Pre-stack and Post-stack Data Compression and its Impact on Seismic Attributes

Márcio Spínola, Raphael Melo, Carlos Henrique, HALLIBURTON LANDMARK
Juan Soldo, PETROBRAS

Abstract

With the multiplication of seismic data being generated each year for exploration and reservoir purposes, storage and management of pre-stack and post stack seismic files are still a challenge for infrastructure management. Even today, compressed seismic files make storage and management of large original volumes highly practical. At the end of the day, geoscientists and interpreters want to integrate all available data and, preferably, with the highest possible fidelity.

This work attempts to present the implications of compressing seismic files on both pre-stack and post-stack data analysis. First compressing methodologies are presented, and then a series of analyses of compressed volumes using different parameters were compared with original non compressed data. The results are discussed with examples.

Introduction

In the Brazil national repository of exploration and Production Database (BDEP), the collection of data in December 2010 was about 3.12 petabytes of seismic data [1].

Presently, there is a increase of data and information being generated on a daily basis in the geophysical community from both 2-D and 3-D surveys, which are designed and recorded for different purposes, including basin exploration, 4-D reservoir monitoring, high-resolution 3-D, multi-component seismic (2C, 3C, 4C), etc.

In this digital era, disk space rationalization is becoming an increasing issue, even though prices per GB of hard drive [2] are continually decreasing, as shown below.

Year	Estimated price/GB
1981	$300,000
1987	$50,000
1990	$10,000
1994	$1000
1997	$100
2000	$10
2004	$1
2010	$0.10

Pre-stack data tends to be the villain in terms of data volume and is, thus, often stored on tapes. Despite a decrease in the hard-drive cost, it seems not yet practical to store all of the acquired pre-stacked data on disk. This is not the case for post-stack data that, after processing, become available to interpret on fast accessed disks.

Although the post-stack dataset is quite small compared to pre-stack, it still requires some compression, as discussed in Spinola et al. 2009.

Regarding post-stack, this work only addresses connectivity issues when using compressed files and tries to give more emphasis on another method of compression and how it works on pre-stack data.

Compression methods

a) The Fidelity Factor Compression

Spinola et al. 2009 addresses compression on post-stack data using a fidelity factor [3] (FF) that is expressed by Eq. 1:

$$FF = 1 - \left[\frac{\sigma}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}} \right] \times 100 \qquad (1)$$

Where, $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} xerr_i^2}$ is the RMS Error, which is also called standard deviation, if normal distribution, 68.26%, of the population has less than RMS Error, *x* is the measure (signal), *xerr* is the error, and *n* is the number of samples.

The compressed files using FF produces exceedingly high compression rates (from 5 to 72 times). Depending on the data, small variations in FF can produce huge differences in volume size. This abstract will show that interpreters must be very careful when applying FF values lower than 99.9% because phase information is being lost, which may influence the interpretation of subtle stratigraphic or structural features (Spinola 2009).

b) The SeisPEG Compression

The SeisPEG compression algorithm is a lossy technique, similar to the widely-known JPEG standard but with adaptations for seismic data. The essence of JPEG and JPEG-like algorithms, such as SeisPEG, is that the data is broken into blocks during compression, and lapped transforms are used to avoid blocking artifacts. The advantages of a blocked approach for seismic data include computation efficiency, more effective handling of poorly-gained data, and the ability to decompress subsets of the data.

There are three steps in the SeisPEG compression process, including:

1. Lapped Orthogonal Transform (LOT): The purpose of the lapped orthogonal transform is to pack most of the signal energy into a minimum number of coefficients.

2. Quantization: This step reduces the number of unique values that occur in the data in the transformed domain.

3. Minimum redundancy coding: This step represents commonly occurring values in the transformed data with fewer bits than values that do not occur commonly.

The **LOT** is used in SeisPEG instead of the more commonly-used DCT (Discrete Cosine Transform) because the LOT better mitigates blocking artifacts and shows greater coding efficiency. **Quantization** essentially consists of scaling data to smaller values and then rounding to the nearest integer so that fewer possible discrete values exist in the data, which can, therefore, be represented with fewer bits. **Minimum redundancy coding** is the final step in SeisPEG compression. It is used to squeeze out redundancies in the data, which is a perfectly lossless operation.

The JPEG and SeisPEG algorithms can be adjusted to different quality levels, gaining higher accuracy in exchange for less-effective compression. In SeisPEG, the approximate amount of distortion to tolerate is specified during the compression process by the level of distortion parameter.

Methodology

In this work, we used two different methods for the compression that was previously discussed. The original amplitude 3-D SEGY 32-bit seismic dataset was compressed using a fidelity factor with different parameters, and the pre-stack data was compressed using the SeisPEG format with different distortions. Both algorithms are provided by Landmark. Illustrating how compression of post-stack seismic data may impact the quality of seismic attribute, we selected a connectivity volume to analyze. Connectivity can illustrate effects on subsurface geological mapping and the volumetrics of geobodies.

The same idea was applied to pre-stack data. Therefore, the original pre-stack seismic has compacted using SeisPEG compression algorithm with different distortions parameters.

Some tests of compressing and writing seismic data to disk were performed for post-stack and pre-stack data using this distortion parameter. Figures 1 and 2 present these different parameters of distortion versus size in megabytes of original file. Note the similar shape in pre-stack and post-stack curves using SeisPEG compression. Also, for comparison, a simple reduction in the original file converted from 32-bit to 16-bit and 8-bit are plotted.

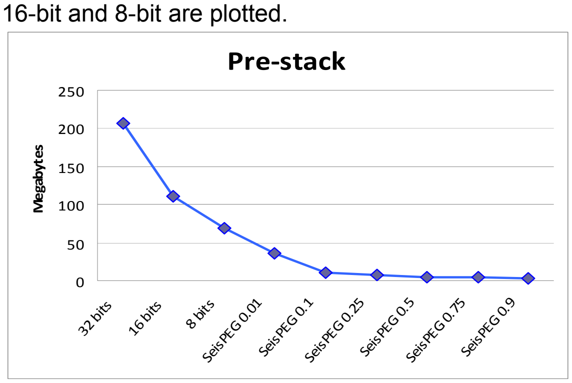

Figure 1: Original post-stack file compresses using SeisPEG.

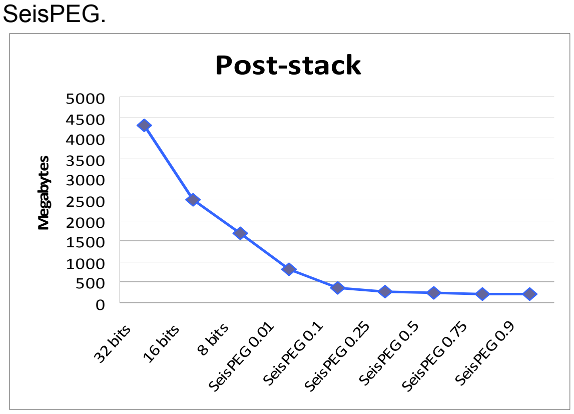

Figure 2: Original pre-stack file compressed using SeisPEG.

Data Analysis and Discussion

a) Post-stack Data

In order to visualize the impact compression of post-stack data causes on interpretation, we analyzed the connectivity attribute, as shown in Figures 3 and 4. Both Figures display geoanomalies (automatic volume segmentation using the same parameters - negative amplitude and volumes greater than 1000 voxels) with color uniquely representing connected geobodies. The volume of each geobody can be measured and shows an increase proportional to the increase in compression rates.

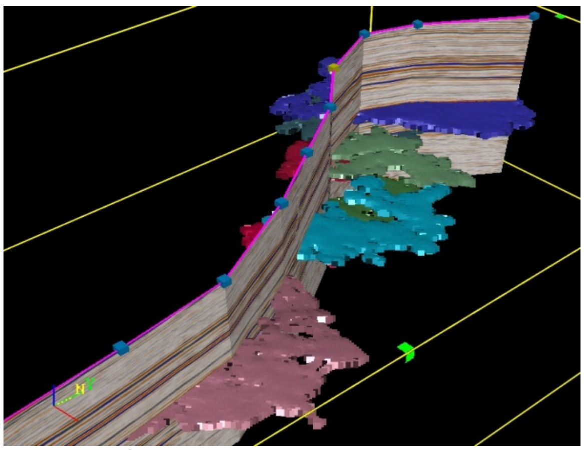

Figure 3: Original seismic data displayed on the ribbon section: geoanomalies computed from this data are based on negative amplitude rage (-128 to -50) and volumes greater than 1,000 voxels. Geobodies in blue, white, blue, and pink are at the same reflector level.

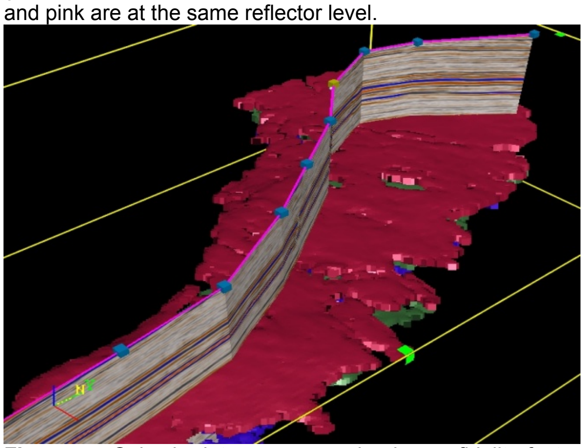

Figure 4: Seismic data compressed using a fidelity factor of 10 displayed on the ribbon section: geoanomalies computed from this data are based on negative amplitude rage (-128 to -50) and volumes greater than 1,000 voxels. Note that the geobody in red is one connected volume instead of three separated volumes, as in Figure 3.

In this case, the three isolated geobodies detected on the original volume become a single, larger geobody on the compressed seismic data. The recommendation here is to reduce compression as much as possible because the high compression, or bad clipping, is compromising the interpretation. Therefore, be careful because small variations in fidelity factor increase the connectivity of geobodies substantially.

It is quite clear from previous paper that FF produces exceedingly high compression rates (from 5 to 72 times). The small variations in FF produce huge differences in volume size in MB. To visualize the impact it causes on interpretation, phase volume differences were studied.

The results from this analysis reveal that care should be taken when applying FF lower than 99.9 because phase information is being lost, which may influence interpretation, especially in subtle stratigraphic or structural features.

b) Pre-stack Data

To visualize the impact of compression on the pre-stack data, a series of differences of amplitude versus offset was computed for compressed and non-compressed data.

Figures 5–10 show differences between pre-stack seismic amplitude of compressed and non-compressed data using SeisPEG of different distortions. Note the random pattern that the computed differences of the pre-stack data using distortions 0.01 and 0.1 displayed in Figures 5 and 6. For these two cases, the compression set seems to be acceptable. In Figures 7, 8, and 9, showing distortions 0.25, 0.5, and 0.75, respectively, is no more acceptable since computed differences fails to show a random pattern.

The amplitude spectrum of pre-stack images for SeisPEG compression using distortions 0.01, 0.1, and 0.5 were computed and compared with non-compressed data (Figures 10, 11, and 12). For distortions 0.01 to 0.1, the amplitude spectrum is still similar, but the low frequencies and the high frequencies begin to be impacted at distortion values of 0.1. Highly-significant amplitude spectra differences can be seen when using a distortion of 0.5

Figure 13 shows amplitude-versus-offset plots for both compressed and non-compressed data. Note that the curves look the same when using a small

SeisPEG distortion parameter, like 0.01 but not acceptable for 0.25 and 0.75.

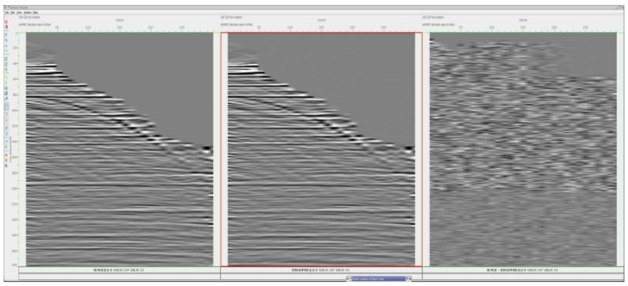

Figure 5: Left pre-stack data without compression, middle distortion 0.01 and on the right differences.

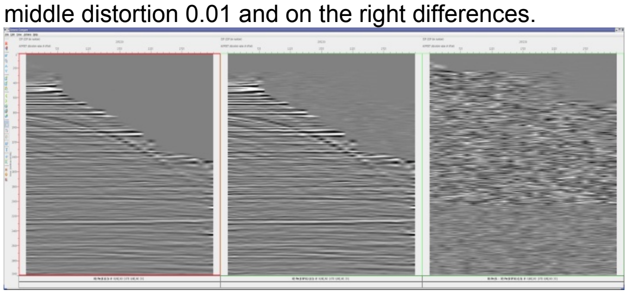

Figure 6: Left pre-stack data without compression, middle distortion 0.1 and on the right differences.

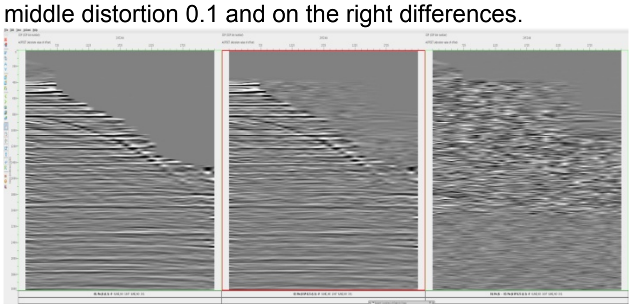

Figure 7: Left pre-stack data without compression, middle distortion 0.25 and on the right differences.

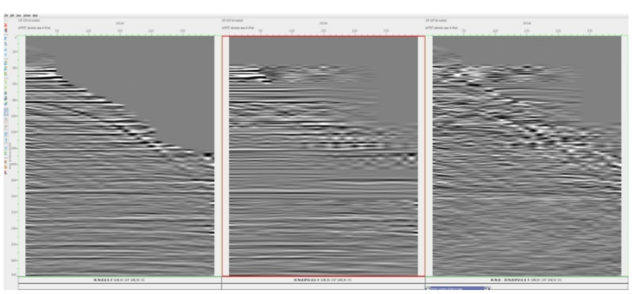

Figure 8: Left pre-stack data without compression, middle distortion 0.5 and on the right differences.

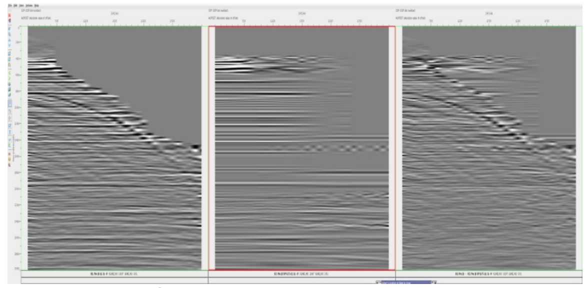

Figure 9: Left pre-stack data without compression, middle distortion 0.75 and on the right differences.

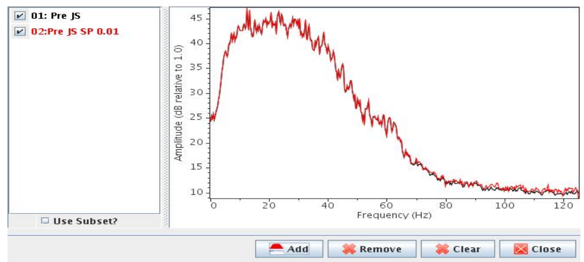

Figure 10: Comparison of amplitude spectrum of pre-stack data: non-compressed data with SeisPEG distortion 0.01.

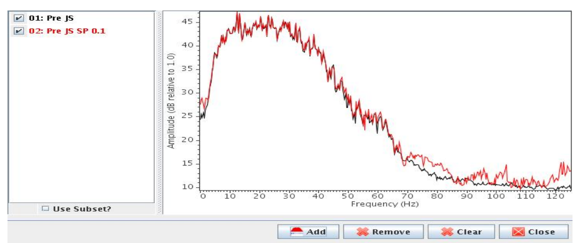

Figure 11: Comparison of amplitude spectrum of pre-stack data: non-compressed data with SeisPEG distortion 0.1. Note low and high-frequency distortions.

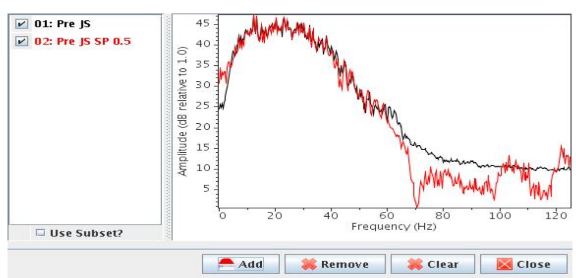

Figure 12: Comparison of amplitude spectrum of pre-stack data: not compressed data with SeisPEG distortion 0.5. Note low and high-frequency distortions.

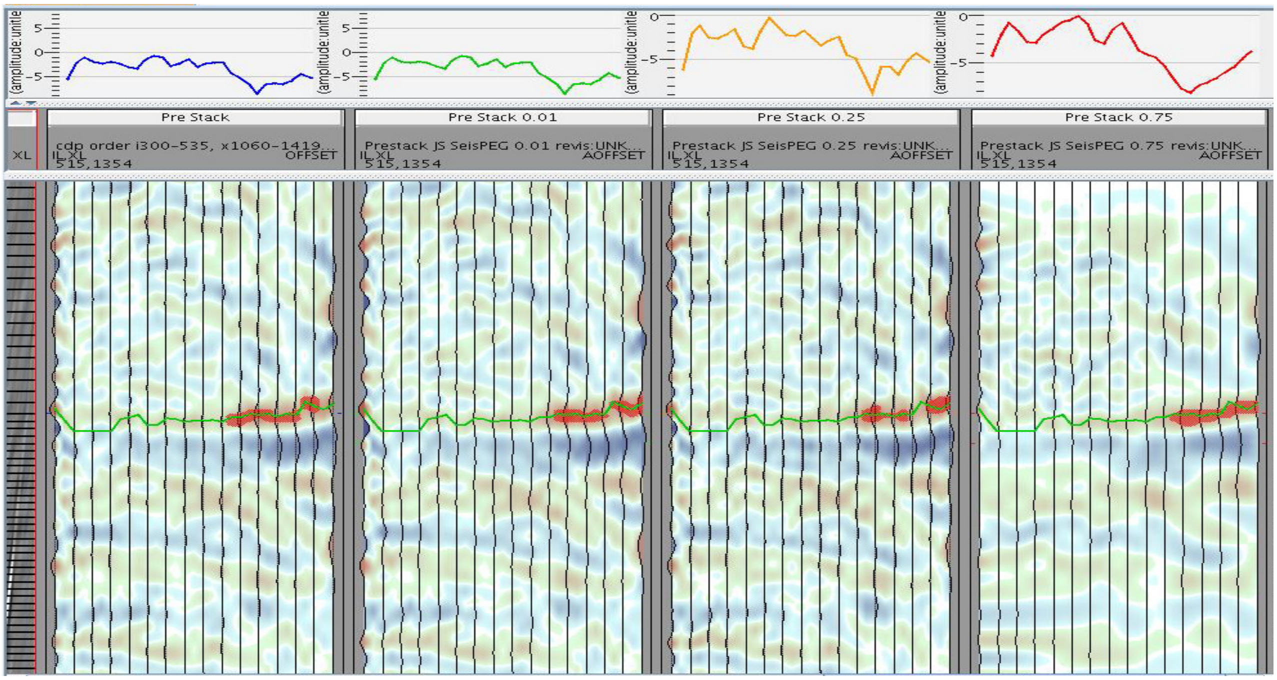

Figure 13: Amplitude-versus-offset at interpreted pre-stack horizon close to 2,400 ms (class 3 type anomaly). Note AVO shape curves blue, green orange and red (top panels) and pre-stack data original and compressed using SeisPEG distortions 0.01, 0.25 and 0.75 (respectively from left to right). Note the AVO shape curves blue and green looks the same (non compressed and compressed using distortions 0.01). Same color map was used for pre-stack data. Solid red on pres-stack section highlights same negative threshold values of AVO obtained from cross-plotted polygon .

Conclusions

Clearly, compressing 32-bit seismic data saves a lot of disk space. There are benefits from the data management point of view; however, over compression can effect interpretation, as demonstrated by the connectivity of geobody differences at FF values less that 99.9%.

An analysis of this dataset has shown that compressing the pre-stack seismic data can be very helpful in reducing data sizes. However, increasing compression factors beyond 0.01 to 0.1 leads to distortions on amplitude spectrum and loss of information, which will also impact the interpretation. The amplitude-versus-offset are very similar for small distortion parameters (0.01) and may not significantly impact the AVO analysis if applied at these low levels.

Acknowledgments

We would like to acknowledge Halliburton for granting permission to publish this work, Larry Fink from Halliburton Landmark Graphics, RMOTC and the U.S. Department of Energy as the data source as well as Statoil for providing the pre-stack training data.

References

Spínola, M., Silva, C. H. , Silva B., 2009 Issues in Seismic Data Compression and Attribute Calculations for Interpretation. Presented in Brazilian Geophysical Society held in Salvador, Brazil, August 24-28.

SeisSpace® Tools Reference**,** 2010 ¨ SeisPEG Compression Test ¨. Houston, EUA

The Department of Energy, US Federal Government, 2007. Teapot Dome Seismic Dataset Rocky Mountain Oilfield Testing Center, Teapot Dome, Natrona County, Wyoming, EUA

[1] http://www.anp.gov.br/?id=600

[2]http://notebooks.com/2011/03/09/hard-drive-prices-over-time-price-per-gb-from-1981-to-2010/ (accessed 03-20-11)

[3] Promax® Reference Manual ¨Summary of Landmark´s Seismic Data Formats¨