

Comparing time series using wavelet-based semblance analysis[☆]

G.R.J. Cooper^{a,*}, D.R. Cowan^b

^a*School of Geosciences, University of the Witwatersrand, 1 Jan Smuts Avenue, Johannesburg 2050, South Africa*

^b*Cowan Geodata Services, 12 Edna Road, Dalkeith, Western Australia, Australia*

Received 5 May 2006; received in revised form 2 March 2007; accepted 13 March 2007

Abstract

Similarity measures are becoming increasingly commonly used in comparison of multiple datasets from various sources. Semblance filtering compares two datasets on the basis of their phase, as a function of frequency. Semblance analysis based on the Fourier transform suffers from problems associated with that transform, in particular its assumption that the frequency content of the data must not change with time (for time-series data) or location (for data measured as a function of position). To overcome these problems, semblance is calculated here using the continuous wavelet transform. When calculated in this way, semblance analysis allows the local phase relationships between the two datasets to be studied as a function of both scale (or wavelength) and time. Semblance analysis is demonstrated on synthetic datasets and on gravity and aeromagnetic data from the Vredefort Dome, South Africa. Matlab source code is available from the IAMG server at www.iamg.org.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Fourier transform; Correlation; Time-series analysis

1. Introduction

Because of the inherent ambiguity in the interpretation of geophysical datasets, most projects use more than one data type, such as magnetics, gravity and EM. The interpretation process can then include looking for correlations (both positive and negative, depending on the target) between the different datasets. There are various methods of doing this, such as cross-correlation or cross-

spectral density. These methods often produce results that are hard to interpret or that do not give information about the relative phases of the datasets. Semblance filtering compares two datasets based on correlations between their phase angles, as a function of frequency. The Fourier transform $H(f)$ of a dataset $h(t)$ is given by (Blackledge 2003, p. 76)

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{-2\pi ift} dt, \quad (1)$$

where f is the frequency and t is the time. In general H is complex, and so has both an amplitude and a phase at each frequency. When the Fourier transforms of two datasets are calculated, the difference in their phase angles at each frequency can be computed simply (von Frese et al., 1997a;

[☆] Code available from server at <http://www.iamg.org/CGEditor/index.htm>.

*Corresponding author. Tel.: +27 11 717 6608;
fax: +27 11 717 6579.

E-mail addresses: cooperg@geosciences.wits.ac.za,
grcooper@iafrica.com (G.R.J. Cooper),
cowangeo@bigpond.net.au (D.R. Cowan).

Christensen, 2003):

$$S = \cos \theta(f) = \frac{R_1(f)R_2(f) + I_1(f)I_2(f)}{\sqrt{R_1^2(f) + I_1^2(f)} \times \sqrt{R_2^2(f) + I_2^2(f)}}, \quad (2)$$

where $R_1(f)$ and $I_1(f)$ are the real and imaginary components of the Fourier transform of dataset 1, expressed as a function of frequency f (R_2 and I_2 are defined similarly for dataset 2). The semblance S can take on values from -1 to $+1$. A value of $+1$ implies perfect phase correlation, 0 implies no correlation, and -1 implies perfect anticorrelation. Semblance filtering splits each input dataset into two output datasets consisting of the portion of the input datasets that is correlated to a given degree, and the portion that is not. This process is performed in the frequency domain. A threshold correlation value is chosen, and the Fourier transform of each dataset is split into two parts, one comprising the Fourier coefficients with a semblance above the threshold and the other comprising the remainder. The missing coefficients are replaced with zeros in each case. The inverse Fourier transform is then applied to each part, thus producing two output datasets for each input dataset. However, care must be taken or the sharp discontinuities in the filter shape will cause ringing in the output datasets (the Gibbs phenomenon). von Frese et al. (1997b) used the method to compare magnetic and gravity anomalies in Ohio, and Christensen (2003) used it to compare airborne magnetic and gravity datasets.

2. The continuous wavelet transform and semblance analysis

Wavelet-based approaches provide the ability to account for temporal (or spatial) variability in spectral character. Although wavelet analysis is relatively new compared with Fourier analysis, its use has become widespread in recent years, and so the theory will be described only briefly here. Mallat (1998) or Strang and Nguyen (1996) contain excellent and detailed summaries of wavelet analysis. The continuous wavelet transform (CWT) of a dataset $h(t)$ is given by (Mallat, 1998, p. 5)

$$\text{CWT}(u, s) = \int_{-\infty}^{\infty} h(t) \frac{1}{|s|^{0.5}} \Psi^* \left(\frac{t-u}{s} \right) dt, \quad (3)$$

where s is scale, u is displacement, Ψ is the mother wavelet used, and $*$ means complex conjugate. The CWT is therefore a convolution of the data with scaled version of the mother wavelet. Of course, the time coordinate t in Eq. (3) could equally well be the spatial coordinate x if profile data were being analysed. In this study, the complex Morlet wavelet was used, which is defined as (Teolis, 1998, p. 62)

$$\Psi(x) = \frac{1}{\pi f_b} e^{2\pi i f_c x} e^{-x^2/f_b}, \quad (4)$$

where f_b controls the wavelet bandwidth and f_c is the wavelet centre frequency. The complex Morlet wavelet is shown in Fig. 1. A value of 1.0 was used for f_c in order that scale became equivalent to wavelength. Unlike Fourier transform-based semblance analysis, the wavelet transform does not

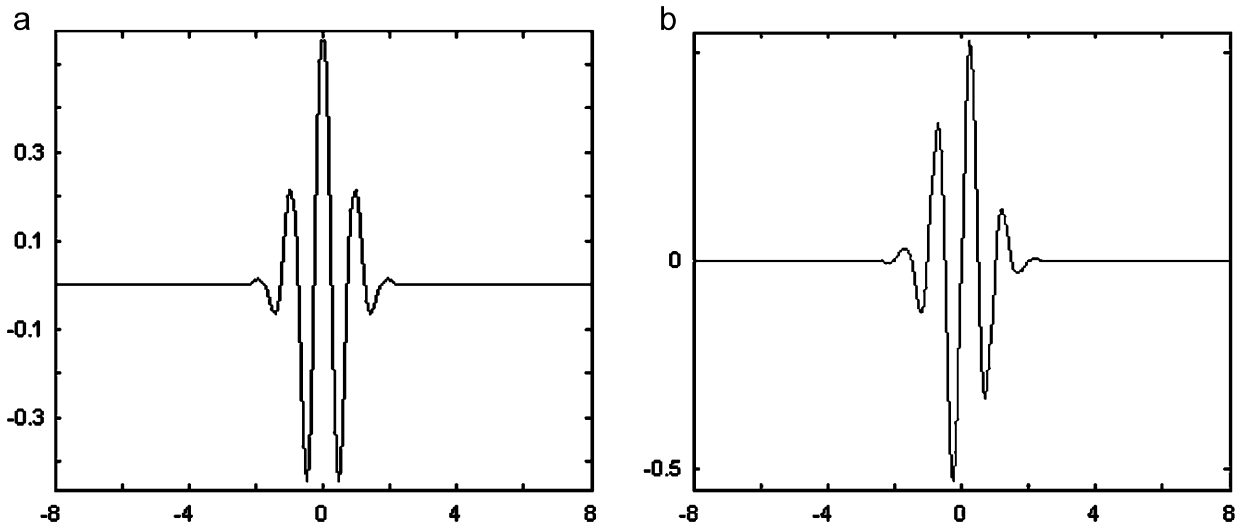


Fig. 1. (a) Real part of complex Morlet wavelet. (b) Imaginary part of complex Morlet wavelet.

assume that the frequency content of a dataset is constant with time (or position), and in fact allows changes in that behaviour to be analysed. Hence the wavelet-transform-based semblance analysis provides much better temporal (or spatial) resolution than Fourier-transform-based semblance methods. The use of different values of s (in Eq. (3)) yields information about the behaviour of the dataset on different scales. In addition, the mother wavelet on which the analysis is based can be chosen for its particular mathematical properties. When the mother wavelet is complex, its real and imaginary parts form a Hilbert transform pair, to ensure orthogonality. The use of a complex wavelet results in the CWT being complex, and therefore having a phase at each time (or position) and scale.

One method of comparing two time series using wavelets is the cross-wavelet transform (Torrence and Compo, 1998), defined as

$$\text{CWT}_{1,2} = \text{CWT}_1 \times \text{CWT}_2^*, \quad (5)$$

which is a complex quantity having an amplitude (the cross-wavelet power) given by

$$A = |\text{CWT}_{1,2}| \quad (6)$$

and local phase θ :

$$\theta = \tan^{-1}(\Im(\text{CWT}_{1,2})/\Re(\text{CWT}_{1,2})). \quad (7)$$

Using θ (which can range between $-\pi$ and $+\pi$) directly as a measure of phase correlation between two datasets was not found to give good results (see below), so the wavelet equivalent of Eq. (2) was devised, namely

$$S = \cos^n(\theta), \quad (8)$$

where n is an odd integer greater than zero. One advantage of the use of S compared with θ is that values now range from -1 (inversely correlated) through zero (uncorrelated) to $+1$ (correlated). In addition, the use of odd values of n greater than 1 enables the wavelet semblance response to be sharpened, as will be demonstrated. Because S and θ compare phase angles rather than amplitude information, they have the advantage that the two datasets being compared do not have to have the same units. However, a disadvantage of the lack of amplitude information is the sensitivity to noise, so we now further define

$$D = \cos^n(\theta)|\text{CWT}_1 \times \text{CWT}_2^*|, \quad (9)$$

which is equivalent to the vector dot product of the two complex wavelet vectors at each point in scale and position, when $n = 1$. D combines the phase

information of S with the amplitude information of A , which can be useful if the phase correlations of the larger amplitude components of the dataset are of the most interest.

3. Application to synthetic data

Unlike the standard Fourier-transform-based semblance analysis, which is calculated solely as a function of frequency, the CWT semblance analysis is calculated as a function of both scale (or wavelength) and time (or position). This enables the changing phase relationships of the two datasets to be visualised and analysed. Fig. 2 shows two synthetic datasets which contain sine waves of wavelengths 150 and 45 units. Both the datasets have three parts, and the phase relationship between any two parts in each dataset is constant for all wavelengths. The parts are indicated as sections A–B, B–C and C–D on Fig. 2, and their boundaries are visible as discontinuities where the phase changes are large (e.g. at B in Fig. 2c).

In the first dataset (Fig. 2a), both of the two sine wave components exist for the duration of the dataset, as can be seen from the real part of the complex CWT shown in Fig. 2b. In the second dataset (Fig. 2c and its CWT in Fig. 2d), the same two sine wave components are present, but their phases differ from those in the first dataset as a function of position. In portion A–B of the datasets, the longer wavelength component is anticorrelated (180° phase difference, $S = -1$) with the longer wavelength component in Fig. 2a, while the shorter wavelength component is perfectly correlated (0° phase difference, $S = 1$). Fig. 2e shows the local phase angle θ , Fig. 2f shows the semblance S (calculated with $n = 1$), and Fig. 2(g) shows the dot product D . The latter two figures show a broad blue patch running throughout portion A–B at a wavelength of approximately 150 units, indicating a negative correlation between the two datasets at that wavelength at those times. However, the phase angle θ pattern is not as clear in this regard as are the plots for S and D . S and D also show a red region at these times at a wavelength of 45 units, indicating a positive correlation between the two datasets at that wavelength. In the portion B–D of the plot, the phase relationship between the sine wave components of the two datasets was altered so that they are correlated, uncorrelated or inversely correlated as a function of position, and this correlation can be read directly from the colours of

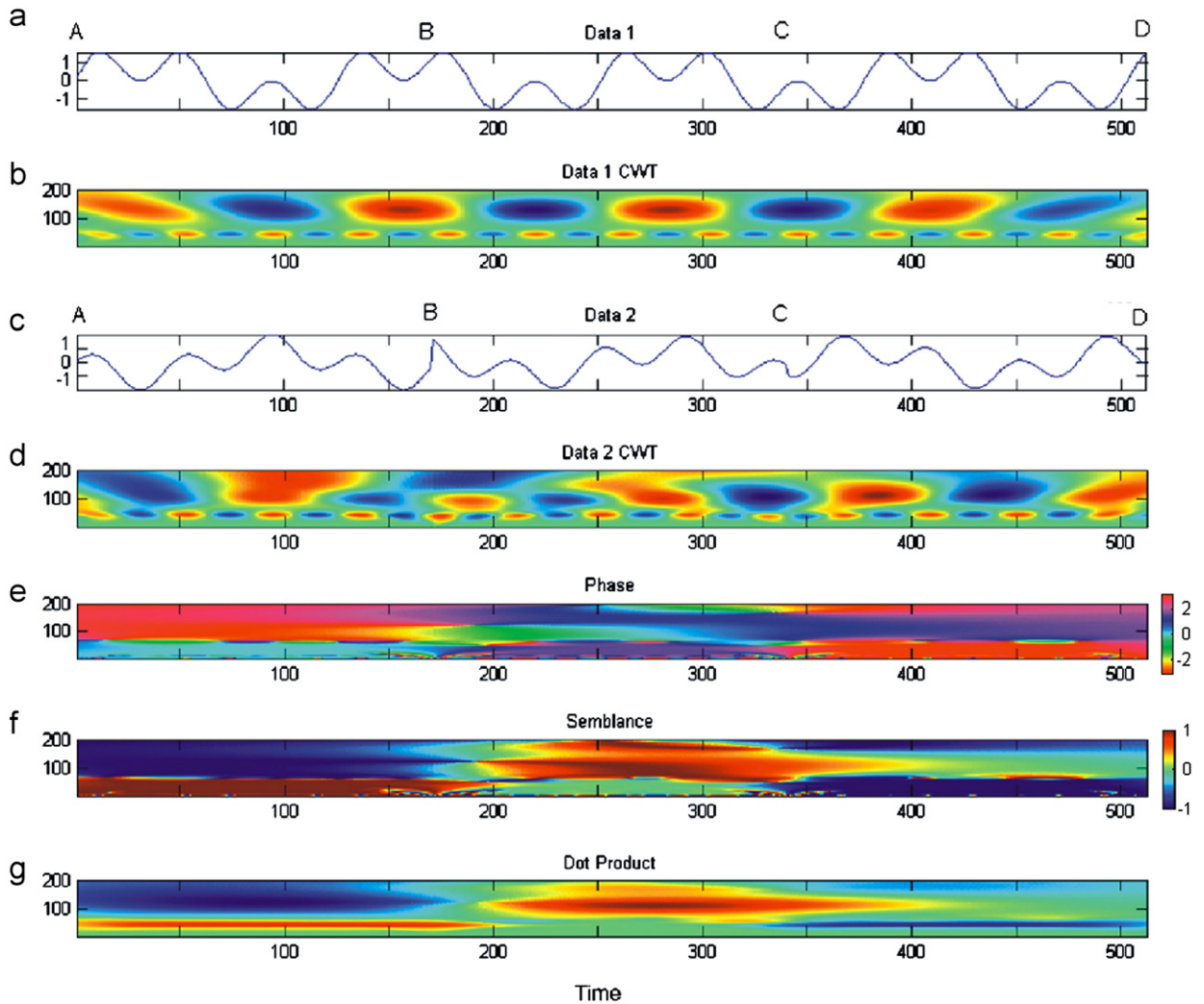


Fig. 2. (a) Dataset containing sine waves of wavelengths 110 and 35 units. (b) Real part of the complex CWT of dataset in (a). White (bright red in online version) indicates a large positive amplitude and black (dark blue in online version) indicates a large negative amplitude. (c) Dataset containing sine waves of wavelengths 110 and 35 units with different phases to those in (a). (d) Real part of complex CWT of dataset in (b). White (bright red in online version) indicates a large positive amplitude and black (dark blue in online version) indicates a large negative amplitude. (e) Local phase angle θ (calculated from Eq. (7)). White (bright red in online version) indicates a phase of $\pm 180^\circ$ and black (dark blue in online version) indicates a phase of 0° . (f) Semblance S (calculated from Eq. (8), with $n = 1$). White (bright red in online version) corresponds to a semblance of +1, 50% grey (green in online version) to a semblance of zero, and black (dark blue in online version) to a semblance of -1. (g) Dot product D (calculated from Eq. (9)). White (bright red in online version) corresponds to large-amplitude signals with positive semblance, while black (dark blue in online version) corresponds to large-amplitude signals with negative semblance. 50% grey (green in online version) corresponds either to low amplitude signals or those with a semblance close to zero.

Fig. 2f. When amplitude information is used simultaneously, the clarity of the plot improves (Fig. 2g); however, if the phase relationship of smaller amplitude components is of interest, then this plot would not be the best way of displaying them. The plot of θ (Fig. 2e) is not as clear as those of S or D , and the phase relationships are harder to see. The spatial resolving limits of the method are wavelength dependant, i.e.

visible spatial changes in semblance are sharper at short wavelengths than at long wavelengths. This is due to the nature of the wavelet transform.

Fig. 3 shows the effect of increasing the parameter n on the plots of S and D . As n is increased to 3 (Figs. 3a and b), and then to 9 (Figs. 3c and d) the smearing of the regions of strong positive and negative correlation is successively reduced. The use

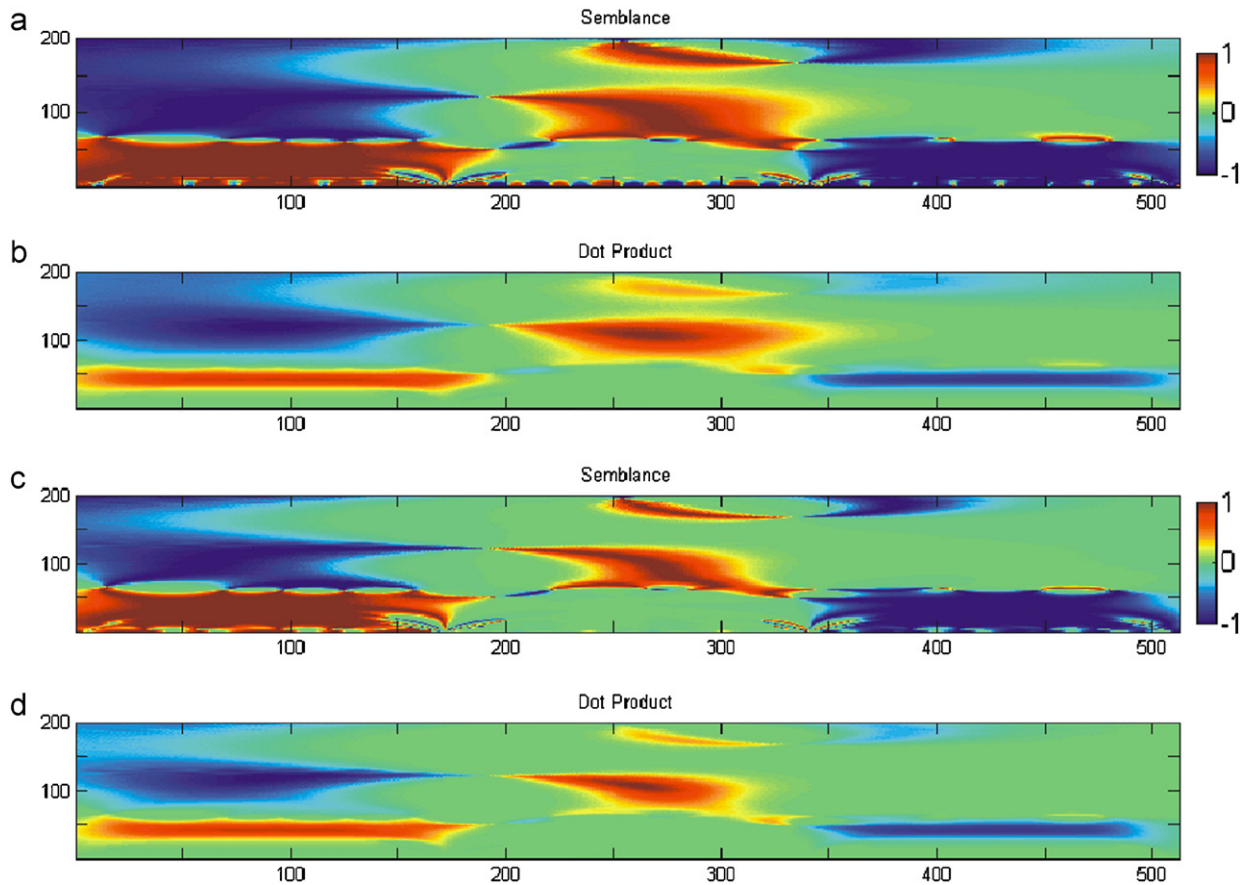


Fig. 3. (a) Semblance of datasets shown in Fig. 2 calculated with $n = 3$. (b) Dot product of datasets shown in Fig. 2 calculated with $n = 3$. (c) Semblance of datasets shown in Fig. 2 calculated with $n = 9$. (d) Dot product of datasets shown in Fig. 2 calculated with $n = 9$.

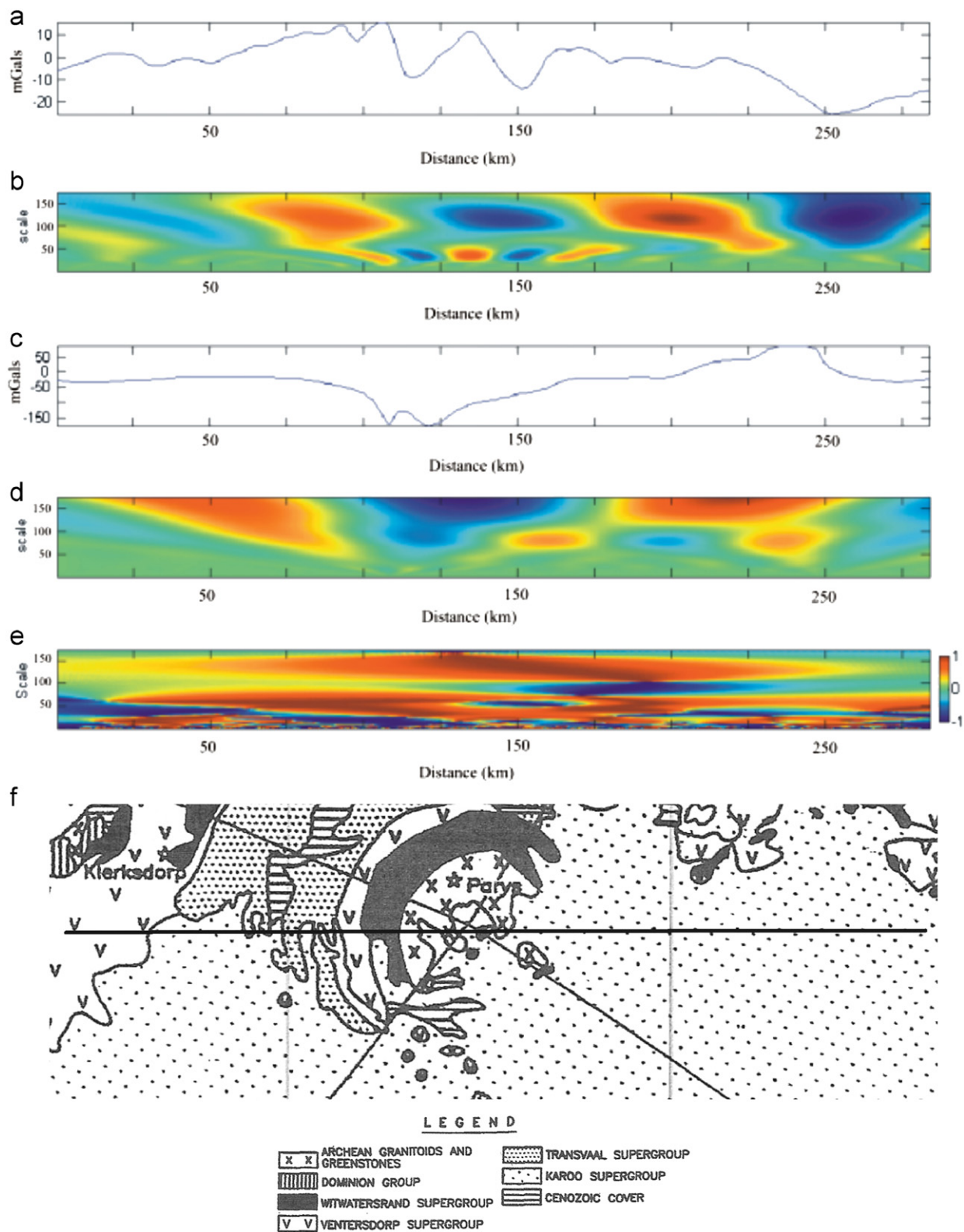
of very large values of n however tends to erode the extent of these regions excessively, particularly at longer wavelengths, making the choice of an optimum value dependant on the dataset.

4. Application to time and spatial datasets

Fig. 4 shows an analysis of EW gravity and pseudogravity profiles over the Vredefort dome impact structure and the Witwatersrand basin, South Africa. The Vredefort Dome, the 80 km wide central uplift of a very large 2 Ga meteorite impact, consists of a core of Archaean crystalline rocks, surrounded by an upturned collar of Archaean-Palaeoproterozoic supracrustals. The aeromagnetic data show a complex picture with clear evidence of remanent magnetisation. The gravity response is also complex. The Archaean core consists of a central gravity high due to granulites surrounded by an annular gravity low due to amphibolite facies

basement gneisses. The supracrustals show multi-ring gravity anomalies.

Pseudogravity converts the magnetic field into the gravity field that would be observed if the magnetisation distribution were to be replaced with an equivalent density distribution (Blakely, 1995, p. 344). It involves two steps; the magnetic data are first reduced to the pole, then their 1st vertical integral is computed. Because the pole reduction step produces incorrect results if remanent magnetisation is present, a correlation analysis between pseudogravity and gravity shows both where and on what scales remanent magnetisation exists, and also where gravity anomalies do not have associated magnetic anomalies. Visually, the CWT analyses (Figs. 4b and d) appear to correlate well at wavelengths of 125–150 km, and this is shown clearly in the broad red patch (indicating strong positive correlation) in the semblance plot (Fig. 4e). On smaller scales, the relationship between the two datasets is more difficult to see, and the correlation



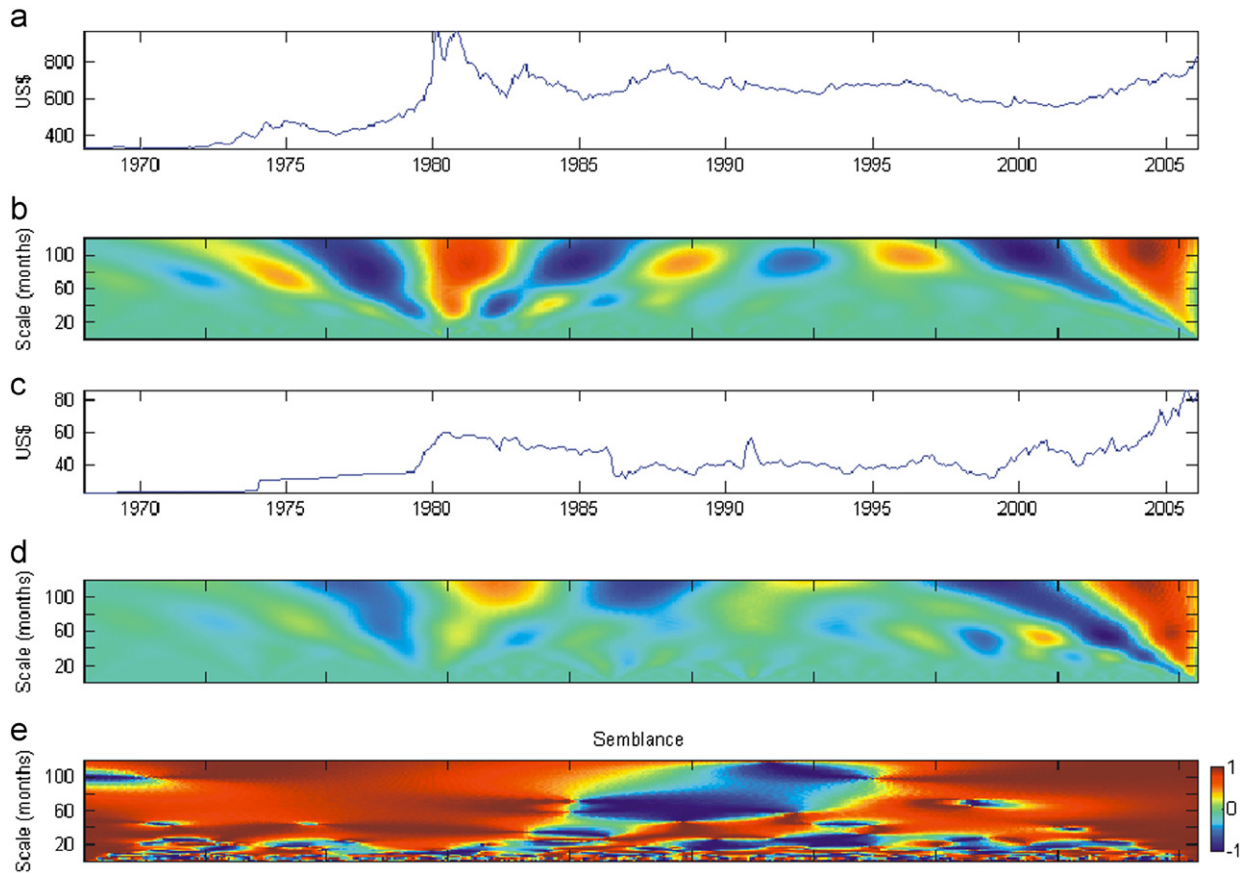


Fig. 5. (a) Price per ounce for Gold since 1968, measured monthly. (b) Real part of complex CWT of the dataset in (a). White (bright red in online version) indicates a large positive amplitude and black (dark blue in online version) indicates a large negative amplitude. (c) Price per barrel for oil since 1968, measured monthly. (d) Real part of complex CWT of the dataset in (c). White (bright red in online version) indicates a large positive amplitude and black (dark blue in online version) indicates a large negative amplitude. (e) Semblance S (calculated from Eq. (8), with $n = 1$). White (bright red in online version) corresponds to a semblance of $+1$, 50% grey (green in online version) to a semblance of zero, and black (dark blue in online version) to a semblance of -1 .

varies depending on the magnetisation of the geological units on the profile.

As an example of the generality of the method, the gold price and the oil price since 1968 are compared in Fig. 5. The data are sampled monthly and were obtained from www.forecasts.org. As can be seen, the datasets are strongly correlated on all scales and at all times, except for the period 1985–1995, when they appear anticorrelated on scales of 5–10 years.

5. Conclusions

Wavelets can be used to perform semblance analysis of time and spatial data series to display their correlations as a function of both scale (wavelength) and time (or position). The calculations involved are not computer intensive, and the visual nature of the results aids in their interpretation. The method has application to a wide range of problems both inside and outside the geosciences.

Fig. 4. (a) Gravity profile EW across the Vredefort dome structure, South Africa. The mean of the data has been removed. (b) Real part of the complex CWT of the dataset in (a). White (bright red in online version) indicates a large positive amplitude and black (dark blue in online version) indicates a large negative amplitude. (c) Pseudogravity profile EW across the Vredefort dome structure, South Africa. The mean of the data has been removed. (d) Real part of the complex CWT of the dataset in (c). White (bright red in online version) indicates a large positive amplitude and black (dark blue in online version) indicates a large negative amplitude. (e) Semblance S (calculated from Eq. (8), with $n = 1$). White (bright red in online version) corresponds to a semblance of $+1$, 50% grey (green in online version) to a semblance of zero, and black (dark blue in online version) to a semblance of -1 . (f) Geology of Vredefort dome area (Henkel and Reimold, 1998), showing location of data profiles used in (a)–(e).

Acknowledgements

Two reviewers (Ralph von Frese and Mitchell Craig) are thanked for their constructive comments which improved the manuscript. The Council for Geoscience, Pretoria, is thanked for permission to use the potential field datasets shown in Fig. 4.

Appendix A. Supplementary materials

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.cageo.2007.03.009](https://doi.org/10.1016/j.cageo.2007.03.009).

References

- Blackledge, J.M., 2003. *Digital Signal Processing*. Horwood Publishing, Chichester, 757pp.
- Blakely, R.J., 1995. *Potential Theory in Gravity and Magnetic Applications*. Cambridge University Press, New York, 435pp.
- Christensen, A.N., 2003. Semblance filtering of airborne potential field data. In: *Extended Abstracts, 16th ASEG Conference and Exhibition*, Adelaide.
- Henkel, H., Reimold, W.U., 1998. Integrated geophysical modelling of a giant, complex impact structure: anatomy of the Vredefort structure, South Africa. *Tectonophysics* 28, 1–20.
- Mallat, S., 1998. *A Wavelet Tour of Signal Processing*. Academic Press, New York, 577pp.
- Strang, G., Nguyen, T., 1996. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, MA, 490pp.
- Teolis, A., 1998. *Computational Signal Processing with Wavelets*. Birkhäuser Boston Inc., Boston, MA, 352pp.
- Torrence, C., Compo, G.P., 1998. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society* 79, 61–78.
- von Frese, R.R.B., Jones, M.B., Kim, J.W., Kim, J.H., 1997a. Analysis of anomaly correlations. *Geophysics* 62 (1), 342–351.
- von Frese, R.R.B., Jones, M.B., Kim, J.W., Li, W.S., 1997b. Spectral correlation of magnetic and gravity anomalies of Ohio. *Geophysics* 62 (1), 365–380.