Abstract—Datasets are usually compared using cross-correlation, often using a moving window to calculate the correlation as a function of time or space. However, signals can be considered as being composed of sinusoids which possess amplitudes, frequencies and phases. All these three attributes can be computed at each point in time, and then used as the basis of a cross-correlation method. By using all the three measures of correlation together, their individual disadvantages can be minimised. By combining these measures with the continuous wavelet transform, information on the correlation as a function of wavelength can be obtained.

1. Introduction

Pearson’s cross-correlation coefficient \( r \) between two datasets \( x \) and \( y \) is given by (MCGREW and MONROE 1993, p. 249)

\[
r = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{N\sigma_x\sigma_y},
\]

(1)

where the datasets each have \( N \) datapoints, \( \bar{x} \) and \( \bar{y} \) are their means, and \( \sigma_x \) and \( \sigma_y \) are their standard deviations.

\( r \) can take on values from +1 to −1 and is a measure of the similarity of the two datasets. In many applications, it is useful to know how \( r \) changes as a function of time or distance, in which case \( r \) can be computed using a sliding window. Unfortunately, the results will depend on the window size that is used.

Datasets can be considered as being composed of sinusoidal functions which are defined by their amplitude \( a \), frequency \( \omega \) and phase \( \varphi \), i.e.

\[
y = a \cdot \sin(\omega \cdot t + \varphi).
\]

(2)

Fourier transform-based semblance analysis compares two datasets based on correlations between their phase angles, as a function of frequency. Semblance \( S \) is defined as the difference between the phase angles of the two datasets at each frequency (VON FRESE et al. 1997; CHRISTENSEN 2003)

\[
S(\omega) = \frac{R_1(\omega)R_2(\omega) + I_1(\omega)I_2(\omega)}{\sqrt{R_1^2(\omega) + I_1^2(\omega)}}
\]

(3)

where \( R_1(\omega) \), and \( I_1(\omega) \) are the real and imaginary components of the Fourier transform of dataset 1, as a function of frequency \( \omega \) (and \( R_2 \) and \( I_2 \) are defined similarly for dataset 2). COOPER and COWAN (2008) produced a multiscale version of Eq. (3) using the continuous wavelet transform.

This paper introduces methods of comparing time series that are based on the amplitude and frequency of the signal, not just its phase.

2. Analysis of Data Using Attributes

Seismic reflection data is used, amongst other applications, to search for oil worldwide. The raw seismic data is extensively processed and enhanced prior to its interpretation. Attributes of the seismic data, such as its coherence (GERSZTENKORN et al. 1999) or curvature (BLUMENRITT et al. 2006) are computed to enhance fine detail within the data. The first attributes to be commonly used were the instantaneous amplitude \( A \), phase \( \theta \), and frequency \( \omega \) of the data (TANER et al. 1979), which are defined as;

\[
A(t) = \sqrt{(y(t))^2 + H(y(t))^2},
\]

(4)

\[
\theta = \tan^{-1}(H(y(t))/y(t)),
\]

(5)

and
where $H(y(t))$ is the Hilbert transform of the data $y(t)$. In practice, $\omega$ is computed directly from the data without the intermediate calculation of $\theta$, using (TANER et al. 1979)

$$\omega(t) = \frac{y(t) \frac{dH(y(t))}{dt} - H(y(t)) \frac{dy(t)}{dt}}{y(t)^2 + H(y(t))^2}. \quad (7)$$

When calculated in this way, $\omega$ can take on positive or negative values, including values greater than the Nyquist limit. Figures 1, 2 and 3 show examples of the instantaneous attributes of synthetic signals. Note the oscillatory behaviour of the instantaneous phase in Fig. 2b, d; if the instantaneous frequency is to be computed using Eq. (6) then the phase must first be unwrapped (see Fig. 11) to avoid introducing discontinuities whenever the phase flips from $+\pi$ to $-\pi$ radians. All three of these attributes can be used to compare two datasets $y_1(t)$ and $y_2(t)$ in different ways:

**Amplitude cross-correlation**

$$Sa(t) = 1 - \frac{|A_1(t) - A_2(t)|}{A_1(t) + A_2(t)}. \quad (8)$$

**Phase cross-correlation**

$$Sp(t) = \cos(\theta_1(t) - \theta_2(t)). \quad (9)$$

**Frequency cross-correlation**

$$So(t) = 1 - \frac{[\omega_1(t) - \omega_2(t)]}{\omega_1(t) + \omega_2(t)}. \quad (10)$$

$Sp$ can have values from $-1$ (anticorrelated) to $+1$ (positively correlated), while $Sa$ and $So$ have values from 0 (uncorrelated) to $+1$ (positively correlated). These measures of phase and frequency comparison are unaffected by the units of the two datasets.

### 3. Application to Synthetic Data

Figures 1, 2 and 3 show two synthetic datasets consisting of four portions, marked AB, BC, CD, and DE. In section AB the datasets are identical, so the amplitude, phase, and frequency correlations $Sa$, $Sp$, $So$ are identical. In section BC, the amplitude of the signal has been increased by a factor of four compared to the dataset shown in a. In section CD, the phase has been altered by $\pi/2$ radians compared to the dataset shown in a. In section DE, the frequency of the signal has been increased to 20 Hz. Overlaid using a dotted line is the instantaneous amplitude computed using Eq. (4).
and $S_{\omega}$ are all very near to +1. In section BC the amplitude of the second dataset has been increased, resulting in a decreased amplitude correlation, and unchanged phase and frequency correlations. In section CD the amplitudes and frequencies of the two datasets are equal but the phase of the second dataset has been changed by $\pi/2$ radians resulting in a phase correlation $S_{\phi}$ of near zero. In section DE the frequency of the second dataset has been increased, resulting in a decreased frequency correlation $S_{\omega}$. Because frequency is the time derivative of phase, the phase correlation is also affected by this frequency change.

Figures 4, 5 and 6 repeat the above analysis, only with uniformly distributed random noise with an amplitude equal to 10% of the data amplitude having been added to each dataset. All of the measures of correlation are affected, but the methods do not appear to be unduly sensitive to noise.

4. The Continuous Wavelet Transform and Attribute Analysis

The continuous wavelet transform (CWT) can be used to display the frequency content of a dataset as a function of time, allowing changes in its behaviour to be studied. The CWT of a dataset $h(t)$ is given by (MALLAT 1998, p. 5)
\[
\text{CWT}(u, s) = \int_{-\infty}^{\infty} h(t) \frac{1}{|s|^{0.5}} \Psi^*(\frac{t-u}{s}) \, dt, \quad (11)
\]

where \(s\) is the scale, \(u\) is the displacement, \(\Psi\) is the mother wavelet, and \(^*\) means the complex conjugate. Different mother wavelets can be used, depending on the type of analysis being performed. The complex Morlet wavelet was used in this case because it is a damped sinusoidal function, and Eq. (2) considers the data as being based on sinusoids. It is defined as

\[
\Psi(t) = \frac{1}{f_0} e^{2\pi i f t} e^{-\pi^2 t^2}, \quad (12)
\]

where \(f_0\) controls the wavelet bandwidth and \(f_c\) is the wavelet centre frequency. Because it is a complex wavelet the CWT of the data is also complex. Figure 7 shows the same two datasets that were used in Figs. 1, 2, 3, 4, 5 and 6, together with the amplitude of their continuous wavelet transforms (i.e. \(A_1 = |\text{CWT}_1|\), \(A_2 = |\text{CWT}_2|\)). In Fig. 7b, the signal appears as coefficients with relatively large values at a wavelength of 0.2 s (5 Hz). In Fig. 7d, the signal appears as coefficients with relatively large values at a wavelength of 0.2 s (5 Hz) in the region AD, after which the wavelength of the signal decreases to 0.05 s (20 Hz).

The amplitude cross-correlation of the CWT coefficients is then defined as

\[
Sa(t, \lambda) = 1 - \left| \frac{A_1(t, \lambda) - A_2(t, \lambda)}{A_1(t, \lambda) + A_2(t, \lambda)} \right|, \quad (13)
\]
a Synthetic dataset consisting of a 5 Hz signal sampled at 100 Hz. The amplitude and frequency are constant throughout. Uniformly distributed random noise with an amplitude equal to 10% of the data amplitude has been added to the dataset. Overlaid using a dotted line is the instantaneous amplitude computed using Eq. (4).
b Synthetic dataset consisting of four regions AB, BC, CD, and DE. Section AB is identical to that of the dataset shown in a. In section BC the amplitude of the signal has been increased by a factor of four compared to the dataset shown in a. In section CD the phase has been altered by \( \pi/2 \) radians compared to the dataset shown in a. In section DE the frequency of the signal has been increased to 20 Hz. Uniformly distributed random noise with an amplitude equal to 10% of the data amplitude has been added to the dataset. Overlaid using a dotted line is the instantaneous amplitude computed using Eq. (4).
c Amplitude correlation factor \( S_a \) computed from Eq. (8).

d Instantaneous phase of the dataset shown in a, computed using Eq. (5).
e Synthetic dataset consisting of four regions AB, BC, CD, and DE. Section AB is identical to that of the dataset shown in a. In section BC the amplitude of the signal has been increased by a factor of four compared to the dataset shown in a. In section CD the phase has been altered by \( \pi/2 \) radians compared to the dataset shown in a. In section DE the frequency of the signal has been increased to 20 Hz. Uniformly distributed random noise with an amplitude equal to 10% of the data amplitude has been added to the dataset. d Instantaneous phase of the dataset shown in c, computed using Eq. (5).
f Phase-correlation factor \( S_p \) computed from Eq. (9).
where $\lambda$ is wavelength. Figure 7 shows the amplitude cross-correlation values, and it can be seen that the correlation is close to one in portions AB and CD, and less in portion BC, at the wavelength of 0.2 s. An alternative method of comparing two time series using the wavelet transform is the amplitude of the cross-wavelet transform (TRENTORE and COMPO 1998)

$$P = |\text{CWT}_{12}|,$$  \hspace{1cm} (14)

where

$$\text{CWT}_{12} = \frac{\text{CWT}_1 \cdot \text{CWT}_2}{|\text{CWT}_1|^{\frac{1}{2}} |\text{CWT}_2|^{\frac{1}{2}}}. \hspace{1cm} (15)$$

For comparison, this is plotted in Fig. 7f. It has a large value when the individual wavelet coefficients are large, whereas Eq. (13) takes on large values when the wavelet coefficients are similar.

The multiscale equivalent of the phase correlation $Sp$ is the semblance $S$ (COOPER and COWAN 2008)

$$S = \cos(\theta), \hspace{1cm} (16)$$

where

$$\theta = \tan^{-1} \left( \Im(\text{CWT}_{12})/\Re(\text{CWT}_{12}) \right). \hspace{1cm} (17)$$

Figure 8 displays the semblance of the synthetic datasets shown in the previous figures for comparison with $Sp$ (Fig. 2). Both $S$ and $Sp$ show that the signals have a high degree of phase correlation in the region between points A and C, a near-zero phase correlation between C and D, and an oscillatory phase relationship between D and E. $S$ has the advantage of showing the wavelengths at which the correlations occur, albeit at a much greater computational cost. $Sp$
can be considered as showing the overall phase relationship at a particular point.

The multiscale frequency correlation is defined as

$$S_{\phi}(t, \lambda) = 1 - \frac{\omega_1(t, \lambda) - \omega_2(t, \lambda)}{\omega_1(t, \lambda) + \omega_2(t, \lambda)}, \quad (18)$$

where $\omega$ is the time derivative of the phase angle $\theta$.

Figure 9 displays $\omega(t, \lambda)$ for each synthetic time series. $S_{\phi}$ shows a strong correlation between the frequency contents of the signals between points A and D around a wavelength of 0.2 s, and a much reduced correlation from points D–E where the frequency contents of the signals differ.

5. Application to Real Time Series Datasets

South Africa can be divided into eleven different regions on the basis of climate. The main climate divides are mountain ranges, with rivers being of
secondary importance (van Zyl 2003, p. 44). Figure 10 shows the monthly rainfall from 1960 to 2002 in Johannesburg and Cape Town. Johannesburg is on the highveld, 1500 m above sea level, and receives very little rain in the winter months (May–August), unlike Cape Town, where it rains all year round, but particularly in winter. The climate in Cape Town is dominated by its proximity to the ocean. Figure 10c shows the amplitude cross correlation $S_a$ (Eq. 8). The values are generally high, between 0.6 and 1.0, reflecting the similar annual average rainfalls of the two regions (713 mm for Johannesburg and 515 mm for Cape Town). The correlation drops considerably during the winter months when there is little rainfall in Johannesburg. Figure 11 shows the phase correlation $S_p$ (Eq. 9) which is mostly negative due to the aforementioned difference in the timing of the rains in the two cities. Figure 12 shows the frequency correlation $S_f$ (Eq. 12), which is mostly close to one because the rates of change of the rainfall in both regions is similar, both regions being dominated by the annual cycle.

Figure 13 compares the two rainfall datasets in a multiscale manner using the amplitudes of their wavelet transforms (which are shown in Fig. 13b, d). The latter are dominated by the annual component, which appears clearly at a wavelength of 12 months in both wavelet transforms. Figure 13e shows the multiscale amplitude correlation $S_A$ (Eq. 13). This displays a strong correlation at an annual wavelength, but on
smaller wavelengths, the correlation is poor, due both to noise and to the small winter rainfall in Johannesburg. The amplitude of the cross-wavelet transform is shown in Fig. 13f. It only has large values at a wavelength of 12 months, when the amplitudes of both of the wavelet transforms have large values.

Wavelet-based semblance analysis (Cooper and Cowan 2008) is applied to the rainfall data in Fig. 14 and clearly shows the strong anticorrelation of the two datasets on an annual scale (Fig. 14e). Finally, Fig. 15 shows the correlation between the instantaneous frequency content of the datasets, which has a maxima at a wavelength of 12 months.

To summarise therefore, a comparative analysis of the rainfall patterns of the Cape and the Highveld regions, South Africa, shows them to be positively correlated in terms of their amplitude and their rates of change, but negatively correlated in phase. Wavelet-transform-based correlation methods showed that the dominant component of the signals has a 12-month wavelength.

As a second example, the rainfall and the temperature datasets from Cape Town (1960–2002) are compared. Because they have different units, only the phase and frequency cross-correlation methods were used. As Cape Town receives most of its rain in the winter, the two datasets are negatively correlated on average (correlation coefficient of \(-0.65\)). Figure 16e shows how the phase correlation of the two datasets varies as a function of time, and it is negative most of

Figure 9

| a | Synthetic dataset consisting of a 5 Hz signal sampled at 100 Hz. The amplitude and frequency are constant throughout. | b | Instantaneous frequency of the dataset shown in a, i.e. the derivative as a function of time of the unwrapped instantaneous phase shown in Fig. 8b. | c | Synthetic dataset consisting of four regions AB, BC, CD, and DE. Section AB is identical to that of the dataset shown in a. In section BC the amplitude of the signal has been increased by a factor of four compared to the dataset shown in a. In section CD the phase has been altered by \(\pi/2\) radians compared to the dataset shown in a. In section DE the frequency of the signal has been increased to 20 Hz. d | Instantaneous frequency of the dataset shown in c, i.e. the derivative as a function of time of the unwrapped instantaneous phase shown in Fig. 8d. | e | Frequency correlation factor \(S_{x_0}\) computed from Eq. (18). |
Figure 10

a Monthly rainfall in Johannesburg, South Africa, from 1960 to 2002. Overlain using a dotted line is the instantaneous amplitude computed using Eq. (4).

b Monthly rainfall in Cape Town, South Africa, from 1960 to 2002. Overlain using a dotted line is the instantaneous amplitude computed using Eq. (4).

c Amplitude correlation factor $S_a$ computed from Eq. (8)
Figure 11

(b) Unwrapped instantaneous phase of the data from (a) computed using Eq. (5).
(c) Monthly rainfall in Cape Town, South Africa, from 1960 to 2002. 
(d) Unwrapped instantaneous phase of the data from (c) computed using Eq. (5).
(e) Phase-correlation factor $S_p$ computed from Eq. (9)
Figure 12

(b) Instantaneous frequency of the data from (a) computed using Eq. (7). 
(c) Monthly rainfall in Cape Town, South Africa, from 1960 to 2002. 
(d) Instantaneous frequency of the data from (c) computed using Eq. (7). 
(e) Frequency correlation factor $S_{xy}$ computed from Eq. (10).
Figure 13

a Monthly rainfall in Johannesburg, South Africa, from 1960 to 2002.
b Amplitude of the CWT of the dataset shown in a, computed using the complex Morlet wavelet.
c Monthly rainfall in Cape Town, South Africa, from 1960 to 2002.
d Amplitude of the CWT of the dataset shown in c, computed using the complex Morlet wavelet.
e Amplitude correlation factor \( S_u \) computed from Eq. (13).
f Amplitude of the cross-wavelet transform computed from Eq. (14).
Figure 14

(a) Monthly rainfall in Johannesburg, South Africa, from 1960 to 2002. (b) Instantaneous phase of the dataset shown in (a), i.e. \( \theta = \tan^{-1}(\text{Im}(CWT)/\text{Re}(CWT)) \). (c) Monthly rainfall in Cape Town, South Africa, from 1960 to 2002. (d) Instantaneous phase of the dataset shown in (c), i.e. \( \theta = \tan^{-1}(\text{Im}(CWT)/\text{Re}(CWT)) \). (e) Semblance \( S \) computed from Eq. (16)
Figure 15

a Monthly rainfall in Johannesburg, South Africa, from 1960 to 2002. b Instantaneous frequency of the dataset shown in a, i.e. the derivative as a function of time of the unwrapped instantaneous phase of Fig. 14b. c Monthly rainfall in Cape Town, South Africa, from 1960 to 2002. d Instantaneous frequency of the dataset shown in c, i.e. the derivative as a function of time of the unwrapped instantaneous phase shown in Fig. 14d. e Frequency correlation factor $S_{xy}$ computed from Eq. (18)
Figure 16

(a) Monthly average temperatures in Cape Town, South Africa, from 1960 to 2002. 
(b) Unwrapped instantaneous phase of the data from (a) computed using Eq. (5).
(c) Monthly rainfall in Cape Town, South Africa, from 1960 to 2002. 
(d) Unwrapped instantaneous phase of the data from (c) computed using Eq. (5). 
(e) Phase-correlation factor $S_p$ computed from Eq. (9)
Figure 17

(a) Monthly average temperatures in Cape Town, South Africa, from 1960 to 2002. (b) Instantaneous frequency of the data from (a) computed using Eq. (7). (c) Monthly rainfall in Cape Town, South Africa, from 1960 to 2002. (d) Instantaneous frequency of the data from (c) computed using Eq. (7). (e) Frequency correlation factor $S_0$ computed from Eq. (10)
Figure 18

a Monthly average temperatures in Cape Town, South Africa, from 1960 to 2002. b Instantaneous phase of the dataset shown in a, i.e. $\theta = \tan^{-1}(\Im(CWT)/R(CWT))$. c Monthly rainfall in Cape Town, South Africa, from 1960 to 2002. d Instantaneous phase of the dataset shown in c, i.e. $\theta = \tan^{-1}(\Im(CWT)/R(CWT))$. e Semblance $S$ computed from Eq. (16)
Figure 19

a Monthly average temperatures in Cape Town, South Africa, from 1960 to 2002. 
b Instantaneous frequency of the dataset shown in a, i.e. the derivative as a function of time of the unwrapped instantaneous phase of Fig. 18b. 
c Monthly rainfall in Cape Town, South Africa, from 1960 to 2002. 
d Instantaneous frequency of the dataset shown in c, i.e. the derivative as a function of time of the unwrapped instantaneous phase shown in Fig. 18d. 
e Frequency correlation factors $\delta_x$ computed from Eq. (18)
the time. Figure 17 shows the frequency correlation of the two datasets. Conversely this is mostly close to one, because both are driven by the same annual cycle. This is made clear by the wavelet-based semblance analysis (Fig. 18) which shows a strong negative correlation between the temperature and rainfall in Cape Town on an annual scale. The instantaneous frequency content of the two datasets are strongly correlated on almost all scales from 1 to 24 months, but particularly on an annual scale.

6. Conclusions

Many measures of cross-correlation between different aspects of data can be computed. Cross-correlations based on the instantaneous amplitude, phase and frequency of the signals have been introduced here which can help to understand the behaviour of the data. Multiscale wavelet-based implementations of these measures have the advantage of showing the wavelength at which the correlations occur, but require considerably more computational effort to calculate.

Acknowledgments

The South African weather bureau is thanked for the data shown in Figs. 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19. The NRF (Pretoria) is thanked for funding this project.

REFERENCES


CHRISTENSEN, A. N. (2003), Semblance filtering of airborne potential field data. Extended abstracts, 16th ASEG conference and exhibition, Adelaide.


TANER, M.T., KOEHLER, F., and SHERIFF, R.E. (1979), Complex seismic trace analysis, Geophysics 44, 1041–1063.

TEOLIS, A. (1998), Computational signal processing with wavelets, Birkhäuser.


(Received April 13, 2015, revised August 15, 2015, accepted September 30, 2015)