# The Effect of Lossy Compression on the Quality of Maps in Geophysics

Henrique Frederico Bucher, AeroGeoPhysica Latino América (AGP-LA)
Francisco Andrade, AeroGeoPhysica Latino América
Éder Molina, Universidade de São Paulo (USP)



This paper was prepared for presentation at the 8th International Congress of The Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 14-18 September 2003.



## Abstract

This paper discusses the effect of applying the wavelet transform, as part of a lossy compression scheme, to the final quality of maps generated from geophysical airborne campaigns. The final goal of this paper is to develop a methodology of compression specific to geophysical data so the final information format (map) is not corrupted or deformed, not leading to any kind of misinterpretation that would not happen without the compression scheme. We start by analysing the current map production pipeline – as usually encountered in the industry – in detail, stressing all transformations performed with data. Then, based in an analytical model of this process, we depict opportunities to eliminate pieces of non-important information, thus introducing compression, that would be otherwise deleted by the transformation pipeline. Compression ratios are then studied using a series of data collected in the field.

## Introduction

The underlying premise on constructing maps for geophysical purposes is that every piece of information is displayed with no directional bias other than true geophysical structures hidden below surface. Otherwise the interpretation of such "biased" maps would be misleaded into false conclusions about imaginary structures, mainly on top of each flight line.

Therefore we would require that any operation performed in the pipeline be direction-independent, which can be translated to the simple requirement that the spectral content of interpolated data be basically the same as of the data that originated it. In the special case of two flight lines separated by a distance *s*, experience shows that in the middle of those lines most of the high-frequency content is simply removed by natural smoothing of the interpolation algorithm [1][2].

Therefore, in the case of maps with geophysical content, high-frequency content must be discarded anyway prior to the plotting phase. This information can be used to boost compression ratios obtained by the simple operation of thresholding non-significant coefficients generated by the wavelet transform.

## The Wavelet Transform

The wavelet theory [3][4][5] has proved to be a great tool when classification of data in multiresolution levels is needed. Numerically, the wavelet transform takes a vector of measurements *x* and produces a vector of coefficients *c* where each element corresponds to a basis function centered in a specific instant of time and within a specific band of frequency. Fourier transform basis functions, by contrast, have only frequency location – they have infinite support in time. A more detailed description on this issue can be found in [6].

The result of this distribution of coefficients on both time and frequency is that very often just a few coefficients have significant values and most of them can be zeroed, operation called *thresholding*. By keeping only those coefficients above a certain threshold, along with their respective location, one can generate a sparse structure much smaller in size than the original vector.

Given that each coefficient is also responsible for a distinct space in the time-frequency plane, one can use wavelet thresholding for filtering data. If, for instance, high frequency content is to be filtered out, one can wavelet-transform the original data, discard all coefficients whose frequency range is within the discardable area and then perform the inverse transform, thus obtaining the filtered data. Therefore we notice that filtering and compression are obtained in the wavelet compression framework by the same operation.

## Analytical study of spectral content

A standard way to convert scattered points as acquisited in the field into an ordered set of points (or grid), is to solve a biharmonic equation with a linear parameter associated with tension as

$$\Delta^2 u(x,y) - t\Delta u(x,y) = 0 \qquad (1)$$

where u(x,y) is the interpolated field. As this equation also describes the behaviour of a thin membrane, the term *t* is usually recognized as the tension factor. This partial differential equation (PDE) Is usually solved using an iterative algorithm for which specific preconditioners are available [1].

For the purpose of this analytical study, we are interested on the spectral contents of data on a data line and its relation to the spectral content of information at the rest of the map. More specifically, we are interested on finding a transfer function in wavelength domain, if exists, in such a way that it would be possible to write

$$U_{map}(\lambda) = H(\lambda) U_f(\lambda) \qquad (2)$$

Where *Uf* refers to the collected data in a flight line, *Umap* represents the interpolated data, *H* is the transfer function and *λ* is the wavelength.

To solve the above PDE (1) we follow a spectral approach. First we trace two horizontal data lines, one at y=0 and the other at y=1. It is important to remember that this choice of axis is easily extended to any arbitrary axis in x-direction by a linear transformation of the *y* variable. This linear transformation does not interfere on the values of any constant $C_i$ or roots of equation (3), both given later, which depend only on the tension factor *t* and the wavelength *λ*.

Second, we apply a Fourier transform in x-direction, which is commonly referred as the k-y transform, obtaining the following fourth-order differential equation (3) for which the variable's dependency (*U*) on the wavelength was hidden for sake of readability

$$\left[\frac{d^4}{dy^4}-\left(\frac{2}{\lambda^2}+t\right)\frac{d^2}{dy^2}+\frac{1}{\lambda^4}+\frac{t}{\lambda^2}\right]U(y)=0 \tag{3}$$

Such equation has a general solution given by

$$U(y)=C_1e^{\omega_1 y}+C_2e^{\omega_2 y}+C_3e^{\omega_3(y-1)}+C_4e^{\omega_4(y-1)} \tag{4}$$

and can be solved provided proper boundary conditions are given. In the above equation $\omega_{1..n}$ are the roots of the polynomial equation associated to (3) given by

$$\omega_1=-\sqrt{\frac{1}{\lambda^2}+t}\cdot e^{i\frac{\arg\left(1/\lambda^2+t\right)}{2}},\omega_2=-\left|\frac{1}{\lambda}\right| \tag{5}$$

$$\omega_3=-\omega_1,\omega_4=-\omega_2 \tag{6}$$

Then we need to stablish four boundary conditions as, for example,

$$U(0)=f_1,U(1)=f_2,\frac{dU(0)}{dy}=f_3,\frac{dU(1)}{dy}=f_4 \tag{4}$$

For such choice, the constants $C_{1..n}$ can be calculated from the following equations

$$C_1=\frac{\begin{vmatrix} f_1 & 1 & e^{-\omega_3} & e^{-\omega_4} \\ f_2 & e^{\omega_2} & 1 & 1 \\ f_3 & \omega_2 & \omega_3e^{-\omega_3} & \omega_4e^{-\omega_3} \\ f_4 & \omega_2e^{\omega_2} & \omega_3 & \omega_4 \end{vmatrix}}{D} \tag{5}$$

$$C_2=\frac{\begin{vmatrix} 1 & f_1 & e^{-\omega_3} & e^{-\omega_4} \\ e^{\omega_1} & f_2 & 1 & 1 \\ \omega_1 & f_3 & \omega_3e^{-\omega_3} & \omega_4e^{-\omega_3} \\ \omega_1e^{\omega_1} & f_4 & \omega_3 & \omega_4 \end{vmatrix}}{D} \tag{5}$$

$$C_3=\frac{\begin{vmatrix} 1 & 1 & f_1 & e^{-\omega_4} \\ e^{\omega_1} & e^{\omega_2} & f_2 & 1 \\ \omega_1 & \omega_2 & f_3 & \omega_4e^{-\omega_3} \\ \omega_1e^{\omega_1} & \omega_2e^{\omega_2} & f_4 & \omega_4 \end{vmatrix}}{D} \tag{5}$$

$$C_4=\frac{\begin{vmatrix} 1 & 1 & e^{-\omega_3} & f_1 \\ e^{\omega_1} & e^{\omega_2} & 1 & f_2 \\ \omega_1 & \omega_2 & \omega_3e^{-\omega_3} & f_3 \\ \omega_1e^{\omega_1} & \omega_2e^{\omega_2} & \omega_3 & f_4 \end{vmatrix}}{D} \tag{5}$$

where the denominator *D* is a constant given by a similar equation

$$D=\begin{vmatrix} 1 & 1 & e^{-\omega_3} & e^{-\omega_4} \\ e^{\omega_1} & e^{\omega_2} & 1 & 1 \\ \omega_1 & \omega_2 & \omega_3e^{-\omega_3} & e^{-\omega_4} \\ \omega_1e^{\omega_1} & \omega_2e^{\omega_2} & \omega_3 & \omega_4 \end{vmatrix} \tag{5}$$

In order to plot some examples, we need to build a particular solution of (3), which can be done by requiring the solution to be symmetric in respect to y=0 and y=1, which sets two first derivatives to zero (*f3=f4=0*) and also couple their absolute values to be the same in these specific points ($f_1=f_2$). It is important to notice at this point that the solution is not anymore in physical domain but in wavelength domain due to the k-y transform that originated equation (3).

The other boundary condition necessary to completely define all four constants $C_i$ can be obtained by normalization of the solutions' value to one in these two points, i.e., making $f_1=f_2=1$. As the solution is linear with respect to the constants, this is done without any loss of generality.

Therefore, the four boundary conditions are, in summary,

$$\frac{dU(0)}{dy}=0,\frac{dU(1)}{dy}=0 \tag{5}$$

and

$$U(0)=U(1)=1 \tag{6}$$

Unfortunately there is no closed solution for such PDE and the solution is found with the help of some (numeric) linear algebra. However, after some coding effort, one is able to write a script that can be called to evaluate the interpolated function in wavelength domain as *U*(*λ*,*y*). The transfer function cited in (2) is simply recognized as

$$H(\lambda,y)=U(\lambda,y) \tag{7}$$

since $U_f(\lambda)=U(\lambda,y)=1$ on y=0 and y=1 (above flight lines) due to the boundary condition given by (6).

**Results**

Figure 1 shows the result of equation 7 for some wave numbers and tension zero. It is noticeable that for wavelengths bigger or equal to the separation distance there is no loss of content in the middle. However, for smaller wavelengths there is a sensible attenuation. For a wavelength five times smaller than the separation, there is almost a 50% attenuation in spectral content.

If the original, acquisited data is not smoothed previously to the plotting process, the resulting map will certainly have an undesired appearance with high frequency bumps easily identified on top of flight lines [7]. As data is to be filtered prior to plotting, one can use such information as a parameter for compression purposes. For a wave number equal to 10.0 there's a 90% reduction in frequency content.

Data from a geophysical campaign in the state of Minas Gerais [8] was used to validate the algorithms introduced in this work. Figure 7 shows the location of the surveyed area. Figure 2 shows data from a specific flight line, relative to the total gamma-ray counting. This flight line is 143km long and the original data was resampled using bicubic interpolation such that the spacing between points was kept constant at 25.0 m. Data was then transformed using the coiflet wavelet family with 10 taps.

Figure 3 shows the wavelet coefficients sorted by modulus. It can be noticed that the first 30 coefficients hold most of the signal's energy. A threshold value was then selected such that only 286 coefficients (from a total of 5810) were kept – all the others were discarded – thus generating a compression ratio of 20:1. We then applied the inverse wavelet transform to this thresholded vector to obtain the restored version of the signal, also shown in Figure 2.

The effect of thresholding is threefold. First, high-frequency content is removed – as desired – which is equivalent to applying a low-pass filter. Second, there is an overall noise removal. Several studies show that wavelet thresholding is an optimal way to remove white noise from signals [9] and several noise estimators are available for use within this framework [10][11][12][13]. Third, data is compressed by keeping just those high-valued coefficients.

The low-pass filtering effect is evidenciated in Figure 4 where the Fourier transform of both the original and compressed versions of the signal are plotted. In this case it is clear that we zeroed all coefficients related to wavelengths smaller than 125m. For bigger wavelengths, however, both graphs converge to the same curve which indicates that no frequency content was discarded for this region of the spectrum.

Figure 5 shows the residual vector obtained by subtracting the wavelet filtered signal from the original one. It can be seen that the maximum error is around 50.0, which is much less than what would be lost by using a traditional moving average filter. Its respective histogram is shown in the graph below, where the standard deviation value of 11.5 is also plotted. The well-distribution of the residuals energy along the signal and its normal-like distribution shape is a strong indicative that most of the discarded energy is in fact noise.

An important feature to be noticed in Figure 2 is that peaks are well preserved, i.e. their values are kept intact as it can be seen in the peak around 30km from the beginning. Figure 6 shows this peak along with a smoothed version of the original signal obtained by applying a moving average filter with 200m of effective extension (8 taps). The wavelet-filtered version correctly keeps the original value of 458.0 but the moving average version cuts it out to 375.0.

**Conclusions**

This work showed that most high-frequency content must be eliminated prior to the plotting phase such that coherence of the resulting maps is obtained. We showed that for wavelengths smaller than 2 times the separation distance between two consecutive data lines there is a increasing loss of spectral content in the middle greater than 10% and more than 50% for wavelengths 5 times smaller.

A data line, correspondent to a total gama-ray counting, was selected and compressed using the methodology presented. A resulting compression ratio of 20:1 was obtained simultaneously to a desired low-pass filtering treatment. It was observed that the discarded energy had noise-like statistical distribution, suggesting that an optimal noise-removal filtering was also obtained as a side effect. Also, peak values were mostly preserved in contrast with the flattening effect of usual moving average filters.

The particular data set used for compression was composed by data lines with average extension of 100km, which is a limiting factor for compression ratios, specially for data aquisited at low rates, which is the case for gamaespectrometric surveys.

**Acknowledgements**

The authors would like to thank the Brazilian National Petroleum Agency (ANP) for funding the project from which context this paper was elaborated. The authors would like also to thank Secretaria de Estado de Minas e Energia (SEME) and Companhia Mineradora de Minas Gerais (COMIG) for providing the raw data processed in this work.

**References**

[1] Smith, W. H. F, and P. Wessel, 1990, Gridding with continuous curvature splines in tension, *Geophysics*, 55, 293-305.
[2] Wessel, P., and W. H. F. Smith, 1995, The Generic Mapping Tools (GMT) version 3.0 Technical Reference & Cookbook, SOEST/NOAA.
[3 Daubechies, I., "Orthonormal bases of compactly supported wavelets", *Commun. Pure Appl. Math.*, v. 41, pp. 909-996, 1988.
[4] Daubechies, I., *Ten Lectures on Wavelets*, SIAM, 1992.
[5] Mallat, S.G., A theory for multiresolution signal decomposition: The wavelet decomposition, *IEEE Trans. Pattern Anal. Mach. Intelligence*, v.11, pp. 674-693, 1989.

[6] Bucher, H.F., Tese de mestrado, COPPE, Universidade Federal do Rio de Janeiro, 1998.
[7] Minty B.R.S., Luyendyk A.P.J & Brodie R.C. (1997). Calibration and data processing for airborne gamma-ray spectrometry. AGSO Journal of Australian geology & Geophysics. Volume 17, n. 2, p. 51-62.
[8] COMIG (1999). Mapa Geológico do estado de Minas Gerais, Brasil.
[9] Donoho, D.L., “Nonlinear Wavelet Methods for Recovery of Signals, Densities, and Spectra from Indirect and Noisy Data”, In: *Proceedings of Symposia in Applied Mathematics*, v. 00, pp. 173-205, Stanford, 1993.
[10] J. Lu, J.B. Weaver, Y. Xu and D.M. Healy, Jr. "Noise reduction with multiscale edge representation and perceptual criteria," *Proc. of IEEE-SP Time-Frequency and Time-Scale Analysis*, Victoria, Oct. 1992.
[11] J. Lu, V.R. Algazi, and R.R. Estes "Comparison of wavelet image coders using the Picture Quality Scale (PQS)," Wavelet Applications II, Proc. of SPIE, v. 2491, pp. 1119-1130, Apr. 1995.
[12] Rioul, O., “On the choice of wavelet filters for still image compression”, *Proceedings of ICASSP’93*, v. 5, pp. 550-553, 1993.
[13] Mandal, M.K., Panchanathan, S., Aboulnasr, T., “Choice of wavelets for image compression”, *Lecture notes in Computer Science,* v. 1133, pp. 239-249, 1993.

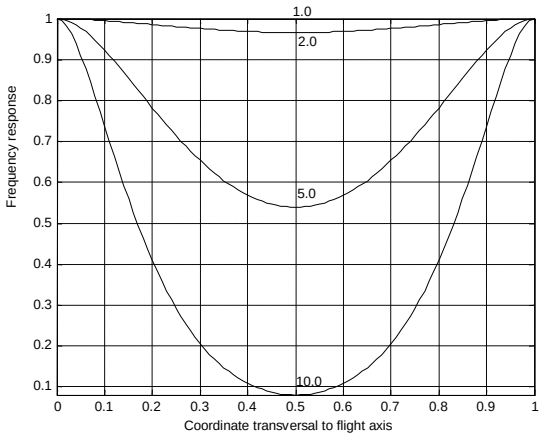


**Figure 1 Frequency content along transversal line between two consecutive data lines**

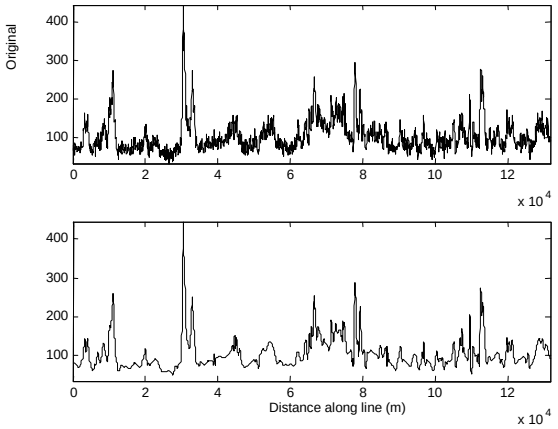


**Figure 2 Original and compressed data from a specific flight line**

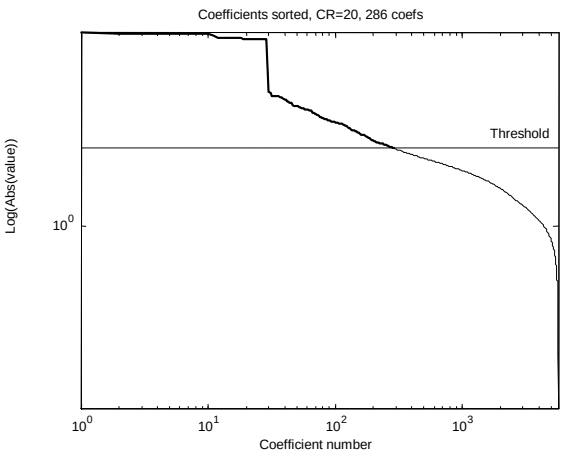


**Figure 3 Wavelet coefficients sorted by modulus**

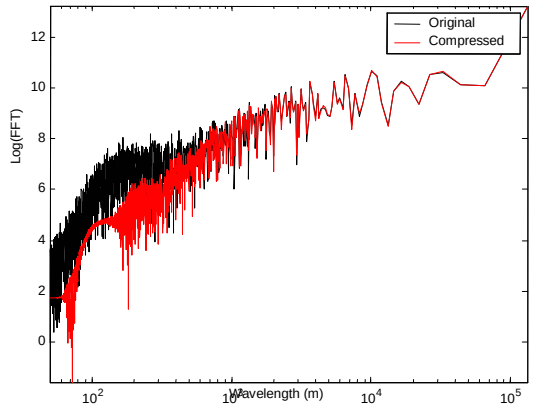


**Figure 4 Modulus of Fourier transform comparing original and compressed data lines**

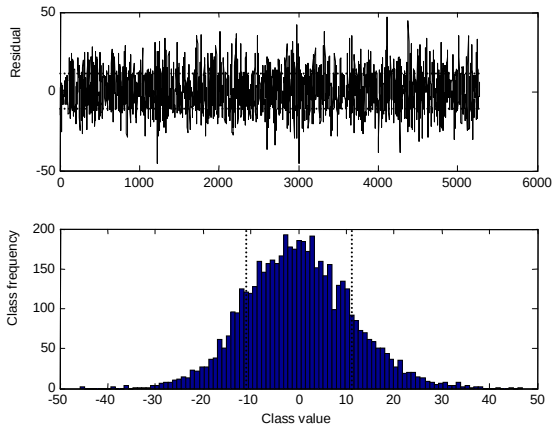


**Figure 5 Residual signal and respective histogram**

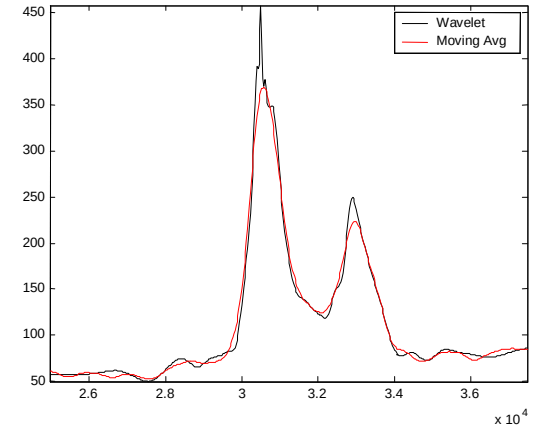


**Figure 6 Wavelet and moving average versions of the original signal showing filtering characteristics of each method**

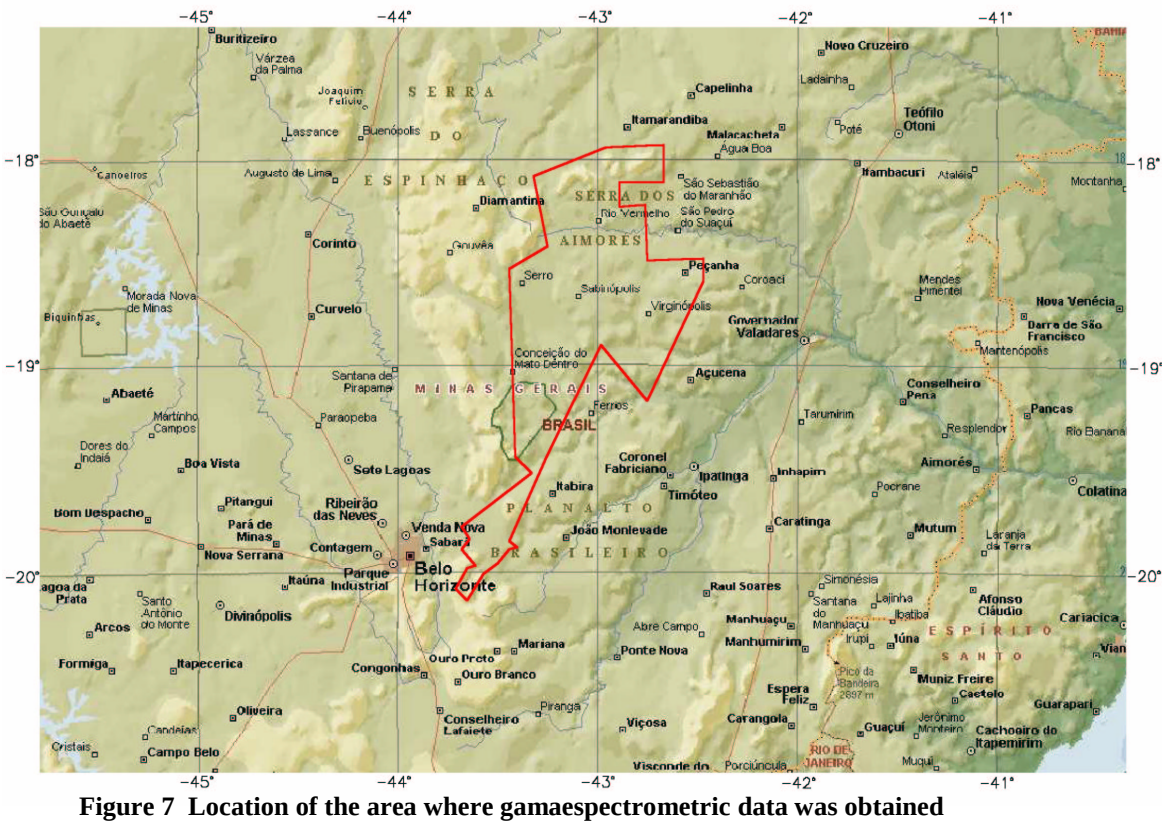


**Figure 7 Location of the area where gamaespectrometric data was obtained**