Module 1: What is a Computer Experiment?

Jerome Sacks and William J. Welch

National Institute of Statistical Sciences and University of British Columbia

Adapted from materials prepared by Jerry Sacks and Will Welch for various short courses

Acadia/SFU/UBC Course on Dynamic Computer Experiments
September–December 2014

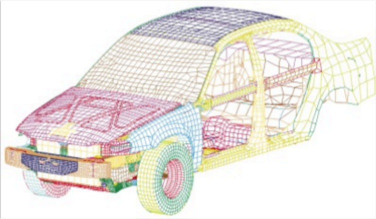
Outline of Topics

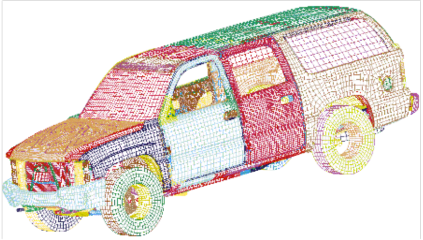
1. Examples of Computer Models
2. Computer codes
3. Role of Statistics
4. Next Steps

Computer Models

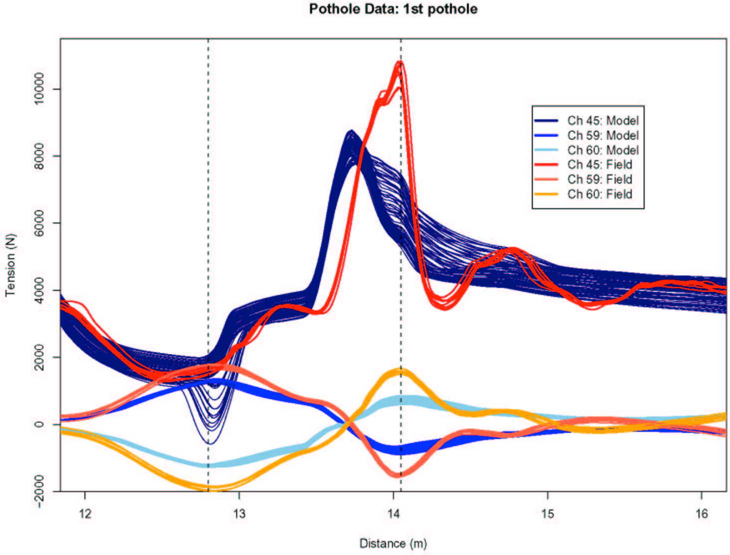
- Scientific / engineering / . . . phenomenon
  - car crash
  - vehicle suspension system meets pothole
- Theory and mathematical formulation
  - mechanical deformation
  - dynamic stress-strain
- Computational implementation (computer model, numerical model, computer code)
  - finite element code; vehicle represented by as many as 300,000 elements; 1+ days per run per work station
  - suspension system less, but 1 hour/run

Car Crash





Vehicle Suspension



Vehicle Suspension Experiment

- Implementation, generally complex
  - characteristics of restraint systems and components
  - inputs for geometry, materials properties
  - inputs for numerical parameters (time-step, mesh size)
- assessment of validity of the code
  - physical data needed
- Objectives
  - prediction (e.g., peak stress); answer what if questions
  - optimize time to deploy air bag
  - validate code against physical data
- Iterative feedback at all stages

Arctic Sea-Ice (Chapman, Welch, Bowman, Sacks, Walsh 1994)

- Objective: estimate sensitivities to input parameters
- Computer model: dynamic formulation based on a momentum balance for a mass of ice within a grid cell
- Model run: daily time step 1960-1988; 110 km grid covering Arctic Ocean and nearby bodies of water
- Inputs (13 variables): drag coefficients (ocean, atmosphere), albedo (snow, ice, open water), surface sensible heat, etc.
- Output(s): Ice mass, ice area, mean drift velocity, range of ice area

Arctic Ocean

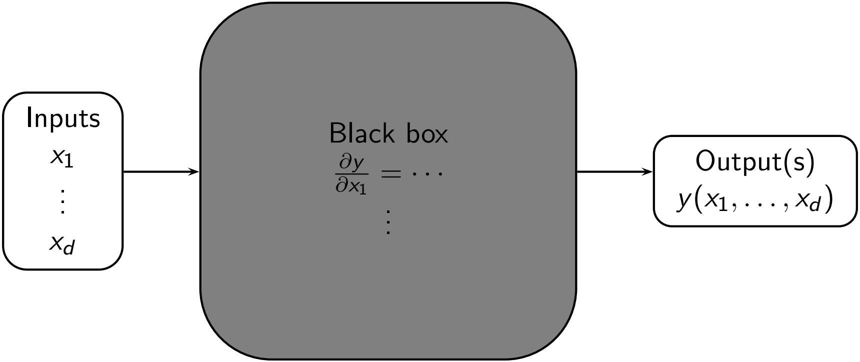
Arctic Sea-Ice Code

See icecode.pdf

Characteristics of Computer Codes

- Numerical solution of differential equations
- Science is complex, codes are complex
- Codes often costly to run
- Deterministic: repeating the inputs produces the same outputs
- Black-box: codes can be proprietary or scientifically/mathematically impenetrable; sometimes partially transparent (gray-box)

Computer Code



Deterministic Codes: Why Statistics?

- Need for statistical approximation of computer output
  - limited number of runs (few data)
  - many inputs (high dimension)
  - predict model output at untried inputs based on limited data
  - there will be uncertainty in predicitions
- Some other sources of uncertainty
  - propagation of variation: variability in inputs induces uncertainty in outputs
  - assess the fidelity of the model to reality (validation); needs field/experimental data which will be measured with error
- Design
  - where to make the runs (selection of inputs)

What's Different?

- The pillars of classical design of experiments are blocking, randomization, and replication
- The codes we will look at are (nearly) deterministic: blocking, randomization, and replication (with the same input values) are largely irrelevant
- Bias dominates in variance-bias trade-offs
- Less concerned with balanced (symmetric) designs for minimizing variance; more with spatial coverage (space filling)
- Experimenters will often have wide ranges for inputs, hence highly nonlinear relationships are often found
- With many input variables and limited computer runs, polynomial regression, neural nets, common nonparametric smoothers are difficult to use

Plan for the Next Few Weeks

- Analysis: Approximation / prediction / emulation of a computer code, mainly via Gaussian process (GP) statistical models, along with a measure of uncertainty
- Design of computer experiments: space-filling designs, sequential designs
- Scientific and engineering objectives
- Combining computer model runs and physical data: calibration, assessment of validity
- See course outline for more details
- Task for Thursday: Write down and bring to class 10 properties of the normal distribution, multivariate normal distribution, or conditional multivariate normal distribution. They will be collected and we will compile a list of such properties.