Importance sampling strategy for oscillatory stochastic processes

Jan Podrouzek
Centre for Water Resource Systems (CWRS), Vienna University of Technology, Austria,
podrouzek@waterresources.at

Abstract: This paper contributes to the structural reliability problem by presenting a novel approach that enables for identification of stochastic oscillatory processes as a critical input for given mechanical models. Identification development follows a transparent image processing paradigm completely independent of state-of-the-art structural dynamics, aiming at delivering a simple and wide purpose method. Validation of the proposed importance sampling strategy is based on multi-scale clusters of realizations of digitally generated non-stationary stochastic processes. Good agreement with the reference pure Monte Carlo results indicates a significant potential in reducing the computational task of first passage probabilities estimation, an important feature in the field of e.g. probabilistic seismic design or risk assessment generally.

Keywords: Stochastic process, Critical excitation, Reliability analysis, Importance sampling, Image processing, Pattern recognition, Identification problem

1. Introduction

The necessity for adopting probabilistic design concepts has become imperative among the structural static problems (Ang and Tang, 1990; Haldar and Mahadevan, 2000; Melchers 2001). On the other hand, structural dynamics is still far from practical utilizations of such concepts despite cheap contemporary computational costs. Among the main reasons is the uncertain nature of environmental loading that has to be modelled as a time-varying phenomena, represented in this paper by non-stationary stochastic oscillatory process as an analogy to earthquake event.

It is a well accepted fact that structures respond in a very uncertain manner to different ground motion events while there is very limited a priori knowledge on the structural behaviour. Same applying for models, an implication is the necessity to perform the structural analysis for each realization of the event separately, which makes the Monte-Carlo based reliability analysis computationally unfeasible for realistic assumptions, i.e. small probabilities and large sample sizes.

There have been several recent attempts to avoid such reliability problem in its full form. Moustafa (2011) proposed a framework for deriving optimal earthquake loads expressed as a Fourier series. More widely, critical excitation methodologists propose to identify critical frequency content of ground motions maximizing the mean earthquake energy input rate to structures, for details see e.g. (Takewaki, 2006). From a different perspective, Barbato et al. (2011) approximates the first passage problem by formulating exact closed form solutions for the spectral characteristics of random processes. Macke et al. (2002) presents an importance sampling technique for randomly excited dynamical systems.

The author of this paper attempt to, unlike the above, maintain the up-to-date most conceptually correct fully probabilistic concept (Ang and Tang, 2007) while reducing the number of required analyses by means of the proposed identification framework. It is based on a non-traditional assumption that there exists a

5th International Conference on Reliable Engineering Computing (REC 2012)
Edited by M. Vořechovský, V. Sadílek, S. Seitl, V. Veselý, R. L. Muhanna and R. L. Mullen

ISBN 978-80-214-4507-9. Published by Ing. Vladislav Pokorný − LITERA

finite set of rules capable of classifying synthetic samples of stochastic processes according to their importance as a critical input for given mechanical model. Whether such set of rules could be formulated for arbitrary system remains an open problem for further research.

2. Development

The identification strategy development follows a transparent image processing paradigm completely independent of state-of-the-art structural dynamics, thus representing a non-traditional option in the field. Reason behind such premise is experimental, aiming at delivering simple and wide-purpose method. The goal can be formulated as follows: find the critical realization ($S_{T,Crit}$) of a stochastic process (S) from a target sample set S_T under defined critical response (C_r) criteria.

Proposed STS strategy steps:

1) Construct a training sample set S_t of size $S_t << S_T$.
2) Solve the mechanical model (i.e. carry out a structural dynamic analysis): S_t -> C_r, usually extremely computationally expensive, therefore the size of S_t should be as small as possible.
3) Select a proper graphical representation G of S_t (in time domain), which should serve for automatic feature extraction in the next step. There are two general options maintaining the physicality of S_t -> G S_t, transformation of S_t into evolutionary spectra (Priestley, 1965) or wavelet-vector coefficients based scalogram (Wolfram, 2011), both as 2D graphical arrays. The computational complexity of this task should be minimized, therefore small resolution is desired.
4) Find a finite set of rules R such that consistently maps R(G S_t) -> C_r, Narrow the search domain by ignoring pixels with constant or random-behaviour. Include pixels into R for which the difference of state values between upper and lower 5th percentile of the ranked G S_t : C_r is maximized.
5) Obtain $S_{T,Crit}$ by applying R -> $S_{T.}$

In the broader context one should use the STS strategy to limit the number of necessary executions of numerical analysis of the mechanical model. It is assumed that mechanisms behind rules extracted from reasonably small samples are applicable to arbitrarily larger sample. Clearly, whenever using a black-box type of approach, there is a risk of extracting mechanisms that apply only to the training sample if its sample size is too small or in cases of “statistical bad luck”. The determination of minimal size of a training set should be based on a requirement for STS’s predictive confidence.

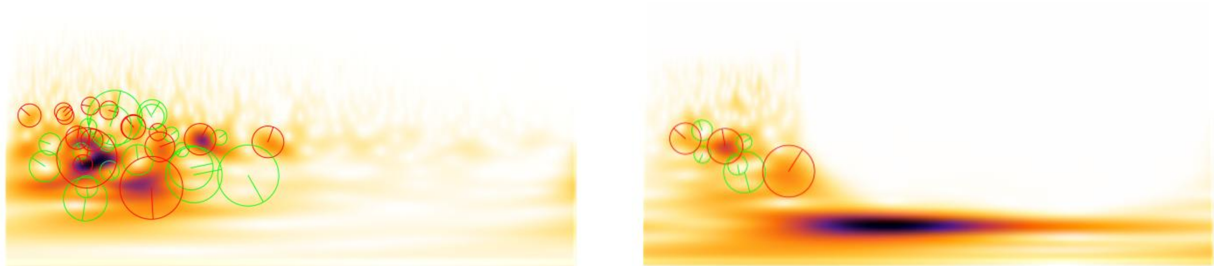


Figure 1. Graphical representation (G) of L1 (left) and L2 (right) in a form of Wavelet Scalogram and visualized detected keypoints (R) using their scale (radius of the circle), orientation and contrast sign (colour).

As stated before, the proposed STS strategy aims at general and automated feature extraction. It should be noted here however that rare instances where experienced when visual comparison of the ranked scalograms G itself enabled for formulation of identification rule R by comparing the number of regions with steep contrast gradient, i.e. image keypoints. For such feature a number of standardized algorithms exists, e.g. implemented SURF (Herbert et al 2008), numerically robust against translation, rotation and scale changes. Such approach can be interpreted as assessment of localized of energy in the time domain and proved to be consistent for configurations of SDOF oscillators loaded by stationary or amplitude modulated processes. In such instances a low number of detected keypoints indicates a critical process, i.e. G has minimal scatter of excitation energy, for example see Fig. 2.

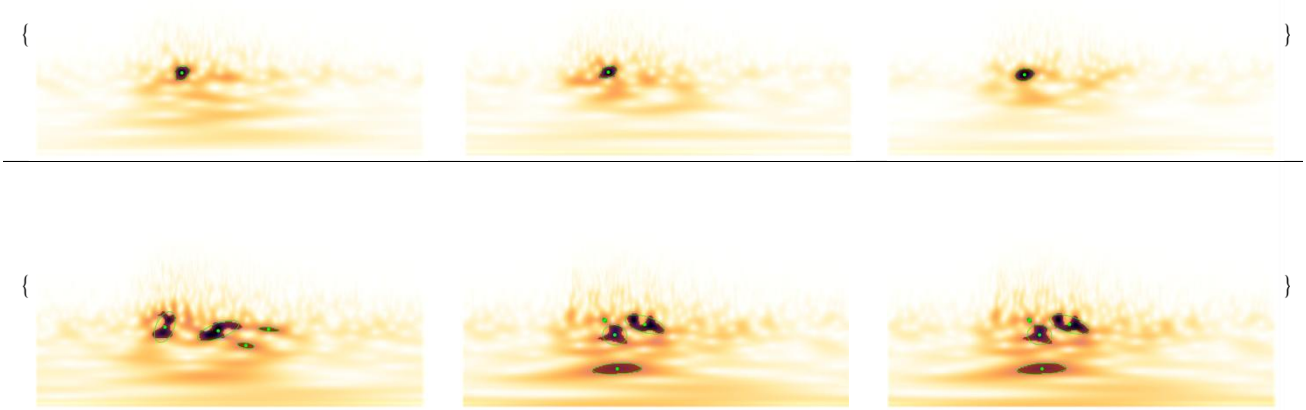


Figure 2. Number of fitted oriented ellipses (based on SURF) as a performance indicator, upper row: 3 ranked maximum and (lower row) 3 ranked minimum response.

The most general non-physical version of STS utilizes several pixels of small-resolution Wavelet Scalograms image for composition of R and R(G S_t) -> C_r mapping (step 4) based on a stochastic sensitivity analysis, returning pixels with state values that varies systematically according to the ranked small sample training sets, see Fig. 3.

The sensitive pixels are usually in clusters forming a line (indicating a dominant scale) and/or points (Fig. 3). Regardless of the attractiveness of emerging questions on physical connections of these clusters to the mechanical models (and dominant frequencies), such debates will not be detailed here due to the limited scope of the paper.

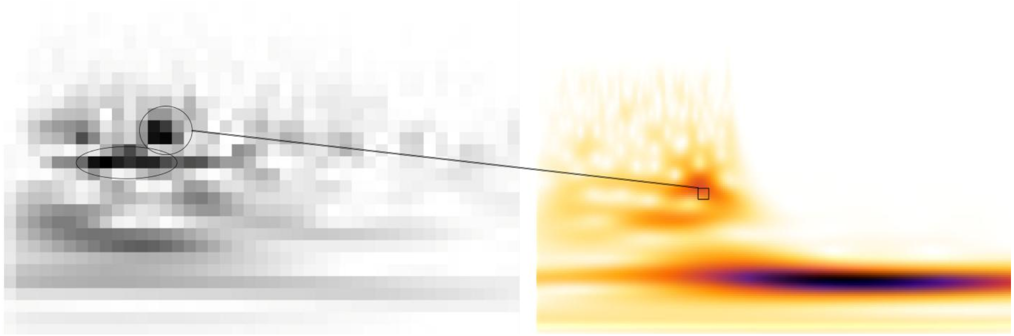


Figure 3. Left: Array of pixels (rescaled) according to their behaviour, darker the colour, more sensitive the pixel is to the ranked Gs, lighter colours indicates random or invariant behaviour; right: corresponding position of the sensitive pixel at the wavelet scalogram.

3. Acceleration and Structural Models

For validation of STS method four distinct combinations of two models (M1 and M2) and loadings (L1 and L2) are considered. The mechanical models represent a single degree of freedom (SDOF) damped linear oscillator (M1) and nonlinear seismically isolated SDOF on a friction pendulum system (M2) subjected to an earthquake loading $F(t) = -m\, a(t)$. Here $a(t)$ is the ground acceleration described as (L1) an amplitude modulated random process

$$a(t) = e(t) \cdot b(t) \tag{1}$$

where $e(t)$ is the amplitude modulating function given by

$$e(t) = 4[\exp(-0.25t) - \exp(-0.5t)] \qquad \text{for } t > 0 \tag{2}$$

and $b(t)$ denotes the stationary zero-mean Gaussian random process with power spectral density

$$S_{bb}(\omega) = S_0 \frac{4\zeta_g^2\omega_g^2\omega^2 + \omega_g^4}{(\omega_g^2 - \omega^2)^2 + 4\zeta_g^2\omega_g^2\omega^2} \tag{3}$$

and as (L2) an amplitude and frequency modulated random process whose objective is to reproduce the general frequency variation characteristics of the acceleration record from the 1964 Niigata earthquake (Shinozuka, 1991) described by the Bogdanhoff-Goldberg-Bernard (1961) envelope function

$$A(t) = a1\, t \exp(-a2\, t) \qquad \text{for } t > 0 \tag{4}$$

and Clough-Penzien acceleration spectrum with parameters S_0, ω_g and $\zeta_g = \zeta_f$ as functions of time:

$$S(\omega,t) = S_0(t)\left[\frac{1+4\zeta_g^2(t)\left[\frac{\omega}{\omega_g(t)}\right]^2}{\left\{1-\left[\frac{\omega}{\omega_g(t)}\right]^2\right\}^2+4\zeta_g^2(t)\left[\frac{\omega}{\omega_g(t)}\right]^2}\right] \times \left[\frac{\left[\frac{\omega}{0.1\omega_g(t)}\right]^2}{\left\{1-\left[\frac{\omega}{0.1\omega_g(t)}\right]^2\right\}^2+4\zeta_f^2(t)\left[\frac{\omega}{0.1\omega_g(t)}\right]^2}\right] \tag{5}$$

$$S_0(t) = \frac{\sigma^2}{\pi\, \omega_g(t)\left(2\zeta_g(t) + \frac{1}{2\zeta_g(t)}\right)} \tag{6}$$

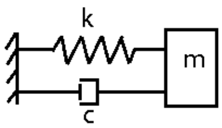
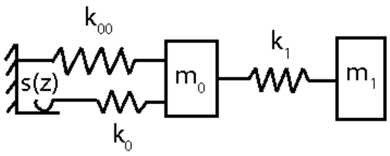
$$\omega_g(t) = \begin{cases} 15.56, & 0 \le t < 4.5 \\ 27.12\,(t-4.5)^3 - 40.68(t-4.5)^2 + 15.56, & 4.5 \le t < 5.5 \\ 2.0, & t \ge 5.5 \end{cases} \tag{7}$$

$$\zeta_g(t) = \begin{cases} 0.64, & 0 \le t < 4.5 \\ 1.25\,(t-4.5)^3 - 1.875(t-4.5)^2 + 0.64, & 4.5 \le t < 5.5 \\ 0.015, & t \ge 5.5 \end{cases} \tag{8}$$

where parameters a_1 = 0.68, a_2 = 0.25 and σ = 100.

The nonlinear mechanical model M2 represents a building (SDOF) combined with a friction based seismic isolation (friction pendulum system) device that introduces another mechanical degree of freedom as well as an internal variable representing plastic slip *z*. The implementation was adopted from (Bucher, 2010) and will not be detailed in this paper. The structural data for both M1 and M2 are provided in Table I, random realizations of L1 and L2 and response characteristics are depicted at Fig. 2.

Table I. Mechanical models and structural data

M1			M2		
	m	400 kg		m_0	6080 kg
	k	80000 KN/m		m_1	79770 kg
	c	120		k_0	42372 KN/m
				k_{00}	2629 KN/m
				k_1	62500 kN/m

Critical response criterion was formulated either as absolute values of top displacement of mass most distant from the application of seismic load or as given percentile of the mean-square values of the displacements. The former criterion led to better identification performance and therefore was adopted.

4. Identification Results

Development and testing of the STS on multiple scales and process-model scenarios showed that it is difficult, perhaps impossible, to formulate a general identification rule of physical interpretability, a fact that corresponds with the structural dynamics paradigm. One of such attempts led to the formulation of R incorporating the image keypoints as a way of quantifying the energy scatter in the loading process. Therefore, soft computing techniques were deployed in search for general black-box type method. The presented state of STS was tested on large number of clusters composed from a total of 4.2×10^4 realizations of Kt and Ni process in combination with various mechanical models. The stochastic simulations revealed the existence of R for every tested process-model scenario. Results presented in Fig. 5 were chosen to demonstrate the variability of performance and do not represent the best nor worst analyzed process-model instances.

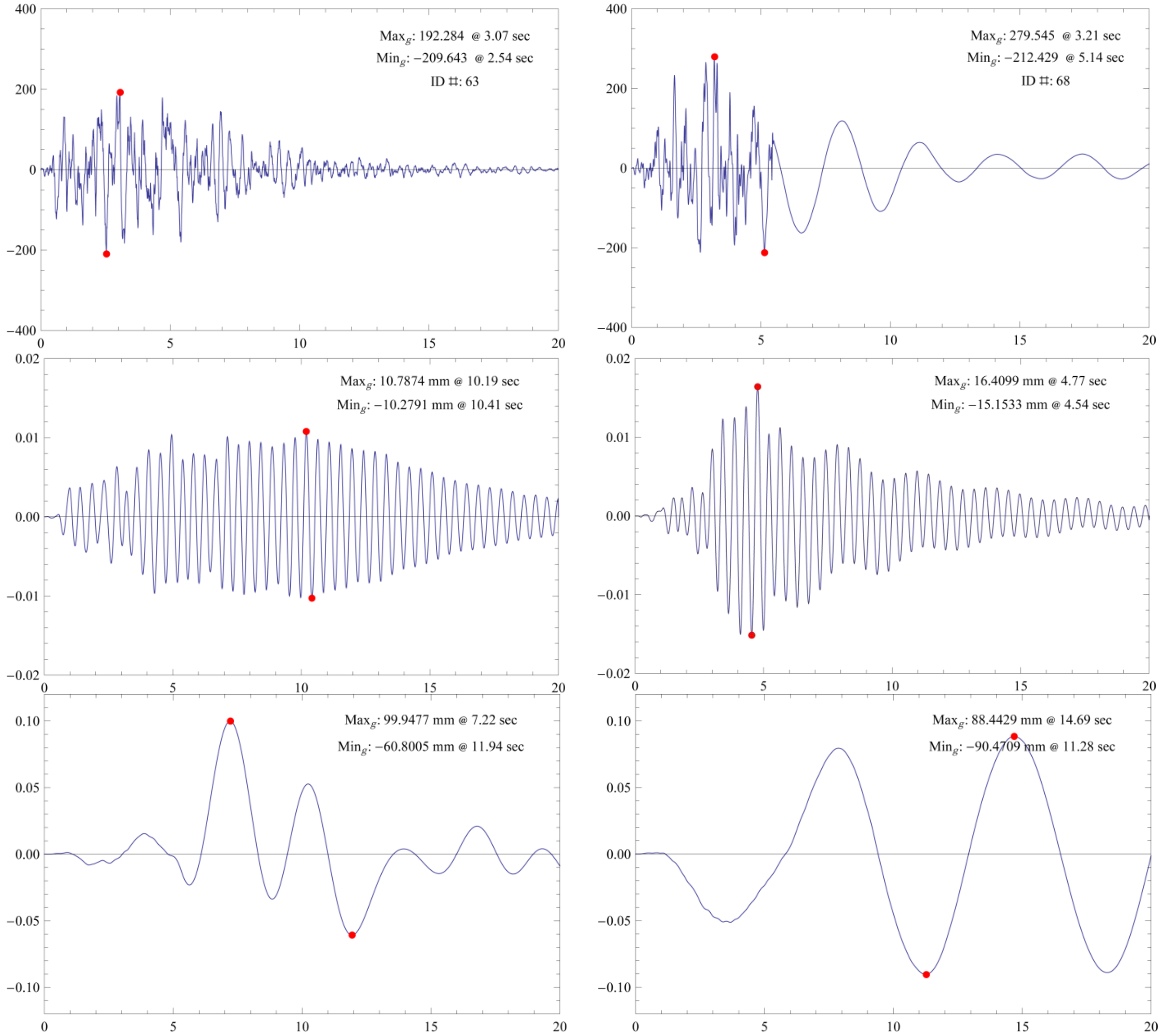


Figure 4. Example of realizations: left column top down: L1 process, M1 and M2 response to L1; right column top down: L2 process, SDOF response to L2, FPS response to L2 (note the abrupt change of frequency content at 5.5 sec); time at horizontal axes, acceleration/displacement on vertical axes.

The performance index was defined according to the following integral

$$P = \int_0^1 PDF_{min}(x).PDF_{max}(x)\,\mathrm{d}x \tag{9}$$

where $PDF_{min/max}$ states for the probability distribution function fitted to the ranked minimum/maximum set, growing isolation of these functions indicates better performance (see Fig. 6). The integration range corresponds to the admissible value of the G pixels.

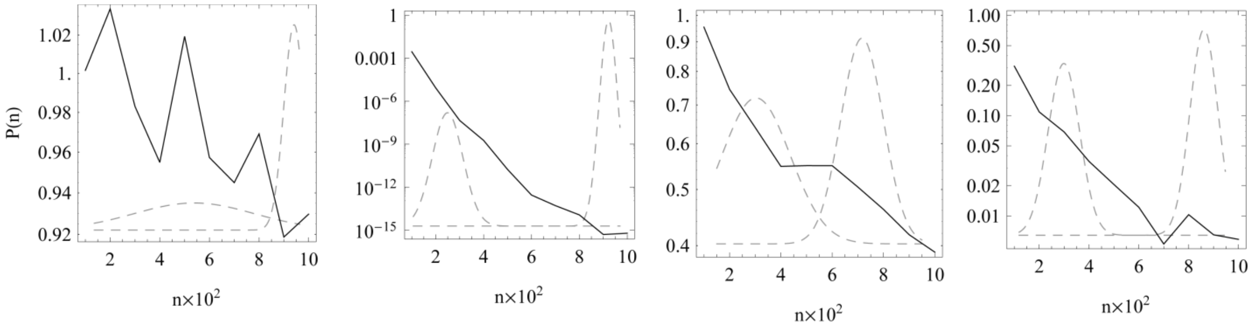


Figure 5. Performance index P as a function of sample size *n*, left to right: KtM1, KtM2, NiM1, NiM2; dashed line represents the Normal distribution PDFs fitted to 21 joined min and max sets for sample size $n = 2.1 \times 10^4$; Note the PDFs spacing effect on performance index.

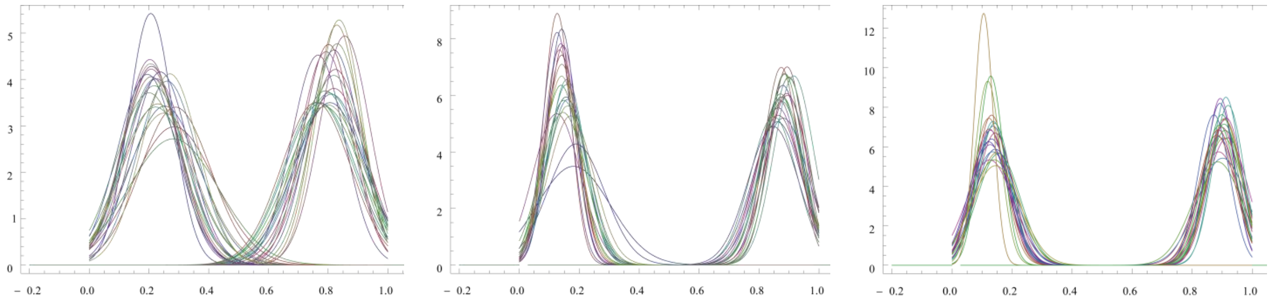


Figure 6. Ranked sets PDF {min, max} Isolation growth with increasing sample size *n* = {100,500,1000}; KtM2 realizations.

5. Importance Sampling

Following a successful formulation and validation of R according to the proposed STS, the importance sampling strategy is based on applying R to the full (original) set of realizations of stochastic processes and sorting the functional values of this product. Finally, the first *n* realizations corresponding to the ranked set are determined as critical input for numerical models. The determination of *n* depends on the required Importance Sampling confidence, e.g. in the presented case study (KtM2 model-process scenario) $n = 10$, i.e. 1% of the full set (1000), see figures 7 and 8.

The importance sampling test scenario, as described above, proved to be a consistent measure for reducing the 1000 sample set to a smaller set while maintaining the same critical response characteristics. The STS utilized 100 sample training set (10%) and the consequent importance sampling required additional 10 analyses (1%), therefore reducing the computational task by 89%. The additional 1% ensured that the important sample (most critical response) was captured by over 91% (within 21 test runs). Note the effect of emergent 2nd branch STS artefact from C_r distribution plot according to ranked R product. The inverse of the same plot (fig. 7) does not exhibit such effect, representing the amount of unaccounted information by STS. This is partly due to (i) incorporating only one sensitive point and (ii) ambiguous C_r -> R_p identifier based RGB channels. The performance of STS could be enhanced by including multiple sensitive points with cross-correlations (i) and by modifying the R products by labels (e.g. random binary sequencing) or by enhancing the colour depth to ensure uniqueness (ii).

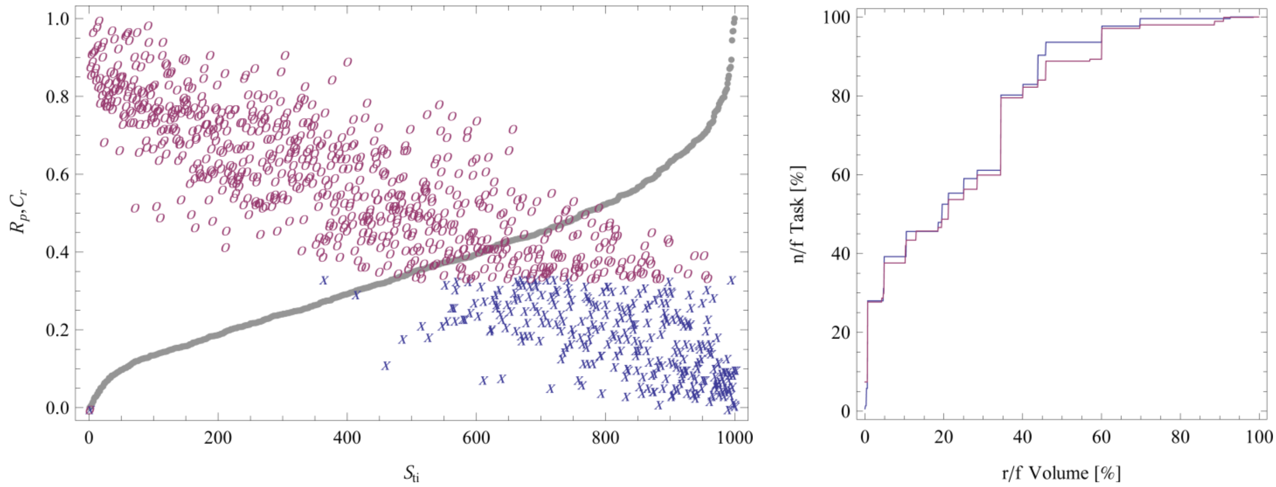


Figure 7. Left: Inverse property of ranked critical response (C_r points) and the R product (both Rescaled to (0,1) vs. sample size 1000, S_{ti}); here for illustration n = 10 and corresponding critical input markers "x", others "o". Right: Percentage of necessary/full computational task as a function of C_r ranked maxima (required/full volume) for 2 colour channels (R**GB**).

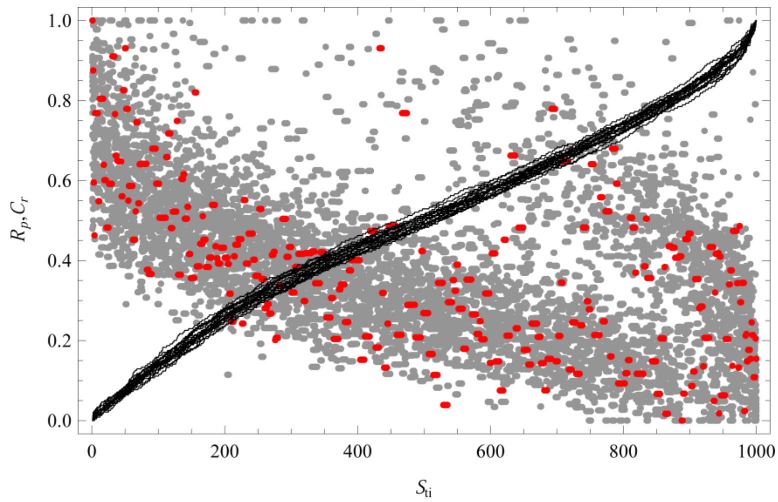


Figure 8. Rescaled distribution of C_r points (gray cloud) according to ranked R product (black line) from individual realizations S_{ti}; 21 repeated runs; particular realization in red points; note the emergent 2nd branch STS artefact.

The effect of unaccounted information does not only exhibit itself via the 2nd branch, but clearly also by the inability to always capture the single C_r maximum, as one might observe on the comparison plot at Fig. 9. Here the goal was to determine the probability of exceeding a critical displacement threshold u_{lim} at different sample size scales and compare it against reference pure Monte Carlo values. In terms of accuracy the maximum reached deviation between the MC reference and SST value was 15%, however, in terms of computational efficiency the STS based importance sampling utilized only 4.6% of the MC computational cost, i.e. 0.6% for feature extraction and the remaining 4% for running the *n* realizations corresponding to the P_r ranked sets. This result indicates that there is significant potential in the application of the STS approach to the estimation of first passage probabilities. Nevertheless, the accuracy in its present form is not comparable to established simulation techniques. Due to its substantial computational advantage, however, the present approach will be suitable especially for reliability-based design optimization in which the reliability analysis has to be repeated frequently.

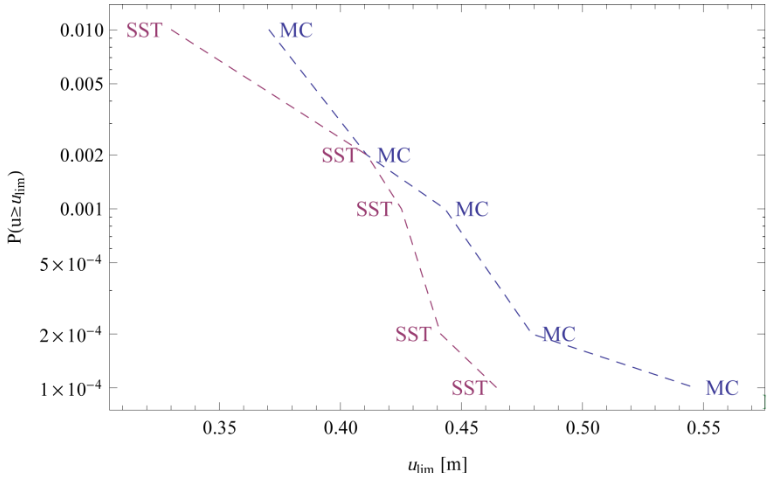


Figure 9. Determination of probability of exceeding a critical displacement threshold u_{lim} at different sample size scales: Comparison of pure Monte Carlo method (100% computational costs) and STS based importance sampling at 4.6% of computational cost.

6. Discussion and Conclusion

A novel **S**mall **T**raining **S**et strategy proposed by the author enables for identification of critical stochastic oscillatory processes with respect to given mechanical model. Such process is understood here as an environmental load acting on a structural system. From a design point of view, it is essential to understand what particular realization of such process has the critical impact on the structure. Traditionally, it is understood that each individual dynamical system has a very unique response to various stochastic loads. Therefore, for Monte-Carlo-based structural reliability considerations, all realizations of the stochastic load must be executed individually, making the task computationally unfeasible for realistic failure probabilities, since no sampling technique capable of reducing such task is available up to current date.

Motivated by the latter statement, an importance sampling strategy is formulated such that it reduces the size of the computational task without sacrificing any of the properties of fully probabilistic approach. As demonstrated on the numerical examples, the identification is feasible with varying performance according to the type of process-model scenario. As one may observe at Fig. 5, there is no relationship between the complexity of the process or model and the performance index.

Positively tested for both stationary and non-stationary processes, linear and non-linear mechanical models, an important implication is that the proposed STS strategy moves the fully probabilistic approach within the context of dynamical systems one step closer to the engineering practitioners, motivated by the ever-growing demand for performance-based design. Besides from the engineering community, STS may be a useful technique in the context of environmental sciences, such as water resources, solving analogous problems, e.g. realistic critical precipitation scenarios.

Further research will focus on possible extensions and improvements regarding the accuracy of the first passage probabilities as well as the treatment of more complex engineering models including structural dynamics and hydrology.

Acknowledgements

Author would like acknowledge financial support from the Austrian Science Funds (FWF) as part of the Vienna Doctoral Programme on Water Resource Systems (DK-plus W1219-N22) and the hospitality of the Department of Civil Engineering and Engineering Mechanics at Columbia University, NY, where part of the research was conducted, especially Prof. Deodatis for inspiring discussions.

References

Ang, A. H.-S. and Tang W.H. *Probability Concepts in Engineering Planning and design*, New York: John Wiley & Sons, Inc., 1990.

Bay, H., Ess, A. Tuytelaars, T., and Van Gool, L., "SURF: Speeded Up Robust Features", *Computer Vision and Image Understanding* (CVIU), Vol. 110, No. 3, 2008, 346-359.

Barbato, M. and Conte, J. P. Structural Reliability Applications of Nonstationary Spectral Characteristics, *Journal of Engineering Mechanics*, ASCE, May 2011, 371-382.

Bucher, Ch. Optimal probabilistic design of friction-based seismic isolation devices, *Safety, Reliability and Risk of Structures, Infrastructures and Engineering Systems* – Furuta, Frangopol & Shinozuka (eds), © 2010 Taylor & Francis Group, London, ISBN 978-0-415-47557-0.

Haldar, A. and Mahadevan, S. *Reliability assessment using finite element analysis*. New York: John Wiley & Sons, Inc. 2000.

Melchers, R.E. *Structural reliability analysis and prediction*. 2nd edn. Chichester: John Wiley & Sons, Inc. 2001.

Macke, M. and Bucher, Ch. Importance sampling for randomly excited dynamical systems, *Journal of Sound and Vibrations* 268, 2003, 269-290.

Moustafa, A. Damage-Based Design Earthquake Loads for Single-Degree-Of-Freedom Inelastic Structures, *Journal of Structural Engineering*, ASCE, March 2011, 456-467.

Novak, D., Strauss, A., Bergmeister, K., Hoffman, S. and Pukl, R. Advanced life-cycle analysis of existing concrete bridges, *Journal of Materials in Civil Engineering*, 20(1), 2008, 9-19.

Podrouzek, J. and Strauss, A. Probabilistic assessment of concrete structures in the context of model uncertainties. *Computer and Concrete* (In Print), 2011.

Priestley, M. B. Evolutionary Spectra and Non-Stationary Processes. *Journal of the Royal Statistical Society*. Series B (Methodological), Vol. 27, No. 2 (1965), 204-237.

Shinozuka, M. and Deodatis, G. Simulation of Stochastic Processes by Spectral Representation. *Appl. Mech. Rev*. 44, 191 (1991), DOI:10.1115/1.3119501.

Takewaki, I. Probabilistic Critical Excitation Method for Earthquake Energy Input Rate, *Journal of Engineering Mechanics*, ASCE, September 2006, 990-1000.

Wolfram Mathematica Documentation Center, ref/WaveletScalogram, 2011.