Working with Inexact Models: The World of Computer Modeling

Combining Information from Computer Models and Data

Jim Berger
Duke University
Statistical and Applied Mathematical Sciences Institute

Symposium: 30 Years of Bootstrap & Recent Advances in Statistics
In memory of Yehuda Vardi
December 4, 2009
Vision: To focus on the *synthesis* of the statistical sciences and the applied mathematical sciences with disciplinary science to confront difficult data- and model-driven scientific challenges.

Location: NISS building
Research Triangle Park
North Carolina

Information: www.samsi.info
Current and Future Programs

• **2009-10**
  – Space-time Analysis in Environmental Mapping, Epidemiology and Climate Change
  – Stochastic Dynamics
  – Semiparametric Bayesian Inference, with Applications in Pharmacokinetics and Pharmacodynamics (July 12-23, 2010);

• **2010-11**
  – Analysis of Object-Oriented Data
  – Complex Networks
  – Summer program: TBD

• **Under consideration for 2011-12**
  – Uncertainty Quantification
Opportunities for Participation at SAMSI

- Proposing and leading a research program
- Visiting researchers (short and long term)
- Workshop participation
- Providing testbed problems and data
- Postdoctoral positions
- Graduate students
  - participate in major research programs
  - summer interdisciplinary workshops
- Undergraduate workshops
Outline

- The world of computer modeling (at SAMSI)
- Views on computer modeling
- Sources of uncertainty
- A pedagogic example
- SAVE methodology and an engineering application
- The modularization issue
- Barriers to implementation
A Small Sample of the World of Computer Modeling

Some SAMSI programs that heavily involved computer models

- **Large-Scale Computer Models for Environmental Systems**
  - Porous media models
  - Atmospheric ozone models
  - Computer models predicting health effects of pollutants

- **Inverse Problem Methodology In Complex Stochastic Models**
  - Models of HIV infection dynamics
  - Models of dielectric materials

- **Network Modeling for the Internet**
  - Models of computer networks

- **Data Assimilation for Geophysical Systems**
  - Ocean models, especially with Lagrangian data

- **Computational Biology of Infectious Diseases**
  - Models of cell transport
National Defense and Homeland Security
  - Social network models
  - Agricultural network models

Development, Assessment & Utilization of Complex Computer Models
  - Air quality models
  - Weather models
  - Climate models
  - Hydrological models
  - Models of jet engines
  - Models of cerebral blood flow
  - Models of infectious disease
  - Models of volcanic flow
  - Terrestrial ecology models
  - Models of high-energy nuclear collisions
  - Models of granular flow
  - Systems biology models
• **Stochastic Dynamics**
  – Models of polymer flows
  – Models of mechanical rupture
  – Compartment model for alcohol abuse
  – Contagion network models
  – Neuronal network models
  – Molecular motors
  – Reaction networks

• **Space-time Analysis for Environmental Mapping, Epidemiology and Climate Change**
  – Biological compartment models
  – Multiscale fluid flow model
  – Model of Dengue fever spread in Peru
Three views of computer modeling:

- It is the future of science, technology and society.
- It can be highly successful, but requires careful statistical validation.
- It is too difficult to be useful in most practical scenarios.

Four views on validation and verification of computer models:

- Hard Core Modeler: “Don’t bother me; I need every waking moment to work on the science/math/computation to improve the model.”
- Hard Core Statistician: “Practical use of the model is irresponsible unless all sources of uncertainty have been properly accounted for.”
- Soft Core Modelers:
  - “I’ll talk to statisticians if it help’s me to improve the model;”
  - “I’ll consign some time and model runs to dealing with uncertainty.”
- Soft Core Statisticians: “What I want in terms of model, field data, and information about uncertainties is not possible, so I’ll take what is available and try to do something.”
Some Sources of Uncertainty in Computer Modeling:

**Reality:** denote the real process being modeled by $y^R(x, Z)$, where

- $x$ are the model inputs (e.g., initial conditions, control variables, ...),
- $Z$ are stochastic elements of the model.

**Computer Model:** $y^M(x_1, x^*_2, u, Z^*)$, where

- $x = (x_1, x^*_2)$, with * indicating those inputs that are uncertain,
- $u$ are unknown parameters of the model (e.g., unknown coefficients of equations or rate constants, or unknown fudge factors),
- $Z^*$ are the modeled stochastic elements (rarely identical to $Z$),
- $b(x, Z) = y^R(x, Z) - y^M(x_1, x^*_2, u, Z^*)$ is unknown bias (convergence, systematics, discrepancy) of the model.

**Field data (if available):** $y^F(x, Z) = y^R(x, Z) + \epsilon$, where $\epsilon$ could have a complicated error structure, depending on inputs or other features.
The State of Consideration of Uncertainty:

- $Z$ is increasingly being dealt with (e.g., there is an increase in the use of stochastic dynamics over deterministic dynamics).

- Uncertainty in initial conditions, $\mathbf{x}^*_2$, is increasingly being incorporated, through techniques such as particle filtering or ensembles.

- Uncertainty in model parameters $\mathbf{u}$ is typically ignored, with the model being used at ‘best fit’ values or values ‘from the literature;’ modern statistical techniques (e.g. MCMC) can, in principle, be utilized here.

- Bias, $b(\mathbf{x}, Z)$, is almost universally ignored (or treated as random error)
  - partly for sociological reasons (“my model is as good as I could make it, so don’t call it biased!”);
  - partly because the field data about reality is often too limited for any sensible estimation of bias (or the field error may contain unknown bias);
  - partly because extrapolation of bias away from the input region where there were field measurements is fraught with peril;
  - partly because bias is severely confounded with other uncertainties.
Uncertainty in the predictions of the climate response to rising levels of greenhouse gases
(Lenny Smith and many others) Nature (2005)
Pedagogic Example

- Dynamics of a chemical reaction process $y(t)$ are thought to be

$$\frac{dy(t)}{dt} = -uy(t); \quad y(0) = 5,$$

with the reaction rate $u$ unknown, and known initial chemical concentration of 5.

- Solution $y^M(u, t) = 5 \exp(-ut)$ is the ‘computer model.’

- Field data from the process, at ten times $t_i$ equally-spaced over the interval $(0.11, 3.01)$, is modeled as

$$y^F(t_i) = y^R(t_i) + \epsilon_i,$$

where $y^R(t)$ is the real process and the $\epsilon_i$ are independently $N(0, \sigma^2)$, with $\sigma^2$ unknown. Three independent replicate observations $y^F_r(t_i)$ were obtained at each time point.
Standard ‘Best Fit’ Analysis: Estimate $u$ from the data, say by least squares, i.e., choose $u$ to minimize

$$
\sum_{r=1}^{3} \sum_{i=1}^{10} \left[ y^F_r(t_i) - 5 \exp(-ut_i) \right]^2 .
$$

Note that, in essence, this is assuming that the computer model is either correct (has no bias) or is “truth plus random error.”

**Answer:** $\hat{u} = 0.63$, with the resulting computer model being

$$
y^M(\hat{u}, t) = 5 \exp(-0.63t) .
$$

This is graphed in Figure 1 along with the data.
Figure 1: Least squares fit of the computer model to the data for the pedagogic example.
Figure 2: Residuals of the fit of the computer model to the data and a linear fit to the residuals.
Issues in proceeding:

1. We are pretending that $u = \hat{u}$. The uncertainty in this should be taken into account.

2. If the computer model is incorrect, i.e. has a systematic bias or discrepancy

$$b(t) = y^R(t) - y^M(u, t)$$

from the real process, ‘over-fitting’ will typically have occurred; the fit tries to ‘make up’ for the model inadequacy by over-shifting $u$ to compensate.

3. This over-fitting makes it problematical to believe any structure found in the residuals, e.g. the linear structure in Figure 3.
Bayesian approach:

1. Acknowledge the uncertainty in $u$ (and $\sigma^2$), by viewing them as random with prior distribution $\pi(\mu, \sigma^2)$; Bayes theorem then gives the posterior distribution of $(\mu, \sigma^2)$, given the data $x$, as

$$
\pi(\mu, \sigma^2 \mid x) \propto \pi(\mu, \sigma^2) \frac{1}{\sigma^3} \exp \left( -\frac{1}{2\sigma^2} \sum_{r=1}^{3} \sum_{i=1}^{10} [y_r^F(t_i) - 5 \exp(-u t_i)]^2 \right).
$$

2. Recognize that the computer model may have a bias (Kennedy and O’Hagan, 2001 JRSSB) and so model field data as

$$
y^F(t_i) = 5 \exp(-u t_i) + b(t_i) + \epsilon_i,
$$

where $b(t)$ is the unspecified function, itself assigned a prior distribution (e.g., a Gaussian process, as in the GaSP of Will Welch). Then also compute the posterior distribution of $b(t)$.

Note: This is very different than the ubiquitous scientific modeling

$$
field \ observation = computer \ model + random \ error
$$
**Major difficulties:**

- There is a severe lack of identifiability of $u$ and $b(\cdot)$.

- There is severe practical confounding of $\sigma^2$ and $b(\cdot)$ if there are no replicate field observations.

Bayesian analysis (in principle) can accommodate this:

- In the computer model scenario, $u$ may have physical meaning (e.g., the reaction rate in the example) or, at least, physical limits, so that experts may be able to construct a fairly tight prior distribution for $u$.

- The prior distribution for the bias ‘encourages’ $b(\cdot)$ to be zero, allowing a correct computer model to emerge with little bias if supported by the data.

But there are potentially severe computational issues, as we shall see later.
Figure 3: *Upper left:* bias-corrected predictions (solid) with 90% confidence bands (dashed). *Upper right:* marginal posterior distribution of $u$. *Bottom:* posterior distribution of bias, given the posterior mean (left), given the least squares estimate (right).
**Truth:** The true model used to generate the data was

\[ y^F(t_i) = (3.5) \exp(-1.7t_i) + 1.5 + \epsilon_i , \]

with \( \sigma^2 = (0.3)^2 \). (The initial chemical concentration was indeed 5, but the reaction had a residual of 1.5 units unreacted.)

- So \( \hat{u} = 0.63 \) was indeed a severe over-fit.
- There is indeed a systematic bias,

\[ b(t) = y^R(t) - y^M(1.7, t) = [(3.5)e^{-1.7t} + 1.5] - 5e^{-1.7t} = 1.5(1 - e^{-1.7t}) . \]
Additional Statistical Challenges in Real Problems

- Small and/or complex data sets
  - maybe no system level field data (e.g. nuclear weapons systems)
  - the “Admiral’s test”
  - field and model-run data are often at different values

- Extremely expensive computer model runs
  - requires careful design in selecting input values for model runs
  - requires development of model approximations (surrogates, emulators) for statistical analysis, optimization, ...

- Difficulties with model inputs
  - uncertainty in inputs is common
  - dimension reduction: which model inputs are key, and which have little effect and can be mostly ignored?
SAVE: Simulation Analysis and Validation Engine

(Software: National Institute of Statistical Sciences; Technometrics, 2007)


• Step 1. Specify uncertainties in the model, called the Input/Uncertainty Map.

• Step 2. Determine evaluation criteria.

• Step 3. Design and collect validation data.

• Step 4. If the computer model is slow, develop a fast (response surface) approximation.

• Step 5. Assess predictive accuracy and tune model.

• Step 6. Iterate as needed.
The Main Focus: Prediction with Tolerance Bounds

Instead of asking “Is the model correct?”, report tolerance bounds for predictions.

Example: With probability 80%, the model prediction 5.17 (at specified input $x$) will be within $\pm 0.44$ of the true process value. This simple reporting device overcomes the difficulties that

- it is often impossible to adequately characterize the regions of model accuracy and inaccuracy;

- the degree of accuracy that is needed can vary from one application of the computer model to another;

- it is usually crucial to incorporate model bias, as well as variance, in assessment of accuracy.
Other Features of SAVE

• It can incorporate all types of uncertainty; from uncertainty in model inputs or parameters to uncertainty in the data to uncertain expert opinion.

• It can be implemented even if data (model-run data and/or field data) is very limited.

• The model-run and field data can be observed at different input values.

• One can ‘tune’ unknown parameters of the computer model based on field data, and at the same time apply the validation methodology.

• Analysis of model bias is included – indeed, is central.

• Optimal predictions involve a synthesis of data-driven and model-based prediction.

• The methodology includes use of a fast approximation (emulator) of the computer model, with computed accuracy, if needed.

• Prediction for a ‘related situation’ is possible.
Validation of a Computer Model of Road-Load Dynamics


Consider a vehicle being driven over a road with two major potholes.

- $\mathbf{x} = (x_1, \ldots, x_7)$ is the vector of key vehicle characteristics, unknown because of manufacturing variability.

- $y^R(\mathbf{x}; t)$ is the time-history curve of resulting forces.

A finite element PDE computer model of the vehicle being driven over the road

- depends on $\mathbf{x} = (x_1, \ldots, x_7)$ and unknown calibration parameters
  $\mathbf{u} = (u_1, u_2)$;

- yields time-history force curve $y^M(\mathbf{x}, \mathbf{u}; t)$. 
**Field Data:** Seven runs of a given test vehicle over the same road containing two potholes.

Denote the $r$-th field time-history curve by $y_r^F(x^*; t), r = 1, \ldots, 7,$ where $x^* = (x_1^*, \ldots, x_7^*)$ refers to the unknown vehicle characteristics of the given test vehicle. We assume

$$y_r^F(x^*; t) = y_r^R(x^*; t) + \epsilon_r(t);$$

the $\epsilon_r(t)$ are independent realizations of a certain error process.

**Model Data:** The computer model of the vehicle was ‘run over the potholes’ at 65 input values of $z = (x, u) = (x_1, \ldots, x_7, u_1, u_2);$ let $z_r = (x_r, u_r), r = 1, \ldots, 65,$ denote the corresponding parameter vectors, which were chosen by a Latin-hypercube design over the input uncertainty ranges.

Let $y_r^M(z_r; t)$ denote the $r^{\text{th}}$ computer model time-history curve, $r = 1, 2, \ldots, 65.$
Figure 4: Forces from field and model data for two potholes.
Data registration: To compare field curves and model-run curves, it is crucial to align them, so that the potholes occur at the same ‘locations.’

- First convert all time-histories to distance-histories (using available vehicle velocity information), i.e., represent the road load as a function of the distance, \(d\) (in meters), that a vehicle travels, instead of time \(t\).

- Next compute the average of all model runs as the ‘reference’ curve.

- Finally align the field data to match the peaks of both major potholes for the field runs to the peaks of this reference curve, using a piecewise linear transformation.
Wavelet Representation of the Curves:

The functions (of $t$) $y^F_r(x^*; t)$, $y^M(z_r; t)$, and $y^R(x^*; t)$ are represented via a wavelet expansion with 289 selected basis elements, $\psi_i(t)$, $i = 1, \ldots, 289$. The 65 model response curves and 7 field response curves are thus represented as

$$y^M(z_j; t) = \sum_{i=1}^{289} w^M_i(z_j) \psi_i(t), \quad j = 1, \ldots, 65,$$

$$y^F_r(x^*; t) = \sum_{i=1}^{289} w^F_{ir}(x^*) \psi_i(t), \quad r = 1, \ldots, 7,$$

where the $w^M_i(z_j)$ and $w^F_{ir}(x^*)$ are the coefficients computed through the wavelet decomposition.
Figure 5: The accuracy of the wavelet decomposition.
Introduction of Bias and Model Approximation:

Following Kennedy and O’Hagan (2001), introduce model bias, e.g.

\[ y^R(x; t) = y^M(x, u; t) + b(x; t). \]

In the transformed wavelet domain, this becomes

\[ w^R_i(x^*) = w^M_i(x^*, u^*) + b_i(x^*), \quad i = 1, \ldots, 289. \]

The replicated field data is modeled, for \( r = 1, \ldots, 7 \), as

\[ w^F_{ir}(x^*) = w^R_i(x^*) + \epsilon_{ir} = w^M_i(x^*, u^*) + b_i(x^*) + \epsilon_{ir}, \]

where the \( \epsilon_{ir} \) are i.i.d. \( N(0, \sigma_i^2) \) random errors.

For computational reasons, replace each \( w^M_i(x, u) \) by an emulator (a response surface approximation); we used the Gaussian stochastic process emulator GaSP (of Welch) with separable covariance function of the form

\[ \lambda_i^{-1} \prod_{j=1}^7 \exp \left( -\beta_{ij} |x_j - x_j'|^{\alpha_{ij}} \right) \prod_{j=1}^2 \exp \left( -\beta_{ij}^* |u_j - u'_j|^{\alpha_{ij}^*} \right). \]
Prior Distributions for Inputs and Calibration Parameters:

The following priors were elicited from the engineers:

\[
\pi(x^*, u^*) = \prod_{i=1}^{2} p(u_i^*) \prod_{i=1}^{7} p(x_i^*),
\]

\[
p(u_i^*) = \text{Uniform}(u_i^* \mid 0.125, 0.875), \quad i = 1, 2,
\]

\[
p(x_i^*) \propto \mathcal{N}(x_i^* \mid 0.5, 0.1111^2)I_{(0.1667, 0.8333)}(x_i^*), \quad i = 1, 2, 3
\]

\[
p(x_4^*) \propto \mathcal{N}(x_4^* \mid 0.5, 0.0641^2)I_{(0.3077, 0.6923)}(x_4^*),
\]

\[
p(x_i^*) \propto \mathcal{N}(x_i^* \mid 0.5, 0.1176^2)I_{(0.1471, 0.8529)}(x_i^*), \quad i = 5, 6,
\]

\[
p(x_7^*) \propto \mathcal{N}(x_7^* \mid 0.5, 0.1026^2)I_{(0.1923, 0.8077)}(x_7^*),
\]

where \( I_A(u) \) is 1 if \( u \in A \) and 0 otherwise.
Prior Distributions for Other Parameters:

$$
\pi(b, \mathbf{x}^*, \mathbf{u}^*, \sigma^2, \tau^2) = \pi(b \mid \mathbf{x}^*, \mathbf{u}^*, \sigma^2, \tau^2)\pi(\tau^2 \mid \mathbf{x}^*, \mathbf{u}^*, \sigma^2)\pi(\sigma^2 \mid \mathbf{x}^*, \mathbf{u}^*)\pi(\mathbf{x}^*, \mathbf{u}^*)
$$

where $b$, $\sigma^2$ and $\tau^2$ refer to the vectors of $b_i$, $\sigma_i^2$ and $\tau_j^2$; and, with $\bar{\sigma}_j^2$ denoting the average of the $\sigma_i^2$ at wavelet level $j$,

$$
\pi(b \mid \mathbf{x}^*, \mathbf{u}^*, \sigma^2, \tau^2) = \prod_{i=1}^{289} N(b_i \mid 0, \tau_j^2),
$$

$$
\pi(\tau^2 \mid \mathbf{x}^*, \mathbf{u}^*, \sigma^2) \propto \prod_{j=0}^{12} \frac{1}{\tau_j^2 + \bar{\sigma}_j^2/7},
$$

$$
\pi(\sigma^2 \mid \mathbf{x}^*, \mathbf{u}^*) \propto \prod_{i=1}^{289} \frac{1}{\sigma_i^2}.
$$

Prior Distributions for the GaSP Parameters: None - here they were all estimated by their MLE’s based on the model run data.
Posterior Distributions for all items of interest can then be determined via MCMC (with a caveat we return to shortly):

- inputs $u^*$ and $x^*$
- bias $b(\cdot)$
- prediction of reality for the tested vehicle
- prediction of reality for a different vehicle of the same type
- prediction of reality for a new vehicle type
Figure 6: Posterior distribution of $u^*$ and $x^*$. 
Figure 7: Posterior bias curve estimate and 90% tolerance bands.
Figure 8: Multiplicative extrapolation of bias to Vehicle B.
Computation was initially done by a standard Markov Chain Monte Carlo analysis:

- Closed form full conditionals are available for $b$, and for the emulator wavelet coefficients $w^M \equiv w^M(x^*, u^*)$.

- Metropolis-Hastings steps were used for $(x^*, u^*, \sigma^2, \tau^2)$; efficient proposal distributions were available, so all seemed fine.

**Shock:** The original computation failed and could not be fixed using traditional methods; the answers were also ‘wrong’!

- **Problem:** Some of the $\sigma^2_i$ (variances corresponding to certain wavelet coefficients of the field data) got ‘stuck’ at very large values, with the effect that the corresponding biases were estimated as near zero.

- **Likely Cause:** Modeling the $b_i$ as hierarchically normally distributed; biases for many wavelet coefficients can be expected to be small, but some are likely to be large.
**Ideal solution:** Improve the hierarchical models; for instance, one could consider use of more robust models (e.g. Cauchy models) for the $b_i$.

**Pragmatic solution:** Cheat computationally, and only allow generation of the $\sigma^2_i$ from the replicate information (i.e., from the $s^2_i = \sum_{r=1}^{7}(w_{ir}^F(x^*) - \bar{w}_i^F)^2$), not allowing transference of information from the $b_i$ to the $\sigma^2_i$.

- We call such cheating *modularization*, the idea being to not always allow Bayesian updating to flow both ways between modules (components) of a complex model.

- Another name given to this idea is *cutting feedback* (Best, Spiegelhalter, ...); related notions are *inconsistent dependency networks* (Heckerman); and *inconsistent Gibbs for missing data problems* (Gelman and others).

- In the road load analysis, the modularization approach gives very similar answers to the improved modeling approach.
Figure 9: Trace plots for $\sigma^2_{170}$. *Upper*: global Normal analysis; *Middle*: global Cauchy analysis; *Lower*: modularized Normal analysis.
Understanding modularization as a computational ‘fix’ for differing levels of model believability

Consider a simplified situation – the random effects model

\[ y_{ir} = \mu_i + b_i + \epsilon_{ir}, \quad i = 1, \ldots, 289; \quad r = 1, \ldots, 7, \]

where \( \mu = \{ \mu_i \} \) are assumed known, \( b_i \sim N(0, \tau^2) \), and \( \epsilon_{ir} \sim N(0, \sigma^2_i) \).

The integrated likelihood of \( \tau^2 \) and \( \{ \sigma^2_i \} \) (integrate out the \( b_i \)) is

\[
L(\tau^2, \{ \sigma^2_i \} ; \{ \bar{y}_i, s^2_i \}) \propto \prod_{i=1}^{289} \frac{\sigma_i^{-6}}{\sqrt{\tau^2 + \frac{1}{7} \sigma_i^2}} \exp \left( -\frac{(\bar{y}_i - \mu_i)^2}{2 (\tau^2 + \frac{1}{7} \sigma_i^2)} - \frac{s_i^2}{2\sigma_i^2} \right),
\]

where \( \bar{y}_i = \frac{1}{7} \sum_{r=1}^{7} y_{ir} \) and \( s_i^2 = \sum_{r=1}^{7} (y_{ir} - \bar{y}_i)^2 \).

- If some \( (\bar{y}_i - \mu_i)^2 \) is ‘extreme,’ one thinks that the hierarchical analysis will compensate by increasing \( \tau^2 \), but instead it here increases \( \sigma^2_i \).
- It follows that the corresponding bias \( b_i \), which has full conditional mean \( \bar{y}_i \tau^2 / (\tau^2 + \frac{1}{7} \sigma_i^2) \), will be inappropriately shrunk towards 0.
A more robust Bayesian model for the $b_i$ (that can accommodate extreme realizations) is the Cauchy distribution

$$\pi(b_i \mid \tau^2) \sim \text{Cauchy} \left(0, \tau^2\right).$$

(1)

Using this, the problem disappears, but computation increases significantly.

**The Modular Solution:** Keep the normality assumption for the $b_i$, but ‘separate’ the MCMC for the $\sigma_i^2$ from the rest of the Bayesian analysis, i.e., instead of using the true posterior

$$\pi(\tau^2 \mid \{\sigma_i^2\}, \{\bar{y}_i, s_i^2\}) \times \pi(\{\sigma_i^2\} \mid \{\bar{y}_i, s_i^2\}),$$

work with the modularized posterior

$$\pi(\tau^2 \mid \{\sigma_i^2\}, \{\bar{y}_i, s_i^2\}) \times \pi(\{\sigma_i^2\} \mid \{s_i^2\}).$$

**Key Condition:** The modularized posterior must be a real probability distribution for the parameters.
Useful modularizations when working with computer models:

- Assuming replications of field measurements are available (important), use only the replications for inference about the field measurement error variances.

- If using emulators (e.g. GaSP), estimate emulator parameters in a modular fashion, using only the computer model runs.

- If the bias term is itself represented via a GaSP, try to isolate the estimation of it’s GaSP parameters from any unknown calibration parameters.
An Aside: Determination of the Probability that the Computer Model Is Correct

- Specify a prior probability, $\pi_0$, that the computer model, $M_0$, is correct.
- Select an alternative model $M_1$ (which here, will be ‘model + bias’).
- Specify prior densities for unknown parameters of $M_0$ and $M_1$ (already done).
- Compute the *posterior probability that $M_0$ is correct*:
  - computationally much harder, since one must compute marginal likelihoods;
  - tends to be a waste of time; for any computer model we ever tried, the posterior probability that the model is correct (in the sense of not being rejectable by field data) is essentially zero (if there is any field data whatsoever).
Summary of SAVE

- Validation of computer/math models should focus on providing *errors of prediction* in use of the model.
- Acknowledging the likely presence of *model bias* is key, and bias-corrected prediction is possible.
- Validation usually requires model-run and field data.
- Bayesian spatial and hierarchical techniques provide a framework for all of the analysis, and provide a fast model approximation.
- A mix of subjective and objective Bayesian methodology (and likelihood methodology) is used.
Additional issues that need to be addressed in model validation

- Handling multiple correlated outputs (hierarchical or correlated spatial)
- Large output functions can require special handling, such as reducing the basis size by PCA (POD, EOF)
- What are we to do if there is no field data to even estimate the bias? Unsatisfactory (but better than nothing) solutions include
  - Use runs of the model at various levels of resolution or development.
  - Use competing models.

Fundamental issues in model building (relating to statistics)

- Detail of model versus feasibility versus statistical modeling
- Statistical methods of detecting model flaws (ex: car engine maps)
- Determination of important inputs for focused refinement of the model (requires good statistical design).
Fundamental issues in model use

- In bias corrected prediction, is one really doing better than a pure statistical analysis of field data? (If so, it’s in the joint posterior.)
- Extrapolation: often the model is to be used in regions where there is no field data (the main purpose for building the computer model)
  - How can we sensibly formalize extrapolation of bias?
  - Are the uncertainties at least believable?
- The model is going to be used to predict something, and so it will be combined with an entirely different layer of statistics/modeling (e.g. impact of climate change on forests, hazard maps for volcanic flows, ...).
  - Often, the uncertainty (e.g. volcanic hazard maps) is a key quantity of interest.
  - Quantities of interest are typically, highly non-linear, in which case ‘best guess’ or ‘best fit’ solutions can be terrible.