Data Analysis and Investigation of Self-Similarity in Oceanographic Sediment Data

Lam Ling Shum and Lionel Sacks

Advanced Communications Systems Engineering Group, Department of Electronic and Electrical Engineering, University College London

Abstract: This paper aims to analyse the oceanographic sediment data obtained from the WaveNet project. In particular we would like to search for any deterministic trends present and explore the degree of self-similarity in the data. The results from the data analysis will form the foundation of the development of adaptive sampling and data fusion strategies in the SECOAS sensor network project.

1. Introduction

The work in this paper is initiated by the Self-Organised Collegiate Sensor Network Project (SECOAS). SECOAS aims to deploy a network of low-cost sensors for the efficient retrieval of oceanographic data with good temporal and spatial resolution. This is to replace the traditional approach of using one or a few expensive and high precision sensors for data collection. The sensors communicate with a wireless ad-hoc network and the network is distributed [1].

The SECOAS sensor network has the specialities of distributed algorithms [2][3], and simple biologically inspired algorithms [4][5][6]. Data analysis is performed prior to the development of data handling algorithms including adaptive sampling and fusion strategies. It is important to understand the nature of data in relation to the costs specific to the project, which are battery power and communication bandwidth. Efficient use of power can extend sensors' lifetime and communication bandwidth is a scarce resource in the ocean environment. The requirement for preserving interesting science, sensor power, network bandwidth and user input regarding data resolution form the specific network engineering scenario.

Turbidity is one of the parameters obtained from the WaveNet project at Scroby Sands area at Great Yarmouth [7]. Other measurements are temperature, conductivity, and pressure. Turbidity, a measurement of *sediment level* is selected in this paper because of two reasons. Firstly, SECOAS monitors the environmental impact of the wind farm on the coastal area and sand banks and sediment is the major measurement for such event. Secondly, out of the four measurements available, sediment is the least predictable and is mostly dependent on the local area, thus, making it the most interesting in engineering terms. The data were gathered in bursts of 1024 points every hour/ bin at 1Hz frequency. The system then slept for the rest of the hour (2576 seconds) and woke up afterwards for the measurement in the next hour.

2. Parameter Description

Sediment, or precipitation refers to both organic and inorganic loose material that is moved from time to time by physical agents including wind, waves, currents and gravity [8]. The short space and time scales of sediment make it difficult to interpret point measurements [9]. There are a variety of methods to measure sediments including the use of chemical tracers, light attenuation and scattering, and oxygen isotope. Light scattering is used to obtain the test data in this project.

Turbidity, or cloudiness of water describes sediment level in a relative term depending on characteristics of the scattering particles, external lighting conditions and the instrument used. Turbidity in this project is measured in Formazin Turbidity Units (FTUs). It is derived from diluted concentrations of 4000-FTU formazin, a murky white suspension that can be purchased commercially [9].

3. Methodology

Basic analysis on the time series and frequency domain using Fast Fourier Transform (FFT) will be performed first. We then proceed to determine if self-similarity exists in the data.

- **Self-similarity**

Self-similarity is one of the characteristics that result from a long-range dependency (LRD) or 1/f random process. A stationary LRD process has the characteristic of:-

$$\lim_{\tau \to \infty} \frac{S_X(f)}{C_s \left| f \right|^{\alpha}} = 1 \qquad -1 < \alpha < 0 \qquad (1)$$

where S_x denotes the spectral density function (SDF), *f* is the frequency and C_s is a constant. $S_X(f) \approx C_s |f|^{\alpha}$ with *f* approaches zero. An alternative definition stated in terms of autocorrelation function (ACF) is such that:-

$$\lim_{\tau \to \infty} \frac{s_{X,\tau}}{C_s |\tau^{\beta}|} = 1 \qquad -1 < \beta < 0 \tag{2}$$

where $s_{X,\tau}$ denotes the ACF and τ denotes the time lag. A stationary process has $s_{X,\tau} \approx C\phi^{\tau}$ where a LRD process has $s_{X,\tau} \approx C_s \tau^{\beta}$ for large τ. In both case $s_{X,\tau} \to 0$ as $\tau \to \infty$, but the rate of decay toward zero is much slower for a long memory process, implying that observations that are widely separated in time can still have a covariance that cannot be neglected [10][12].

Common stationary models for LRD processes includes fractional Gaussian noise (fGn), fractional difference process (FD) and pure power law (PPL) [10][12]. In general, α from Eq. (1) is used to quantify the degree of LRD. The *Hurst parameter H* is particularly used to characterize fGn where *0<H<1* indicates a stationary LRD fGn process. Although fGn is always stationary while LRD processes can be stationary or non-stationary, *H* and α are both used interchangeably in literature for any LRD process with the relation of $\alpha = 1 - 2H$ (Table 1).

Non-stationary LRD process	**Stationary LRD process**	**White noise**	**Stationary not LRD process**
$H \geq 1$ or $\alpha \leq -1$	$1/2 < H < 1$ or $-1 < \alpha < 0$	$H = 1/2$ or $\alpha = 0$	$H < 1/2$ or $\alpha \geq 0$

Table 1: Relation of H, α and type of random process

- **Hurst estimate - Wavelet weighted least square estimator**

Wavelet least squares fit estimator (WLSE) proposed in [14] is used for Hurst estimation. WLSE is based on Discrete Fourier Transform (DWT) [10][11] and is described in [10] for estimating a quantity called *wavelet variance* defined as:

$$\hat{v}_x^2(\tau_j) = \frac{1}{N} \sum_{t=0}^{N_j - 1} W_{j,t}^2 \tag{3}$$

where $W_{j,t}^2$ denotes the wavelet coefficients at octave j or scale τ_j. [10] proposes that $v_x^2(\tau_j) \propto \tau_j^{-\alpha}$ for LRD process such that when v_x^2 is plotted against τ_j a straight line is obtained with a slope of $-\alpha$. Therefore, α, or $H = (1+\alpha)/2$, can be estimated by linearly regressing $\log(v_x^2(\tau_j))$ on $\log(\tau_j)$ using least square fit. [12][14] further proposed the use of weighted least square fit to increase the robustness of the estimator. The weights $S_j = (n \ln^2 2)/2^{2+1}$ used are the inverse of the theoretical asymptotic variance of a chi-square distribution [14].

(a) (b)

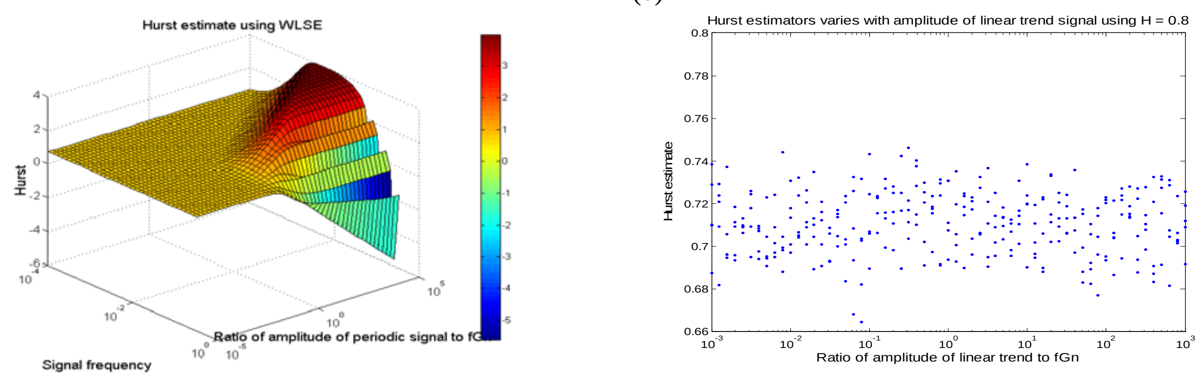


Figure 1: Testing of robustness of WLSE Hurst estimator using fGn H=0.8 for (a) periodic signal with various amplitudes and frequencies, (b) linear trend.

- **Robustness of the estimator**

Since the turbidity data exhibit both linear trends and periodicity, the simpler form of time domain estimators including the R/S statistics, correlogram, periodogram and variance plot [12] have poor statistical performance, notably high bias and sub-optimal variance [14][12]. The effect has been studied in many other research papers

[14][15][16]. We experiment on some non-stationary trends specific to the data characteristics in preparation for the analysis section.

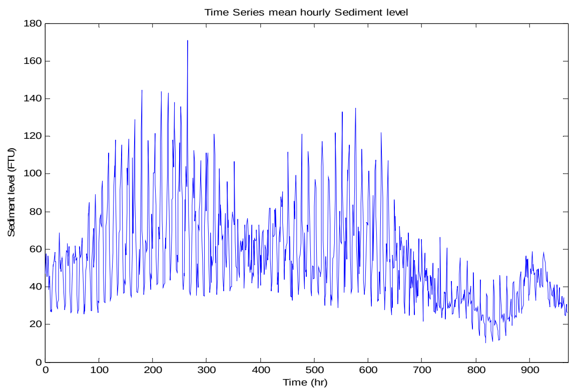
Figure 1 is the result for the robustness test on WLSE estimator. From Figure 1(b) it is observed that the WLSE is robust against linear trend, that it gives consistent estimate of around H = 0.71 for all amplitudes of linear trend. In Figure 1(a), the flat region is the 'safe region' for the estimator where it gives consistent results. We can see that the estimator is only robust with some small amplitude and long period of periodic signals. If the period is long enough ($f < 10^{-4}$) it is robust against all amplitudes since the periodic signal would resemble a linear trend.

4. Results of the Analysis

Figure 2 (a) plots the hourly mean of the burst data in each bin. Periodicity is observed and the time series shows a drifting mean and varying variance overall. Figure 2 (b) confirms the periodicity observed. There is a sharp peak to the spectrum at 12.28 hours. A smaller frequency component appears at 6.06 hour. Since the spectrum is not smoothed the periods obtained are approximations only. The strong power observed at the very low frequencies is due to a non-zero mean of the data.

Figure 3, which plots the frequency spectrum in hour 1 shows that there is no periodicity within a burst. It can be observed that the power decays with the frequency similar to 1/f noise. This may be evidence to suggest the presence of long-range dependency (LRD). Figure 4 shows a slowly decaying ACF. The coefficients are large even after 100 lags. Figure 5 is the Hurst estimates using WLSE for every burst over time. The estimates range from 0.6 to 1.3 with a mean of 0.887. This is a strong evidence for the presence of LRD in the turbidity data in short time scale. Note that the periodicity observed in Figure 3(a) is 10^4 longer than the sampling time so the estimator is robust in that region. Figure 6 shows an interesting result that the Hurst estimates are actually periodic in synchronous with the actual data themselves.

(a)



(b)

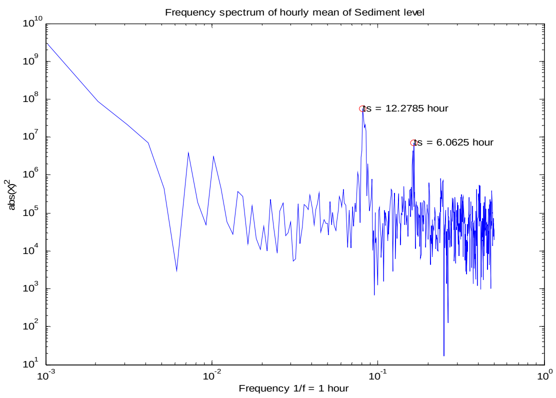


Figure 2: Basic analysis of (a) Time series of hourly mean of Turbidity (b) FFT of hourly mean data

5. Conclusion

In this paper, the turbidity data as a measurement of sediment level has been analysed first in the time and frequency domains. In the time domain periodicity is observed for hourly mean data. The frequency spectrum of hourly mean data confirms there is a dominant periodicity at 12.3 hour and a frequency component of less power at 6.06 hour.

Another quantity we look for is the degree of self-similarity in the data. Self-similarity is an expression for long-range dependency in a random process and is often quantified by the Hurst parameter. In this paper we have demonstrated the use of wavelet weighted least square estimate (WLSE) for Hurst estimation in data. The estimator is robust against linear trends, long periodicity or small amplitudes of periodic signals.

Self-similarity is observed in several ways in the short time-scale turbidity data. The power of the spectrum decays with the frequency similar to 1/f noise. The autocorrelation function shows some strong correlation of data even at 100 time lags. The mean Hurst parameter estimated is 0.887 and is an evidence for self-similarly in the data. Another interesting characteristic is that the Hurst estimates shows periodic behaviour similar to the time series of hourly mean data.

In the future we would like to explore the modelling and parameter estimation with the test data and develop appropriate sampling and fusion strategies for the SECOAS sensor network.

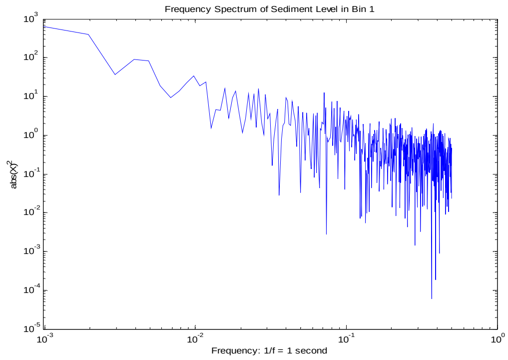


Figure 3: FFT of sediment data in bin 1

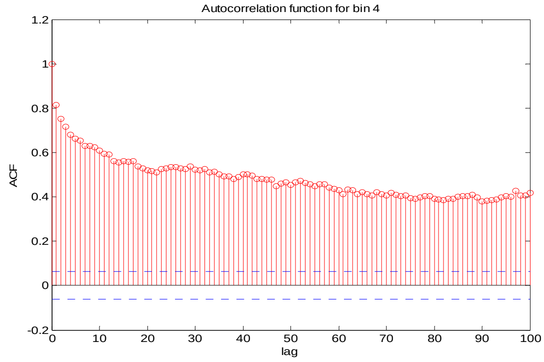


Figure 4: Autocorrelation of sediment data in bin 4

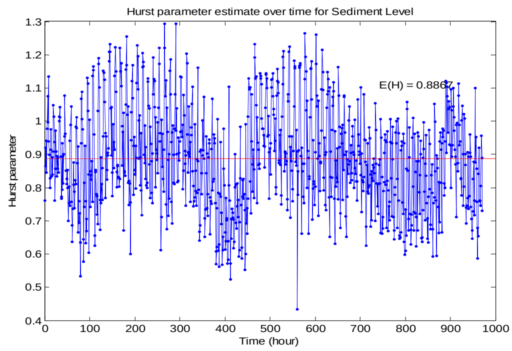


Figure 5: Hurst estimate using WLSE for each bin over time

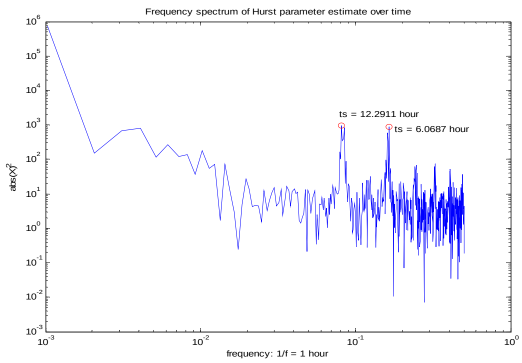


Figure 6: FFT for Hurst estimate in Figure 5

Acknowledgments.

We would like to thank Adrian Li Mow Ching and Ibiso Wokoma for their help with this paper.

References.

[1] L. Sacks, M. Britton, I. Wokoma, A. Marbini, T. Adebutu, I. Marshall, C. Roadknight, J. Tateson, D. Robinson, A. Velazquez, "*The development of a robust, autonomous sensor network platform for environmental monitoring*", IoP Sensors & their Applications (S&A XII) University of Limerick, Ireland, Sep 2-4, 2003.

[2] L. Shum, I. Wokoma, T. Adebutu, A. Marbini, "*Distributed Algorithm implementation and interaction in wireless sensor network*", International workshop on Sensor and Actor Network Protocols and Applications, August 2004 *(to appear)*.

[3] L. Shum, Presentation on "*Self-organisation in SECOAS sensor network*", Multi-service networks conference, July 2004.

[4] I. Wokoma, L. Sacks, I. Marshall, "*Biologically inspired model for sensor network design"*, London Communications Symposium 2002.

[5] I. Wokoma, L. Sacks, I. Marshall, "*Clustering in sensor networks using quorum sensing"*, London Communications Symposium 2003.

[6] I. Wokoma, I. Liabotis, O. Prnjat, L. Sacks, I. Marshall, "*A Weakly Coupled Adaptive Gossip Protocol for Application Level Active Networks"*, International workshop on policies for distributed systems and networks, June 2002.

[7] WaveNet Long-term Wave monitoring around the English and Welsh Coast http://www.cefas.co.uk/wavenet/default.htm.

[8] Gerhard Masselink, Michael Hughes, "*Introduction to coastal processes & geomorphology*", Oxford University Press Inc., 2003.

[9] William J. Emery and Richard E. Thomson, "*Data analysis methods in physical oceanography*", 2nd and revised edition, Elsevier Science B.V., 2001.

[10] Donald B. Percival & Andrew T. Walden, "*Wavelet Methods for Time Series Analysis*", Cambridge University Press 2000.

[11] Amara Graps, "An introduction to wavelets", IEEE computational science and engineering vol.2, num. 2, 1995.

[12] Jan Beran, "*Statistics for Long-Memory Processes*", Chapman & Hall CRC Press, USA, 1998.

[13] Darryl Veitch, Patrice Abry, "A wavelet based joint estimator of the parameters of long-range dependence", IEEE Trans. Info. Th. April 1999, Vol 45, No.3, 1999.

[14] P. Abry, D. Veitch, "*Wavelet Analysis of Long Range Dependent Traffic*", Trans. Info. Theory, Vol.44, No.1 pp.2-15, Jan 1998.

[15] Mohammed F. Hawa, "A Study of Deterministic Trends in Web-Cache Traffic Using Wavelet Analysis", London Communications Symposium, 1999.

[16] V. Teverovsky and M. Taqqu. "Testing for long range dependence in the presence of shifting means or a slowly declining trend, using a variance type estimator". Journal of Times Series Analysis, Vol 18 No3 1997.