Article:

Pering, T.D., Tamburello, G., McGonigle, A.S. et al. (2 more authors) (2014) Correlation of oscillatory behaviour in Matlab using wavelets. Computers & Geosciences, 70. 206 - 212. ISSN 0098-3004

https://doi.org/10.1016/j.cageo.2014.06.006

Correlation of Oscillatory Behaviour in Matlab® using Wavelets

Pering T.D.[a]*, Tamburello G.[b], McGonigle A.J.S.[a,c], Hanna, E.[a], Aiuppa A.[b,c],

DOI: 10.1016/j.cageo.2014.06.006

* Corresponding author, ggp12tdp@sheffield.ac.uk, Department of Geography, Winter Street University of Sheffield, Sheffield, S10 2TN, United Kingdom, +447838219369

[a]University of Sheffield, Dept. of Geography, Winter Street, S10 2TN, United Kingdom

[b]DiSTeM, Università di Palermo, via Archirafi, 22, 90123 Palermo, Italy

[c]Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Palermo, Via Ugo La Malfa, 153, 90146, Palermo, Italy

Abstract - Here we present a novel computational signal processing approach for comparing two signals of equal length and sampling rate, suitable for application across widely varying areas within the geosciences. By performing a continuous wavelet transform (CWT) followed by Spearman's rank correlation coefficient analysis, a graphical depiction of links between periodicities present in the two signals is generated via two or three dimensional images. In comparison with alternate approaches, e.g., wavelet coherence, this technique is simpler to implement and provides far clearer visual identification of the inter-series relationships. In particular, we report on a Matlab® code which executes this technique, and examples are given which demonstrate the program application with artificially generated signals of known periodicity characteristics as well as with acquired geochemical and meteorological datasets.

Continuous Wavelet Transform; Wavelets; Spearman's Rank Correlation; Periodicity; Oscillation; De-noising

1. Introduction

Given the significant increase in computational power over the last decades, signal processing techniques such as wavelet analysis have become commonplace in their application within the geosciences. In particular, wavelets are applied, via a process of convolution, to reveal information on periodicities present in data series, and their stability as a function of time, in contrast to Fourier transforms, which only probe frequency characteristics (Welch, 1967; Harris, 1978). The exception here is with the Short Fourier Transform (e.g., spectrogram), which is applied to reveal spectral frequency variations with time (Oppenheim et al. 1999). Whereas, a continuous wavelet transform (CWT) operates over a continuous range of scales, providing potentially more detailed information than the discretely sampled discrete wavelet or Short Fourier Transform (Torrence and Compo, 1998; Oppenheim et al. 1999). Hence, wavelets are more suited to investigation of transient or unstable periodic phenomena.

Oscillatory behavior is widely manifest in datasets acquired from across the geo and environmental sciences, for example concerning the 11-year sunspot cycle (e.g. Hoyt and Schatten, 1997; Frohlich and Lean, 2004), the El Niño Southern Oscillation (Torrence and Compo, 1998) and the North Atlantic Oscillation (NAO) (Hurrell, 1995). These phenomena can change significantly in strength and period as a function of time and are an integral part of climate variability (e.g. Hurrell et al. 2003; Lockwood 2012; Philander 1990). Oscillations are also present over much shorter timescales of seconds to hours, for example within geochemical datasets concerning volcanic degassing (Tamburello et al. 2012). The links between fluctuations present in environmental data series can wax and wane dramatically, providing a motivation for the application of wavelet analysis. Here we present a straightforward and new approach to investigating the correlation between oscillations present in two or more environmental datasets; this technique is based on CWT analysis using Matlab® and the Matlab Wavelet Toolbox® followed by Spearman's rank correlation coefficient analysis.

2. Technique Overview

The Matlab® function (available in the auxiliary materials) was written in Matlab® 2010b and has been tested on the 2008a, 2011b and 2013a versions, with correct operation demonstrated in each case. The program uses the CWT function (part of the Matlab Wavelet Toolbox®) for two separate signals. These signals should be normalised prior to processing by this code, performance is independent of normalisation technique as long as signal amplitude is preserved, the code normalises through division by the maximum value. This is followed by linear correlation (using Spearman's rank correlation coefficient, which accounts for non-linearity and variable amplitude of the wavelet coefficients), to generate a visual representation of the links between the coefficients generated by the wavelet transforms (e.g. Fig. 1b, 3d, 4, 5a, 5b). For the examples illustrated in this paper the Morlet wavelet was applied as the mother wavelet (Morlet et al. 1982; Grinstead et al. 2004):

$$\Psi_0(\eta) = \pi^{-1/4} e^{i\omega_0\eta} e^{-\eta^2/2}.$$

where $\Psi_0(\eta)$ is the wavelet function, η is a non-dimensional parameter representing a time component and ω_0 refers to the wavelets' non-dimensional frequency. This particular class of wavelet is implemented here, given its similarity to naturally occurring oscillations manifest in data series spanning the geosciences (e.g. Torrence and Compo, 1998). This said, the code could also use non-complex alternates, e.g., Gaussian wavelets from the Matlab Wavelet Toolbox® if these are judged more suitable for the application in question. Indeed, the Matlab Wavelet Toolbox® provides a comprehensive overview and visualisation of available mother wavelets. In general, wavelet analysis works best with selection of a mother wavelet which closely resembles the target oscillation. The CWT itself is defined as (e.g. Grinstead et al. 2004):

$$W_n(s) = \sqrt{\frac{\delta t}{s}} \sum_{n'=1}^{N} x_{n'} \Psi^* \left[(n' - n) \frac{\delta t}{s} \right],$$

where δt is a uniform time-step, x_n is the subject signal, $W_n(s)$ represents the changing wavelet scale on the left-hand-side and similarly as s on the right-hand-side, * is the complex conjugate, N the maximum scale, and n the points of the time series, (Morlet et al. 1982; Colestock, 1993; Grinstead et al. 2004). The result is the conjugation of the scaled selected wavelet with the subject signal and outputs, which demonstrates the stability and power of any periodic features which match the scaled wavelet. We refer to the extensive literature for more in-depth descriptions of the CWT (e.g. Morlet et al. 1982; Daubechies, 1990; Colestock, 1993; Huang et al. 1998; Torrence and Compo, 1998).

The next step is to correlate the output of the CWT at each scale (W_{ni}) using Spearman's Rank (r_s) correlation coefficient (Spearman, 1904; Zar, 1972):

$$r_s(W_{ni}) = 1 - \frac{6 \sum d_i^2 (W_{ni})}{n(n^2 - 1)},$$

where d_i^2 is the ranked difference between the outputs of each CWT. The code, therefore, determines the degree of match between oscillations present in the two different signals over a broad scale range. This is particularly useful where signals are highly variable or 'noisy' and where links are difficult to discern from comparison of the individual standard wavelet transforms. Likewise, this provides clearer scope for visual identification of links between the series than alternates such as wavelet coherence (e.g., Grinstead et al. 2004; Cannata et al. 2013) by virtue of generating a single plot whose axes are the scales of the compared datasets, rather than two discrete plots of scales vs. time. This approach also requires less computational power, in addition to the primary benefits of the technique, namely: simplicity of operation and ease in interpretation. This is a code and display approach, which to the authors' knowledge, has not previously been applied or documented in the literature, with the exception of a brief overview given in Pering et al. (2014).

3. The Matlab® Function

In summary, the Matlab function 'corrplot.m' is displayed below, including only those elements related to the production and extraction of data. The full code is available online in the supplementary materials. The code requires a number of inputs: signals x and y (e.g., the data series which are to be compared, which must be of identical sampling frequency and length); wavelet type (e.g., the class of mother wavelet, for example 'morl' for Morlet); scales (e.g., the maximum scale for the CWT - the default setting is to run the CWT in steps of 1, from 1 up to this value); and finally, the sampling rate of the dataset in Hertz (Hz). The dominant oscillation(s) in each of the input series are also determined as part of the code, using Welch's power spectral density (PSD) method (Welch, 1967), as an additional means of identifying similarities in the series. Furthermore, an automatic code-interruption error message is incorporated to avoid analysis above the Nyquist criterion (Nyquist, 2002).

```
function [a,b] = corrplot( x,y,wavelet,scales,fs )
    if scales>((length(x)/2))
        error('Scales above Nyquist limit')
    end
 % Wavelet Transform
```

```
  cwt1=cwt(x/max(x),1:scales,wavelet);
  cwt2=cwt(y/max(y),1:scales,wavelet);
 % Shift the data
  cwt1=ctranspose(cwt1); cwt2=ctranspose(cwt2);
 % Correlate the data
  a=corr(cwt1,cwt2,'type','Spearman');
 % Extract the "best-fit" line
  b=diag(a);
% Extract max and min correlation location
  [max_corr,loc_max_corr]=max(b)
  [min_corr,loc_min_corr]=min(b)
  [M1,N1]=ind2sub(size(b),loc_max_corr);
  [M2,N2]=ind2sub(size(b),loc_min_corr);
% Individual coefficients at max and min location
  wave_coeff1_max=cwt1(:,M1);  wave_coeff1_min=cwt1(:,M2);
  wave_coeff2_max=cwt2(:,M1); wave_coeff2_min=cwt2(:,M2);
% Power spectral densities
  [b1,freq1]=pwelch(x/max(x),scales,0,scales,fs);
  [b2,freq1]=pwelch(y/max(y),scales,0,scales,fs);
```

```
% Xcorr lag plot
    cwt1=ctranspose(cwt1);
    cwt2=ctranspose(cwt2);
        for ls=1:scales;
            s1=cwt1(ls,:);
             s2=cwt2(ls,:);
            maxlags=scales/2;
            lag_corr=xcorr(s1,s2,maxlags, 'coeff');
            c(ls,:)=horzcat(lag_corr);
        end
    c=ctranspose(c);
```

The code generates the following outputs: of which, the first, fourth and sixth can be exported to the Matlab® workspace:

i) a correlation image with colour scale;

ii) power spectral densities of signals ‘x’ and ‘y’;

iii) a 3D visualisation of the correlation image;

iv) correlation coefficients along the 1:1 line in the correlation image;

v) plots of the wavelet coefficients, which correspond to the points of maximum positive and negative correlations, along with 1:1 line;

vi) a plot with colour scale showing the correlation coefficients of the wavelet coefficients at each individual scale, over a defined range of lags.

4. Example applications

Firstly, we present an example application of the code on a pair of synthetic signals to illustrate this approach for establishing the presence of common periodicities. Fig. 1a shows these signals: two sinusoids of period 125 s, with noise added, using a normally distributed random number generator. The generated 2D correlation image (Fig 1b) shows a clear positive correlation between ≈ 75 – 150 s, with a peak value > 0.8, and the dominant series frequencies are further manifest in the Welch's PSD curves in Figs. 1c and 1d showing a clear peak at 125 s (0.008 Hz) in each case. The correlation plot also demonstrates that there are no other sources of significant correlation on any other timescales. For reference, a correlation image showing perfect correlations across all scales is presented in Fig 2. Probability values for observed correlations can be easily estimated using in-built Matlab® algorithms, see Kendall (1970), Best and Roberts (1975), Ramsey (1989), and references therein for additional information.

The 1:1 line is included in Figs. 1b and 2 to highlight the region in which one would expect relationships to occur e.g., where periods are common to both series. Fig. 3a shows the coefficient profile along this line, auto-generated by 'corrplot.m' from the correlation image (Fig. 1b): revealing the scales at which correlation is manifested in this case. It is then for the user to investigate the cause of such links, e.g., through analysis of whether the series are in or out of phase or shifted in phase relative to one another. To expedite this, the code also extracts the wavelet coefficient time series for the scales along the 1:1 line which present the strongest points of maximum and minimum correlation; these outputs are shown in Figs. 3b and 3c, respectively, for our sample synthetic data. In this case, the in-phase nature of the two

125 s period sinusoids is clearly manifested in Fig. 3a. For series which are out of phase, the lag could be determined by visual inspection of these two wavelet coefficient time series. As an additional aid, the code outputs the cross-correlation coefficient at each wavelet coefficient scale over the maximum possible range of lags. The code produces an image (e.g., Fig. 4) which clearly indicates the maximum or minimum lag between series at each scale. This is of particular use when the signals are not perfectly in phase or antiphase. This section of the code is illustrated on a cosinusoidal (s1) and sinusoidal (s2) signal (Fig. 4a), both generated with the same frequency of 90 s, amplitude, and with added random noise. The possible lags can be identified in Fig. 4b clearly corresponding to the known frequency value. These particular functions are of particular use for investigating the links and lags between oscillations and periodicity in natural contexts, where raw signals can demonstrate considerable temporal variability.

We also applied the ‘corrplot.m’ code to measurements of temperatures and relative humidity collected hourly from the Department of Geography, University of Sheffield automatic weather station during June, July and August 2013. The raw data are presented in Fig. 5a and the resulting correlation image is shown in Fig. 5b, facilitating straightforward identification of the links present between the two data series. As expected, strong relationships are present at periods >200 hours (e.g., >8 days), with peak correlation values at ≈ 600-800 hours (e.g., ≈ 25-33 days). This demonstrates that our technique clearly resolves the inter-series links related to synoptic meteorological changes occurring on timescales of weeks. Furthermore, a strong link, of r_s = -0.94 at ≈24 hours is evident, capturing the relationships between changes in temperature and humidity over the diurnal cycle.

For comparison, the continuous wavelet transform plots of these two series are presented in

Figs. 5c and 5d. The cross wavelet coherence and the cross wavelet spectrum are also shown in Figs 5e and 5f, respectively, as generated from the Matlab® wavelet coherence function 'wcoher'. Relative to visual inter-comparison of the wavelet plots, or inspection of either of the other two technique outputs, the correlation plot (Fig. 5b) provides scope for far clearer and more intuitive visualisation of the inter-series links, e.g., illustrating the key benefit of the approach over alternates.

Finally, we present the application of our code on volcanic gas signals: Hydrogen Sulphide (H_2S) and Carbon Monoxide (CO) concentration time series, acquired using a 'Multi-GAS' sensor (Shinohara, 2005; Aiuppa et al., 2005) placed in the plume of the North East Crater of Mount Etna (Sicily, Italy). Fig. 6a shows the correlation image generated. The most significant features are positive links between the datasets at ≈300-400 s, ≈500-700s, and at > 900 s. These are similar to the periodicities in sulphur dioxide SO_2 emission rates reported by Tamburello et al. (2012) indicating that a variety of volcanic gases fluctuate rapidly in their fluxes, with similar periodicity characteristics. In addition, several weak negative correlation areas also appear at ≈100-300 s, ≈400-500 s, and ≈700-900 s, revealing points worthy of further investigation. This technique is particularly useful on data such as these as links between the series are resolvable, even where sensors might have differing response characteristics (Aiuppa et al. 2005). In Fig. 6b, this correlation image is displayed in 3D.

5. Summary and Conclusions

Here, we have presented a new use of CWT analysis combined with correlation to determine the similarity between oscillations present in two separate signals. This paper reports on a straightforward to implement Matlab® code, which executes this approach, providing a more readily interpretable visualisation of these links than available from existing alternate techniques, and the coupled capacity to resolve connections between noisy and transient signals. A number of example applications have been presented, via the analysis of synthetic signals and those acquired from various disciplines within the geosciences, which demonstrate the above benefits.

Acknowledgements

TDP and AMcG acknowledge the support of a NERC studentship, the University of Sheffield and a Google Faculty Research award. AA acknowledges support from the European Research Council Starting Independent Research Grant (agreement number 1305377). We acknowledge logistical support in the fieldwork by colleagues from the Università degli Studi di Palermo and INGV Catania and Palermo.

References

Aiuppa, A., Federico, C., Paonita, A., Giudice, G., Valenza, M., 2005. Chemical mapping of a fumarolic field: La Fossa Crater, Vulcano Island (Aeolian Islands, italy). Geophysical Research Letters 13 (L13309), doi:10.1029/2005GL023207

Best, D. J., Roberts, D. E., 1975. Algorithm AS 89: The Upper Tail Probabilities of Spearman's rho. Applied Statistics 24, 377-379

Cannata, A., Montalto, P., Patanè, D., 2013. Joint analysis of infrasound and seismic signals by cross wavelet transform: detection of Mt. Etna explosive activity. Natural Hazards Earth System Science 13, 1669-1677

Colestock, M. A., 1993. Wavelets – A New Tool for Signal Processing Analysts. Digital Avionics Systems Conference, 1993. 12th DASC., AIAA/IEEE, pp. 54-59

Daubechies, I., 1990. The wavelet transform time-frequency localization and signal analysis. IEEE Transactions Information Theory 36, 961-1004

Frohlich, C., Lean, J., 2004. Solar radiative output and its variability: evidence and mechanisms. Astronomy and Astrophysics Review 12 (4), 273-320, doi:10.1007/s00159-004-0024-1

Grinstead, A., Moore, J. C., Jevrejeva, S., 2004. Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Processes in Geophysics 11, 561-566

Harris, F. J., 1978. Use of windows for harmonic-analysis with Discrete Fourier-Transform. Proceedings of the IEEE 66 (1), 51-83, doi:10.1109/PROC.1978.10837

D. V. Hoyt, and K. H. Schatten, 1997. “The Role of the Sun in Climate Change. Oxford, U.K.: Oxford Univ. Press.

Huang, N. E., Shen, Z., Long, S. R., Lu, M. C., Shih, H. H., Zheng, Q., Yen, N-C., Tung, C. C., Liu, H. H., 1998. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-stationary Time Series Analysis. Proceedings: Mathematical, Physical and Engineering Sciences 454 (1971), 903-995

Hurrell, J. W., 1995. Decadal trends in the North-Atlantic Oscillation – Regional Temperatures and Precipitation. Science 269 (5224), 676-679

Hurrell, J.W., Y. Kushnir, G. Ottersen, and M. Visbeck, Eds., 2003. The North Atlantic Oscillation: Climate Significance and Environmental Impact. American Geophysical Union Geophysical Monograph Series, **134**, 279pp.

Kendall, M. G., 1970. Rank Correlation Methods. London: Griffin.

Lockwood, M.,2012. Solar influence on global and regional climates. Surveys in Geophysics, 33 (3-4). pp. 503-534. doi: 10.1007/s10712-012-9181-3

Morlet, J., Arens, G., Fourgeau, E., Giard, D., 1982. Wave propagation and sampling theory – Part 1: Complex signal and scattering in multilayered media. Geophysics 47 (2), 203-221

Nyquist, H., 2002. Certain topics in telegraph transmission theory (Reprinted from Transactions of the A. I. E. E., February, pg 617-644, 1928). Proceedings IEEE. 90 (2), 280-305, doi:10.1109/5.989875

Oppenheim, A. V., Schafer, R. W., Buck, J.R., 1999. Discrete-Time Signal Processing. 2nd Ed. New Jersey: Prentice Hall

Pering, T. D., Tamburello, G., McGonigle, A. J. S., Aiuppa, A., Cannata, A., Giudice, G., Patanè, D., 2014. High time resolution fluctuations in volcanic carbon dioxide degassing from Mount Etna. Journal of Volcanology and Geothermal Research 270, 115-121, doi:10.1016/j.jvolgeores.2013.11.014

Philander, S.G., 1990. El Niño, La Niña, and the Southern Oscillation. Elsevier International Geophysics series Vol. 46, pp. iii-ix, 1-289.

Ramsey, P. H., 1989. Critical Values for Spearman’s Rank Order Correlation. Journal of Educational and Behavioral Statistics 14 (3), 245-253, doi:10.3102/10769986014003245

Shinohara, H., 2005. A new technique to estimate volcanic gas composition: Plume measurements with a portable multi-sensor system. Journal of Volcanology and Geothermal Research. 143, 319– 333.

Spearman, C., 1904. The Proof and Measurement of Association between Two Things. The American Journal of Psychology 15 (1), 72-101

Tamburello, G., Aiuppa, A., Kantzas, E.P., McGonigle, A. J. S., Ripepe, M., 2012. Passive vs. active degassing modes at an open-vent volcano (Stromboli, Italy). Earth and Planetary Science Letters 359-360, 106-116

Torrence, C., Compo, G. P., 1998. A practical guide to wavelet analysis. Bulletin of the American Meteorological Society 79 (1), 61-78

Welch, P.D., 1967. The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. IEEE Transactions on Audio Electroacoustics AU-15 (2), 70–73

Zar, J. H., 1972. Significance Testing of the Spearman Rank Correlation Coefficient. Journal of the American Statistical Association 67, 578-580

Figure 1 – An example application of the code on synthetic signals showing: a) the signals themselves (two sinusoids of period 125 s with noise added); b) the correlation image generated by the code, with the 1:1 line marked in white, indicating where mutual oscillations are present; c) and d) Welch's power spectral densities of the two series, which show the dominant oscillation at 125 s in each case.

Figure 2 – A sample correlation image for perfect correlation over all scales.

Figure 3 – Three plots auto-generated by the code: a) correlation coefficients along the diagonal 1:1 line extracted from the correlation image in Fig.1b, showing the scales at which correlation is manifested; the wavelet coefficient time series corresponding to scales of maximum b) and minimum c) correlation coefficients in a). The latter plots allow the user to investigate temporal lags between the series, in this case confirming that the two series have a mutual in phase oscillation at 125 s.

Figure 4 – An example application of the code on: a) a cosinusoid (s1) and sinusoid (s2), out of phase with each other, but with matching period of 90 s and added random noise. In b) the last auto-generated plot by the code shows the correlation coefficients at the given lag value and wavelet coefficient scale. The latter plot is of particular use for determining lags, in addition to those in Fig. 3, and also when signals are not in perfect phase or antiphase.

Figure 5 – An example application of our code on temperature and relative humidity measurements, acquired hourly at the automatic weather station of the Department of Geography, at the University of Sheffield, showing: a) the raw data; b) the correlation plot, revealing positive correlation on scales > 200 hours indicative of synoptic meteorological trends and negative correlation on scales of a day in line with diurnal changes; c) and d) continuous wavelet transforms for the two series and e) and f) the cross wavelet coherence and cross wavelet spectrum plots for the data, indicating that the approach presented here provides more intuitive and straightforward visual identification of the inter-series links, than available from these alternatives.

Figure 6 – Output from the code, applied to data concerning Hydrogen Sulphide and Carbon Monoxide emissions from the North East Crater of Mount Etna, showing: a) the 2D correlation image and b) the 3D correlation image.

Figure 1

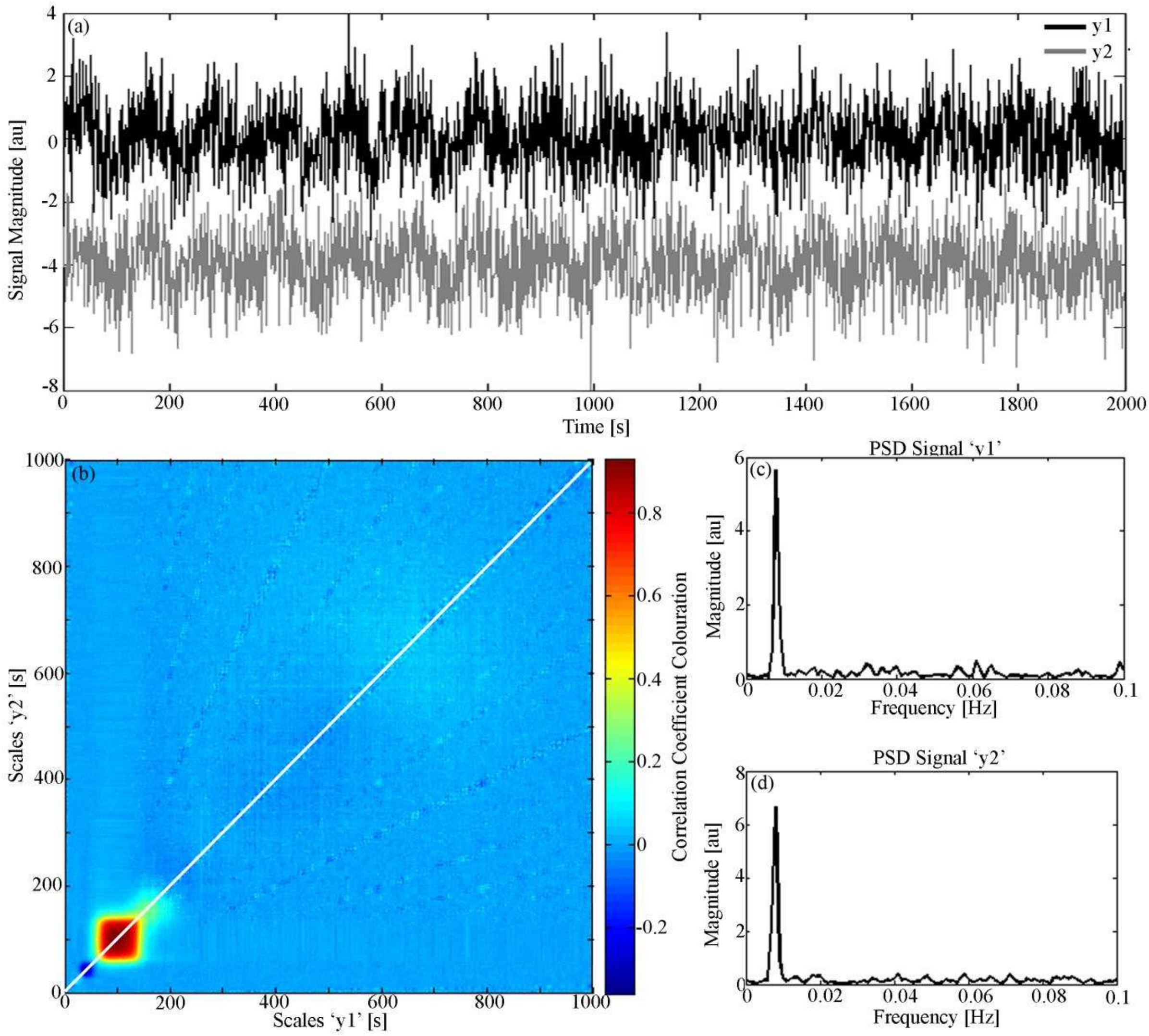

(a)
y1
y2
Signal Magnitude [au]
Time [s]
(b)
Scales 'y2' [s]
Scales 'y1' [s]
Correlation Coefficient Colouration
PSD Signal 'y1'
(c)
Magnitude [au]
Frequency [Hz]
PSD Signal 'y2'
(d)
Magnitude [au]
Frequency [Hz]

Figure 2

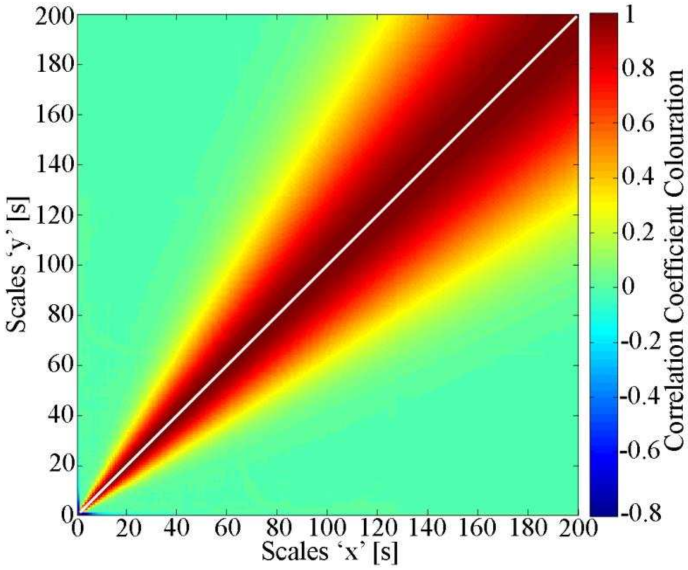


Figure 3

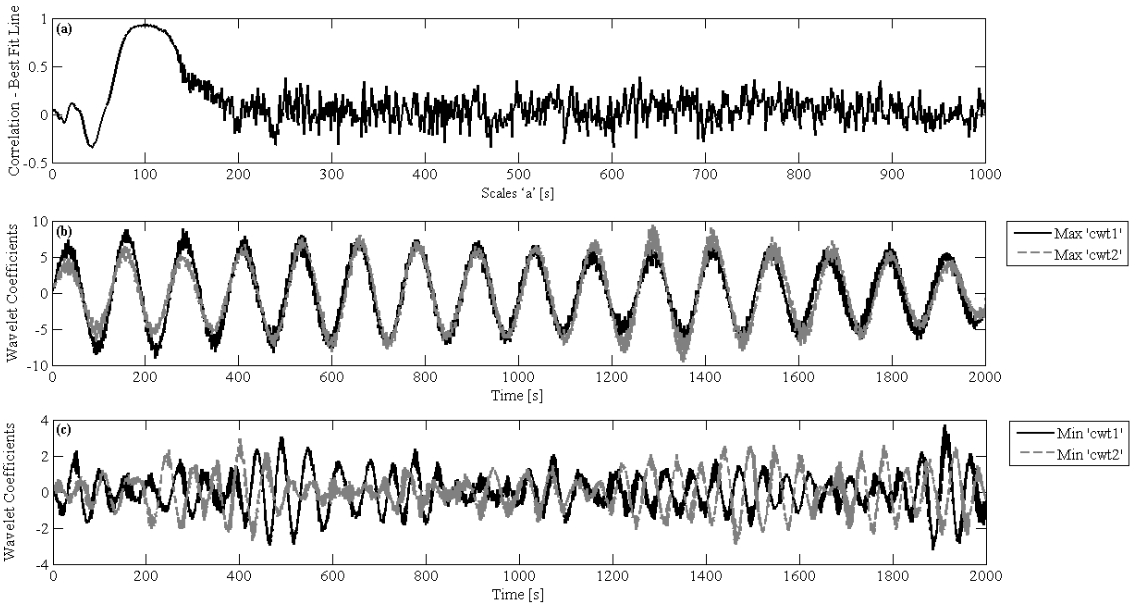


Figure 4

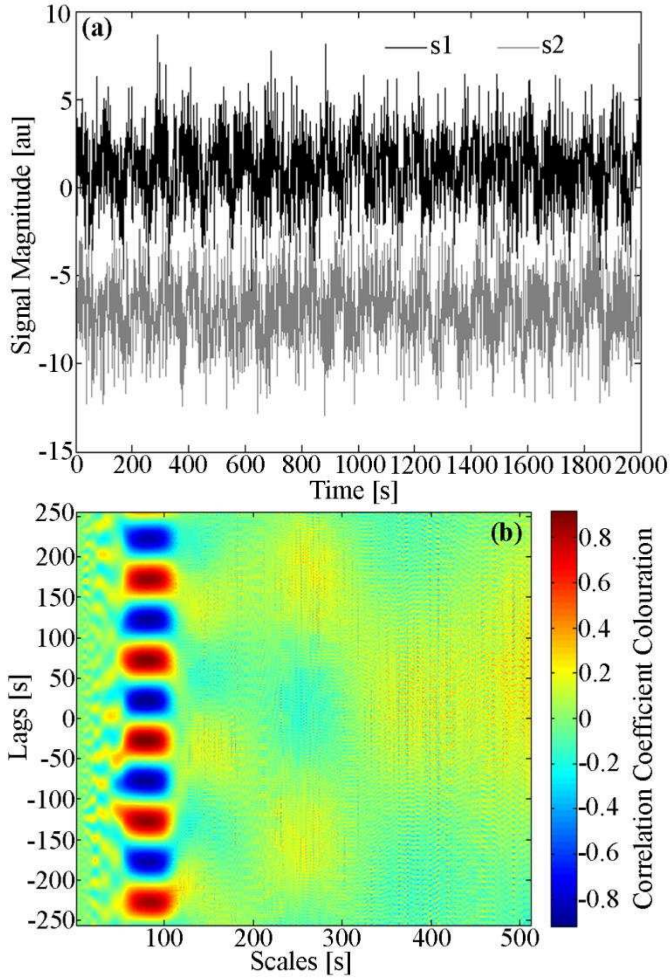
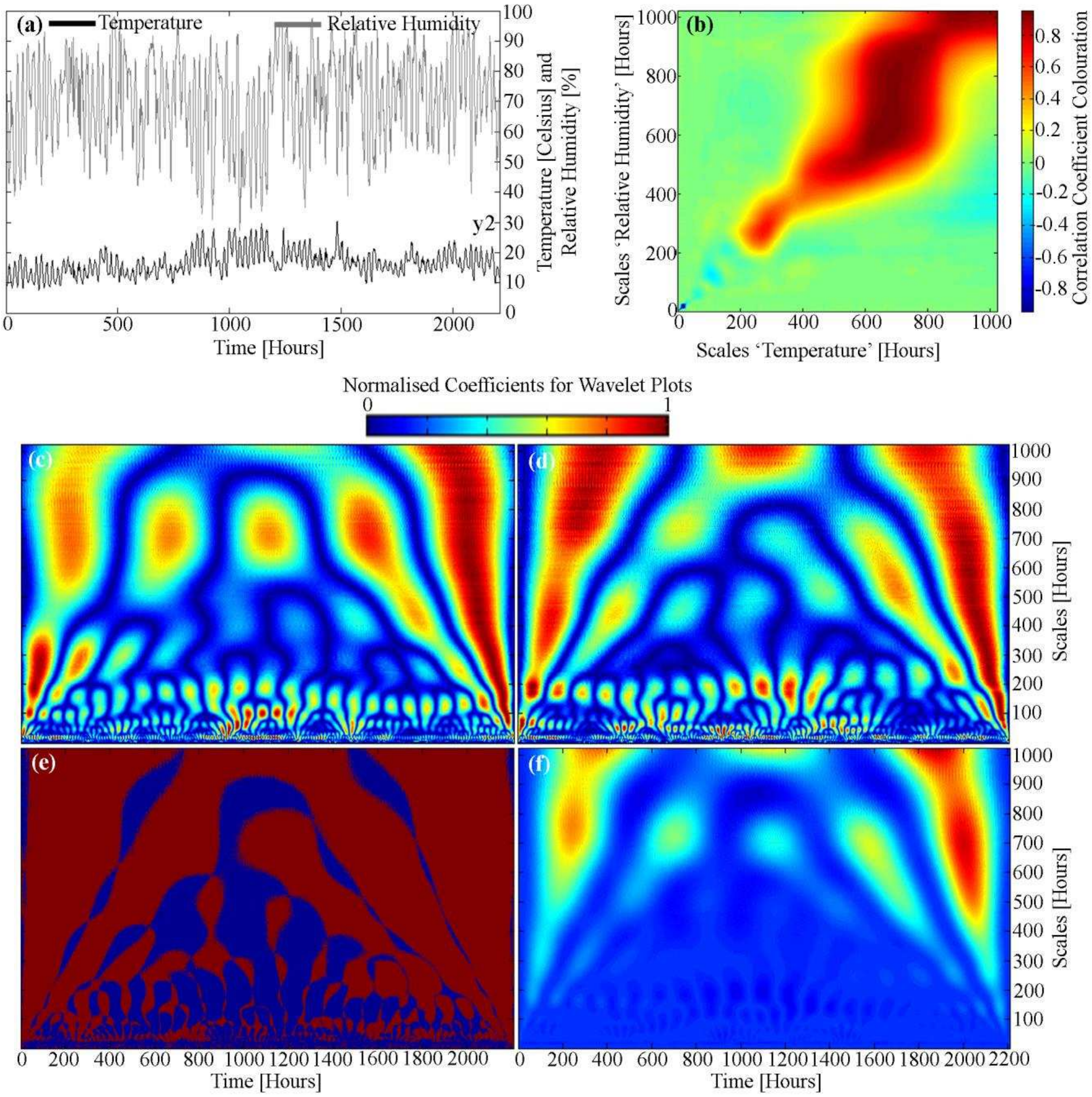
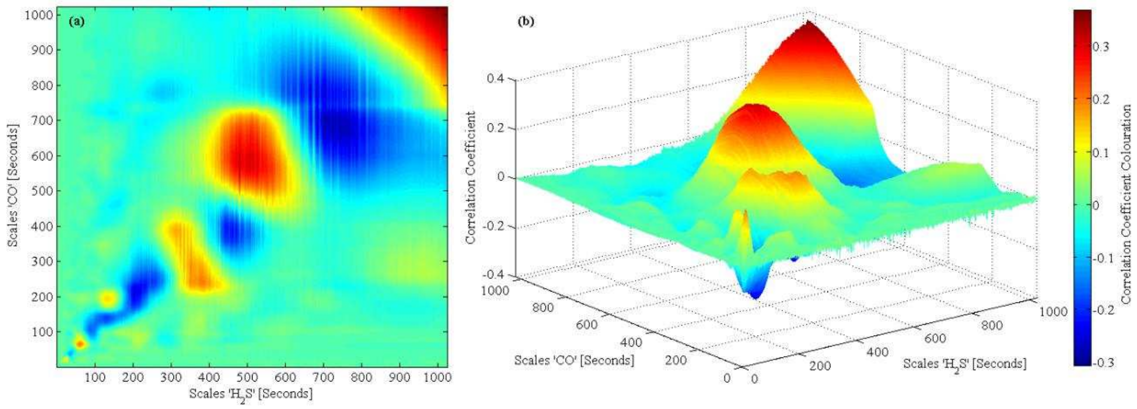

(a)
s1
s2
Signal Magnitude [au]
10
5
0
-5
-10
-15
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Time [s]
(b)
Lags [s]
250
200
150
100
50
0
-50
-100
-150
-200
-250
100
200
300
400
500
Scales [s]
Correlation Coefficient Colouration
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8

Figure 5



(a)
Temperature
Relative Humidity
Temperature [Celsius] and Relative Humidity [%]
y2
Time [Hours]
(b)
Scales 'Relative Humidity' [Hours]
Scales 'Temperature' [Hours]
Correlation Coefficient Colouration
Normalised Coefficients for Wavelet Plots
(c)
(d)
(e)
(f)
Scales [Hours]
Time [Hours]

Figure 6


(a)
1000
900
800
700
600
500
400
300
200
100
Scales 'CO' [Seconds]
100 200 300 400 500 600 700 800 900 1000
Scales 'H2S' [Seconds]
(b)
0.4
0.2
0
-0.2
-0.4
Correlation Coefficient
1000
800
600
400
200
0
Scales 'CO' [Seconds]
0
200
400
600
800
1000
Scales 'H2S' [Seconds]
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
Correlation Coefficient Colouration