Issues in seismic data compression and attribute calculations for interpretation

Márcio Spínola*, Carlos Henrique, Bruno Silva, HALLIBURTON LANDMARK, Brasil

This paper was prepared for presentation during the 11th International Congress of the Brazilian Geophysical Society held in Salvador, Brazil, August 24-28, 2009.

Abstract

The 3D postack seismic data is known to be one of the most important tool for interpreters in order to identify structural and stratigraphic framework that leads to oil findings. Nowadays there is a tsunami of information being generated on a daily basis and the compressed seismic files make storage and management of large original and attribute volumes practical. Although interpreters might not be interested on hard disk drive space rationalization, it is necessary for oil operators because implies on capital allocation. This works attempts to discuss the issues when compressing seismic data and the impact it causes qualitatively and quantitatively on quality data for attribute generation and interpretation. Series of analysis of compressed volumes using different parameters were compared with original non compressed data. Resulted are discussed with examples.

Introduction

Despite of the fact of existing 2D and 3D seismic surveys are already in the national repository (exploration and production data banks) representing hundreds of Tera Bytes, nowadays there is a tsunami of information being generated on a daily basis in geophysical community with 2D and 3D surveys designed and recorded for different purposes, recurrent 3D acquisitions (time lapse surveys), high resolution 3D over known areas, multicomponent seismic (2C, 3C, 4C), etc.

In a digital era disk space rationalization has became an increasing demand although prices per GB of Hard Drives seems to be decreased.

This leads to the needs of getting these monsters seismic datasets being compressed after processing to be available on the network before being delivered to the interpreters as poststack seismic data sets ready for interpretation.

This works attempts to discuss the issues when compressing seismic data. The parametrization of compression and the impact it causes on quality data for attribute generations and interpretation. We took an original 32bit format seismic dataset and compressed using available commercial algorithms to Float 32 bits with different parameters and studied the impact it produces on the amplitude preservation and attribute extraction.

Rather than randomize the compression parameters, it is better to preserve as much as possible amplitude information making reliable workflows of reservoir characterization or prospect generation. Usually the subtle amplitude changes and phase information preservation are essential for target definitions and well planning.

Fractures, sequence stratigraphy, horizon terminations and discontinuities represent seismic facies changes. These changes have to be found and understood in the dataset, with the hope that all available amplitude information be available and preserved.

Interpreters might not be interested on disk space at first instance but data base management people should care about it, because it implies on capital allocation associated with storage and network traffic specially when dealing with huge datasets. Then which would be the best parametrization on order to compress data? Which are the valid parameters and how it impacts on the quality of data and interpretation? How to preserve disk space while not destructing subtle information data set may contain? This work attempts to ilustrate some of these questions.

Seismic Data Formats and Compression

Before continue some words on data formats and definitions are addressed. Data can be organized in different ways. Seismic section can be in vertical files indexed as line, traces or time slices in vertical or horizontal indexed files. Data can also be organized as bricks with the advantage of bricks being accessed to generate arbitrary lines in compute less intensive way.

The brick dimensions can also be designed to suite workflows needs. The dimension of bricks implies on number of inline or crossline and time/depth which are the number of samples in horizontal dimension (crossline, inline) and time/depth dimension.

Tools provided by vendors offer compression algorithms which work on small blocks of volume independently, also using overlapping block. The algorithm uses bricks to organize data and these bricks have 8x8x8 samples in dimension. During compression each sample contributes to 8 blocks with a sample. By providing amplitude-range-preserving compression ratios and parameters different file sizes can be generated.

Controlling compression is given by an indirect factor called Fidelity Factor (FF) that is expresses by the Equation (1):

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

where • = $\sqrt{\frac{1}{n}\sum_{i=1}^{n} xerr_i^2}$ is RMS Error is also called standard deviation (Sheriff, 2006), if normal distribution 68,26% of the population has less than RMS Error and *x* is the measure (signal) and *xerr* is the error and *n* is the number of samples.

Methodology

Data set used in this work is public domain published by the Department of Energy - US Federal Government, and was produced by the Rocky Mountain Oilfield Testing Center. It is named Teapot Dome and was released on June 13th, 2007. It was acquired in Natrona County, Wyoming - USA.

The original amplitude 3D SEGY file (Figure 1) is in 32bits floating point and has about 400MB file size. It was compressed (.cmp file type) using different compression rates (FF) and their resulting file size were measured (Table 1). FF 100% represents original non compressed volume.

Figure 2 shows the rate of compression in MB by the Fidelity Factor based on Table 1. Note small variations in Fidelity Factor produces huge differences in the file size in MB. For instance FF=99.9 results seismic files 6 times smaller than original and 99.5 a file 9 times smaller.

With purpose to make comparisons, interpretation and volume attribute calculation analysis was started and each compressed file was loaded to volume visualization and interpretation package in 8 bits. Neither decimation or amplitude clipping was applied. During scaling to 8 bits all extreme amplitude range were preserved (no clipping).

Three volumetric attributes derived from amplitude commonly used by interpreters were selected: semblance, phase and a combination of instantaneous amplitude and instantaneous frequency, namely Sweetness (Amp/$\sqrt{InstFreq}$).

These results were tested against original attribute derived from non compressed volume. Total of 50 volumes were generated including difference volumes and results are discussed in the next section. Comparisons were made in terms of statistics, volume differences and interpreted horizon slices both qualitatively and quantitatively.

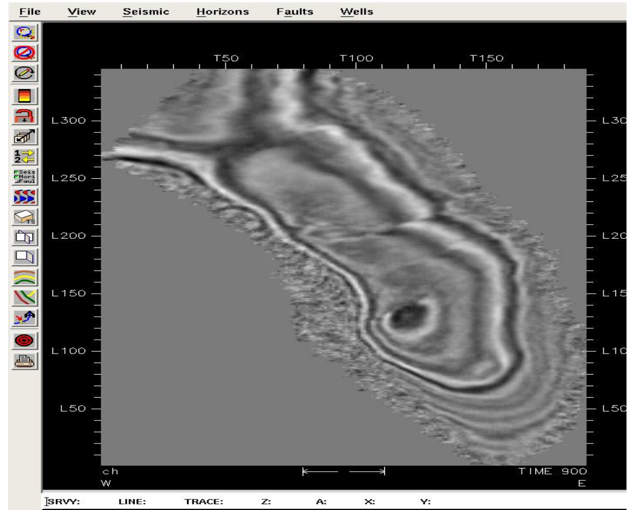

Figure 1 - Time slice at 900ms with a nice closing structure.

Fidelity Factor (FF)	File Size MB	Number of Times Smaller
100.0	389.4	1.0
99.9	68.6	5.7
99.5	43.1	9.0
99.0	35.2	11.1
98.0	28.7	13.6
95.0	21.7	17.9
90.0	17.1	22.7
75.0	11.7	33.2
50.0	8.1	48.2
10.0	5.4	72.4

Table 1: Fidelity Factor and resulting file size in MB. Third column displays the number of times the resulting compressed file is smaller than original.

Data analysis & Discussion

It is very clear from Figure 2 and Table 1 that FF produces compression rates very high (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 first.

Figure 3a shows computed phase volume differences between reference phase volume, obtained from non compressed seismic amplitude and phase volume computed from amplitude compressed with FF 99.9. It can be seen slightly random points in this seismic section, but no eloquent differences can be found.

Same comparison was performed for Fidelity Factor 99.5 (Figure 3b), Fidelity Factor 98 (Figure 3c) and Fidelity Factor 75 (Figure 3d). Note that little variations in Fidelity Factor increase substantially the differences and distortion is produced. Note also coherent noise introduced in Figure 3c and 3d.

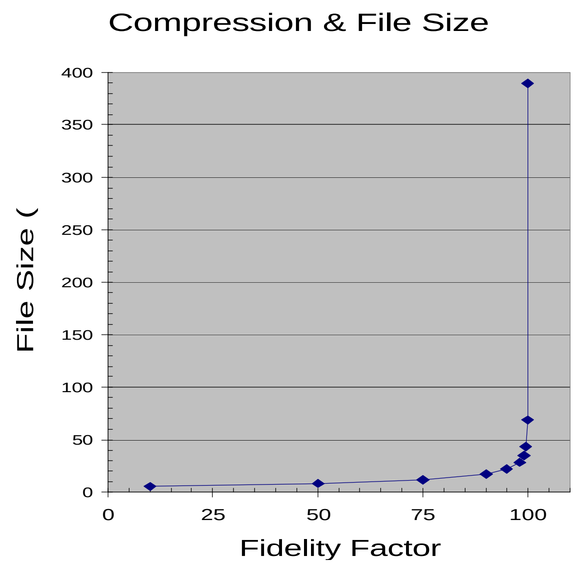

Figure 2: Compression rates and resulting files sizes. Few variations in Fidelity Factor produce huge differences in the file size in MB.

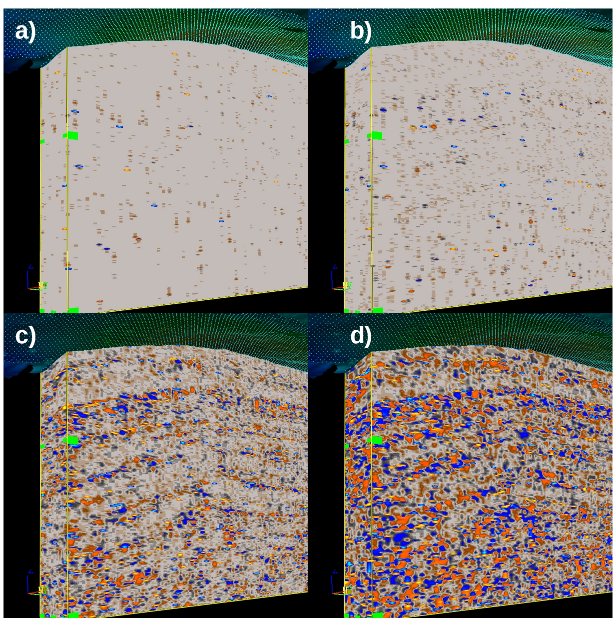

Figure 3 - Difference phase volumes between original phase computed from amplitude no compressed volume and phase computed from amplitude volumes with Fidelity Factor a) 99.9; b) 99.5; c) 98; d) 75.

So very careful must be taken when applying FF lower than 99.9 because phase information is being lost and this may influence interpretation specially in subtle stratigraphic or structural features.

Semblance volumes were also investigated in respect to their differences. Figure 5 displays a horizon slice of the semblance differences for each fidelity factor. Note in Figure 5a and 5b some information is lost and FF 90 and FF 10 difference increases and become not practical for interpretation purposes.

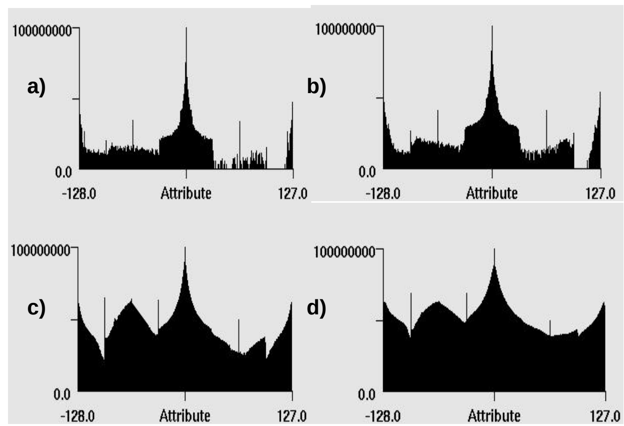

Figure 4: Histogram of difference (original minus compressed) phase volumes. a) compression parameter respectively a) 99.9; b) 99.5; c)99.0 and d) 75.

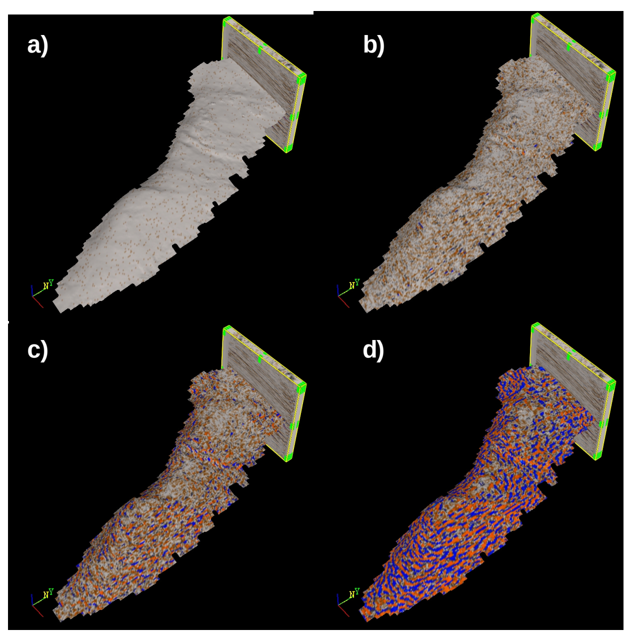

Figure 5 - a) Difference on semblance 99.9; b) Difference on semblance 98; c) Difference on semblance 90; d) Difference on semblance 10.

Quantitative investigation can be seen on computed histogram statistics of phase volume differences (Figure 4). Note that logarithimic scale was applied to the histogram in order to enhance frequency classes. It

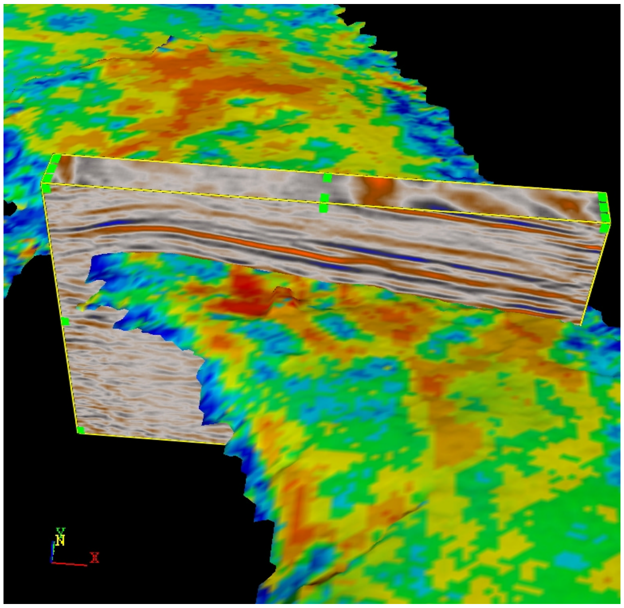

Figure 6 - Shows Sweetness attribute over interpreted horizon computed based on original non compressed volume. Original non compressed amplitude data is displayed on the probe.

became very clear that small increases in compression (decreasing Fidelity Factors) produces higher frequency classes on the histogram thus increasing variance and being an indicative of noise creation.

So what is the best compression rate? Based on the histograms and volume differences, results lead to as smaller as possible, because smaller the compression smaller the phase disturb as well variance. There are benefits on compressing data on data management point of view, but phase information is distorted. Compression rate are also dependent on each particular dataset. Note in Figure 1 that many zero traces are present. Other data set was tested using same compression rates and resulted file size changes with the same Fidelity Factor.

Qualitative visualization on the impact over connectivity can be seen in Figure 6 and 7. Both figures shows sweetness attribute as horizon slice. Figure 7 shows more continuous channels (reddish) when compared with Figure 6 (original sweetness data computed without compression).

Conclusions

Compressing 32-bit seismic data saves a lot of disk space but few work has been done discussing what are the impact of performing these operations on the quality of data, amplitude analysis, phase preservation and attribute extraction for interpretation.

Phase volume differences and histogram analysis shows that information is distorted when compressing data. Increase in compression factors leads to changes in phase that impacts on interpretation on subtle structural and stratigraphic features. There are benefits on

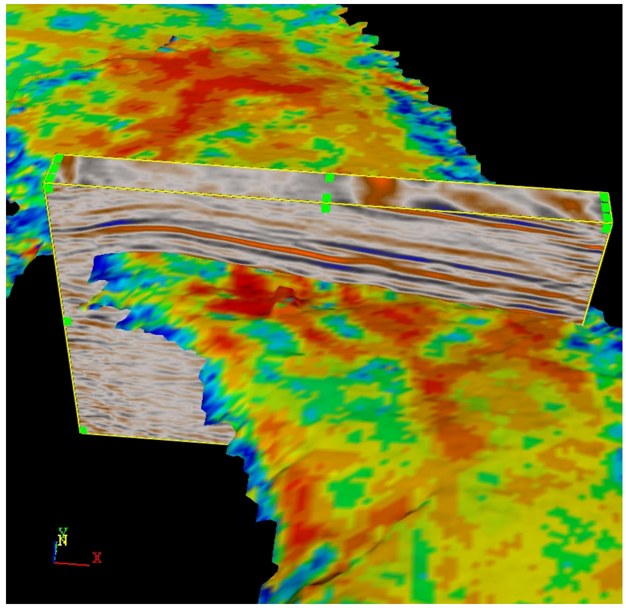

Figure 7 - Shows Sweetness attribute over interpreted horizon computed based on compressed data using 99 fidelity factor. Original non compressed amplitude data is displayed on the probe.

compressing data on data management point of view but it has been shown that phase information is disturbed.

Careful must me taken when applying Fidelity Factors lower than 99.9 because phase information is being lost and attribute (semblance amplitude and frequency) computed are distorted.

Wearing geoscientist glasses recommendation here is to reduce compressing as much as possible rather than a high compression or bad clipping.

Acknowledgments

We would like to acknowledge friends from Landmark Érica Filgueiras, Stevenson Hall and Steaven Miler, Halliburton for allowing to publish this work as well as RMOTC and the U.S. Department of Energy as the data source.

References

Duarte, O de Oliveira. 2003. Dicionário Enciclopédico Inglês Português Geofísica e Geologia 2ª. Edição, Rio de Janeiro, SBGf

ProMAX Reference Manual, 2008 ¨New 3D Data Formats¨. Houston, EUA

Sheriff, Robert E., 2006. Encyclopedic Dictionary of Applied Geophysics – 4th edition (Geophysical Reference; v14), SEG, Tulsa, OK, EUA. 2006.

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