UQ6 Ch 9 1

Uncertainty propagation in models

Uncertainty bounds for QoI: propagate input uncertainties through model while accounting for measurement errors

Direct evaluation for linearly parameterized models

Sampling methods, e.g. MC, for nonlinear problems, best method for large parameter dimension p

Perturbation methods, truncated Taylor series of QoI at parameter mean

Linear models

$$Y_i = Q_1 + \sum_{j=2}^{p} x_{ij} Q_j + \varepsilon_i \quad , \quad i = 1 \cdots n$$

$$Y = \underbrace{XQ}_{f(Q)\text{ model response}} + \varepsilon$$

$$E(f_i(Q)) = \sum_{j=1}^{p} x_{ij} \bar{q}_j \quad , \quad x_{i1} = 1$$

$$Var(f_i(Q)) = Var\left(\sum_{j=1}^{p} x_{ij} Q_j\right) =$$

$$= \sum_{j=1}^{p} x_{ij}^2 Var(Q_j) + 2 \sum_{j<k} x_{ij} x_{ik} Cov(X_j, X_k)$$

Given density for $Q \Rightarrow$ propagation to Y explicitly

$$E(Y_i) = E(f_i(Q))$$
$$Var(Y_i) = Var(f_i(Q)) + Var(\varepsilon_i)$$

model response and errors are mutually independent

Q_i are mutually independent and $N(\bar{q}_i, \sigma_i^2)$
$V = diag(\sigma_1^2 \cdots \sigma_p^2)$

$$f(Q) \sim N(X\bar{q}, X^T V X)$$

Sampling methods

Input uncertainties determined e.g. by Bayesian techniques

Random sampling of measurement errors and joint input distributions to construct ensemble of responses for statistics
Disadvantage: slow convergence, MC has $O(1/\sqrt{M})$ convergence for M realizations

Perturbation methods

For large or complex nonlinearly parametrized methods, sampling-based uncertainty propagation is infeasible
Instead: Taylor series expansion about mean $\bar{q}$

$$Q = \bar{q} + \delta Q$$

$$f(Q) = f(\bar{q}) + \sum_{i_1=1}^{p} \frac{\partial f}{\partial Q_{i_1}}\Big|_{\bar{q}} \delta Q_{i_1} + \frac{1}{2}\sum_{i_1,i_2=1}^{p} \frac{\partial^2 f}{\partial Q_{i_1}\partial Q_{i_2}}\Big|_{\bar{q}} \delta Q_{i_1}\delta Q_{i_2} + \dots$$

linear expansion

$$f(Q) = \bar{y} + \sum_{i=1}^{p} s_i \delta Q_i \quad , \quad \bar{y} = f(\bar{q})$$
$$s_i = \frac{\partial f}{\partial Q_i}(\bar{q})$$

Q has joint pdf $\rho_Q(q)$

$$E(Q_i) = \bar{q}_i \quad , \quad \delta q_i = q_i - \bar{q}_i$$

$$Var(Q_i) = \int_{\mathbb{R}^p} \delta q_i^2 \, \rho_Q(q) \, dq$$

$$Cov(Q_i, Q_j) = \int_{\mathbb{R}^p} \delta q_i \, \delta q_j \, \rho_Q(q) \, dq$$

$$\Rightarrow E(f(Q)) = \bar{y} \underbrace{\int_{\mathbb{R}^p} \rho_Q(q) \, dq}_{1} + \sum_i s_i \underbrace{\int_{\mathbb{R}^p} \delta q_i \, \rho_Q(q) \, dq}_{0} = \bar{y}$$

$$Var(f(Q)) = E\left((f(Q) - \bar{y})^2\right)$$

$$= \int_{\mathbb{R}^p} \sum_i s_i \delta q_i \sum_j s_j \delta q_j \, \rho_Q(q) \, dq$$

$$= \sum_i \sum_j s_i s_j Cov(Q_i, Q_j) = \sum_i s_i^2 Var(Q_i) + \sum_i \sum_{j \neq i} s_i s_j Cov(Q_i, Q_j)$$

$$S^T = (s_1 \cdots s_p) \Rightarrow Var(f(Q)) = S^T V S$$
$$V_{ij} = Cov(Q_i, Q_j)$$

Q_i, $i = 1 \cdots p$, mutually independent, $Q_i \sim N(\bar{q}_i, \sigma_i^2)$
$V = diag(\sigma_1^2, \cdots \sigma_p^2) \Rightarrow f(Q) \sim N(\bar{y}, S^T V S)$

Exercise 9.1

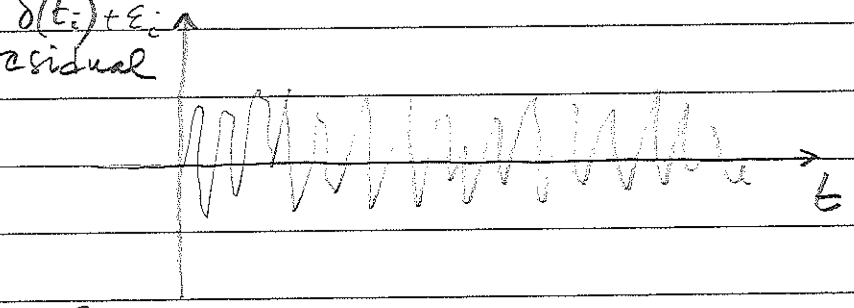
Prediction in the Presence of Model Discrepancy (§12)

$$Y_i = f(t_i, q) + \delta(t_i) + \varepsilon_i$$

Earlier: sum of model and measurement errors
ε_i iid $\sim N(0, \sigma^2)$
Here: combined errors: biased, correlated

Neglecting discrepancy terms $\delta(t_i)$ may hamper model calibration and prediction in different ways

1. Validity of estimates q may diminish since the optimized q attempts to compensate for missing physics
2. Correlated errors contrary to hypotheses for likelihoods $\Rightarrow$ poor estimated parameter distributions
3. Predictions with other inputs than those used for calibration may be inaccurate

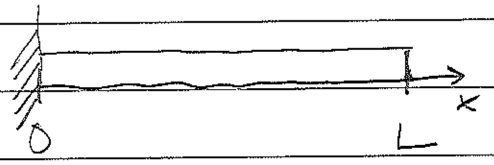


periodic error
not iid, unbiased

Solutions: Add more physics, Quantify model error

Incorporate missing physical mechanisms

Example
$$\frac{d^2 T_s}{dx^2} = \frac{2(a+b)}{ab}\frac{h}{k}\left(T_s - T_{amb}\right)$$



$$\frac{dT_s}{dx}(0) = \frac{\Phi}{k}$$

$$\frac{dT_s}{dx}(L) = \frac{h}{k}\left[T_{amb} - T_s(L)\right]$$

T_{amb}: ambient temp
h: convective heat transfer coefficient
k: thermal conductivity
Φ: source heat flux
a, b: cross section of rod
Φ, h unknown

Change boundary condition $\frac{dT_s}{dx}(0) = \frac{\eta}{k}\left(T_s(0) - T_{source}\right)$
parameters $q = [T_{source}, h, \eta]$
$\Rightarrow$ better model fit to data

To reduce model discrepancies: add more physical mechanisms

Not always possible! Why not include all from the beginning?

Techniques to quantify model errors

$$Y_i = f(x_i, q) + \delta(x_i) + \varepsilon_i \, , \quad x \in \mathbb{R}$$

spatial statistical model

(1) Polynomial model : $\delta(x_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$

Estimate hyperparameters $\beta_0, \beta_1, \beta_2$ together with model parameters q using frequentist or Bayesian techniques

(2). Gaussian process model (Kriging)

$$\delta(x, \mu, \sigma^2, \theta) = [\delta(x_1, \mu, \sigma^2, \theta), \cdots \delta(x_n, \mu, \sigma^2, \theta)]$$

$$\delta \sim N(\mu I, \sigma^2 R) \, , \quad R_{ij} = \exp(-\theta |x_i - x_j|)$$

hyperparameters σ^2, μ, θ
augmented parameter set $q_{aug} = [q, \sigma^2, \mu, \theta]$

(1) and (2) general techniques, no knowledge about background of model is necessary

Problems: 1. New parameters not identifiable
Cure: * better prior information in Bayesian approach
* Solve for q first, then for δ-parameters
iterate until convergence

2. Extrapolation with model outside calibration domain

$$Y_i = f(x_i, q) + \delta(x_i, q_{dis}) + \varepsilon_i$$

δ ensures that q is estimated in a statistically consistent manner in the calibration domain, may be inaccurate outside domain without restrictions on δ or prior information.
Combined model $f(x_i, q) + \delta(x_i, q_{dis})$ is accurate inside calibration domain but has minimal predictive capability outside this domain ("overfitting")

<u>Surrogate models</u> (§13)

Construct representations quantifying primary features of the high-fidelity model while being computationally efficient for Bayesian model calibration, uncertainty quantification, design, optimization ....

<u>Regression or interpolation-based models</u>

Data-fit models, response surface models, emulators, meta-models, approximation models

Model $\quad y = f(q), \quad q \in \Gamma \subset \mathbb{R}^p$