Reducing the Effects of Model Errors by Augmenting the Numerical Model with a Machine Learning Component

Istvan Szunyogh

Texas A&M University, Department of Atmospheric Sciences
College Station, TX

6th WGNE Workshop on Systematic Errors in Weather and Climate Prediction
ECMWF, Reading, UK, October 31-November 4, 2022

Coauthors and Publications

Coauthors:

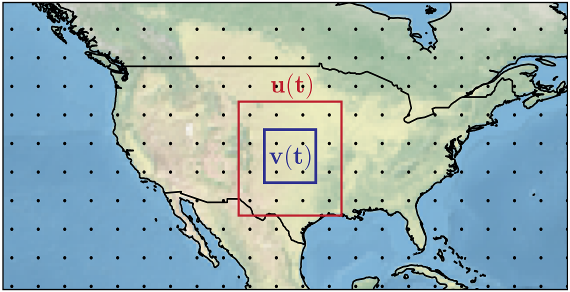
- Troy Arcomano (Texas A&M University)
- Brian Hunt (University of Maryland)
- Edward Ott (University of Maryland)
- Alex Wikner (University of Maryland)

Related Publications:

- Wikner, A., J. Pathak, B. Hunt, M. Girvan, T. Arcomano, I. Szunyogh, A. Pomerance, and E. Ott, 2020: Combining machine learning with knowledge-based modeling for scalable forecasting and subgrid-scale closure of large, complex, spatiotemporal systems. *Chaos,* **30,** 053111.
- Arcomano, T., I. Szunyogh, J. Pathak, A. Wikner, B. Hunt, and E. Ott, 2020: A machine learning-based global atmospheric forecast model. *Geophys. Res. Lett.,* **47,** e2020GL087776.
- Arcomano, T., I. Szunyogh, A. Wikner*, J. Pathak*, B. Hunt, and E. Ott, 2022: A hybrid approach to atmospheric modeling that combines machine learning with a physics-based numerical model. *JAMES,* **14,** e2021MS002712.
- Arcomano, T., 2022: *Machine Learning Applications for Weather and Climate Modeling.* Ph. D. Thesis, Texas A&M University, College Station, Texas. United States.

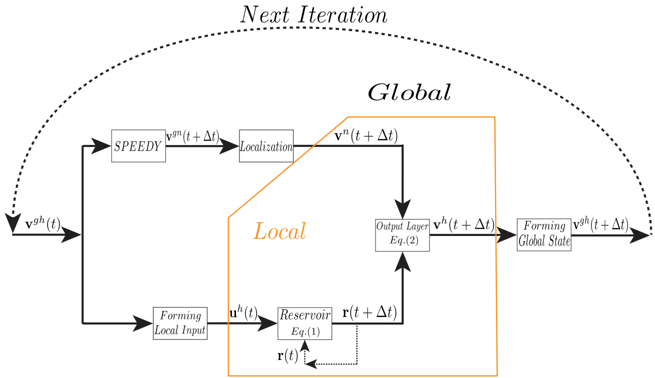
Combined Hybrid-Parallel Prediction (CHyPP)

Our **hybrid model** (Arcomano et al. 2021) implements the **Combined Hybrid-Parallel Prediction (CHyPP)** technique of (Wikner et al. 2020) on Version 41 of the **Simplified Parameterization, primitive-Equation Dynamics (SPEEDY)** AGCM



ML calculations are done for local subdomains (e.g., blue rectangle) in parallel

Flow Chart of the Hybrid Model



$$\mathbf{r}(t+\Delta t) = \tanh\left[\mathbf{A}\mathbf{r}(t) + \mathbf{B}\mathbf{u}^h(t)\right], \tag{1}$$

A: a **weighted adjacency matrix** of a low-degree, directed, random graph, and the **output layer** is

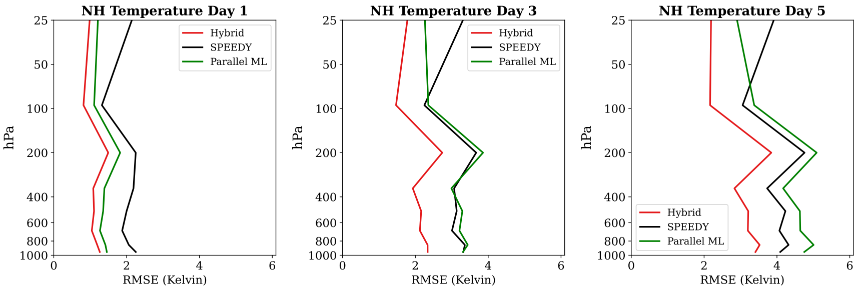
$$\mathbf{v}^h(t+\Delta t) = \mathbf{W}_{mod}\mathbf{v}^n(t+\Delta t) + \mathbf{W}_{res}\mathbf{r}(t+\Delta t), \tag{2}$$

where $\mathbf{W} = (\mathbf{W}_{mod}\mathbf{W}_{res})$ is a **matrix of parameters determined by training** (minimizing a quadratic cost function of the difference between $\mathbf{v}^h(t+\Delta t)$ and the training data $\mathbf{v}^a(t+\Delta t)$)

Forecast Experiments

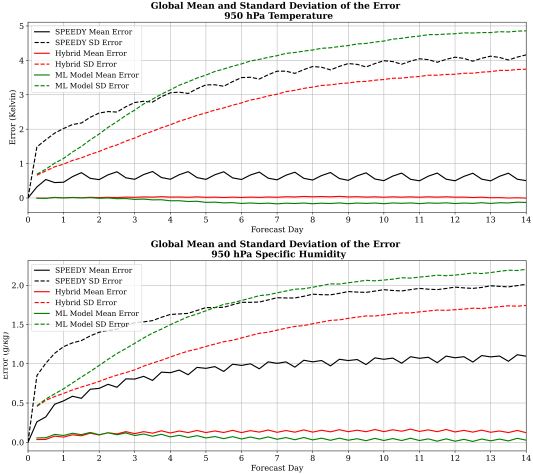
- **Training Data:** ERA5 reanalyses (from 1 January, 1990 to 26 June, 2011)
- **"Time Step":** 6 h
- **Forecasts:** 100 21-day forecasts equally spaced in time between 27 June, 2011 and 28 July, 2012
- **Forecast Verification Data:** ERA5 reanalyses
- **Benchmark Forecasts:**
  - **SPEEDY**
  - **ML-only**

Forecast Verification Results: NH Midlatitudes Temperature



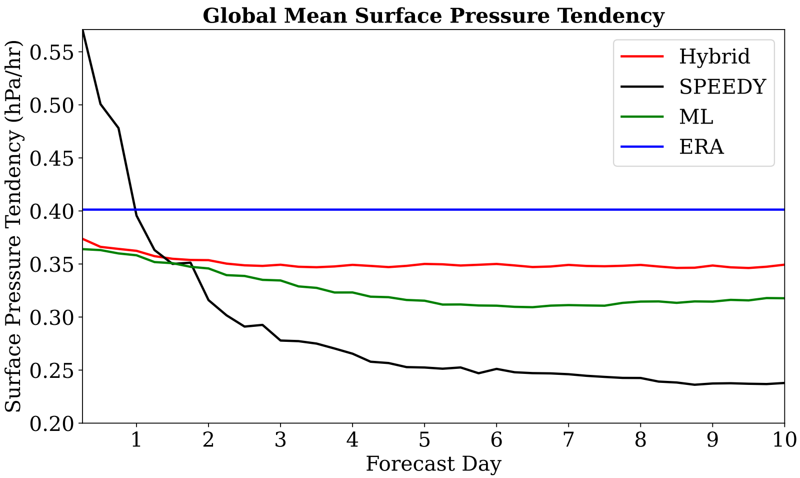
The hybrid forecasts are more accurate than either the numerical forecasts or the ML forecasts.

Forecast Verification Results: Global Error Statistics



The hybrid approach reduces the magnitudes of both the systematic and transient component of the forecast errors

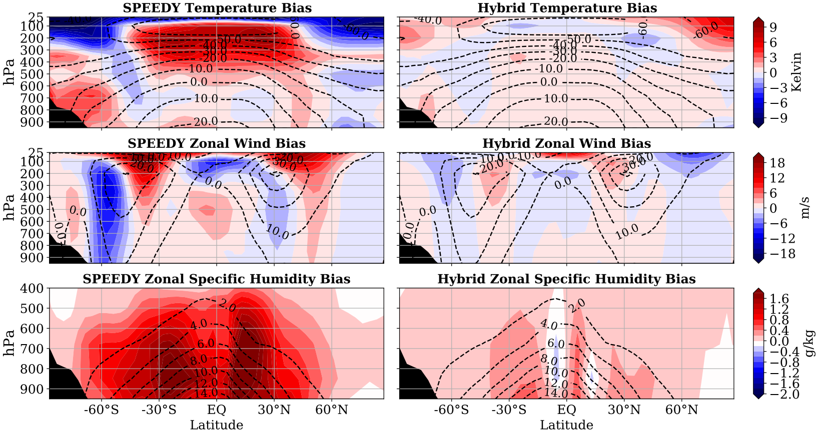
Forecast Verification Results: Atmospheric Balance



The hybrid model state is **well balanced** throughout the forecasts and produces the **more realistic surface pressure tendencies** than the benchmarks

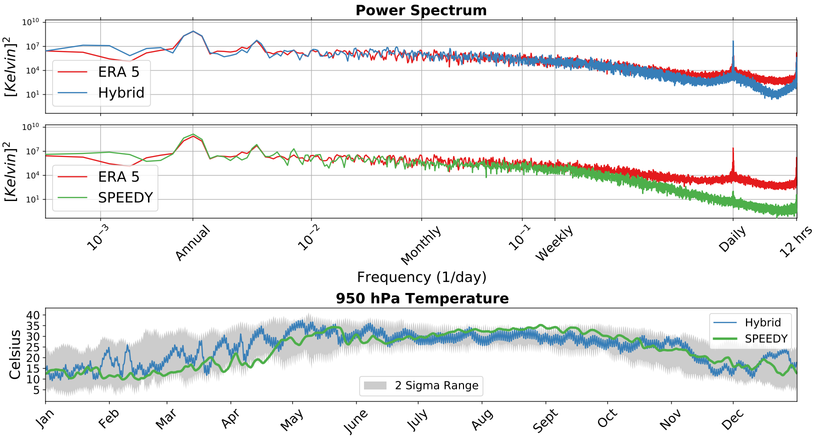
Climate Simulation Experiment: Biases

- Training: 19 years of ERA5 data (January 1981-December 1999)
- Simulation: 11-year free run with hybrid model (first year is discarded)



Boreal Winter (DJF)

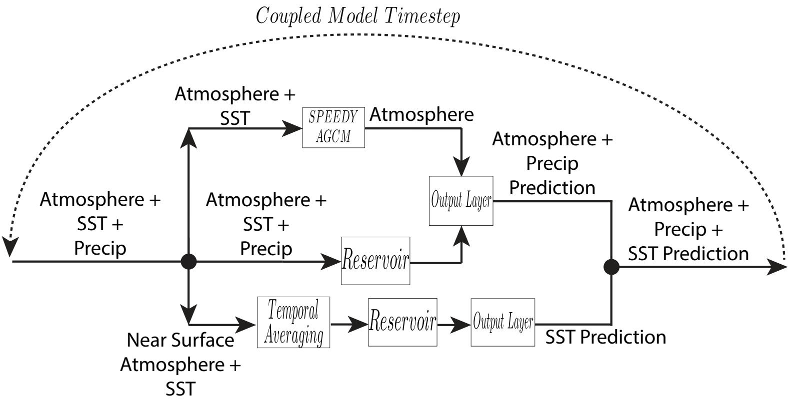
Climate Simulation Experiment: Atmospheric Variability



925 hPa Temperature in the Sahara Desert

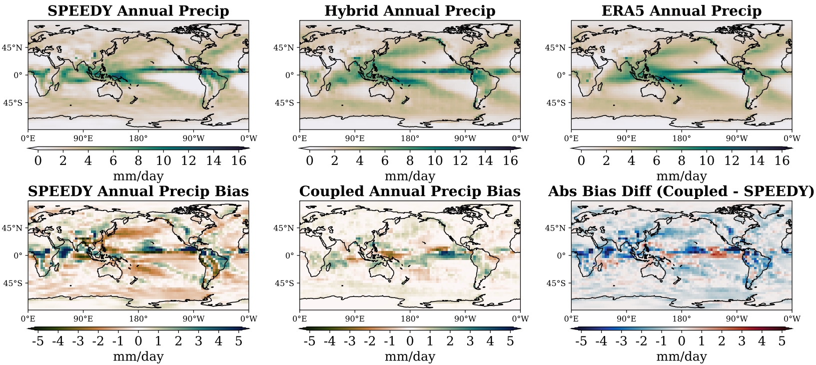
Observed frequency: 95%, Frequency for **Hybrid: 98%**, Frequency for **SPEEDY: 88.2%**

Adding ML-based Prognostic Variables: Illustration for Precipitation and SST



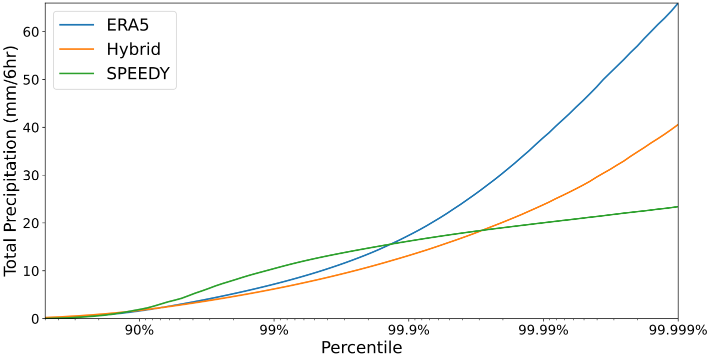
Flowchart of the approach

Annual Precipitation Totals



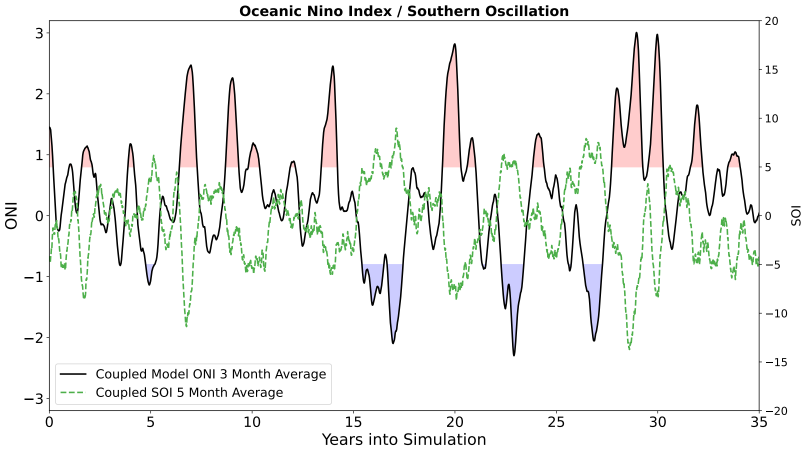
Correlations between ERA5 and **hybrid: 0.96**, and ERA5 and **SPEEDY: 0.82**

Extreme Precipitation Events



The hybrid captures the frequency of the heavier precipitation events better than the numerical model

The Model Produces an ENSO Cycle



By adding the ML-based SST as a prognostic variable, the model produces an ENSO cycle

Concluding Remarks

- Even though the ML components were **trained to minimize short-term forecast errors** (6 hours for atmosphere and 7 days for SST), they **effectively reduce the systematic model errors** even in a decades-long simulations
- The proposed approach offers a computationally efficient, simple technique to **add ML-based prognostic variables** to a numerical (or hybrid numerical-ML) model
- **Caveat:** It is yet to be seen that the technique also works effectively for a state-of-the-art numerical model