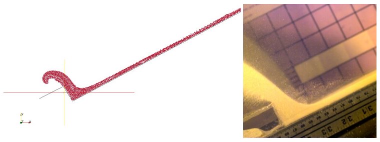
Efficient Computational Modeling and UQ with Reusable Tools

Abani Patra, P. Bauman and V. Chandola
University at Buffalo, Buffalo, NY 14260
{abani,pbauman,vchandola}@buffalo.edu

Topics: D-G crosscutting

As evidenced by the broad literature in the sciences and several major workshop/panel reports (from the PITAC report [1] on computational science to the many fusion science specific reports) we are making great progress to "*well founded extrapolation*" using "*verified and validated models and data"* of unprecedented quality leading to quantitative **predictive science**. A central prerequisite for a prediction capability for plasma physics processes is the development of a systematic and reliable workflow for the incorporation of uncertainty into all stages of the prediction process. This includes advanced algorithms for the solution of complex mathematical models of the systems of interest, solution methodologies for inverse problems that leverage high-order information from the complex forward model, including gradient and Hessian information, leveraging that high order information for dimensional reduction, and a robust software infrastructure seamlessly unifying all stages of analysis and exploiting the hardware.

Central to all stages of the UQ workflow is the mathematical model of the physical process of interest. This includes not only sampling for the forward UQ problem, but also algorithms used in the solution of inverse problems. Advanced algorithms for the forward problem are vital to achieving both high-performance and accuracy. Such algorithms include goal-oriented error estimation, adaptive and higher order mesh refinement, geometric multigrid on unstructured discretizations and *advanced particle-based discretizations like the Godunov-SPH methods* we have been developing, adaptive modeling, and adjoint-based gradient and Hessian computations. The Godunov-SPH method eliminates the need to use numerical viscosity and at the same time allows larger time--steps when compared with classical SPH. These methods have to be harnessed to effective tools like the libmesh toolkit that take advantage of extreme scale computing and data analysis capabilities coming online at the exascale and beyond on the complex new architectures. Complexity of architectures requires that application systems effectively make the best use of the dynamic status of the machine. In recent work we have collected performance and resource (network, CPU, power, I/O, cache etc.) data on all jobs running on the Stampede computer at Texas Advanced Computing Center over the last two years. This data is now being mined to reveal issues in resource contention due to application or interaction among applications and system. In the next generation simulations the application will have to dynamically monitor such resource usage and adapt to making the best use of what is available. Resilience, to machine performance and component failure will also be a necessity.



Godunov SPH simulation of step flow of granular materials

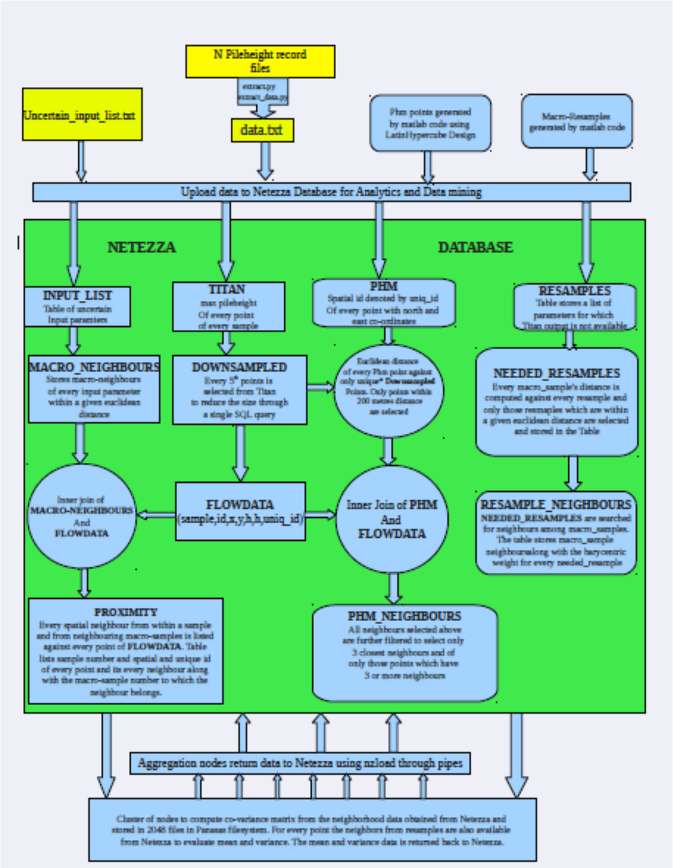
Vital to any UQ endeavor is the characterization of uncertainty in the parameters of the mathematical model. This is accomplished using Bayesian inference whereby prior information on the material parameters is supplied as a probability density function; the prior is updated using the likelihood, comparing the model and any available experimental data, yielding the sought posterior distribution. The high-dimensional nature of the parameter space necessitates sampling based algorithms, such as MCMC. Nevertheless, these sampling approaches can be accelerated using available gradient and Hessian information, if it is available. Additionally, this information can be used to construct surrogate models of the forward model. These surrogate models encapsulate the parameter-to-data map, but are computationally more efficient to evaluate. Even with surrogate models, very high-dimensional problems pose many challenges. A key aspect, then, is to reduce the dimension of the problem to those that are actually informed by the available data. These dimensional reduction strategies, again, can take advantage of available

Efficient Computational Modeling and UQ with Reusable Tools

Abani Patra, P. Bauman and V. Chandola
University at Buffalo, Buffalo, NY 14260
{abani,pbauman,vchandola}@buffalo.edu

gradient and Hessian information to facilitate more rapid and accurate dimensional reductions and facilitate surrogate model developments on the reduced space. A key aspect that we have recently started investigating is *the confounding of the numerical error and uncertainty*. In a recent calculation of catastrophe from a surrogate based UQ using an AMR based shallow water type code – computed probability of catastrophe shifted from 0.6 to 0.49 when the numerical error was reduced from a normally acceptable rate to machine precision.

All of the above techniques require the instantiation of scientific software and workflows. A central component of the research programs being developed at UB is the deployment of freely available, open-source software to achieve all of the above goals for complex, multiphysics PDEs, such as those arising in plasma physics. Advanced software infrastructure has been developed to rapidly deploy modern finite element methods on complex multiphysics problems including goal-oriented error estimation and AMR, adjoint-based computation of gradient and Hessian information, all on modern parallel computing platforms. This infrastructure has been exploited to interface with solvers for inverse problems, including those based on Bayesian formulations, as well as forward UQ packages such as DAKOTA. The UQ workflows with the need to build complex models and process vast amounts of data are non-trivial. Alongside we present a very complex workflow that uses a high-end supercomputer for simulation and analytics processing and uses a special purpose data warehouse appliance (IBM/Netezza) to process the very large data generated. Mastering such workflows will need leveraging both specialized hardware and careful attention to software design.



Research Challenges

Many research challenges are thus readily identified. We summarize major ones.

- High Dimensionality
- Computational Methodology to integrate extreme scale computing, dynamic machine information
- New modeling and discretization methodologies
- Dealing with very large volumes of data

Efficient Computational Modeling and UQ with Reusable Tools

Abani Patra, P. Bauman and V. Chandola
University at Buffalo, Buffalo, NY 14260
{abani,pbauman,vchandola}@buffalo.edu

References

[1] www.nitrd.gov/**pitac**/reports/20050609_**computational**/**computational**.pdf

[2]http://extremecomputing.labworks.org/nationalsecurity/whitepapers/UncertaintyQuantificationandErrorAnalysis.pdf

[3] http://science.energy.gov/%7E/media/ascr/pdf/program-documents/docs/Crosscutting_grand_challenges.pdf