Reduction of random variables in the Stochastic Harmonic Function representation via spectrum-relative dependent random frequencies

Jianbing Chen[1], Liam Comerford[2,3], Yongbo Peng[1], Michael Beer[2,3], Jie Li[1]

[1] *State Key Laboratory on Disaster Reduction in Civil Engineering, Tongji University, China*

[2] *Institute for Risk and Reliability, Leibniz University Hannover, Germany*

[3] *Institute for Risk and Uncertainty, University of Liverpool, Liverpool, L69 3GH, UK*

Abstract

Two significant developments pertaining to the application of the Stochastic Harmonic Function representation of stochastic processes are presented. Together, they allow for Gaussian records to be simulated within the bounds of the representation with the fewest number of random variables. Specifically, independent random frequencies that form a staple component of the Stochastic Harmonic Function are replaced by dependent random frequencies, along with a specific scheme for choosing frequency interval widths. Numerical examples demonstrating spectrum reconstruction accuracy and estimated PDF convergence to the Gaussian are presented to support the work.

Keywords: Stochastic Harmonic Function, Stochastic Process, Power Spectrum, System Response Spectrum, Random Frequencies

1. Introduction

Propagation of randomness through non-linear systems can be highly computationally demanding where numerical schemes are utilised. For many problems, a brute force Monte Carlo simulation is not feasible when the time required to simulate the system output for a single case is not insignificant. Possible approaches to such problems include simplification of the system response, for example through statistical linearisation [1, 2]; response probability density function (PDF) determination through Wiener Path integrals

[3, 4, 5], recently extended to utilise PDF sparsity in a monomial basis [6]; and application of the Probability Density Evolution Method (PDEM) [7, 8, 9] to a track probability masses through the system. Where stochastic input processes need to be propagated through a non-linear system, discounting the linearisation option cited above, process representations that are able to rely on fewer random variables could further reduce computational demand.

To address these issues, Ref. [10] introduced the stochastic harmonic function (SHF) method of stochastic process representation as an extension of the established spectral representation approach [11, 12] to be used to describe an input process for a PDEM problem. SHFs are defined as a summation of harmonics with associated power from a described power spectrum with both random phase and random frequency. The addition of random frequencies over deterministic frequencies actually allows a stochastic process to be represented with fewer random components than the standard spectral representation approach. Further, due to the relaxed restrictions on the discretisation of the frequency space, an ensemble of records of an SHF defined process can more accurately approximate a target power spectrum function. Although SHF is considered an extension of the spectral representation method, the idea of using random frequencies as part of a Fourier series to simulate random processes is not new. Shinozuka [12] proposed using independent identically distributed random frequencies to simulate random processes, with distribution function based upon the scaled power spectrum. This is similar to the SHF-I (given in Section 2), but without frequency partitioning. Shinozuka also used small random frequencies in combination with the familiar deterministic frequencies in the spectral representation method [13], essentially nudging the deterministic integer frequencies slightly to reduce periodicity in the simulated records.

Although the SHF already presents as a practical tool for process representation, particularly in the context of PDEM and has been extended for non-stationary process representation [14], there remain significant potential improvements that could further advance and generalise the method. Two distinct alterations to the definition of the SHF are explored in this work for the purpose of reducing the number of random components, and increasing convergence rate of simulated records to the target process statistics:

1. Dependency will be introduced between random frequencies of the SHF representation. By making the frequencies of the harmonics completely dependent, the number of required random variables can be effectively

halved.

2. The SHF formulation allows for a range of frequency interval definitions including overlapping intervals. By focusing on the convergence of the PDF of simulated records to the Gaussian, the optimal interval definition is derived.

To begin with, an overview of classical spectral representation work upon which SHF is based, followed by basic SHF-I and SHF-II representations is provided in Section 2. This background is necessary to appreciate the new developments that focus on reduction of random variables. Sections 3 and 4 will focus on the above points 1 and 2 above respectively. These are explored with a focus on the SHF-II formulation only. Finally, the hypotheses presented in Sections 3 and 4 are then confirmed by way of numerical experiments in Section 5.

2. Overview of spectral representation and Stochastic Harmonic Functions

The established SHF process description and extensions thereof discussed in this work are founded on spectral representation work summarised in [15], essential features of which are presented here for completeness, and is henceforth referred to as "classical spectral representation".

2.1. Spectral representation of stochastic processes

Consider a stationary stochastic process $X(t)$ with the autocorrelation function $R^0(\tau) = \mathrm{E}[X(t)X(t+\tau)]$ and the power spectrum density (PSD) function $S^0(\omega)$, where

$$S^0(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R^0(\tau)e^{-i\omega\tau}d\tau \tag{1}$$

$$R^0(\tau) = \int_{-\infty}^{\infty} S^0(\omega)e^{i\omega\tau}d\omega \tag{2}$$

Such a process may be expressed as the sum of a series of harmonic functions with random phases, i.e.,

$$X^{\mathrm{SR}}(t) = \sum_{j=1}^{N} A(\omega_j)\cos(\omega_j t + \phi_j) \tag{3}$$

where ϕ_j are independent identically distributed (i.i.d) uniform random variables over $[0, 2\pi]$. The amplitudes of each harmonic, $A(\omega_j)$, may be derived from a given spectral density function, $S^0(\omega)$, via the equivalence of the autocorrelation function of $X^{\mathrm{SR}}(t)$ with the target $R^0(\tau)$.

$$A(\omega_j) \approx 2\sqrt{S^0(\omega_j)\Delta\omega_j}. \tag{4}$$

2.2. Stochastic Harmonic Function representation of stochastic processes

The simulated process $X^{\mathrm{SR}}(t)$ tends toward the target process $X(t)$ as N tends to infinity. However, in the context of reproducing the target power spectrum, the SHF formulation allows $X(t)$ to be represented exactly with as few as 2 random variables. The SHF representation is denoted by,

$$X^{\mathrm{SHF}}(t) = \sum_{j=1}^{N} A(\tilde{\omega}_j)\cos(\tilde{\omega}_j t + \phi_j), \tag{5}$$

where again ϕ_i are i.i.d random variables uniformly distributed over $[0, 2\pi]$. In contrast to Eq.3, the frequencies are random rather than deterministic variables. To determine the PDF of each $\tilde{\omega}_i$ and their amplitudes $A(\tilde{\omega}_j)$, as with the previous spectral representation method, the equivalence of the autocorrelation function of $X^{\mathrm{SHF}}(t)$ and the target $R^0(\tau)$ is considered. First, the frequency domain of interest is partitioned into non-overlapping intervals. If said frequency domain is bounded by $[\omega_{\mathrm{L}}, \omega_{\mathrm{U}}]$, then the partitioned sub intervals lie between the points $\omega_{\mathrm{L}} = \omega_0^{(\mathrm{P})} < \omega_1^{(\mathrm{P})} < \omega_2^{(\mathrm{P})} < \cdots < \omega_N^{(\mathrm{P})} = \omega_{\mathrm{U}}$ such that $\tilde{\omega}_j \in [\omega_{j-1}^{(\mathrm{P})}, \omega_j^{(\mathrm{P})}]$. Then the correlation function of $X^{\mathrm{SHF}}(t)$ may be obtained as,

$$\begin{aligned} R_{X^{\mathrm{SHF}}}(\tau) &= \mathrm{E}[X^{\mathrm{SHF}}(t)X^{\mathrm{SHF}}(t+\tau)] \\ &= \sum_{j=1}^{N} \frac{1}{2} \int_{\omega_{j-1}^{(\mathrm{P})}}^{\omega_j^{(\mathrm{P})}} p_{\tilde{\omega}_j}(\omega) A^2(\omega)\cos(\omega\tau)d\omega, \end{aligned} \tag{6}$$

where $p_{\tilde{\omega}_j}(\omega)$ is the PDF of $\tilde{\omega}_j$ with support $[\omega_{j-1}^{(\mathrm{P})}, \omega_j^{(\mathrm{P})}]$. Further, utilising Eq.1, and under the assumption that $S^0(\omega) = 0; \omega \notin [\omega_{\mathrm{L}}, \omega_{\mathrm{U}}]$,

$$R^0(\tau) = 2\sum_{j=1}^{N} \int_{\omega_{j-1}^{(\mathrm{P})}}^{\omega_j^{(\mathrm{P})}} S^0(\omega)\cos(\omega\tau)d\omega. \tag{7}$$

Combining Eq.6 and Eq.7 yields

$$\frac{1}{2}p_{\tilde{\omega}_j}(\omega)A^2(\omega) = 2S^0(\omega). \tag{8}$$

Therefore, the amplitudes are given by

$$A(\tilde{\omega}) = 2\sqrt{\frac{S^0(\tilde{\omega})}{p_{\tilde{\omega}_j}(\tilde{\omega})}}. \tag{9}$$

This means that for a given PDF $p_{\tilde{\omega}_j}(\omega)$, the amplitudes for the SHF process representation may be determined. In [10], two standards are established for the shape of $p_{\tilde{\omega}_j}(\omega)$. The so called SHF of the first kind (SHF-I) uses the shape of the normalised power spectrum for each sub-interval, whereas the SHF of the second kind (SHF-II) uses a uniform distribution for each sub-interval i.e., for SHF-I,

$$p^{(\text{SHF-I})}_{\tilde{\omega}_j}(\tilde{\omega}) = \frac{S^0(\tilde{\omega})}{\int_{\omega^{(P)}_{j-1}}^{\omega^{(P)}_j} S^0(\omega)d\omega}\text{I}\left\{\tilde{\omega} \in [\omega^{(\text{P})}_{j-1}, \omega^{(\text{P})}_j]\right\}, \tag{10}$$

$$A^{(\text{SHF-I})}(\tilde{\omega}) = 2\sqrt{\int_{\omega^{(\text{P})}_{j-1}}^{\omega^{(\text{P})}_j} S^0(\omega)d\omega} \tag{11}$$

and for SHF-II,

$$p^{(\text{SHF-II})}_{\tilde{\omega}_j}(\tilde{\omega}) = \frac{1}{\omega^{(\text{P})}_j - \omega^{(\text{P})}_{j-1}}\text{I}\left\{\tilde{\omega} \in [\omega^{(\text{P})}_{j-1}, \omega^{(\text{P})}_j]\right\} \tag{12}$$

$$A^{(\text{SHF-II})}(\tilde{\omega}_j) = 2\sqrt{S^0(\tilde{\omega}_j)[\omega^{(\text{P})}_{j-1} - \omega^{(\text{P})}_j]}, \tag{13}$$

where $I\{.\}$ is the indicator function with the value being 1 if the event is true and 0 otherwise. Both methods are shown to yield similar results in [10]. It is noted here that both SHF methods do not simulate ergodic process records even in the limit (assuming a finite number of harmonics). The previously introduced spectral representation method will yield records that are ergodic in the first two statistical moments over multiples of the period. Conversely, the SHF records are almost certainly not periodic, and even if all of the random frequencies were chosen with a common multiple, the

period would vary from record to record. Although SHF-II records would not exhibit ergodicity in the first two statistical moments, as the length of the record approaches infinity, where various frequency components in the non-periodic signal may be considered orthogonal, SHF-I records would be ergodic in mean and variance, but not in covariance (where $\tau > 0$), due to constant amplitude across frequency intervals throughout an ensemble but not constant frequency intervals. In this work, extensions are applied specifically to the SHF-II formulation, primarily due to the ease of random frequency generation.

3. Stochastic Harmonic Function representation with dependent frequencies

The SHF formulations shown in the previous section made use of independent random variables for frequencies in non-overlapping bands. In this setting, two random variables are needed per harmonic. While Eq.5 may account for the entire second order statistics of a process over the infinite ensemble of X^{SHF}, even with only a single frequency interval (encompassing $[\omega_{\mathrm{L}}, \omega_{\mathrm{U}}]$), individual realisations are likely highly unrepresentative of the target process. This makes using the absolute minimum number of random variables unsuitable for most practical applications. Indeed in [10], to produce records that exhibit Gaussianity, around $N = 10$ harmonics are needed, and for real scenarios such as earthquake ground motion simulation, $N = 50$ are recommended. Although this is a vast reduction on the number of harmonics usually needed to represent a process using classical spectral representation methods (where N is usually equal to half the number of sample points in digital realisations or half the Nyquist rate), further reduction of random variables without loss of statistical features for individual records would be of significant benefit. Example use cases include large MCS applications and analysis of uncertainty propagation such as PDEM.

In this section, reduction of random variables will be achieved by introducing dependencies between the random frequencies, with all frequencies ultimately generated from a single seed variable. To realise this, all random frequencies $\tilde{\omega_j}$ are defined through a function of a single random variable, i.e.,

$$\tilde{\omega_j} = f_j(\lambda), \tag{14}$$

where $f_j(0) = \omega_{j-1}^{(P)}$ and $f_j(1) = \omega_j^{(P)}$, with λ being a random variable uni-

formly distributed over $[0, 1]$. In this case, the SHF representation is

$$X^{\text{SHF}}(t) = \sum_{j=1}^{N} A(f_j(\lambda)) \cos\left(f_j(\lambda)t + \phi_j\right). \tag{15}$$

Hence, the autocorrelation function of the SHF process (counterpart to Eq.6) becomes

$$\begin{aligned} R_{X^{\text{SHF}}}(\tau) &= \sum_{j=1}^{N} \frac{1}{2} \text{E}\left[A^2(f_j(\lambda)) \cos\left(f_j(\lambda)\tau\right)\right] \\ &= \sum_{j=1}^{N} \frac{1}{2} \int_0^1 p_\lambda(\lambda) A^2(f_j(\lambda)) \cos\left(f_j(\lambda)\tau\right) d\lambda. \end{aligned} \tag{16}$$

Similarly to Eq.7, the target autocorrelation function may be re-written,

$$R^0(\tau) = 2 \sum_{j=1}^{N} \int_0^1 S^0(f_j(\lambda)) \cos\left(f_j(\lambda)\tau\right) f_j'(\lambda) d\lambda, \tag{17}$$

where $f_j'(\lambda) = \frac{df_j(\lambda)}{d\lambda}$. Equating the summed terms in Eq.16 with Eq.17 yields

$$A(f_j(\lambda)) = 2\sqrt{\frac{S^0(f_j(\lambda)) f_j'(\lambda)}{p_\lambda(\lambda)}}, \tag{18}$$

where $p_\lambda(\lambda) = 1$ for the case where all random frequencies are generated from the same random variable. Further, the most straightforward way to define the dependency on the random variable is by

$$f_j(\lambda) = \omega_{j-1}^{(\text{P})} + [\omega_j^{(\text{P})} - \omega_{j-1}^{(\text{P})}]\lambda, \tag{19}$$

and hence the harmonic amplitudes may be written as,

$$A(f_j(\lambda)) = 2\sqrt{S^0\left(f_j(\lambda)\right)\left[\omega_j^{(\text{P})} - \omega_{j-1}^{(\text{P})}\right]} \tag{20}$$

By setting up $f_j(\lambda)$ as indicated in Eq.19, not only does it simplify the formulation of the harmonic amplitudes $A(f_j(\lambda))$, but it also guarantees a minimum distance (in the frequency space) between neighbouring harmonics as $f_j(\lambda) - f_{j-1}(\lambda) \geq \min[\omega_{j+1}^{(\text{P})} - \omega_j^{(\text{P})}, \omega_j^{(\text{P})} - \omega_{j-1}^{(\text{P})}]$. This feature leads

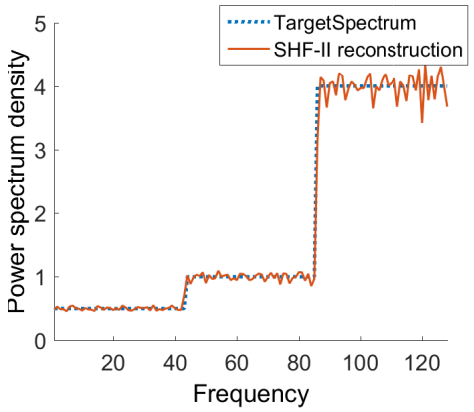
to greater orthogonality between neighbouring harmonics for individual process records, and hence statistics estimated from individual records will more closely match the target. This is particularly true for simulated short time records where overwhelming constructive or destructive interference could otherwise take place between neighbouring harmonics. However, as shown in Section 5, where the frequency intervals remain constant, a corresponding effect of guaranteed separation of harmonics in the frequency space can introduce significant periodicity into the record.

4. Stochastic Harmonic Function representation with variable frequency interval widths

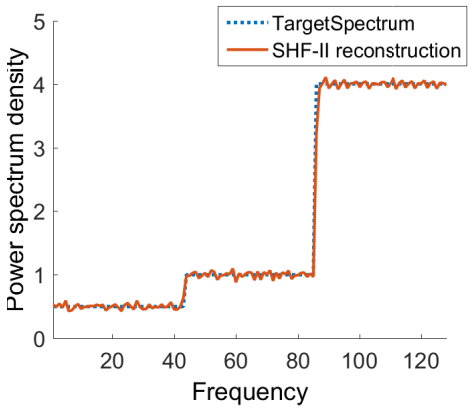
When simulating stochastic processes via SHF with a constant $\Delta\omega$ it can be observed that for any power spectrum where spectral density varies over the frequency domain (i.e., non white noise processes), error in the estimated spectrum increases with increased power. Although the extent of the error is dependent on the choice of spectrum representation from the ensemble of process records, generally if averaging a series of periodograms drawn from individual records, the higher the power spectrum is at a particular frequency, the greater the variance in estimated power will be at that frequency over the ensemble, and hence the larger the error. Figure 1 shows target spectra with increasing power steps to demonstrate this effect. An averaged periodogram based upon an ensemble of 1000 records from an SHF-II representation with constant $\Delta\omega$ is shown in Figure 1a. The increase in error at higher frequencies as the power increases is clearly visible, note that this phenomenon also occurs when utilising classical spectral representation methods.

To mitigate the variation in error across the spectrum, it is suggested that variable frequency widths are defined as inversely proportional to the power spectrum bounded by the partitions. The resulting effect increases ensemble spectrum estimation accuracy at higher powers, equalising the average error over all frequencies, and increasing overall rate of convergence to the target. Considering again Figure 1, when reconstructing the spectrum from SHF-II simulated records with spectrum-relative frequency intervals (shown in Figure 1b), the variation in reconstruction error is visibly more consistent.

It can also be shown that by utilising SHF-II with the aforementioned spectrum-relative frequency widths, not only does the bi-variate distribution of the simulated records converge to the Gaussian (by way of the Lyapunov Central Limit Theorem, e.g., [16, 17]), but does so at a faster rate with in-



(a) Example target spectrum with SHF-II reconstruction using constant $\Delta\omega$. Note that the reconstruction variance increases with power spectrum density



(b) Example target spectrum with SHF-II reconstruction using spectrum-relative $\Delta\omega$. Here reconstruction variance remains constant with increasing power spectrum density

Figure 1: Comparisons of 40-harmonic SHF-II spectrum reconstructions using different definitions for frequency width

creasing number of harmonics than established spectral representation methods with fixed $\Delta\omega$. This is demonstrated in Section 4.3, however, first, the convergence to the Gaussian PDF for classical spectral representation and for SHF-II are provided for completeness.

4.1. Gaussianity of simulated process records via classical spectral representation

In [15], the convergence of the bi-variate PDF of simulated process records between two time instants via classical spectral representation to the Gaussian is proved by way of the Lyapunov CLT which implies that a sequence of independent random variables $G = \{G_1, G_2 \cdots G_N\}$ are Gaussian as N tends to infinity if the following condition is met:

$$\lim_{N\to\infty} \frac{1}{\sigma_G^{2+\delta}} \sum_{j=1}^{N} \mathrm{E}\left[\left|G_j - \mathrm{E}\left[G_j\right]\right|^{2+\delta}\right] = 0, \tag{21}$$

where σ_G is the standard deviation of G and $\delta > 0$ (in this case $\delta = 1$ is used). If Eq.21 holds for $G = X^{\mathrm{SR}}(t)$ for any two time instants then pairs of variables $Z^{\mathrm{SR}} = \left(X^{\mathrm{SR}}(t_m), X^{\mathrm{SR}}(t_n)\right)$ of the simulated processes via

classical spectral representation exhibit a bi-variate Gaussian PDF with first and second moments given by:

$$\mathrm{E}\left[Z^{\mathrm{SR}}\right] = \left(\sum_{j=1}^{N} \mathrm{E}\left[X_j^{\mathrm{SR}}(t_m)\right], \sum_{j=1}^{N} \mathrm{E}\left[X_j^{\mathrm{SR}}(t_n)\right]\right), \tag{22}$$

$$\sigma^2_{X^{\mathrm{SR}}(t_m)} = \sum_{j=1}^{N} \sigma^2_{X_j^{\mathrm{SR}}(t_m)}, \quad \sigma^2_{X^{\mathrm{SR}}(t_n)} = \sum_{j=1}^{N} \sigma^2_{X_j^{\mathrm{SR}}(t_n)}, \tag{23}$$

$$\mathrm{Cov}\left[X^{\mathrm{SR}}(t_m), X^{\mathrm{SR}}(t_n)\right] = \sum_{j=1}^{N} \mathrm{Cov}\left[X_j^{\mathrm{SR}}(t_m) X_j^{\mathrm{SR}}(t_n)\right], \tag{24}$$

where $X_j^{\mathrm{SR}}(t)$ are the individual harmonic components that make up $X^{\mathrm{SR}}(t)$ as given in Eq.3, i.e.,

$$X_j^{\mathrm{SR}}(t_m) = A(\omega_j)\cos(\omega_j t_m + \phi_j), \tag{25}$$

$$X_j^{\mathrm{SR}}(t_n) = A(\omega_j)\cos(\omega_j t_n + \phi_j). \tag{26}$$

These components are independent for different j as required for Eq.21 on account of the independence of the phase angles ϕ_j. Due to the stationarity of the process, it is sufficient to test the Gaussianity of the one dimensional PDF and then assume Gaussianity of the bi-variate. Also, as the individual harmonics are zero-mean, the term $\mathrm{E}[G_j]$ may be discounted. Therefore, the expected value $\mathrm{E}\left[|G_j - \mathrm{E}[G_j]|^{2+\delta}\right] = \mathrm{E}\left[|G_j|^3\right] = \mathrm{E}\left[\left|X_j^{\mathrm{SR}}(t_m)\right|^3\right]$ is given as:

$$\begin{aligned}
\mathrm{E}\left[\left|X_j^{\mathrm{SR}}(t_m)\right|^3\right] &= \mathrm{E}\left[\left|A(\omega_j)\cos\left(\omega_j t_m + \phi_j\right)\right|^3\right] \\
&= \frac{A(\omega_j)^3}{2\pi}\int_0^{2\pi} \left|\cos\left(\omega_j t_m + \phi_j\right)\right|^3 d\phi_j \\
&= \frac{4A(\omega_j)^3}{3\pi}.
\end{aligned} \tag{27}$$

In the above, the fact that $A(\omega_j)$ is positive has been used. Hence, assuming that $\sigma_G^3 = \sigma^3_{X^{\mathrm{SR}}(t_m)}$ is finite, Eq.21 becomes:

$$\frac{4}{3\pi\sigma^3_{X^{\mathrm{SR}}(t_m)}} \lim_{N\to\infty} \sum_{j=1}^{N} A(\omega_j)^3. \tag{28}$$

The limit may be evaluated as follows,

$$\begin{aligned}\lim_{N\to\infty}\sum_{j=1}^{N}A(\omega_j)^3 &= \lim_{N\to\infty}\sum_{j=1}^{N}\left[2\sqrt{S^0(j\Delta\omega)\Delta\omega}\right]^3\\ &= 8\lim_{N\to\infty}\left(\frac{\omega_u}{N}\right)^{\frac{3}{2}}\sum_{j=1}^{N}\left[S^0(j\Delta\omega)\right]^{\frac{3}{2}}\\ &\leq 8\omega_{\mathrm{U}}^{\frac{3}{2}}\lim_{N\to\infty}\left(\frac{1}{N}\right)^{\frac{3}{2}}N(S^0_{\max})^{\frac{3}{2}}\\ &= 8(\omega_{\mathrm{U}}S^0_{\max})^{\frac{3}{2}}\lim_{N\to\infty}\frac{1}{\sqrt{N}}=0.\end{aligned} \tag{29}$$

In Eq.29, ω_{U} is the ultimate frequency considered in the decomposition (above which the process is assumed to have zero power), hence $\left(\frac{\omega_{\mathrm{U}}}{N}\right)=\Delta\omega$, and $S^0_{\max}$ is the maximum power the considered frequency spectrum.

4.2. Gaussianity of simulated process records via SHF-II representation

In the same way as in the previous section, the Gaussianity of SHF-II processes can be investigated, however due to the additional random variable, the evaluation of Eq.21 is more involved. Further, as will be shown, not all SHF-II simulated processes will exhibit Gaussian PDFs in the limit. Once again, due to the individual harmonics of Eq.5 having zero-mean, the term $\mathrm{E}\left[G_j\right]$ in Eq.21 is discounted. Here, the expected value $\mathrm{E}\left[|G_j|^3\right]=\mathrm{E}\left[\left|X_j^{\mathrm{SHF}}(t_m)\right|^3\right]$ is given as:

$$\begin{aligned}\mathrm{E}\left[\left|X_j^{\mathrm{SHF}}(t_m)\right|^3\right] &= \mathrm{E}\left[\left|A(\tilde{\omega_j})\cos\left(\tilde{\omega_j}t_m+\phi_j\right)\right|^3\right]\\ &= \int_{\omega_{j-1}^{(p)}}^{\omega_j^{(p)}}\int_0^{2\pi}\left|A(\tilde{\omega_j})\cos(\tilde{\omega_j})t_m+\phi_j\right|^3\frac{1}{2\pi}\frac{1}{\Delta\omega_j}d\tilde{\omega_j}d\phi_j\\ &= \frac{1}{2\pi\Delta\omega_j}\left(\frac{8}{3}\right)\int_{\omega_{j-1}^{(p)}}^{\omega_j^{(p)}}A(\tilde{\omega_j})^3d\tilde{\omega_j}.\end{aligned} \tag{30}$$

Again, use has been made of the fact that $A(\tilde{\omega_j})$ is positive. Assuming that $\sigma_G^3=\sigma^3_{X^{\mathrm{SHF}}(t_m)}$ is finite and using the SHF-II definition of A in Eq.13, Eq.21 becomes:

$$\frac{32}{3\pi\sigma^3_{X^{\mathrm{SHF}}(t_m)}}\lim_{N\to\infty}\sum_{j=1}^{N}\sqrt{\Delta\omega_j}\int_{\omega_{j-1}^{(p)}}^{\omega_j^{(p)}}S^0(\tilde{\omega_j})^{\frac{3}{2}}d\tilde{\omega_j} \tag{31}$$

Upon attempting to evaluate the limit in Eq.31, it becomes clear that the condition of Gaussianity is dependent upon the nature of the frequency intervals $\Delta\omega_j$. Therefore, to guarantee Gaussianity for an SHF-II process, restrictions must be imposed upon $\Delta\omega_j$. The most familiar case, akin to the classical spectral representation method, is to use $\Delta\omega_j = \frac{\omega_U}{N}$. In this case,

$$\lim_{N\to\infty} \int_{\omega_{j-1}^{(p)}}^{\omega_j^{(p)}} S^0(\tilde{\omega}_j)^{\frac{3}{2}} d\tilde{\omega}_j = S^0(\tilde{\omega}_j)^{\frac{3}{2}} \Delta\omega_j, \tag{32}$$

and hence,

$$\begin{aligned} \lim_{N\to\infty} \sum_{j=1}^{N} \sqrt{\Delta\omega_j} \int_{\omega_{j-1}^{(p)}}^{\omega_j^{(p)}} S(\tilde{\omega}_j)^{\frac{3}{2}} d\tilde{\omega}_j &= \lim_{N\to\infty} \left(\frac{\omega_U}{N}\right)^{\frac{3}{2}} \sum_{j=1}^{N} S^0(\tilde{\omega}_j)^{\frac{3}{2}} \\ &\leq \omega_U^{\frac{3}{2}} \lim_{N\to\infty} \left(\frac{1}{N}\right)^{\frac{3}{2}} N (S^0_{\max})^{\frac{3}{2}} \\ &= \left(S^0_{\max}\omega_U\right)^{\frac{3}{2}} \lim_{N\to\infty} \frac{1}{\sqrt{N}} = 0. \end{aligned} \tag{33}$$

For amplitudes with dependent frequencies in Eq.20, the result of Eq.33 holds. Despite the dependencies between frequencies, individual harmonics are still independent on account of the independence of ϕ_j. The frequency dependence is lost when $S^0_{\max}$ is used to replace $S(\tilde{\omega}_j)$ in Eq.33, however the dependence will effect the convergence rate of the simulated process PDF to the Gaussian. Specifically, the convergence rate is dependent also on the target power spectrum. For example, a white noise (constant spectrum) process will converge to the Gaussian provably faster with a single seed variable for all random frequencies. This result can be considered as a special case of the following section.

4.3. Gaussianity of simulated process records via SHF-II representation with spectrum-relative frequency interval widths

It has been shown that by using a constant $\Delta\omega_j$ for SHF-II process simulation, as the number of harmonics tend to infinity, the simulated record tends to exhibit a Gaussian PDF. However, it can also be shown in a similar manner that by using variable $\Delta\omega_j$ to maintain constant power across all individual harmonics, not only is the process Gaussian, but it also converges

to the Gaussian at a higher rate. Consider a new set of frequency intervals of width $\Delta\psi_j$ defined by

$$\Delta\psi_j = \psi_j - \psi_{j-1}, \tag{34}$$

where,

$$\psi_j = Q^{-1}\left(\frac{1}{N}Q\left(\omega_{\mathrm{U}}\right) + Q\left(\psi_{j-1}\right)\right), \tag{35}$$

and,

$$Q(\omega) = \int_0^{\omega} S^0(\omega)d\omega. \tag{36}$$

This new set of frequency intervals may be evaluated numerically for a given target power spectrum and covers the same overall space as $\Delta\omega_j$ from $j = 0$ to ω_{U} with no overlap, therefore,

$$\lim_{N\to\infty}\sum_{j=1}^{N} S^0(\tilde{\omega_j})\Delta\omega_j = \lim_{N\to\infty}\sum_{j=1}^{N} S^0(\tilde{\psi_j})\Delta\psi_j, \tag{37}$$

and,

$$S^0(\tilde{\psi_j})\Delta\psi_j \approx \frac{1}{N}\sum_{j=1}^{N} S^0(\tilde{\omega_j})\Delta\omega_j = \mathrm{E}\left[\rho\right] \tag{38}$$

where ρ is the vector of random variables $S^0(\tilde{\omega})\Delta\omega$. Next, referring back to Eq.33, where Gaussianity in the case of constant frequency intervals was confirmed through demonstrating that $\lim_{N\to\infty}\sum_{j=1}^{N}(S^0(\tilde{\omega_j})\Delta\omega_j)^{\frac{3}{2}} = 0$, we define the following function,

$$h(x) = x^{\frac{3}{2}}, \tag{39}$$

and observe that,

$$\lim_{N\to\infty}\sum_{j=1}^{N}(S^0(\tilde{\omega_j})\Delta\omega_j)^{\frac{3}{2}} = N\mathrm{E}\left[h(\rho)\right], \tag{40}$$

and,

$$\lim_{N\to\infty}\sum_{j=1}^{N}(S^0(\tilde{\psi_j})\Delta\psi_j)^{\frac{3}{2}} = Nh(\mathrm{E}\left[\rho\right]). \tag{41}$$

As $h''(x)$ is positive for any $x > 0$ (which is always true for Eq.40 and Eq.41), $h(x)$ may be considered strictly convex in this case and so Jensen's inequality (e.g., [18]) holds, i.e.,

$$Nh\left(\mathrm{E}\left[\rho\right]\right) \leq N\mathrm{E}\left[h\left(\rho\right)\right]. \tag{42}$$

Hence, not only does this show that Eq.41 tends to zero, it also does so at a higher rate than Eq.40.

5. Numerical examples

The proposed SHF-II variants in Sections 3 and 4 will be compared to standard SHF-I and SHF-II in terms of convergence to the target spectrum and to Gaussianity. For all of the examples, the target spectrum is defined by the following stationary earthquake model:

$$S^0(\omega) = S_0 \frac{\omega_g^4 + 4\zeta_g^2\omega_g^2\omega^2}{(\omega_g^2 - \omega^2)^2 + 4\zeta_g^2\omega_g^2\omega^2}, \tag{43}$$

where $\omega_g = 25.6$, $\zeta_g = 0.6$, and $S_0 = 0.5$. This spectrum and associated parameters are chosen to provide a realistic example with a significant change in power spectrum density over the frequency domain of interest. To compare rates of convergence to the target spectrum, the following error function is introduced:

$$e_S = \frac{\sum_{j=1}^{M} \left| S^{(\mathrm{SHF})}(\omega_j) - S^0(\omega_j) \right|}{\sum_{j=1}^{M} S^0(\omega_j)}, \tag{44}$$

where $S^{(\mathrm{SHF})}(\omega_j)$ is the spectrum reconstructed via SHF. The choice of error function is largely arbitrary, serving only as a point of reference for comparison of different reconstruction methods. In the first numerical example, the number of generated processes required for the error between the target and reconstructed spectrum to drop below 2% in Eq.44 for the four different SHF approaches are presented. The target spectrum is shown in Figure 2 along with an SHF-II reconstruction with 2% error as defined by Eq.44. The number of averaged records required to achieve this level of accuracy over 256 point samples for 1, 10, and 40 harmonics are shown in Figure 3. Note that the tests were repeated 100 times with the mean values displayed in the figures (to reduce variance in the results).

The first observation from Figure 3 is that, as expected, more harmonics per record result in less records required to accurately approximate the spectrum of the process. It can also be determined that when using only a single harmonic, SHF-I has an advantage over SHF-II approaches (all three of which are effectively identical in the case of a single harmonic). This is due to a similar effect as that demonstrated in Eq.42, and quickly dissipates as the number of harmonics increase. Next, the same data is shown, normalised

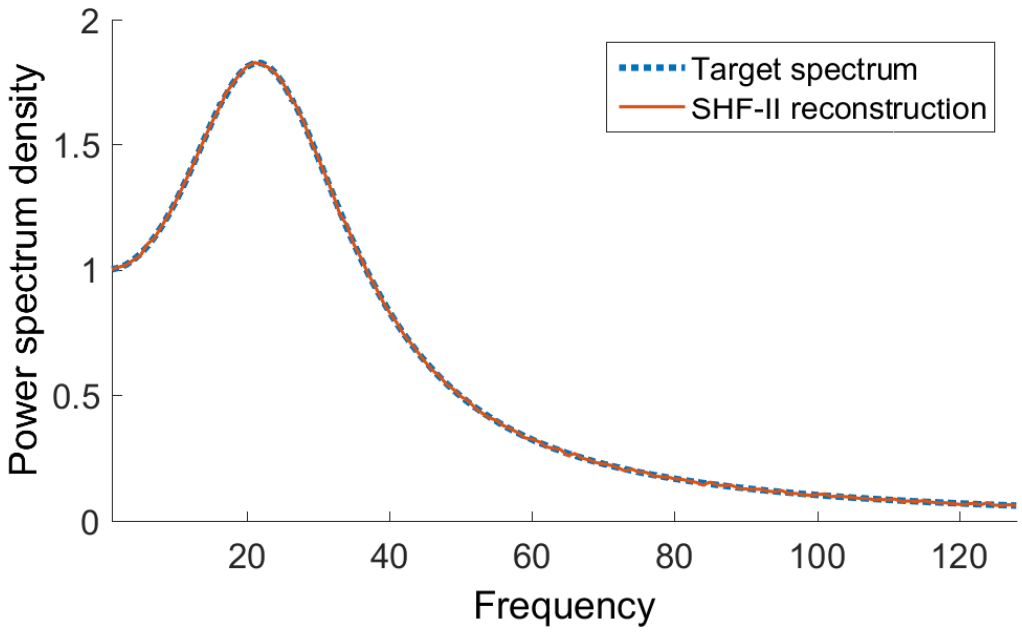


Figure 2: Target power spectrum $S^0(\omega)$ defined by Eq.43 with SHF-II reconstruction with 2% error defined by Eq.44.

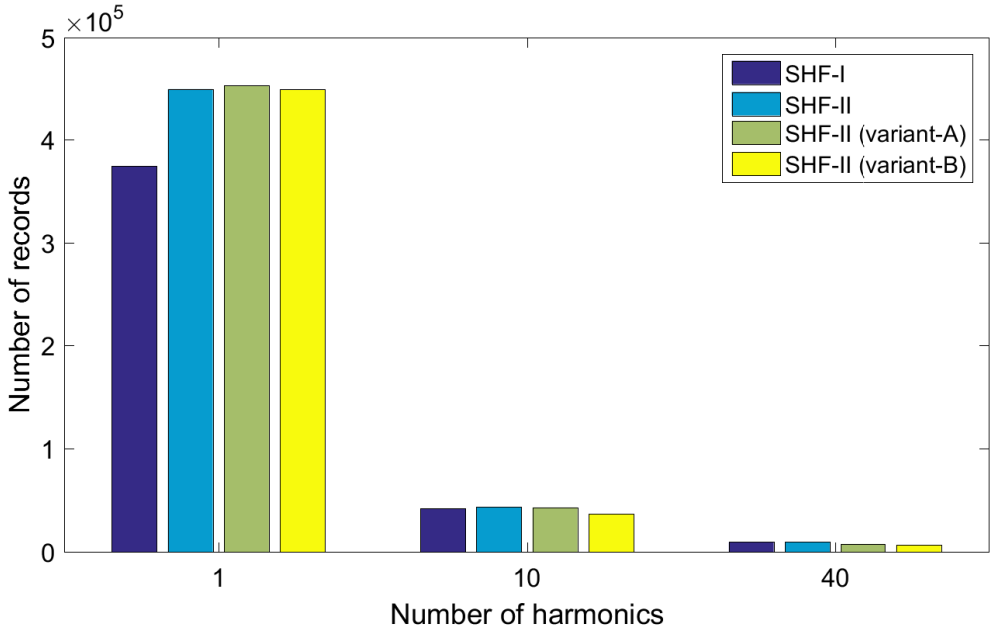


Figure 3: Number of generated process records required to achieve 2% error defined by Eq.44 for numerous SHF approaches. *Note that Variant-A refers to SHF with dependent random frequencies and Variant-B to SHF with dependent random frequencies over intervals of spectrum-relative width*

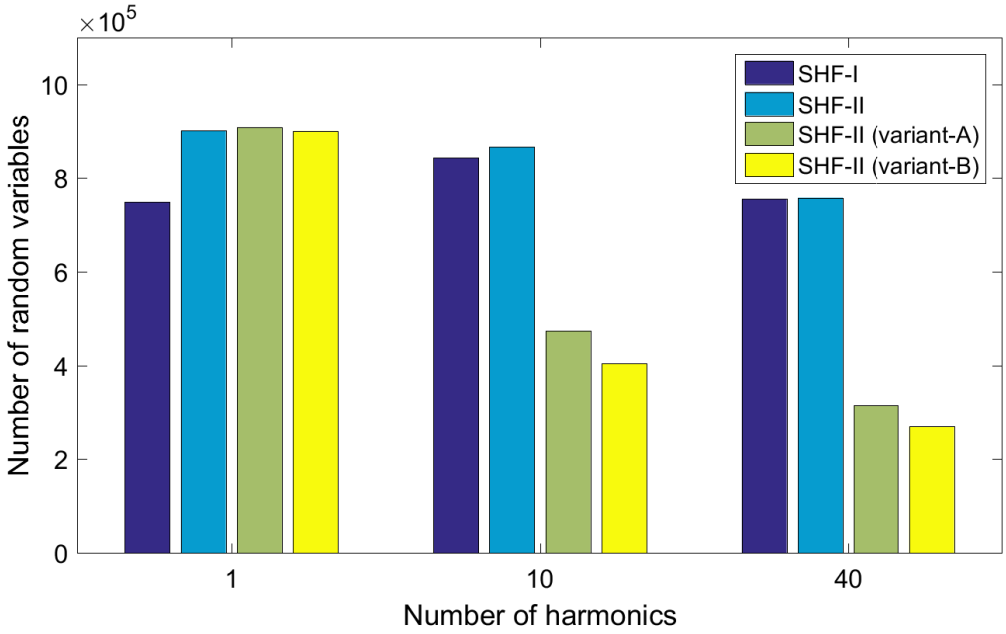


Figure 4: Number of samples of random variables (*number of records* × *number of random variables per record*) required to achieve 2% error defined by Eq.44 for numerous SHF approaches. *Note that Variant-A & B are as in Figure 3*

by number of random variables generated, in Figure 4. This presentation follows the same trend as the computational times required to generate the processes, and is highly relevant for probabilistic analyses that benefit from random variable dimension reduction such as PDEM.

Where the number of random variables required for standard SHF-I and SHF-II to meet the target accuracy remains relatively unchanged as number of harmonics increases, Figure 4 now clearly demonstrates the advantage to using a single random variable as a seed for all random frequencies (SHF-II variant-A). Further, although SHF-II variant-A is utilising the same number of random variables per record as SHF-II variant-B, the latter consistently requires fewer random variables overall (on account of fewer records) to meet the accuracy requirement for 10 and 40 harmonics.

5.1. Comparison of PDF convergence to the Gaussian

Here, the PDFs of the simulated processes are estimated though scaled histograms which are then compared to corresponding Gaussian distributions. The comparison Gaussian distributions are Normal with mean and

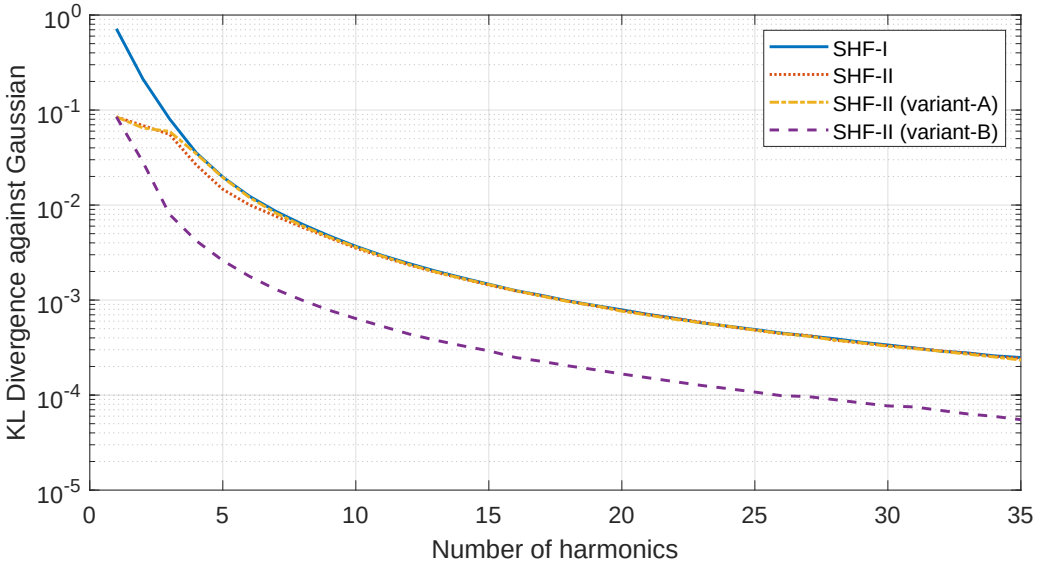


Figure 5: Convergence of the estimated PDF for a single time instant over 5×10^6 records to the Gaussian for numerous SHF approaches. *Note that Variant-A & B are as in Figure 3*

standard deviation based upon the simulated process records. Gaussianity in this case will be measured by the estimated PDF's Kullback–Leibler Divergence [19] from its comparison Gaussian distribution, i.e.,

$$D_{\mathrm{KL}}(j) = \sum_{j=1}^{M} p_{X^{\mathrm{SHF}}(t)}(x_j) \log \left(\frac{p_{X^{\mathrm{SHF}}(t)}(x_j)}{\mathcal{N}_{X^{\mathrm{SHF}}(t)}(x_j)} \right) \tag{45}$$

where $\mathcal{N}_{X^{\mathrm{SHF}}(t)} = \mathcal{N}(\mu_{X^{\mathrm{SHF}}(t)}, \sigma^2_{X^{\mathrm{SHF}}(t)})$ is the Normal distribution. In Figure 5, statistics from ensembles of 5×10^6 process records are processed for a single point in time for each of the four highlighted SHF representation methods over 35 harmonics. The high numbers of records are necessary to numerically distinguish between histogram approximations of the Gaussian PDF. This is due to the fact that in most cases, for signals with more than ~ 10 harmonics, the simulated records are already highly Gaussian.

From Figure 5 it is clear that, as hypothesised in Section 4.3, the spectrum-relative frequency intervals increased the rate of convergence to the Gaussian for SHF-II (i.e., SHF-II Variant-B has lower K-L Divergence for all frequencies). Also of note is that > 2 harmonics are required before the standard SHF-I shows similar Gaussian properties to the standard SHF-II. This is due

to the fact that the SHF-I approach favours harmonics with higher power (see Eq.10). This results in individual simulated signals with low numbers of harmonics having similar amplitude and frequency, hence, the estimated PDF is more likely to resemble that of a single harmonic. In contrast, the SHF-II is equally likely to produce high and low amplitude harmonics which, for ensembles of signals with small numbers of harmonics, skews the estimated PDF mass towards its centre (and making it appear more Gaussian).

The advantage of the SHF-II variant-B over variant-A in approximating a Gaussian process is clear, however, there is yet another compelling reason to choose this approach over variant-A when attempting to reduce the number of random variables required to effectively model a spectrum-defined process: A significant shortfall of the SHF-II variant-A is identified when estimating the simulated process PDF over a single, long record, rather than over an ensemble at a single time point. Figure 6 shows the K-L Divergence against the optimal Gaussian distributions for single simulated records. Due to low frequency tendencies in the signals, more data was needed to estimate the PDF than in the ensemble case, hence signals of length 200×10^6 were simulated. Further, the resulting K-L Divergence was calculated for 10 signals for every data point, with Figure 6 displaying the mean. This has the effect of smoothing the result as some simulated signals could be more Gaussian than others. SHF-II variant-A is significantly less Gaussian than all alternatives. This is due to the fact that as the frequency intervals are constant, and all of the random frequencies are generated from the same random variable, the frequency distance between adjacent harmonics are equal and thus generated processes are likely to be near-periodic over a small time window (far less than the total record length). High levels of repetition in the individual record inhibit efficient convergence to Gaussian statistics over time. This periodicity can be clearly seen when comparing individual simulated time records for SHF-II variant-A and variant-B in Figure 7.

6. Conclusion

An overview of classical spectral representation and SHF schemes is provided alongside two novel extensions of said schemes aiming to not only use fewer random variables, but also to more efficiently represent Gaussian processes. Numerical examples demonstrating spectrum reconstruction accuracy and estimated PDF convergence to the Gaussian support the arguments presented in Section 4. The advantages of enhancing the SHF-II representation

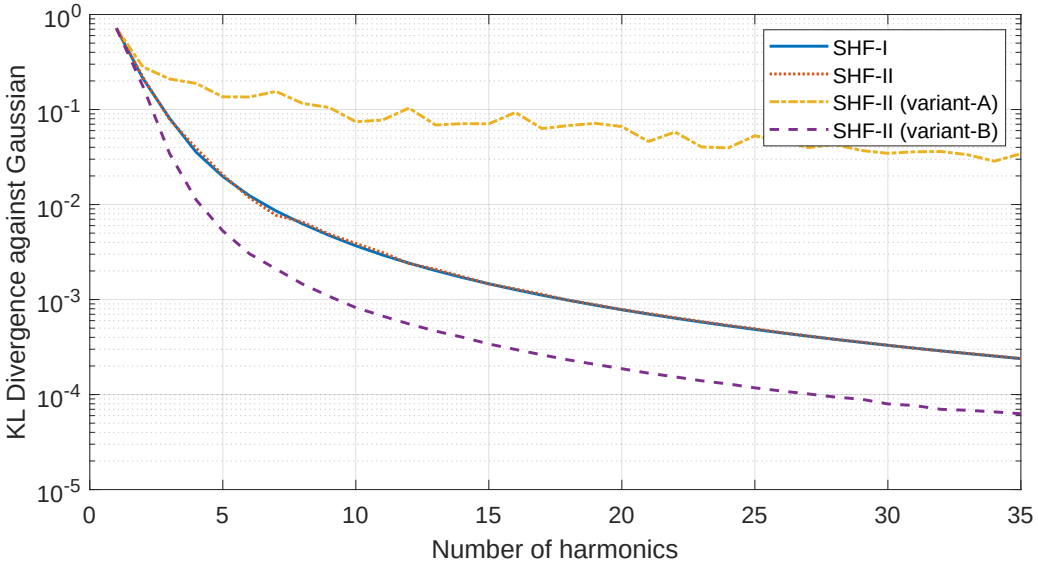


Figure 6: Convergence of the estimated PDF for a single record over 200×10^6 time points to the Gaussian for numerous SHF approaches. *Note that Variant-A & B are as in Figure 3*

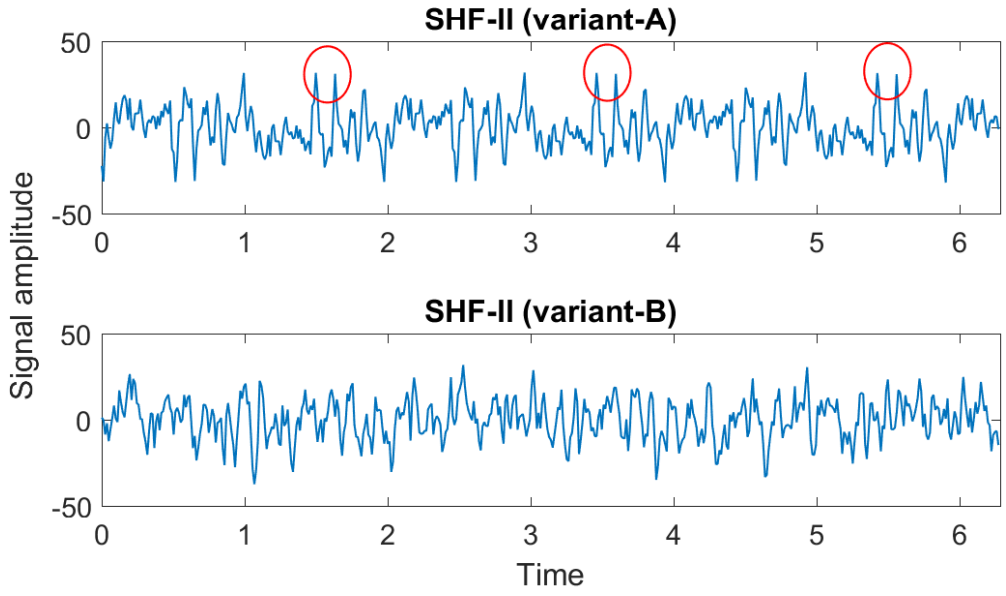


Figure 7: Example 500-point time histories simulated from Eq.43 via SHF-II variant-A and variant-B with 40 harmonics. The red circles draw attention to periodicity in the signal. *Variant-A & B are as in Figure 3*

through random frequency dependency and optimal selection of frequency intervals are clearly presented.

Of particular interest would also be the effect of such frequency selection schemes when applied with SHF for representation of non-stationary processes and is suggested as a potential future investigation.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) and National Natural Science Foundation of China (NSFC) under the Sino-German research project: BE 2570/4-1, CO 1849/1-1 (DFG) 11761131014 (NSFC). The work was additionally supported by NSFC projects: 51725804, 11672209 & 51678450.

References

[1] J. B. Roberts, P. D. Spanos, Random vibration and statistical linearization, Courier Corporation, 2003.

[2] P. D. Spanos, I. A. Kougioumtzoglou, Harmonic wavelets based statistical linearization for response evolutionary power spectrum determination, Probabilistic Engineering Mechanics 27 (1) (2012) 57–68.

[3] I. A. Kougioumtzoglou, P. D. Spanos, An analytical wiener path integral technique for non-stationary response determination of nonlinear oscillators, Probabilistic Engineering Mechanics 28 (2012) 125–131, pT: J; SI: SI; CT: 6th International Conference on Computational Stochastic Mechanics (CSM); CY: JUN 13-16, 2010; CL: Rhodes, GREECE; TC: 0; UT: WOS:000301561800016.

[4] I. A. Kougioumtzoglou, P. D. Spanos, Response and first-passage statistics of nonlinear oscillators via a numerical path integral approach, Journal of Engineering Mechanics 139 (9) (2012) 1207–1217.

[5] A. Di Matteo, I. A. Kougioumtzoglou, A. Pirrotta, P. D. Spanos, M. Di Paola, Stochastic response determination of nonlinear oscillators with fractional derivatives elements via the wiener path integral, Probabilistic Engineering Mechanics 38 (2014) 127–135.

[6] A. F. Psaros, I. A. Kougioumtzoglou, I. Petromichelakis, Sparse representations and compressive sampling for enhancing the computational efficiency of the wiener path integral technique, Mechanical Systems and Signal Processing 111 (2018) 87–101.

[7] J. Li, J. Chen, Probability density evolution method for dynamic response analysis of structures with uncertain parameters, Computational Mechanics 34 (5) (2004) 400–409.

[8] J. Li, J. Chen, W. Sun, Y. Peng, Advances of the probability density evolution method for nonlinear stochastic systems, Probabilistic Engineering Mechanics 28 (2012) 132–142.

[9] J. Li, J. Chen, The principle of preservation of probability and the generalized density evolution equation, Structural Safety 30 (1) (2008) 65–77.

[10] J. Chen, W. Sun, J. Li, J. Xu, Stochastic harmonic function representation of stochastic processes, Journal of Applied Mechanics 80 (1) (2013) 011001.

[11] G. Deodatis, M. Shinozuka, Simulation of stochastic processes by spectral representation, American Society of Mechanical Engineers 44 (4) (1991) 191–204.

[12] M. Shinozuka, Simulation of multivariate and multidimensional random processes, The Journal of the Acoustical Society of America 49 (1B) (1971) 357–368.

[13] M. Shinozuka, C.-M. Jan, Digital simulation of random processes and its applications, Journal of sound and vibration 25 (1) (1972) 111–128.

[14] J. Chen, F. Kong, Y. Peng, A stochastic harmonic function representation for non-stationary stochastic processes, Mechanical Systems and Signal Processing 96 (2017) 31–44.

[15] M. Shinozuka, G. Deodatis, Simulation of stochastic processes by spectral representation, Applied Mechanics Reviews 44 (4) (1991) 191–204.

[16] P. Billingsley, Probability and measure, John Wiley & Sons, 2008.

[17] R. B. Ash, C. Doleans-Dade, Probability and measure theory, Academic Press, 2000.

[18] M. Kuczma, An introduction to the theory of functional equations and inequalities: Cauchy's equation and Jensen's inequality, Springer Science & Business Media, 2009.

[19] S. Kullback, R. A. Leibler, On information and sufficiency, The annals of mathematical statistics 22 (1) (1951) 79–86.