Introductory Overview Lecture on Computer Experiments - Modeling

Thomas Santner
Ohio State University
Columbus OH

Joint Statistical Meetings
Seattle, WA
August 2006

Outline

1. Experiments

2. Experimentation using Computer Codes

3. A Taxonomy of Problems

4. Gaussian Stochastic Process (GaSP) Models

5. Prediction based on the GSP Model

4. Example

5. Conclusions-Take Home Messages

1. Experiments

A. Physical Experiments

- **Gold** standard for establishing cause and effect relationships
- Mainstay of Agricultural, Industry, Medicine
- Principles of randomization, blocking, choice of sample size, and stochastic modeling of response variables all developed in response to needs of physical experiments

1. Experiments

A. **Physical Experiments**

- **Gold** standard for establishing cause and effect relationships
- Mainstay of Agricultural, Industry, Medicine
- Principles of randomization, blocking, choice of sample size, and stochastic modeling of response variables all developed in response to needs of physical experiment**s**

B. **Simulation Experiments** Complex physical system each of whose parts interact in a known stochastic manner but whose ensemble is not understood analytically. Heavily used in Industrial Engineering--e.g., compare job shop set ups

1. Experiments

A. Physical Experiments

- **Gold** standard for establishing cause and effect relationships
- Mainstay of Agricultural, Industry, Medicine
- Principles of randomization, blocking, choice of sample size, and stochastic modeling of response variables all developed in response to needs of physical experiments

B. Simulation Experiments Complex physical system each of whose parts interact in a known stochastic manner but whose ensemble is not understood analytically. Heavily used in Industrial Engineering--e.g., compare job shop set ups

C. Computer Experiments relatively new (below)

1. Experiments

A. **Physical Experiments**

- **Gold** standard for establishing cause and effect relationships
- Mainstay of Agricultural, Industry, Medicine
- Principles of randomization, blocking, choice of sample size, and stochastic modeling of response variables all developed in response to needs of physical experiments

B. **Simulation Experiments** Complex physical system each of whose parts interact in a known stochastic manner but whose ensemble is not understood analytically. Heavily used in Industrial Engineering--e.g., compare job shop set ups

C. **Computer Experiments** relatively new (below)

D. **Combinations** of **A** to **C**

2. Experimentation using Computer Codes

• In some situations performing a physical experiment is not feasible

1. Physical process is technically too difficult to study
2. Number of variables is too large
3. Too expensive to study directly (it's all money)
4. Ethical considerations

2. Experimentation using Computer Codes

- In some situations performing a physical experiment is not feasible
  1. Physical process is technically too difficult to study
  2. Number of variables is too large
  3. Too expensive to study directly
  4. Ethical considerations
- When physical experiments are not possible, it may still be feasible to conduct a **computer experiment**

2. Experimentation using Computer Codes

• In some situations performing a physical experiment is not feasible

1. Physical process is technically too difficult to study
2. Number of variables is too large
3. Too expensive to study directly
4. Ethical considerations

• When physical experiments are not possible, it may still be feasible to conduct a **computer experiment**

• ***IF*** the physical process relating the inputs $\boldsymbol{x}$ to the response(s)

a. Can be described by a mathematical model, $y(\boldsymbol{x})$, relating the output to $\boldsymbol{x}$

b. Numerical methods exist for solving the mathematical model

c. The numerical methods can be implemented with computer code (in finite time!)

THEN one can run the computer code to produce one or more "responses" $y(\boldsymbol{x})$ at any input $\boldsymbol{x} \in \boldsymbol{\mathcal{X}} \subset \mathbb{R}^d$, i.e., one can conduct a **computer experiment**

$$\boldsymbol{x} \longrightarrow \boxed{\textbf{Code}} \longrightarrow y(\boldsymbol{x})$$

The computer code is a **proxy** for the physical process.

Features of Computer Experiments

$$\boldsymbol{x} \longrightarrow \boxed{\textbf{Code}} \longrightarrow y(\boldsymbol{x})$$

- $y(\boldsymbol{x})$ is deterministic

Features of Computer Experiments

$$\boldsymbol{x} \longrightarrow \boxed{\textbf{Code}} \longrightarrow y(\boldsymbol{x})$$

- $y(\boldsymbol{x})$ is deterministic
- Our interest in settings where very few number of computer runs are possible due to
  1. Complex codes (fine-grid FEA codes)
  2. High--dimensional input $\boldsymbol{x}$

Features of Computer Experiments

$$\boldsymbol{x} \longrightarrow \boxed{\textbf{Code}} \longrightarrow y(\boldsymbol{x})$$

- $y(\boldsymbol{x})$ is deterministic

- Our interest in settings where very few number of computer runs are possible due to
  1. Complex codes (fine-grid FEA codes)
  2. High--dimensional input $\boldsymbol{x}$

- Traditional principles used in designing physical experiments to balance the effects of non-identical experimental units (randomization, blocking, etc) are irrelevant.

Examples of Computer Experiments

(1) Design of VLSI circuits

(2) Modeling weather or climate

(3) Determine optimum operating conditions for a compression molding process

(4) Determine the performance of controlled nuclear fusion devices

(5) Describe the **temporal evolution of contained and wild fires**

(6) Design of helicopter rotor blades

(7) Biomechanics – Explain behavior of (or even Design) prosthetic devices

Example **Zone Computer Models** used to predict the evolution of a fire in an **enclosed** room (eg, the computer code **ASET** **=** **A**vailable **S**afe **E**gress **T**ime)

In particular, ASET-B describes the *temporal evolution* of a fire in a single room with closed doors and windows that contains an object at some point below the ceiling that has been ignited.

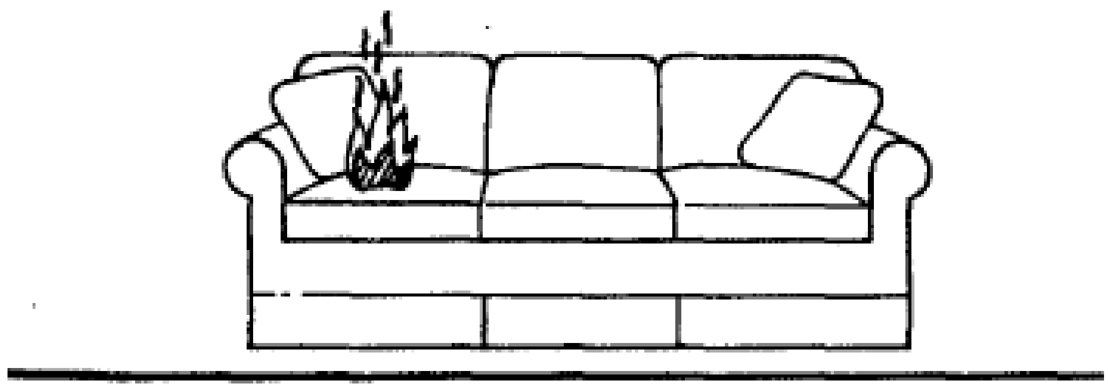


Fig. 3-10.1. Events immediately after ignition.

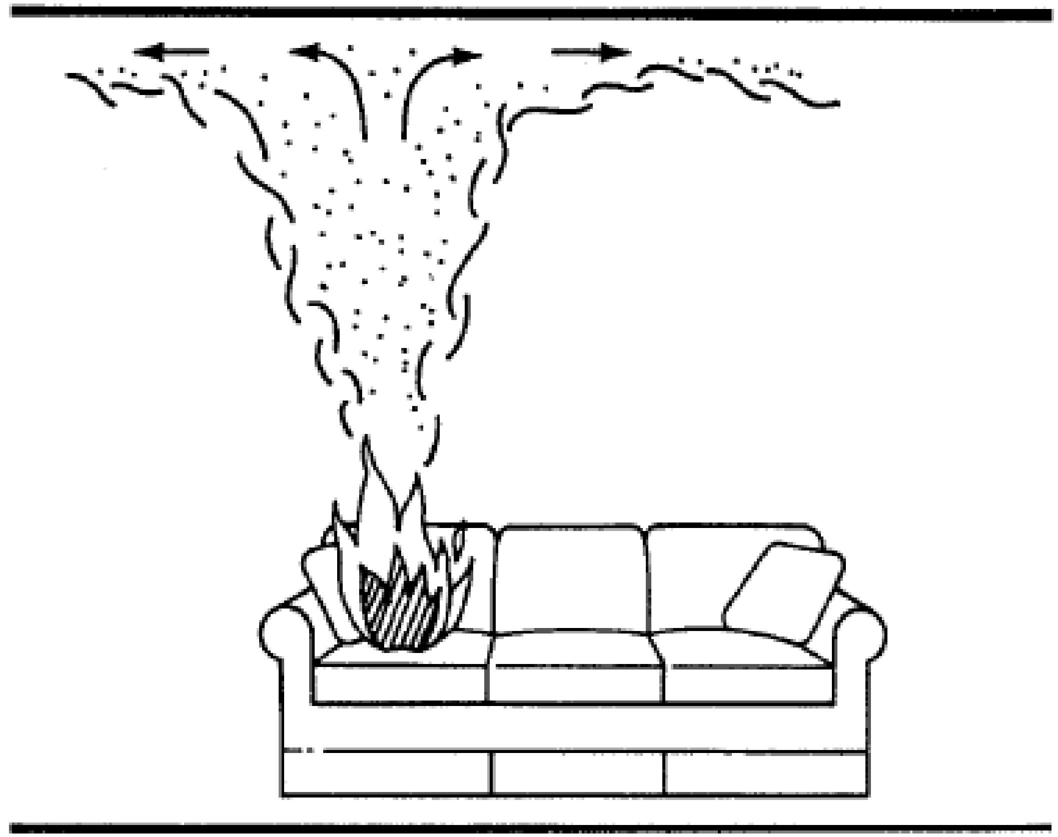


Fig. 3-10.3. The plume-ceiling interaction.

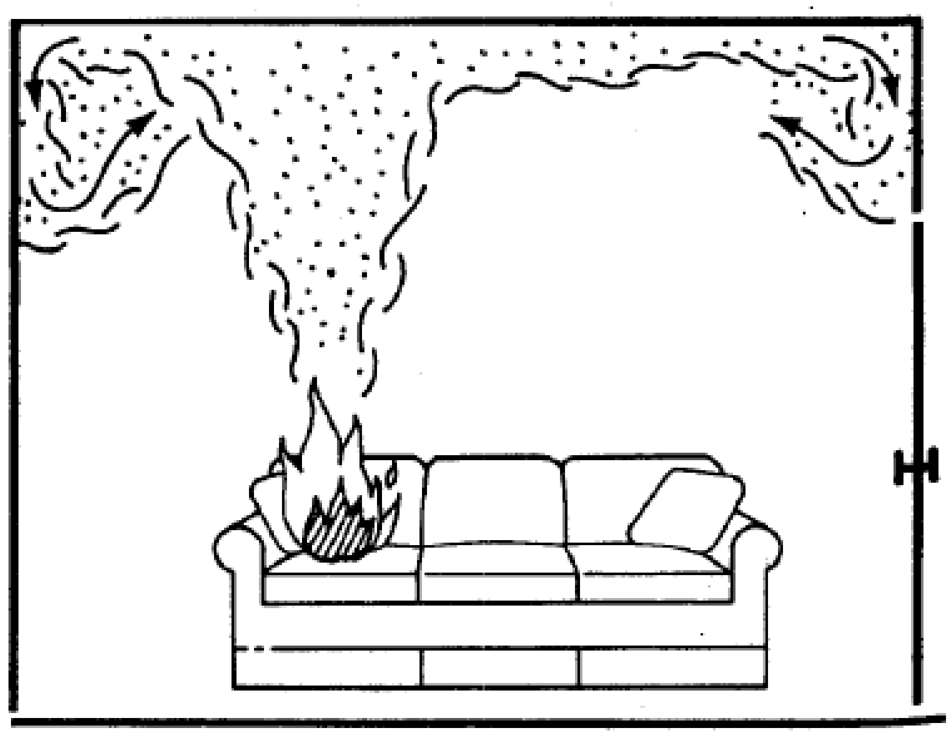


Fig. 3-10.4. Ceiling jet-wall interaction.

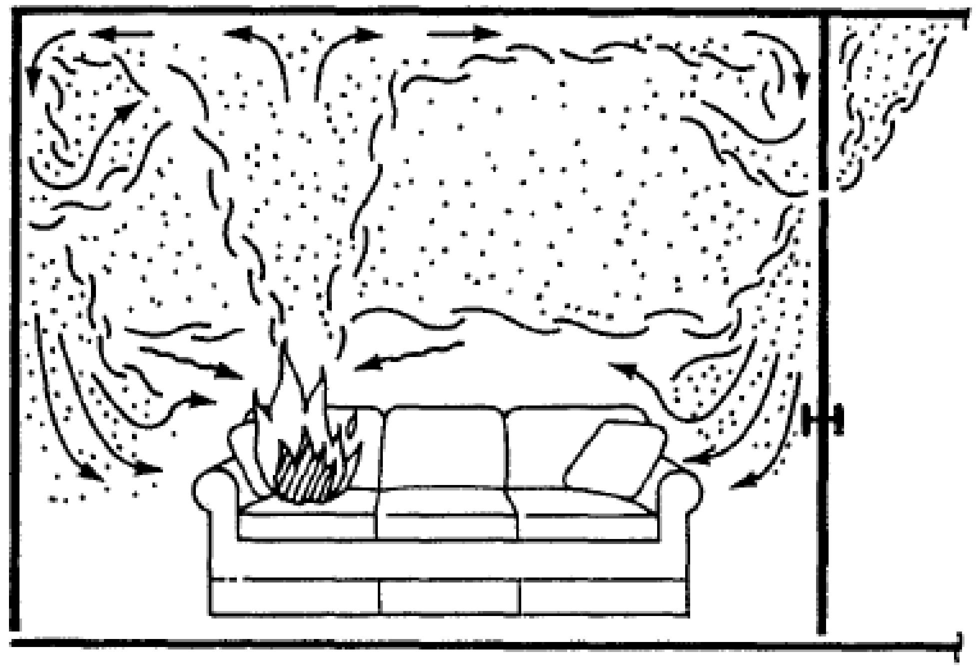
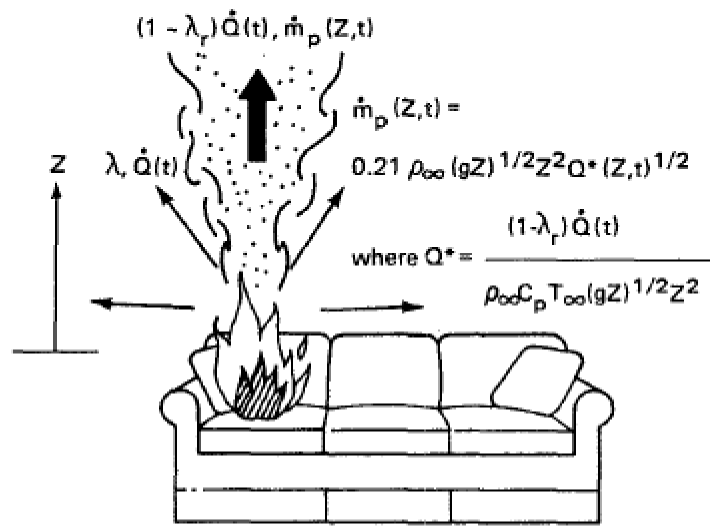


Fig. 3-10.6. Further "smoke filling."

Mathematical Model



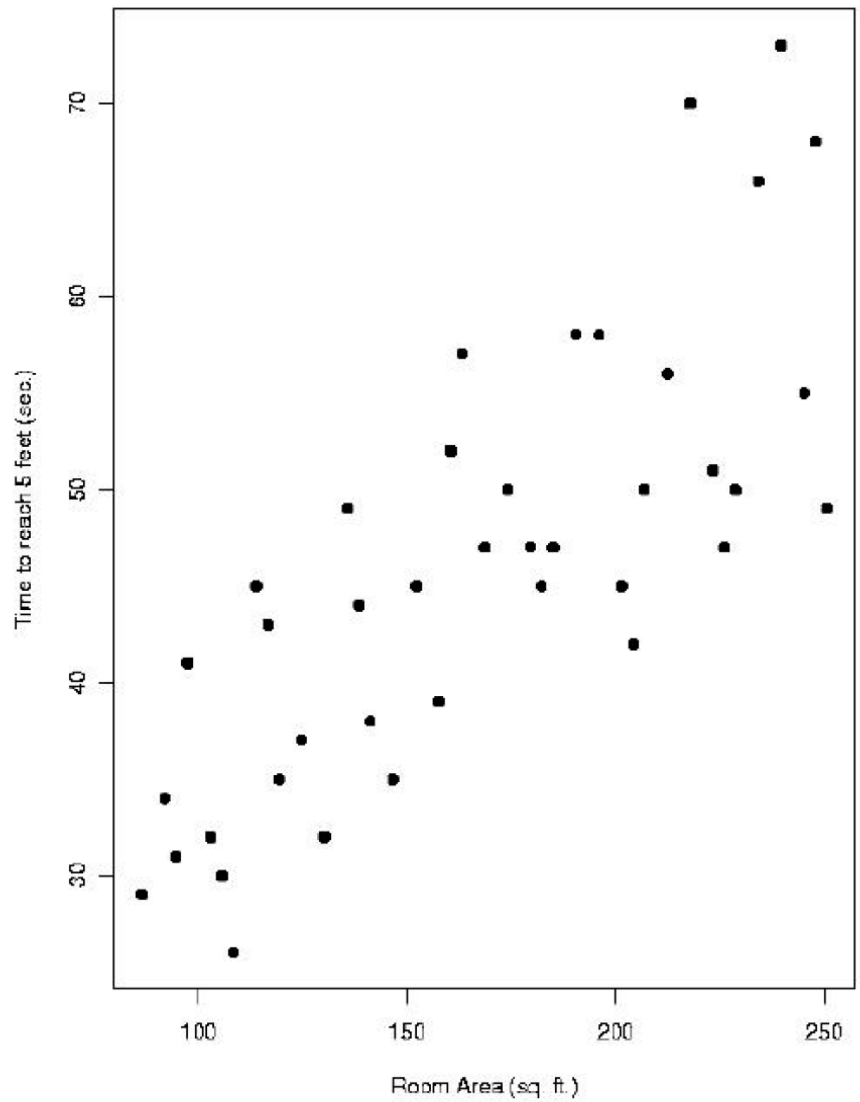
Inputs to ASET-B

- Room ceiling height
- Room floor area
- Height of the fire source (the burning object) above the floor
- Heat loss fraction for the room (which depends on the insulation in the room)
- (etc, material-specific heat release rate)

Computed Response

$$y(\text{RmCeHgt,RmFlArea, FireHgt, HeatLFrac})$$
$$= \text{time required by smoke layer to reach 5 ft above fire source}$$

Objective Predict the time for the fire plume to reach 5 ft above the ground for untried combinations



Features of Computer Experiments

- $y(\boldsymbol{x})$ is deterministic
- Our interest in settings where very few number of computer runs are possible due to
  1. Complex codes (fine-grid FEA codes)
  2. High--dimensional input $\boldsymbol{x}$
- Traditional principles used in designing physical experiments to balance the effects of non-identical experimental units (randomization, blocking, etc) are irrelevant.

- Sometimes output from a **physical experiments** is also available. Usual philosophy physical experiment is a **noisy** measurement of the **true** input-output relationship $(\boldsymbol{x} \longrightarrow \mu^T(\boldsymbol{x}))$. Model

$$Y^p(\boldsymbol{x}) = \mu^T(\boldsymbol{x}) + \epsilon(\boldsymbol{x})$$

where the $\{\epsilon(\boldsymbol{x})\}_{\boldsymbol{x}}$ are independent measurement errors with zero mean and unknown variance (usually, white noise).

Warning

- Sometimes physical experiments are available only for **components** of the ensemble process, eg, code that emulates an auto crash test.
- In other cases, only experiments that **approximate** reality are available, e.g., a knee simulator

3. A Taxonomy of Problems

Setup

1. **Inputs** $\boldsymbol{x} = (\boldsymbol{x}_c, \boldsymbol{x}_e, \boldsymbol{x}_m)$ where

$\boldsymbol{x}_c$ = **control** (**manufacturing, engineering design**) variables

$\boldsymbol{x}_e$ = **noise** (**field**, **enviromental**) variables

$\boldsymbol{x}_m$ = **model** variables

(Not all types of inputs need be present in every application.)

3. A Taxonomy of Problems

Setup

1. Inputs $\boldsymbol{x} = (\boldsymbol{x}_c, \boldsymbol{x}_e, \boldsymbol{x}_m)$ where

$\boldsymbol{x}_c$ = **control (manufacturing, engineering design)** variables

$\boldsymbol{x}_e$ = **noise (field, enviromental)** variables

$\boldsymbol{x}_m$ = **model** variables

(Not all types of inputs need be present in every application.)

2. Outputs

Real-valued : $y(\boldsymbol{x})$ ***or***

Multivariate: $(y_1(\boldsymbol{x}),\ y_2(\boldsymbol{x}), \ldots, y_k(\boldsymbol{x}))$ ***or***

Functional: $(t, y(t, \boldsymbol{x}))$

3. A Taxonomy of Problems

Setup

1. **Inputs** $\boldsymbol{x} = (\boldsymbol{x}_c, \boldsymbol{x}_e, \boldsymbol{x}_m)$ where

$\boldsymbol{x}_c$ = **control** (**manufacturing, engineering design**) variables

$\boldsymbol{x}_e$ = **noise** (**field**, **enviromental**) variables

$\boldsymbol{x}_m$ = **model** variables

(N*ot all types of inputs need be present in every application.)*

2. **Outputs**

Real-valued : $y(\boldsymbol{x})$ ***or***

Multivariate: $(y_1(\boldsymbol{x}),\ y_2(\boldsymbol{x}), \ldots, y_k(\boldsymbol{x}))$ ***or***

Functional: $(t, y(t, \boldsymbol{x}))$

3. $\mathbf{X}_e \sim F(\bullet)$ describes "target field conditions"

$\mathbf{X}_m \sim \pi(\bullet)$ describes prior knowledge about the model specifications

3. A Taxonomy of Problems

Setup

1. Inputs $\boldsymbol{x} = (\boldsymbol{x}_c, \boldsymbol{x}_e, \boldsymbol{x}_m)$ where

$\boldsymbol{x}_c$ = **control (manufacturing, engineering design)** variables

$\boldsymbol{x}_e$ = **noise (field, enviromental)** variables

$\boldsymbol{x}_m$ = **model** variables

(N*ot all types of inputs need be present in every application.)*

2. Outputs

Real-valued : $y(\boldsymbol{x})$ ***or***

Multivariate: $(y_1(\boldsymbol{x}),\ y_2(\boldsymbol{x}), \ldots, y_k(\boldsymbol{x}))$ ***or***

Functional: $(t, y(t, \boldsymbol{x}))$

3. $\mathbf{X}_e \sim F(\bullet)$ describes "target field conditions"

$\mathbf{X}_m \sim \pi(\bullet)$ describes prior knowledge about the model specifications

4. Summary (assuming no $\mathbf{X}_m$) of $y(\boldsymbol{x}_c, \boldsymbol{X}_e)$ distribution

$$\mu(\boldsymbol{x}_c) = E_F\{y(\boldsymbol{x}_c, \boldsymbol{X}_e)\};$$

$$\xi(\boldsymbol{x}_c):\ \mathrm{P}_F\{y(\boldsymbol{x}_c, \boldsymbol{X}_e) \le \xi(\boldsymbol{x}_c)\} = \alpha \text{ (median)};$$

$$\sigma^2(\boldsymbol{x}_c) = \mathrm{Var}_F(y(\boldsymbol{x}_c, \boldsymbol{X}_e))$$

$\boldsymbol{s}$

3. A Taxonomy of Problems

Setup

1. Inputs $\boldsymbol{x} = (\boldsymbol{x}_c, \boldsymbol{x}_e, \boldsymbol{x}_m)$ where

$\boldsymbol{x}_c$ = **control (manufacturing, engineering design)** variables

$\boldsymbol{x}_e$ = **noise (field, enviromental)** variables

$\boldsymbol{x}_m$ = **model** variables

(N*ot all types of inputs need be present in every application.)*

2. Outputs

Real-valued : $y(\boldsymbol{x})$ ***or***

Multivariate: $(y_1(\boldsymbol{x}),\ y_2(\boldsymbol{x}), \ldots, y_k(\boldsymbol{x}))$ ***or***

Functional: $(t, y(t, \boldsymbol{x}))$

3. $\mathbf{X}_e \sim F(\bullet)$ describes "target field conditions"

$\mathbf{X}_m \sim \pi(\bullet)$ describes prior knowledge about the model specifications

4. Summary (assuming no $\mathbf{X}_m$) of $y(\boldsymbol{x}_c, \boldsymbol{X}_e)$ distribution

$$\mu(\boldsymbol{x}_c) = E_F\{y(\boldsymbol{x}_c, \boldsymbol{X}_e)\};$$

$$\xi(\boldsymbol{x}_c):\ \mathrm{P}_F\{y(\boldsymbol{x}_c, \boldsymbol{X}_e) \leq \xi(\boldsymbol{x}_c)\} = \alpha \text{ (median)};$$

$$\sigma^2(\boldsymbol{x}_c) = \mathrm{Var}_F(y(\boldsymbol{x}_c, \boldsymbol{X}_e))$$

5. "Design" of a computer experiment ≡ choice of $\boldsymbol{x}_1^t, \ldots, \boldsymbol{x}_n^t$ at which to evaluate computer code, where, eg, $\boldsymbol{x}_i^t = \left(\boldsymbol{x}_{c,i}^t, \boldsymbol{x}_{e,i}^t\right),\ i = 1, \ldots, n$

3. Taxonomy of Problems

Problem 1 **Interpolation/Emulation** – Given computer code output at a set of training inputs,

$$\left(\boldsymbol{x}_1^t,\, y(\boldsymbol{x}_1^t)\right), \ldots, \left(\boldsymbol{x}_n^t,\, y(\boldsymbol{x}_n^t)\right)$$

predict the response at a new input $\boldsymbol{x}_0$ (predictor $\equiv$ **metamodel**)

3. Taxonomy of Problems

Problem 1 **Interpolation/Emulation** – Given computer code output at a set of "training" inputs,

$$\left(\boldsymbol{x}_1^t,\, y(\boldsymbol{x}_1^t)\right),\, \ldots,\, \left(\boldsymbol{x}_n^t,\, y(\boldsymbol{x}_n^t)\right)$$

predict the response at a new input $\boldsymbol{x}_0$ (predictor $\equiv$ **metamodel**)

Problem 2 **Experimental design** – Determine a set of inputs at which to carry out the sequence of code runs (a "good" design of a physical or computer experiment depends on the **scientific objective** of the research)

- Exploratory Designs ("space-filling")
- Prediction-based Designs
- **Optimization-based Designs** (e.g., find $\boldsymbol{x}_c^{opt} = \text{argmin } \boldsymbol{y}(\boldsymbol{x})$)

3. Taxonomy of Problems

Problem 1 **Interpolation/Emulation** – Given computer code output at a set of "training" inputs,

$$\left(\boldsymbol{x}_1^t,\, y(\boldsymbol{x}_1^t)\right),\, \ldots,\, \left(\boldsymbol{x}_n^t,\, y(\boldsymbol{x}_n^t)\right)$$

predict the response at a new input $\boldsymbol{x}_0$ (predictor $\equiv$ **metamodel**)

Problem 2 **Experimental design** – Determine a set of inputs at which to carry out the sequence of code runs (a "good" design of a physical or computer experiment depends on the **scientific objective** of the research)

- **Exploratory Designs** ("space-filling")
- **Prediction-based Designs**
- **Optimization-based Designs** (e.g., find $\boldsymbol{x}_c^{opt} = \text{argmin } \boldsymbol{y}(\boldsymbol{x})$)

Problem 3 **Uncertainty/Output Analysis** – Determine the distribution of the random variable $y(\boldsymbol{x}_c, \boldsymbol{X}_e)$. (Determine the variability in the performance measure $y(\bullet)$ for design $\boldsymbol{x_c}$ when applied to the population defined by the distribution of $\boldsymbol{X}_e$, eg, patient specific variables (patient weight or bone material properties) or surgeon specific variables (measuring surgical skill)

Example In his Cornell PhD thesis, Kevin Ong studied the effect of **Surgical, Patient, and Fluid Effects** on the Stability of Uncemented Acetabular Components

3. Taxonomy of Problems

Problem 1 **Interpolation/Emulation**

Problem 2 **Experimental design**

Problem 3 **Uncertainty/Output Analysis**

Problem 4 **Sensitivity Analysis** – Determine how variation in $y(\boldsymbol{x})$ can be apportioned to the different inputs of $\boldsymbol{x}$ (which inputs is $y(\boldsymbol{x})$ not sensitive to? which ones is $y(\boldsymbol{x})$ most sensitive to?)

Philosophy Inputs that have relatively little effect on the output can be set to some nominal value; additional investigation can be restricted to determining how the output depends on the active inputs

3. Taxonomy of Problems

Problem 1 **Interpolation/Emulation**

Problem 2 **Experimental design**

Problem 3 **Uncertainty/Output Analysis**

Problem 4 **Sensitivity Analysis**

Problem 5 **Calibrate the computer code** – Use outputs from a physical experiment represented by the computer code to set the computer code **calibration variables** (or to update the uncertainty regarding these parameters)

Example Set FEA Mesh Density = ?, Load Discretization = ?, etc

3. Taxonomy of Problems

Problem 1 **Interpolation/Emulation**

Problem 2 **Experimental design**

Problem 3 **Uncertainty/Output Analysis**

Problem 4 **Sensitivity Analysis**

Problem 5 **Calibrate the computer code**

Problem 6 **Prediction** – Using the calibrated simulator to give predictions (including uncertainty bounds) for an associated physical system.

3. Taxonomy of Problems

Problem 1 **Interpolation/Emulation**

Problem 2 **Experimental design**

Problem 3 **Uncertainty/Output Analysis**

Problem 4 **Sensitivity Analysis**

Problem 5 **Calibration**

Problem 6 **Prediction**

Problem 7 **Find Robust Inputs** – In experiments with engineering design and patient-specific, enviromental variables, determine robust choices of the engineering design variables. If

$$\mu(\boldsymbol{x_c}) = E_F\{y(\boldsymbol{x_c}, \boldsymbol{X}_e)\}$$

then a **robust set of inputs** $\boldsymbol{x_c}$ is an engineering "design" whose output is **minimally sensitive** to the assumed distribution $F(\bullet)$ of $\boldsymbol{X}_e$

Bottom Line Many of the problems above have "natural" solutions obtained by approximating $y(\boldsymbol{x_c}, \boldsymbol{x}_e)$ by a fast (linear in the training data) predictor, a metamodel

4. Gaussian Stochastic Process (GaSP) Models
(used as basis for both prediction and some design choices)

Idea Regard $y(\boldsymbol{x})$ as a realization, a "draw," of a random function $Y(\boldsymbol{x})$

The simplest possible (prior) model for $Y(\boldsymbol{x})$ is

$$Y(\boldsymbol{x}) = \underbrace{\sum_j \beta_j f_j(\boldsymbol{x})}_{\text{"large scale trends"}} \quad + \quad \underbrace{Z(\boldsymbol{x})}_{\text{"smooth deviations"}}$$

$$= \beta^\top f(\boldsymbol{x}) + Z(\boldsymbol{x})$$

where

$f_1(\boldsymbol{x}), \ldots, f_k(\boldsymbol{x})$ are *known* regression functions,
$\boldsymbol{\beta}$ is an unknown regression vector, *and*
$Z(\boldsymbol{x})$ is a stationary Gaussian Stochastic Process (GaSP)

- $Z(\boldsymbol{x})$, $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ satisfies
  - E$\{Z(\boldsymbol{x})\} = 0$ (zero mean) ($\Rightarrow E\{Y(\boldsymbol{x})\} = \boldsymbol{\beta}^\top f(\boldsymbol{x}) + 0 = \boldsymbol{\beta}^\top f(\boldsymbol{x})$)
  - Var$(Z(\boldsymbol{x})) = \sigma_Z^2$
  - Correlation Function: symmetric $R(\bullet)$ with $R(0) = 1$,

$$\text{Cov}(Z(\boldsymbol{x}_1), Z(\boldsymbol{x}_2)) = \sigma_Z^2 \times R(\boldsymbol{x}_1 - \boldsymbol{x}_2)$$

  - Typically $R(\bullet) = R(\bullet|\boldsymbol{\xi})$ is a function of a finite number of *unknown* parameters
  - GaSP: For any $\boldsymbol{x}_{1,\ldots,}\boldsymbol{x}_s$, $(Z(\boldsymbol{x}_1),\ldots,Z(\boldsymbol{x}_s))$ has the multivariate normal distribution

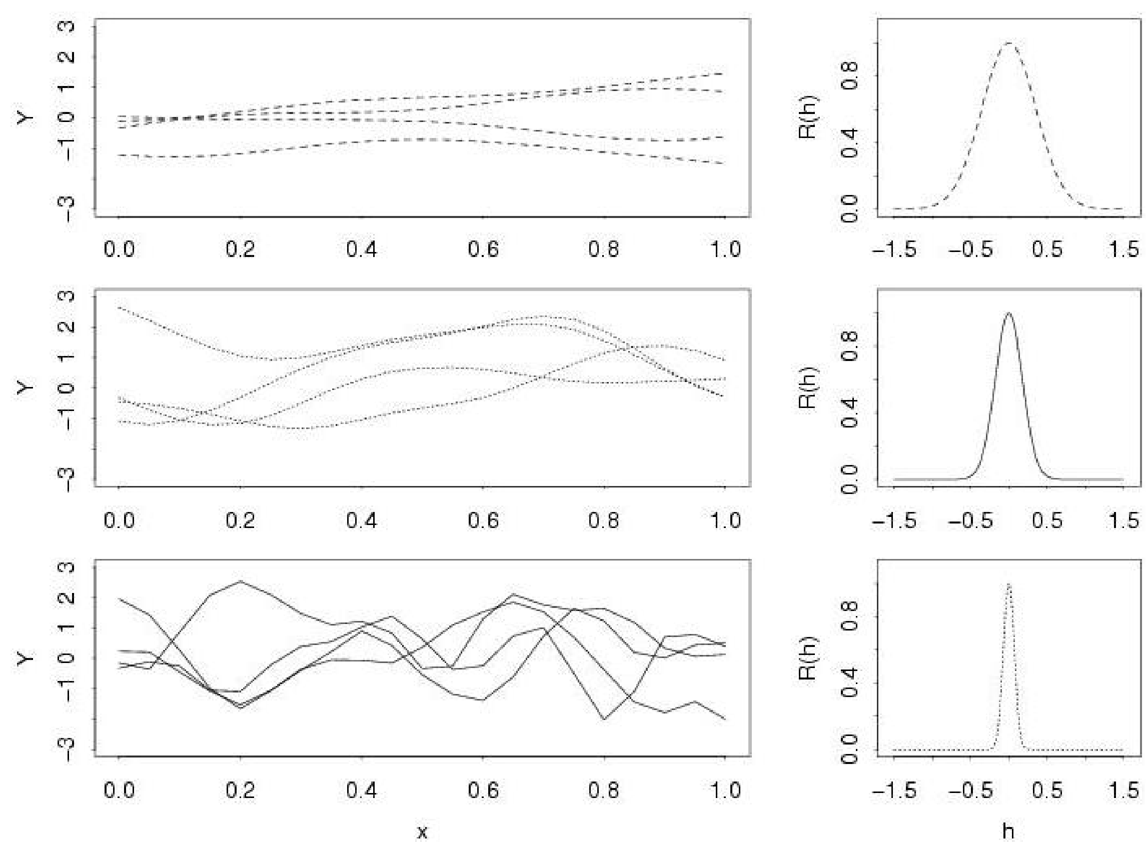
- Usually, taking $\boldsymbol{\beta}^\top f(\boldsymbol{x}) = \beta_0$ with a data-selected parametric correlation function R$(\bullet|\xi)$

4. Gaussian Stochastic Process (GaSP) Models

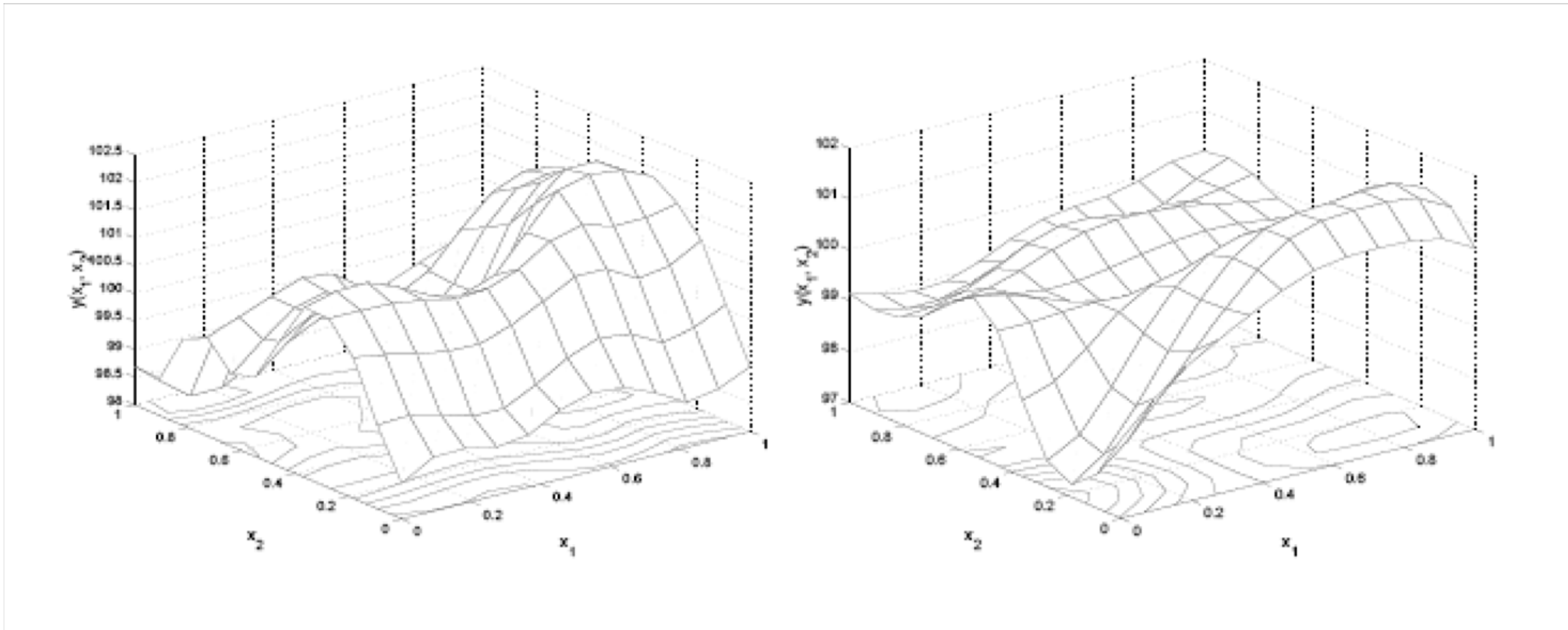
GaSP Models Are Flexible Four draws from $Z(x)$, a zero mean, unit variance GSP with inputs $x \in$ [0,1] and having correlation function

$$R(h) = exp(-\theta h^2)$$

for $\theta =$ 0.5 (solid lines), $\theta =$ 1.0 (dotted lines), and $\theta =$ 10.0 (dashed lines)



• Some draws from a GaSP with inputs $\boldsymbol{x} \in [0,1]^2$



4. Gaussian Stochastic Process (GaSP) Models

- **GaSP Models Are Flexible**

- **GaSP Models are computationally "simple"**

4. Gaussian Stochastic Process (GaSP) Models

- **GaSP Models Are Flexible**

- **GaSP Models are computationally "simple"**

- **GaSP Models are Bayesian in character**

5. Prediction based on the GSP Model

- Given (**training**) data

$$\big(\boldsymbol{x}_1^t,\, y(\boldsymbol{x}_1^t)\big),\, \ldots,\, \big(\boldsymbol{x}_n^t,\, y(\boldsymbol{x}_n^t)\big)$$

predict $y(\boldsymbol{x}_0)$, where $\boldsymbol{x}_0$ is an untried new input.

5. Prediction based on the GSP Model

- Given (**training**) data

$$\big(\boldsymbol{x}_1^t,\, y(\boldsymbol{x}_1^t)\big),\, \ldots,\, \big(\boldsymbol{x}_n^t,\, y(\boldsymbol{x}_n^t)\big)$$

predict $y(\boldsymbol{x}_0)$, where $\boldsymbol{x}_0$ is an untried new input.

- **Notation** $\boldsymbol{Y}^n = (Y(\boldsymbol{x}_1^t), \ldots, Y(\boldsymbol{x}_n^t))$ and $\boldsymbol{y}^n = (y(\boldsymbol{x}_1^t),\, \ldots, y(\boldsymbol{x}_n^t))$

5. Prediction based on the GSP Model

- Given (**training**) data

$$\left(\boldsymbol{x}_1^t,\, y(\boldsymbol{x}_1^t)\right),\, \ldots,\, \left(\boldsymbol{x}_n^t,\, y(\boldsymbol{x}_n^t)\right)$$

predict $y(\boldsymbol{x}_0)$, where $\boldsymbol{x}_0$ is an untried new input.

- **Notation** $\boldsymbol{Y}^n = (Y(\boldsymbol{x}_1^t), \ldots, Y(\boldsymbol{x}_n^t))$ and $\boldsymbol{y}^n = (y(\boldsymbol{x}_1^t),\, \ldots, y(\boldsymbol{x}_n^t))$

- The minimum MSE predictor of $y(\boldsymbol{x}_{\mathbf{0}})$ is

$$\widehat{y}(\boldsymbol{x}_0) = E\{Y(\boldsymbol{x}_0) \,|\, \boldsymbol{Y}^n = \boldsymbol{y}^n\}$$

5. Prediction based on the GSP Model

- Given (**training**) data

$$\left(\boldsymbol{x}_1^t,\, y(\boldsymbol{x}_1^t)\right),\, \ldots,\, \left(\boldsymbol{x}_n^t,\, y(\boldsymbol{x}_n^t)\right)$$

predict $y(\boldsymbol{x}_0)$, where $\boldsymbol{x}_0$ is an untried new input.

- **Notation** $\boldsymbol{Y}^n = (Y(\boldsymbol{x}_1^t), \ldots, Y(\boldsymbol{x}_n^t))$ and $\boldsymbol{y}^n = (y(\boldsymbol{x}_1^t),\, \ldots, y(\boldsymbol{x}_n^t))$

- The minimum MSE predictor of $y(\boldsymbol{x}_{\mathbf{0}})$ is

$$\widehat{y}(\boldsymbol{x}_0) = E\{Y(\boldsymbol{x}_0) \,|\, \boldsymbol{Y}^n = \boldsymbol{y}^n\}$$

Example If $Y(\boldsymbol{x})$ follows the GaSP$(\beta_0, \sigma_Z^2, R(\bullet))$, then

$$\widehat{y}(\boldsymbol{x}_0) \equiv E\{Y(\boldsymbol{x}_0) \,|\, \boldsymbol{Y}^n = \boldsymbol{y}^n\} = \widehat{\beta}_0 + \boldsymbol{r}^{\boldsymbol{T}}(\boldsymbol{x}_0)\boldsymbol{R}^{-1}\left(\boldsymbol{y}^{\boldsymbol{n}} - \widehat{\beta}_0 \mathbf{1}_n\right)$$

where

- $\mathbf{R} = (R(\boldsymbol{x}_i - \boldsymbol{x}_j))$ is $n \times n$
- $\boldsymbol{r}^{\boldsymbol{T}}(\boldsymbol{x}_0) = (R(\boldsymbol{x}_0 - \boldsymbol{x}_i))$ is $1 \times n$
- $\widehat{\beta}_0 \equiv$ WLSE of $\beta_0 = (\mathbf{1}_n^T \mathbf{R}^{-1} \mathbf{1}_n)^{-1} (\mathbf{1}_n^T \mathbf{R}^{-1} \boldsymbol{y}^{\boldsymbol{n}})$

5. Prediction based on the GSP Model

- Given (**training**) data

$$\left(\boldsymbol{x}_1^t,\, y(\boldsymbol{x}_1^t)\right),\, \ldots,\, \left(\boldsymbol{x}_n^t,\, y(\boldsymbol{x}_n^t)\right)$$

predict $y(\boldsymbol{x}_0)$, where $\boldsymbol{x}_0$ is an untried new input.

- **Notation** $\boldsymbol{Y}^n = (Y(\boldsymbol{x}_1^t), \ldots, Y(\boldsymbol{x}_n^t))$ and $\boldsymbol{y}^n = (y(\boldsymbol{x}_1^t),\, \ldots, y(\boldsymbol{x}_n^t))$
- The minimum MSE predictor of $y(\boldsymbol{x}_\mathbf{0})$ is

$$\widehat{y}(\boldsymbol{x}_0) = E\{Y(\boldsymbol{x}_0) \,|\, \boldsymbol{Y}^n = \boldsymbol{y}^n\}$$

Example If $Y(\boldsymbol{x})$ follows the GaSP$(\beta_0, \sigma_Z^2, R(\bullet))$, then

$$\widehat{y}(\boldsymbol{x}_0) \equiv E\{Y(\boldsymbol{x}_0) \,|\, \boldsymbol{Y}^n = \boldsymbol{y}^n\} = \widehat{\beta}_0 + \boldsymbol{r}^{\boldsymbol{T}}(\boldsymbol{x}_0)\boldsymbol{R}^{-1}\left(\boldsymbol{y}^{\boldsymbol{n}} - \widehat{\beta}_0 \mathbf{1}_n\right)$$

where

- $\mathbf{R} = (R(\boldsymbol{x}_i - \boldsymbol{x}_j))$ is $n \times n$
- $\boldsymbol{r}^{\boldsymbol{T}}(\boldsymbol{x}_0) = (R(\boldsymbol{x}_0 - \boldsymbol{x}_i))$ is $1 \times n$
- $\widehat{\beta}_0 \equiv$ WLSE of $\beta_0 = (\mathbf{1}_n^T \mathbf{R}^{-1} \mathbf{1}_n)^{-1} (\mathbf{1}_n^T \mathbf{R}^{-1} \boldsymbol{y}^{\boldsymbol{n}})$

- **IF** only the moment assumptions holds, $\widehat{y}(\boldsymbol{x}_0) \equiv$ **BLUP** of $Y(\boldsymbol{x}_0)$

5. Prediction based on the GSP Model

- **Model Prediction uncertainty** at $\boldsymbol{x}_0$

$$\sigma^2(\boldsymbol{x}_0) = \mathrm{E}\left\{ \left(Y(\boldsymbol{x}_0) - \widehat{y}(\boldsymbol{x}_0)\right)^2 \middle| \boldsymbol{Y}^n = \boldsymbol{y}^n \right\}$$

5. Prediction based on the GSP Model

- **Model Prediction uncertainty** at $\boldsymbol{x}_0$

$$\sigma^2(\boldsymbol{x}_0) = \mathrm{E}\{(Y(\boldsymbol{x}_0) - \widehat{y}(\boldsymbol{x}_0))^2 | \boldsymbol{Y}^n = \boldsymbol{y}^n\}$$

- **Empirical BLUP** **If** the correlation is unknown (the usual case),

$$\Rightarrow \ \mathbf{R}, \boldsymbol{r}^T(\boldsymbol{x}_0), \text{ and } \widehat{\beta}_0 \text{ are also unknown : – (}$$

If, further, $R(\bullet) = R(\bullet \,|\, \boldsymbol{\xi})$ is **parametric**, and we estimate $\boldsymbol{\xi}$ by $\widehat{\boldsymbol{\xi}}$, say, we can predict using the corresponding **empirical BLUP**

$$\widehat{y}(\boldsymbol{x}_0) \equiv E\left\{Y(\boldsymbol{x}_0) \,|\, \boldsymbol{Y}^n = \boldsymbol{y}^n, \widehat{\boldsymbol{\xi}}\right\} = \widehat{\beta}_0 + \widehat{\boldsymbol{r}}(\boldsymbol{x}_0)\widehat{\boldsymbol{R}}^{-1}\left(\boldsymbol{y}^n - \widehat{\beta}_0 \mathbf{1}_n\right)$$

$\widehat{\boldsymbol{\xi}} =$ MLE, REML, penalized likelihood, or other estimator of $\boldsymbol{\xi}$

5. Prediction based on the GSP Model

- **Model Prediction uncertainty** at $\boldsymbol{x}_0$

$$\sigma^2(\boldsymbol{x}_0) = \mathrm{E}\{(Y(\boldsymbol{x}_0) - \widehat{y}(\boldsymbol{x}_0))^2 | \boldsymbol{Y}^n = \boldsymbol{y}^n\}$$

- **Empirical BLUP** **If** the correlation is unknown (the usual case),

$\Rightarrow$ **R**, $\boldsymbol{r}^T(\boldsymbol{x}_0)$, and $\widehat{\beta}_0$ are also unknown : – (

If, further, $R(\bullet) = R(\bullet | \boldsymbol{\xi})$ is **parametric**, and we estimate $\boldsymbol{\xi}$ by $\widehat{\boldsymbol{\xi}}$, say,
we can predict using the corresponding **empirical BLUP**

$$\widehat{y}(\boldsymbol{x}_0) \equiv E\left\{Y(\boldsymbol{x}_0) \,|\, \boldsymbol{Y}^n = \boldsymbol{y}^n, \widehat{\boldsymbol{\xi}}\right\} = \widehat{\beta}_0 + \widehat{\boldsymbol{r}}(\boldsymbol{x}_0)\widehat{\boldsymbol{R}}^{-1}\left(\boldsymbol{y}^n - \widehat{\beta}_0\mathbf{1}_n\right)$$

$\widehat{\boldsymbol{\xi}} =$ MLE, REML, penalized likelihood, or other estimator of $\boldsymbol{\xi}$

- **Fully Bayesian Predictor** (giving **all** parameters priors)

$$\widehat{y}(\boldsymbol{x}_0) = E\{Y(\boldsymbol{x}_0) \,|\, \boldsymbol{Y}^n = \boldsymbol{y}^n\} = E\left\{E\left\{Y(\boldsymbol{x}_0) \,|\, \beta_0, \sigma_Z^2, \boldsymbol{\xi}, \boldsymbol{y}^n\right\}\right\}$$

$$\widehat{y}(\boldsymbol{x}_0) \equiv \int y_0 \times [y_0 | \beta_0, \sigma_Z^2, \boldsymbol{\xi}, \boldsymbol{y}^n] \times [\beta_0, \sigma_Z^2, \boldsymbol{\xi} | \boldsymbol{y}^n] \, dy_0$$

Properties of $\widehat{\boldsymbol{y}}(\boldsymbol{x_0}) = \boldsymbol{E}\{Y(\boldsymbol{x}_0) \,|\, \boldsymbol{Y}^n = \boldsymbol{y}^n\}$

- Simple to compute (linear in $\boldsymbol{y}^n$)

$$\widehat{y}(\boldsymbol{x}) = c_0(\boldsymbol{x}) + \sum_{j=1}^{n} c_j(\boldsymbol{x})\, y(\boldsymbol{x}_j^t)$$

Properties of $\widehat{y}(\boldsymbol{x}_0) = \boldsymbol{E}\{Y(\boldsymbol{x}_0) \,|\, \boldsymbol{Y}^n = \boldsymbol{y}^n\}$

- Simple to compute (linear in $\boldsymbol{y}^n$)

$$\widehat{y}(\boldsymbol{x}) = c_0(\boldsymbol{x}) + \sum_{j=1}^{n} c_j(\boldsymbol{x})\, y(\boldsymbol{x}_j^t)$$

  - but not the Empirical BLUP **:-(**
  - GASP (W. Welch)
    SAS Proc Mixed
    PErK (B. J. Williams)
    BACCO (Hankin)
    and others.......
  - $\mathbf{R}^{-1}$ can be computationally demanding

Properties of $\widehat{\boldsymbol{y}}(\boldsymbol{x}_0) = \boldsymbol{E}\{Y(\boldsymbol{x}_0) \,|\, \boldsymbol{Y}^n = \boldsymbol{y}^n\}$

- Simple to compute (linear in $\boldsymbol{y}^n$)

$$\widehat{y}(\boldsymbol{x}) = c_0(\boldsymbol{x}) + \sum_{j=1}^{n} c_j(\boldsymbol{x})\, y(\boldsymbol{x}_j^t)$$

- Viewed as a function of $\boldsymbol{x}$,

$$\widehat{y}(\boldsymbol{x}) = d_0 + \sum_{j=1}^{n} d_j\, R\big(\boldsymbol{x} - \boldsymbol{x}_j^t\big)$$

Properties of $\widehat{\boldsymbol{y}}(\boldsymbol{x}_0) = \boldsymbol{E}\{Y(\boldsymbol{x}_0) \mid \boldsymbol{Y}^n = \boldsymbol{y}^n\}$

- Simple to compute (linear in $\boldsymbol{y}^n$)

$$\widehat{y}(\boldsymbol{x}) = c_0(\boldsymbol{x}) + \sum_{j=1}^{n} c_j(\boldsymbol{x})\, y(\boldsymbol{x}_j^t)$$

but not the Empirical BLUP **:-(**

- Viewed as a function of $\boldsymbol{x}$,

$$\widehat{y}(\boldsymbol{x}) = d_0 + \sum_{j=1}^{n} d_j\, R\big(\boldsymbol{x} - \boldsymbol{x}_j^t\big)$$

- $\widehat{y}(\boldsymbol{x})$ interpolates data, i.e.,

$$\widehat{y}(\boldsymbol{x}_j^t) = y(\boldsymbol{x}_j^t),\ j = 1, \ldots, n$$

Properties of $\widehat{y}(\boldsymbol{x}_0) = \boldsymbol{E}\{Y(\boldsymbol{x}_0) \,|\, \boldsymbol{Y}^n = \boldsymbol{y}^n\}$

- Simple to compute (linear in $\boldsymbol{y}^n$)

$$\widehat{y}(\boldsymbol{x}) = c_0(\boldsymbol{x}) + \sum_{j=1}^{n} c_j(\boldsymbol{x})\, y(\boldsymbol{x}_j^t)$$

- Viewed as a function of $\boldsymbol{x}$,

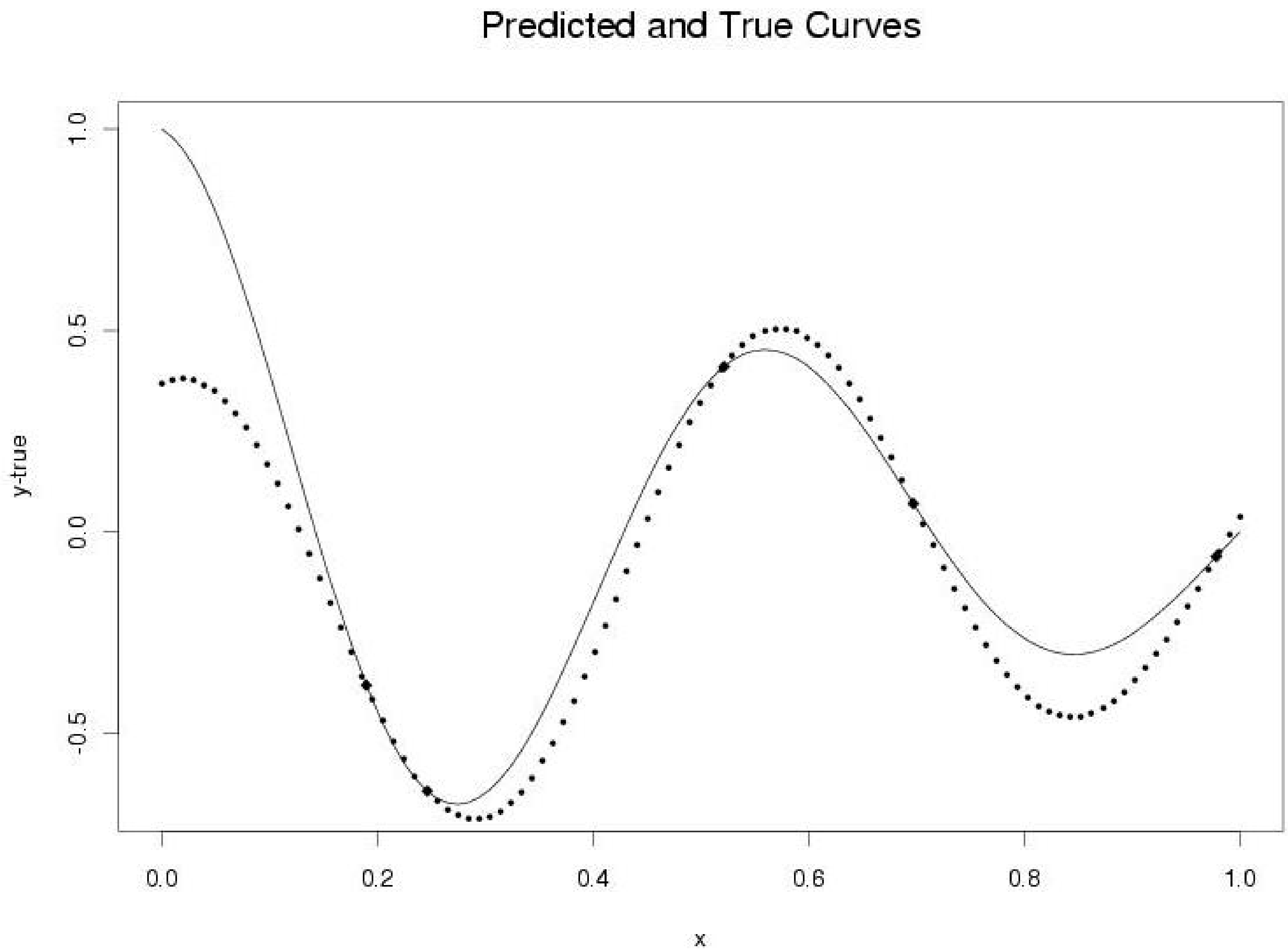
$$\widehat{y}(\boldsymbol{x}) = d_0 + \sum_{j=1}^{n} d_j\, R\big(\boldsymbol{x} - \boldsymbol{x}_j^t\big)$$

- $\widehat{y}(\boldsymbol{x})$ interpolates data, i.e.,

$$\widehat{y}(\boldsymbol{x}_j^t) = y(\boldsymbol{x}_j^t),\; j = 1, \ldots, n$$

- Splines, neural networks and other well-known interpolators correspond to specific choices of regressors and correlation function $R(\bullet)$

Illustration True Curve (solid); five training data points (diamonds); REML-EBLUP with exponential correlation function (dotted)



5. Conclusions-Take Home Messages

1. An increasing number of phenomenon that could previously be studied only by physical experiments, can now be investigated using "computer experiments" or combinations of computer and physical experiments

2. Modeling the responses from computer experiments must account for the (highly) correlated nature of the output $y(\boldsymbol{x})$ over the input space.

3. Prediction of the output function $y(\boldsymbol{x})$ based on Gaussian (or other) stochastic processes can be used to ***interpolate*** trainng data

4. GaSP models are the basis for

- assessing error bands based on model uncertainty,
- for much targeted experimental design, and
- for solving calibration problems