Inverse Problem Methodology in Complex Stochastic Models

In many fields, including engineering, physics, material sciences and biology, there is a growing need for use of complex dynamical systems to describe and capture essential features of experimental findings. In diverse applications ranging from characterization of individual parameters in HIV modeling to characterization of polarization mechanisms in dielectric materials, statistical methods associated with computational algorithms for estimation of parameters are needed. These methods entail treatment of (generally unobservable) individual "parameters" (e.g., functions representing growth and mortality rates) as random variables to be estimated from data on the overall dynamics. An approach combining applied mathematics and statistics in a synergistic treatment of the fundamental issues offers ground breaking potential.

Even simple forward solutions (i.e., solutions of the dynamics when the parameters are specified) of the required dynamical systems (examples are discussed below) often necessitate sophisticated mathematical, statistical and computational techniques. Inverse problem methodologies for these systems are even more challenging. Nonetheless, there is a large literature available as long as parameters in these complex dynamical systems are viewed as deterministic variables. This may be sufficient when interest is on fitting them to data for a single individual, though intra-individual variability may cause difficulties. Moreover, data are routinely collected from a number of individuals from a population with a broader focus on understanding mechanistic behavior both across the population and within individuals. In this case, simple aggregation of data ignoring individuals is inappropriate. A relevant alternative treats unknown (generally unobservable) system parameters as random quantities whose distribution is to be estimated.

There is a substantial statistical literature on frameworks in which one treats finite-dimensional parameters characterizing a nonlinear model as random, and associated computational techniques for fitting are available. Efforts on treating function-valued parameters in dynamical systems are on-going. These random effects techniques have been used successfully with linear and simple nonlinear ODE models by statisticians, typically with little input from mathematicians. However, application in the context of complex dynamical systems is largely unexplored.

When combined, the computational and theoretical challenges posed by both mathematical and statistical issues are substantial. Their resolution requires integration of statistical and mathematical modeling considerations. A major thrust of this SAMSI program entails facilitating the essential cooperative effort required and promoting the development of a jointly-derived mathematical and statistical theoretical framework for estimation in complex nonlinear dynamical systems. This development should include estimation for unobservable function-space-(infinite-dimensional)-valued random parameters. For such a framework, it will be necessary to combine core mathematical components (e.g., PDE theory, functional analysis, approximation theory, optimization, computational algorithms) with probabilistic foundations (e.g., empirical process theory), and statistical methodology and foundations (e.g., computational fitting algorithms, nonparametric function and density estimation).

As a natural part of this SAMSI program, development of reduced order modeling techniques in the context of inverse problems will be pursued. Recent developments in model (and hence computational time) reduction, that preserve essential dynamic features, are based on the Karhunen-Loève or Proper Orthogonal Decomposition (POD) methods for reduced basis construction arising in feature extraction in statistical data analysis. These methods have recently been applied with tremendous success in electromagnetic inverse problems (resulting in speed-ups on the order of 4000 with little or no loss in accuracy). Computational issues posed by complex structure random parameter models in this proposed program should be approachable through development of a POD-based inverse problem computational methodology. There are many fundamental issues to be addressed (e.g., how to best choose snapshots to generate good POD basis elements in inverse problems), and these must be integrated with statistical estimation considerations.

Further refinement of goals will be pursued through the following 3 test-bed examples to illustrate the types of challenges involved. These examples are not necessarily to be pursued by participants at workshops as the focus for specific research activities. Rather, they serve to illustrate the needs for new ideas and methodology and to exemplify the types of questions requiring mathematical and statistical scientists to work together.

Polymers [Interrogating light beam through dilute polymer-sample]:
Here interest focuses on the distribution of distance between random points on molecules to yield information about shape of molecules; it is characterized by simple applied math/inverse problem/system formulation. Inference on this distribution is based on the model for intensity of scattered light at different angles taking into account measurement error. This embodies a complicated statistical problem (estimation over function space of distributions).

Dielectric Materials [Interrogating dielectric materials with microwave pulses to determine polarization (P) of materials from reflected signals]:
The electric polarization for such systems is described by linear and nonlinear dynamical systems which are coupled (as internal dynamics) to the usual Maxwell system via constitutive laws. A cutting edge research problem consists of determining a characterizing constitutive polarization (and/or conduction) law from observations of reflections of interrogating microwave pulses. Because any polarization mechanism in a realistic material (including living tissue) is a complex combination of mechanisms, even in data collected from a single individual or material sample, a “mixing distribution” of polarization laws is required to treat intra-individual variability. For general characterizations using aggregate data taken across samples, the usual inter-individual variability must also be taken into account, thus resulting in an inverse problem where the object being sought is a random variable distributed over a class of linear and/or nonlinear dynamical systems; i.e., the unknown parameter is a dynamical-systems-valued random parameter in the Maxwell equations.

HIV dynamics [modeling of population HIV dynamics from intermittent longitudinal data from each of a sample of infected individuals]:
A critical problem in the study of HIV disease is elucidation of the mechanisms governing the evolution of resistance to potent antiretroviral therapy, viral eradication or remission, and the potential effects of vaccination. There is a growing realization among HIV scientists that real progress in uncovering underlying mechanisms will require consideration of complex dynamical systems (e.g., nonlinear delay differential and integro–partial differential equations with unknown functional parameters and kernels). Application of models to data across patients to draw population-level inferences must be carried out in a statistical framework that recognizes both intra- and inter-patient variation in underlying mechanisms.

The Program Committee is chaired by H.Thomas Banks (Mathematics, NCSU; contact htbanks@unity.ncsu.edu). Other members are Richard Albanese (M.D., Air Force Research Labs), Marie Davidian (Statistics, NCSU), Sarah Holte (Fred Hutchinson Cancer Center), Joyce McLaughlin (Mathematics, Rensselaer Polytechnic Institute), Alan Perelson (Biology and Biophysics at Los Alamos National Labs), George Papanicolaou (Mathematics, Stanford), John Rice (Statistics, Berkeley), and Robert Wolpert (Statistics, Duke).