DE-FG02-10ER2602

DOE Award Number DE-SC0005176

PRINCETON UNIVERSITY, Princeton, NJ

Prof. Ioannis G. Kevrekidis, PI

(sister grants: Prof. N. Zabaras, Cornell University; Dr. Guang Lin, PNNL)

Modeling and Simulation of High Dimensional Stochastic Multiscale PDE Systems at the Exascale

FINAL REPORT

for Dr. Karen Pao

attn. Ms. Teresa Beachley

Office of Advanced Scientific Computing Research

Office of Science

US DOE

PROGRESS REPORT

The thrust of the proposal was to exploit modern data-mining tools in a way that will create a systematic, computer-assisted approach to the representation of random media -- and also to the representation of the solutions of an array of important physicochemical processes that take place in/on such media. A parsimonious representation/parametrization of the random media links directly (via uncertainty quantification tools) to good sampling of the distribution of random media realizations. It also links directly to modern multiscale computational algorithms (like the equation-free approach that has been developed in our group) and plays a crucial role in accelerating the scientific computation of solutions of nonlinear PDE models (deterministic or stochastic) in such media – both solutions in particular realizations of the random media, and estimation of the statistics of the solutions over multiple realizations (e.g. expectations).

A particular and important example was to study the dynamics of complex networks (and especially heterogeneous networks – heterogeneous BOTH in the nature of the nodes and in the connectivity between the nodes). This is because networks themselves can be thought of as a realization of a random medium, and the randomness of the medium become the heterogeneity of the network (both in the nature of the nodes and the degree of connectivity between the nodes).

The focus of our work has been consistently on the development of computational tools associated with data mining, and, in particular, diffusion maps. What we are most proud of (see Figure 1) is our first success in being able to link diffusion maps with being able to do data-mining when the data come in the form of graphs; this is a paper that we are now writing with Mr. Karthik Rajendran, a fourth year graduate student in the Department of Chemical and Biological Engineering, who was also honored for his work by Princeton University with the Ray Grimm memorial Prize for computational science.

The PI just completed several visits to PNNL, to work with the co-PI, Dr. Lin, and his postdocs on two subjects: the statistics of model reduction (and, in particular, on exploring ways to test the hypothesis that high dimensional data may be reducible), and also on linking our work with several PNNL groups working on data in the form of graphs (cyber-security, energy smart grid applications) to explore “technology transfer” of our algorithms for graph data reduction.

Ms. Carmeline Dsilva, the DOE-CSGF graduate student that was a member of our group and worked on the research supported by this proposal visited Argonne National Laboratory, doing her summer fellowship under the supervision of Dr Mihai Anitescu. There have been three items that Carmeline has been worked on – the first and most important one scientifically is the factoring out of symmetries before the data mining/data reduction stage (or, how to do diffusion maps taking symmetries into account). We had a paper by Ben Sonday had a first manuscript on this nd Carmeline had two important papers on multiscale data mining and on the parsimonious representation of complex dynamical systems (in problems of interacting particles, e.g.

chemotaxis). She also worked on another important problem: the ability to "map" different observations of a dynamical system to each other, so that we can **fuse** data between different realizations/measurements of the same dynamical system. We also worked on molecular dynamics and on agent-based models on complex dynamic problems. Along with L. Nedialkova, Carmeline worked on an interesting project: we were invited for a presentation in collaboration with another group (that of Prof. M. Grover at Georgia Tech) – the subject was a functional comparison of two different reduction techniques: diffusion maps, and local feature analysis (LFA) on both some textbook and some molecular dynamics data. There is a presentation and a paper from this work that is almost accepted (under revision, 5 below). We also worked (in collaboration with G. Hummer, at NIDDK/NIH) on the use of diffusion maps and data clustering techniques to get useful representations (and reduced Markov models) for macromolecular MD data (simulations of alanine pentapeptide) – this has resulted in a paper in J. Chem. Phys. We also explored the effect of using different pairwise distances –more intelligent metrics-). Issues of anomaly detection in exploring glassy dynamics in MD simulations were are also explored – anomaly detection is a natural extension of the nonlinear data mining techniques we study.

We have also worked on the reduction of stochastic simulations with a group of collaborators from England (see publication 3 and Figure 2 below) – using brief bursts of constrained SSA simulation to solve effective Fokker-Planck equations for the coarse-grained variables in chemical simulations. The publication works with a 2D coarse-grained example, and we are continuing work on a 3D coarse-grained mechanism.

In our overall study of coarse-graining data that come in the form of graphs, we have several projects that involve both dynamic simulations as well as illustrations of nontrivial dynamics for graph evolution problems (4,9,11). Most importantly, we have worked on creating a link between uncertainty quantification and heterogeneous network descriptions (10) – we use tools developed in the UQ context –both PC methods as well as sparse collocation grids) to develop effective models for entire populations of heterogeneous oscillators. We are currently, with Mr. Rajendran, exploring the use of these techniques when the heterogeneity is not an intrinsic parameter, but rather has to do with the local network structure.

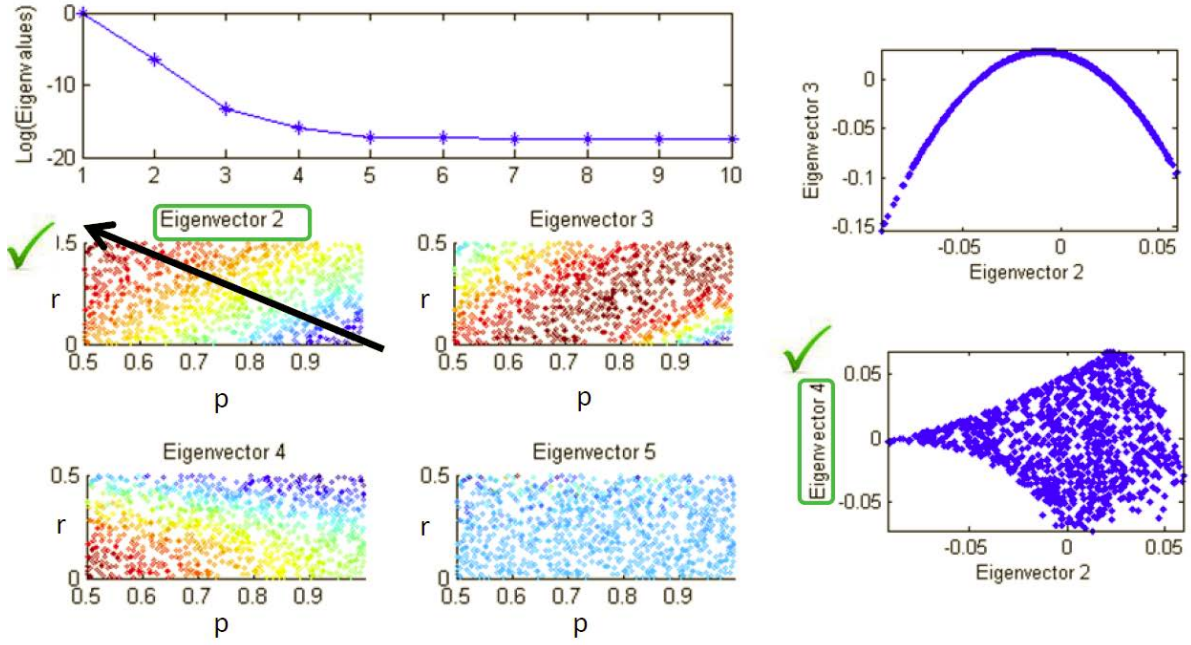
We have also worked (with Prof. C. W. Gear) on the extension of our equation-free methods to also make them mesh-free, observing the simulation data in the form of radial basis functions. In the process, we developed the idea of a new type of basis functions (time-varying RBFs, that is, RBFs whose coefficients are local linear polynomials in time). This "twist" alleviates the discontinuities that can potentially arise when several sequential snapshots of data are represented in terms of RBFs (8).

We worked in collaboration with Prof. Zabaras on the problem of factoring out of spatial symmetries when representing complex spatial data. We explored the formal testing of hypotheses: when is it that high dimensional data can successfully be reduced / modeled in terms of a reduced effective model ? The third and last one is something that we had hoped we would have resolved

this year but we were only able to do several years later (just recently): how to successfully extend as far as possible "off-sample" data taking advantage of their low-dimensionality. We resolved this through a combination of (global) diffusion map processing linked with (local) PCA.

UNEXPENDED FUNDS There were no unexpended funds (~ $8,000) by the end of the funding period (August 31, 2011) in Princeton.

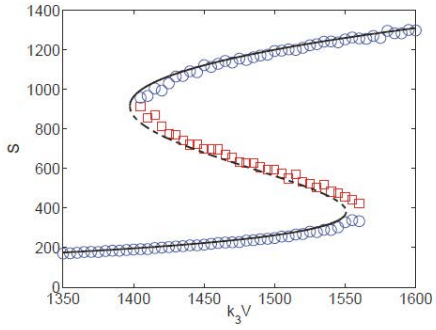
Two Representative Figures from the Research. TOP: Manifold learning for data in the form of graphs. Each colored point (bottom left) is a graph belonging to a two-parameter family. We were able to perform diffusion maps for these data, and the leading eigenvalues are plotted at the top left. The colors correspond to the component of the graph data on the leading eigenvector (eigenvector 2) and the next important eigenvector (eigenvector 4) discovered by the process – this means the data mining does discover that the family of graphs is a two-dimensional one, and also discovers the two "important observables). A careful study of the eigenvector 3 (plotted versus eigenvector 2 at the top right) indicates that it does not contain new information – it is like a second harmonic to eigenvector 2

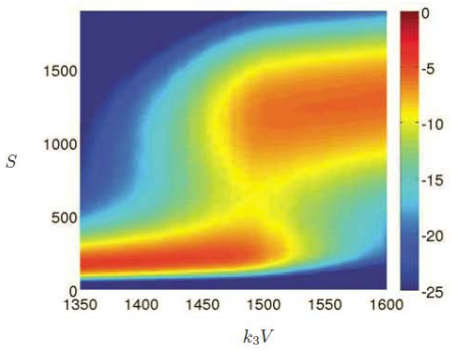


2

(what cos2x is for cosx – a higher harmonic, not a new independent variable).

BOTTOM: (from [3]). LEFT: stochastic bifurcation diagram for a chemical system as a parameter, estimated using a coarse-grained master equation (the solid lines are mean-field ODE results); RIGHT: log probability density of the system steady states again using a coarse grained master


S
k3V
1400
1200
1000
800
600
400
200
0
1350
1400
1450
1500
1550
1600


S
k_3V
1500
1000
500
0
1350
1400
1450
1500
1550
1600
0
-5
-10
-15
-20
-25

equation.

PUBLICATIONS acknowledging DOE support from this grant

1. A common approach to the computation of coarse-scale steady states and to consistent initialization on a slow manifold, C. Vandekerckhove, B. Sonday, A. Makeev, D. Roose, C. W. Gear and I. G. Kevrekidis *Comp. CHem. Eng.* **35**(10) pp.1949-1958 (2011)

2. Noisy dynamic simulations in the presence of symmetry: data alignment and model reduction, B. Sonday, A. Singer and I. G. Kevrekidis, *CAMWA* **65**(10) pp.1535-1557 (2013)

3. A constrained approach to multiscale stochastic simulation of chemically reacting systems, S. Cotter, K. C. Zygalakis, I. G. Kevrekidis and R. Erban, J. Chem. Phys. **135** (9) Article No. 094102 (2011)

4. Coarse-graining the dynamics of network evolution: the rise and fall of a networked society, Tsoumanis, A. C., Rajendran, K., Siettos, C. I., and Kevrekidis I. G., *N. J. Phys.* **14** (083037) DOI 10.1088/1367-2630/083037 (2012)

5. State reduction in molecular simulations, Y. Xue, P. J. Ludovice, M. A. Grover, L. V. Nedialkova, C. J. Dsilva and I. G. Kevrekidis, *Comp. Chem. Eng.* **51** pp.102-110 (2013)

6. An equation-free approach to agent-based computation: bifurcation analysis and control of stationary states, C. I. Siettos, C. W. Gear and I. G. Kevrekidis, , *Europhys. Lett.* **99**(4) (48007) DOI 10.1209/0295- 5075/99/48007 (2012)

7. Diffusion Maps, Clustering and Fuzzy Markov Modeling in Peptide Folding Transitions, L. V. Nedialkova, M. Amat, I. G. Kevrekidis and G. Hummer, *J. Chem. Phys.* **141** 114102 (2014)

8. Time varying radial basis functions and their use in mesh free/equation free computation, A. A. Jamshidi, C. W. Gear and I. G. Kevrekidis, *J. Comp. App. Math.* **266** pp.61-72 (2104)

9. Chimeras in Sparse Networks of Phase Oscillators, C. Laing, K. Rajendran and I. G. Kevrekidis Chaos 22 013132 (2012); doi:10.1063/1.3694118

10. Managing heterogeneity in the study of neural oscillator dynamics, C. Laing, Y. Zou, B. Smith and I. G. Kevrekidis, *J. Math. Neurosci.* **2**(5) DOI 10.1186/2190-8567-2-5 (2012)

11. An equation-free approach to coarse-graining the dynamics of networks, K. A. Bold, K. Rajendran, B. Rath and I. G. Kevrekidis, *J. Comp. Dyn* **1**(1) (2014)

12. E. Chiavazzo, C. W. Gear, C. J. Dsilva, N. Rabin and I. G. Kevrekidis, Reduced Models in Chemical Kinetics via nonlinear data-mining, *Processes* **2**(1) pp.112-140 (201

Plenary/Invited Talks, Presentations

There was a plenary lecture that included material from this work, at the 2011 annual SIAM control meeting in Baltimore; there was also two contributed presentations at the SIAM Dynamical Systems meeting in Utah, in May 2011, an invited talk at the Conference on Chemical Complexity (Fritz-Haber Institut der MPG, Berlin, July 2011), the Gutzwiller Colloquium at the MPIPKS-Dresden in December 2011, an invited talk at a Soft Condensed Matter Conference, Kyoto University, September 2011, a plenary lecture at ICNAAM 2011, Chalkidiki, Greece, September 2011, two contributed presentations at the AIChE annual meeting in Minneapolis, October 2011 and one contributed as well as one invited talk at MAPCON12, MPIPKS Dresden, May 2012.

There were also several invited presentations where work supported by this grant was presented: Department of Mathematics, University of Paderborn, Germany, June 2011; Department of Chemical Engineering, Kyoto University, September 2011; Department of Chemical Engineering, NJIT, Newark, NJ, November 2011; PTB, Berlin, December 2011; PICScie, Princeton University, March 2012 Mathematics Department, University of Delaware, April 2012; Applied Mathematics, University of Washington, May 2012; and Pacific Northwest National Lab, May 2012

It is gratifying to write that, in part as a result of this work, the PI has garnered the following distinctions over the last few years:

- Einstein Foundation and Zuse Institute, Distinguished Visiting Fellowship (2016-2018)
- Isaac Newton Institute, Cambridge, Rothschild Visiting Distinguished Fellowship (2016)
- Institute for Advanced Study, T. U. Muenchen and European Union – Hans Fischer Senior Fellow, 2015-2017
- American Institute of Chemical Engineers, Fellow, Elected 2016
- Academy of Athens, Corresponding Member, Elected 2015
- Institute for Advanced Study, Technical University of Munich, and EEC: Hans Fischer Senior Fellowship (2015-2018)
- Isaac Newton Institute, Cambridge, Microsoft Visitorship (2013)
- Max Planck Society, MPIPKS, Dresden, Martin Gutzwiller Fellowship (2010-2011)