AN OVERVIEW OF COMPUTER EXPERIMENTS

David M. Steinberg

Department of Statistics and OR

Tel Aviv University

ENBIS-EMSE
July 2009

Partners

The TITOSIM Project: Fiat, LSE, CNRS, Politecnico Torino, Blue Eng., KPA, TAMAM, Easi Eng., Snecma

Students: Dizza Bursztyn, Sigal Levy, Einat Neumann Ben-Ari

Collaborators: Ron Kenett, Ron Bates, Henry Wynn, Dennis Lin, Gideon Leonard, Tamir Reisin, Eyal Hashavia, Zeev Somer

SAMSI Focus Year in 2006-7

Software: JMP, PERK (Brian Williams)

PREVIEW

1. What are Computer Experiments?
2. Applications
3. Designing a Computer Experiment
4. Validation and Verification
5. Data Analysis and Modeling
6. Challenges

What are Computer Experiments?

Experiments where a process is studied using a *computer simulator* rather than doing laboratory or field testing.

The computer code implements a mathematical model of the process.

The Experimental Lab

Traditional

Expt run in lab, test bench or pilot plant

Today

Experiment run on computer simulator

What are Computer Experiments?

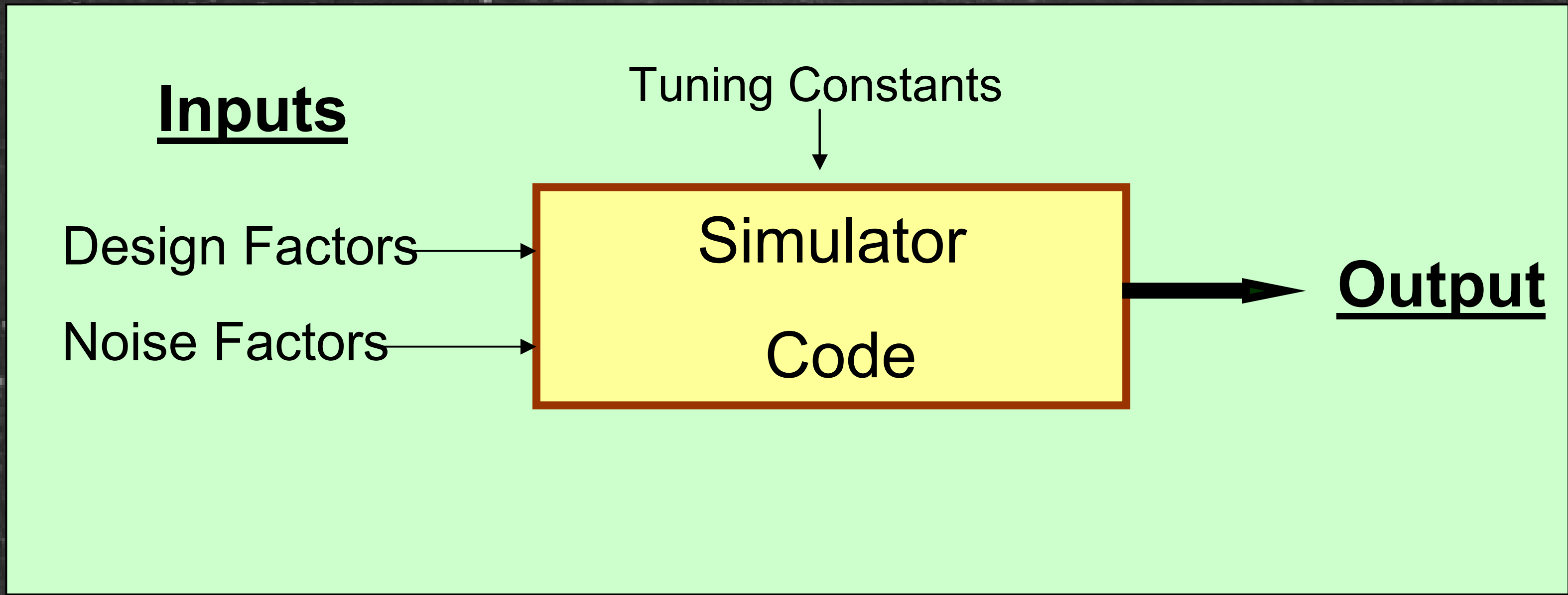
"In the twenty-first century computer simulation experiments will replace physical experiments for many applications."

Jeff Wu

International Conference on Design of Experiments

Nankai University, 2006

What are Computer Experiments?



There may also be field data that permit validation of the code.

What are Computer Experiments?

The goals might be:

- To identify the important factors.
- To better understand the system.
- To assess the consequences of uncertain inputs.
- To develop a “prediction equation”.
- To locate “good operating conditions”.
- To improve system robustness to noise.

What are Computer Experiments?

Inputs to the code can be classified into three primary types:

- Design Factors – which we can control and set to desired values.
- Noise Factors – which we cannot control.
- Tuning Constants – physical parameters whose value is not known in advance.

What are Computer Experiments?

Background reading.

R.S. Kenett, D.M Steinberg, New frontiers in design of experiments. *Quality Progress,* August (2006), 61-65.

S. Thomke, Enlightened experimentation – the new imperative for innovation. Harvard Business Review (2001), 66-72.

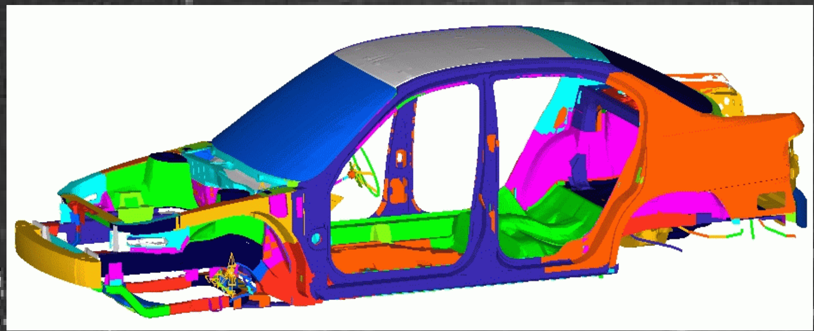
J. Sacks, W.J. Welch, T.J. Mitchell, H.P. Wynn, Design and analysis of computer experiments (with discussion). *Statistical Science* 4 (1989), 409–435.

V.C.P. Chen, K_L. Tsui, R.R. Barton and M. Meckesheimer, A review on design, modeling and applications of computer experiments, *IIE Transactions* (2006), 273- 291.

T.J. Santner, B.J. Williams, W.I. Notz, *The Design and Analysis of Computer Experiments.* Springer Verlag, New York (2003).

K.T. Fang, R. Li, A. Sudjianto, A. *Design and Modeling for Computer Experiments*, Taylor & Francis Group, Boca Raton, (2006).

Example: Automotive Crash Tests



Auto manufacturers want to improve crash resistance without adding excess weight to vehicles.

Field testing is limited due to expense.

Simulator codes make it much easier and faster to study design options.

Example: Automotive Crash Tests

Inputs

- Raw materials used
- Physical dimensions like weight, thickness
- Variations about nominal dimensions
- Angle of impact
- Velocity at impact

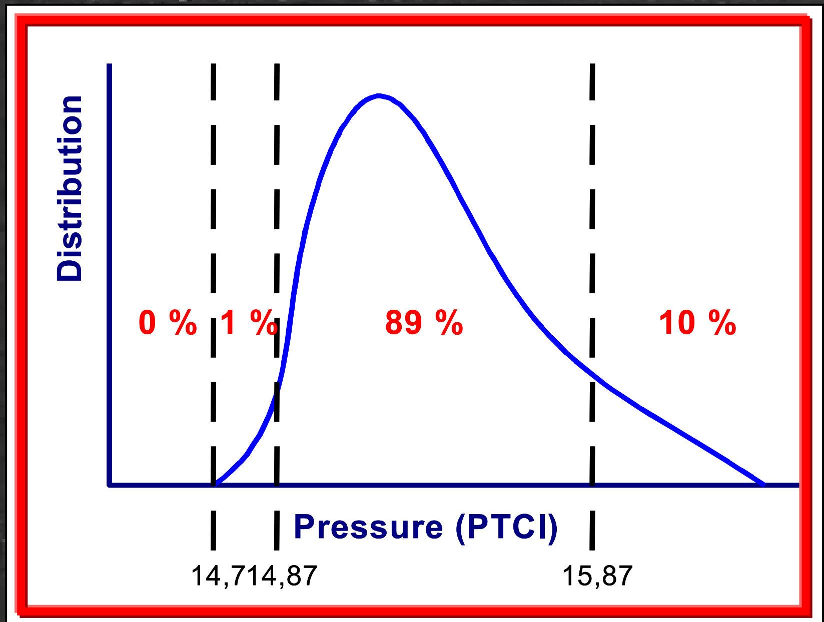
Outputs

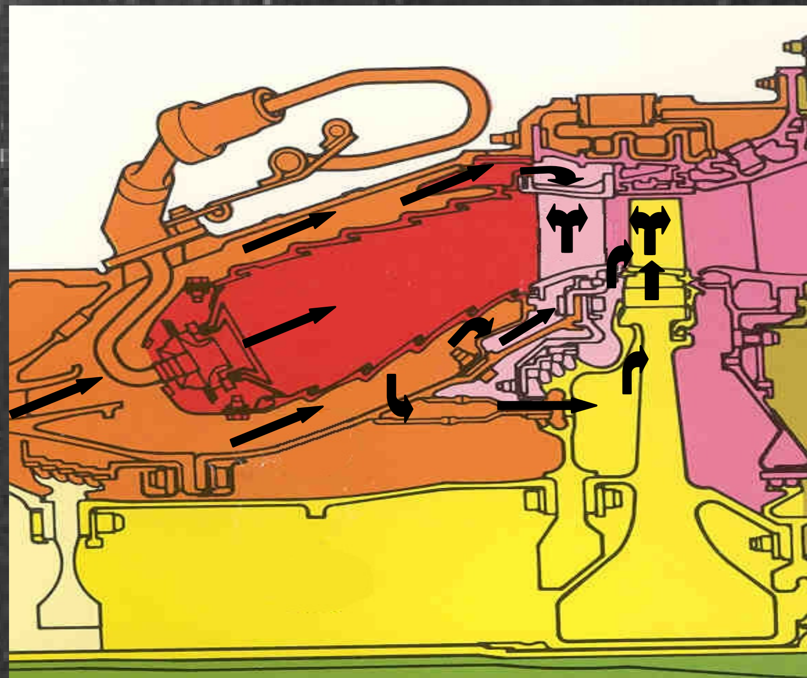
- Maximum penetration

Example: Air Cooling in High Pressure Turbine Blades

Inputs

- One design factor.
- Five noise factors with known distributions.

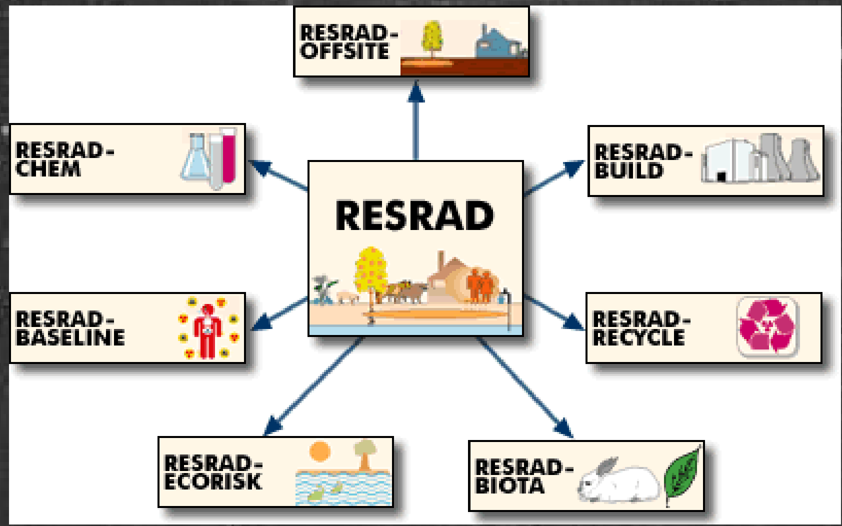




Outputs

- Temperature
- Pressure

Example: Nuclear Waste Repository



- RESRAD is a computer model designed to estimate radiation doses and risks from RESidual RADioactive materials.
- RESRAD simulates radiation doses and cancer risks for a variety of pathways in the environment (e.g. drinking water, food chain, atmosphere).
- Time frame is thousands of years, so field study is impossible.

Example: Nuclear Waste Repository

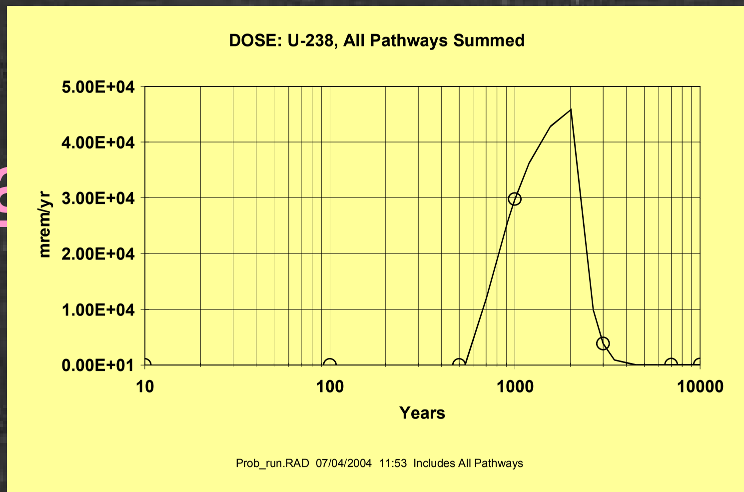
Inputs

- Initial isotope amounts
- Distribution coefficients of the isotopes
- Lithology of the repository

Outputs

- Maximal dose during 10,000 years

Goa sis.



Example: Climate Forecasts

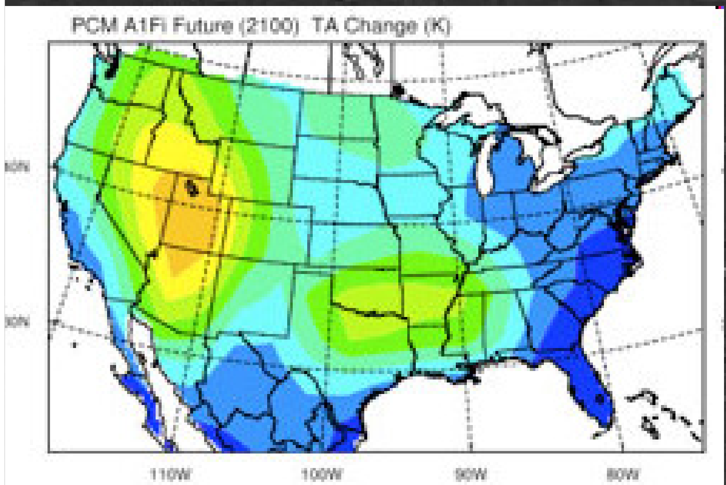
Inputs

- Atmospheric conditions.
- Physical constants.
- Multiple models – ensemble forecasting.

Outputs

- Reaction to increase in CO2.

See THORPEX project.



Example: Robust Design of Hip Implants

Goal: Achieve acetabular cup stability accounting for patient and surgical variability

Inputs

- Cup equatorial diameter and eccentricity
- Ten “field” variables that represent patient- and surgeon-specific factors.

Outputs

- Four performance measures

Designing a Computer Experiment

Guiding principles for "physical" experiments:

- Factorial structure (e.g. 2^{k-p} designs).
- Blocking.
- Randomization.
- Replication.

Designing a Computer Experiment

The only relevant issue is to choose the set of input vectors where the simulator will be run.

Often levels can be chosen on a fine grid; no need to limit to a small set of levels.

BUT …. This is *not* always possible.

Designing a Computer Experiment

Possible design criteria:

- Integrated MSE ↔ Bayes Posterior Variance
- Maximum Entropy
- Minimum Discrepancy
- Maximin Distance

Designing a Computer Experiment

Latin Hypercubes are the most popular class of experimental plan.

LHD's place the input levels for each factor on a uniform grid.

Then "mate" the levels across factors by randomly permuting the column for each factor.

McKay, Beckman and Conover, *Technometrics*, 1979.

Designing a Computer Experiment

Example of an LHD for 3 factors.

Initial Grids				Shuffled Grids		
1	1	1		1	0.3	0.5
0.9	0.9	0.9		0.9	0.4	0.2
0.8	0.8	0.8		0.8	1	0.7
0.7	0.7	0.7		0.7	0.6	0
0.6	0.6	0.6		0.6	0.2	1
0.5	0.5	0.5		0.5	0.7	0.9
0.4	0.4	0.4		0.4	0	0.1
0.3	0.3	0.3		0.3	0.9	0.6
0.2	0.2	0.2		0.2	0.5	0.4
0.1	0.1	0.1		0.1	0.8	0.8
0	0	0		0	0.1	0.3

Designing a Computer Experiment

With many factors, **Latin Hypercube Designs** typically have some large pairwise factor correlations.

Steinberg and Lin (Biometrika, 2006) developed a class of LHD's with many input factors, all of them orthogonal.

For example, we can generate a 256-run LH with 248 orthogonal factor columns.

Designing a Computer Experiment

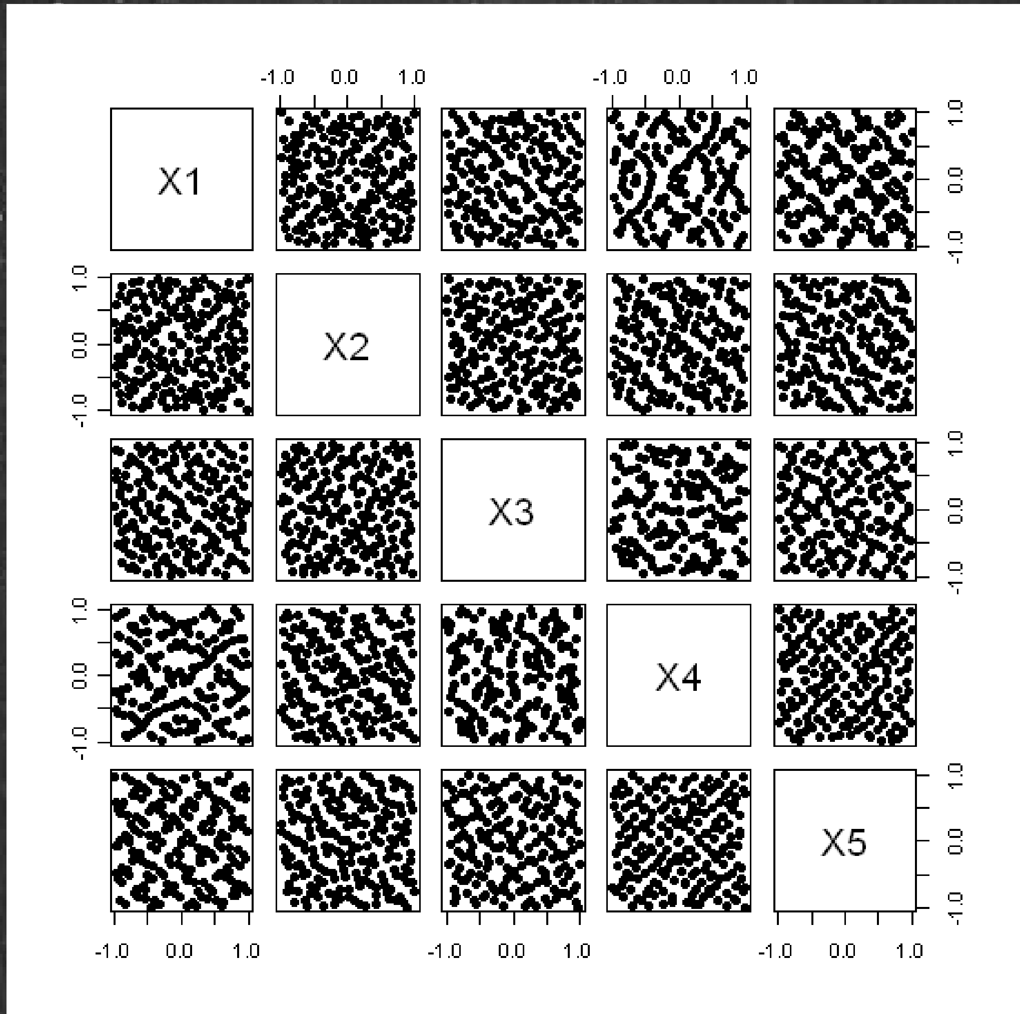
- Maximin is a popular method for choosing a LHD: maximize the minimum distance between any two points.
  - The intuition is that predictions are based mostly on "nearby" outcomes. This is a practical way to guarantee that all points have neighbors.

Designing a Computer Experiment

Other plans include "space filling" sequences, developed by Sobol, Niederreiter, Halton and Faure and the "uniform designs" of Fang and Winkler.

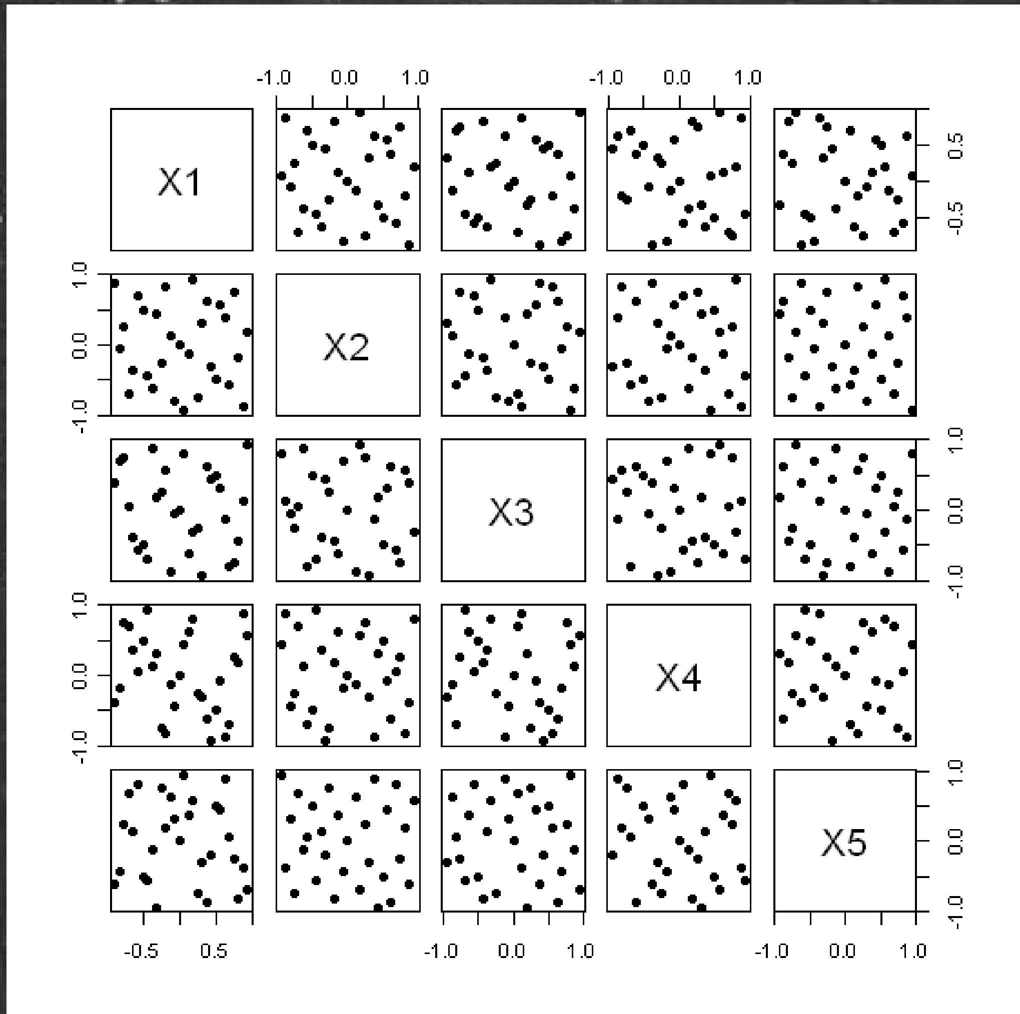
Designing a Computer Experiment

Two-factor projections from a 200-run, 5-factor Sobol sequence.



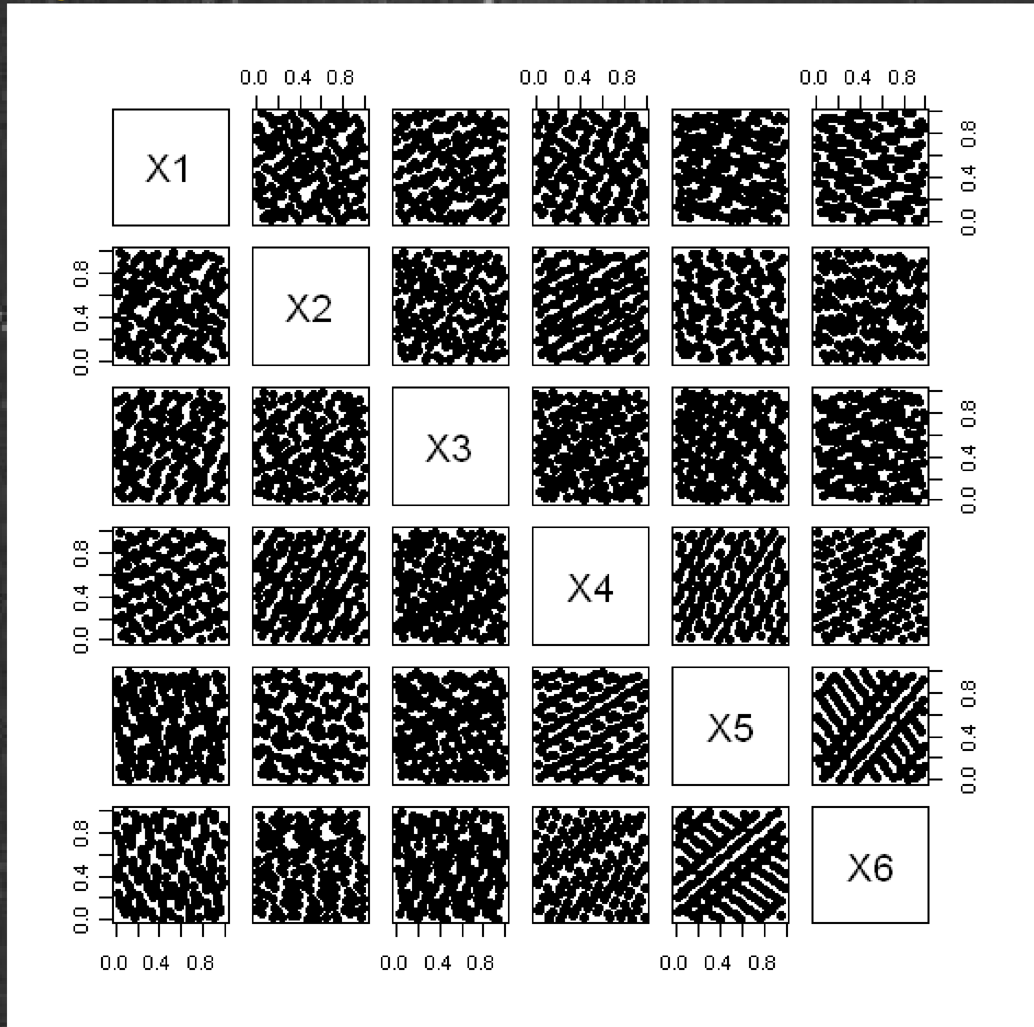
Designing a Computer Experiment

Two-factor projections from a 30-run, 5-factor Sobol sequence.



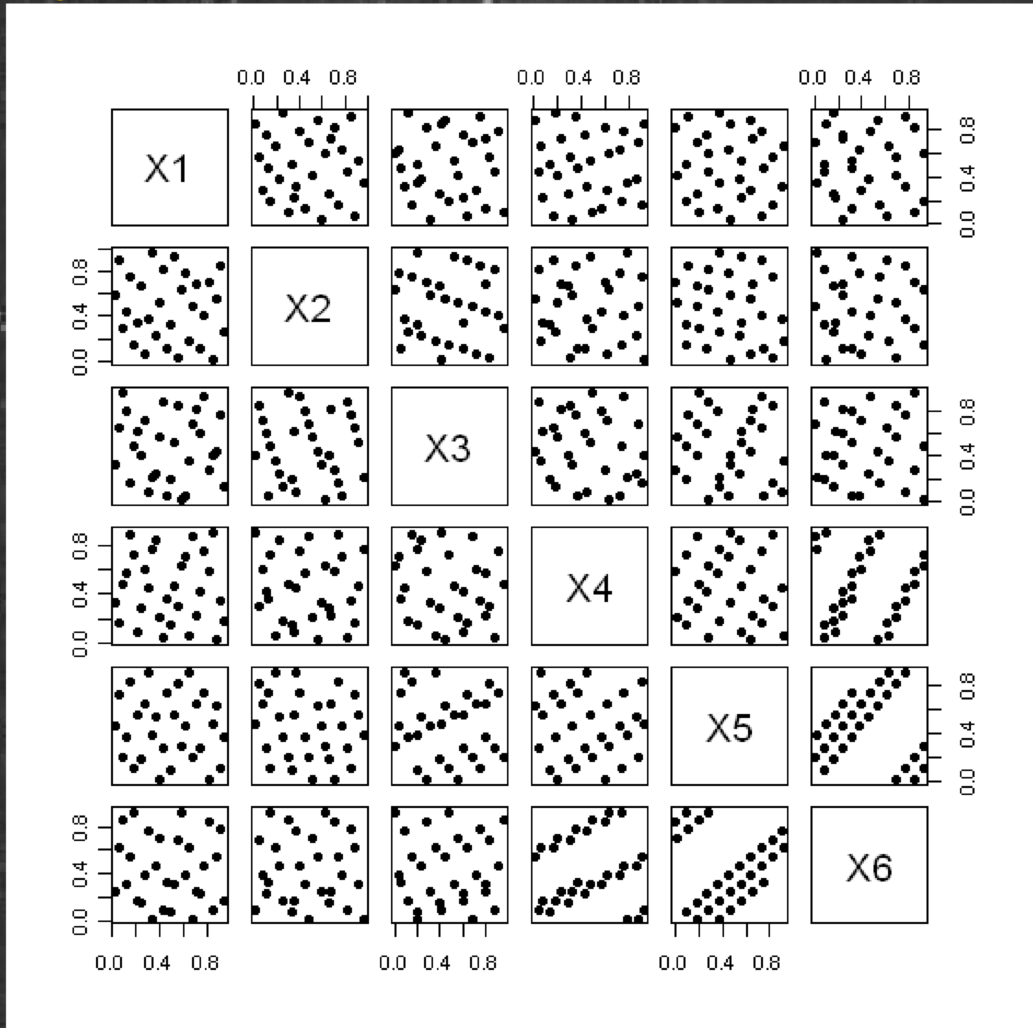
Designing a Computer Experiment

Two-factor projections from a 200-run, 6-factor Halton sequence.



Designing a Computer Experiment

Two-factor projections from a 30-run, 6-factor Halton sequence.



Validation and Verification

Validation:

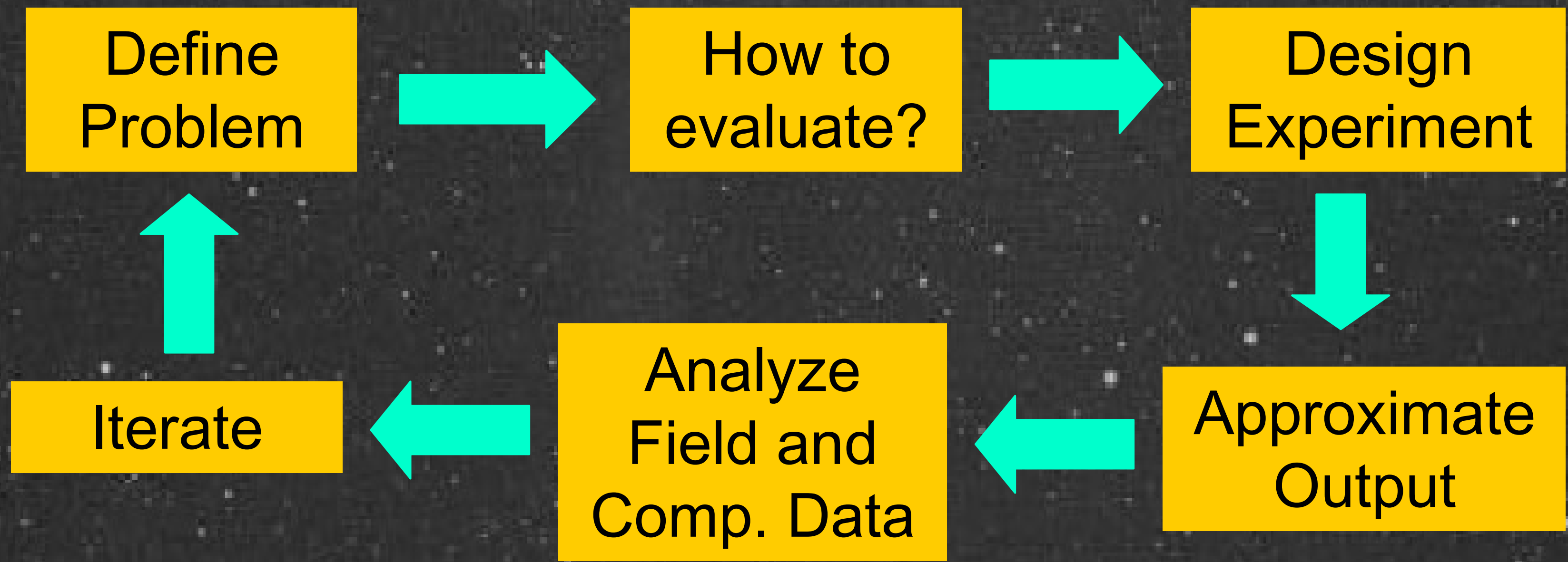
Does the code match actual experimental or field data?

Verification:

Does the code correctly represent the theory?

Validation and Verification

The six stage validation scheme in Bayarri et al. (*Technometrics*, 2007).



Data Analysis And Modeling

What kind of model should be used for data from computer experiments?

We need to consider:

- Attention to bias, not to variance.
- Nonlinear effects and interactions.
- High-dimensional inputs.

Data Analysis And Modeling

Methods for high-dimensional smoothing or interpolation are natural candidates.

- Gaussian Process (GP) model.
- Treed GP model. *Herbie Lee.*
- Kernel methods. *Stéphane Canu.*
- Interpolating polynomials. *Wynn & Bates.*
- Spline ANOVA decomposition.
- MARS.
- ACE.
- Projection Pursuit.

Data Analysis and Modeling

The GP model has its roots in kriging, a method of spatial interpolation or smoothing common in geostatistics.

Much of the fundamental theory was developed at the École des Mines de Paris by Georges Matheron.

1958

1988

Data Analysis and Modeling

The idea is that the outcomes at two "nearby" input sites should be similar, and to model this via high correlation.

Kriging

- 2-3 dimensions
- Euclidean distance
- Emphasis on form of correlation function.

Computer Expts

- Many dimensions
- No natural distance
- Emphasis on form of distance function.
- Limited choice of correlation functions

Data Analysis and Modeling

Let y denote a response and x a vector of input factor settings.

Treat y as the realization of a *Gaussian Process* with a fixed regression component:

$$y(x) = \beta_0 + \sum \beta_j f_j(x) + \eta(x)$$

The regression part is often limited to just the constant term.

For stochastic data, one can add a white noise term (or “nugget”).

Data Analysis and Modeling

The random field $\eta(x)$ is used to represent the departure of the true response function from the regression model.

Typical assumptions:

$$E\{\eta(x)\} = 0.$$

$$E\{\eta(x_1)\eta(x_2)\} = C(X_1,X_2) = \tau^2 R(X_1,X_2)$$

Data Analysis and Modeling

- Typically the correlation function *R* includes parameters that can be estimated by maximum likelihood or by cross-validation.

- One popular recommendation:

$$R(x_1, x_2) = \exp\{ -\sum \theta_j \mid x_{1,j} - x_{2,j} \mid^{p(j)} \}$$

Distance function

Data Analysis and Modeling

- We can estimate the response at a new input site using the Best Linear Unbiased Predictor.
- The estimator is also the posterior mean if we assume that all random terms have normal distributions.
- The estimator is much more flexible than the standard regression model. It smoothly interpolates the output data.

Data Analysis and Modeling

The GP model has been found to generate good predictors in applications.

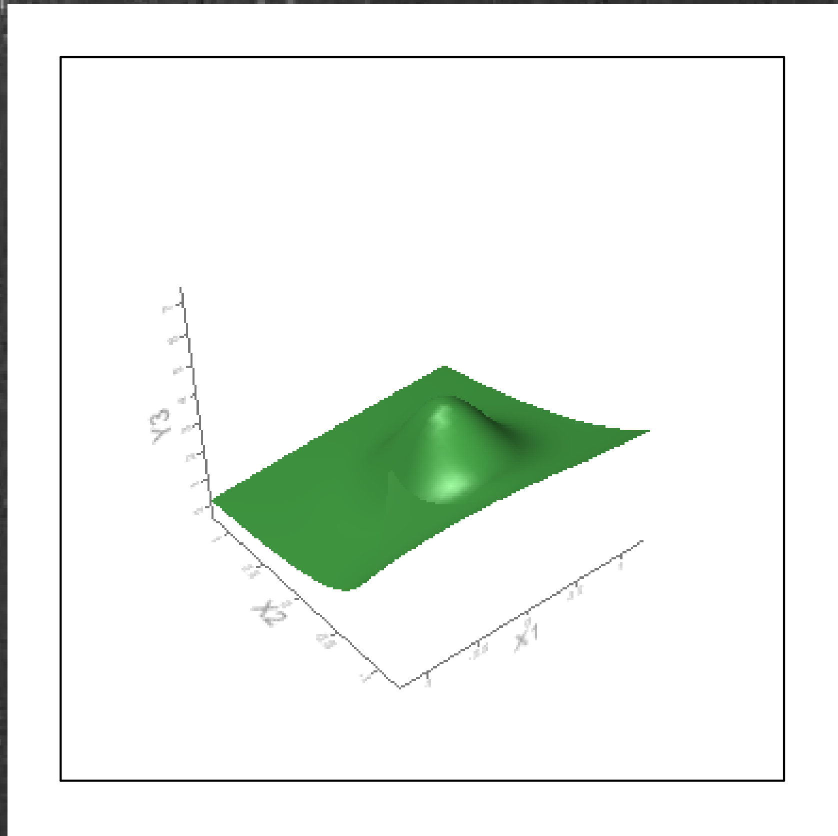
Neumann Ben-Ari compared the GP model to MARS and projection pursuit on a suite of examples.

The GP results were consistently better, sometimes much better.

Data Analysis and Modeling

GP Predictor on simulated data.

- Six factors.
- LHC design with 25 runs.
- Response depends only on X1 and X2.



Data Analysis and Modeling

The fitted model places almost all the weight on X1 and X2.

The scale coefficients are as follows:

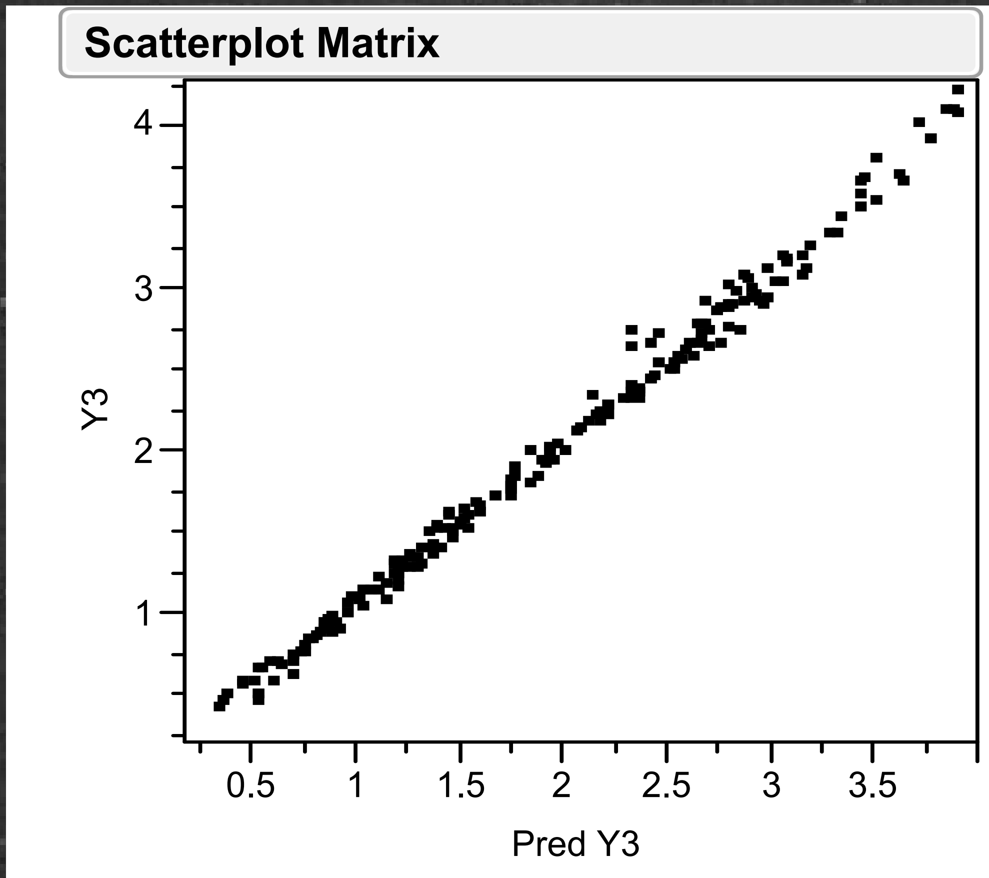
X1: 1.85	X2: 1.80	X3: 0
X4: 0.0006	X5: 0	X6: 0.0004

Correlations for Y at points 0.5 apart along the corresponding factor axis:

X1: 0.630	X2: 0.638	X3: 1.000
X4: 0.999	X5: 1.000	X6: 0.999

Data Analysis and Modeling

True vs. predicted values on an independent test set of 200 points.



Data Analysis and Modeling

The GP model has a natural Bayesian interpretation.

The correlation function can be seen as a prior distribution on the response function.

Data Analysis and Modeling

Steinberg and Bursztyn showed that the GP model can be understood as a Bayesian regression model.

Linkletter et al. developed a Bayesian approach to estimating the parameters in the correlation function via MCMC. They also considered averaging predictions over those parameters.

O'Hagan, Oakley and Kennedy have applied Bayesian ideas to address further problems in data analysis. Much is available on his web site.

Data Analysis and Modeling

The Steinberg-Bursztyn model is based on a series expansion to represent the outcome:

$$y(x) = \beta_0 + \sum \beta_j f_j(x)$$

Assign a vague prior to the constant.

Assume that the remaining terms are independent, with

$$\beta_j \sim \mathrm{N}(0, \tau_j^2).$$

Data Analysis and Modeling

We now have

$$y(x) = \beta_0 + \sum \beta_j f_j(x)$$

$$= \beta_0 + \eta(x)$$

Data Analysis and Modeling

The term $\eta(x)$ is a GP whose distribution is induced by the prior assumptions on the regression coefficients.

$$E\{\eta(x)\} = 0.$$

$$E\{\eta(x_1)\eta(x_2)\} = C(X_1,X_2) = \sum \tau_j^2 f_j(X_1) f_j(X_2)$$

Data Analysis and Modeling

The GP model does a good job of interpolating in the range of the data.

Sometimes we need to *extrapolate* outside the range of the data.

Extrapolations require having a fixed component to the regression model – but which terms?

The *blind kriging* method of Joseph, Hung and Sudjianto (J. of Mech. Design, 2008) is a good solution.

Data Analysis and Modeling

The GP model is built around continuous predictors. What can we do if there are also qualitative factors?

There are two recent proposals for defining and estimating appropriate correlation functions:

Qian and Wu, *Technometrics*, Aug., 2008.

Han and Santner, *Technometrics*, Aug., 2009.

Data Analysis and Modeling

Often data are multivariate or functional.

For example, a simulator of ground response to an earthquake gives accelerations as a function of location on a hillside.

For some useful approaches, see:

Bayarri et al. (*Annals of Statistics*, 2007).

Higdon et al. (in *Bayesian Statistics 8*, 2007)

Rougier (*J. Comp. Graph. Stat.*, 2008)

Data Analysis and Modeling

Often data are multivariate or functional.

For time series output, a better approach might be to emulate the autoregressive steps of the process.

This has been termed dynamic emulation.

Several papers by Marco Ratto, Peter Young and colleagues.

Data Analysis and Modeling

Data may have multiple levels.

For example, three codes for the same process.

Fast	Low Accuracy
Medium	Med. Accuracy
Slow	High Accuracy

Or simulator output together with field data.

Data Analysis and Modeling

Bayesian models have again proved a very effective way of "tying together" multi-level output.

Kennedy and O'Hagan (*Biometrika*, 2000)

Qian and Wu (*Technometrics*, 2008)

Han, Santner and Rawlinson (*Techn.*, 2009)

Data Analysis and Modeling

There is a natural interface with optimization.

Gramacy and Lee (*Technometrics*, 2009)

Lee, Taddy, Gray and Griffin (*Technometrics*, 2009)

Bates, Kenett, Steinberg and Wynn (*QTQM*, 2006)

Challenges

- Designs that are tied to our methods of analysis.
- Sample size guidelines.
- Methods for sequential experimentation.
- Design and analysis when factors are quantitative or have few levels.
- Effective design and analysis are needed for initial factor screening.
- Analysis methods for experiments with many factors.
- Combining multi-resolution output.
- Use of field data together with simulation output.
- Enhanced cooperation with scientists and engineers who are using computer models.
- More user-friendly software.