Contribution ID: **34**   Type: **Oral presentation**

S2S forecasting using large ensembles of data-driven global weather prediction models

Wednesday, 7 October 2020 17:00 (30 minutes)

We develop an ensemble prediction system (EPS) based on a purely data-driven global atmospheric model that uses convolutional neural networks (CNNs) on the cubed sphere. Mirroring practices in operational EPSs, we incorporate both initial-condition and model-physics perturbations; the former are sub-optimally drawn from the perturbed ECMWF ReAnalysis 5 members, while the latter are produced by randomizing the training process for the CNNs. Our grand ensemble consists of 320 perturbed members, each of 32 CNNs run with 10 perturbed initial conditions. At lead times up to two weeks, our EPS lags the state-of-the-art 50-member ECMWF EPS by about 2-3 days of forecast lead time, and is modestly under-dispersive, with a spread-skill ratio of about 0.75.

For weekly-averaged forecasts in the sub-seasonal-to-seasonal forecast range (2-6 weeks ahead), a particularly challenging window for weather forecasting, our data-driven EPS consistently outperforms persistence forecasts of 850-hPa temperature and 2-meter temperature, with useful skill relative to climatology as measured by the ranked probability skill score and the continuous ranked probability score (CRPS). Over twice-weekly forecasts in the period 2017-18, the CRPS of our model matches that of the ECMWF EPS to within 95% statistical confidence bounds for T850 at week 4 and weeks 5-6. While our model performs similarly for T2 compared to T850, the ECMWF EPS includes a coupled ocean model, which results in much better T2 forecasts, especially over tropical oceans. Our EPS is closest to parity with the ECMWF EPS in the extra-tropics, especially during spring and summer months, where the ECMWF ensemble is weakest. Notably, our model, while only predicting six 2-dimensional atmospheric variables, runs extremely efficiently, computing all 6-week forecasts in under four minutes on a single GPU. Nevertheless, further work remains to incorporate ocean and sea-ice information into our data-driven model to improve its representations of large-scale climate dynamics.

Thematic area

1. Machine Learning for Model Identification and Development - Including Model identification, Fast Emulation of Parameterisations, Data driven Parameterisations

Primary authors: WEYN, Jonathan (University of Washington); DURRAN, Dale (University of Washington, Seattle, USA); Dr CARUANA, Rich (Microsoft)

Presenter: WEYN, Jonathan (University of Washington)

Session Classification: Session 5 (cont.): ML for Model Identification and development

Track Classification: ECMWF-ESA Workshop on Machine Learning for Earth System Observation and Prediction