Post-Fisherian Experimentation: from Physical to Virtual

C. F. Jeff Wu
School of Industrial and Systems Engineering
Georgia Institute of Technology

- Fisher's legacy in experimental design.
- Post-Fisherian work in Factorial experiments:
  - principles for factorial effects.
  - conditional main effect analysis.
- Computer (virtual) experiments:
  - numerical approach.
  - stochastic approach via kriging.
- Summary remarks.

R. A. Fisher and his legacy

- In Oct 1919, Fisher joined Rothamsted Experimental Station. His assignr
  "examine our data and elicit further i hat
  we had missed." (by John Russell, St r)

- And the rest is history!
- By 1926 (a mere 7 years ☺), Fishe
  invented ANalysis Of VAriance and
  Design Of Experiments as new methods to
  design and analyze agricultural experiments.

Fisher's Principles in Design

- **Replication**: to assess and reduce variation.
- **Blocking**.
- **Randomization**.

"Block what you can,

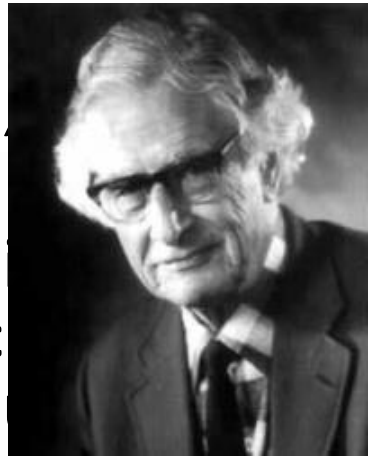
and randomize what you cannot."

- Originally motivated by agricultural expts, have been widely used for any physical expts.

Factorial Experiments

- Factorial arrangement to accommodate factorial structure of treatment/block, by Fisher (1926) . Originally called “complex experiments”.
- Major work on factorial design by F. Yates (1935, 1937),

  and fractional factorials by D. Finney (1945), both worked with Fisher.



- Major development after W icat industrial exp y t n Sc G. Box and co S. G. H

- What princip ove experiments?

Guiding Principles for Factorial Effects

- **Effect Hierarchy Principle**:
  - Lower order effects more important than higher order effects;
  - Effects of same order equally important.
- **Effect Sparsity Principle**: Number of relatively important effects is small.
- **Effect Heredity Principle**: for an interaction to be significant, at least one of its parent factors should be significant.

(Wu-Hamada book "Experiments", 2000, 2009)

Effect Hierarchy Principle

- First coined in Wu-Hamada book; was known in early work in data analysis.
- "From physical considerations and practical experience, (interactions) may be expected to be small in relation to error - - " (Yates, 1935); "higher-order interactions - - are usually of less interest than the main effects and interactions between two factors only." (Yates, 1937).
- The more precise version is used in choosing *optimal* fractions of designs; it can be used to justify maximum resolution criterion (Box-Hunter, 1961) and minimum aberration criterion (Fries-Hunter, 1980).

Effect Heredity Principle

- Coined by Hamada-Wu (1992); originally used to rule out incompatible models in model search.
- Again it was known in early work and used for analysis: "- - factors which produce small main effects usually show no significant interactions." p.12 of Yates (1937): "The design and analysis of factorial experiments",

  Imperial Bureau of Soil Science, No. 35.

More on Heredity Principle

- Strong (both parents) and weak (single parent) versions defined by Chipman (1996) in bayesian framework. Strong heredity is the same as the marginality principle by McCullagh-Nelder (1989) but with different motivations.
- Original motivation in HW: application to analysis of experiments with complex aliasing.

Design Matrix OA(12, 27) and Cast Fatigue Data

Full Matrix: $OA(12,2^{11})$

Run	Factor A	B	C	D	E	F	G	8	9	10	11	Logged Lifetime
1	+	+	−	+	+	+	−	−	−	+	−	6.058
2	+	−	+	+	+	−	−	−	+	−	+	4.733
3	−	+	+	+	−	−	−	+	−	+	+	4.625
4	+	+	+	−	−	−	+	−	+	+	−	5.899
5	+	+	−	−	−	+	−	+	+	−	+	7.000
6	+	−	−	−	+	−	+	+	−	+	+	5.752
7	−	−	−	+	−	+	+	−	+	+	+	5.682
8	−	−	+	−	+	+	−	+	+	+	−	6.607
9	−	+	−	+	+	−	+	+	+	−	−	5.818
10	+	−	+	+	−	+	+	+	−	−	−	5.917
11	−	+	+	−	+	+	+	−	−	−	+	5.863
12	−	−	−	−	−	−	−	−	−	−	−	4.809

Partial and Complex Aliasing

- For the 12-run Plackett-Burman design OA(12, 2^{11})

$$\mathrm{E}\hat{\beta}_i = \beta_i + \frac{1}{3}\sum_{j,k\neq i} \pm \beta_{jk}$$

partial aliasing: coefficient $\pm\frac{1}{3}$

complex aliasing: $45\,(=\binom{10}{2})$ partial aliases.

- Traditionally complex aliasing was considered to be a disadvantage (called "hazards" ☹ by C. Daniel).
- Standard texts pay little attention to this type of designs.

Analysis Strategy

- Use effect sparsity to realize that the size of true model(s) is much smaller than the nominal size.
- Use effect heredity to rule out many *incompatible* models in model search.
- Frequentist version by Hamada-Wu (1992); Bayesian version by Chipman (1996)
- Effective if the number of significant interactions is small.

Analysis Results

- Cast Fatigue Experiment:

Main effect analysis:	F	(R^2=0.45)
	F, D	(R^2=0.59)
HW analysis:	F, FG	(R^2=0.89)
	F, FG, D	(R^2=0.92)

- Blood Glucose Experiment (3-level factors):

Main effect analysis:	E_q, F_q	(R^2=0.36)
HW analysis:	B_l, $(BH)_{lq}$, $(BH)_{qq}$	(R^2=0.89)

- Bayesian analysis also identifies B_l, $(BH)_{ll}$, $(BH)_{lq}$, $(BH)_{qq}$ as having the highest posterior model probability.

A Fresh Look at Effect Aliasing

- The two-factor interactions (2fi's) AB and CD are said to be aliased (Finney, 1945) because they represent the *same* contrast (same column in matrix); mathematically similar to *confounding* between treatment and block effects (Yates, 1937).
- Example: a 2^{4-1} design with I = ABCD,

  generated by Col D=(Col A)(Col B)(Col C).

A	B	C	D	AB	= CD
-	-	-	-	+	+
-	-	+	+	+	+
-	+	-	+	-	-
-	+	+	-	-	-
+	-	-	+	-	-
+	-	+	-	-	-
+	+	-	-	+	+
+	+	+	+	+	+

De-aliasing of Aliased Effects

- The pair of effects cannot be disentangled, and are thus *not estimable*. They are said to be fully aliased.
- Can they be de-aliased *without* adding runs??
- *Hint*: an interaction, say AB, should be viewed together with its parent effects A and B.
- Approach: view AB as part of the 3d space of A, B, AB; similarly for C, D, CD; because AB=CD, joint space has 5 dimensions, not 6; then reparametrize each 3d space.

Two-factor Interaction via Conditional Main Effects

- Define the conditional main effect of A given B at level +: $ME(A|B+) = \bar{y}(A+|B+) - \bar{y}(A-|B+)$

  similarly, $ME(A|B-) = \bar{y}(A+|B-) - \bar{y}(A-|B-)$.
- Then $AB = [ME(A|B+) - ME(A|B-)]/2$
- View the conditional main effects $ME(A|B+)$, $ME(A|B-)$

  as *interaction components.*

De-aliasing via CME Analysis

- Reparametrize the 3d space as A, B|A+, B|A-; the three effects are orthogonal but not of same length; similarly, we have C, D|C+, D|C-; in the joint 5d space, some effects are not orthogonal ➡ some conditional main effects (CME) can be estimated via variable selection, call this the *CME Analysis*.
- *Non-orthogonality* is the saving grace ☺.
- Potential applications to social and medical studies which tend to have fewer factors.

Matrix Representation

- For the 2^{4-1}design with I = ABCD

A	B	C	D	B|A+	B|A-	D|C+	D|C-
-	-	-	-	0	-	0	-
-	-	+	+	0	-	+	0
-	+	-	+	0	+	0	+
-	+	+	-	0	+	-	0
+	-	-	+	-	0	0	+
+	-	+	-	-	0	-	0
+	+	-	-	+	0	0	-
+	+	+	+	+	0	+	0

Car marriage station simulation experiment (GM, Canada, 1988)

- A 2_{IV}^{6-2} expt with I=ABCE=ABDF=CDEF,
  note: all 2fi's are fully aliased, i.e., not estimable
- Six factors:

  A: no of lanes in brake cell (3,4)

  B: % of cars with ABS (0, 100)

  C: lane selection logic (FIFO, free flow)

  D: no of Automatic Guided Vehicles (24,34)

  E: % repair in marriage (8, 16)

  F: marriage base cycle time (124, 124+29)
- y: throughput of a simulation run (40 hrs)

Data

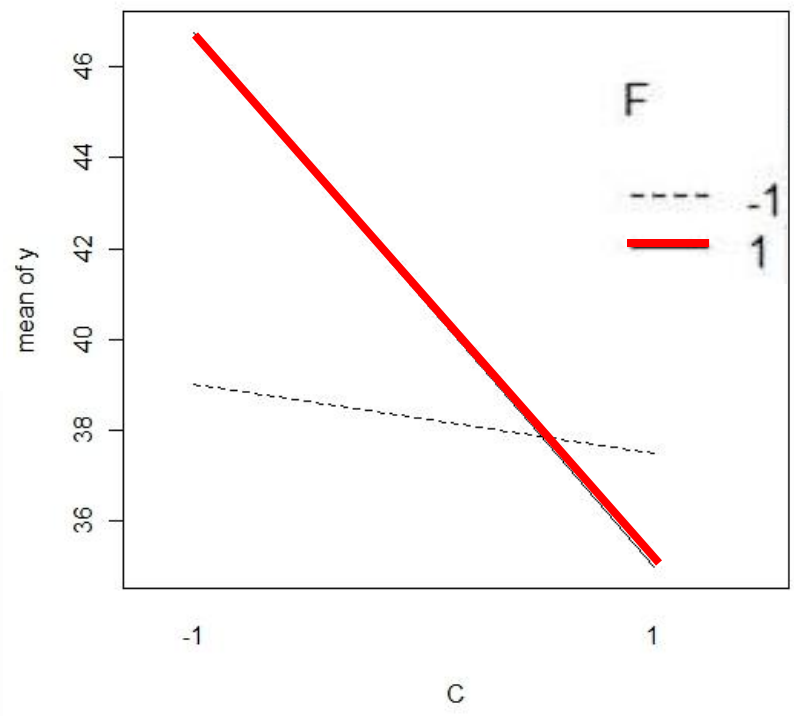
Factors						y
A	B	C	D	E	F	
-	-	-	-	-	-	13
+	+	-	-	-	-	5
-	-	+	-	+	-	69
-	-	-	+	-	+	16
+	-	+	+	-	-	5
+	-	+	-	-	+	7
+	-	-	+	+	-	69
+	-	-	-	+	+	69
-	+	+	+	-	-	9
-	+	+	-	-	+	11
-	+	-	+	+	-	69
-	+	-	-	+	+	89
+	+	+	-	+	-	67
+	+	-	+	-	+	13
-	-	+	+	+	+	66
+	+	+	+	+	+	56

CME vs Standard Analysis

- Model 1(Standard): E(***), C(0.018*), A(0.022*), CF(0.056); R^2=98.29%, C_p=5.0
- Model 2(CME): E(***), C|F+(0.0037**), A(0.017*); R^2=98.26%, C_p=4.0
- Model 3(CME): E(***), C|F+(0.003**), A(0.014*), D(0.15); R^2=98.57%, C_p=5.0
- Model 2 $\succ$ Model 3 $\succ$ Model 1

Interpretation of C|F+

- Lane selection C has a significant effect for larger cycle time F+, a more subtle effect than the obvious effect of E (i.e., % repair affects throughput).



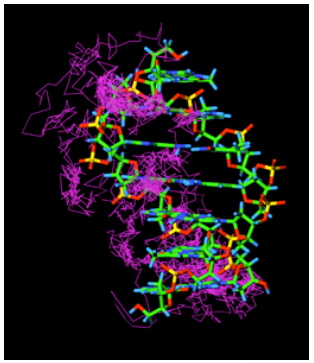
From Physical to Virtual (Computer) Experiments

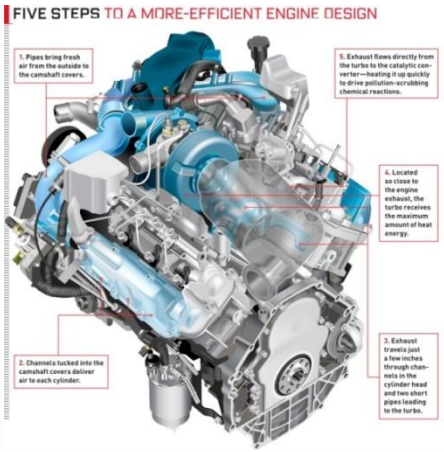
Chemical &Biology: nanoparticle and Polymer synthesis…

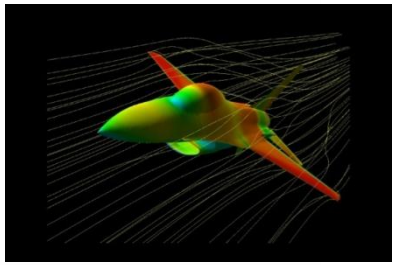
Mechanical: machining, material…

Computer Experiments /Simulations

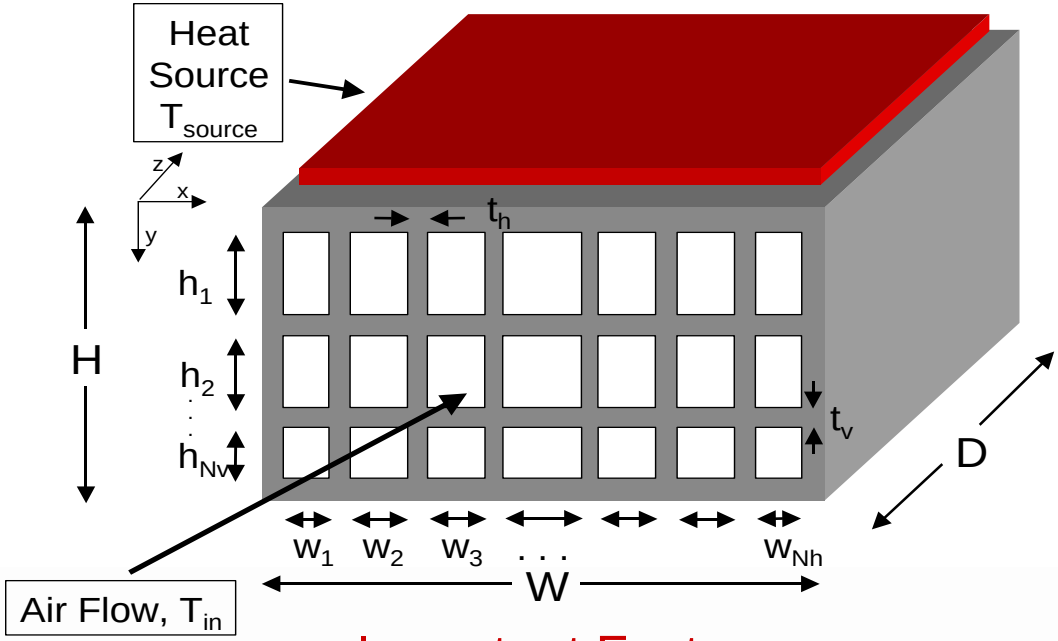
Aerospace: Aircraft design, dynamics…







Example of Computer Simulation: Designing Cellular Heat Exchangers



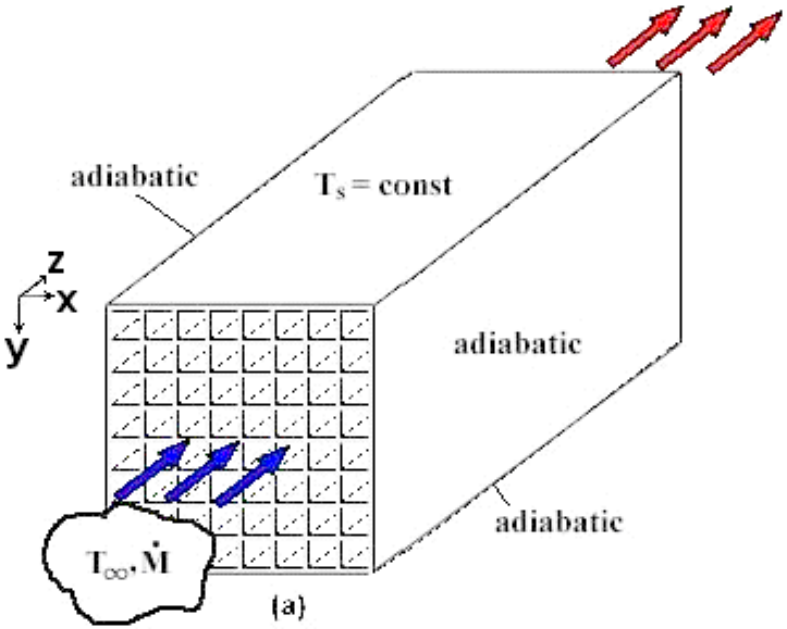
Important Factors

- Cell Topologies, Dimensions, and Wall Thicknesses
- Temperatures of Air Flow and Heat Source
- Conductivity of Solid
- Total Mass Flowrate of Air

Response

- Maximum Total Heat Transfer

Heat Transfer Analysis



ASSUMPTIONS

–Forced Convection

–Laminar Flow: Re < 2300

–Fully Developed Flow

–Three Adiabatic (Insulated) Sides

–Constant Temperature Heat Source on Top

–Fluid enters with Uniform Temp

–Flowrate divided among cells

GOVERNING EQUATIONS

$$Q_c = k_s A_c \frac{\Delta T}{\Delta x} = A_c q_c \quad (\text{in walls})$$

$$Q_h = hA_h \Delta T = A_h\ q_h \quad (\text{convection from walls to fluid})$$

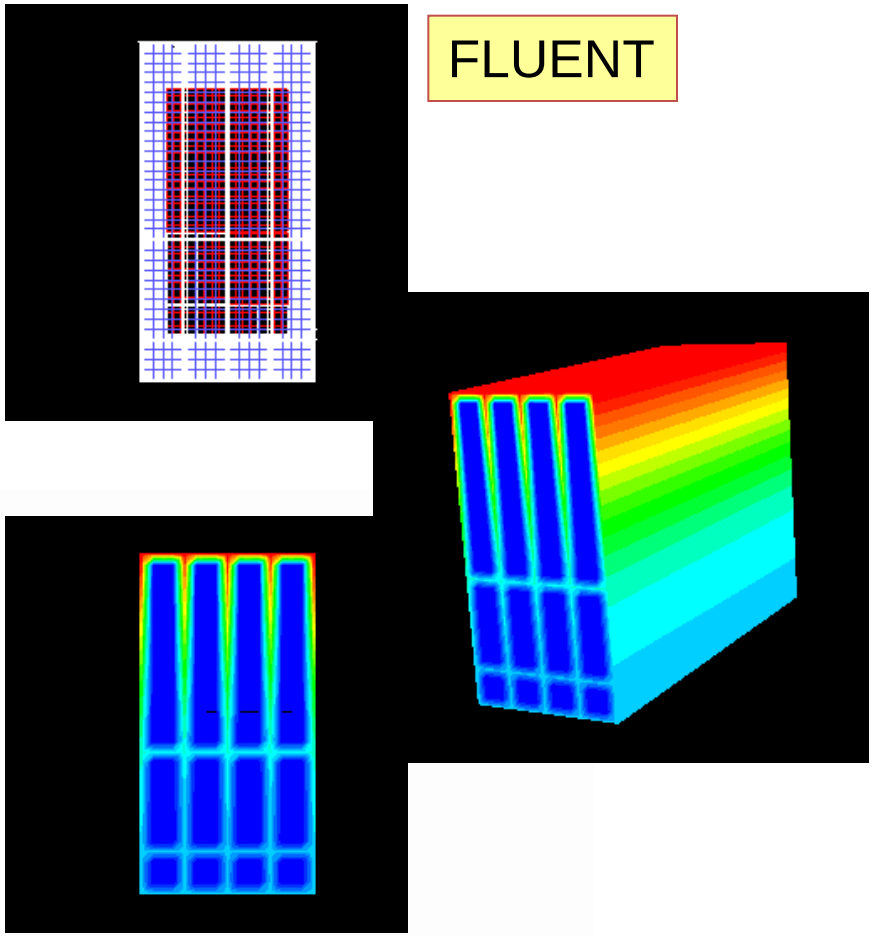
$$Q_f = \dot{m} c_p \Delta T \quad (\text{fluid heating})$$

***B. Dempsey, D.L. McDowell**
ME, Georgia Tech

Heat Transfer Analysis

A Detailed Simulation Approach--FLUENT

- FLUENT solves fluid flow and heat transfer problems with a computational fluid dynamics (CFD) solver.
- Problem domain is divided into thousands or millions of elements.
- Each simulation requires hours to days of computer time on a Pentium 4 PC.



Why Computer Experiments?

- Physical experiments can be time-consuming, costly or infeasible (e.g., car design, traffic flow, forest fire).
- Because of advances in numerical modeling and computing speed, computer modeling is commonly used in many investigations.
- A challenge: Fisher's principles not applicable to deterministic (or even stochastic) simulations. Call for new principles!
- Two major approaches to modeling computer expts:
  - stochastic modeling, primarily the kriging approach,
  - numerical modeling.

Gaussian Process (Kriging) Modeling

- Gaussian process:

$$y(x) \sim GP(\mu(x), \sigma^2 \phi(.)).$$

— $\mu(x)$: mean, linear model $\mu(x) = f(x)'\beta$,

— $\phi(.)$: *positive* correlation function, $\phi(x_1 - x_2, \theta)$, e.g., Gaussian correlation function,

$$\phi(x_1 - x_2, \theta) = \exp\left(-\sum_{i=1}^{p} \theta_i (x_{1,i} - x_{2,i})^2\right), \theta \in \mathbb{R}_+^p$$

— σ^2: variance.

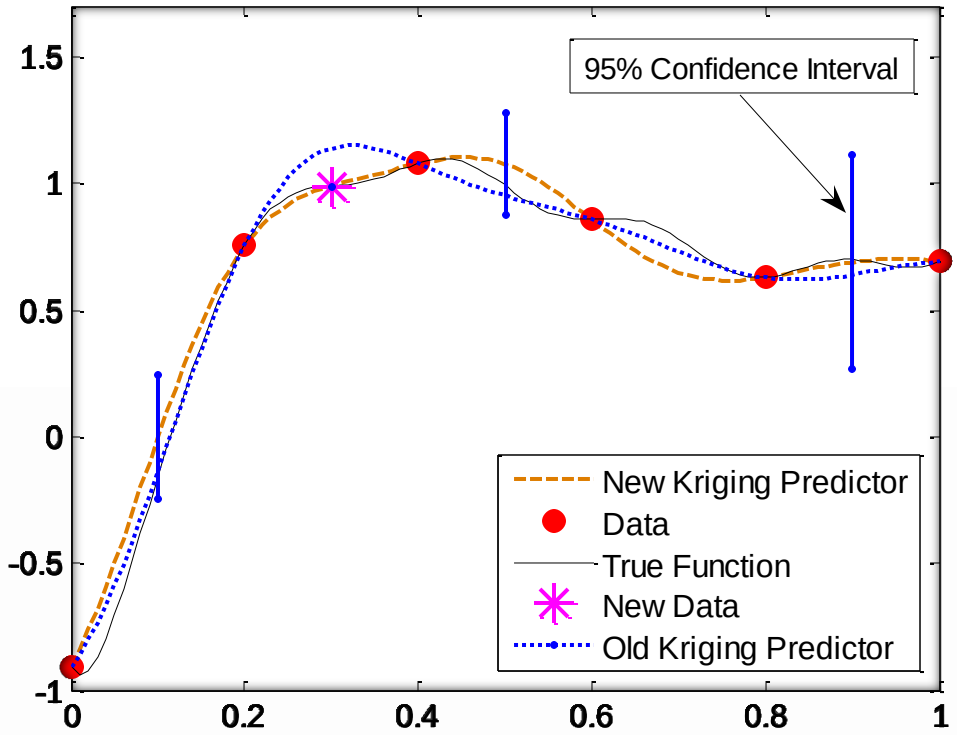
Kriging Predictor

- Best Linear Unbiased Predictor (BLUP):

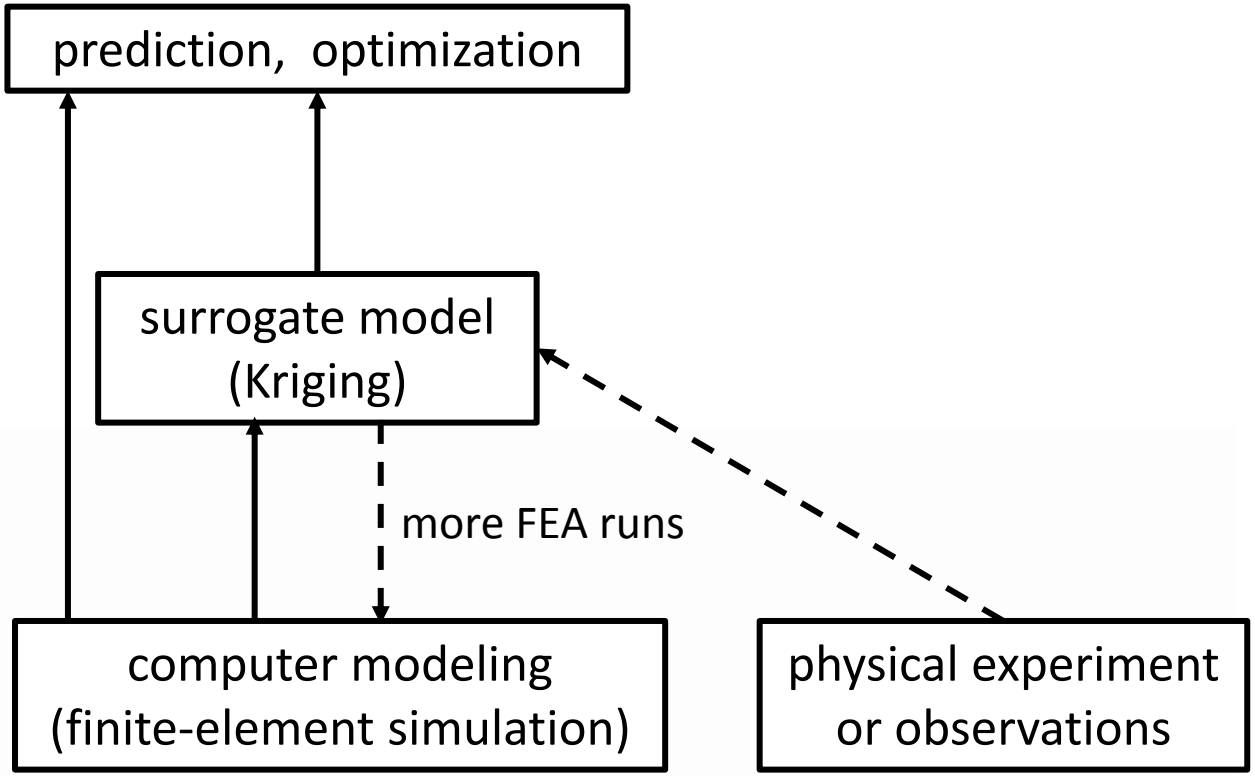
$$\hat{y}(x) = f(x)'\hat{\beta} + r(x)'R^{-1}\left(y - F\hat{\beta}\right),$$

  - $\hat{\beta} = (F'R^{-1}F)^{-1}F'R^{-1}y$ is the generalized least squares estimation, $\mathrm{F} = n \times k$ model matrix.
  - R is the correlation matrix, $R_{i,j} = [\phi\left(x_i - x_j, \theta\right)]$.
  - $r(x) = \left(\phi(x - x_1, \theta), \ldots, \phi(x - x_n, \theta)\right)' =$ vector of correlation between prediction points and observed points.
  - $\hat{y}(x_i) = y_i$, *interpolating* property.

Kriging as Interpolator and Predictor



Statistical Surrogate Modeling of Computer Experiments



More on Kriging

- Matheron (1963) named "kriging" after the mining engineer D. G. Krige (1950). Earlier work by Wiener and Kolmogorov. Widely used in spatial statistics. Use of kriging for computer expts came from Sacks, Welch, Ylvisacker, etc. in the 80's.
- Main *differences* from spatial statistics:
  - Interpolation is a desirable property in deterministic simulation.
  - Design of experiments is a more acute issue.
  - Emphasis on variable selection and response optimization.
- Problem with kriging: the $n \times n$ matrix R (n=# of x_i) is prone to ill-conditioning for large n.

Interplay Between Design and Modeling

- Computer simulations with different levels of accuracy (Kennedy-O'Hagan, 2000; Qian et al., 2006; Qian-Wu, 2008) ➡

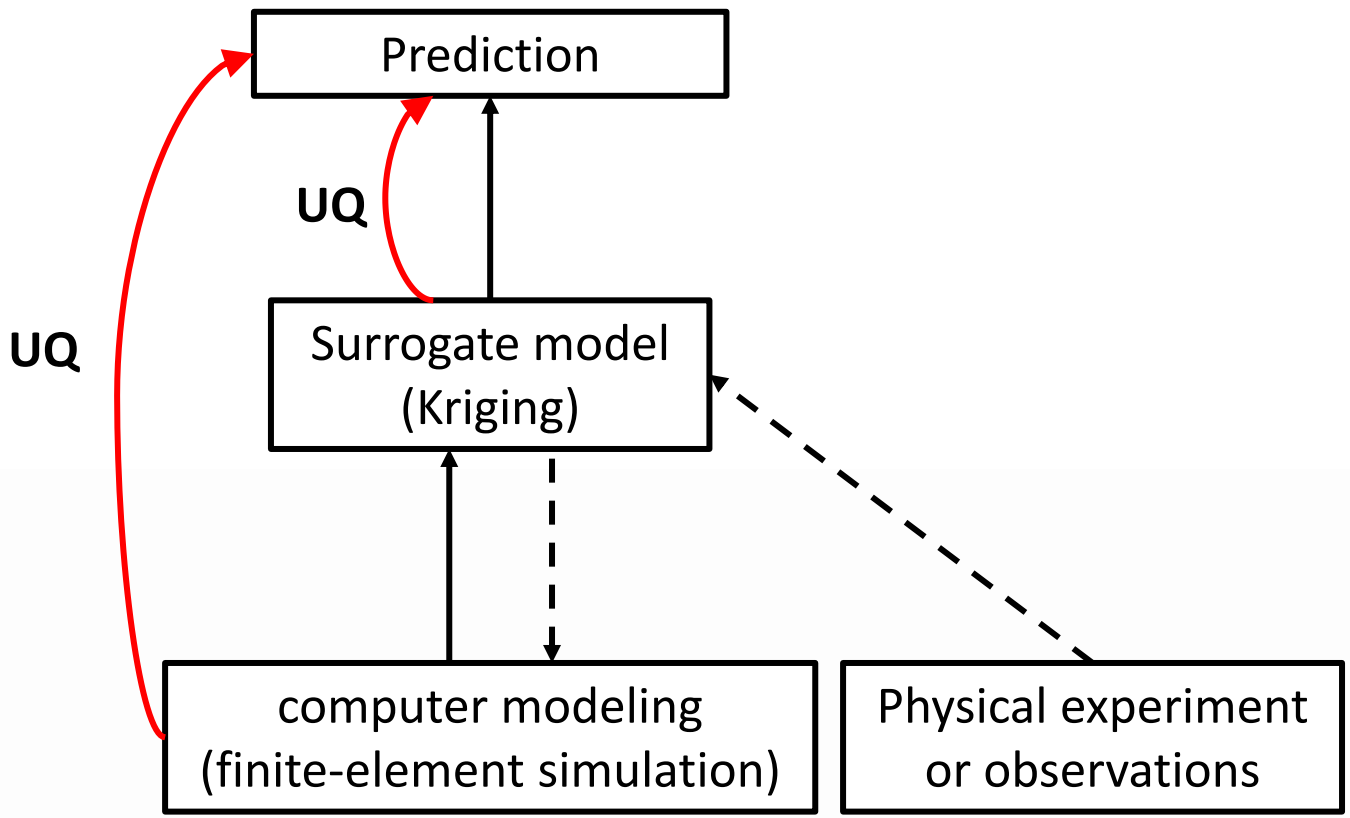
  construction of nested space-filling (e.g., Latin hypercube) designs (Qian-Ai-Wu, 2009, various papers by Qian and others, 2009-date).
- GP model with quantitative and qualitative factors (Qian-Wu-Wu, 2008, Han et al., 2009) ➡

  construction of sliced space-filling (e.g., Latin hypercube) designs (Qian-Wu, 2009, Qian, 2010).

Numerical Approach

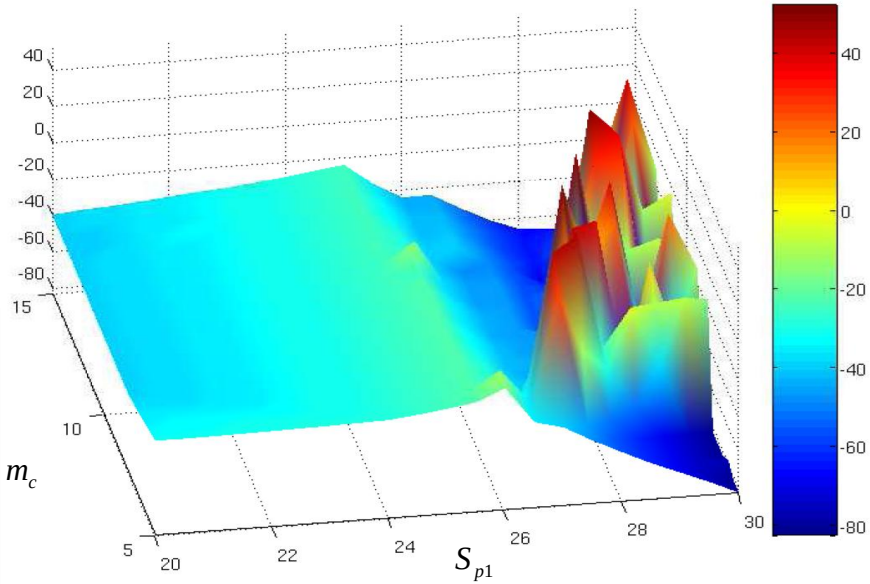
- Can provide faster and more stable computation, and fit non-stationary surface with proper choice of basis functions.
- Some have inferential capability: **R**adial **B**asis interpolating **F**unctions (closely related to kriging), smoothing splines (Bayesian interpretation).
- Others do *not*: MARS, Neural networks, regression-based inverse distance weighting interpolator (var est, but no distribution), sparse representation from overcomplete dictionary of functions. Need to impose a *stochastic structure* to do Uncertainty Quantification. One approach discussed next.

Uncertainty Quantification



Response Surface for Bistable Laser Diodes

- The true surface over a pre-specified grid over pump rate S_{p1} and modulation current m_c:



- Response value (vertical) is the Lyapunov exponent in dynamic system, positive Lyapunov exponents (PLEs) are in red. Computations of PLEs are time-consuming.

Scientific Objectives in Laser Diode Problem

- Each PLE corresponds to a chaotic light output, which can accommodate a secure optical communication channel; finding more PLEs would allow more secure communication channels.
- Objectives: Search all possible PLE (red area) and obtain predicted values for PLEs.
- A numerical approach called OBSM (next slide) can do this. *Question*: how to attach error limits to the predicted values?

Overcomplete Basis Surrogate Model

- Use an overcomplete dictionary of basis functions, *no* unknown parameters in basis functions.
- Use linear combinations of basis functions to approximate unknown functions; linear coefficients are the only unknown parameters.
- Use Matching Pursuit to identify nonzero coefficients; for fast and greedy computations.
- Choice of basis functions to "mimic" the shape of the surface. Can handle nonstationarity.

Chen, Wang, and Wu (2010)

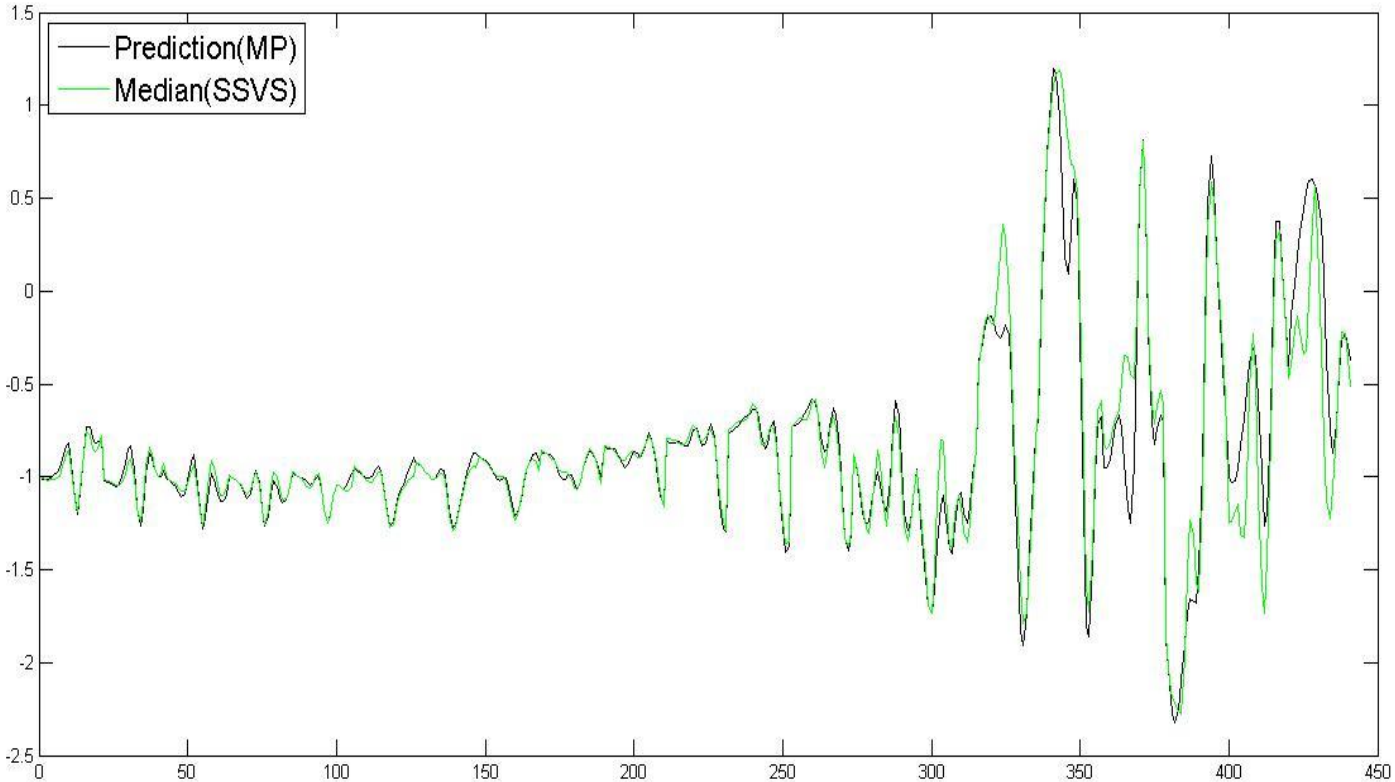
Imposing a Stochastic Structure

- Original numerical model:

$$f(x) = \sum_{j=1}^{M} c_j \phi_j(x), \text{ for large } M.$$

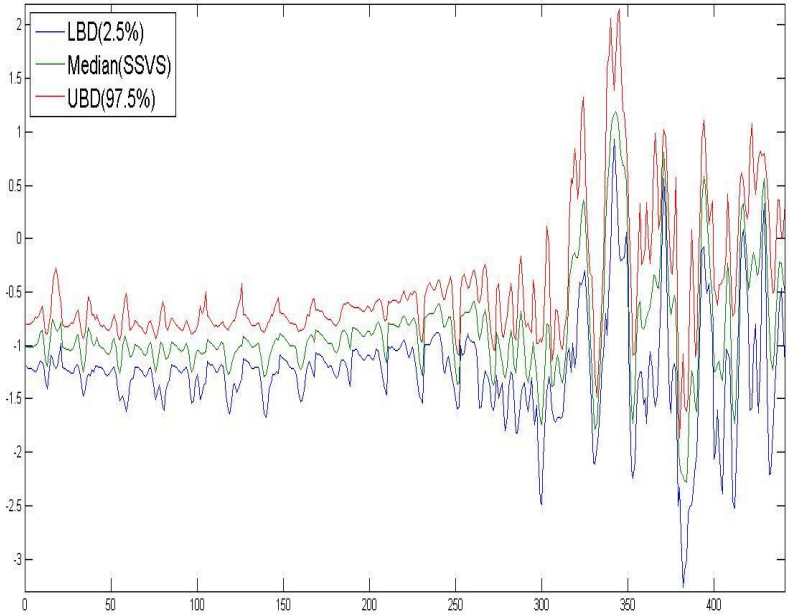
- Impose a Bayesian prior on the coefficients c_j, i.e., $c_j \sim N(0, \sigma^2)$, $j = 1, \cdots, M$; since # of $\{c_j \neq 0\} \ll M$, i.e., effect sparsity ☺, we need a *collapsing* scheme in variable selection.
- Bayesian variable selection: choose an appropriate prior for c_j to represent effect sparsity, e.g., Stochastic Search Variable Selection (George and McCulloch 1993).
- Use MCMC to compute posteriors for $\{c_j\}$ and to select significant coefficients. MCMC outputs can be used for prediction inference, etc.

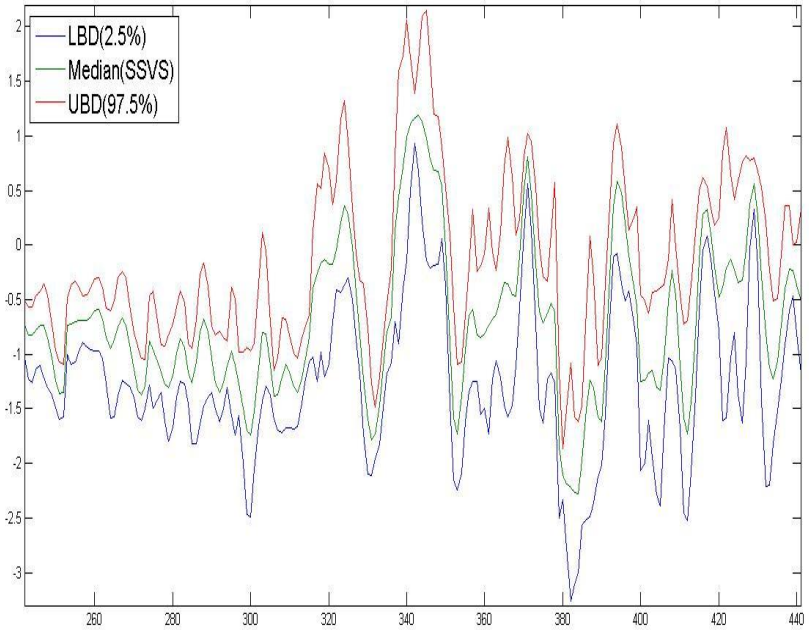
Simulation Results I



Comparison between MP and SSVS

Simulation Results II





- Left figure shows the medians and credible intervals for prediction points.
- Right figure gives a detailed plot for the last 200 points.

Summary Remarks

- Fisher's influence continued from agricultural expts to industrial expts; motivated by the latter, new concepts (e.g., hierarchy, sparsity, heredity) and methodologies (e.g., response surface methodology, parameter design) were developed, which further his legacy.
- Because Fisher's principles are less applicable to virtual experiments, we need new guiding principles.
  - Kriging can have numerical problems; tweaking or new stochastic approach?
  - Numerical approach needs Uncertainty Quantification, a new opportunity between stat and applied math.
  - Design construction distinctly different from physical expts; need to exploit its interplay with modeling.