Uncertainty Quantification (UQ)
Bayesian approach

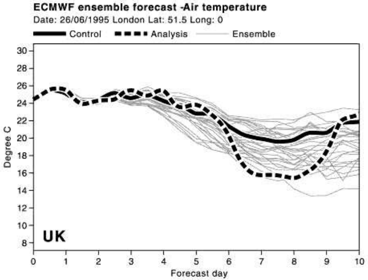
Ranaivo M. Razakanirina

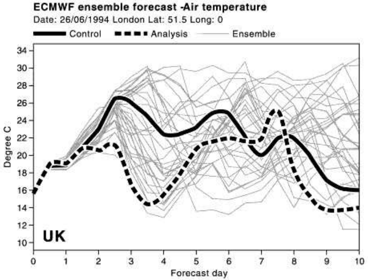
University of Geneva
Scientific and Parallel Computing Group, Switzerland

Thesis Exam

July 31st, 2014

Examples: Weather forecasting





Surface temperature in London

- Top: Relatively predictable period (June 26th, 1995)
- Bottom: Unpredictable period (June 26th, 1994)

- Forecast (Solid line)
- Verified (Dashed line)

Gaussian plume model (Clarke, 1979): Model the dispersion of radioactive material

Examples: Radioactive material dispersion

Gaussian plume model (Clarke, 1979): Model the dispersion of radioactive material

- Model inputs
  - Atmospheric conditions (wind direction, wind speed)
  - Source of radioactive material characteristics (source location, release weight, release duration)
- Lack of knowledge, uncertainty
  - The detailed input information unknown
  - The process is not observable
  - Simplified model assumption

Uncertainty vs margin vs error

"We might say that the length of a certain stick measures 20 centimetres plus or minus 1 centimetre, at the 95 percent confidence level. This result could be written:"

20 cm $\pm$ 1 cm, at a level of confidence of 95%.

Uncertainty is a quantification of the doubt about the measurement result.

From "A Beginner's Guide to Uncertainty of Measurement"
Stephanie Bell

Type of uncertainties
Kennedy and O'Hagan (2001)

- Parameter uncertainty: True values of model inputs are unknown
- Parametric variability: Additional conditions to model input are unspecified
- Model inadequacy: Model is biased and approximative, and does not describe perfectly the true process at the true values of input
  - Falling object model: air friction is not considered
  - Weather forecasting: does the chosen model remain consistent with all data?
- Algorithmic or numerical uncertainty:
  - $\frac{0.1}{0.2} = 0.5$ and $\frac{0.000001}{0.000002} = ?$

Type of uncertainties
Kennedy and O'Hagan (2001) - cont

- Epistemic uncertainty: Due to the lack of knowledge of the underlying process, system
- Aleatoric uncertainty: Each experience is different even with the same model inputs and parameters

Formal uncertainty model

- Consider the measurements $\mathbf{y}^e(x)$ obtained from true inputs $\mathbf{x}$

$$\mathbf{y}^e(\mathbf{x}) = \zeta(\mathbf{x}) + \varepsilon \quad (1)$$

  - ε: Measurements are subject to measurement errors
  - $\zeta(\mathbf{x})$: The true value of physical system
- Consider the outputs of the computer model $\mathbf{y}^m(x, \theta)$ obtained from true inputs $\mathbf{x}$

$$\zeta(\mathbf{x}) = \mathbf{y}^m(x, \theta) + \delta(\mathbf{x}) \quad (2)$$

  - θ: vector of unknown inputs and parameters
  - $\delta(\mathbf{x})$: Model discrepancy (differences between reality and computer model outputs)
- Uncertainty model with parameter uncertainty θ, model discrepancy and error

$$\mathbf{y}^e(\mathbf{x}) = \mathbf{y}^m(x, \theta) + \delta(\mathbf{x}) + \varepsilon \quad (3)$$

Formal uncertainty model

- Consider the measurements $\mathbf{y}^e(x)$ obtained from true inputs $\mathbf{x}$

$$\mathbf{y}^e(\mathbf{x}) = \zeta(\mathbf{x}) + \varepsilon \tag{1}$$

  - ε: Measurements are subject to measurement errors
  - $\zeta(\mathbf{x})$: The true value of physical system

- Consider the outputs of the computer model $\mathbf{y}^m(x, \theta)$ obtained from true inputs $\mathbf{x}$

$$\zeta(\mathbf{x}) = \mathbf{y}^m(x, \theta) + \delta(\mathbf{x}) \tag{2}$$

  - θ: vector of unknown inputs and parameters
  - $\delta(\mathbf{x})$: Model discrepancy (differences between reality and computer model outputs)

- Uncertainty model with parameter uncertainty θ, model discrepancy and error

$$\mathbf{y}^e(\mathbf{x}) = \mathbf{y}^m(x, \theta) + \delta(\mathbf{x}) + \varepsilon \tag{3}$$

Formal uncertainty model

- Consider the measurements $\mathbf{y}^e(x)$ obtained from true inputs $\mathbf{x}$

$$\mathbf{y}^e(\mathbf{x}) = \zeta(\mathbf{x}) + \varepsilon \quad (1)$$

  - ε: Measurements are subject to measurement errors
  - $\zeta(\mathbf{x})$: The true value of physical system
- Consider the outputs of the computer model $\mathbf{y}^m(x, \theta)$ obtained from true inputs $\mathbf{x}$

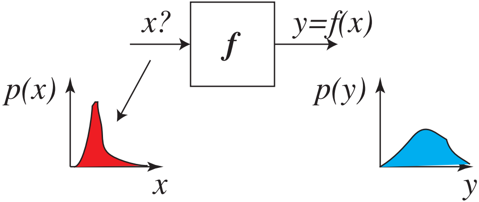
$$\zeta(\mathbf{x}) = \mathbf{y}^m(x, \theta) + \delta(\mathbf{x}) \quad (2)$$

  - θ: vector of unknown inputs and parameters
  - $\delta(\mathbf{x})$: Model discrepancy (differences between reality and computer model outputs)
- Uncertainty model with parameter uncertainty θ, model discrepancy and error

$$\mathbf{y}^e(\mathbf{x}) = \mathbf{y}^m(x, \theta) + \delta(\mathbf{x}) + \varepsilon \quad (3)$$

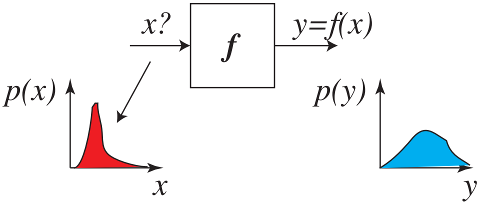
Method to quantify uncertainty

- Forward uncertainty propagation



  - Predict outputs distributions from uncertain inputs
  - Model uncertain inputs using distribution
  - Monte Carlo method (based on frequentist approach)
- Inverse uncertainty quantification: Bayesian inverse problem
  - Assume a prior distribution of inputs
  - Use Bayesian inference to build the posterior distribution of inputs

- Forward uncertainty propagation



  - Predict outputs distributions from uncertain inputs
  - Model uncertain inputs using distribution
  - Monte Carlo method (based on frequentist approach)

- Inverse uncertainty quantification: Bayesian inverse problem
  - Assume a prior distribution of inputs
  - Use Bayesian inference to build the posterior distribution of inputs

Bayesian inverse problem

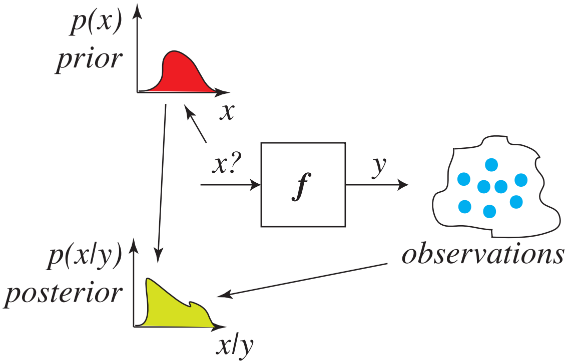
Bayes' rule (x and y are random variables).

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \tag{4}$$

Main idea: Find the solution of the inverse problem $p(x|y)$. The best density matching the observations y.

- $p(x)$: *prior* probability. A priori knowledge of the uncertain inputs x
- $p(y|x)$: *likelihood*. Our observations. Distribution of data given the input.
- $p(y) = \int p(y|x)p(x)$: marginal distribution of data.

Bayesian inverse problem: Illustration



- f: system model
- y: set of n observations
- $p(x)$: a priori knowledge of x
- $p(x|y)$: posterior distribution of x considering observations y

Bayesian inverse problem
Estimate mean of normally distributed data

- Collect n normally distributed observations y_i $(i = 1, \dots, n)$ with unknown mean μ and known variance σ^2. $y_i \in \mathbb{R}$. y_i are i.i.d. Therefore,

$$p(y_i, \sigma^2 | \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(y_i - \mu)^2}{2\sigma^2} \right\} \tag{5}$$

The likelihood of these n observations is:

$$L(y, \sigma^2 | \mu) = \prod_{i=1}^{n} p(y_i, \sigma^2 | \mu) \tag{6}$$

- Consider that the prior distribution of μ is $p(\mu) = \mathcal{N}(\lambda, \tau^2)$
- The posterior distribution of μ is (from Bayes' rule):

$$p(\mu | y, \sigma^2) \propto L(y, \sigma^2 | \mu) p(\mu) \tag{7}$$

Gaussian process
Definition

- 1d Gaussian: Completely defined by mean μ and variance σ^2

$$p(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \tag{8}$$

- Multi-variate Gaussian: Completely defined by mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$

$$p(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{2\pi|\boldsymbol{\Sigma}|}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \tag{9}$$

- Gaussian process (GP): Collection of random variables $f(\mathbf{x}) = (f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots, f(\mathbf{x}_n))$ that have joint Gaussian distribution. A GP is completely defined by its mean function $m(\mathbf{x})$ and its covariance function or kernel $k(\mathbf{x},\mathbf{x}')$

- 1d Gaussian: Completely defined by mean μ and variance σ^2

$$p(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \tag{8}$$

- Multi-variate Gaussian: Completely defined by mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$

$$p(\mathbf{x}|\boldsymbol{\mu},\mathbf{\Sigma}) = \frac{1}{\sqrt{2\pi|\mathbf{\Sigma}|}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \tag{9}$$

- Gaussian process (GP): Collection of random variables $f(\mathbf{x}) = (f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_n))$ that have joint Gaussian distribution. A GP is completely defined by its mean function $m(\mathbf{x})$ and its covariance function or kernel $k(\mathbf{x},\mathbf{x}')$

- Mean function $m(\mathbf{x})$ ($\mathbf{x} \in \mathbb{R}^d$, d is the number of inputs)
  - Any function $\mathbb{R}^d \to \mathbb{R}$
  - Habitually, $m(\mathbf{x}) = 0$: No prior knowledge
- Kernel function $k(\mathbf{x}, \mathbf{x}')$
  - Function $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$
  - Building the covariance matrix with kernel function

$$\mathbf{\Sigma} = \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & k(\mathbf{x}_1, \mathbf{x}_2) & \dots & k(\mathbf{x}_1, \mathbf{x}_n) \\ k(\mathbf{x}_2, \mathbf{x}_1) & k(\mathbf{x}_2, \mathbf{x}_2) & \dots & k(\mathbf{x}_2, \mathbf{x}_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(\mathbf{x}_n, \mathbf{x}_1) & k(\mathbf{x}_n, \mathbf{x}_2) & \dots & k(\mathbf{x}_n, \mathbf{x}_n) \end{pmatrix} \tag{10}$$

Kernel function $k(\mathbf{x}, \mathbf{x}')$

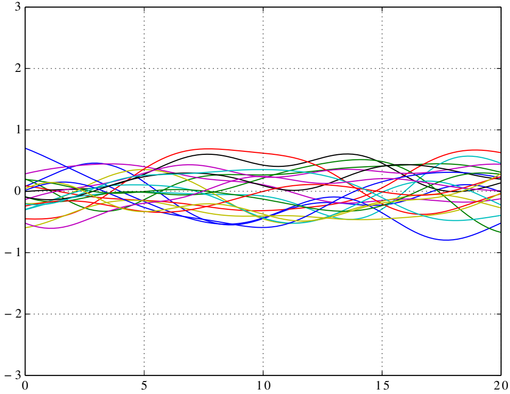
- Measure of correlations:
  - If $\mathbf{x}$ and $\mathbf{x}'$ are correlated, $k(\mathbf{x}, \mathbf{x}') \rightarrow 1$
  - If $\mathbf{x}$ and $\mathbf{x}'$ are not correlated, $k(\mathbf{x}, \mathbf{x}') \rightarrow 0$
- Example of kernel for d=1 (Squared exponential kernel)

$$k(x, x') = \sigma^2 \exp\left\{-\frac{(x - x')^2}{2\ell^2}\right\} \tag{11}$$

- Hyperparameters
  - σ^2: output variance
  - ℓ: length scale

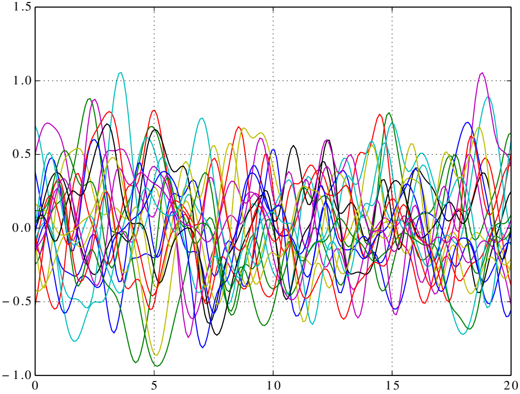
Gaussian process
Kernel function $k(\mathbf{x}, \mathbf{x}')$

- Impact of hyperparameters: $\sigma^2 = 0.1$, $\ell = 3$

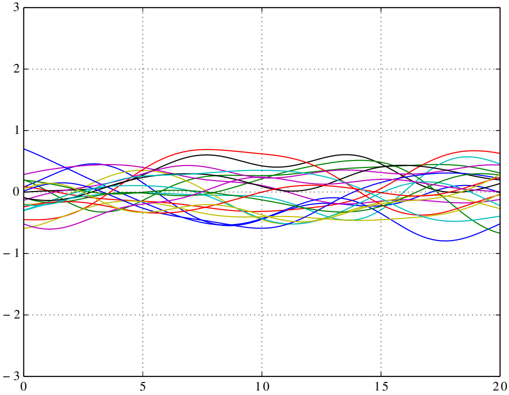


Scientific
& Parallel Computing
Group

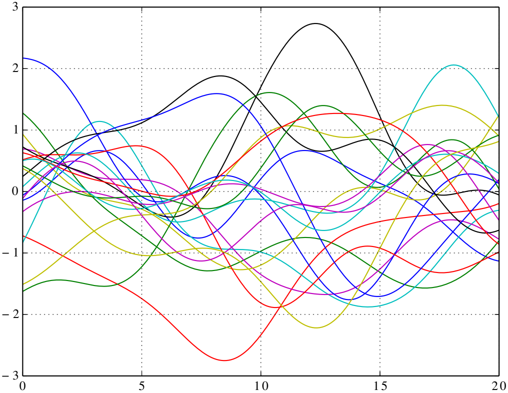
- Impact of hyperparameters: $\sigma^2 = 0.1$, $\ell = 0.5$



- Impact of hyperparameters: $\sigma^2 = 0.1$, $\ell = 3$



- Impact of hyperparameters: $\sigma^2 = 1$, $\ell = 3$



Gaussian process
Conditional Gaussian distribution

Important conditional property

- $\mathbf{y} = \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n \sim GP(\mu, C(\mathbf{x}, \mathbf{x}))$
- $\mathbf{y}_* \sim GP(\mu_*, C(\mathbf{x}_*, \mathbf{x}_*))$

Therefore

$$(\mathbf{y}, \mathbf{y}_*) \sim GP\left(\begin{pmatrix} \mu \\ \mu_* \end{pmatrix}, \begin{pmatrix} C(\mathbf{x}, \mathbf{x}) & C(\mathbf{x}, \mathbf{x}_*) \\ C(\mathbf{x}_*, \mathbf{x}) & C(\mathbf{x}_*, \mathbf{x}_*) \end{pmatrix}\right) \tag{12}$$

And the conditional property (posterior) is

$$\mathbf{y}_* | \mathbf{y} \sim GP(\mathbf{M}, \mathbf{\Sigma}) \tag{13}$$

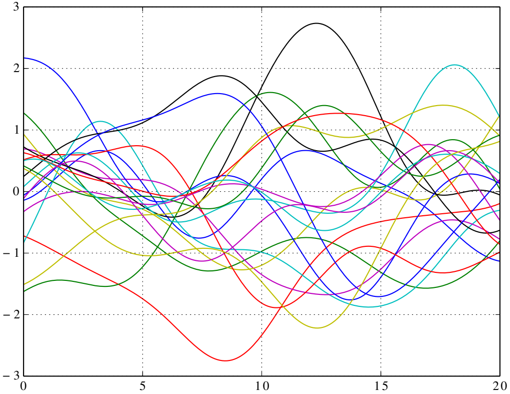
where

$\mathbf{M} = \mu + C(\mathbf{x}, \mathbf{x}_*) C(\mathbf{x}_*, \mathbf{x}_*)^{-1} (\mathbf{y} - \mu_*)$

$\mathbf{\Sigma} = C(\mathbf{x}, \mathbf{x}) - C(\mathbf{x}, \mathbf{x}_*) C(\mathbf{x}_*, \mathbf{x}_*)^{-1} C(\mathbf{x}, \mathbf{x}_*)^T$

- Kernel: Squared exponential. $\sigma^2 = 1$ and $\ell = 3$
- $\mathbf{y}_*$: 20 samples from $xsin(x)$, $x \in [0, 20]$

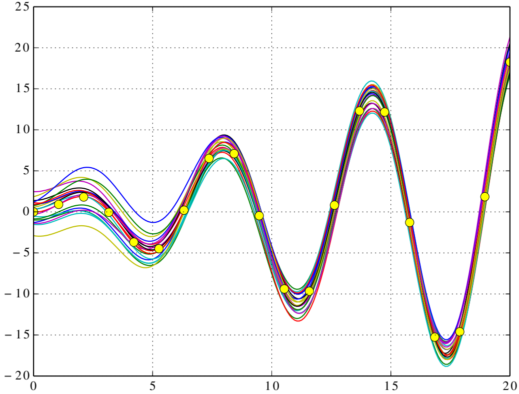
Prior GP



Gaussian process
Example

- Kernel: Squared exponential. $\sigma^2 = 1$ and $\ell = 3$
- $\mathbf{y}_*$: 20 samples from $x sin(x)$, $x \in [0, 20]$

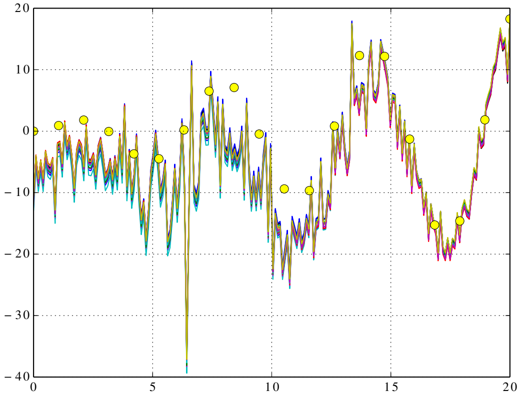
Posterior $\mathbf{y}_* | \mathbf{y}$



Gaussian process
Example

- Kernel: Squared exponential. $\sigma^2 = 1$ and $\ell = 3$
- $\mathbf{y}_*$: 20 samples from $xsin(x)$, $x \in [0, 20]$

Impacts of hyperparameters: Posterior $\mathbf{y}_*|\mathbf{y}$, $\sigma^2 = 0.1$ and $\ell = 7$



Scientific & Parallel Computing Group

Modular Bayesian approach of UQ
Kennedy and O'Hagan (2001)

$$\mathbf{y}^e(\mathbf{x}) = \mathbf{y}^m(\mathbf{x}, \theta) + \delta(\mathbf{x}) + \varepsilon \qquad (14)$$

Goal: Find the posterior distribution of θ (calibration parameter) and $\delta(\mathbf{x})$ (model discrepancy) using GP and Bayes' rule
$\rightarrow$ Modular Bayesian approach proposed by Kennedy and O'Hagan

- Module 1: Model computer model $\mathbf{y}^m$ using GP prior
- Module 2: Model discrepancy $\delta(\mathbf{x})$ using GP prior
- Module 3: Compute posterior distribution of calibration parameters

Modular Bayesian approach of UQ
Module 1: Prior model of computer model

$$y^m(x,\theta) \sim GP(h_m(x,\theta)^T\beta_m, \sigma_m^2 R_m((x,\theta),(x',\theta'))) \tag{15}$$

where (chosen kernel)

$$R_m((x,\theta),(x',\theta')) = \exp\left\{-\sum_{k=1}^{d}\omega_k^m(x_k - x_k')^2\right\}\exp\left\{-\sum_{k=1}^{r}\omega_r^m(\theta_k - \theta_k')^2\right\} \tag{16}$$

- $x \in \mathbb{R}^d$ and $\theta \in \mathbb{R}^r$
- $h_m(x,\theta)^T$ is a predefined function
- Get hyperparameters $\phi_m = (\beta_m, \sigma_m, \omega_m)$ by maximizing log likelihood.
  If $y \sim GP(0, K)$ and θ the set of its hyperparameters:

$$\log(p(y|X,\theta)) = -\frac{1}{2}y^T K y^{-1} - \frac{1}{2}\log|K| - \frac{n}{2}\log(2\pi) \tag{17}$$

Modular Bayesian approach of UQ
Module 2: Prior model of model discrepancy

$$\delta(x) \sim GP(h_\delta(x)^T \beta_\delta, \sigma_\delta^2 R_\delta(x)) \tag{18}$$

where (chosen kernel)

$$R_\delta(x, x') = \exp\left\{ -\sum_{k=1}^{d} \omega_k^\delta (x_k - x_k')^2 \right\} \tag{19}$$

- $h_\delta(x)^T$ is a predefined function
- Hyperparameters $\phi_\delta = (\beta_\delta, \sigma_\delta, \omega_\delta)$

Prior model of experimental response with $\varepsilon \sim N(0, \sigma^2)$

$$y^e(x, \theta) \sim GP(m_e(x, \theta), V_e((x, \theta), (x', \theta'))) \tag{20}$$

where from i.i.d and property of GP addition

- $m_e(x, \theta) = h_m(x, \theta)^T \beta_m + h_\delta(x)^T \beta_\delta$
- $V_e((x, \theta), (x', \theta')) = \sigma_m^2 R_m((x, \theta), (x', \theta')) + \sigma_\delta^2 R_\delta(x) + \sigma^2$

Modular Bayesian approach of UQ
Module 3: Posterior distribution of calibration parameters

Use Bayes' rule

$$p(\theta|y^e, y^m, \phi_e, \phi_m) \propto L(y^e, y^m, \phi_e, \phi_m|\theta)p(\theta) \tag{21}$$

with prior distribution of θ

$$p(\theta) = N(\mu_\theta, \sigma_\theta^2) \tag{22}$$

Summary

- Key roles of uncertainty quantification
- Uncertainty related to probability density
- Two main approaches:
  - Forward propagation: Monte Carlo method
  - Bayesian approach: Bayes' rule and Gaussian Process
- Key issues:
  - Selection of kernel function
  - Computing hyperparameters

References

- M. C. Kennedy and A. O'Hagan, "Bayesian Calibration of Computer Models," Journal of the Royal Statistical Society. Series B (Statistical Methodology), vol. 63, no. 3, pp. 425–464, Jan. 2001.
- C. E. Rasmussen, Gaussian Processes for Machine Learning. Cambridge, Mass: The MIT Press, 2005.
- H.-P. Wan, Z. Mao, M. D. Todd, and W.-X. Ren, "Analytical uncertainty quantification for modal frequencies with structural parameter uncertainty using a Gaussian process metamodel," Engineering Structures, vol. 75, pp. 577–589, Sep. 2014.
- J.Brynjarsdottir and A. O'Hagan, "Learning about Physical Parameters: The importance of model discrepancy" To appear in Inverse Problems. 2014.
- M. J. Bayarri, J. O. Berger, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish, C.-H. Lin, and J. Tu, "A Framework for Validation of Computer Models," Technometrics, vol. 49, no. 2, pp. 138–154, May 2007.

THANK YOU
Questions?