Deep Learning Weather Prediction: Epistemology and New Scientific Horizons

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

Machine Learning & Science

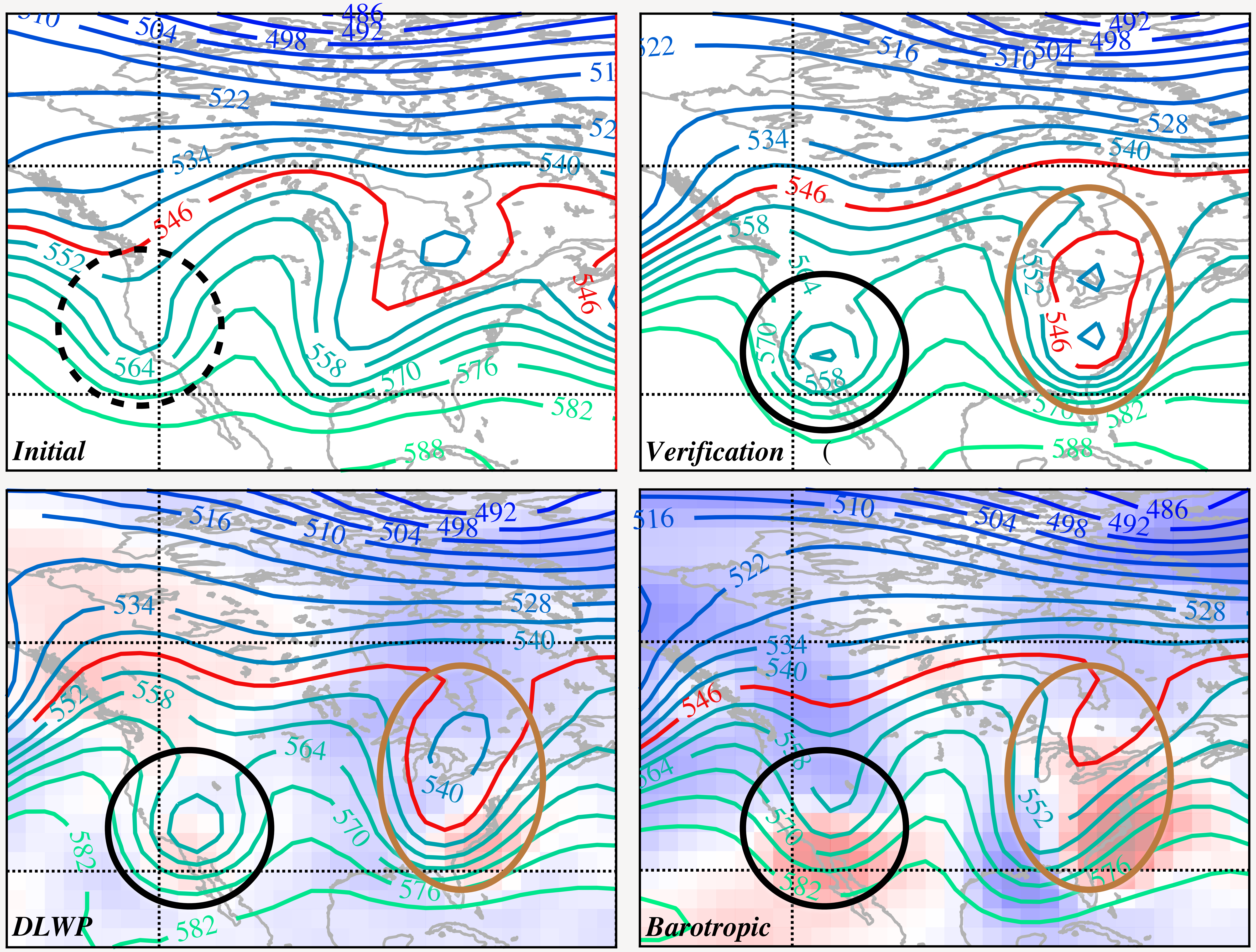
xkcd

- Direct discovery
- Controlled experiments

Direct Discovery

- 24-hr predictions of 500 hPa height using *only* 500 hPa height
- Machine learning beats the barotropic model
  - Color fill shows error

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

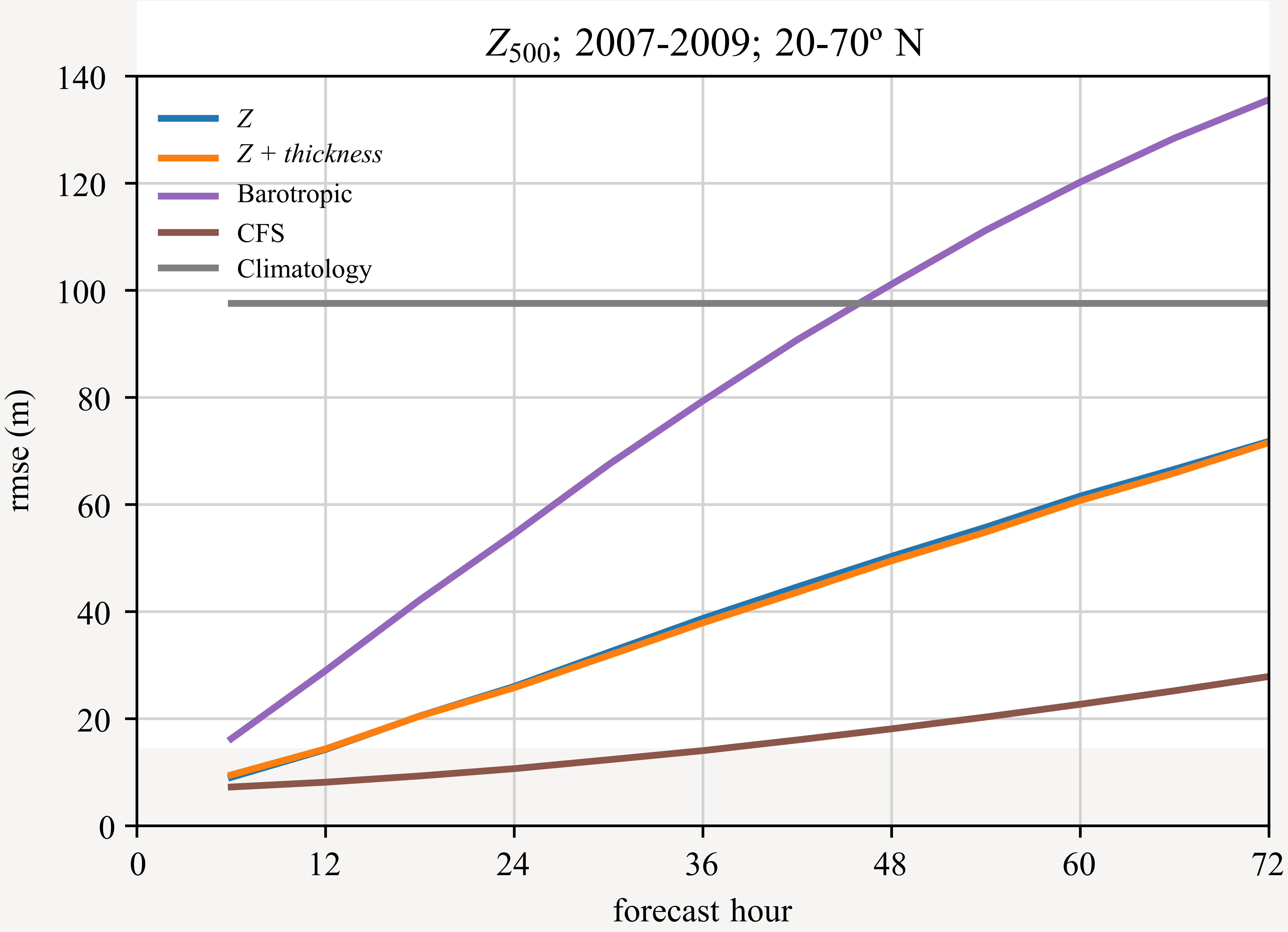


Weyn, et al., 2019: Can machines learn to predict weather….

Adding Baroclinity Data (Thickness)

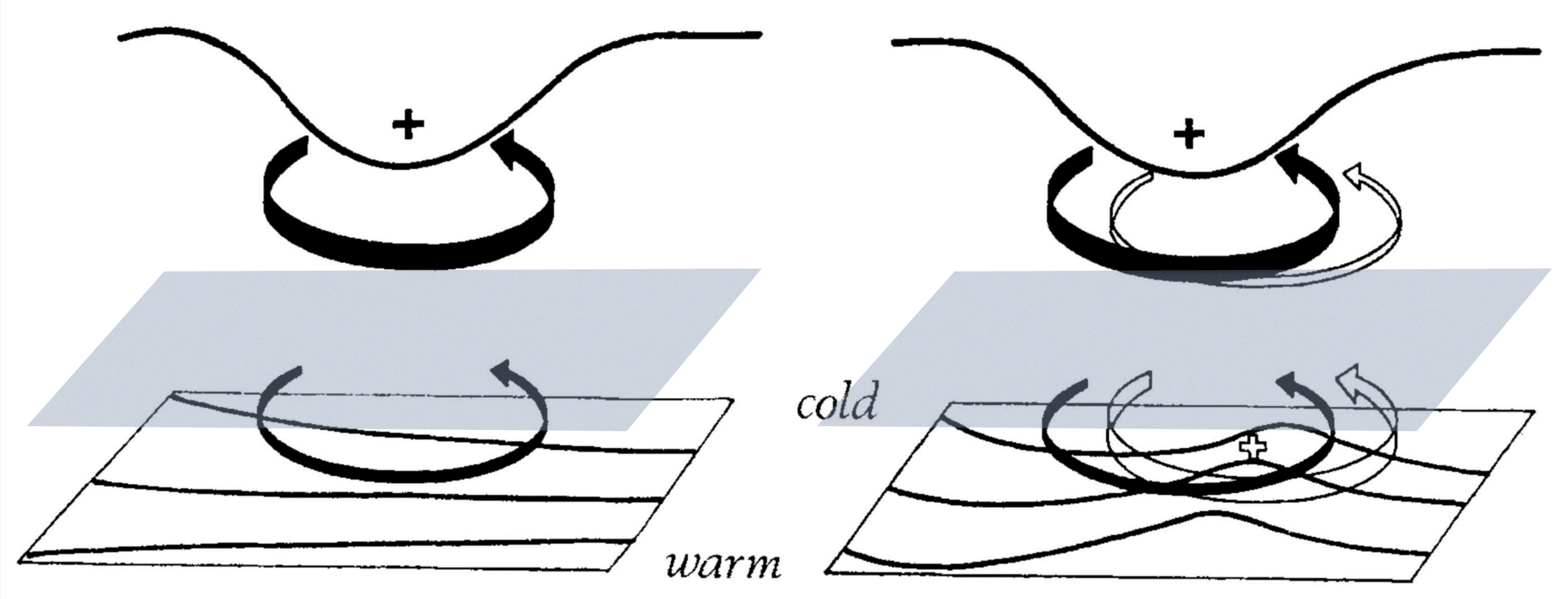
- Z_{500} and 700-300 hPa thickness
- Test set is 2007-2009
- RMSE over 20–70° N
- 2.5 x 2.5° resolution

700-300-hPa thickness adds little skill



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

Intensification Determined by Single-Level Data ???

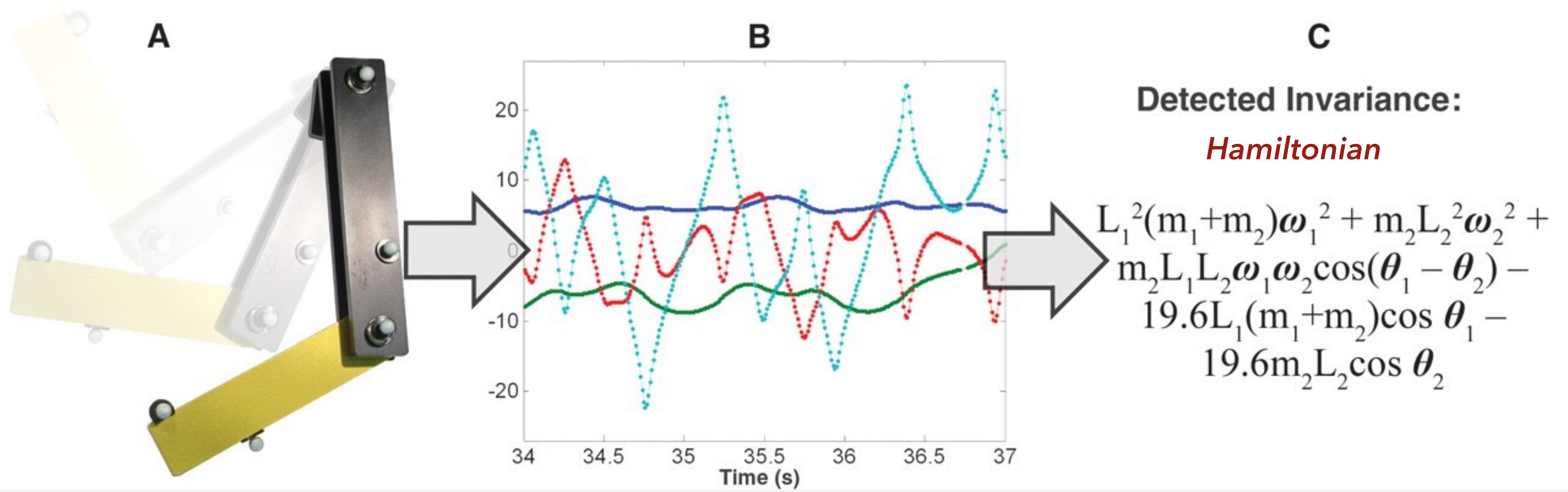


Hoskins et al., 1985: On the use and significance of isentropic potential vorticity maps

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

DO WE NEED AI TO BETTER *UNDERSTAND* MIDLATITUDE SYNOPTIC-SCALE DYNAMICS?

"Symbolic Regression"



Schmidt & Lipson, 2009: Distilling free-form natural laws from experimental data

Controlled Experiments

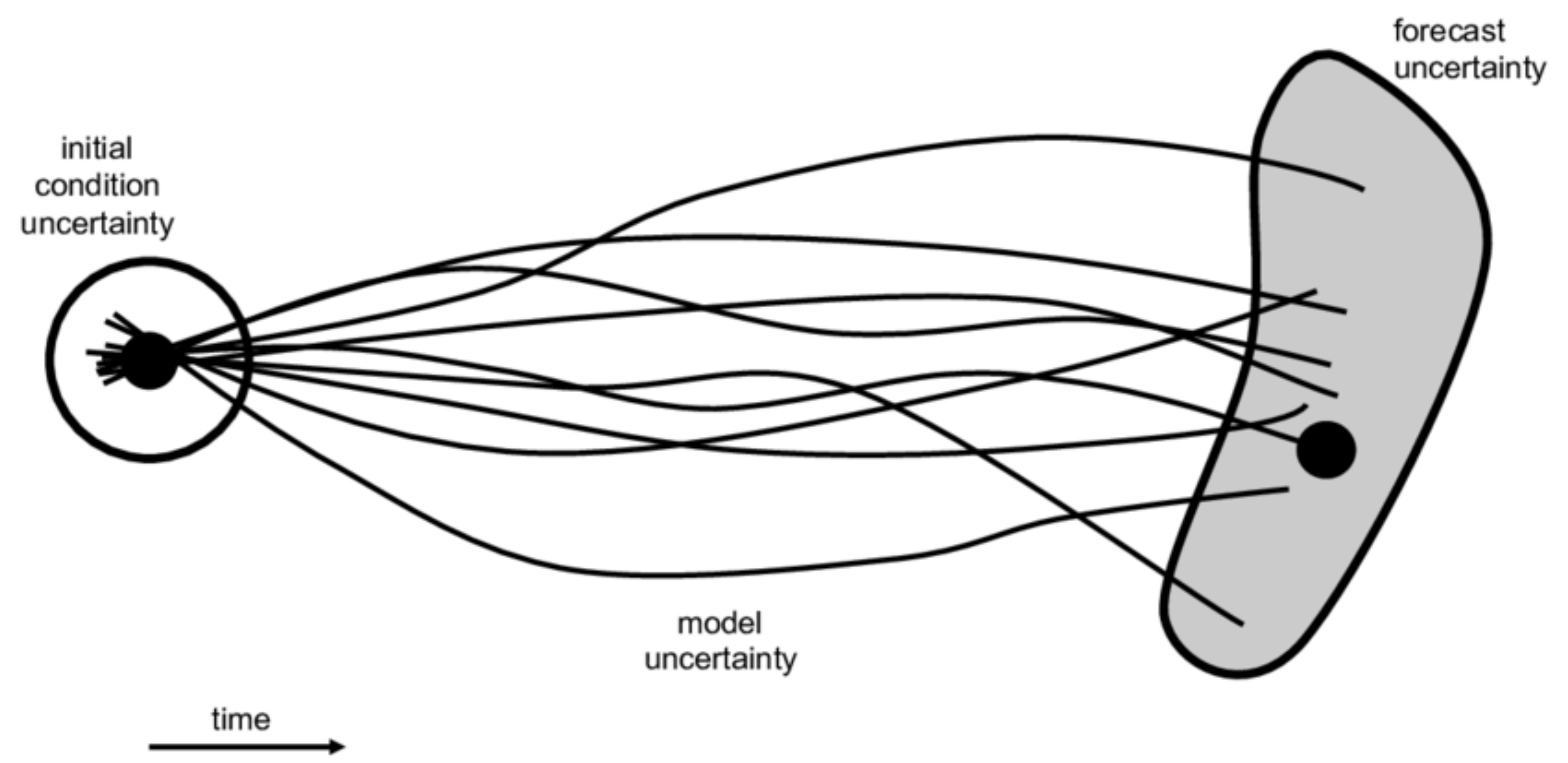
- To understand processes governing surface temperatures, winds, precipitation
  - Including a wide range of time an space scales
- Use hierarchy of models
  - NWP models to explicitly compute surface temperature, precipitation, winds
  - Idealized models with “simple” equations
    - Insight is not automatic: Coriolis force acting on E-W motions
  - Express understanding in words

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, …
- Global NWP models rely on parameterizations
  - Clouds and precipitation
  - Influence of the earth's surface
- *Parameterizations are evaluated empirically!*

ATMOSPHERE IS CHAOTIC

- Use ensembles
  - For prediction and attribution
- IC ensembles are under-dispersive
  - *Empirically* increase the spread
  - SKEB, SPPT



Empirically validated DLWP is a reasonable alternative to NWP in controlled experiments with a hierarchy of models.

Advantages of Deep-Learning Weather Prediction

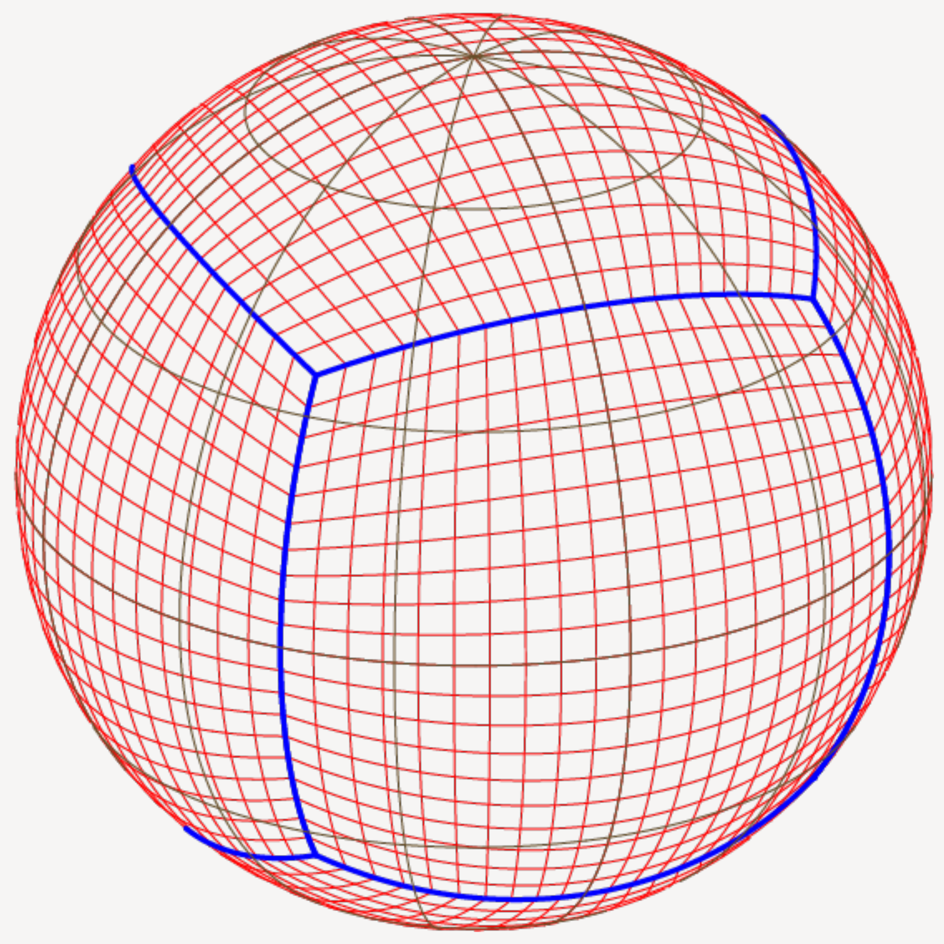
- Reduces the time required for each simulation by orders of magnitude
  - Allowing a *large* number O(1000) of ensemble members
    - Potentially better defining the probable distribution of future atmospheric states
    - Better estimating the chances of extreme events
- *Efficient computation of gradients with respect to the initial state*
  - Back propagation is a very efficient implementation of the chain rule
  - Gradients of what?
- Can replace empirical NWP-style parameterizations with `holistic' machine learning.
  - Potentially crucial for the sub-seasonal and seasonal forecasting and analysis

Gradients of What?

- Any differentiable function of the NN output (not just the loss function for training)
  - Typical cost functions for the adjoint in data assimilation
  - Specialized cost functions for analysis
    - Projection onto leading modes of variability
- Generate Jacobian and singular vectors for IC ensembles
- "Gradients aren't all you need" (Metz et al., 2022)
  - Can't avoid limits from predictability horizons
  - May be noisy (use ensemble mean?)
  - Linearize to create Tangent Linear Model and associated adjoint
    - Differentiable non-orographic GWD (Hatfield et al., 2021: Building tangent-linear and adjoint models ...)

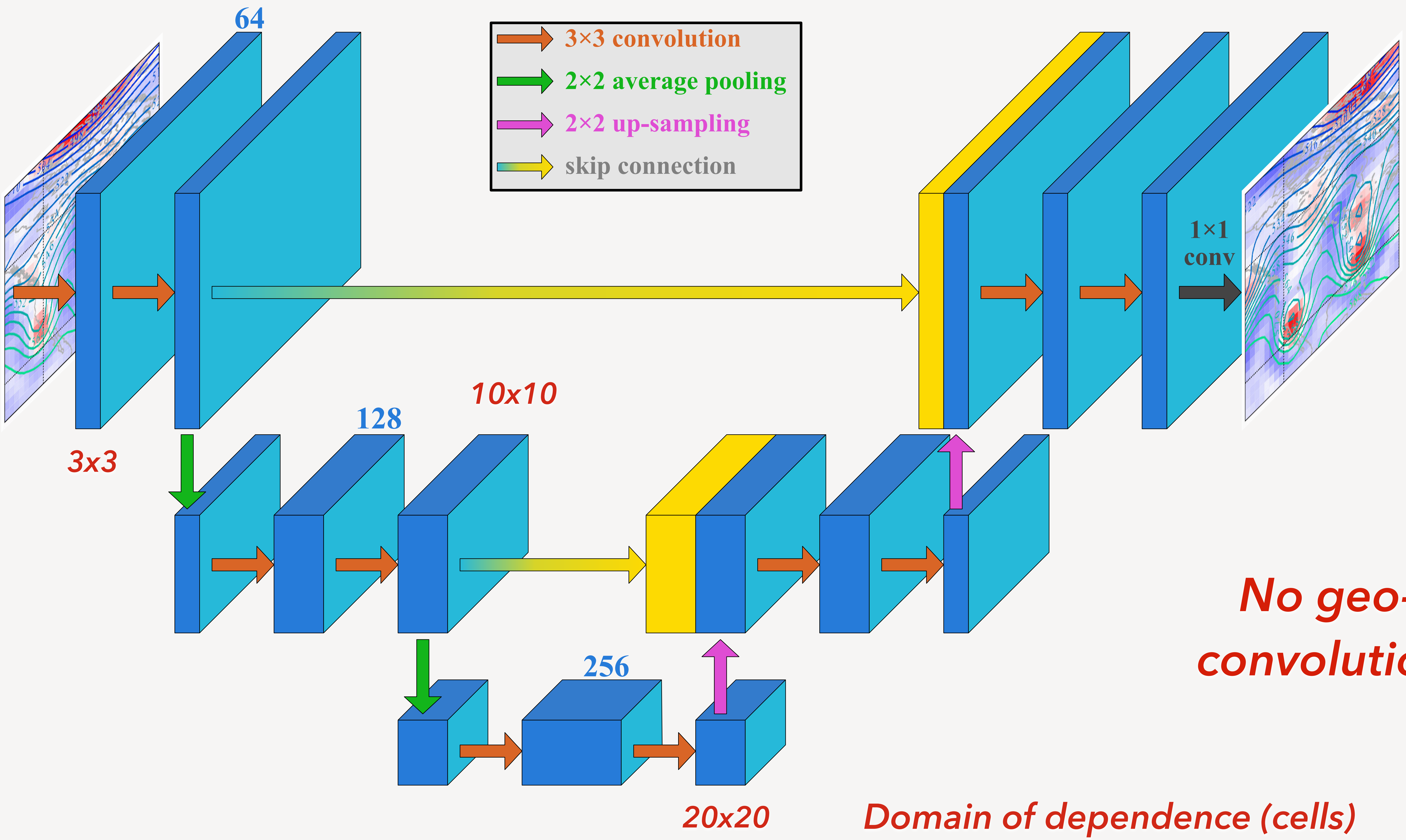
HOLISTIC PHYSICS PARAMETERIZATION: SURFACE TEMPERATURE

- 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)
  - 150 x 150 km quasi-uniform

U-Net Architecture

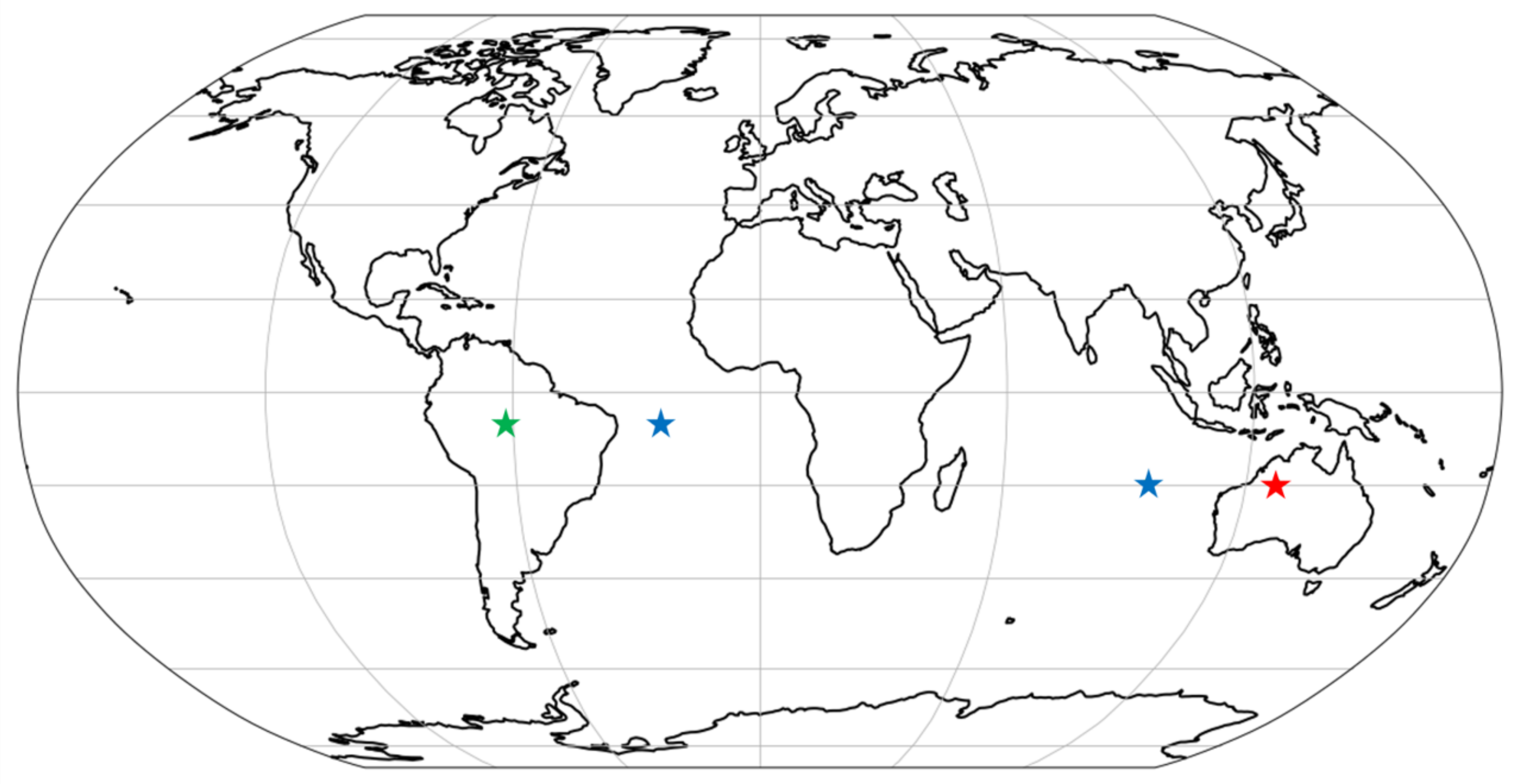


No geo-specific convolutional filters

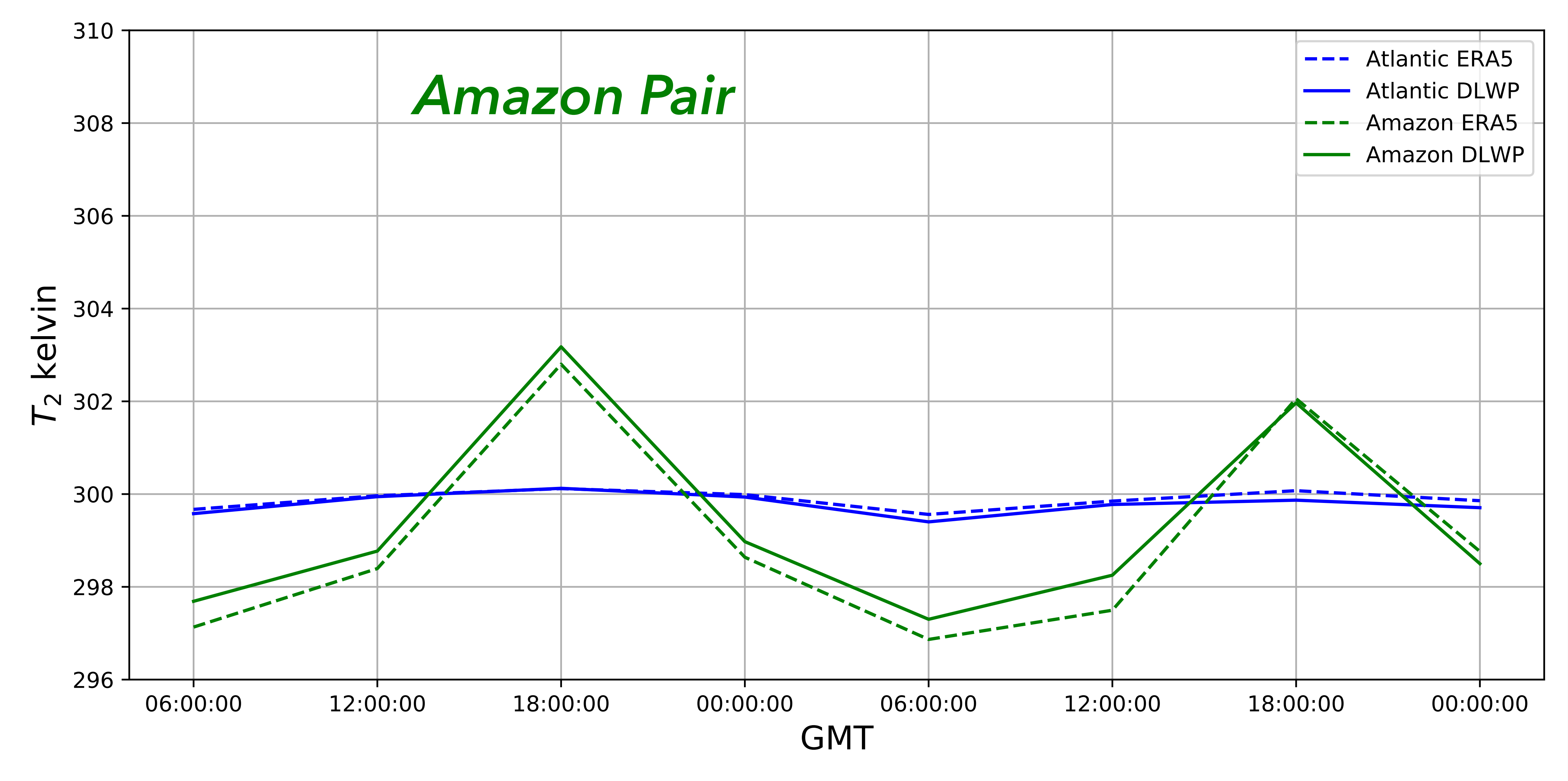
Domain of dependence (cells)

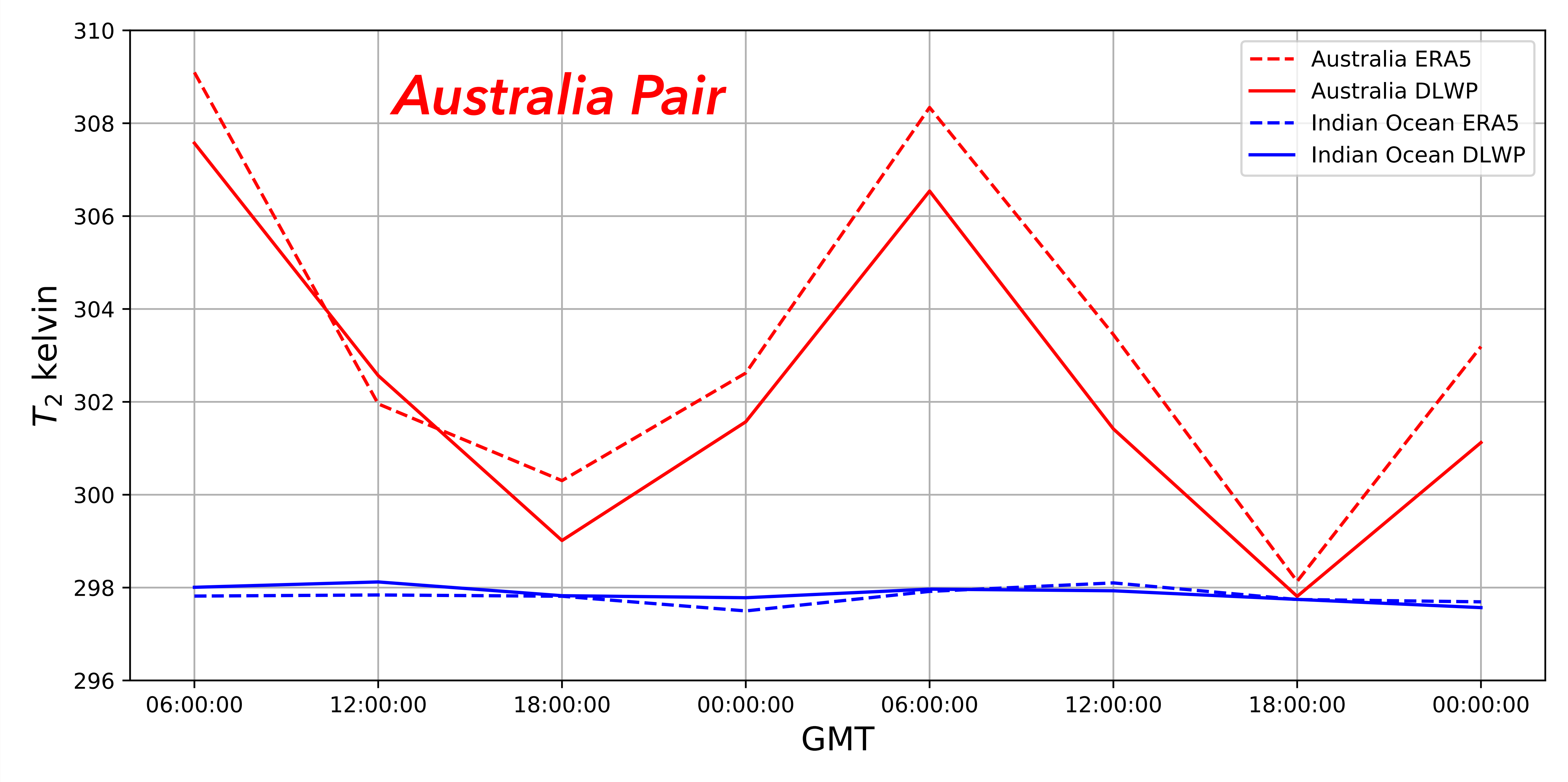
COMPARE DAILY TEMPERATURE CYCLES

- 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
- Larger diurnal variations over Australia than the Amazon
  - Total-column water vapor?

No geo-specific filter coefficients!

Conclusions

- *DLWP has the potential to improve our scientific understanding*
  - Direct discovery
    - Even mid-latitude synoptic scale dynamics
  - As part of a hierarchy of models in controlled experiments
- *Ease of computing gradients for novel cost functions opens many possibilities for analysis.*
  - Research required to determine suitable approaches for gradient computations.
- *DLWP can learn dynamics and physical parameterizations at the same time*.
  - Integrated approach to Earth-system modeling for sub-seasonal & seasonal forecasts
  - May facilitate the computation of smooth gradients.

References:

Hatfield et al., 2021: Building tangent-linear and adjoint models for data assimilation with neural networks. *J. Adv. Modeling Earth Sys*, 13, doi.org/10.1029/2021MS002521

Hoskins et al., 1985: On the use and significance of isentropic potential vorticity maps. QJRMS, 111. doi/abs/10.1002/qj.49711147002

Metz et al., 2022: Gradients are not all you need. arxiv.org/abs/2111.05803

Schmidt & Lipson, 2009: Distilling free-form natural laws from experimental data. *Science*, 324. doi/full/10.1126/science.1165893

Weyn et al., 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. doi/10.1029/2019MS001705