DOE BER Earth and Environmental Systems Science Division (EESSD) Call for White Papers to Advance an Integrative Artificial Intelligence Framework for Earth System Predictability: AI4ESP

1. Title: Integration of AI/ML with Data Assimilation for Earth System Prediction

2. Authors/Affiliations: Stephen G. Penny, Hsin-Yi Lin, Michael Goodliff, Tse-Chun Chen, Timothy Smith. *Affiliation: Cooperative Institute for Research in Environmental Sciences at the University of Colorado Boulder*

3. Focal Area(s): Development of dynamically consistent AI/ML methods that can integrate with the data assimilation cycle to improve the efficiency and effectiveness of state estimation and prediction at short to medium range forecast horizons.

4. Science Challenge: This white paper addresses a first step toward complete integration of AI/ML and DA. Here, we address model emulation and its application to simple coupled atmosphere-ocean dynamics for use in data assimilation applications.

5. Rationale: Coupled data assimilation (CDA) is an emerging research area in which observations are used to initialize the complete Earth system model, while accounting for interactions across multiple coupled sub-component models. Currently, large scale coupled Earth system models are incredibly computationally expensive to run. This makes testing of new CDA methods extremely difficult and time-consuming. New methods are needed that accelerate some of the most limiting bottlenecks within the data assimilation cycle. Here we focus on the cost of model integration. By replacing numerical model integration with AI/ML emulators, we can reduce the cost of generating forecasts and estimating model uncertainty, while still producing the key information needed for data assimilation to produce accurate state estimates.

6. Narrative: Advances in coupled Earth system modeling, such as higher grid resolutions and a growing number of modeled component systems, are making the naive application of state-of-the-art data assimilation methods increasingly difficult. For example, the data assimilation approaches used by leading operational weather prediction centers leverage both variational and ensemble-based techniques. However, the increased costs associated with the Earth system models make ensemble approaches infeasible outside of large carefully planned exploratory experiments. And, the increased complexity of the Earth system models limit the application of variational methods such as 4D-Var, for which both a tangent linear and adjoint model are required but unavailable for such systems.

Promising data-driven attempts to emulate Earth system dynamics range from Linear Inverse Models (LIMs; *Penland and Sardeshmukh, 1995*) and their extension through variants of Dynamic Mode Decomposition (DMD; *Brunton et al., 2020*), to variants of recurrent neural networks including reservoir computing *(Acromano et al., 2020)* and convolutional Long Short Term Memory (LSTM) networks *(Sonderby et al., 2020)*. Recent advances improve understanding of recurrent neural networks such as reservoir computing as a type of

DOI: https://doi.org/10.2172/1769728
DOE BER, AI4ESP 2021 #AI4ESP1102 (https://www.ai4esp.org)

generalized synchronization between the AI/ML emulator and the original dynamics *(Platt et al., 2021)*; there are close parallels here with the underlying principles of data assimilation itself.

Our research team is testing and advancing any of these data-driven techniques for application to coupled Earth system models and for use in coupled data assimilation. Recent results from *Lin and Penny (2021)* indicate that a simple coupled atmosphere-ocean model using quasi-geostrophic dynamics *(De Cruz et al., 2016)* can be reasonably well modeled using a scalable reservoir computing framework that is provided only pointwise state information as training data (see Figure 1). These results provide the groundwork for transitioning the same techniques to be applied to long Earth system model simulations and observationally guided historical reanalyses, which are all typically stored as gridded fields of physical quantities to facilitate human analysis.

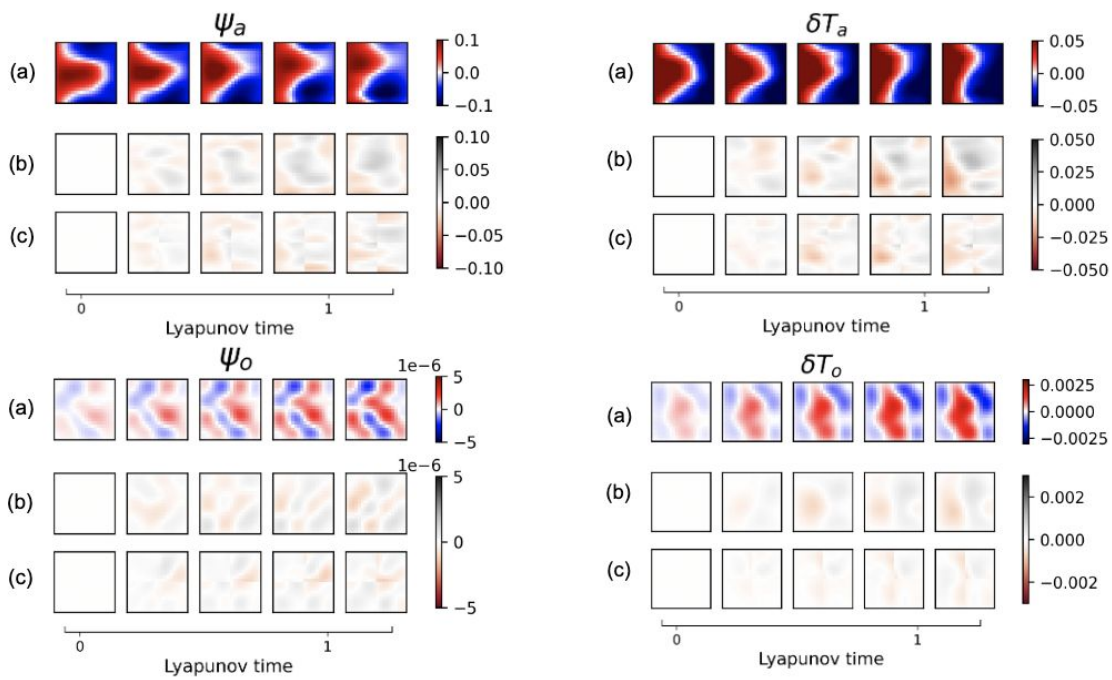

Figure 1. Prediction of stream function and temperature for atmosphere (top) and ocean (bottom). Each column shows five consecutive days, based on the timescale of the model. The rows show: (a) the target (or 'truth'), (b) target-minus-prediction by a modified reservoir computing approach, and (c) target-minus-prediction applying localization combined with the modified reservoir computing approach. (Reproduced from *Lin and Penny, 2021*).

7. Suggested Partners/Experts: Henry D.I. Abarbanel (UCSD/Scripps); Brian R. Hunt (UMD); Jason Hickey (Google Research); Stephan Hoyer (Google Research); Cecile Penland (CIRES); Prashant Sardeshmukh (CIRES); Matt Newman (CIRES); Greg Hakim (UW); Clark Rowley (NRL); Patrick Hogan (NOAA).

8. References

Arcomano, T., I. Szunyogh, J. Pathak, A. Wikner, B. R. Hunt, and E. Ott (2020). A machine learning-based global atmospheric forecast model. *Geophysical Research Letters*, **47**(9):e2020GL087776.

Brunton, S.L., B. R. Noack, and P. Koumoutsakos. Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics*, **52**:477–508, 2020.

De Cruz, L., J. Demaeyer, and S. Vannitsem. The modular arbitrary-order ocean-atmosphere model:maooamv1.0. *Geoscientific Model Development*, **9**(8):2793–2808, 2016. doi: 10.5194/gmd-9-2793-2016.

Lin, H.-Y., and S.G. Penny (2021): Fourier Reservoir Computing for effective prediction of multi-scale coupled quasi-geostrophic dynamics. *(Submitted for publication)*

Penland, C., & Sardeshmukh, P. D. (1995). The Optimal Growth of Tropical Sea Surface Temperature Anomalies, *Journal of Climate*, **8**(8), 1999-2024.

Penny, S.G., E. Bach, K. Bhargava, C-C. Chang, C. Da, L. Sun, T. Yoshida (2019). Strongly coupled data assimilation in multiscale media: experiments using a quasi-geostrophic coupled model. *Journal of Advances in Modeling Earth Systems*, **11**. https://doi.org/10.1029/2019MS001652

Platt, J.A., L. Fuller, A. Wong, R. Clark, S.G. Penny, H.D.I. Abarbanel (2021). Forecasting Using Reservoir Computing: The Role of Generalized Synchronization. *(Submitted for publication)*

Sønderby, C., Espeholt, L., Heek, J., Dehghani, M., Oliver, A., Salimans, T., Agrawal, S., Hickey, J., & Kalchbrenner, N. (2020). MetNet: A Neural Weather Model for Precipitation Forecasting. ArXiv, abs/2003.12140.