

11. Adaptive model selection, validation, and uncertainty quantification in complex multiscale systems (UT-Austin)

Research thrusts: *Validation, adaptation, and management of models*

Research sub-thrusts: *Goal-driven adaptive model reduction; Goal-driven model adaptivity; Bayesian coarse graining; Model selection; Inadequacy modeling*

This work addresses the development of theory, algorithms, and applications for modeling *complex multiscale systems*, involving model selection, model calibration, model validation, and uncertainty quantification [45, 46, 47, 78, 101, 120]. We have addressed the age-old problem not dealt with often with mathematical rigor: the problem of producing coarse-grained models of atomic systems. The question is: exactly how does one construct coarse-grained (molecular) models of very large atomic systems so as to predict quantities of interest (Qols) within preset levels of accuracy? Here, the basic problem of model selection is first encountered: how does one aggregate atoms into molecules or superatoms to obtain a reduced model capable of representing with sufficient accuracy the key Qol in simulations? Once a family of possible coarse-grained models is identified, how are the most plausible models selected, given relevant observational data? Beyond this extremely important issue, how are uncertainties in the data managed and what processes and principles are involved to determine the validity of the models under consideration? These questions are further complicated by the presence of uncertainties that manifest themselves at different scales in multiscale models of materials. We have used the so-called “All Atom” model furnished by hardened molecular dynamics (MD) codes (LAMMPS, in this case) as the “truth” and developed tools to analyze such uncertainty issues using the MD calculations as surrogates to reality.

We introduced a new algorithm we call OPAL (Occam Plausibility Algorithm) that provides a systematic approach for selection and validation of models of very complex systems. The reference to “Occam” in this process refers, of course, to Occam’s Razor, the classical idea that only the simplest theory (or model in the present context) be used in making a prediction. The algorithm begins with the assumption that a (possibly large) set M of possible parametric models $P_i(\theta_i)$ can be identified, each with unknown and possibly random parameters θ_i . A sensitivity analysis is performed in a unit-level calibration scenario to determine which model parameters affect or do not affect the target Qol in a predictions scenario, involving the full system. The models that do not influence appreciably the Qol are eliminated. Those with the fewest parameters are put in a Category 1; those with the next highest number of parameters are put in the next category, and so on, so that the “simplicity” in the spirit of Occam’s Razor is related specifically to the number of parameters in each model. Given (possibly uncertain) calibration data, a Bayesian approach is used to compute the posterior model plausibility of each surviving model, and the most plausible model (or models) is (are) retained. These are then subjected to validation tests to determine which, if any, satisfy preset validation criteria. The parameter distributions of models deemed valid in a series of validation scenarios are used in the forward problem, and then solved for the Qols. The process thus addresses data uncertainty, parameter sensitivity, model inadequacy, model selection, and uncertainty quantification. Multiscale applications of OPAL involve using validated solutions at one scale to inform models of larger scale. Applications to date have involved analysis of complex molecular models of polyethylene.