Automated Custom Calibration for E3SM

Authors

Benjamin M. Wagman (Atmospheric Science Dept, SNL), Kenny Chowdhary (Extreme Scale Data Science and Analytics Dept, SNL), Andy Salinger (Computational Science Dept, SNL)

Focal Area(s)

Automated custom calibration for the Energy Exascale Earth System Model (E3SM) applies primarily to focal area 2: Predictive modeling through the use of AI techniques, and includes elements of focal area 3: Insight gleaned from complex data (both observed and simulated).

Science Challenge

Earth System Models (ESM) contain uncertain parameters that cannot be uniquely determined by data or theory, and can affect model predictions. The current state-of-the-art is to precede each release of an ESM with an expert tuning, also known as calibration, that takes 6-12 months, requires a large supercomputer allocation, and is deterministic, i.e. a single choice for each uncertain parameter is set in the model code. A single deterministic tuning inevitably degrades some predictions and underestimates uncertainty for all predictions. **Those looking to make decisions on, or answer science questions with, climate predictions will best be served by custom, probabilistic calibrations.**

In recognition of the need for E3SM to make reliable, probabilistic predictions of the future climate for a wide variety of quantities of interests (QOI's), we seek to develop a toolset to enable custom calibration and probabilistic prediction in E3SM. The toolset would enable the user to either optimize or calibrate uncertain parameters in the E3SM model with respect to their own preferred metrics. A user's choice of metrics may reflect personal or broadly-accepted hypotheses of which aspects of model skill relative to observed climate will yield reliable model predictions of future climate for their own QOI.

Rationale

The nation needs Earth system modeling to evolve into a decision-support tool. Decision makers and policy makers across a huge spectrum of economic and national security realms need the best information possible to make sound investments and preparations on potential threats from a changing climate. This may be predictions about availability of arctic shipping lanes, cooling-water availability at power plants, droughts that lead to mass migration, and storm surges on vulnerable coasts. The immense complexity of Earth system processes leads to a large number of uncertain parameters in a global Earth system model like E3SM. We need a model calibration strategy that (1) recognizes that the set of uncertain model parameters that yields the best predictions for droughts may not be the same as the set

DOI: https://doi.org/10.2172/1769677
DOE BER, AI4ESP 2021 #AI4ESP1140 (https://www.ai4esp.org)

that best predicts availability of arctic shipping lanes and (2) the uncertainty in the calibration process is accounted for in the uncertainty of the predictions.

To best position E3SM as a decision-support tool for the wide variety of potential climate impact questions that may be asked, two big leaps are needed.

1. The model tunings need to be *customized* for the science, economic, or national security question being asked. Different model tunings can excel at answering different questions.
2. The model tunings must be *probabilistic*, such as a Bayesian approach, so that climate predictions are fully integrated with uncertainties from the calibration to the observations in addition to intrinsic variability.

To reach the goals of *customized and probabilistic* tunings, for a usable decision support tool, there are major improvements needed in the calibration process. The tuning process needs to be largely automatic, affordable, and robust, so it can be customized for individual analyses. A probabilistic approach adds whole new dimensions of complexity around the tuning process.

To reach the highly efficient and reliable approaches needed to achieve this vision, there are numerous research and development gaps that would need to be overcome. The research gaps that can potentially be spanned by ML techniques will be to create a highly efficient and reliable approach to (1) generating model surrogates and (2) performing the parameter optimization.

Narrative

Calibration is a complex, interwoven collection of machine learning (ML) and AI tools for which there has been an inconsistent and disparate development of technologies. Our approach is to provide a single, robust customizable tool for E3SM that brings together the entire spectrum of calibration techniques, including, but not limited to, AI feature selection, construction of ML surrogates, linear and non-linear dimension reduction techniques, and deterministic and Bayesian parameter optimization. By bringing these disparate techniques together into a single consistent framework, we can select the best models for calibration across a range of state-of-the-art and classical ML methods, aiding the decision making process of tuning and selecting the best model for the QoI. For example, two popular methods for constructing ML surrogates, which may be used for both feature selection and/or parameter optimization, are polynomial chaos expansions (PCEs) and neural networks (NNs). Currently, there is no consistent framework or library that allows for a direct comparison of these two methods when applied to E3SM tuning. Our automated decision tool aims to make this possible so that an end-user can run multiple models and choose the best one for the job. A preliminary prototype will be built in Python using the scikit-learn API which will allow seamless integration with one of the most ubiquitous and popular ML libraries. We envision that such a tool could be added to the E3SM package for the purposes of augmenting and automating calibration for E3SM diagnostics. Details of the development and implementation of the tool are discussed below.

Automated Custom Calibration for E3SM

A typical calibration workflow is as follows: (1) Running ensembles of simulations, e.g., gathering training data, to discover the relation between the feature space and the output quantity of interest, (2) using feature selection and dimensionality reduction techniques to determine which input and output parameters are most important, (3) building a surrogate or emulator in lieu of the prohibitively expensive model using both (1) and (2) to improve its accuracy, and (4) finding the optimal setting, or a probability distribution of, parameters based on the discrepancy between the model prediction over the historical period containing the observed data. These four areas will be the focus of the development of our calibration tool, with a particular focus on the surrogate model construction and the parameter optimization using Bayesian inference methods. In order to add robustness to the calibration tools, we will employ multiple modalities and technologies in each step. This includes modelling structural error and bias in the construction of surrogates and incorporating Bayesian techniques to estimate and propagate uncertainty. To further the automation process, we will include hyper-parameter tuning and model selection techniques, such as cross-validation, so that the end user can more easily select the best model or technique for each step. While E3SM's upcoming ability to run efficiently on GPUs will open the door to running large ensembles very efficiently on DOE exascale machines, the management of large ensembles is not a trivial task. To automate this process, we will employ the Dakota toolkit with an interface to E3SM to organize and streamline the simulations.

One additional challenge comes from recognizing that the utility of using automated custom calibration for improving Earth system predictions rests on the assumption that well-tuned models make better predictions than poorly-tuned models. Although this assumption is intuitive, it is not necessarily valid because long-term predictions are sensitive to climate feedbacks, which can be hard to constrain in observed datasets. The small but growing set of observable metrics that are believed to constrain feedbacks and therefore constrain climate prediction are known in the literature as "emergent constraints" (for examples, see Klein et al., 2015). We envision that the custom calibration capability would eventually include the option of selecting from a dictionary of emergent constraints relevant to their prediction.

Suggested Partners/Experts (Optional)

Charles Jackson (The University of Texas at Austin), Khachik Sargsyan (SNL),
Chris Golaz (LLNL), Peter Caldwell (LLNL)

References (Optional)

Klein, SA, and A Hall. 2015. "Emergent constraints for cloud feedbacks." Current Climate Change Reports 1 (4): 276–287.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND2021-1655 O