Toward Replacing Current NWP with Deep Learning Weather Prediction

DALE DURRAN, JONATHAN WEYN, NATHANIEL CRESSWELL-CLAY, RICH CARUANA
APRIL 2022

Modern NWP: Dynamical Core + Parameterizations

- Dynamical core: equations for conservation of mass, energy and momentum …
  - Large-scale air currents
  - Numerical approximation can be evaluated for order of accuracy, stability, …
- Operational Models Rely on Parameterizations
  - Clouds and precipitation
  - Influence of the earth's surface
- *Parameterizations are evaluated empirically!*

Advantages of Deep-Learning Weather Prediction

- NWP is *very* compute intensive
  - Difficult to produce many simulations of likely future states (large *ensembles*)
- DLWP
  - Reduces the time required for each forecast by 2-3 orders of magnitude
    - Allowing a *large* number O(1000) of ensemble members
    - Potentially better defining the probable distribution of future atmospheric states
      - Better estimating the chances of extreme events
  - Can replace empirical NWP-style parameterizations with `holistic' machine learning.
    - Potentially crucial for the S2S forecasting problem

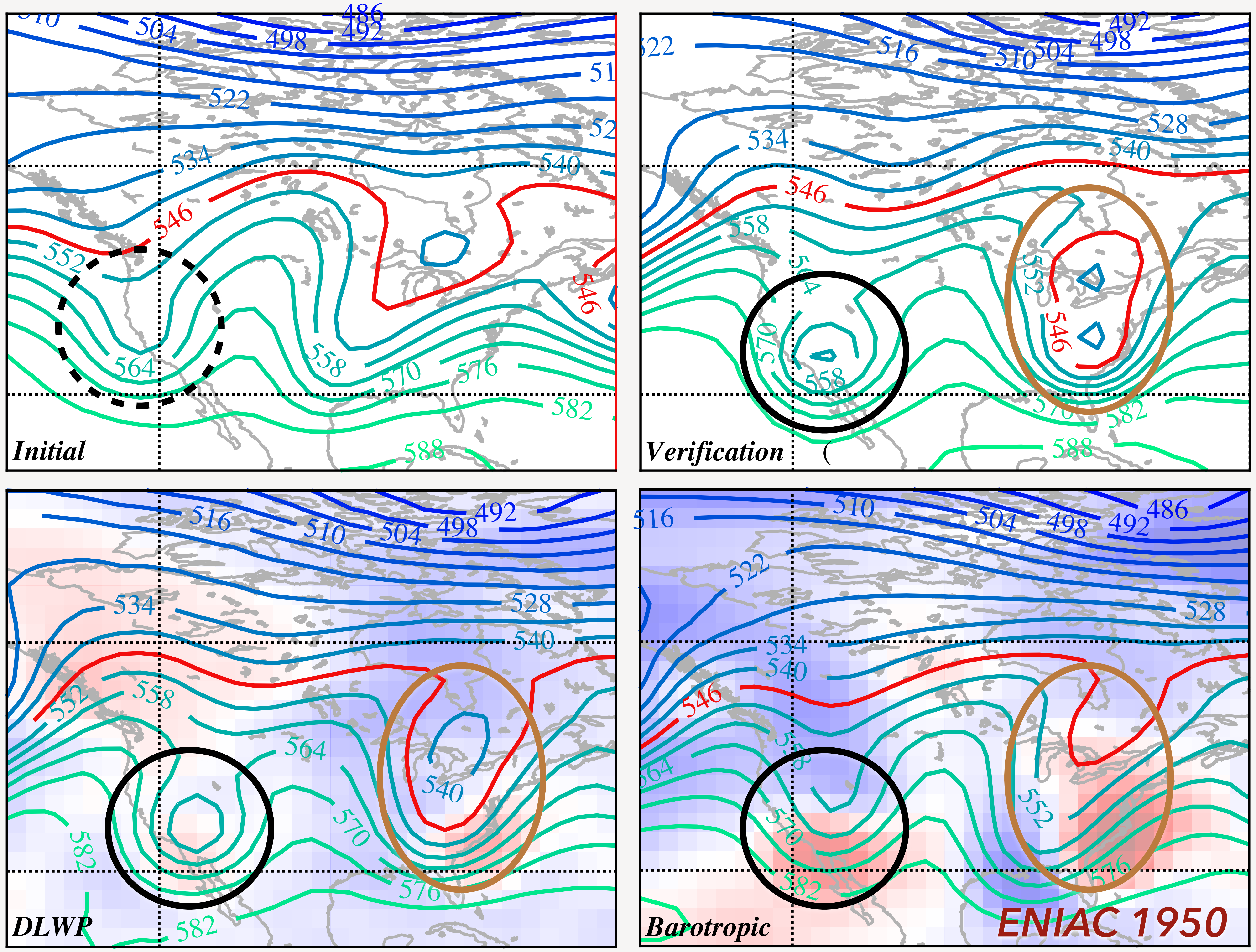
Our DLWP Strategy

- Like NWP, create forecasts by recursively stepping forward in time
- Use just a few variables and coarse resolution to characterize the atmospheric state
    - It’s a starting point for DLWP
    - Limitations on atmospheric predictability suggest the space-time resolution required for NWP likely introduces *more degrees of freedom than necessary* to characterize the atmospheric state
        - For *convergence*: $\Delta z < 200$ m if $\Delta x < 15$ km (Skamarock et al., 2019)

The potential of DLWP

- Predicting 500 hPa height using *only* 500 hPa height data
- Machine learning beats basic equation-based NWP
  - Color fill shows error

Neural net learned to do better than the incomplete dynamical model.

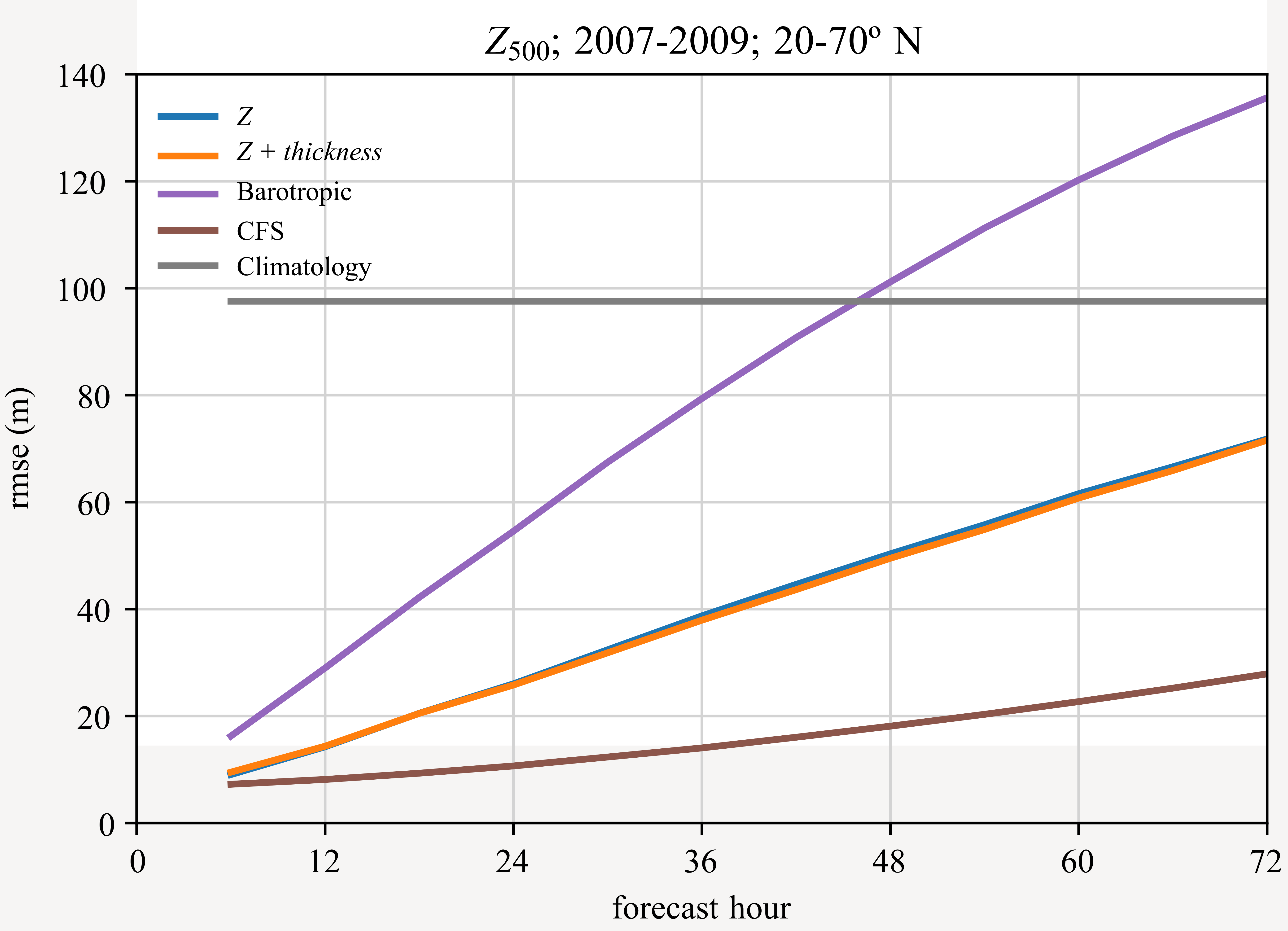


Weyn, et al., 2019

Adding Baroclinity Data (Thickness)

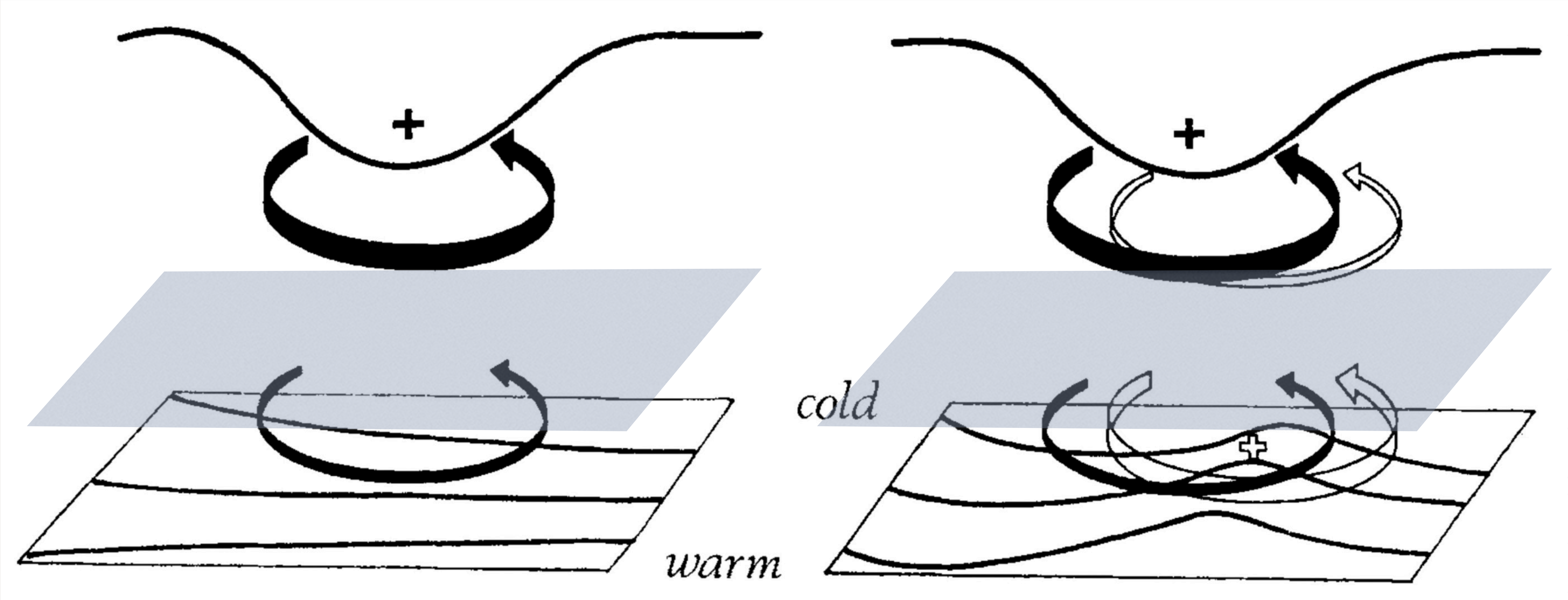
- Z_{500} and 700-300 hPa thickness
- Test set is 2007-2009
- RMSE over 20–70° N
- 2.5 x 2.5° resolution

700-300-hPa thickness adds little skill



Weyn, et al., 2019: Can machines learn to predict weather.…

Intensification Determined by Single-Level Data ???

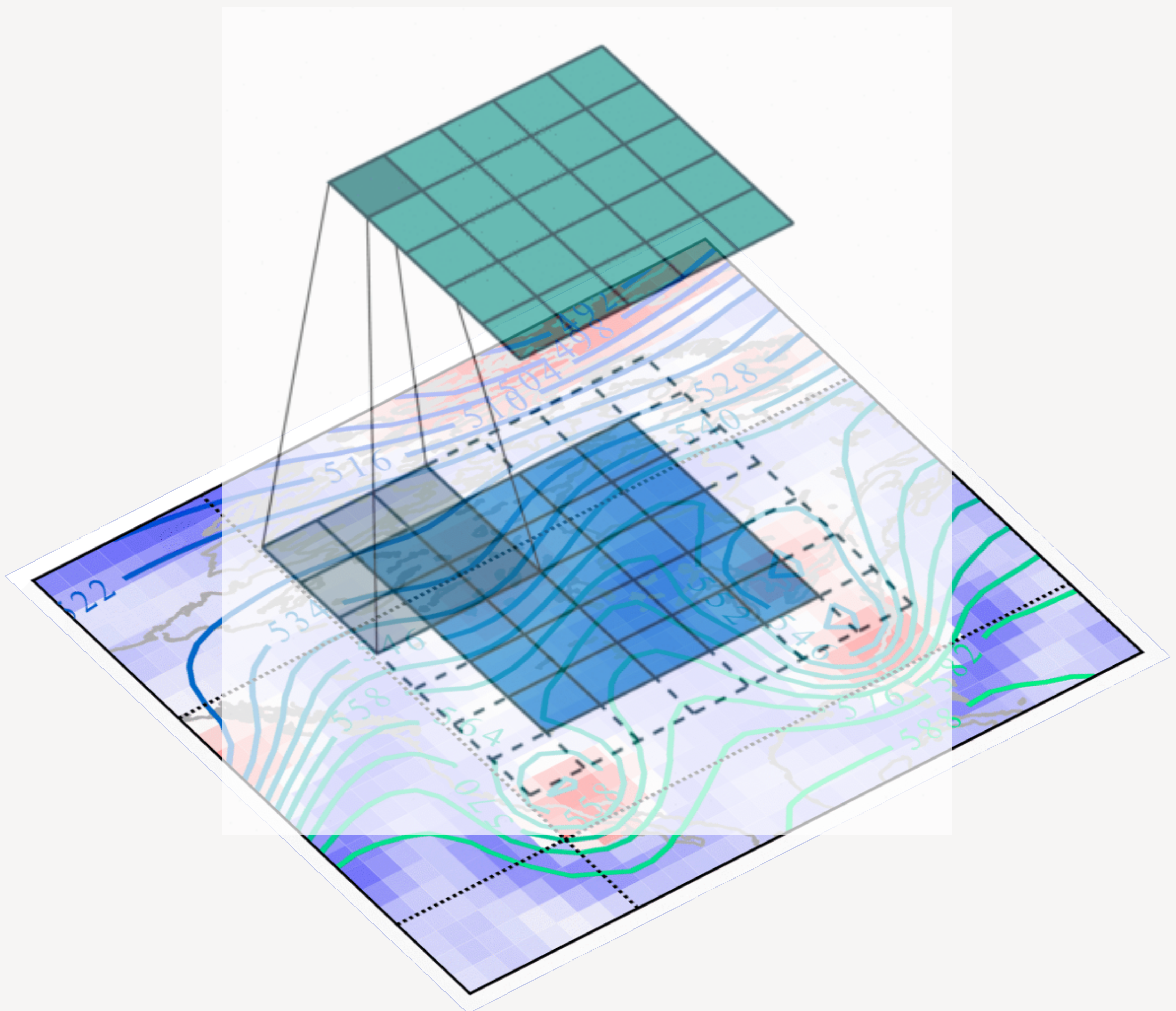


Hoskins et al., 1985: On the use and significance of isentropic potential vorticity maps

500 hPa heights don't obviously determine upper and lower level phasing

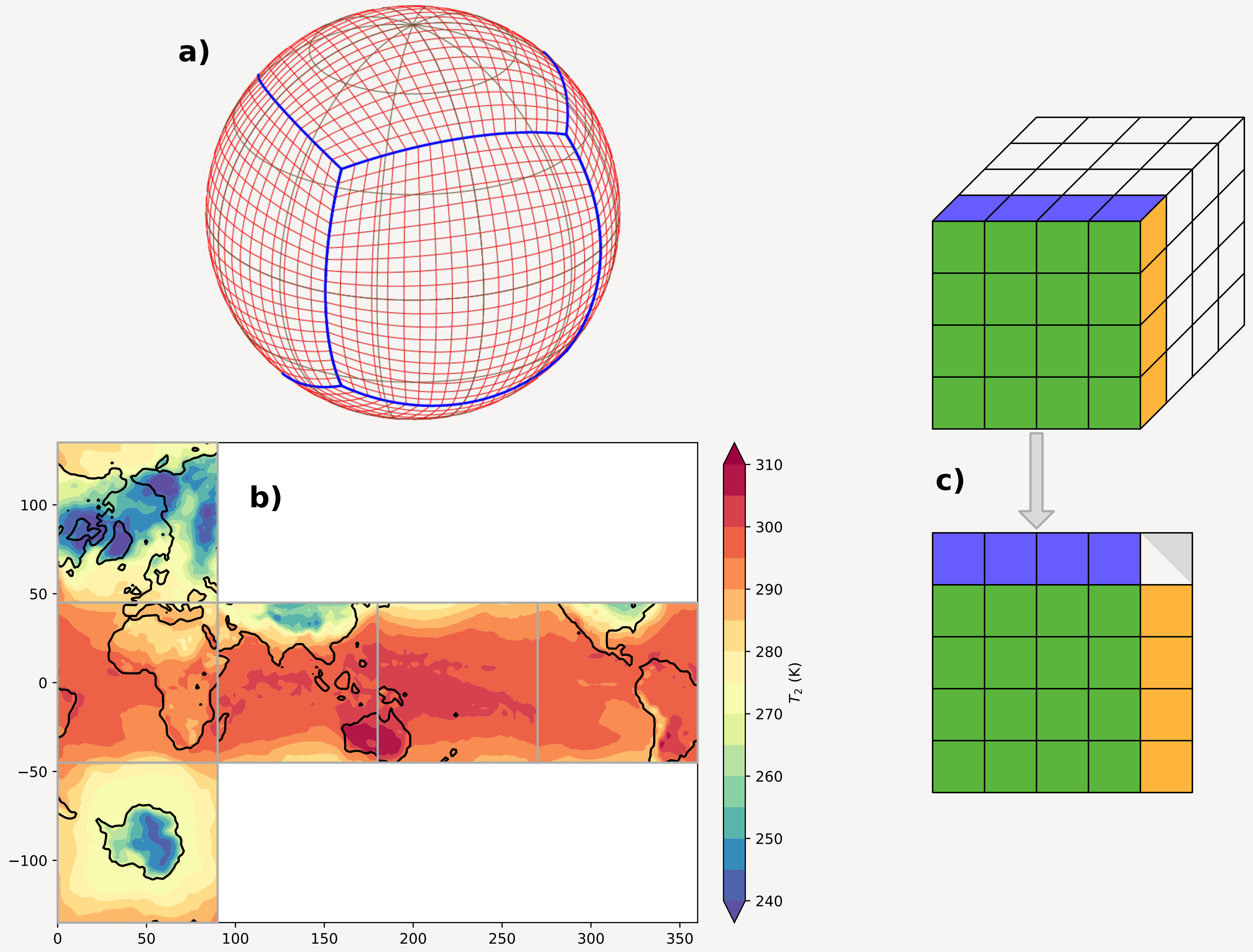
DLWP Building Blocks: Convolutional Neural Networks

- Same filter coefficients multiply the input data at every point
  - 3x3 horizontal stencil
- Learn many sets of these filter coefficients (64/128/256)



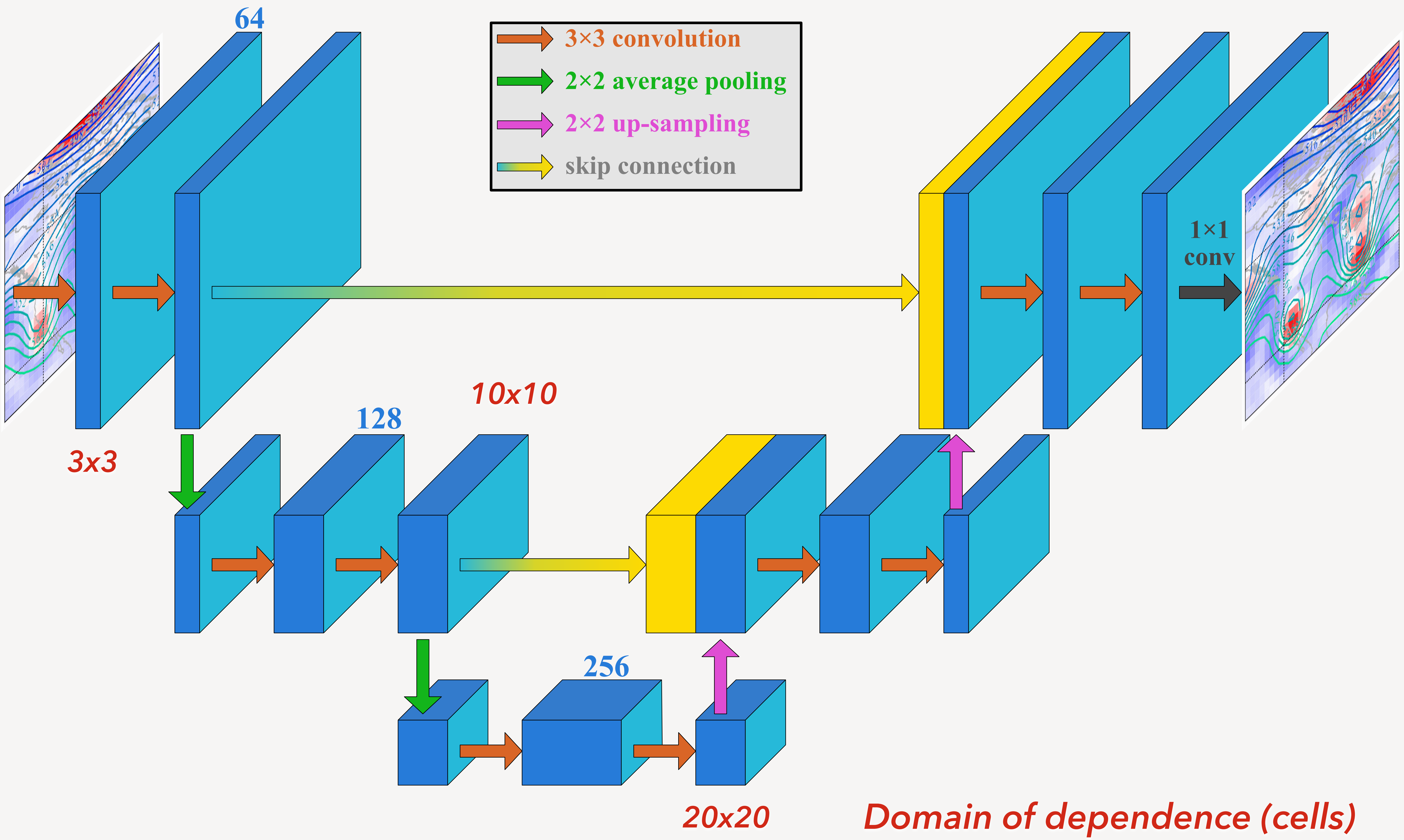
DLWP Building Blocks: Cubed Sphere Grid

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



Weyn, et al., 2020

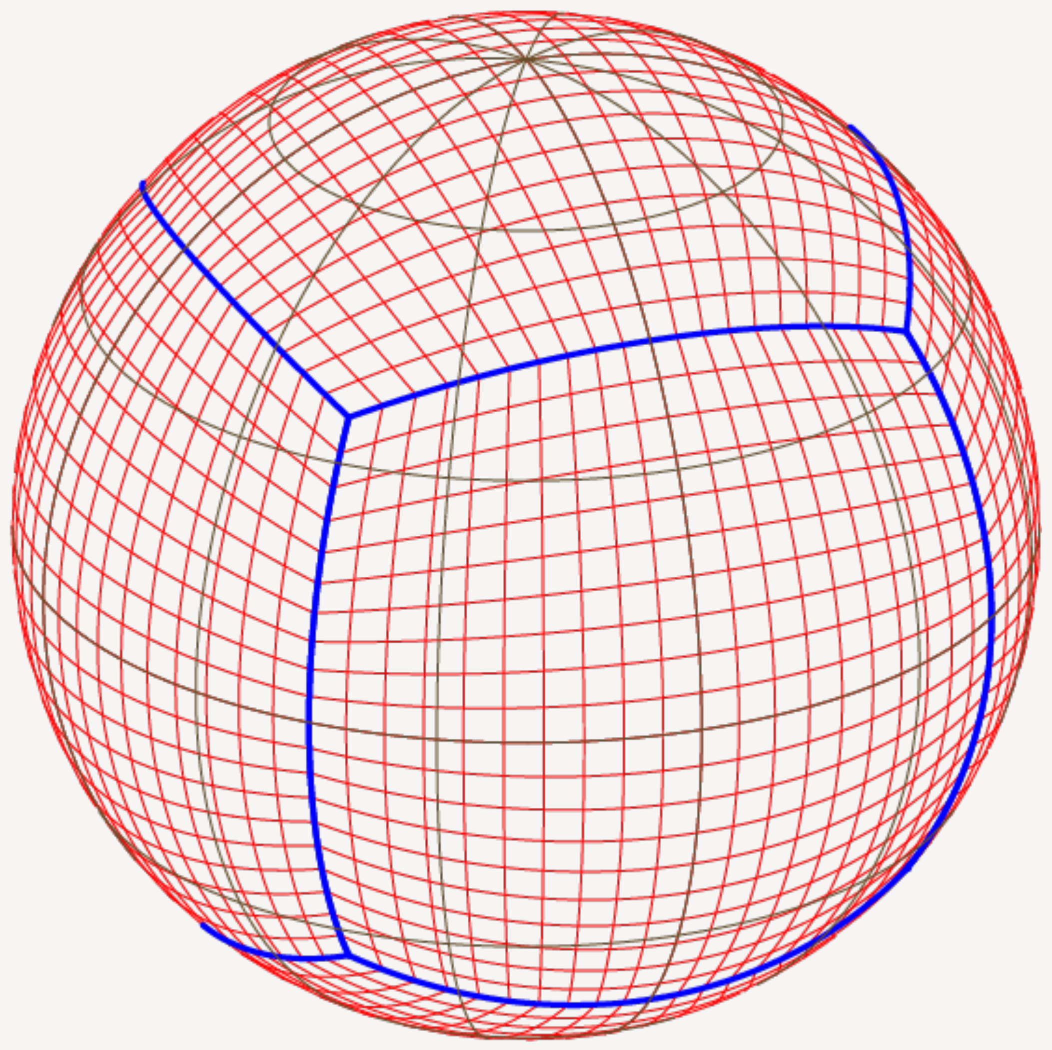
U-NET ARCHITECTURE: NO GEO-SPECIFIC TRAINING



Domain of dependence (cells)

OUR FIELDS ARE 2D ON SPHERICAL SHELLS

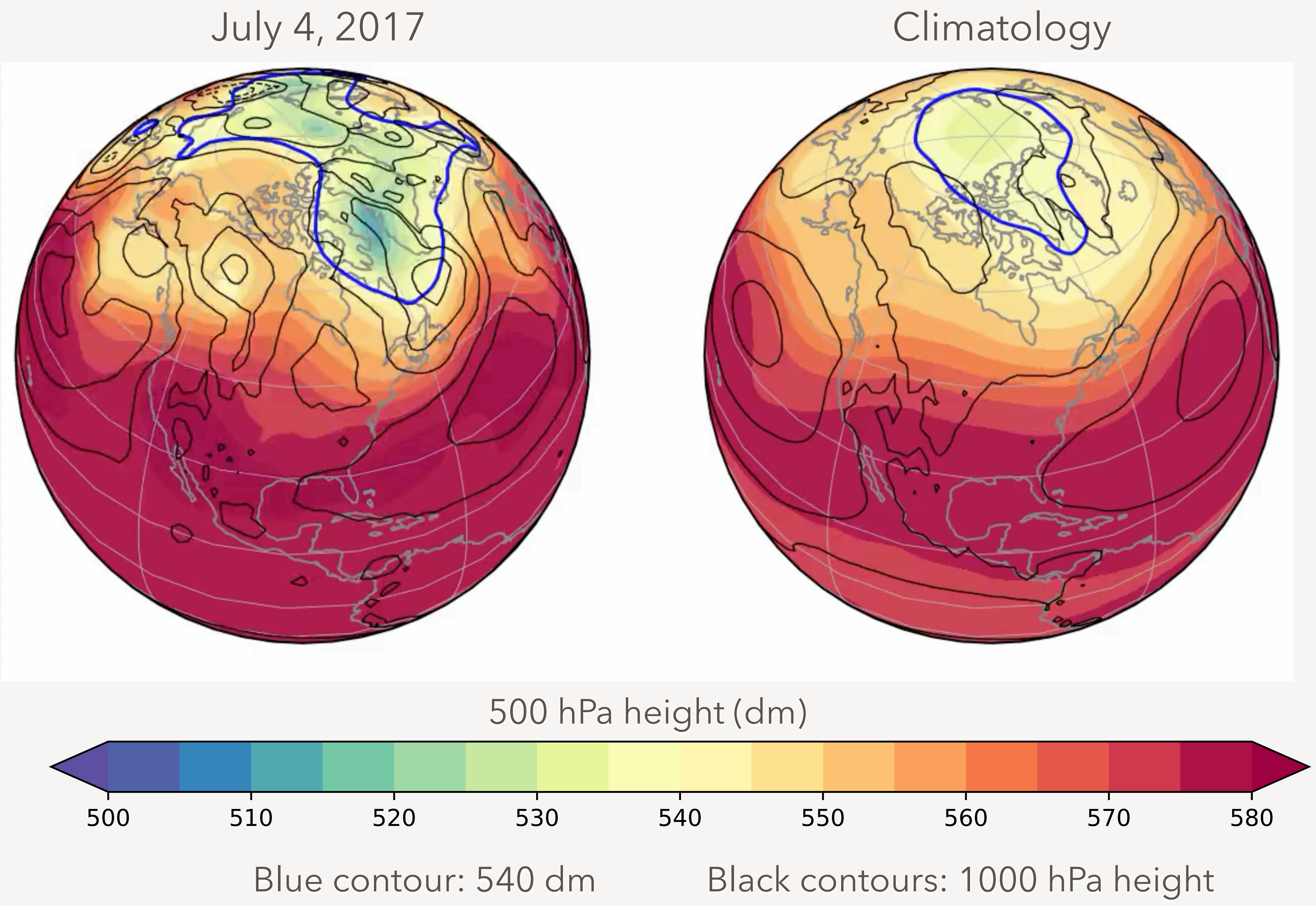
- 4 prognostic variables
  - 1000-hPa height
  - 500-hPa height
  - 300-700-hPa thickness
  - 2-m temperature
- 3 prescribed fields
  - TOA incoming solar radiation
  - land-sea mask
  - topographic height



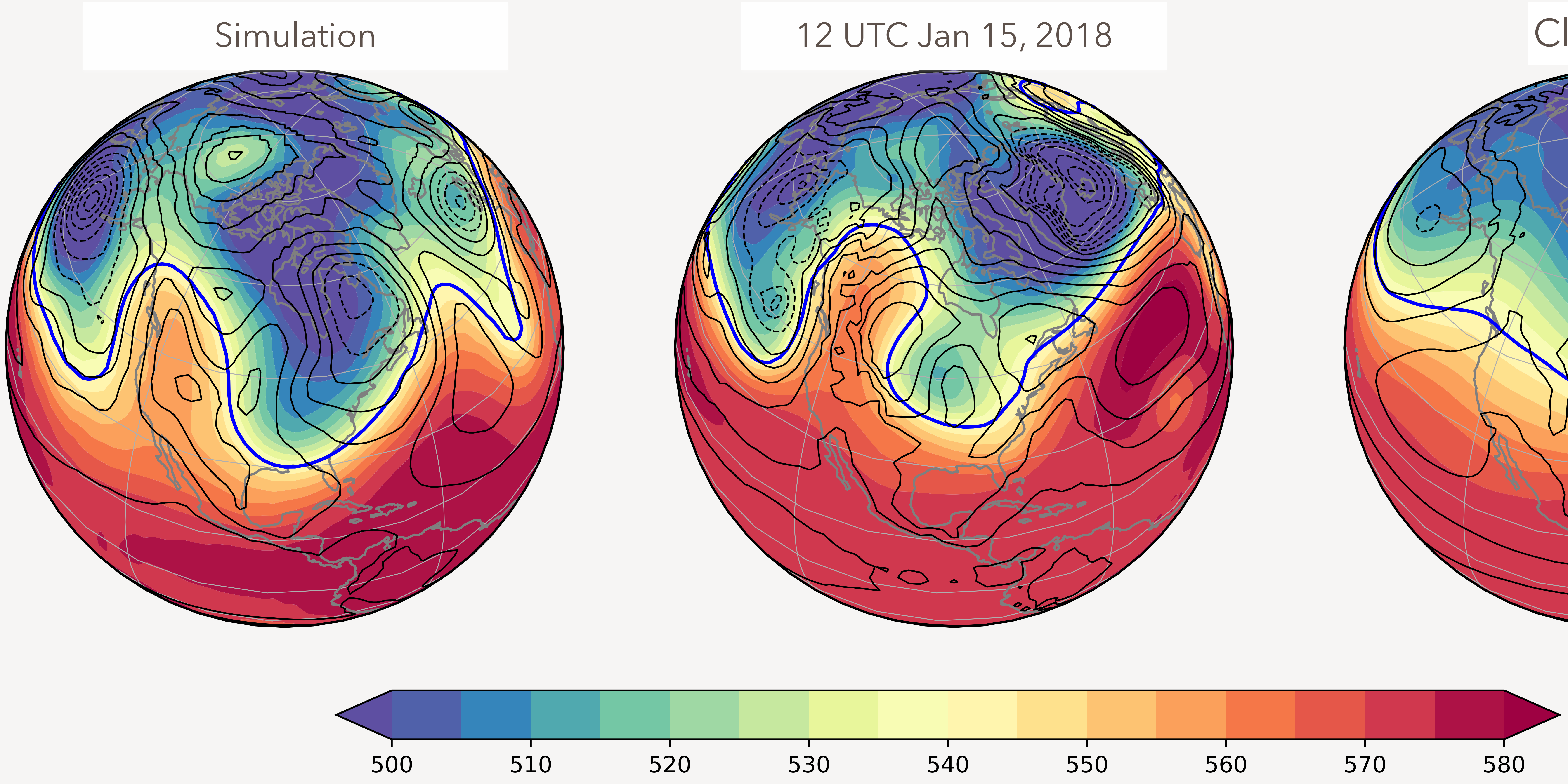
- 48 x 48 points on each face of the cube sphere (figure is 20x20)
- ~200 x 200 km quasi-uniform over the globe

Initial Condition: Free Running 195 Day Simulation

- Loss function
  - MSE every 6 hrs over 1 day
- Time step: 12 hrs
- No physical constraints



You Be the Discriminator! Free Running 195 Day Simulation

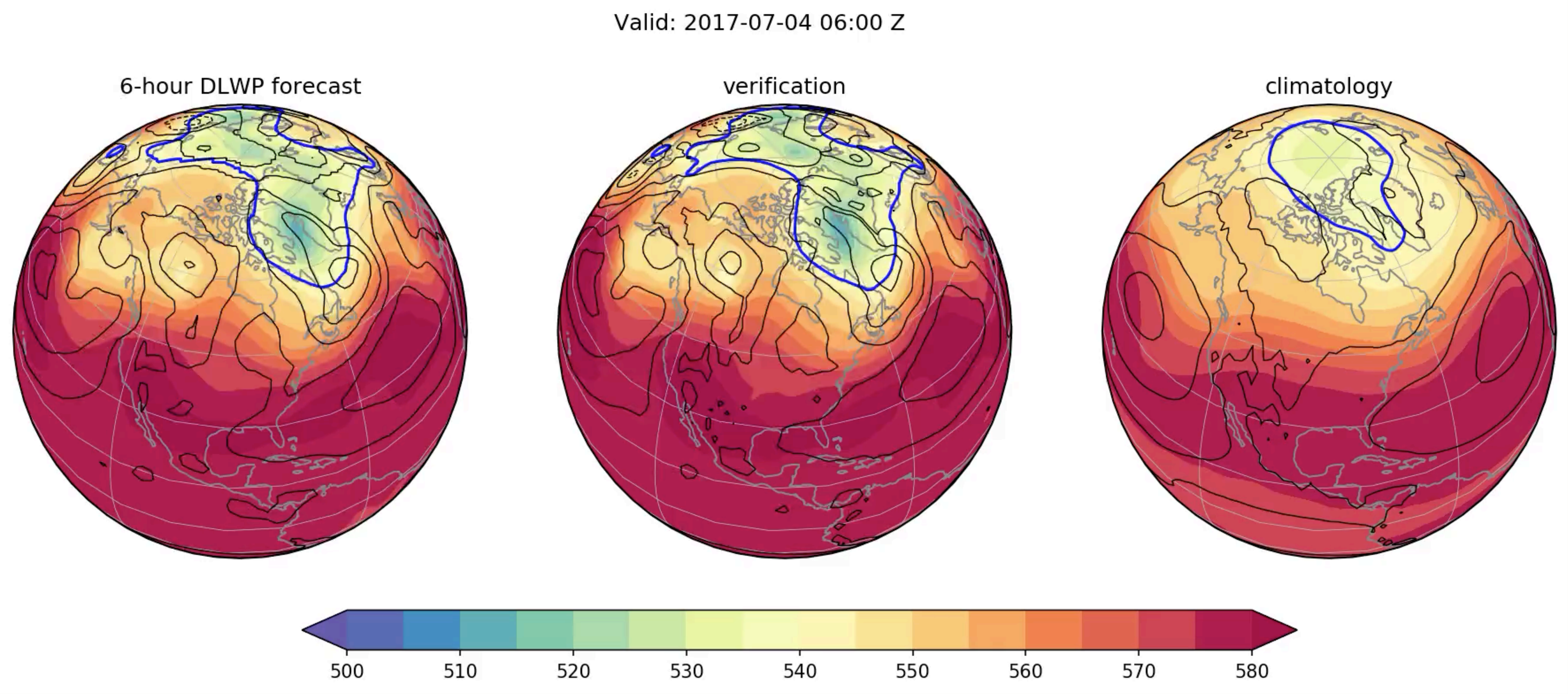


Blue contour: 540 dm

500 hPa height (dm)

Black contours: 1000 hPa height

Animation: Free-Running 1-Year Forecast

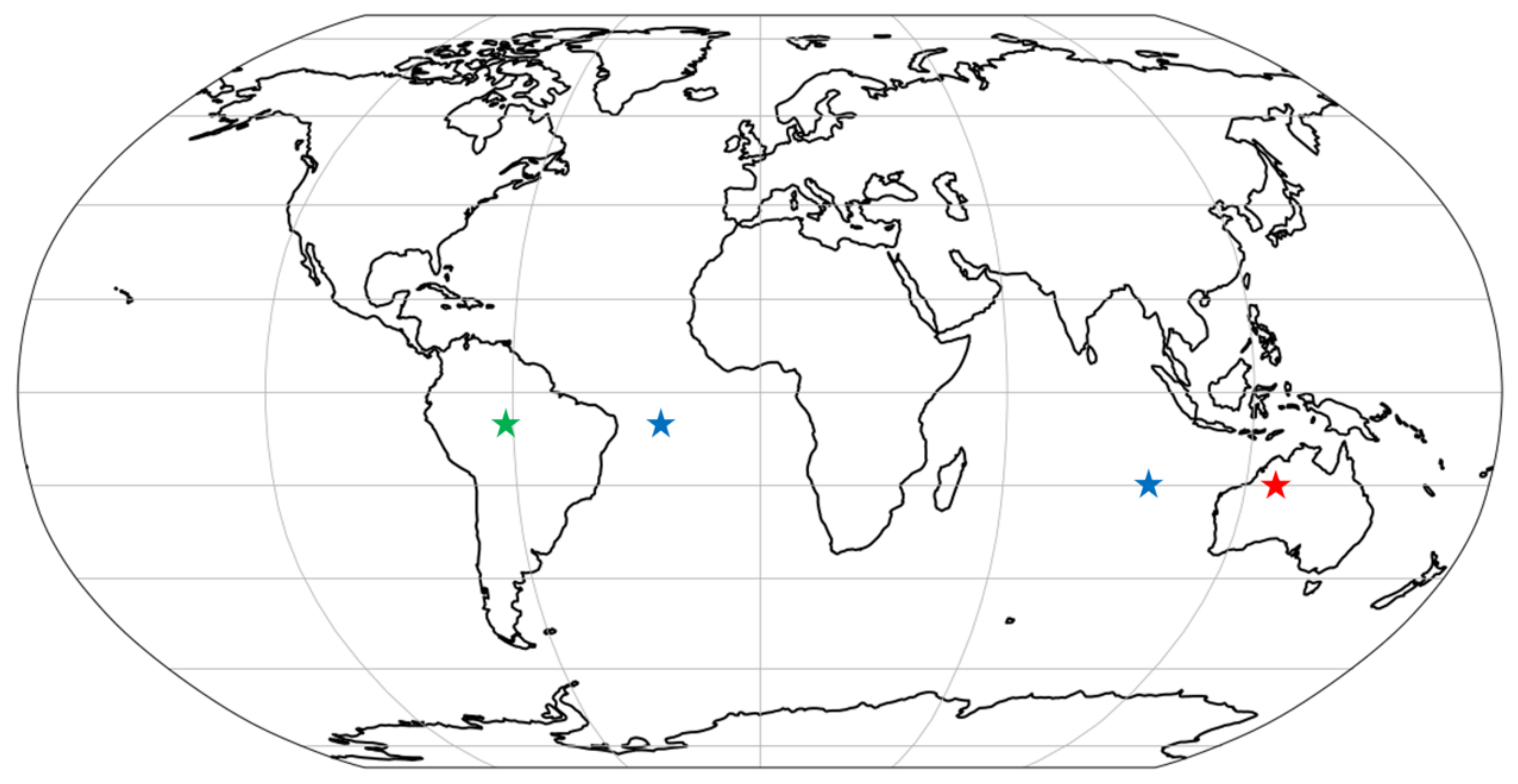


NEAR-SURFACE PREDICTIONS WITHOUT A BOUNDARY-LAYER PARAMETERIZATION

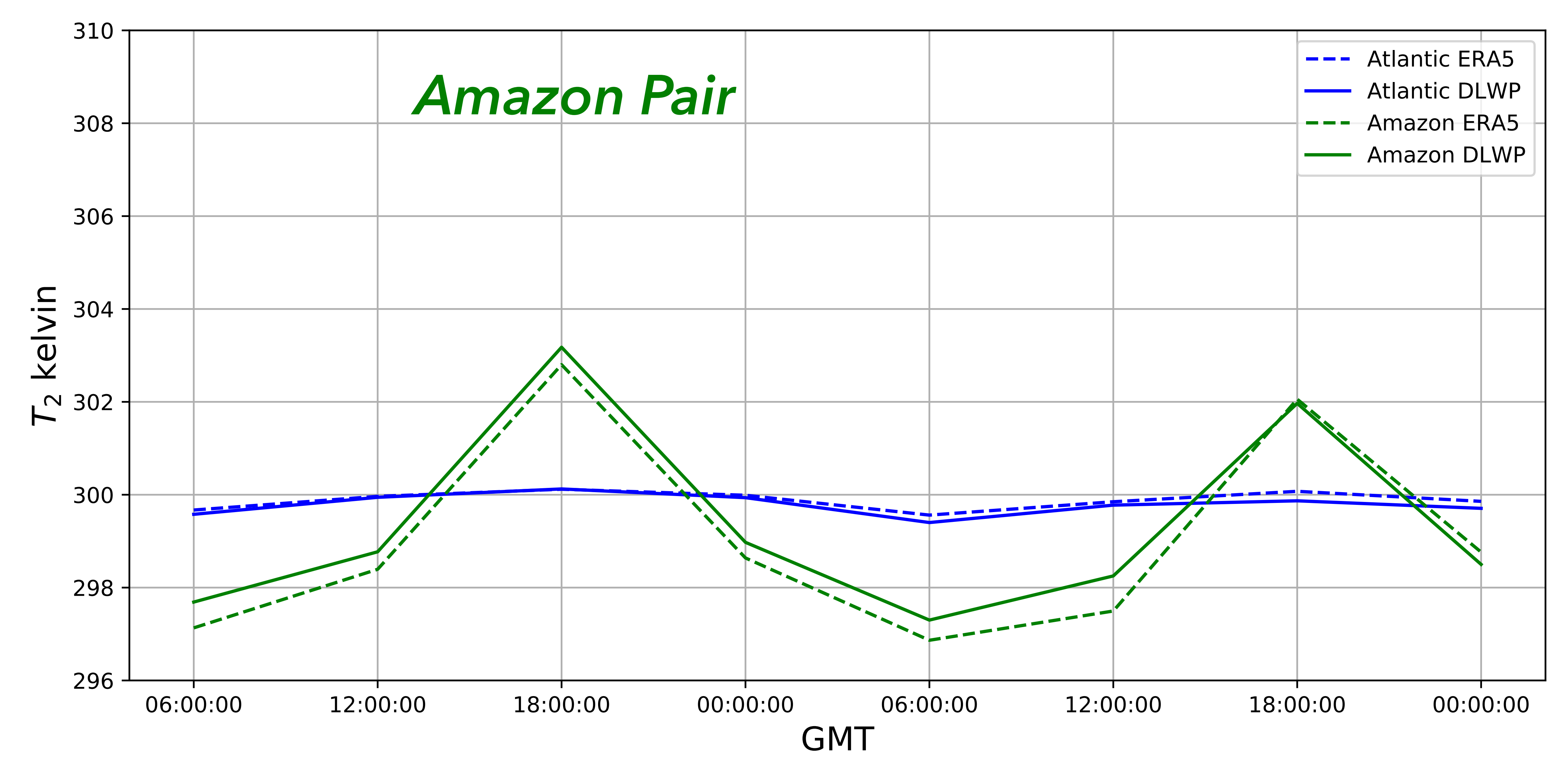
- *Model with 7 spherical shells of data*
  - *Total column water vapor*
- *150 x 150 km horizontal resolution*

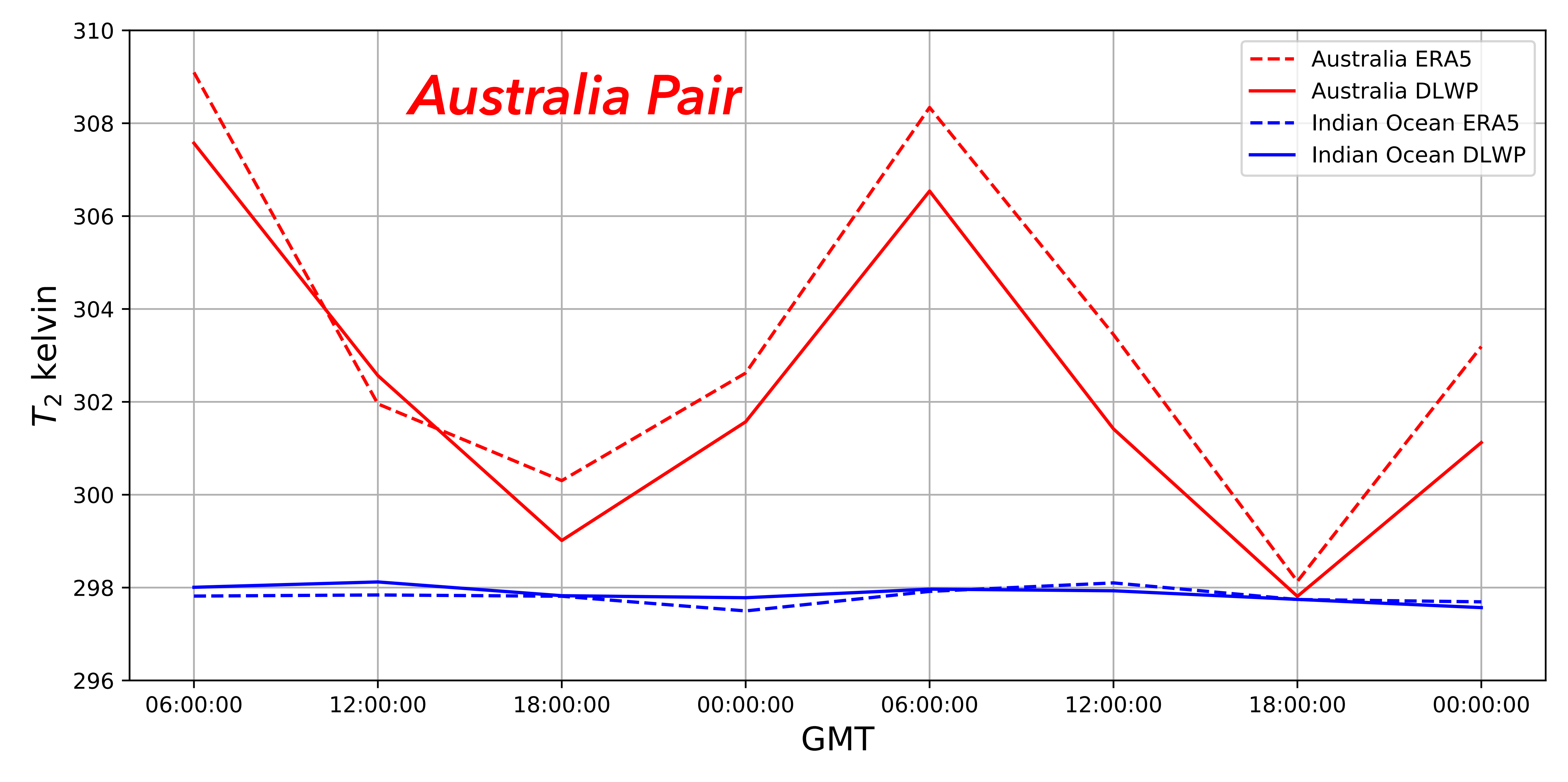
COMPARE DAILY TEMPERATURE CYCLES

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



2-m Temperature Forecasts



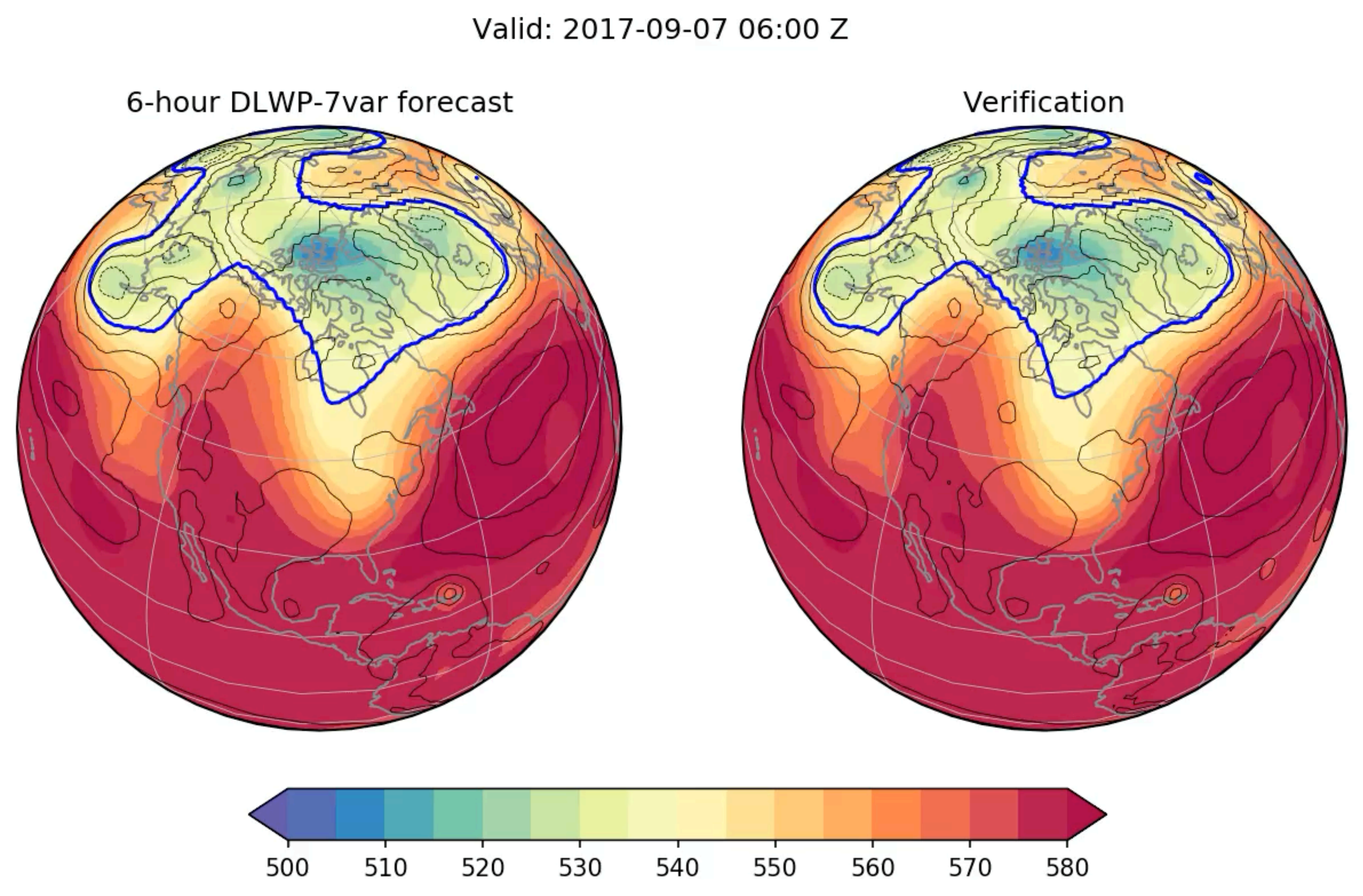


- Little temperature variation over oceans (land-sea mask)
  - Without sea-surface temperature input!
- Larger diurnal variations over Australia than the Amazon
  - Total column water vapor?

No geo-specific data!

Hurricanes Irma & Jose

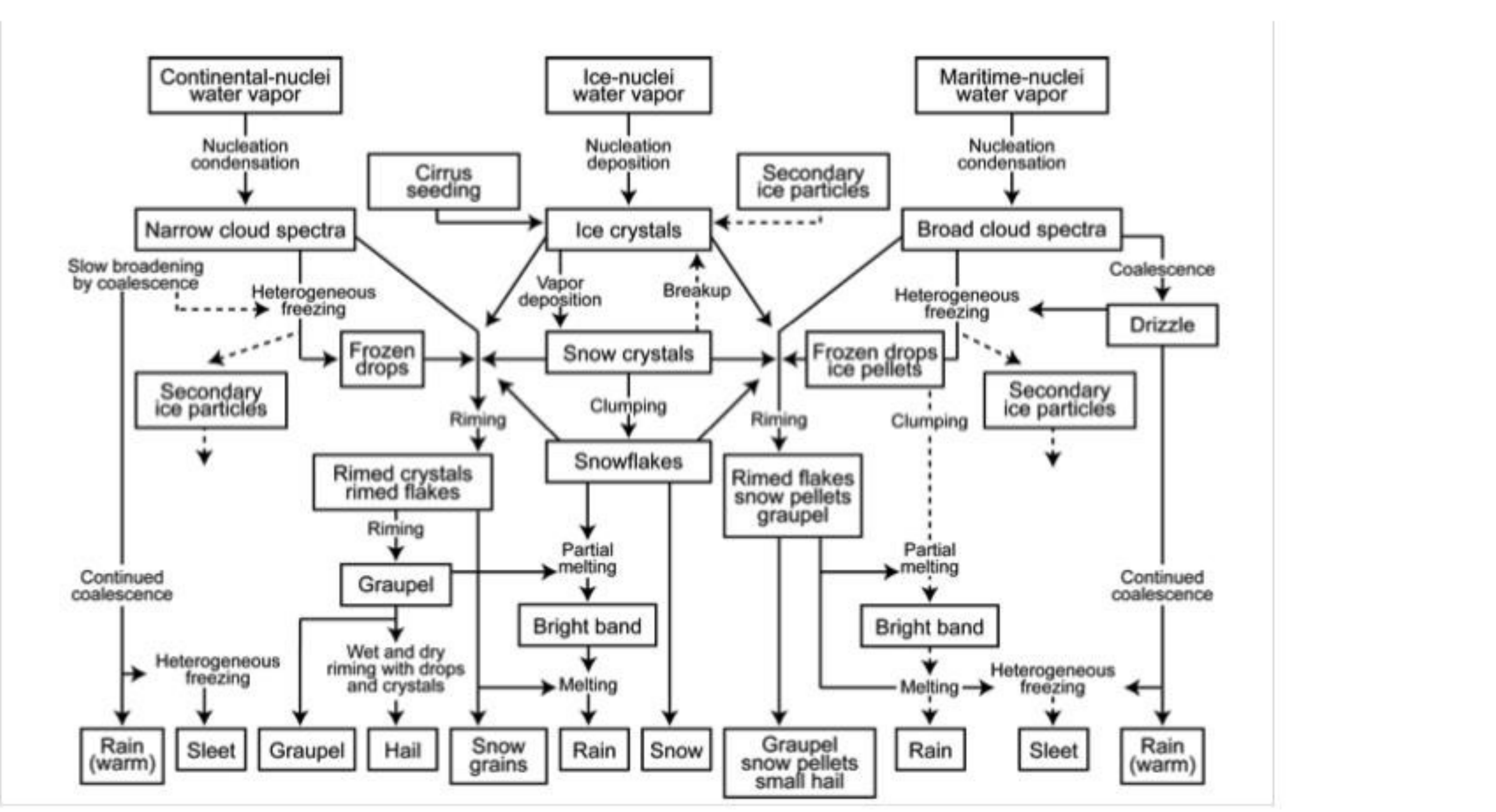
- 4-day single model forecast
- 150 x 150 km resolution
- 7 prognostic variables
- Showing
  - 1000-hPa height (black)
  - 500-hPa height (color fill)



DIAGNOSING PRECIPITATION WITHOUT A CLOUD-MICROPHYSICS PARAMETERIZATION

- *6 spherical shells of data*
- *150 x 150 km horizontal resolution*

Cloud and Precipitation Processes: Underlying Physics

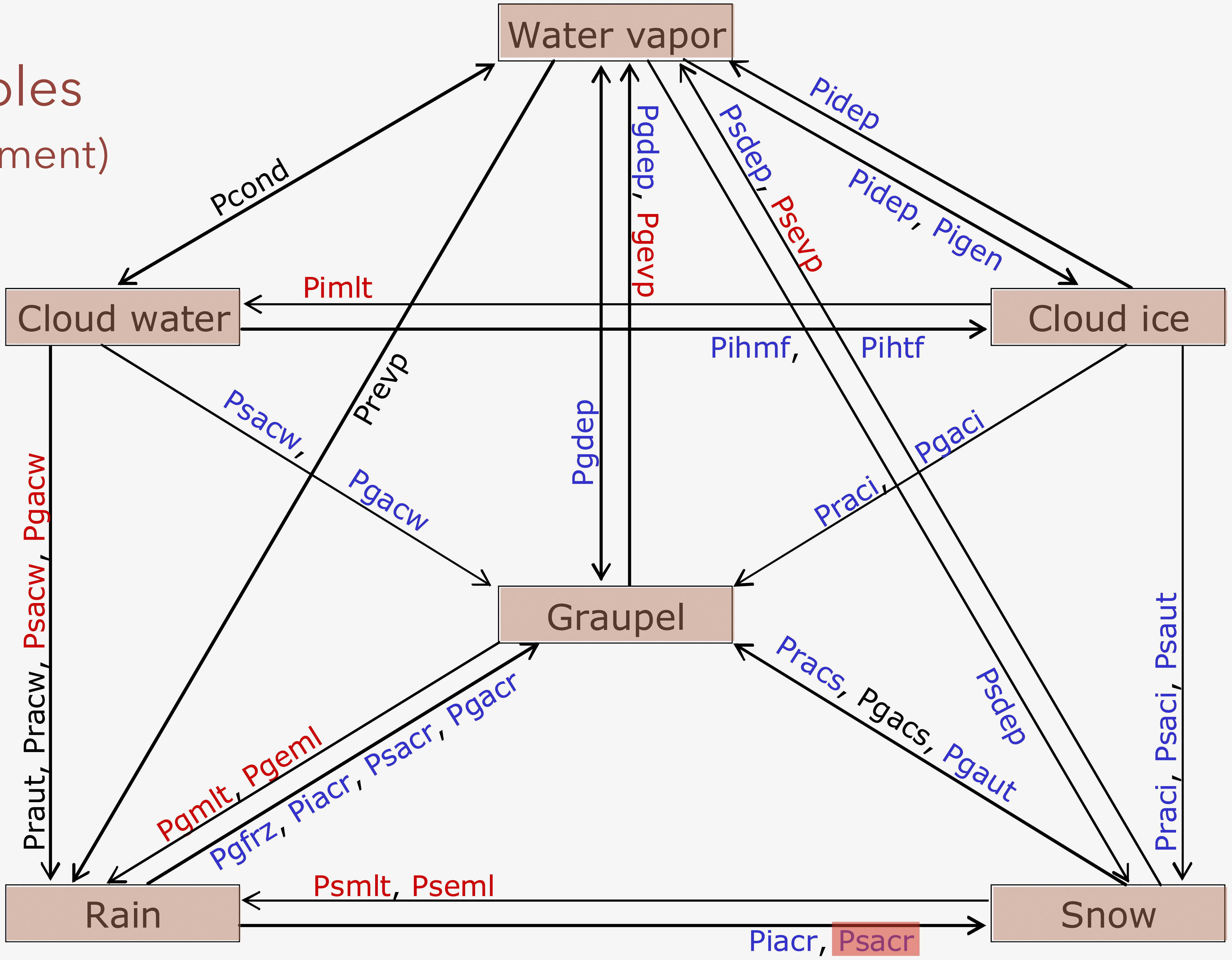


from Dudhia, Overview of WRF Physics

WSM6 Parameterization of Cloud and Precipitation Processes in NWP

6 variables

(single moment)



Snow accreting rain

$$\mathrm{Psacr}[\mathrm{kg\,kg^{-1}s^{-1}}] = \pi^2 E_{SR} n_{0R} n_{0S} \left(\frac{\rho_W}{\rho}\right) \times \left|V_S - V_R\right| \left(\frac{5}{\lambda_R^6 \lambda_S} + \frac{2}{\lambda_R^5 \lambda_S^2} + \frac{0.5}{\lambda_R^4 \lambda_S^3}\right),$$

Snow size distribution

Intercept parameter

$$n_{0S}\ (\mathrm{m}^{-4}) = 2 \times 10^6 \exp[0.12(T_0 - T)]$$

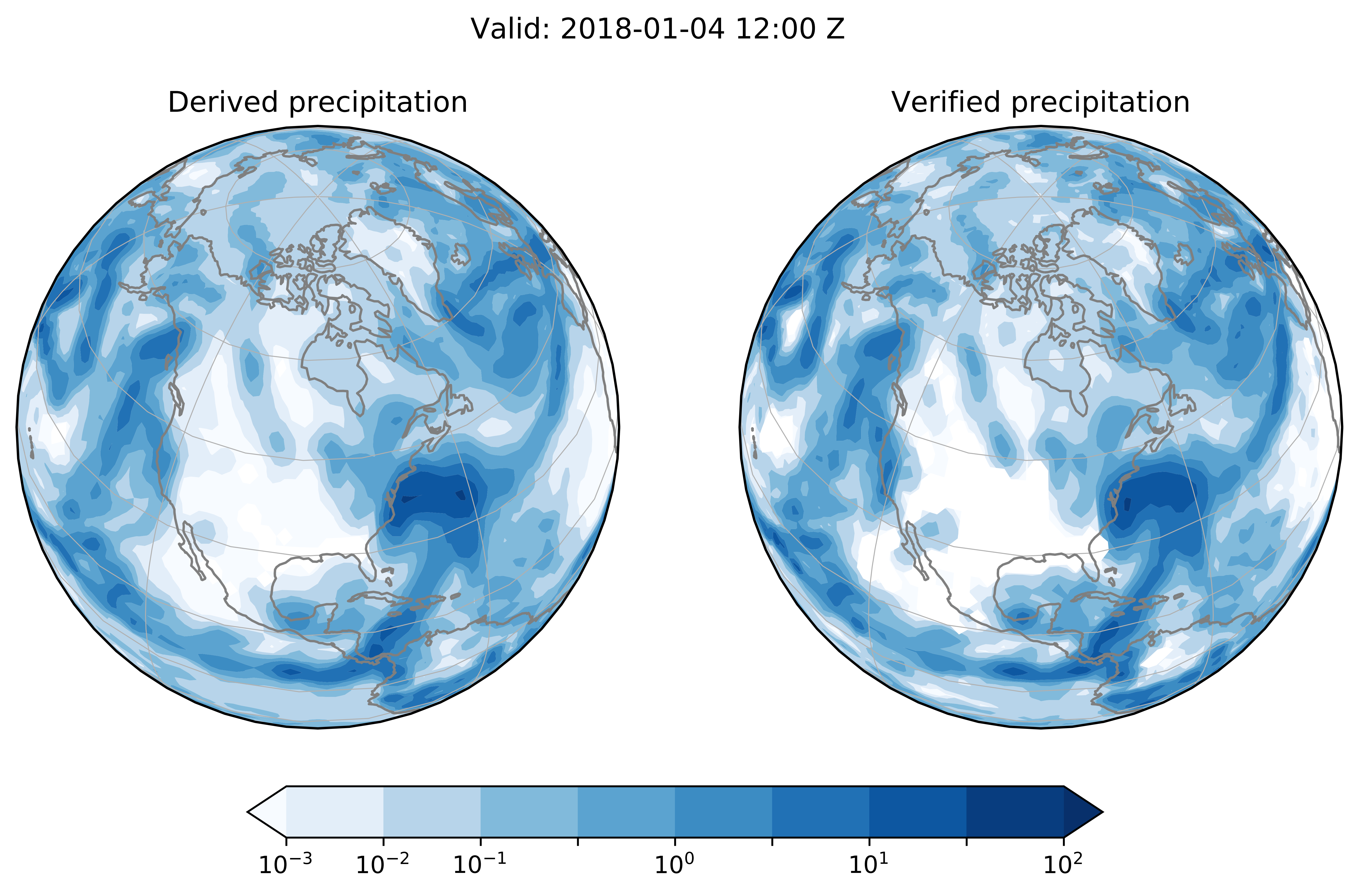
Slope parameter

$$\lambda_S = (\pi \rho_S n_{0S} / \rho q_S)^{1/4}$$

Hong and Lim, 2006

Deep Learning Diagnostic Precipitation Test – 1

- Precipitation is diagnosed from the ERA5 analysis of
  - 1000-hPa height
  - 500-hPa height
  - 300-700-hPa thickness
  - 2-m temperature
  - 850-hPa temperature
  - Total column water vapor
- In the spirit of Larraondo et al., 2019
  - U-net: 1000, 800, 400 hPa height



DEEP LEARNING DIAGNOSTIC PRECIPITATION TEST – 2

Same

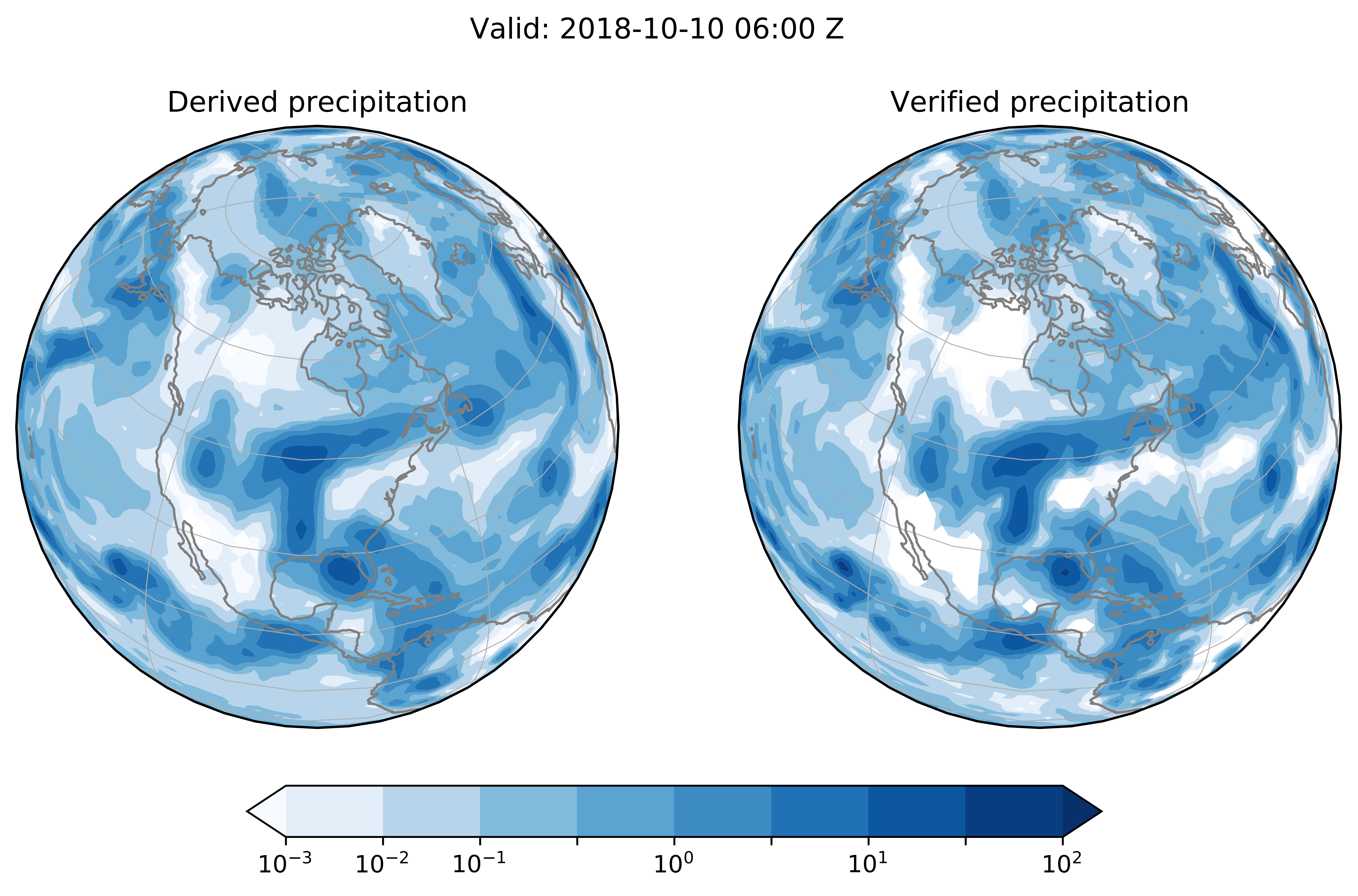
- U-net architecture
- 6 2D variables

Training loss is a function of

Rasp & Thuerey (2021)

$$\ln(p + \epsilon) - \ln(\epsilon)$$

When do we need microphysics?



Ensemble Performance

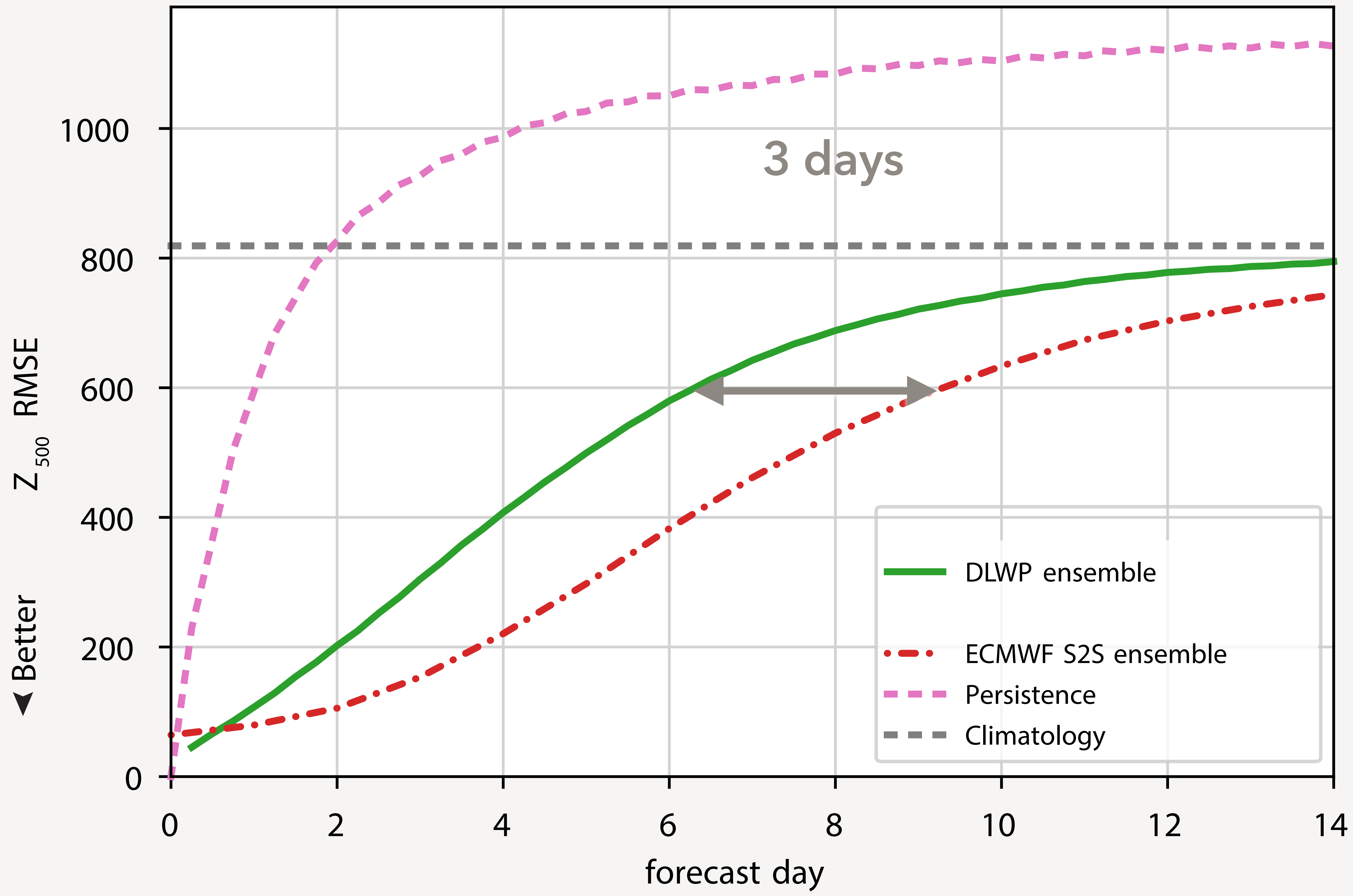
- DLWP vs European Centre for Medium Range Weather Forecasts (ECMWF)
  - 208 forecasts over 2-year period

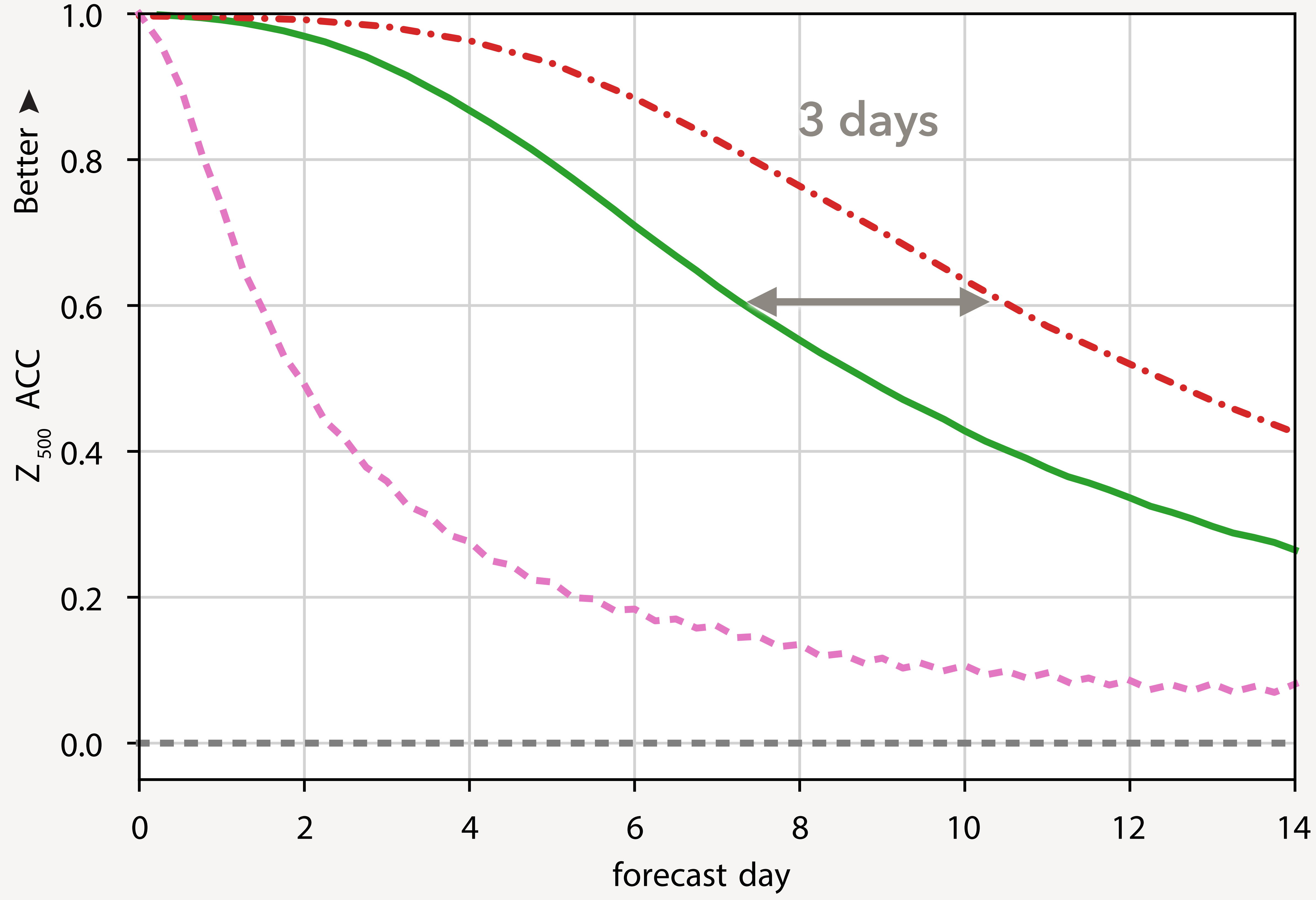
Comparing key model attributes

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 spherical shells	819 spherical shells
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

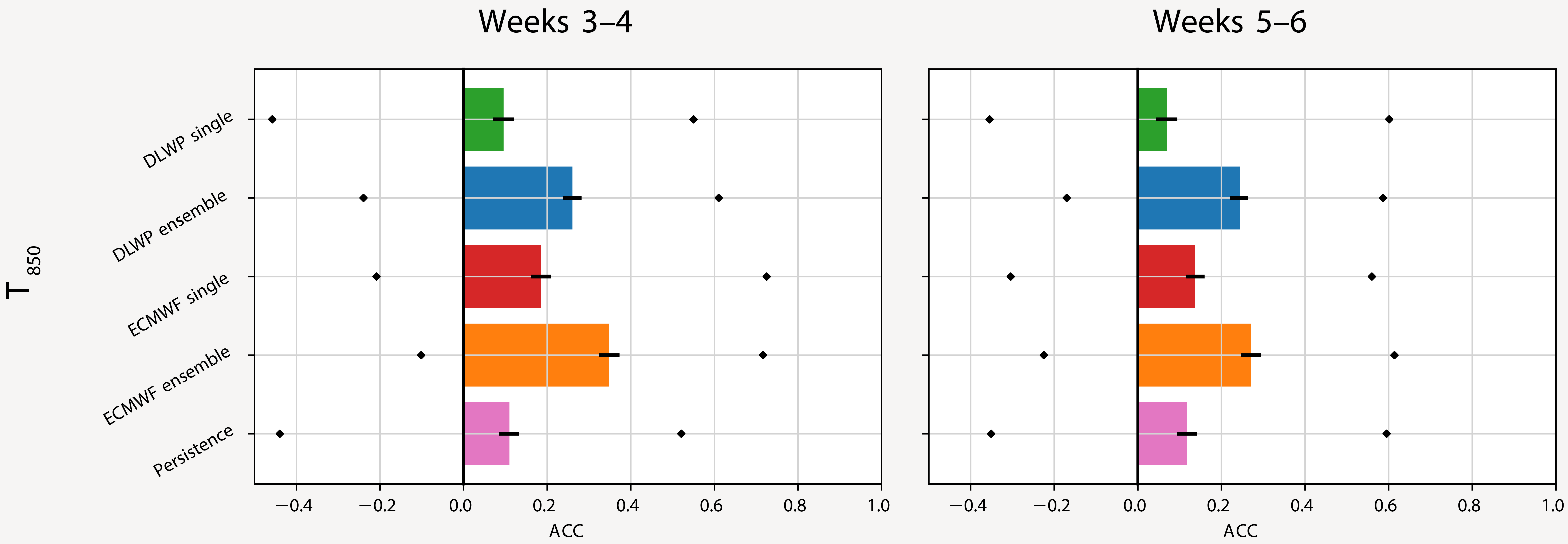




DLWP ensemble: 32 stochastically perturbed models x 10 initial conditions = 320 members

ENSEMBLE PERFORMANCE: S2S LEAD TIMES

Anomaly correlation coefficient of the ensemble mean

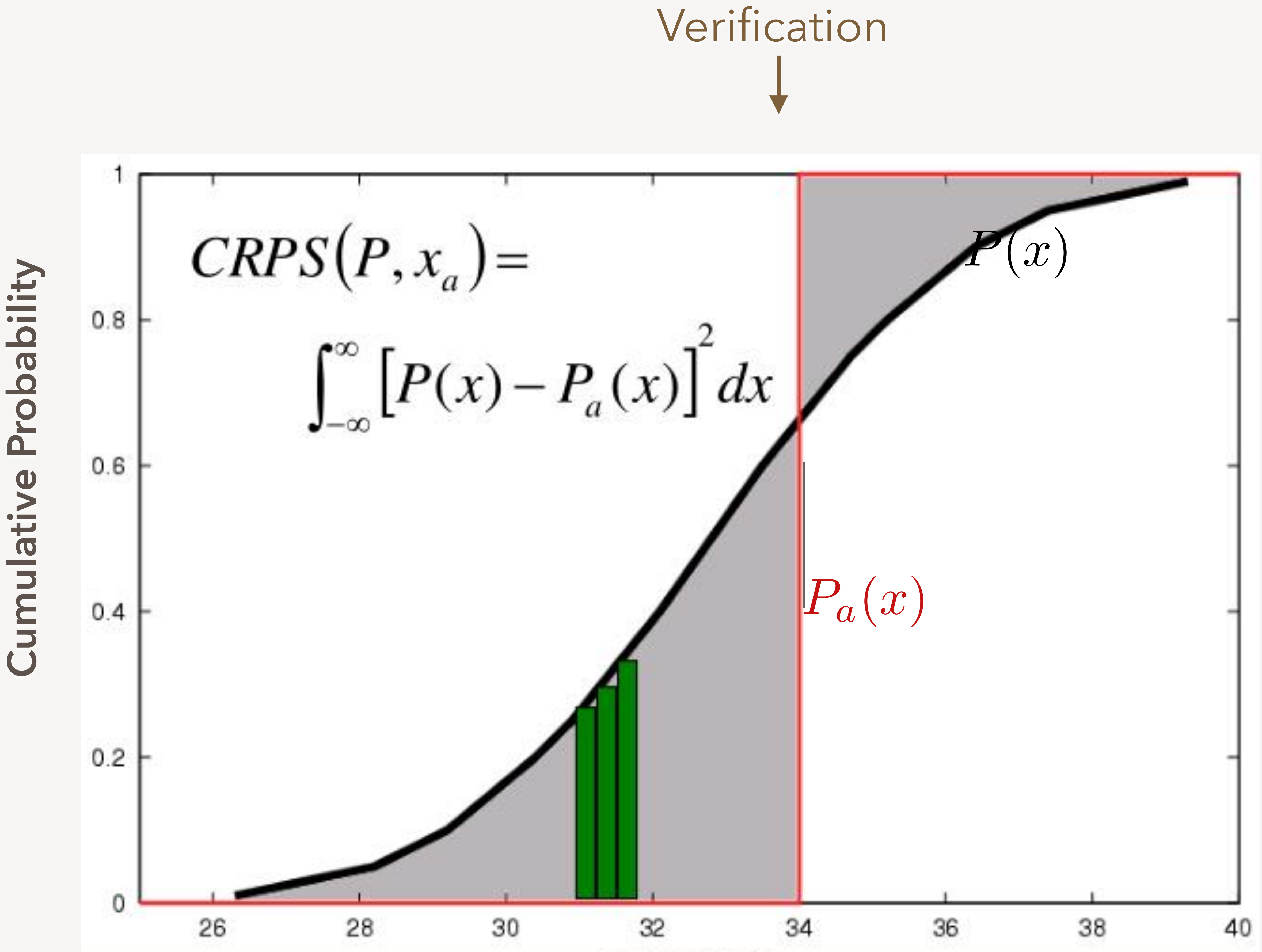


Persistence is computed as the 2-week-averaged anomaly just prior to the initialization

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

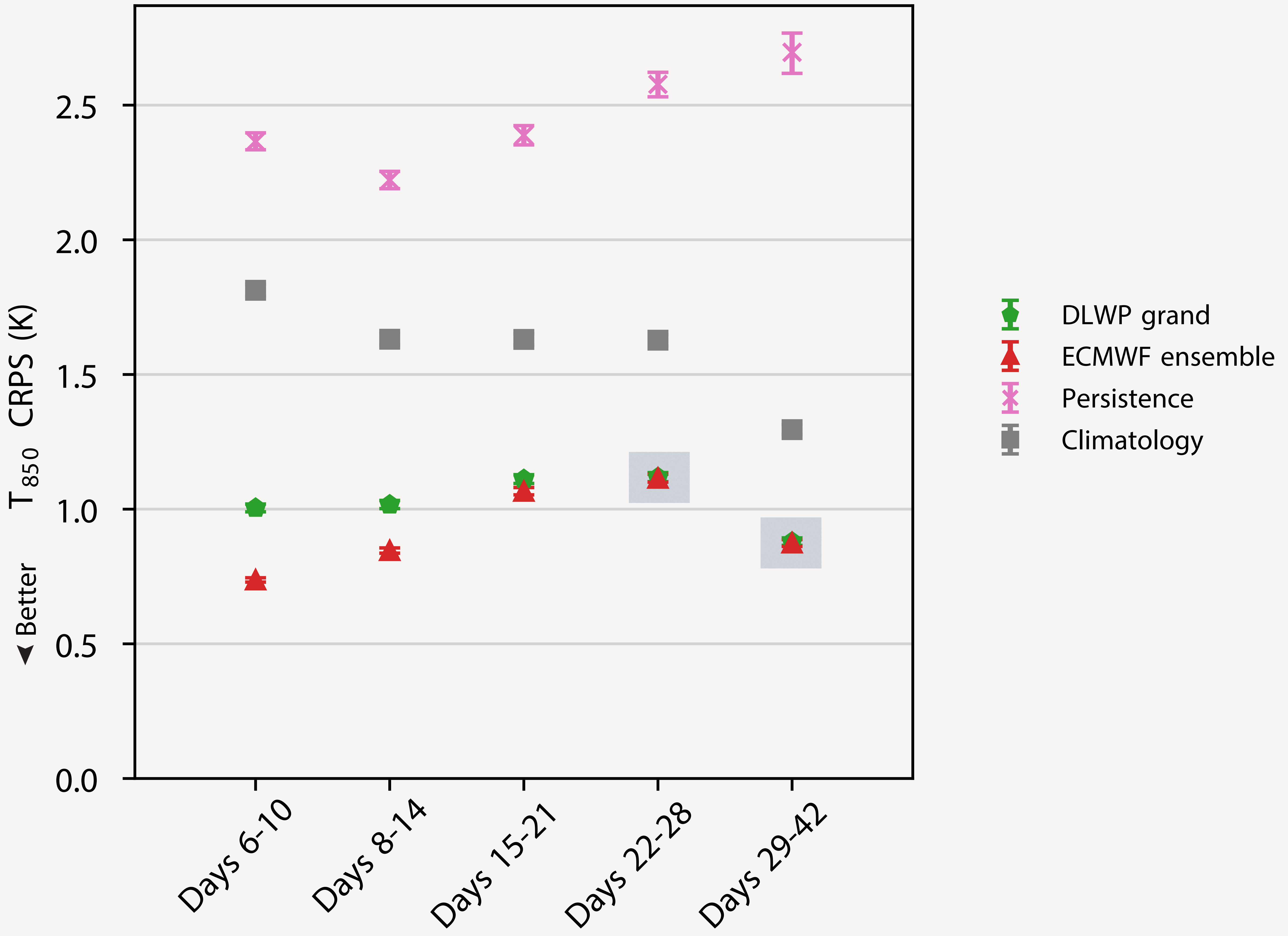
Probablistic Scores of Ensemble Forecasts

- *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*



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.
  - *DLWP may be particularly appropriate for sub-seasonal forecasting.*
- *DLWP can learn dynamics and physical parameterizations at the same time*.
  - Integrated approach to Earth-system modeling for sub-seasonal & seasonal forecasts
- The *speed of DLWP allows use of much larger forecast ensembles* O(1000).
  - Large well-calibrated ensemble would
    - Better define the probable distribution of future atmospheric states
    - Better capture extreme events.
  - 320-member ensemble of 1-month forecasts stepped forward with 12-hr time step (and 6-hr resolution) requires ~2 seconds 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