Development of Predictive Models through Coupled Application of Simulations and Uncertainty Quantification on Exascale Systems

Ramanan Sankaran (sankaranr@ornl.gov)
Oak Ridge National Laboratory, Oak Ridge Leadership Computing Facility, Oak Ridge, TN

Joseph C. Oefelein (oefelei@sandia.gov)
Sandia National Laboratories, Combustion Research Facility, Livermore CA

Aggressive national goals for reducing petroleum use by 25 percent by 2020 and greenhouse gas emissions by 80 percent by 2050 will require major improvements in all aspects of our nation's energy use. At the same time, the U.S. transportation and energy sectors are under tremendous pressure from international competitors and challenging economic conditions. Achieving reduced fuel usage and emission goals will require significantly shortened product development cycles for cleaner, more efficient engine technologies. Concurrently, fuels will also be evolving, creating additional complexity and further underscoring the need for efficient product development cycles. Under the current cut-and-try approach, design cycles simply take too long. These challenges present a unique opportunity for U.S. leadership in supercomputing to develop predictive simulation tools for combustion. Simulations that reliably predict engine efficiency and pollutant emissions using both conventional and new fuels require higher fidelity modeling than is possible with current computing resources. Exascale platforms will facilitate model development and validation, while providing the necessary speedups in the time to solution and fidelity of numerical data required to make revolutionary advances in the design of combustion systems.

While substantial progress has been made to understand basic principles of turbulent combustion through advances in experimental and computational capabilities, to-date this understanding is still limited due to the inherent highly nonlinear multiscale/multiphysics nature of the phenomena. Direct Numerical Simulation (DNS) can only be applied for moderate turbulence levels and small combustion volumes. Thus, some form of modeling will always be necessary to access device-relevant conditions and geometry, even with exascale computing. Recognizing the limitations for application of DNS, Large-Eddy-Simulation (LES) approaches have been advanced for both science and engineering and have significant potential for design. Specific research priorities for development of predictive models have been defined in recent workshops such as the SC-BES sponsored "Workshop on Clean and Efficient Combustion of 21st Century Transportation Fuels," and the jointly sponsored SC-BES and EERE-VT "Workshop to Identify Research Needs and Impacts in Predictive Simulations for Internal Combustion Engines."

One of the key objectives for LES as an approach for model-based design is to make predictions of complex multiphysics/multiscale processes with quantified uncertainty. The scale of the problems, both in terms of the size of individual jobs, and the need for running an ensemble of jobs over wide parameter spaces, requires exascale level computer resources along with corresponding algorithmic developments and demonstrations. This applies to both the solvers used for LES, and the UQ tools developed to quantify the accuracy of the simulations. The accuracy of the required simulations is complicated by the interdependence between different models, numerical implementation, and competition between model and numerical errors. Many uncertainties exist in addition to model accuracy such as error-prone numerical methods, poor grid quality, lack of appropriate spatial or temporal resolution, and ill-posed boundary or initial conditions. In addition, rare, but potentially catastrophic, events can occur.

One such example of a catastrophic event is flame flashback in a stationary gas turbine combustor, which can damage the sensitive fuel injection equipment upstream. Predictive modeling of such occurrences requires, not only an accurate model for the response of the flame speed to varying mixture and flow conditions, but also adequate statistical confidence in the tail regions of the probability distributions. An ensemble of simulations would provide the statistical descriptions of these highly intermittent phenomena, along with the ability to quantify the associated uncertainties.

An important aspect of performing an ensemble of simulations to explore the parameter space is deciding the problems that need to be simulated based on their potential to improve our knowledge and reduce the uncertainty. We need to make decisions on the next problem to be simulated that would provide the most useful information based on what was learned from the simulations performed till that point, while also taking into account the uncertainty associated with our choice of analysis parameters. The ranking and selection process needs to be automated if these experiments are to be carried out on batch scheduled HPC user facilities. It is necessary to develop an automated and optimal learning framework to rank and select the simulation experiments to create the most valuable ensemble of simulation dataset. Optimal learning techniques are being used in the design and selection of experiments in fields such as systems biology and drug discovery. An optimal learning strategy would start with an initial belief on the output characteristics of the experiments and their dependence on the parameters being studied. As additional experiments are performed, the results are used to update the initial beliefs. Further experiments are selected based on their potential to improve the precision of the beliefs and deliver the necessary statistical confidence in the desirable portion of the operation regime, while also satisfying the criteria such as covering a wide range of parameter conditions. An online learning methodology that can be used for ranking and selection of the next experiment is essential to be able to efficiently create a dataset composed of a large ensemble of simulations.

It is also important for the selected set of simulation experiments to balance both an exploitative and explorative strategy. An exploitative strategy would be where a particular experiment is found to have a finite probability for a certain phenomena of interest. In that case, an exploitative strategy would be to continue to advance that experiment in time much longer or to perform additional experiments to obtain multiple statistical realizations at the same condition. An explorative strategy would try to identify a new set of input parameters that is likely to produce statistics at the desired conditions at a much higher rate, but will also come with the risk that it might place the simulation in a different regime. It is anticipated that a combination of exploration and exploitation would be necessary to improve the information content in an ensemble dataset as opposed to a single strategy. The current simulation practices are usually limited to a pure strategy from one of these two, unless the strategy is guided manually during the setup of the simulation experiments. An automated method for designing the simulations based on a chosen balance between these two strategies would be essential for performing automated ensemble simulations on exascale systems.