Assessing Teleconnections-Induced Predictability of Regional Water Cycle on Seasonal to Decadal Timescales Using Machine Learning Approaches

Salil Mahajan[1], Linsey S. Passarella[1,2], Forrest M. Hoffman[1], Murali G. Meena[1], and Min Xu[1]

[1]Oak Ridge National Laboratory, Oak Ridge, TN, USA; [2]University of Tennessee, Knoxville, Knoxville, TN, USA

Focal Area

(3) Insight gleaned from complex data (both observed and simulated) using AI, big data analytics, and other advanced methods, including explainable AI and physics- or knowledge-guided AI

Science Challenge

Predictive understanding of the regional water cycle based on teleconnections of climate modes of variability using AI methods.

Rationale

Low-frequency climate modes of variability like the El Nino Southern Oscillation (ENSO) offer potential predictability of the regional water cycle and its extremes on seasonal to inter-annual timescales across the globe. However, there is a continuum of ENSO flavors, broadly divided into two kinds—Eastern Pacific (EP) and Central Pacific (CP, Modoki)—each with its varied anomalous sea surface temperature (SST) patterns (Timmerman et al. 2018). These different SST patterns induce varied Rossby wave-trains that impact remote hydrological cycles in different ways as they interact with background atmospheric flow. Moreover, the background atmospheric flows exhibit strong intrinsic variability on sub-seasonal scales and respond to other forcing sources like the Arctic, making water cycle predictions challenging during ENSO events. A prominent example of the knowledge gap in water cycle predictability during ENSO is the failed prediction of above-normal winter and spring precipitation over California (after a severe multi-year drought) by both statistical and dynamical models during the 2015–2016 strong El Nino event, when Nino3.4 region SSTs featured strong positive SST anomalies in November 2015 (Cohen et al. 2017). This was contrary to the 1982–1983 and 1997–1998 canonical (EP) El Nino events that brought in heavy precipitation over California (L'Heureux et al. 2017). Studies suggest that the failure to predict the CP ENSO type by full complexity seasonal forecast systems (like the North American Multimodel Ensemble, which includes models like NCEP CFSv2 and GFDL CM) resulted in large water cycle forecasting errors, despite skillful Nino3.4 predictions.

The emerging understanding of the diversity of ENSO (e.g., Johnson et al. 2013) forebodes that a future realization of a non-canonical ENSO—not necessarily the CP El Nino—from the wide spectrum of the ENSO flavors, along with a warming climate, could again result in severely flawed predictions. The above suggests two major issues with the current assessment of ENSO-induced predictability. First, a unified approach to quantifying the climate impacts of ENSO that accounts for the full ENSO diversity remains demonstrably lacking. Second, the predictive skill of seasonal (to decadal) forecasting models for predicting the ENSO flavor is poor even after the spring barrier. We present ideas based on machine learning (ML) methods to address these issues for an improved predictive understanding.

Recent work suggests that ML approaches can predict the Nino3.4 index more skillfully than dynamical forecast systems with lead times of more than a year (Yan et al., 2020, He and Eastman 2020, Ham et al. 2019, Dijkstra et al. 2019), largely using different types of neural networks, for

DOI: https://doi.org/10.2172/1769676
DOE BER, AI4ESP 2021 #AI4ESP1086 (https://www.ai4esp.org)

example, deep neural networks, convolution neural networks (CNN), and recurrent neural networks like the convolution long short-term memory. The lack of sufficient observational data was overcome in one study (Ham et al. 2019) by using Earth system model (ESM) simulations: a CNN network was trained with global SST and heat content data from historical simulations of 21 CMIP5 models to predict the Nino3.4 index. The estimated network weights were then used as the initial weights for retraining the network with a 103-year segment of a reanalysis dataset—a technique known as transfer learning. A classification CNN also demonstrated more skill than dynamical systems in predicting ENSO types (CP, EP or mixed type, distinguished using a combination of Nino3 and Nino4 index) in hindcast experiments, with a network that was trained solely on CMIP5 data. These works attempt only to predict an ENSO index and not the full tropical Pacific SST pattern, an important factor for assessing the predictability of ENSO diversity and its impacts.

Narrative

We premise that ML techniques can be useful in both

1. finding effective teleconnections of climate modes of variability to local water cycles and
2. assessing the predictability of ENSO and its diversity

ML-based Unified ENSO Index and Teleconnections

ENSO indices have traditionally been identified using linear models like principal components (PCs). Nonlinear physical approaches to identifying ENSO events have continually been explored, too (e.g., Williams and Patricola 2018). Deep generative networks, like autoencoders-decoders (AE), in combination with explainable AI methods (tools like heat maps), offer a new avenue to nonlinearly identifying variability patterns of tropical Pacific SSTs (and thus an index). Shallow AEs have been shown to replicate spatial and temporal characteristics of PCs (He and Eastman 2020). A more targeted approach to identifying generative networks that elicit statistically significant ENSO teleconnections to remote regions globally—able to cut through the noise of intrinsic atmospheric variability—could enhance our predictive understanding of ENSO's remote impacts. Evaluating AE modes between and across ESMs would also provide deeper insights into model biases over the tropical Pacific that might be limiting the predictive skill of these dynamical systems and would inform future observational site selection processes. Likewise, ML networks in which the predictand is regional precipitation (say, over California) and the predictors are tropical Pacific (or global) SSTs can also isolate important relationships with a regional focus that the community may have missed because of its reliance on linear regression approaches.

ESM Simulation-Informed ML for Predictive Understanding of ENSO Diversity

Transfer learning approaches using ESM simulation data to pretrain a network are in the nascent stages. ESMs have strong biases both in the mean background state and in ENSO simulation that are consistent across models. Network weights pre-trained with such model data would be optimized for loss-function minima that may not necessarily correspond to the global minima for observed data, yielding networks with poor predictive skill.

A high-resolution digital twin of the Earth system based on the DOE Energy Exascale Earth System Model (E3SM), assimilating atmosphere, land, ocean, and sea-ice observed conditions and contributing to routine seasonal to decadal forecasting, is foreseeable in the next few years. Recent work suggests that high-resolution ESMs, including E3SM, simulate the water cycle more accurately. Some ESMs also can capture the diversity of the ENSO continuum (Capotondi and

Wittenberg 2013). However, seasonal and decadal forecasting dynamical models suffer from coupled model drift. Posteriori linear bias correction of fully initialized models (with observations), anomaly initialization, and flux correction (enthalpy, momentum, and freshwater fluxes) approaches are commonly used for seasonal and decadal forecasting (Magnusson et al. 2013).These have been shown to improve seasonal prediction skills, for example, of tropical cyclones (Vecchi et al. 2014). Short, bias-corrected E3SM digital twin simulations (using advanced AI techniques for posteriori bias correction or anomaly initializations) can be used to generate large ensembles of ENSO evolution of different ENSO types across the ENSO continuum (including different scenarios of radiative forcings from greenhouse gases, aerosols, and so on). Conducting large ensembles packed together is one of the strategies for effectively exploiting exascale computing by the E3SM project. Such an ensemble, combined with control and historical simulations, would provide a wider pool of pre-training data sets more closely related to observed data. An optimized sampling from the ensemble may reduce the possibility that the computed transfer-learning network weights will be attracted to the minima of free running models, while increasing the likelihood of finding a global minimum pertinent to the real world when the network is retrained with observed data.

It is critical for ML-based ENSO forecasts to accurately predict not only the Nino3.4 index but also the entire tropical Pacific SST to capture ENSO diversity. Also, a more comprehensive use of atmospheric and upper ocean data, for both tropical Pacific and remote regions, can enhance predictive skill. Further, most studies use CNNs, which create filters only for local features. Networks that can learn about the spatio-temporal correlations (physics induced) should also be investigated for an improved understanding of ENSO predictability. One recent study (Cachay et al. 2020) using graph neural networks (GNNs), which learns such correlations, suggests that they show improved performance in predicting the Nino3.4 index up to a lead time of 3 months using only global SSTs as predictors.

An ensemble of such ML-based (say GNN) seasonal hindcasts of the entire tropical Pacific, generated by perturbing initial conditions or by using network weights generated by different samples, would also allow for uncertainty quantification of the ML-based system. Preliminary work in ensemble forecasting with ML-based weather forecasting systems shows promise but indicates weaker error growth in ML-based systems that fail to capture the full chaotic nature of weather (Scher and Messori 2018). Investigating the improvements in predictive skills with these ML-based prediction models would provide new insights (from explainable AI-based approaches) into the biases in dynamical forecasting that result in its poor skill. Other hybrid approaches combining AI with simulation data (like deep reinforcement learning), successfully applied in other fields (e.g., autonomous driving), can also provide useful tools for understanding ENSO predictability.

Combining ML-based predictions with an ML nonlinear ENSO index and teleconnection patterns would allow for a comprehensive exploration of the predictability of the ENSO-induced regional water cycle and its extremes globally. Here, we have presented ideas on the use of AI for ENSO-induced predictability as an example; but note that similar approaches could also be used for other modes of natural variability such as the Atlantic Nino, Atlantic Meridional Mode, Indian Ocean Dipole, and others.

References

Cachay, S. R., Erickson, E., Bucker, A., Pokropek, E., Potosnak, W., Osei, S., and Lütjens, B. 2020. *Graph Neural Networks for Improved El Niño Forecasting*.

Capotondi, A., and Wittenberg, A. 2013. ENSO diversity in climate models. *U.S. CLIVAR Variations*, **11**, 10–14.

Cohen, J., Pfeiffer, K., and Francis, J. 2017. Winter 2015/16: A Turning Point in ENSO-Based Seasonal Forecasts. *Oceanog.*, **30**, 82–89, https://doi.org/10.5670/oceanog.2017.115.

Dijkstra, H. A., Petersik, P., Hernández-García, E., and López, C. 2019. The Application of Machine Learning Techniques to Improve El Niño Prediction Skill. *Front. Phys.*, **7**, https://doi.org/10.3389/fphy.2019.00153.

Ham, Y.-G., Kim, J.-H., and Luo, J.-J. 2019. Deep learning for multi-year ENSO forecasts. *Nature*, **573**, 568–572, https://doi.org/10.1038/s41586-019-1559-7.

He, J., and Eastman, J. R. 2020. A Sequential Autoencoder for Teleconnection Analysis. *Remote Sensing*, **12**, 851, https://doi.org/10.3390/rs12050851.

Johnson, N. C. 2013. How Many ENSO Flavors Can We Distinguish? *Journal of Climate*, **26**, 4816–4827, https://doi.org/10.1175/JCLI-D-12-00649.1.

L'Heureux, M. L., et al., 2017. Observing and Predicting the 2015/16 El Niño. *Bulletin of the American Meteorological Society*, **98**, 1363–1382, https://doi.org/10.1175/BAMS-D-16-0009.1.

Magnusson, L., Alonso-Balmaseda, M., Corti, S., Molteni, F., and Stockdale, T. 2013. Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Clim Dyn*, **41**, 2393–2409, https://doi.org/10.1007/s00382-012-1599-2.

Scher, S., and Messori, G. 2018. Predicting weather forecast uncertainty with machine learning. *Quarterly Journal of the Royal Meteorological Society*, **144**, 2830–2841, https://doi.org/https://doi.org/10.1002/qj.3410.

Timmermann, A., et al. 2018. El Niño–Southern Oscillation complexity. *Nature*, **559**, 535–545, https://doi.org/10.1038/s41586-018-0252-6.

Vecchi, G. A., et al. 2014. On the Seasonal Forecasting of Regional Tropical Cyclone Activity. *Journal of Climate*, **27**, 7994–8016, https://doi.org/10.1175/JCLI-D-14-00158.1.

Williams, I. N., and Patricola, C. M. 2018. Diversity of ENSO Events Unified by Convective Threshold Sea Surface Temperature: A Nonlinear ENSO Index. *Geophysical Research Letters*, **45**, 9236–9244, https://doi.org/https://doi.org/10.1029/2018GL079203.

Yan, J., Mu, L., Wang, L., Ranjan, R., and Zomaya, A. Y. 2020. Temporal Convolutional Networks for the Advance Prediction of ENSO. *Scientific Reports*, **10**, 8055, https://doi.org/10.1038/s41598-020-65070-5.