MODEL CREDIBILITY PLAN

To ensure model credibility, all simulations of multiscale models will be benchmarked against limiting cases where analytic or previous simulation results exist. When possible, we will use machine learning (e.g. evolutional algorithms) and/or Bayesian approaches (e.g. Approximate Bayesian Computation) to estimate model parameters from experimental data. In these cases, we will start with synthetic data generated from the model to determine which model parameters are learnable from the data and what additional experimental measurements are needed to constrain model parameters. We also will add "noise" to the synthetic data to investigate how robust our methods are to fluctuations from internal and experimental sources.

An important feature of machine-learning and Bayesian approaches for parameter estimation is they do not return a single "optimal" parameter set, but generate a distribution for the model parameters. These distributions provide a measure of how sensitive the model behavior is to the individual parameters and also allow dependencies between the parameters to be determined. To further quantify uncertainty and parameter sensitivity, we will perform standard single parameter sensitivity analysis and make use of Fisher information metric. We understand that as our models become more complex, these systematic approaches for parameter estimation and sensitivity analysis will not be computationally feasible. However, by using a bottom-up approach in which sub-components of the models are individual validated, we hope to ensure model credibility.

Other important aspects of model credibility are proper documentation and making software freely available for external use. Major software newly developed or substantially modified/extended within this project will be released under a permissive open-source license. These software packages will be:

- Freely available to researchers and educators in the non-profit sector, such as institutions of education, research institutions, and government laboratories.
- In the event that the original investigators are unable or unwilling to continue developing the software, copyright will be assigned to third-party investigators who are committed to taking responsibility for the continued development of the software.
- The terms of availability will permit researchers the ability to modify the source code, and to share such modifications freely with colleagues.

To the extent possible, software development will use open platforms such as GitHub that fosters broader community involvement in the software development and distribution process. Open mailing lists using Google Groups will be established.

Mathematical models developed in the project will be described by a combination of standardized binary file formats (when available within a particular domain) or well-documented internal file formats. When widely used standards are not available, we shall use human-readable text files when feasible. Models will be freely available for research use following initial publication. The results of time-intensive numerical simulations used for publication will be stored for at least three years beyond the award period through UNC ITS Research Computing's mass storage, and will be made available upon request.