Biomedical Statistics and Informatics
2022; 7(3): 41-48
http://www.sciencepublishinggroup.com/j/bsi
doi: 10.11648/j.bsi.20220703.11
ISSN: 2578-871X (Print); ISSN: 2578-8728 (Online)

Wavelet Analysis of Aberrant Observations in the Rate of Inflow of Patients in Some Diseaes in Kogi State, Nigeria

Aideyan Donald Osaro[1], Efuwape Biodun Tajudeen[2]

[1]Dept of Mathematical Sciences, Taraba State University, Jalingo, Nigeria

[2]Dept of Mathematical Sciences, Olabisi Onabanjo University, Ago-Iwoye, Nigeria

Email address:
daideyan2006@yahoo.com (A. D. Osaro)

To cite this article:
Aideyan Donald Osaro and Efuwape Biodun Tajudeen: Wavelet Analysis of Aberrant Observations in the Rate of Inflow of Patients in Some Diseaes in Kogi State, Nigeria. *Biomedical Statistics and Informatics*. Vol. 7, No. 3, 2022, pp. 41-48. doi: 10.11648/j.bsi.20220703.11

Received: February 12, 2022; **Accepted**: March 4, 2022; **Published**: July 28, 2022

Abstract: In recent years, the method of wavelet analysis has been opened to researchers. Wavelet analysis analyses data at different level of decomposition and can capture the characteristics of data series in all decomposition level. In this research work, data was collected on the medical records of the inflow of patients for medication on Malaria fever and Anemia from Grimard Catholic Hospital Anyigba, Kogi State, Nigeria (1993 to 2014). The data was analysed by wavelet methods to detect the aberrant observations over the period under study for the two diseases respectively using a proposed threshold. A total of ten and nine Aberrant Observations (AOs) were detected from the analysis of the original data collected on Malaria Fever and Anemia respectively. At the first and second level of decomposition (resolution), a total of seven and one AO(s) were respectively detected for both Malaria Fever analysis and Anemia analysis. The results obtained showed that the AOs detected in the analysis of the original data maintain the same or closely the same positions as that obtained from the analysis of the decomposed data for the two diseases. It was observed that the inflow of patients in the months of September, October and November into the hospital for medication on the two diseases were more. The Time plot for Malaria Fever and Anemia in the appendix respectively showed that there was no month that fewer patients reported to the hospital for medication.

Keywords: Wavelet, Decomposition, Resolution, Aberrant Observations, Diseases

1. Introduction

Wavelet analysis (also called wavelet theory or just wavelet) has attracted much attention recently in data processing. It has been successfully applied in many field such as transient data analysis, image analysis, communications systems and other data processing applications. Most of the data in practice are time domain data in their raw format. That is, whatever the data is measuring is a function of time. Wavelet analysis techniques provide multi-scale analysis of the data as a sum of orthogonal data corresponding to different time scales. So, it is called time-scale analysis. It provides multi-level analysis to analyse the data at different levels.

Section 2 looks at the research method of wavelet analysis which uses both resolution and location in analyzing data completely, defines Aberrant Observations (AOs) and how thesewill be detected using a proposed threshold which is the main goal of this paper. Section 3 discusses the analysis and interpretation of results while Section 4 dwell on the summary of findings, conclusion and recommendations.

1.1. Purpose of the Study

In many institutions where real data are collected over time, analysing the data to detect AOs has always not been in effect. Basically this is as a result of many of these institutions not knowing the importance of this. The purpose of this study is to use wavelet analysis to discover the rate and period within the year when these diseases occur more or less frequently.

1.2. Aim and Objectives of the Study

The aim of this study is to detect the rate of inflow of patients to Grimard Catholic Hospital Anyigba, Kogi State for medical attention on Malaria Fever and Anemia with the following objectives; To

1. Propose attest statistic or threshold for the detection of aberrant observations.
2. Use wavelet analysis to detect the period in which these diseases occur frequency (AOs) within the year.

1.3. Scope and Source of Data

The data collected for this research work is a primary data on Malaria Fever and Anemia from Grimard Catholic Hospital Anyigba, Kogi State Nigeria from September 1993 to December 2014 on monthly bases.

2. Research Methodology

2.1. Wavelet Analysis

Wavelets are fairly new family of basic functions that are used to express and approximate other functions. Wavelet coefficients are capable of revealing aspects of the data that other techniques might miss, aspects such as changes in variance, level changes, and discontinuities in functions. Thus, due to the essentialities of detecting AOs in signals (or a sequence of data), wavelet analysis is well suited and comes in handy for AO detection [1, 4, 10, 16, 20, 34].

The traditional way of analysing a data in the frequency domain is the well-known Fourier analysis which applies sinusoidal waves as the transformation filter [11, 13]. The main drawback of this transformation is that it cannot maintain the information of the time domain and will be unsuitable for data with irregular behaviour such as spikes or data breaks. The wavelet transformation adopts a basis of spatially localized functions as its transformation filter [3, 5]. Then based on wavelet filtering of the original data through shifting and dilations, the wavelet transformation can capture the characteristics of data series both in the frequency domain and the time domain. It is an excellent tool for the analysis of the non-stationary data showing time-localized discontinuities or abrupt changes. By wavelet multi-resolution analysis (MRA) which combines resolutions from both time and frequency domains [13, 19, 23] the data can be decomposed into different scales where the non-stationarity of the data can be analysed according to their own resolution levels: long run trends correspond to the low frequency resolution and the spikes such as the AOs can be captured in the high frequency resolution.

Aberrant Observations (AOs) are commonly encountered in data and their presence can seriously distort model identification, parameter estimation and forecasting. [12] defined an AO as “an outlying observation”, or is one that appears to deviate markedly from other members of a sample from which it occurs. In almost every (if not all) real data, there are presences of AOs: and most noticeably in large data sets, AOs are inevitable. AOs are described as observations which are unusual, but not necessarily errors. Detection of AOs helps reveal important and valuable information from large data sets. In the field of meteorology, for example, spatial AOs can be associated with disastrous natural events such as tornadoes, hurricanes, and forest fires.

One of such fields where detection of AOs can be applied is the medical field [25, 29]. The number of patients coming to the hospital for medical attention can be viewed to follow a fluctuating pattern over time. These fluctuations can be as a result of AOs. AOs as pointed out earlier, does not necessarily connotes errors but contain valuable information about the data hence, detecting the aberrant observations becomes paramount. Many method of AOs detection exist, such as statistical based AO detection, deviation based AO detection, wavelet based AO detection [26] and others. In this research work, the wavelet based AO detection is of interest.

Wavelet is a waveform of limited duration that has an average value of zero. Wavelet analysis allows data analysis with different resolution match to its scale [32]. It is used to analyze aspects like trends, break points, discontinuity at higher derivatives and self–symmetry, compression or denoising of the data without appreciable degradation [31].

The term *wavelets* is used to refer to a set of orthonormal basis functions generated by dilation and translation of a compactly supported scaling function (or father wavelet), Ø, and a mother wavelet, ψ, associated with an r-regular multi-resolution analysis of $L^2(\mathbb{R})$ (the space of square integrable function) [9, 15, 17]. A variety of different wavelet families now exist that combine compact support with various degrees of smoothness and numbers of vanishing moments, and these are now the most intensively used wavelet families in practical applications in statistics. Hence, many types of functions encountered in practice can be sparsely (i.e. parsimoniously) and uniquely represented in terms of a wavelet series. Wavelet bases are therefore not only useful by virtue of their special structure, but they may also be (and have been!) applied in a wide variety of contexts.

The special structure of wavelet bases may be appreciated by considering the generation of an orthonormal wavelet basis for functions g$\in L^2(\mathbb{R})$ (the space of square integrable real functions) [27].

Following the approach of [6] which is that most often adopted in applications of wavelets in statistics, we start with two related and specially chosen, mutually orthonormal, functions or parent wavelets: the scaling function, Ø, and the mother wavelet ψ. Other wavelets in the basis are then generated by translations of the scaling function Ø and dilations and translations of the mother wavelet ψ by using the relationships

$$\emptyset_{jo,k}(t)=2^{jo/2}\emptyset(2^{jo}t-k),$$

$$\psi_{jk}(t)=2^{j/2}\psi(2^{j}t-k),j=j_o,j_o+1,\ldots,k\in\mathbb{Z} \qquad (1)$$

For some fixed $j_o \in\mathbb{Z}$, where $\mathbb{Z}$ is the set of integers.

The scaling function Ø resembles a kernel function and the mother wavelet ψ is a well-localized oscillation (hence the name wavelet). A unit increase in j in expression (1) (i.e. dilation) has no effect on the scaling function ($\emptyset_{jo,k}$ has a fixed width) but packs the oscillations of $\psi_{j,k}$ into half the width (doubles its ‘frequency’ or, in strict wavelet terminology, its scale or resolution) [28]. A unit increase in k in expression (1) (i.e. translation) shifts the location of both

$\emptyset_{jo,k}$and $\psi_{j,k}$, the former by a fixed amount (2^{-jo}) and the latter by an amount proportional to its width (2^{-j}).

Given the above wavelet basis, a function g∈ $L^2(\mathbb{R})$ is then represented in a corresponding wavelet series as

$$g(t) = \sum_{k\in Z} c_{jo,k}\, \emptyset_{jo,k}(t) + \sum_{j=jo}^{\infty} \sum_{k\in Z} w_{j,k}\psi_{j,k}(t), \quad (2)$$

with$c_{jo,k}$=$< g, \emptyset_{jo,k} >$ and $w_{j,k}$=$< g, \psi_{j,k} >$, where $<.,.>$ is the standard L^2 inner product of two functions:

$$< g_1, g_2 > = \int_R g_1(t) g_2(t) \mathrm{dt}$$

The wavelet expansion (2) represents the function g as a series of successive approximations. The first approximation is achieved by the sequence of scaling terms $c_{jo,k}\emptyset_{jo,k}$ (each intuitively being a smoothed 'average' in the vicinity of $2^{-jo,k}$). The oscillating features which cannot be represented with sufficient accuracy in this way are approximated in 'frequency' and in correspondingly fine detail by sequence of wavelet terms $w_{j,k}\psi_{j,k}$ (each intuitively representing 'smooth wiggly structure' of 'frequency'2^j in vicinity of $2^{-j}k$).

In many practical situations, the function to be represented as a wavelet series may be defined to be zero outside a finite interval, such as the unit interval [0,1]. Adapting wavelets to a finite interval requires some modifications. The obvious approach of simply vanishing the underlying function outside the interval will introduce artificial discontinuities at the end points [30]. However, in practice the most commonly used approaches to adapting wavelet analysis to the interval are based on periodic, symmetric or anti-symmetric boundary handling.

For simplicity in exposition, we shall assume that we are working with periodized wavelet bases on [0,1].

$$\emptyset^p_{jo,k}(t) = \sum_k \emptyset_{jo,k}(t-k) \text{ and } \psi^p_{j,k} = \sum_k \psi_{j,k}(t-k), \text{ for } t\in[0,1],$$

Where $\emptyset_{jo,k}(t)$ and $\psi_{j,k}(t)$ are defined in (1) above.

For any $j_o \geq 0$, the collection $\{\emptyset^p_{jo,k}$,k=0,1,…,2^{jo}-1; $\psi^p_{j,k}$, j≥ $j_o \geq 0$, k=0,1,…,2^j-1$\}$ is then an orthonormal basis of L^2([0,1]). The superscript "p" will be suppressed from the notation for convenience.

The idea underlying such an approach is to express any function g∈ L^2([0,1]) in the form;

$$g(t) = \sum_{k=0}^{2^{jo}-1} c_{jo,k}\emptyset_{jo,k}(t) + \sum_{j=j_o}^{\infty}\sum_{k=0}^{2^j-1} w_{j,k}\psi_{j,k}(t), j_o \geq 0, \quad t\in[0,1],$$

where

$$c_{jo,k} = < g, \emptyset_{jo,k} > = \int_0^1 g(t)\, \emptyset_{jo,k}(t)dt,\ j_o \geq 0,\ k=0,1,\ldots,2^{jo}-1$$

and

$$w_{j,k} = < g, \psi_{j,k} > = \int_0^1 g(t)\, \psi_{j,k}(t)dt,\ j\geq j_o \geq 0,\ k=0,1,\ldots,2^j-1.$$

A usual assumption underlying the use of periodic wavelets is that the function to be expanded is assumed to be periodic. However, such an assumption is not always realistic and periodic wavelet exhibit a poor behaviour near the boundaries (they create high amplitude wavelet coefficients in the neighborhood of the boundaries when the function is not periodic). However, periodic wavelets are commonly used because the numerical implementation is particularly simple. While as [18] has pointed out, this computational simplification affects only a fixed number of wavelet coefficients at each level of resolution.

Wavelet analysis can be classified into the discrete wavelet transform and the continuous wavelet transform.

2.1.1. Wavelet Bases

The selection of any arbitrary pair of mutually orthogonal parent wavelets followed by the reproducing process (1) discussed above, will not automatically result in a basis for $L^2(\mathbb{R})$, let alone a 'good' basis in the sense of providing a parsimonious representation of functions in terms of their corresponding wavelet series clearly, the parent wavelets need to be specially chosen if that is to be the case. Different wavelet bases exist, such as the [6, 14, 21, 22, 33], etc.

The simplest wavelet basis for $L^2(\mathbb{R})$ is the Haar basis [14] which uses a parent couple given by

$$\emptyset(t) = \begin{cases} 1, 0 \leq t \leq 1 \\ 0, otherwise \end{cases}$$

$$\Psi(t) = \begin{cases} 1, 0 \leq t \leq 1/2 \\ -1, 1/2 \leq t \leq 1 \\ 0, otherwise \end{cases}$$

The Haar basis is applied to the discrete wavelet transform discussed above to obtain the discrete scaling coefficients, $c_{jo,k}$ and the discrete wavelet coefficients, $w_{j,k}$.

2.1.2. Multi-Resolution Analysis (MRA)

Multi-resolution analysis, as implied by its name, analyzes the data at different frequencies with different resolutions. Multi-resolution analysis is designed to give good time resolution and poor frequency resolution at high frequencies and good frequency resolution and poor time resolution at low frequencies [13]. This approach makes sense especially when the data at hand has high frequency components for short durations and low frequency components for long durations.

"Wavelet analysis examines the data at different frequencies with different resolutions. That is to say, it uses wider window for low frequency and uses narrower window for high frequency analysis. This feature especially works well for data whose high frequency components have short durations and low frequency components have long durations" [24].

2.1.3. Advantages of Wavelet

1. Denoising of the Signal or data: Approximate coefficients are containing low frequency components while detail coefficients contain the high frequency components [30]. When, we discard all high frequency information, we lose many of the information of original signal's sharpest features. Denoising is done by

using suitable approach called 'thresholding'. In this approach detail coefficients are discarded when it exceeds certain limit. Denoised signal is reconstructed by using both the coefficients.

2. Detecting long term evaluation: Wavelet analysis may be used to detect the overall tread of the signal. As the approximation level increases the trend becomes clearer. Trend represents the slowest part of the signal. In terms of wavelet analysis, as the scale increases, resolution decreases. So, it produces better estimate of unknown trend. In terms of frequency, successive approximations possess progressively less high frequency information. With the higher frequencies removed, what is left is the overall trend of the signal.
3. Splitting signal components: The wavelet transform is used to split the signal in terms of its detail and approximate coefficients. The approximate coefficients represent the outlines and the detail coefficients represent detailed information. The detail coefficients are used to notice high frequency components and approximate coefficients are used to notice low frequency components.
4. Detecting discontinuities and breakdown points: These analyses are used to know at what exact instance the signal change occurs i.e. site of change, type of change, amplitude of change and discontinuities. By using detail coefficients at different level we can identify that the measurement and state noise. So by using wavelet transform we are able to detect the break down points.
5. Multiscale analysis: Wavelet technique provides Multiscale analysis of the signal as a sum of orthogonal signals corresponding to different time scale. So it is time scale analysis.
6. Compression: The wavelet transform denoised the signal by applying appropriate thresholding rule. Thresholding means removing the coefficients which are responsible for noise. So because of reduction in coefficients the signal is compressed without any original signal degradation.

2.2. Thresholding

Donoho, D. L., et al. (1993) [7] propose a threshold τ based on the following result:

Result: Let Z_i be independently, identically distributed (iid) standard normal random variables. Define

$$A_n=\{max_{i=1,n}|Z_i|\leq \sqrt{2logn}\}.$$

Then

$$\pi_n=\mathrm{P}(A_n)\rightarrow 0,\ \mathrm{n}\rightarrow \infty$$

In addition, if $B_n(t)=\{max_{i=1,n}|Z_i|> t+\sqrt{2logn}\}$.

Then $\mathrm{P}(B_n(t))< e^{-\frac{t^2}{2}}$. That motivates the following threshold:

$$\tau^u=\hat{\sigma}\sqrt{2logn},$$

Which [7, 8] call *universal*. This threshold is one of the first proposed and provides an easy, fast, and automatic thresholding. The rationale is to remove all wavelet coefficients that are smaller than the expected maximum of an assumed independently and identically distributed normal noise sequence of given size. There are several possibilities for the estimator, $\hat{\sigma}$. Almost all methods involve the wavelet coefficients of the finest scale. The signal-to-noise ratio is smallest at high resolutions in a wavelet decomposition for almost all reasonably behaved signals.

2.2.1. Some Standard Estimators Are

1. $\hat{\sigma}^2=\frac{1}{\frac{N}{2}-1}\sum_{i=1}^{N/2}(d_{n-1,i}-\bar{d})^2$, or a more robust.
2. $\hat{\sigma}^2=\frac{1}{0.6745}\mathrm{MAD}(\{d_{n-1,i}, i=1,N/2\})$, where n-1 is the highest level.

2.2.2. Proposed Wavelet Threshold Method

$\tau=\sigma(2\log(n))^{1/2}$ is the threshold originally suggested by [2, 8, 29]. Whereas the proposed threshold is

$$\mathrm{Z}=\frac{X_j-\bar{X}_j}{\sigma^j\sqrt{2logn^j}}$$

Which can be obtained as follows

Given the data series $\{X_t\}$,

1. Calculate the mean of X_t.
2. Calculate the standard deviation σ of X_t.
3. Calculate Z as above and j =1,2,3,……..J representing the level of wavelet decomposition.
4. Perform step 1 to 3 on the original series and on different level of wavelet coefficient J of interest.

It should be noted that Z is assumed to follow a normal distribution and can have possible values in the range [-1,1]. Any observation outside this range is considered as an AO. Observations lesser than -1 are said to be AOs with lower values and observations greater than 1 are said to be AOs with higher values when compared with the threshold value. This is also true for the compressed data by wavelet analysis.

3. Analysis and Interpretation

In this chapter, a concise analysis of the data collected for this study is analyzed by wavelet methods. This analysis is carried out with the aid of a statistical package called R and an appropriate interpretation follows.

A complete rundown of the analysis by R is presented in the appendix.

3.1. Analysis for Malaria Fever

The results obtained from the analysis of the Malaria fever data at different resolution levels are summarized in the table below:

Table 1. *Results of the analysis of Malaria fever.*

Resolution level	AO figure	AO position	AO original value	Month of occurrence
256	1.065422297	27	179	Nov 1995
	1.212408775	38	194	Oct 1996
	1.094819593	39	182	Nov 1996
	1.055623199	62	178	Oct 1998
	1.290801563	86	202	Oct 2000
	1.369194352	98	210	Oct 2001
	1.143815086	111	187	Nov 2002
	1.447587140	122	218	Oct 2003
	1.222207874	135	195	Nov 2004
	1.104618692	146	183	Oct 2005
128	1.352158400	14	179	Nov 1995
	1.674957161	20	182	Nov 1996
	1.043394367	32	149	Nov 1998
	-1.160058050	49	210	Oct 2001
	1.464436230	56	187	Nov 2002
	-1.440752625	61	218	Oct 2003
	1.338123671	68	195	Nov 2004
64	1.03187710	22	202	Oct 2000

At the first level of decomposition, the set of the AO location is S={14, 20, 32, 49, 56, 61, 68}. These are the location of potential AOs.

3.2. Analysis for Anemia

The results obtained from the analysis of the Anemia data at different resolution levels are summarized in the table below:

Table 2. *Results of the analysis of Anemia.*

Resolution level	AO figure	AO position	AO original value	Date of occurrence
256	1.006763100	38	79	Oct 1996
	1.006763100	85	79	Sep 2000
	1.278808211	97	92	Sep 2001
	1.195102023	111	88	Nov 2002
	1.195102023	122	88	Oct 2003
	1.132322382	133	85	Sep 2004
	1.069542741	145	82	Sep 2005
	1.153248929	147	86	Nov 2005
	1.006763100	206	79	Oct 2010
128	1.352158400	14	59	Nov 1995
	1.674957161	20	71	Nov 1996
	1.043394367	32	69	Nov 1998
	-1.16005805	49	92	Sep 2001
	1.464436230	56	88	Nov 2002
	-1.440752625	61	88	Oct 2003
	1.338123671	68	49	Nov 2004
64	1.03187710	22	79	Sep 2000

At the first level of decomposition, the set of the AO location is S={14, 20,32,49,56,61,68}. These are the locations of potential AOs.

3.3. Interpretation

3.3.1. Malaria Fever

The set of the AO location for the original data is S={27,38,39,62,86,98,111,122,135,146}, corresponding to the observations 179, 194, 182, 178, 202, 210, 187, 218, 195 and 183 respectively. These observations can be traced to their various years and months of occurrence. Observation 179 occurred in November of 1995, observation 194 occurred in October of 1996, observation 182 occurred in November of 1996, observation 178 occurred in October of 1998, observation 202 occurred in October of 2000, observation 210 occurred in October of 2001, observation 187 occurred in November of 2002, observation 218 occurred in October of 2003, observation 195 occurred in November of 2004 and observation 183 occurred in October of 2005.

At the first level of decomposition, it can be observed that the aberrant observations from the compressed data on the original data are 179, 182, 149, 210, 187, 218 and 195 respectively. All of these aberrant observations are also picked out as AOs from the analysis of the original data except for the observation 149 which is close to the aberrant observation 178.

At the second level of decomposition, there is only one aberrant observation present. The AO is at location 22. Therefore, the location of the AO is 44 or 43 in the first level of decomposition and 85 or 86 (corresponding to location 43 in the first level of decomposition) or, 87 or 88

(corresponding to location 44 in the first level of decomposition) in the original data. Out of these points, location 86 corresponding to observation 202 in the original data is an AO.

3.3.2. Anemia

The set of the AO location for the original data is S={38,85,97,111,122,133,145,147,206}, corresponding to the observations 79, 79, 92, 88, 88, 85, 82, 86 and 79 respectively. These observations can be traced to their various years and months of occurrence. Observation 79 occurred in October of 1996, September of 2000 and October of 2010. Observation 92 occurred in September of 2001, observation 88 occurred in November of 2002 and October 2003, observation 85 occurred in September of 2004, observation 82 occurred in September of 2005 and observation 86 occurred in November of 2005.

At the first level of decomposition, it can be observed that the aberrant observations from the compressed data on the original data are 59, 71, 69, 92, 88, 88 and 49 respectively. The observations 92, 88, and 88 respectively are aberrant observations in the original data. However, observations 71 and 49 are not aberrant points in the original data but are close to the aberrant points 79 and 85 respectively. Furthermore, observations 59 and 69 which are picked as AOs are not outlying points in the original data and are farther away from an aberrant point in the original data.

At the second level of decomposition, there is only one aberrant observation present. The AO is at location 22. Therefore, the location of the AO is 44 or 43 in the first level of decomposition and 85 or 86 (corresponding to location 43 in the first level of decomposition) or, 87 or 88 (corresponding to location 44 in the first level of decomposition) in the original data. Out of these points, location 85 corresponding to observation 79 in the original data is an AO.

4. Summary of Findings, Conclusion and Recommendations

In this chapter, we present a summary of the analysis carried out, make conclusions based on our findings and recommend potential solutions to reduce the rate of infections of the disease, Malaria Fever and Anemia.

4.1. Summary of Findings

This research work was carried out on two diseases (namely; Malaria fever and Anemia), with data collected from Grimard Catholic Hospital Anyigba, Kogi State from September 1993 to December 2014. Wavelet based approach to outlier detection was carried out on the data as presented in chapter four of this research work. This approach was carried out separately on each of the disease.

The results obtained from the analysis of Malaria fever indicates that there are ten aberrant points when the original data was analyzed. The months and years of their occurrence were stated in the interpretation following the analysis, with October 2003 recording the most inflow of patients for medication in the hospital. The first level of decomposition on the original data picked seven aberrant points, with six of them exactly the same as that from the analysis on the original data. The second level of decomposition on the original data picked only one AO which was traced to be observation 202 in the original data.

The result obtained from the analysis of Anemia indicated that there were nine AOs for the analysis of the original data. The months and corresponding years of occurrence were duly recorded in the interpretation that followed it. The analysis from the first level decomposition of the original data picked seven aberrant points, where three of them are exactly the same as that obtained from the analysis of the original data. Two of those points are close to an aberrant observation in the original data while the other two are not. The second level of decomposition analysis on the original data yielded only one aberrant location which was traced to be observation 79 in September of 2000.

For both analyses, there are no aberrant locations to the left (that is no negative aberrant position) on the analyzed original data. It then means that there is not a month that fewer persons than expected came for medication over the observed years.

4.2. Conclusion

It is evident from the analysis of this study that the inflow of patients for medication on Malaria fever and Anemia are similar. There was no time that lower number of patients came in for medication over the period under study. This can be vividly seen from Figures 1 and 2 in the appendix, again it can be observed that the aberrant points fell between the months September, October and November.

The analysis on Malaria fever showed that the aberrant observations are only those where more patients came in for medication. Since the original data at different level of decomposition maintain the same or approximately the same location of AO with that of the analysis for the original data, the importance of wavelet via the multi-resolution analysis has been established. The AOs were detected in the months of October and November. This is due to the fact that water logs can be found in every place possible since there is no much rain at that time for the water to run. Mosquitoes can easily breed in such water if not treated or drained and thereby multiplying the risk of infection by the malaria parasite.

The analysis on Anemia also indicates only those where more patients came in for medication as AOs. By implication, no fewer numbers of patients came in for medication at any time under the stipulated time of the study. Since the original data at different level of decomposition maintain the same or approximately the same location of AO with that of the analysis for the original data, the practicability of wavelet analysis via multi-resolution analysis is emphasized. The AOs were detected in the months of September, October and November. This could be as a

result of poor dietary intake and/or absorption of iron or absorption of vitamin (A, B-12, folic acid and others) because of poverty, malaria, reproduction (excessive blood loss during menstruation, delivery and postpartum period; too many pregnancies), bacterial or viral infections (peptic ulcers, gastritis, diarrhea), contamination by heavy metals (leads) in the environment, lack of access to services like sanitation services, trained birth attendants to manage bleeding during delivery, etc.

However, from the study it is evident that in more recent year, there is a considerable decline in the inflow of patients to the hospital for medication owing to the fact that health programs for awareness of these diseases are in effect at least to some extent.

4.3. Recommendation

In situations where there are data that are transient and dyadic which is always the with medical data, wavelet method of multi-resolution analysis can be adapted to detect AOs in such data to reduce the cumbersomeness.

Appendix

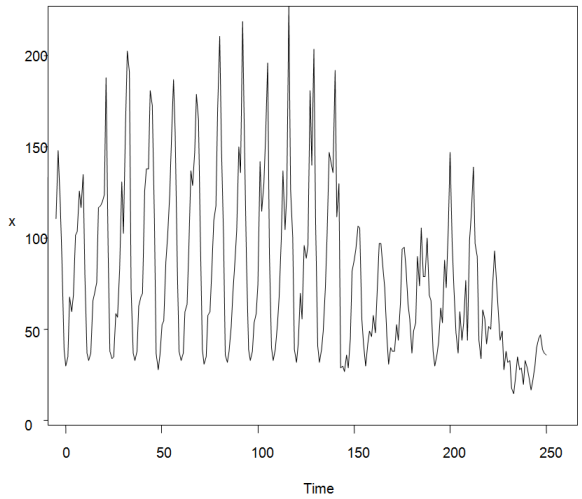

Figure 1. *Time plot for Malaria.*

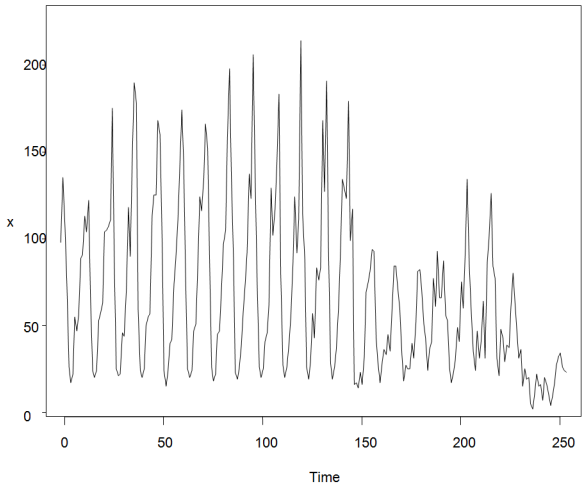

Figure 2. *Time plot for Anemia.*

References

[1] Barnet, V. & Lewis, T. (1994), Outliers in statistical data, John Wiley, ISBN 0-471-93094-6, Chichester.

[2] Bilen, C., Huzurbazar, S. (2002), "Wavelet-Based Detection of Outliers in Time Series", Journal of Computational and Graphical Statistics, Volume 11, Number 2, 311-327.

[3] Caraballo H, "Emergency department management of mosquito-borne illness: malaria, dengue, and west nile virus", 2014.

[4] Chen, Z.; Fu, A. & Tang, J., (2002). Detection of outliered Patterns, Dept. of CSE, Chinese University of Hong Kong.

[5] Combes, J. M.; Grossman A. & Tchamitchian P. (1989) Wavelets: Time-Frequency Methods and Phase Space, Second Edition, Springer-Verlag, New York.

[6] Daubechies, I. (1992). Ten Lectures on Wavelets. Philadelphia, Pa.: SIAM.

[7] Donoho, D. L., Johnstone I. M, Kerkyacharian G, and Picard D, 1993, April, Density estimation by wavelet thresholding, Preprint, Department of Statistics, Stanford University.

[8] Donoho D. L. and Johnstone I. M. (1994). "Ideal Spatial Adaptation by Wavelet Shrinkage", Biometrika, Vol. 81, pp. 425-455.

[9] Ziou D and Tabbone S. Edge detection techniques: Anoverview. International Journal of Pattern Recognition and Image Analysis, 8 (4): 537–559, 1998.

[10] Eric Poulin and Colin Yu, 2005: Outlier Detection & Analysis.

[11] Grossmann, A. & Morlet, J. (1984) Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape, SIAM J. Math. Anal. Vol. 15, N° 4, pp. 723-736.

[12] Grubbs, F. E. (1969), Procedures for detecting outlying observations in samples, Technometrics 11, pp. 1-21.

[13] G. Sheikholeslami, S. Chatterjee, and A. Zhang. Wavecluster: A multi-resolution clustering approach for verylarge spatial databases. In Proceedings of 24th VLDB Conference, 1998.

[14] Haar, A. (1910). Zur Theorie der orthogonal unktionen systeme. Math. Ann., 69, 331–371.

[15] Han, J. & Kamber M. (2001) Data Minings Concepts and Techniques, Morgan Kauffman Publishers.

[16] Hawkins, D. (1980), Identification of Outliers, Chapman and Hall, London.

[17] Hodge, V. J. (2004), A survey of outlier detection methodologies, Kluver Academic Publishers, Netherlands, January 2004.

[18] Johnstone, I. (1994). Minimax bayes, asymptotic minimax and sparse wavelet priors. In Statistical Decision Theory and Related Topics, V (S. Gupta andJ. Berger, eds.). Springer-Verlag, New York, 303–326. MR1286310.

[19] Kantardzic, M. (2003). Data mining Concepts, Models, Methods and Algorithms. IEEE Transactions on neural networks, Vol. 14, N. 2, March 2003.

[20] Kern, M.; Preimesberger, T; Allesch, M.; Pail, R.; Bouman, J. & Koop, R. (2005). Outlier detection algorithms and their performance in GOCE gravity field processing, Journal of Geodesy, Vol. 78, pp. 509-519, January 2005.

[21] Mallat, S. (1999). A Wavelet Tour of Signal Processing. 2nd ed. Academic Press, San Diego. MR1614527.

[22] Moore, D. S. and McCabe G. P. (1999), Introduction to the Practice of Statistics., Freeman & Company.

[23] Ramasmawy R.; Rastogi R. & Kyuseok S. (2000). Efficient algorithms for mining outliers from large data sets. Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 427-438, ISBN 1-58113-217-4, Dallas, Texas, United States.

[24] Polikar, R. 2001. The wavelet Tutorial. Fundamental concepts and an overview of the wavelet theory.

[25] Rodak, Bernadette F. (2007). Hematslogy: clinical principles and applications (3rd ed.). Philadelphia. Saunders.

[26] Shekhar S, Huang Y, Wu W, Lu C, and Chawla S. What's special about spatial data mining: three case studies. In Data Mining for Scientific and Engineering Applications. V. Kumar, R. Grossman, C. Kamath, R. Namburu (eds.), 2001.

[27] Shekhar S, Lu C, and Zhang P. Detecting graph-basedspatial outliers. International Journal of Intelligent Data Analysis (IDA), 6 (5): 451–468, 2002.

[28] Li T, Li Q, Zhu S, and Ogihara M. A survey on wavelet applications in data mining. SIGKDD Explorations, 4 (2): 49–67, 2002.

[29] Wang, Y. (1995) Jump and sharp cusp detection by wavelets. Biometrika, Vol. 82, pp. 385-397.

[30] Weiss, L. G. (1994). Wavelets and Wideband Correlation Processing, Signal Processing magazine IEEE, Vol. 11, pp. 13-32, ISSN 1053-5888, January 1994.

[31] WHO. March 2014: Malaria Fact sheet W "94".

[32] World Health Organisation (2010). Guidelines for the treatment of malaria (2nd ed.). Geneva: World Health Organisation P. ix. ISBN 9789241547925.

[33] Meyer Y. Wavelet and Operators. Cambridge University Press, 1992.

[34] Yu, D.; Sheikholeslami G. & Zhang, A. (2002) Find Out: Finding Outliers in Very Large Datasets. Knowledge and Informations Systems, vol. 4, pp. 387-412, Springer-Verlag, London.