Bayesian calibration of computer models - Kennedy & O'Hagan (2001)

Shivi Vaidyanathan

April 17, 2014

Outline

1. Calibration
   - Problem definition
   - Notation
2. Model
3. Application
4. Refrences

Outline

Framework

- Physical Process of interest - difficult to obtain observations
- Mathematical Model
  - Deterministic
  - Can be expensive in terms of computing time
- To make useful predictions, need to **calibrate** the computer model with observed data

Motivating Example

Dose regime for a new drug (size, frequency and release rate of tablets)

- Pharmacokinetic model -
  - Need to specify rates with which it moves between different body compartments
  - Experiments provide - outputs of the pharmacokinetic model such as conc. of drug in blood or urine at certain time points
  - Calibration here : adjusting the unknown rate parameters till the model fits the observed data
- Calibration is using observed data to learn about context specific inputs to the Computer model.

Calibration Inputs and Variable Inputs

- Distinguish between two types of inputs
  - Calibration inputs : Context specific parameters that are unknown in the true process
    - In Dose Regime example, the rate parameters
  - Variable inputs : All other inputs that vary in the model that can be 'controlled' or observed for the true process
    - In example : Size, frequency, release rate of tablets

Notation

- Variable inputs $\mathbf{x} = \{x_1, \ldots, x_{q_1}\}$
- Calibration parameters $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_{q_2}\}$
- Denote **Calibration Inputs** to computer model $\mathbf{t} = \{t_1, \ldots, t_{q_2}\}$
- Computer model response at inputs $\mathbf{x}$ and $\mathbf{t}$ by $\eta(\mathbf{x}, \mathbf{t})$
- True Process $\zeta(\mathbf{x})$

Notation, contd

- Data
  - Observations $\mathbf{z} = \{z_1, \ldots, z_n\}$
  - Can only use a limited number of model runs
  - Computer model runs $\mathbf{y} = \{y_1, \ldots, y_N\}$ where $y_j = \eta(\mathbf{x}_j, \mathbf{t}_j)$
  - Full data $\mathbf{d}^\mathbf{T} = \{\mathbf{z}^\mathbf{T}, \mathbf{y}^\mathbf{T}\}$

Outline

Data model

Assume that the data relates to the true process

$$z_i = \zeta(\mathbf{x}_i) + e_i \quad (1)$$

$$\zeta(\mathbf{x}_i) = \rho\eta(\mathbf{x}_i, \boldsymbol{\theta}) + \delta(\mathbf{x}_i) \quad (2)$$

Assume

- e_i's are independent $N(0, \lambda)$
- ρ unknown regression parameter
- $\delta(.)$ independent of $\eta(.,.)$

Prior Distributions

- Prior distributions of the unknown functions $\eta(.,.)$ and $\delta(.)$

$$\begin{aligned}
\eta(\cdot\,,\cdot\,) &\sim N(m_1(\cdot\,,\cdot\,), c_1((\cdot\,,\cdot\,),(\cdot\,,\cdot\,)));\\
m_1(\mathbf{x},\boldsymbol{\theta}) &= h_1(\mathbf{x},\boldsymbol{\theta})^T\beta_1; \qquad (3)\\
\delta(\cdot\,) &\sim N(m_2(\cdot\,), c_2((\cdot\,,\cdot\,)))\\
m_2(\mathbf{x}) &= h_2(\mathbf{x})^T\beta_2 \qquad (4)
\end{aligned}$$

- Define ψ_1 and ψ_2 as parameters corresponding to covariance functions $c_1((\cdot\,,\cdot\,),(\cdot\,,\cdot\,))$ and $c_2((\cdot\,,\cdot\,))$
- Note that we specify (3) only when the computer model is expensive to run and hence we need to interpolate $\eta(\cdot\,,\cdot\,)$ at unseen values of $(\mathbf{x},\mathbf{t})$

- Prior distributions on parameters

$$\pi(\beta_1, \beta_2) \propto 1 \tag{5}$$

- Denote $\boldsymbol{\phi} = \{\rho, \lambda, \boldsymbol{\psi}\}$ and $\boldsymbol{\beta} = (\beta_1^T, \beta_2^T)^T$
- Complete set of parameters $\{\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\phi}\}$

$$\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\phi}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{\phi}) \tag{6}$$

Posterior distribution

$$\mathbf{d}^\mathbf{T} = \{\mathbf{z}^\mathbf{T}, \mathbf{y}^\mathbf{T}\}$$

$$\begin{aligned} z_i &= \rho\eta(\mathbf{x}_i, \boldsymbol{\theta}) + \delta(\mathbf{x}_i) + e_i, \quad i = 1, \ldots, n \\ y_i &= \eta(\mathbf{x}_i, \mathbf{t}_i), \quad i = 1 \ldots, N \end{aligned}$$

- Can write down data likelihood

$$\mathbf{d}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\phi} \sim N(m_d(\boldsymbol{\theta}), V_d(\boldsymbol{\theta})) \tag{7}$$

- Posterior Distribution

$$\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\phi}|\mathbf{d}) \propto \pi(\mathbf{d}|m_d(\boldsymbol{\theta}), V_d(\boldsymbol{\theta}))\pi(\boldsymbol{\theta})\pi(\boldsymbol{\phi}) \tag{8}$$

Estimating hyperparameters

- We have, $\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\phi}|\mathbf{d})$, Interested in $\pi(\boldsymbol{\theta}|\mathbf{d})$
- Integrate out $\boldsymbol{\beta}$
- Estimate $\boldsymbol{\phi}$ in two steps
  - Use **y** to estimate ψ_1 of $c_1((\cdot\,,\cdot),(\cdot\,,\cdot))$
  - Fix ψ_1 and use **z** to estimate ρ, λ and ψ_2 of $c_2(\cdot\,,\cdot)$

Calibration and prediction

- With estimates of ϕ we can then write
- $\pi(\boldsymbol{\theta}|\phi = \hat{\phi}, \mathbf{d}) \propto \pi(\boldsymbol{\theta}, \hat{\phi}|\mathbf{d})$
- Predicting true process at unobserved locations $\zeta(\mathbf{x})$ can be done by computing
$$\zeta(\mathbf{x})|\boldsymbol{\theta}, \phi, \mathbf{d} \tag{9}$$
- We can then make inference on $\zeta(\mathbf{x})|\hat{\phi}, \mathbf{d}$ using (9) and $\pi(\boldsymbol{\theta}|\hat{\phi}, \mathbf{d})$

Modeling choices

- Need to specify $h_1(\mathbf{x}, \mathbf{t})$, $h_2(\mathbf{x})$,$c_1((\mathbf{x}, \mathbf{t}), (\mathbf{x}', \mathbf{t}'))$ and $c_2((\mathbf{x}, \mathbf{x}'))$.
- Set $h_1(\mathbf{x}, \mathbf{t}) = 1$ and $h_2(\mathbf{x}) = 1$
- Then, β_1 and β_2 are scalars that represent an unknown constant mean.
- For the covariance functions, they choose

$$\begin{aligned} c_1((\boldsymbol{x}, \boldsymbol{t}), (\boldsymbol{x}', \boldsymbol{t}')) &= \sigma_1^2 \exp\{-(\boldsymbol{x} - \boldsymbol{x}')^T \boldsymbol{\Omega}_x (\boldsymbol{x} - \boldsymbol{x}')\} \exp\{-(\boldsymbol{t} - \boldsymbol{t}')^T \boldsymbol{\Omega}_t (\boldsymbol{t} - \boldsymbol{t}')\}, \\ c_2(\boldsymbol{x}, \boldsymbol{x}') &= \sigma_2^2 \exp\{-(\boldsymbol{x} - \boldsymbol{x}')^T \boldsymbol{\Omega}_x^* (\boldsymbol{x} - \boldsymbol{x}')\}, \end{aligned}$$

- with diagonal forms for Ω_t, Ω_x and Ω_x^*

Outline

Gaussian Plume Model(GPM)

Radiological Protection

- GPM used to predict dispersion and subsequent deposition of radioactive material following an accidental release
- Code inputs
  - Atmospheric conditions at release time (wind direction, wind speed, atmospheric stability)
  - Nature of release (source term, source location, release height, release duration, deposition velocity)
- Very cheap to run

Example: Tomsk Data

- Accident at the Tomsk-7 chemical plant Russia (1993)
- Deposition of ruthenium 106 (^{106}Ru)
- 695 measurements of ^{106}Ru deposition were made at locations shown in figure

Tomsk Data

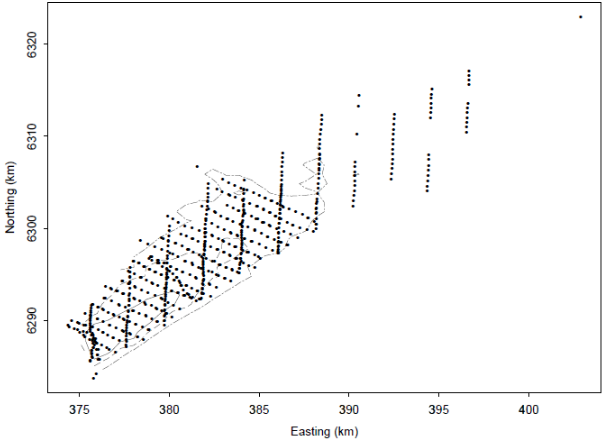
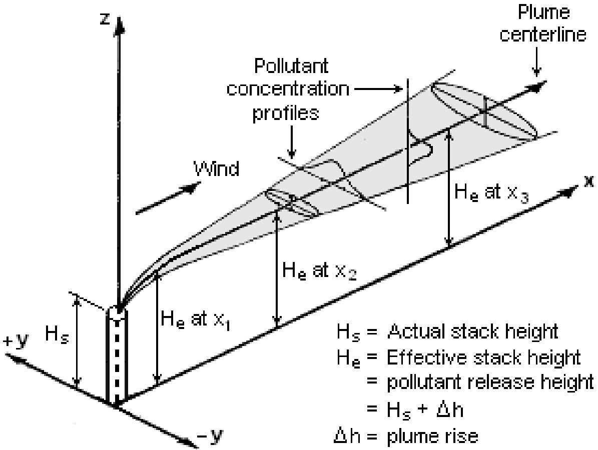


Fig. 1. Tomsk aerial survey of 695 Ru106 deposition measurements, with contours at heights of 11 (solid line), 10 (— — —) and 9 (— - —).

Prior specifications

- $\eta(\cdot,\cdot)$ use Gaussian Plume model
- Calibration parameters $\boldsymbol{\theta}$: Logarithm of source term and deposition velocity
- Variable inpute: Two orthogonal linear functions of the northing and easting co-ordinates such that $x(0,0)$ represents the source point and $x(x_1,x_2)$ represents point at distance x_1 downwind and distance x_2 from plume center line.
- Assume normal priors for $\boldsymbol{\theta}$
- Prior means obtained from National Radiological Protection Board
- Prior variance set to 5

Gaussian plume model



Model assumptions

- Since GPM is very cheap to run, $\eta(\cdot\,,\cdot)$ treated as known
- Simplifies model significantly, only covariance function to be specified is $c_2(\cdot\,,\cdot)$
- Assume product Gaussian form $\sigma^2 r(\mathbf{x} - \mathbf{x}')$

$$r(x - x') = \exp\{-\sum_{j=1}^{q} \omega_j (x_j - x'_j)^2\}.$$

- So the roughness parameters $\psi_2 = \{\omega_1, \omega_2\}$

Experimental setup

To conduct the analysis,

- Observed data : Use subset of size $\{n = 10, 15, 20 \text{ and } 25\}$ of the 695 measurements
- Computed posterior means and variances of $z(x)$ for 670 'unobserved' locations
- Accuracy was assessed on the basis of true values at these points.
- Three strategies used
  - Use GP interpolation of the physical observations alone
  - Use Bayesian Calibration technique described here
  - Use Gaussian plum model with 'plug in' parameters. Physical data are not interpolated in anyway, choose 'plug in' estimates by minimizing sum of squared differences between model and data.

Results

$RMSE$	n=10	n=15	n=20	n=25
Strategy 1	0.75	0.76	0.86	0.79
Strategy 2	0.42	0.41	0.37	0.36
Strategy 3	0.82	0.79	0.76	0.66

Results contd

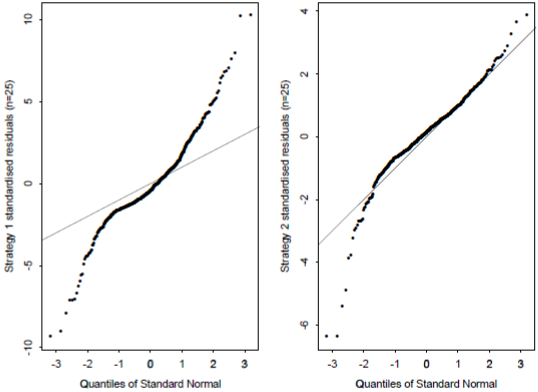


Fig. 2. Quantile-Quantile plots for Strategies 1 and 2 with $n = 25$

Outline

Marc C. Kennedy and Anthony O'Hagan (2001) *Bayesian Calibration of Computer Models*. *J. R. Statist. Soc. B*, **63**,Part 3. 425-464.

Thank you!