Advancing Regional Climate Predictability through ML-enabled Dynamical System Approach

Jian Lu[1], Zhangshuan (Jason) Hou[1], and Yi Deng[2],

[1]Pacific Northwest National Laboratory
[2]Georgia Institute of Technology

Focal Area Predictive modeling through the use of AI techniques and AI-derived model components; the use of AI and other tools to design a prediction system comprising of a hierarchy of models.

Science Challenge The lack of systematic framework and understanding of the operating dynamics and forcing-response relationship for the climate system has been long hindering the informed decision-making in addressing the imminent climate change crisis. The fast progresses in deep machine learning tools and their great promise of representing the dynamical system of the climate offer a novel opportunity to tackle the said challenge.

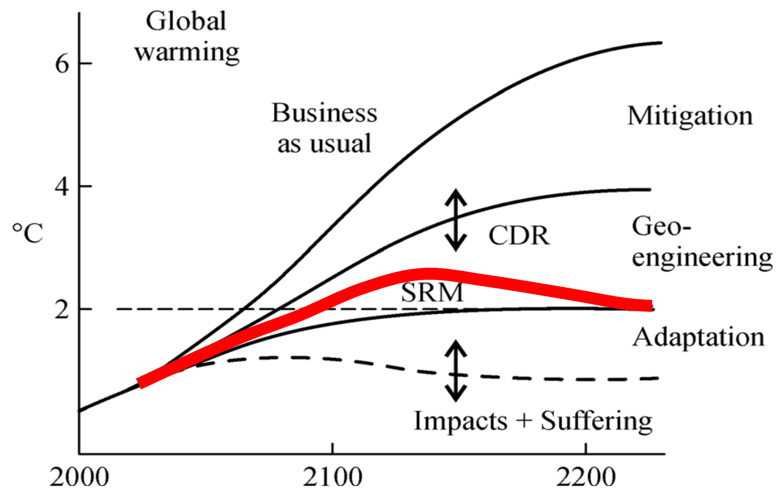
A Crisis: Catastrophic Climate Change

As climate crisis is ramping up, the challenges the climate research community is facing are profound and manyfold. Chief among them is the lack of confidence in the regional details in the projected changes of the hydrological cycle and the related circulation anomalies (e.g., Neelin et al. 2003). Without the regional details and the confidence therein, it is hard for stakeholders and policy makers to make decisions to mitigate or adapt to climate change (at state or even municipal levels). As the time window to act to keep the global warming from reaching a dangerous level (like the 2°C target set by the Paris Agreement) shrinks, the community is contemplating solar radiation management (SRM) as a complementary strategy to avert a catastrophic climate change scenario, or, at least to buy some time (see Shepherd's Napkin Diagram of SRM in the next page). However, SRM would be ethically and politically no-go if scientists could not solve the geoengineering conundrum of "what if one country's geoengineering negatively affects another country?" At the root of the conundrum is the climate change *regionality* problem, related to the interconnectedness between forcing everywhere and climate response everywhere. Heaven forbid, if the climate change impacts were so dire that humanity must take intervening actions, we still need to figure out where and how to place the radiation anomalies so that the regional side effects can be minimized—a *climate forcing optimization* problem.

A boost in our capability to address the scientific challenges above would be to build a dynamical system emulator of the climate and the associated forcing-response relationship function (FRR) linking the external climate perturbations with the climate impacts with significant social and economic consequences. The FRR functions as a highly efficient surrogate model, predicting the response for a given forcing, or vice versa. Depending on how the emulator is configured and trained, it may also help unveil the feedback relationships, both linear and nonlinear, from the variables characterizing a climate state or amongst themselves. However, given the chaotic nature, high dimensionality, nonlinearity, and hierarchical structure of the dynamical system of the earth's climate, the challenges are steep to obtain reliable and operable FRRs for the climate. Past research has been overwhelmingly occupied by the forcing-response-

DOI: https://doi.org/10.2172/1769654
DOE BER, AI4ESP 2021 #AI4ESP1083 (https://www.ai4esp.org)

feedback problem of the global mean temperature and hydrological cycle, whereas the science for solving the scientific challenges identified above is falling far behind. Nevertheless, the recent theoretical advance in dynamical system theories, particularly the generalization of the fluctuation-dissipation theorem (FDT[1]) towards non-Gaussian system, and the rapid growth in the machine learning (ML) and deep learning (DL) applications on big data and complex systems present a novel opportunity to inter-catalyze the progress on both fronts to establish quantitative FRRs for our climate system as well as the embedded hydrological cycle, thereby we can advance understanding of the regional climate feedbacks, build consensus and confidence in the regional climate response to external climate perturbations, and provide the quantitative information of the optimal forcing the solar geoengineering research community sorely needs.



John Shepherd's Napkin Diagram of SRM Red line indicates the SRM bridge to get to the 2°C target. The red line will always be the "red line" that should not be crossed, if the necessary international governance and science-informed compensation scheme are not in place, with the science part being predicated upon the quantitative understanding between the forcing and response of our climate system (which is far behind).

A solution: ML-enabled Dynamical System for Real Climate

The forcing-response problem in a broad class of physical systems can be addressed using the FDT. The recent advances of extending the FDT to quasi-Gaussian (through coarse-graining the system states, Majda et al. 2010) and non-Gaussian systems (through non-parametric statistics, Cooper and Haynes, 2011) have brought promise in applying FDT to climate response problem. One remarkable aspect of the FDT is that it does not require the explicit knowledge of the unperturbed governing equations of the system in question. This makes it especially suitable for data-driven deep learning approaches. Moreover, the applied math community is building the groundwork to bridge the exceedingly powerful ML tools with the continuous dynamical system: deep learning neural networks can be thought of as discrete dynamical systems or neutral-ODEs (Chen et al. 2019). Particularly, recent studies have succeeded in emulating prototype climate systems and predicting the phase space trajectories and their PDF using reservoir computing and recurrent neural network with long short-term memory (LSTM) (Lu et al. 2017; Pathak et al.

[1] The FDT originated from statistical physics and was introduced by Leith (1975) to climate problem. FDT gives an expression for the corresponding linear operator or linear response function, relating the response to the forcing, in terms of the statistics of the unforced system.

2018; Vlachas et al. 2018; Chattopadhyay et al., 2020). A similar emulator for the real-world climate, achievable through data-driven dynamical system emulation leveraging a very long climate simulation or a large set of perturbed simulations by Green's function-type forcing with the state-of-the-art climate model (e.g., Lu et al. 2020), would equip us with powerful tools with important applications as follows.

i. Map out the sources and pathways of predictability for a large variety of climate phenomena of interest, including the hydrological extremes. The resultant understanding of the source of predictability will help pinpoint the processes or components essential for prediction, so as to offer strategic guidance to model development. It will also advise the optimal use and design of observational networks such as ARM's facilities.
ii. Forward and backward FRR for the equilibrium response. The forward FRR can be used for climate response prediction/projection for a large number of samples of forcings; the backward FRR can be used to identify the optimal forcing for a given anomalous target climate state.
iii. FRR can be further interrogated through examining its dynamical modes, which serve as a powerful avenue for building emergent constraints and inter-model consensus (consensus is more readily to be built for modes than specific geographic locations).

To achieve the objectives above by building climate emulators for the real climate or the state-of-the-art climate models such as DOE E3SM, challenges but also opportunities abound. To name a few:

i. Data-driven versus physical law-abiding. As the space-dependent response and feedbacks are the cornerstone of the FRR, a good balance should be stricken between developing efficient, accurate, and reliable ML algorithms and the abidance of the physical constraints (conservation of mass/energy, etc., e.g. Kumar et al. 2020).
ii. Dilemma between high dimensionality and interpretability. For instance, reservoir computing is especially efficient for large-dimension systems, however the complexity of the architecture of the reservoir hinders the interpretability.
iii. High computational cost. To represent sufficient regional details, the dimension of the dynamical system for the climate will be inevitably large. Innovative ways of dimension reduction and system classification to represent the high-dimensional system should be explored.
iv. Structural instability of the climate system, because of the multiscale interactions between the coarse-grained, slow-scale dynamics and fast scale processes (such as convection and gravity waves). It can potentially obstruct the validity of the response prediction. Hierarchical structure representation (e.g., by using graphical networks) is needed for characterizing and modeling multiscale, multi-physics interactions.

These crosscutting challenges demand the collective wisdom from experts in the fields of the mainstream climate science, applied mathematics (especially dynamical system), and computational science. The research portfolio of DOE EESSD places itself in a unique and advantageous position to lead the charge to tackle these challenges with supreme importance and societal impacts. Therefore, we propose *Advancing climate predictability through ML-enabled dynamical system* as a key constituting theme for the related Focal Area. We are happy to provide a list of potential partners and experts who are the pioneers in the related areas, if the proposed theme for the workshop is selected.

References:

Chattopadhyay, A., P. Hassanzadeh, and D. Subramanian, 2020: Data-driven predictions of a multiscale Lorenz 96 chaotic system using machine-learning methods: reservoir computing, artificial neural network, and long short-term memory network. *Nonlin. Processes Geophys.,* **27**, 373-389.

Chen, R. T. Q., Y. Rubanova, J. Bettencourt, and D. Duvenaud, 2019: Neural ordinary differential equations. arXiv:1806.07366v5.

Cooper, F. C. and P. H. Haynes, 2011: Climate sensitivity via a nonparametric fluctuation-dissipation theorem. *J. Atmos. Sci.,* **68**, 937-953.

Kumar, A., T. Islam, T., Y. Sekimoto, Y., Mattmann, C., and B. Wilson, 2020: Convcast: An embedded convolutional lstm based architecture for precipitation nowcasting using satellite data. *PLOS ONE*, **15** (3), 1-18.

Leith C. E., 1975: Climate response and fluctuation dissipation. *J. Atmos. Sci.,* **32**, 2022–2026.

Lu, J., F. Liu, R. Leung, and L. Huan, 2020: Neutral modes of surface temperature and the optimal ocean thermal forcing for global cooling. *npj Climate and Atmospheric Science,* **9**, doi:10.1038/s41612-020-0112-6

Lu, Z., Pathak, J., Hunt, B., Girvan, M., Brockett, R., and Ott, E., 2017: Reservoir observers: Model-free inference of unmeasured variables in chaotic systems, *Chaos*, **27**, 041102, https://doi.org/10.1063/1.4979665

Majda, A. J., R. V. Abramov, and B. Gershgorin, 2010a: High skill in low-frequency climate response through fluctuation dissipation theorems despite structural instability. *Proc. Natl. Acad. Sci.* USA, **107**, 581–586.

Neelin J. D., C. Chou, and H. Su, 2003: Tropical drought regions in global warming and El Niño teleconnections. *Geophys. Res. Lett.,* **30**, L2275, doi:10.1029/2003GL018625

Pathak, J., Hunt, B., Girvan, M., Lu, Z., and Ott, E., 2018: Model free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach, *Phys. Rev. Lett.,* **27**, https://doi.org/10.1063/1.5010300

Vlachas, P., Byeon, W., Wan, Z., Sapsis, T., and Koumoutsakos, P., 2018: Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks, *P. Roy. Soc. A-Math.,* **474**, 20170844, https://doi.org/10.1098/rspa.2017.0844