

Maxima estimate of non gaussian process from observation of time history samples

L. Borsoi

Framatome, Département EEC, Paris la Défense, France

1. STATEMENT OF THE PROBLEM

Structural engineers have not a lot of resources when they have to analyse nonlinear structures dynamically excited by stochastic processes. The first idea is obviously to make linear the mathematical model. But it is not always possible nor at least satisfactory especially when the nonlinearities are due to the presence of stops with gaps, as it is the case in this paper.

The second idea, the most consistent, is to adopt a complete random vibration approach. This is well known and common for linear structures excited by gaussian processes. Unfortunately it is drastically more complicated for nonlinear structures even if the excitation is gaussian. One possible way is based on the Fokker-Planck-Kolmogorov equation which can be derived for markovian processes. But, on the one hand, engineers are not very familiar with this equation they consider too mathematical and on the other hand practical industrial applications are limited by the required computational effort. However as super-computer capability is blowing up this approach could become tractable in a next future.

Finally it appears that the only way, presently practicable, consists in performing time history simulations. For example, let us consider the event E "Earthquake" modeled by a nonstationary stochastic process W. W can be defined in the time domain by an ensemble of sample functions $\Gamma_i(t)$ - ground acceleration versus time - which all present the same basic characteristics: envelope shape (including the duration T), frequency content, ... Each acceleration sample function $\Gamma_i(t)$, also called here realization of event E, provides via nonlinear calculations a realization of the structure response $X_i(t)$. This step does not set special problems since nonlinear time history computations are now widely used and numerically well controlled. For design considerations, the Analyst is often interested in the "maximum" value of the response. In fact this has no meaning if not connected to a duration T and a probability of non exceedance Prob_{NE} . For illustration: the design maximum value $X_{max}(T=15.s, \text{Prob}_{NE}=0.9)$ means that the value X_{max} will be exceeded only in one out ten realizations of 15 seconds.

For extracting the design maximum value X_{max} through simulations, the Analyst could proceed as following:

- 1/ To generate a great number of time history excitations from the stochastic process W,
- 2/ To calculate the corresponding time history responses of the nonlinear structure and note the observed maxima - one maximum per

realization - ,

3/ To perform statistics on these maxima in order to carry out the design value X_{max} .

This rough but correct approach is often unacceptable for economical reasons. Thus the actual problem for the Analyst can be resumed in:

1/ To minimize the number of time histories, or the duration of one single time history for ergodic processes,

2/ To capture all available informations from these time histories (more than only the observed maximum),

3/ To estimate from all these informations the design maximum value related to a duration and a probability of non exceedance.

This paper deals with this topic closely connected to the first passage problem. The studied structures are simple SDOF structures but highly nonlinear due to presence of stops with gaps. The focused processes - excitation and response - are stationary with zero mean value. The excitation is a gaussian and wide band process (physical white noise with a cut off frequency). But obviously the response is not: the nonlinearities destroy the gaussian nature, moreover the structure makes the response process rather narrow band.

Although attention is focused here on non gaussian process, an application of Vanmarcke formulation (involving gaussian process) is first presented in order to illustrate the difficulties of the stated problem.

2. STATIONARY GAUSSIAN PROCESS

In order to introduce later used notations the first passage problem (with 2 symmetrical barriers) is briefly recalled here .

Let $X(t)$ a time history response of a component with zero mean value and $+a, -a$ a double barrier. For simplicity it is assumed that at the initial time $t=0$, $X(0)$ is always found between the double barrier which defines the safe domain. The first passage problem is to know how long time $X(t)$ is contained in the safe domain. In other words after how many times will occur the first passage of barrier (a) ? Of course this problem is set in probability terms.

A completely consistent mathematical solution has not yet been found even for the gaussian case. However relatively satisfactory methods exist, which involve physical assumptions like sometimes semi-empirical coefficients for a better adjustment of simulation results.

The oldest idea is to assume that the barrier (a) crossings occur according to a Poisson model. This implies mainly:

1/ the (a) crossings are independent one from each other;

2/ the (a) crossing rate presents an average value ν_a ,

(with $\nu_a = 2 \nu_{+a} = 2 \nu_{-a}$ for symmetry reasons; only safe to unsafe crossings are taken into account).

The probability not to exceed barrier (a) during the period $(0, T)$ is :

$$\text{Prob}_{NE}(a, T) = \exp(-\nu_a T) \quad (1)$$

The design problem is posed in inverse terms: what is the "maximum" design value (a) knowing the duration T and the probability Prob_{NE} ?

It is well known that the Poisson assumption is justified for high level barriers - the derived solution is asymptotically exact -, but this assumption is generally too poor for barriers of practical interest. As the crossings tend to occur in clumps especially for narrow band processes, the independence condition is not well respected. In order to

take into account this effect and others (for example time of excursion in unsafe domain) Vanmarcke replaces ν_a in equation (1) by a more sophisticated coefficient α which depends on barrier level (a) and on a measure of process bandwidth δ (ref.1). The formulation is available for stationary gaussian processes: it involves the Rayleigh nature of gaussian process envelopes, and the crossing frequency ν_a is connected to the "average" frequency ν_0 according to the classical exponential law :

$$\nu_a = \nu_0 \exp(-r^2/2) \quad (2)$$

with $r=a/\sigma$ the peak factor, σ the process standard deviation. The unknown quantities σ , ν_0 , δ are derived from the three first moments of the one side Power Spectral Density (PSD) of the process. It is recognized that the Vanmarcke formulation works quite well for process whose PSD is simple: narrow band - one peak PSD - or wide band process (for 2 peaks PSD it appears less efficient due to average frequency concept). However its application to solve the focused design problem is not evident.

For example, let us consider a SDOF whose frequency is $f=5$.hz and percentage of critical damping $\xi=0.04$. The oscillator base is subjected to a white noise gaussian acceleration. Design value is related to a stationary duration $T=20$.seconds. (The transient part of the response is removed). This example is representative of earthquake engineering excluding the nonstationary aspect.

The response process is ergodic, that simplifies the work of the Analyst who can be satisfied with running a single time history. For mathematicians, ergodicity states that a infinite time history provides all the characteristics of the process. But this definition is "rejected" by Engineers who want reduce as more as possible time history duration.

Let us focus on the first and simplest PSD moment i.e. the variance. The time step of seismic time history is usually $\Delta t=0.01$ s. Thus a 20.s time history carries out 2000 response values, but of course not all independent each other. The number of available informations contained in the 20.s time history is not connected to the time step Δt but to a certain duration T_I which ensures the independence of two response values $X(t)$ and $X(t+T_I)$. T_I can be derived from the autocorrelation function $R_{XX}(\tau)$ of the process. In the SDOF case:

$$T_I = \beta (1/\xi f) \quad (3)$$

where β depends on the wanted "independence" level ($R_{XX}(\tau)=0.1$ if $\beta=0.37$ N.A. $T_I=1.85$ sec). Roughly speaking it means that a 20.s time history gives only 11 (20./1.85) informations actually independent for process variance estimation. This is too poor and is confirmed by simulations which show a great dispersion on variance estimates each based on a 20.s run. (For 25 simulations σ is found between 0.68 and 0.94 with 0.80 average ; this relative dispersion is unacceptable since it is translated to maxima for constant peak factors).

Therefore one has to prolong time history duration. A sufficient length can be determined from convergence tests or from simple probabilistic arguments: let σ the unknown process standard deviation and N the number of independent informations $N=T/T_I$, the σ^2 estimate is a gaussian random variable with a mean σ^2 and a variance close to σ^4/N .

An other way is used here with the help of periodic stochastic processes. The excitation white noise is built by addition of sinus

functions with initial phase angles randomly and equally distributed. As each sinus frequency is an integrate multiple of an elementary frequency Δf , the excitation process, and consequently the response process, are periodic (period $T=1/\Delta f$). Due to joint properties of ergodicity and periodicity, it is clear that the observation of a single time history on a complete period should be sufficient to provide reliable estimates of the process characteristics.

Results of simulations are shown figure 1. (100 runs - $\Delta f=0.05$ hz \Rightarrow $T=20$.s). The figure gathers the histogram and the Cumulative Distribution Function (CDF) of the 100 observed maxima - one per time history run -, and the 100 maxima predictions based on the Vanmarcke formula. Each maximum is estimated from one single time history for a 20.s duration and three probability levels (0.5,0.9,0.99).

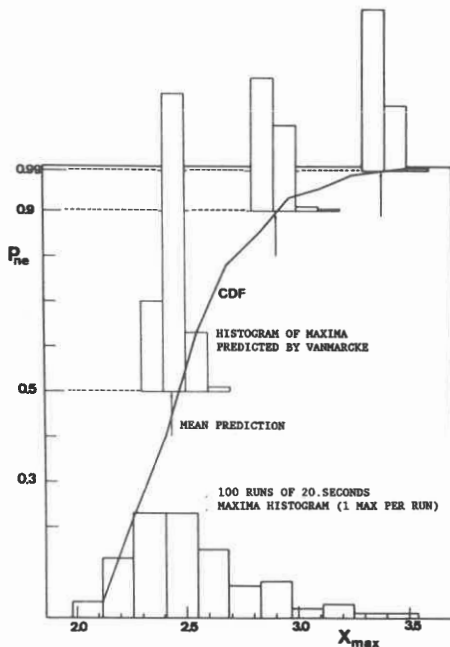


Figure 1: Linear SDOF Simulation

3. THE GUMBLE-POISSON MODEL

When confronted to nonlinear structures which destroy the gaussian nature of excitation, the Analyst must change his strategy generally in performing extensive time history simulations. However it appears interesting to adjust from limited simulation a semi-empirical probabilistic model which is more tractable.

For this purpose the simplest way is to come back to the Poisson model (eq.1). The problem consists in the determination of crossing rates ν_a which have no longer an analytical formulation like in equation 2 (in most cases the joint probability density function $P(X, \dot{X}) - \dot{X}$ for velocity - is unknown, that prevents the use of Rice formula).

The different frequencies ν_a can be estimated by counting the crossing events. But the more the barrier (a) is high, the less crossings occur and the less estimate is reliable. Thus it seems better to consider an observation threshold (b), neither too high for previously reasons, nor

too low for respecting the Poisson independence assumption, and then estimate ν_b by counting. The distribution of peaks above threshold (b) can often be approached by a Gumble law - Type 1 - which is currently used in extreme value statistics:

$$F_b(x) = \exp(-\exp(-\alpha(x-u))) \quad (4)$$

for the CDF. u is the mode of distribution and α a coefficient of dispersion. Both u and α depend on the threshold (b) for a given process.

All frequencies ν_a are now connected to ν_b by:

$$\nu_a = \nu_b (1 - F_b(a)) \quad (5)$$

Introducing (5) into (1) gives a "Gumble-Poisson" model entirely defined by the 3 parameters ν_b , u and α .

Let N the number of (b) crossings, and X_j ($j=1, N$) the levels of peaks observed above threshold (b) during the time T_N . The 2 parameters u and α can be deduced from the peaks estimated average m_N , the peaks estimated standard deviation σ_N , and the Euler constant γ according to the moment method (ref. 2):

$$\alpha_N = \pi / \sqrt{6} \sigma_N \quad \text{and} \quad u_N = m_N - \gamma / \alpha_N \quad (6)$$

Parameters, especially the more sensitive α , can also be estimated through a Bayesian technique. Doing this, the reference 3 approach is applied: a hierarchy of the different maxima X_j is erected, X_1 is the largest maximum, X_2 the second one, ...

For simplicity let the 2 parameters u and ν_b be fixed. The α likelihood function related to the observation of the M largest maxima ($X_j, j=1, M$) is:

$$L_M(\alpha/X_1, X_2, \dots, X_M) = \prod P_j(X_j, \alpha) \quad (7)$$

where $P_j(X_j, \alpha)$ is the probability to observe X_j as the j^{th} largest maximum, for α given. The Probability Density Function (PDF) of α involving the observation of the M largest maxima $Pr_M(\alpha)$ is thus updated with respect to the previous $Pr_{M-1}(\alpha)$ according to the Bayes theorem:

$$Pr_M(\alpha) = C \cdot L_M(\alpha/X_1, X_2, \dots, X_M) \cdot Pr_{M-1}(\alpha) \quad (8)$$

where C is a normalization constant.

The PDF $P_j(x)$ related to the j^{th} maximum is deduced by the derivation of the corresponding CDF $Q_j(x)$. If the $(j+1)^{\text{th}}$ largest maximum is lower than or equal to x (event A), it clearly implies that either the j^{th} largest maximum is also lower or equal to x (event B), or there are exactly j crossings of the barrier x (event C). Let $Q_{Dj}(x)$ the probability of the event C; in probability terms:

$$A = B \cup C \quad \text{and} \quad B \cap C = \emptyset \Rightarrow Q_{j+1}(x) = Q_j(x) + Q_{Dj}(x) \quad (9)$$

The Poisson model gives classically:

$$Q_{Dj}(x) = \frac{(1 - F_b(x))^j}{j!} (\nu_b T)^j \exp(-\nu_b T(1 - F_b(x))) \quad (10)$$

Incidentally equation 10 is available even for $j=0$ and gives the initial CDF $Q_1(x)$. The derivation with respect to x of the recurrent equation 9 on CDF provides a recurrent equation on PDF which is not developed here due to limited space:

$$P_{j+1}(x_{j+1}) = P_j(x_j) + dQ_{Dj}(x)/dx \quad (11)$$

Figure 2 shows an example of PDF hierarchy given by eq.11. It must be noted that the PDF become more and more peaked. Due to the adopted mathematical model, it is clear that eq.11 is not more correct as soon as the PDF significantly cross the observation threshold. This can easily be detected by calculating the PDF integral which is not longer equal to 1.

Simulations are made with a nonlinear SDOF (figure 3) excited by a gaussian white noise. Figure 4 gathers rough maxima CDF (one maximum per realization of 20.s) and predictions for a 20.s duration given by the Gumble-Poisson model from two observation thresholds. Of course the results are better for the highest threshold ($b=1.50$) than for the other one ($b=1.25$). Nevertheless, even in such limit case which does not respect the independence assumption, the Gumble-Poisson results are not so bad.

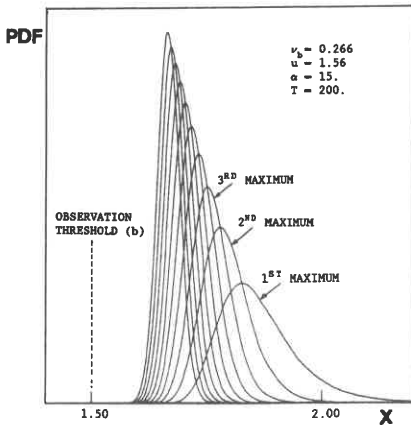


Fig.2: Maxima PDF Hierarchy

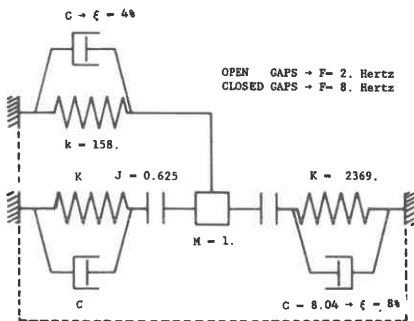


Fig.3: Nonlinear SDOF

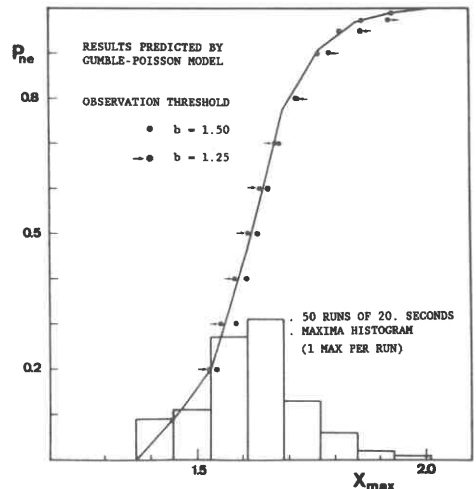


Fig.4: Nonlinear SDOF Simulation

Unfortunately figure 4 shows only the consistency of the model since the 3 parameters ν_b , u and α are estimated from all simulations. It must be recognized that the stated problem is not presently solved in a convincing way. Indeed the quality of parameters estimates is ameliorated using a Bayesian technique, but is not sufficient to make reliable maxima predictions from a reduced time history simulation as it is wanted (for example 50 seconds).

4. CONCLUSION

The problem stated here about the prediction of maxima from time history samples constitutes a formidable task but is essential for industrial applications: extreme value design, fatigue analysis,...

Even for the linear gaussian case, the process ergodicity does not prevent the observation duration to be long enough to make reliable estimates. As well known, this duration is closely related to the process autocorrelation. A subterfuge, which distorts a little the problem, consists in considering periodic random process and in adjusting the observation duration to a complete period.

In the nonlinear case, the stated problem is as much important as time history simulation is presently the only practicable way for analysing structures. Thus it is always interesting to adjust a tractable model to rough time history observations. In some cases this can be done with a Gumble-Poisson model. Then the difficulty is to make reliable estimates of the 3 parameters ν_b , u , α involved in the model (especially the more sensitive α). Unfortunately it seems that even the use of sophisticated Bayesian method does not permit to reduce as wanted the necessary observation duration.

One of the difficulties lies in process ergodicity which is often assumed to be based on physical considerations but which is not always rigorously stated. An other difficulty is the confusion between hidden informations - which can be extracted - and missing informations - which cannot be extracted -.

Finally it must be recalled that the obligation of considering time histories long enough is not always embarrassing due to the current computer cost reduction.

REFERENCE

- (1) E.H. VANMARCKE: Structural Response to Earthquakes. Ch. 8 in Seismic Risk and Engineering Decisions. C. Lomnitz & E. Rosenblueth Eds. Elsevier.
- (2) A.H.S. ANG, W.H. TANG: Probability Concepts in Engineering Planning and Design. John Wiley & Sons.
- (3) A. PREUMONT: Estimate of the most probable value of the peak factor of a stationary random process. Private Communication.