STUDY OF LEAST SQUARES SUPPORT VECTOR REGRESSION FILTERING TECHNOLOGY WITH A NEW 2D RICKER WAVELET KERNEL

XIAOYING DENG[1], DINGHUI YANG[2], TAO LIU[3] and BAOJUN YANG[4]

[1] Department of Electronic Engineering, Beijing Institute of Technology, Beijing 100081, P.R. China. xydeng@bit.edu.cn
[2] Department of Mathematical Sciences, Tsinghua University, Beijing 100084, P.R. China. dhyang@math.tsinghua.edu.cn
[3] Military Representative Office of PLA in Second Research Institute of CASIC, Beijing 100854, P.R. China.
[4] Department of Geophysics, Jilin University, Changchun 130026, P.R.China. yangbaojun@jlu.edu.cn

(Received July 22, 2010; revised version accepted December 21, 2010)

ABSTRACT

Deng, X., Yang, D., Liu, T. and Yang, B., 2011. Study of least squares support vector regression filtering technology with a new 2D Ricker wavelet kernel. *Journal of Seismic Exploration*, 20: 161-176.

To suppress the random noise in seismic data, the least squares support vector regression (LS-SVR) filtering technology with a new 2D Ricker wavelet kernel is proposed in this paper. Firstly, we prove that the 2D Ricker wavelet kernel is an admissible support vector kernel. The proposed 2D Ricker wavelet kernel takes into account the characteristics of seismic data in the time-space domain. And the kernel parameters of the 2D Ricker wavelet kernel reflect the dominant frequency of seismic data in time domain and the wavenumber of seismic data in space domain, which will help the difficult problem of parameters selection for LS-SVR. Then by solving a quadratic optimization problem with constrains, we can obtain the regression function so as to compute the filtered output. The simulation experiments on synthetic records show that compared with the LS-SVR using 1D Ricker wavelet kernel and the common f-x prediction filtering method, the proposed method can suppress the random noise more efficiently, and enhance the continuity of events greatly. An example on a real seismic data processing also proves the effectiveness of the proposed method. So the LS-SVR with the 2D Ricker wavelet kernel can be used to attenuate the random noise in seismic data.

KEY WORDS: 2D Ricker wavelet kernel, least squares support vector regression, seismic data, random noise reduction.

0963-0651/11/$5.00

INTRODUCTION

For both the prestack and stacked seismic data, random noise reduction is a very important part of seismic data processing. Many filtering methods have been studied for the random noise suppression of seismic data. The f-x prediction filtering (Canales, 1984; Sacchi and Kuehl, 2000) introduced to extract linear features and suppress random noise for seismic data. However, such an algorithm is not successful in processing seismic data with nonlinear events. The median filter (Bednar, 1983; Duncan and Beresford, 1995; Liu et al., 2009) is a nonlinear processing method, which can efficiently suppress the spike-like noise. But it works less well for the white Gaussian noise. Local singular value decomposition proposed by Bekara et al. (2007) is better than f-x deconvolution and median filtering in removing background noise, but it performs less well in enhancing weak events or events with conflicting dips. Lu (2006) proposed an adaptive SVD filter to enhance the non-horizontal events by detection of seismic image texture direction and then horizontal alignment of the estimated dip through data rotation. In addition, some technologies such as Karhunen-Loeve transform (Jones and Levy, 1987), forward-backward linear prediction (Wang, 1999), complex-trace analysis (Karsli et al., 2006), and empirical mode decomposition (Bekara and Baan, 2009) have also been presented and given good results. The above different methods have different principles, application conditions and limits.

SVM is a machine learning method based on the statistical learning theory (Vapnik, 1995). It has been widely applied to the classification and regression. SVM for regression is also named as support vector regression (SVR). Least squares SVR (LS-SVR) (Sunkens and Vandewalle, 1999) is a simplified version for the standard SVR. We have proposed the LS-SVR with 1D Ricker wavelet kernel (marked with 1D SVR) for noise reduction (Deng et al., 2009). The results show that the 1D SVR method works better than the adaptive Wiener filtering and wavelet transform-based method. Based on the 1D SVR method, we develop the LS-SVR with 2D Ricker kernel (marked with 2D SVR) in this paper. Firstly, we create a 2D Ricker wavelet kernel according to the features of seismic data in time-space domain. Then the theoretical frame of 2D SVR is given. Finally, for testing the performance of the proposed method, we compare the proposed method with the 1D SVR method and the commonly used f-x prediction filtering by experiments on the synthetic and real seismic data.

SVR WITH 2D RICKER WAVELET KERNEL

2D Ricker wavelet kernel

The Ricker wavelet is often used as the seismic wavelet in the seismic simulation computations. According to the expression of Ricker wavelet function in the time domain $y(t) = (1 - 2\pi^2f^2t^2)e^{-\pi^2f^2t^2}$ (f is the dominant frequency), we have proposed the Ricker wavelet kernel

$$K(t_1,t_2) = (1 - 2\pi^2f^2\|t_1-t_2\|^2)e^{-\pi^2f^2\|t_1-t_2\|^2} , \tag{1}$$

(f is the kernel parameter) and we have proved that it is an admissible support vector kernel (Deng et al., 2009). Because t = d/v and f = vk (v is wave speed, d is offset, and k is the wavenumber), we have

$$\begin{aligned} y(t) &= [1 - 2\pi^2f^2t^2]e^{-\pi^2f^2t^2} \\ &= [1 - 2\pi^2(vk)^2(d/v)^2]e^{-\pi^2(vk)^2(d/v)^2} \\ &= [1 - 2\pi^2k^2d^2]e^{-\pi^2k^2d^2} . \end{aligned} \tag{2}$$

So in space domain, we get

$$y(d) = [1 - 2\pi^2k^2d^2]e^{-\pi^2k^2d^2} \quad \text{(k is the wavenumber)}. \tag{3}$$

Now we make the function in the space domain

$$K(d_1,d_2) = [1 - 2\pi^2k^2\|d_1 - d_2\|^2]e^{-\pi^2k^2\|d_1-d_2\|^2} . \tag{4}$$

(k is the kernel parameter)

Obviously, eq. (4) has the same expressional forms as eq. (1), so this function in the space domain is also an admissible support vector kernel. The following proposition (Cristianini and Shawe-Taylor, 2000) allows us to create more complicated kernels from simple kernels.

Proposition 1. Let K_1 and K_2 be kernels over $X \times X$, $X \in R^n$. Then the function $K(\mathbf{x},\mathbf{z}) = K_1(\mathbf{x},\mathbf{z})K_2(\mathbf{x},\mathbf{z})$ is kernel.

In accordance with proposition 1, from the kernels (1) and (4), we create the following 2D Ricker wavelet kernel

$$\begin{aligned} K(\mathbf{x},\mathbf{z}) = &\{[1 - 2\pi^2f^2(t_1 - t_2)^2]e^{-\pi^2f^2(t_1-t_2)^2}\} \\ &\times \{[1 - 2\pi^2k^2(d_1 - d_2)^2]e^{-\pi^2k^2(d_1-d_2)^2}\} , \end{aligned} \tag{5}$$

where the 2D input vector $\mathbf{x} = (t,d)$ and the kernel parameter vector $\mathbf{p} = (f,k)$. This 2D Ricker wavelet kernel embodies the features of the seismic wave in the time-space domain. Furthermore, the kernel parameter vector reflects the physical meaning of the seismic wave in the frequency domain and the wavenumber domain, which will avoid the blind selection of kernel parameters such as the exhaustive search method. The appropriate parameter selection has been a vexed question ever since the SVM was proposed in 1995 (Cherkassky and Ma, 2004).

LS-SVR with 2D Ricker wavelet kernel

Given a seismic data set of l points $\{\mathbf{x}_i, y_i\}_{i=1}^{l}$ with input data $\mathbf{x}_i = (t_i, d_i) \in R^2$ and output $y_i \in R$, a regression function for LS-SVR is given by

$$f(\mathbf{x}) = \omega \cdot \varphi(\mathbf{x}) + b \quad , \tag{6}$$

where the nonlinear mapping $\varphi(\mathbf{x})$ maps the input space to a so-called feature space, $\omega \cdot \varphi(\mathbf{x})$ is the inner product between the weight vector ω and $\varphi(\mathbf{x})$, and $b \in R$ is the bias term. The optimization problem of the LS-SVR (Suykens and Vandewalle, 1999) is defined as follows

$$\min\{\tfrac{1}{2}\|\omega\|^2 + (\gamma/2)\sum_{i=1}^{l} e_i^2\} \quad , \tag{7}$$

$$\text{s.t. } y_i = \omega \cdot \varphi(\mathbf{x}_i) + b + e_i, \quad i = 1,\cdots,l.$$

where γ is a regularization parameter. The Lagrangian function of the above problem is

$$L(\omega, e, b, \alpha) = \tfrac{1}{2}\|\omega\|^2 + (\gamma/2)\sum_{i=1}^{l} e_i^2 - \sum_{i=1}^{l} \alpha_i[\omega \cdot \varphi(\mathbf{x}_i) + b + e_i - y_i] \quad , \tag{8}$$

with Lagrange multipliers $\alpha_i \in R$ $(i = 1,2,\cdots,l)$. Using the optimality conditions of problem (7) and solving the corresponding system of linear equations, we obtain

$$b = \mathbf{1}^T(\mathbf{A}^{-1})^T\mathbf{y}/\mathbf{1}^T\mathbf{A}^{-1}\mathbf{1} \quad , \quad \alpha = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{1}b) \quad , \tag{9}$$

where $\mathbf{A} = \mathbf{\Omega} + \gamma^{-1}\mathbf{I}$, $\mathbf{\Omega}$ is a kernel matrix in which the element $\Omega_{ij} = \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$ $(i,j = 1,2,\ldots,l)$, $\mathbf{I}$ is a unit matrix, $\mathbf{y} = [y_1, y_2, \ldots, y_l]^T$, and $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_l]^T$. Then the regression function can be expressed as

$$f(\mathbf{x}) = \omega\cdot\varphi(\mathbf{x}) + b = \sum_{i=1}^{l} \alpha_i\varphi(\mathbf{x}_i)\cdot\varphi(\mathbf{x}) + b \quad . \tag{10}$$

Note that we can replace the inner product in the above formula by a kernel function that satisfies Mercer's condition (Mercer, 1909), i.e., $K(\mathbf{x}_i,\mathbf{x}_j) = \varphi(\mathbf{x}_i)\cdot\varphi(\mathbf{x}_j)$. In such a case, the nonlinear regression function becomes

$$f(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i K(\mathbf{x}_i,\mathbf{x}) + b \quad . \tag{11}$$

Then for all the inputs $\mathbf{x}_i = (t_i,d_i)$ $(i = 1,2,\ldots,l)$, we compute the regression outputs $f(\mathbf{x}_i)$ $(i = 1,2,\ldots,l)$ using the regression function, namely the denoised results.

SIMULATION EXPERIMENTS

To investigate the denoising performance of the proposed 2D SVR, we test the method on synthetic seismic records, and compare it with the 1D SVR and the f-x prediction filtering. Based on the parameters listed in Table 1, we synthesize a noise-free record with 100 traces. The geophone interval is 10 m. The record simulates a model with a faulted layer and a thin layer as corresponding events. Fig. 1a draws the synthetic record every three traces. Then random noise with different types and levels are added on the original record, including the white Gaussian noise added on traces from 1 to 35, the spike-like noise added on traces from 36 to 70, and the uniformly distributed noise added on traces from 71 to 100. The above three different denoising methods are used to suppress the random noise. Note that a windowing algorithm is used in all three methods. The selected parameters are as follows:

Table 1. Parameters used to generate the synthetic record.

Event	t_0 (s)	A (V)	v (m/s)	f (Hz)
1	0.5	1	1000	32
2	1	0.9	1500	30
3	1.04	0.85	1540	28
4	1.4	0.8	2500	22

for the 1D SVR, the window length is 200 samples (overlapping 10-samples border), $\gamma = 1$ and $f = 30$; for the 2D SVR, the window size is 200 samples by 10 traces (overlapping 10-by-1 border), $\gamma = 1$, $f = 30$, and $k = 0.05$; for the f-x prediction filtering, we only process the frequencies fallen in [0, 100] Hz and the amplitudes for other frequencies components are set as zero, and the window size is 200 samples by 10 traces (overlapping 10-by-1 border) and the bilateral prediction filter length is 7. Here for the 1D and 2D SVR methods, we use the SVR with implicit bias term (Deng et al., 2010).

Table 2 lists the mean square errors (MSE) between the noise-free record and the denoised record and the average signal-to-noise ratios (SNR) of the whole record before and after denoisng. From Table 2, the proposed method can always obtain the lowest MSE and the highest SNR when the SNR of noisy records gradually declines.

Table 2. Comparisons for MSE and SNR obtained by using different denoising methods.

SNR before denoising (dB)	1D SVR		2D SVR		*f-x* prediction	
	MSE(V^2)	SNR(dB)	MSE(V^2)	SNR(dB)	MSE(V^2)	SNR(dB)
6.43	0.0033	18.28	0.0021	19.56	0.010	10.59
4.66	0.0044	16.22	0.0030	17.49	0.011	10.42
2.53	0.0061	14.61	0.0042	15.92	0.015	8.58
0.73	0.0087	11.88	0.0055	13.90	0.016	8.24

The following gives a set of experimental results. Fig. 1b shows the noisy record with SNR 2.53 dB. The strong random noise masks much useful information such as events and the position of the fault layer, especially for the traces from 1 to 35, where the white Gaussian noise is so strong that the events are completely invisible. Figs. 1c-e show the results denoised by using the 1D SVR, 2D SVR and f-x prediction filtering, respectively. In Figs. 1c and 1d, the random noise is greatly suppressed, and more completely in Fig. 1d than Fig. 1c. Obviously, though most of the random noise in Fig. 1e is also suppressed, some useful seismic wavelets are badly attenuated, such as traces 1, 22, 40 and 100.

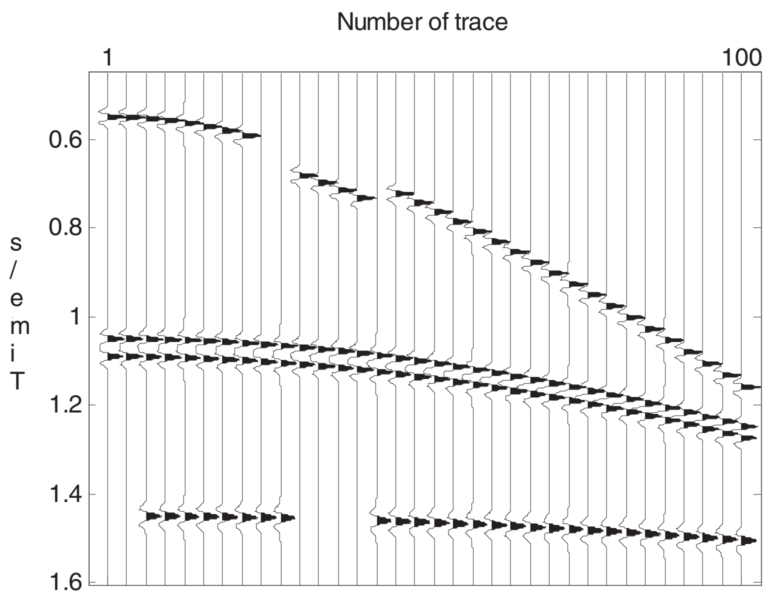


Figure 1a

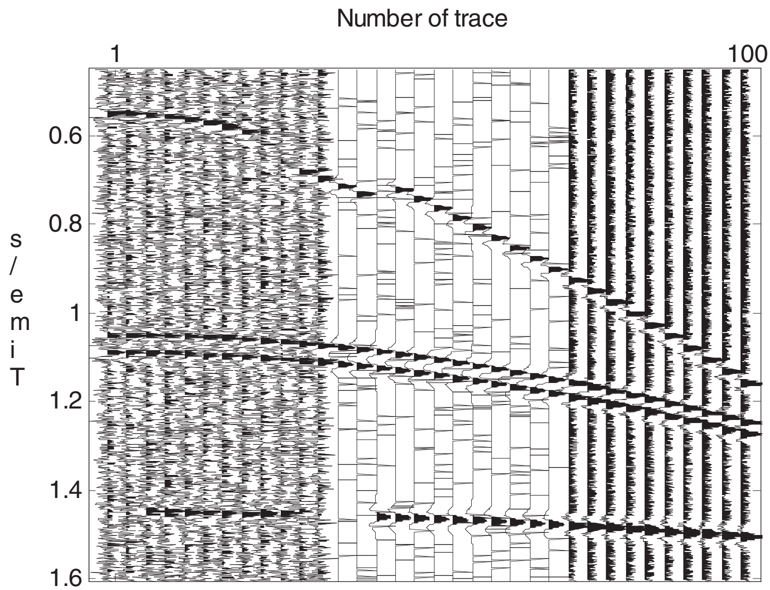


Figure 1b

Fig. 1. An example for random noise reduction on a synthetic record. Figures are drawn every three traces. (a) Noise-free synthetic record; (b) Noisy record.

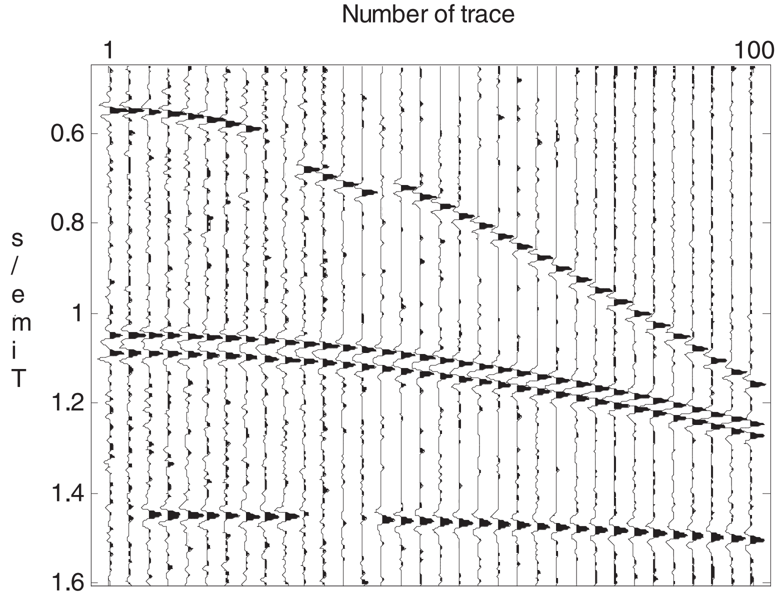


Figure 1c

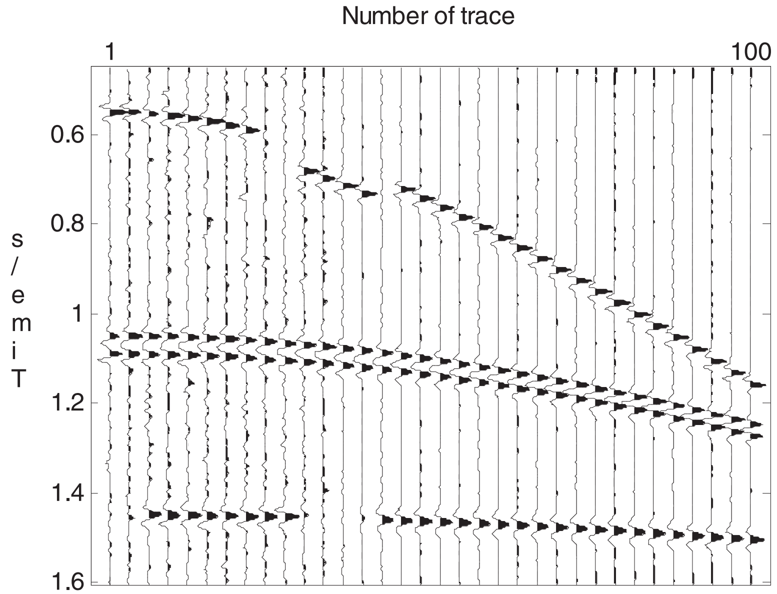


Figure 1d

Fig. 1. An example for random noise reduction on a synthetic record. Figures are drawn every three traces. (c) Result obtained by using 1D SVR; (d) Result obtained by using 2D SVR.

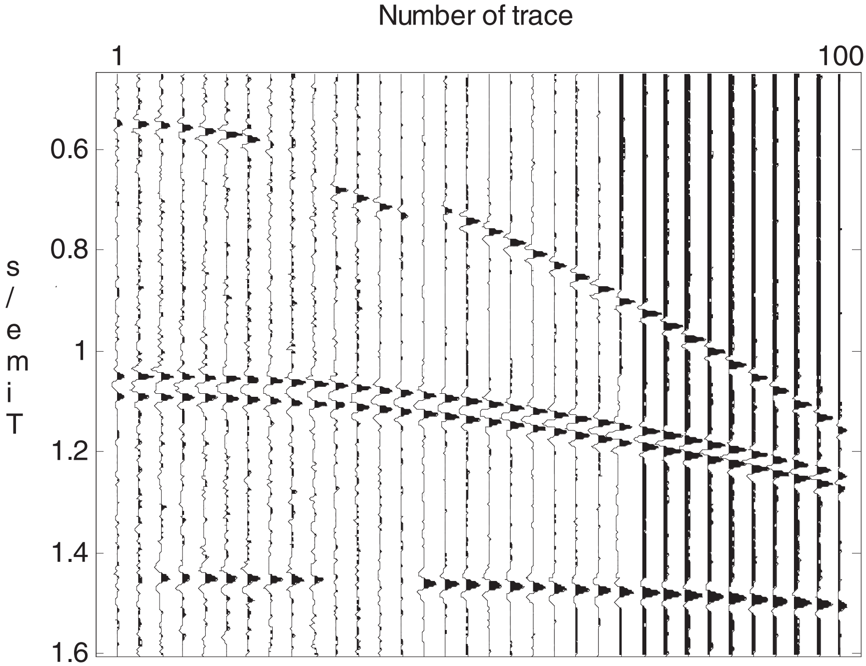


Figure 1e

Fig. 1. An example for random noise reduction on a synthetic record. Figures are drawn every three traces. (e) Result obtained by using f-x prediction filtering.

For a further comparison, Figs. 2a and 2b draw the values of all the MSEs and SNRs trace by trace, respectively. From Fig. 2a, the curve for the MSEs of all traces obtained by using the 2D SVR is in the lowest position, that is to say, the denoised signals of all traces furthest approximates to the desired signals. From Fig. 2b, we can see that the curve for the SNRs of all traces obtained by using the 2D SVR is in the highest position. It is noted that for the traces from 71 to 100, the MSEs and SNRs obtained by using the f-x prediction filtering is much worse than the other two methods, which indicates that the widely used f-x prediction filtering is not suitable for the reduction of the uniformly distributed noise. That is the reason why in Table 2 the SNRs and MSEs of the whole records obtained by using the f-x prediction filtering are very bad.

Consequently, from the above the proposed method works better than the other two methods. And the proposed denoising method is effective on conditions that the seismic data is corrupted by the white Gaussian noise, the spike-like noise or uniformly distributed noise.

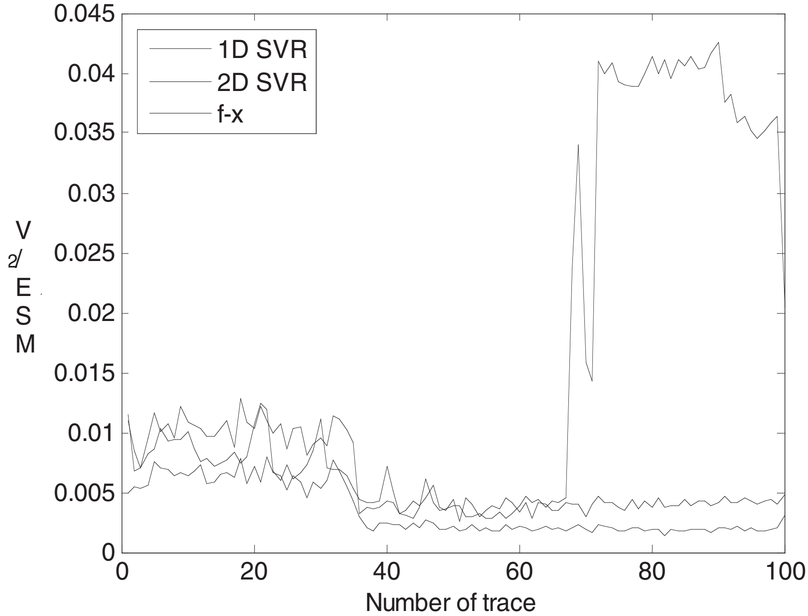


Figure 2a

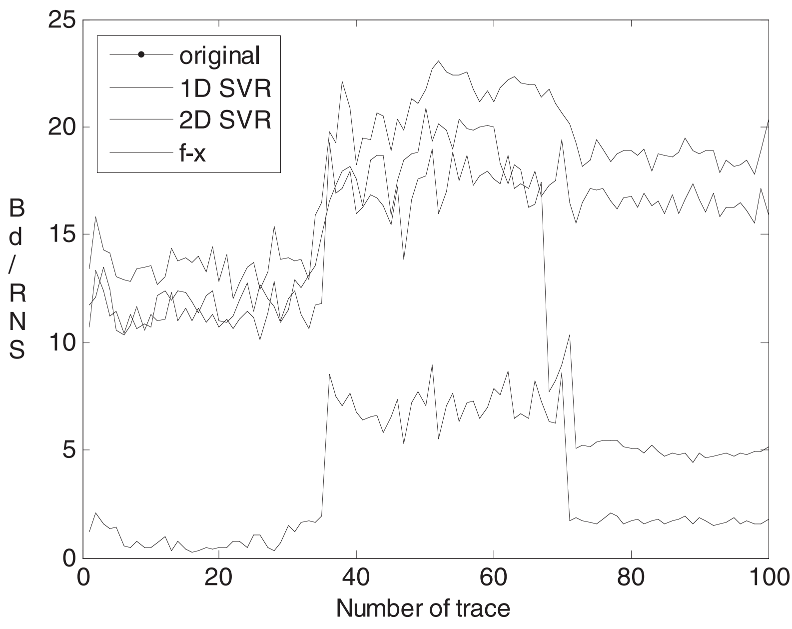


Figure 2b

Fig. 2. Comparisons on MSE and SNR obtained by using different denoising methods. (a) MSE; (b) SNR.

EXAMPLE WITH REAL SEISMIC DATA

To test the effectiveness of the proposed 2D SVR in real seismic data processing, we apply it to a field seismic record and compare with the 1D SVR and the f-x prediction filtering. Fig. 3a shows one of three dimensional common shot records with 168 traces, record length 6 s, time sampling interval 1 ms, survey line interval 160 m and geophone interval 40 m. There is much random noise in the raw record such as the ranges N_1-N_4. The presence of the boring noise seriously masks the reflection events and makes the whole record obscure and disordered.

Figs. 3b-d show the results obtained by using the 1D SVR, 2D SVR and f-x prediction filtering, respectively. For the 1D SVR, the window length is 800 samples (overlapping 20-samples border), the regularization parameter γ is set as 10, and the kernel parameters are set as 30. For the 2D SVR, the window size is 800 samples by 10 traces (overlapping 20-by-1 border), $\gamma = 10$, $f = 30$, and $k = 0.0025$. For the f-x prediction filtering, we only process the frequencies fallen in [0, 100] Hz and the amplitudes for other frequencies components are set as zero, and the window size is 800 samples by 10 traces (overlapping 20-by-1 border) and the bilateral prediction filter length is 7. Here for the 1D and 2D SVR methods, we use the SVR with implicit bias term (Deng et al., 2010).

From Fig. 3c, the whole record becomes very neat and tidy, and some events such as those in the ranges E_1-E_4 become clear and visible, especially events in ranges E_3 and E_4 which are almost invisible in the original record. But the results obtained by using the other two methods are not as good as our method. For example, the continuity of events in Fig. 3b is not as good as that in Fig. 3c, and the signals are attenuated seriously in Fig. 3d. For a detailed comparison, Fig. 4 shows the zooms of range E_4 in Fig. 3. In Fig. 4a, the noise deforms the seismic wavelets and destroys the continuity of events. After denoising by three methods, the qualities of the waveforms and the SNR are greatly improved. Compared with Fig. 4b, the result obtained by using the proposed method (Fig. 4c) has better continuity, especially in the three ranges marked by rectangles. Compared with Fig. 4d, the enhanced signals in Fig. 4c are stronger. So in the real seismic data processing, the proposed method works better than the 1D SVR method and the common f-x prediction filtering.

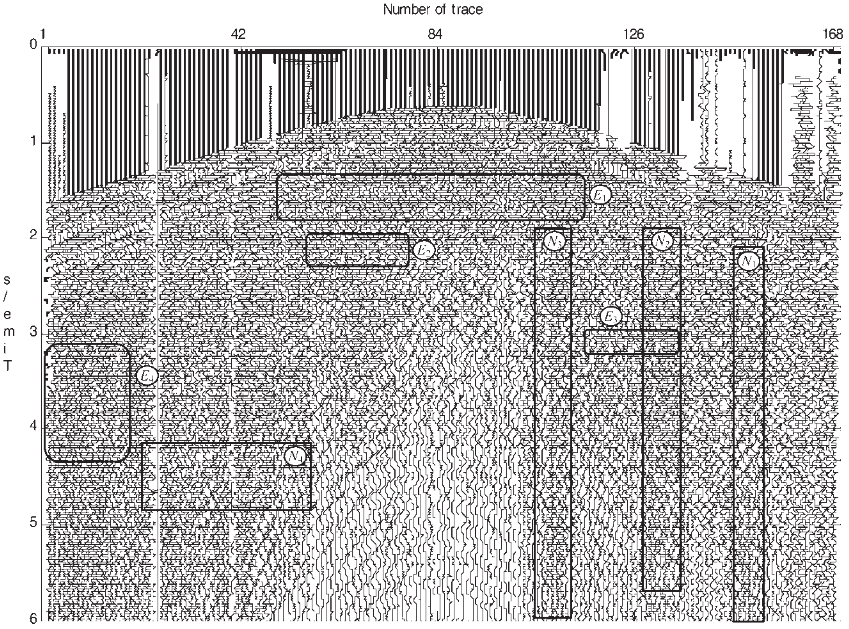
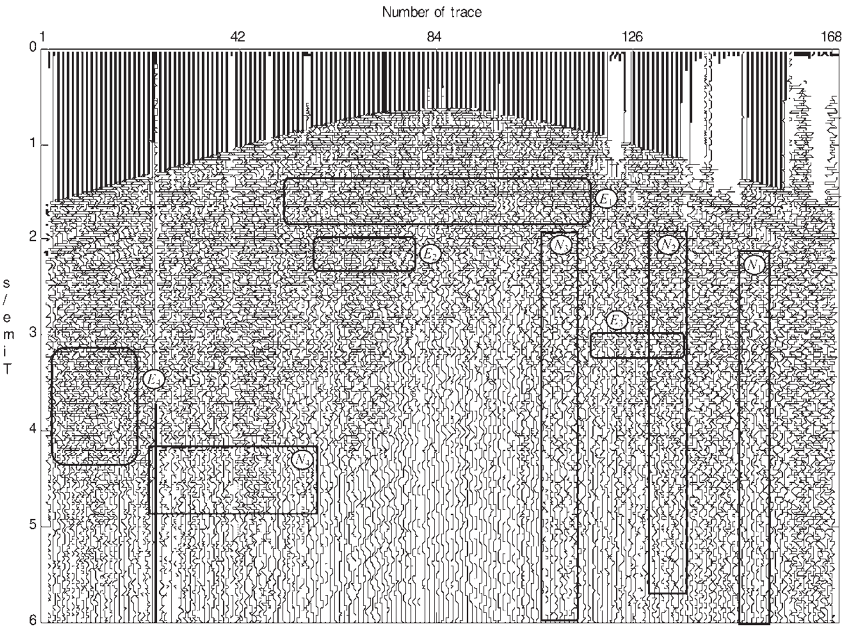


Figure 3a



Fgiure 3b

Fig. 3. An example for random noise reduction on a real record. (a) Raw record; (b) Result obtained by using 1D SVR.

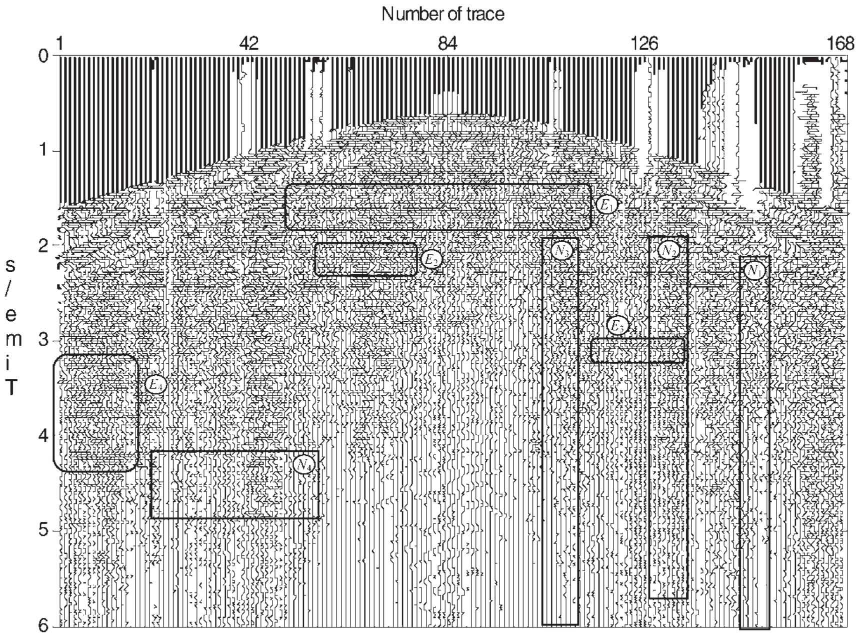


Figure 3c

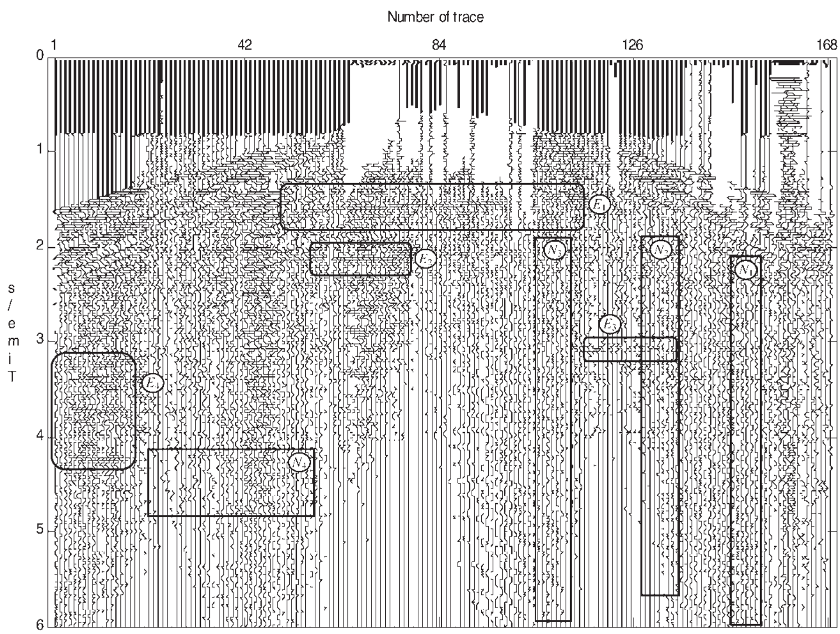


Figure 3d

Fig. 3. An example for random noise reduction on a real record. (c) Result obtained by using 2D SVR; (d) Result obtained by using f-x prediction filtering.

a)

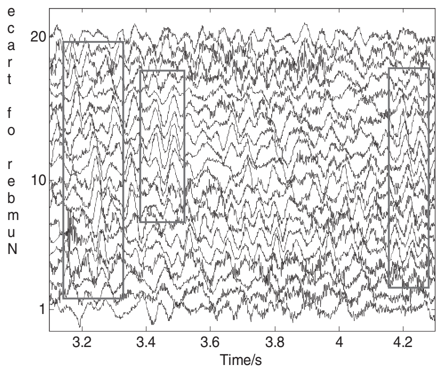


Figure 4a

b)

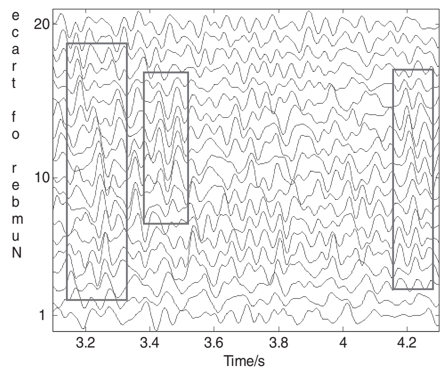


Figure 4b

c)

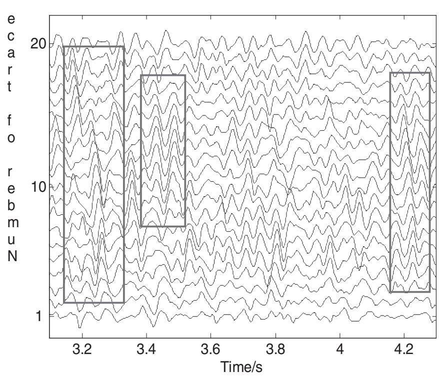


Figure 4c

d)

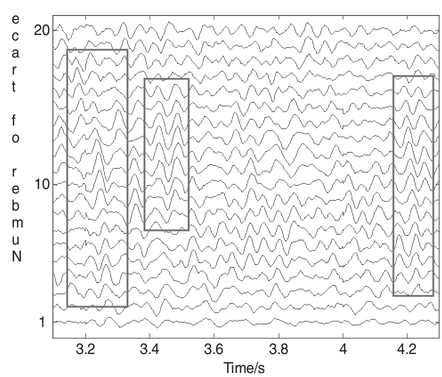


Figure 4d

Fig. 4. Comparison on zooms of range E_4 in Figs. 3(a)-(d) correspond to range E_4 in Fig. 3, respectively.

CONCLUSIONS

According to the characteristics of the seismic data in the time-space domain, we propose a new 2D Ricker wavelet kernel. For the 2D Ricker wavelet kernel, the selection of the two kernel parameters is relating to the dominant frequency of seismic data in time domain and the wavenumber of seismic data in space domain. By experiments on the synthetic and real seismic records, we compare the SVR based on the 2D Ricker wavelet kernel with the SVR based on 1D Ricker wavelet kernel and the common f-x prediction filtering. The results show that the proposed method can suppress the random noise more efficiently, and enhance the continuity of events greatly. So the SVR with the 2D Ricker wavelet kernel is an efficient method for reducing the random noise in seismic data. And in the future, after considering sufficiently the properties of 3D seismic data, it is possible to propose a 3D SVR by constructing a 3D kernel function so as to make further improvement on the quality of 3D seismic data.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 40804022) and by the Postdoctoral Science Foundation of China (Grant No. 20080430354).

REFERENCES

Bednar, J.B., 1983. Applications of median filtering to deconvolution, pulse estimation, and statistical editing of seismic data. Geophysics, 48: 1598-1610.

Bekara, M., Baan, M.V.D. and Mirko, V.D.B., 2007. Local singular value decomposition for signal enhancement of seismic data. Geophysics, 72: 59-65.

Bekara, M. and Baan, M.V.D., 2009. Random and coherent noise attenuation by empirical mode decomposition. Geophysics, 74: 89-98.

Canales, L., 1984. Random noise reduction. Expanded Abstr., 54th Ann. Internat. SEG Mtg., Atlanta: 525-527.

Cherkassky, V. and Ma, Y.Q., 2004. Practical selection of SVM parameters and noise estimation for SVM regression. Neural Networks, 17: 113-126.

Deng, X., Yang, D. and Xie, J., 2009. Noise reduction by support vector regression with Ricker wavelet kernel. J. Geophys. Engin., 6: 177-188.

Deng, X., Yang, D., Peng, J., Guan, X. and Yang, B., 2010. Noise reduction and drift removal using least squares support vector regression with implicit bias term. Geophysics, 75: 119-127.

Duncan, G. and Beresford, G., 1995. Median filter behavior with seismic data. Geophys. Prosp., 43: 329-345.

Jones, I.F. and Levy, S., 1987. Signal-to-noise ratio enhancement in multichannel seismic data via the Karhunen-Loeve transform. Geophys. Prosp., 35: 12-32.

Karsli, H., Dondurur, D. and Cifci, G., 2006. Application of complex-trace analysis to seismic data for random-noise suppression and temporal resolution improvement. Geophysics, 71: 79-86.

Liu, Y., Liu, C. and Wang, D., 2009. A 1D time-varying median filter for seismic random, spike-like noise elimination. Geophysics, 74: 17-24.

Lu, W., 2006. Adaptive noise attenuation of seismic images based on singular value decomposition and texture direction detection. J. Geophys. Engin., 3: 28-34.

Mercer, J., 1909. Functions of positive and negative type and their connection with the theory of integral equations. Philosoph. Transact. Roy. Soc. London, Series A 209: 415-446.

Sacchi, M.D. and Kuehl, H., 2000. ARMA formulation of FX prediction error filters and projection filters. J. Seismic Explor., 9: 185-197.

Suykens, J.A.K. and Vandewalle, J., 1999. Least squares support vector machine classifiers. Neural Process. Lett., 9: 293-300.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York.

Wang, Y., 1999. Random noise attenuation using forward-backward linear prediction. J. Seismic Explor., 8: 133-142.