Available online at www.eccomasproceedia.org
Eccomas Proceedia UNCECOMP (2017) 154-164

ECCOMAS
Proceedia

UNCECOMP 2017
2nd ECCOMAS Thematic Conference on
Uncertainty Quantification in Computational Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou (eds.)
Rhodes Island, Greece, 15–17 June 2017

SAMPLING SCHEMES FOR HISTORY MATCHING USING SUBSET SIMULATION

Z. Gong[1], F. A. DiazDelaO[1], and M. Beer[2]

[1]Institute for Risk and Uncertainty, School of Engineering
University of Liverpool, Liverpool, United Kingdom
e-mail: {gong7, f.a.diazdelao}@liverpool.ac.uk

[2] Institute for Computer Science in Civil Engineering
Leibniz University Hannover, Hannover, Germany
e-mail: beer@bauinf.uni-hannover.de

Keywords: Subset Simulation, Rare Event Simulation, History Matching, Non-implausibility, Bayesian Emulation.

Abstract. *History Matching (HM) is a form of model calibration suitable for high-dimensional and computationally expensive numerical models. It sequentially cuts down the input space to find the non-implausible domain that provides a reasonable match between the output and experimental data. The non-implausible domain can be orders of magnitude smaller than the original input space and it can have a complex topology. This leads to one of the most challenging open problems in implementing HM, namely, the efficient generation of samples in the non-implausible set. Previous work has shown that Subset Simulation can be used to solve this problem. Unlike Direct Monte Carlo, Subset Simulation progressively decomposes a rare event (here is the non-implausible set), which has very small failure probabilities, into sequential less rare nested events. The original Subset Simulation uses a Modified Metropolis algorithm to generate the conditional samples that belong to intermediate less rare failure events. Generating samples moving forwards to the target space is the heart for Subset Simulation. This work considers different sampling strategies to generate samples and compares their performance in the context of expensive model calibration. A numerical example is provided to show the potential of HM using different Subset Simulation sampling schemes.*

Peer-review under responsibility of the organizing committee of UNCECOMP 2017.
doi: 10.7712/120217.5359.16948

1 INTRODUCTION

Numerical models, also known as simulators, are universally designed and employed to represent and study complex real-world systems. In order for a simulator to be reliable, it requires calibration using experimental data. Unfortunately, high-dimensional input and computational cost often hinder the calibration process. This results in simple goodness-of-fit methods such as distance-based methods or likelihood functions to become impracticable [1].

History matching (HM) is a form of model calibration that is suitable for high-dimensional and expensive simulators. Anterion [2] first applied it to the oil industry to diminish computation time in inverse problems for reservoir simulation. Since then, more applications using HM to get reservoir characteristics have emerged [3, 4, 5]. HM sequentially cuts down the initial input space using an implausibility threshold that includes various of uncertainty sources. At each iteration, HM rules out samples in the implausible input domain, and generates samples that eventually define the non-implausible domain providing a reasonable match between the model output and experimental data. Evaluating the model output at each sample point is typically expensive, therefore HM also involves Bayesian emulation [6], which reduces the running cost of complex simulators and quantifies the variance of the predictive posterior output.

Recently, HM has been successfully applied to highly nonlinear geophysical simulators [7], to galaxy formation models [8], and to large climate systems modeling [9] amongst other applications. In the above literature, the non-implausible domain can reduce to orders of magnitude smaller than the initial input space, and might present a complex topology or be disconnected. Therefore, generating samples from such non-implausible domain has remained an open and challenging problem.

Traditionally, there have been at least three classes of methods to generate samples in the non-implausible set: an acceptance-rejection strategy [10], an implausibility driven evolutionary Monte Carlo algorithm [11], and a perturbation approach [1]. However, their efficiency depends on several requirements. Recent work [12] has shown that Subset Simulation can be employed as an efficient sampling scheme.

Unlike Direct Monte Carlo, Subset Simulation progressively decomposes a rare event, which has a small failure probability, into sequential less rare nested events. Samples providing better non-implausibilities in intermediate events remain as seeds to generate more samples towards the final target space. Eventually, rare samples of interests can be targeted down given the occurrence of less rare events.

The original Subset Simulation algorithm uses Modified Metropolis algorithm (MMA) [13] to generate conditional samples that belong to intermediate less rare failure events. During the last decade, multiple versions of sampling schemes for Subset Simulation have emerged in different contexts [14, 15]. Santoso et al. [16] repeat the sample generation process until the samples are accepted at the first step. Miao and Ghosn [17] proposed a delayed rejection approach. Recently, Papaioannou et al [18] proposed a conditional sampling from the current sample, avoiding the first step in the traditional Markov Chain Monte Carlo (MCMC) algorithm. Au and Patelli [19] interpreted this sampling method from a different perspective and validated its powerful efficiency in reducing correlation between samples. This work considers different MCMC strategies and compares their performance in the context of model calibration.

The paper is organized as follows. Section 2 presents important components of HM: initial design for simulation, Bayesian emulation, uncertainty quantification, implausibility threshold, and sampling on the non-implausible domain. Section 3 presents different sampling algorithms for Subset Simulation. In section 4, an example combining Subset Simulation and HM is pre-

sented. Finally, section 5 provides some conclusions.

2 HISTORY MATCHING

HM aims to find the input domain (representing value ranges of parameters) that provides a match between computer model outputs and observations for a complex physical system. HM sequentially cuts down the non-implausible input domain, which leads to a higher concentration of non-implausible samples in each subsequent iteration (also known as *wave*) , using an implausibility threshold that takes into account diverse sources of uncertainty. To reduce the computation time, HM adopts emulation to make inference about the simulator's output. If a large number of non-implausible samples are found in a dramatically reduced input space, HM can be stopped. non-implausible in a wave. If the stopping criterion is not met, HM refocuses on the non-implausible domain via sampling on it and conduct another wave of model running and implausibility testing. A typical HM workflow is summarized in Figure 1.

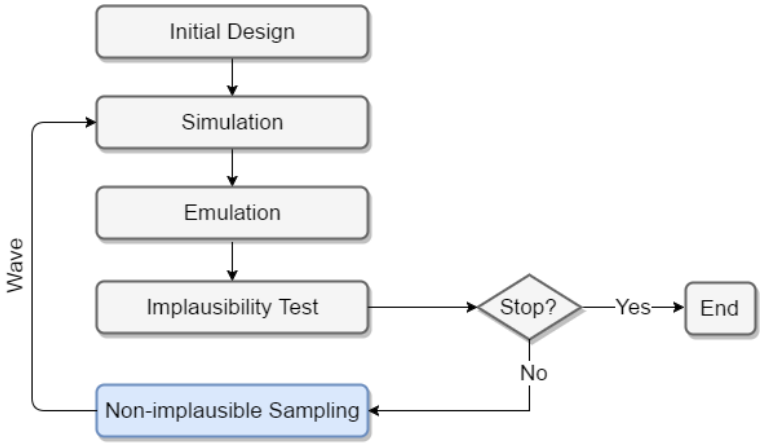

Figure 1: Typical workflow of HM. This work focuses on the sampling from the non-implausible domain efficiently.

An initial design to run the simulator is the first step for a typical HM workflow. Let z be an observation of a physical system y and $g(\cdot)$ be a simulator. Given z, wide possible ranges of the input domain are considered at the beginning. As a rule of thumb, Loeppky et al [20] suggest $n = 10d$ as the number of sample points, where d is the input dimension. We resort to Latin hypercube sampling (LHS) to efficiently explore the input space. Once an LHS design is specified, $g(\cdot)$ is run at each input point $\mathbf{x}_i \in \mathbb{R}^d$. Hence, the initial design as training data for the emulation is $\{(\mathbf{x_1}, g(\mathbf{x}_1)), ..., (\mathbf{x_n}, g(\mathbf{x}_n))\}$.

Based on the simulation data, a Bayesian emulator [21, 22, 23] interpolates the model output to reduce the computational cost. A Bayesian emulator (also called Gaussian process emulator or Kriging) is a stochastic approximation to the output of an expensive computer model, which is widely used as a surrogate for complex simulators. Bayesian emulators have been found effective in multiple disciplines: modeling for structural dynamic analyses [24, 25], stochastic mechanical responses [26] and reliability assessments [27], amongst any others.

In this paper, we use a Bayesian emulator of the form:

$$g(\mathbf{x}) = \sum_{i=1}^{d} h_i(\mathbf{x})^\intercal \beta_i + Z(\mathbf{x}) \tag{1}$$

where $g(\mathbf{x})$ is the emulator's inference based on the training data; each β_i is a regression coefficient; the function $h(\mathbf{x})^\intercal\beta$ models the global trend of the output, whereas a Gaussian process $Z(\mathbf{x})$ models local variations. Note that, the emulator output interpolates the simulator output, for which $g(\cdot)$ is used to denote both.

The emulator's output $g(\mathbf{x})$ follows the student t distribution: $t_{n-p}(E^*(g(\mathbf{x})), \sqrt{V_c(\mathbf{x})})$, where $n-p$ is the degree of freedom, $E^*(g(\mathbf{x}))$ is the predive mean and $V_c((x))$ is the predictive variance.

HM takes into account three sources of uncertainty: V_o is the observation uncertainty (OU), which stands for the experimental error, such as the finite accuracy of measurement instruments or human operations; V_c is the code uncertainty (CU), which stands for the error generated by the emulation (implemented as computer codes) being an approximation; V_m is the model discrepancy (MD), which stems from the inability to model a physical system y perfectly. Summarizing, the total uncertainty considered is as follows:

$$V(\mathbf{x}) = V_o + V_c(\mathbf{x}) + V_m \tag{2}$$

Let $I(\mathbf{x})$ denote the implausibility that an output matches the specified observation, quantified via the difference between them with uncertainty margins [1]:

$$I(\mathbf{x}) = \frac{|z - E^*(g(\mathbf{x}))|}{\sqrt{V(\mathbf{x})}} \tag{3}$$

If at a sample point $\mathbf{x}_i$, the implausibility measure $I(\mathbf{x}_i)$ returns a small value, it is very likely that this input is an acceptable match between the model output and the experimental data. A criterion of the implausibility threshold $I(\mathbf{x}) \leq 3$ is applied [1] following the Pukelsheim three sigma rule [28] (around 99%, or to say nearly all values in a distribution lie within a three-fold standard variance band on either hand of the mean). Sample points that fail the criterion are considered implausible.

Once the non-implausible samples at the current wave are defined by the implausibility measure, if the stopping criterion of HM is not met, the new wave begins by sampling from the non-implausible domain. As previously mentioned, the initial input design can be generated easily by an LHS plan. However after the first wave, sampling from a non-implausible domain with a complex topology can be challenging. It has been recently proposed [12] to use Subset Simulation to sample new emulation points since the non-implausibility domain can very naturally be interpreted as a failure set, both requiring the definition of a threshold.

The conceptual core of Subset Simulation is: generate a sequence of subsets $F = \{\mathbf{x} : I(\mathbf{x}) < 3\} = F_d \subset F_{d-1} \subset \cdots \subset F_1$, so that the probability for the event of interest F (here is the non-implausible set) can be calculated as $P(F_1) \times P(F_2|F_1) \times \cdots \times P(F_d|F_{d-1})$. This decomposes the event F of small probability into conditional events that are more likely to happen and easier to sample from. Every intermediate event corresponds to a level in Subset Simulation. An MCMC algorithm is applied to populate samples for intermediate events and eventually for the target event.

There are two important parameters for Subset Simulation: the level probability p_L and the number of samples in each level n_L. Both of them are determined by the user. The level probability is usually chosen as 0.1 in the literature [29]. The number of samples for each level n_L should be decided by a balance considering between the dimensions and the computation speed. It must also ensure integer values for both the number of chains in the MCMC algorithm $n_c = n_L \times p_L$ and the number of new samples in an event $n_s = (1 - p_L)/p_L$.

At the initial 0^{th} level, Subset Simulation selects the first $n_c = n_L \times p_L$ samples $\mathbf{x}_1, ..., \mathbf{x}_{n_c}$ that have the smallest implausibility, according to Equation 3. Then, for the implausibility measure I, the intermediate threshold defining event $F_1 = \{\mathbf{x} : I(\mathbf{x}) < Y_{SS_1}\}$ is:

$$Y_{SS_1} = \frac{I_0^{(n_c)} + I_0^{(n_c+1)}}{2} \tag{4}$$

The goal then becomes the generation of samples from F_1, on which the subsequent levels are conditioned. The Modified Metropolis algorithm (MMA) [13] algorithm applied to sample in F_1 is as follows:

1. Propose a random standard Normal move $\alpha \sim N(\mathbf{x}_k, 1)$ in each dimension from each seed $\mathbf{x}_1, ..., \mathbf{x}_k, ..., \mathbf{x}_{n_c}$, or alternatively adaptive MMA: a Normal move using the variance of the current sampling seeds $\mathbf{x}_1, ..., \mathbf{x}_{n_c}$[29];

2. Accept the candidate movement $\xi = \alpha$ with probability $min\{1, \frac{\phi(\alpha)}{\phi(\mathbf{x}_k)}\}$, otherwise stay at $\mathbf{x}_k$, $\xi = \mathbf{x}_k$;

3. Remain $\mathbf{x}_k$ if the movement $\xi \notin F_1$.

The process is repeated until:

$$n_F = \frac{\sum_{k=1}^{n_L} \mathcal{I}_{I_{\mathbf{x}\in(I<3)}}}{n_L} > p_L \tag{5}$$

where $\mathcal{I}_I(\cdot)$ is an indicator function that counts the number of samples in the relevant set. Figure 2 summarises the workflow for Subset Simulation combined with HM found in [12].

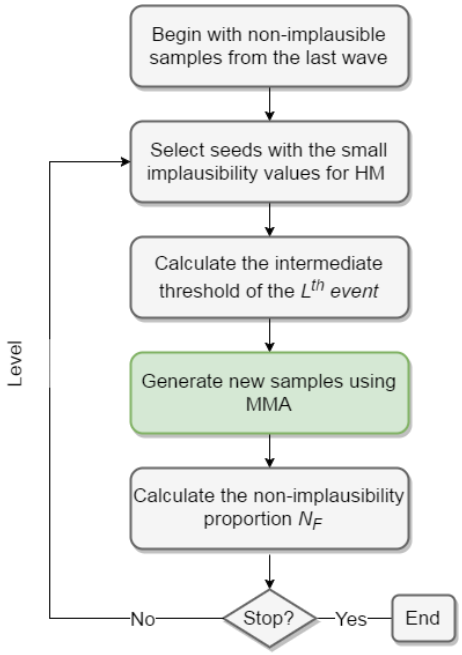

Figure 2: The workflow for Subset Simulation sampling embedded in HM. This work focuses on the sampling generation step, which is implemented and compared with different schemes in following sections.

3 Sampling schemes for Subset Simulation

The original MMA and adaptive MMA are not the only sampling schemes available in the literature [18]. The aim of this paper is to compare the performance of different sampling schemes for Subset Simulation within the HM framework. To this end, some alternative sampling schemes are outlined.

3.1 Delayed rejection

To improve the acceptance rate, Miao and Ghosn [17] repeat the sample generation process by a second proposal PDF if the initial candidate is rejected, thus delaying rejection. The approach can be described as follows:

1. Propose a random standard Normal move $\alpha \sim N(\mathbf{x}_k, 1)$ in each dimension from each seed $\mathbf{x}_1, ..., \mathbf{x}_k, ..., \mathbf{x}_{n_c}$;
2. Accept the candidate movement $\xi = \alpha$ with probability $min\{1, \frac{\phi(\alpha)}{\phi(\mathbf{x}_k)}\}$, and go straight to step 5, otherwise go to step 3;
3. Propose another move e.g. $\alpha' \sim U(\mathbf{x}_k, 1)$ in each dimension from each seed $\mathbf{x}_1, ..., \mathbf{x}_k, ..., \mathbf{x}_{n_c}$;
4. Accept the candidate movement $\xi = \alpha'$ with probability $min\{1, \frac{\phi(\alpha')}{\phi(\mathbf{x}_k)}\}$, otherwise stay at $\mathbf{x}_k$;
5. Remain at $\mathbf{x}_k$ if the movement $\xi \notin F_1$.

3.2 Adaptive MCMC with optimal scaling

Roberts et al. [30, 31] found that the optimal efficiency is achieved when the final sample acceptance rate is around 0.44. Based on this, Papaioannou et al. [18] scale standard deviation of the proposal PDF adaptively (using a scaling parameter λ) to stabilize the acceptance probability. At each level, all n_c seeds are randomly divided into R groups, each with N_a seeds. Given an initial standard deviation for the proposal PDF, the initial N_a seeds are used to generate sample candidates conditioned on them. The value of the scaling parameter λ is determined by the difference between the acceptance rate of the first N_a seeds and the optimal value 0.44. The adapted standard deviation is then plugged into the next group of N_a seeds. The adaptive standard deviation algorithm for conditional sampling is as follows:

1. Randomly divide n_c seeds into R groups of N_a samples;
2. For r^{th} group ($r = 1, ..., R$), set the initial standard deviation of the proposal distribution to $s_i = 1, i = 1, ..., d$, and the initial scaling parameter λ to 0.6 [18] ;
3. Compute the coefficient: $a_i = \sqrt{1 - (\lambda s_i)^2}$, where $a = [a_1, ..., a_d]$ and $s = [s_1, ..., s_d]$;
4. Generate the sample candidates: $\mathbf{x}'_k \sim N(a\mathbf{x}_k, \lambda s)$, $k = 1, ..., N_a$;
5. Accept $\mathbf{x}'_k$ if $\mathbf{x}'_k \in F_1$, otherwise the chain remains at $\mathbf{x}_k$.
6. Compute the average acceptance rate from $N_a \times n_s$ chains: $A = \frac{\mathcal{I}_A(\cdot)}{N_a \times n_s}$, where $\mathcal{I}_A(\cdot)$ is an indicator function which counts the number of acceptable samples;
7. Upgrade the adaptive scaling parameter: $\lambda = 10^{(\log 10(\lambda) + \zeta(A - 0.44))}$, with $\zeta = r^{-1/2}$;
8. Repeat steps 2 to 9 until finishing all R groups of n_c seeds.

3.3 Subset-infinity

Au and Patelli [19] extended the conditional sampling to decomposing the sample $\mathbf{x}_i$ of standard Gaussian distribution by an arbitrary number ($1 \leq N \leq \infty$) of i.i.d. standard Gaussian variables. Let

$$\mathbf{x}_i = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} Z_j \tag{6}$$

where $Z_j \sim N(0,1)$. When $N \to \infty$, sample candidates x_i' conditioned on x_i can be generated by a proposal PDF ($i = 1, ..., d$):

$$p(x_i'|x_i) = \frac{1}{\sqrt{2\pi}s_i} exp\left[-\frac{1}{2s_i^2}(x_i' - ax_i)^2\right] \tag{7}$$

where

$$a_i = 1 - 2\kappa_i \tag{8}$$

$$s_i^2 = 4\kappa_i - 4\kappa_i^2 \tag{9}$$

$$\kappa_i = \int_0^\infty w^2 \Phi(-\frac{w}{2}) p_i^*(w) dw \tag{10}$$

It can be shown that $a_i^2 + s_i^2 = 1$, and $0 \leq \kappa' \leq 1$, hence $a_i \in [-1, 1]$ and $s_i \in [0, 1]$. To sample in F_1, the conditional sampling algorithm for Subset-infinity is:

1. Assign values for parameters of the proposal PDF: e.g. $s_i = 0.5$ (suggested by [32]), $a_i = \sqrt{1 - s_i^2}$, $i = 1, ..., d$;
2. Generate the sample candidate dimension-wise: $\mathbf{x}_k' \sim N(a_i \mathbf{x}_k, s_i)$, $k = 1, ..., n_c$;
3. Remain at $\mathbf{x}_k$ if the candidate $\mathbf{x}_k' \notin F_1$.

4 Numerical example

The different sampling schemes described in the previous section are now compared under the HM framework using an 8-dimensional robot arm model [33].

$$\begin{aligned} &f(x) = (u^2 + v^2)^{0.5} \\ &where \quad u = \sum_{i=1}^{4} L_i \, cos\left(\sum_{j=1}^{i} \theta_j\right), \quad v = \sum_{i=1}^{4} L_i \, sin\left(\sum_{j=1}^{i} \theta_j\right) \end{aligned} \tag{11}$$

The response $f(x)$ is the distance from the end of the robot arm to the origin, on the (u, v)-plane. The input variables and their ranges are shown in Table 1.

In order to implement HM, an observation of 4 units is assumed. The observational uncertainty and model discrepancy were both given fixed values of 0.01. For Subset Simulation, the parameters $p_L = 0.1$ and $n_L = 8000$ were chosen. Table 2 compares the results of the different sampling schemes. It is worth noting that all sampling schemes required 4 levels of Subset Simulation. Also, since by construction MCMC rejects samples from proposal, this yields repetitive sample points. Our aim is to determine which scheme produces more unique non-implausible

Input variables' range	Input variables' meaning
$\theta_1 \in [0, 2\pi]$	angle of the first arm segment
$\theta_2 \in [0, 2\pi]$	angle of the second arm segment
$\theta_3 \in [0, 2\pi]$	angle of the third arm segment
$\theta_4 \in [0, 2\pi]$	angle of the fourth arm segment
$L_1 \in [0, 1]$	length of the first arm segment
$L_2 \in [0, 1]$	length of the second arm segment
$L_3 \in [0, 1]$	length of the third arm segment
$L_4 \in [0, 1]$	length of the fourth arm segment

Table 1: The input variables and their ranges for the robot arm function

Sampling Scheme	Unique non-implausible samples
Original MMA	1956
MMA with adaptive variance	3129
Delayed rejection	1131
Adaptive optimal scaling ($N_a = 80$)	3158
Subset-infinity ($s = 0.5$)	2211

Table 2: The performance difference among differnet MCMC sampling approaches, $p_L = 0.1$ and $n_L = 8000$.

samples. The MMA with adaptive variance and the adaptive MCMC with optimal scaling produced the greatest number of such samples.

The resulting non-implausible input domains given by different sampling schemes are shown in the upper triangle panels of Figure 3. The lower triangle panels of Figure 3 show optical depth plots [10]: on each plane of inputs' pairs, a 20×20 grid is created, with the color denoting the proportion of the non-implausible samples over each grid, which gives plausibility information perpendicular to projective input planes.

5 CONCLUSIONS

The combination of Subset Simulation and HM is possible due to the natural analogy between the non-implausible space in HM and the failure set in reliability analysis. In this paper, different variants of Subset Simulation were presented and illustrated in the numerical example.

In this work, the performance of the sampling scheme was judged by the number of unique non-implausible samples. It was observed that the MMA with adaptive variance and the adaptive MCMC with optimal scaling ranked the highest under this simple criterion.

Future research will concentrate on different metrics of performance, as well as the application of the best-performing sampling schemes to calibrate more complex and realistic models.

REFERENCES

[1] I. Andrianakis, I. R. Vernon, N. McCreesh, T. J. McKinley, J. E. Oakley, R. N. Nsubuga, M. Goldstein, and R. H. White. Bayesian history matching of complex infectious disease models using emulation: A tutorial and a case study on HIV in Uganda. *Plos Computational Biology*, 11(1), 2015.

[2] F. Anterion. History matching: a one day long competition: classical approaches versus gradient method. *First International Forum On Reservoir Simulation, Alpbach, Austria*, 1988.

Figure 3: The implausibility plot and optical depth plot of the input after the second wave using different sampling strategies.

[3] P. Craig, M. Goldstein, A. Seheult, and J. Smith. Bayes linear strategies for matching hydrocarbon reservoir history. *Bayesian Statistics*, pages 69–95, 1996.

[4] R. Baker. Streamline technology, reservoir history matching and forecasting; its success, limitations, and future. *Journal of Canadian Petroleum Technology*, 40:23–27, 2001.

[5] M. Alfi and S. A. Hosseini. Integration of reservoir simulation, history matching, and 4D seismic for CO2-EOR and storage at Cranfield, Mississippi, USA. *Fuel*, 175:116–128, 2016.

[6] R. Schulze-Riegert and S. Ghedan. Modern techniques for history matching. *9th International Forum On Reservoir Simulation, Abu Dhabi, United Arab Emirates*, 2007.

[7] R. Tokmakian, P. Challenor, and Andrianakis Y. On the use of emulators with extreme and highly nonlinear geophysical simulators. *Journal of Atmospheric And Oceanic Technology*, 29:1704–1715, 2012.

[8] I. Vernon, M. Foldstein, and R. Bower. Galaxy Formation: Bayesian History Matching for the Observable Universe. *Statistical Science*, 29(1):81–90, May 2014.

[9] D. Williamson and A. T. Blaker. Evolving Bayesian emulators for structured chaotic time series, with application to large climate models. *Society for Industrial and Applied Mathematics*, 2:1–28, 2014.

[10] I. Vernon, M. Goldstein, and R.G. Bower. Galaxy formation: a bayesian uncertainty analysis. *Bayesian Analysis*, 5(4):619–669, 2010.

[11] D. Williamson and I. Vernon. Efficient uniform designs for multi-wave computer experiments. *arXiv:1309.3520 [stat.ME]*, 2013.

[12] Z. T. Gong, F. A. DiazDelaO, and M. Beer. Bayesian model calibration using subset simulation. *ESREL Risk, Reliability and Safety: Innovating Theory and Practice, University of Strathclyde, Scotland, UK*, pages 293–298, 2016.

[13] S. K. Au and J. L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4):263–277, 2001.

[14] J. Ching, S.K. Au, and J. L. Beck. Reliability estimation for dynamical systems subject to stochastic excitation using subset simulation with splitting. *Computer Methods in Applied Mechanics and Engineering*, 194(Special Issue on Computational Methods in Stochastic Mechanics and Reliability Analysis):1557–1579, 2005.

[15] E. Zio and N. Pedroni. "how to effectively compute the reliability of a thermal hydraulic nuclear passive system". *Nuclear Engineering and Design*, 241:310–327, 2011.

[16] A. M. Santoso, K. K. Phoon, and S. T. Quek. Modified Metropolis-Hastings algorithm with reduced chain correlation for efficient subset simulation. *Probabilistic Engineering Mechanics*, 26:331–341, 2011.

[17] F. Miao and M. Ghosn. Modified subset simulation method for eliability analysis of structural systems. *Struct. Saf.*, 33:233–241, 2011.

[18] I. Papaioannou, W. Betz, K. Zwirglmaier, and D. Straub. MCMC algorithms for subset simulation. *Probabilistic Engineering Mechanics*, 41:89–103, 2015.

[19] S. K. Au and E. Patelli. Rare event simulation in finite-infinite dimensional space. *Reliability Engineering and System Safety*, 148:67–77, 2016.

[20] J. L. Leoppky, J. Sacks, and W. J. Welch. Choosing the sample size of a computer experiment: A practical guide. *Technometrics*, (4):366–378, 2009.

[21] G. Wahba. *Spline models for observational data.* CBMS-NSF Regional Conference series in applied mathematics: vol. 59. Society for Industrial & Applied Maths, 1990.

[22] E. Oakley, J. and A. O'Hagan. Probabilistic sensitivity analysis of complex models: A bayesian approach. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, (3):751–770, 2004.

[23] A. O'Hagan. Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering and System Safety*, 91:1290–1300, 2006.

[24] F. A. DiazDelaO, S. Adhikari, E. I. Saavedra Flores, and M. I. Friswell. Stochastic structural dynamic analysis using Bayesian emulators. *Computers and Structures*, 120:24–32, 2013.

[25] J. Ghosh, J. E. Padgett, and L. Dueas-Osorio. Surrogate modeling and failure surface visualization for efficient seismic vulnerability assessment of highway bridges. *Probabilistic Engineering Mechanics*, 34:189–199, 2013.

[26] E. I. Saavedra Flores, F. A. DiazDelaO, M. I. Friswell, and J. Sienz. A computational multi-scale approach for the stochastic mechanical response of foam-filled honeycomb cores. *Composite Structures*, 94:1861–1870, 2012.

[27] L. Pichler, H.J. Pradlwarter, and G.I. Schueller. A mode-based meta-model for the frequency response functions of uncertain structural systems. *Computers and Structures*, 87:332–341, 2009.

[28] F. Pukelsheim. The three sigma rule. *The American Statistician*, (2):68–95, 1994.

[29] S. K. Au and Y. Wang. *Engineering Risk Assessment with Subset Simulation.* Wiley, 2014.

[30] G. O. Roberts, A. Gelman, and W.R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7:110–120, 1997.

[31] G. O. Roberts and J. S. Rosenthal. Optimal scaling for various MetropolisHastings algorithms. *Stat. Sci.*, 16:351367, 2001.

[32] E. Patelli and S. K. Au. Efficient monte carlo algorithm for rare failure event simulation. *The 12th International Conference on Applications of Statistics and Probability in Civil Engineering held in Vancouver, Canada on July 12-15.*, 2015.

[33] S. Surjanovic and D. Bingham. Virtual library of simulation experiments: Test functions and datasets. 2016.