Data-Driven Exploration of Climate Attractor Manifolds For Long-Term Predictability

Carlo Graziani (ANL), Robert Jacob (ANL), Sandeep Madireddy (ANL), Romit Maulik (ANL), Salman Habib (ANL), Roser Matamala (ANL), James Franke (University of Chicago), and Ian Foster (ANL)

Focal Area

This white paper responds to Focal Area 3. We seek to gain insight into decadal-scale climate predictability by applying novel manifold-finding probabilistic AI techniques to the complex data produced by Earth System models (ESMs) such as E3SM. The associated portfolio of research activities leverages DOE’s asset mix of HPC platforms, climate expertise, climate simulation codes, and AI expertise.

Science Challenge

Climate and climate models are dynamical systems exhibiting properties that are interpretable through chaos theory. The theory contains an important concept that is relevant to multi-decade-scale climate prediction: a chaotic attractor. While the space containing all the possible states of the Earth’s atmosphere and ocean, the possible weather, is large, the realized states tend to stay near the smaller-dimensioned attractor. This behavior is responsible for the “order behind the irregularity” [1] of climate phenomena. Climate change can be thought of as a change in the properties of the attractor, and predicting the climate over years to decades is equivalent to predicting how those properties will change. To date, the attractor has been a useful conceptual tool, but has not been amenable to direct characterization. A new development is the advent of efficient high-dimensional manifold-finding probabilistic AI techniques, which permit a data-driven characterization of the ESM attractor and its probability distribution over weather states. Such a characterization would result in a natural dimensional reduction — a “non-linear Principal Components Analysis (PCA) adapted to climate simulation data” — leading to important advances in scenario-based long-term climate prediction, long-term prediction of water cycle extremes, ESM verification, inter-model comparison, and process model development.

Rationale

Climate prediction at multi-decade scales is an urgent priority in Earth System Science. Predictions of sea level and precipitation changes, and of average temperature rise due to greenhouse gas (GHG) emission scenarios, are attended by substantial modeling uncertainties [2]. Analysis of the causes of these discrepancies is challenging due the very complex and high-dimensional nature of ESM output. In order to reduce and better quantify the associated uncertainties we require new approaches to the interpretation of climate model output data. In particular, information-preserving dimension
reduction of such data is highly desirable. A promising approach to the required reduction is to seek a data-driven representation of the dynamical attractor.

In terms of chaos theory, an attractor is a finite-dimensional subspace in the infinite-dimensional space of possible climate states to which all real weather is confined. In the same language, the "invariant measure" on the attractor is the probability of each state’s occurrence. The promise that these ideas hold for climate simulation is that while simulation output consists of very high-dimensional data (all climate variables on an Earth surface-, ocean-, and atmosphere-spanning mesh), there is likely a much lower-dimensional surface embedded in this space to which all the states are confined. It would be well-worth locating and characterizing this manifold, since it would furnish a maximally-informative dimensional reduction of the simulation output — effectively a "non-linear PCA adapted to climate simulation data." Analogously to PCA, this dimensional reduction could indicate clarifying and important features, correlations, covariances, and causalities.

Climate prediction is a qualitatively different enterprise from weather and seasonal/sub-seasonal forecasting: unlike such shorter-term forecasts, detailed current conditions have little bearing on decadal scale climate predictions and none at all on multi-decade scales, because memory of such conditions is progressively destroyed by chaotic dynamics on such timescales. Regions that are influenced by parts of the climate with a longer memory, such as the ocean, retain some specific predictability based on current conditions, but on multi-decade timescales even this memory is lost. Rather, we must seek to characterize how the distribution of weather states itself evolves under various forcing scenarios. The theoretical issue that underpins the problem of long-term climate prediction is summarized by the question what is the structure of ESM attractors, and how does those attractors evolve under exogenous and random forcings?

The Opportunity: The theoretical study of climate from the dynamical systems perspective is an active field with abundant literature [see, e.g. 3, and references therein]. Nonetheless, the climate attractor has remained a rather abstract conception, useful for clarifying discussion but not amenable to explicit representation. With the advent of new probabilistic AI/ML manifold-finding methods and ever more powerful HPC resources, a new opportunity has arisen: data-driven exploration of the climate attractor. We can now deploy at scale computational AI models with expressive capacity comparable to the expected number of degrees of freedom associated with the attractor of a computational ESM, or that of a model subsystem. Using these AI models we expect enhanced predictability of ESMs in virtue of: (1) the capability to sample weather states from the resulting AI-based distribution, resulting in a novel kind of weather emulator, and in a new tool for the study of the statistics of water cycle extremes; (2) new information concerning what variables act together in important and possibly causal ways affecting climate prediction beyond seasonal scales; (3) new tools for the study of the differences between climate model outputs, and of the sensitivity of any climate model to its constituent process models; and (4) separate study of the attractors of climate subsystems, whether geographical (e.g. Tropical Pacific, Great Lakes) or variable projections (e.g. water vapor fluxes, stratospheric wind speeds), and of relations of such attractor structures to each other and to the properties of the models from which they arise.

Narrative

AI/ML Approaches such as autoencoders [4] have proven adept at locating data submanifolds in high-dimensional spaces. Probabilistic variants — variational autoencoders (VA) [5], or variational information bottleneck (VIB) machines [6, 7, 8], possibly chained to powerful distribution-learning tools such as normalizing flows [9, 10] — can both find submanifold and approximate the data distribution on the submanifold. These methods can allow direct characterization of the structure
and evolution of dynamical attractors of ESMs and their subsystems.

Many approaches are feasible depending on the question to be answered, and the search for optimal approaches is an important research question. A “reference implementation” approach is to gather simulation output from ensembles of runs of a single model, and feed the complete data set, or a geographically-delimited dataset, or a physical subsystem dataset, to a probabilistic autoencoder-type manifold-learning AI code, such as a VIB system [6, 7, 8]. After training, the output is a low(er)-dimensional representation of the input data which may be directly queried for geometric and statistical properties, as well as sampled. We may, in this way, obtain compact weather state distributions that describe sea-surface temperature (SST) distributions, or tropical depression statistics.

V&V: Attractors may be very useful tools for studying sensitivities and GHG response differences between models. Inter-model comparison and reconciliation assisted by such studies could reduce spreads in decadal-scale scenario predictions, which are considerable at this time [2].

Model validation using data from weather stations, LIDAR, optical, UV and radar observations from ground-based and satellite platforms can play a key role in constraining attractor models, both through data assimilation-mediated process model parameter estimation and through direct application of observation operators to weather states sampled from the data-driven model. Examples of the latter are weather variable interpolations from model mesh boxes to station locations, or line-of-sight integration of optical signals across a model mesh. In this way the data-driven models can be compared to current and past observations, offering novel verification opportunities.

Dynamical Theory: Ideas from theoretical climate dynamics such as pull-back and stochastic attractor concepts [3] may be used to fashion attractor models to compare to the data-driven representation, or used to suggest specific analyses that extract information from that representation. Further fruitful interactions between dynamical theory and the data-driven attractor model may include linear resolvent analyses [11, 12], approachable through automatic differentiation of climate codes, or through proper orthogonal decomposition (POD) approaches [13].

Subsystems: Subsystem attractor analysis also makes possible the identification of key processes (or factors) of attractor geometry and state distribution. Noteworthy anomalies, such as those associated with tropical precipitation can be put under the microscope, to aid the search for variable interrelations associated with such anomalies. For example by exploring the importance of different factors, such as those associated to vegetation dynamics, production of volatile organic compounds (VOCs), and others, in driving precipitation events and other phenomena by analyzing their effects when they are included or not included in different models. This may also be a valuable approach for the study of anthropogenic effects due to land-use or agricultural policy.

Impacts: Decadal-scale prediction of climate can be closely coupled to the study of economic and social impacts. Climate change is already negatively impacting many economic sectors[14] including food production [15] as robust changes in the hydrology cycle [16, 17] impact rainfed agricultural in many regions around the world. Climate variability explains a large fraction of the historically variability in agricultural yields [18, 19] and the near-term changes in precipitation variability and extremes (which remain underrepresented in models[20]) need to be characterized to improve risk assessments to food production[21]. Improving understanding of expected climatological of the variability in the hydrological cycle including e.g. the El Niño Southern Oscillation (ENSO)[22], the Indian Ocean Dipole (IOD)[23], and the South Asian Monsoon more broadly [24] can therefore can directly lead to improved decadal-scale prediction of food and economic security. Other hydrology-related impact domains include infections diseases [25], vector-borne diseases [26], flooding [27], hydropower [28], and wildfires [29].
Suggested Partners/Experts

1. Michael Ghil and the UCLA Theorerial Climate Dynamics Group.

2. Leonard Smith Virginia Tech.

References


