Bayesian inference for expensive computer models in chemical engineering

Peter L. W. Man, Andreas Braumann and Markus Kraft

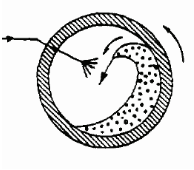
CoMo Group
Department of Chemical Engineering and Biotechnology
University of Cambridge

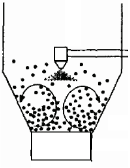
7th March 2011

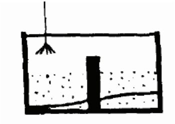
Granulation

- Products

- Devices

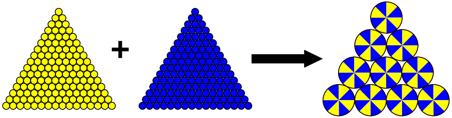




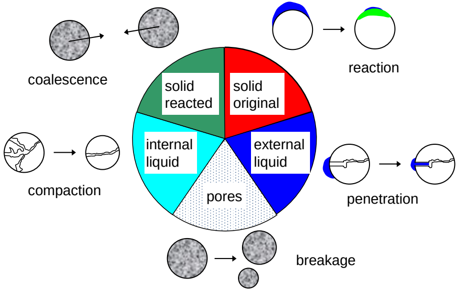


Granulation

- Lump small particles into bigger entities
  - Improve handling (storage, transport, safety,...)
  - Creation of micro mixtures (segregation, distribution of active components)
- Desire knowledge about:
  - Influences of precursors, process design and conditions on final product



Granulation model



transformation	unknown parameter
coalescence	k_{coag}
compaction	k_{compact}
breakage	k_{break}
penetration	k_{pen}
reaction	k_{reac}

Regression I

Aim: Regress for some data points $\boldsymbol{y}_{\text{obs}}$ over input space $\mathcal{Z}$

- Assume

$$y_{\text{obs}}(z) = y(z) + \epsilon \quad \text{where} \quad \epsilon \sim \mathcal{N}(0, \tau^2) \tag{1}$$

- Underlying $y(z)$ is *unknown*, so endow with prior distribution :
  - $y(.) \sim \text{GP}(\,\mu(.), \sigma^2\Sigma(.,.)\,)$
  - mean function $\mu(z)$
  - correlation function $\Sigma(z_1, z_2)$

Regression II

- Make observations $\boldsymbol{y}_{\text{obs}} := y(\boldsymbol{z}) + \boldsymbol{\epsilon}$ at $\boldsymbol{z} = (z_1, \ldots, z_n)^\top$
- Want predictions at $\widetilde{\boldsymbol{z}}$

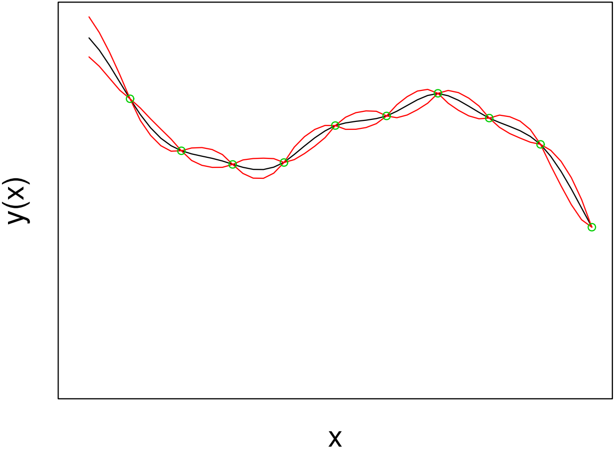
Posterior of $y(\widetilde{\boldsymbol{z}}) \,|\, \boldsymbol{y}_{\text{obs}}$:

- $y(\widetilde{\boldsymbol{z}}) \,|\, \boldsymbol{y}_{\text{obs}}$ has a multivariate normal distribution with

$$\mathbb{E}\left[y(\widetilde{\boldsymbol{z}}) \,|\, \boldsymbol{y}_{\text{obs}}\right] = \mu(\widetilde{\boldsymbol{z}}) + \sigma^2\Sigma(\widetilde{\boldsymbol{z}}, \boldsymbol{z}) \left[\sigma^2\Sigma(\boldsymbol{z}, \boldsymbol{z}) + \tau^2 I\right]^{-1} (\boldsymbol{y}_{\text{obs}} - \mu(\boldsymbol{z})) \quad (2)$$

$$\text{Var}\left(y(\widetilde{\boldsymbol{z}}) \,|\, \boldsymbol{y}_{\text{obs}}\right) = \sigma^2\Sigma(\widetilde{\boldsymbol{z}}, \widetilde{\boldsymbol{z}}) - \sigma^4\Sigma(\widetilde{\boldsymbol{z}}, \boldsymbol{z}) \left[\sigma^2\Sigma(\boldsymbol{z}, \boldsymbol{z}) + \tau^2 I\right]^{-1} \Sigma(\widetilde{\boldsymbol{z}}, \boldsymbol{z})^\top \quad (3)$$

Regression IV



Correlation function

Choices of $\Sigma(z_1, z_2)$:

- Exponential :

$$\Sigma(z_1, z_2) = \prod_{k=1}^{\dim(\mathcal{Z})} \exp\left(-\phi_k \cdot |z_{1k} - z_{2k}|^{\phi_0}\right) \quad (4)$$

- Matérn :

$$\Sigma(z_1, z_2) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\nu^{\frac{1}{2}}||z_1 - z_2||}{\rho}\right)^{\nu} \mathcal{K}_{\nu}\left(\frac{2\nu^{\frac{1}{2}}||z_1 - z_2||}{\rho}\right) \quad (5)$$

where $\mathcal{K}$ is modified Bessel function of 2^{nd} kind.

Notation

- $\mathcal{X}$ is space of *process conditions* or covariates
- Θ is space of unknown parameters of a model
- $\mathcal{Z} := \mathcal{X} \times \Theta$

Regression Models

- Underlying **Simulator** function $\eta_s(z)$:
  - $\eta_s(z) \sim \text{GP}\left(\mu_s(.), \sigma_s^2 \Sigma_s(.,.)\right)$
- Underlying **Experimental** function $\eta_e(x)$:
  - Fix calibrated value of $\theta = \theta_c$ (unknown)
  - $\eta_e(x) := \eta_s(x, \theta_c) + \delta(x)$
  - $\delta(x) \sim \text{GP}\left(\mu_e(.), \sigma_e^2 \Sigma_e(.,.)\right)$
- $\implies \eta_e \sim \text{GP}\left(\mu_s(.,.) + \mu_e(.)\,,\, \sigma_s^2 \Sigma_s(.,.) + \sigma_e^2 \Sigma_e(.,.)\right)$

Available data

- Observed **Experimental** data :
  - Have $\boldsymbol{y}_{\mathsf{e}} := (\eta_e(x_1), \ldots, \eta_e(x_{n_e}))^\top + \boldsymbol{\epsilon}_e$
- Observed **Simulator** data :
  - Have $\boldsymbol{y}_{\mathsf{s}} := \left(\eta_s(x_1^*, \theta_1^*), \ldots, \eta_s(x_{n_e}^*, \theta_{n_e}^*)\right)^\top + \boldsymbol{\epsilon}_s$
- Full data $\boldsymbol{d} := (\boldsymbol{y}_s^\top, \boldsymbol{y}_e^\top)^\top$
- Observation errors ϵ_e, ϵ_s for simplicity:
  - $\boldsymbol{\epsilon}_e \sim \mathcal{N}(0, \tau_e^2\, I)$ and $\boldsymbol{\epsilon}_s \sim \mathcal{N}(0, \tau_s^2\, I)$

Priors

- GP Priors
  - GP Matérn parameters ϕ_e, ϕ_s—$U[0, \phi_{\text{upper}}]$
  - GP mean linear parameters β_e, β_s—constant prior
  - GP variance parameters σ_e^2, σ_s^2—inverse gamma
- Observation error priors
  - τ_e^2, τ_s^2—inverse gamma
- Prior for θ is given by user
- All unknown values — represent as $(\theta, \beta_s, \beta_e, \boldsymbol{\xi})$

θ-estimation

- Posterior for $(\theta, \beta_s, \beta_e, \boldsymbol{\xi})$:

$$\pi(\theta, \beta_s, \beta_e, \boldsymbol{\xi} \,|\, \boldsymbol{d}) \propto \underbrace{f(\boldsymbol{d} \,|\, \theta, \beta_s, \beta_e, \boldsymbol{\xi})}_{\text{gaussian likelihood}} \times \underbrace{\pi(\theta)\pi(\beta_s)\pi(\beta_e)\pi(\boldsymbol{\xi})}_{\text{prior}} \tag{6}$$

- Integrate out β_s and β_e *analytically*
- Integrating out $\boldsymbol{\xi}$ is harder - do this via Wang-Laudau sampling
- Do similar for prediction of $\eta_e(x)$ and $\eta_s(x, \theta)$.

Example - simulator function $\eta_s(x, \theta)$

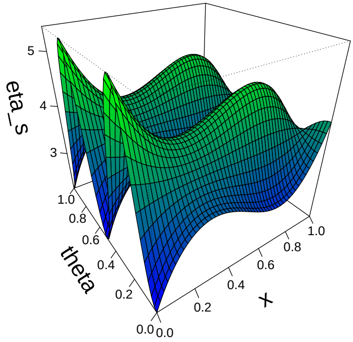
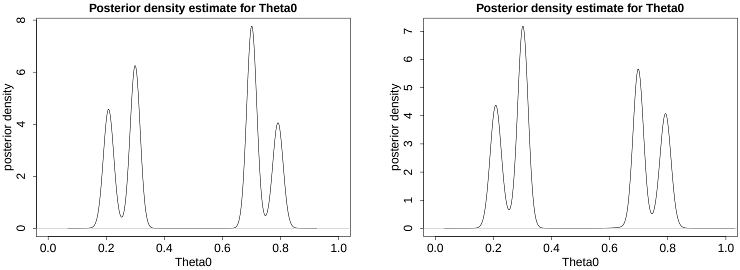


Figure: True $\eta_s(x, \theta)$

$$\mu_s(x, \theta) := \frac{1}{4}\left[6(x - 0.15)(x - 0.5)(x - 0.85) - 0.27\right]\left[10 \cos(4\pi\theta) + 0.3\right] + \text{const}$$

$$\theta_c = 0.2$$

Example - θ-posterior

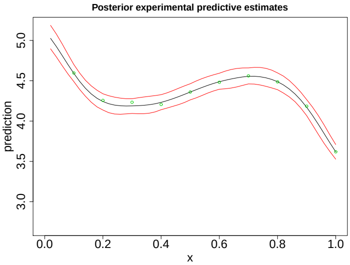


(a) assuming $\tau_s^2 = 0$, τ_e^2 unknown, $\delta(x) \equiv 0$

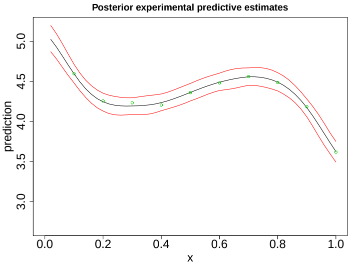
(b) assuming $\tau_s^2 = 0$, τ_e^2 and $\delta(x)$ unknown

Figure: θ posterior plots making different assumptions

Example - $y_e(x)$-posterior prediction + noise



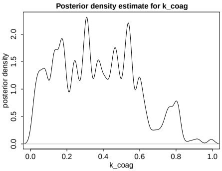
(a) assuming $\tau_s^2 = 0$, τ_e^2 unknown, $\delta(x) \equiv 0$



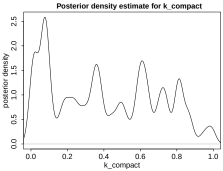
(b) assuming $\tau_s^2 = 0$, τ_e^2 and $\delta(x)$ unknown

Figure: $y_e(x)$ (+noise) posterior plots making different assumptions

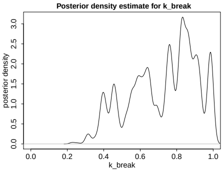
Granulation - $(k_{\text{coag}}, k_{\text{compact}}, k_{\text{break}}, k_{\text{pen}}, k_{\text{reac}})$-posteriors



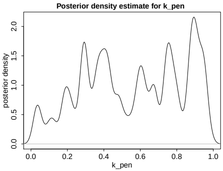
(a) k_{coag} posterior



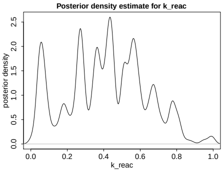
(b) k_{compact} posterior



(c) k_{break} posterior



(d) k_{pen} posterior



(e) k_{reac} posterior

Figure: Assuming $\tau_s^2 = 0$, $\tau_e^2, \delta(x)$ unknown