Arab J Geosci (2016) 9: 580
DOI 10.1007/s12517-016-2578-4

ORIGINAL PAPER

Automatic identification of fake patterns caused by short-width wavelets in seismic data

Ahmed Mohamed Tawfiek[1,2] **· Guanzheng TAN**[1] **· Ali G. Hafez**[2,3] **· Abdullah Al-Amri**[4] **· Nassir Alarif**[4] **· Kamal Abdelrahman**[2,4]

Received: 19 August 2015 / Accepted: 23 June 2016 / Published online: 6 August 2016

Abstract Despite the popularity of using the Haar wavelet filter in many applications, it sometimes introduces fake patterns into the multi resolution analysis (MRA) of seismic data. In this work, we compared different wavelet filters to demonstrate that these patterns are fake and not part of the original waveforms and to show that they are a result of using the Haar wavelet filter as a short-width wavelet. To achieve this, many seismic waveforms from two different sources: the Egyptian National Seismic Network (ENSN) and the High Sensitivity Seismograph Network Japan (Hi-net) are used with different wavelet filters. We propose an algorithm based on an autoregressive (AR) model to detect these patterns automatically and fully.

Keywords Seismic data · Haar wavelet · Multi resolution analysis (MRA) · Spectrum leakage · Fake patterns

✉ Ahmed Mohamed Tawfiek
ahmed_tawfiek@csu.edu.cn

✉ Guanzheng TAN
tgz@csu.edu.cn

Ali G. Hafez
aligamal@ltlab.com

Abdullah Al-Amri
amsamri@ksu.edu.sa

Nassir Alarif
nalarifi@ksu.edu.sa

Kamal Abdelrahman
khassanein@ksu.edu.sa

1 School of Information Science and Engineering, Central South University, Changsha, China

2 Seismology Department, National Research Institute of Astronomy and Geophysics, Cairo, Egypt

3 Department of Electrical Engineering, Faculty of Engineering, Nahda University, Bani Sweif, Egypt

4 Geology and Geophysics Department, King Saud University, Riyadh, Saudi Arabia

Introduction

The primary issue in wavelet analysis is the criteria through which to pick a specific wavelet filter. A sensible decision depends principally on the underlying case of study. The subject of choosing the wavelet filter has been the focus of increasing attention from researchers recently. The fundamental properties of wavelets are generally used to determine the suitability of particular wavelets for particular application fields. (Miao and Moon 1999) used non-orthogonal Morlet wavelets to eliminate the ground roll noise from seismic waveforms, taking advantage of their characteristics which are clear in having appropriate localization both in frequency and time domains. Symmetrical wavelets had been used in auditory evoked potentials by (Bradley and Wilson 2004). Due to its orthogonality, compact support, and vanishing moment, coiflet 4 wavelet filter had been adopted on electromyograms data to extract burst and tonic activities by (Wang et al. 2004). (Beenamol et al. 2012) used Daubechies (Dau 4) wavelets for de-noising seismic signals using Shannon and Tsallis entropy. (Hafez et al. 2013), meanwhile, showed that Dau 2 gives better results than the Haar wavelet in detecting precursory signals in front of impulsive P-waves.

There are many kinds of wavelet filters, such as Daubechies, Coiflets, Symlets, Discrete Meyer, Morlet, Haar, and Mexican hat. For the purposes of this paper, we can divide wavelets in two groups according to their width: short and long. Each one has some advantages and disadvantages.

The very short width wavelet, such as Haar, is widely known, not only for its simplicity but also for other advantages such as speed and memory efficiency. Furthermore, it can be calculated in place without a temporary array and is reversible without the edge effects introduced by the other wavelets, it is also very suitable for some applications, as mentioned by (Capilla 2006). Despite these advantages, the Haar wavelet filter also has limitations, which can introduce problems for some applications see (Percival and Walden 2000). One such limitation is the appearance of fake patterns in MRA of seismic waveforms.

Wavelets with large widths are characterized by their good resolution for smoothly changing time series. Unfortunately, however, they have many drawbacks such as the following:

(i) The boundary conditions have severe impacts on many coefficients.
(ii) Some decrease in wavelet coefficients' localization degree.
(iii) Long processing time.

A sensible approach is to use wavelet filter with short width that gives good results, iteratively increasing the width until one comes to an analysis that has the lowest number of errors. For applications that have acceptable results with short width wavelet filters, we therefore avoid decisions caused by these errors.

Inspired by such aspects, we propose a procedure to automatically recognize the undesirable patterns by using the autoregressive (AR) approach. The procedure can determine the locations of these patterns automatically in the time series under investigation. In order to construct this new algorithm, we first introduce MRA, we then collect data from two networks of seismic monitoring stations, to which we apply MRA. This allows us to identify fake patterns and assess their response to various wavelet filters. Building on this data, we then construct an algorithm for the automatic identification of fake patterns.

Multi resolution analysis (MRA)

Many commonly known signals are non-stationary such as radar, music, biomedical, speech, and seismic. Indeed, seismic records were the application used by (Grossmann and Morlet 1984) and (Goupillaud et al. 1984) to test the wavelet transform. This transform is a very suitable for processing non-stationary signals, since the Fourier transform cannot give information about the time of occurrence of a frequency component within the window under investigation. In addition, the wavelet transform (unlike Fourier transform) has the ability to interpret the fine structures of a signal, by decomposing it into details and smooth features, high and low frequencies, respectively. These smooth and detailed features can be obtained by passing the signal through a scaling filter and wavelet filter, respectively. This process has to be done at different stages in order to get the details and smooth at many scales, as shown below (for more information please review (Hafez and Kohda 2009)).

Let i the number of the stage, n the width of the vector, $u_{i,n}$ $n = 0, 1, 2, \ldots, N-1$ the scaling vector, and $v_{i,n}: n = 0, 1, 2, \ldots, N-1$ the wavelet vector. The scaling vector u_n is a quadrature mirror vector has the relation with the wavelet vector v_n as shown in equation (1)

$$u_n = (-1)^{n+1} v_{N-1-n} \tag{1}$$

Hence:

$$v_n = (-1)^n u_{N-1-n} \tag{2}$$

After passing the input signal $x(n)$ through scaling and wavelet vectors, the scaling coefficients $C_{1,m}$ and the wavelet coefficients $W_{1,m}$ will be obtained. This is shown by Eqs. (3) and (4)

$$C_{1,m} = \sum_{n=0}^{M-1} u^o_{1,n} x_{2m+1-n \mod M} \tag{3}$$

$$W_{1,m} = \sum_{n=0}^{M-1} v^o_{1,n} x_{2m+1-n \mod M} \tag{4}$$

where $m = 0, 1, \ldots, \frac{M}{2}-1$, both $\{u^o_n\}$ and $\{v^o_n\}$ are periodized to length M of $\{u_n\}$ and $\{v_n\}$, respectively, and "*a mod b*" denotes "*a modulo b*."

Equations (5) and (6) show the smooth S_1 and detail T_1

$$S_1 = \sum_{n=0}^{M-1} u^o_{1,n} C_{1,m+n \mod M} \tag{5}$$

$$T_1 = \sum_{n=0}^{M-1} v^o_{1,n} W_{1,m+n \mod M} \tag{6}$$

The following relationship demonstrates how to reconstruct in original input signal $x(n)$ from the details and approximations of the first stage

$$x(n) = S_1 + T_1$$

At the second stage, the values of the scaling vector $C_{1,m}$ which is generated from the first stage equation (3) is again passed through scaling vector $u_{2,n}$ and wavelet filter $v_{2,n}$ with the resultant output being the stage scaling; $C_{2,m}$ and wavelet; $W_{2,m}$ coefficients. This is shown by the equations (7) and (8)

$$C_{2,m} = \sum_{n=0}^{M-1} u^o_{2,n} x_{4(m+1)-n \mod M} \tag{7}$$

$$W_{2,m} = \sum_{n=0}^{M-1} v^o_{2,n} x_{4(m+1)-n \mod M} \tag{8}$$

Equations (9) and (10) show the smooth S_2 and detail T_2

$$S_2 = \sum_{n=0}^{M-1} u^o_{2,n} C_{2,m+n \mod M} \tag{9}$$

$$T_2 = \sum_{n=0}^{M-1} v^o_{2,n} W_{2,m+n \mod M} \tag{10}$$

After applying the following relationship, the original input signal $x(n)$ can be restored

$$x(n) = S_2 + T_2 + T_1$$

By extending the same work, equations 11–14 demonstrate the values of scaling vector $C_{i,m}$ the wavelet vector values $W_{i,m}$, the smooth S_i and the detail T_i at *ith* stage

$$C_{i,m} = \sum_{n=0}^{M_i-1} u_{i,n} x_{2^i(m+1)-n \mod M} \tag{11}$$

$$W_{i,m} = \sum_{n=0}^{M_i-1} v_{i,n} x_{2^i(m+1)-n \mod M} \tag{12}$$

$$S_i = \sum_{n=0}^{N_i-1} u^o_{i,n} C_{i,m+n \mod M} \tag{13}$$

$$T_i = \sum_{n=0}^{N_i-1} v^o_{i,n} W_{i,m+n \mod M} \tag{14}$$

In the same manner, the original input signal $x(n)$ can be recovered again by using the following relationship

$$x(n) = S_i + T_i + \ldots + T_1$$

Fake patterns in seismic records

Discovery

As mentioned before, during our analysis of seismic records, strange patterns were evident in the multiresolution analyses (MRA) of seismic noise, while the earthquake records had been also checked to investigate if there are fake patterns or not. We have noticed that there are not any fake patterns appear in earthquake records. Figure 1 shows seismic noise records and the MRA of seismic segments recorded at two stations: SAF (ENSN) and Wakkanai (Hi-net). The appearance of the undesirable patterns can be clearly recognized in each of the three details. Note that the sampling rate is 100 and each count is equal to 1 nm/s^2.

In an effort to determine the cause of these, we first sought to confirm the existence of these patterns in different seismic stations and then sought to justify the relationship between these patterns and the used wavelet filter. The data used in this study are seismic records from many stations belonging to the Egyptian National Seismic Network (ENSN) and the Japanese High Sensitivity Seismograph Network (Hi-net). Figure 2 shows the study area and used data from ENSN.

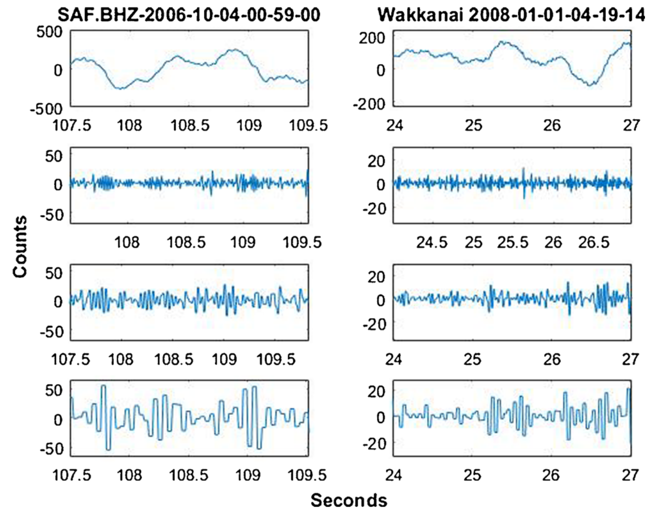

Fig. 1 Fake patterns appear in SAF and Wakkanai stations

To address the first point, we checked many seismic records extracted from different stations belong to ENSN and Hi-net to see if their MRA contained similar strange patterns. We noticed the appearance of these patterns in most of the stations from both networks. To address the second point, meanwhile, we used many wavelet filters to calculate the MRA for every seismic record. We noticed that these patterns appeared more clearly with short width wavelet filters, such as the Haar wavelet filter. With longer width wavelet filters, such as Daubechies 4, 6 and coiflet 4, these patterns did not appear clearly in the MRA of seismic records. Figure 3 shows an example of this from one station, revealing the existence of many patterns in the first detail of the Haar wavelet, but also showing how these patterns decrease when long width

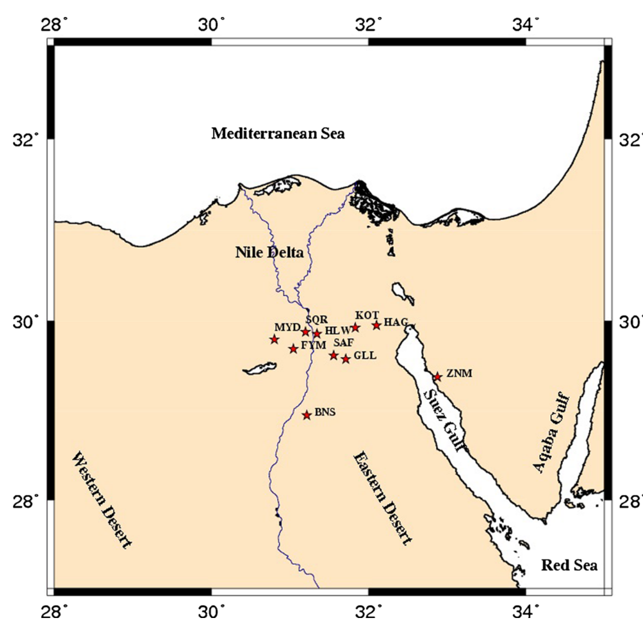

Fig. 2 Study area and used data from ENSN

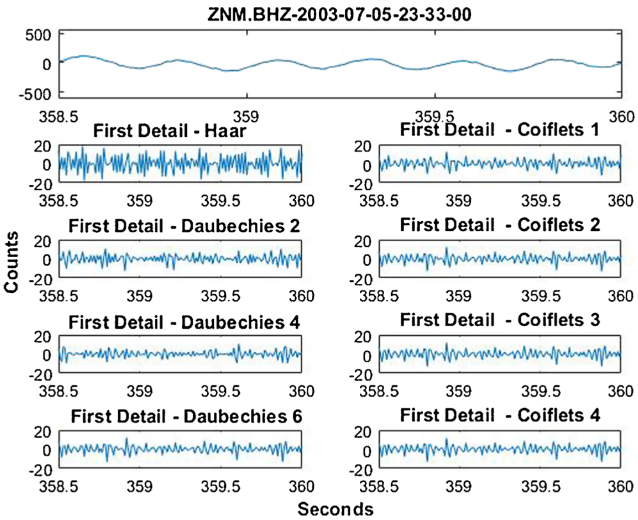

Fig. 3 First detail of MRA of seismic waveform recorded at ZNM station (ENSN) based on different wavelet filters

wavelet filters, such as Daubechies (Dau4) are used. This is a strong indication that the Haar wavelet filter (or more specific the short width wavelet filters) is the main cause of these patterns.

Wavelet variance

Wavelet variance estimate of wavelet coefficients is an important tool that can be used to compare between the coefficients that are calculated using different wavelet filters at different scales. Using such tool will enable us to decide about specific wavelet filter and check whether it is suitable or not, where the wavelet variance estimates of this filter should have near values of the estimates of other wavelets. Calculation of the discrete wavelet transform (DWT) variance had been done in the literature for many applications including by (Unser 1995), (Nason et al. 2000), (Percival and Walden 2000), (Tsakiroglou and Walden 2002), (Whitcher et al. 2002), (Stoev and Taqqu 2003), and (Capilla 2006).

Assuming that $\{\mathcal{X}_t, t\in\mathbb{Z}\}$ is an innately stationary signal of order d, where d is a positive integer and its $\mathcal{D}^{th}$ order increments $(1-\mathcal{B})^{\mathcal{D}}\mathcal{X}_t$ are stationary.

S_x is the spectral density function (SDF) of the signal. Wavelet coefficients of each stage (scale j) can be calculated by equation (4) in section 2. At scale $\tau_i = 2^{i-1}$. The variance of $W_{i,m}$ is the wavelet variance at level i.

Supposing that $L \geqslant 2d$, $\{W\}_{i,m}$ stationary with spectral density function $H_i(f)\ S_x(f)$, then we can calculate the variance of the coefficients of the wavelet by

$$\mathcal{V}_X^2(\tau_i) = \int_{\frac{-1}{2}}^{\frac{1}{2}} H_i(f)\ S_X(f)df \qquad (15)$$

In case a wavelet of width, L = 2, is used, such as the Haar wavelet filter, the DWT variance for each scale τ_i of a stationary signal, X_t, can be calculated as follows,

$$var(X_t) = \sum_{i=1}^{\infty} \mathcal{V}_X^2(\tau_i) \qquad (16)$$

This representation can replace the old calculation,

$$var(X_t) = \int_{\frac{-1}{2}}^{\frac{1}{2}} S_X(f)df \qquad (17)$$

Using the decomposition of $var(X_t)$, which can be obtained from the DWT variance, together with the spectral density function (SDF) of each stage is an important issue for many applications especially in the geosciences. Assuming an observation $X_0 \ldots X_{M-1}$, when calculating the wavelet coefficients using a wavelet filter of width $L > 2d$ where $\{W_{i,m}\}$, has a zero mean, the unbiased estimator of $\mathcal{V}_X^2(\tau_i)$ can be given by equation (18), [8, 16], Where $M_i = L - N_i + 1 > 0$

$$\hat{v}_X^2(\tau_i) = \frac{1}{M_i} \sum_{m=N_i-1}^{N-1} W_{i,m}^2 \qquad (18)$$

In the case we have in this study, where fake patterns appear in the MRA components of the seismic noise records, we decided to use the DWT variance estimates, $\mathcal{V}_X^2(\tau_i)$ as a measure to compare between the MRA variance estimates when using different kinds of wavelet filters of different widths. The wavelet filters used in this study are Haar, Daubechies 2, Daubechies 4, and Coiflet 4, as shown in Fig. 4 and from it

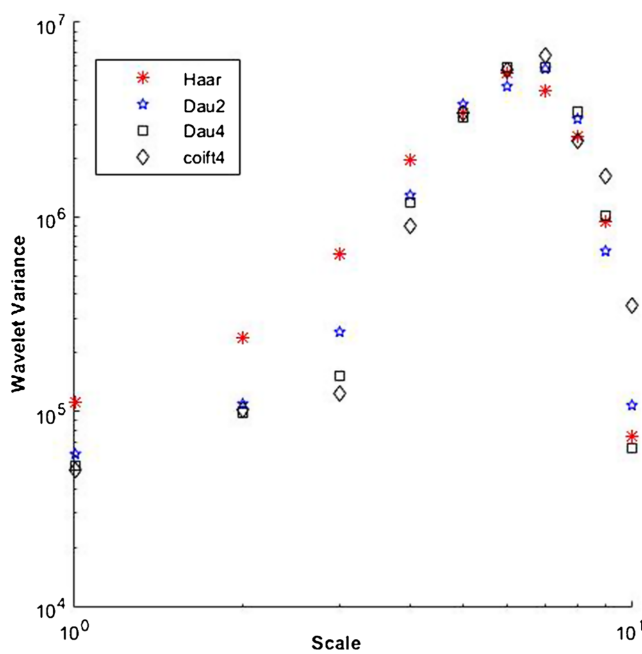

Fig. 4 Estimation of the variance of wavelet coefficients for seismic noise records using Haar, Daubechies 2 (Dau2), Daubechies 4 (Dau4), and Coiflet 4 wavelet filters

we can observe that there is an agreement between the variance estimates of all the wavelet filters except that of the Haar wavelet, which has a completely different value to the estimates from the other wavelets. This difference gives us a good implication that the use of wavelet vector of type Haar for analysing seismic waveforms will introduce errors in the coefficients of wavelet vector.

Automatic identification of fake patterns

In our work, we introduce a procedure monitors the coefficients of an autoregressive (AR) model, which has the ability to recognize fake patterns automatically. The proposed algorithm will determine the exact places of these patterns within the time series. This automatic identification will be a great tool that can aid automatic picking systems of seismic data by pointing out the fake patterns from original signals. In addition to that, this algorithm will help in showing the leakage which caused by short width wavelet filters.

Autoregressive (AR) model

The term autoregressive appears in many fields such as statistics and signal processing, referring to a method that tries to expect many natural phenomena in which the current output of the system (X_t) is based on the previous outputs ($X_{t-1}, X_{t-2}, \ldots$). The term AR(p) means an autoregressive model of order p which can be defined by equation (19)

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t \tag{19}$$

where $\varphi_1, \ldots, \varphi_p$ are the model parameters and ε_t is white noise.

There are many methods to estimate the AR parameters, such as Burg, Covariance, Modified Covariance, and Yule–Walker. These methods have similar characteristics, although with a few differences. We can divide them into two categories: the first applies windowing to the data (Yule–Walker) while the second does not apply windowing (Burg, Covariance, and Modified Covariance). We can further classify them into two groups; the first group uses least square to reduce the forward prediction error (Covariance and Yule–Walker); and the second group (Burg and Modified Covariance) also uses the least square to minimize the prediction errors of backward and forward (Mathworks 2015). Burg algorithm is the most widely known AR procedure since it always produces a stable model by forcing the parameters of the AR model to satisfy Levinson-Durbin recursion, see (Burg 1975), (Kay and Makhoul 1983), and (Kay 1988).

In this study, we used MATLAB software for the estimation of AR model parameters based on Burg model.

Methodology of the algorithm

The goal of this study is to build an algorithm able to identify the patterns caused by using Haar wavelet filter, not to remove these patterns. Removing these patterns is not a good strategy since it would cause a change in the original time series. Furthermore, since these patterns are errors from using the Haar wavelet, it is not easy to find coefficients that would replace the gaps.

In order to build this algorithm, visual inspection of the windows that have these patterns had been carefully done. Then, the characteristics of these windows in both time and frequency domains had been checked. During these investigations, we found that the AR coefficient φ_2 for the first scale detail of this window ranges between 0.97 and 0.99. In order to confirm that, other windows where fake patterns did not appear, the values of φ_2 are not in this range. Therefore, we used this range as a significant feature to judge the existence of the fake patterns. We will describe the proposed algorithm in more details in the following section.

Algorithm description

We have divided the proposed algorithm into steps to approach the automatic identification of the fake patterns of the Haar wavelet filter.

- The first step is the calculation of the first detail of the multiresolution analysis for the seismic record under consideration.
- In the second step, we have calculated the second-order AR parameters for a moving window of the first detail, which we have calculated in the first step.
- The third step is to check if the AR coefficient φ_2 is in the range from 0.97 to 0.99 or not:

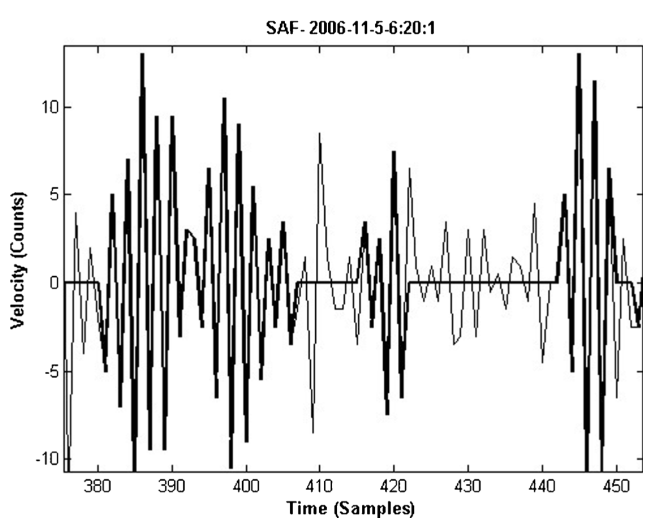

Fig. 5 The first detail of MRA of seismic noise record. *Thin trace* is the original seismic record. *Thick trace* is the identified fake patterns

Table 1 Percentage of existence of fake patterns in seismic records using different wavelet filters

	Haar	Dau2	Dau4	Coiflet4
Percentage	35.4 %	15.5 %	12.4 %	12 %

– If not, the window is moved by one-sample wise for a further check regarding φ_2.
– If yes, calculate the window length and declare the existence of fake patterns in this window.

• Repeat until the end of time series.
• Calculate the ratio of fake patterns compared to the original time series.

In Fig. 5, we have applied our algorithm into a seismic record from SAF station, this station belongs to the Egyptian National Seismic Network (ENSN). We can see that the repeated fake patterns are identified by a thick line using the proposed algorithm.

Simulation results

We have applied our algorithm into many seismic waveforms from ENSN and Hi-net and with different wavelet filters. We have calculated the percentage of fake patterns identified for each wavelet filter. It was proved that the Haar wavelet filters gave a higher percentage of fake pattern windows in the checked segments. The values of these percentages for four different wavelet filters are shown in Table 1.

Conclusion

A study is made of seismic records to investigate the existence of fake patterns recognized in MRA of these records. Different stations from the ENSN and Hi-net Japan were used in this study. From this study, we can conclude that using the Haar wavelet filter will introduce more fake patterns in MRA of seismic noise records than using other wavelet filters, where it give us the conclusion to not to use Haar wavelet filter when analysing the seismic noise. Despite this, the Haar wavelet filter remains the most widely used wavelet in many applications due to the simplicity of its implementation. Motivated by these facts, an algorithm is proposed to identify these fake patterns automatically and fully. Simulation results indicate that the fake patterns introduced by the Haar wavelet constitute around 35 % of the seismic record.

Acknowledgments The authors would like to thank the staff in the Egyptian National Seismic Network (ENSN) and in the High Sensitivity Seismograph Network Japan (Hi-net) for their efforts in preparing the data used in this work. Authors record their heartiest gratitude and thanks to the Egyptian Government and Chinese Scholarship Council for the financial support during the stay of Mr. Ahmed Mohamed Tawfiek in China. The authors extend their sincere appreciations to the Deanship of Scientific Research at the King Saud University for its funding the Prolific Research Group (PRG-1436-21) which shared in this article.

References

Beenamol M, Prabavathy S, Mohanalin J (2012) Wavelet based seismic signal de-noising using Shannon and Tsallis entropy. Comput Math Appl 64:3580–3593. doi:10.1016/j.camwa.2012.09.009

Bradley AP, Wilson WJ (2004) On wavelet analysis of auditory evoked potentials. Clin Neurophysiol 115:1114–1128. doi:10.1016/j.clinph.2003.11.016

Burg, JP. (1975) Maximum entropy spectral analysis. Ph.D Thesis, Stanford University

Capilla C (2006) Application of the Haar wavelet transform to detect microseismic signal arrivals. J Appl Geophys 59:36–46. doi:10.1016/j.jappgeo.2005.07.005

Goupillaud P, Grossmann A, Morlet J (1984) Cycle-octave and related transforms in seismic signal analysis. Geoexploration 23:85–102

Grossmann A, Morlet J (1984) Decomposition of hardy functions into square integrable wavelets of constant shape. SIAM J Math Anal 15: 723–736

Hafez AG, Kohda T Accurate P-wave arrival detection via MODWT. In: Computer Engineering & Systems, 2009. ICCES 2009. International Conference on, 14–16 Dec. 2009 2009. pp 391–396. doi:10.1109/ICCES.2009.5383235

Hafez AG, Rabie M, Kohda T (2013) Detection of precursory signals in front of impulsive P-waves. Digital Signal Process 23:1032–1039. doi:10.1016/j.dsp.2012.12.018

Kay S, Makhoul J (1983) On the statistics of the estimated reflection coefficients of an autoregressive process. IEEE Trans Acoust Speech Signal Process 31:1447–1455

Kay SM (1988) Modern spectral estimation : theory and application. Prentice-Hall signal processing series, Prentice Hall, Englewood Cliffs

Mathworks (2015) Compute estimate of autoregressive (AR) model parameters using Burg method - Simulink. http://www.mathworks.com/help/dsp/ref/burgarestimator.html. 2015

Miao X-G, Moon WM (1999) Application of wavelet transform in reflection seismic data analysis. Geosci J 3:171–179. doi:10.1007/BF02910273

Nason GP, Von Sachs R, Kroisandt G (2000) Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. J R Stat Soc Ser B (Stat Methodol) 62:271–292. doi:10.1111/1467-9868.00231

Percival DB, Walden AT (2000) Wavelet methods for time series analysis vol 4. Cambridge University Press

Stoev S, Taqqu MS (2003) Wavelet estimation for the Hurst parameter in stable processes. In: Rangarajan G, Ding M (eds) Processes with long-range correlations, vol 621. Lecture notes in physics. Springer, Berlin Heidelberg, pp. 61–87. doi:10.1007/3-540-44832-2_4

Tsakiroglou E, Walden AT (2002) From Blackman–Tukey pilot estimators to wavelet packet estimators: a modern perspective on an old

spectrum estimation idea. Signal Process 82:1425–1441. doi:10.1016/S0165-1684(02)00282-7
Unser M (1995) Texture classification and segmentation using wavelet frames. IEEE Trans Image Process 4:1549–1560
Wang S-Y, Liu X, Yianni J, Aziz TZ, Stein JF (2004) Extracting burst and tonic components from surface electromyograms in dystonia using adaptive wavelet shrinkage. J Neurosci Methods 139:177–184. doi:10.1016/j.jneumeth.2004.04.024
Whitcher B, Byers SD, Guttorp P, Percival DB (2002) Testing for homogeneity of variance in time series: long memory, wavelets, and the Nile River. Water Resour Res 38:12-11–12-16. doi:10.1029/2001wr000509