1. Title:
A Modular System for Increasing Predictiveness for Extreme Climate Predictions

2. Authors/Affiliations:
Christopher Rackauckas (MIT Mathematics, UMD)
Alan Edelman (MIT Mathematics, ComputerScience & AILabs), Raffaele Ferrari (MIT EAPS),

3. Focal Area 2: 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.

4. Science Challenge:

We know that climate change is poised to reshape our world, but we lack clear enough predictions about precisely how. The preponderance of these changes is associated with human activity, specifically the emission of CO_2 and other greenhouse gases. Problematically, projections of climate change continue to be marred by unacceptably large uncertainties which hamper informed decision-making and cost society a chance to adapt proactively and effectively. These uncertainties stem from deficiencies in predictions of future greenhouse gas emissions, but also from inaccuracies in the representation of the physical models used to predict the climate response to such emissions. The uncertainties in projections associated with the inaccurate representation of climate physics, chemistry and biology are similar to those that plagued the first global climate models developed fifty years ago, despite more than a factor 10^8 increase in computer performance. Our transformational question is then, how can the accuracy of climate projections be dramatically improved by applying recent advances in the computational and data sciences to train the models with the wealth of data being constantly collected about the ongoing changes in the climate system?

The deficiencies in the representation of the Earth system do not result primarily from lack of understanding of the basic laws of physics, chemistry or biology; rather they arise from the poor representation of the processes that occur at scales smaller than the computational grid of the models, such as atmospheric clouds and ocean turbulence. These processes are represented through semi-empirical schemes called parameterizations, which are typically developed independently of the model into which they are incorporated and are tested against sparse observations from field studies in a few limited locations. As a result these parameterizations are often quite inaccurate especially in predicting extreme, rare events. We propose to upend this conventional approach by exploiting advances in machine learning and real world data assimilation to train parameterizations with global observations from satellites,

DOI: https://doi.org/10.2172/1769647
DOE BER, AI4ESP 2021 #AI4ESP1038 (https://www.ai4esp.org)

ocean floats, and other platforms. This will allow us to reduce and quantify uncertainties in climate predictions.

5. Rationale: Description of the research needs/gaps, the barriers to progress, and the justification for and benefits associated with the proposed approach

While some research groups have been showing the potential benefit that can be offered by mixing machine learning into ocean models, this technique has not been perfected enough to be used at scale at this time. Prototypes have shown that course models could potentially have orders of magnitude higher fidelity, allowing accurate and direct predictions of climate events to utilize for policy decisions. “How high of a sea wall does New York need?” is a question where a precise answer can save billions of dollars, and it rests on this foundational change and the accuracy that follows. Yet, some group needs to lead the charge to put such a technique into actual production.

The core piece which is required for the deployment of such machine learning into climate modeling practice is the development of a modular system. At any time the true physics needs to be able to be swapped in for checking accuracy, and the system needs to be able to easily handle such model possibilities to allow for easily testing and tuning models. Such a system needs to be compatible with differentiable programming, as without it the models themselves would not be able to fit neural networks within the context of their true predictions. This modular system will allow for testing the accuracy of all kinds of parameterizations, including those developed by national labs, giving a common test bed to further the research. While simple examples have shown that physics-informed machine learning techniques may be able to give the suspected accuracy improvements, this will be the first system to allow testing that hypothesis at the full scale and directly influence policy.

We do not suspect that the current methods will fully work on their first try when applied to the full scale model, but by giving our mathematicians and scientists the platform to assess the difficulties of the problem will greatly accelerate the process of discovery.

6. Narrative: Scientific and technical description of the opportunities and approach; activities that will advance the science; and specific field, laboratory, model, synthesis, and/or analysis examples

In the burgeoning field known as scientific machine learning, physics-informed learning, physics-guided machine learning, and more, such a project can establish climate modeling as a core challenge problem to focus the field. This will allow direct comparisons of accuracy and performance, similar to the popular MNIST dataset of deep learning, to turn proposed game changing methods into standard practice. Research in physics-informed neural networks from Brown, University of Pennsylvania, and University of Colorado, Boulder will be directly compared to the universal differential equation techniques of MIT to discover the best variants for accelerating high fidelity models of ocean dynamics. We plan to showcase our approach using the ocean component of a climate model that our group is developing as part of the

Climate Modelling Alliance, a multi-institutional collaboration between Caltech, MIT, the Naval Postgraduate School of Monterey, and the NASA Jet propulsion Laboratory. The advantage of starting with ocean models is that we do know the equations that govern ocean physics and the only limitation is that we do not have computers sufficiently powerful to resolve those equations in global models with mesh grids sufficiently fine to resolve all relevant motions. The situation is different in the atmosphere, where many aspects of the basic physics, like the microphysics of clouds, are not well understood. We will begin by studying convection in the upper ocean, a key process for the exchange of heat and carbon between the ocean and the atmosphere. (Ocean convection is a simpler analogue of atmospheric convection which sets precipitation patterns around the globe. The simplification arises from the lack of phase transition which are instead ubiquitous in air through for water vapor.) Large eddy simulations which are run at very high resolution, but only over limited ocean areas, will be the fertile ground for providing ground truth to train the models.

These techniques will allow the use of unprecedented data while having a physical backing to check the accuracy. These physics-informed learning techniques can simultaneously be used to learn for the large sources of incoming data, becoming a digital twin matching the Earth's true behavior possibly better than pure physical abstractions can provide. While these techniques for digital twinning have been used to perform precision cardiology and for the development of next generation spacecraft at NASA, this will be the first time a fully digital replica of a system as massive as the Earth's ocean systems will be accurately captured and controlled.

Our group has been building a fully functional climate model, the Climate Machine [3]. Critical to our approach is the interoperability of components and people [2]. Part of the interoperability story involves synergistic research on the high performance and edge computing problem, whereby productivity is improved when codes do not need to be rewritten for multiple architectures. The concept is code what you want to compute, but not how it is to be computed so it can scale and be adaptable to any computational architecture.

Our code will be more than available on GitHub. Most open source codes these days are available. The code that we produce will be composable with other software. Our goal [2] is to heal the planet together, not just throw a code out there. This means it should be easy for, say, researchers at universities, or at national labs, or new people to join and contribute.

References for the lay public

[1] Ferrari TEDxMIT Talk: Ocean, Climate, and Climate Change
[2] Edelman TEDxMIT Talk: A programming language to heal the planet together
[3] CSM Article January 22, 2021 Meet the team shaking up climate models