Journal of Applied Geophysics 117 (2015) 60–66

Contents lists available at ScienceDirect

Journal of Applied Geophysics

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

The L-DVV method for the seismic signal extraction

Pengjun Yu [a], Yue Li [a,*], Yuanlin Cheng [b], Hongbo Lin [a]

[a] Department of Information Engineering, Jilin University, Changchun, Jilin, China
[b] The Overall Design Part Seventh Research Institute, CASC, Chengdu, Sichuan, China

ARTICLE INFO

Article history:
Received 17 November 2014
Received in revised form 11 March 2015
Accepted 12 March 2015
Available online 13 March 2015

Keywords:
L-DVV
Nonlinearity
Surrogate data
QDV
Seismic signal processing

ABSTRACT

The delay vector variance (DVV) method enables analyzing the nonlinearity of a time series by the so-called scatter diagram, where the original time series is compared to the linearized version, the so-called surrogate. The surrogate is often constructed by the iterative amplitude adjusted Fourier transformation (iAAFT) method, so a linear signal in the DVV method can be generated by a linear time-invariant system in a white Gaussian noise environment. Owing to the iAAFT method and the linearity criterion, the DVV method is not sensitive to stochastic noise and the results are unstable, which may cause some confusion between the desired signal and the stochastic noise when processing the seismic signal. The now proposed method, the delay vector variance based on the straight line sequence (L-DVV) method, is essentially an extension of the DVV method, so that the straight line sequence is selected from the original signal and defined as the surrogate time series, which is used to construct the L-DVV scatter diagram based on the target variances calculated from both time series. This new scatter diagram is taken as criterion for testing the linearity of the analyzed time series. The analyses as to the stability of the method and its sensitivity to the stochastic noise are based on the quantification of the delay vector variance (QDV) or deviation with respect to bisector in the scatter diagram. These are the two key points – L-DVV scatter diagram and QVD – that we bear in mind when studying some examples of synthetic seismic signals contaminated by different noise levels and also real field data. The L-DVV method has high stability and strong sensitivity to the stochastic noise, and is able to discriminate accurately the seismic signal masked by stochastic noise, thus being a useful tool for the seismic signal processing in the exploratory practice.

1. Introduction

The now proposed method, the delay vector variance based on the straight line sequence (L-DVV) method, is essentially an extension of the delay vector variance (DVV) method and could discriminate accurately the seismic signal masked by the stochastic noise according to the criterion of the linearity. Analyzing the nonlinear nature of a time series has received considerable attention in recent years and has been widely applied in the fields of medicine, system analysis and so on (Yuan et al., 2009; Chen et al., 2008). The existing methods are twofold. In one case, different models, such as the AR model and the Volterra series, are built to determine the nonlinearity depending on the model error (Galka and Ozaki, 2001; Barahona and Poon, 1996). The models have high accuracy and low computational complexity only for a set order, while the performance would be poor as to an uncertain one. In the other case the time series is compared to the obtained for the linearized version, the so-called surrogate (Schreiber and Schmitz, 1996, 2000). The performance difference in some statistics could reflect the nonlinear nature (Schreiber and Schmitz, 1997; Manetti and Ceruso, 1997). Without modeling for the time series, the methods have stronger applicability. The kernel of these methods is the selection of the surrogate, which determines the accuracy and the criterion of the linearity.

The surrogate time series in the DVV method, which is generated by the iterative amplitude adjusted Fourier transformation (iAAFT) method (Gautama et al., 2004a), shares the exactly same probability density distribution and the roughly same power spectrum with the original one. As a result, the criterion of the linearity could be equivalent to a linear time-invariant system driven by white Gaussian noise, and it helps to distinguish the disease electroencephalogram and electrocardiogram signals from the normal ones (Chen et al., 2008). However, the robustness (which is not sensible in nature) to the stochastic noise and the instability of this method prevent the seismic signal extraction (Gautama et al., 2004b).

In the now proposed L-DVV method, the surrogate time series is a straight line sequence selected from the original time series to overcome the drawbacks in the seismic signal processing. The scatter diagram makes a comparison for the target variances of both times series and acts as the criterion for texting the linearity, which is the rectilinearity in this method. With strong sensitivity to the stochastic noise

* Corresponding author at: College of Department of Information Engineering, Jilin University, Changchun, Jilin 130012, China.
E-mail addresses: yupengjun1015@gmail.com (P. Yu), liyue84@gmail.com (Y. Li), 1739146223@qq.com (Y. Cheng), hblin@jlu.edu.cn (H. Lin).

http://dx.doi.org/10.1016/j.jappgeo.2015.03.016
0926-9851/

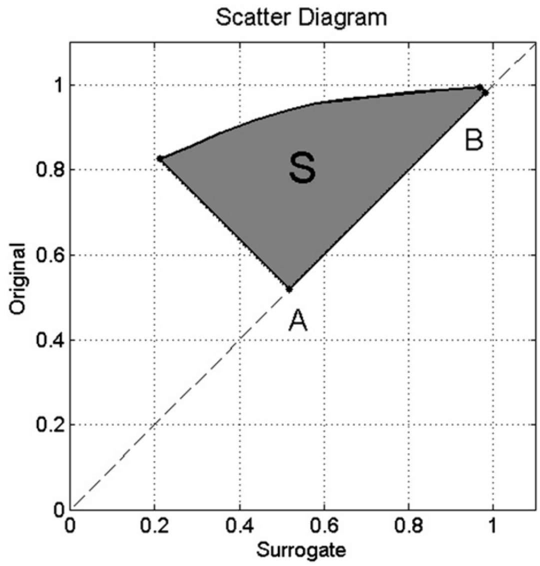

Fig. 1. Schematic representation of the scatter diagram used to quantify QDV, i.e., the delay vector variance or deviation with respect to bisector line.

and high stability, the L-DVV method could distinguish the noisy seismic signal with different noise intensities and extract the desired signal in the seismic records.

The rest of this paper is organized as follows. In Section 2, we address the introduction of the DVV and L-DVV methods. Section 3 addresses the stability analysis and the sensitivity analysis to the stochastic noise of the L-DVV method when compared with the DVV method. Section 4 focuses on the signal extraction in the seismic record. Finally, we present the conclusions in Section 5.

2. The DVV and L-DVV methods

The “surrogate” time series is generated as the realization of a certain null hypothesis, which is the linearity in the DVV method. If the DVV

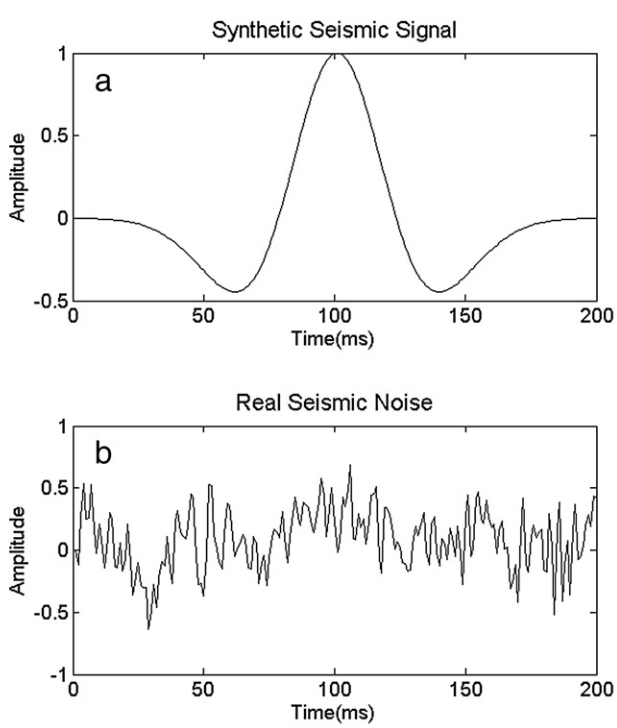

Fig. 2. Two time series graphs: (a) the Ricker wavelet; (b) real seismic noise.

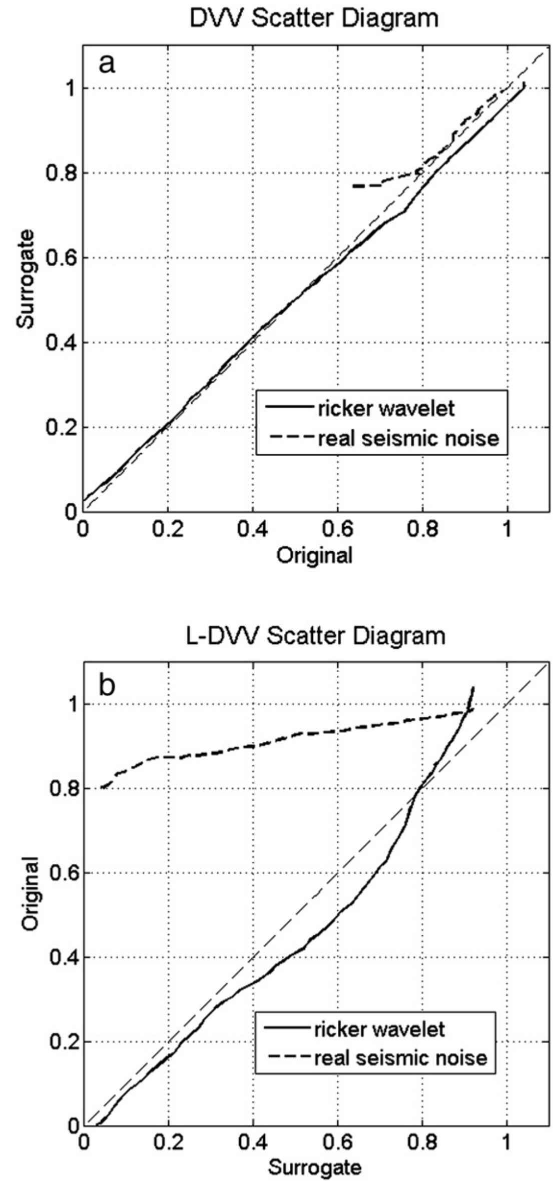

Fig. 3. Scatter diagrams obtained from the two time series shown in Fig. 2: (a) the DVV scatter diagram; (b) the L-DVV scatter diagram.

plot of the surrogate time series and that of the original time series are almost the same which results in coinciding with the bisector line in the scatter diagram, the null hypothesis is correct and the original time series is considered linear in nature, while the scatter diagram of a nonlinear signal has an obvious deviation with respect to the bisector line. The surrogate time series, which is the kernel of this method, is often generated by the iAAFT method. As a result, the DVV method may cause some confusion between the desired seismic signal and the

Table 1
Stability analysis based on quantification of the delay vector variance (QDV) estimated by the DVV and L-DVV methods.

	Ricker wavelet		Real seismic noise	
	QDV	L-QDV	QDV	L-QDV
1st	0.0169	0.0431	0.0103	0.2583
2nd	0.0182	0.0430	0.0118	0.2585
3th	0.0199	0.0430	0.0134	0.2583
4th	0.0203	0.0430	0.0097	0.2585
5th	0.0185	0.0433	0.0103	0.2584

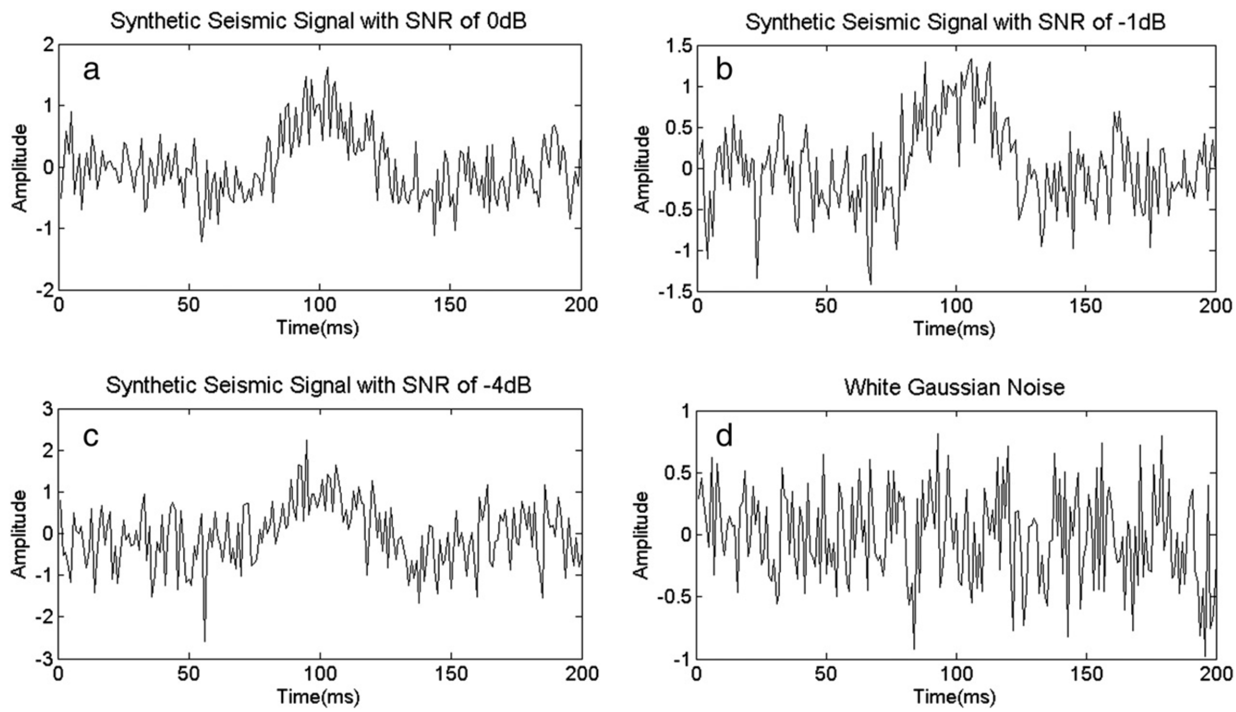

Fig. 4. Time series graphs of noisy signals with different SNRs: (a) noisy signal with SNR of 0 dB; (b) noisy signal with −1 dB; (c) noisy signal with −4 dB; (d) pure white Gaussian noise.

stochastic noise. The L-DVV method selects the straight line sequence as the surrogate time series and guarantees the stability and the sensitivity to the noise, which has huge potential in the seismic signal extraction. We next start with the description of the DVV and L-DVV methods and proceed with the theoretical proof of the criterion of the linearity in theory for the now proposed method.

For the optimal embedding dimension m and time lag τ, a time series $\{x_i\}(i = 1, 2, \cdots, N)$ can generate the delay vectors (DVs):

$$\vec{x}(k) = \left[x_{k-m\tau}, x_{k-(m-1)\tau}, \cdots, x_{k-\tau}\right]^T. \tag{1}$$

$\vec{x}(k)$ is the kth delay vector in the reconstructed phase space, while x_k is the corresponding target. So the DVV method can be summarized as follows.

Step 1) In the reconstructed phase space, the pairwise Euclidean distances between DVs can be computed as follows:

$$d = \left\|\vec{x}(i) - \vec{x}(j)\right\|(i \neq j). \tag{2}$$

Then the mean μ_d and the standard deviation σ_d of the distances are computed. A set $\Omega_k(r_d)$ is generated by grouping those DVs that are within a certain Euclidean distance, r_d, to the $\vec{x}(k)$:

$$\Omega_k(r_d) = \left\{\vec{x}(i) \middle| \left\|\vec{x}(k) - \vec{x}(i)\right\| \leq r_d\right\} \tag{3}$$

where r_d is taken from the interval $[\max\{0, \mu_d - n_d\sigma_d\}, \mu_d + n_d\sigma_d]$, being n_d the parameter controlling the span.

Step 2) For a set $\Omega_k(r_d)$, the variance of the corresponding targets σ_k^2 is computed, normalized by the variance of the time series σ_s^2 and yields the "target variance", where N is the number of the $\Omega_k(r_d)$:

$$\sigma_{(rd)}^2 = \left(1/N\sigma_s^2\right)\sum_{k=1}^{N}\sigma_k^2. \tag{4}$$

The DVV plot (the target variance, $\sigma_{(rd)}^2$) is a function of the standardized distance$(r_d - \mu_d)/\sigma_d$.

Step 3) For the given original time series, the surrogate time series of the DVV method is generated by the iAAFT method (Schreiber and Schmitz, 2000). Then the target variance of the surrogate time series $\sigma_{(rd)}^{*2}$ is computed according to the first two steps.

Step 4) The scatter diagram is constructed in the way where the horizontal axis corresponds to the DVV plot of the original time series, and the vertical to that of the surrogate time series. A linear signal results in the scatter diagram coinciding with the bisector line, whereas a nonlinear signal has a deviation from the bisector line.

It is the phase randomization surrogate time series generated by the iAAFT method that makes the scatter diagram unstable. The surrogate time series shares the exactly same probability density distribution and the roughly same power spectrum with the original one, which makes some confusion between the desired seismic signal and the noise since the stochastic noise (such as white Gaussian noise) has roughly even power spectrum (Schreiber and Schmitz, 2000).

The now proposed L-DVV method is an extension of the DVV method. The first two steps are the same with the traditional method, and Steps 3 and 4 could be summarized as follows.

Step 3) For the set original time series $\{x_i\}(i = 1, 2, \cdots, N)$, the surrogate time series of the L-DVV method $\{s_i\}(i = 1, 2, \cdots, N)$ is selected as the follows:

$$s_i = i \cdot \max(\{x_i\})/N, 1 \leq i \leq N. \tag{5}$$

With the same embedding dimension m and time lag τ, the DVs of the surrogate time series are constructed as follows:

$$\vec{s}(k) = \left[s_{k-m\tau}, s_{k-(m-1)\tau}, \cdots, s_{k-\tau}\right]^T. \tag{6}$$

$\vec{s}(k)$ is the kth delay vector in the reconstructed phase space, while s_k is the corresponding target. Sharing the same n_d and

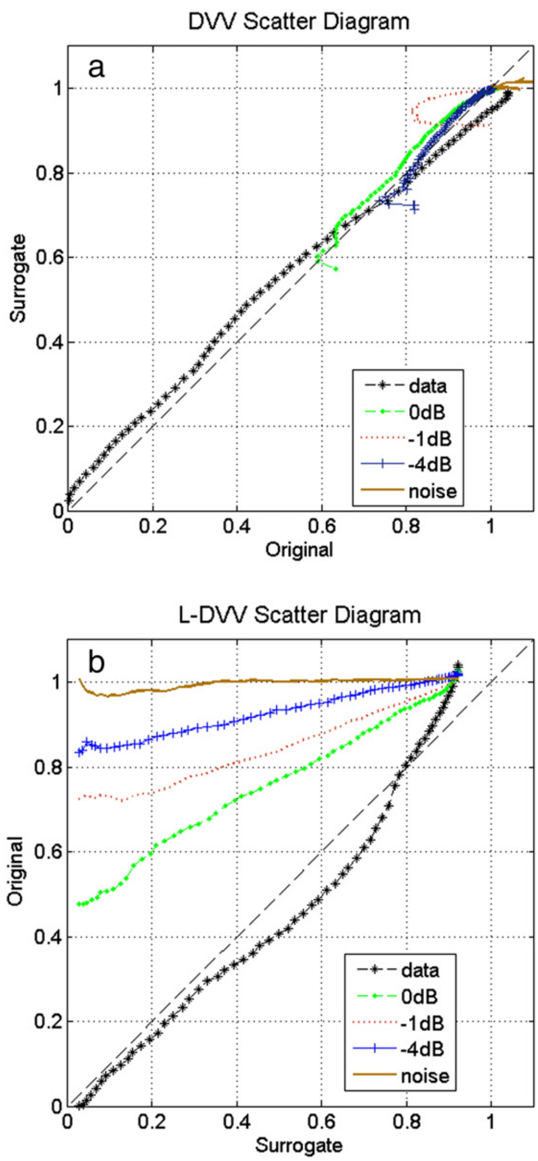

Fig. 5. Scatter diagrams obtained from the time series shown in Fig. 4 by applying: (a) the DVV method; (b) the L-DVV method.

r_d with the original time series, the $\Omega_k'(r_d)$ is constructed by grouping those DVs that are within r_d to the $\vec{s}(k)$:

$$\Omega_k'(r_d) = \left\{\vec{s}(i)\middle|\left\|\vec{s}(k)-\vec{s}(i)\right\|\le r_d\right\}. \quad (7)$$

With the variance of the corresponding targets $\sigma_k'^2$ and that of the time series σ_s^2, the "target variance" of the surrogate time series is calculated as follows:

$$\sigma'^2_{(rd)} = \left(1/N\sigma_s^2\right)\sum_{k=1}^{N}\sigma'^2_k. \quad (8)$$

Step 4) The scatter diagram of the L-DVV method is constructed in the way opposite to the DVV method, where the horizontal axis corresponds to the DVV plot of the surrogate time series and the vertical to that of the original time series. This has a reasonable vision effect without influencing the analysis. The deviation with respect to bisector line still reflects the degree of the nonlinearity.

Table 2
QVD's and L-QVD's of the synthetic noisy signals (Fig. 4) after being estimated by the DVV and L-DVV methods, respectively.

	White Gaussian noise		Real seismic noise	
	QDV	L-QDV	QDV	L-QDV
Signal	0.0306	0.0432	0.0203	0.0431
0 dB	0.0184	0.1694	0.0204	0.0860
−1 dB	0.0475	0.2221	0.0143	0.1001
−4 dB	0.0112	0.2712	0.0183	0.1427
Noise	0.0134	0.3099	0.0103	0.2583

A linear time series could be composed of similar slope straight lines, while a nonlinear one could not. So the L-DVV method is sensitive to the stochastic noise and is able to discriminate accurately the seismic signal masked by the stochastic noise.

For the given embedding dimension m and time lag τ, a linear time series $\{x_1, x_2, \cdots x_n\}$ can be described as $x_n = a \cdot n + C$, where $a \neq 0$ and C are constant. The theoretical proof for the criterion of the linearity selected by the L-DVV method can be described as follows:

Step 1) The pairwise Euclidean distance between any two distinct DVs can be described as follows:

$$\begin{aligned} d_{ij} = \|\vec{x}(i) - \vec{x}(j)\| &= \sqrt{\sum_{n=1}^{m}\left(x_{i-n\tau}-x_{j-n\tau}\right)^2} \\ &= \sqrt{\sum_{n=1}^{m}\{[a(i-n\tau)+C]-[a(j-n\tau)+C]\}^2} \\ &= \sqrt{m}|a(i-j)|. \end{aligned} \quad (9)$$

The mean μ_d and the standard deviation σ_d of all the distances could be calculated and are proportional to a. The influence of a can be eliminated after normalizing the distance by $(r_d - \mu_d)/\sigma_d$ in Step 2 of the L-DVV method.

Step 2) Then we focus on the selection of the $\Omega_k(r_d)$:

$$\Omega_k(r_d) = \left\{x_i\middle|\left\|\vec{x}(k)-\vec{x}(i)\right\|\le r_d\right\} = \left\{x_i\middle||k-i|\le r_d/|a|\sqrt{m}\right\}. \quad (10)$$

Since the value of $r_d/a\sqrt{m}$ does not change with a, a and C are irrelevant to the selection of the $\Omega_k(r_d)$.

Step 3) Finally, the final target variances of both time series would have no effect upon a and C after normalizing by the variance of the time series through Eq. (8).

So any time series, which can be described as $x_n = a \cdot n + C$, $a \neq 0$with the same length, has the scatter diagram coinciding with the bisector line and meets the criterion of the linearity selected by the L-DVV method.

3. Performance analysis of the L-DVV method

The analyses for the stability and the sensitivity to the stochastic noise are introduced. To have the value of the parameters for the DVV and L-DVV methods, the Cao method is used to get the optimal m, and $\tau = 1$ is selected with the need of the continuity in time domain (Cao, 1997).

Stability plays a crucial role in the nonlinearity analysis. The scatter diagram makes the nonlinear nature visual, but does not quantify the analysis. The quantification of the delay vector variance (QDV) or deviation with respect to bisector in the scatter diagram makes the analysis

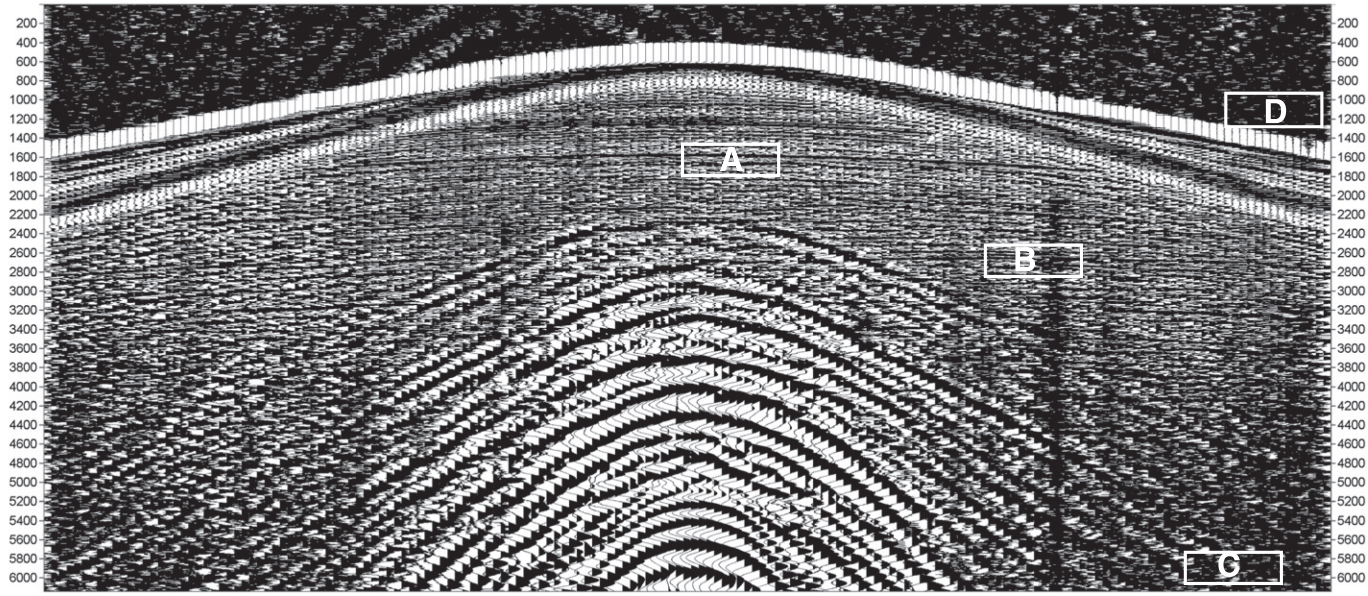

Fig. 6. Real seismic record composed by 168 traces with recording length of 6 s. The rectangular boxes A, B, C and D enclose the field data that are analyzed in this study.

more accurate (Yuan et al., 2009). It can be summarized as follows:

$$QDV = S/d. \tag{11}$$

As shown in Fig. 1, S expresses the area surrounded by the scatter curve and the bisector line, while d expresses the distance between A and B that are the projected points to the endpoints of the scatter curve. So QDV could quantify the level of the nonlinearity and simplify the analysis.

The synthetic seismic record is often composed of Ricker wavelets for the seismic signal processing (Zhuang et al., 2015). As shown in Fig. 2, the Ricker wavelet and the real seismic noise of 200 points are selected to establish the stability of the L-DVV method. The results of the DVV method are also list as a reference. The scatter diagrams of both methods are shown in Fig. 3. Table 1 shows the quantifiable results of the 5 repetitive nonlinearity analyses for both sequences. The DVV and L-DVV scatter diagrams, the QDVs and L-QDVs are named to distinguish the DVV and L-DVV methods in the qualitative and quantitative analyses.

The erratic fluctuation of the QDVs shown in Table 1 reflects the instability of the DVV method, which would lead to the significant error when extracting the desired signals in the seismic records. Correspondingly, the L-DVV method has the similar L-QDVs and the error is so small that could be supposed as the quantization error. Then the scatter diagram in Fig. 3(b) shows that the L-DVV method could discriminate between the Ricker wavelet and the real seismic noise accurately, while

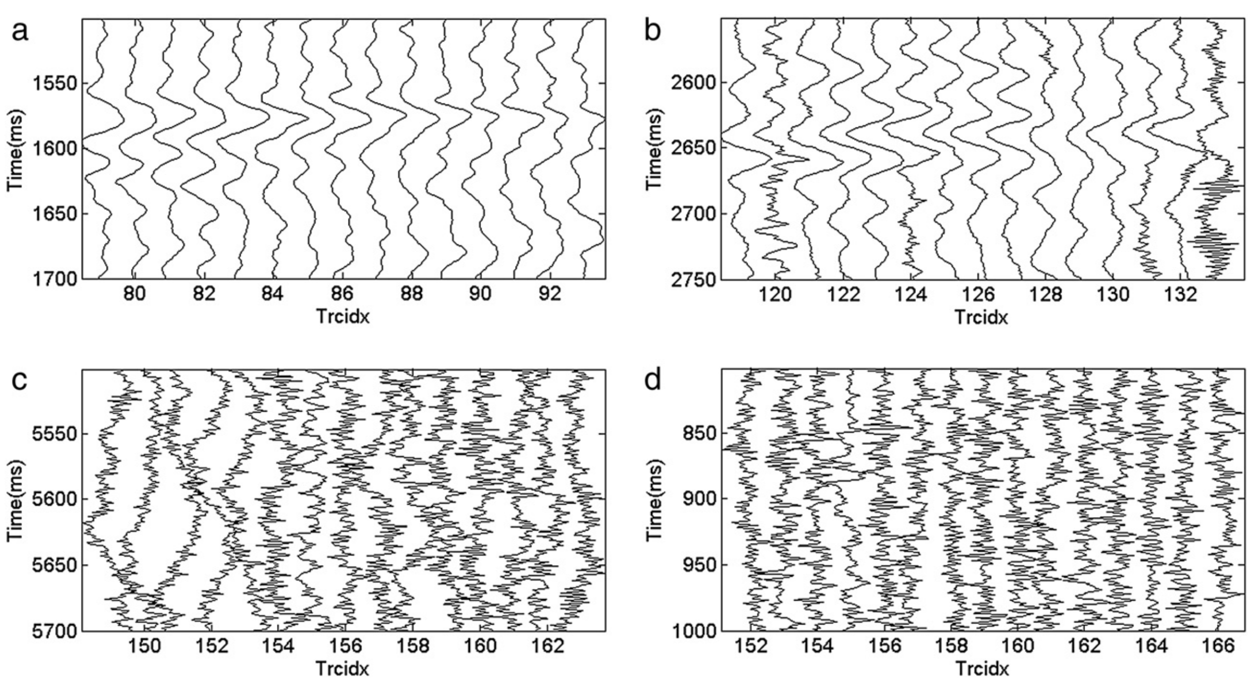

Fig. 7. Zoomed view of the seismic signals contained in the boxes indicated in Fig. 6: the plots a, b, c and d correspond to the seismic portions A, B, C and D, respectively.

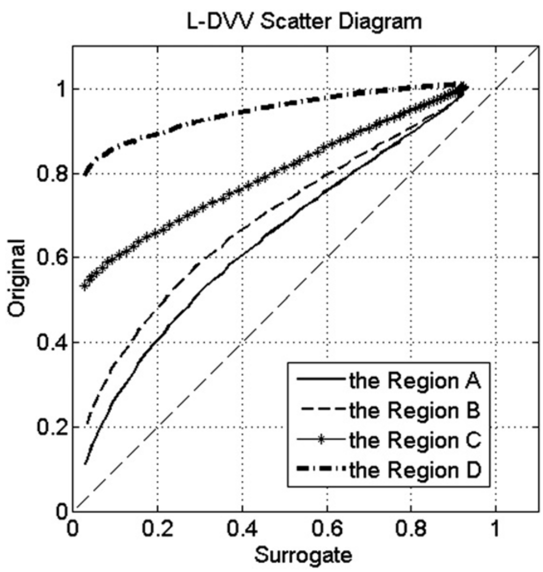

Fig. 8. The L-DVV scatter diagrams integrated by the averaged scatter curves calculated for each of the four previously selected regions, which measure the nonlinearity of the traces within these regions.

the DVV method almost confuses these sequences in Fig. 3(a). The L-DVV method has a better performance in the stability aspect and lead potential for the seismic signal extraction.

The Ricker wavelet shown in Fig. 2(a) is selected as the synthetic seismic signal and respectively embedded in white Gaussian noise with the signal to noise ratio (SNR) of 0 dB, −1 dB, −4 dB (shown in Fig. 4). The DVV and L-DVV methods measure the nonlinearity for the pure signal, noisy signals and pure noise. The scatter diagrams of both methods are shown in Fig. 5. Then white Gaussian noise is changed into the real seismic noise shown in Fig. 2(b) and the experiments above are repeated. The quantitative results are shown in Table 2. According to the scatter diagrams and the quantitative results, the sensitivity to the stochastic noise for the both methods is shown obviously.

The DVV method confuses the curves of the noisy signals with those of the pure signal and noise, since the deviation level with respect to bisector could not be compared to the naked eye. The difference of the QDVs is too subtle to discriminate the noisy signals with different SNRs, especially when considering the poor stability of the DVV method. The robustness of this method is still suitable to the seismic noise. The synthetic seismic signal is linear in nature so that this method is ineffective in the seismic signal extraction. Correspondingly, the curve of the pure signal in the L-DVV scatter diagram coincides with the bisector line most which corresponds to the most linear, while the curve of the pure noise has the biggest deviation and reflects the most nonlinear in nature. The curves of the noisy signals fall in the region bounded by the two curves above and the deviation level increases with the falling of SNR. The L-QDVs shown in Table 2 are more obvious to indicate the relationship between the nonlinearity and SNR. The L-QDVs of the noisy signals decrease with the rising SNR of the both noise and are still intermediate between those of the pure signal and noise. The difference of the L-QDVs between the noisy signals and the pure noise is still apparent with the SNR of −4 dB, which meets the need of the seismic signal processing.

The analyses above show that the L-DVV method is more stable and has a stronger sensitivity to the stochastic noise, which could distinguish the desired seismic signal masked by the stochastic noise.

Table 3
QVD's and L-QVD's of the seismic traces within the boxes A, B, C, D (Fig. 6) estimated by the L-DVV method.

Trace	L-QDV			
	A	B	C	D
1st	0.1079	0.0936	0.1233	0.2571
2nd	0.1187	0.2057	0.0933	0.2901
3th	0.1191	0.0957	0.1255	0.2609
4th	0.1421	0.1065	0.1113	0.2045
5th	0.0711	0.1321	0.1293	0.2781
6th	0.0603	0.1396	0.2330	0.2556
7th	0.0644	0.1566	0.2570	0.3179
8th	0.0942	0.1312	0.2801	0.2595
9th	0.0781	0.1365	0.1642	0.3044
10th	0.0712	0.0893	0.1583	0.2950
11th	0.0855	0.0997	0.2372	0.2905
12th	0.1123	0.0934	0.2857	0.3487
13th	0.0787	0.1144	0.2454	0.2468
14th	0.1274	0.0537	0.2161	0.2943
15th	0.1052	0.2053	0.2243	0.2912
Average	0.0957	0.1235	0.1923	0.2796

4. Field data processing

The field seismic data, a common-shot-point record with 168 traces, record length 6 s, sampled frequency 1000 Hz and the geophone interval 30 m, is acquired from a certain zone of China shown in Fig. 6 to test the practicality of the L-DVV method. The rectangle boxes in Fig. 6 are representative in the record, each of which has 15 traces and 200 points in each trace as shown in Fig. 7 in detail. The record in region A contains the reflection events in all the traces, while there are a few pure noise traces in region B. The few traces of the desired signals are masked in the strong stochastic noise in region C and the reflection events almost could not be recognized by the naked eye. Finally, the record in region D contains the seismic noise only. The nonlinearity of the traces in the four regions is measured by the L-DVV method, and the averaged scatter curves for each region are shown in Fig. 8, while the quantifiable results are listed in Table 3. The curve of region A coincides with the bisector line most, while the curves of regions B and C deviate from the bisector line with the increase of the noise intensity. These three curves are all quite closer to the bisector line than that of region D. The data shown in Table 3 also supports the analysis above. So the L-DVV method could allow extracting the seismic signal from the original record.

5. Conclusion

The now proposed L-DVV method is an extension of the DVV method, where the straight line sequence is selected from the original signal and defined as the surrogate time series to measure the nonlinearity. This method is sensitive to the stochastic noise and the results are more stable, so that it could discriminate accurately the seismic signal masked by stochastic noise according to QDV or deviation with respect to bisector in the scatter diagram. This is an exploratory practice to reduce the misjudgments caused by the naked eye for the seismic signal extraction.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant 41130421 and Grant 41274118).

References

Barahona, M., Poon, C.S., 1996. Detection of nonlinear dynamics in short, noisy time series. Nature 381, 215–232. http://dx.doi.org/10.1038/381215a0 (May).

Cao, L., Feb. 1997. Practical method for determining the minimum embedding dimension of a scalar time series. Physica D 110 (1), 43–50. http://dx.doi.org/10.1016/S0167-2789(97)00118-8.

Chen, M., Gautama, T., Mandic, D.P., Jan. 2008. An assessment of qualitative performance of machine learning architectures: modular feedback networks. IEEE Trans. Neural Netw. 19 (1), 183–189. http://dx.doi.org/10.1109/TNN.2007.902728.

Galka, A., Ozaki, T., Oct. 2001. Testing for nonlinearity in high-dimensional time series from continuous dynamics. Physica D 158, 32–41. http://dx.doi.org/10.1016/S0167-2789(01)00318-9.

Gautama, T., Mandic, D.P., Van Hulle, M., 2004a. A novel method for determining the nature of time series. IEEE Trans. Biomed. Eng. 51 (5), 728–736. http://dx.doi.org/10.1109/TBME.2004.824122 (May).

Gautama, T., Mandic, D.P., Van Hulle, M., Apr. 2004b. The delay vector variance method for detecting determinism and nonlinearity in time series. Physica D 190, 167–176. http://dx.doi.org/10.1016/j.physd.2003.11.001.

Manetti, C., Ceruso, M.A., Jan. 1997. Recurrence quantification analysis as a tool for the characterization of molecular dynamics simulations. Phys. Rev. Lett. 59 (1), 992–998. http://dx.doi.org/10.1103/PhysRevE.59.992.

Schreiber, T., Schmitz, A., Jul. 1996. Improved surrogate data for nonlinearity tests. Phys. Rev. Lett. 77 (4), 635–638. http://dx.doi.org/10.1103/PhysRevLett.77.635.

Schreiber, T., Schmitz, A., 1997. On the discrimination power of measures for nonlinearity in a time series. Phys. Rev. E 55 (5), 5443–5447. http://dx.doi.org/10.1103/PhysRevE.55.5443 (May).

Schreiber, T., Schmitz, A., Aug. 2000. Surrogate time series. Physica D 142, 346–382. http://dx.doi.org/10.1016/S0167-2789(00)00043-9.

Yuan, Y., Li, Y., Yu, L., Guo, H., Jun. 2009. Difference of degree of nonlinearity between normal and epileptic EEG signals, in Proc. Int. Conf. Signal Process(9th ICSP'08). pp. 2162–2165 http://dx.doi.org/10.1109/ICOSP.2008.4697575.

Zhuang, G., Li, Y., Liu, Y., Lin, H., Feb. 2015. Varying-window-length TFPF in high-resolution radon domain for seismic random noise attenuation. IEEE Geosci. Remote Sens. Lett. 12 (2), 404–408. http://dx.doi.org/10.1109/LGRS.2014.2345126.