Stanford Exploration Project, Report 92, November 12, 1997, pages 299–319

Stanford Exploration Project, Report 92, November 12, 1997, pages 299–319

Test of wavelet-based seismic data compression software

Yalei Sun and Biondo Biondi[1]

keywords: *not available*

ABSTRACT

In this paper, we test a new wavelet-transform based seismic data compression technique developed by Chevron. We apply this technique to two synthetic datasets and one field dataset. Our results show that this new compression approach can virtually retain all of the important seismic information at high compression ratios. We summarize several empirical rules which will help the high performance of this software.

INTRODUCTION

Seismic exploration has entered the 3-D age. With the help of high-performance computers, we can image the structure of the subsurface more and more accurately. Meanwhile, the size of 3-D seismic dataset is also increasing dramatically. The typical volume of 3-D seismic dataset has reached terabyte levels. The huge volume not only taxes the storage and I/O capacities of computer systems, but also makes managing large datasets a formidable problem. However, the large difference between the volume of prestack dataset and poststack dataset implies there is some redundancy in the prestack dataset. This redundancy provides us the possibility of compressing the prestack dataset at high compression ratios.

At SEP, we also face the problems of handling large dataset. One is that we usually do not have enough disk space to store the whole 3-D prestack dataset. The other is the retrieval of a dataset from tapes to disk needs a large amount of time. Moreover, since we have begun to issue a CD-ROM version of the SEP reports, compressing the dataset means we can store more on one compact disk.

Ergas and Donoho et al. from Chevron (1995) (1995) have developed a new seismic data compression technique based on wavelet transform. This compression algorithm performs wavelet decomposition on the seismic data and represents them in the wavelet domain characterized by a number of subbands consisting of different

[1]**email:** yalei@sep.stanford.edu, biondo@sep.stanford.edu

temporal and spatial frequency components. The coherency existing in the multidimensional seismic datasets, ranging from prestack to poststack data, allows efficient quantization of the data in each wavelet-transform subband, such that the original data can be precisely represented by a very small average number of bits per sample. Many tests show that this technique can compress the 3-D prestack dataset at a compression ratio as high as 100 : 1.

WAVELET-TRANSFORM BASED COMPRESSION ALGORITHM

Data compression has been investigated in the field of digital communication for a long time. Generally, data compression techniques can be divided into two major families (Nelson, 1995):

- *lossless* compression
  Lossless compression consists of those techniques guaranteed to generate an exact duplication of the input dataset after a compress/decompress cycle. Lossless compression is essentially a coding technique. There are many different kinds of coding algorithms, such as Huffman coding (Huffman, 1952), run-length coding (Storer, 1988), and arithmetic coding (Witten et al., 1987).

- *lossy* compression
  Lossy data compression concedes a certain loss of accuracy in exchange for high compression ratio. Lossy compression proves effective when applied to digitized representations of analog phenomena. By their very nature, these representations are not perfect to begin with, so the idea of output and input not matching exactly is somewhat more acceptable. Most lossy compression techniques can be adjusted to different quality levels, gaining high accuracy in exchange for less effective compression. Lossless compression is a necessary component of every lossy compression approach.

Most of the lossy data compression algorithms follow similar methodology: the original data are mathematically transformed to a new domain in which they are better organized for data compression than in the normal spatial-temporal domain. Therefore, the choice of mathematical transformation is crucial to the performance of compression algorithms.

Among many different kinds of transformation, the wavelet transform (Daubechies, 1992) has been chosen to develop data compression algorithms. There is a large difference between wavelet transform and Fourier transform. In the Fourier domain, all the elements of the basis are active for all time t, i.e., they are *non-local*. Consequently, Fourier series converge very slowly when approximating a localized function (Cohen, 1992). Wavelet transform makes up for the defficiencies of Fourier transform. *Wavelet basis* function is a novel basis *localizing* in both time domain and frequency domain. Therefore, *wavelet basis* function can provide a good approximation for a localized function with only a few terms.

Seismic data have the characteristic of *localization*, which is the main reason of choosing wavelet transform in seismic data compression. Wavelet transform organizes the seismic data into subbands, each of which shows a different level of temporal and spatial characteristics. In general, large subbands consist of high frequency data in temporal and spatial dimensions. Because of coherency of the data along the spatial dimensions, most of the data in this subband represent noise of little geophysical significance. However, many small subbands at lower frequencies contain more seismic information that should be retained. Figure 1 shows the main procedures of Chevron's compression package.

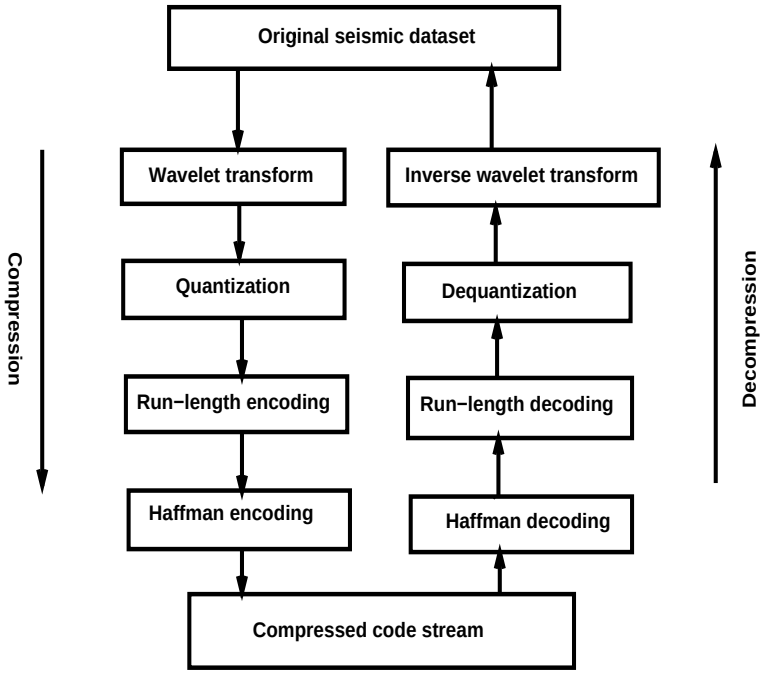

Figure 1: The main procedures of Chevron's compression package. yalei-procedure [NR]

- *Wavelet transform*
  Mathematically, there are numerous kinds of *wavelet basis* function. In the compression applications, the choice of wavelet is not very critical (Ergas, personal communication), as long as it is reasonably smooth. Chevron compression package chooses the wavelet introduced by Bradley and Brislawn (1994). Biorthogonal wavelet filters are employed along each dimension after a study of a wide range of different wavelet filters (Villasenor et al., 1995).

- *Quantization*
  Among the four compression procedures, the key to the success of this lossy compression technique is the choice of quantization methods. It is also this quantization step which introduces loss of information.

- *Run-length encoding* and *Huffman encoding*
  Chevron's package uses two kinds of coding algorithms, run-length encoding and Huffman encoding. Each of them has different advantages when we are dealing with different components of datasets.

In Figure 2, we show a 2-D schematic representation of wavelet subbands of the seismic data. Different subbands contain different seismic information. It is crucial to choose different quantization procedures for each subband. For example, the low-frequency and low-wavenumber subband contains most of the seismic reflection energy. We should be very careful when dealing with this part. In other words, most of this part should be preserved. In the high-frequency and high-wavenumber subband, there is only some fraction of seismic information, such as high-frequency diffraction hyperbolas, which is nonetheless very important in seismic interpretation. Most of the data samples in this subband belong to uncorrelated noise, losing them only introduces nonobservable influence to the whole dataset.

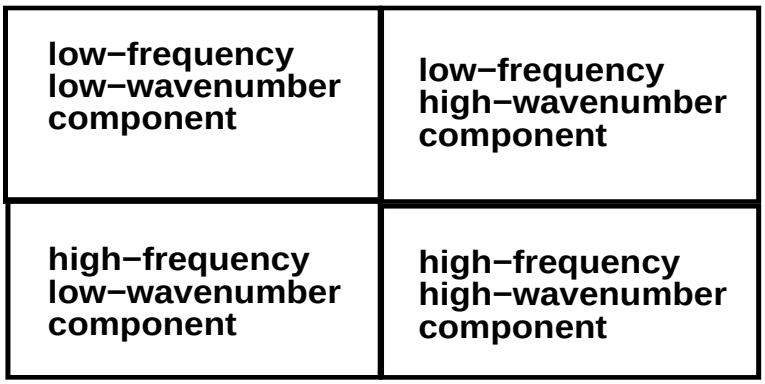

Figure 2: Schematic representation of the wavelet subbands of the seismic dataset. yalei-subband [NR]

In practice, if we want to reach a high compression ratio with good quality, we must implement this technique in a high-dimensional space. Chevron's algorithm performs compression and decompression on 4-D blocks of seismic data (which includes 2- and 3-D blocks as subsets). Each block must be of the same size of $n_1 * n_2 * n_3 * n_4$. n_1 is the time axis, but n_2, n_3, and n_4 can represent any spatial axis, such as offset, shot, CDP, streamer, inline, or crossline direction. n_1 and n_2 must be greater than one, and should generally be more than 16 for a reasonable performance. n_3 and n_4 can be one or greater.

APPLICATION OF WAVELET-BASED COMPRESSION

In this section, we discuss our testing of three datasets in different manners. We explore these datasets to investigate the different aspects of this compression algorithm.

Specifically, the first is a 2-D synthetic shot gather contaminated by randdom noise. Using this example, we demonstrate how the random noise is removed from the shot gather. The second is a 3-D synthetic dataset consisting of many dipping reflectors. We show that making use of coherency will bring high compression ratio. The third one is a 3-D marine dataset which was recorded in the North Sea. In this example, NMO correction is applied before compression to improve the performance of this compression technique.

Synthetic shot gather

The synthetic shot gather consists of three hyperbolas. The dataset has been contaminated by random noise. For this 2-D dataset, $n_1 = 512$ and $n_2 = 128$. Figure 3, 4, 5, and 6 are the t-x domain and f-k domain results.

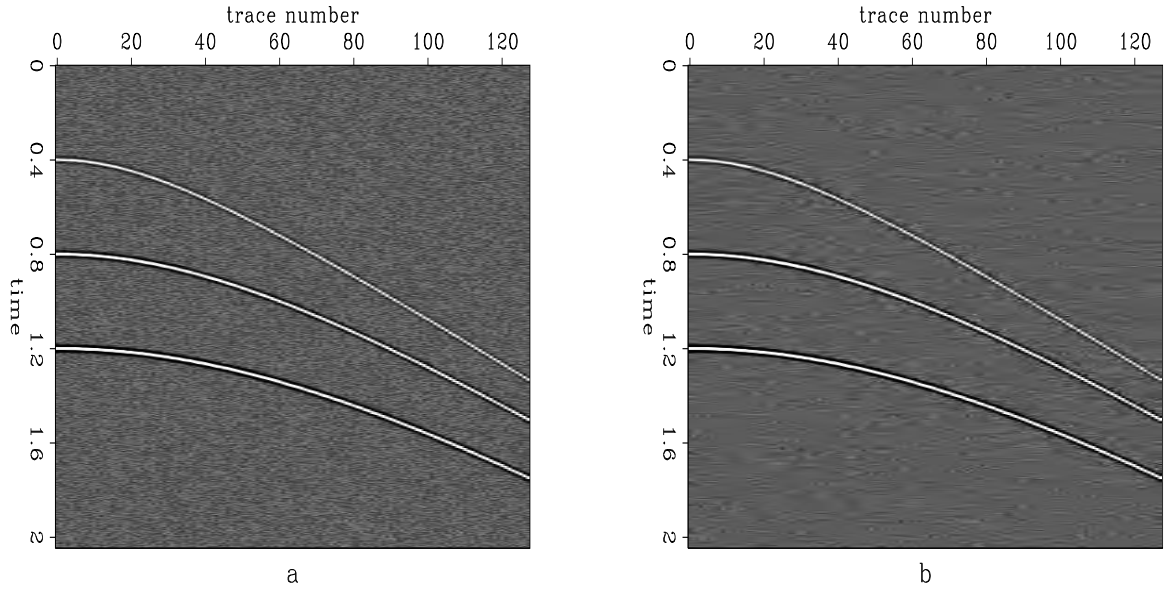

Figure 3: Synthetic shot gather in the t-x domain ($scale = 0.715$). In this case, the compression ratio is 50 and SNR is 7.69dB. The two figures are plotted on the same scale.
a: Original dataset.
b: Compressed/decompressed dataset. yalei-compcsg [CR]

After one compress/decompress cycle, random noise is removed from the input dataset. It means that this technique attains high compression ratios by filtering uncoherent component. We conduct a series of tests using different *scale* values. As shown in Figure 7 and Figure 8, with the increase of *scale* value, the compression ratio increases more and more sharply and SNR decreases more and more slowly.

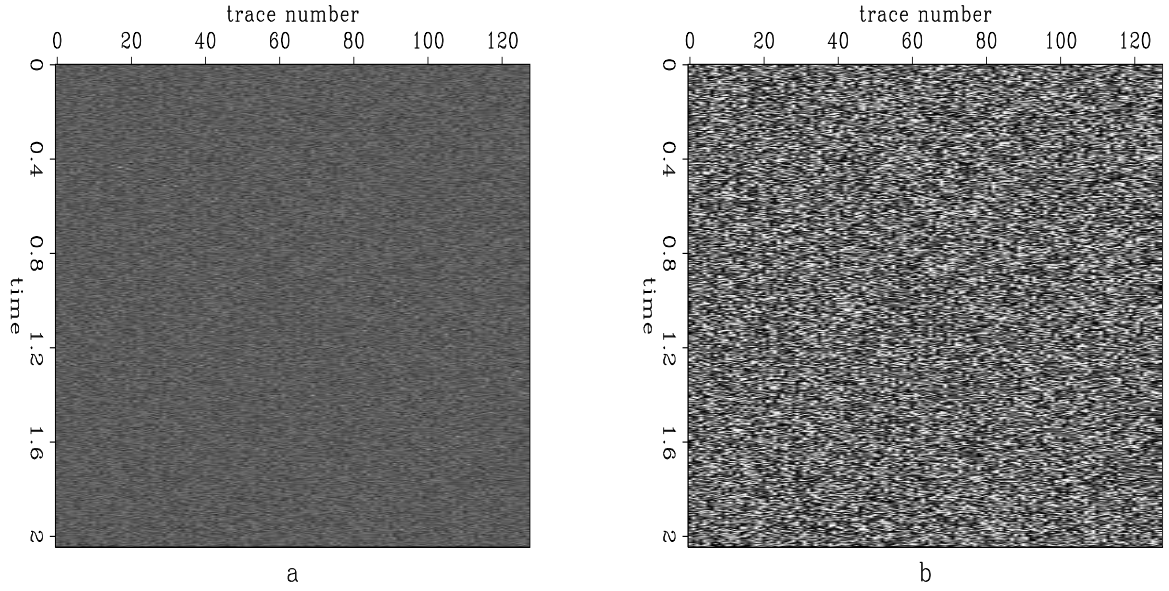

Figure 4: Difference of the two results in last figure. There is little remnant of the three hyperbolas in the figure.
a: Difference shown on the absolute scale.
b: Difference shown on the relative scale. yalei-compcsgdiff [CR]

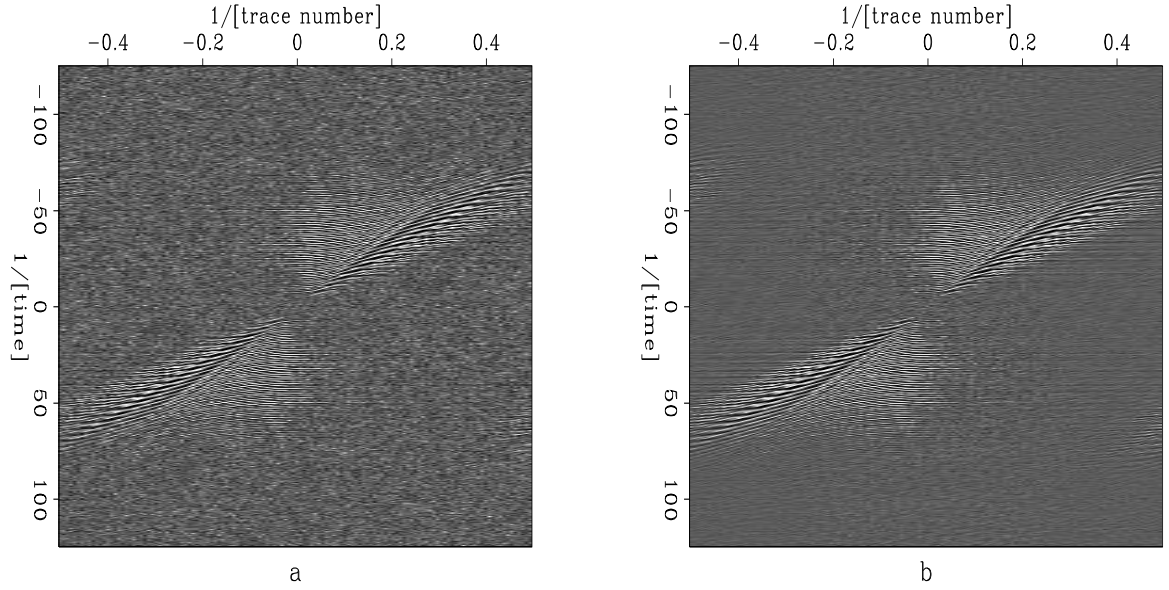

Figure 5: Synthetic shot gather in the *f-k* domain.
a: Original dataset.
b: Compressed/decompressed dataset. yalei-ftcompcsg [CR]

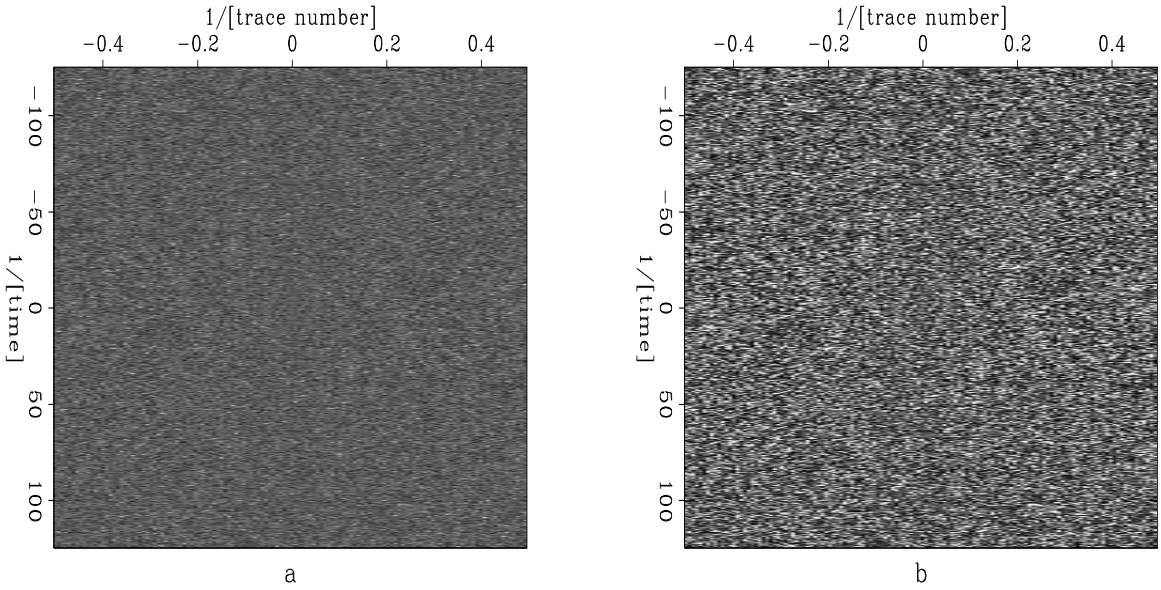

Figure 6: Difference of the two results in last figure. High frequency noise is the main component of this figure.
a: Difference shown on the absolute scale.
b: Difference shown on the relative scale. yalei-ftcompcsgdiff [CR]

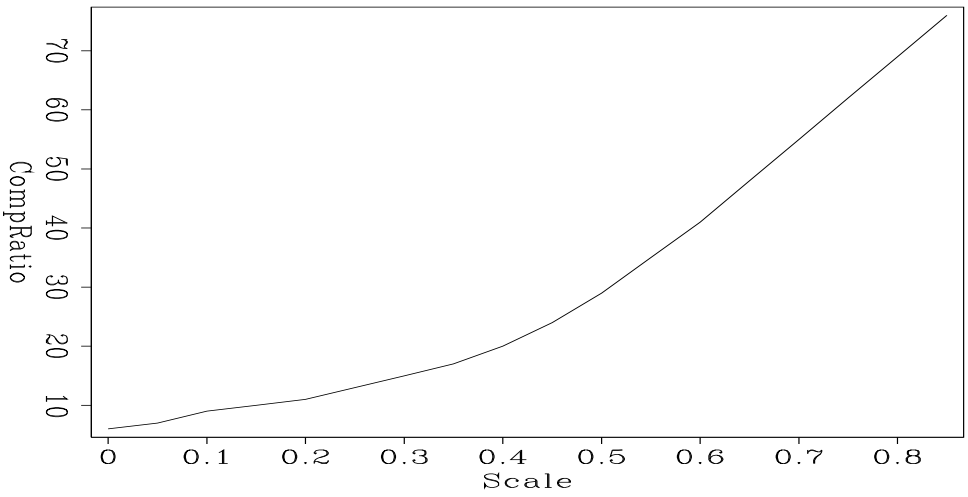

Figure 7: Compression ratio increases from 6 to 76, when *scale* value increases from 0.05 to 0.9. yalei-csgcr [CR]

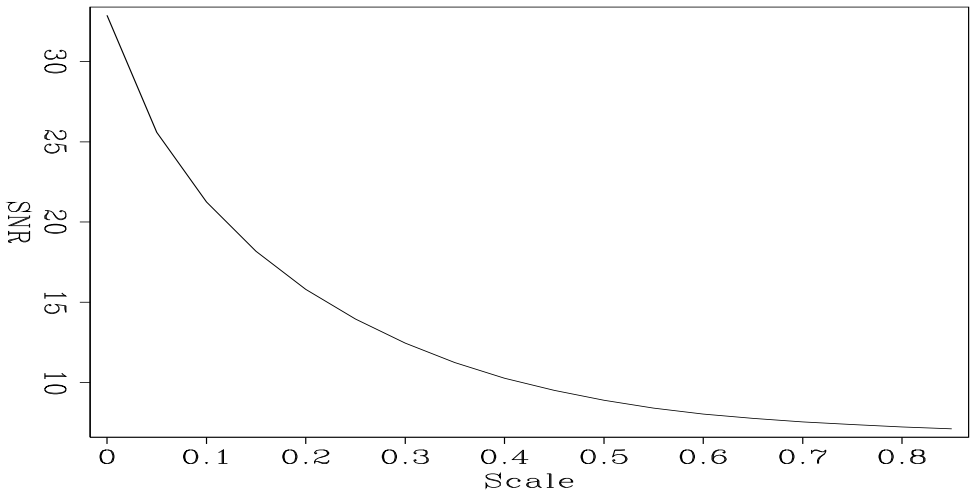

Figure 8: SNR decreases from 32.88dB to 7.11dB, when *scale* value increases from 0.05 to 0.9. yalei-csgsnr [CR]

Synthetic 3-D data

We generate a synthetic 3-D dataset using the formula

$$t_{shift} = t_0 + a \cdot i_2 + b \cdot i_3, (i_2 = 1, 2, ..., n_2; i_3 = 1, 2, ..., n_3)$$

First, a white noise trace is generated and passed through a low-pass filter. Then, according to the position of the trace on the n_2 by n_3 plane, we shift the trace by t_{shift}. As a special case, if we set $a = 0$ and $b = 0$, we will get a dataset consisting of many horizontal reflections, as shown in Figure 9. Otherwise, we will generate a series of dipping reflections, as shown in Figure 10 ($a = 1, b = 1$). Finally, we mix three datasets of different dipping reflections together ($0°, 45°, 135°$) and get a new dataset, as shown in Figure 11. The size of this cube is $n_1 = 256$, $n_2 = 128$, and $n_3 = 128$.

This dataset mainly consists of coherent component. Therefore, the result from horizontal reflections is nearly perfect. However, with the increase of dipping angle, the extent of coherency decreases and the compression ratio also decreases from several thousand to less than two hundred. This verifies our analysis in last section, i.e., steep dipping reflections are more difficult to compress. This is why we prefer to compress the dataset after NMO correction.

One interesting thing is, the result of the mixing case is better than the case of 45° dipping reflections (difference of SNR is $1.17dB$). This is because some high-frequency, high-wavenumber coherent noise is introduced into the dataset after compression. In the mixing case, the coherent noise cancels each other. Therefore the influence is not so serious. As shown in Figure 12 and 13, it is very easy to notice the high-frequency, high-wavenumber coherent component in the *f-k* domain. In the real seismic dataset,

there exist reflections with many different dipping angles. This coherent noise usually will not cause big problem.

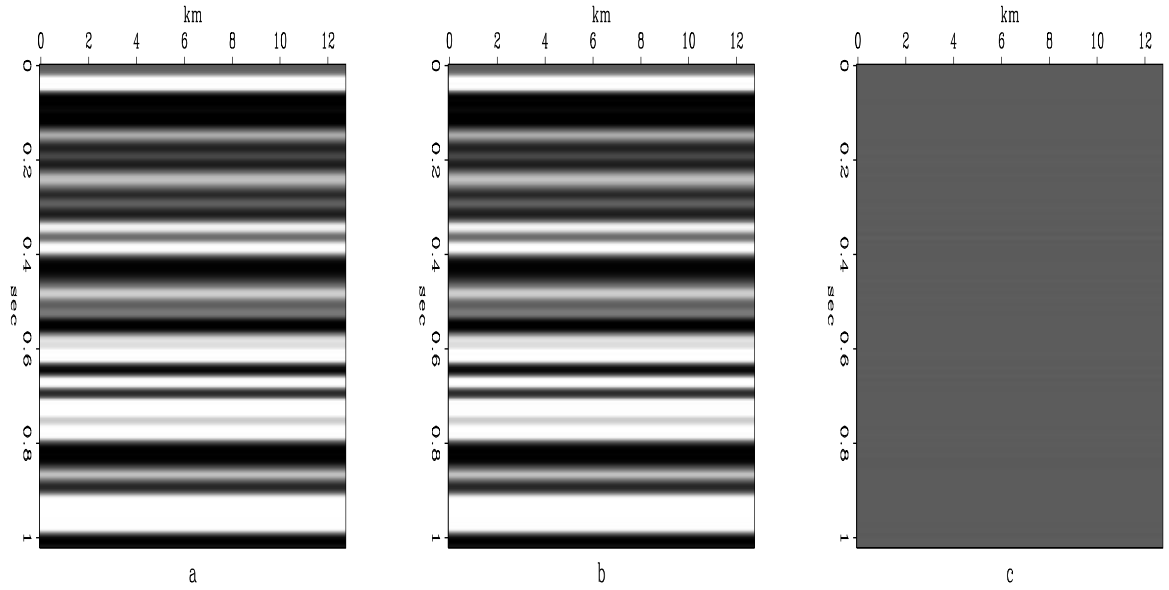

Figure 9: Synthetic horizontal reflection section (*scale* = 0.2). In this case, the compression ratio is 3009 and SNR is 46.91dB.
a: Original dataset.
b: Compressed dataset.
c: Difference of the two datasets. yalei-compdip0 [CR]

North Sea prestack data

The field data is a 3-D marine seismic sail line from the North Sea. The survey system consists of two sources and three streamers. There are many conflicting dipping reflections and high frequency diffraction hyperbolas in the dataset, which provide a critical challenge to this compression technique. In this application, we compress the dataset in two ways: before NMO correction (case A) and after NMO correction (case B). The size of this dataset is $n_1 = 500$, $n_2 = 112$, $n_3 = 6$, and $n_4 = 150$. We compress/decompress the two different kinds of dataset at three compression ratio levels, 40, 70, and 100. Figure 14 is the original dataset. Figures 15 and 16 are the results of $CompRatio = 40$. Figures 17 and 18 are the results of $CompRatio = 70$. Figures 19 and 20 are the results of $CompRatio = 100$.

As shown in the figures, the result of after NMO correction is better than that of before NMO correction for each of the three compression levels. This is because NMO correction can strengthen the extent of coherency in the dataset. The SNR of the whole dataset is higher in case B. We choose the result of $CompRatio = 70$ to analyze the change of SNR in the time domain, frequency domain, and wavenumber domain, as shown in Figures 21, 22, and 23. In the time domain, the shallow zone result

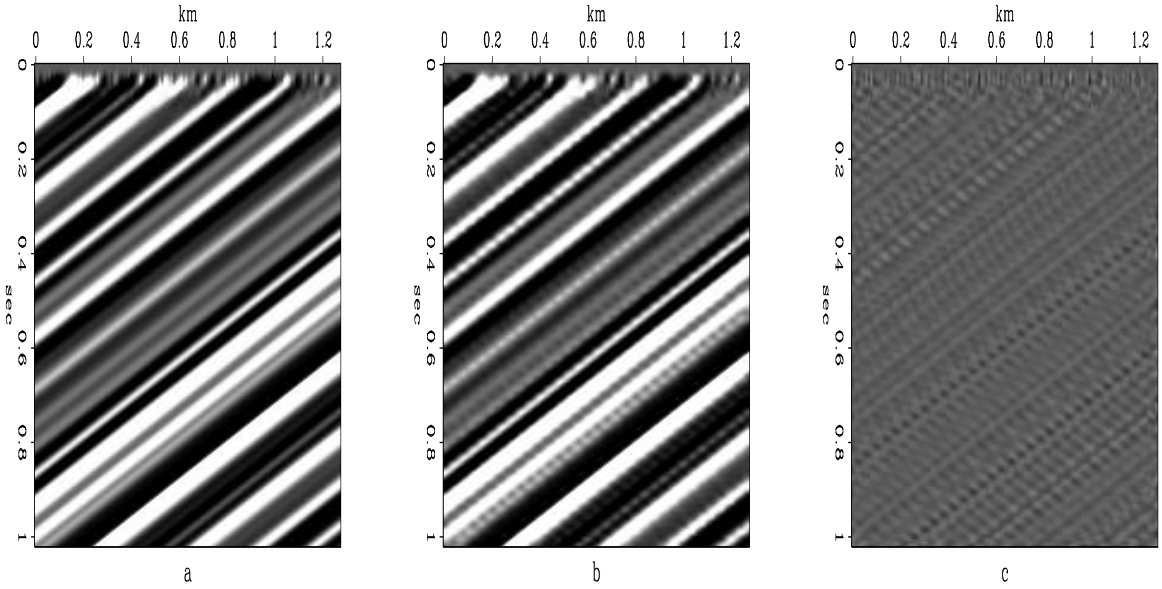

Figure 10: Synthetic dip reflection section (*scale* = 0.28, *dipping angle*=45°). In this case, the compression ratio is 149 and SNR is 21.49dB.
a: Original dataset.
b: Compressed dataset.
c: Difference of the two datasets. yalei-compdip45 [CR]

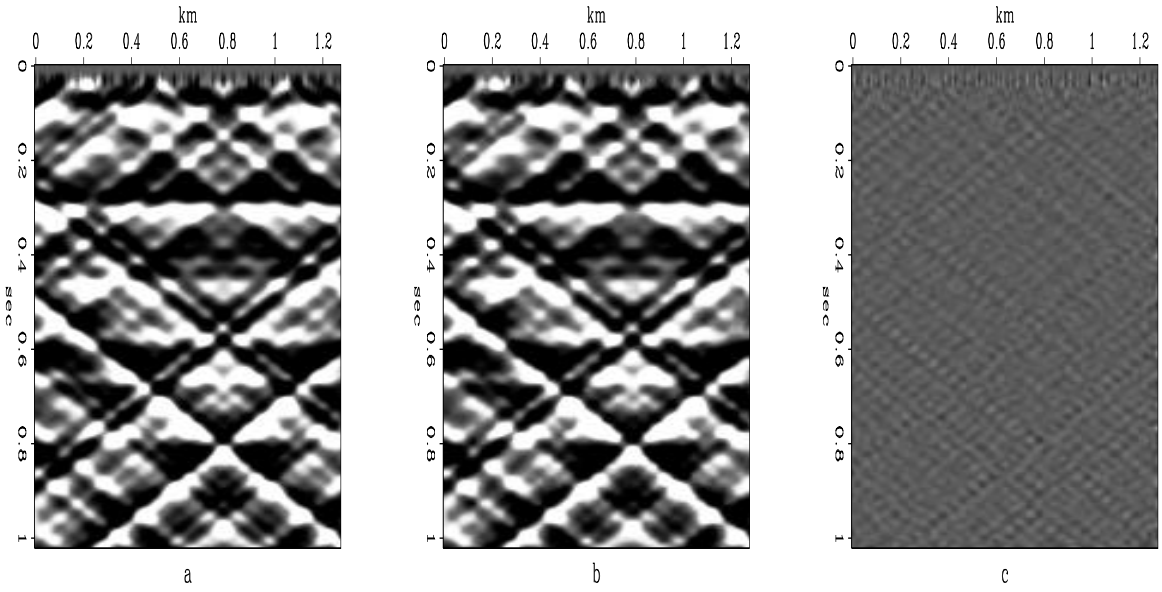

Figure 11: Mixed section consisting of three different dipping events (*scale* = 0.24, *dipping angle*=0°, 135°, 45°). In this case, the compression ratio is 151 and SNR is 22.66dB.
a: Original dataset.
b: Compressed dataset.
c: Difference of the two datasets. yalei-compdipmix [CR]

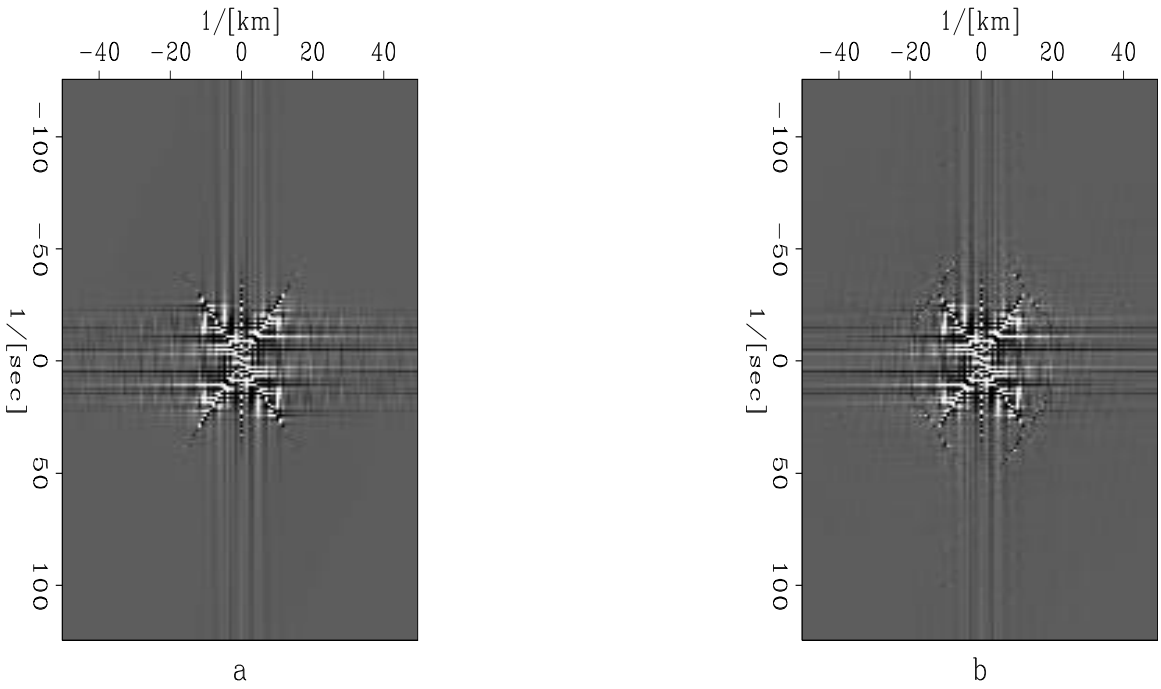

Figure 12: Mixed section consisting of three different dipping events in *f-k* domain.
a: Original dataset.
b: Compressed dataset. yalei-ftcompdipmix [CR]

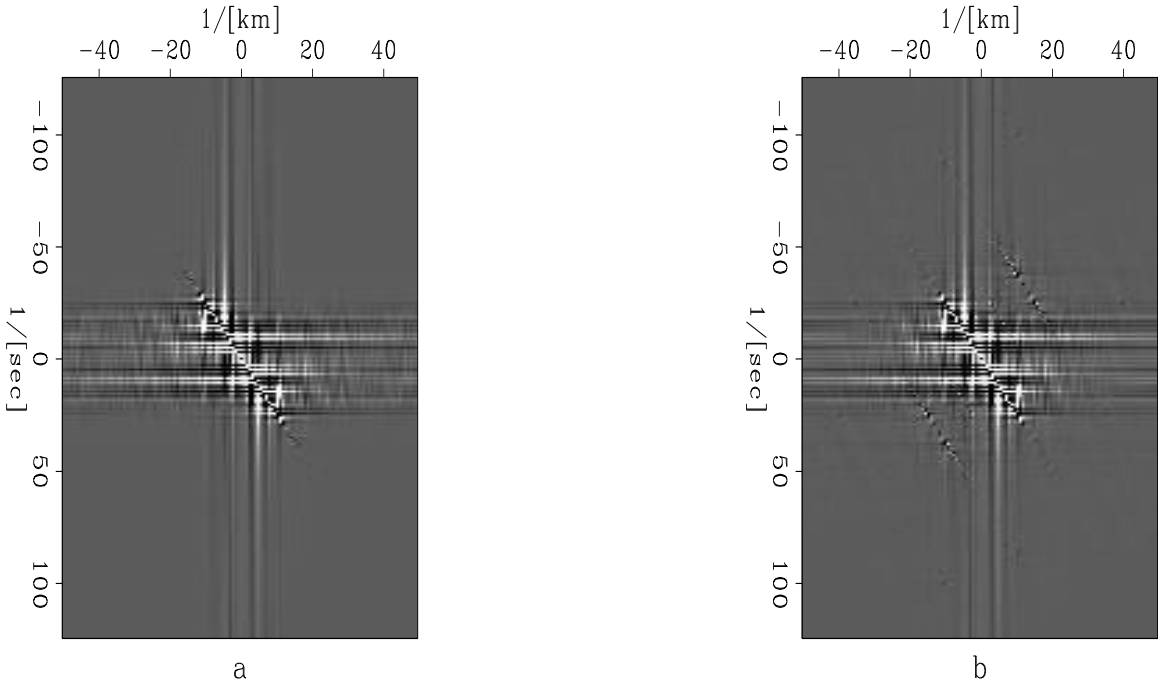

Figure 13: 45° dip reflection section in *f-k* domain.
a: Original dataset.
b: Compressed dataset. yalei-ftcompdip45 [CR]

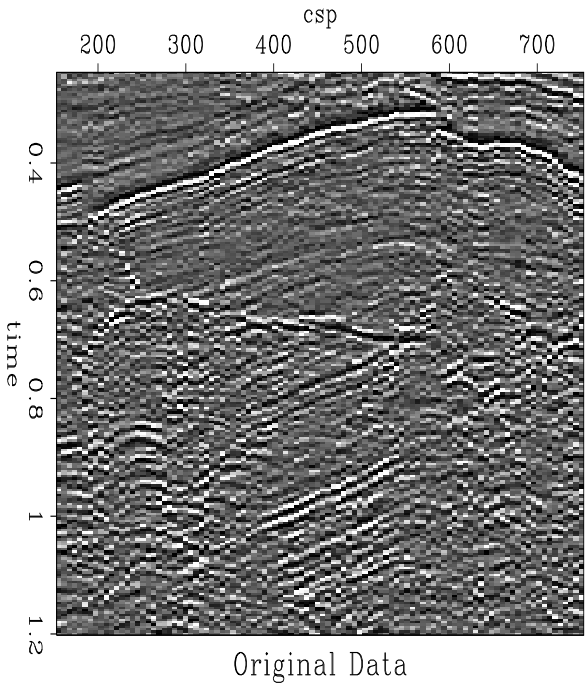

Figure 14: Original 2-D near-offset section of the North Sea dataset (extracting the near-offset trace from each CSP gather). yalei-syl12DW [CR]

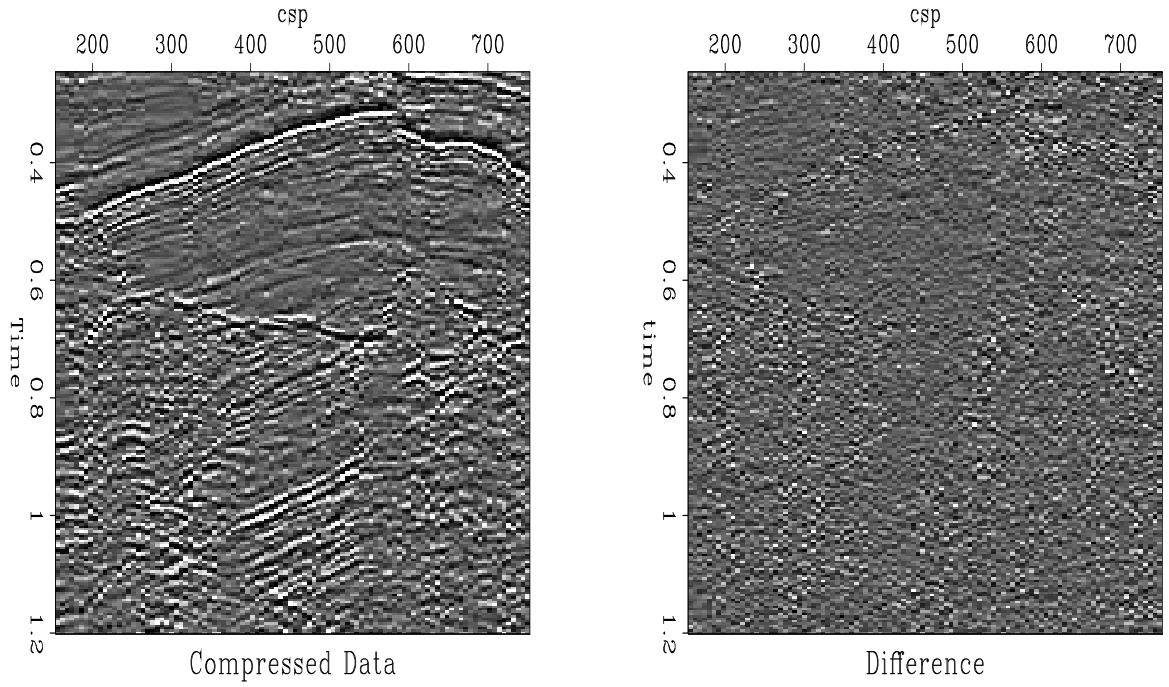

Figure 15: Result from case A (CompRatio=40, SNR=3.95dB). yalei-syl12D31W [CR]

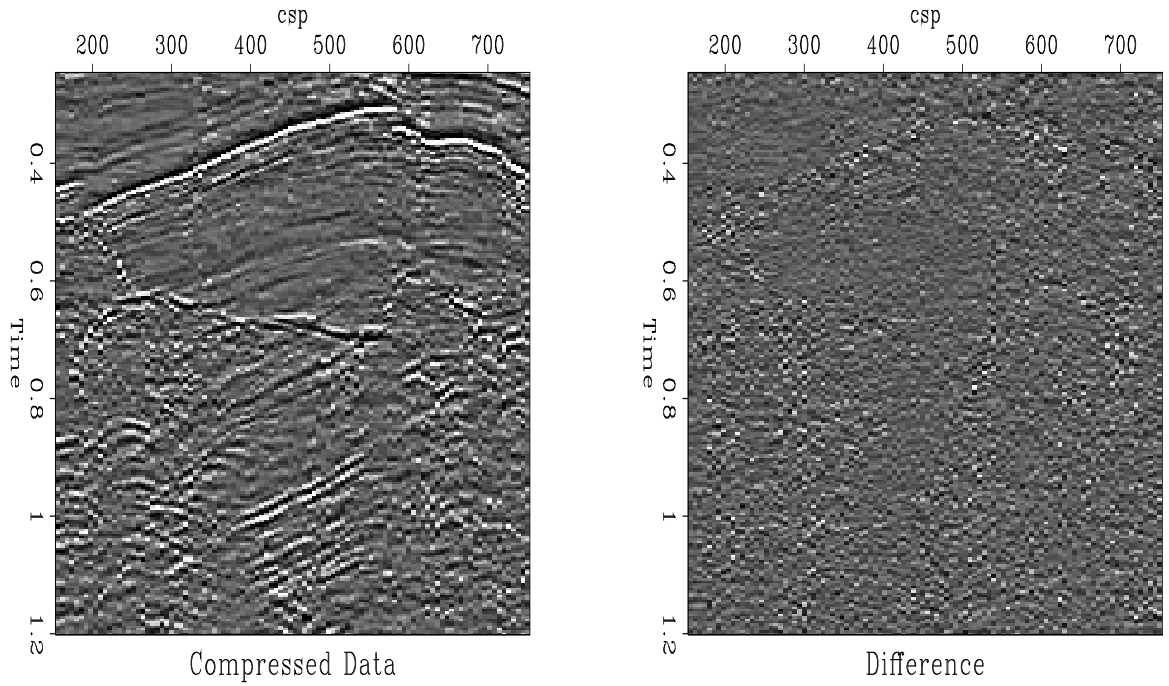

Figure 16: Result from case B (CompRatio=40, SNR=4.01dB). yalei-syl1nmo2D25W [CR]

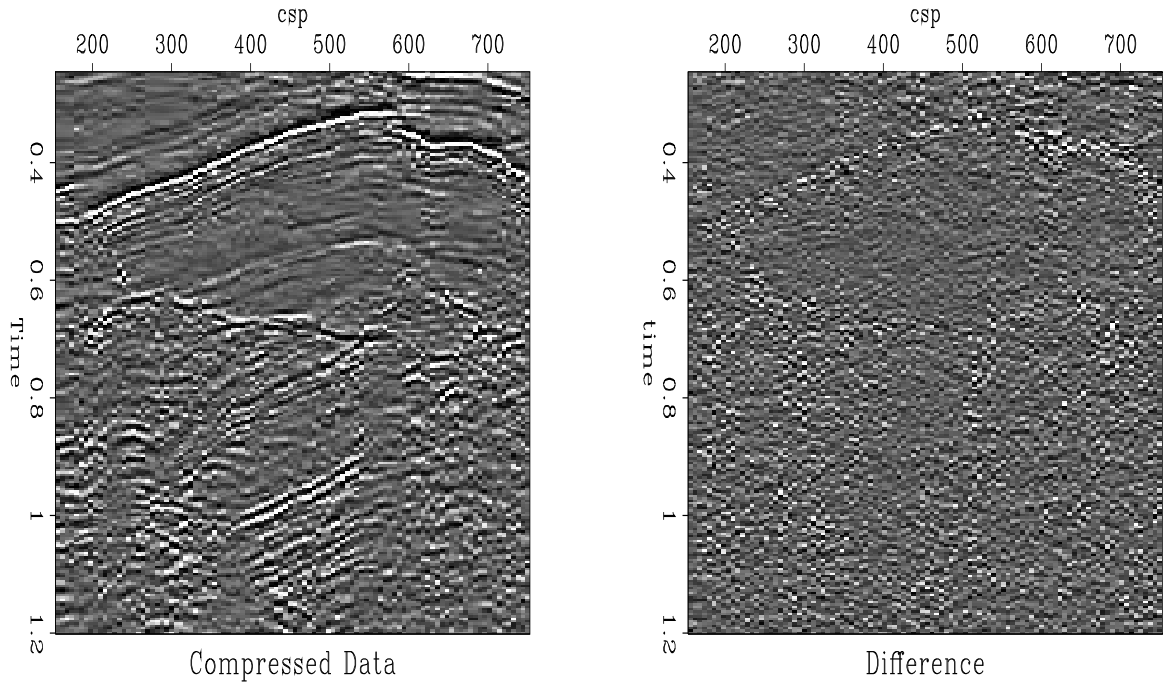

Figure 17: Result from case A (CompRatio=70, SNR=2.94dB). yalei-syl12D51W [CR]

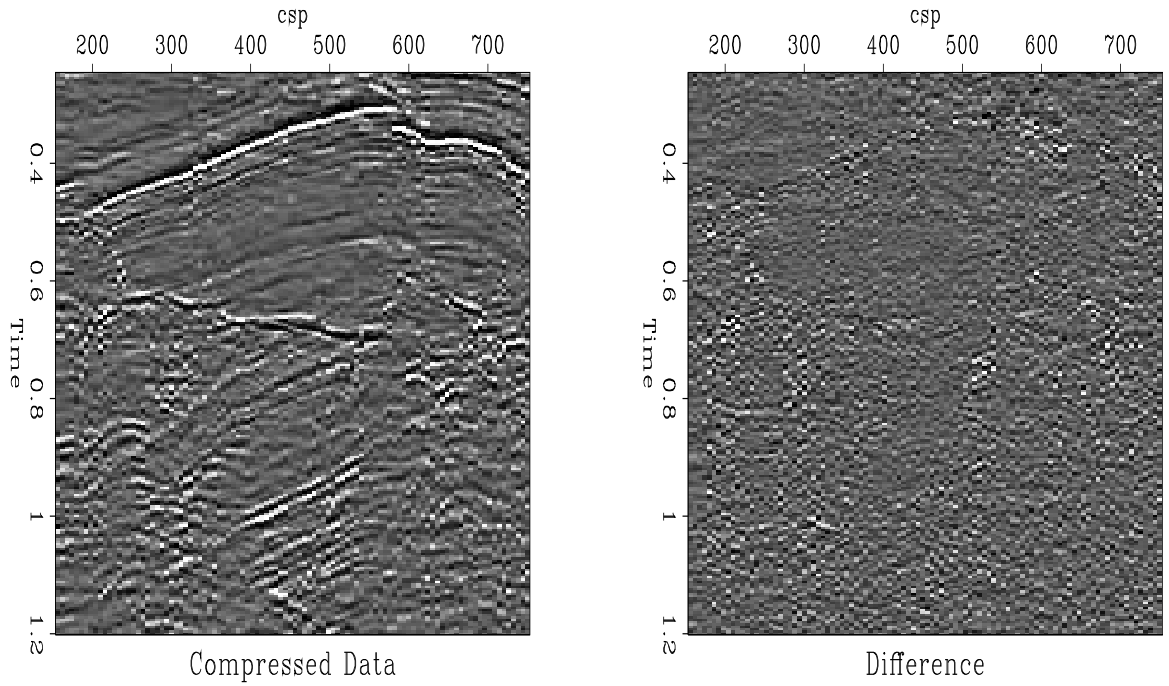

Figure 18: Result from case B (CompRatio=70, SNR=3.09dB). yalei-syl1nmo2D41W [CR]

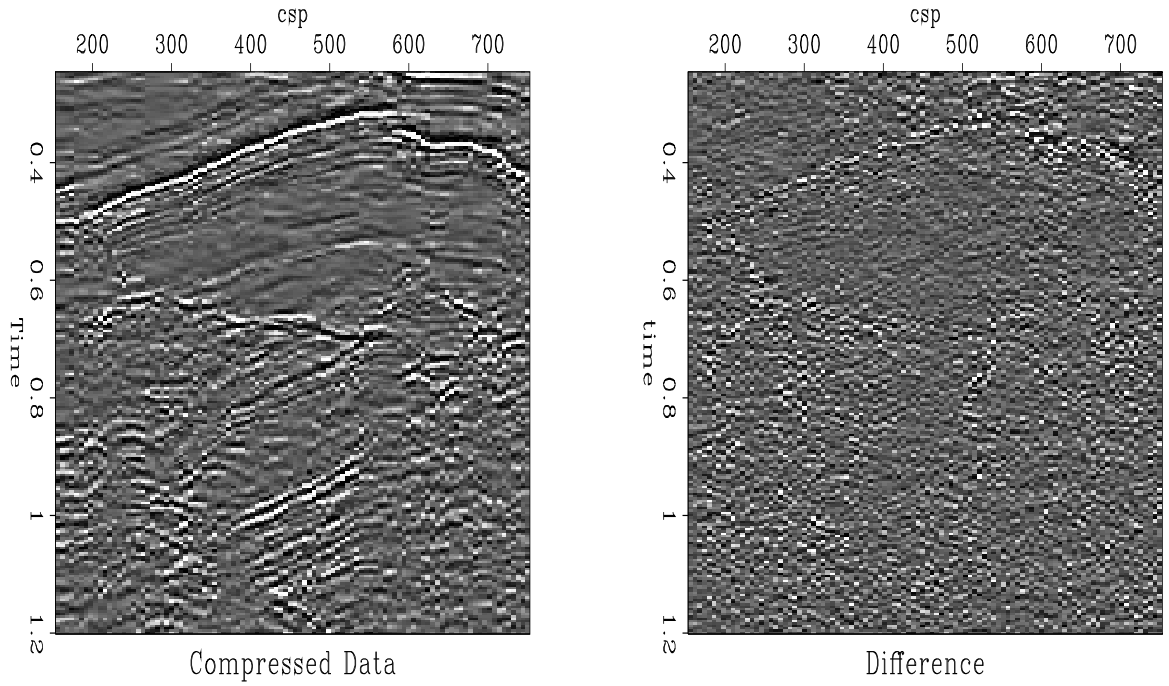

Figure 19: Result from case A (CompRatio=100, SNR=2.65). yalei-syl12D65W [CR]

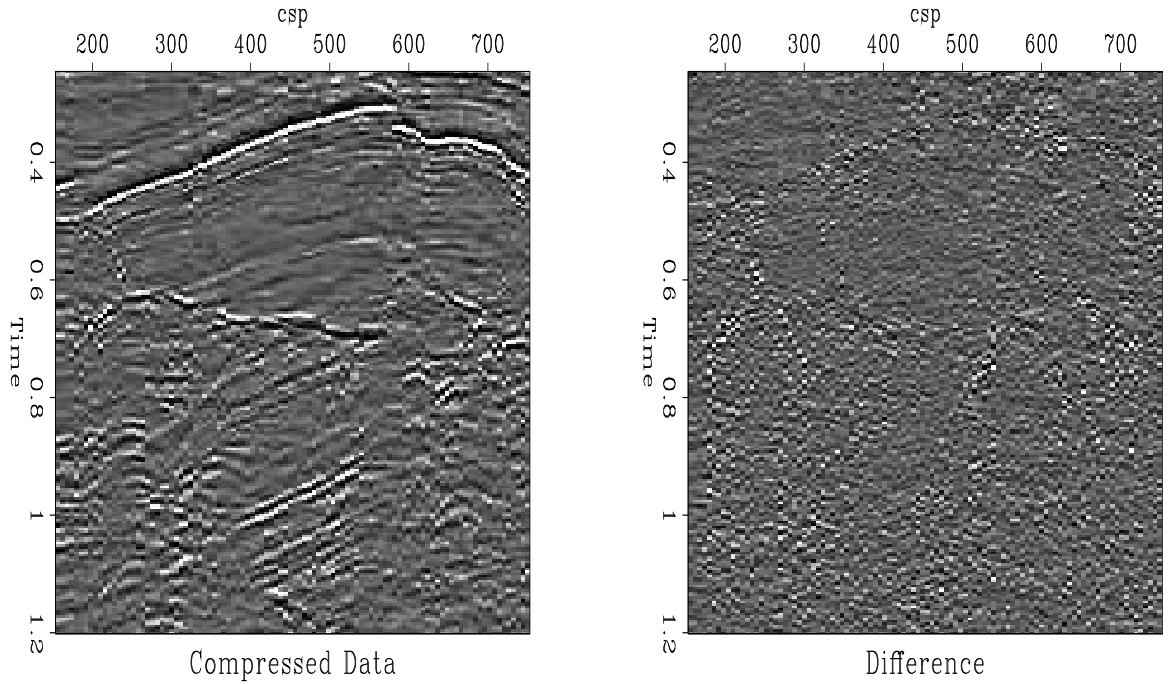

Figure 20: Result from case B (CompRatio=100, SNR=2.78dB). yalei-syl1nmo2D54W [CR]

of post-NMO correction is prominently better than case A. In the deep zone, which is mainly composed of high-frequency uncorrelated noise, the SNR has decreased greatly in both cases. In the frequency and wavenumber domain, they show similar results, i.e., the result of case B is better than that of A in the low-frequency and low-wavenumber areas. In the high-frequency and high-wavenumber ranges, the accuracy of this compression algorithm is lowered.

CONCLUSIONS

Using different datasets, we test the performance of Chevron's seismic data compression software. The results show that it is a very promising approach. Especially this new technique provides a possible solution to the contradiction between the increase of 3-D seismic dataset and the trend toward distributed and networked computing enviroments. The most prominent characteristic is this method's high compression ratio, which is achieved in the high-dimensional wavelet domain.

Although this compression method is a general technique, there are still some empirical rules that will help us to attain better results:

- The conducting of some tests on small-sized dataset to find an acceptable compression ratio and the corresponding *scale* value. Since there is no fixed quantitative relation between *scale* value and compression ratio, same *scale* value can

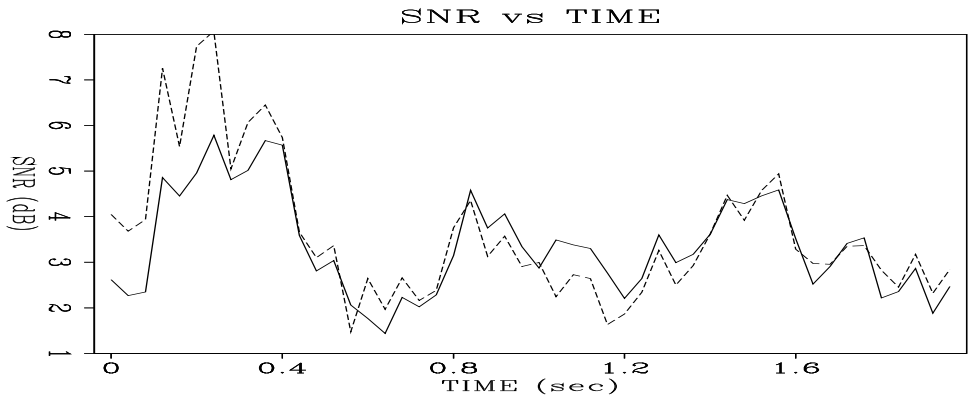

Figure 21: SNR *vs* TIME. The solid line represents case A and the dashed line case B. yalei-TIMESNR44 [CR]

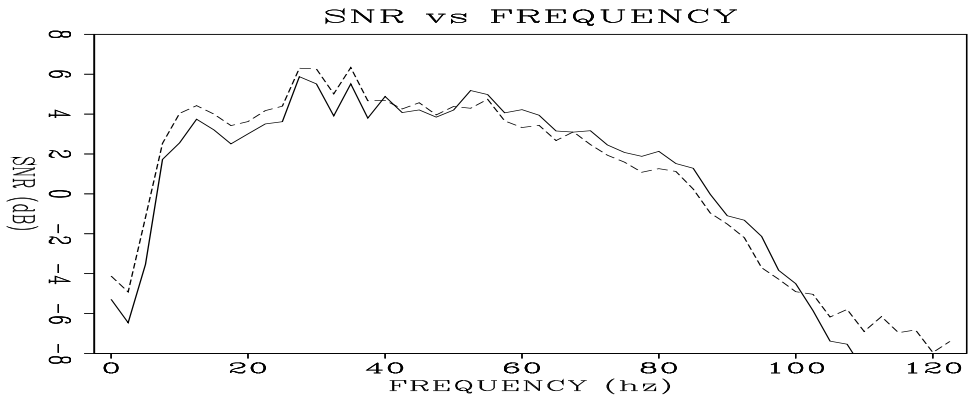

Figure 22: SNR *vs* FREQUENCY. The solid line represents case A and the dashed line case B. yalei-FREQSNR44 [CR]

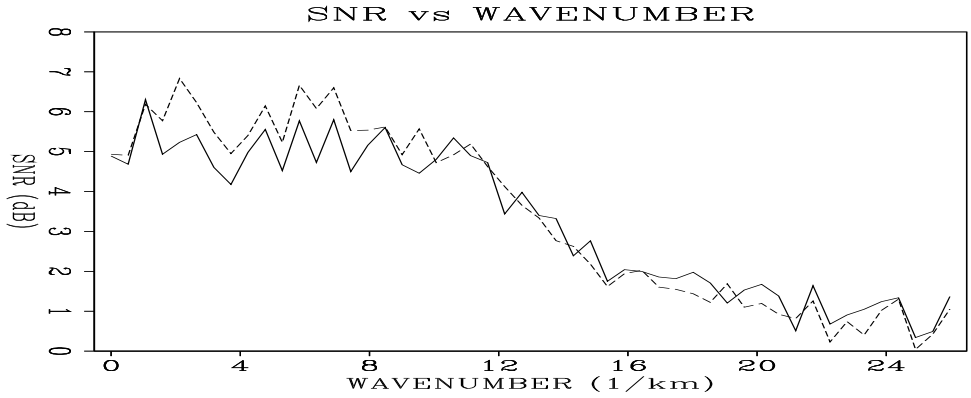

Figure 23: SNR *vs* WAVENUMBER. The solid line represents case A and dashed line case B. yalei-WAVENUMBERSNR44 [CR]

result in different compression ratios when applied to different datasets. Even at the same compression ratio, the compression quality (for example, SNR) also varies from one dataset to another. Therefore, it is necessary to perform some tests and compare the corresponding results to select an appropriate *scale* value.

- The organization of the dataset into different groups to facilitate the compression. In order to achieve a better result, we should make use of the coherency of the dataset as much as possible. The reasonable organization of the dataset is essential to the exploration of the coherency. Usually, for 3-D datasets, the common-shot gather and common-midpoint gather are both acceptable input forms for the compression session.

- The application of some conventional processing, such as NMO correction and AGC gain, to the original dataset prior to carrying out wavelet transform. As shown in last section, NMO correction can enhance the coherency of the original dataset and improve the performance of this compression technique. Thereafter, applying some processing procedures which can strengthen the coherency is undoubtedly beneficial to the final result. But there is a trade-off between the time spent by preprocessing and the effect achieved by the same procedure.

ACKNOWLEDGMENTS

We would like to thank Chevron for allowing us to use their software. Ray Ergas from Chevron gave a lot of help on using this package. Jon Claerbout suggested we use the second synthetic dataset. Matthias Schwab has provided many insights about this project that have been proved to be very useful. Many thanks to Bob Clapp and Sean Crawley for their help on how to use SEPlib.

We would like to thank Conoco and its partners BP and Mobil, for agreeing to release the real data that we showed in this paper.

REFERENCES

Bradley, J. N., and Brislawn, C. M., 1994, Wavelet/scalar quantization compression standard for digital fingerprint images: Proceedings of the 1994 IEEE International Symposium on Circuits and Systems, 205–208.

Cohen, J. K., 1992, Wavelets – a new orthonormal basis: Colorado School of Mines.

Daubechies, I., 1992, Ten lectures on wavelets: Philadelphia, SIAM.

Donoho, P. L., Ergas, R. A., and Villasenor, J. D., 1995a, High-performance seismic trace compression: 65th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 160–163.

Huffman, D. A., 1952, A method for the construction of minimum-redundancy codes: Proceedings of the IRE, **40**, no. 9, 1098–1101.

Nelson, M., 1995, The data compression book: The M&T Books.

Stigant, J. P., Ergas, R. A., Donoho, P. L., Minchella, A. S., and Galibert, P. Y., 1995b, Field trial of seismic compression for real-time transmission: 65th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 960–962.

Storer, J. A., 1988, Data compression: Computer Science Press.

Villasenor, J. D., Belzer, B., and Liao, J., 1995, Wavelet filter evaluation for image compression: IEEE Transactions on Image Processing, **4**, 1053–1060.

Witten, I. H., Neal, R. M., and Cleary, J. G., 1987, Arithmetic coding for data compression: Communications of the ACM, **30**, no. 6, 520–540.