BAYESIAN HIERARCHICAL MODELS FOR INFORMATION COMBINATION

OBJECTIVES

- Introduce a statistical method for use in analyzing combined computer and physical experimental data with expert opinion
- Illustrate the method by means of an example
- Discuss the proposed method and its application

OUTLINE

- Motivation/Background
- Present the model
- Example
- Conclusions/Discussion

BACKGROUND

- Computer/Physical experimental data
- Same (or a subset of the same) factors, but possibly different factor values
- Different responses – transfer function
- Expert opinion
- Simultaneously analyze the combined data using *recursive Bayesian hierarchical model* (RBHM)

MOTIVATION

- Why bother? What do we gain?

  1. More precisely estimated model
  2. Validation of computer experiments
  3. Better predictions

- Cost savings (design?)

MOTIVATION

- The RBHM recognizes important differences between different data sources (expert opinion, computer model and physical data)

  1. Both location and scale biases in computer models (see Uncertainty and Reliability), allowed to be different for each run of the computer model
  2. Both location and scale biases in individual experts, allowed to be different for each expert opinion (same or different experts)

MODEL

- RBHM is multi-stage Bayesian modeling
- Recall:

$$\pi(\theta|data) \propto f(data|\theta)\pi(\theta)$$

  – reads: posterior distribution (distribution of parameters given the data) is proportional to the likelihood (joint distribution of the data) times the prior distribution

MODEL

- Stage 1
  - Define initial priors on all unknown parameters, including the biases
  - Update these priors using the expert opinions to form the posterior distributions (using Bayes theorem)

MODEL

- Stage 2
  - Use the posteriors from Stage 1 as the priors at Stage 2.
  - Update these priors using the computer model output to form new posterior distributions (again by Bayes theorem).

MODEL

- Stage 3
  - Use the posteriors from Stage 2 as the priors at Stage 3.
  - Update these priors using the physical experimental data to form new posterior distributions (Bayes theorem).
  - This yields the fully updated or final posterior distributions of interest (e.g., regression coefficients, or parameters of a reliability distribution)

DISCUSSION

- We can assess the effect of each data source by comparing the posteriors as they evolve from Stages 1 to 3 (this will be illustrated in the example)

- RBHM can be applied in a linear model framework as well as a reliability context. We will illustrate it in a linear model framework.

MODEL DETAILS

- Physical experimental data
  - $\underline{Y}_p \sim N(X\underline{\beta}, \sigma^2 I)$, where the physical data $\underline{Y}_p$ are normally distributed with mean $X\underline{\beta}$, X is a model matrix of factor values, and $\underline{\beta}$ is a vector of unknown regression parameters. The notation $\sigma^2 I$ indicates that each physical observation is independent of the others and has variance σ^2.

MODEL DETAILS

- Goal
  - The primary goal is to estimate $\underline{\beta}$ and σ^2 and make inferences about them; namely, which components of $\underline{\beta}$ are non-zero or "significant"
  - More appropriately, we want to know which covariates affect the performance metric.

MODEL DETAILS

- Computer experimental data
  - Comes from complex computer models of physical phenomena, e.g., finite element models.
  - $\underline{Y}_c \sim N(X\underline{\beta} + \underline{\delta}_c, \sigma^2\Sigma_c)$, where $\underline{\delta}_c$ is a vector of model run specific location biases and Σ_c is a matrix of scale biases (again computer model run specific)

MODEL DETAILS

– Usually

$$\Sigma_c = \begin{pmatrix} 1/k_{c_1} & 0 & \cdots & 0 \\ 0 & 1/k_{c_2} & 0 & \cdots \\ \vdots & 0 & \ddots & \cdots \\ 0 & \cdots & \cdots & 1/k_{c_C} \end{pmatrix}.$$

MODEL DETAILS

- Expert opinion data (expert judgment)
  - $\underline{Y}_o \sim N(X\underline{\beta} + \underline{\delta}_o, \sigma^2\Sigma_o)$, where $\underline{\delta}_o$ is a vector of possible location biases and Σ_o is a matrix of possible scale biases.

MODEL DETAILS

- How do these biases arise?
  - Location bias: an expert's average value is often either higher or lower than the true mean.
  - Scale bias: when an expert provides, say, a 0.90 quantile on the true response, this elicited value is often in reality a 0.60 or 0.70 quantile (over-valuation of information)

MODEL DETAILS

- How are these expert opinions elicited?
  - An expected response, y_o.
  - A quantile q_ξ for a prespecified probability ξ (e.g., $\xi = 0.9$, and thus the expert believes that 90% of the responses will be below q_ξ).
  - The "worth" of the expert opinion, m_o

MODEL DETAILS

- What is meant by the worth of expert opinion?
  - The corresponding number of physical experimental observations equivalent to the opinion.
  - May be fractional (e.g., may be less than 1)
  - Uncertainty about m_o is expressed through a prior distribution, which is then marginalized (integrated out) when applying the *RBHM*.

MODEL DETAILS

- Why is it called RB**H**M?
  - Hierarchical model

$$X_i \sim F_{X_i}(x;\Theta_i)$$
$$\Theta_i \sim F_{\Theta_i}(\theta;\Omega)$$
$$\Omega \sim F_{\Omega}(\omega;\tau),$$

    where τ is a (possibly) vector-valued constant.
  - Individual specific parameters
  - "Borrowing of Strength"
  - Results in shrinkage
- We have the hierarchical structure in the biases.

MODEL DETAILS

- Why has this not been done before?
  - Recall the posterior distribution:

$$\pi(\theta|data) \propto f(data|\theta)\pi(\theta)$$

but

$$\pi(\theta|data) = \frac{f(data|\theta)\pi(\theta)}{\int f(data|\theta)\pi(\theta)d\theta}$$

and the denominator is *hard* to calculate.

MODEL DETAILS

- MCMC methods to simulate observations from the posterior distribution.
- Our method uses Gibbs sampling which involves simulation from complete (or full) conditional distributions.
  - Distribution of each parameter conditional on all other parameters and the data
  - When the complete conditional can't be found in closed form, we simulate from the complete conditional distribution using Metropolis-Hastings algorithm.

MODEL DETAILS

- Prior distributions

$$
\begin{aligned}
\underline{\beta}|\sigma^2 &\sim N(\underline{\mu}_o, \sigma^2 \mathbf{C}_o) \\
\sigma^2 &\sim IG(\alpha_o, \gamma_o), \\
m_{o_i} &\sim Uniform(0.5 m_{o_i}^{(e)}, 2.0 m_{o_i}^{(e)}) \\
\delta_{o_i} &\overset{iid}{\sim} N(\theta_o, \xi_o^2) \\
k_{o_i} &\overset{iid}{\sim} G(\phi_o, \omega_o)
\end{aligned}
$$

MODEL DETAILS

- Hyperprior distributions
  - For $\underline{\delta}_o$:

$$\begin{aligned} \theta_o &\sim N(m_{\theta_o}, s^2_{\theta_o}) \\ \xi^2_o &\sim IG(a_{\xi^2_o}, b_{\xi^2_o}) \end{aligned}$$

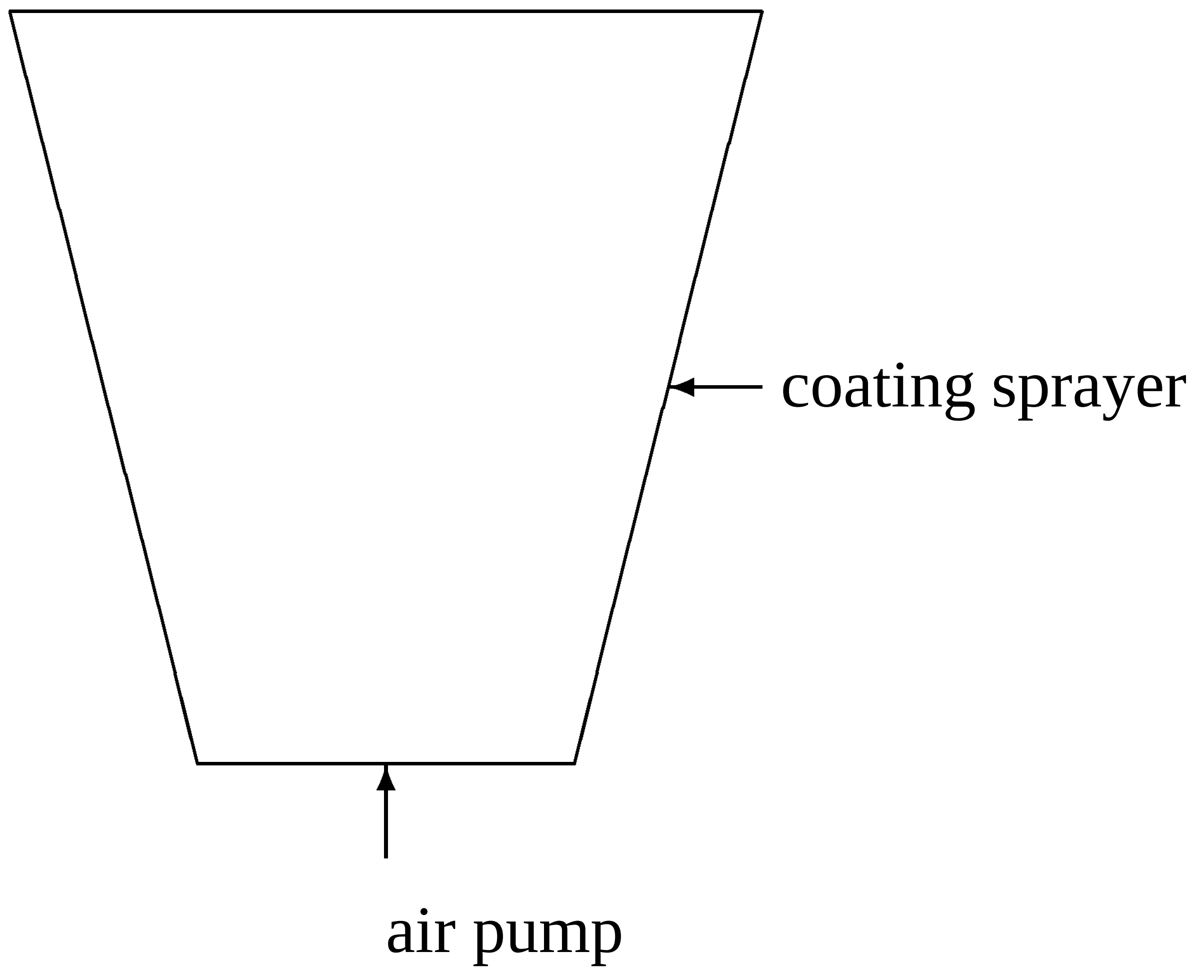
  - For $\underline{k}_o$:

$$\begin{aligned} \phi_o &\sim G(a_{\phi_o}, b_{\phi_o}) \\ \omega_o &\sim G(a_{\omega_o}, b_{\omega_o}), \end{aligned}$$

EXAMPLE

- Fluidized Beds used to coat food products
- Air is used to “float” the product through for even coating

EXAMPLE



EXAMPLE

- Three thermodynamic computer models (with increasing fidelity) were developed.
- Response: Steady-state thermodynamic operating point (Y)
- Input variables:
  - Pump air temperature (A)
  - Fluid velocity (V)
  - Coating solution flow rate (R)
  - Atomization air pressure (P)
  - Room Humidity (H)
  - Room temperature (T)

EXAMPLE

- 28 runs of each computer model (at different combinations of input variables) for a total of 28×3 computer model runs.

- 28 runs of the physical machine at each of the combinations of input variables.

- There are differences between “data” sources

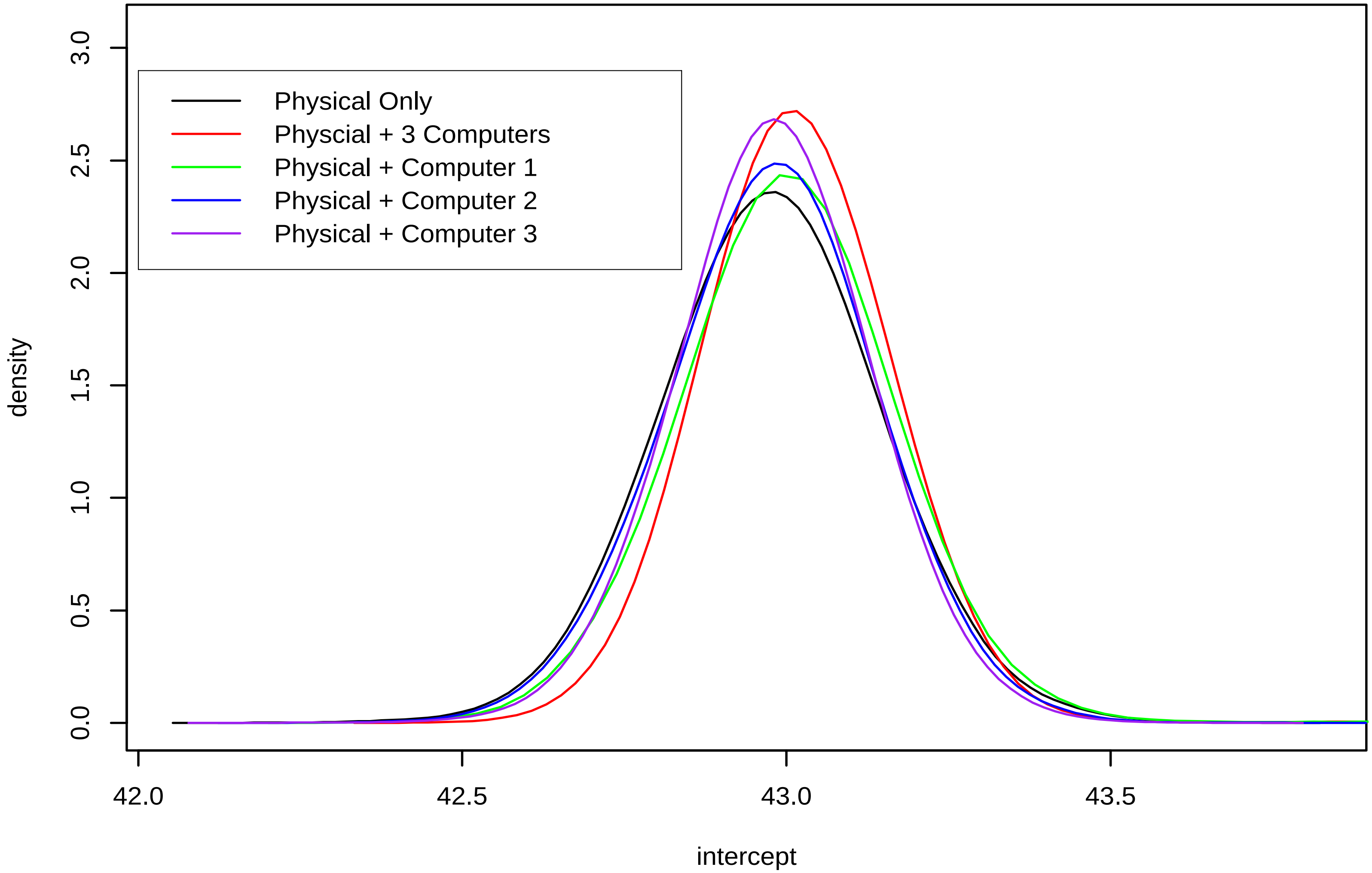
EXAMPLE

- Model

$$E(Y_p) = X\underline{\beta} = \beta_0 + \beta_1 A + \beta_2 R + \beta_3 V + \beta_4 (R \times V)$$

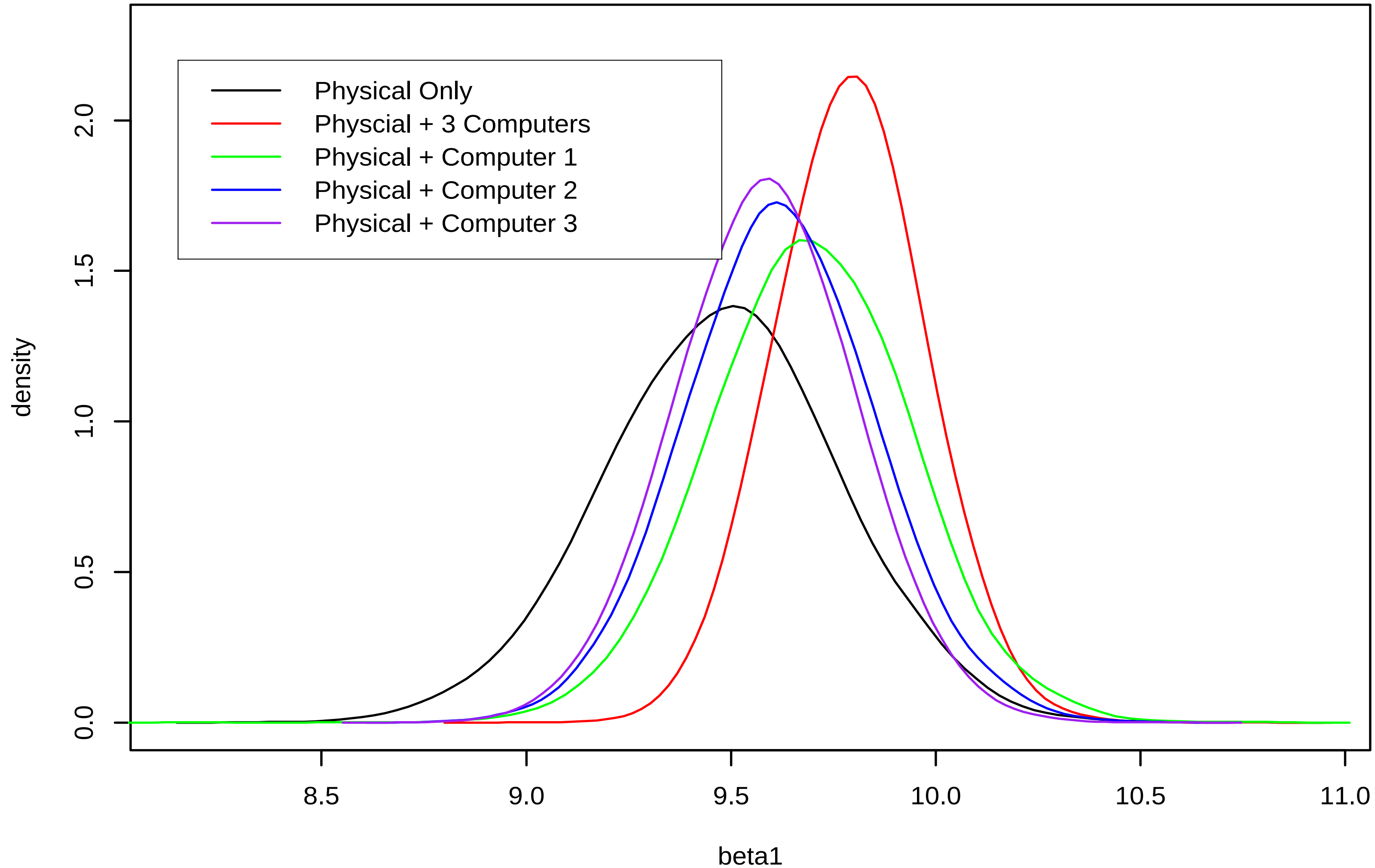
- Interest lies in estimation of $\underline{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ and σ^2.

RESULTS



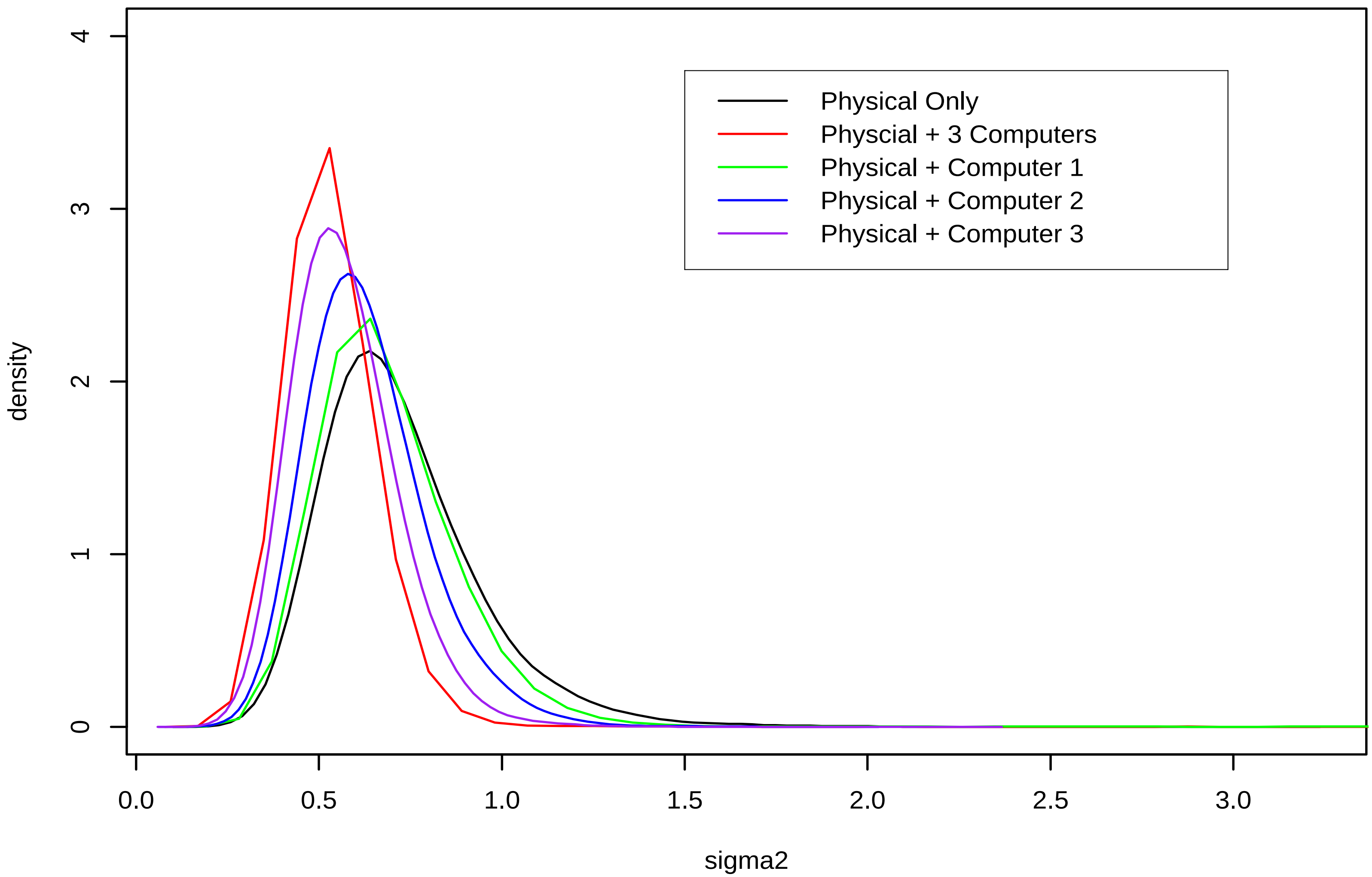
Physical Only
Physcial + 3 Computers
Physical + Computer 1
Physical + Computer 2
Physical + Computer 3
density
intercept
0.0
0.5
1.0
1.5
2.0
2.5
3.0
42.0
42.5
43.0
43.5

RESULTS



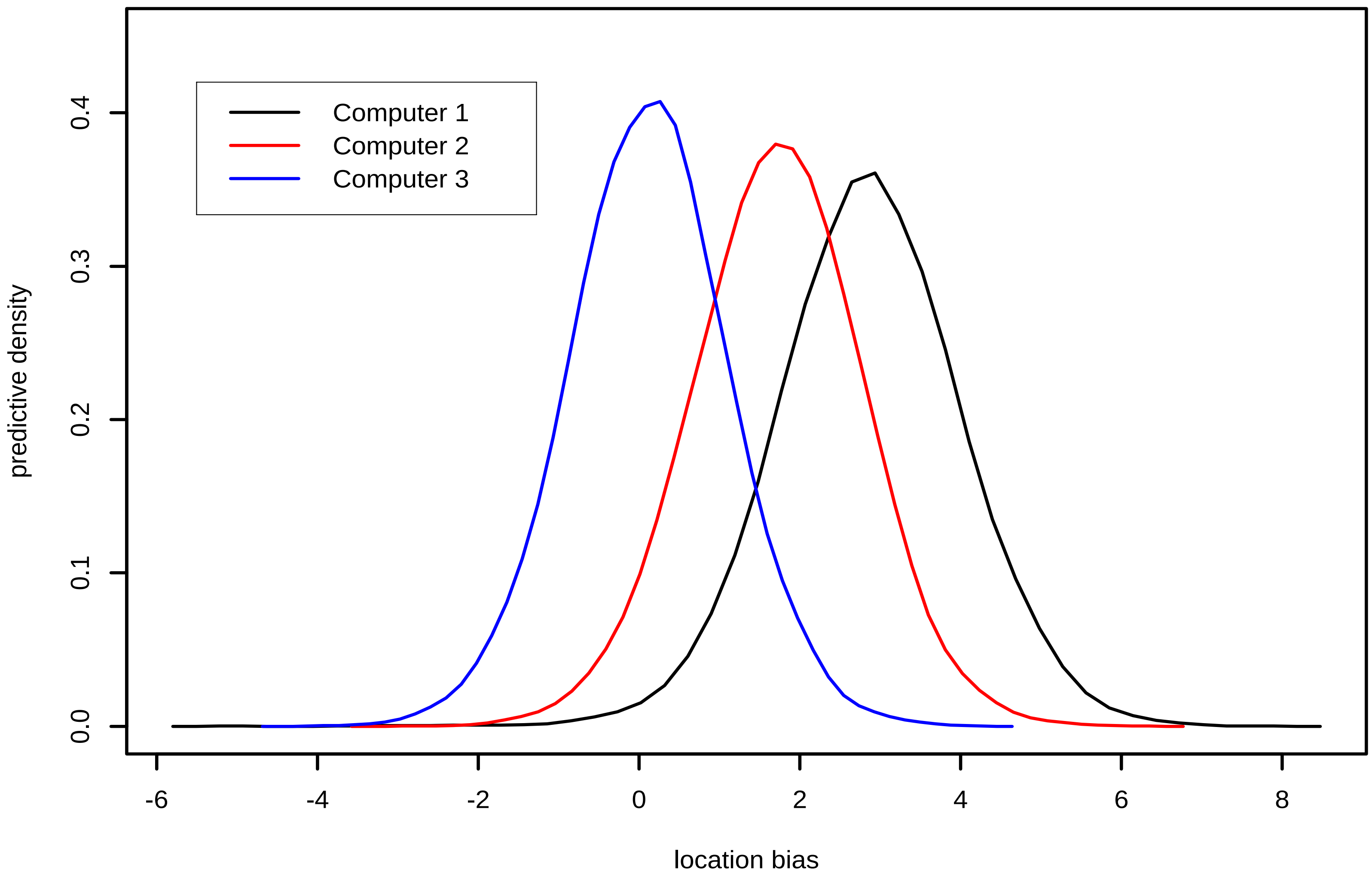
Physical Only
Physcial + 3 Computers
Physical + Computer 1
Physical + Computer 2
Physical + Computer 3
density
0.0
0.5
1.0
1.5
2.0
8.5
9.0
9.5
10.0
10.5
11.0
beta1

RESULTS



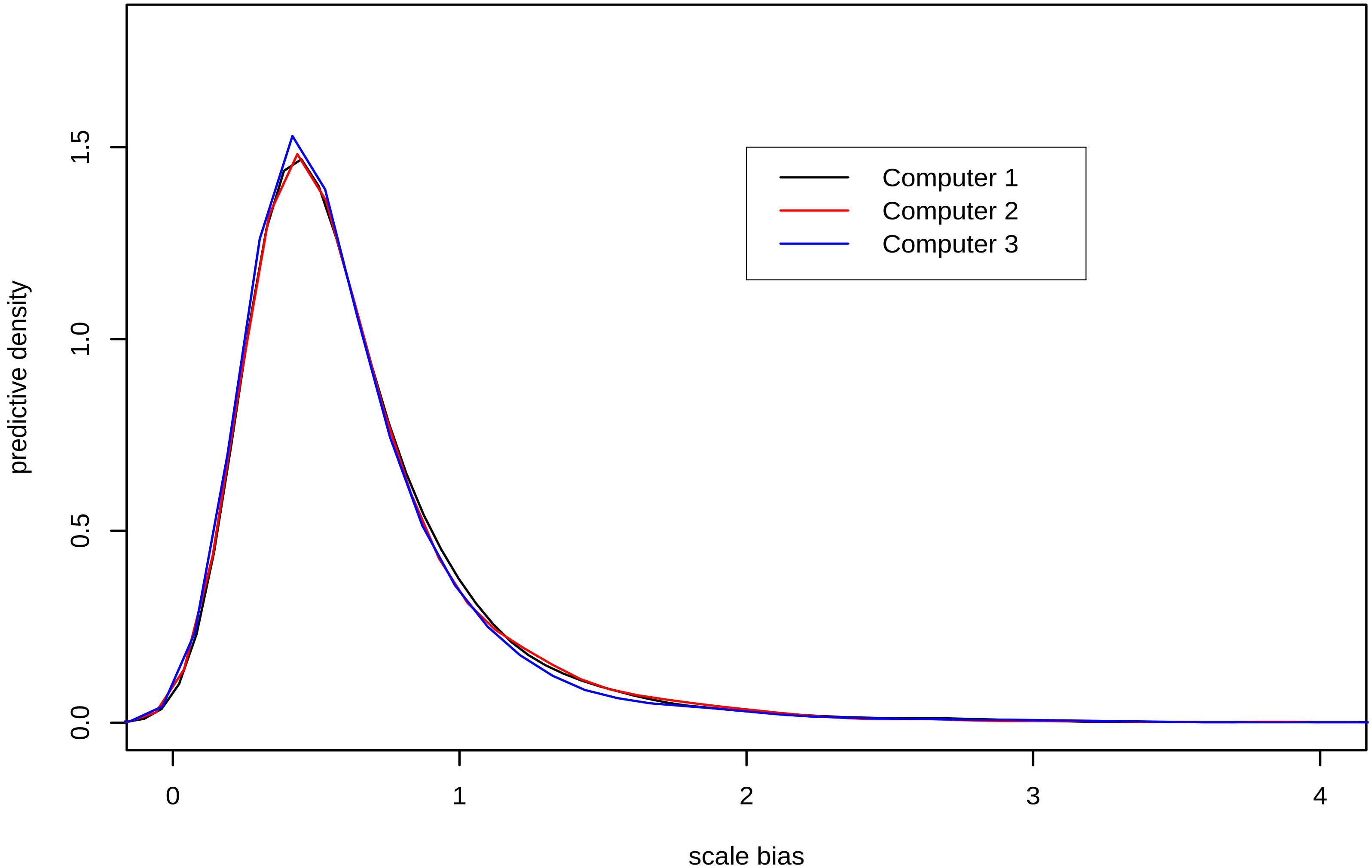
Physical Only
Physcial + 3 Computers
Physical + Computer 1
Physical + Computer 2
Physical + Computer 3
density
sigma2
0
1
2
3
4
0.0
0.5
1.0
1.5
2.0
2.5
3.0

RESULTS



Computer 1
Computer 2
Computer 3
predictive density
0.0
0.1
0.2
0.3
0.4
-6
-4
-2
0
2
4
6
8
location bias

RESULTS



Computer 1
Computer 2
Computer 3
predictive density
0.0
0.5
1.0
1.5
0
1
2
3
4
scale bias

DISCUSSION

- More precise estimation of parameters
- Predictive distribution of biases provides validation of computer models
- Wide applicability
  - Example is for performance metrics in linear models framework.
  - Reliability distributions are minor modification
  - RDMS could be populated with biases recognized.

DISCUSSION

- Generalizes fundamentals section (covariances and framework)
- Complicated models can be handled