STAT 6544: Surrogate Modeling

Instructor: Robert B. Gramacy (Bobby)
e–mail: `rbg@vt.edu`
web: `http://bobby.gramacy.com`
Office hours: by appointment in Hutcheson 403G

Lectures:
Dates and times: Fall 2020 M/W 2:30–3:45pm in SEITZ 313 (virtual)

Prerequisites:
Mathematical Statistics, (Computational) Linear Algebra, Design of Experiments, experience with Bayesian Inference, and comfort with a programming language (e.g, R, Matlab, Python, C or Fortran)

Grading Breakdown:
75% Homework; 25% Take home project.
There will not be an in–class final exam.

Required Text:
R.B. Gramacy. *Surrogates: Gaussian Process Modeling, Design and Optimization for the Applied Sciences.* Chapman Hall/CRC 2020. `http://bobby.gramacy.com/surrogates/`

Optional Texts:

- R.H. Myers, D.C. Montgomery, C.M. Anderson-Cook. *Response Surface Methodology.* 4th edition; Wiley 2016
- T. Santner, B. Williams, W. Notz. *The Design and Analysis of Computer Experiments*, 2nd edition; Springer 2018
- G.E. Box, N.R. Draper, *Empirical Model-Building and Response Surfaces.* Wiley 1987; superceded by *Response Surfaces, Mixtures, and Ridge Analyses*, 2007, by the same authors
- A. Forrester, A. Sòbester, A. Keane, *Engineering Design via Surrogate Modeling, a practical guide*; Wiley 2008

Synopsis: This course details statistical techniques at the interface between mathematical modeling via computer simulation, computer model meta-modeling (i.e., emulation/surrogate modeling), calibration of computer models to data from field experiments, and model-based sequential design and optimization under uncertainty. The treatment will include some of the historical methodology in the literature, and canonical examples, but will concentrate on modern statistical methods, computation and implementation, as well as modern application/data type and size. The course will return at several junctures to real-word experiments coming from the physical and engineering sciences, such as studying the aeronautical dynamics of a rocket booster re-entering the atmosphere; modeling the drag on satellites in orbit; desigining a hydrological remediation acheme for water sources threatened by underground contaminants; studying the formation of super-nova via radiative shock hydrodynamics. The course material will emphasize deriving and implementing methods over proving theoretical properties.

Tentative Schedule/List of Topics:

- Overview: mathematical models, numerical approximation, simulation, computer experiments and (field) data, uncertainty quantification, and where statistics fits in.
- Four motivating examples, revisited throughout, illustrating some of the "goals" of statistical modeling and decision making via computer simulation:
  1. Rocket-booster dynamics: exploring the parameter space to learn a non-stationary (potentially multidimensional) input-output relationship
  2. Groundwater remediation: optimization under uncertainty and under constraints
  3. Radiative shock hydrodynamics: calibrating computer simulation to real data
  4. Satellite drag: challenges from a modern scale (big-data) computer experiment
- Space-filling design
- Gaussian process (GP) spatial models
  - Modeling (its all in the covariance function), inference and prediction (kriging)
  - Properties: smoothness, stationarity, isotropy, separable models, etc.
  - Interpolating deterministic computer model runs
  - Optimal design v. space-filling design; sequential design for surface exploration
  - Interpretation and alternative formulation, e.g., process convolutions
  - Sequential design for optimization (a.k.a. Bayesian Optimization) and optimization under constraints
  - Calibration to field data, fully propagated uncertainty quantification
  - Input sensitivity analysis
- Pushing the envelope: dealing with computational hurdles and rigid (uncheckable) assumptions (in GPs)
  - Big-n solutions: big computation, approximation via sparse matrices, data sub-setting, divide-and-conquer, hybrids
  - Nonstationary GP modeling
  - Non-GP alternatives: trees, treed-GPs, dynamic trees
  - Revisit four motivating applications
  - Heteroskedastic Gaussian process modeling and sequential design