

# Personal Statement

Manav Vohra

The NSF-funded CyberTraining program at UMBC seems like an excellent opportunity to gain credible research experience in the field of data science, high performance computing (HPC), and a big data application. My work has predominantly focused on developing and implementing efficient computational methods for model uncertainty quantification for a variety of scientific applications such as nanocomposites, porous media transport, and more recently; chemical kinetics, molecular dynamics, and additive manufacturing. The methods are aimed at optimal design of experiments for model calibration, validation, and parameter dimension reduction for quantifying predictive uncertainty in simulations in an efficient manner. Although the methods have been shown to be remarkably useful, they can become computationally prohibitive in the presence of big data especially in the absence of an underlying physics-based model. Moreover, they are limited in terms of scalability to be able to exploit modern HPC architectures. The knowledge of Machine Learning techniques for big data analysis using advanced HPC systems through participation in the CyberTraining program would help me enhance the applicability of existing methods aimed at model uncertainty quantification. Additionally, it would help guide the development of novel, more advanced data driven approaches for large-scale scientific computing applications where a reliable model formulation may not be available. I am also excited about the prospects of collaborating with researchers from diverse backgrounds and gaining new perspectives to problem-solving.

I have gained intensive experience in performing numerical simulations using Fortran and Python. Additionally, I have developed tools in Matlab for sensitivity analysis, surrogate modelling, and parameter estimation using Bayesian inference. Furthermore, I have contributed to the development of QUESO, a C++ software package for Bayesian calibration [1]. I have also participated in training sessions on MPI and OpenMP using C, organized by the Texas Advanced Computing Center (TACC) as a postdoctoral fellow at UT Austin. While the training sessions were insightful, I am truly motivated to gain practical experience in performing parallel computations using C and Python for a pertinent real-world application involving big data, during the program.

As mentioned earlier, my research thus far has mainly focused on physics-based model uncertainty quantification. Through this unique opportunity, I hope to gain relevant experience in applying Machine Learning techniques for big data applications using parallel computations. This would help me broaden the scope of my work significantly. Specifically, I expect to develop advanced methods for real-time decision making using computer simulations, in the presence of uncertainty for a broad range of scientific and engineering applications. These methods would combine heterogeneous sources of information based on observed data, model predictions, errors, and expert knowledge to enable reliable decision making in an efficient manner enabled by high performance computing. Moreover, through this program, I aim to gain valuable knowledge and insights pertaining to the impact of greenhouse emissions on climate change and global warming. I firmly believe that my research background in uncertainty quantification and scientific computing can help contribute immensely in this area of research. Specifically, the satellite data used for calibrating the climate models could be noisy, sparse, and erroneous due to faulty sensors. A Bayesian approach to calibration provides a systematic probabilistic framework to account for measurement error, model error, and prior knowledge. Sensitivity analysis could help identify key contributors to the variability in the model predictions, and dimension reduction techniques could be applied to reduce computational effort associated with forward propagation of the uncertainty from model parameters to the quantity of interest as well as Bayesian calibration.

I would be keen to assimilate ideas discussed in the modules with my background to develop innovative approaches aimed at addressing the technical challenges in the project work. In ad-

dition, I would seek inputs from the team members and the faculty advisor on ways to improve these approaches and integrate them into an overall strategy proposed by the team. I would also endeavour to provide any possible assistance or guidance to my team mates as needed. Lastly, I also consider this program an opportunity to foster a long-term research collaboration with the faculty mentors at UMBC to pursue our shared interests in this field of research.

## References

- [1] K.C. Estacio-Hiroms, E.E. Prudencio, N.P. Malaya, M. Vohra, and D. McDougall. The QUESO library. *Users Manual. arXiv preprint*, 2016.