A Probabilistic Characterization of Simulation Model Uncertainties

Victor Ontiveros
Mohammad Modarres

Center for Risk and Reliability
University of Maryland

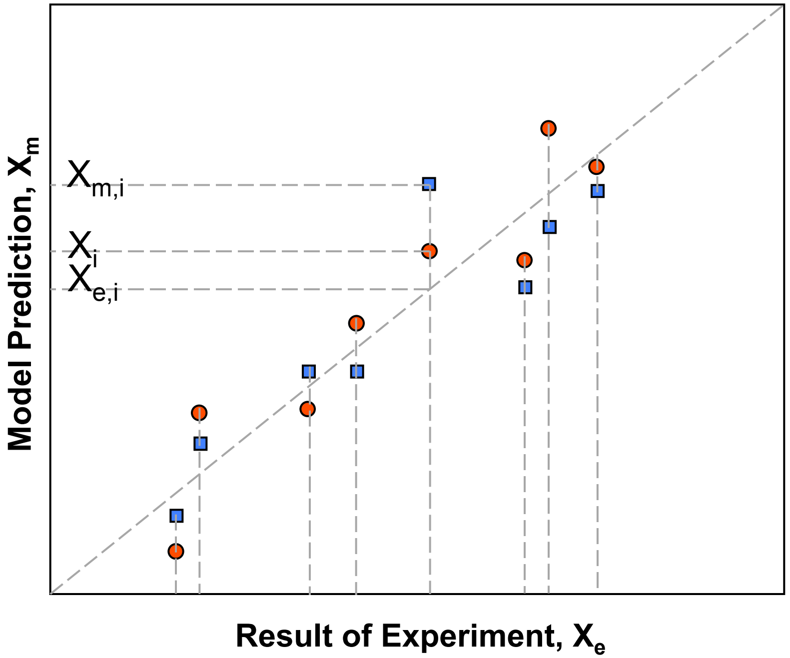
Introduction

- There is uncertainty in model predictions as well as uncertainty in experiments
- The uncertainty in the experiment and uncertainty in the model predictions are considered independent
- Any comparison between the results of experiment and model predictions may be used to estimate uncertainty of the real value of interest
- The model uncertainty is the only uncertainty that should be considered when the distribution of real value given a model prediction is estimated.
- Three different model errors are presented: the additive error, percentage error model and multiplicative error model
- The multiplicative error model shows a better agreement

Additive Error Model: Assumptions

- The additive error of the results of experiment compared to the real values of interest is a normally distributed random number with given mean and standard deviation
- The additive error of model predictions compared to the real values of interest is also a normally distributed random number
- Error between model and experiment can now be assessed

Additive Error Model



$$X_i = X_{e,i} + E_{e,i} \quad ; \quad E_e \sim N(b_e, s_e)$$

$$X_i = X_{m,i} + E_{m,i} \quad ; \quad E_m \sim N(b_m, s_m)$$

where:

X : Real Quantity

X_e : Result of experiment

X_m : Model prediction

b_e, σ_e : Mean and SD of experimental additive error

b_m, σ_m : Mean and SD of model additive error

$$E_{me} \sim N\left(b_m - b_e, \sqrt{s_m^2 + s_e^2}\right)$$

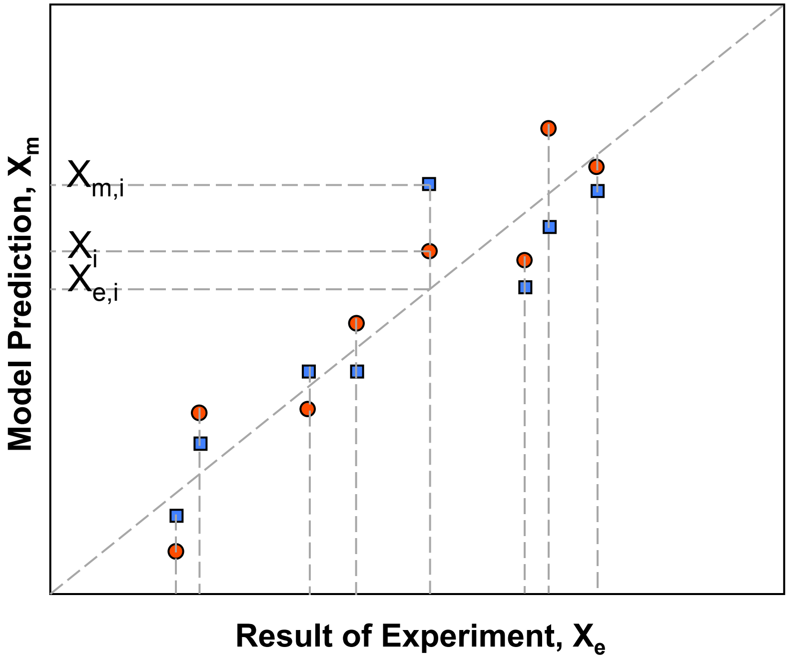
Pitfalls

- Error plotted between model prediction and experimental measurements must be identically distributed which is not always true as they may increase at higher ranges.
- The introduced additive error can be negative, zero or positive
- The E_{me} which is the additive error between model prediction and experiment can be negative, positive or zero. This limits the choices of likelihood function for this random variable to normal distribution.
- When the data is widely scattered the normal distribution assumption results in negative lower bounds with no meaningful physical interpretation

Percentage Error Model: Assumptions

- The percentage error of the results of experiment compared to the real values of interest is a normally distributed random number with given mean and standard deviation
- The percentage error of model predictions compared to the real values of interest is a normally distributed random number
- The percentage error of model predictions compared to the results of experiment is a function of the two random variables introduced earlier. The distribution of this random number will be used to represent the likelihood of data

Percentage Error Model



$$\frac{X_i - X_{e,i}}{X_{e,i}} = E_{e,i} \quad ; \quad E_e \sim N(b_e, \sigma_e)$$

$$\frac{X_i - X_{m,i}}{X_{m,i}} = E_{m,i} \quad ; \quad E_m \sim N(b_m, \sigma_m)$$

$$\frac{X_{e,i} - X_{m,i}}{X_{m,i}} = E_{em,i}, (1), (2) \Rightarrow E_{em} = \frac{E_m - E_e}{E_e + 1}$$

where:

X : Real Quantity

X_e : Result of experiment

X_m : Model prediction

b_e, σ_e : Mean and SD of experimental percentage error

b_m, σ_m : Mean and SD of model percentage error

b_{em}, σ_{em} : Mean and SD percentage errorof Experiment compared to model

$$\text{Independency of } E_m, E_e \} \Rightarrow E_{em} \sim N\left(\frac{b_m - b_e}{1 + b_e} + \frac{\sigma_e^2 (1 + b_m)}{(1 + b_e)^3}, \sqrt{\frac{\sigma_m^2}{(1 + b_e)^2} + \frac{\sigma_e^2 (1 + b_m)^2}{(1 + b_e)^4}}\right)$$

Approximation!

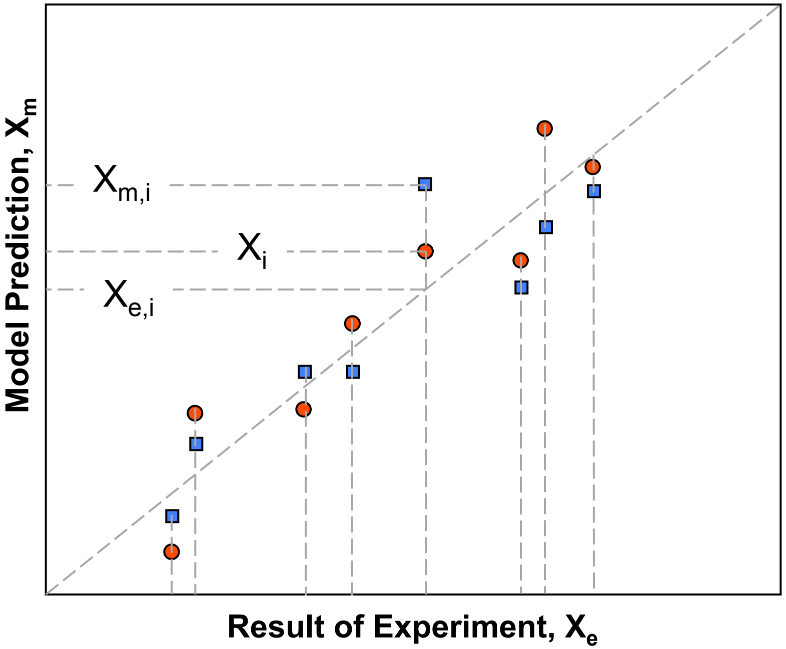
Pitfalls

- The introduced percentage error can be negative, zero or positive. This basically forces the normal distribution assumption for percentage errors, E_m & E_e.
- The E_{em} which is the percentage error between model prediction and experiment can be negative, positive or zero. This limits the choices of likelihood function for this random variable to normal distribution.
- The exact distribution of E_{em} can not be analytically derived
- When the data is widely scattered the normal distribution assumption results in negative lower bounds with no meaningful physical interpretation

Multiplicative Error: Assumptions

- The model prediction, result of experiment and real value of interest have the same sign (all positive or all negative)
- The ratio of real value and experimental results is a random variable with lognormal distribution for which confidence bounds are known
- The ratio of real value and model prediction is a random variable with lognormal distribution with parameters to be determined
- The ratio of model predictions and results of experiment is a function of the two random variables introduced earlier. The distribution of this random variable is lognormal and will be used to represent the likelihood of data
- Having the above assumptions the distribution of real quantity of interest given a model prediction will be a lognormal distribution

Multiplicative Error Model



$$\frac{X_i}{X_{e,i}} = F_{e,i} \quad ; \quad F_e \sim LN(b_e, \sigma_e)$$

$$\frac{X_i}{X_{m,i}} = F_{m,i} \quad ; \quad F_m \sim LN(b_m, \sigma_m)$$

where :

X : Real Quantity

X_e : Result of experiment

X_m : Model prediction

F_e : The error factor for experimental data

F_m : The error factor for model predictions

b_e, σ_e : Mean and SD of experimental error factor

b_m, σ_m : Mean and SD of model error factor

$$\left.\begin{array}{l} F_{e,i} X_{e,i} = F_{m,i} X_{m,i} \\ \dfrac{X_{e,i}}{X_{m,i}} = \dfrac{F_{m,i}}{F_{e,i}} = F_{t,i} \\ \text{Independency of } F_m, F_e \end{array}\right\} \Rightarrow F_t \sim LN\left(b_m - b_e, \sqrt{\sigma_m^2 + \sigma_e^2}\right)$$

Multiplicative Error: Bayesian Posterior

$$f(b_m,\sigma_m \mid X_{e,i},X_{m,i},b_e,\sigma_e)=\frac{f_0(b_m,\sigma_m)\times L(X_{e,i},X_{m,i},b_e,\sigma_e \mid b_m,\sigma_m)}{\int\limits_{\sigma_m}\int\limits_{b_m} f_0(b_m,\sigma_m)\times L(X_{e,i},X_{m,i},b_e,\sigma_e \mid b_m,\sigma_m)db_m d\sigma_m} \qquad (3)$$

$where:$

$$L(X_{e,i},X_{m,i},b_e,\sigma_e \mid b_m,\sigma_m)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\left(\frac{X_{e,i}}{X_{m,i}}\right)\sqrt{\sigma_m^2+\sigma_e^2}}e^{-\frac{1}{2}\times\frac{\left[\ln\left(\frac{X_{e,i}}{X_{m,i}}\right)-(b_m-b_e)\right]^2}{\sigma_m^2+\sigma_e^2}}$$

$f_0(b_m,\sigma_m)$: Prior Joint Distirbution of Parameters

$f(b_m,\sigma_m \mid X_{e,i},X_{m,i},b_e,\sigma_e)$: Posterior Joint Distirbution of Parameters

Given a model prediction such as X_m the distribution of X will be estimated as following:

$$\left.\begin{array}{l} X_m \text{ given as model prediction} \\ F_m \sim LN(b_m,\sigma_m) \\ X = F_m X_m \end{array}\right\} \Rightarrow X \sim LN\left(\ln(X_m)+b_m,\sigma_m\right)$$

Example – FIVE Radiant Heat Flux
Bayesian Approach

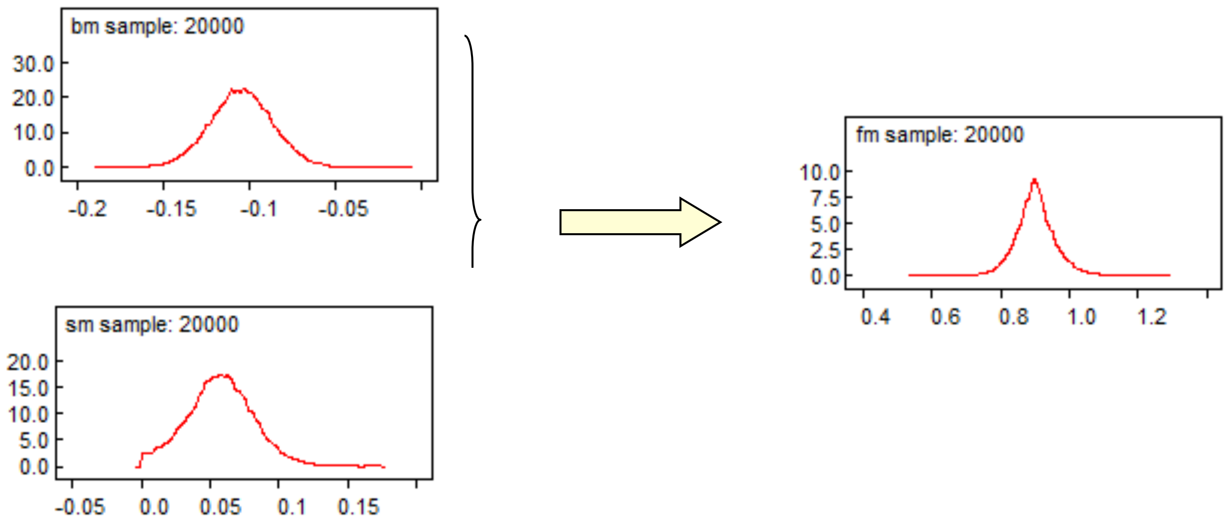
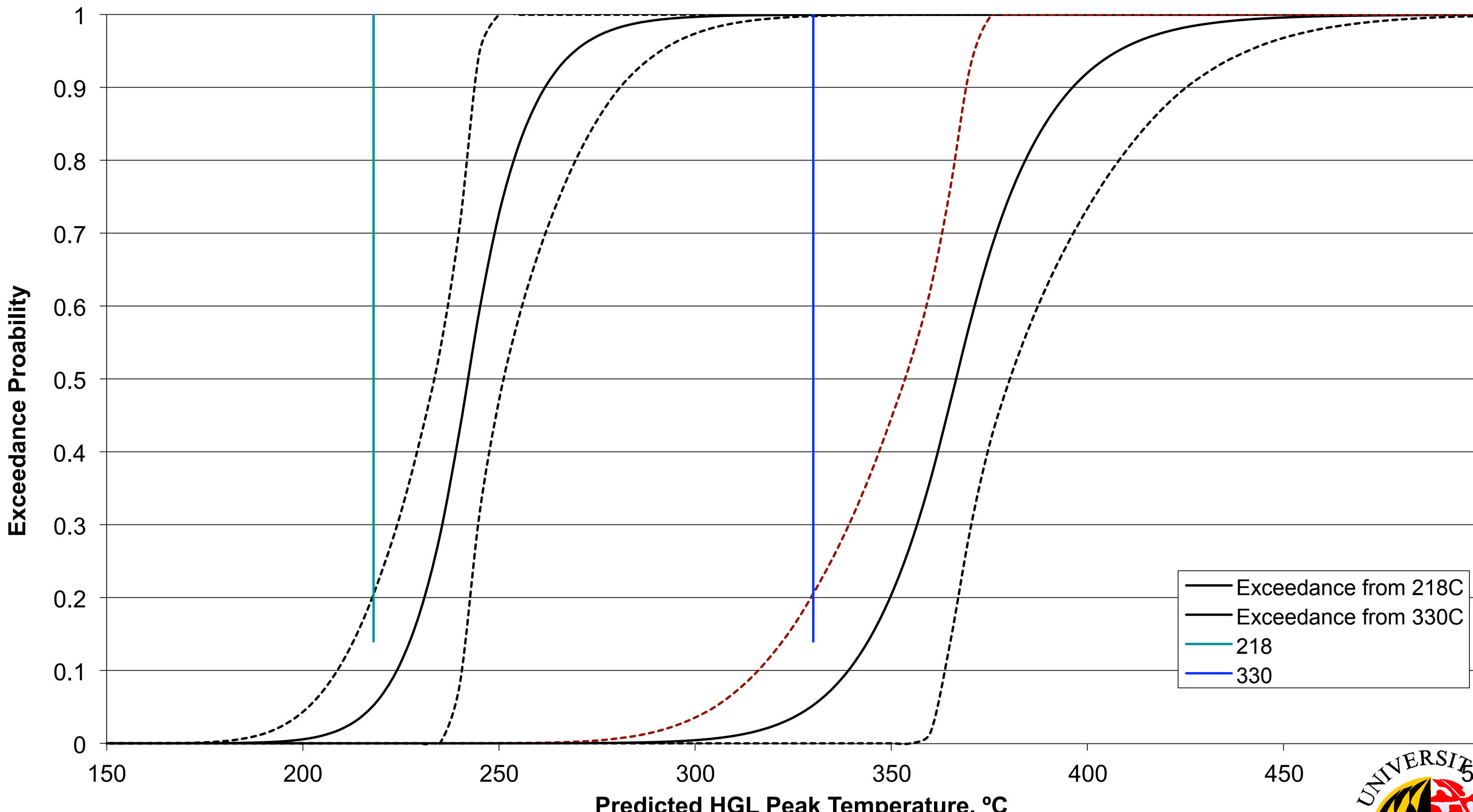


Table I. Summary Statistics of Parameters

Parameter	Mean	STD	2.5%	97.5%
b_m	-0.1052	1.87E-02	-0.1422	-6.78E-02
s_m	5.72E-02	2.40E-02	8.62E-03	0.1049
F_m	0.902	5.90E-02	0.7885	1.03

Probability of Exceedance: HGL Temperature



Conclusions

- The error estimated by comparing experiments and model predictions is not the uncertainty of the model. It is rather a combination of experimental and model uncertainties.
- The distribution of the real value of interest given a model prediction depends only on the uncertainty of the model.
- The Bayesian framework allows different weights and expert judgments to be later considered when dealing with non-homogeneous population of experiments or model predictions
- Posteriors from Bayesian analysis can be used as prior to be updated by new data points when become available
- Multiplicative error model provides good results