# Can deep learning replace current numerical weather prediction models?

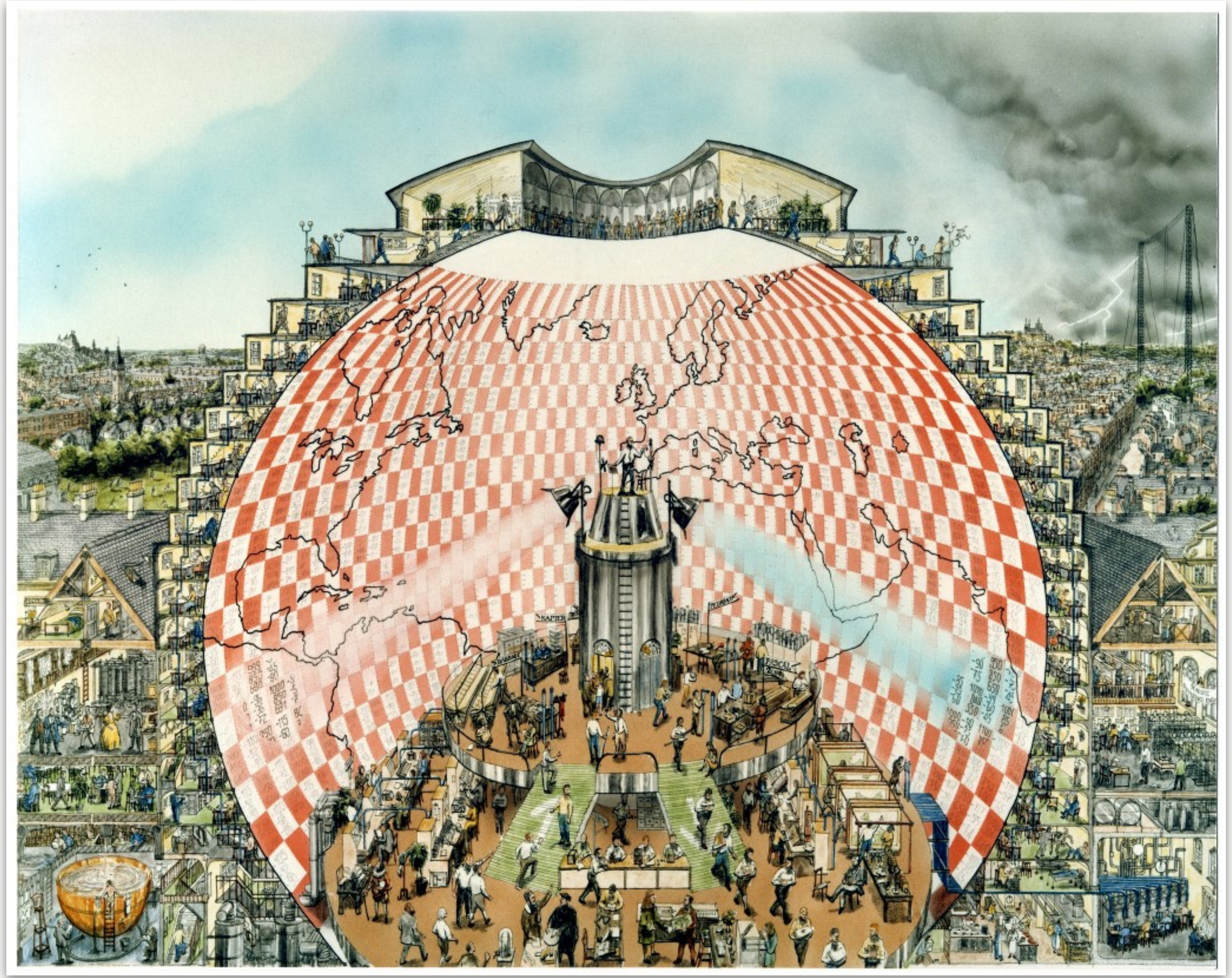
DALE DURRAN, JONATHAN WEYN, NATHANIEL CRESSWELL-CLAY, RICH CARUANA

OCTOBER 8, 2021

# A Vision of Numerical Weather Prediction (NWP)

"Imagine a large hall like a theatre… the walls of this chamber are painted to form a map of the globe…. A myriad computers are at work upon the weather of the part of the map where each sits, but each computer attends only to one equation or part of an equation."

*–Lewis Fry Richardson, Weather Prediction by Numerical Process,1922*



"Weather Forecasting Factory" by Stephen Conlin, 1986

# Made Practical by Advancements in Computing and Numerics

- Jule Charney and John Von Neumann led the first numerical weather prediction experiment in 1950
- They integrated the barotropic vorticity equation on 500-hPa surface

$$\frac{\partial \nabla^2 \psi}{\partial t} = \frac{1}{a^2}\left[\frac{\partial \psi}{\partial \mu}\frac{\partial \nabla^2 \psi}{\partial \lambda} - \frac{\partial \psi}{\partial \lambda}\frac{\partial \nabla^2 \psi}{\partial \mu}\right] - \frac{2\Omega}{a^2}\frac{\partial \psi}{\partial \lambda}$$

- 24-hour forecast took about 24 hours to compute on ENIAC computer

# FUNDAMENTAL PHYSICS AND NWP MODELS?

- Dynamical core: equations for conservation of mass, energy and momentum …
  - Inviscid motions and wave propagation
  - Its numerical approximation can be evaluated for order of accuracy, stability, …
- Operational Models Rely on Parameterizations
  - Clouds and precipitation
  - Influence of the Earth's surface (surface temperatures)
  - Heat transfer by electromagnetic radiation
- Parameterizations
  - *Have a major impact on forecast skill*
  - *Are tested empirically*

# AI AND NWP

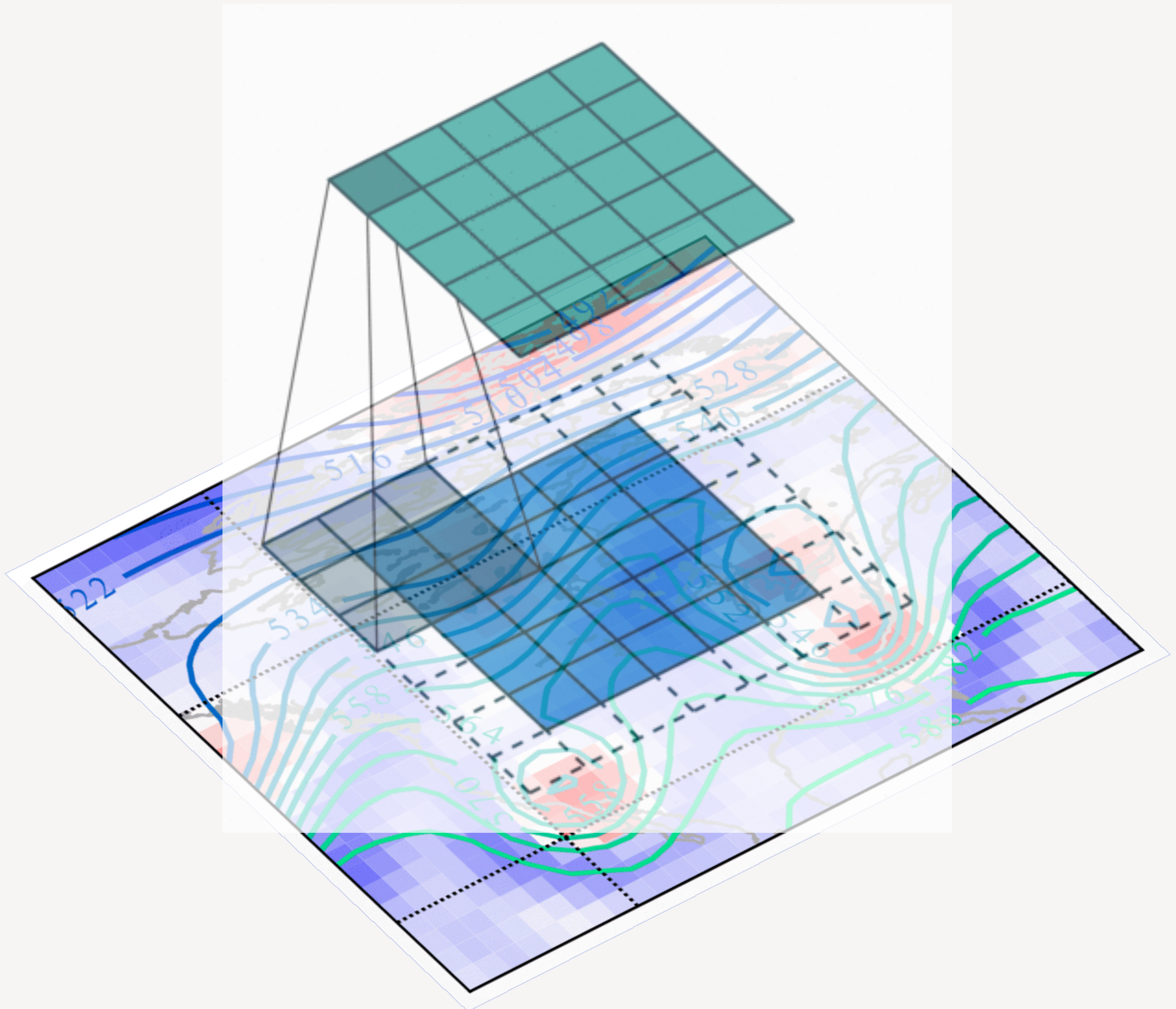
- Parameterizations are empirical and major limitations in the accuracy of NWP
  - Many groups are trying to improve parameterizations using AI
- State-of-the-art NWP models require enormous computer resources for each forecast
- Completely replacing NWP with Deep Learning Weather Prediction (DLWP) may
  - 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

# OUR DLWP STRATEGY

- Like NWP, we create forecasts by recursively stepping forward in time
- We will use far fewer variables to characterize the atmospheric state
- We use just a few variables and coarse horizontal resolution because
    - It's a starting point for DLWP
    - Numerical resolution in NWP may be greater than the important degrees of freedom in a given atmospheric state.
        - For convergence: $\Delta z < 200$ m if $\Delta x < 15$ km (Skamarock et al., 2019)

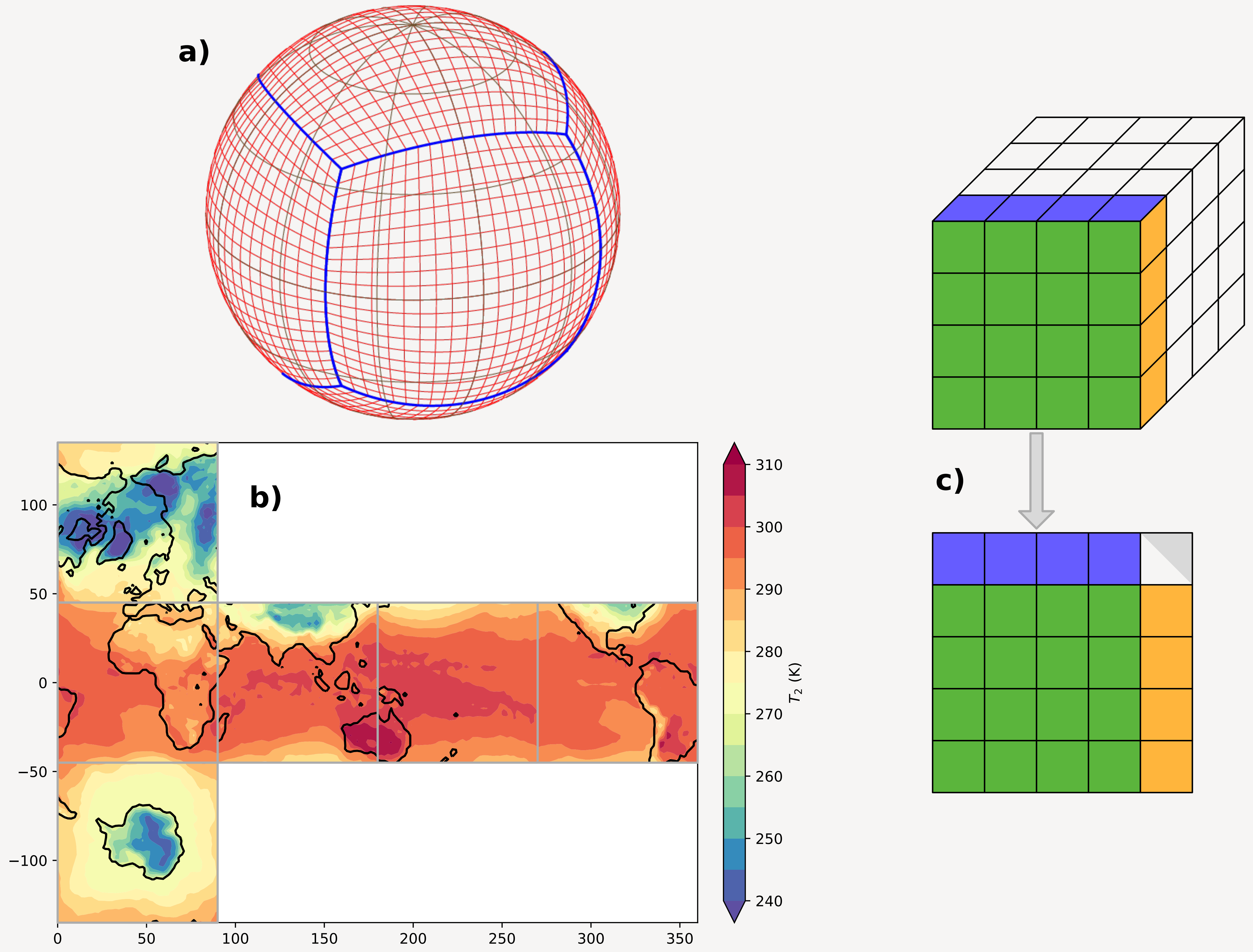
# DLWP Building Blocks: Convolutional Neural Networks

- Same filter coefficients multiply the input data at every point
  - 3x3 horizontal stencil
  - 3rd dimension is number of fields
  - Output is single number
- Learn many sets of these filter coefficients (64/128/256)

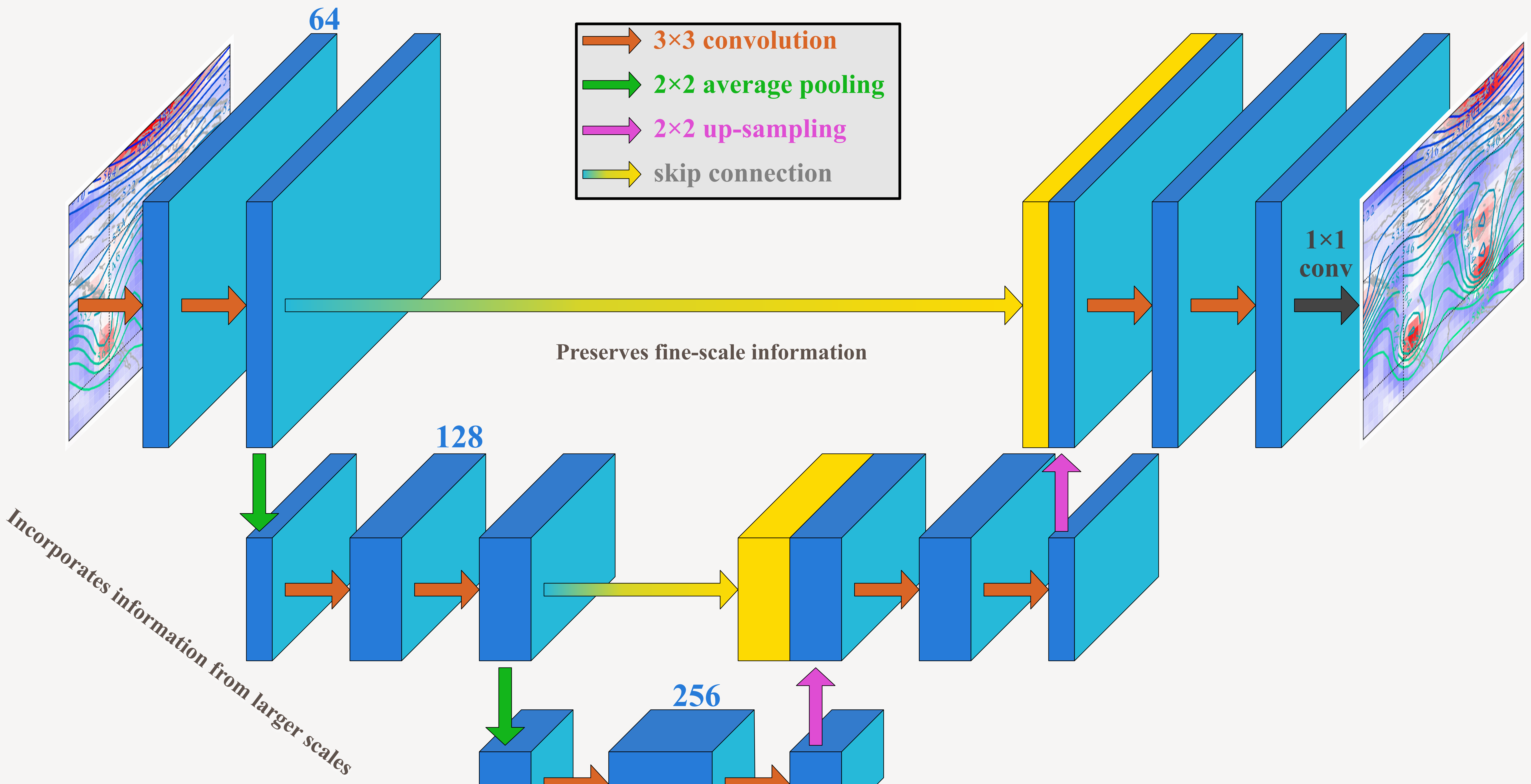


# DLWP Building Blocks: Cubed Sphere Grid

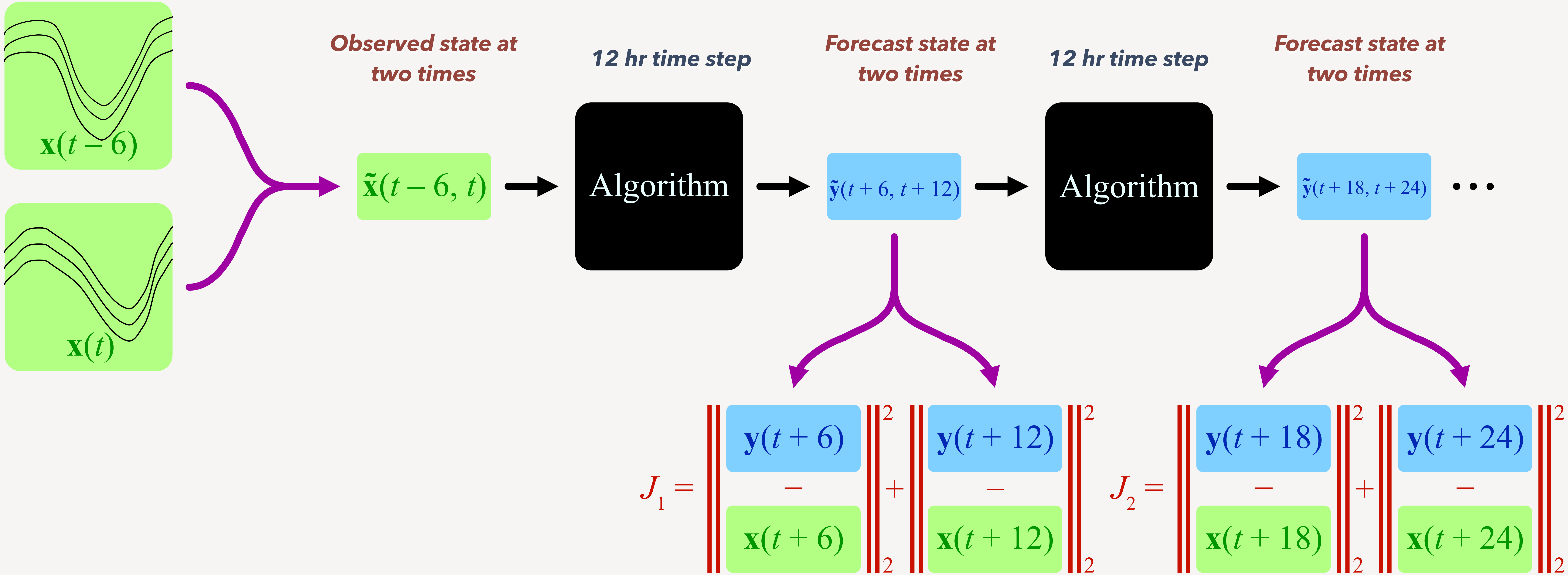
- Convenient for 3x3 spatial stencil
- Train identical filters for
  - 4 equatorial-centered faces
  - 2 polar faces
    - sense of rotation reversed between polar faces



# DLWP Building Blocks: U-Net Architecture



# Time Stepping



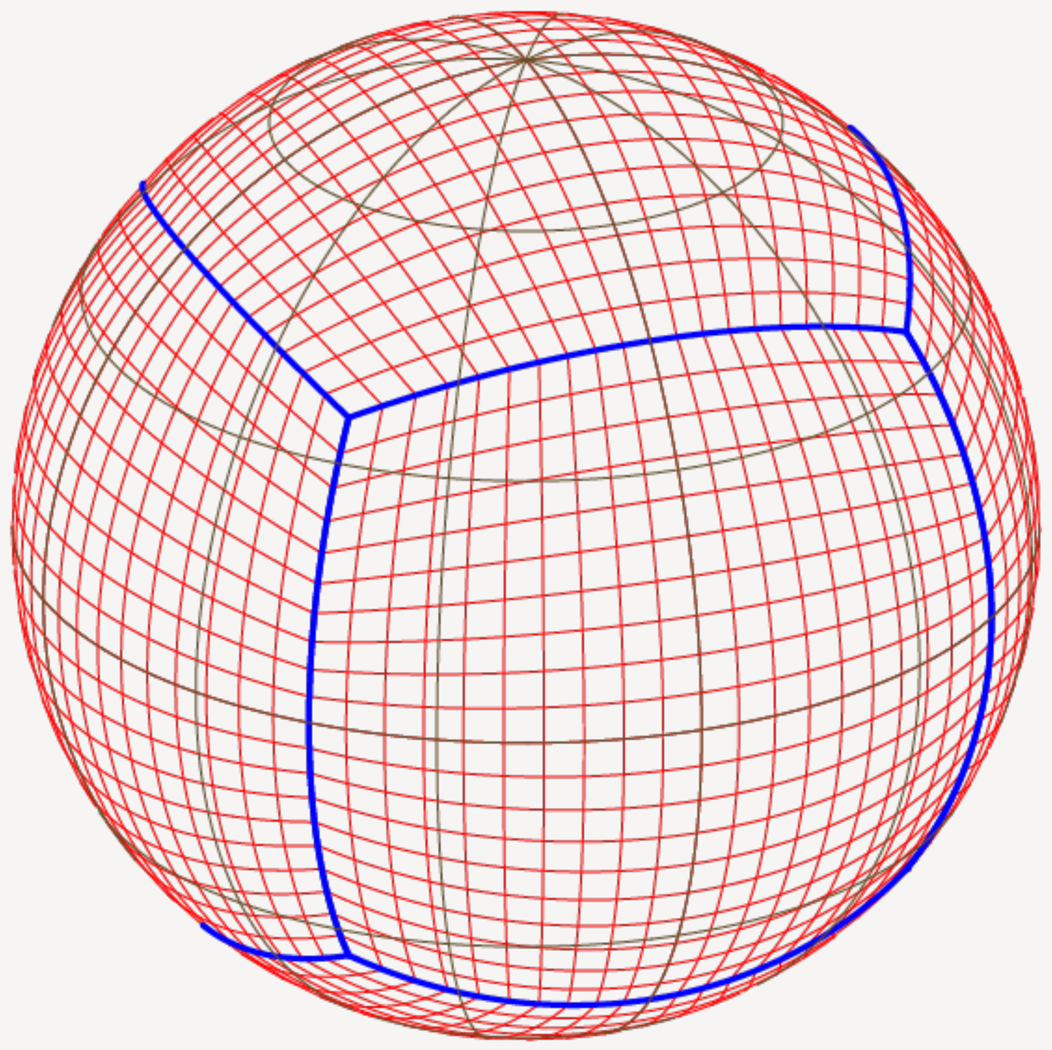

$J = J_1 + J_2$ ***Loss function is the sum of the MSE over 24 hrs***

# DLWP Building Blocks: Data

- ERA5: observations blended with NWP model output
  - Retrieved on 1° lat-lon grid
  - Re-gridded to cubed sphere (Ullrich & Taylor, 2015)
- Model *training*: 1979-2012
  - ~100,000 samples
- Model *validation* set: 2013-2016
- *Test* set: final performance evaluation: 2017-2018
  - twice weekly: 208 cases

# 2D Fields on Spherical Shells

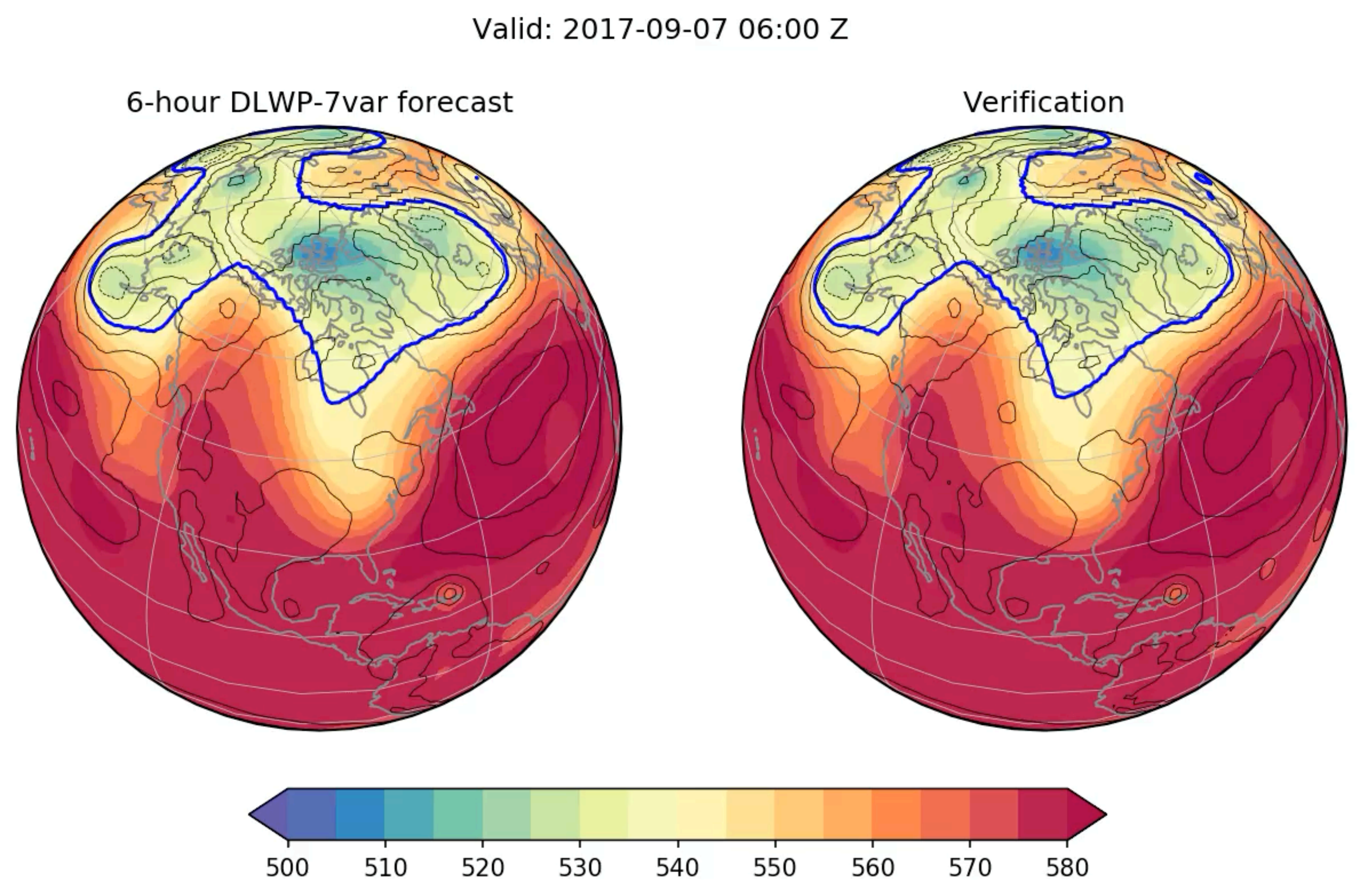
- 6 or 7 prognostic variables
  - 1000-hPa height
  - 500-hPa height
  - 300-700-hPa thickness
  - 2-m temperature
  - 850-hPa temperature
  - Total column water vapor
  - 250-hPa height
- 3 prescribed fields
  - TOA incoming solar radiation
  - land-sea mask
  - topographic height



- Resolution
  - 64x64 points on each face of the cube sphere (figure is 20x20)
  - ~1.4° x 1.4° at the equator

# Hurricanes Irma & Jose

- 4-day single model forecast
- 1.4° x 1.4° resolution
- 7 prognostic variables
- Showing
  - 1000-hPa height (black)
  - 500-hPa height (color fill)

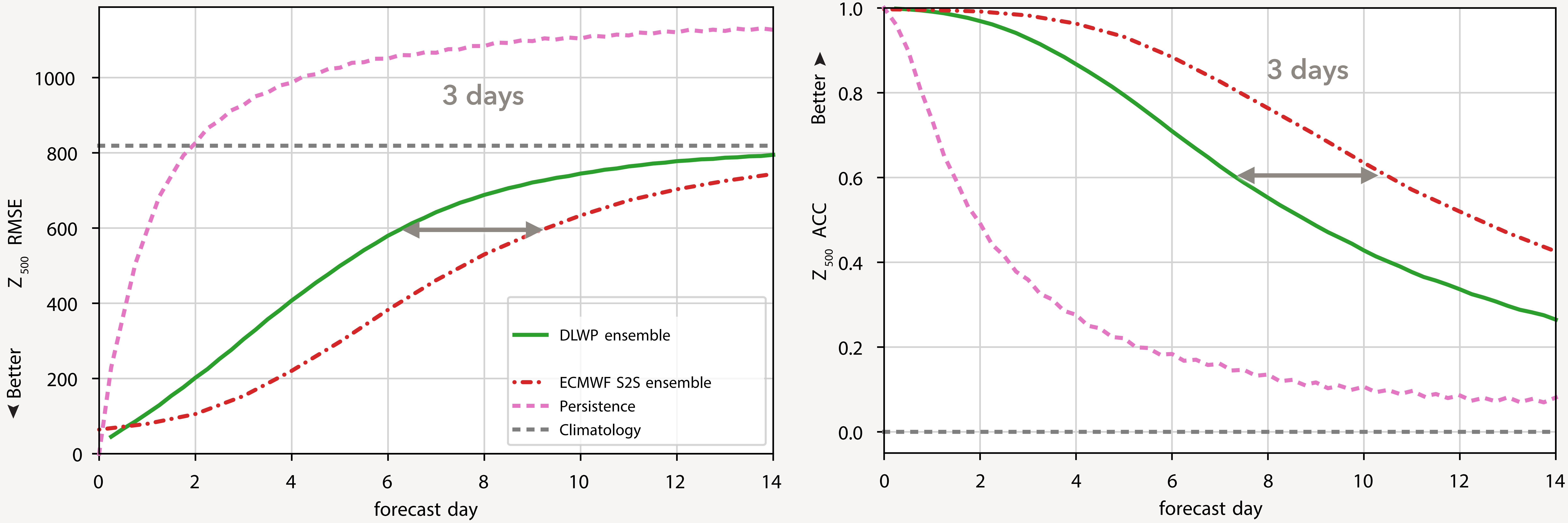


# DLWP-NWP Comparison

*Comparison of Key Attributes of Our DLWP Ensemble and Those of the State-of-the-Art ECMWF Ensemble for Extended-Range Forecasting*

| | DLWP | ECMWF |
|---|---|---|
| Atmospheric fields | 6 2-D variables | 9 prognostic 3-D variables; 91 vertical levels |
| Horizontal resolution | 150 km | 18 km (36 km after day 15) |
| Atmospheric physics | 3 prescribed inputs | Many physical parameterizations |
| Coupled models | None | Ocean, wave, and sea ice models |
| Initial condition perturbations | 10 (ERA5 uncertainty) | 50 (SVD/4DVAR) |
| Model perturbations | Perturbed CNN weights | Stochastic physics |
| Ensemble members | 320 (+control) | 50 (+control) |

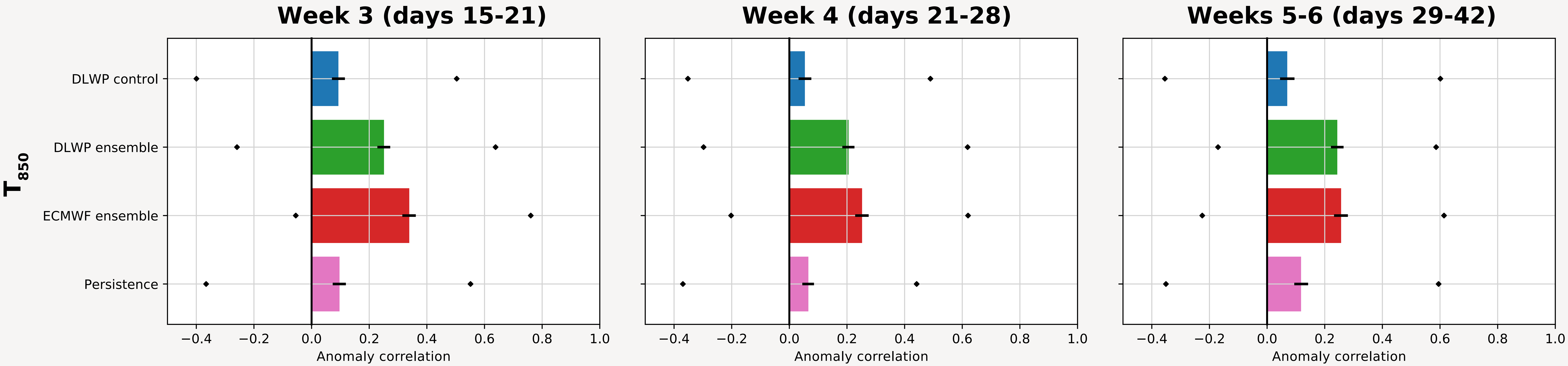
# Ensemble Performance: Deterministic Lead Times




*DLWP grand ensemble: 32 stochastically perturbed models x 10 initial conditions = 320 members*

# Ensemble Performance: S2S Lead Times

## Anomaly correlation coefficient of the ensemble mean

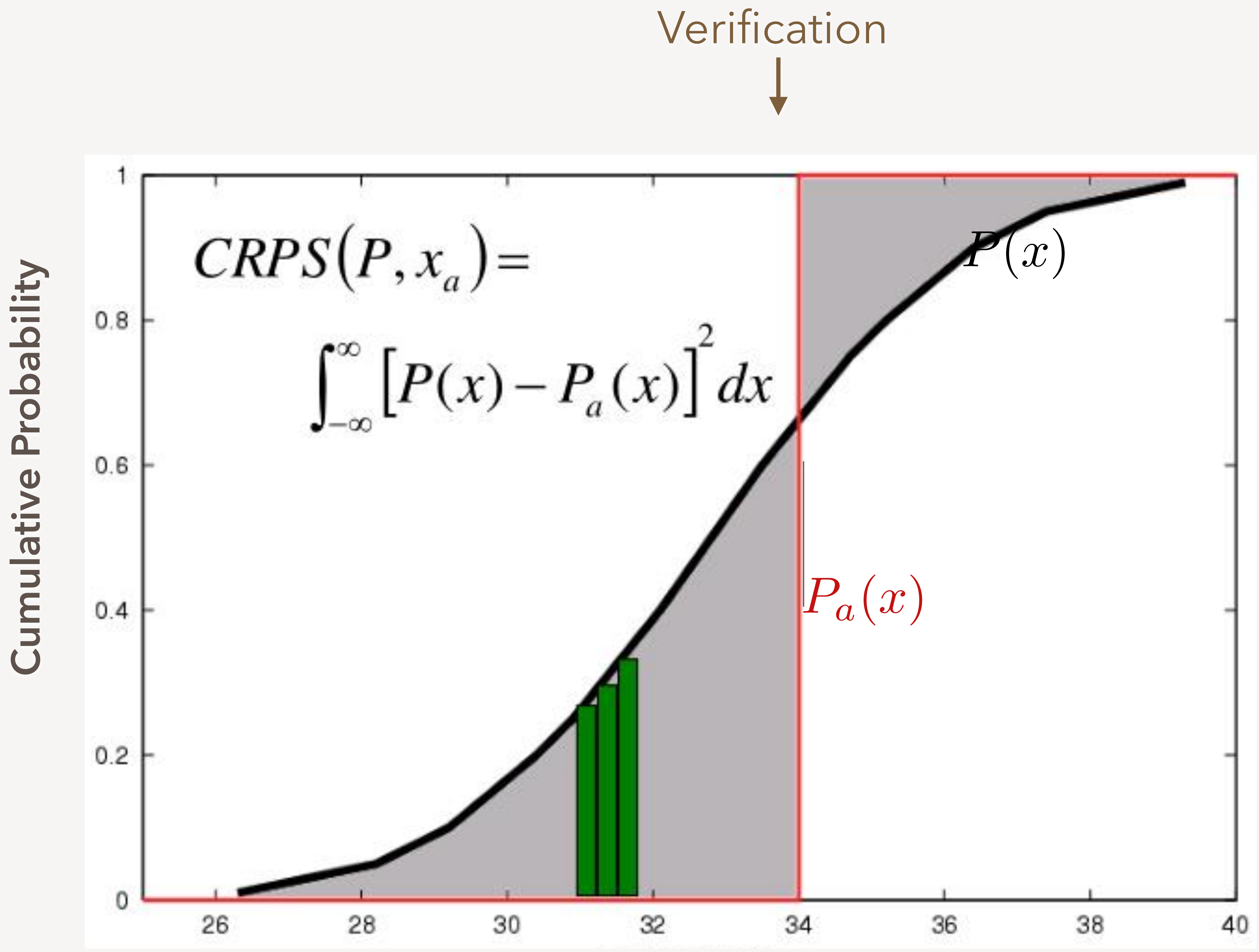



Persistence is computed as the 1- or 2-week-averaged anomaly just prior to the initialization

Black bar: 95% confidence interval. Black dots: best and worst forecast.

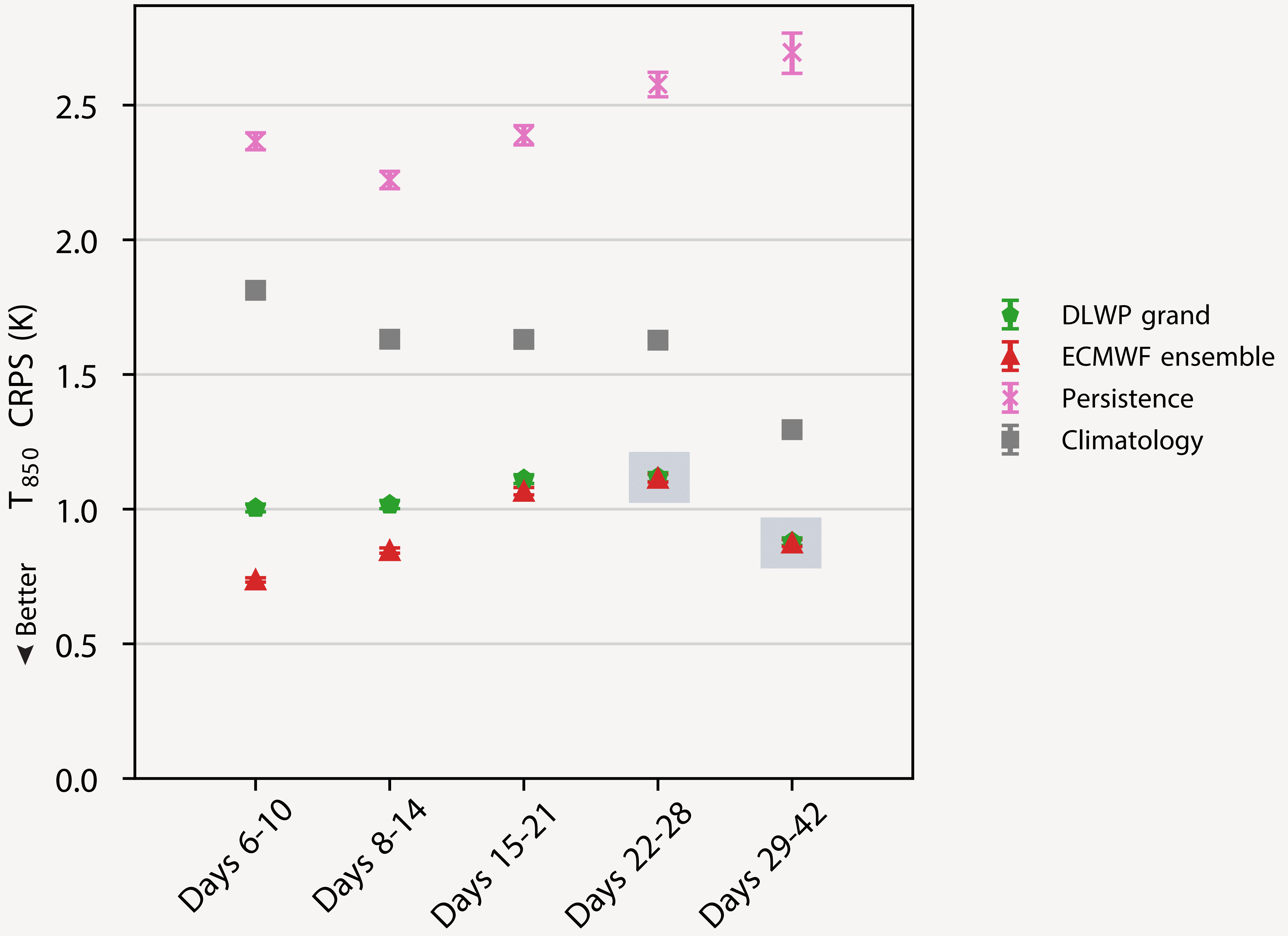
# PROBABLISTIC SCORES OF ENSEMBLE FORECASTS

- *Continuous ranked probability score* (CRPS)
  - Evaluates the integrated square error between the forecast and observed cumulative probability distribution
  - Reduces to *mean absolute error* for a deterministic forecast
- Dimensional score, lower is better.



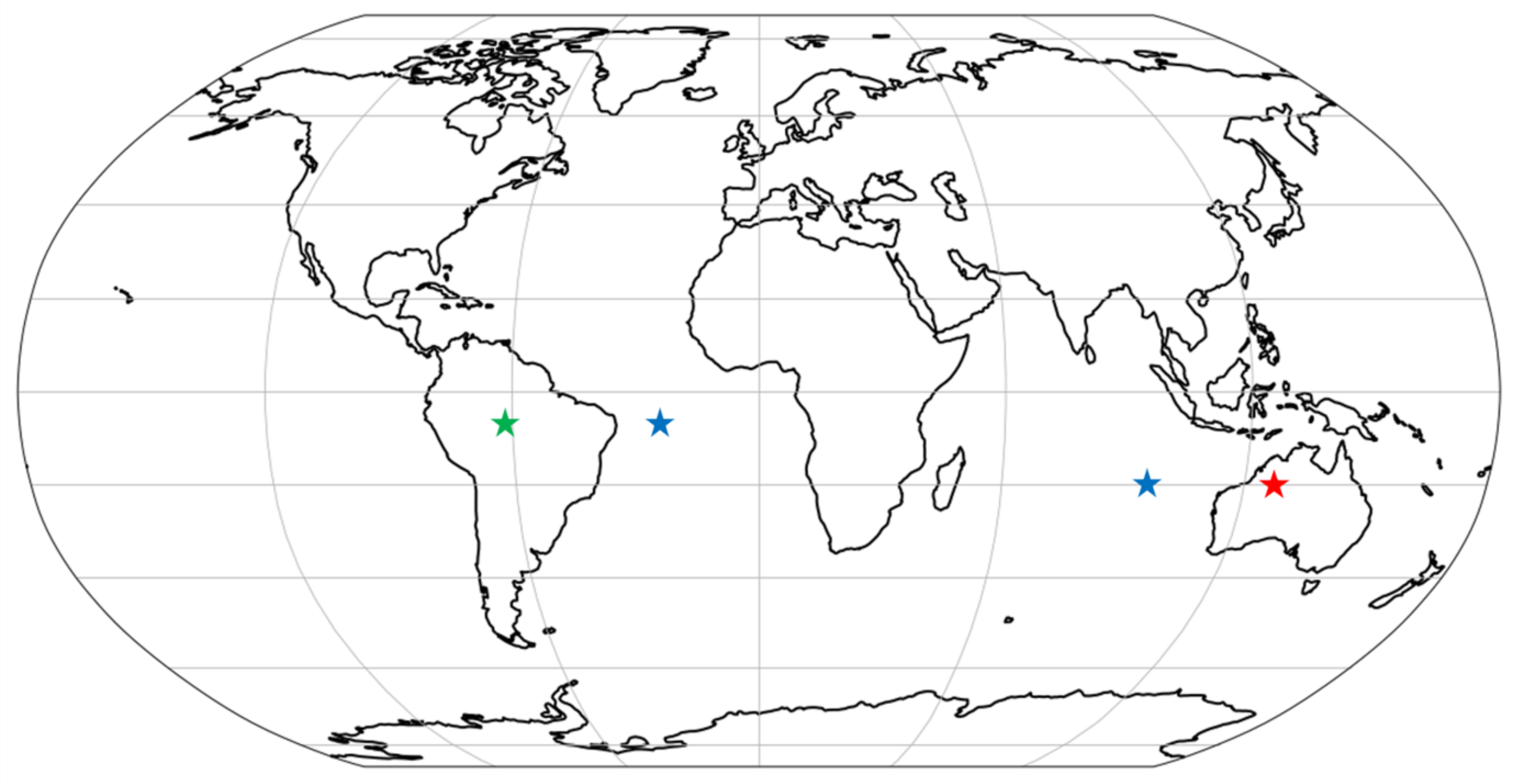
# CRPS: DLWP vs Current ECMWF S2S Ensemble

- Global, annual average
- *DLWP & ECMWF tied in week 4 and weeks 5-6*
- *Both ensembles beat persistence and climatology*

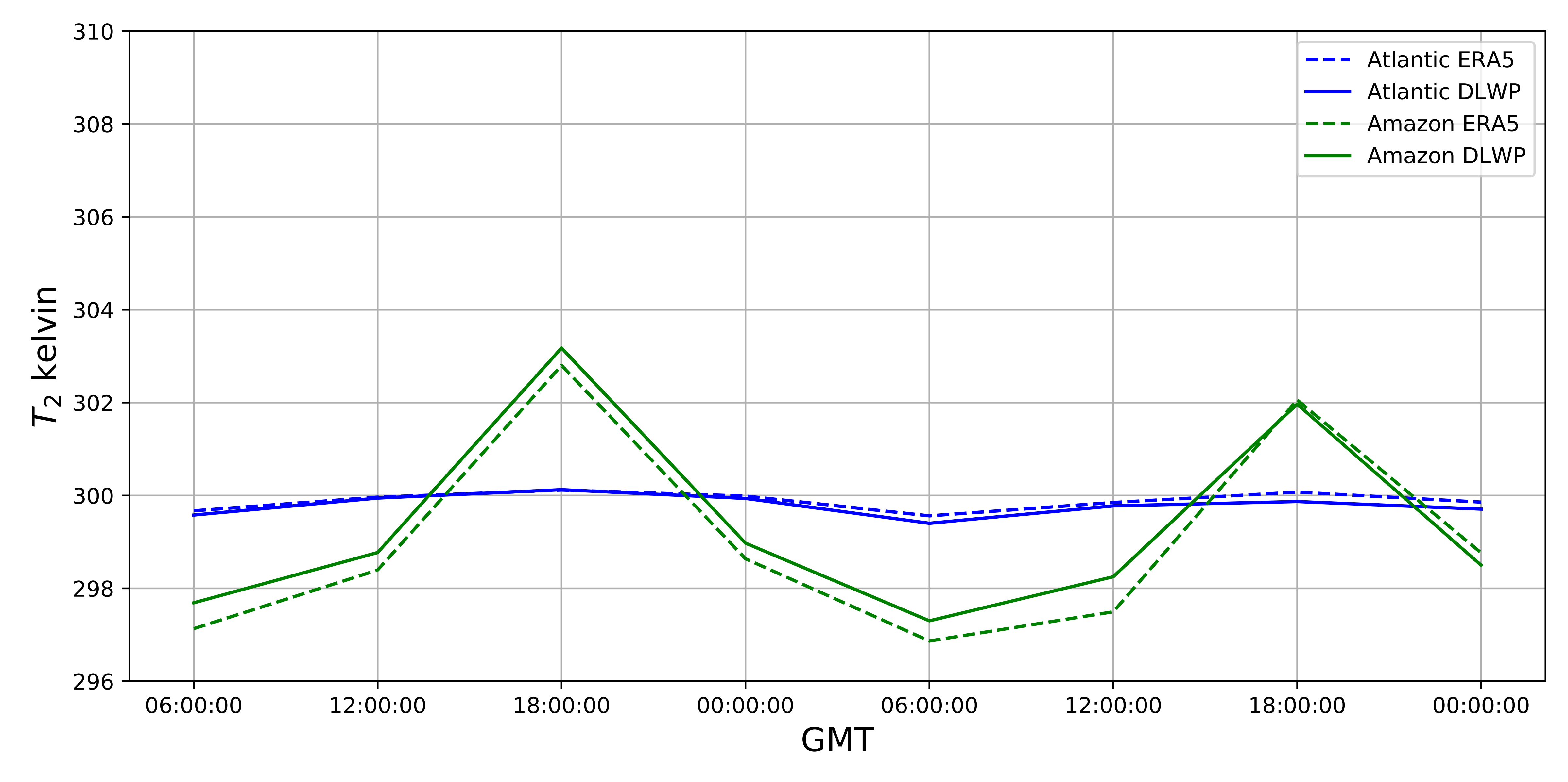


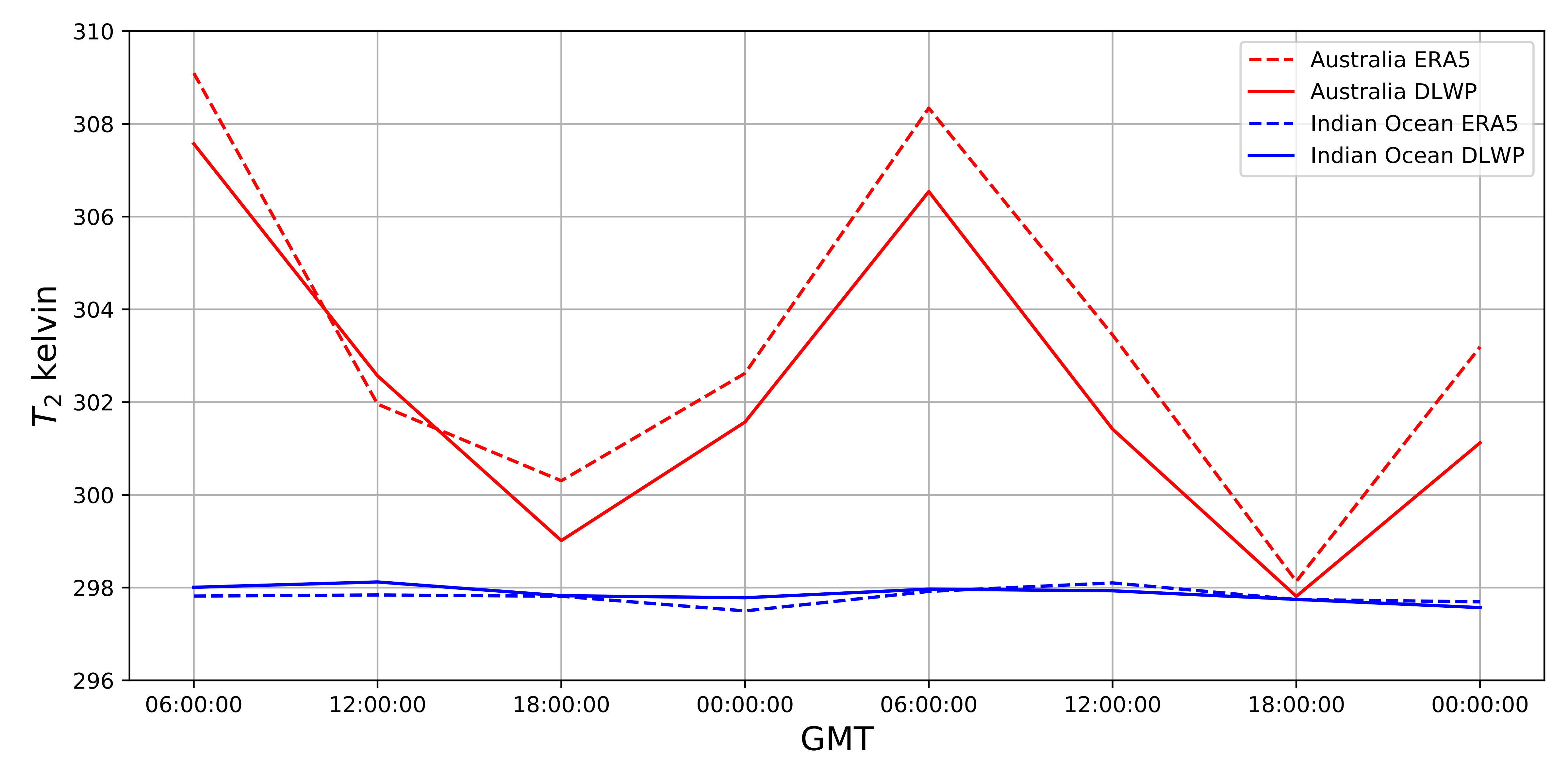
# Near Surface Predictions without Planetary Boundary Layer Parameterization

- 2-m temperature
- 2-day forecast
- Initialized March 11, 2018 at 00 UTC
- 2 paired sites
  - Amazon & ocean
  - Australia & ocean



# 2-m Temperature Forecasts

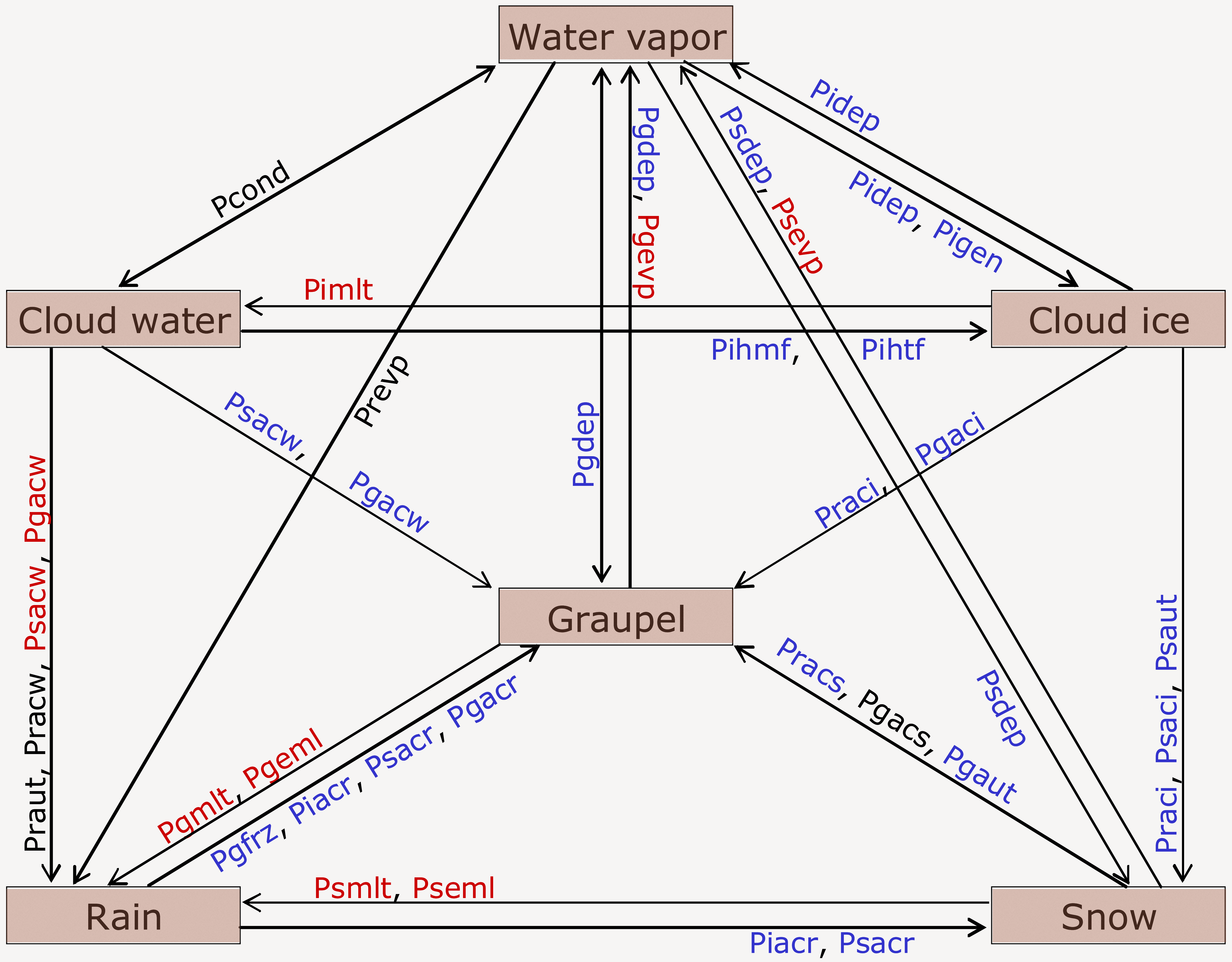



- Little temperature variation over oceans (land-sea mask)
- Larger diurnal variations over Australia than the Amazon (total column water vapor)

# Parameterization of Cloud and Precipitation Processes in NWP

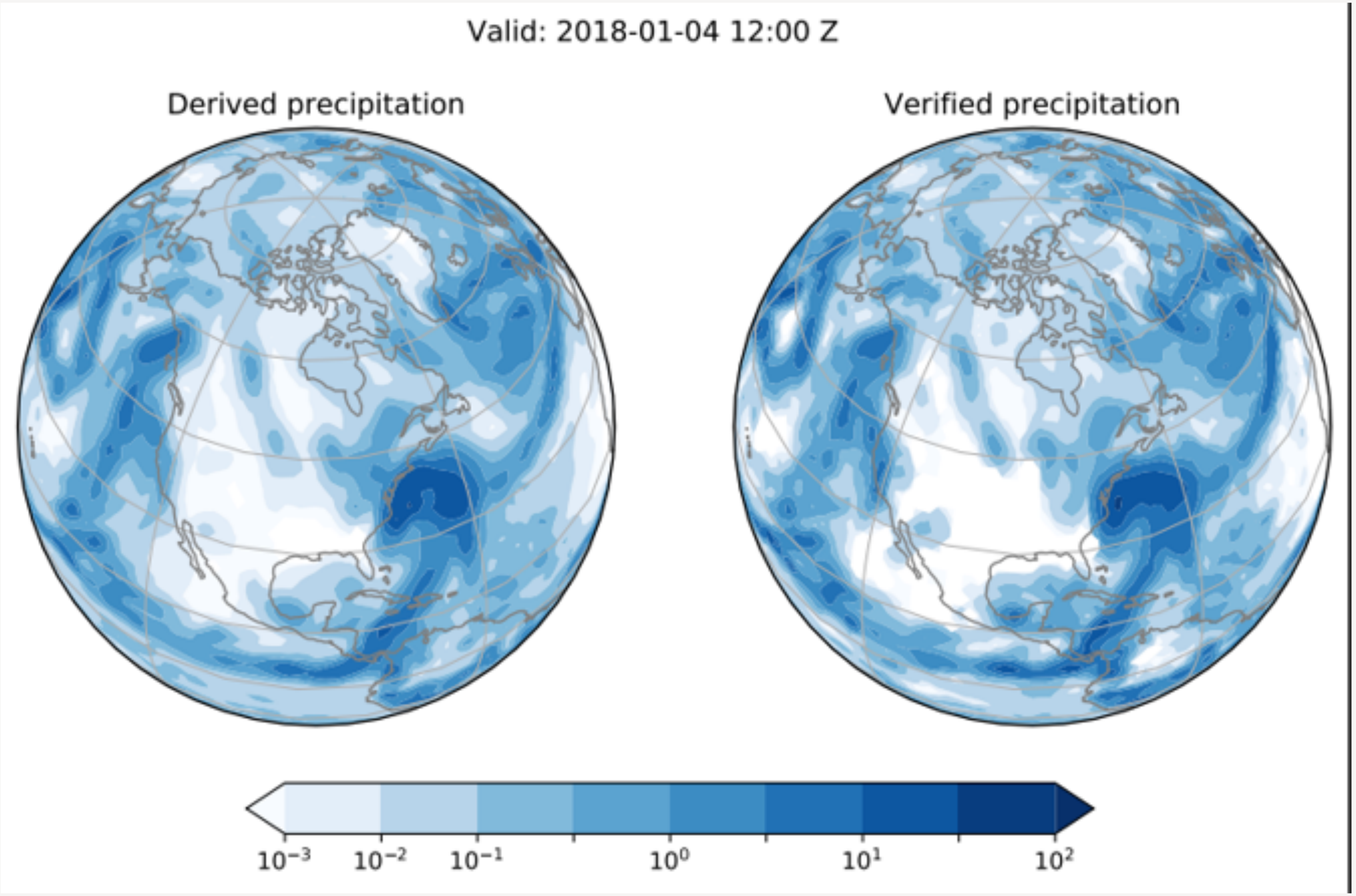
6 variables




*Hong and Lim, 2006*

# U-NET DIAGNOSIS OF PRECIPITATION

- Same 6 variables as prognostically forecast in DLWP model
- But precipitation is diagnosed from the ERA5 analysis
- Can be used to diagnose precipitation in DLWP forecasts



# Conclusions

- DLWP has the potential to revolutionize weather forecasting, echoing of the impact produced by the introduction of NWP in the 1950's
  - Data-driven AI-based weather prediction has been enabled by advances in algorithms and hardware.
- Can learn dynamics and physical parameterizations at the same time.
- The speed of DLWP allows use of much larger "ensembles" of near-twin forecasts.
  - Large well-calibrated ensemble would
    - Better define the probable distribution of future atmospheric states
    - Better capture extreme events.
  - 1-week forecast stepped forward with 12-hr time step (and 6-hr resolution) requires just 1/10 of a second on one Nvidia V100 GPU

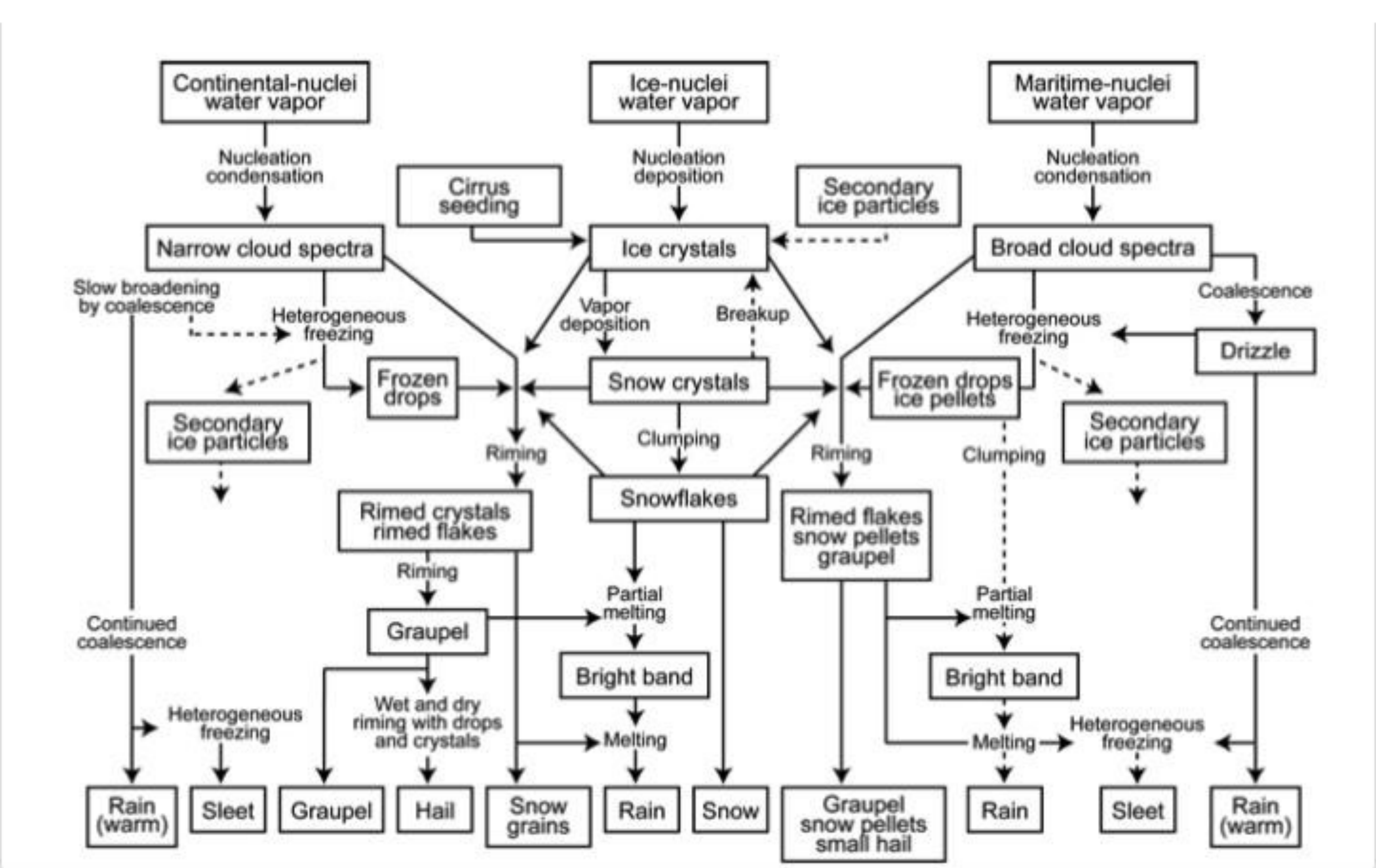
# Further Reading:

Weyn, J.A., Durran, D. R., Caruana, R., and Cresswell-Clay, N. (2021). Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. *J. Adv. Modeling Earth Sys*, 13, e2021MS002502 https://doi.org/10.1029/2021MS002502

Weyn, J. A., Durran, D. R., & Caruana, R. (2020). Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *J. Adv. Modeling Earth Sys*, 12, e2020MS002109 https://doi/10.1029/2020MS002109

Weyn, J. A., Durran, D. R., & Caruana, R. (2019). Can machines learn to predict weather? Using deep learning to predict 500 hPa geopotential height from historical weather data. *J. Adv. Modeling Earth Sys*, 11, 2680-2693. https://doi/10.1029/2019MS001705

# CLOUD AND PRECIPITATION PROCESSES: UNDERLYING PHYSICS




*from Dudhia, Overview of WRF Physics*