Geophysical Research Abstracts
Vol. 20, EGU2018-5555, 2018
EGU General Assembly 2018

SPUX: Scalable high performance uncertainty quantification framework for stochastic models in environmental data sciences

Jonas Šukys (1), Mira Kattwinkel (2), and Peter Reichert (1)

(1) Eawag, SIAM, Dübendorf, Canton of Zürich, Switzerland, (2) University of Koblenz-Landau, Germany

In environmental science stochastic models, e.g. individual based models (IBMs) and stochastic differential equations (SDEs), are becoming increasingly successful in simulations of complex dynamical systems, where intrinsic epistemic uncertainties are present. Examples include numerous research fields such as ecology, hydrological catchment modeling, subsurface ground water flows and urban flood risk assessment. Scarce raw experimental data is complemented by wide availability of extensive expert knowledge on processes and plausible model parameter values. However, such prior knowledge is often not treated systematically within environmental modeling efforts.

Emerging uncertainty quantification techniques will significantly advance scientific understanding of such complex systems and will enhance the credibility of predictive model capabilities. In particular, Bayesian inference can be used for both parameter estimation incorporating prior expert knowledge and uncertainty quantification. Despite obvious methodological advantages, its successful application to realistic scenarios has been hindered by the complex stochastic nature of IBMs and SDEs, leading to technical difficulties in likelihood estimation and entailing high computational costs. Likelihood estimation issues have been recently tackled by means of computationally expensive techniques such as Approximate Bayesian Computation (ABC) and Particle Markov Chain Monte Carlo (PMCMC), the latter capable of exploiting hidden-Markov model structures. In the PMCMC approach, Particle Filter (PF) is employed for efficient marginal likelihood approximations using time-series observations, where multiple model simulations (particles) are required, with periodic ensemble adaptation steps.

Additionally, due to the large number of MCMC samples required and the hundreds or thousands of independent particles (model simulations) within each of them as well, the resulting total computational effort in PMCMC inference becomes prohibitively large for computationally demanding numerical models. High performance computing (HPC) is hence paramount in addressing this issue, especially since alternative approaches such as stochastic emulation, despite their success in some domains, have a restricted range of applicability, introduce additional statistical errors and increase model approximation bias. Unstructured particle ensemble adaptation stages within PMCMC methodology, complemented by the need for modern parallel distributed memory computing clusters, increase the algorithmic complexity of numerical Bayesian inference techniques and pose a great challenge for efficient concurrent high performance implementation. Our recently developed Python-based prototype framework SPUX aims to accelerate PMCMC by using parallel clusters and significantly reducing the runtime required. SPUX mitigates high computational costs by distributing particles (model evaluations) over multiple computational units in a parallel compute cluster. Such efficient use of computational resources makes the SPUX framework accessible for domain scientists interested in Bayesian inference for complex stochastic environmental models.