Statistica Sinica: Preprint
doi:10.5705/ss.202016.0403

Inference on large-scale models is of great interest in modern science.
Examples include deterministic simulators of fluid dynamics to recover the source of a pollutant, or stochastic agent-based simulators to infer features of consumer behaviour. When computational constraints prohibit model evaluation at all but a small ensemble of parameter settings, exact inference becomes infeasible. In such cases, emulation of the simulator enables the interrogation of a surrogate model at arbitrary parameter values. Combining emulators with observational data to estimate parameters and predict a real-world process is known as computer model calibration. The choice of the emulator model is a critical aspect of calibration. Existing approaches treat the mathematical model as implemented on computer as an unknown but deterministic response surface. However, in many cases the underlying mathematical model, or the simulator approximating the mathematical model, are not determinsitic and in fact have some uncertainty associated with their output. In this paper, we propose a Bayesian statistical calibration model for stochastic simulators. The approach is motivated by two applied problems: a deterministic mathematical model of intra-cellular signalling whose implementation on computer nonetheless has discretization uncertainty, and a stochastic model of river water temperature commonly used in hydrology. We show the proposed approach is able to map the uncertainties of such non-deterministic simulators through to the resulting inference while retaining computational feasibility. Supplementary computer code and datasets are provided online.