Utilizing simulations and experiments for inference in Engineering Applications

Dave Higdon, Los Alamos National Laboratory
Brian Williams, Los Alamos National Laboratory
Jim Gattiker, Los Alamos National Laboratory
Charles Nakhleh, Los Alamos National Laboratory
Katrin Heitmann, Los Alamos National Laboratory
Salman Habib, Los Alamos National Laboratory

LA-UR-06-6667
Application: Assessing launch risk in prototype aircraft

Posterior missile trajectories for a particular flight condition $x_0$
Example: Linear charged particle accelerator
Experimental Data and Simulations are Radiographic Images
Example: Spotwelding – Combining experimental data and simulations (with Marc Kennedy Univ Sheffield)

COPPER ELECTRODE - H₂O COOLED
LOW ELECTRIC RESISTANCE
HIGH THERMAL CONDUCTIVITY

FAYING SURFACE
TOP WORKPIECE

FUSED WELD NUGGET
BOTTOM WORKPIECE

INDENTATION
HIGH CURRENT

WORKPIECE
HIGHER ELECTRICAL RESISTANCE
LOWER THERMAL CONDUCTIVITY
Calibration of Flyer Plate Calculations to Observational Data

- Velocity profile a function of material constitutive behavior
- Preston-Tonks-Wallace (PTW) model utilized in calculations to describe stress-strain relationship
- Calibrate free PTW parameters (7) to observational data
Certification Issues at LANL

Bombs for super geniuses

- Detonation
- Implosion
- Nuclear yield

Implosion experiments

Sub-critical experiments

Historical nuclear tests

Off-line experiments
Materials, equations of state (EOS), high explosive (HE)
Implosion Experiments at Los Alamos

1943, Neddermeyer’s initial implosion experiments

A 2005 experiment at the Dual Axis Radiographic Hydrodynamic Test Facility
Bayesian analysis of an inverse problem

- A simple example...

  - $x$: experimental conditions
  - $\theta$: model calibration parameters
  - $\zeta(x)$: true physical system response given inputs $x$
  - $\eta(x, \theta)$: forward simulator response at $x$ and $\theta$.
  - $y(x)$: experimental observation of the physical system
  - $e(x)$: observation error of the experimental data

Assume:

$$y(x) = \zeta(x) + e(x) = \eta(x, \theta) + e(x) \quad \theta \text{ unknown.}$$
Bayesian formulation

Sampling model:

\[ y_i = \eta(x_i, \theta) + e_i, \quad \text{where } e_i \stackrel{iid}{\sim} N(0, 1/\lambda_y) \]

which gives likelihood:

\[
L(y|\theta, \lambda_y) \propto \lambda_y^{n/2} \exp \left\{ -\frac{1}{2 \cdot 25^2 \lambda_y} \sum_{i=1}^{n} (y_i - \eta(x_i, \theta))^2 \right\}
\]

Priors

\[
\pi(\theta) \propto I[0 \leq \theta \leq 1]
\]
\[
\pi(\lambda_y) \propto \lambda_y^{a_y-1} \exp\{-b_y \lambda_y\}, \quad a_y = 5, \quad b_y = 5
\]

\[
\pi(\theta, \lambda_y|y) \propto L(y|\eta(x, \theta), \lambda_y) \times \pi(\theta) \times \pi(\lambda_y)
\]
\[
\propto \lambda_y^{n/2} \exp \left\{ -\frac{1}{2 \cdot 25^2 \lambda_y} \sum_{i=1}^{n} (y_i - \eta(x_i, \theta))^2 \right\} \times I[0 \leq \theta \leq 1] \times
\lambda_y^{a_y-1} \exp\{-b_y \lambda_y\} \]
Bayesian formulation

\[ L(y | \eta(z)) \propto |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (y - \eta(z))^T \Sigma^{-1} (y - \eta(z)) \right\} \]

\[ \pi(z | \lambda_z) \propto \lambda_z^{m} \exp \left\{ -\frac{1}{2} z^T W_z z \right\} \]

\[ \pi(\lambda_z) \propto \lambda_z^{a_z - 1} \exp \left\{ b_z \lambda_z \right\} \]

\[ \pi(z, \lambda_z | y) \propto L(y | \eta(z)) \times \pi(z | \lambda_z) \times \pi(\lambda_z) \]
Posterior realizations of $\theta$
Accounting for limited simulator runs


\( x \) model or system inputs
\( \theta \) calibration parameters
\( \zeta(x) \) true physical system response given inputs \( x \)
\( \eta(x, \theta) \) simulator response at \( x \) and \( \theta \).

simulator run at limited input settings
\[
\eta = (\eta(x_1^*, \theta_1^*), \ldots, \eta(x_m^*, \theta_m^*))^T
\]
treat \( \eta(\cdot, \cdot) \) as a random function
use GP prior for \( \eta(\cdot, \cdot) \)

\( y(x) \) experimental observation of the physical system
\( e(x) \) observation error of the experimental data

\[
y(x) = \zeta(x) + e(x)
\]
\[
y(x) = \eta(x, \theta) + e(x)
\]
Accounting for limited simulation runs

Again, standard Bayesian estimation gives:

\[
\pi(\theta, \eta(\cdot, \cdot), \lambda, \rho_{\eta}, \lambda_{\eta}|y(x)) \propto L(y(x)|\eta(x, \theta), \lambda_{\epsilon}) \times \pi(\theta) \times \pi(\eta(\cdot, \cdot)|\lambda_{\eta_{\cdot}}, \rho_{\eta_{\cdot}}) \times \pi(\lambda_{\epsilon}) \times \pi(\rho_{\eta}) \times \pi(\lambda_{\eta})
\]

- Posterior means and quantiles shown.
- Uncertainty in \(\theta, \eta(\cdot, \cdot)\), nuisance parameters are incorporated into the forecast.
- Gaussian process models for \(\eta(\cdot, \cdot)\).
A Schematic Outline of the Cosmic History

- **The Big Bang**
  - The Universe filled with ionized gas
- **~ 300 thousand years**
  - The Universe becomes neutral and opaque
  - The Dark Ages start
- **~ 500 million years**
  - Galaxies and Quasars begin to form
  - The Reionization starts
- **~ 1 billion years**
  - The Cosmic Renaissance
  - The Dark Ages end
  - Reionization complete, the Universe becomes transparent again
- **~ 9 billion years**
  - Galaxies evolve
- **~ 13 billion years**
  - The Solar System forms
- **Today:**
  - Astronomers figure it all out!

S.G. Djorgovski et al. & Digital Media Center, Caltech
Comparing Simulations with Observations

Dark matter simulation, two dimensional, projected density field, \((90 \text{ Mpc})^3\) box

Projected galaxy distribution measured by the Sloan Digital Sky Survey, total survey volume: \((500 \text{ Mpc})^3\)
Our analyses use statistical methods to combine different simulation codes and observational data.

Define problem: identify data, parameters and ranges, outputs of interest, codes.

Design simulation campaign over parameter ranges.

Do 64, 128, ..., runs of simulation code(s).

Statistical code (GPM).

- Response surface for simulation code
- Calibration distributions
- Model inadequacy
- Predictive distributions

Observed data.
Data, parameter ranges, simulations

Synthetic data were generated from a “true” cosmology using both linear perturbation theory and the particle mesh code MC²

Calibration parameter ranges
Spectral index  0.8 to 1.4
Hubble parameter  0.5 to 1.1
Sigma 8  0.6 to 1.6
Omega CDM  0.051 to 0.6
Omega baryon  0.02 to 0.12
Model of the data

\[ y(k) = \eta(\theta; k) + \epsilon(k) \]

Posterior density:

\[ \pi(\eta(\cdot; k), \theta, \xi|y) \propto L(y|\eta(\cdot, k), \theta, \Sigma_\epsilon) \times \pi(\eta(\cdot; k)|\xi) \times \pi(\theta) \times \pi(\xi) \]

\( \Sigma_\epsilon \) is known, \( \xi \) controls statistical parameters governing \( \eta(\cdot; k) \).

Posterior for cosmological parameters computed via MCMC

\[ \pi(\theta|y) \propto \int \pi(\eta(\cdot; k), \theta, \xi|y) \, d\eta \, d\xi \]
Basis representation of simulated spectra

The power spectra resulting from the 128 simulations are used to construct a mean-adjusted principal component representation.

Power spectra are represented as a function of the 5-d input parameters $\theta$ and PC basis functions $\phi_j(k)$:

$$
\hat{\eta}(\theta; k) = \sum_{j=1}^{p_\eta} w_j(\theta) \phi_j(k)
$$
Gaussian process model to *emulate* simulation output

Gaussian process (GP) models are used to estimate the weights $w_j(\theta)$ at untried settings

$$\hat{\eta}(\theta; k) = \sum_{j=1}^{p_{\eta}} w_j(\theta) \phi_j(k)$$
Response surface accuracy

holdout response surface fits

log P

4.5

5.5

3.5

2.5

log k

-3.0

-2.0

-1.0

holdout response surface residuals

residual (sim - resp.surf.)

-0.2

-0.1

0.0

0.1

0.2

-3.0

-2.0

-1.0

log k
Simulator emulation and sensitivity

Changes in the emulator prediction as each parameter is varied while holding the others at their midpoint.

Note: $\sigma_8$ and $\Omega_{CDM}$ have the largest effect on log P.

Only $\sigma_8$ has a substantial effect on nonlinear part of the mass power spectrum ($\log k < -1$).
Calibration results for test problem

Two separate analyses:
• Using data which lie on the linear part of the spectrum
• Using data over the entire spectrum
Accounting for limited simulation runs

Again, standard Bayesian estimation gives:

\[
\pi(\theta, \eta(\cdot, \cdot), \lambda_\epsilon, \rho_\eta, \lambda_\eta | y(x)) \propto L(y(x) | \eta(x, \theta), \lambda_\epsilon) \times \\
\pi(\theta) \times \pi(\eta(\cdot, \cdot) | \lambda_\eta, \rho_\eta) \times \\
\pi(\lambda_\epsilon) \times \pi(\rho_\eta) \times \pi(\lambda_\eta)
\]

- Posterior means and quantiles shown.
- Uncertainty in \(\theta, \eta(\cdot, \cdot)\), nuisance parameters are incorporated into the forecast.
- Gaussian process models for \(\eta(\cdot, \cdot)\).
Approximate the simulator with a linear function

- expand $\eta(x, \theta)$ about $(x_0, \theta_0)$

$$\eta(x, \theta) \approx \hat{\eta}(x, \theta) = \eta(x_0, \theta_0) + \nabla \eta(x_0, \theta_0)^T \left( x - x_0 \right) \theta - \theta_0$$

Now assume:

$$y(x) = \zeta(x) + e(x)$$

$$= \hat{\eta}(x, \theta) + e(x) \quad \theta \text{ unknown.}$$

Standard Bayesian estimation gives:

$$\pi(\theta|y(x)) \propto L(y(x)|\hat{\eta}(x, \theta)) \times \pi(\theta)$$

- in best case, this becomes linear regression
- numerous alternative versions (eg linearize about $\theta$ only).
Using low-fidelity simulations
Discussion

• Incorporating detailed information in simulation models can greatly improve inference.

• Extrapolation

• A number of other approaches are possible:
  Filtering methods
  Utilizing Derivatives
  Crude models
  Response surface models for modeling simulation output

• Choices depend on a number of factors:
  number of parameters in simulation model
  initial condition uncertainty
  speed of simulator
  complexity of simulator (can one get inside?)