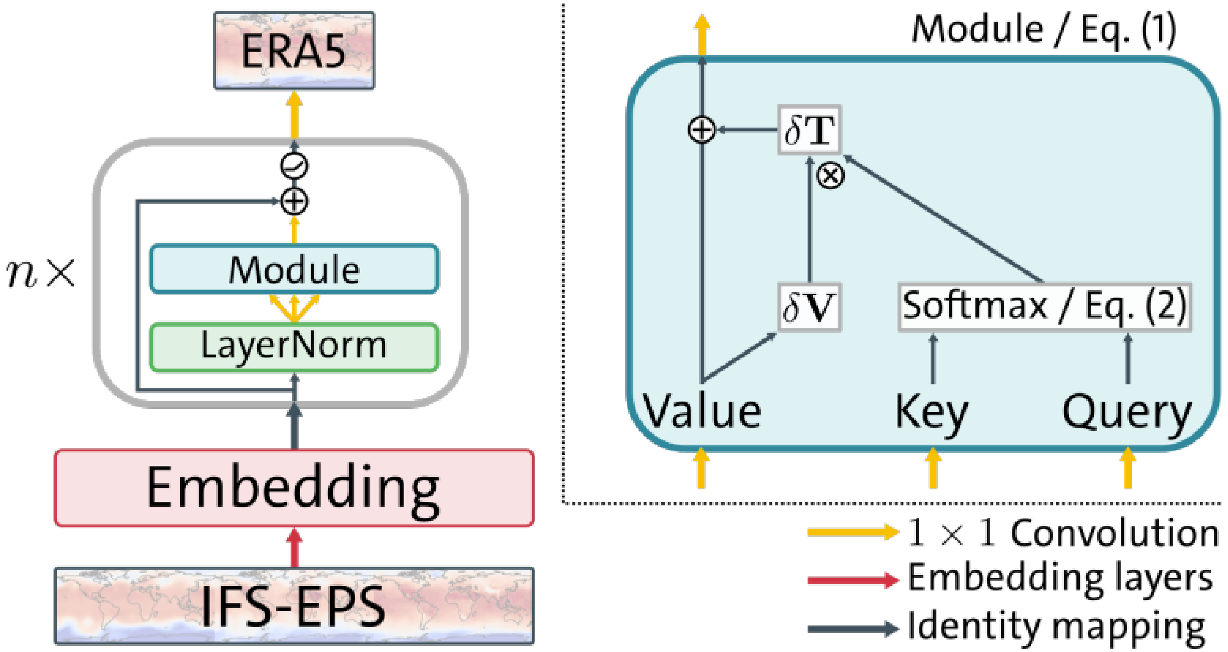
- Task: Predict ensemble of 48-hour forecasts for Z500, T850, T2 on a global, 0.5-degree grid
- ECMWF IFS model hindcasts (10 members) available for 20 years of training data



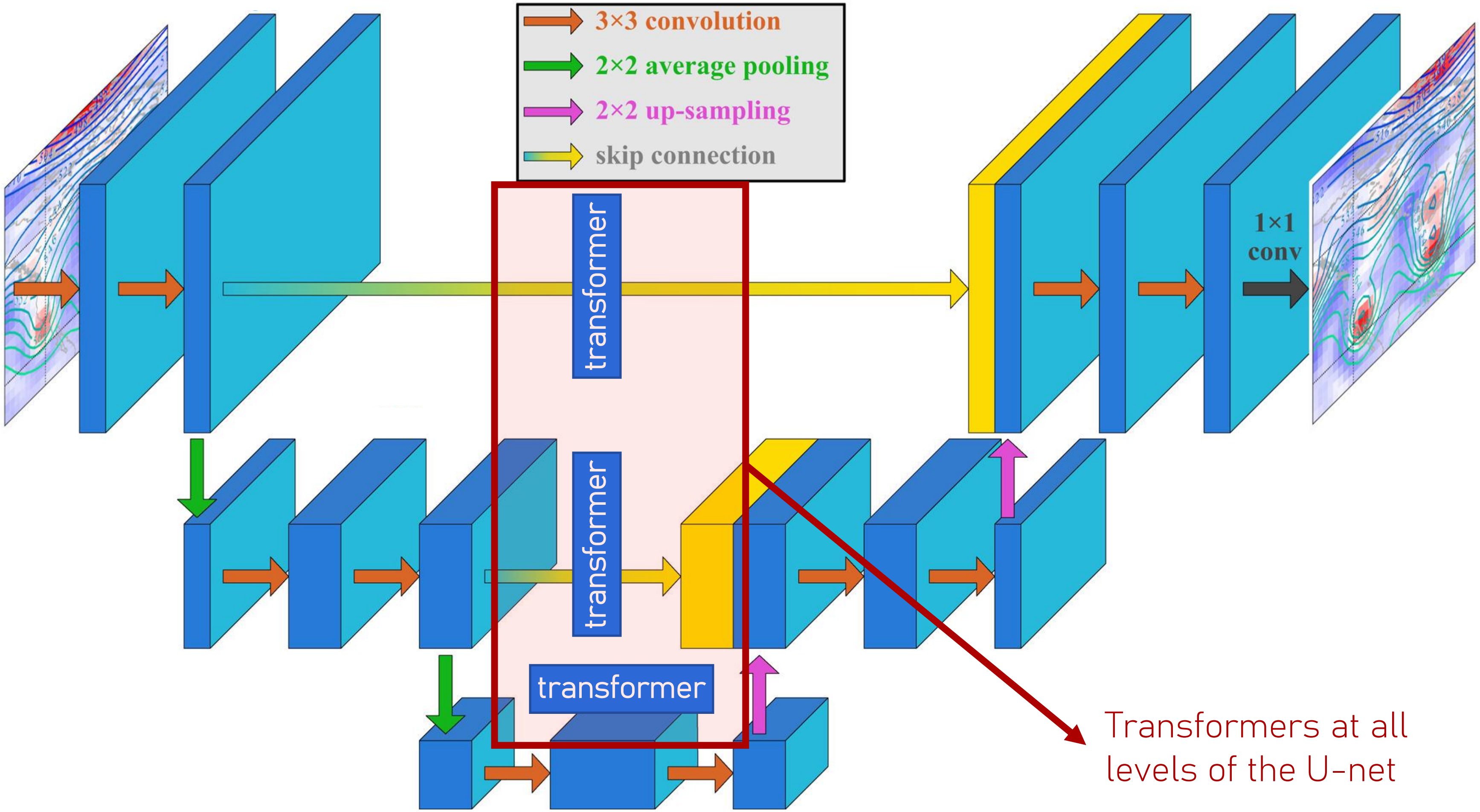
With ECMWF:
Zied Ben Bouallegue, Matthew Chantry, Mariana Clare, Peter Dueben, Jesper Dramsch

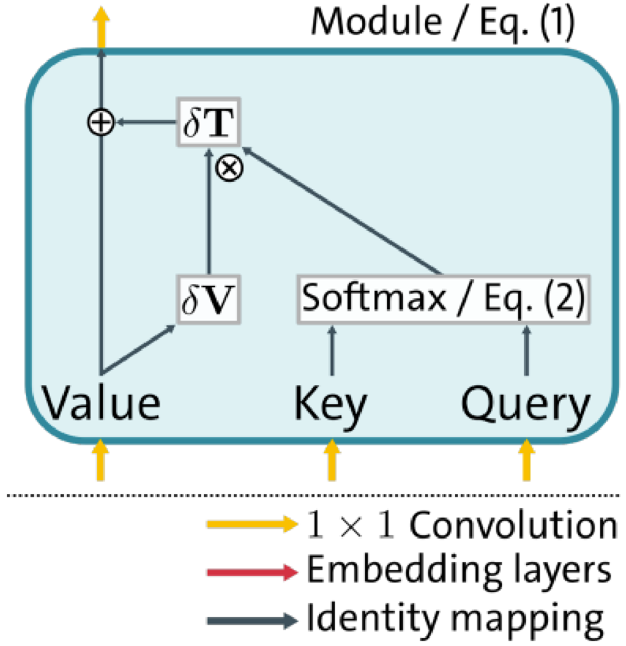
Mean + Std is not enough: The ensemble transformer

- Finn, "Self-Attentive Ensemble Transformer", 2021
- Apply the transformer along the *ensemble member dimension*
- Member-by-member approach that applies corrections based on interactions with all other members
- Maintains physical realism of each ensemble member



PoET: a transformer U-net





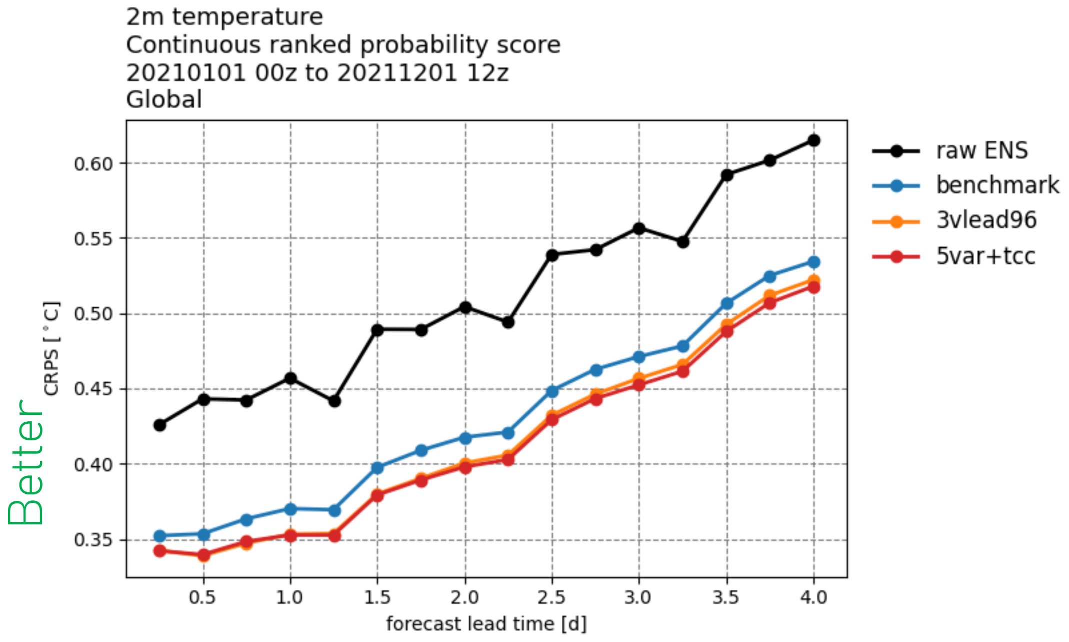
Rankings

Table 2: Global mean CRPS and EECRPS on the ENS-10 test set (2016–2017) for baseline models with five (5-ENS) or ten (10-ENS) ensemble members.

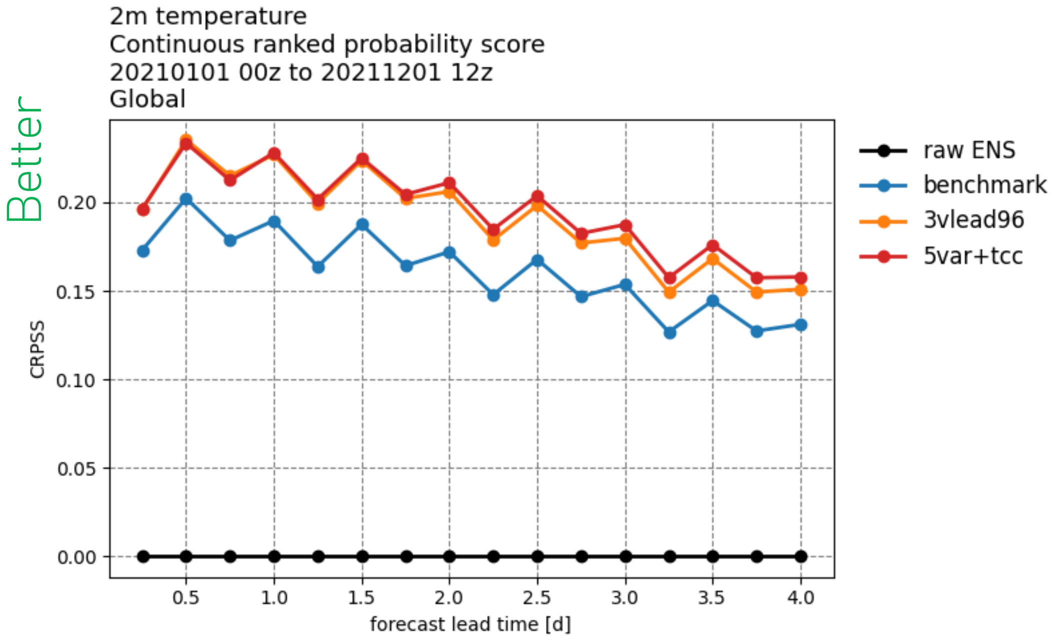
		Z500 $[m^2 s^{-2}]$		**T850** [K]		**T2m** [K]	
Metric	**Model**	5-ENS	10-ENS	5-ENS	10-ENS	5-ENS	10-ENS
CRPS	Raw	81.03	78.24	0.748	0.719	0.758	0.733
	EMOS	$79.08^{\pm 0.739}$	$81.74^{\pm 6.131}$	$0.725^{\pm 0.002}$	$0.756^{\pm 0.052}$	$0.718^{\pm 0.003}$	$0.749^{\pm 0.054}$
	MLP	$75.84^{\pm 0.016}$	$74.63^{\pm 0.029}$	$0.701^{\pm 2e-4}$	$0.684^{\pm 4e-4}$	$0.684^{\pm 6e-4}$	$0.672^{\pm 5e-4}$
	LeNet	$\mathbf{75.56}^{\pm 0.101}$	$\mathbf{74.41}^{\pm 0.109}$	$0.689^{\pm 2e-4}$	$0.674^{\pm 2e-4}$	$0.669^{\pm 7e-4}$	$0.659^{\pm 4e-4}$
	U-Net	$76.66^{\pm 0.470}$	$76.25^{\pm 0.106}$	$0.687^{\pm 0.003}$	$0.669^{\pm 0.009}$	$0.659^{\pm 0.005}$	$0.644^{\pm 0.006}$
	Transformer	$77.30^{\pm 0.061}$	$74.79^{\pm 0.118}$	$\mathbf{0.686}^{\pm 0.002}$	$\mathbf{0.665}^{\pm 0.002}$	$\mathbf{0.649}^{\pm 0.004}$	$\mathbf{0.626}^{\pm 0.004}$
PoET U-net Transformer			**73.97**		**0.650**		

Expanded forecast performance

CRPS



CRPS SCORE



Ensemble post-processing: Summary

- The ensemble transformer can run on the full operational IFS ensemble, while
  - preserving inter-member calibration
  - improving ensemble metrics.
- While the operationalization process might be modeled on existing post-processing work, there are still notable challenges:
  - Manipulation and serving of very large datasets, especially at high resolution
  - Communicating to the public the risks and uncertainties of extreme events

ML + Weather: Two strategies?

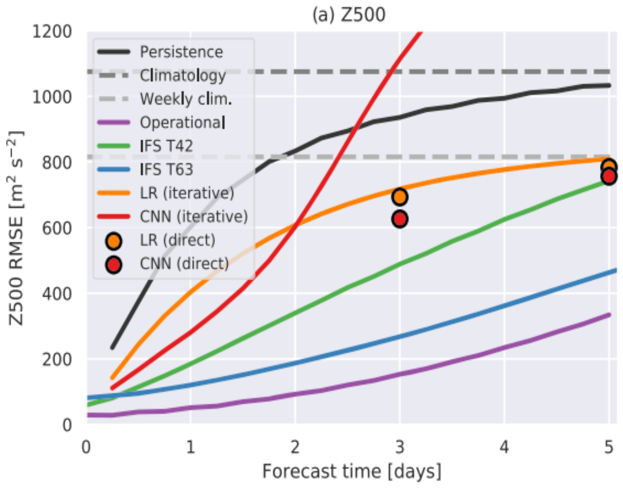
Augmenting NWP

- Post-processing and calibration
- Super-resolution
- ...

New models or replacement

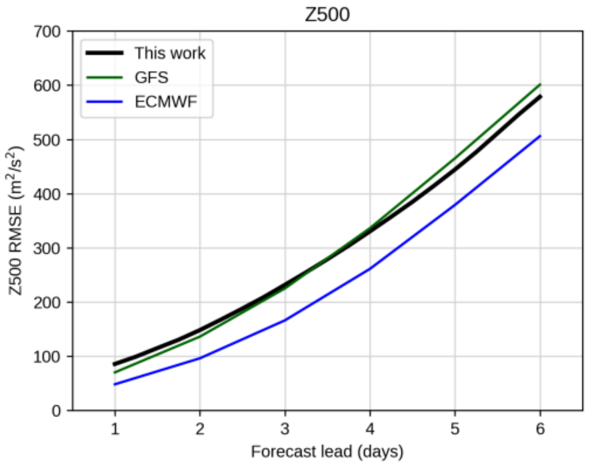
- Physics-free deep learning models
- Precipitation nowcasting
- ...

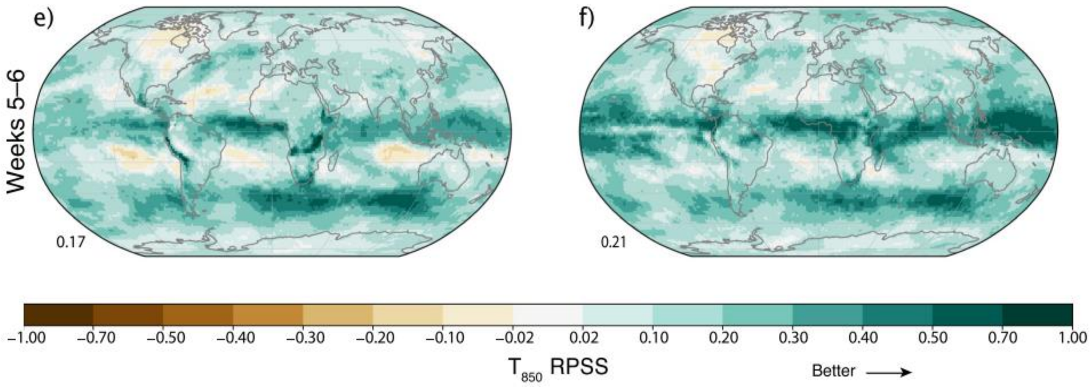
Growing work on NWP replacement



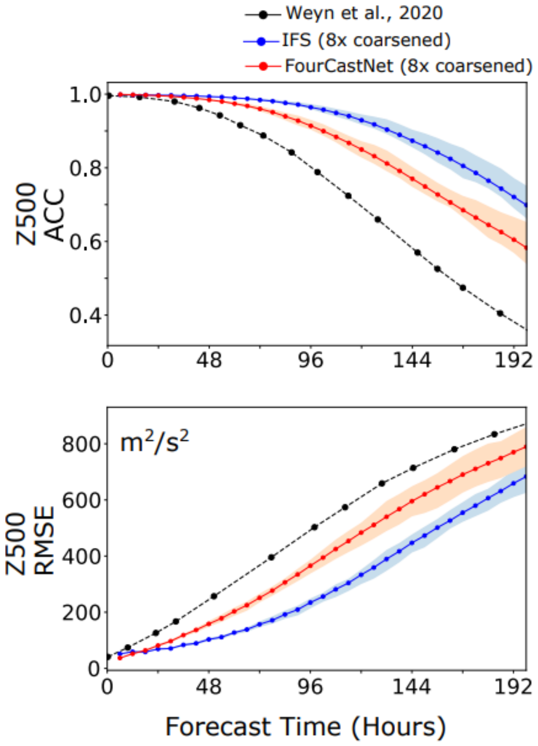
WeatherBench (Rasp et al. 2020)

Keisler (2022)





Weyn et al. (2020, 2021)



Pathak et al. (2022)
Kurth et al. (2022)

Why replace NWP?

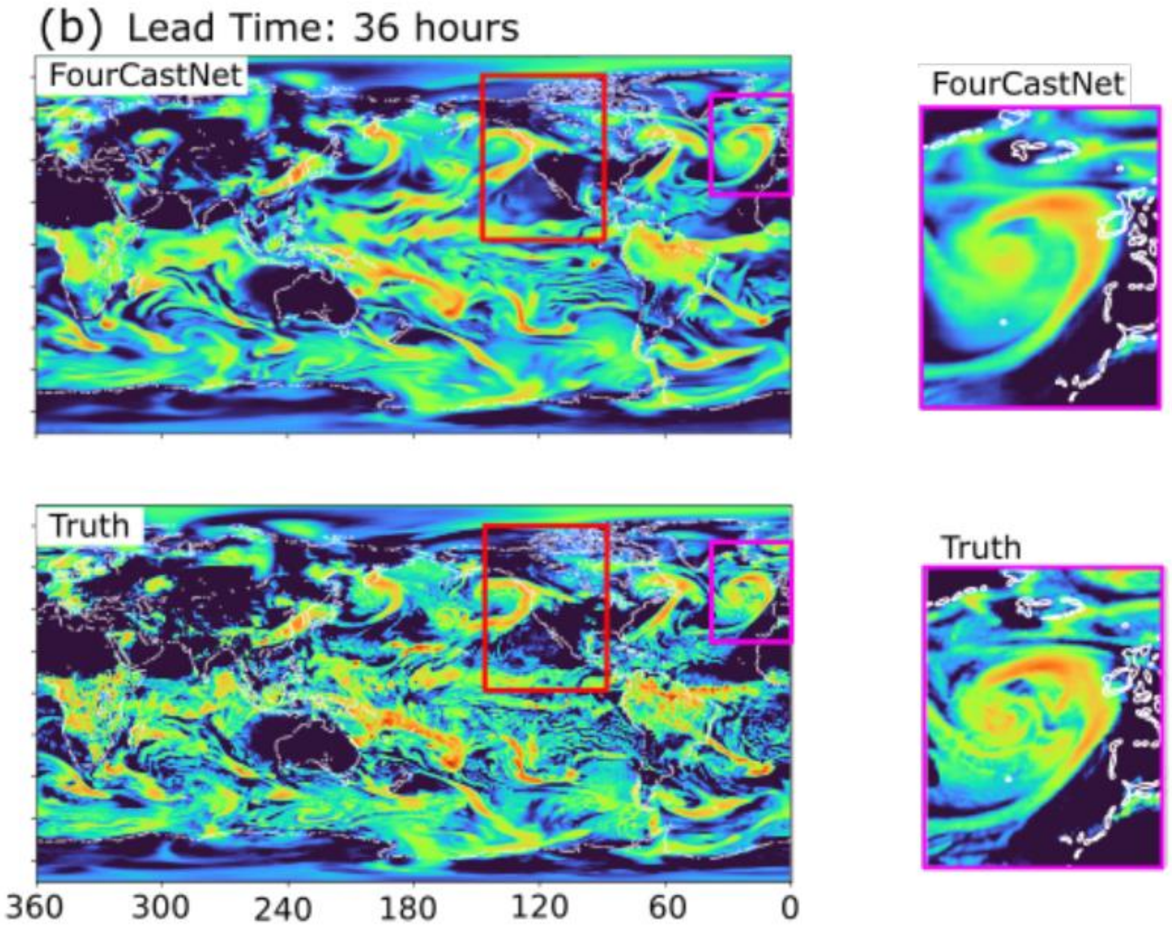
- Efficiency
- Efficiency
- Did I mention efficiency?
- Seriously, NVIDIA's FourCastNet is *80,000*x faster (per compute node) and *10,000*x more energy efficient than a comparable ECMWF IFS simulation
- This enables

  Very large ensemble forecasts

  Data reproduction in near-real-time instead of archiving petabytes of data on slow storage

Active challenges

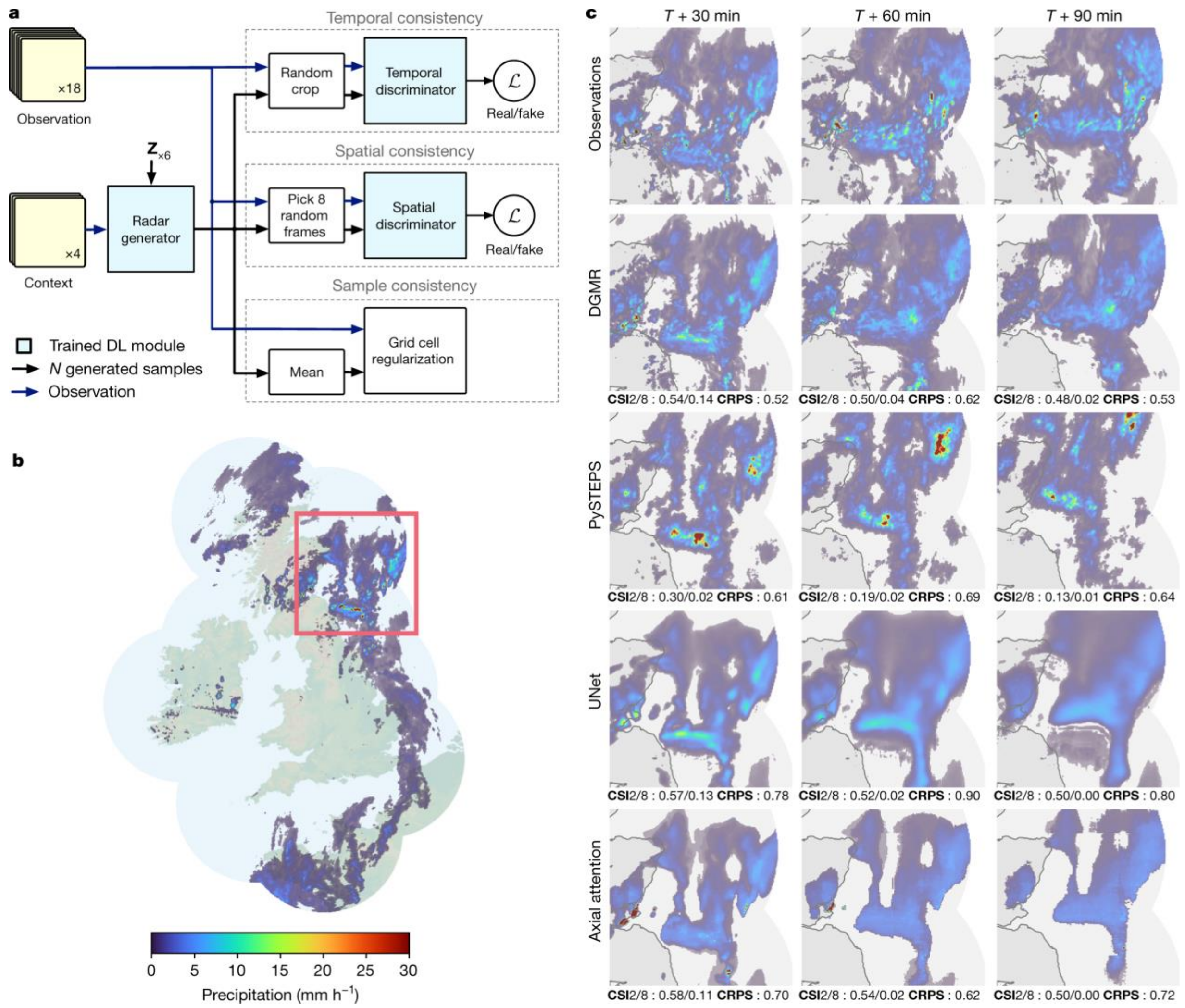
- Increasing physical realism
- Evidence of large ML-based ensembles picking up extreme events missed by NWP
- Adaptation to future climate



Tackling new problems: Precipitation nowcasting

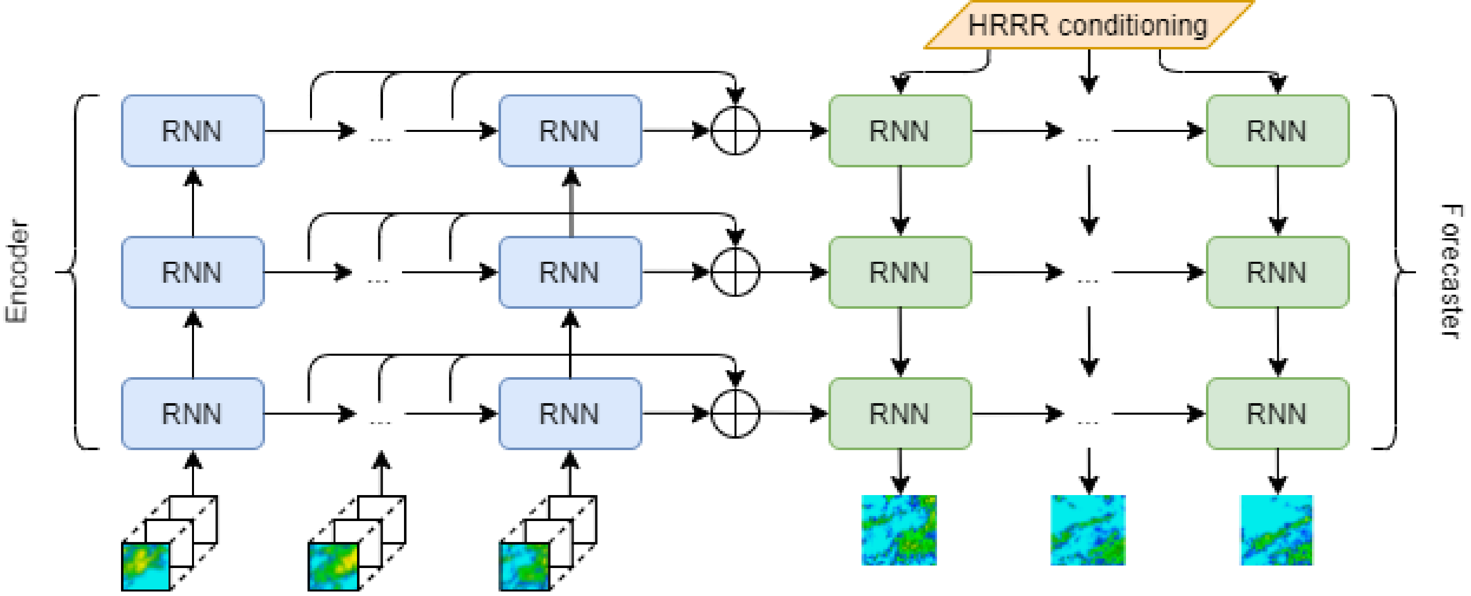
- The convolutional LSTM architecture was created for radar nowcasting (Shi et al. 2015)
- Google's MetNet-2 integrated HRRR high-resolution regional NWP model for a hybrid approach (Espeholt et al. 2021)
- DeepMind + UK Met Office: Deep generative models of radar (Ravuri et al. 2021)

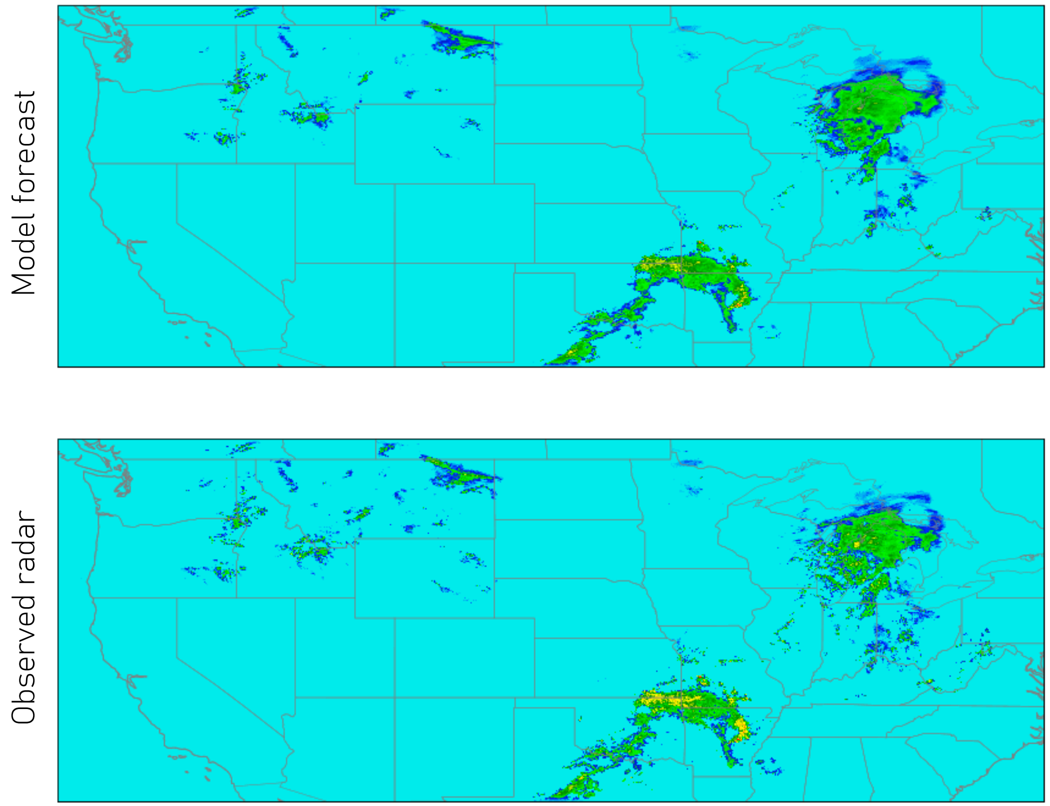
  Chosen 89% of the time over competing methods by expert meteorologists



MS-Nowcasting

- ConvLSTM-style generator architecture
- Added adversarial loss using DGMR-style discriminators
- Condition forecasting layers with HRRR reflectivity forecast to inject hybrid model
- Several techniques such as channel stacking and dilated convolutions used to reduce memory/computation footprint



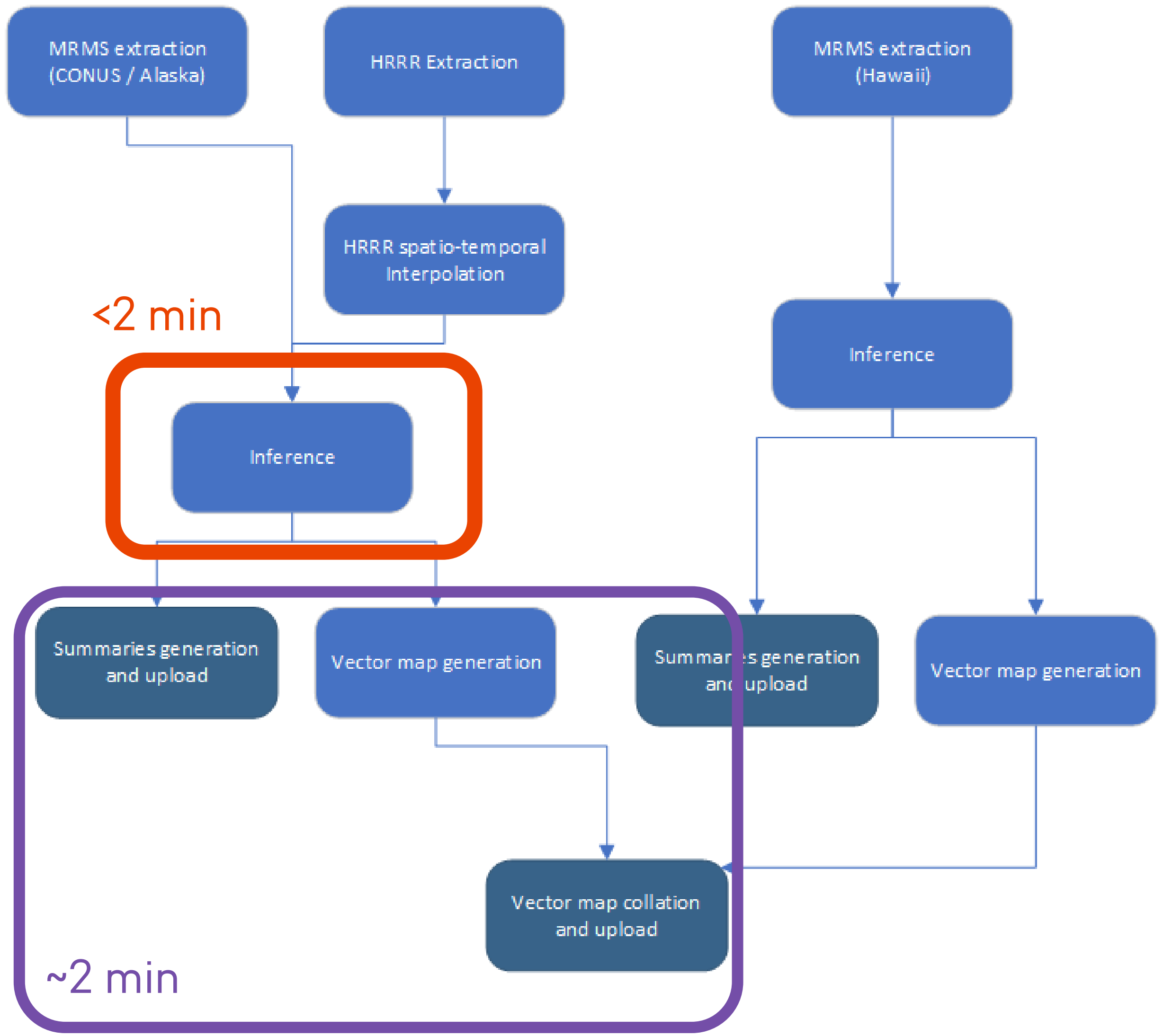


Operations

- Latest radar images from MRMS (in US) retrieved with about 2 min latency
- Model inference is controlled by a streaming pipeline based on a message queueing framework

  streaming is not bound by performance of any single component

  no additional I/O costs
- CONUS region is tiled into 8 sectors with small overlap
- Final product generation consists of making text summaries and shapes for maps



Concluding thoughts…

- When designing an operational machine-learning-based weather solution, it's important to always keep the end user in mind
  - Even for simple post-processing, it's necessary to have automation to continually improve forecasts
  - For nowcasting, speed is vital, as is real-time detection of issues
- There is very promising and exciting research in ensemble post-processing and model replacement, but to operationalize for the public, it will need effective communication

jweyn@microsoft.com
msn.com/weather