Use of AI to perform inline bias correction of NOAA UFS medium-range forecasts

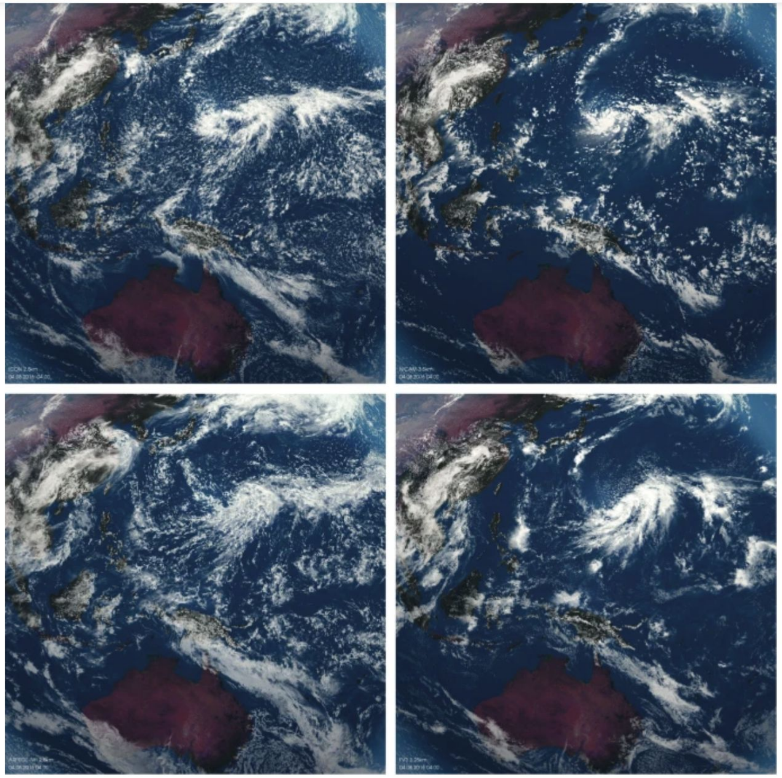
Stefan Tulich[1,2], S. Frolov[2], L. Slivinski[2], P. Pegion[2], J. Whitaker[2], T.-C. Chen[3]

1) CU/CIRES; 2) NOAA/PSL; 3)PNNL

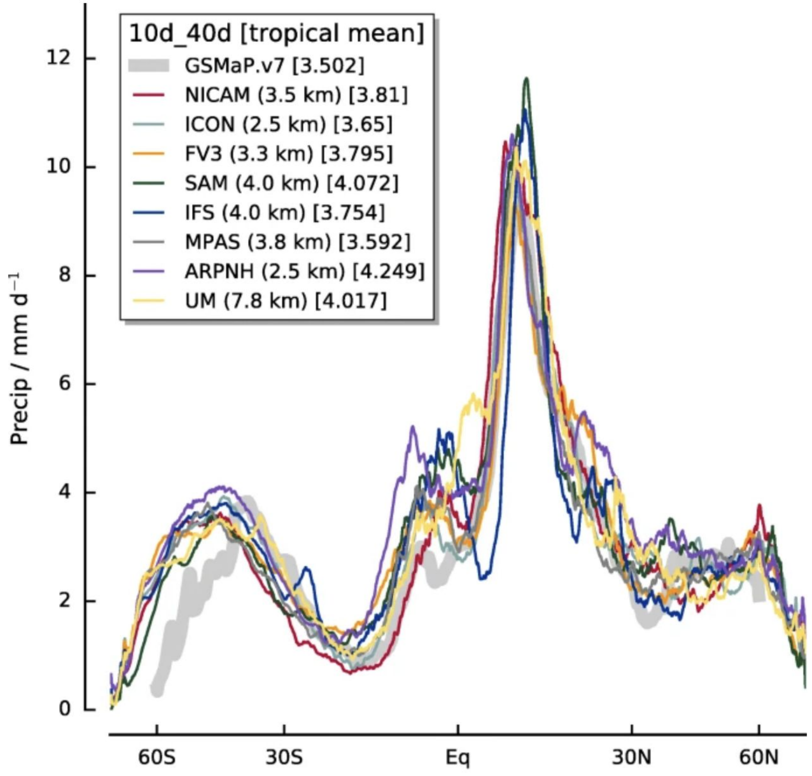
S2S Community Workshop, June 2024, Boulder, CO

Problem of systematic errors in NWP models will likely never go away

Results of DYAMOND (Stevens et al. 2019) show systematic biases even in 40-dy convection-permitting hindcasts

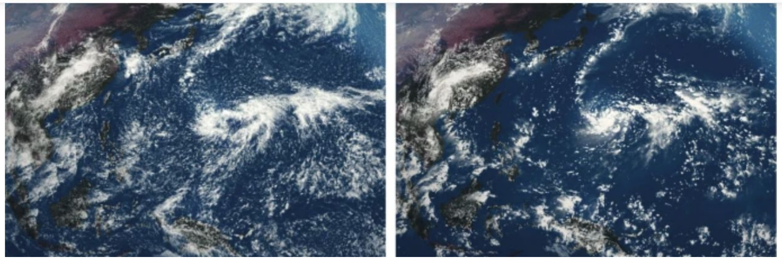


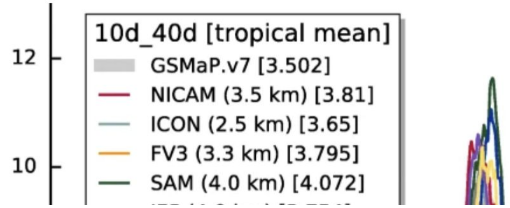
Aug. 4, 2016



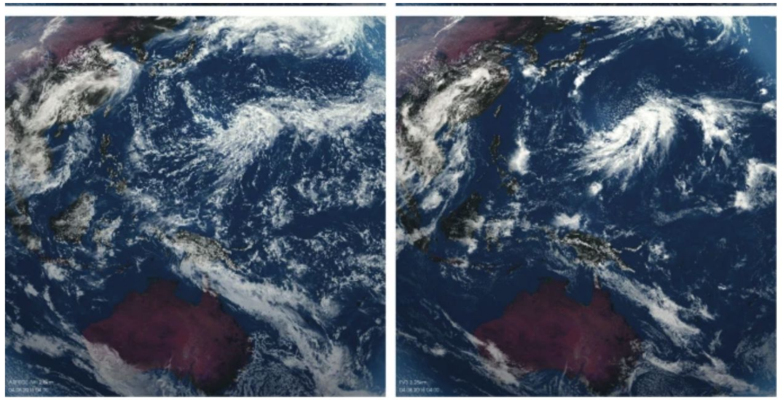
Problem of systematic errors in NWP models will likely never go away

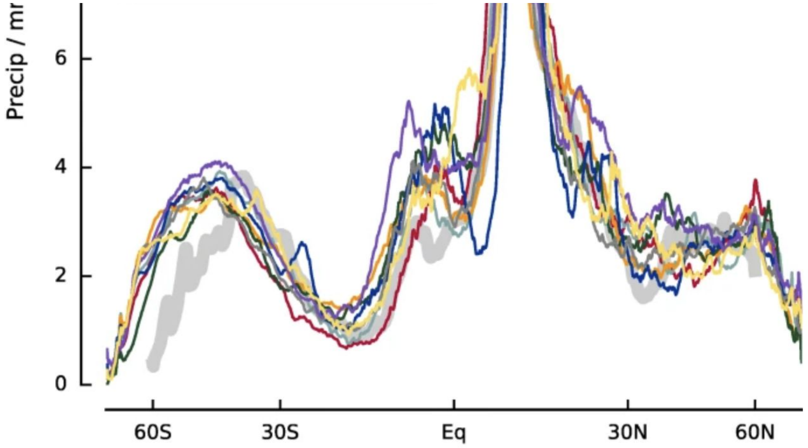
Results of DYAMOND (Stevens et al. 2019) show systematic biases even in 40-dy convection-permitting hindcasts





Increasing resolution/complexity will likely not save us!





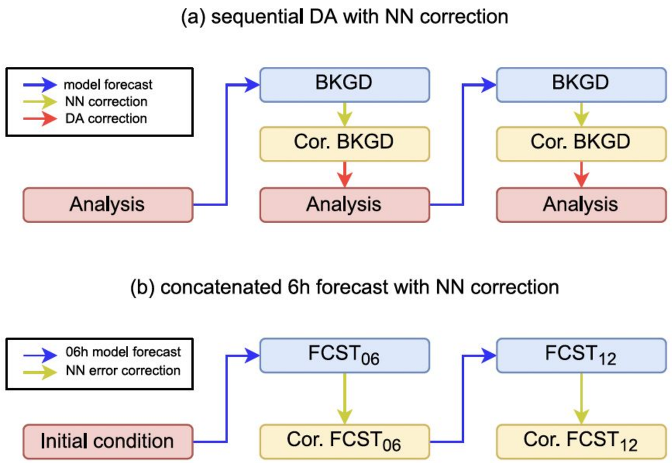
Aug. 4, 2016

Aim of this project is to build on our recent published work:

Journal of Advances in Modeling Earth Systems 10.1029/2022MS003309

Correcting Systematic and State-Dependent Errors in the NOAA FV3-GFS Using Neural Networks

Tse-Chun Chen[1,2], Stephen G. Penny[1,3], Jeffrey S. Whitaker[2], Sergey Frolov[2], Robert Pincus[4], and Stefan Tulich[1,2]

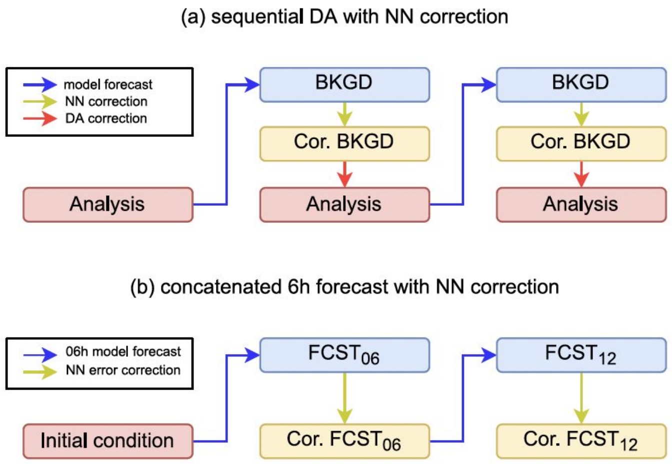


Aim of this project is to build on our recent published work:

Journal of Advances in Modeling Earth Systems 10.1029/2022MS003309

Correcting Systematic and State-Dependent Errors in the NOAA FV3-GFS Using Neural Networks

Tse-Chun Chen[1,2], Stephen G. Penny[1,3], Jeffrey S. Whitaker[2], Sergey Frolov[2], Robert Pincus[4], and Stefan Tulich[1,2]



Drawback: IBC applied external to the GFS model (file-system based)

Open questions about the sensitivity to sample size in training and choice of inputs

Outline of this talk

1) Methods and some highlights of Chen et al. 2022

2) New results looking at sensitivity to the length of the training dataset and performance of inline bias-corrected (IBC) forecasts with NN embedded directly in the GFS code repository

Architecture and training of the NN

Chen et al. explored two types of NN architectures:

1) 1D (column-based) that relies only on information local to a model grid column; due to memory constraints, subsampling of grid columns used in training

2) Convolutional NN that relies on information from a 2d-horizontal stencil of columns; horizontal smoothing used in training

Here the focus is on the column-based approach, since its easier to implement and no substantial advantage found for convolutional approach

Details about the training data and NN predictors

- Training data taken from the first year of a 15-month “replay” integration of the uncoupled FV3-GFS at C192 (roughly ½-deg. grid spacing) using a 6-hour replay window; last 3 months used for validation

- Target analysis for replay is operational IFS for the period Nov. 20, 2019 to March 1, 2020

- The NN predictors are set of analysis corrections (increments) derived from the replay calculation for: 1) horizontal wind, 2) meridional wind, 3) temperature, and 4) specific humidity (one NN for each variable)

Training inputs

- Inputs for each of the four NNs are:

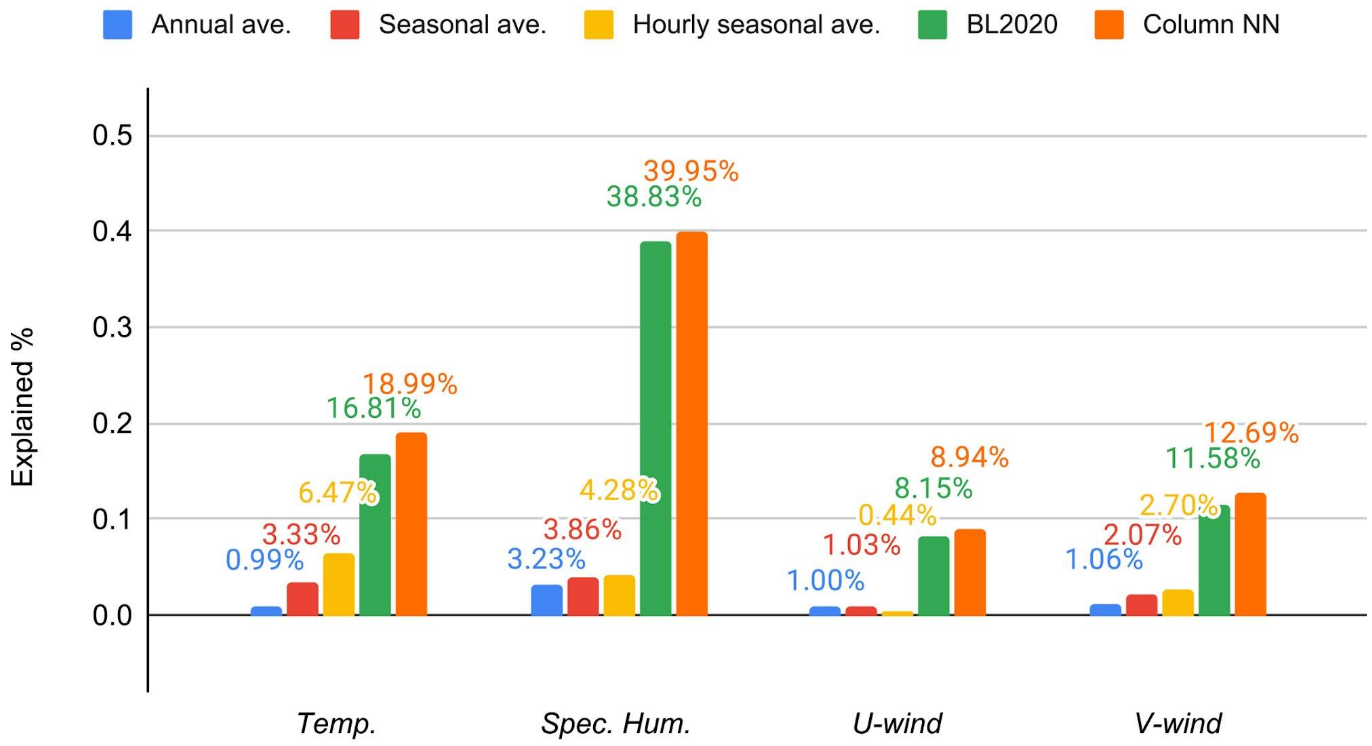
  1) Vertical profiles of u, v, T, and q (127 levels x 4 variables = 508)

  2) log(surface pressure) and 20 physical quantities pertaining to conditions at the surface (e.g. land-sea-ice mask, surface roughness, downwelling/upwelling clear-sky LW/SW radiation, albedo, etc.) and top of the atmosphere (downwelling/upwelling clear-sky LW/SW radiation)

  3) lat, sin(lon), cos(lon) sin(hour), cos(hour), sin(day of year), cos(day of year)

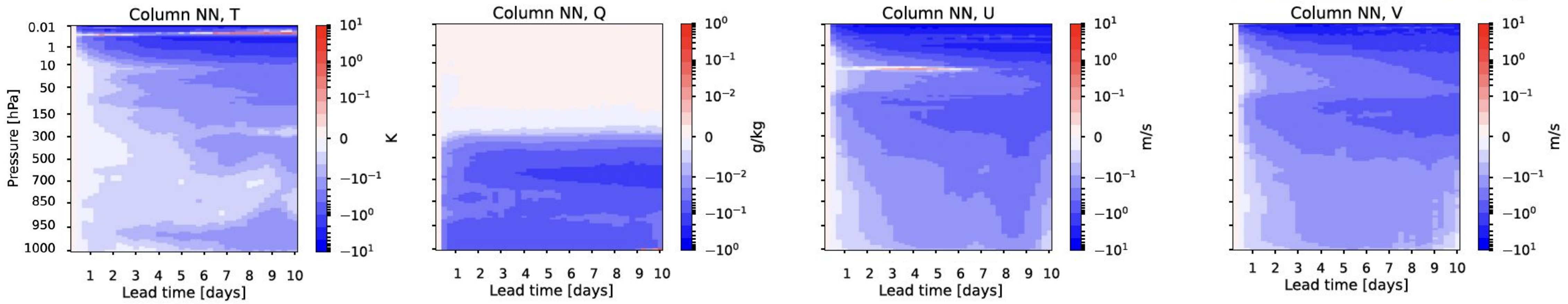
- Total size of input vector = 536 (larger than BL20, who ignored covariance and physical surface information)

Offline evaluation of NN performance



Key results: 1) Column NN (orange) greatly outperforms linear baselines of varying complexity
2) Performance slightly better than a column NN formulated as in BL20 (green)

Inline evaluation: reduced global RMSE out to 10 days!



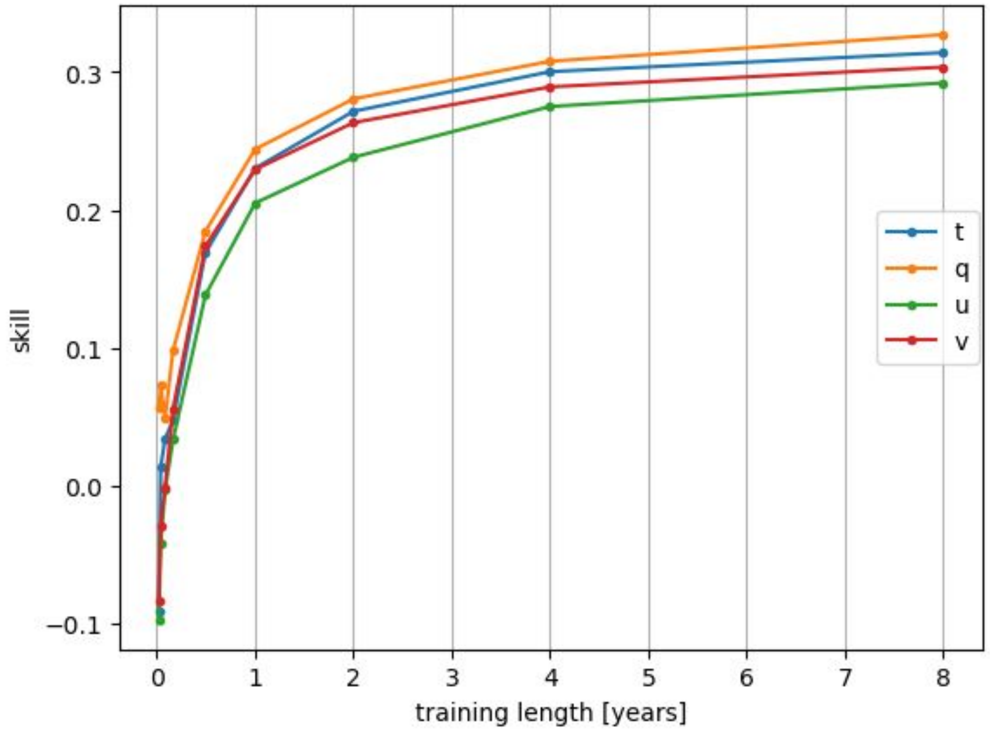
Objectives of this follow on work:

- Translate the external NN-correction software from Python to Fortran code internal to the UFS (at the same level as the replay increment software), to enable more efficient IBC forecasts

- Leverage more extensive 30-year archived coupled replay dataset; generated by driving coupled UFS “HR1” towards ERA5 in the atmosphere and ORAS5 in the ocean (2 datasets available: 1/4-deg. and 1-deg.)

- Numerous questions: How much training data is needed for NN performance to saturate? Sensitivity to training inputs? What happens after 10 days and what if in-line corrections are applied to the ocean as well?

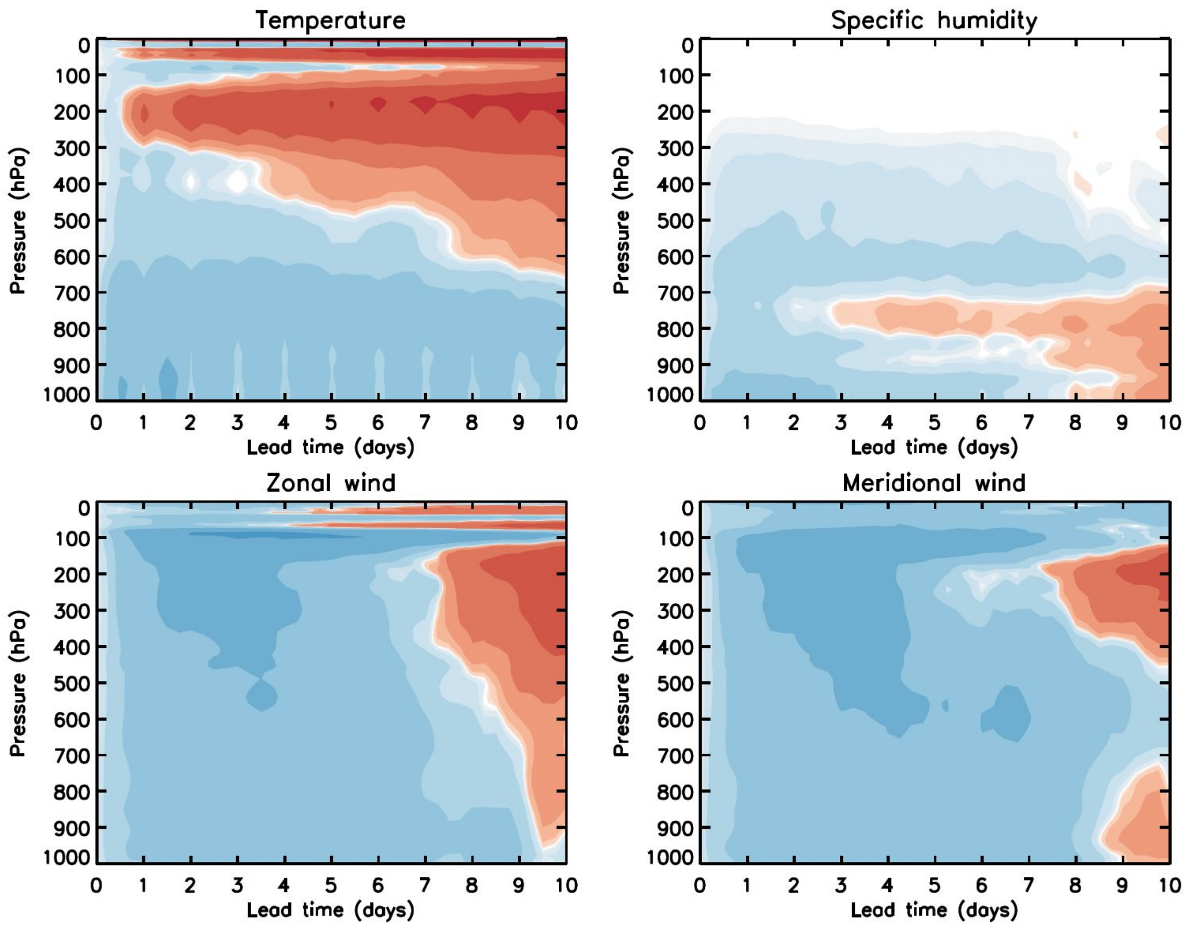
How much training data is needed for saturation?



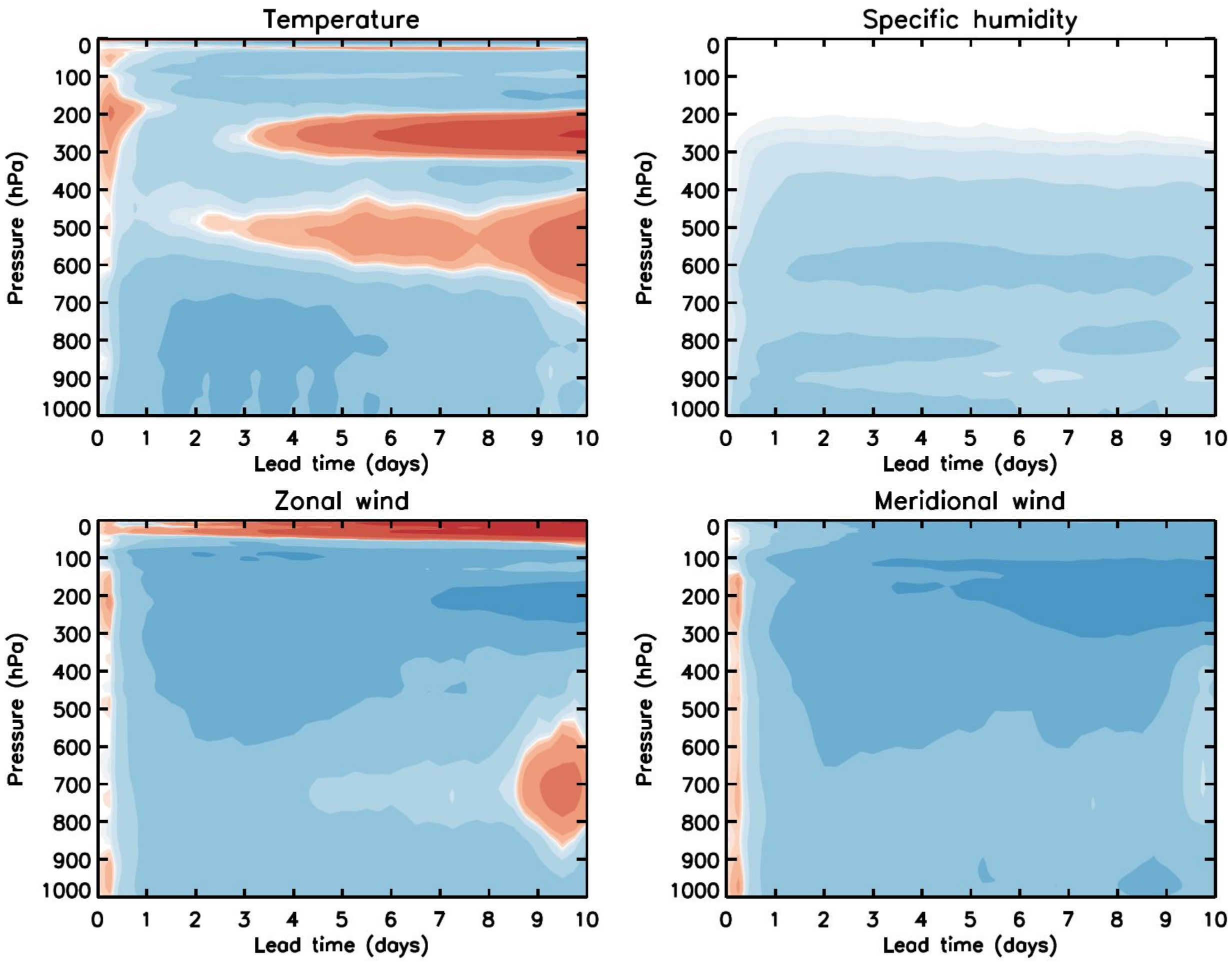
Training input vector reduced to retain only clear-sky radiation variables and information about calendar day, time of day, and latitude/longitude

Result based on sub-sampling 1-deg. data to 8-deg.

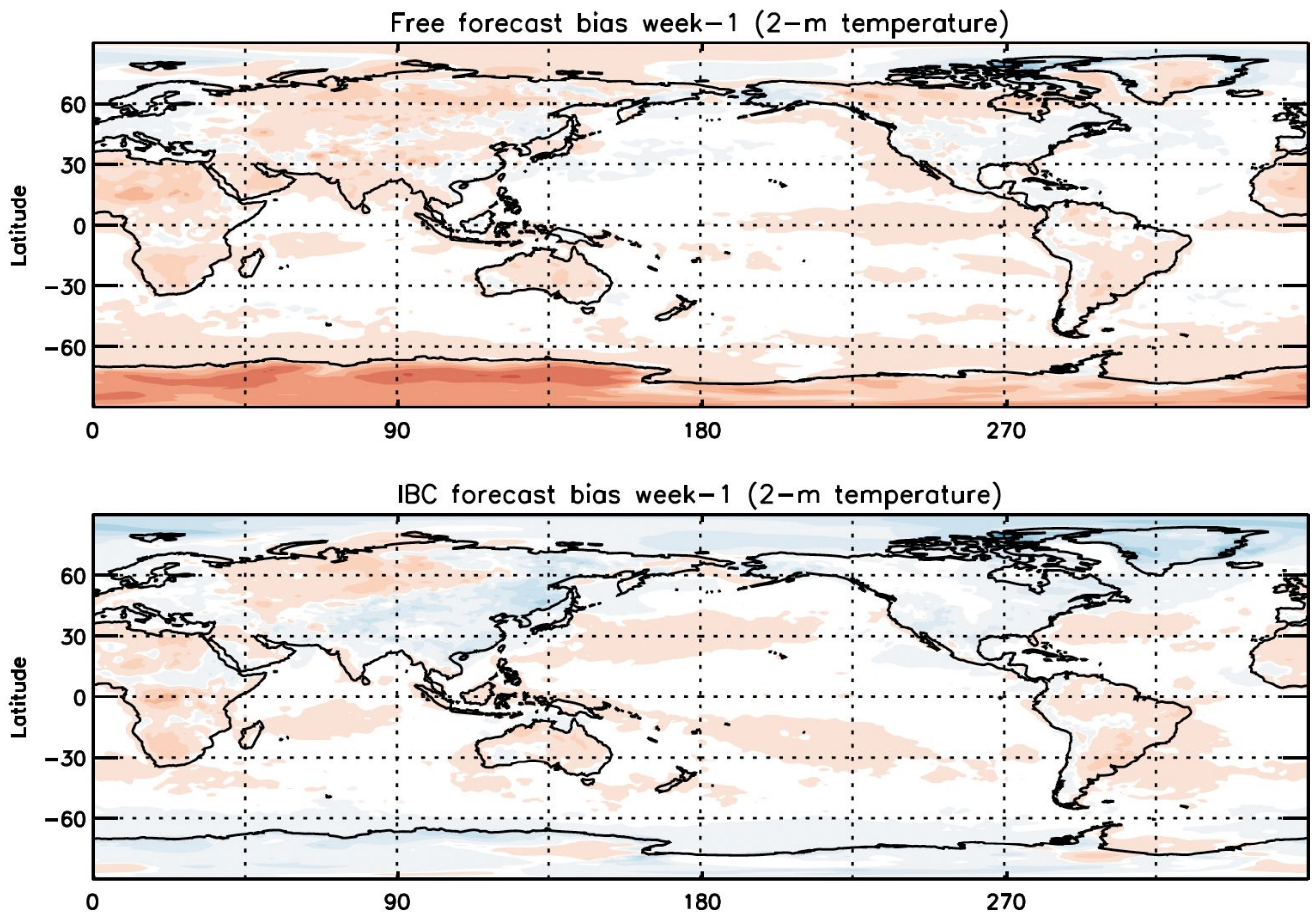
Initial attempt to reproduce Tse-Chun's result was unsuccessful



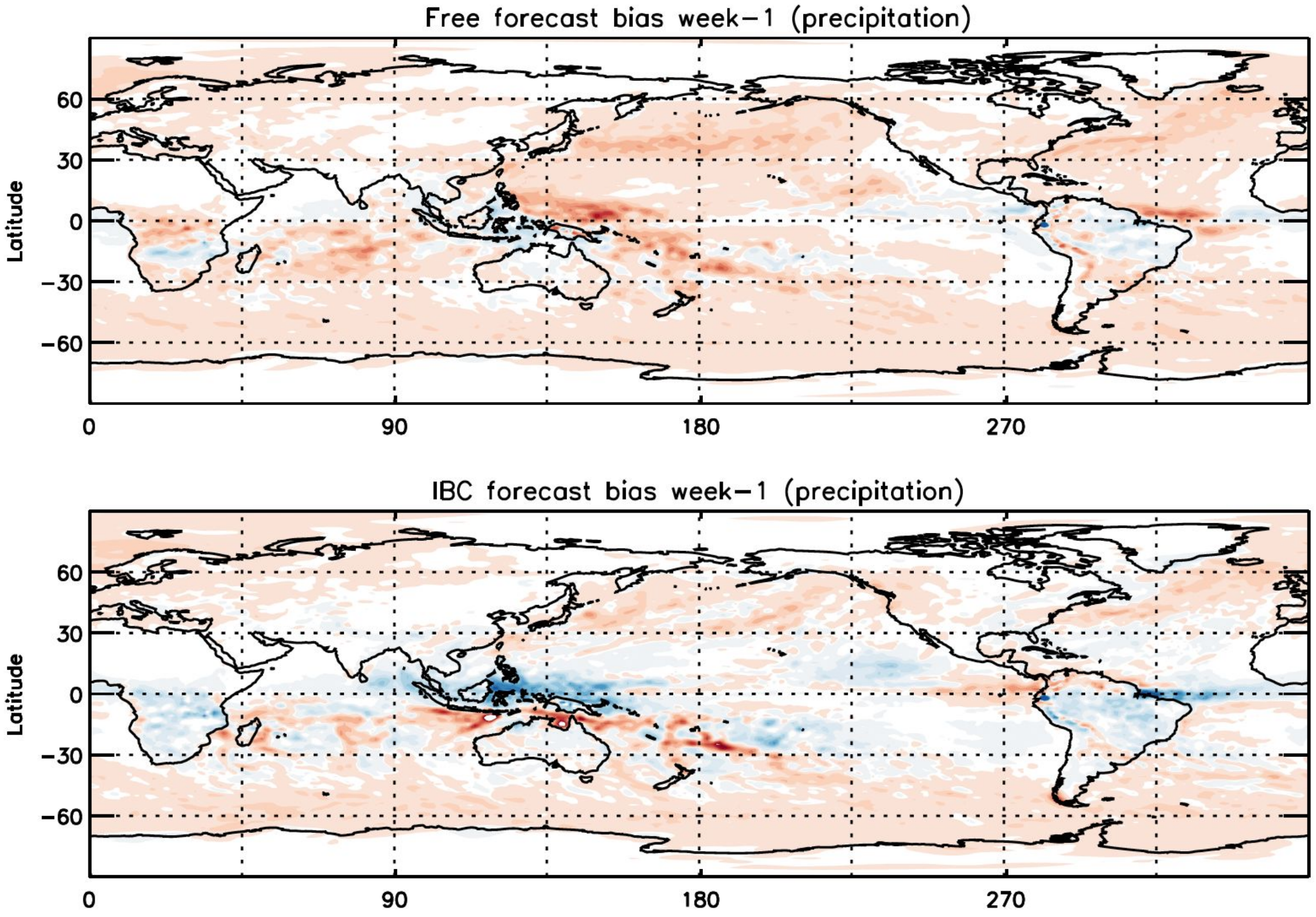
Reverting to a set of inputs similar to Chen et al. 2022 gets us closer



Meanwhile, bias maps of T-2m shows evidence of overcorrection



The story is similar for precipitation



Conclusions and next steps

- Considerably more work is needed to fully evaluate the potential of IBC for advancing numerical weather and S2S predictions
- Here, the choice of inputs was seen to critically affect model performance, despite having less impact on NN predictive capability
- Future work will further examine sensitivity to inputs, as well as the effect of trickling in the IBC corrections.

Thanks!