SAND2011-5673C

Uncertainty Quantification in Climate Modeling

Khachik Sargsyan, Cosmin Safta, Robert Berry, Bert Debusschere, Habib Najm
Sandia National Laboratories, Livermore, CA

August 3, 2011

We address challenges that sensitivity analysis and uncertainty quantification methods face when dealing with complex computational models. In particular, climate models are computationally expensive and typically depend on a large number of input parameters. We consider the Community Land Model (CLM), which consists of a nested computational grid hierarchy designed to represent the spatial heterogeneity of the land surface. Each computational cell can be composed of multiple land types, and each land type can incorporate one or more sub-models describing the spatial and depth variability. Even for simulations at a regional scale, the computational cost of a single run is quite high and the number of parameters that control the model behavior is very large. Therefore, the parameter sensitivity analysis and uncertainty propagation face significant difficulties for climate models. This work employs several algorithmic avenues to address some of the challenges encountered by classical uncertainty quantification methodologies when dealing with expensive computational models, specifically focusing on the CLM as a primary application.

First of all, since the available climate model predictions are extremely sparse due to the high computational cost of model runs, we adopt a Bayesian framework that effectively incorporates this lack-of-knowledge as a source of uncertainty, and produces robust predictions with quantified uncertainty even if the model runs are extremely sparse. In particular, we infer Polynomial Chaos spectral expansions that effectively encode the uncertain input-output relationship and allow efficient propagation of all sources of input uncertainties to outputs of interest.

Secondly, the predictability analysis of climate models strongly suffers from the curse of dimensionality, i.e. the large number of input parameters. While single-parameter perturbation studies can be efficiently performed in a parallel fashion, the multivariate uncertainty analysis requires a large number of training runs, as well as an output parameterization with respect to a fast-growing spectral basis set. To alleviate this issue, we adopt the Bayesian view of compressive sensing, well-known in the image recognition community. The technique efficiently finds a sparse representation of the model output with respect to a large number of input variables, effectively obtaining a reduced order surrogate model for the input-output relationship. The methodology is preceded by a sampling strategy that takes into account input parameter constraints by an initial mapping of the constrained domain to a hypercube via the Rosenblatt transformation, which preserves probabilities. Furthermore, a sparse quadrature sampling, specifically tailored for the reduced basis, is employed in the unconstrained domain to obtain accurate representations.

The work is supported by the U.S. Department of Energy's CSSEF (Climate Science for a Sustainable Energy Future) program. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.