Introduction to Design and Analysis of Computer Experiments

Thomas Santner

Department of Statistics
The Ohio State University

October 2010
Physical Experiments (agriculture, industrial, medicine) are the gold-standard for determining the relationship between a endpoint of scientific interest and potential factors that affect it.
Empirical Experimentation

- **Physical Experiments** (agriculture, industrial, medicine) are the gold-standard for determining the relationship between a endpoint of scientific interest and potential factors that affect it.

- **Viewpoint**

  Experimental output = true I/O relationship + “noise”

\[
Y^p(x) = \eta^{\text{true}}(x) + \epsilon(x)
\]

where

\[
x \rightarrow \eta^{\text{true}}(x)
\]

is the unknown true I/O relationship.
Some Challenges for Physical Experiments

- There are scientifically unimportant nuisance factors—model these factors via $x_i$
Empirical Experimentation

- **Some Challenges for Physical Experiments**
  - There are scientifically unimportant nuisance (blocking) factors- model these factors via $x_i$
  - Some factors affecting the outcome are not recognized—**randomize** $R_x$s to experimental units to prevent unrecognized factors from systematically affecting treatment comparisons
Some Challenges for Physical Experiments

- There are scientifically unimportant nuisance (blocking) factors- model these factors via $x_i$
- Some factors affecting the outcome are not recognized– randomize $R_x$s to experimental units to prevent unrecognized factors from systematically affecting treatment comparisons
- **Choice of sample size** determine “right-sized experiments” to detect “important” treatment differences
Some Challenges for Physical Experiments

- There are scientifically unimportant nuisance (blocking) factors—model these factors via $x_i$
- Some factors affecting the outcome are not recognized—randomize $R_x$ to experimental units to prevent unrecognized factors from systematically affecting treatment comparisons
- Choice of sample size determine “right-sized experiments” to detect “important” treatment differences
- Stochastic Models relating the inputs $x$ to the output $Y^p(x)$
Empirical Experimentation

- **Stochastic Simulation Experiments** popular tool for studying complex physical systems each of whose parts behave in a known stochastic manner but whose ensemble behavior is not understood analytically. (Heavily used in Industrial Engineering—e.g., compare job shop set ups)
Computer Experiments and Physical Experiments

- **Computer Experiment** use of (deterministic) computer simulators that implements a **detailed mathematical model** to study a input/output relationship of scientific interest (micro model vs macro model used in SS Experiments)
Computer Experiments and Physical Experiments

- **Computer Experiment** use of (deterministic) computer simulators that implements a *detailed mathematical model* to study a input/output relationship of scientific interest (micro model vs macro model used in SS Experiments)
- Mathematical model is typically a PDE/ODE. Implementation via FE, CFD, …
Computer Experiments and Physical Experiments

- **Computer Experiment** use of (deterministic) computer simulators that implements a *detailed mathematical model* to study a input/output relationship of scientific interest (micro model vs macro model used in SS Experiments)
- Mathematical model is typically a PDE/ODE. Implementation via FE, CFD, …
- Some applications have data from both (one or more) computer experiment(s) and (one or more) physical experiment(s) to investigate the true input/output relationship of scientific interest. Thus computer experiments are used as **adjuncts** or **surrogates** for physical experiments for studying input/output relationships
Designing an Acetabular Cup
Ong et al (2006) conducted an uncertainty analysis of the effects of engineering cup design, surgical, and patient variables on the stability of uncemented acetabular cups ("porous ingrowth" cups which rely on growth of the bone into the prosthesis).
Designing an Acetabular Cup

- **Engineering Cup design Inputs** Length stem, Stem cross section, sphericity of acetabular cup, etc
Designing an Acetabular Cup

- **Engineering Cup design Inputs** Length stem, Stem cross section, sphericity of acetabular cup, etc
- **Patient Inputs** elastic modulus of the bone, friction between prosthesis stem and the bone
Designing an Acetabular Cup

- **Engineering Cup design Inputs** Length stem, Stem cross section, sphericity of acetabular cup, etc
- **Patient Inputs** elastic modulus of the bone, friction between prosthesis stem and the bone
- **Surgical Inputs** accuracy of reamed acetabular cup in pelvis (gouges area between rear of acetabular cup insert and bone provide space for later debris buildup)
Designing an Acetabular Cup

- **Engineering Cup design Inputs** Length stem, Stem cross section, sphericity of acetabular cup, etc
- **Patient Inputs** elastic modulus of the bone, friction between prosthesis stem and the bone
- **Surgical Inputs** accuracy of reamed acetabular cup in pelvis (gouges area between rear of acetabular cup insert and bone provide space for later debris buildup)
- **Output** is one of three measures of the amount of ingrowth of the bone into the acetabular cup
Other Disciplinary Examples of Computer Experiments

- **Biomechanics** Rawlinson, et al. (2006) simulate the stresses and strains experienced by knee prostheses of different engineering designs.
Other Disciplinary Examples of Computer Experiments

- **Biomechanics** Rawlinson, et al. (2006) simulate the stresses and strains experienced by knee prostheses of different engineering designs.
- **Aeronautical Engineering**
Other Disciplinary Examples of Computer Experiments

- **Biomechanics** Rawlinson, et al. (2006) simulate the stresses and strains experienced by knee prostheses of different engineering designs.
- **Aeronautical Engineering**
- **Drug Development**
Other Disciplinary Examples of Computer Experiments

- **Biomechanics**  Rawlinson, et al. (2006) simulate the stresses and strains experienced by knee prostheses of different engineering designs.
- **Aeronautical Engineering**
- **Drug Development**
- **Economics**  Lempert, et al. (2000) introduce the *Wonderland model* which simulates world economic development under given scenarios of innovation and pollution.
Other Disciplinary Examples of Computer Experiments

- **Biomechanics** Rawlinson, et al. (2006) simulate the stresses and strains experienced by knee prostheses of different engineering designs.

- **Aeronautical Engineering**

- **Drug Development**

- **Economics** Lempert, et al. (2000) introduce the *Wonderland model* which simulates world economic development under given scenarios of innovation and pollution.

- **Tissue Engineering** Spilker (2009) modeled the kinetics of fluid flow in two-phase model of cartilage under sinusoidal loading.
Features of Computer Simulators

- \( x \rightarrow \text{Code} \rightarrow y(x) \) is a proxy for the physical process (based on simplified physics or biology)
Features of Computer Simulators

- $\mathbf{x} \rightarrow \text{Code} \rightarrow y(\mathbf{x})$ is a **proxy** for the physical process (based on simplified physics or biology)
- $y(\mathbf{x})$ is presumed to be a **biased** representation of the true input-response relationship (the true relationship would be obtainable by a physical experiment)
Features of Computer Simulators

- $\mathbf{x} \rightarrow \text{[Code]} \rightarrow y(\mathbf{x})$ is a **proxy** for the physical process (based on simplified physics or biology)
- $y(\mathbf{x})$ is presumed to be a **biased** representation of the true input-response relationship (the true relationship would be obtainable by a physical experiment)
- $y(\mathbf{x})$ need not be **replicated** (the code is deterministic) (in contrast to stochastic computer simulations)
Features of Computer Simulators

- \( x \rightarrow \text{Code} \rightarrow y(x) \) is a **proxy** for the physical process (based on simplified physics or biology)
- \( y(x) \) is presumed to be a **biased** representation of the true input-response relationship (the true relationship would be obtainable by a physical experiment)
- \( y(x) \) need not be **replicated** (the code is deterministic) (in contrast to stochastic computer simulations)
- \( x \) can often be **high-dimensional** (often functional)
Features of Computer Simulators

- $x \rightarrow \text{Code} \rightarrow y(x)$ is a proxy for the physical process (based on simplified physics or biology)
- $y(x)$ is presumed to be a biased representation of the true input-response relationship (the true relationship would be obtainable by a physical experiment)
- $y(x)$ need not be replicated (the code is deterministic) (in contrast to stochastic computer simulations)
- $x$ can often be high-dimensional (often functional)
- When $x$ is high-dimensional, screening is important
Features of Computer Simulators

- \( x \rightarrow \text{Code} \rightarrow y(x) \) is a **proxy** for the physical process (based on simplified physics or biology)
- \( y(x) \) is presumed to be a **biased** representation of the true input-response relationship (the true relationship would be obtainable by a physical experiment)
- \( y(x) \) need not be **replicated** (the code is deterministic) (in contrast to stochastic computer simulations)
- \( x \) can often be **high-dimensional** (often functional)
- When \( x \) is high-dimensional, **screening** is important
- Computer codes used in practice have **running times** that can range from **minutes** to **days**.
Prediction Based on Computer Output

- Suppose we have **only computer output** and no data from an associated physical experiment (momentarily forget about the possibility of data from an associated physical experiment)
Prediction Based on Computer Output

• Suppose we have **only computer output** and no data from an associated physical experiment (momentarily forget about the possibility of data from an associated physical experiment)

• **Warning** Can’t determine code bias in such a setting
Prediction Based on Computer Output

- Suppose we have **only computer output** and no data from an associated physical experiment (momentarily forget about the possibility of data from an associated physical experiment)
- **Warning** Can’t determine code bias in such a setting
- **Notation** Let the inputs and outputs for our training data be

\[
(x_{1}^{tr}, y^{c}(x_{1}^{tr})), \ldots, (x_{n}^{tr}, y^{c}(x_{n}^{tr})),
\]
Suppose we have **only computer output** and no data from an associated physical experiment (momentarily forget about the possibility of data from an associated physical experiment)

- **Warning** Can’t determine code bias in such a setting

- **Notation** Let the inputs and outputs for our training data be

\[
(x^{tr}_1, y^c(x^{tr}_1)), \ldots, (x^{tr}_n, y^c(x^{tr}_n)),
\]

- **Goal** Predict \( y^c(x_0) \) where \( x_0 \) is an untried ("new") input
• **Idea** Regard $y^c(x)$ as a realization (a “draw”) from a random function $Y^c(x)$.

• This is a **Bayesian** model; the desired “smoothness” properties of $y^c(x)$ must be built into all functions drawn from $Y^c(x)$. 
• Provides flexibility—draws from three urns ($Y^c(x)$ processes)
Prediction Based on Computer Output

- We use the training data to pick the exact $Y^c(x)$ urn we draw from.
Prediction Based on Computer Output

- We use the training data to pick the exact $Y^c(x)$ urn we draw from.
- A naive predictor

$$\hat{y}^c(x_0) = E\{Y^c(x_0)|\text{training data}\}$$
Prediction Based on Computer Output

- We use the training data to pick the exact $Y^c(x)$ urn we draw from.
- A naive predictor

$$\hat{y}^c(x_0) = E\{Y^c(x_0)|\text{training data}\}$$

- A measure of prediction uncertainty is

$$\text{Var}(Y^c(x_0)|\text{training data})$$

(due to uncertainty about the model)
Prediction Based on Computer Output
• The **simplest** possible model for $Y(x)$ is the regression-like model

$$Y(x) = \sum_{j=1}^{k} \beta_j f_j(x) + Z(x) = \beta^T f(x) + Z(x)$$

($Y(x) = \text{Large Scale Trends} + \text{Smooth Deviations}$)
The simplest possible model for $Y(x)$ is the regression-like model

$$Y(x) = \sum_{j=1}^{k} \beta_j f_j(x) + Z(x) = \beta^T f(x) + Z(x)$$

($Y(x) =$ Large Scale Trends + Smooth Deviations)

- $f_1(x), \ldots, f_k(x)$ are known regression functions,
GaSP Models

- The **simplest** possible model for $Y(x)$ is the regression-like model

$$Y(x) = \sum_{j=1}^{k} \beta_j f_j(x) + Z(x) = \beta^T f(x) + Z(x)$$

($Y(x) = \text{Large Scale Trends} + \text{Smooth Deviations}$)

- $f_1(x), \ldots, f_k(x)$ are known regression functions,
- $\beta_1, \ldots, \beta_k$ are unknown regression coefficients, and
GaSP Models

- The **simplest** possible model for $Y(x)$ is the regression-like model

$$Y(x) = \sum_{j=1}^{k} \beta_j f_j(x) + Z(x) = \beta^T f(x) + Z(x)$$

$$(Y(x) = \text{Large Scale Trends} + \text{Smooth Deviations})$$

- $f_1(x), \ldots, f_k(x)$ are known regression functions,
- $\beta_1, \ldots, \beta_k$ are unknown regression coefficients, and
- Usually, $\beta^T f(x) = \beta_0$ in computer experiments
GaSP Models

\[ Y(x) = \sum_{j=1}^{k} \beta_j f_j(x) + Z(x) = \beta^T f(x) + Z(x) \]

- **In regression** we regard deviations from the regression mean as “measurement errors” meaning that \( Z(x_1) \) and \( Z(x_2) \) are unrelated (“white noise”).
- **In computer experiments** we want a \( Z(x) \) model that exhibits smooth deviations (from the Large Scale Trends). The simplest such model for \( Z(x) \) is a **stationary Gaussian Stochastic Process** (“GaSP”).
Stationary GaSPs

- $Z(x)$ is **stationarity** if for any $x$, the random variable $Z(x)$ has **constant mean** and **constant variance**, and $\text{Cor}(Z(x_1), Z(x_2))$ **depends only** on $x_1 - x_2$, i.e.,
Stationary GaSPs

- $Z(x)$ is **stationarity** if for any $x$, the random variable $Z(x)$ has **constant mean** and **constant variance**, and $\text{Cor}(Z(x_1), Z(x_2))$ **depends only** on $x_1 - x_2$, i.e.,

\[
E\{Z(x)\} = 0 \quad \iff \quad E\{Y(x)\} = \beta^T f(x)(= \beta_0)
\]
Stationary GaSPs

- $Z(\mathbf{x})$ is **stationarity** if for any $\mathbf{x}$, the random variable $Z(\mathbf{x})$ has **constant mean** and **constant variance**, and $\text{Cor}(Z(\mathbf{x}_1), Z(\mathbf{x}_2))$ **depends only** on $\mathbf{x}_1 - \mathbf{x}_2$, i.e.,

  - $E\{Z(\mathbf{x})\} = 0 \quad (\implies E\{Y(\mathbf{x})\} = \beta^T \mathbf{f}(\mathbf{x}) (= \beta_0))$
  - $\text{Var}(Z(\mathbf{x})) = \sigma_Z^2 \quad (\implies \text{Var}(Y(\mathbf{x})) = \sigma_Z^2)$
Stationary GaSPs

• $Z(x)$ is **stationarity** if for any $x$, the random variable $Z(x)$ has **constant mean** and **constant variance**, and $\text{Cor}(Z(x_1), Z(x_2))$ **depends only** on $x_1 - x_2$, i.e.,

\[
\begin{align*}
\text{(i) } & E\{Z(x)\} = 0 \quad (\iff \ E\{Y(x)\} = \beta^T f(x)(= \beta_0)) \\
\text{(ii) } & \text{Var}(Z(x)) = \sigma_Z^2 \quad (\iff \text{Var}(Y(x)) = \sigma_Y^2) \\
\text{(iii) } & \text{Cor}(Z(x_1), Z(x_2)) \text{ is determined by a } \text{correlation function}, \text{ i.e., an } R(\cdot) \text{ that satisfies } R(0) = 1 \text{ and}
\end{align*}
\]

\[
\begin{align*}
\text{Cor}(Z(x_1), Z(x_2)) &= R(x_1 - x_2) \\
(\iff \text{Cor}(Y(x_1), Y(x_2)) &= R(x_1 - x_2)) \\
\text{(and Cov}(Y(x_1), Y(x_2)) &= \sigma_Y^2 \times R(x_1 - x_2))
\end{align*}
\]
Stationary GaSPs

For any $s \geq 1$ and $x_1, \ldots, x_s$

$$Z \equiv (Z_1, \ldots, Z_s) \equiv (Z(x_1), \ldots, Z(x_s)) \sim \text{MultVar Normal}$$
Stationary GaSPs

- For any $s \geq 1$ and $x_1, \ldots, x_s$

\[ Z \equiv (Z_1, \ldots, Z_s) \equiv (Z(x_1), \ldots, Z(x_s)) \sim \text{MultVar Normal} \]

- The process $Y(\cdot)$ is completely determined once we identify $(\beta_0, \sigma_Z^2, R(\cdot))$, i.e.,

\[ Y \equiv (Y_1, \ldots, Y_s) \equiv (Y(x_1), \ldots, Y(x_s)) \sim N_s \left( F \beta, \sigma_Z^2 \times R \right) \]

where $F = F(x_1, \ldots, x_s)$ is $s \times k$ and $R_{i,j} = R(x_i - x_j)$ is $s \times s$
Stationary GaSPs

- For any $s \geq 1$ and $x_1, \ldots, x_s$

  $$Z \equiv (Z_1, \ldots, Z_s) \equiv (Z(x_1), \ldots, Z(x_s)) \sim \text{MultVar Normal}$$

- The process $Y(\cdot)$ is completely determined once we identify $(\beta_0, \sigma_Z^2, R(\cdot))$, i.e.,

  $$Y \equiv (Y_1, \ldots, Y_s) \equiv (Y(x_1), \ldots, Y(x_s)) \sim N_s \left( F\beta, \sigma_Z^2 \times R \right)$$

  where $F = F(x_1, \ldots, x_s)$ is $s \times k$ and $R_{ij} = R(x_i - x_j)$ is $s \times s$

  - Typically assume $R(\cdot) = R(\cdot | \xi)$ is known up to a finite vector of parameters $\xi$, i.e., $R(\cdot)$ is \textit{parametric}
Popular Correlation Functions

Let \( z \equiv x_1 - x_2 = (z_1, \ldots, z_d) \)
Popular Correlation Functions

Let $\mathbf{z} \equiv \mathbf{x}_1 - \mathbf{x}_2 = (z_1, \ldots, z_d)$

- **White Noise**

  $$R(\mathbf{z}) = \begin{cases} 
  1, & \mathbf{z} = \mathbf{0} \\
  0, & \mathbf{z} \neq \mathbf{0}
  \end{cases} \implies R = I_s$$
Let \( \mathbf{z} \equiv \mathbf{x}_1 - \mathbf{x}_2 = (z_1, \ldots, z_d) \)

- **White Noise**

  \[
  R(\mathbf{z}) = \begin{cases} 
  1, & z = 0 \\
  0, & z \neq 0 
  \end{cases} \implies R = I_s
  \]

- **Product Power Gaussian Correlation Function**

  \[
  R(\mathbf{z}) = \prod_{j=1}^{d} \exp(-\xi_j z_j^2) = \exp\left(-\sum_{j=1}^{d} \xi_j z_j^2\right), \quad \mathbf{z} \in \mathbb{R}^d
  \]
Popular Correlation Functions

- **Product Cubic Correlation Function**

\[
R(z|\xi) = \prod_{j=1}^{d} R(z_j|\xi_j)
\]

where \(\xi = (\xi_1, \ldots, \xi_d) > 0\) and

\[
R(z|\xi) = \begin{cases} 
1 - 6 \left( \frac{z}{\xi} \right)^2 + 6 \left( \frac{|z|}{\xi} \right)^3, & |z| \leq \xi/2 \\
2 \left( 1 - \frac{|z|}{\xi} \right)^3, & \xi/2 < |z| \leq \xi \\
0, & \xi < |z|
\end{cases}
\]

for \(\xi > 0\) and \(z \in \mathbb{I}R\).
The Multivariate Normal Distribution

- **Definition** \( \mathbf{X} = (X_1, \ldots, X_d)^\top \) has the **multivariate normal distribution** with mean \( \mu \) and covariance matrix \( \Sigma \), denoted \( \mathbf{X} \sim N_d(\mu, \Sigma) \), if \( \mathbf{X} \) joint pdf

\[
\frac{1}{(2\pi)^{d/2} \sqrt{\det(\Sigma)}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - bt\mu)^\top \Sigma^{-1} (\mathbf{x} - bt\mu)^\top \right\}
\]
The Multivariate Normal Distribution

- **Definition** \( X = (X_1, \ldots, X_d)^\top \) has the **multivariate normal distribution** with mean \( \mu \) and covariance matrix \( \Sigma \), denoted \( X \sim N_d(\mu, \Sigma) \), if the joint pdf

\[
\frac{1}{(2\pi)^{d/2} \sqrt{\det(\Sigma)}} \exp\left\{-\frac{1}{2} (x - bt\mu)^\top \Sigma^{-1} (x - bt\mu)^\top \right\}
\]

- Recall that if

\[
W = (W_1, W_2) \sim N_d \left( \left( \begin{array}{c} \mu_1 \\ \mu_2 \end{array} \right), \left( \begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array} \right) \right)
\]

then given that \( W_2 = w_2 \) the conditional density of \( W_1 \) given \( W_2 = w_2 \), denoted \([W_1|W_2 = w_2]\) is

\[
N_{d_1} \left( \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2), \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \right)
\]
Naive Predictor

\[ \hat{y}_c(x_0) = E\{Y_c(x_0) | Y^{tr}\} = \sum_{j=1}^{k} \beta_j f_j(x_0) + r_0^\top R^{-1} (Y^{tr} - F \beta) \]
Application

- **Naive Predictor**

\[
\hat{y}^c(x_0) = E\{Y^c(x_0)|Y^{tr}\} = \sum_{j=1}^{k} \beta_j f_j(x_0) + r_0^T R^{-1} (Y^{tr} - F\beta)
\]

- **Prediction Uncertainty** is

\[
Var(Y^c(x_0)|Y^{tr}) = \sigma_Z^2 \left(1 - r_0^T R^{-1} r_0\right)
\]
Application

- **Naive Predictor**

\[
\hat{y}^c(x_0) = E\{Y^c(x_0)|Y^{tr}\} = \sum_{j=1}^{k} \beta_j f_j(x_0) + r_0^\top R^{-1} (Y^{tr} - F\beta)
\]

- **Prediction Uncertainty** is

\[
\text{Var}(Y^c(x_0)|Y^{tr}) = \sigma_Z^2 \left(1 - r_0^\top R^{-1} r_0\right)
\]

- Usually don’t know \((\beta, \sigma_Z^2, \xi)\) :-(
Inference

- **Frequentist Inference** Choice of urn equivalent to selection of \((\beta_0, \sigma^2_Z, \xi)\). Use training data to estimate \((\beta_0, \sigma^2_Z, \xi)\) (the “urn”) which is the basis for our predictor

\[
\hat{y}^c(x_0) = E\{Y^c(x_0) | \text{training data, } \hat{\beta}_0, \hat{\sigma}^2_Z, \hat{\xi}\}
\]
Inference

- **Frequentist Inference** Choice of urn equivalent to selection of \((\beta_0, \sigma_Z^2, \xi)\). Use training data to estimate \((\beta_0, \sigma_Z^2, \xi)\) (the “urn”) which is the basis for our predictor

\[
\hat{y}^c(x_0) = E\{Y^c(x_0)|\text{training data}, \hat{\beta}_0, \hat{\sigma}_Z^2, \hat{\xi}\}
\]

- **Bayesian Inference** Places a prior distribution on \((\beta_0, \sigma_Z^2, \xi)\) and uses the posterior \([\beta_0, \sigma_Z^2, \xi|\text{training data}]\) to select a collection of urns and averages the predictors from each urn

\[
\hat{y}^c(x_0) = E\{E\{Y^c(x_0)|\text{training data}, \beta_0, \sigma_Z^2, \xi}\}\}
\]

where the outer \(E\{\cdot\}\) is wrt \([\beta_0, \sigma_Z^2, \xi|\text{training data}]\).
True Curve (solid); \( n = 5 \) training data (diamonds); 
REML-EBLUP with Gaussian correlation function (dotted)
Conclusions

- The frequentist “plug-in” \( \hat{y}(x_0) \) is the basic off-the-shelf predictor of \( y(x_0) \) based on training data.
The frequentist “plug-in” $\hat{y}(x_0)$ is the basic off-the-shelf predictor of $y(x_0)$ based on training data.

When $(\beta_0, \sigma^2_Z, \xi)$ are known, $\hat{y}(x_0)$ has many theoretical optimality properties although when these parameters are estimated from the data, the list of optimality properties is much shorter.
Conclusions

- The frequentist “plug-in” \( \hat{y}(x_0) \) is the basic off-the-shelf predictor of \( y(x_0) \) based on training data.

- When \( (\beta_0, \sigma_2^2, \xi) \) are known, \( \hat{y}(x_0) \) has many theoretical optimality properties, although when these parameters are estimated from the data, the list of optimality properties is much shorter.

- How to use the data to estimate \( (\beta_0, \sigma_2^2, \xi) \)? MLE? REML? PMLE? XV? No plug-in method overcomes the “smoothing” problem seen in the example, although Bayesian predictors can...
Conclusions

• The frequentist “plug-in” $\hat{y}(x_0)$ is the basic off-the-shelf predictor of $y(x_0)$ based on training data

• When $(\beta_0, \sigma^2_Z, \xi)$ are known, $\hat{y}(x_0)$ has many theoretical optimality properties although when these parameters are estimated from the data, the list of optimality properties is much shorter

• How to use the data to estimate $(\beta_0, \sigma^2_Z, \xi)$? MLE? REML? PMLE? XV? No plug-in method overcomes the “smoothing” problem seen in the example, although Bayesian predictors can

• Bayesian methodology can be used to make more legitimate predictors and honest estimates of prediction uncertainty
Conclusions

- Other limitations: there are many cases where the $y(\cdot)$ does not “look” like a draw from a GaSP. Need predictors based on non-stationary models.
Conclusions

• Other limitations: there are many cases where the $y(\cdot)$ does not “look” like a draw from a GaSP. Need predictors based on non-stationary models.

• Two approaches to the **Design** of a computer experiment, i.e., choice of $\{x_i^{tr}\}_{i=1}^n$. 
Conclusions

- Other limitations: there are many cases where the $y(\cdot)$ does not “look” like a draw from a GaSP. Need predictors based on non-stationary models.

- Two approaches to the Design of a computer experiment, i.e., choice of $\{x_{i}^{tr}\}_{i=1}^{n}$.

- **Approach 1** geometric-choose $v$ to be “space-filling.”
Conclusions

• Other limitations: there are many cases where the \( y(\cdot) \) does not “look” like a draw from a GaSP. Need predictors based on non-stationary models.

• Two approaches to the Design of a computer experiment, i.e., choice of \( \{x_i^{tr}\}_{i=1}^n \).

• **Approach 1** geometric-choose \( v \) to be “space-filling.”

• **Approach 2** criteria-based designs for computer experiments, eg, find an input which achieves

\[
\arg\min_x y^c(x)
\]
Questions?