Bayesian Hierarchical Modeling for Integrating Low-accuracy and High-accuracy Experiments

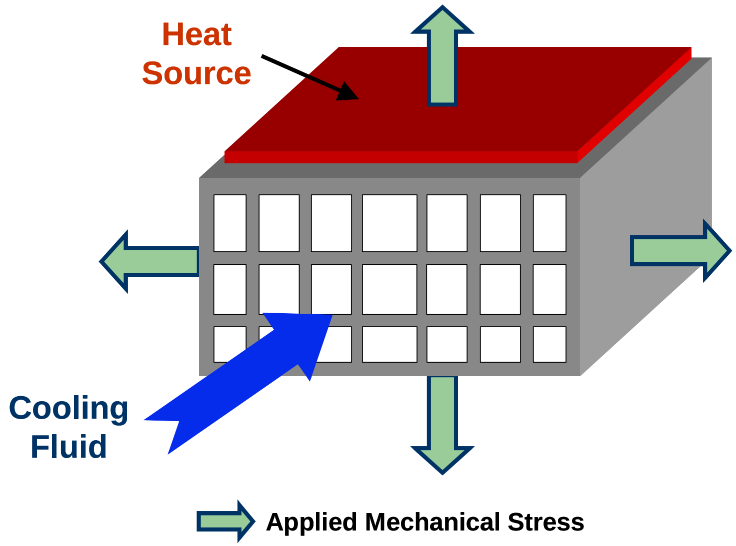
C. F. Jeff Wu
School of Industrial and Systems Engineering,
Georgia Institute of Technology
(Joint work with Zhiguang Qian)

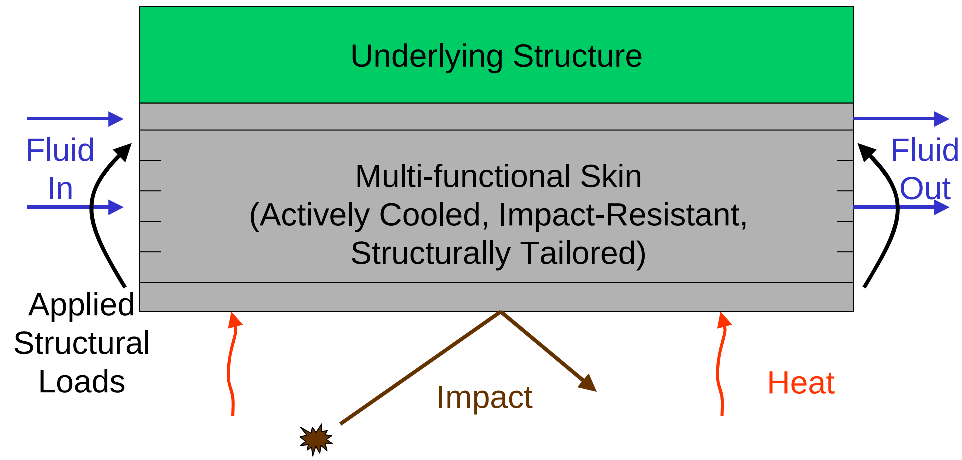
Papers available at www.isye.gatech.edu/~jeffwu

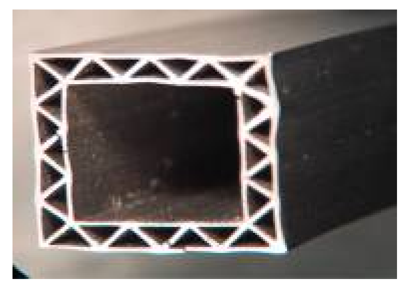
Outline

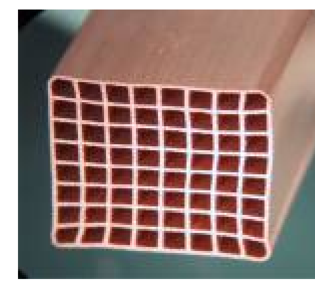
- Motivating examples:
  - cellular heat exchangers for electronic cooling.
  - data centers.
- A general strategy to integrate data from high-accuracy and low-accuracy experiments .
- Bayesian hierarchical Gaussian process modeling and estimation.
- Illustrations.
- Conclusions.

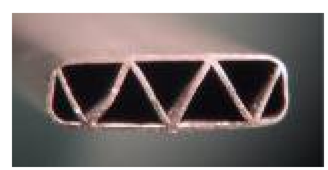
Example 1: Designing Cellular Heat Exchangers for an Electronic Cooling Application











Linear Cellular Alloys

Important Factors:

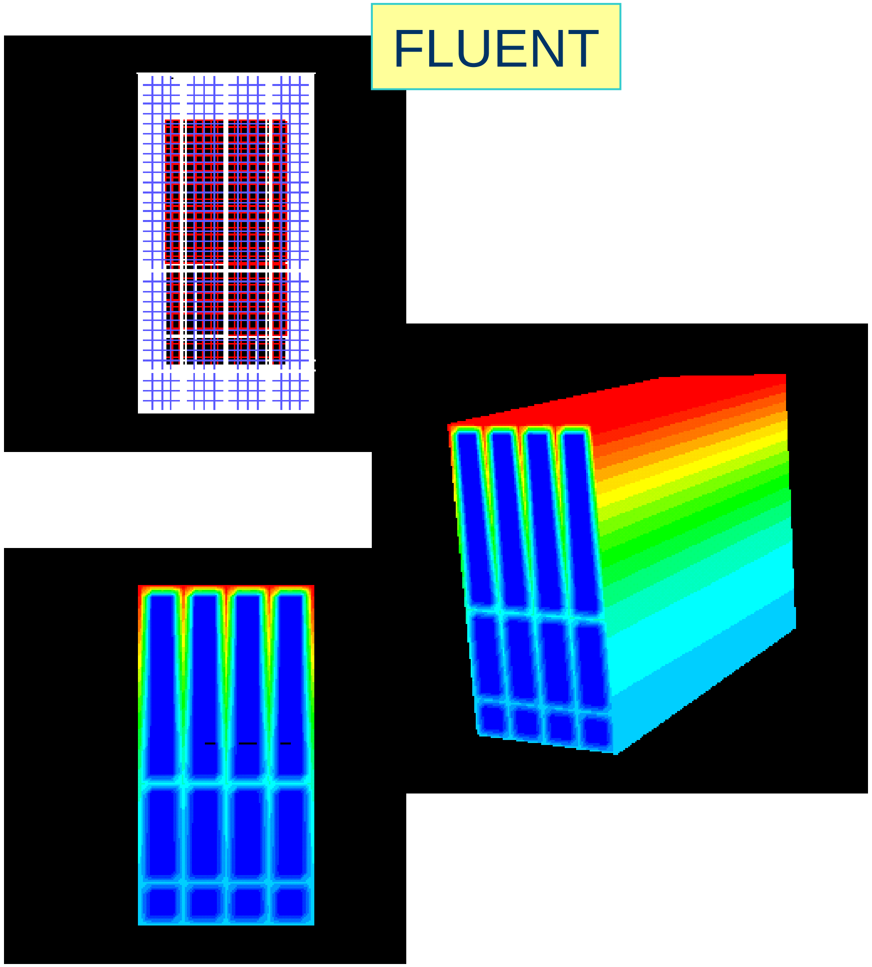
- Flow-rate of Air
- Inlet Temp of Air
- Conductivity of Solid
- Temp of Upper Wall

Response: Total Heat Transfer Rate from Solid to Air

Courtesy of Systems Realization Lab at Georgia Tech

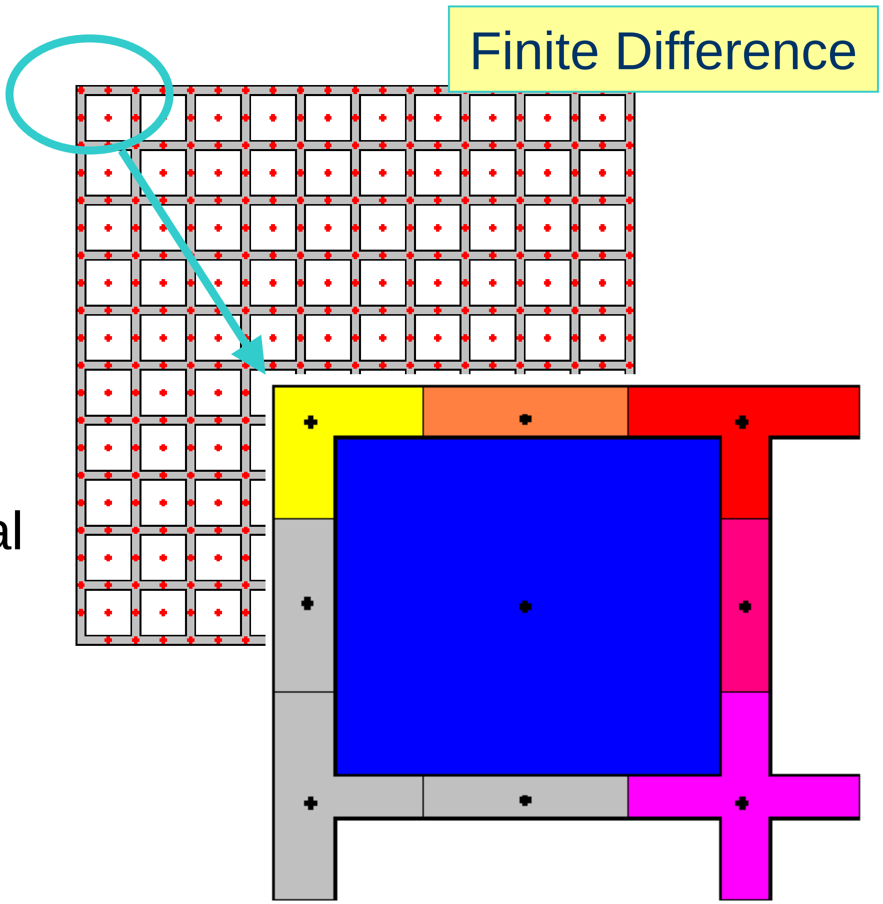
Heat Transfer Analysis
HE: Detailed Computer Simulation – Finite Element Analysis (FEA) Method

- Using the computational fluid dynamic solver FLUENT
- Problem domain is divided into thousands or millions of elements.
- Each run requires hours to days to complete.

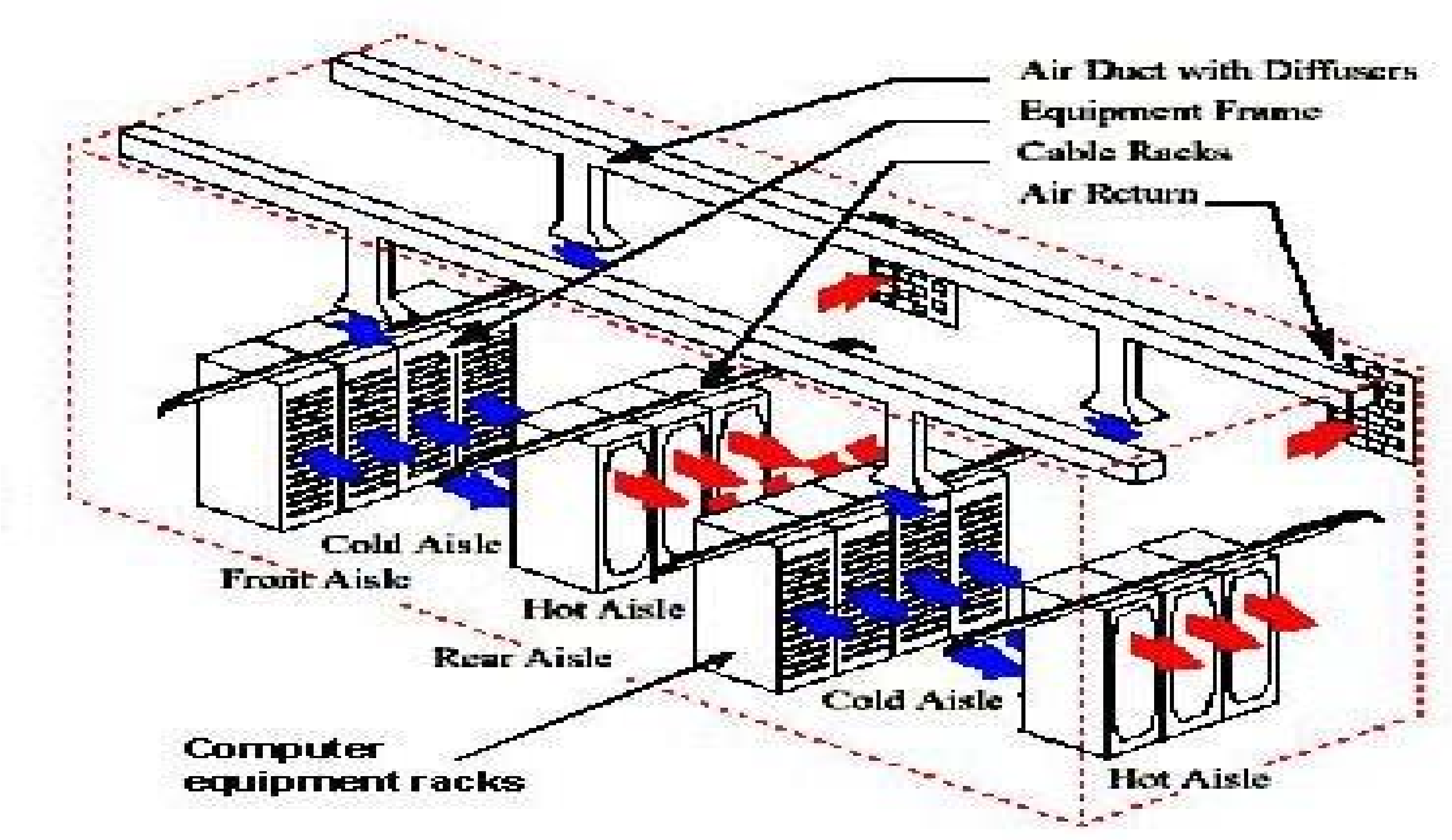


Heat Transfer Analysis
LE: Approximate Computer Simulation —Finite Difference Method

- The finite difference technique is a numerical technique for solving 2- or 3-D steady state heat transfer problems.
- Temperature distribution approximated via numerical solution of 3D heat transfer equations using forward or central difference methods.
- Each run takes minutes to complete.
- Less accurate than FEA.



Example 2: Modeling Thermal Distribution of a Data Center



Courtesy of IBM T. J. Waston Research Center

HE and LE for Modeling Data Center Thermal Distribution

HE: Physical experiment

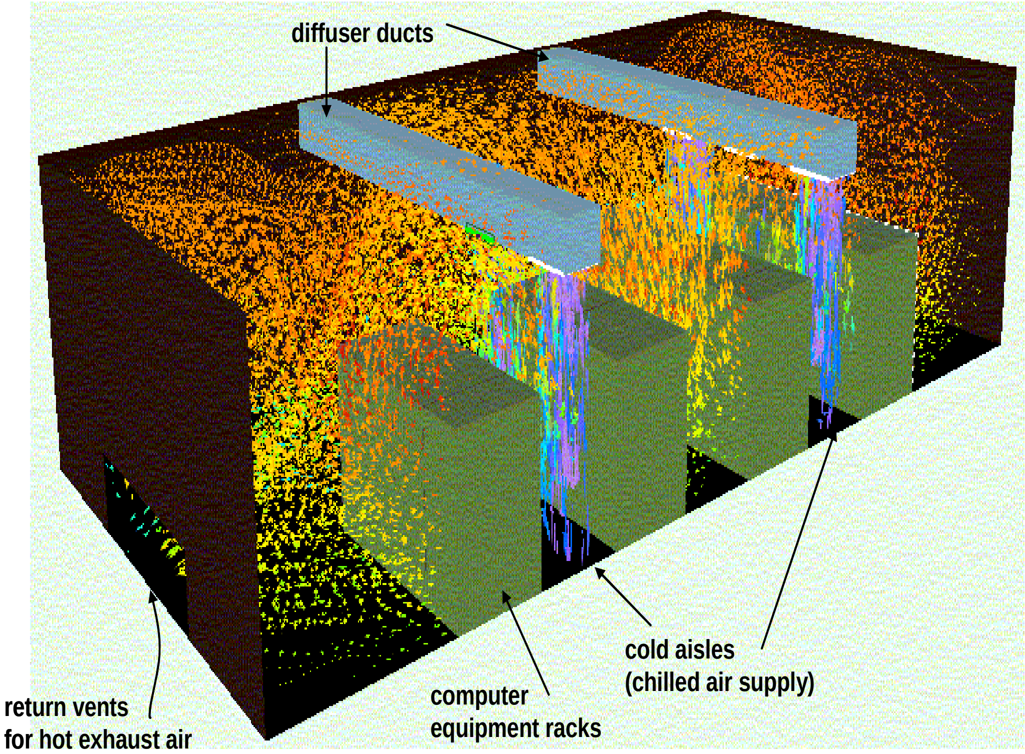
• Building a large data center costs a few million dollars and takes several months to complete.

LE: Computer simulation based on computational fluid dynamics, Flotherm

• Price for Flotherm: less than $2000.

• Each run takes hours to complete.

• Result is not as accurate as the physical experiment.



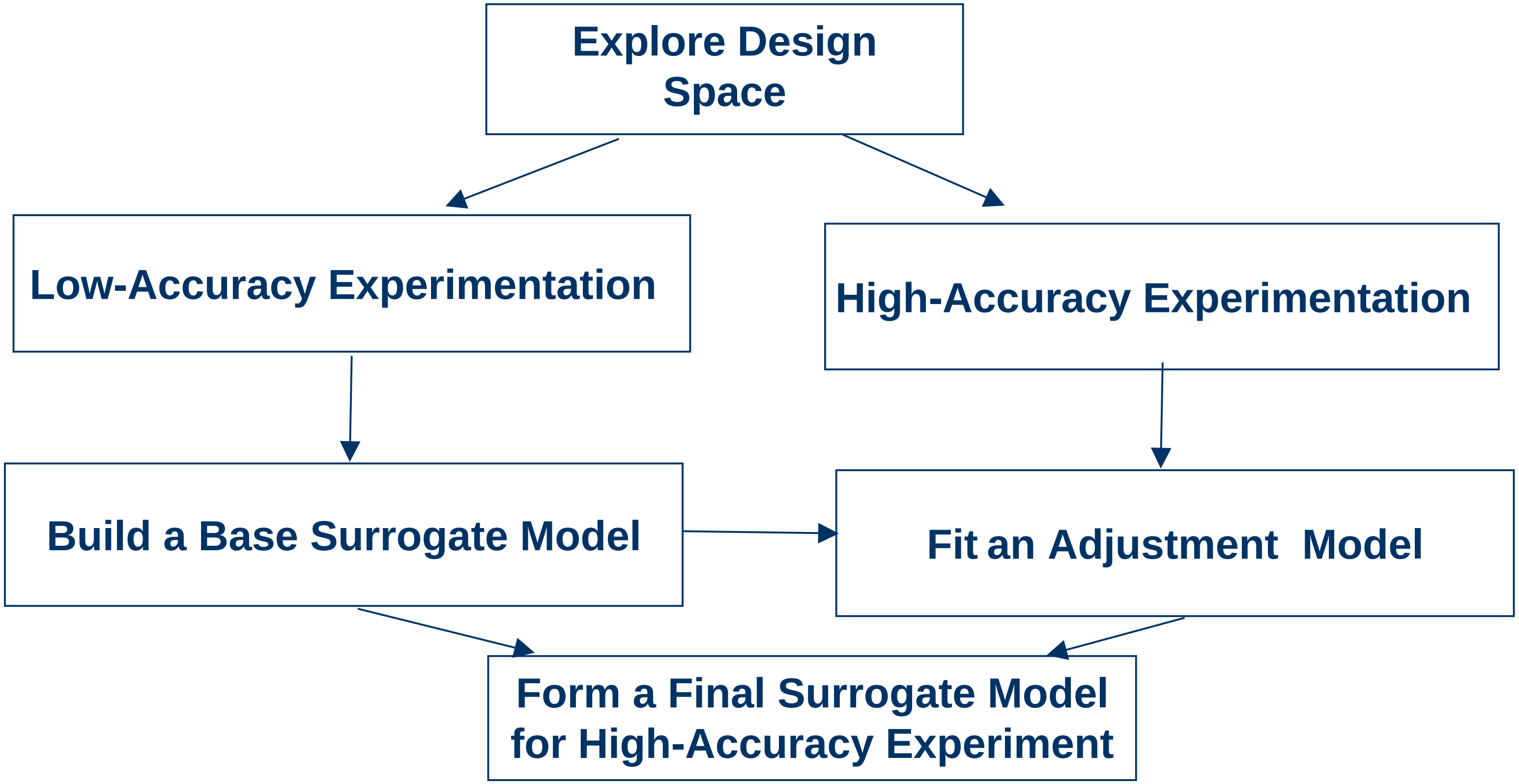
HE and LE

- A generic pair: high-accuracy experiment (HE) and low-accuracy experiment (LE).
- Pair one: detailed computer experiment vs. approximate computer experiment
  - Hydrocarbon reservoir example in Kennedy and O'Hagan (2000).
  - Cellular heat exchangers example in Qian, Seepersad, Joseph, Allen and Wu (2006).
- Pair two: physical experiment vs. computer experiment
  - Fluidized bed processes example in Reese et al. (2004).

Integration of HE and LE Data

- HE vs. LE:
  - HE runs (much) *slower* than LE.
  - HE is more *accurate* than LE.
- A general strategy:
  1. *Fit* a flexible model for LE data $\Rightarrow$ *Base Surrogate Model*.
  2. *Adjust* the fitted model in step 1 with HE data $\Rightarrow$ *Final Surrogate Model*.
- What methods for fitting and adjustment?

Schematic of Integrated Analysis of HE and LE Data



Gaussian Process Modeling

- Popular tool in computer experiment, global optimization and machine learning. Suitable for modeling **non-linear** phenomena.

- Data: k: the number of variables, n: the number of points, $\mathbf{x}_i = (x_{i1}, \cdots, x_{ik})$: sampled point i, $y_i = y(\mathbf{x}_i)$: response value. $\mathbf{y} = (y_1, \ldots, y_n)^t$.

- Model:

$$y(\mathbf{x}_i) = \sum_m \beta_m f_m(\mathbf{x}_i) + \varepsilon(\mathbf{x}_i),\ i = 1, \ldots, n,$$

  - $f_m(\mathbf{x})$'s: functions of $\mathbf{x}$, β_m's: unknown coefficients,
  - $\varepsilon \sim GP(0, \sigma^2, \phi)$: Gaussian process with mean zero, variance σ^2 and correlation parameters ϕ.

- Gaussian correlation function:

$$R(\mathbf{x}_i, \mathbf{x}_j) = \mathrm{corr}[\varepsilon(\mathbf{x}_i), \varepsilon(\mathbf{x}_j)] = \exp[-\sum_{h=1}^{k} \phi_h |\mathbf{x}_{ih} - \mathbf{x}_{jh}|^2].$$

Best Linear Unbiased Predictor (BLUP)

- At $\mathbf{x}$,

$$\widehat{y}(\mathbf{x}) = \mathbf{f}^t\widehat{\beta} + \mathbf{r}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{F}\widehat{\beta}).$$

  – $\mathbf{r} = (R(\mathbf{x}, \mathbf{x}_1), \ldots, R(\mathbf{x}, \mathbf{x}_n))^t$.

  – $\mathbf{f} = \mathbf{f}(\mathbf{x})$.

  – $\widehat{\beta} = (\mathbf{F}^t\mathbf{R}^{-1}\mathbf{F})^{-1}\mathbf{F}^t\mathbf{R}^{-1}\mathbf{y}$.

  – $\mathbf{R}$ is the $(n \times n)$ matrix with entries $R(\mathbf{x}_i, \mathbf{x}_j)$ for $i, j = 1, \ldots, n$.

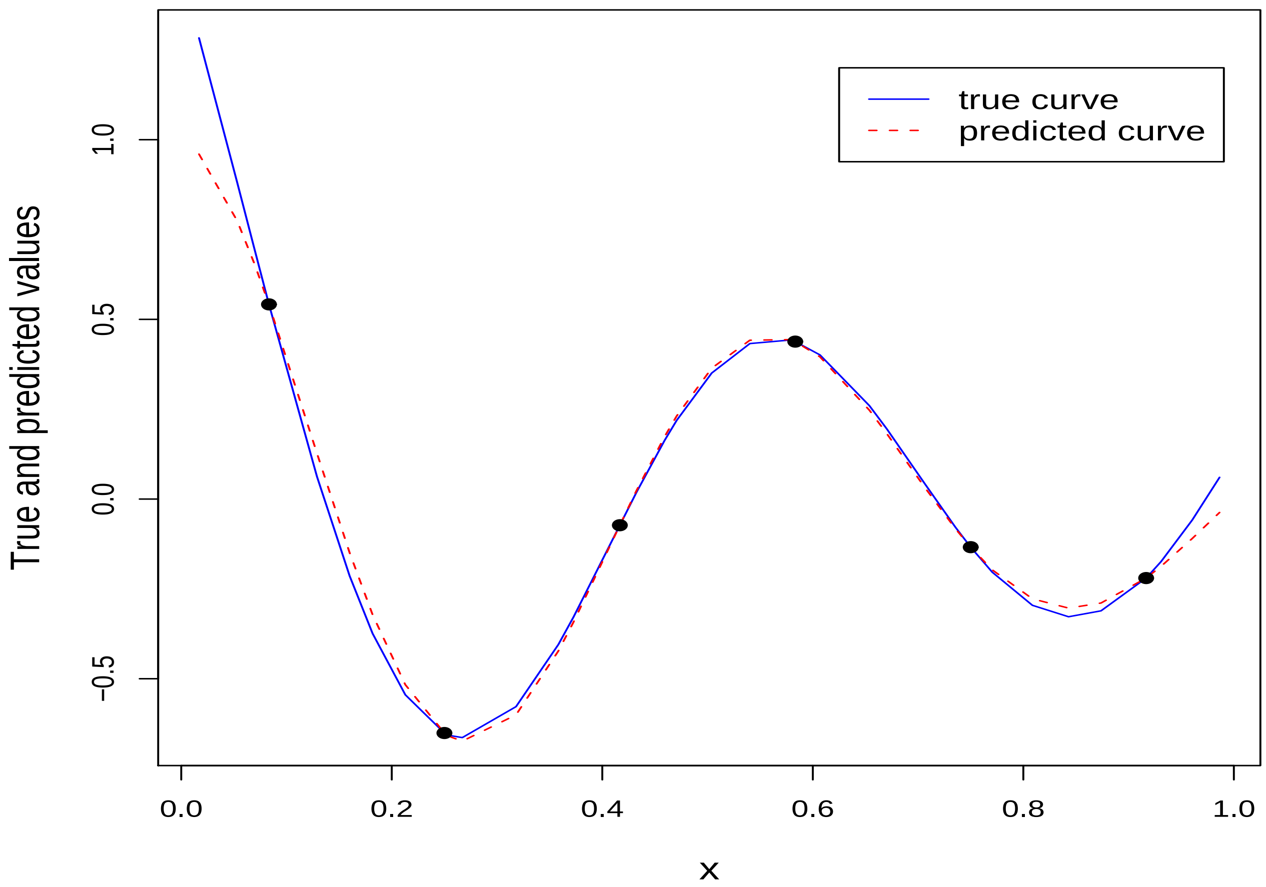
  – $\mathbf{F} = (\mathbf{f}(\mathbf{x}_1)^t, \ldots, \mathbf{f}(\mathbf{x}_n)^t)^t$.

- At observed $\mathbf{x}_i$, $\widehat{y}(\mathbf{x}_i) = y_i$.

- It has the property that it can **interpolate** observed data points $y_i, 1 = 1, \ldots, n$.

- This is desirable for computer experiments as there is no measurement error in y_i (i.e., same y_i for fixed x_i).

BLUP as an Interpolator



The true curve $y(x) = \exp(-1.4x)\cos(3.5\pi x)$ (solid blue line), a six point input design (dots), and the BLUP for $x \in (0,1]$ (dashed red line).

Notation

- D_l: set of design points ($\mathbf{x}_i$) for LE.
- D_h: set of design points ($\mathbf{x}_i$) for HE.
- Assume $D_h \subset D_l$.
- y_l: output from LE.
- y_h: output from HE.
- $\mathbf{y}_l$: LE data.
- $\mathbf{y}_h$: HE data.

Frequentist approach by Qian, Seepersad, Joseph, Allen and Wu (2006)

- Base Surrogate Model:

$$y_l = \beta_{l0} + \sum_j \beta_{lj}\mathbf{x}_i + \varepsilon_l(\mathbf{x}_i), \;\; \mathbf{x}_i \in D_l,$$

  - $\varepsilon_l \sim GP(0, \sigma_l^2, \phi_l)$.

- Adjustment Model:

$$y_h(\mathbf{x}_i) = \rho(\mathbf{x}_i) y_l(\mathbf{x}_i) + \delta(\mathbf{x}_i), \;\; \mathbf{x}_i \in D_h,$$

  - scale adjustment: $\rho(\mathbf{x}_i) = \rho_0 + \sum_j \rho_j x_{ij}$,
  - location adjustment: $\delta \sim GP(\delta_0, \sigma_\delta^2, \phi_\delta)$.

- Fitting:

$$\left.\begin{array}{r} \text{Base surrogate model: } \widehat{y}_l \\ \text{Adjustment models: } \widehat{\rho} \text{ and } \widehat{\delta} \end{array}\right\} \Longrightarrow \text{Final surrogate model: } \widehat{y}_h = \widehat{\rho}\widehat{y}_l + \widehat{\delta}$$

Bayesian approach by Qian and Wu (2006)

- Use flexible Bayesian hierarchical Gaussian process models.
- Can accommodate uncertainty in model parameters.
- Provide more flexible adjustment.
- Perform prediction using Bayesian predictive density function.

Bayesian Hierarchical Gaussian Process (BHGP) Models

- Base Surrogate Model:

$$y_l = \beta_{l0} + \sum_j \beta_{lj} \mathbf{x}_i + \varepsilon_l(\mathbf{x}_i), \ \mathbf{x}_i \in D_l,$$

  - $\varepsilon_l \sim GP(0, \sigma_l^2, \phi_l)$.

- Flexible Adjustment Model:

$$y_h(\mathbf{x}_i) = \rho(\mathbf{x}_i) y_l(\mathbf{x}_i) + \delta(\mathbf{x}_i) + \varepsilon(\mathbf{x}_i), \ \mathbf{x}_i \in D_h,$$

  - scale adjustment $\rho \sim GP(\rho_0, \sigma_\rho^2, \phi_\rho)$,
  - location adjustment $\delta \sim GP(\delta_0, \sigma_\delta^2, \phi_\delta)$,
  - $\varepsilon(\mathbf{x}) \sim N(0, \sigma_\varepsilon^2)$. No measurement error: $\varepsilon = 0$.

- Model parameters:
  $\theta_1 = (\beta_l, \rho_0, \delta_0)$, $\theta_2 = (\sigma_l^2, \sigma_\rho^2, \sigma_\delta^2, \sigma_\varepsilon^2)$ and $\theta_3 = (\phi_l, \phi_\rho, \phi_\delta)$.

Priors for the BHGP Models

$$
\begin{aligned}
p(\sigma_l^2) &\sim IG(\alpha_l, \beta_l), \\
p(\sigma_\rho^2) &\sim IG(\alpha_\rho, \beta_\rho), \\
p(\sigma_\delta^2) &\sim IG(\alpha_\delta, \beta_\delta), \\
p(\sigma_\varepsilon^2) &\sim IG(\alpha_\varepsilon, \beta_\varepsilon), \\
p(\beta_l | \sigma_l^2) &\sim MVN(\mathbf{u}_l, v_l \mathbf{I}_{(k+1)\times(k+1)} \sigma_l^2), \\
p(\rho_0 | \sigma_\rho^2) &\sim N(u_\rho, v_\rho \sigma_\rho^2), \\
p(\delta_0 | \sigma_\delta^2) &\sim N(u_\delta, v_\delta \sigma_\delta^2), \\
p(\phi_{l_i}) &\sim G(a_l, b_l),\ i = 1, \ldots, k, \\
p(\phi_{\rho_i}) &\sim G(a_\rho, b_\rho),\ i = 1, \ldots, k, \\
p(\phi_{\delta_i}) &\sim G(a_\delta, b_\delta),\ i = 1, \ldots, k.
\end{aligned}
$$

Bayesian Prediction

- Main goal: make prediction of y_h at an untried point $\mathbf{x}_0$. WLOG, assume $\mathbf{x}_0 \in D_l / D_h$.
- Motivations:
  - The posterior density of θ_3 is messy. Difficult to draw samples from the posterior density of $(\theta_1, \theta_2, \theta_3)$.
  - Empirical evidence: prediction of GP models not very sensitive to the specification of correlation parameters.
- Two-step procedure:
  - 1. Fit correlation parameters θ_3 at their posterior modes. **Use Stochastic Programming methods**.
  - 2. Make prediction conditionally on θ_3. **Use MCMC methods**.
- Bayesian predictive density function: $p(y_h(\mathbf{x}_0)|\mathbf{y}_h, \mathbf{y}_l)$.
  - Predicted value: $E(y_h(\mathbf{x}_0)|\mathbf{y}_h, \mathbf{y}_l)$.
  - Uncertainty of prediction: $Var(y_h(\mathbf{x}_0)|\mathbf{y}_h, \mathbf{y}_l)$.

Fitting Correlation Parameters θ_3 (1)

- Posterior mode $\widehat{\theta}_3 = \arg\max_{\theta_3} p(\theta_3|\mathbf{y}_h, \mathbf{y}_l)$.
- $p(\theta_3|\mathbf{y}_h, \mathbf{y}_l) \propto p(\theta_3) \int_{\theta_1,\theta_2} p(\theta_1, \theta_2) p(\mathbf{y}_l, \mathbf{y}_h|\theta_1, \theta_2, \theta_3) d\theta_1 d\theta_2$

$$
\begin{aligned}
L_1 \propto p(\theta_3) \int_{\tau_1,\tau_2} & \tau_1^{-(\alpha_\delta+\frac{3}{2})} \tau_2^{-(\alpha_\varepsilon+1)} |\mathbf{a}_1 \mathbf{R}_l \mathbf{M}|^{-\frac{1}{2}} (a_2 a_3)^{-\frac{1}{2}} \\
& \cdot (\gamma_l + \frac{4c_1 - \mathbf{b}_1^t \mathbf{a}_1^{-1} \mathbf{b}_1}{8})^{-(\alpha_l+\frac{n}{2})} \\
& \cdot (\gamma_\rho + \frac{\gamma_\delta}{\tau_1} + \frac{\gamma_\varepsilon}{\tau_2} + \frac{4a_3c_3 - b_3^2}{8a_3})^{-(\alpha_\rho+\alpha_\delta+\alpha_\varepsilon+\frac{n_1}{2})} d\tau_1 d\tau_2.
\end{aligned}
$$

-

$$
(P): \max_{\phi_l,\phi_\rho,\phi_\delta} L_1 \Leftrightarrow \left\{ \begin{array}{c} (P_1): \max_{\phi_l} p(\phi_l)|\mathbf{R}_l \mathbf{a}_1|^{-\frac{1}{2}} (\gamma_l + \frac{4c_1 - \mathbf{b}_1^t \mathbf{a}_1^{-1} \mathbf{b}_1}{8})^{-(\alpha_l+\frac{n}{2})} \\ (P_2): \max_{\phi_\rho,\phi_\delta} L_2 \end{array} \right.
$$

Fitting Correlation Parameters θ_3 (2)

- (P_1) can solved by some quasi-Newton algorithms.

-

$$L_2 = \int_{\tau_1,\tau_2} p(\phi_\rho)p(\phi_\delta)\tau_1^{-(\alpha_\delta+\frac{3}{2})}\tau_2^{-(\alpha_\varepsilon+1)}|\mathbf{M}|^{-\frac{1}{2}}(a_2a_3)^{-\frac{1}{2}}$$

$$\cdot(\gamma_\rho+\frac{\gamma_\delta}{\tau_1}+\frac{\gamma_\varepsilon}{\tau_2}+\frac{4a_3c_3-b_3^2}{8a_3})^{-(\alpha_\rho+\alpha_\delta+\alpha_\varepsilon+\frac{n_1}{2})}d\tau_1 d\tau_2.$$

-

$$(P_2)\ \max_{\phi_\rho,\phi_\delta} L_2 \Leftrightarrow (P_2')\ \max_{\phi_\rho,\phi_\delta} L_2 = E_{\tau_1,\tau_2} f(\tau_1,\tau_2),$$

$$f(\tau_1,\tau_2) = \frac{p(\phi_\rho)p(\phi_\delta)\exp(\frac{2}{\tau_1})\exp(\frac{2}{\tau_2})}{|\mathbf{M}|^{\frac{1}{2}}(a_2a_3)^{\frac{1}{2}}(\beta_\rho+\frac{\beta_\delta}{\tau_1}+\frac{\beta_\varepsilon}{\tau_2}+\frac{4a_3c_3-b_3^2}{8a_3})^{\alpha_\rho+\alpha_\delta+\alpha_\varepsilon+\frac{n_1}{2}}},$$

$p(\tau_1) \sim IG(\alpha_\delta+\frac{1}{2},2)$, $p(\tau_2) \sim IG(\alpha_\varepsilon,2)$, and $p(\tau_1)\perp p(\tau_2)$.

Solving (P_2') : $\max_{\phi_\rho,\phi_\delta} L_2 = E_{\tau_1,\tau_2} f(\tau_1,\tau_2)$

- Stochastic programming: optimization under uncertainty.
- Sample Average Approximation (SAA) algorithm:
  - 1. Generate random samples (τ_1^s,τ_2^s) from $p(\tau_1,\tau_2), s=1,\cdots,S$.
  - 2. Solve the approximate problem:

$$(\widetilde{\phi}_\rho,\widetilde{\phi}_\delta) = \arg\max_{\phi_\rho,\phi_\delta}[\widehat{L}_2 = \frac{1}{S}\sum_{s=1}^{S} f(\tau_1^s,\tau_2^s)].$$

- Validations of SAA solutions:
  - Statistical verification of KKT conditions.
  - Construction of statistical bounds on optimal values.
- Faster computation than the fully Bayesian approach for large k.

More on SAA

Reference: Qian and Shapiro (2006).

Convergence analysis.

- Consistency of SAA solutions.
- Exponential convergence:

  Theorem 1. For any $\varepsilon > 0$, there exist positive constants $C(\varepsilon)$ and $\beta(\varepsilon)$, independent of S, such that

$$\text{Prob}\left\{ \sup_{\phi_\rho, \phi_\delta} \left| \hat{L}_2 - L_2 \right| \geq \varepsilon \right\} \leq C(\varepsilon) \exp\{-S\beta(\varepsilon)\}.$$

Validation of SAA solutions.

- Calculation of optimality gaps.
- Statistical verification of the Karush-Kuhn-Tucker (KKT) conditions.

MCMC Sampling from $p(\theta_1, \theta_2 | \mathbf{y}_l, \mathbf{y}_h, \theta_3)$

- Full conditional distributions for $(\beta_l, \delta_0, \rho_0, \sigma_l^2, \sigma_\rho^2, \tau)$:
  - $p(\beta_l | \mathbf{y}_l, \mathbf{y}_h, \theta_3, \overline{\beta_l}) \sim Normal$,
  - $p(\rho_0 | \mathbf{y}_l, \mathbf{y}_h, \theta_3, \overline{\rho_0}) \sim Normal$,
  - $p(\delta_0 | \mathbf{y}_l, \mathbf{y}_h, \theta_3, \overline{\delta_0}) \sim Normal$,
  - $p(\sigma_l^2 | \mathbf{y}_l, \mathbf{y}_h, \theta_3, \overline{\sigma_l^2}) \sim IG$,
  - $p(\sigma_\rho^2 | \mathbf{y}_l, \mathbf{y}_h, \theta_3, \overline{\sigma_\rho^2}) \sim IG$,
  - $p(\tau_1, \tau_2 | \mathbf{y}_l, \mathbf{y}_h, \overline{\tau_1, \tau_2}) \propto \frac{1}{\tau_1^{\alpha_\delta + \frac{3}{2}}} \frac{1}{\tau_2^{\alpha_\varepsilon + 1}} \exp\{-\frac{1}{\tau_1}(\frac{\gamma_\delta}{\sigma_\rho^2} + \frac{(\delta_0 - u_\delta)^2}{v_\delta \sigma_\rho^2}) - \frac{\gamma_\varepsilon}{\tau_2 \sigma_\rho^2} \frac{1}{|\mathbf{M}|^{\frac{1}{2}}}\}$

    $$\cdot \exp\{-\frac{(\mathbf{y}_h - \rho_0 \mathbf{y}_{l_1} - \delta_0 \mathbf{1}_{n_1})^t \mathbf{M}^{-1} (\mathbf{y}_h - \rho_0 \mathbf{y}_{l_1} - \delta_0 \mathbf{1}_{n_1})}{2\sigma_\rho^2}\}$$

- Use the Metropolis-within-Gibbs algorithm.
- Can now compute Bayesian predictive density function.

Case Study 1: Detailed and Approximate Computer Experiments for Cellular Heat Exchangers

- Response: Q = total heat transfer rate from solid to air.
- 4 input variables

$\dot{m}$ (flow-rate of air)	[0.00055, 0.001]
Tin (inlet temp of air)	[270.00, 303.15]
k (thermal conductivity of solid)	[202.4, 360.0]
T_{wall} (temp of upper wall)	[330, 400]

- 32 HE runs (using Finite Difference software).
- 32 LE runs (using FLUENT software).
- Training set: 24 randomly selected HE runs + 32 LE runs.
- Objective: Predict y_h at the remaining 8 runs.

Data of Cellular Heat Exchangers Example

Run #	$\dot{m}(kg/s)$	$T_{in}(K)$	$k(W/mk)$	$T_{wall}(K)$	y_l	y_h	Status
1	0.0005	293.15	362.73	393.15	25.601	23.54	Test
2	0.00055	315	310	365	21.23	20.15	Train
3	0.000552	293.53	318.63	388.29	11.44	10.17	Train
4	0.000557	290.18	298.27	377.49	15.03	15.29	Test
5	0.00056	277.01	354.98	374	18.55	18.39	Train
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
31	0.000751	287.99	326.02	354.08	36.56	47.05	Train
32	0.000757	300.64	235.03	391.68	27.24	25.82	Train

- Input values are standardized.

Hyperparameters of Priors

Par	Value	Par	Value
α_l	2	u_δ	0
γ_l	1	v_δ	1
α_ρ	2	a_l	2
γ_ρ	1	b_l	0.1
α_δ	2	a_ρ	2
γ_δ	1	b_ρ	0.1
$\mathbf{u}_l$	$(0,0,0,0,0)^t$	a_δ	2
v_l	1	b_δ	0.1
u_ρ	1		
v_ρ	1		

Implementation of the BHGP model

- Step 1: Use SAA to compute posterior modes of θ_3.

Parameter	Posterior Mode
ϕ_l	(2.83, 2.13,22.65,12.87)
ϕ_ρ	(3.22,7.23,1.26,1.38)
ϕ_δ	(2.26,0.74,6.92,7.24)

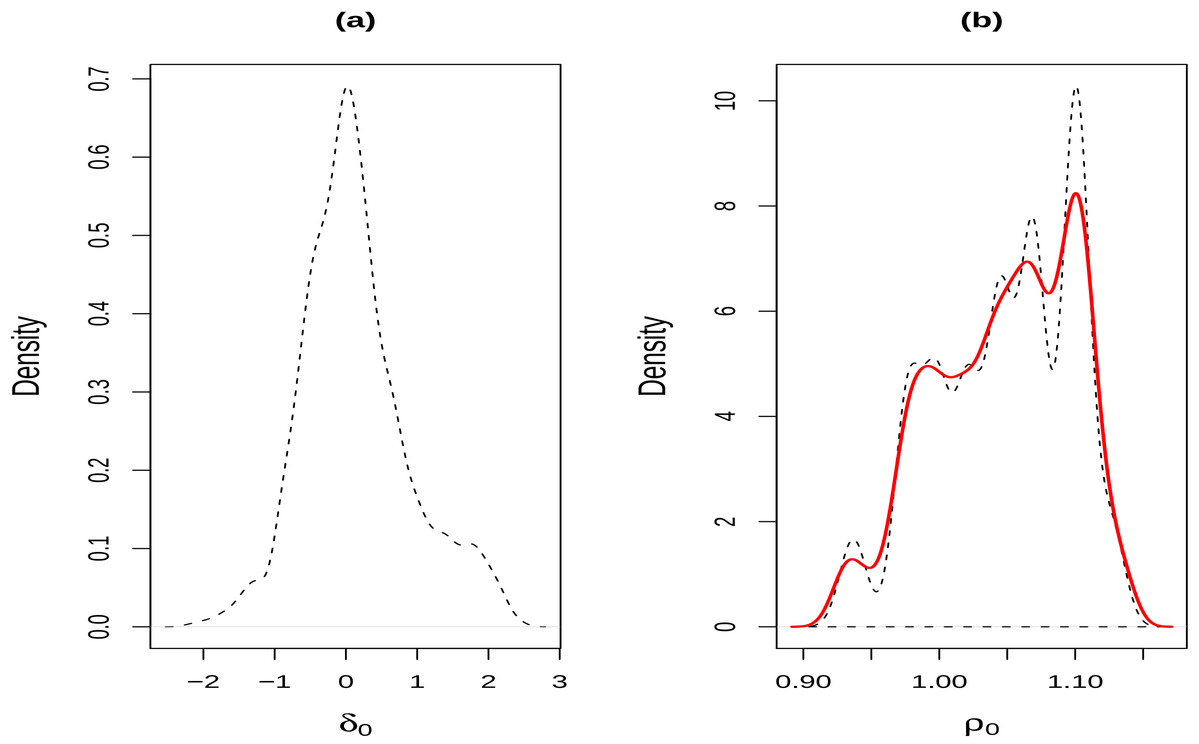
Table 1. Posterior modes of θ_3

- Step 2: Draw MCMC samples of θ_1 and θ_2. First 5000 runs as burn-in. Another 5000 runs for calculation.

	Posterior mean	Lower Bound	Upper Bound
ρ_0	**1.05**	0.94	1.13
σ_ρ^2	**0.29**	0.16	0.49
δ_0	**-0.14**	-1.24	1.93
σ_δ^2	**0.78**	0.18	2.70

Table 2. Posterior means and 95% confidence intervals of adjustment terms

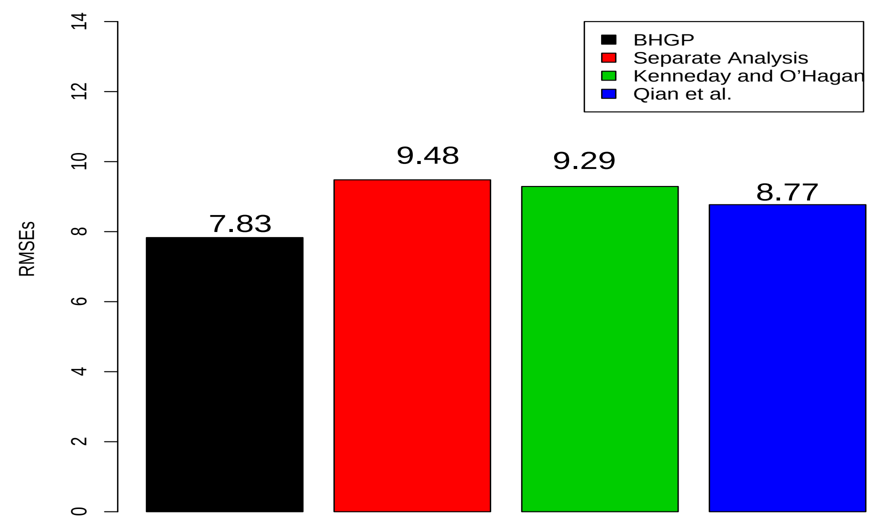
Intricate Location and Scale Change



- δ_0 is symmetric with center at -0.14.
- ρ_0 has multiple modes. ρ_0 may come from a **mixture model**.
- Scale adjustment has both **shrinkage** and **enlargement** effects.
- Scale adjustment is governed by **multiple laws**.

Prediction Results

- Four methods for predicting y_h for the eight testing runs: 1. BHGP model, 2. Separate analysis, 3. Kennedy-O'Hagan (2000) and 4. Qian et al. (2005).
- Compute RMSE (root-mean-square-error)$= \sqrt{\sum_{j=1}^{8} (y_h(\mathbf{x}_j) - \widehat{y}_h(\mathbf{x}_j))^2/8}$.



- Three combined methods beat the separate analysis. BHGP outperforms the other two by 16% and 11% respectively.

Case Study 2: Physical and Computer Experiments for Fluidized Bed Processes

- Source: Reese et al. (2004).
- Response: steady-state thermodynamic operation point.
- Six input variables

H_r	humidity
T_r	temperature
T_a	temperature of the air from the pump
R_f	flow rate of the coating solution
P_a	pressure of atomized air
V_f	fluid velocity of the fluidization air

- 28 LE runs (from computer experiment).
- 28 HE runs (from physical experiment).
- Training set: 20 randomly selected HE runs + 28 LE runs.
- Detailed analysis in Qian and Wu (2006).

Scale and Location Change

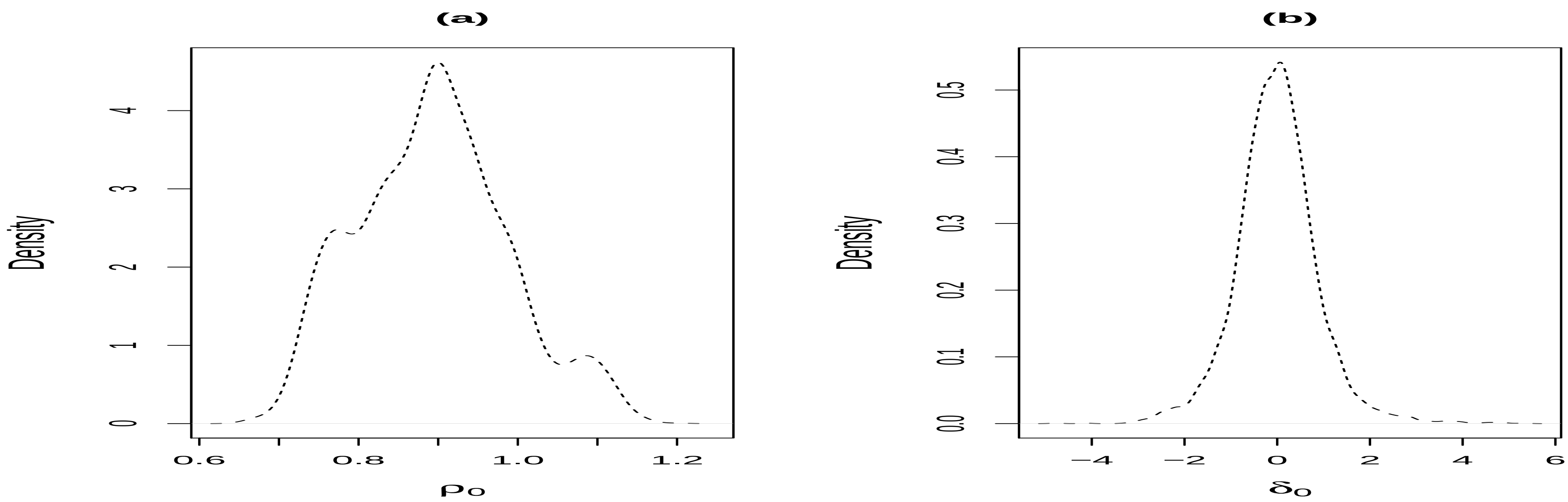


Figure 1. Posterior density of ρ_0 and δ_0.

- **Symmetric** location change δ_0 with center at 0.
- ρ_0 has multiple modes. ρ_0 may come from a **mixture model**.
- Scale adjustment has both **shrinkage** and **enlargement** effects.
- Scale adjustment is governed by **multiple laws**.

Planning of HE and LE

- Key to efficiently allocate resources and acquire information from HE and LE.
- $\mathbf{x} = (x_1, \ldots, x_k)$: design variables in $[0,1]^k$.
  D_l: set of design points ($\mathbf{x}_i$) for LE. D_h: set of design points ($\mathbf{x}_i$) for HE.
- Three principles for constructing D_l and D_h:

  Principle of economy: The size of D_h is less than the size of D_l.

  Principle of nested relationship: D_h is a subset of D_l.

  Principle of uniformity: Points in D_h and D_l are uniformly distributed.
- How to construct **multiple experiments** with respect to **multiple requirements**?

Summary and Conclusions

- Discuss **modeling** and **planning** perspectives for HE and LE.
- Model LE-HE data by using **hierarchical Gaussian process models**.
- Use **Stochastic programming** and **Markov chain Monte Carlo** methods in modeling building.
- Broadly applicable to problems and phenomena from engineering and sciences.

Kennedy and O'Hagan (2000)

- Models:

$$
\begin{aligned}
y_l &= \beta_{l0} + \varepsilon_l(\mathbf{x}_i), \;\; \mathbf{x}_i \in D_l, \\
y_h(\mathbf{x}_i) &= \rho y_l(\mathbf{x}_i) + \delta(\mathbf{x}_i), \;\; \mathbf{x}_i \in D_h,
\end{aligned}
$$

  - $\varepsilon_l \sim GP$ and $\delta \sim GP$.

- Non-informative priors.
- Empirical Bayes estimates.
- Prediction model: $\widehat{y}_h = \widehat{\rho}\widehat{y}_l + \widehat{\delta}$.

Calculating Bayesian Predictive Density Function

- Definition:

$$p(y_h(\mathbf{x}_0)|\mathbf{y}_h,\mathbf{y}_l) = \int_{\theta_1,\theta_2} p(y_h(\mathbf{x}_0)|\mathbf{y}_l,\mathbf{y}_h,\theta_1,\theta_2,\theta_3)p(\theta_1,\theta_2|\mathbf{y}_l,\mathbf{y}_h,\theta_3)d\theta_1 d\theta_2.$$

- Compute $p(y_h(\mathbf{x}_0)|\mathbf{y}_l,\mathbf{y}_h,\theta_1,\theta_2,\theta_3)$:
  - $p(y_h(\mathbf{x}_0)|\mathbf{y}_l,\mathbf{y}_h,\theta_1,\theta_2,\theta_3) = \frac{p(y_h(\mathbf{x}_0),\mathbf{y}_h|\mathbf{y}_l,\theta_1,\theta_2,\theta_3)}{p(\mathbf{y}_h|\mathbf{y}_l,\theta_1,\theta_2,\theta_3)}$.
  - $p(y_h(\mathbf{x}_0),\mathbf{y}_h|\mathbf{y}_l,\theta_1,\theta_2,\theta_3) \sim Normal$.
  - $p(\mathbf{y}_h|\mathbf{y}_l,\theta_1,\theta_2,\theta_3) \sim Normal$.
- Sample from $p(\theta_1,\theta_2|\mathbf{y}_l,\mathbf{y}_h,\theta_3)$:
  - Introduce new parameters: $\tau_1 = \frac{\sigma_\delta^2}{\sigma_\rho^2}$ and $\tau_2 = \frac{\sigma_\varepsilon^2}{\sigma_\rho^2}$.
  - $p(\sigma_\rho^2,\tau_1,\tau_2)$.
  - Use MCMC.