Combining Simulations and Physical Observations to Estimate Cosmological Parameters

Dave Higdon, Statistical Sciences, Los Alamos National Laboratory
Katrin Heitmann, Space Science and Applications, Los Alamos National Laboratory
Salman Habib, Nuclear & Particle Physics, Astrophysics, and Cosmology, Los Alamos National Laboratory
Brian Williams, Statistical Sciences, Los Alamos National Laboratory
Charlie Nakhleh, Pulsed Power Sciences Center, Sandia National Laboratories
Earl Lawrence, Statistical Sciences, Los Alamos National Laboratory
Brief history: simulation-aided inference

@ LANL

Implosion calculations motivate first physics computer simulations

Computing evolves to carry out complex implosion simulations

Analysis algorithms evolve from LANL computing (e.g. Monte Carlo method)

Supercomputing, ASCI, computational models in the era of no nuclear testing
A Schematic Outline of the Cosmic History

- The Big Bang
  - The Universe filled with ionized gas
- The Universe becomes neutral and opaque
  - The Dark Ages start
- Galaxies and Quasars begin to form
  - The Reionization starts
- The Cosmic Renaissance
  - The Dark Ages end
- Reionization complete, the Universe becomes transparent again
- Galaxies evolve
- The Solar System forms
- Today: Astronomers figure it all out!

S.G. Djorgovski et al. & Digital Media Center, Caltech
The ΛCDM Model

• The “Standard Model of Cosmology”
• Controls content of universe
  – ~73% dark energy
  – ~23% dark matter
  – ~4% baryons
• Additional parameters such as the Hubble constant, optical depth, spectral index,…
• ~ 20 parameters in the standard model, known to ±10%
• For comparison, standard model for particle physics known to ±0.1%
Sloan Digital Sky Survey

data & simulated power spectra

log P(k)

wavenumber (log10 k)
Statistical framework

$$y(x_i) = \eta(x_i, \theta) + \epsilon_i$$
Gaussian Process Models in 1-d

An example of $z(s)$ of a Gaussian process model on $s_1, \ldots, s_n$

$$z = \begin{pmatrix} z(s_1) \\ \vdots \\ z(s_n) \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma \\ \vdots \\ \vdots \\ \Sigma \end{pmatrix} \right), \text{ with } \Sigma_{ij} = \exp\{-||s_i - s_j||^2\},$$

where $||s_i - s_j||$ denotes the distance between locations $s_i$ and $s_j$.

$z$ has density $\pi(z) = (2\pi)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}}\exp\{-\frac{1}{2}z^T\Sigma^{-1}z\}$. 
Gaussian Process Models in 1-d

Draws from \( \pi(z) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} z^T \Sigma^{-1} z \right\} \)
Conditioning on data (or simulations)

\[
\begin{pmatrix}
  z_1 \\
  z_2
\end{pmatrix}
\sim
N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)
\]

\[z_2 | z_1 \sim N\left(\Sigma_{21} \Sigma_{11}^{-1} z_1, \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}\right)\]

**Conditional mean**

**Conditional realizations**
A GP emulator of simulation output

Simulation output is typically smooth and noise free ideal for a GP model.
Gaussian Process models for combining field data and complex computer simulators

$$y = \begin{pmatrix} y(x_1) \\ \vdots \\ y(x_n) \end{pmatrix} \quad \text{field data} \quad \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p_x} \\ \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np_x} \end{pmatrix} \quad \text{input settings (spatial locations)}$$

$$\eta = \begin{pmatrix} \eta(x_1^*, \theta_1^*) \\ \vdots \\ \eta(x_m^*, \theta_m^*) \end{pmatrix} \quad \text{sim data} \quad \begin{pmatrix} x_{11}^* & \cdots & x_{1p_x}^* & \theta_{11}^* & \cdots & \theta_{1p_\theta}^* \\ \vdots & \vdots & \vdots \\ x_{m1}^* & \cdots & x_{mp_x}^* & \theta_{m1}^* & \cdots & \theta_{mp_\theta}^* \end{pmatrix} \quad \text{input settings } x; \text{ params } \theta^*$$

Model sim response $\eta(x, \theta)$ as a Gaussian process

$$y(x) = \eta(x, \theta) + \epsilon$$

$$\eta(x, \theta) \sim GP(0, C^\eta(x, \theta))$$

$$\epsilon \sim \text{iid } N(0, 1/\lambda_\epsilon)$$

$C^\eta(x, \theta)$ depends on $p_x + p_\theta$-vector $\rho_\eta$ and $\lambda_\eta$
Vector form – restricting to $n$ field observations and $m$ simulation runs

$$y = \eta + \epsilon$$

$$\eta \sim N_m(0_m, C^n(\rho_\eta, \lambda_\eta))$$

$$\Rightarrow \begin{pmatrix} y \\ \eta \end{pmatrix} \sim N_{n+m}\left( \begin{pmatrix} 0_n \\ 0_m \end{pmatrix}, C_y = C^{\eta} + \begin{pmatrix} 1/\lambda_\epsilon I_n & 0 \\ 0 & 1/\lambda_s I_m \end{pmatrix} \right)$$

where

$$C^{\eta} = 1/\lambda_\eta R^{\eta}\left( \begin{pmatrix} x \\ x^* \end{pmatrix}, \begin{pmatrix} 1\theta \\ \theta^* \end{pmatrix}; \rho_\eta \right)$$

and the correlation matrix $R^{\eta}$ is given by

$$R^{\eta}((x, \theta), (x', \theta'); \rho_\eta) = \prod_{k=1}^{p_x} \rho_{\eta_k}^4 (x_k - x'_k)^2 \times \prod_{k=1}^{p_\theta} \rho_{\eta(k+p_x)}^4 (\theta_k - \theta'_k)^2$$

$\lambda_s$ is typically set to something large like $10^6$ to stabilize matrix computations and allow for numerical fluctuation in $\eta(x, \theta)$.

note: the covariance matrix $C^{\eta}$ depends on $\theta$ through its “distance”-based correlation function $R^{\eta}((x, \theta), (x', \theta'); \rho_\eta)$.

We use a 0 mean for $\eta(x, \theta)$; an alternative is to use a linear regression mean model.
Likelihood

\[ L(y, \eta | \lambda_\epsilon, \rho_\eta, \lambda_\eta, \lambda_s, \theta) \propto \]
\[ |C_{y\eta}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (y^T C_{y\eta}^{-1} y) \right\} \]

Priors

\[ \pi(\lambda_\epsilon) \propto \lambda_\epsilon^{-1} e^{-b_\epsilon \lambda_\epsilon} \] perhaps well known from observation process

\[ \pi(\rho_\eta k) \propto \prod_{k=1}^{p_x+p_\theta} (1 - \rho_\eta k)^{-\frac{1}{2}}, \text{ where } \rho_\eta k = e^{-\frac{1}{2} \beta_k^2} \text{ correlation at dist } = .5 \sim \beta(1,.5). \]

\[ \pi(\lambda_\eta) \propto \lambda_\eta^{-1} e^{-b_\eta \lambda_\eta} \]

\[ \pi(\lambda_s) \propto \lambda_s^{-1} e^{-b_s \lambda_s} \]

\[ \pi(\theta) \propto I[\theta \in C] \]

- could fix \( \rho_\eta, \lambda_\eta \) from prior GASP run on model output.
- Many prefer to reparameterize \( \rho \) as \( \beta = -\log(\rho) \)/.5^2 in the likelihood term
Posterior Density

\[
\pi(\lambda_\epsilon, \rho_\eta, \lambda_\eta, \lambda_s, \theta | y, \eta) \propto |C_{y\eta}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left( \frac{y}{\eta} \right)^T C_{y\eta}^{-1} \left( \frac{y}{\eta} \right) \right\} \times \\
p_{x+p_\theta} \prod_{k=1}^{p_x+p_\theta} (1 - \rho_{\eta_k})^{-0.5} \times \lambda_{\eta}^{a_{\eta}-1} e^{-b_\eta \lambda_\eta} \times \lambda_{s}^{a_{s}-1} e^{-b_s \lambda_s} \times \\
\lambda_{\epsilon}^{a_{\epsilon}-1} e^{-b_\epsilon \lambda_\epsilon} \times I[\theta \in C]
\]

If \( \rho_\eta, \lambda_\eta, \) and \( \lambda_s \) are fixed from a previous analysis of the simulator data, then

\[
\pi(\lambda_\epsilon, \theta | y, \eta, \rho_\eta, \lambda_\eta, \lambda_s) \propto |C_{y\eta}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left( \frac{y}{\eta} \right)^T C_{y\eta}^{-1} \left( \frac{y}{\eta} \right) \right\} \times \\
\lambda_{\epsilon}^{a_{\epsilon}-1} e^{-b_\epsilon \lambda_\epsilon} \times I[\theta \in C]
\]
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 \( \eta(\cdot, \cdot) \).
Predicting a new outcome: \( \zeta = \zeta(x') = \eta(x', \theta) \)

Given a MCMC realization \((\theta, \lambda, \rho, \lambda)\), a realization for \(\zeta(x')\) can be produced using Bayes rule.

\[
v = \begin{pmatrix} y \\ \eta \\ \zeta \end{pmatrix} \Sigma_v = \begin{pmatrix} \lambda \epsilon I_n & 0 & 0 \\ 0 & \lambda_s I_m & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \mu_z = \begin{pmatrix} 0_n \\ 0_m \\ 0 \end{pmatrix} \quad C_\eta = \lambda^{-1}_\eta R^n \begin{pmatrix} x \\ x^* \\ \theta^* \\ \theta \end{pmatrix} \rho_\eta
\]

Now the posterior distribution for \(v = (y, \eta, \zeta)^T\) is

\[
v \mid y, \eta \sim N(\mu^{v \mid y, \eta} = V \Sigma_v v, V), \quad \text{where } V = (\Sigma_v + C_\eta^{-1})^{-1}
\]

Restricting to \(\zeta\) we have

\[
\zeta \mid y, \eta \sim N(\mu^{\zeta \mid y, \eta} = V_n+m+1, V_{n+m+1,n+m+1})
\]

Alternatively, one can apply the conditional normal formula to

\[
\begin{pmatrix} y \\ \eta \\ \zeta \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda^{-1}_\epsilon I_n & 0 & 0 \\ 0 & \lambda^{-1}_s I_m & 0 \\ 0 & 0 & 0 \end{pmatrix} + C_\eta \right)
\]

so that

\[
\zeta \mid y, \eta \sim N \left( \Sigma^{-1}_{21 \Sigma_{11}} \begin{pmatrix} y \\ \eta \end{pmatrix}, \Sigma_{22} - \Sigma_{21} \Sigma^{-1}_{11} \Sigma_{12} \right)
\]
Statistical framework \[ y(x_i) = \eta(x_i, \theta) + \delta(x_i) + \epsilon_i \]
Statistical framework \( y(x_i) = \eta(x_i, \theta) + \delta(x_i) + \epsilon_i \)
Data, parameter ranges, and simulations

Physical observations and simulations

OA-LHS 128 run design

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
Basis representation of simulated spectra

Basis representation for matter power spectra.

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 multivariate simulation output

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

\[
\eta(\theta; k) = \sum_{j=1}^{p_{\eta}} w_j(\theta) \phi_j(k)
\]

Prediction at new \( \theta \)
Response surface is sufficiently accurate for this application
Simulator emulation and sensitivity

- Changes in emulator prediction as each parameter is varied while holding the others at their midpoint
- Note, $\sigma_8 \Omega_{CDM}$ have the largest effect on $\log P$

Posterior distribution of cosmological parameters

Resulting fit and uncertainty for the matter power spectrum
Methodology allows combining multiple computational models and data sources

Sloan Digital Sky Survey

Wilkinson microwave anisotropy probe

![CMB TT power spectrum: WMAP data & simulations](image)
Combined WMAP & SDSS analysis

Combining information from both data sources sharpens inference on ΛCDM parameters
This basic template for cosmology is applicable in other scientific investigations.

Observation/experiment: large scale structure of universe

Hydrodynamic behavior

Calibration: finding parameter settings consistent with observations

Prediction uncertainties

notional data & simulations

0.1 2 3 4 x 10^-5 s

0.1 2 3 4 x 10^-5 s
Simple, 1-d inverse problem: \( y = \eta(\theta) + \epsilon \)

Prior model
\[
\pi(\theta) \propto \exp\left\{ -\frac{1}{2} \theta^2 \right\}
\]

Observation model:
\[
L(y|\theta) \propto \exp\left\{ -\frac{1}{2\sigma^2} (y - \eta(\theta))^2 \right\}
\]

Computational model
\[
\eta(\cdot)
\]

Posterior
\[
\pi(\theta|y) \propto L(y|\theta) \times \pi(\theta)
\]
Using a Gaussian Process Emulator

Prior model
\[ \pi(\theta) \propto \exp \left\{ -\frac{1}{2} \theta^2 \right\} \]

Observation model:
\[ L(y|\theta) \propto \exp \left\{ -\frac{1}{2\sigma^2} (y - \eta(\theta))^2 \right\} \]

Computational model
\[ \eta(\cdot) \sim GP(\mu(\cdot), C(\cdot, \cdot)) \]

Posterior
\[ \pi(\theta, \eta(\cdot)|y) \propto L(y|\theta, \eta(\cdot)) \times \pi(\theta) \times \pi(\eta(\cdot)) \]
Ensemble Kalman smoother: normal version

Prior model

\[
\begin{pmatrix}
\theta \\
\eta(\theta)
\end{pmatrix}
\sim N(\mu_{pr}, \Sigma_{pr})
\]

Observation model:

\[
\begin{pmatrix}
\theta \\
\eta(\theta)
\end{pmatrix}
\sim N(\mu_{obs} = \begin{pmatrix} \star \\ y \end{pmatrix}, \Sigma_{obs} = \begin{pmatrix} 0 & 0 \\ 0 & \sigma_y^{-2} \end{pmatrix})
\]

Posterior

\[
\begin{pmatrix}
\theta \\
\eta(\theta)
\end{pmatrix}
\sim N(\mu_{post}, \Sigma_{post})
\]

\[
\Sigma_{post}^{-1} = \Sigma_{pr}^{-1} + \Sigma_{obs}^{-1}
\]

\[
\mu_{post} = \Sigma_{post} (\Sigma_{pr}^{-1} \mu_{pr} + \Sigma_{obs}^{-1} \mu_{obs})
\]
Ensemble Kalman smoother: ensemble version

Perturb each pair \( \begin{pmatrix} \theta_k \\ \eta(\theta_k) \end{pmatrix} \rightarrow \begin{pmatrix} \theta \\ \eta(\theta) \end{pmatrix}_k \) so that the sample \( \begin{pmatrix} \theta \\ \eta(\theta) \end{pmatrix}_1, \ldots, \begin{pmatrix} \theta \\ \eta(\theta) \end{pmatrix}_m \) is an approximate collection of draws from the posterior distribution.

Use posterior mean induced by combining these two (random) information sources: 
\[
\begin{pmatrix} \theta_k \\ \eta(\theta_k) \end{pmatrix} \sim N(\mu_{pr}, \Sigma_{pr}) \quad \text{and} \quad y_k \sim N(y, \sigma_y^2) .
\]
Ensemble Kalman smoother: ensemble version

Perturb each pair \( \begin{pmatrix} \theta_k \\ \eta(\theta_k) \end{pmatrix} \rightarrow \begin{pmatrix} \theta \\ \eta(\theta) \end{pmatrix}_k \) so that the sample \( \begin{pmatrix} \theta \\ \eta(\theta) \end{pmatrix}_1, \ldots, \begin{pmatrix} \theta \\ \eta(\theta) \end{pmatrix}_m \) is an approximate collection of draws from the posterior distribution.

Use posterior mean induced by combining these two (random) information sources:

\[
\begin{pmatrix} \theta_k \\ \eta(\theta_k) \end{pmatrix} \sim N(\mu_{pr}, \Sigma_{pr}) \quad \text{and} \quad y_k \sim N(y, \sigma_y^2)
\]
Ensemble Kalman smoother: ensemble version

Perturb each pair \( \left( \frac{\theta_k}{\eta(\theta_k)} \right) \rightarrow \left( \frac{\theta}{\eta(\theta)} \right)_k \) so that the sample \( \left( \frac{\theta}{\eta(\theta)} \right)_1, \ldots, \left( \frac{\theta}{\eta(\theta)} \right)_m \) is an approximate collection of draws from the posterior distribution.

Use posterior mean induced by combining these two (random) information sources:

\[
\begin{align*}
\left( \frac{\theta_k}{\eta(\theta_k)} \right) &\sim N(\mu_{pr}, \Sigma_{pr}) \quad \text{and} \quad y_k \sim N(y, \sigma_y^2) \\
\end{align*}
\]
Ensemble Kalman smoother: ensemble version

Prior model: centered @ $k^{th}$ ensemble

\[
\begin{align*}
\left( \begin{array}{c} 
\theta \\
\eta(\theta)
\end{array} \right)_k & \sim N \left( \begin{array}{c} 
\theta_k \\
\eta(\theta_k)
\end{array} , \Sigma_{pr} \right) \\
\left( \begin{array}{c} 
\theta \\
\eta(\theta)
\end{array} \right)_k & \sim N \left( \begin{array}{c} 
\mu_{obs} = \left( \begin{array}{c} 
* \\
y_k
\end{array} \right) \\
\Sigma_{obs} = \left( \begin{array}{cc} 
0 & 0 \\
0 & \sigma_y^{-2}
\end{array} \right) \right)
\end{align*}
\]

Observation model with perturbed data $y_k \sim N(y, \sigma^2)$

Set $k^{th}$ ensemble value to the posterior mean

\[
\begin{align*}
\left( \begin{array}{c} 
\theta \\
\eta(\theta)
\end{array} \right)_k & = \left( \Sigma_{pr}^{-1} + \Sigma_{obs}^{-1} \right)^{-1} \left( \Sigma_{pr}^{-1} \left( \begin{array}{c} 
\theta_k \\
\eta(\theta_k)
\end{array} \right) + \Sigma_{obs}^{-1} \left( \begin{array}{c} 
* \\
y_k
\end{array} \right) \right)
\end{align*}
\]
Ensemble Kalman smoother: ensemble version

Prior model: centered @ $k^{\text{th}}$ ensemble  
Observation model with perturbed data $y_k \sim N(y, \sigma^2)$

\[
\begin{align*}
\left( \begin{array}{c} \theta \\ \eta(\theta) \end{array} \right)_{k} & \sim N \left( \left( \begin{array}{c} \theta_k \\ \eta(\theta_k) \end{array} \right), \Sigma_{\text{pr}} \right) \\
\left( \begin{array}{c} \theta \\ \eta(\theta) \end{array} \right)_{k} & \sim N \left( \mu_{\text{obs}} = \left( \begin{array}{c} y_k \\ \ast \end{array} \right), \Sigma_{\text{obs}}^{-1} = \left( \begin{array}{cc} 0 & 0 \\ 0 & \sigma_{y}^{-2} \end{array} \right) \right)
\end{align*}
\]

Set $k^{\text{th}}$ ensemble value to the posterior mean (using Kalman gain)

\[
\begin{align*}
\left( \begin{array}{c} \theta \\ \eta(\theta) \end{array} \right)_{k} & = \left( \begin{array}{c} \theta_k \\ \eta(\theta_k) \end{array} \right) + \Sigma_{\text{pr}} \Sigma' (\Sigma_{\text{pr}} \Sigma' + \Sigma_{\text{obs}} \Sigma')^{-1} (y_k - \eta(\theta_k))
\end{align*}
\]
Ensemble Kalman smoother: ensemble version

Prior model: centered @ \( k \)th ensemble

\[
\begin{align*}
\left( \begin{array}{c}
\theta \\
\eta(\theta)
\end{array} \right)_k & \sim N \left( \left( \begin{array}{c}
\theta_k \\
\eta(\theta_k)
\end{array} \right), \Sigma_{pr} \right) \\
\left( \begin{array}{c}
\theta \\
\eta(\theta)
\end{array} \right)_k & \sim N \left( \mu_{obs} = \left( \begin{array}{c}
y_k \\
\end{array} \right), \Sigma_{obs} = \left( \begin{array}{cc}
0 & 0 \\
0 & \sigma_y^{-2}
\end{array} \right) \right)
\end{align*}
\]

Set \( k \)th ensemble value to the posterior mean

\[
\begin{align*}
\left( \begin{array}{c}
\theta \\
\eta(\theta)
\end{array} \right)_k &= \left( \Sigma_{pr}^{-1} + \Sigma_{obs}^{-1} \right)^{-1} \left( \Sigma_{pr}^{-1} \left( \begin{array}{c}
\theta_k \\
\eta(\theta_k)
\end{array} \right) + \Sigma_{obs}^{-1} \left( \begin{array}{c}
y_k \\
\end{array} \right) \right)
\end{align*}
\]
Cosmology application: large scale structure

Ensemble parameter design

Resulting Matter Spectra & Data
Cosmology application: large scale structure

Posterior Parameter Density

Matter Spectrum Estimate

GP-based Emulator
Cosmology application: large scale structure

**Posterior Parameter Density**

**Matter Spectrum Estimate**

**GP-based Emulator vs. EnKS (normal)**
Cosmology application: large scale structure

Posterior Parameter Density

Matter Spectrum Estimate

GP-based Emulator vs. EnKS (ensemble)
Ending thoughts on calibration & UQ

• Design – choosing input settings at which to run model
  – Adaptive approaches
  – Searching for high-consequence events

• Dealing with extrapolations
  – Need to understand what predictions are/are not extrapolations
  – Need to mark out conditions in which model predictions can be trusted
  – Multiple model approaches
  – Theoretical approaches (e.g. bounding tail probabilities, bounding model errors)

• Making use of lower fidelity and/or reduced models to speed up exploration of input space

• Resource allocation
  – What new data sources impact uncertainties? By how much?
  – What is the most cost effective way to reduce our uncertainties?

• Other settings:
  – Physically constrained models and limited amounts of data vs. weaker, empirical models with large amounts of physical observations.