Detection of Anomalous Observations Using Wavelet Analysis in Frequency Domain

Aideyan D.O.

Dept of Mathematical Sciences, Kogi State University, Anyingba, Kogi State, Nigeria

Abstract- This study used wavelet analysis to detect Anomalous Observations (AOs) in non-stationary and non-periodic simulated data while extending it to the parametric method using a developed test statistic. In the three series analysed, it was discovered that Turkey's method (TM), the Modified Turkey's method (MTM) and the developed test statistic were found to be very effective and efficient in the detection of these anomalous observations even when the data are compressed, they still retained their statistical properties

Index Terms- Anomalous, Coefficients, Resolution, Spectrum, Software,

I. Introduction

Wavelet analysis is known as non-parametric orthogonal series estimator which are capable of providing the necessary time and frequency information on time series data simultaneously in a highly flexible fashion. It is an alternative to the spectral method and it reduces the size of the series into resolutions without losing the statistical properties of the series. In statistics, (AOs) are observations that are numerically distinct from the rest of the data. They occur by chance in any distribution but are often indicative either of measurement error or that the population is heavy tailed. Section 2 will dwell on the overview of wavelet analysis using matrix representation, key advantages of wavelet analysis and the aim and objectives of this paper. Section 3 shows the tables of the analysis and its discussions while Section 4 includes summary of findings, interpretation of results, conclusion, recommendations and suggested area for further work.

II. WAVELET ANALYSIS

Wavelet analysis is a statistical tool that can be used to extract information from any kind of data and are generally needed to analyze data fully at different resolution (scale) and location. Eckley,A.I. et. al (2005)

Discrete Wavelet Transform re - expresses a time series in terms of coefficients that are associated with a particular time and a particular dyadic scale 2^{J}. These coefficients are fully equivalent to the original series from its Discrete Wavelet Transform coefficients. Nason, G.P. (2002), Armando D. M et. al (2003)

The Discrete Wavelet Transform allows us to partition (decompose) the information in a time series into pieces that are associated with different scales and time. This decomposition is very close to the statistical technique known as the Analysis of variance (ANOVA), so DWT leads to a scaled – based ANOVA that is quite analogous to the frequency – based ANOVA provided by the power spectrum Graps A.(1995),

2.1 MATRX REPRESENTATION

Like the orthonormal discrete Fourier transform, the discrete wavelets transform (DWT) of X_t is an orthonormal transform Neill P. (2012. Let $[W_n; n = 0 \ldots\ldots.. N-1]$ be the DWT coefficients then, we can write $W = w_n$ where W is a column vector of length $N = 2^2$ whose n^{th} DWT and satisfying $w^T w = I_N$ orthornormality implies that $X = w^T w$ and $\|W\|^2 = \|X\|^2$. Hence W_n^2 represents the contribution to the energy attributable to the DWT coefficient with index n.

Whereas ODFT coefficients are associated with frequencies the n^{th} wavelet coefficient W_n is associated with a particular scale and with a particular set of times Daubechies, I (1988, 1992).

Explicitly, the rows of this mature for n=0, 8, 12, 14, and 15 are

$$w_0^T = \left[-{}^{1}/_{\sqrt{2}}, {}^{1}/_{\sqrt{2}}, \underbrace{0 \ldots\ldots\ldots\ldots\ldots.0}_{14\ zero}\right]$$

$$w_8^T = \left[-{}^{1}/_{2}, {}^{-1}/_{2}, {}^{1}/_{2}, {}^{1}/_{2}, \underbrace{0 \ldots\ldots\ldots.0}_{12\ zero}\right]$$

$$w_{12}^T = \left[-{}^{1}/_{\sqrt{8}}, \ldots\ldots\ldots\ldots.., -{}^{1}/_{\sqrt{8}}, {}^{1}/_{\sqrt{8}} \ldots\ldots\ldots {}^{1}/_{\sqrt{8}}, \underbrace{0 \ldots\ldots..0}_{8\ zero}\right]$$

$$w_{14}^T = \left[-{}^1/_{\sqrt{4}}, \dots \dots .., -{}^1/_{\sqrt{4}}, {}^1/_{\sqrt{4}} \dots \dots \dots {}^1/_{\sqrt{4}}\right]$$

$$w_{15}^T = \left[{}^1/_{\sqrt{4}} \dots \dots \dots {}^1/_{\sqrt{4}}\right]$$

The remaining eleven rows are shifted version of the above;
$w_1 = T^2 w_0$, $w_2 = T^4 w_0 \dots \dots \dots \dots \dots \dots \dots \dots \dots w_7 = T^{14} w_0$
$w_9 = T^4 w_8$, $w_{10} = T^8 w_8$ $w_{11} = T^{12} w_8$
$w_{13} = T^8 w_{12}$

Let us now, define exactly what the notation of scale means for a positive integer k let

$$\bar{X}_t(k) = \frac{1}{k}\sum_{l=0}^{k-1} X_{t-1} \qquad (2.13)$$

Percival, et. al (2000)

2.2 The Key advantages of Wavelet Analysis over Other Methods. (Nason, 2008)

- Sparsity of representation for a wide range of data as resolution decreases including those with discontinuities guaranties the presence of required statistics.
- The ability to analyze data at a number of resolutions and also to work with information at such resolutions.
- Ability to detect aberrant observations and represent neighbourhood features and also to create localized features on synthesis (the process of combining differences into a new whole)
- Efficiency in terms of compilation speed and storage.

2.3 AIM AND OBJECTVES

The major goal of this paper is to compare the efficiency of wavelet coefficients as a tool for detecting aberrant observations in both simulated and real data.

Below are the underlying objectives are to :

- Use Turkey's and modified Turkey's methods in wavelet analysis for detection of aberrant observations even at lower resolutions.
- Compare the performance of these two methods with a view of identifying the methods with a better detective mechanism of aberrant observations time series data.
- Use a derived test statistic to analyse such data and compare their results.

Section 3

III. DATA ANALYSIS AND DISCUSSION OF RESULTS

Since wavelet analysis is dyadic, the data analysed were three simulated data from normal distribution using R-Software involving 512, 1024 and 2048 data sets containing four, four, and eight injected AOs.

3.1 DATA ANALYSIS

Table 3.1: Wavelet Analysis of simulated data N = 512, A.O. = 4 using Turkey's Method

Resolutions Level (No of Observations)	Location (L) of Aberrant Observations	Aberrant Observation Values	T L	TU
9 (512)	1, 256, 257, 512	43, -44, 41, 47	-2.61	2.68
8 (256)	1, 128, 129, 256	30.35, 31.66, 28.94, -32.68	-2.44	2.63
7 (128)	1, 64, 65, 128	21.41, -21.89, 20.41, 23.61	-2.65	2.93
6 (64)	1, 32, 33, 64	16.43, 15.45, 15.72, -16.73	-2.84	3.42

5 (32)	1, 16, 17, 32	12.25, -10.69, 11.75, 12.06	-2.53	2.02

Note: T_U and T_L represent upper and lower limits for both Turkey's and Modified Turkey's Methods.

Table 3.2: Wavelet analysis using modified Turkey's method by Neil Patterson (2012)

Resolutions Level (No of Observations)	Location (L) of Aberrant Observations	Aberrant Observation Values	T L	TU
9 (512)	1, 256, 257, 512	43, -44, 41, 47	-5.85	5.92
8 (256)	1, 128, 129, 256	30.35, 31.66, 28.94, -32.68	-5.55	5.74
7 (128)	1, 64, 65, 128	21.41, -21.89, 20.41, 23.61	-6.06	6
6 (64)	1, 32, 33, 64	16.43, 15.45, 15.72, -16.73	-6.77	7.25
5 (32)	1, 16, 17, 32,	12.25, -10.69, 11.75, 12.06	-5.32	4.81

Series: B: Simulated data N =1024,A.O. = 4

Table 3.3: Wavelet analysis using Turkey's method

Resolutions Level (No of Observations)	Location (L) of Aberrant Observations	Aberrant Observation Values	T L	TU
10 (1024)	1, 512, 513, 1024	32,41,40,-37	-2.75	2.75
9 (512)	1, 256, 257, 512	22.57,-28.89, 28.23, 26.26.	-2.92	3
8 (256)	1, 128, 129, 256	15.91 , 21.15, 19.91, -17.85	-3.14	3.35
7 (128)	1, 64, 65,128	12.54, -14.36, 15.36,13.21	-3.18	3.44
6 (64)	1, 32, 33, 64	9.50, 10.10, 11.50, -9.40	-2.4	2.41

Table 3.4: Wavelet analysis using modified Turkey's method by Neil Patterson (2012)

Resolutions Level (No of Observations)	Location (L) of Aberrant Observations	Aberrant Observation Values	T L	TU
10 (1024)	1, 512, 513, 1024	32,41,40,-37	-6.11	6.11
9 (512)	1, 256, 257, 512	22.57,-28.89, 28.23,26.26.	-6.54	6.63
8 (256)	1, 128, 129, 256	15.91 , 21.15, 19.91, -17.85	-7.11	7.32
7 (128)	1, 64, 65,128	12.54, -14.36, 15.36,13.21	-7.23	7.49
6 (64)	1, 32, 33, 64	9.50, 10.10, 11.50,-9.40	-5.34	5.35

Series: C Simulated data N = 2048, A.O. = 8

Table 3.5: Wavelet analysis using Turkey's method

Resolutions Level (No of Observations)	Location (L) of Aberrant Observations	Aberrant Observation Values	T L	TU
11 (2048)	1, 512, 513, 1024, 1025,1536, 1537, 2048	45,-51, 50, -52, 48, -49,47, 46	-2.75	2.75

10 (1024)	1, 256, 257, 512, 513, 768,769, 1024	31.76,36.16,35.30,36.8,33.89,34.75, 33.18,32.43	-2.92	3.00
9 (512)	1, 128, 129, 256, 257, 384,385, 512	22.41, -24.85,24.91,25.35, 23.91, -23.85,23.41, 23.65	-3.14	3.35
8 (256)	1, 64, 65, 128, 129, 192,193, 256	17.13, 18.16, 18.90,18.52, 18.19, 17.46,17.84, -16.13	-3.18	3.45
7 (128)	1, 32, 33, 64, 65, 96, 97,128	12.75,-12.90,14.00,13.15, 13.50,-12.40, 13.25,11.35	-2.4	2.41

Table 3.6: Wavelet analysis using modified Turkey's method by Neil Patterson (2012)

Resolutions Level (No of Observations)	Location (L) of Aberrant Observations	Aberrant Observation Values	T L	TU
11 (2048)	1,512,513,1024,1025,1536, 1537, 2048	45,-51, 50, -52, 48, -49, 47, 46	-6.11	6.11
10 (1024)	1, 256, 257, 512, 513, 768,769, 1024	31.76,36.16,35.30,36.87,33.89,34.75,33.18, -32.43	-6.54	6.63
9 (512)	1, 128, 129, 256, 257, 384,385, 512	22.41, -24.85,24.91, -25.35, 23.91, -23.85,23.41, 23.65	-7.11	7.32
8 (256)	1, 64, 65, 128, 129, 192,193, 256	17.13, 18.16, 18.90,18.52, 18.19, 17.46,17.84, -16.13	-7.23	7.49
7 (128)	1, 32, 33, 64, 65, 96, 97, 128	12.75,-12.90, 14.00,-13.15, 13.50,-12.40,13.25,11.35	-5.34	5.35

3.2 Using The Developed Test Statistic

Series: A Simulated data N=512, α=5%, and 10% where U=1.96, and 1.28

Table 3.7: Wavelet analysis of parametric method

Resolutions Lev. (No of Obsns)	Location (L) of AOs	AOs Values	Location (L) of U Values	U Values
9 (512)	1, 256, 257, 512	43, -44, 41, 47	1, 256, 257, 512	4.61, -4.78, 4.40, 5.04
8 (256)	1, 128, 129, 256	30.35, 31.66, 28.94, -32.68	1, 128, 129, 256	3.43, 3.58, 3.27 -3.74
7 (128)	1, 64, 65, 128	21.41, -21.89, 20.41, 23.61	1, 64, 65, 128	2.56, 2.71, 2.43, 2.83
6 (64)	1, 32, 33, 64	16.43, 15.45, 15.72, -16.73	1, 32, 33, 64	2.01, 1.89, 1.78, -2.19

Series: B N=1024, α= 5%, and 10% where U= 1.96, and 1.28

Table 3.8: Wavelet analysis of parametric method

Resolutions Lev. (No of Obsns)	Location (L) of AOs	AOs Values	Location (L) of U Values	U Values

10 (1024)	1, 512, 513,1024	32,41,40,-37	1, 512, 513, 1024	5.10, 6.54, 6.38,-5.92
9 (512)	1, 256, 257,512	22.57,-28.89, 28.23, 26.26	1, 256, 257, 512	3.75, -4.86, 4.70, 4.37
8 (256)	1, 128, 129, 256	15.91 , 21.15 19.91, 17.85	1, 128, 129, 256	2.80, 3.73, 3.51, -3.21
7 (128)	1, 64, 65,128	12.54, -14.36 15.36, 13.21	1, 64, 65, 128	2.25, -2,71 2.77, 2.37
6 (64)	1, 32, 33, 64	9.50, 10.10,11.49, -9.40	1, 32, 33, 64	1,99,1.97, 2.18, -1.99

Series: C N=2048, α= 5%, and 10% where U= 1.96, and 1.28

Table 3.9: Wavelet analysis of parametric method.

Resolutions Lev. (No of Obsns)	Location (L) of AOs	AOs Values	Location (L) of U Values	U Values
11 (2048)	1, 512, 513,1024, 1025,1536, 1537,2048	45,-51, 50, -52, 48, -49,47, 46	1, 512, 513, 1074, 1075, 1536, 1537, 2048	5.48, -6.22, 6.09, -6.34, 5.84, -5.97, 5.72, 5.60
10 (1024)	1, 256, 257,512,513, 768, 769,1024	31.76,36.16,35.30,36.87,33.89,34.75, 33.18,-32.43	1, 256, 257, 512, 513, 768, 769, 1024	4.02 4.58, 4.474.67, 4.29, 4.40, 4.20, -4.17
9 (512)	1, 128, 129,256, 257, 384, 385,512	22.41 , 24.85, 24.91, -25.35, 23.91,-23.85, 23.41, 23.65	1, 128, 129, 256,257, 384, 385, 512	3.02,-3.38, 3.36,-3.45, 3.22, -3.24, 3.15, 3.19
8 (256)	1, 64, 65, 128,129, 192, 193,256	17.13,18.16,18.52,18.19,17.46,17.84, -16.13	1, 64, 65, 128,129, 192, 193,256	2.32 2.46, 2.56,2.51, 2.46, 2.36, 2.4144 -2.3286
7(128)	1, 32, 33, 64, 65, 96, 97,128	12.75, 12.90, 14.00, -13.15, 13.50, -12.40, 13.25, 11.35	1,32,33,64,65,96,97,128	2.82, 2.97, 2.05, 2.98, 2.98, -2.80, 2.01, 2.99
6 (64)	1, 16, 17, 32,33	9.36, 8.92, 10.24, 9.09, 9.89, 8.56, 9.71, -8.23	1,16,17,32,33,48,49, 64	2.05, 2.28, 1.96, 2.30, 2.43, 2.22, 2.40, -2.47

3.3 Discussions on the Tables

Table 3.1:It can be seen that the four aberrant observations injected randomly were identified up to the fifth resolution at the same location using Turkey's method.

Table 3.2: Also shows that the four aberrant observations injected randomly were identified up to the fifth resolution at the same location using Modified Turkey's method.

Table: 3.3. The results obtained using 1024 observations with four aberrant observations injected shows that Turkey's method detected the aberrant observations up to the sixth resolution level at the same location.

Table: 3.4. Confirms the result obtained in table 3.3. using Modified Turkey's method.

Table: 3.5. Contain 2048 observation with eight aberrant observations injected shows that Turkey's method detected the aberrant observations up to the seventh resolution level at the same location.

Table: 3.6. Confirms the result obtained in table 3.6. using Modified Turkey's method.

Table 3.7 provides information on simulated data of 512 with four aberrant observations injected . The aberrant observations were detected at first and second resolution, seventy five percent of the third(3 out of 4) and fifty percent (2 out of 4) of fourth resolution at α=0.05. At α=0.10, the aberrant observations were all detected up to the fourth resolution

Table 3.8 contains analysis of simulated data of 1024 with four aberrant observations injected. These aberrant observations were detected up to the fifth resolution level at α = 0.05 and 0.01 level of significance.

Table 3.9 provides information on the detection of eight aberrant observations injected in the simulated series of 2048 data set. It was obscured that these eight aberrant observations were detected up to the seventh resolution levels in both α – values of 0.05 and 0.10 respectively

4.1 SUMMARY OF FINDINGS

4.1.1 The non-parametric approach

In the non-parametric setting, containing three simulated series of 512, 1024 and 2048 observations in which four (43,-44, 41, 47), four (32, 41, 40,-37) and eight (45,-51, 50,-52, 48,-49, 47, 46) anomalous observations were injected into the series respectively. Turkey's and modified Turkey's methods were able to detect all the AOs in the three series.

4.1.2 The Parametric Approach

In the parametric approach, using the developed test statistics, it was observed that these AOs were detected at same location even when the series has been compressed to the fifth resolution level in series A, B and up to the seventh resolution level in series C at α =5% and 10% level of significance. It was also observed that the more the series, the more efficient wavelet method is in detecting these aberrant observations at more resolution levels.

4.2 Conclusion

In the non-parametric method, Turkeys method detected the anomalous observations in the three data sets as well as modified Turkey's method irrespective of the increase in the multiplier from 1,50 to 3.95. and at lower resolutions (when the data were compressed).

In the parametric method, the developed test statistic was also able to detect anomalous observations in all the three data sets, hence; it is as effective as the modified Turkeys and the Turkey's methods in the non-parametric method

4.3 Recommendations

Anomalous observations present in a data set cannot be determined apriori, it is recommended that every data set most importantly time series data be diagnosed for anomalous observations using the proposed test statistic which has been proved to be more efficient than other existing methods.

From both the parametric and non-parametric setting, it was discovered that even when data is compressed, wavelet analysis can be used to obtain the required information at same location, still preserving its properties, wavelet analysis could be used especially when the series is non-stationary and non-periodic. It is therefore recommended that wavelet analysis be used were data has to be compressed, where issue of stationary and period is not important.

4.4 Suggested Areas for Further Research.

This research work focused on aberrant observations detection in the frequency domain using wavelet analysis. After this work, some research areas for feature academic work has now been opened. They are to:

- Extend the detection of anomalous observations in the frequency domain using the Non- decimated wavelet transform.
- Comparing our results in Haar wavelet analysis objectively with what may be obtained using the Non-decimated wavelet transform.

Using multiple wavelet transform for the detection and modeling of anomalous observations in time series data.

REFERENCES

[1] Armando D. M., Gracia J. F. and Ramon M. R. (2003). A Practical Procedure to Estimate the shape Parameter in the Generalized Gaussian Distribution {\it ITESM Campus. Monterry}

[2] Daubechies, I. (1988). Orthonormal bases of compactly supported wavelet. {\it Communication Pure and SIAM}, XLI, 909-966.

[3] Daubechies, I. (1992). Ten lectures on wavelet. {\it SIAM}, Philadelphia.

[4] Daubechies, I. and Lagarias, J.C. (1992). Two-scale difference equations II: Local regularity, infinite products of matrices and fractals, {\it SIAM, J. Math. Ann}. 23, 1031-1079.

[5] Eckley, I.A. (2001). {\it Wavelet methods for Time Series and Spatial Data}. Ph.D. Thesis, University of Bristol, U.K.

[6] Eckley, I.A. and Nason, G.P. (2005). Computation of the Discrete Autocorrelation Wavelet Inner Product Matrix, {\it Statistics and Computing}, 15, 83-92.

[7] Graps A.(1995). In introduction to wavelets, {\it IEEEcomputer}, pages 50-61, Vol 2, issue 2, June, 1995.

[8] Nason, G.P. (2002). Choice of wavelet smoothness, primary resolution and threshold in wavelet shrinkage, {\it statistics and computing}, 129; 219-227.

[9] Nason, G.P. (2006). Stationary and non-stationary time series. Madder and S. eds., Statistics in volcanology, pp. 129-142. {\it Geological Society of London}, London.

ISSN 2250-3153

[10] Nason, G.P. and Sapatinas, T. (2005). Wavelet Packet Transfer Function Modeling of nonstationary time series. {\it Statistics and Computing}, 12, 19-229.

[11] Neill P. (2012). A robust, non parametric methodsto identify outliers and improve final yield and quality. {\it C.S MANTECHconference} April 23-26, 2012 Boston Massachusette.

[12] Percival, D. B. and Walden, A. T. (2000)wavelet method for time series Analysis, Cambridge university press, Cambridge.

AUTHORS

First Author – Aideyan Donald Osaro, P.D.S, B.Sc, M.Sc. Ph.D