SUMMER SCHOOL ON DESIGN AND ANALYSIS OF COMPUTER EXPERIMENTS

THE IRMACS CENTRE, SIMON FRASER UNIVERSITY
AUGUST 11-16, 2006

CONFERENCE PROGRAM

SUMMER SCHOOL ON DESIGN AND ANALYSIS OF COMPUTER EXPERIMENTS

THE IRMACS CENTRE, SIMON FRASER UNIVERSITY
AUGUST 11-16, 2006

We are very pleased to welcome you to the Summer School on the Design and Analysis of Computer Experiments. There are many aims of the Summer School, ranging from the training of highly qualified personnel to encouraging the interaction between graduate students, post-doctoral fellows, industry and academic researchers. This event helps lay the groundwork for the 2006-07 Program on Development, Assessment and Utilization of Complex Computer Models at the Statistical and Applied Mathematical Sciences Institute (SAMSI).

We are grateful to the IRMACS Centre for hosting the Summer School, and to all our sponsors for their generous support:

The National Program on Complex Data Structures (NPCDS)
Pacific Institute for the Mathematical Sciences (PIMS)
Statistical and Applied Mathematical Sciences Institute (SAMSI)

We hope you find the conference enjoyable and stimulating.

Jim Berger
Derek Bingham
Randy Sitter
Boxin Tang
Will Welch

Summer School on the Design and Analysis of Computer Experiments

The IRMACS Centre, Simon Fraser University
August 11-16, 2006

Conference Schedule

Friday, August 11 – IRMACS Presentation Studio

8:00 - 9:00	Registration & IRMACS Centre Tour
8:15 - 9:00	BREAKFAST, IRMACS ATRIUM
9:00 - 9:30	Opening Remarks: Derek Bingham, Conference Coordinator James Berger, Director, SAMSI Pam Borghardt, Associate Director, The IRMACS Centre Brian Corrie, IRMACS Demo
9:30 - 10:00	Short course in computer experiments
10:00 - 10:15	COFFEE BREAK, IRMACS ATRIUM
10:15 - 12:00	Short course in computer experiments
12:00 - 1:15	LUNCH BREAK – BOX LUNCH PROVIDED
1:15 - 2:30	Short course in computer experiments
2:30 - 2:45	COFFEE BREAK, IRMACS ATRIUM
2:45 - 4:15	Short course in computer experiments
4:30	RECEPTION, IRMACS ATRIUM

Saturday, August 12 – IRMACS Presentation Studio

8:30 - 9:00	BREAKFAST, IRMACS ATRIUM
9:00 - 10:00	Short course in computer experiments
10:00 - 10:15	COFFEE BREAK, IRMACS ATRIUM
10:15 - 11:30	Short course in computer experiments
11:30 - 12:15	Computer Experiments at NCAR: Applications and Opportunities
12:45	Coach leaves **SFU Bus Loop** for excursion to Stanley Park
6:00	First coach returns to SFU from Stanley Park
9:30	Second Coach returns to SFU from Stanley Park

Summer School on the Design and Analysis of Computer Experiments

The IRMACS Centre, Simon Fraser University
August 11-16, 2006

Sunday, August 13 – IRMACS Presentation Studio

8:30 - 9:00	BREAKFAST, IRMACS ATRIUM
9:00 - 10:30	Anthony O'Hagan, University of Sheffield ***Building and using an emulator with GEM-SA*** Brian Williams, Los Alamos National Laboratory ***GPM: Software for Calibrating Computer Models to Experimental Data***
10:30 -10:45	COFFEE BREAK, IRMACS ATRIUM
10:45 - 11:30	Laura Swiler, Sandia National Laboratories ***The DAKOTA Toolkit and its use in Computational Experiments***
11:30 - 12:00	Tom J. Santner, The Ohio State University ***A Tutorial on the PErK Program***
12:00 - 1:30	LUNCH BREAK
1:30 - 2:30	Jim Berger and Fei Lui, Duke University Rui Paulo, University of Bristol Jerry Sacks, NISS ***SAVE-1 and SAVE-2 for Computer Models***
2:30 - 2:45	COFFEE BREAK, IRMACS ATRIUM
2:45 - 3:45	Jim Berger and Fei Lui Duke University Rui Paulo, University of Bristol Jerry Sacks, NISS ***SAVE-1 and SAVE-2 for Computer Models***
3:45 - 4:45	Tutorial on software for cosmology problem

Summer School on the Design and Analysis of Computer Experiments

The IRMACS Centre, Simon Fraser University
August 11-16, 2006

Monday, August 14 – IRMACS Presentation Studio

8:30 - 9:00	BREAKFAST, IRMACS ATRIUM
9:00 - 10:30	Fei Liu, Duke University ***Simulator Analysis and Validation Engine 2*** Matt Taddy, University of California, Santa Cruz ***Multi-Resolution Treed Gaussian Processes*** Elaine Spiller, University at Buffalo ***Rare Events in Nonlinear Lightwaves Systems***
10:30 -10:45	COFFEE BREAK, IRMACS ATRIUM
10:45 - 12:15	Pritam Ranjan, Simon Fraser University ***Sequential Experiment Design for Contour Estimation from Computer Simulators*** Yan Lan, University of Michigan ***A Two-Stage Procedure for Change Point Estimation*** Dianne Bautista, The Ohio State University ***Nonparametric Estimation of the Covariance Function of Stationary Gaussian Processes***
12:15 - 1:45	LUNCH BREAK
1:45 - 2:45	Gang Han, The Ohio State University ***Calibration and Prediction for Computer Experiment Output Having Qualitative and Quantitative Input Variables*** Jason Loeppky, UBC Okanagan ***Successful Calibration: A Practitioners Guide***
2:45 - 3:00	COFFEE BREAK, IRMACS ATRIUM
3:00 - 4:30	Bela Nagy, University of British Columbia ***Fast Bayesian Implementation (FBI) of Gaussian Process Regression*** Ying Hung, Georgia Tech. University ***Blind Kriging: A New Method for Developing Metamodels*** James D. Delaney, Georgia Tech. University ***"Functionally Induced Priors for the Analysis of Physical Experiments"***
5:30	CONFERENCE DINNER AT THE DIAMOND ALUMNI CENTRE

Summer School on the Design and Analysis of Computer Experiments

The IRMACS Centre, Simon Fraser University
August 11-16, 2006

Tuesday, August 15

8:30 - 9:00	BREAKFAST, IRMACS ATRIUM
9:00 - 9:15	Opening Remarks: Derek Bingham
9:15 - 10:45	Application in Cosmology
10:45 - 11:00	COFFEE BREAK, IRMACS ATRIUM
11:00 - 12:30	Problem Solving
12:30 - 2:00	LUNCH BREAK
2:00 - 3:30	Problem Solving
3:30 - 3:45	COFFEE BREAK, IRMACS ATRIUM
3:45 - 5:15	Problem Solving

Wednesday, August 16

8:30 - 9:00	BREAKFAST, IRMACS ATRIUM
9:00 - 10:30	Problem Solving
10:30 - 10:45	COFFEE BREAK, IRMACS ATRIUM
10:45 - 12:15	Problem Solving
12:15 - 1:45	LUNCH BREAK
1:45 - 3:15	Presentations
3:15 - 3:30	COFFEE BREAK, IRMACS ATRIUM
3:30 - 5:45	Presentations
5:45	Closing Remarks

Summer School on the Design and Analysis of Computer Experiments

The IRMACS Centre, Simon Fraser University

August 11-16, 2006

Abstracts, in order of presentation

SUNDAY, AUGUST 13, 9:00 – 10:30 AM

Anthony O'Hagan, University of Sheffield

Building and using an emulator with GEM-SA

GEM-SA is user-friendly, Windows-based software for building a Gaussian process emulator and carrying out uncertainty and sensitivity analyses. This talk will illustrate how to use the software to build and validate an emulator of a computer code, and how to interpret the diagnostics and analyses it produces.

Brian Williams, Los Alamos National Laboratory

GPM: Software for Calibrating Computer Models to Experimental Data

GPM (Gaussian Process Modeling) is software in MATLAB for calibrating computer models to experimental data using a version of the Kennedy and O'Hagan model. Univariate and multivariate outputs are accommodated. For multivariate outputs, users have the flexibility to establish basis representations for the code output and discrepancy model, that are suitable for each individual application. GPM facilitates the use of information from multiple data sources to inform on a common set of parameters to be calibrated. In addition, optimization options are available to assist users in baselining their computer models. GPM offers users some basic sensitivity analysis tools for assessing model output sensitivity to the input parameters. Future directions for GPM include hierarchical modeling options for calibration parameters and discrepancy across separate physical experiments.

SUNDAY, AUGUST 13, 10:45 – 11:30 AM

Laura Swiler, Sandia National Laboratories

The DAKOTA Toolkit and its use in Computational Experiments

The DAKOTA toolkit (Design Analysis Kit for Optimization and Terascale Applications) provides a flexible, extensible interface between analysis codes and iterative system analysis methods. DAKOTA contains algorithms for optimization with gradient and nongradient-based methods; uncertainty quantification with sampling, reliability, and stochastic finite element methods; parameter estimation with nonlinear least squares methods; and sensitivity/variance analysis with design of experiments and parameter study capabilities. These capabilities may be used on their own or as components within advanced strategies such as surrogate-based optimization, mixed-integer nonlinear programming, or optimization under uncertainty. This talk will provide an overview of DAKOTA's capabilities with a focus on the uncertainty analysis and experimental design capabilities. Some examples will be presented.

Tom J. Santner, The Ohio State University

A tutorial on the PErK program

We will give an overview of the capabilities of the PErK (Parametric Empirical Kriging) software. Then we describe how to find and install it on Unix (Linux or Cygwin) systems. Finally we will run a series of PErK jobs that illustrate the program in action.

Summer School on the Design and Analysis of Computer Experiments

The IRMACS Centre, Simon Fraser University
August 11-16, 2006

SUNDAY, AUGUST 13, 1:30 – 2:30 & 2:45 PM – 3:45 PM

Jim Berger and Fei Lui Duke University
Rui Paulo, University of Bristol
Jerry Sacks, NISS

SAVE-1 and SAVE-2 for Computer Models

The Simulator Analysis and Validation Engine (SAVE) is a set of research software applications that implements the validation framework for computer models of Bayarri et al. (2005) and subsequent extensions. In this presentation, we will describe the main ideas behind this validation strategy, the scope of applicability of each of the software applications, and some of the implementation details, including its use in the context of specific validation problems.

We will start by describing the basic ideas of the general validation framework, which is comprised of six steps. Each of these steps may or may not involve computational work, and this exposition will emphasize the computational tasks rather than the methodological aspects of the framework. In essence, one has to deal with the construction of an approximation to the output of the computer model, one has to estimate the unknown parameters in the statistical model relating reality and computer output, and one has to predict both reality and the output of the computer model in untried combinations of the inputs.

The first module of the software bundle, SAVE-1, implements the strategy in the situation where the output of the computer model is a scalar, and there is potentially an uncertain parameter that takes the same value in all the field experiments. This is detailed precisely, and the utilisation of the code is exemplified using a specific real-world problem, which will help describing the more technical details of the software.

The second module, SAVE-2, deals with the situation where the output of the computer code is a highly irregular function of the inputs. The functional data is decomposed using wavelet representation techniques, and the validation strategy then proceeds by applying a hierarchical version of the scalar validation methodology to the wavelet coefficients, followed by transforming back to the functional data realm. The issue of uncertain controllable inputs in the field experiments is also tackled in this software. Again, technical details of the construction and utilization of the software are described using specific real-world applications.

MONDAY, AUGUST 14, 9:00 – 10:30 AM

Fei Liu, Duke University
Simulator Analysis and Validation Engine 2

A key question in evaluation of computer models is ``Does the computer model adequately represent reality?" A six-step process for computer model validation is set out based on comparison of computer model runs with field data of the process being modeled. The methodology is particularly suited to treating the major issues associated with the validation process: quantifying multiple sources of error and uncertainty in computer models; combining multiple sources of information; and being able to adapt to different -- but related-- scenarios.

Two complications that frequently arise in practice are the need to deal with highly irregular functional data and the need to acknowledge and incorporate uncertainty in the inputs. We develop methodology to deal with both complications. A key part of the approach utilizes a wavelet representation of the functional data, applies a hierarchical version of the scalar validation methodology to the wavelet coefficients, and transforms back, to ultimately compare computer model output with field output. The generality of the methodology is only limited by the capability of a combination of computational tools and the appropriateness of decompositions of the sort (wavelets) employed here.

Matt Taddy, University of California, Santa Cruz
Multi-Resolution Treed Gaussian Processes

Coupling Gaussian Processes with treed partitioning is an efficient way to model non-stationary behavior. I will discuss how this idea can be extended to deal with the common computer experiment situation where the data come from more than one model, and the models can be ordered in terms of fidelity. By partitioning over Gaussian Process models that incorporate multiple fidelity output, we maintain the structure and efficiency of the original TGP methodology. The methodology will be illustrated on example datasets.

Elaine Spiller, University at Buffalo
Rare Events in Nonlinear Lightwaves Systems

The nonlinear Schroedinger equation (NLS) with a periodic, varying dispersion coefficient models the dynamics of light in optical communication systems and mode-locked lasers. The dispersion-managed nonlinear Schroedinger equation (DMNLS) is an averaged version of NLS which restores some symmetries that are lost in NLS when the dispersion coefficient is not constant. I will discuss these symmetries, the corresponding conservation laws, and modes of the linearized DMNLS. I will also discuss how these linearized modes can be utilized to guide importance-sampled Monte-Carlo simulations of rare events in dispersion-managed lightwave systems subject to noise. This study is pertinent because the performance of lightwave systems is limited by the occurrence of rare events, i.e., noise-induced errors.

MONDAY, AUGUST 14, 10:45 – 12:15 PM

Pritam Ranjan, Simon Fraser University

Sequential Experiment Design for Contour Estimation from Computer Simulators

In many engineering applications, one is interested in identifying the inputs to a computer simulator that lead to a pre-specified output. In this talk we introduce statistical methodology that identifies the desired contour in the input space. The proposed approach has three main components. Firstly, a stochastic model is used to approximate the global response surface. The model is used as a surrogate for the underlying computer model and provides an estimate of the contour together with a measure of uncertainty, given the current set of computer trials. Then, a strategy for choosing subsequent computer experiments to improve the estimation of the contour is outlined. Finally, we discuss how the contour is extracted and represented. The methodology is illustrated with an example from a multi-class queuing system.

Yan Lan, University of Michigan

A Two-Stage Procedure for Change Point Estimation

Consider a constant regression model for a bounded covariate that has a single discontinuity (change point). It is assumed that one can sample the covariate at different values and measure the corresponding responses. Budget constraints dictate that total of n such measurements can be obtained. The goal is to estimate accurately the location of the change-point. A two-stage procedure is proposed and its properties examined, where at the first stage a proportion of the n points is sampled and the location of the change-point estimated. Subsequently, the remaining proportion of points are sampled from an appropriately chosen neighborhood of the initial estimate of the change point and a new estimate is obtained. The asymptotic distribution of the least squares estimate is derived using ideas from empirical processes. The improved efficiency of the procedure is demonstrated using real and synthetic data. The problem is motivated by problems in engineering systems, where the response corresponds to cost functionals and the covariate to stress or loading levels of the underlying system.

Dianne Bautista, The Ohio State University

Nonparametric Estimation of the Covariance Function of Stationary Gaussian Processes

The estimation of the covariance function is of interest in predicting the outcome of a computer experiment using the Emprical Best Linear Unbiased Predictor (EBLUP). A valid covariance function must be positive definite. To guarantee this, conventional parametric estimation arbitrarily assumes that the covariance function belongs to a certain family indexed by a parameter,$\boldsymbol{\theta}$, which is consequently estimated via maximum or penalized likelihood. To circumvent this arbitrariness, several non/semi-parametric approaches have been proposed. Four such estimators based on a single realization of a stationary Gaussian process are discussed. The data consist of $(\boldsymbol{x}_i, y_i)$, $i=1,2,\ldots,n$, where $\boldsymbol{x}_i \in \Re^d$, $d \geq 1$, and $y(\mathbf{x}_i) \in \Re$. These estimators are those introduced by Shaprio and Botha (1991), Hall and Patil (1994), Ong et al. (2002), and Elogne et al. (2003). These methods are compared to the Restricted Maximum Likelihood (REML) estimation procedure and also to each other with respect to mean square predictive error.

MONDAY, AUGUST 14, 1:45 – 2:45 PM

Gang Han, The Ohio State University

Calibration and Prediction for Computer Experiment Output Having Qualitative and Quantitative Input Variables

We propose statistical models for prediction and calibration that allow both qualitative and quantitative input variables. The model allows prediction of a computer code at an untested set of qualitative and quantitative inputs as well as quantifying the uncertainty in the prediction. In the case of calibration, both the physical experiment and computer code are allowed to depend on both types of variables. A Bayesian Qualitative and Quantitative Variable (QQV) model is constructed and implemented by Markov Chain Monte Carlo methodology. This model is compared with a frequentist approach and a Bayesian independence model in several examples.

This is joint work with Thomas Santner and William Notz.

Jason Loeppky, UBC Okanagan

Successful Calibration: A Practitioners Guide

Computer models to simulate physical phenomena are now widely available in engineering and science. Before relying on a computer model, a natural first step is often to compare its output with physical or field data, to assess whether the computer model reliably represents the real world. Field data, when available, can also be used to calibrate unknown parameters in the computer model. Calibration can be particularly problematic in the presence of systematic discrepancies between the computer model and field observations. In this talk we present results on a simulation study that is designed to assess how well the calibration parameter has been estimated, and the conditions under which calibration is possible. By simulating both computer model data, and physical observations from a Gaussian process the uncertainty due to using the incorrect model does not arise. This allows us a more accurate picture of the problems that can arise when attempting to calibrate the model in the presence of systematic discrepancy.

Joint work with William Welch and Brian Williams

MONDAY, AUGUST 14, 3:00 – 4:30 PM

Bela Nagy, University of British Columbia

Fast Bayesian Implementation (FBI) of Gaussian Process Regression

The traditional prediction variance formula for Gaussian Process Regression (Kriging) underestimates the true uncertainty because it doesn't incorporate the variability due to estimating the model parameters. This leads to overly optimistic prediction bands about the predictor. We propose a computationally cheap Bayesian alternative in the absence of subjective prior distributions of the parameters. Simulations show that the resulting prediction bands have better frequentist properties (in terms of coverage probabilities) than the ones based on the traditional method.

Join work with Stella Karuri, Jason Loeppky, William J. Welch

Ying Hung, Georgia Tech. University

Blind Kriging: A New Method for Developing Metamodels

Kriging is a useful method for developing metamodels for product design optimization. The most popular kriging method, known as ordinary kriging, uses a constant mean in the model. In this article, a modified kriging method is proposed, which has an unknown mean model. Therefore it is called blind kriging. The unknown mean model is identified from experimental data using a Bayesian variable selection technique. Many examples are presented which show remarkable improvement in prediction using blind kriging over ordinary kriging. Moreover, blind kriging predictor is easer to interpret and seems to be more robust to misspecification in the correlation parameters.

This is joint work with V. Roshan Joseph and Agus Sudjianto.

James D. Delaney, Georgia Tech. University

Functionally Induced Priors for the Analysis of Physical Experiments

Specifying a prior distribution for the large number of parameters in the linear statistical model is a difficult step in the Bayesian approach to the design and analysis of experiments. Here we address this difficulty by proposing the use of functional priors and then by working out important details for three and higher level experiments. One of the challenges presented by higher level experiments is that a factor can be either qualitative or quantitative. We propose appropriate correlation functions and coding schemes so that the prior distribution is simple and the results easily interpretable. The prior incorporates well known experimental design principles such as effect hierarchy and effect heredity, which helps to automatically resolve the aliasing problems experienced in fractional designs. (Joint work with Dr. V. Roshan Joseph.)