Coupled Gaussian Process Models

V. Roshan Joseph
(Joint work with Shan Ba)

Supported by NSF CMMI-1030125

Georgia Institute of Technology

INDUSTRIAL & SYSTEMS ENGINEERING

Gaussian Process Model

- GP model (ordinary kriging)

$$Y(\mathbf{x}) = \mu + Z(\mathbf{x}),$$

where $Z(\mathbf{x}) \sim GP(0, \sigma^2 R)$ is a stationary GP with mean 0 and covariance $\sigma^2 \mathrm{cov}(Y(\mathbf{x}+\mathbf{h}), Y(\mathbf{x})) = \sigma^2 R(\mathbf{h})$.

GP Model-continued

- Data from computer experiment:

$$(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)$$

- Predictor: $\hat{y}(\mathbf{x}) = \hat{\mu} + \mathbf{r}^\top(\mathbf{x})\mathbf{R}^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1})$,

$$\mathbf{r}(\mathbf{x}) = (R(\mathbf{x} - \mathbf{x}_1), \ldots, R(\mathbf{x} - \mathbf{x}_n))^\top, \ \mathbf{R} = [R(\mathbf{x}_i - \mathbf{x}_j)]_{i,j=1,\ldots,n},$$
$$\hat{\mu} = (\mathbf{1}^\top \mathbf{R}^{-1}\mathbf{1})^{-1}(\mathbf{1}^\top \mathbf{R}^{-1}\mathbf{y}).$$

- Can we improve the prediction?

GP Model-continued

- Prediction intervals:

$$\hat{y}(\boldsymbol{x}) \pm z_{\frac{\alpha}{2}} \hat{\sigma} \left\{ 1 - \boldsymbol{r}(\boldsymbol{x})' \boldsymbol{R}^{-1} \boldsymbol{r}(\boldsymbol{x}) + \frac{(1 - \boldsymbol{r}(\boldsymbol{x})' \boldsymbol{R}^{-1} \mathbf{1})^2}{\mathbf{1}' \boldsymbol{R}^{-1} \mathbf{1}} \right\}^{1/2}$$

- Can we improve the coverage?

GP Model Assumptions

1. Constant mean: μ
2. Constant variance σ^2
3. Stationary correlation function:

$$\mathrm{cor}(Y(\mathbf{x}+\mathbf{h}), Y(\mathbf{x})) = R(\mathbf{h})$$

4. Gaussian distribution

GP Model Assumptions

1. Constant mean: μ
   – Improve the mean model
2. Constant variance: σ^2
   – Improve the variance model

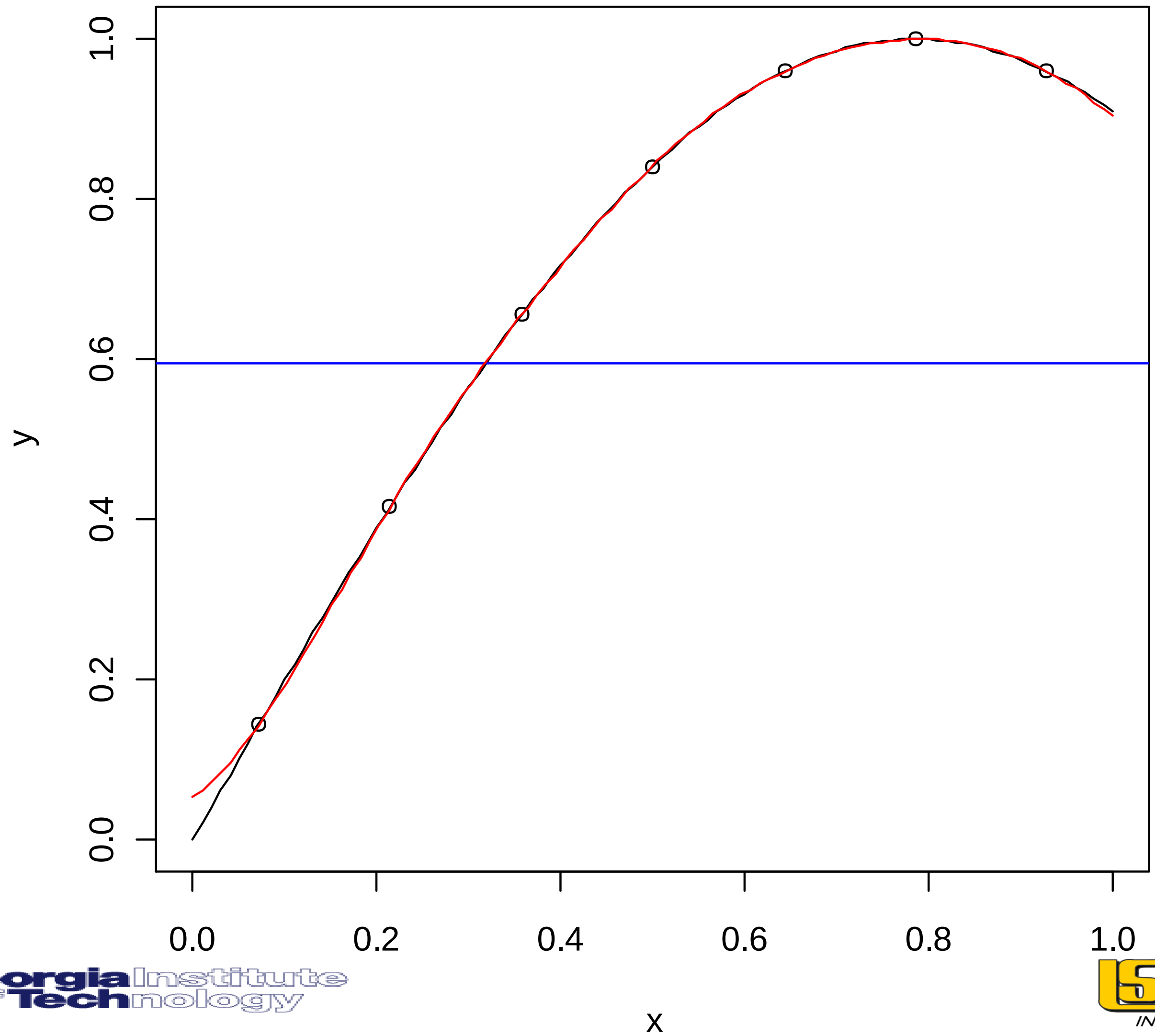
Two Earlier Attempts

1. Limit Kriging (Joseph 2006)
2. Blind Kriging (Joseph, Hung, Sudjianto 2008)

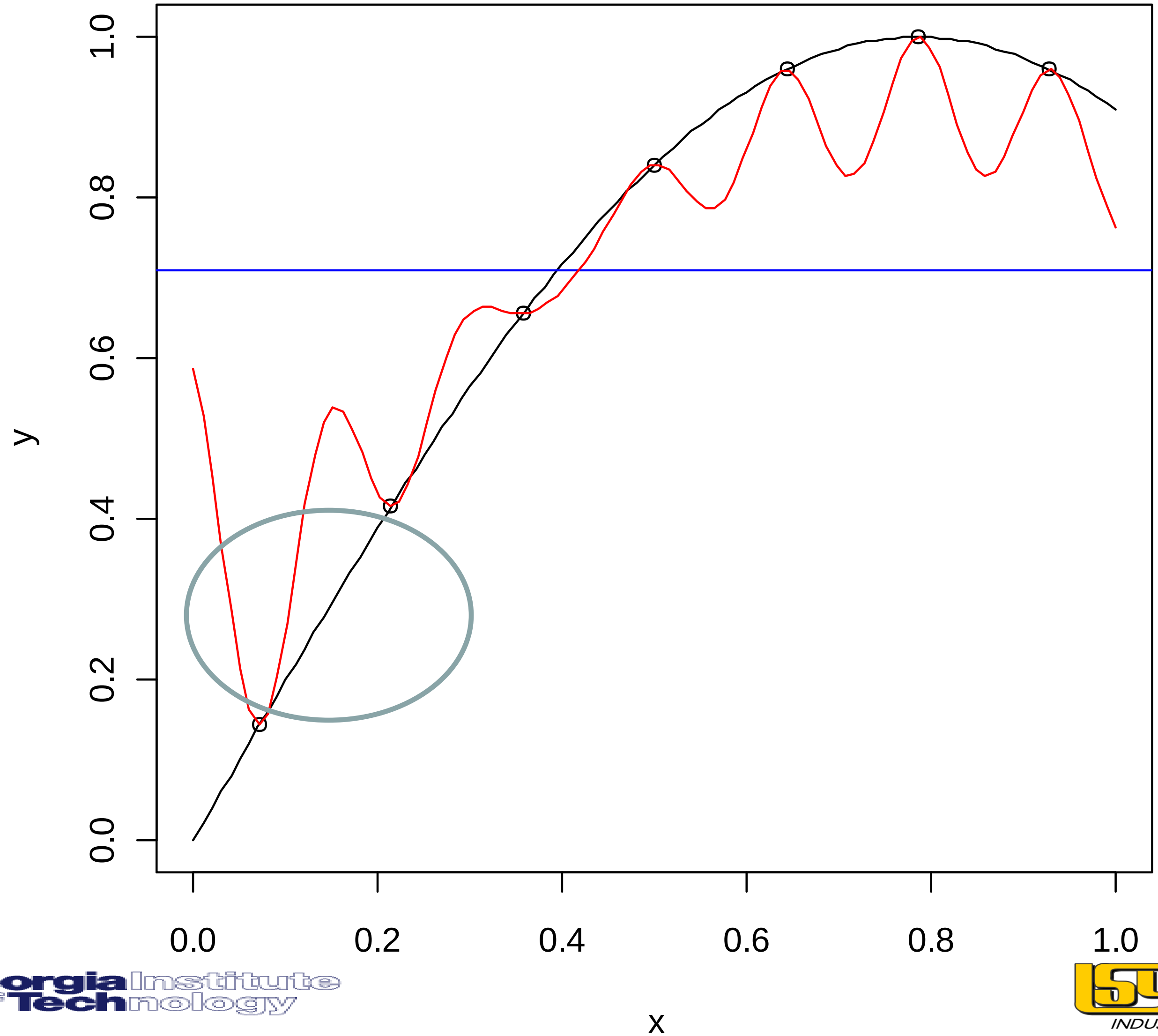
Other Approaches:

- Nonstationary covariance functions
  - Sampsons and Guttorp (1992), Schmidt and O'Hagan (2003), Higdon et al. (1999), Paciorek and Schervish (2006)
- BayesianTreed Gaussian processes
  - Gramacy and Lee (2008)

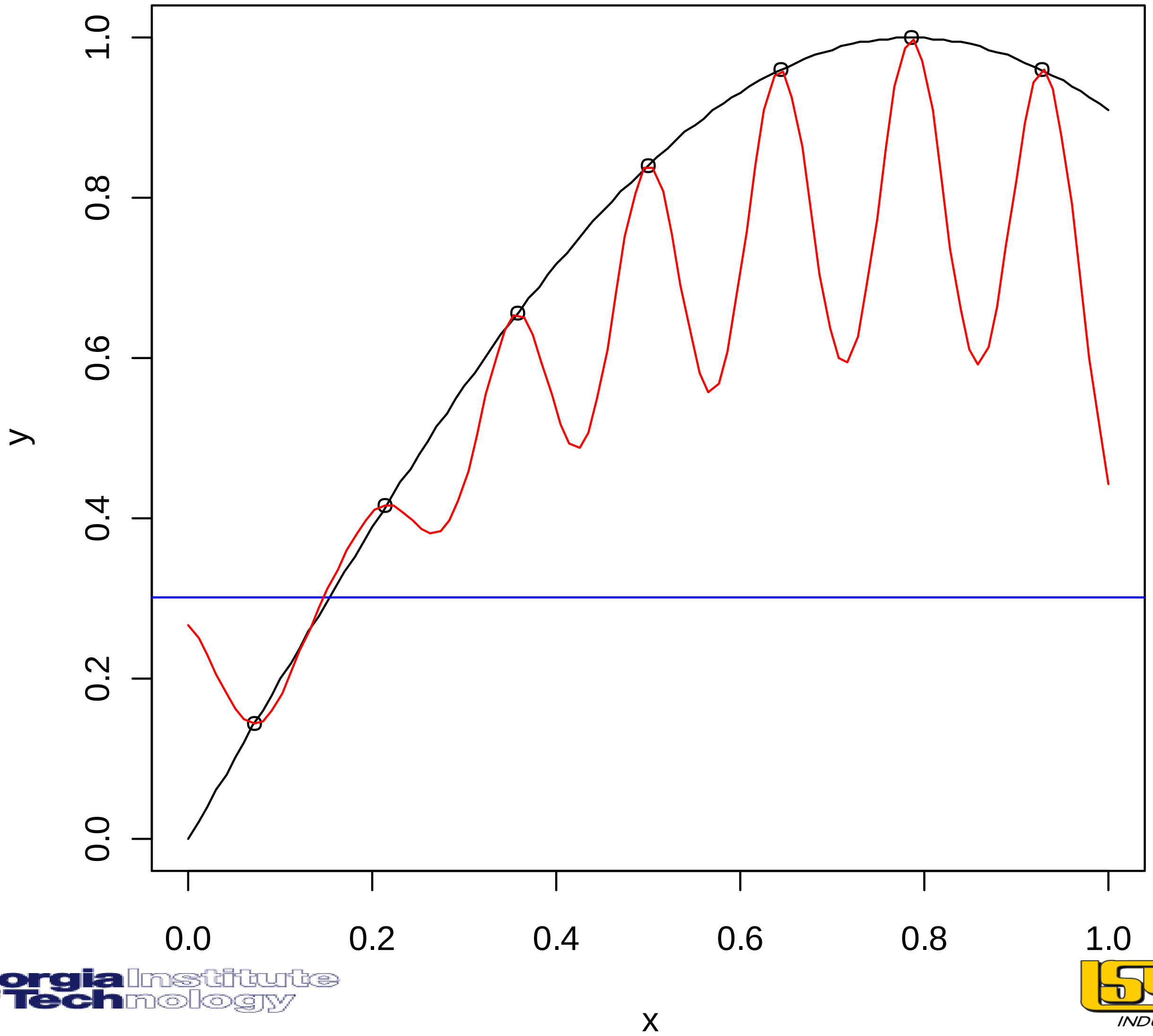
Limit Kriging Idea



Limit Kriging Idea



Limit Kriging Idea



Georgia Institute of Technology

ISyE INDUSTRIAL & SYSTEMS ENGINEERING

Limit Kriging Idea

- Change the mean at each prediction point

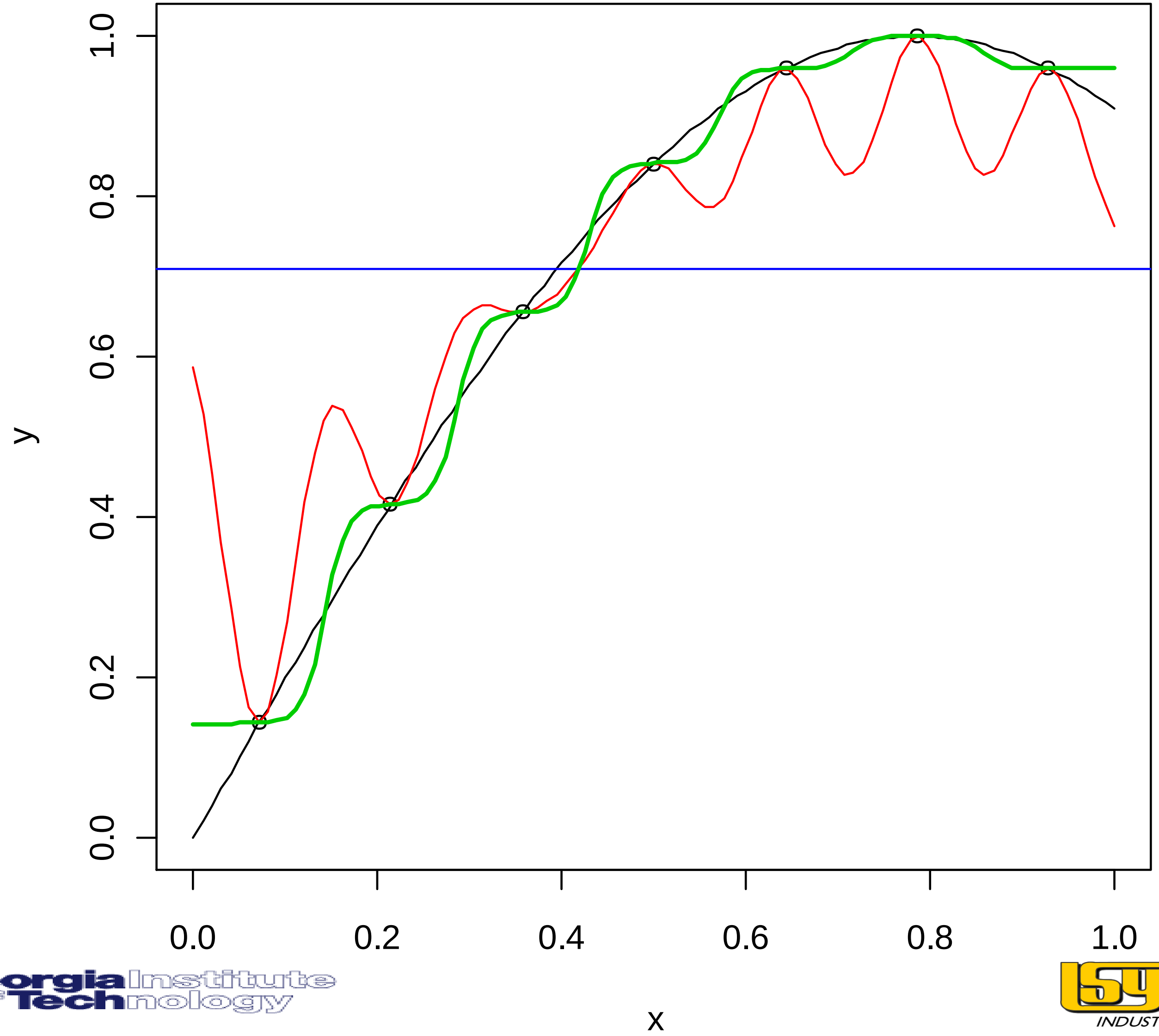
$$\hat{y}_k(\boldsymbol{x}) = \hat{y}_{k-1}(\boldsymbol{x}) + \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \hat{y}_{k-1}(\boldsymbol{x})\mathbf{1})$$

- As $k \to \infty$,

$$\hat{y}(\boldsymbol{x}) = \hat{y}(\boldsymbol{x}) + \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \hat{y}(\boldsymbol{x})\mathbf{1})$$

$$\hat{y}(\boldsymbol{x}) = \frac{\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\boldsymbol{y}}{\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mathbf{1}}$$

Limit Kriging Idea



Georgia Institute of Technology

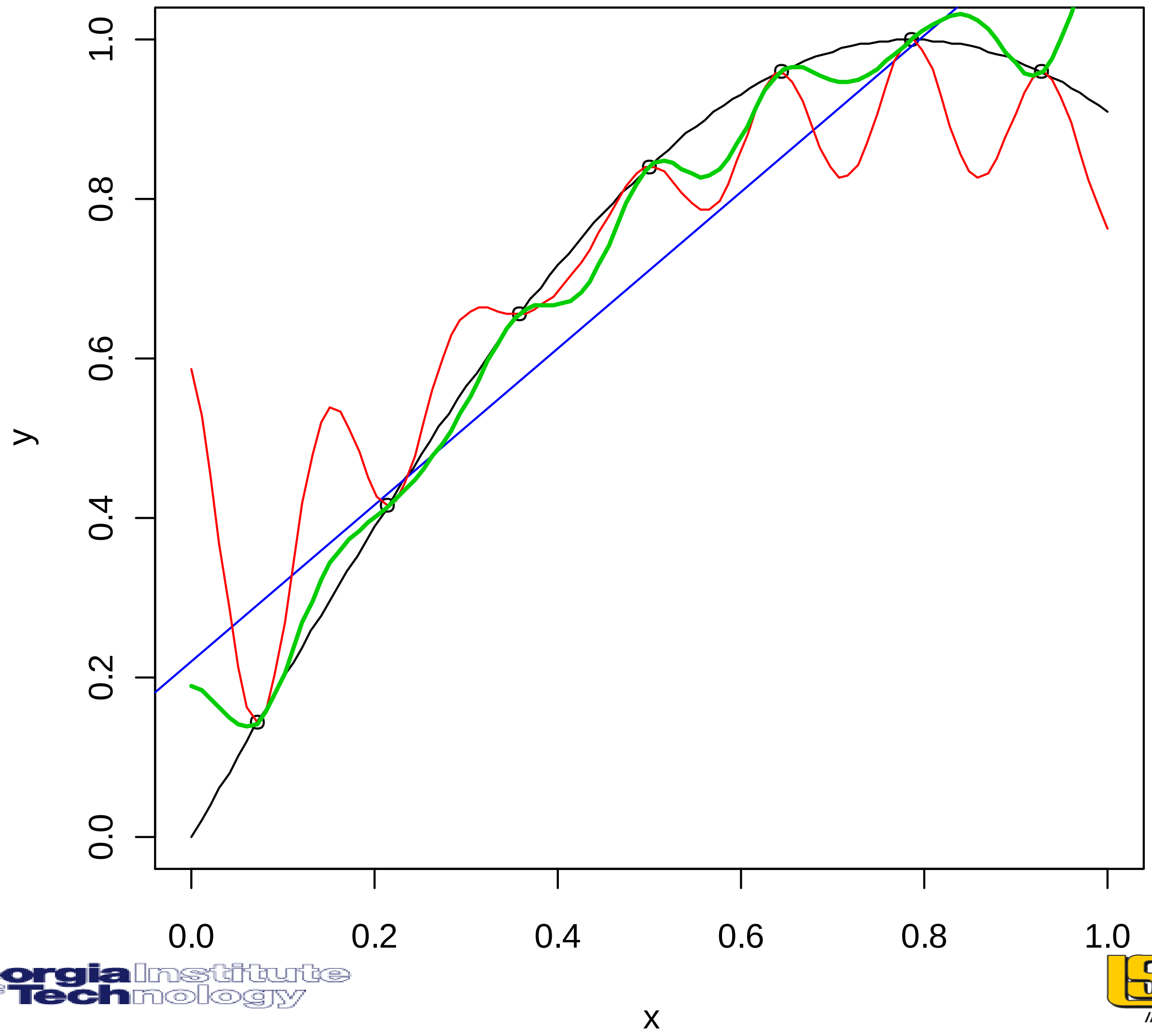
ISyE INDUSTRIAL & SYSTEMS ENGINEERING

Blind Kriging Idea

- Universal kriging

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

Blind Kriging Idea



INDUSTRIAL & SYSTEMS ENGINEERING

Blind Kriging Idea

- Morris, Mitchel, Ylvisaker (1993)

$$y = \frac{2\pi T_u(H_u - H_l)}{ln(r/r_w)\left[1 + \frac{2LT_u}{ln(r/r_w)r_w^2 K_w} + \frac{T_u}{T_l}\right]}$$

Blind Kriging Idea

Method	m	RMSPE
Ordinary Kriging	0	9.7
Universal Kriging (linear)	8	11.3
Universal Kriging (linear and quadratic)	16	18.0

Blind Kriging Idea

- Perform variable selection using a candidate set of functions

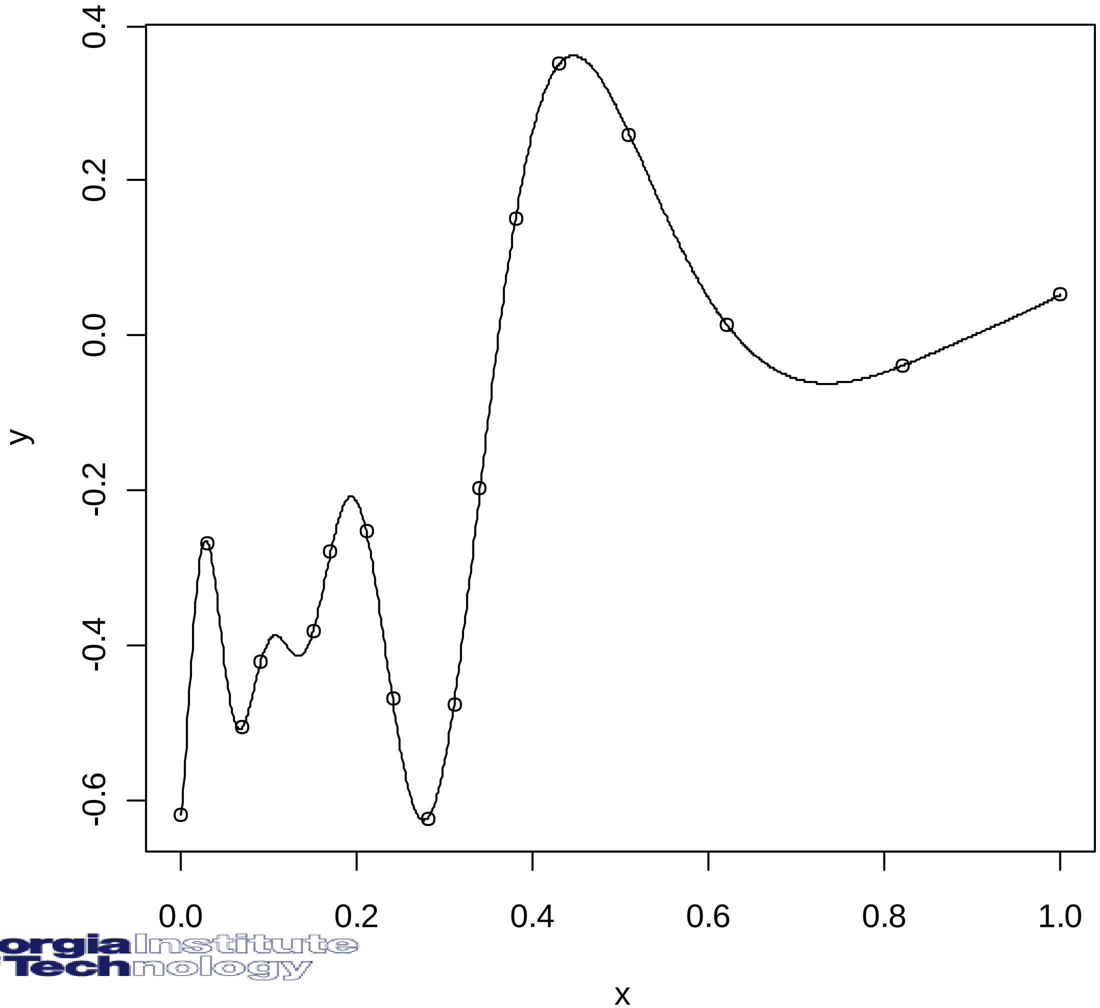
Method	m	RMSPE
Ordinary Kriging	0	9.7
Blind Kriging	1	**5.5**
Universal Kriging (linear)	8	11.3
Universal Kriging (linear and quadratic)	16	18.0

Shortcomings

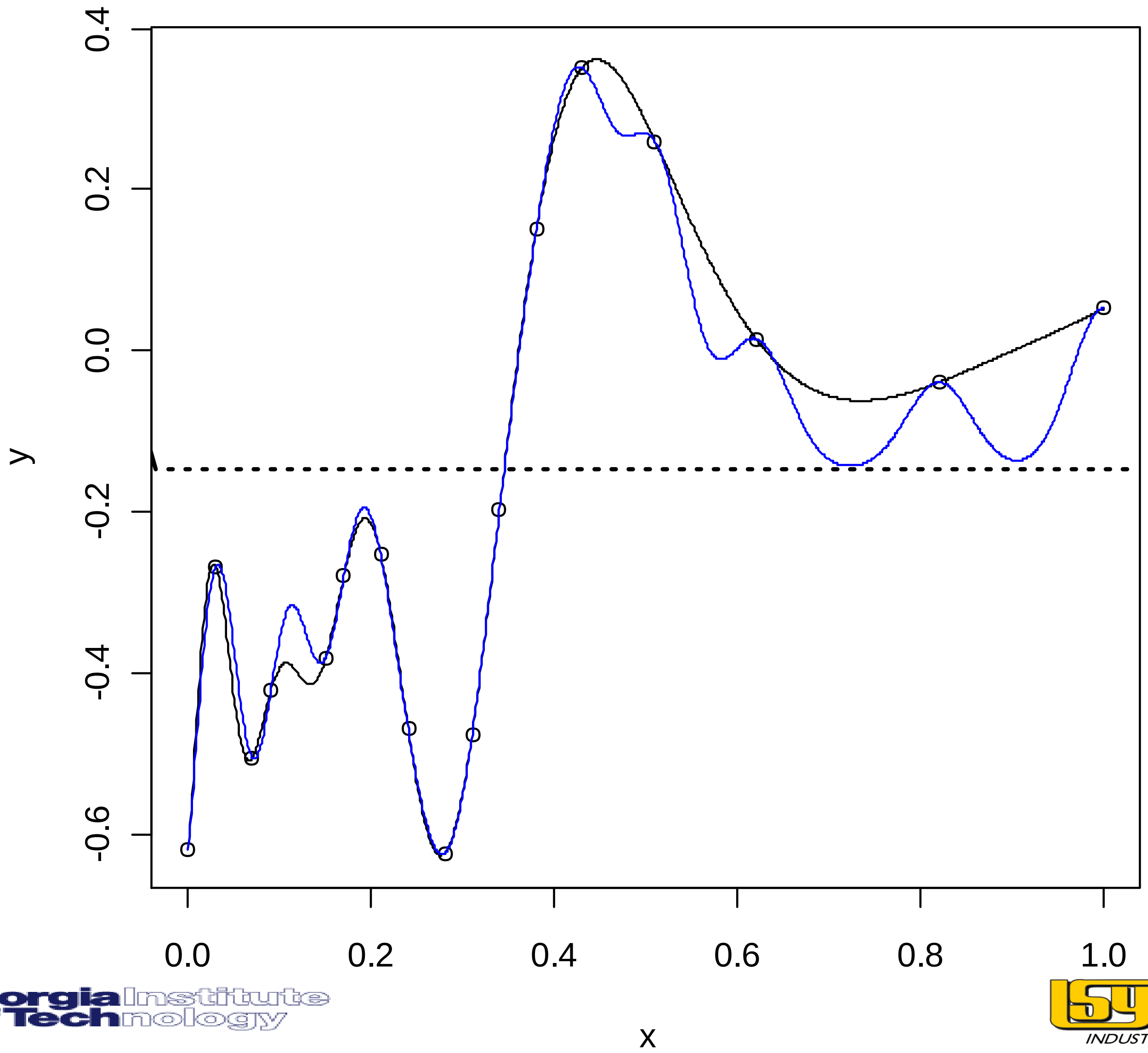
- Variable selection is unstable.
- Not easy to capture the global trend in more complex problems.

An Example

- Xiong et al. (2007)



GP Model



Georgia Institute of Technology

ISyE INDUSTRIAL & SYSTEMS ENGINEERING

Coupled Gaussian Process

- Use another kriging model to capture the global trend.

$$Y(\mathbf{x}) = Z_{global}(\mathbf{x}) + Z_{local}(\mathbf{x}),$$
$$Z_{global}(\mathbf{x}) \sim GP(\mu, \tau^2 g),$$
$$Z_{local}(\mathbf{x}) \sim GP(0, \sigma^2 I),$$

where $Z_{global}(\mathbf{x})$ and $Z_{local}(\mathbf{x})$ are stationary and independent

CGP

- The model reduces to

$$Y(\mathbf{x}) \sim GP(\mu, \tau^2 g + \sigma^2 I)$$

- Predictor:

$$\hat{y}(\mathbf{x}) = \hat{\mu} + (\mathbf{g}(\mathbf{x}) + \lambda \mathbf{l}(\mathbf{x}))^\top (\mathbf{G} + \lambda \mathbf{L})^{-1} (\mathbf{y} - \hat{\mu}\mathbf{1})$$

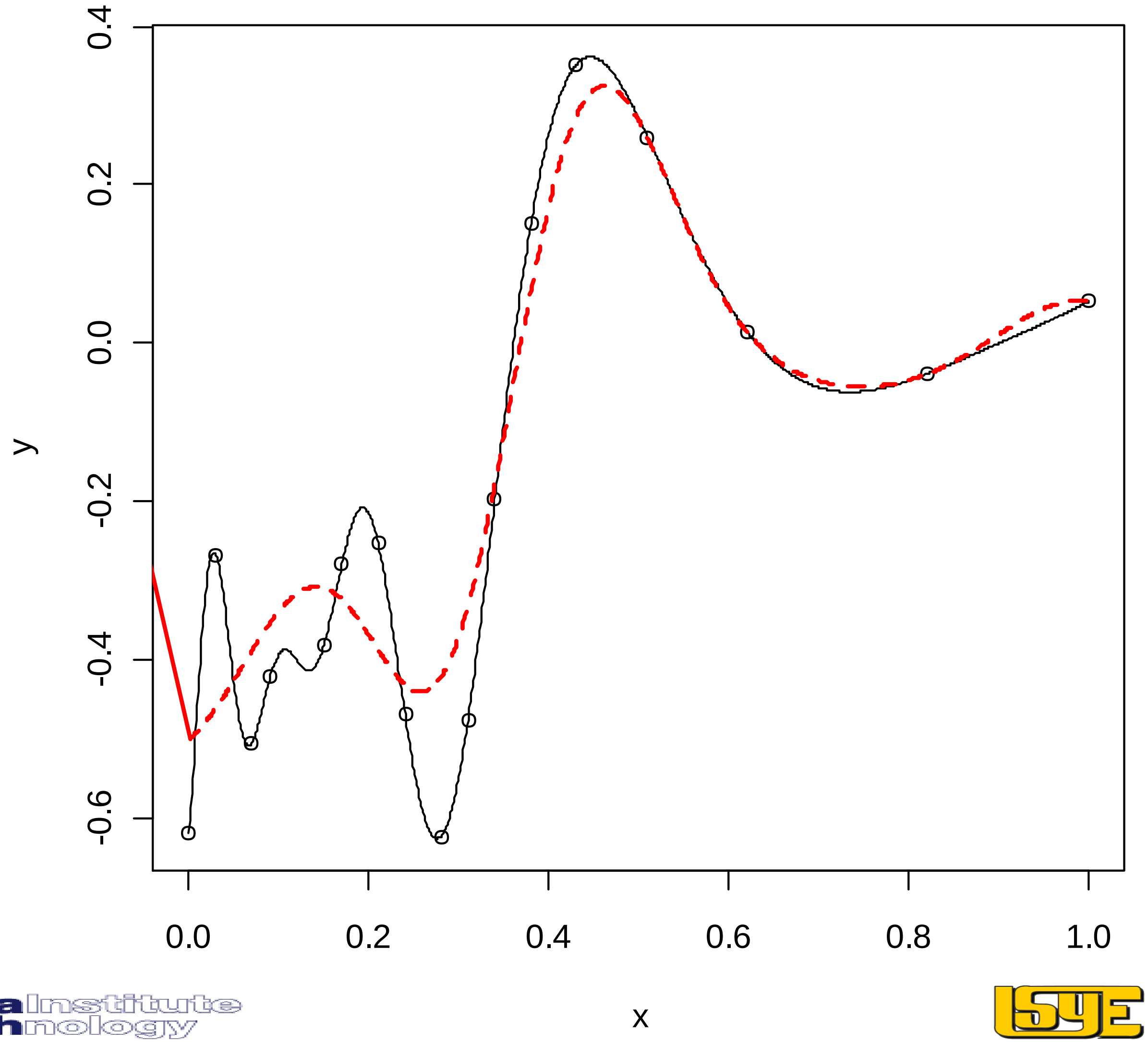
where $\lambda = \sigma^2 / \tau^2$

CGP

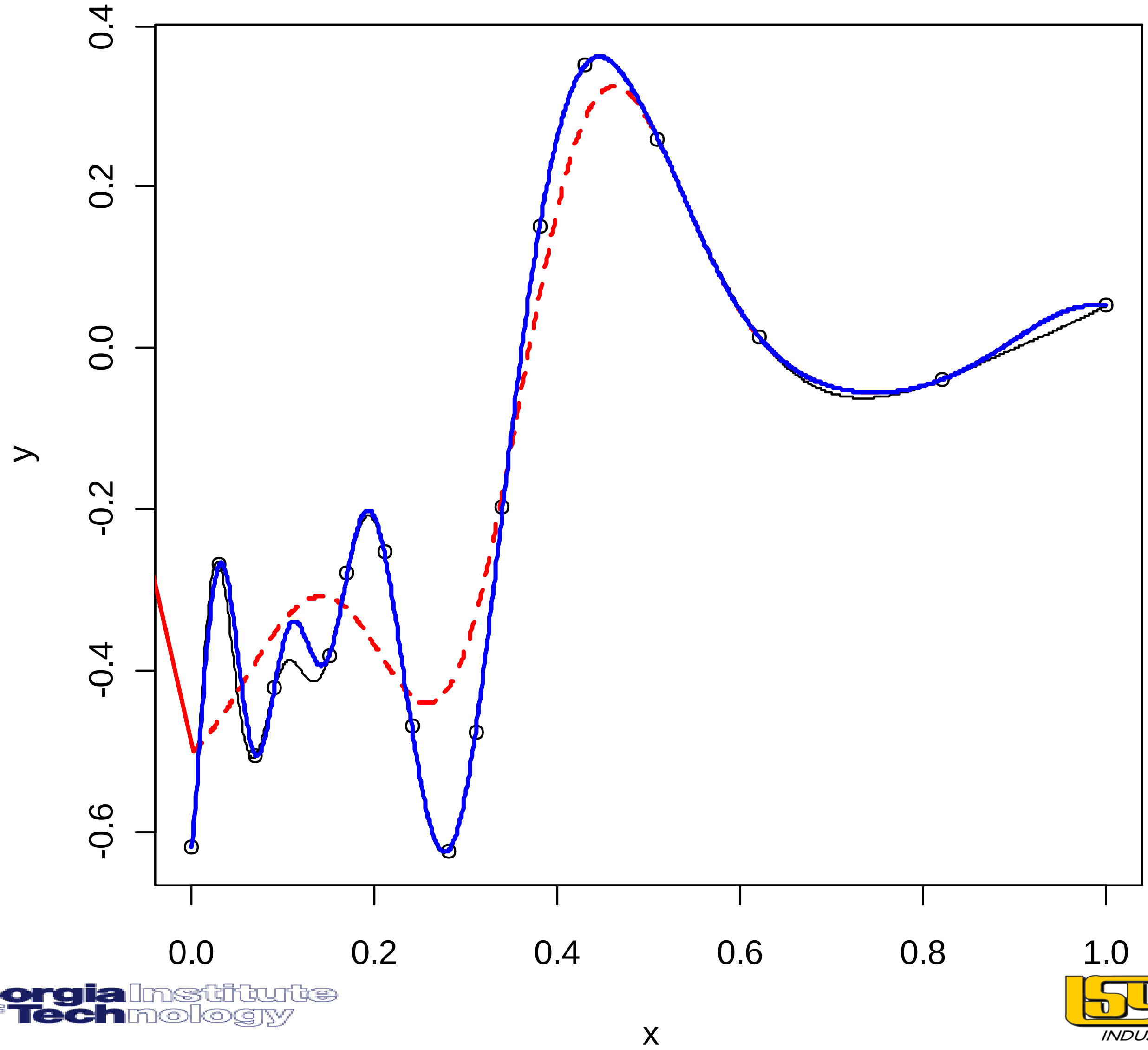
- Sum of two predictors

$$\hat{y}(\mathbf{x}) = \hat{y}_{global}(\mathbf{x}) + \hat{y}_{local}(\mathbf{x}),$$
$$\hat{y}_{global}(\mathbf{x}) = \hat{\mu} + \mathbf{g}^\top(\mathbf{x})(\mathbf{G} + \lambda\mathbf{L})^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}),$$
$$\hat{y}_{local}(\mathbf{x}) = \lambda\mathbf{l}^\top(\mathbf{x})(\mathbf{G} + \lambda\mathbf{L})^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}).$$

Example: Global Trend



Example: Global+Local



Georgia Institute of Technology

ISyE INDUSTRIAL & SYSTEMS ENGINEERING

Connections with Nugget Predictor

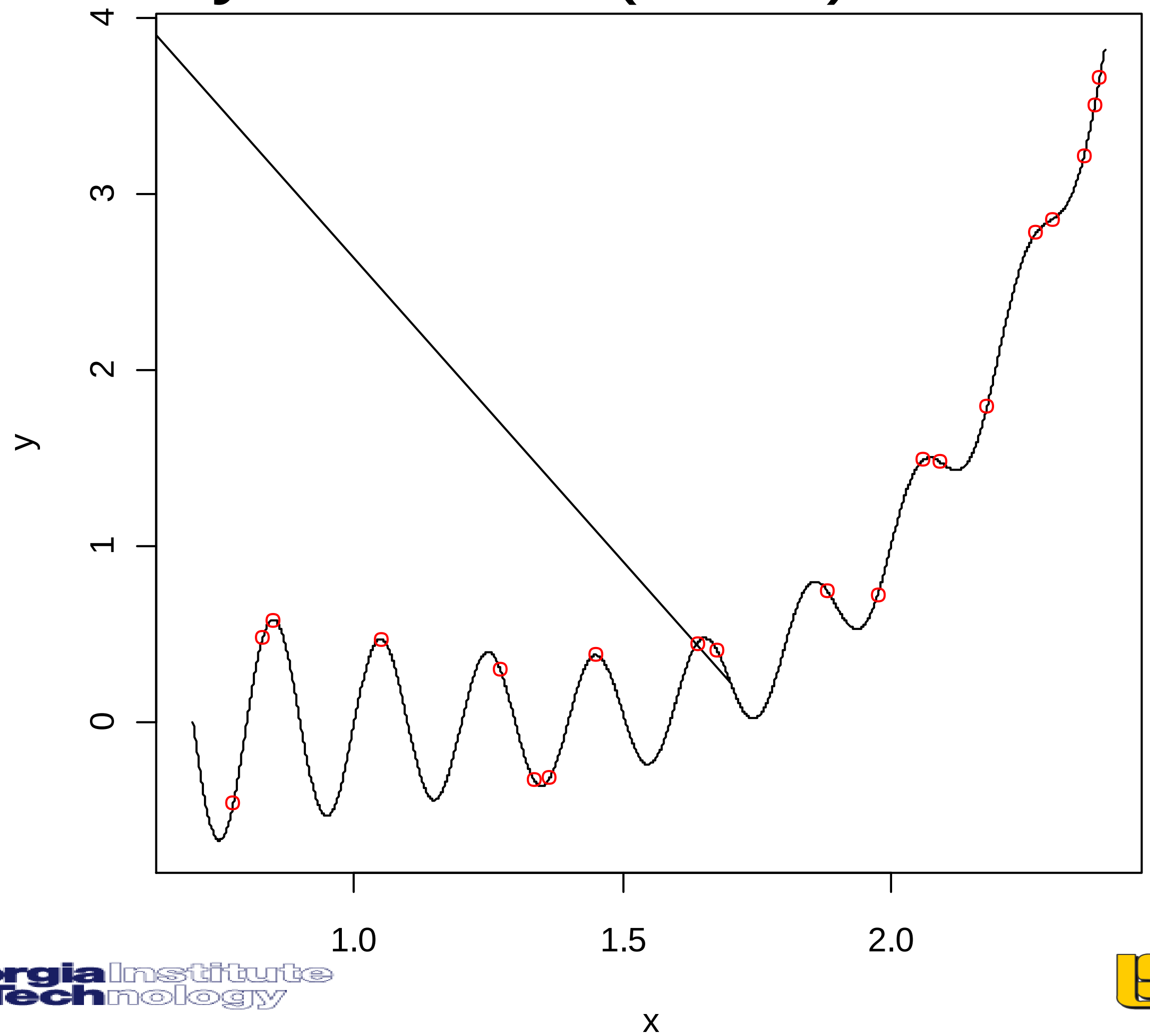
- If the local model has zero correlations, then

$$\mathbf{L} = \mathbf{I}$$

$$\hat{y}(\boldsymbol{x}) = \hat{\mu} + \boldsymbol{g}(\boldsymbol{x})'(\boldsymbol{G} + \lambda \boldsymbol{I})^{-1}(\boldsymbol{y} - \hat{\mu}\mathbf{1})$$

An Example

- Gramacy and Lee (2010)

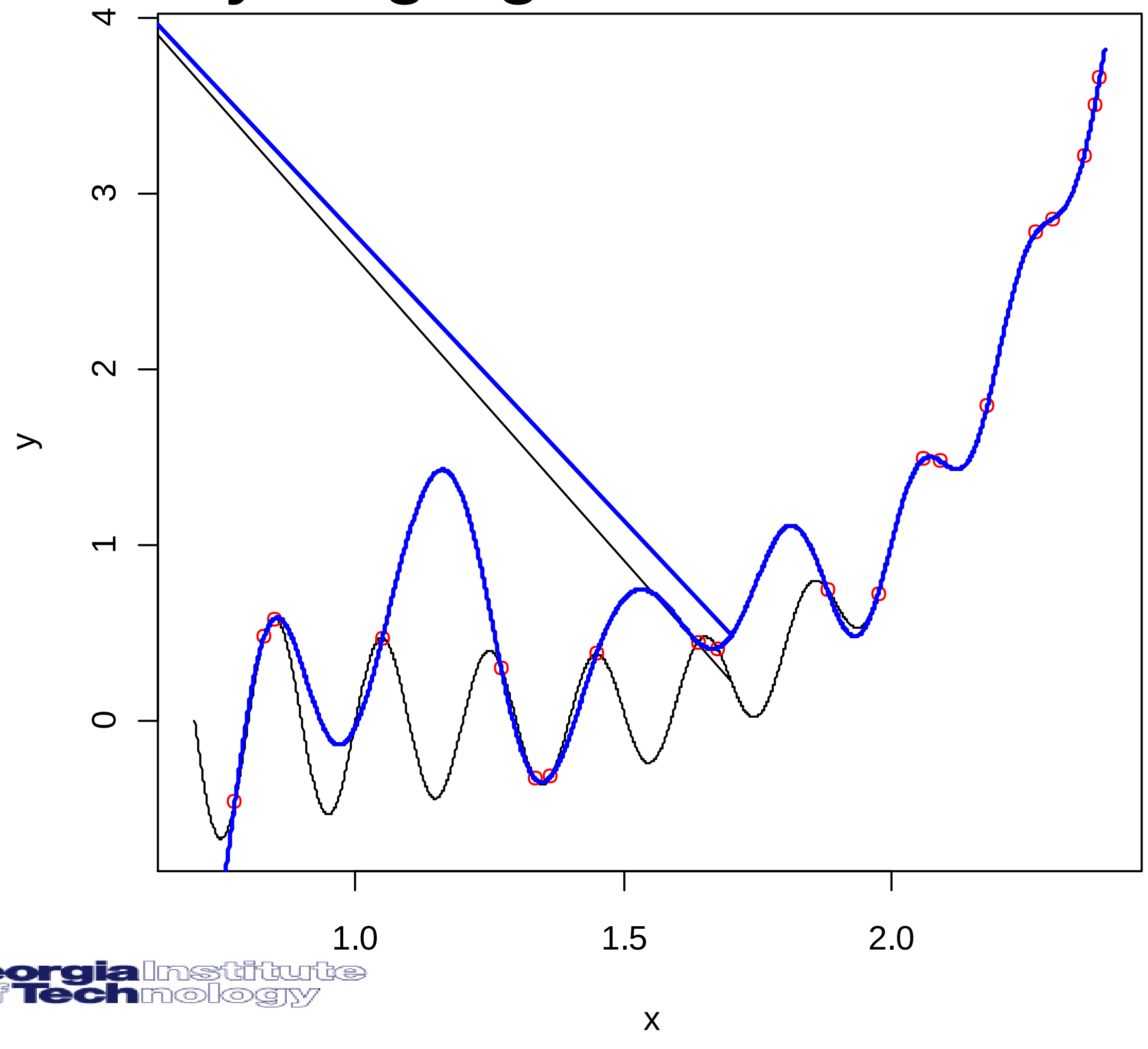


Georgia Institute of Technology

ISyE INDUSTRIAL & SYSTEMS ENGINEERING

An Example

- Ordinary kriging

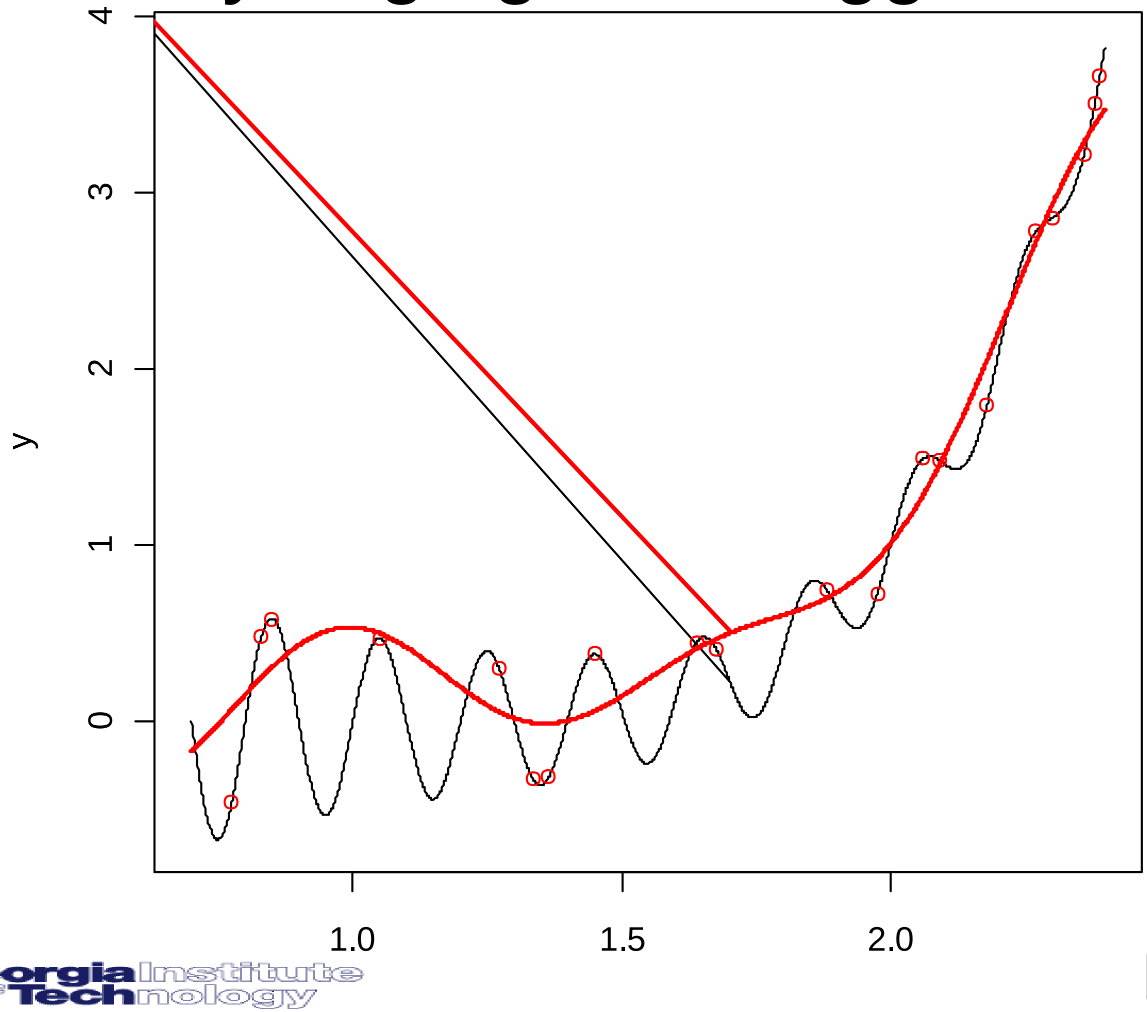


Georgia Institute of Technology

ISyE INDUSTRIAL & SYSTEMS ENGINEERING

An Example

- Ordinary kriging with nugget

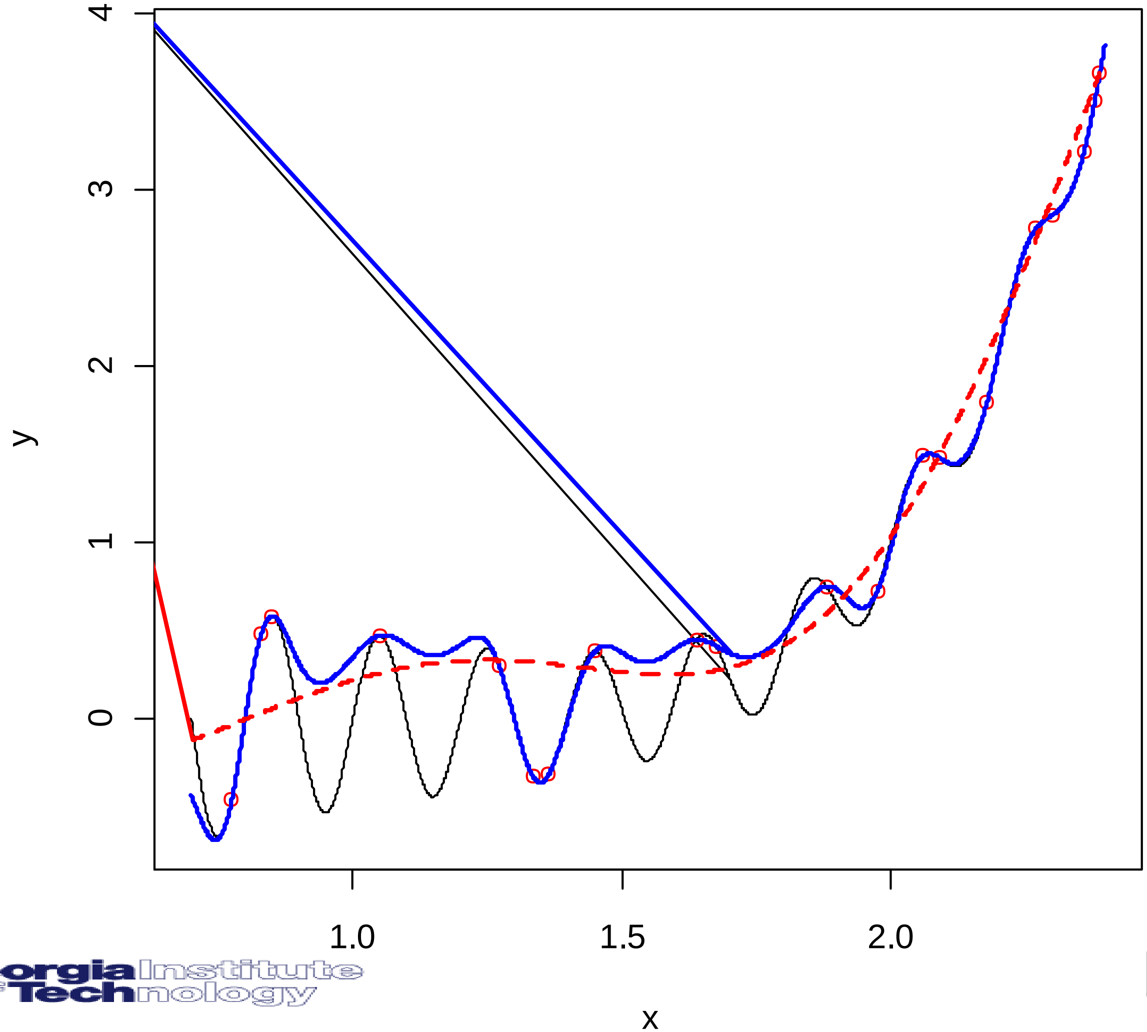


Georgia Institute of Technology

ISyE INDUSTRIAL & SYSTEMS ENGINEERING

An Example

- CGP



Georgia Institute of Technology

INDUSTRIAL & SYSTEMS ENGINEERING

Improving Variance Model

$$Y(\mathbf{x}) = Z_{global}(\mathbf{x}) + \sigma^*(\mathbf{x}) Z_{local}(\mathbf{x}),$$
$$Z_{global}(\mathbf{x}) \sim GP(\mu, \tau^2 g),$$
$$Z_{local}(\mathbf{x}) \sim GP(0, l).$$

How to Obtain Variance Model?

- Obtain residuals from the global trend

$$\mathbf{s}^2 = (s_1^2, \ldots, s_n^2)$$

- Kernel regression

$$\sigma^2(\mathbf{x}) = \frac{\mathbf{g}_b^\top(\mathbf{x})\mathbf{s}^2}{\mathbf{g}_b^\top(\mathbf{x})\mathbf{1}},$$

where $\mathbf{g}_b(\mathbf{x}) = \mathbf{g}(\mathbf{x})$ if $b = 1$, and $\mathbf{g}_b(\mathbf{x}) \to \mathbf{1}$ as $b \to 0$.

$$\sigma^{*2}(\mathbf{x}) = \lambda^* \sigma^2(\mathbf{x})$$

- Joseph and Kang (2011)

Prediction

$$\hat{y}(\mathbf{x}) = \hat{\mu} + (\tau^2 \mathbf{g}(\mathbf{x}) + \lambda^* \sigma(\mathbf{x}) \mathbf{\Sigma}^{1/2} \mathbf{l}(\mathbf{x}))^\top (\tau^2 \mathbf{G} + \lambda^* \mathbf{\Sigma}^{1/2} \mathbf{L} \mathbf{\Sigma}^{1/2})^{-1} (\mathbf{y} - \hat{\mu} \mathbf{1})$$

$$\mathbf{\Sigma} = diag\{\sigma_1^2, \ldots, \sigma_n^2\}$$

Prediction Interval

$$\hat{y}(\mathbf{x}) \pm Z_{\alpha/2} v_{0|n}(\mathbf{x})$$

$$v_{0|n}^2(\mathbf{x}) = \tau^2 \{1 + \lambda\sigma^2(\mathbf{x}) - \mathbf{q}^\top(\mathbf{x})\mathbf{Q}^{-1}\mathbf{q}(\mathbf{x}) + \frac{(1 - \mathbf{q}^\top(\mathbf{x})\mathbf{Q}^{-1}\mathbf{1})^2}{\mathbf{1}^\top\mathbf{Q}^{-1}\mathbf{1}}\}.$$

where $\mathbf{Q} = \mathbf{G} + \lambda\boldsymbol{\Sigma}^{1/2}\mathbf{L}\boldsymbol{\Sigma}^{1/2}$ and $\mathbf{q}(\mathbf{x}) = \mathbf{g}(\mathbf{x}) + \lambda\sigma(\mathbf{x})\boldsymbol{\Sigma}^{1/2}\mathbf{l}(\mathbf{x})$.

Estimation

Maximum likelihood method

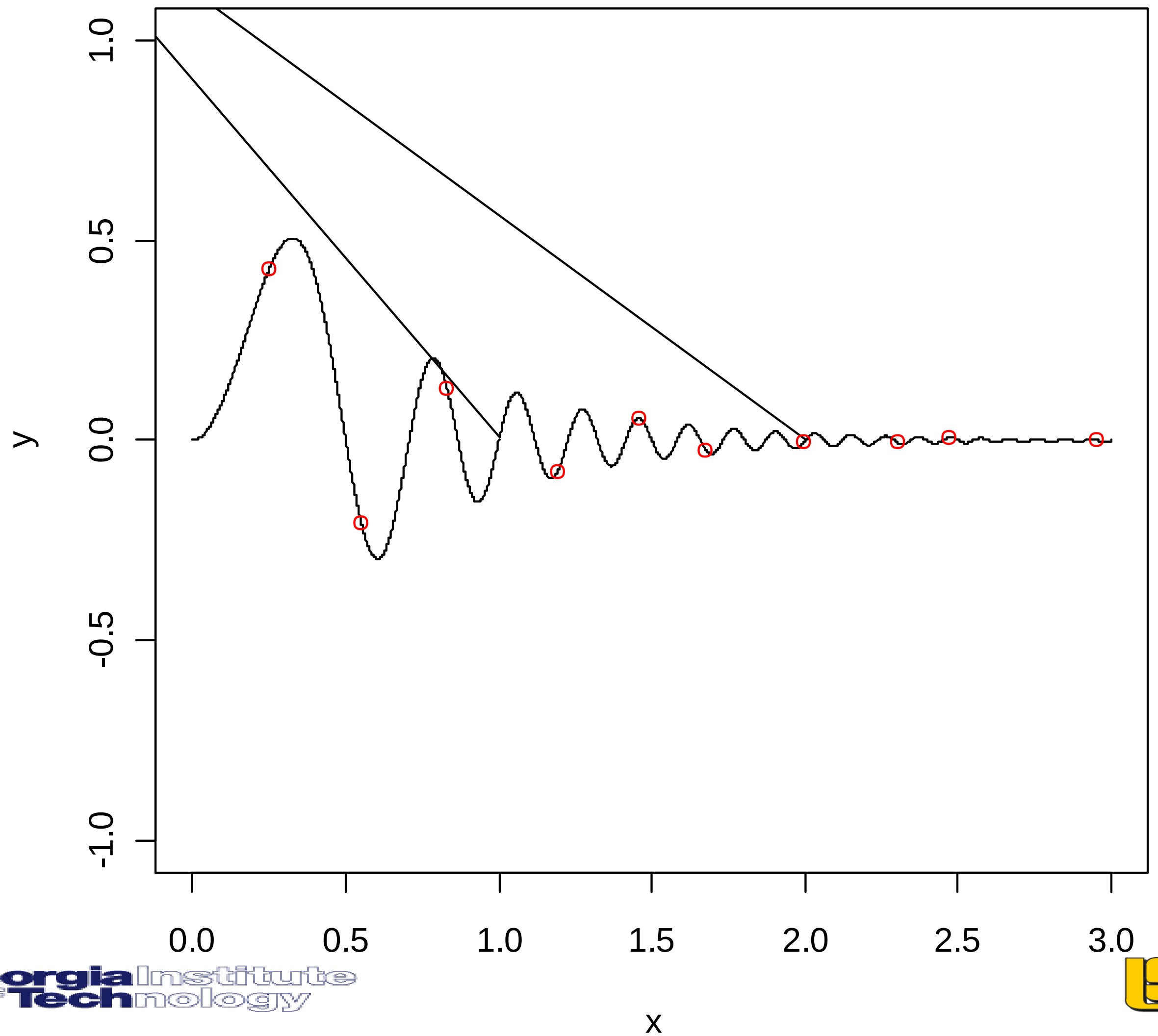
For each set of values $(\lambda, \boldsymbol{\theta}, \boldsymbol{\alpha}, b)$, we can

1. Obtain $\boldsymbol{\Sigma} = diag\{\sigma_1^2, \ldots, \sigma_n^2\}$ and $\sigma^2(\mathbf{x})$.
2. $\hat{\mu}(\lambda, \boldsymbol{\theta}, \boldsymbol{\alpha}, b) = (\mathbf{1}^\top(\mathbf{G} + \lambda\boldsymbol{\Sigma}^{1/2}\mathbf{L}\boldsymbol{\Sigma}^{1/2})^{-1}\mathbf{1})^{-1}(\mathbf{1}^\top(\mathbf{G} + \lambda\boldsymbol{\Sigma}^{1/2}\mathbf{L}\boldsymbol{\Sigma}^{1/2})^{-1}\mathbf{y})$
   $\hat{\tau}^2(\lambda, \boldsymbol{\theta}, \boldsymbol{\alpha}, b) = \frac{1}{n}(\mathbf{y} - \hat{\mu}\mathbf{1})^\top(\mathbf{G} + \lambda\boldsymbol{\Sigma}^{1/2}\mathbf{L}\boldsymbol{\Sigma}^{1/2})^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1})$.

Estimate $(\lambda, \boldsymbol{\theta}, \boldsymbol{\alpha}, b)$ by minimizing

$$\phi(\lambda, \boldsymbol{\theta}, \boldsymbol{\alpha}, b) = n\log(\hat{\tau}^2(\lambda, \boldsymbol{\theta}, \boldsymbol{\alpha})) + \log(\det(\mathbf{G} + \lambda\boldsymbol{\Sigma}^{1/2}\mathbf{L}\boldsymbol{\Sigma}^{1/2})).$$

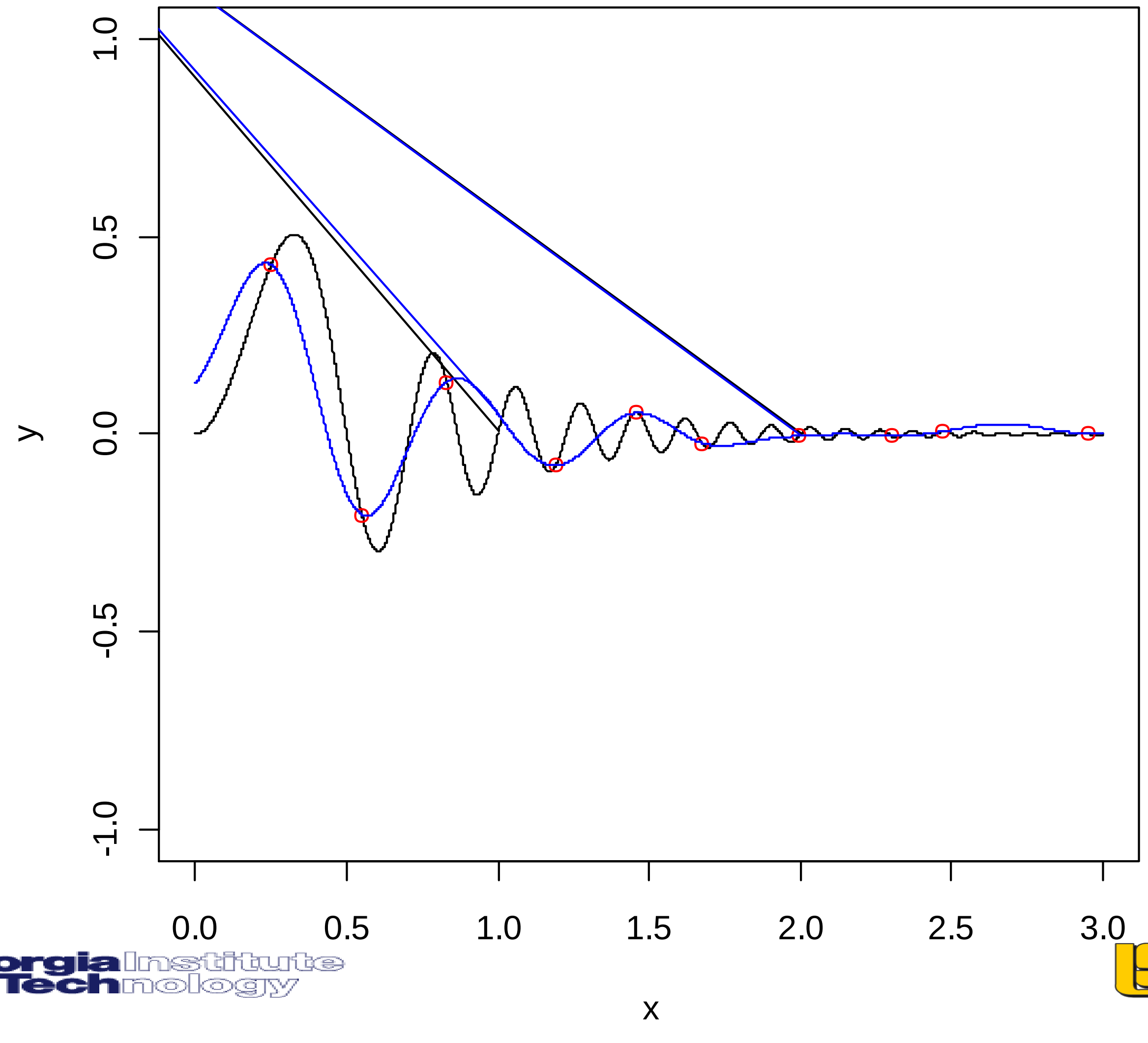
An Example



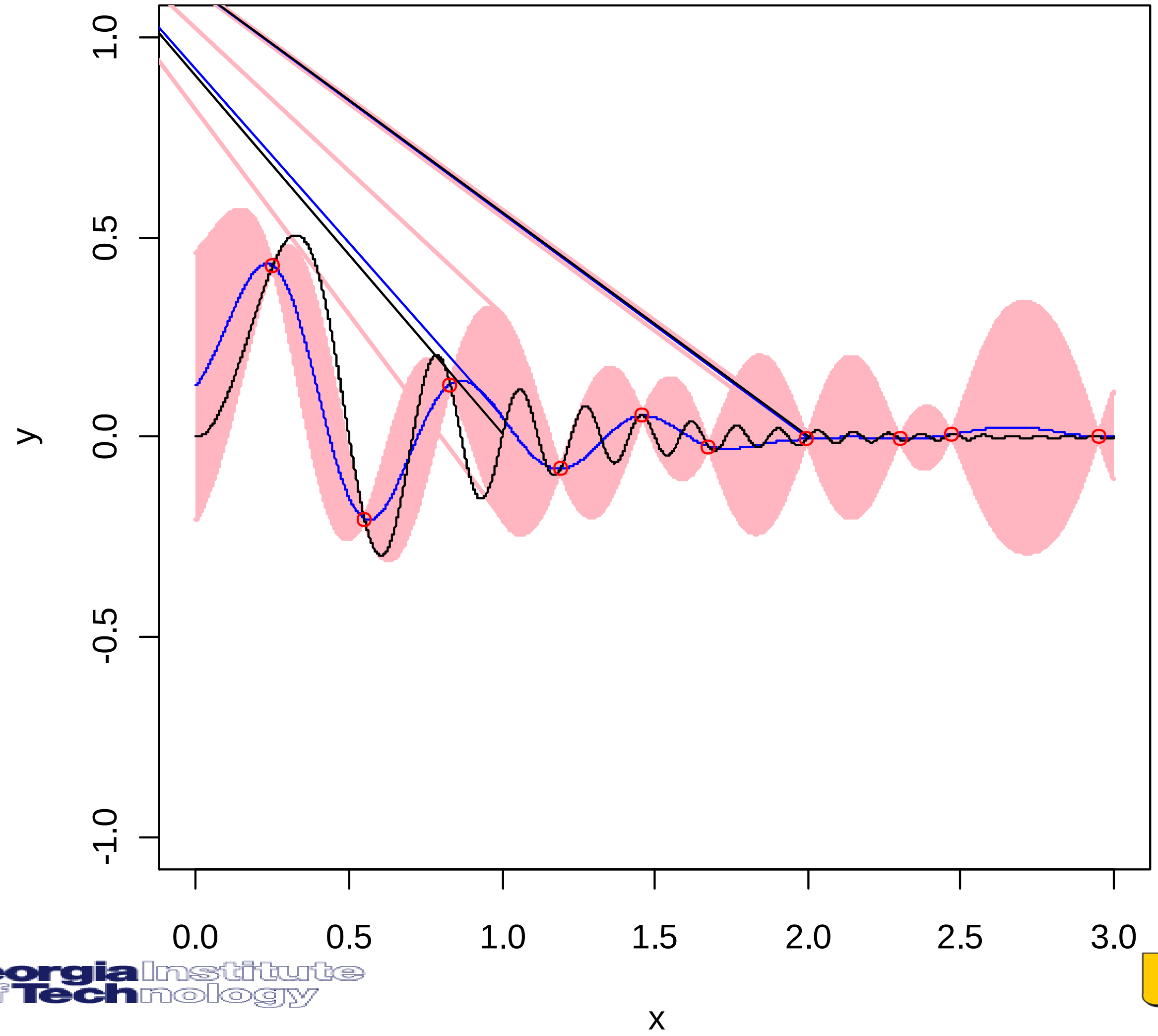
Georgia Institute of Technology

ISyE INDUSTRIAL & SYSTEMS ENGINEERING

GP Model



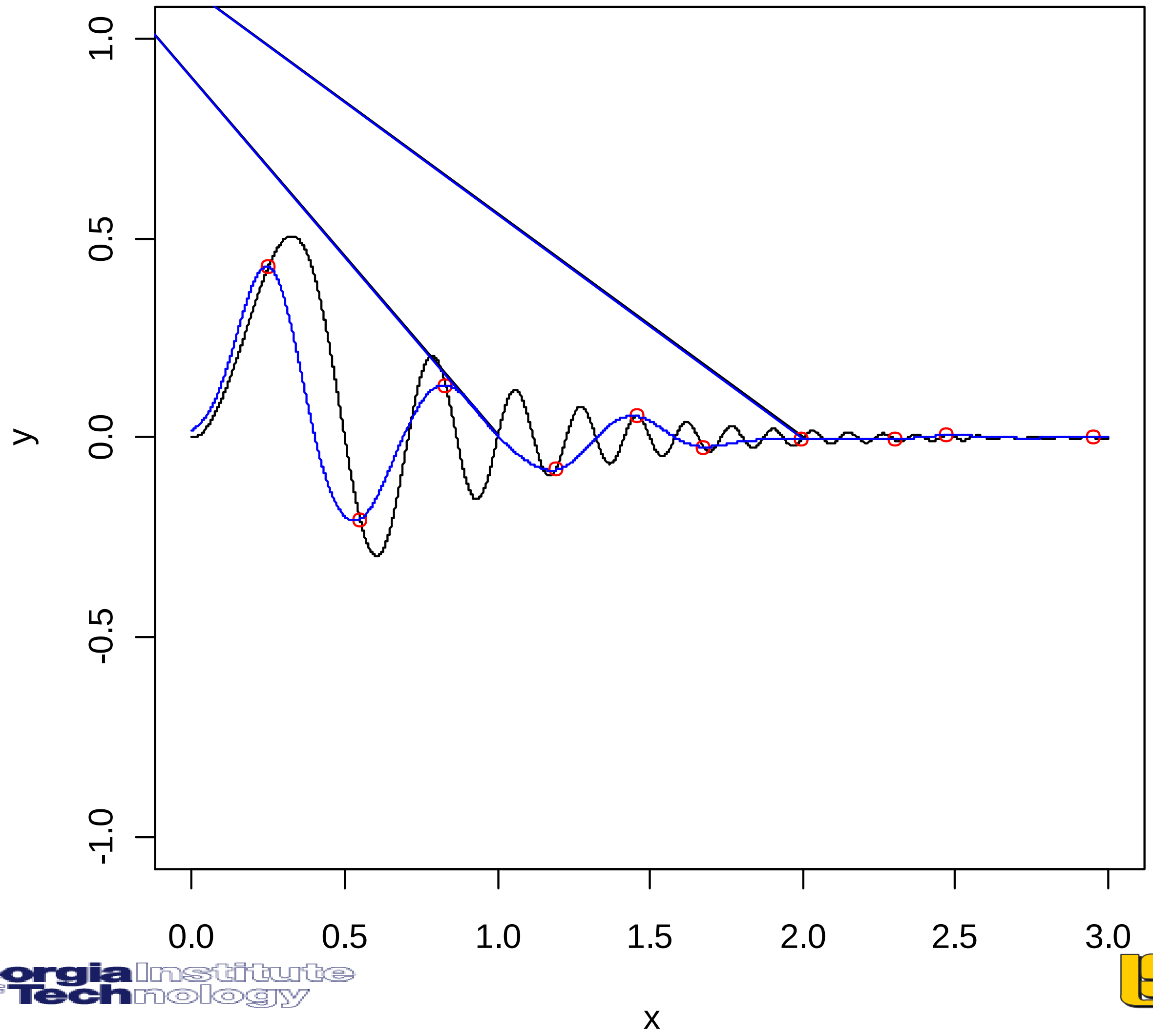
GP: Prediction Interval



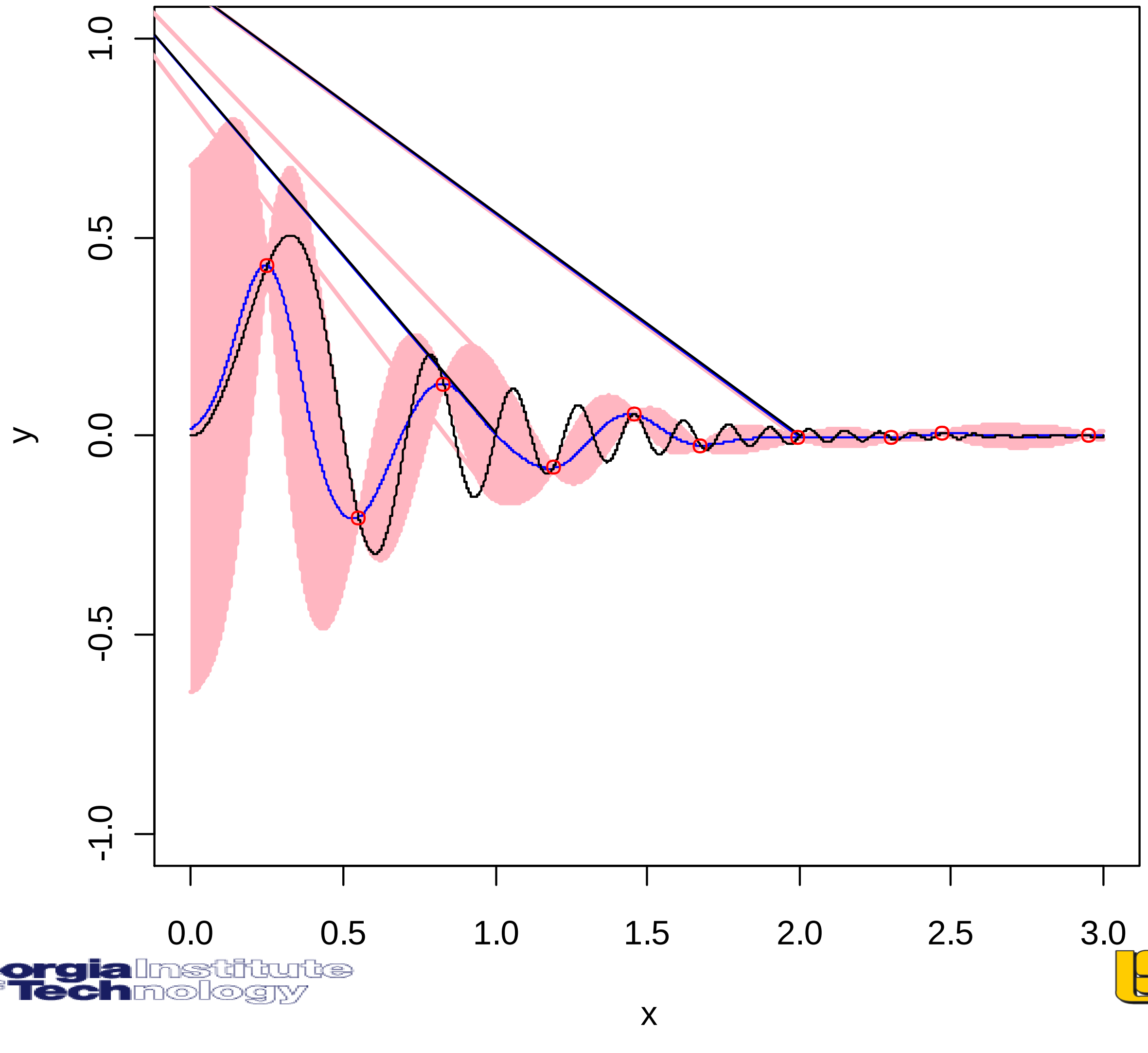
Georgia Institute of Technology

INDUSTRIAL & SYSTEMS ENGINEERING

CGP



CGP: Prediction Interval



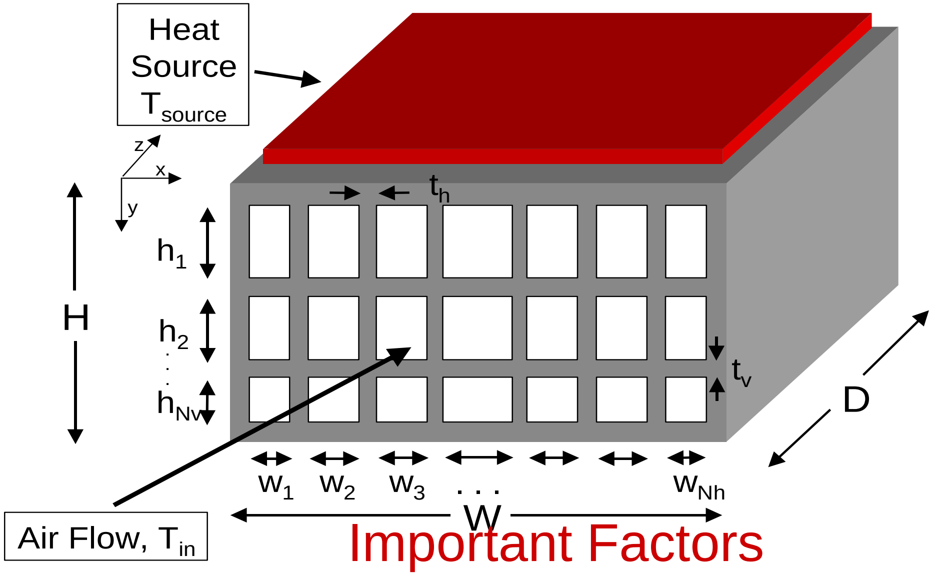
Borehole Example

$$y = \frac{2\pi T_u(H_u - H_l)}{ln(r/r_w)[1 + \frac{2LT_u}{ln(r/r_w)r_w^2 K_w} + \frac{T_u}{T_l}]}$$

Method	27-run OA	64-run MLD	80-run LHD
Ordinary kriging	5.96	1.68	0.46
Blind kriging	5.17	1.54	0.66
TGP	12.81	21.88	3.59
CGP	2.40	1.24	0.46

Designing Cellular Heat Exchangers for an Electronic Cooling Application

Qian, Seepersad, Joseph, Allen, Wu (2006)



Important Factors

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

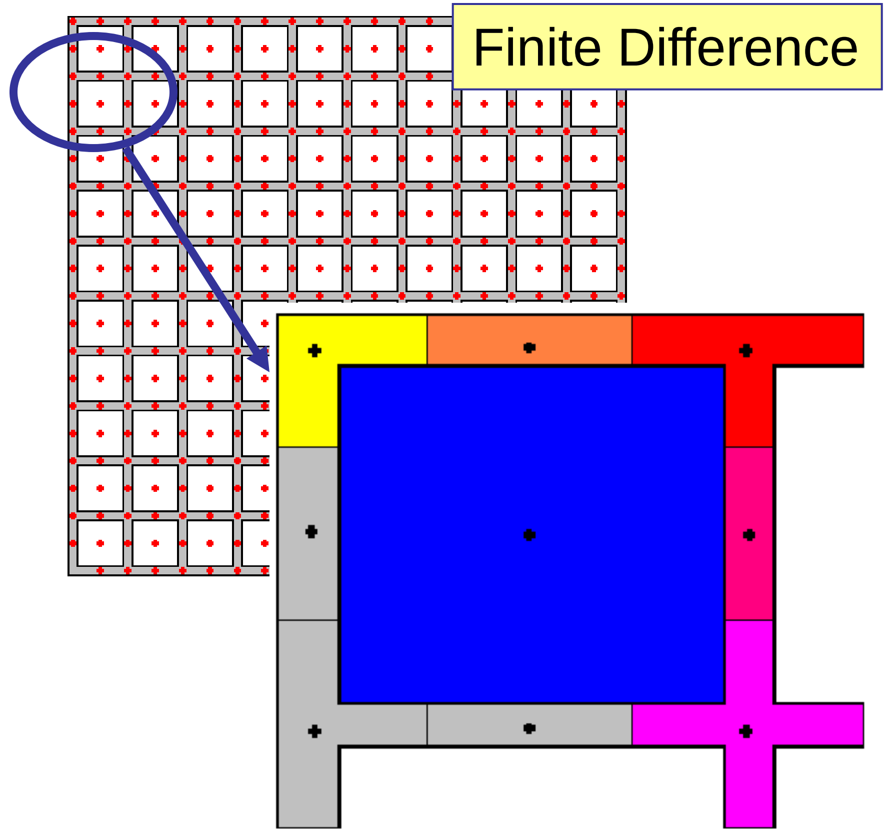
Objective

- Maximum Total Heat Transfer

Heat Transfer Analysis

A Quick Simulation Approach—Finite Difference*

- The finite difference technique is a numerical technique for solving 2- or 3-D steady state heat transfer problems.
- Planar cross section is discretized with a matrix of nodes. Nodal matrix repeated at regular intervals along length.
- Temperature distribution approximated via numerical solution of 3D heat transfer equations using forward or central difference methods.



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

Georgia Institute of Technology

ISyE INDUSTRIAL & SYSTEMS ENGINEERING

Results

- OA-based LHD (64,5)
- 14 test data

Method	RMSPE
Ordinary Kriging	5.15
Blind Kriging	2.59
CGP	2.31

Conclusions

- Coupled Gaussian Processes
  - Improved predictions
  - Improved prediction intervals
- Disadvantage
  - Increased computational cost
  - Small datasets