Soil Dynamics and Earthquake Engineering 29 (2009) 742–751

Contents lists available at ScienceDirect

Soil Dynamics and Earthquake Engineering

journal homepage: www.elsevier.com/locate/soildyn

Wavelet approximation of earthquake strong ground motion-goodness of fit for a database in terms of predicting nonlinear structural response

Maria I. Todorovska *, Hadi Meidani, Mihailo D. Trifunac

Department of Civil Engineering, University Southern California, Los Angeles, CA 90089-2531, USA

ARTICLE INFO

Article history:
Received 6 December 2007
Received in revised form
27 July 2008
Accepted 3 August 2008

Keywords:
Earthquake strong ground motion
Wavelets
Wavelet approximation
Nonlinear structural response
Energy

ABSTRACT

Reduced dimensionality representation of strong ground motion records as a superposition of a relatively small number of pulses is studied. Such representation is obtained by the expansion of velocity in orthogonal wavelet series using the Fast Wavelet Transform, and approximation by only the largest energy terms in the series. The Coiflet 5 wavelet family is used, which is orthogonal, smooth and nearly symmetric. The goodness of the approximation is examined for the EQUINFOS for USA Part I database, as representative of a large variety of strong motion records (it consists of 494 three-component records from 106 earthquakes recorded in the western US between 1933 and 1984). The goodness of fit is measured in terms of closeness of predicting several input and output characteristics of a nonlinear oscillator representing a structure: (i) energy of the input ground motion (proportional to integral of velocity squared), (ii) the peak input power, and (iii) the time of collapse of a bi-linear oscillator (considering also collapse due to dynamic instability). The results show very high degree of correlation of these characteristics as estimated from the actual record and from its approximations, even for small number of pulses (relative to the number that would represent the ground velocity exactly). Such reduced representation of strong ground motion is useful for extracting such pulses from strong motion records to study their nature, and for development of new algorithms for the synthesis of artificial earthquake strong motion records.

1. Introduction

The linear response of structures to earthquake shaking can be estimated from a reduced representation of the input ground motion, consisting only of the response spectrum ordinates at the modal frequencies. The spectral values, however, are not sufficient for estimation of nonlinear response, which is sensitive to the time-dependent *time rate* of the seismic input energy. An analogous reduced representation of the ground motion, such that preserves the time information, can be in terms of a *small number of pulses*. Such a representation can be motivated, for example, by the asperity model of the earthquake source, per which accumulated seismic energy is released via radiation from highly stressed patches (asperities) on the fault surface radiating pulses of energy. The ground motion at a site can be viewed as a sum of these pulses, delayed at the source relative to each other, and modified by attenuation, dispersion, scattering and diffraction along the propagation path. The sum of the largest energy pulses would represent an *approximation* of the actual ground motion.

In this paper, we explore representation of strong ground motion in a relatively small number of wavelets of the Coiflet 5 orthogonal family, which we use as proxies of the pulses that have arrived at the site. In particular, we examine the goodness of the approximation, as a function of the level of the approximation, in terms of the ability to predict the response of a nonlinear oscillator representing a structure. The level of the approximation is defined by the number of pulses used in the approximation, as fraction of the number of pulses that would represent the ground motion exactly, which is equal to the number of samples in a critically sampled record. The goodness of fit is defined by the closeness of: (i) the total energy of the input seismic motion (defined in terms of the L_2 norm of the ground velocity), (ii) the peak of the input power, and (iii) the time of collapse of the oscillator. As a trial set representative of a variety of strong motion records, we use all records of the EQINFOS for USA Part I database [1], which consists of 494 three-component strong motion records of 106 earthquakes between 1933 and 1984.

The objective of this work is to gain insight in the ways of approximating the strong motion time series that could be useful

* Corresponding author.
E-mail addresses: mtodorov@usc.edu, todorovska@usc.edu (M.I. Todorovska).

0267-7261/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.soildyn.2008.08.001

for a synthesis of artificial strong ground motion using wavelets. In this paper, we examine only what is the number of wavelets required to approximate strong motion and how this can be done to satisfy the elementary engineering aspects of nonlinear and collapsing response. The development of empirical scaling equations for prediction of the number, amplitudes and arrival times of the pulses is well beyond the scope of this paper.

Despite their popularity, in many applications in earthquake engineering and strong motion seismology, wavelet analysis has turned out to be just an alternative method, not necessarily superior to the methods used previously (see e.g. Todorovska and Trifunac [17]). In that sense, the existence of *bases* of functions of wavelets, which can be moreover *orthogonal*, and the *efficiency* of such bases to represent transient signals (in the sense that very good approximation is obtained with a small number of terms, as compared to Fourier series), is an unmatched combination of properties by which wavelet analysis is superior to other representations. In digital signal and image processing, the latter property is referred to as *data compression* and the method of preserving only the highest energy terms in the series (with "top" coefficients) is referred as *thresholding*. Wavelet representation of strong ground motion records, as well as of structural response, has been considered earlier, e.g. by Iyama and Kuwamura [2], Mukherjee and Gupta [3], Iyama [4]. This paper differs from the previous studies in: (i) the measures of goodness of fit (which include peak power of the incident motion, and prediction of collapse of a nonlinear oscillator representing structures, including failure due to dynamic instability), (ii) the trial set used (which is a comprehensive database representing a broad variety of strong motion records), and (iii) the wavelet family used (most commonly used wavelets in earthquake engineering are the real or complex harmonic wavelet, or the d4 Daubechies family).

In this paper, we first present the methodology, in which we (i) review briefly the discrete wavelet transform and expansion of signals in wavelet series and (ii) describe the model of a nonlinear oscillator (with dynamic instability) that we use. Then, we illustrate detailed results of the goodness of fit for the S00E component of the El Centro record, and statistical results for the entire database in terms of correlation coefficient. Finally, a discussion and the conclusions are presented.

2. Theoretical background

2.1. Wavelet series representation of discrete time signals

A level J dyadic wavelet basis expansion of a discrete time signal $s[n]$ is

$$s[n] = \sum_{j=1}^{J} \sum_{k=1}^{N/2^j} d_{j,k}\psi_{j,k}[n] + \sum_{k=1}^{N/2^J} s_{J,k}\varphi_{J,k}[n] \tag{1}$$

where $\psi_{j,k}[n]$, $j = 1,\ldots,J$, are the wavelet basis functions (zero mean), and $\varphi_{J,k}[n]$ are the scaling functions (nonzero mean) in which the remainder is expanded. The coefficients of expansion constitute the discrete wavelet transform of the signal, and represent orthogonal projections of the signal on the corresponding wavelet/scaling function

$$\begin{aligned} d_{j,k} &= \langle \psi_{j,k}, s \rangle \\ s_{J,k} &= \langle \varphi_{J,k}, s \rangle \end{aligned} \tag{2}$$

In Eqs. (1) and (2), index j is related to the scale variable of the transform (inversely proportional to frequency), and k is related to the time variable. For a basis of compactly supported wavelets, the coefficients of the expansion are computed by the pyramid algorithm (due to [5]), which consists of splitting the original signal in a low- and high-frequency component followed by downsampling by a factor of two, and further recursively splitting the lower frequency component, J times total. In each splitting, the low-frequency component is a lower-level resolution approximation, and the high-frequency component contains the detail of the signal that was removed. Hence, the wavelet expansion is nothing else but splitting the signal in subbands, and expanding each subband in a series of wavelet functions, which are shifts of one another, and all have central frequency corresponding to the one of the subband. Larger j corresponds to a lower-frequency subband, and k describes the time shift of the wavelet. The subbands have equal width on a logarithmic frequency axis, while on a linear frequency axis the width is progressively smaller, by a factor of two, with increasing j. The pyramid algorithm, also called Fast Wavelet Transform, is shown graphically in Fig. 1. The top part shows the analysis part, or the forward transform, and the bottom part shows the synthesis part, or the inverse transform.

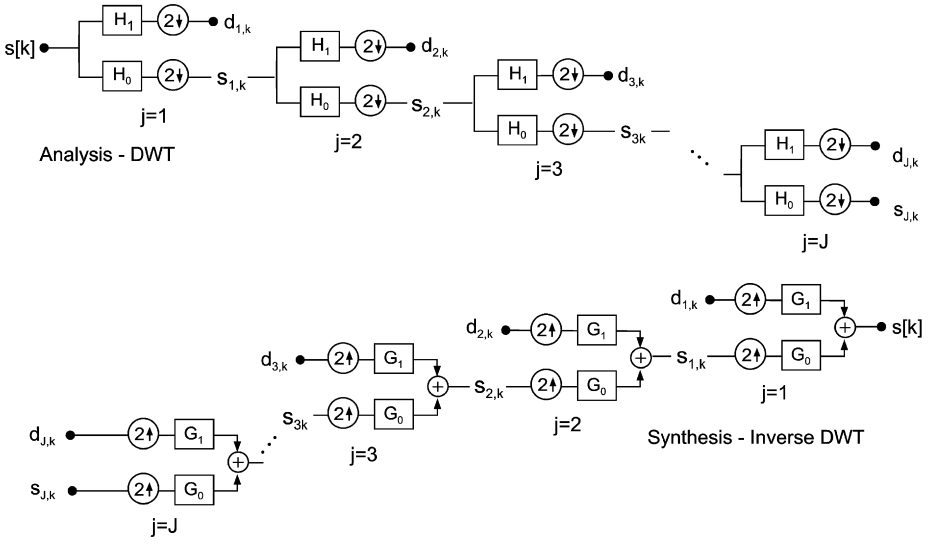

Fig. 1. The forward (top) and inverse (bottom) pyramid algorithms for Fast Wavelet Transform.

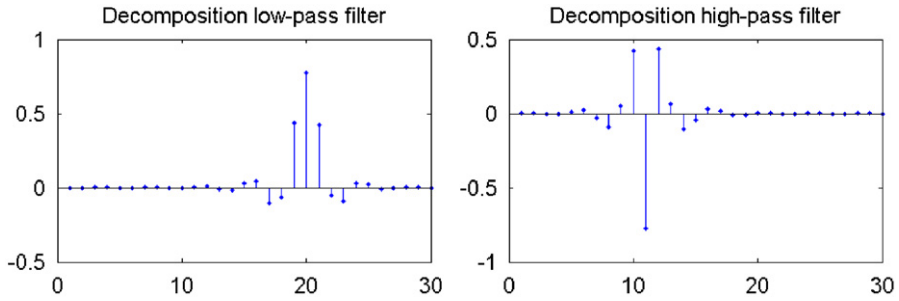

Fig. 2. Decomposition low-pass (left) and high-pass (right) filters for the discrete time Coiflet 5 wavelet shown in the time domain. The units of the *x*-axis are samples.

An important advantage of the orthogonal wavelet bases is the convenience to represent the energy of the signal, $\|s\|^2$, which by the Parseval's equality is

$$\|s\|^2 = \sum_{n=-\infty}^{\infty} |s[n]|^2 = \sum_{j=1}^{J} \sum_{k=1}^{N/2^j} |d_{j,k}|^2 + \sum_{k=1}^{N/2^J} |s_{J,k}|^2 \tag{3}$$

For a high-pass filtered signal, like an earthquake accelerogram, and sufficiently many subbands (large J), the last low-frequency subband contains practically no energy, and the coefficients of expansion $s_{J,k} \approx 0$.

A full expansion consists of N terms, which exactly represent the signal. A subset of these terms represents a lower dimensionality approximation. Dropping all coefficients of the higher frequency subbands creates a low-frequency approximation. The dimensionality of the signal can also be reduced, while preserving the significant high-frequency information, by approximation by thresholding, which consists essentially of dropping the low-energy terms (those with small $|d_{j,k}|$; hard thresholding), and possibly modifying the smallest remaining terms to insure a smooth transition (soft thresholding). In this paper, we apply hard thresholding, i.e. we approximate the signal by the highest energy ("top") coefficients. More on the wavelet transform in signal processing can be found in various textbooks on wavelets, e.g. in Vetterli and Kovačević [6], and on different applications to earthquake records in Todorovska [7], Todorovska and Hao [8].

2.2. Selection of wavelet basis for this study

Our objective is to find an approximation of a strong motion record by a relatively small number of pulses (i.e. wavelets in an orthogonal wavelet family) such that: (1) would produce a reasonably good visual fit for both acceleration, velocity and displacement, and (2) would predict closely the response of a nonlinear oscillator. To satisfy the first requirement, we searched among Coiflet wavelets [9], which are orthogonal, nearly symmetric, and relatively smooth. The orthogonality property is convenient for the evaluation of the energy directly in the wavelet transform domain (see Eq. (3)), the symmetry is desirable to reduce phase distortions in the approximation, and the smoothness is desirable because it helps achieve better fit for strong motion records with smaller number of wavelets. We examined the fit for several representative records, and for thresholding applied either on acceleration, velocity or displacement. Based on the visual fit, we chose the Coiflet 5 family, and thresholding applied on the velocity signal. Fig. 2 shows the decomposition low-pass (left) and high-pass (right) filters for the prototype Coiflet 5 wavelet.

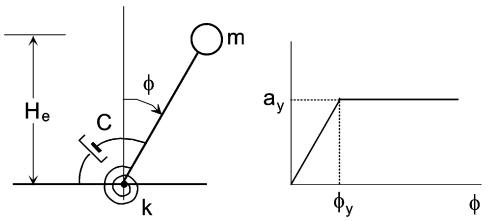

Fig. 3. Nonlinear oscillator.

2.3. Measures of goodness of fit in terms of nonlinear oscillator response

As measures of goodness of the approximation, we compare the energy and peak power of the input ground motion from the exact signal and form the approximation, and the corresponding times of collapse of a nonlinear oscillator excited by such motions. These quantities were estimated as follows.

The energy of the seismic wave motion at a site is proportional to the integral over time of velocity squared [10]. Hence, as measure of the energy and power of the input motion, we compute

$$\mathrm{en}(t_0) = \int_0^{t_0} v(t)^2 \, \mathrm{d}t \tag{4}$$

and

$$p(t) = \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{en}(t) \tag{5}$$

where $v(t)$ is the ground velocity, and t_0 is the duration of shaking. We consider the power of strong ground motion because it controls the destructiveness of nonlinear waves which propagate into the building [11,12].

The equivalent single-degree-of-freedom (SDOF) nonlinear oscillator representing a structure is an inverted pendulum with mass m_b and height H_e, connected to the base by a rotational spring with stiffness k and a rotational dashpot with damping C connected in parallel, as shown in Fig. 3 (left). The parameters of the oscillator are defined assuming that it represents an N-story building of height H, deforming primarily in shear, and with *linear* fundamental period and viscous damping ratio T and ζ. Then, k and C are estimated such that $(k/m_b = (2\pi/T)^2$ and $C/m_b = 2(2\pi/T)\zeta$. The height of the equivalent SDOF oscillator is estimated for the fundamental period T, assuming that $H_e \backsimeq 0.64H$ (from the equivalence of the fundamental frequency of vibration of a shear beam with that of the SDOF oscillator), that T is approximately $T \backsimeq N/10$, and that the average story height is 3.5 m, which gives

$$H_e \simeq 0.64 \times 3.5 \times 10 \times T \tag{6}$$

The nonlinear behavior of the oscillator is modeled by an elasto-plastic stress–strain relationship, as shown in Fig. 3 (right), where ϕ_y is the yield rocking angle, and a_y is the acceleration that statically causes the oscillator to yield. The yield angle ϕ_s can be determined from the static equilibrium of the oscillator, which is

$$k\phi_y = mgH_e \sin \phi_s \tag{7}$$

or in terms of natural frequency ω_0, $\sin\phi_s = \omega_0^2 H_e \phi_y / g$.

Angle ϕ_y in Eq. (7) can be written also as

$$\phi_y = \frac{a_y}{\omega_0^2 H_e} \tag{8}$$

We express our *dynamic* failure criterion in terms of the *static* failure criterion. For small rotations, $\sin\phi_s \backsimeq \phi_s$, and the *static* failure criterion is

$$\phi > \phi_s \tag{9}$$

Because a sudden pulse in the ground motion may reverse the motion of the oscillator, forcing it to return to its elastic range of response, for convenience and simplicity we define as *dynamics* collapse criterion

$$\phi > 2\phi_s$$

to insure that *dynamic* collapse has occurred.

We excite the oscillator by simultaneous horizontal and vertical ground motion and solve the geometrically and materially nonlinear differential equation of its response $\phi(t)$. Because our method of solution is not limited to small deflections, our formulation describes exactly the gravity effects and dynamic instability during all stages of collapse. During nonlinear response, elasto-plastic spring dissipates hysteretic energy.

3. Results and analysis

The oscillator response was computed numerically using a fourth-order Runge–Kutta algorithm, with zero initial conditions, and time step $\Delta t = 5 \times 10^{-5}$ s. The Fortran computer program NLR was used to compute the results. We used nonlinear oscillators with fundamental periods in the linear range of response $T = 0.2$, 0.5, 1 and 2 s, with linear damping ratio $\zeta = 5\%$. For convenience, a small value of the yield acceleration was adopted, $a_y = 1\ \mathrm{cm/s^2}$, so that many oscillators would yield for most of the records in the database. An equivalent alternative would have been to scale up all strong motion records. The purpose has been to create a large database on collapse times, in the presence of gravity effects and dynamic instability.

Due to limitations in recording, the strong motion data are released as band pass filtered. The low-frequency cut-off is a result of baseline correction and is variable, depending on the signal-to-noise ratio [13]. The high-frequency cut-off is determined by the digitization noise, and by the characteristics of the recording transducer, as described in Trifunac [14], Lee and Trifunac [15], and was performed for the EQINFOS database by applying an Ormsby filter with roll-off at 25 Hz and cut-off at 27 Hz. Consequently, the sampling at which the data were released (50/s or at $\Delta t = 0.02$ s) is practically the critical sampling rate.

The wavelet expansion was computed using the Matlab wavelet toolbox [16], as follows. The velocity records were expanded in a six-level orthonormal basis (in L_2) of Coiflet 5 wavelets, and then reconstructed using only the largest amplitude ("top") coefficients. The acceleration of the approximated (by truncated wavelet series) records needed as input to compute the nonlinear oscillator response was calculated from the reconstructed time series by differentiation. The level of the approximation is expressed as a percentage of the number of wavelets that would represent the signal "exactly", which is equal to the number of samples in the critically sampled signal. Hence, the number of wavelets needed to represent a record exactly is proportional to its duration. We tried approximations with 1–8% of the "top" coefficients, which correspond to 0.5–4 wavelets per second—on the average. The approximated signal can be viewed as unevenly subsampled (in the wavelet domain), depending on the energy distribution in the signal, in time and in frequency [8].

3.1. Detailed results for a sample record

We illustrate detailed results for component S00E of the record of the Imperial Valley, California, earthquake of May 18, 1940 ($M = 6.7$) recorded at El Centro, at epicentral distance of 9.3 km [1]. We refer to this record in this paper as "the El Centro record". The duration of this record is 53.7 s.

Fig. 4 shows a wavelet map of the coefficients for the S00E of the El Centro record of the top 8% of the wavelet coefficients. Such maps show the wavelet coefficients (coefficients of expansion in wavelet series) plotted as vertical bars versus the central time (centroid in the time domain) of the corresponding wavelet, for each of the detail subbands (d1–d6) and the remaining smooth subband (s6). The frequency bounds indicated for each subband are those for ideal (box) filters. The actual filters decay gradually across the ideal bounds to avoid the Gibbs effect, and consequently there is some partial overlap between these intervals for the actual filters used. The wavelet maps, therefore, show the position of the wavelets used for the approximation both in time and in frequency. According to the Parseval equality (Eq. (3)), the square of these coefficients would be the contribution of the corresponding term in the series to the energy of the (velocity) signal. Such a map, therefore, indicates the distribution of the energy of the signal on the time–frequency plane. At the top of the figure, the original and approximated velocity signals are also shown for visual comparison.

Table 1 shows the distribution of coefficients in subbands for approximations of the S00E component of the El Centro record, for approximations by 1%, 2%, 3%, 4% and 8% of top coefficients in velocity over the full length of the record. The total number of coefficients used for the approximation is shown in the bottom line. It can be seen that none of the coefficients from the highest detail subband (d1, 12.5–25 Hz) contributed to these low levels of approximation, and that the next detailed subband (d2) contributed only to the 8% approximation.

Fig. 5 shows a comparison of the exact and approximated motions for the 1% and 2% level approximations (parts (a) and (b), respectively). For each case, the plot on the top shows the acceleration and the one on the bottom shows the velocity time history. To avoid clutter, only the first 30 s of the record are shown. It can be seen that the agreement is better for the velocity than for the acceleration signals, which is due to the fact that the thresholding was applied to the velocity signal, from which the acceleration was then derived by differentiation, and that the velocity signal has less energy in the higher-frequency subbands, leading to all or most coefficients in these subbands being eliminated by the thresholding (see Table 1). These plots also show that, *while the low-amplitude high-frequency pulses are smoothed in such low approximation levels, the largest amplitude pulses are still represented quite well even in the acceleration signals*. This is characteristic to data compression by thresholding, in which *the high-frequency components are filtered where they are small but are preserved where they are significant*.

Fig. 6 compares $\mathrm{en}(t) = \int_0^t v(\tau)^2\,\mathrm{d}\tau$, which shows the growth of energy of the input ground motion with time, for the actual (solid

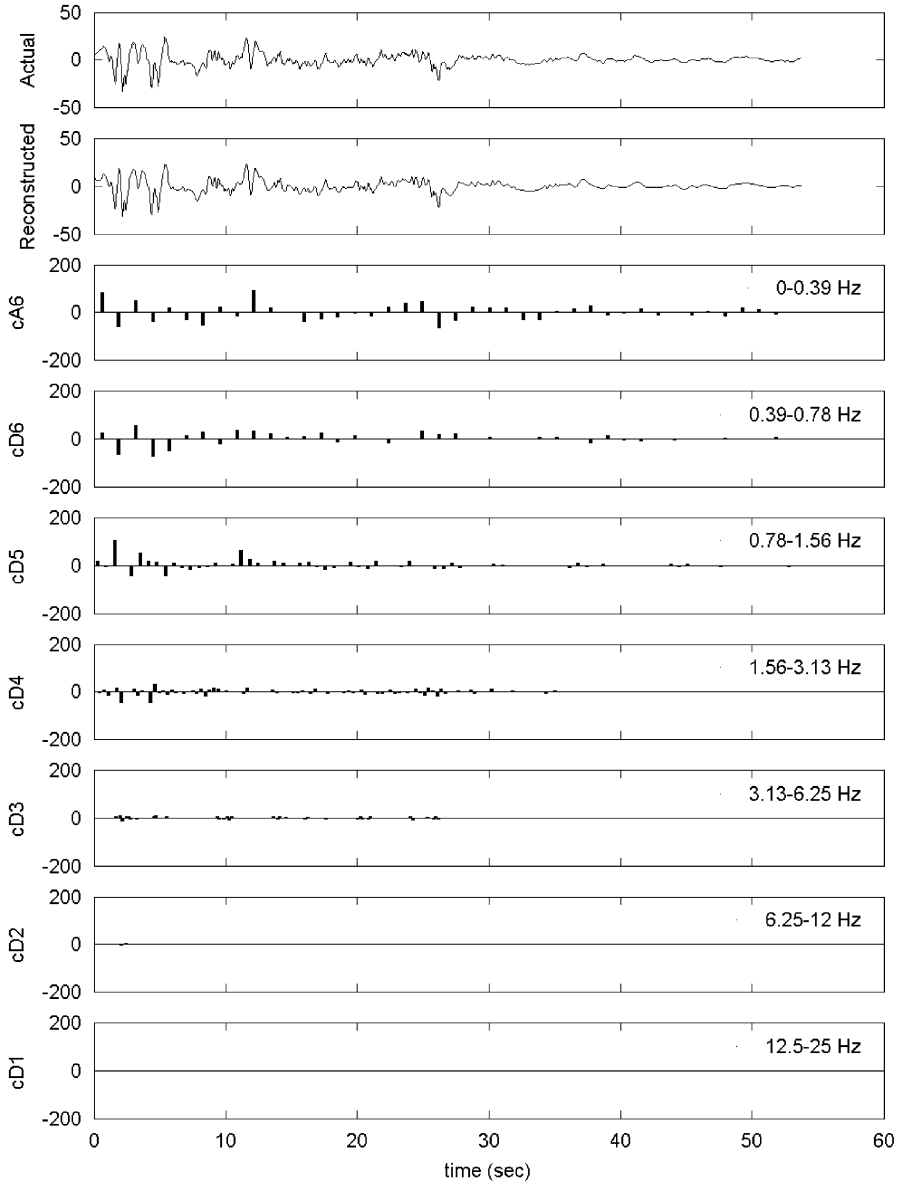

Fig. 4. The top 8% wavelet coefficients of the expansion of the velocity of component S00E of the El Centro strong motion record, plotted versus their central time for each subband. The original signal and the reconstruction using the top 8% coefficients (214 total) are also shown on the top.

Table 1
Distribution of the top coefficients in the approximation of the velocity of the El Centro record (S00E component) for different approximation levels

Subband	Frequency range (Hz)		Level of approximation				
	Lower	Upper	1%	2%	3%	4%	8%
d1	12.50	25.00	–	–	–	–	–
d2	6.25	12.50	–	–	–	–	2
d3	3.13	6.25	–	–	3	14	35
d4	1.56	33.1	3	5	23	46	60
d5	0.78	1.56	5	9	25	39	47
d6	0.39	0.78	5	13	21	26	30
a6	0.00	0.39	13	26	35	36	40
Total number of coefficients used			26	53	107	161	214

line) and approximations (dashed lines) by 1%, 2%, 4% and 8% of the coefficients. It can be seen that the 1, 2 and 4 approximations represent 72%, 86% and 94% of the energy of the signal, and by further increasing the number of coefficients by a factor of 2 (to 8% approximation), the energy increases only by 3%, to 97% of the energy of the exact signal.

Fig. 7 compares the power versus time for the exact signal and the 4% approximation. The pulses in these curves represent bursts of energy "pumped" into the oscillator. Only the first 15 s are shown to avoid clutter. It can be seen that this approximation represents quite well the peaks in the power time history, and in particular the largest peaks.

In the linear range of response, small difference in the input motion implies small difference in the response of the oscillator. However, that does not hold for the nonlinear range. For the bi-linear oscillator considered in this study, the agreement of the time of collapse was chosen as a measure of the goodness of fit, and a weak oscillator was chosen for this test that would fail for most of the records in the database. Fig. 8 shows a comparison of the relative rocking response of the oscillator, up to the time of collapse, for the actual record (100% approximation) and several approximations of the excitation (1%, 4% and 8%). The excitation

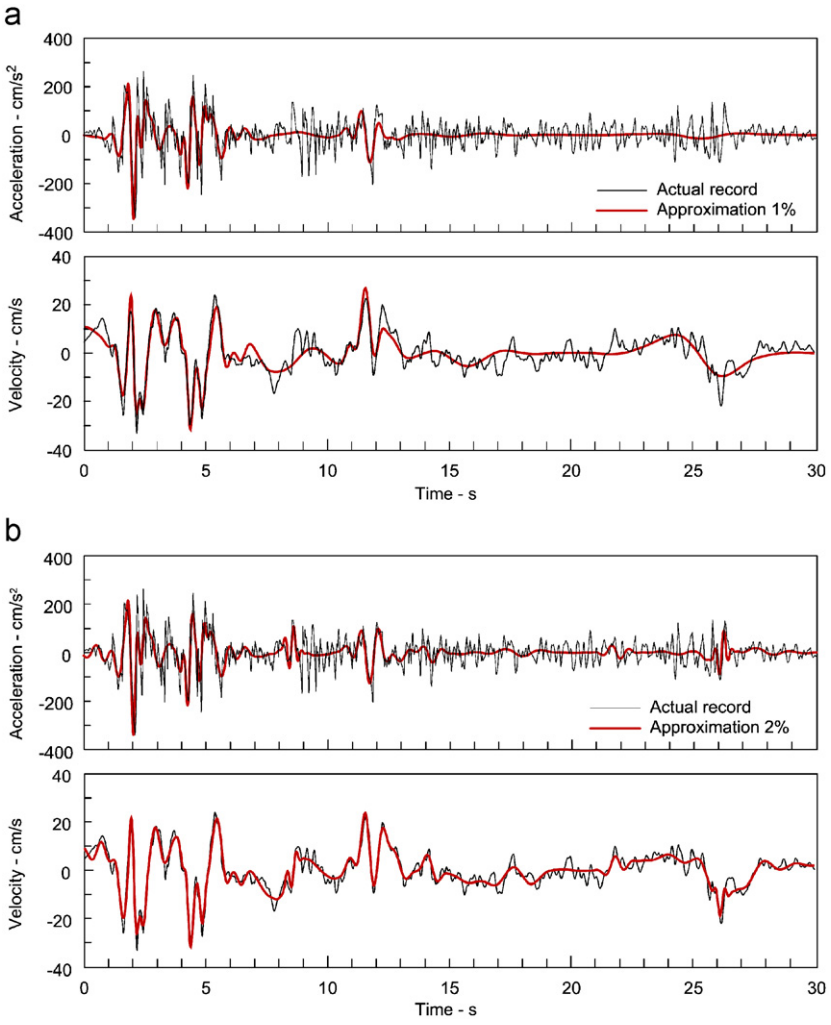

Fig. 5. Comparison of the actual and approximated motions of component S00E of the El Centro record for approximations by (a) the top 1% and (b) the top 2% wavelets in the expansion of velocity.

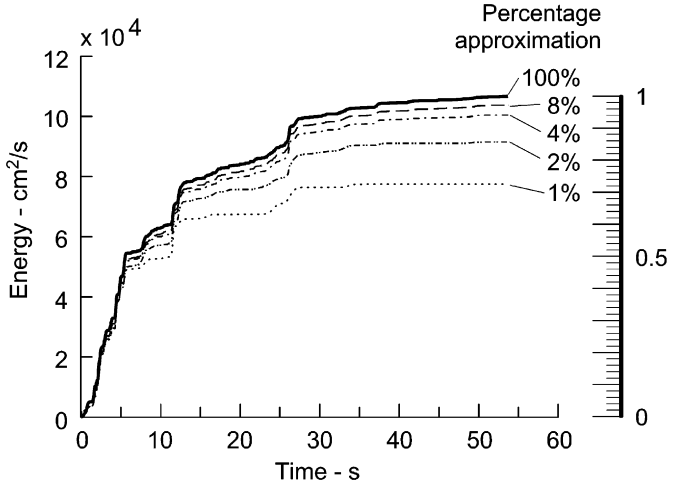

Fig. 6. Energy versus time for the actual record (100% approximation) and approximated El Centro record by 1%, 2%, 4% and 8% of the wavelets.

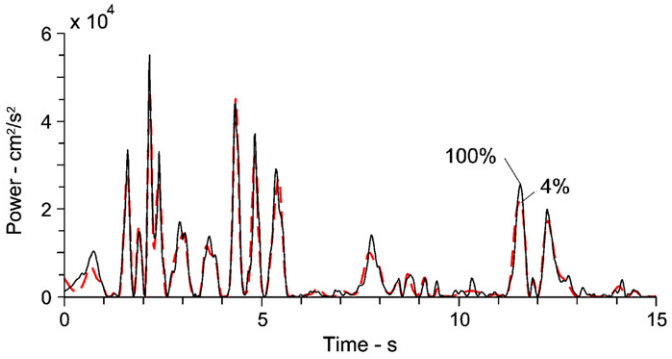

Fig. 7. Comparison of the power of the actual (solid line) and approximated (dashed line) El Centro record by 4% of the wavelets, during the initial 15 s after trigger.

for this case is the S90W component of the record of Kern County, California, earthquake of 1952 at Caltech Athenaeum in Pasadena, which is weaker than the El Centro record, producing collapse at a later time and an opportunity to monitor the differences in the response over a longer time period. The oscillator has period $T = 1$ s, and the yield acceleration $a_y = 1\,\mathrm{cm/s^2}$. The top part of the figure shows the ground velocity for the different level approximations, and the bottom part shows the oscillator response. The peak acceleration for this record is $-52\,\mathrm{cm/s^2}$ occurring at 16.7 s, and collapse of this oscillator for the exact record occurs near

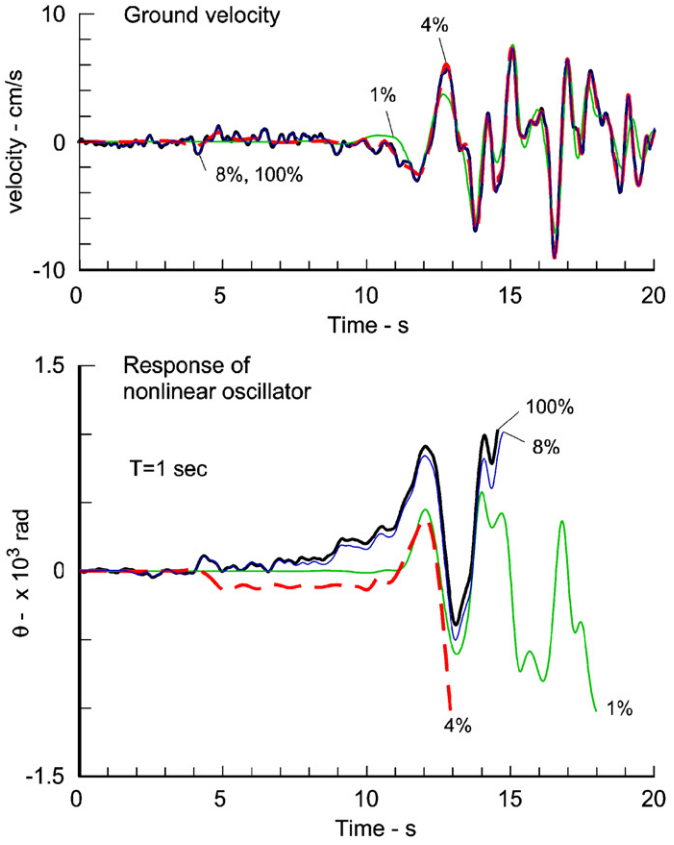

Fig. 8. Comparison of a nonlinear oscillator response (up to time of collapse) to excitation using approximations by 1%, 4%, 8% and 100% of the wavelets. The excitation is the S90W component of the record of the 1952 Kern County, California earthquake at Caltech Athenaeum in Pasadena (peak acceleration—52.063 cm/s^2 at 16.7 s). The top part shows the ground velocity, and the bottom part shows the response angle θ. The oscillator has period is $T = 1$ s and yield acceleration $a_y = 1$ cm/s^2.

Table 2
Number of components of motion in the database for which the error in the total energy exceeds 10%

Top terms in the series approximating the records (%)	Components for which the error exceeds 10%	
	Total number	Percentage (%)
1	1373	92
2	1059	71
4	612	41
6	451	30
8	350	24

$t = 15$ s. It can be seen that the 8% approximation predicts very close to the time of collapse. The 4% approximation collapses the oscillator sooner than the exact record by less than 2 s, while the 1% approximation collapses the oscillator about 3 s later.

3.2. Statistics for a sample database

Statistical results for the goodness of fit are shown for the EQINFOS for USA Part I database [1], which consists of 494 three-component accelerograms of 106 earthquakes between 1933 and 1984 recorded mainly in the western continental US. Table 2 shows different approximations, the number of components of motion for which the error in energy exceeded 10%. It can be seen that for 75% of the components, the top 8% terms in the series represent 90% or more of the energy of the input motion.

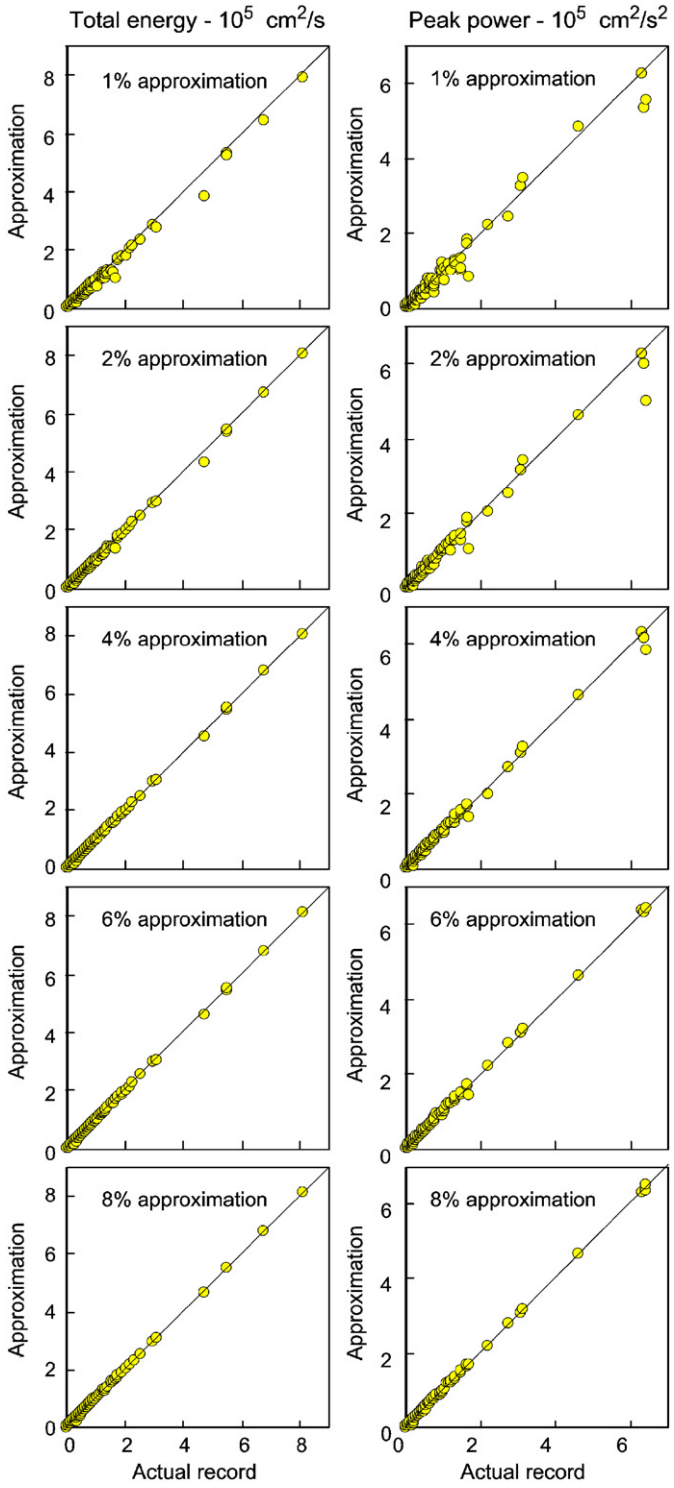

Fig. 9. Correlation plots for the total energy (left) and peak power (right) for the EQINFOS for USA I strong motion database. Parts (a)–(e) correspond to approximation by 1%, 2%, 4%, 6% and 8% of the wavelets in the expansion.

Fig. 9 shows the correlation between the total energy (left) and peak power (right) of the actual accelerogram and that of its approximation. Parts (a)–(e) correspond to approximations by 1%, 2%, 4%, 6% and 8% of the top terms in the series. It can be seen that the correlation is better for the total energy than for the peak power, and that it is very good even for the crudest approximation

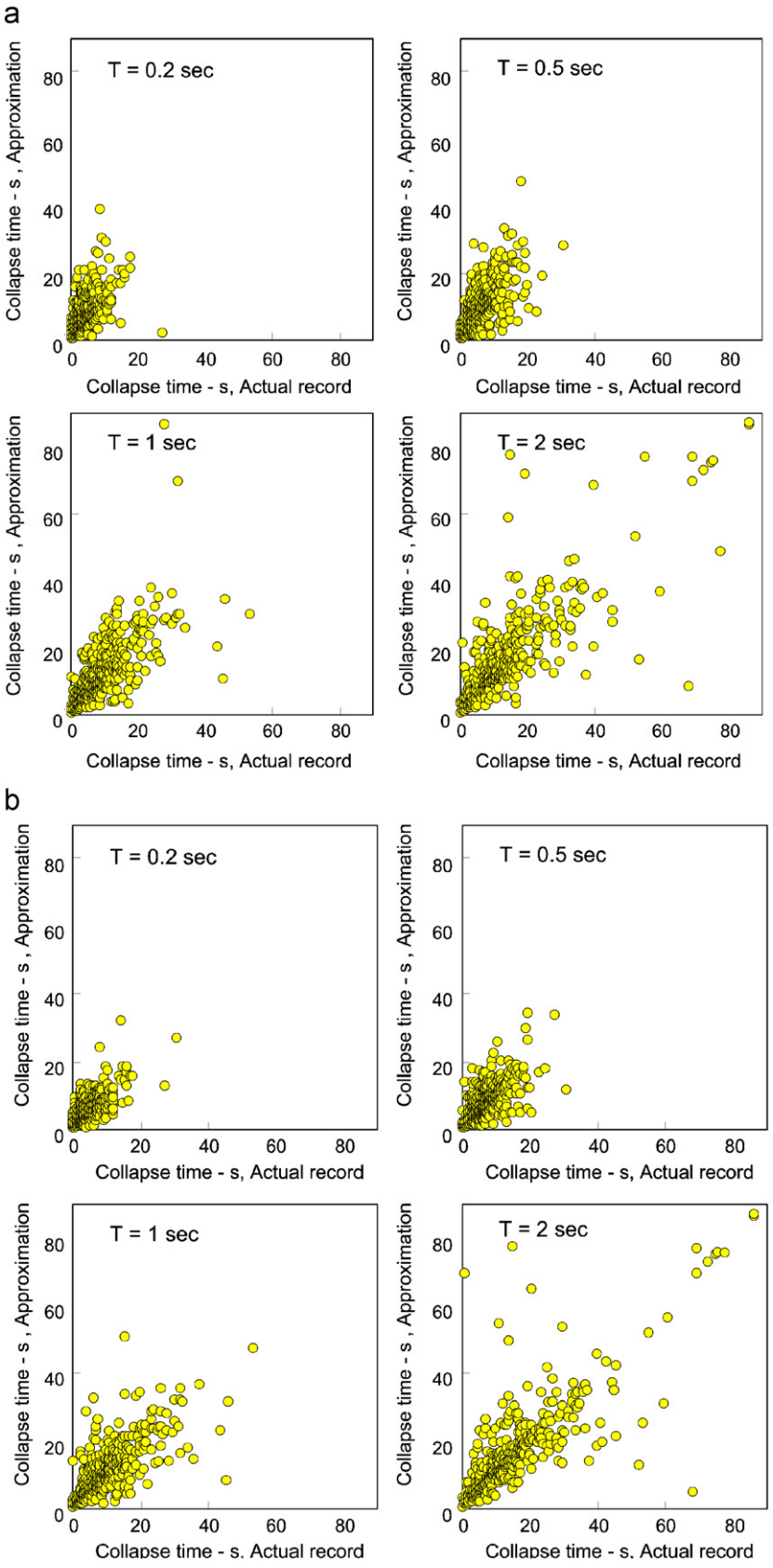

Fig. 10. Correlation plots of collapse time of a nonlinear oscillator with periods $T = 0.2$ s (top-left), 0.5 s (top-right), $T = 1$ s (bottom-left), and $T = 2$ s (bottom-right). (a) Approximation by 1% of the wavelets. (b) Approximation by 8% of the wavelets.

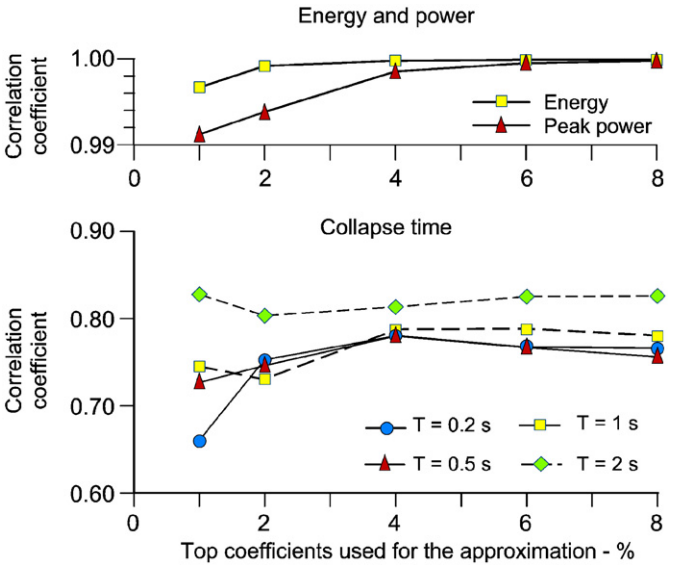

Fig. 11. Correlation coefficients versus percentage of wavelets used for the approximation for the total energy and peak power (top) and for the time of collapse for different oscillator periods of vibration (bottom). The corresponding correlation plots are shown in Figs. 9 and 10.

considered (by 1% of the terms in the series). Fig. 10 shows correlation plots between the collapse times of nonlinear oscillators excited by actual record and by its approximation. Parts (a) and (b) correspond to approximations by 1% and 8% terms in the series. In each part, the different plots correspond to oscillators with periods $T = 0.2$ s (top-left), 0.5 s (top-right), 1 s (bottom-left) and 2 s (bottom-right).

Fig. 11 shows the trends of the coefficients of correlation, plotted versus the percentage of terms in the series used for the approximation, for the total energy and peak power (top) and for the collapse time (bottom). It can be seen that the correlation for the total energy and peak power of the oscillator input is very high (>0.99) even for the approximation by only 1% of the terms. The correlation increases very fast with increasing level of approximation, approaching 1 even for 4% approximation level. Hence, including more terms in the approximation has an incremental effect on the improvement of the goodness of fit in terms of total energy and peak power. For the time of collapse, the correlation is smaller than for the energy and peak power, but still exceeds about 0.75 for approximations better than 2% for all the periods considered. The correlation is best for the longest period oscillator ($T = 2$ s), for which it exceeds 0.8 even for the 1% approximation of the input.

4. Discussion and conclusions

The broader objective of this work is representation of earthquake strong motion records by a relatively small number of pulses, which would be helpful for gaining insight in the nature of the high-frequency radiation from the earthquake source, and for the purpose of physically based synthesis of artificial strong ground motion. In this paper, wavelets of the Coiflet 5 orthogonal wavelet family were used as proxies for the pulses arriving at a site. This family was used because the wavelets are orthogonal (making it simple to compute the energy directly in the wavelet transform domain), nearly symmetric (hence, minimizing phase distortion in the approximated signal), smooth to a high degree (hence, achieving better approximation of strong motion records with smaller number of wavelets), and compactly supported (hence can be used with the Fast Wavelet Transform).

The Coiflet 5 family constitutes a basis in L_2, and a strong motion record can be expanded in such a basis. Representation of a strong motion record by "few" of the basis functions is an approximation, the number of basis functions that would represent the record exactly being equal to the number of points in a critically sampled record. In this paper, the wavelets that contain most of the energy of the ground motion were used for the approximation. These are the wavelets in the expansion of *velocity* (because the energy of ground motion is proportional to the time integral of velocity squared) that are multiplied by the largest amplitude coefficients (because the basis is orthonormal in L_2). The level of approximation was defined by the percentage of the wavelets used for the approximation.

The goodness of the approximation was measured in terms of the ability to represent both the input and output of a *nonlinear* oscillator, reasonably well, representing a structure. For the input, the total energy and peak power were compared, and for the output—the time of collapse (low-yield acceleration was chosen, so that most of the records in the database would cause collapse). The correlation between these quantities obtained from the actual signal and from its approximation was computed for the EQINFOS Part I for USA database. The results show very high correlation for the energy and peak power of the ground motion (coefficient of correlation >0.99), even if only 1% terms were used in the approximation. This would imply also high degree of correlation for the output of an oscillator in the *linear* range of response. For the output of a nonlinear oscillator that would collapse (due to gravity effects and dynamic instability), the coefficient of correlation for the time of collapse was smaller, but still exceeded about 0.75 for the oscillators considered (with period $T = 0.2$, 0.5, 1 and 2 s, 5% viscous damping, and yield acceleration $a_y = 1\ \mathrm{cm/s^2}$), if at least 4% of the terms in the series were used for the approximation. For the yield acceleration chosen, the correlation was higher for the longest period oscillator ($T = 2$ s).

It is concluded that expansion of strong motion records in a wavelet basis is an efficient tool for extraction of pulses from a strong motion record, and representation of strong motion records as a sum of a relatively small number of pulses. This efficiency (good approximation by a small number of pulses) is due to the fact that the basis functions are localized in time (besides in frequency), resembling in nature the strong motion records.

Acknowledgement

This work was in part supported by a grant from the USC WiSE program.

References

[1] Lee VW, Trifunac MD. Strong earthquake ground motion data in EQUINFOS: Part I. Report no. 87-01, Department of Civil Engineering, University of Southern California, Los Angeles, CA, 1987.
[2] Iyama J, Kuwamura H. Application of wavelet to analysis and simulation of earthquake motions. Earthquake Eng Struct Dynam 1999;28(3):255–72.
[3] Mukherjee S, Gupta VK. Wavelet-based characterization of design ground motions. Earthquake Eng Struct Dynam 2002;31:1173–90.
[4] Iyama J. Estimate of input energy for elasto-plastic SDOF systems during earthquakes based on discrete wavelet coefficients. Earthquake Eng Struct Dynam 2005;34(15):1799–815.
[5] Mallat SG. Multiresolution approximations and wavelet orthonormal bases of $L_2(R)$. Trans Am Math Soc 1989;315:69–87.
[6] Vetterli M, Kovačević J. Wavelets and sub-band coding. Upper Saddle River, NJ: Prentice-Hall PTR; 1995.
[7] Todorovska MI. Estimation of instantaneous frequency of signals using the continuous wavelet transform. Report CE 01-07, Department of Civil Engineering, University of Southern California, Los Angeles, CA, 2001.

[8] Todorovska MI, Hao T-Y. Information granulation and dimensionality reduction of seismic vibration monitoring data using orthonormal discrete wavelet transform for possible application to data mining. Report CE 03-02, Department of Civil Engineering, University of Southern California, Los Angeles, CA, 2003.
[9] Daubechies I. Ten lectures on wavelets. Society for Industrial Application of Mathematics (SIAM), Philadelphia, Pennsylvania, 1992.
[10] Trifunac MD. Energy of strong motion at earthquake source. Soil Dyn Earthquake Eng 2008;28(1):1–6.
[11] Gičev V, Trifunac MD. Energy and power of nonlinear waves in a seven story reinforced concrete building. Indian Soc Earthquake Technol J 2007;44(1):305–23.
[12] Gičev V, Trifunac MD. Rotations in a shear beam model of a seven-story building caused by nonlinear waves during earthquake excitation. Struct Control Health Monit 2008, in press, doi:10.1002/stc.264.
[13] Trifunac MD. Zero baseline correction of strong-motion accelerograms. Bull Seism Soc Am 1971;61:1201–11.
[14] Trifunac MD. A note on correction of strong-motion accelerograms for instrument response. Bull Seism Soc Am 1972;62(1):401–9.
[15] Lee VW, Trifunac MD. Automatic digitization and processing of accelerograms using PC. Report no. 90-03, Department of Civil Engineering, University of Southern California, Los Angeles, CA, 1990.
[16] Misiti M, Misiti Y, Oppenheim G, Poggi J-M. Wavelet toolbox 4, user's guide. Natick, MA: The MathWorks Inc.; 2005.
[17] Todorovska MI, Trifunac MD. Earthquake damage detection in the Imperial County Services Building I: the data and time-frequency analysis. Soil Dyn Earthquake Eng 2007;27(6):564–76.