**Workshop on Knowledge Extraction via Comparison
of Complex Computational Models to Massive Data Sets**

July 29-31, 2013.

1   Background

The was a joint workshop with the NSF funded MADAI collaboration (Models and Data Analysis Initiative, http:// madai.us) which is devoted to applying and developing new techniques for the statistical analysis of massively complex models and the application of cutting edge visualization tools to drive data exploration. The workshop proposal to SAMSI resulted from the involvement of two leaders of working groups in the Massive Datasets program, Steffen Bass (Duke, Physics) and Robert Wolpert (Duke, Stats).

1.1   Purpose and objectives

The objective of the workshop was to introduce a broader base of domain scientists in Heavy-Ion Physics, Cosmology and Climate Sciences to statistical and visualization tools that facilitate knowledge extraction via complex model to data comparisons. The workshop was also meant to provide opportunities for the Statistical Science community to learn about recent developments in complex modeling and computer experiments as well as engage in new collaborative ventures. Hands-on tutorials were used to showcase a modular visualization platform (based on Paraview) that allows for advanced visualization of complex model dynamics as well as statistical analysis tools while statistical tools were based on Gaussian process surrogate models for rapid exploration of a model's parameter space.

1.2   Organizers and key personnel

Organizers: S.A. Bass (Physics, Duke University, S. Pratt (Physics and Astronomy, Michigan State Univ), R.M. Taylor II(Computer Science, UNC-Chapel Hill) and R. Wolpert (Statistical Science, Duke University).
SAMSI Directorate Liaison: Snehalata Huzurbazar (Stats, NCSU/U of WY)

2   Workshop

The workshop (see schedule at http://www.samsi.info/workshop/2013-knowledge-extraction-comparison-complex-computational-models-massive-data-sets-july-29) lasted for two and half days. The first day consisted of talks by domain scientists on various research problems that require methods and visualization as envisioned by MADAI. The scientists ranged

from physicists to climate scientists and 6 talks were presented. The second day, the morning session covered 4 talks on methodology ranging from computer experiments and models to topological data analysis. The afternoon session covered scientific visualization methods. The participants spent the last day working on tutorials, the MADAI visualization tutorial in the morning and the MADAI stats tutorial in the afternoon.

3   Participants and Diversity

Of the 15 presenters, 2 were women and 1 was Hispanic. Of the total of 40 participants, 21 were new researchers or graduate students, and 9 were women. The graduate students and new researchers greatly appreciated the tutorials, especially the visualization tutorials.