2020-01-01

Lévy Processes: Characterizing Volcanic And Financial Time Series

Peter Kwadwo Asante
University of Texas at El Paso

LÉVY PROCESSES: CHARACTERIZING VOLCANIC AND FINANCIAL TIME SERIES

PETER KWADWO ASANTE

Master's Program in Computational Science

APPROVED:

Maria C. Mariani, Ph.D., Chair

Thompson Sarkodie-Gyan, Ph.D.

Granville Sewell, Ph.D.

Elsa Villa, Ph.D.

Stephen Crites, Ph.D.
Dean of the Graduate School

to my

Wife, Parents and Siblings

with love

LÉVY PROCESSES: CHARACTERIZING VOLCANIC AND FINANCIAL TIME SERIES

by

PETER KWADWO ASANTE, M.S.

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Computational Science Program

THE UNIVERSITY OF TEXAS AT EL PASO

May 2020

Abstract

In this work, we use the Diffusion Entropy Analysis (DEA) to analyze and detect the scaling properties of time series from both emerging and well established markets as well as volcanic eruptions recorded by a seismic station, both financial and volcanic time series data are known to have high frequencies (i.e they are collected at an extremely fine scale). The objective is to determine the characterization i.e whether they follow a Gaussian or Lévy distribution. If they do follow a Lévy distribution we are then interested in finding if they are characterized by a Lévy walk which has a finite second moment or a Lévy flight which has an infinite second moment. We also seek to establish the existence of long-range correlations in these time series.That is we seek to determine if both time series are persistent (i.e have long-range correlation), anti-persistent or random.

The results obtained from the DEA technique are compared with the Hurst R/S analysis and Detrended Fluctuation Analysis (DFA) methodologies. We conclude that given the scaling exponents δ derived from the DEA and H, α derived from the Hurst R/S analysis and DFA respectively, if $0.5 < H, \alpha, \delta < 1$ the time series is said to exhibit long-range correlations and if $0 < H, \alpha, \delta < 0.5$ the time series is said to be anti-persistent. Also for characterization, if δ is related to H or α by the relation $\delta = \dfrac{1}{3 - 2(H, \alpha)}$, the time series is characterized by a Lévy walk. If $\delta = (H, \alpha)$, the time series may be characterized by Fractional Brownian Motion (FBM) (i.e the time series is random), and finally if $\delta \neq (H, \alpha)$, the time series cannot be characterized by an FBM and this implies that the time series has an infinite second moment and is thus characterized by a Lévy flight.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The collection and analysis of time series data is a very important area of research. Inferences drawn from these data sets have helped in forecasting as well as various industrial product improvements. One important inference usually sought for is whether the time series exhibits persistence (long-range correlations), randomness or anti-persistence. Long-range correlations (also refered to as Long-memory effects or Long-range dependence (these would be used interchangeably)) refers to the slow decay of the temporal or spatial correlation function defined as

$$\gamma_{xy}(\delta) = \langle X(t)Y(t+\delta)\rangle. \tag{1.1}$$

A time series data which exhibits long-range correlations implies that the evolution of the system is affected by previous system states over long periods of time ([1],[2]-[5]). This makes the need to determine long-range correlations in time series data very important for various fields. However to determine the existence of long-range correlations using the formula in equation 1.1 poses challenges due to its sensitivity to noise. This in addition to other factors has pushed research into the development of a number of scaling methods ([6]-[11],[12]-[15],[2]-[5]).

Various scaling methods exist and have been utilized by many researchers in detecting the persistence or anti-persistence in time series. Most notable applications are in financial and geophysical time series . Some examples of these scaling methods are the Rescaled Range Analysis (R/S), the Detrended Fluctuation Analysis (DFA), the Relative Dispersion Analysis (RDA) and the fairly recent Diffusion Entropy Analysis (DEA) which was developed by Scafetta ([6]-[8]). Scafetta used the DEA to detect the scaling behavior of DNA sequences. The R/S, DFA, and RDA are examples of variance scaling methods and their scaling exponent is called the Hurst exponent, named after Hurst who first studied it in hydrology while the DEA on the other hand is a pdf scaling method.

The variance scaling methods however encounter various challenges when faced with time series data that exhibit anomalous behaviors. The R/S analysis in particular is usually unable to detect correctly the scaling exponent of non-stationary time series data while the DFA is known to overestimate the scaling exponent. Thus two short comings of these variance scaling methods are their inability to detect the exact value of the exponent though they may be able to detect the scale invariance and their unavailability for processes with infinite variances like the Lévy flight [6].

This is what makes the DEA, our main focus of this paper an important method in detecting the scaling exponent within a time series data. The DEA detects the scaling parameter δ using the pdf of the diffusion process derived from the time series. The advantage of DEA over the variance scaling methods is that it is able to establish the possible existence of scaling in time series data with normal or anomalous properties efficiently without any data alteration due to detrending as well as being available for processes with infinite variances ([6]-[8], [16]). Thus the DEA overcomes the shortcommings of the variance based methods.

Knowing the characterization is important in helping us understand the scaling behavior determined by a scaling method due to a possibility of wrong conclusion based on some underlining assumptions that may be present in the model. If the time series follows a Gaussian distribution, then the traditional Scaling methods are able to get the correct scaling due the the underlining Gaussian assumption in their models. If it follows a Lévy walk, then the traditional methods may produce correct conclusion with regards to long-range dependence but would not get the correct scaling exponent [4]. And Finally if the time series follow a Lévy flight, then the traditional methods are neither able to get correct scaling or conclusion due to the fact that Lévy flights have infinite variances [6].

Thus knowing the characterization of the time series may help in making informed decisions pertaining to the choice of forecasting models one would use for the purpose of future analysis.

Researches focused on long-range correlations have made it possible to gain more insight into long range evolution patterns of complex and chaotic occurrences both in nature (geophysical time series) and other equally important fields including financial markets, traffic analysis, bioengineering, and others. The results from these researches have provided various approaches to minimize risk and forecast or predict future dynamical trends ([6]-[9], [25]).

In this study we consider several financial time series data as well as some geophysical time series data and analyze their long-range correlations using R/S analysis, the DFA and the DEA. The continuous time-varying Lévy process is effective for capturing the stochastic volatility (SV) and fat tails of data distribution. It is known that the volatilities of high frequency data are correlated, and they vary stochastically over time. We seek to determine the characterization of the time series data (i.e whether it follows the Gaussian or Lévy distribution) by comparing the relation between the scaling exponent derived with the R/S and DFA against that of the DEA. This Thesis is organized as follows: Chapter one introduces the subject matter, a review of relevant literature is presented in chapter 2, in chapter 3 we introduce the R/S and DFA scaling methods and give a detailed background of the DEA with the procedure used to detect the scaling exponent (δ) for both a stationary and non-stationary time series data. In chapter 4 we present the data used for analysis and look at it's stationarity. Numerical results are shown in chapter 5 and chapter 6 presents our conclusions and future work. Figures obtained are presented in the appendixes.

Chapter 2

Lévy Processes And Scaling Methods Applied To Time Series

In this section a review of some existing scaling methods are given as well as some areas of research in which these methods have been applied. Also we look at some similarities and differences of the Lévy walk and the Lévy flight in bounded domains.

2.1 Lévy Walks and Lévy Flights

We begin this section with a definition of the Lévy process. We further lay out some similarities and dissimilarities of the Lévy flight and the Lévy walk. As this work explores the characterization of time series analysis using scaling methods, we would soon observe that the characterization of the time series gives an idea as to why some scaling methods explored in this work perform in certain ways when applied to a particular time series data.

Definition 2.1.1. *A Lévy process is a stochastic process* $\{X_t : t \geq 0\}$ *on* $\mathbf{R}^n$ *if the following conditions are satisfied.*

1 For any choice of $n \geq 1$ *and* $0 \leq t_0 < t_1 < ... < t_n$*, the random variables* X_{t_0}, $X_{t_1} - X_{t_0}$, $X_{t_2} - X_{t_1}$, ..., $X_{t_n} - X_{t_{n-1}}$ *are independent. The process has independent increments.*

2 $X_0 = 0$

3 The distribution of $X_{s+t} - X_s$ *does not depend on* s *therefore the process has the stationary increments.*

4 It is stochastically continuous.

5 There is a $\Omega_0 \in F(\sigma - algebra)$ *with* $\mathbb{P}[\Omega_0] = 1$ *such that for every* $\omega \in \Omega_0$, $X_t(\omega)$ *is right-continuous on* $t \geq 0$ *and has left limits in* $t > 0$

Lévy flights and Lévy walks serve as two paradigms of random walks resembling common features but also bearing differences [29]. They are two well known stochastic models that have been shown to exhibit anomalous diffusion. By analyzing two models of stochastic motion in bounded domains, Bartlomiej et al., [29] showed that both Lévy flights and Lévy walks assume a random walker performs long-jumps distributed according to a heavy-tailed power law-density. However Lévy walks showed continuous trajectories and finite velocity while Lévy flights showed discontinuous trajectories and infinite propagation velocity. This show's some common features as well as fundamental differences between the two.
Though Lévy walks and Lévy flights were shown to have such fundamental difference it was further found that Lévy flights can serve as an approximation to Lévy walks with an improper prediction of the moments of the jump length distribution.

2.2 Some Applications

In this section we look at some example areas where Lévy walks and Lévy flights have been applied in recent researches.

2.2.1 Volcanic Eruptions

We now present a research work on Volcano-seismic data using Lévy flights and Wavelet techniques.

Overview

In this work Beccar-Varela et al. looked at the α scaling parameter of Volcano-seismic data collected from a seismic station. The aim of this work was to detect the scaling properties of the volcanic eruptions by looking at the scaling exponent α. An α value less than 2.0 indicates the evolution of the released energy exhibits long-range dependence [5].

Conclusion

Based on the results obtained from the analysis made, the α exponents for each volcanic eruption was found to be less than 2.0 which indicated that the evolution of the released data exhibited long-range dependence. This result implies that current information is highly correlated with past information at different levels, thus helping in the prediction of future volcanic activity in that region.

2.2.2 Transportation

Overview

It is by no means a surprise that there is ongoing research on transportation and how it can be improved owing to the fact that it has become an integral part of society. Whether you walk, take public transport or drive a personal car, you are prone to seek for the best possible means to arrive on time with less stress and delays.
Pavement maintenance is thus an issue that many government agencies and consulting companies must deal with.
This work focused on the ability to maintain an in-service pavement structure in an acceptable condition from the structural and functional points in relation to many factors which often are not explicit and change over time [32].
In this work Mariani et al. applied the normalized truncated Lévy walk (TLW) to flexible pavement performance so as to forecast the change in serviceability level offered by the structure of pavements together with traffic and climate conditions.

Results and Conclusion

The numerical results obtained by the authors in this work showed that the TLW offered an alternative in representing pavement serviceability trends. They concluded that the TLW function parameters and time gaps could be tailored to represent the stochastic characteristics of the factor interaction determining pavement degradation, maintenance policies, and recurring maintenance programs [32]. Such information could then be used to describe lifelong responses of pavement structures by using larger data sets collected over longer periods of time.

2.3 Scaling Methods

Evidence of long-range dependence has been shown to be strong in many time series data collected from various fields. Being able to detect this phenomenon is a result of the application of scaling methods on these time series. Hurst was a pioneer in this field of scaling detection and as such it is no surprise that the exponent of scaling was named after him by another well known figure in this field, Mandelbrot.

In this section we present a brief history of long-range dependence and the emergence of scaling methods as well as some interesting applications to some time series data from various fields.

2.3.1 Brief history

In the wake of the industrial revolution in the nineteenth century, there arose a need to build large scale reservoirs formed by damming of river valleys. The natural solution to this problem was the a dam that will never overflow or empty. Rippl (1883) gave a compelling solution to this problem, however this was compromised by a requirement to know or assume the future variability of the river flows [30]. Hazen (1914) discovered the breakthrough which brought about the birth of stochastic hydrology by using the simplest model which happens to be iid Gaussian process.

In 1965, Hurst introduced a method for studying fractal properties in his book, Long-Term Storage:An experimental study. He developed this method as he studied the water storage of the nile river in order to design a reservoir which never overflows or empties given the record of observed discharge. Mandelbrot and Wallis, 1969 will later coin the word "Ideal Dam" for such a reservoir. Hurst's work would usher in a wave of interest as well as controversy. Mandelbrot later introduced his first stationary model-fractional Gaussian noise (FGN), which could explain Hursts phenomenon after almost a decade of controversy [30]. Long-range dependence was later then incorporated via a fractional differencing parameter d, into the traditional ARMA(p,q) models, through Hosking and Granger's ARFIMA(p,d,q) model [30].

2.4 Some aplications of Scaling Methods

This section presents some applications of scaling methods in various fields. We present some research works, both recent and old and briefly discuss the result and conclusion from the work.

2.4.1 Hydrology

Overview

Hurst (1951) in this study used a method similar to Rippl's. In his method Hurst analysed a particular statistic of the cumulative flows of rivers overtime called the "adjusted range", R.

Conclusion

After examining 690 different time series, covering 75 different geophysical phenomena spanning varying quanties as river levels, rainfall, temperature, atmospheric pressure, tree rings, mud sediment thickness and sunspots, Hurst found out that in each case the statistic behaved as $R/S(n) \propto n^k$.

He then estimated the variable k and found it to be to be approximately normal with mean 0.72 and standard deviation 0.006. His estimation approach for k was however found to be derived from an inappropriate analysis, namely by assuming a known constant of proportionality which was the assymptotic law that $R/S(n) = (n/2)^k$ .This was later addressed by Mandelbrot when he dropped the fixed point assumption by Hurst and performed a two-parameter log regression to obtain slope parameter k [30].

Hursts work showed that under the independent Gaussian assumption the exponent $k = 0.5$, which implied that contemporary hydrological models did not agree with empirical evidence. The value of the exponent k which Hurst derived was greater than 0.5, and suggested "long-range dependence" of the time series thereby sparking a debate in the hydrology community owing to the implication's it had on Hazen's model which assumed an iid Gaussian process. The discrepancy detected in Hurst's work with regards to the available theory at that time and his empirical result would later be known as the "Hurst Phenomenon" [30].

2.4.2 Stock Markets/Finance

Overview

In past years there have been recordings of stock market crashes in which the behavior of the market prior could not predict it, an example being the infamous global financial recession in 2008. This increased the interest in researches aimed at providing approaches to predict potential crushes based on past and current data.

Assuming normal market conditions, we expect all participating agents to have diverse views which translates to pricing processes not exhibiting long-memory effects. It is known

though that under normal market conditions when working with high-frequency data the price process is far from the log-normal specification [31].
In this work Barany et al. used scaling methods to detect market crashes by analysing long-range dependence using high-frequency data collected minute by minute for several stock markets. In addition they looked at the relationship between the Lévy parameter α and the resulting H parameter characterizing the self-similar property. Data used in this work was collected from stock markets in the areas of entertainment, technology, retail, oil and financial sectors.

Conclusion

The results of this work showed that most of the stock market data studied exhibited long memory effects, of particular interest was the fact that on normal trading days devoid of major events the data exhibited long-memory effects. However in the presence of a stock market crash, the parameter estimating the long-range dependence seems to increase and approach a random behavior during the crash. The results further found that after the crash the processes return to exhibiting long-range dependence[2]. This was a very interesting finding by the authors of this work which evidently made it possible to predict a possible crash of a market by looking at the behavior of the crash parameter.

2.4.3 DNA Sequencing

Overview

The recent progress in experimental techniques of molecular genomics has made available a wealth of genome data [8]. In this work Scafetta et al. sought to address the problem of the statistical analysis of a time series generated by complex dynamics using Diffusion Entropy Analysis (DEA) a fairly recent scaling method introduced by Scafetta which unlike the traditional scaling methods uses the Shannon entropy of the diffusion process derived from the time series to estimate the correct scaling exponent.

They looked at time series data derived from DNA sequences, considering both coding and non-coding sequences. Since pioneer papers focused on the controversial issue of whether the property of long-range dependence was shared by both coding and non-coding sequences.

This controversy could be attributed to the limitations of the traditional methods, which made it unable to detect the right scaling for time series that exhibited anomalous behaviours [8]. This paper thus sought among other things to show that the DEA unlike the traditional scaling methods is able to detect the correct scaling of a time series regardless of anomalous behavior.

Conclusion

The application of the DEA in conjunction to the traditional methods showed that the traditional methods were prone to produce misleading conclusions and that both the coding and non coding DNA sequences generate Lévy statistic in the long-time limit.

The results obtained in this work proved to be ground breaking, essentially revealing the reason behind the controversy concerning coding and non coding DNA sequences and providing a way to characterize time series.

This is what motivates the current research being done with financial and volcanic time series, by taking it further to determine if the time series characterized by a Lévy process is in particular characterized by a Lévy flight or a Lévy walk since there is a fundamental difference in these two forms of random walk as mentioned above.

2.5 Time Series data Characterization: Gaussian or Lévy (Lévy flight or Lévy walk)

In the past sections we have looked at the Lévy walk and Lévy flights with some applications as well as some interesting applications of scaling methods. It is clear how important Lévy processes are in various researches in time series analysis of data and it's forcasting.

Since the traditional scaling methods have an underling Gaussian assumption it presents a possiblility of drawing wrong conclusions when used to derive the scaling exponents of time series that are characterized by a Lévy distribution. This is what motivated Scaffeta to introduce the Diffusion Entropy Analysis (DEA).

What we seek to address in this work is that for time series that have Long-memory effects and follow Lévy distribution are they characterized by a Lévy walk or a Lévy flight. This question is important to answer since the Lévy walk is known to have a finite second moment while the Lévy flight has an infinite second moment.

Chapter 3

Methods For Determining Long-Range Correlations In Time Series

3.1 Variance Scaling Methods

In this section we briefly introduce the Rescaled Range Analysis and the Detrended Fluctuation Analysis. The Diffusion Entropy Analysis is then discussed with more detail.

3.1.1 Rescaled Range Analysis

The idea of the Rescaled-Range analysis (R/S) was presented by Hurst in the framework of his study on the long-run variations of the water level of the Nile river [17]. It has become very popular since then, and has been applied to a wide range of disciplines, including traffic analysis, bioengineering, physics, geology, biology and geophysics.

The name H for the parameter derived from this technique was coined by Mandelbrot in tribute to the hydrologist Hurst and the mathematician Holder. The parameter H also known as index of dependence represents the relative trend of a time series and always lies between 0 and 1, it is equal to $\frac{1}{2}$ in the case of processes with independent increments. Of particular interest for our work is the case in which $0.5 < H < 1$ since it is an indicator of long-range correlations.

3.1.2 Detrended Fluctuation Analysis

In order to study the self-similarity and long-range dependence of time series Peng et.al [18] proposed the Detrended Fluctuation Analysis (DFA) while examining a series of DNA nucleotides. From the moment it was proposed to date, DFA has become a widely used method for the determination of fractal scaling properties and the detection of long-range correlations in non-stationary time series. It has been applied for example in biology, meteorology, geophysics and economics ([18]-[24]).

The principal advantage of the DFA lies in its ability to differentiate the intrinsic auto-correlations of the time series from those imposed by non-stationary external trends. That is, the method focuses on the intrinsic structure of the correlations of market fluctuations at different time scales, leaving aside non-stationary trends.

The application of the DFA method allows obtaining a scale exponent α from estimating the slope of function $F(s)$ that measures the mean square deviation from an optimal linear approximation around the trend signal in segments of length s. The fluctuation function vs s behaves as a power law. Therefore it is possible to compute the value of the exponent α from the slope of the function in a log-log scale plot of $F(s)$ vs s. The DFA exponent α and the Hurst parameter H are related by

$$H = \begin{cases} \alpha & \text{if } 0 < \alpha < 1 \\ \alpha - 1 & \text{if } \alpha \geqq 1 \end{cases} \tag{3.1}$$

However, due to its sensitivity to abnormal values in the series, the rescaled range analysis method is not suitable for analyzing long-range auto-correlation for non-stationary series.

3.1.3 Diffusion Entropy Analysis

Based on the direct evaluation of the Shannon entropy ([6]-[8],[13],[14]), the DEA is a pdf scaling method which perceives the numbers in a time series as the trajectory of a diffusion process [12].

The scaling property for the stationary time series takes the form

$$p(x,t) = \frac{1}{t^{\delta}} F(\frac{x}{t^{\delta}}). \tag{3.2}$$

where x denotes the diffusion variable, $p(x,t)$ is its probability density function (pdf) at time t, and $0 < \delta < 1$ is the scaling exponent.

The scaling property for the non-stationary time series takes the form

$$p(x,t) = \frac{1}{t^{\delta(t)}} F(\frac{x}{t^{\delta(t)}}). \tag{3.3}$$

As derived in ([6]-[7]), a diffusion process generated by Lévy walk is characterized by the following relation:

$$\delta = \frac{1}{3 - 2(H, \alpha)} \tag{3.4}$$

If $\delta = (H, \alpha)$, the time series can be characterized by Fractional Brownian Motion (FBM), since the variance methods are based subtly on the Gaussian assumption ([1], [17]). However if $\delta \neq (H, \alpha)$, and equation (3.4) holds true, the noise can be characterized by Lévy statistics in particular a Lévy walk. (H, α) in Eq.(3.4) refers to the scaling exponent derived from the two variance scaling methods used in this work. Now if $\delta \neq (H, \alpha)$ and equation (3.4) does not hold, then the noise can be characterized by a Lévy flight.

3.1.4 Estimation Procedure

In this subsection, we describe the estimation technique for the scaling exponent, δ. We first present a brief background on the Shannon Entropy that is used for estimating δ.

The Shannon Entropy

The concept of entropy was developed by Rudolph Clausius in 1865, a few years after he stated the laws of thermodynamics ([25] -[26]). The entropy is an indicator of the lack of

information about the measure of an event that occurs with propability p [25].
Other types of entropies are the Kolmogorov-sinai entropy, the Renyi entropy and the Tsallis entropy ([6]- [8], [25]). The Shannon entropy measures information of a probability distribution as follows:

$$S(t) = -\sum_{1}^{N} p_i \log p_i \tag{3.5}$$

The summation is replaced by the integral in the case of continuous probability distributions. The above equation is used to derive the log equation that will be used to determine the DEA δ scaling. We present below the process for estimating δ:

- The time series data is first transformed into a diffusion process.

- Shannon's entropy of the diffusion process is calculated. A log-linear equation or log-quadratic equation is derived from the Shannon entropy by substituting equation 3.2 and 3.3 respectively. Simplifying the result from the substitutions, we have the following relation for stationary time series:

$$S(t) = A + \delta ln(t) \tag{3.6}$$

For the non-stationary series, the relation is as follows:

$$S(t) = A + \delta(t)\tau \tag{3.7}$$

where $\delta(t) = \delta_0 + \eta \log(t)$ and $\tau = \log(t)$ with $\eta \log(t) < 1 - \delta_0$. After some simplifications, equation 3.7 becomes

$$S(t) = A + (\delta_0 - K)\log(t) + (1 - \delta_0)(\log(t))^2 \tag{3.8}$$

where $K < 0$ and $\delta_0 \equiv \delta$ from the stationary pdf. Thus, by fitting a log-quadratic model in the non-stationary series and a log-linear model in the stationary series we are able to determine the δ (δ_0) scaling. At $t = 1$, it is clear that the constant A in both equations 3.6 and 3.7 is given by $S(1)$.

Thus δ (or δ_0) is derived by an estimation of the slope of the above linear-log equation or by the coefficients from the quadratic-log equation. For details of the algorithm used when transforming the series into a diffusion process, we refer the reader to [6].

Chapter 4

Time Series Data

In this work we have applied two variance scaling methods (R/S analysis and DFA) and a pdf scaling method (DEA) on financial and volcanic time series data. This section gives a brief background of the data sets used and also presents the stationarity tests. Augmented Dickey Fuller test (ADF) was used for checking the stationarity of the time series [27].

4.0.1 Financial time series

The financial data used was taken from: Mexico (MXX), from November 8, 1991 to October 22, 2001; Brazil (BOVESPA), from April 27, 1993 to June 24, 2005; Argentina (MERVAL), from October 8, 1996 to June 24, 2005; Hong Kong (HSI), from January 2, 1991 to June 24, 2005; Phillipines (PSI), from 1997 to 2001; Thailand (SETI), from 1997 to 2001; New York (SP500) from January 3, 1950 to June 23, 2005; USA (SPC), from 1991 to 2001; Turkey (XU100) from 1997 to 2001 and USA (NASDAQ), from 1997 to 2001.

4.0.2 Volcanic time series

The Volcanic data used was recorded by seismic stations belonging to the Bezymianny Volcano Campaign Seismic Network (PIRE). Data was requested for 10 days before and 5 days after the published time of the volcanic eruptions. The seismic stations used were BEZB and BELO. Volcanic eruptions 1 and 2 were from BEZB and Volcanic eruptions 3-8 were from BELO.

4.0.3 Stationarity of the Financial and Volcanic time series

In this section the stationarity of the Financial and Volcanic data is determined by using the Augmented Dickey-Fuller test (ADF). We implemented the method in R and Python for comparison.

Augmented Dickey-Fuller

The Augmented Dickey-Fuller test is a type of statistical test called *unit root test*. The null hypothesis of the test is that if the time series can be represented by a unit root, thus it is not stationary (has some time-dependent structure). The alternate hypothesis (rejecting the null hypothesis) is that the time series is stationary.

Financial time series

After implementing the ADF tests to the financial data the following results were obtained for p-values at $\alpha = 0.05$.

Table 4.1: **ADF test applied to the financial time series: p-values [33]**

Market	p-value
BVSP	0.015
SPC	0.034
HSI	0.033
IGPA	0.03
MERV	0.014
MXX	0.024
Nasdaq	0.04
PSI	<0.01
SETI	<0.01
SP500	<0.01
XU100	0.01

Volcanic time series

After implementing the ADF test to the Volcanic time series the following results were obtained for p-values at $\alpha = 0.05$.

Table 4.2: **ADF test applied to the Volcanic time series: p-values [33]**

Eruption Number	p-value
1	0.3568
2	0.6747
3	0.3024
4	0.095
5	0.2064
6	0.3271
7	0.2374
8	0.4059

The above tables summarize the results obtained for the two time series, it is clear from the ADF test that the volcanic time series is non-stationary while the financial time series is stationary.

Chapter 5

Results And Discussion

5.1 Results

This section describes the analysis of financial indices and volcanic time series when our models are applied to the data sets. Tables 5.1 and 5.2 show the scaling exponents derived from applying the three scaling methods. The δ, H, and α exponents are used to obtain $\delta_{Levy}(R/S)$ and $\delta_{Levy}(DFA)$. The Hurst analysis of financial indices and volcano time series are shown in Figs. 1 - 5 and Figs. 8 - 9. The slope of the best straight line fitted on the logarithmic plot of rescaled range (R/S) versus time is the Hurst exponent (see Table 5.1). Figs. 10 - 14 and Figs. 15 - 18 summarize the DFA analysis of financial indices and volcanic eruption, showing the linear trend when plotting (n) and $F(n)$ on a log-log scale. A linear relationship on a double log graph indicates that there is a scaling or self-similarity in the graph, and the fluctuations can be characterized by scaling exponent. Tables 5.1 - 5.2, Figs. 6 - 14, and Figs. 15 - 18 show that the scaling exponent (α) is less than 1, which confirms the presence of long-range correlations, i.e.: the large values are likely to be followed by large values and vice versa. So the DFA allows us to study the correlations in data, without disturbance of seasonality or trend. In Figs. 19 - 23 and Figs. 24 - 27, we notice that there is a considerable difference between the DEA analysis of financial indices and volcanic eruptions data. Unlike the financial indices, $S(t)-S(1)$ of the volcanic eruption data is increased almost exponentially with the logarithm of time scale.

Table 5.1: **Scaling exponents for emerging and established markets time series [33]**

Market	R/S(H)	DFA (α)	DEA(δ)	δ_{Levy} (R/S)	δ_{Levy} (DFA)
BVSP	0.59	0.72	0.57	0.56	0.63
SPC	0.59	0.62	0.60	0.56	0.56
HSI	0.65	0.7	0.60	0.56	0.63
IGPA	0.74	0.65	0.53	0.63	0.56
MERV	0.62	0.62	0.56	0.56	0.56
MXX	0.64	0.66	0.59	0.56	0.56
Nasdaq	0.6	0.72	0.56	0.56	0.56
PSI	0.66	0.71	0.55	0.63	0.56
SETI	0.64	0.70	0.54	0.56	0.56
SP500	0.63	0.66	0.65	0.58	0.60
XU100	0.64	0.70	0.54	0.56	0.56

Table 5.2: **Scaling exponents of Volcanic Data time series [33]**

Eruption Number	R/S(H)	DFA (α)	DEA(δ)	δ_{Levy} (R/S)	δ_{Levy} (DFA)
1	0.45	0.74	0.934	0.4756	0.6547
2	0.51	0.92	0.934	0.5093	0.8682
3	0.38	0.85	0.934	0.4472	0.7636
4	0.39	0.66	0.934	0.4509	0.5957
5	0.39	0.76	0.934	0.4513	0.6729
6	0.37	0.67	0.934	0.4433	0.6002
7	0.42	0.81	0.934	0.4634	0.7194
8	0.504	0.75	0.934	0.5018	0.6684

5.2 Discussion

For the financial series data all three scaling methods correctly detect the existence of long-range correlations. Comparing δ with the relation in equation 3.4, we see that the relation holds (with adjustments within the interval (0,0.06)) since equality is almost always impossible by virtue of the fact that each scaling method derives it's scaling exponent through approximations. Thus we are able to deduce that the financial time series is characterized by a Lévy walk. With the Volcanic data however the R/S analysis is unable to correctly detect the existence of long-range correlations since the volcanic data is non-stationary. However the DEA and DFA correctly detects long-range correlations. Equation 3.4 is however not satisfied and clearly $\delta \neq (H, \alpha)$. Hence the volcanic series can neither be characterized by FBM nor Lévy walk. The volcanic time series is thus characterized by a Lévy flight (i.e it has an infinite variance).

Chapter 6

Concluding Remarks

In this study, we have used high frequency financial and volcanic time series to analyze their scaling and dynamic behavior. We have implemented some scaling techniques, namely Diffusion Entropy Analysis, Detrended Fluctuation Analysis and the R/S analysis that incorporates exponential and Hurst parameters. The techniques allow us to characterize the data distribution and their long-range correlations. To obtain a good fit for the data, we first analyze their stationary behavior using unit root tests (see subsection 4.0.3). Tables 5.1 - 5.2 to confirm that the p-values are significant at all specified levels for financial data, so the high frequency financial indices used in this paper are stationary. In subsection 4.0.3, we see that the volcano time series data shows non-stationary behavior. We fit three scaling exponent techniques into our financial and geophysical data in order to estimate the exponent parameters.

Tables 5.1 and 5.2 summarize the estimation of parameters α, δ, and H for financial and volcano data, respectively. We see that the estimated values (α, δ, and H) fall between 0 and 1, which means that the high frequency stock market data and volcanic eruption data show long memory behavior. The long memory supports that the present information is highly correlated with past information at specified levels, which may facilitate prediction. We conclude that for the high frequency stock market data, the Hurst coefficient is near to 0.65, Detrended fluctuation parameter is near to 0.65, and Diffusion entropy parameter is near to 0.59. For the high frequency volcanic time series, the Hurst coefficient is near to 0.39 and Diffusion entropy parameter is 0.6837. In addition we have shown with a combination of DEA and the variation scaling methods that the financial time series can be characterized by a Lévy walk while the volcanic time series is characterized by a Lévy flight. The Lévy process is useful to detect a financial crash of the stock market or the risky seismic events. Since the high frequency data follow almost log-normal distribution, for any finite-variance Lévy process, randomizing time is equivalent to randomizing variance. Thus the time-varying Lévy process generates stochastic volatility (SV) by randomizing time, which may improve the forecasting performance. The reason is that the SV model takes into account a stochastic component of the data volatility and estimates the time-varying parameters using filtering techniques in order to predict future volatility [28].

6.1 Future Work

1 Based on the current work we are able to determine the characterization of a time series data. Using this information we seek to classify efficiency of predictive forecasting models based on our the knowledge of the characterization of the time series data. We consider in particular stochastic models such as the Simple Moving Average (MA), The Auto Regresive Moving Average (ARMA) and the Auto Regressive Integrated Moving Average (ARIMA) models for this analysis.

2 We would like to explore other entropy measures besides the Shannon entropy in the DEA methodology to see how we can improve on the precision of the δ scaling. We will consider the Kolmogorov-Sinai entropy, the Renyi entropy and the Tsallis entropy for this analysis.

3 We also explore the correlation between the rate parameters obtained from forecasting different time series data from the same field by applying a system of stochastic differential equations to them. We use here a system of superposed Ornstein Uhlenbeck models to model two different complex systems and determine if there is a correlation between the rate parameters for both complex systems.

6.2 Time line

April 2020- December 2020	First Future work
January 2021- August 2021	Second Future work
September 2021- February 2022	Third Future work

References

[1] https://www.physionet.org/tutorials/fmnc/node5.html

[2] Barany, E., Beccar Varela, M.P., Florescu, I., Sengupta, I. (2012), Detecting market crashes by analysing long-memory effects using high-frequency data, *Quantitative Finance*, **12(4)**, 623-634, DOI 10.1080/14697688.2012.664937

[3] Mariani M.C. and Liu Y. (2007), Normalized truncated Lévy walks applied to the study of financial indices, *Physica A: Statistical Mechanics and its Applications*, 377, 590-598.

[4] Qi Jingchao and Yang Huijie (2011), Hurst exponents for short time series, *Physical Review E*, 84, 066114.

[5] Beccar Varela, M.P., Huizar, H.G., Mariani, M.C. and Tweneboah O.K. (2019), Lévy Flights and Wavelets Analysis of Volcano-Seismic Data, *Pure and Applied Geophysics*, 1420-9136.

[6] Scafetta, N. (2003), An Entropic Approach to the Analysis of Time Series, University of North Texas Libraries.

[7] Scafetta, N. and Grigolini, P. (2002), Scaling detection in time series: diffusion entropy analysis, *Physical Review E*, **66(3)**.

[8] Scafetta, N., Latora, V. and Grigolini, P. (2002), Lévy scaling: The diffusion entropy analysis applied to DNA sequences, *Physical review. E, Statistical, nonlinear, and soft matter physics*, 66. 031906. 10.1103/PhysRevE.66.031906.

[9] Vasile V. M., Luiza, B.I., Vamo, C., oltuz, S.M. (2007), Detrended Fluctuation Analysis of Autoregressive Processes, *arXiv:0707.1437*.

[10] Mariani, M.C. and Tweneboah, O.K. (2016), Stochastic differential equations applied to the study of geophysical and financial time series, *Physica A: Statistical Mechanics and its Applications*, **443**, 170-178.

[11] Kritoufek, Rescaled range analysis and Detrended Fluctuation Analysis:Finite sample properties and confidence intervals, *Czech Economic Review*, Charles University Prague, **4(3)**, 315-329.

[12] Huang, J., Shang, P., Zhao, X. (2012), Multifractal diffusion entropy analysis on stock volatility in financial markets, *Physica A: Statistical Mechanics and its Applications*, **391(22)**, 5739-5745.

[13] Haubold, H. J., Mathai, A. M., Saxena, R. K. (2014). Analysis of solar neutrino data from Super-Kamiokande I and II. *Entropy*, 16(3), 1414-1425.

[14] Brooks, C. (1995), A measure of persistence in daily pound exchange rates, *Applied Economics Letters*, **2(11)**, 428-431, ISSN 1466-4291.

[15] *https://en.wikipedia.org/wiki/Rescaled_range*

[16] *http://fusionwiki.ciemat.es/wiki/Long-range_correlation*

[17] Hurst, H. E. (1951), Long term storage capacity of reservoirs, *Trans. Am. Soc. Eng.*, 116, 770-799.

[18] Peng, C.K., Sergey V. B., Shlomo H., Simons, M., Stanley, H.E. and Goldberger, A.L. (1994), Mosaic organization of DNA nucleotides, *Physical review. E*, 49, 1685.

[19] Buldyrev, S.V., Goldberger, A.L., Havlin, S., Mantegna, R.N., Matsa, M.E., Peng, C.K., Simons, M. and Stanley, H.E. (1995), Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis, *Physical review. E*, 51, 5084.

[20] Heneghan, C. and McDarby, G. (2000), Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes, *Physical review. E*, 62, 6103.

[21] Siwy, Z., Ausloos, M. and Ivanova, K. (2000), Correlation studies of open and closed state fluctuations in an ion channel: Analysis of ion current through a large-conductance locust potassium channel, *Physical review. E*, 65, 031907.

[22] Janosi, I.M. and Muller, R. (2005), Empirical mode decomposition and correlation properties of long daily ozone records., *Phys Rev E Stat Nonlin Soft Matter Phys.*, 71, 056126.

[23] Santhanam, M.S., Bandyopadhyay, J.N. and Angom, D. (2006), Quantum spectrum as a time series: Fluctuation measures, *Physical review. E*, 73, 015201.

[24] Talkner, P., Lutz, E. and Hnggi, P. (2007), Fluctuation theorems: Work is not an observable, *Physical review. E*, 75, 032903.

[25] *http://www.ueltschi.org/teaching/chapShannon.pdf*

[26] *http://galileo.phys.virginia.edu/classes/152.mf1i.spring02/Entropy.pdf*

[27] Mariani, M.C., Bhuiyan, M.A.Masum, Tweneboah, O.K., Gonzalez-Huizar, H. and Florescu, I. (2018), Volatility models applied to geophysics and high frequency financial market data, *Physica A: Statistical Mechanics and its Applications*, 503, 304-321.

[28] Mariani, M.C., Bhuiyan, M.A.Masum, Tweneboah, O.K. (2018), Estimation of stochastic volatility by using OrnsteinUhlenbeck type models, *Physica A: Statistical Mechanics and its Applications*, 491, 167-176.

[29] Bartlomiej Dybiec, Ewa Gudowska-Nowak, Eli Barkai, Alexandre A. Dubkov, (2017), Lévy flights versus Lévy walks in bounded domains. *PhysRevE*,052102(13),1-9.

[30] Graves Timothy, Gramacy B. Robert, Watkins W. Nicholas, Franzke L.E. Christian, (2016), A brief history of long memory: Hurst, Mandelbrot and the road to ARFIMA. *Entropy*,19(9),437

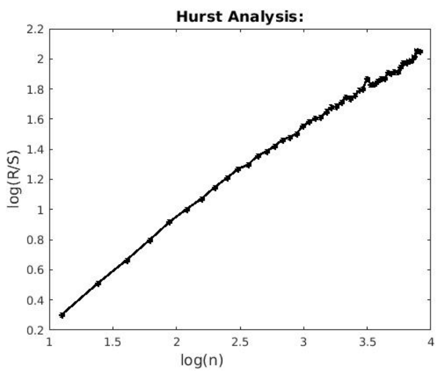
[31] Berestycki,H.,Busca,J., Florent,I., (2004), Computing the implied volatility in stochastic volatility models. *Commun.Pure Appl.Math.*,57(10),1352-1373.

[32] Mariani, M. C., Bianchini, A., Bandini, P. (2012). Normalized truncated Levy walk applied to flexible pavement performance. *Transportation Research Part C: Emerging Technologies*, 24, 1-8.

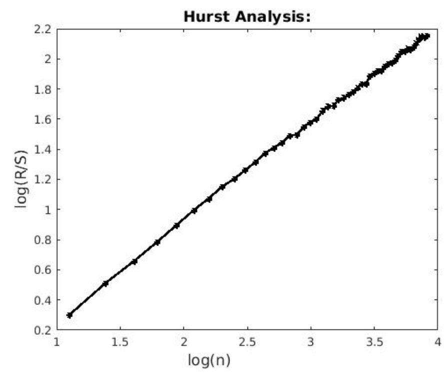
[33] Mariani, M.C.; Asante, P.K.; Bhuiyan, M.A.M.; Beccar-Varela, M.P.; Jaroszewicz, S.; Tweneboah, O.K. *Long-Range Correlations and Characterization of Financial and Volcanic Time Series.*, Mathematics 2020, 8, 441.

Appendix A

This appendix presents the figures obtained from our numerical proceedure using the Rescaled range analysis (R/S). Figures 1 - 5 are the plots obtained from the (R/S) applied to the various financial markets and figure 6-9 are the plots obtained from the (R/S) applied to the various volcanic eruptions recorded at different times [33].

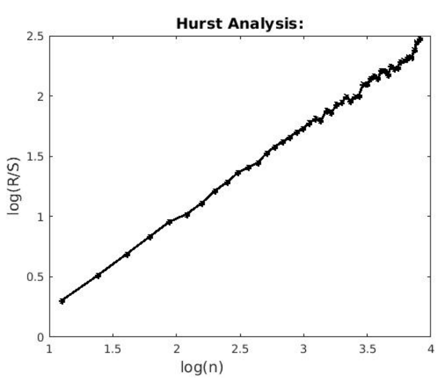


(a) BVSP

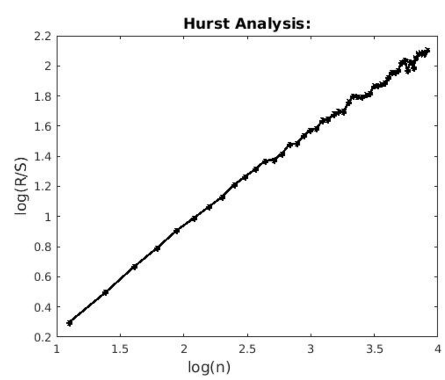


(b) HSI

Figure 1: R/S analysis for BVSP and HSI

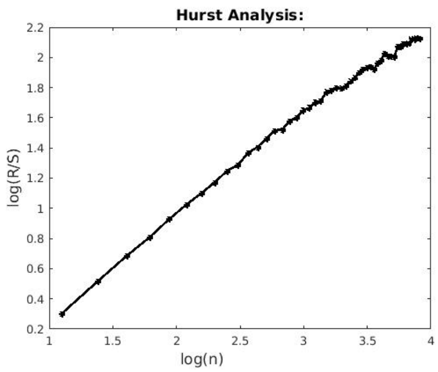


(a) IGPA

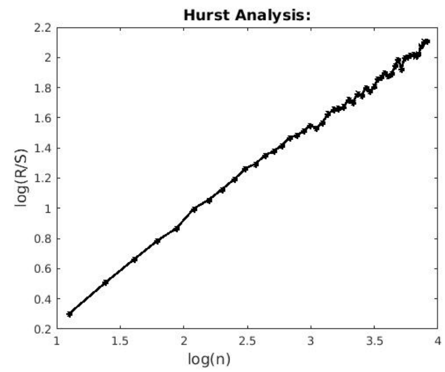


(b) MERV

Figure 2: R/S analysis for IGPA and MERV

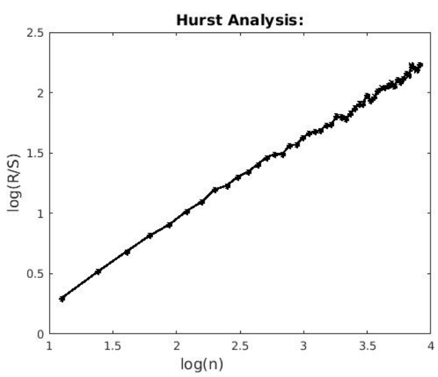


(a) MXX

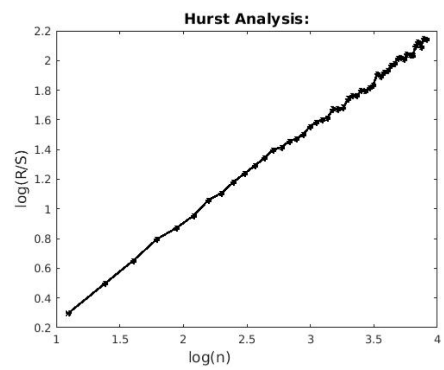


(b) Nasdaq

Figure 3: R/S analysis for MXX and Nasdaq

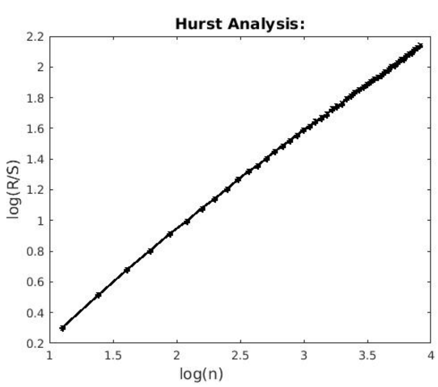


(a) PSI

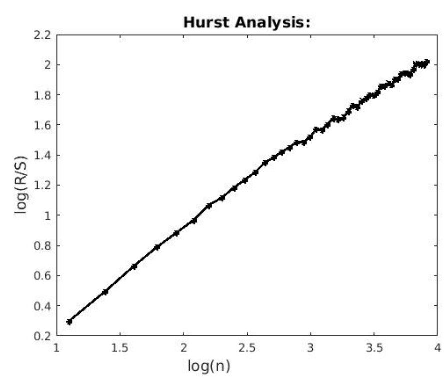


(b) SETI

Figure 4: R/S analysis for PSI and SETI

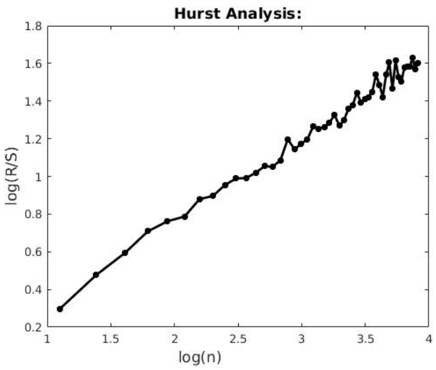


(a) SP500

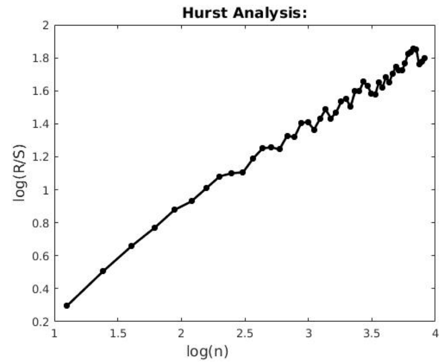


(b) SPC USA

Figure 5: R/S analysis for SP500 and SPC USA

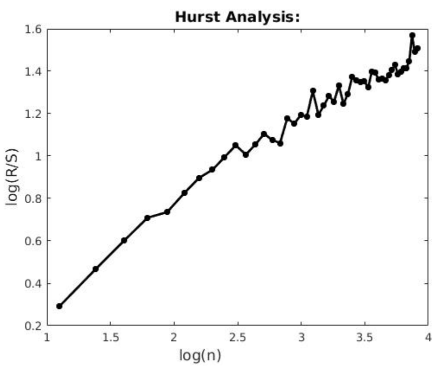


(a) Volcanic Eruption 1

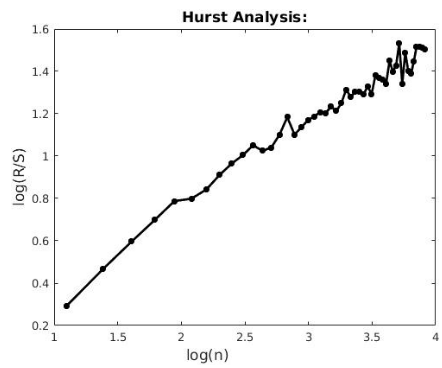


(b) Volcanic Eruption 2

Figure 6: R/S for Volcanic Eruptions 1 and 2

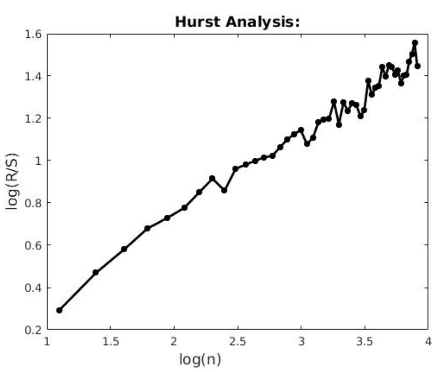


(a) Volcanic Eruption 3



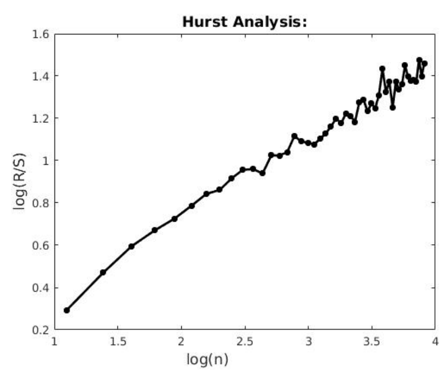
(b) Volcanic Eruption 4

Figure 7: R/S for Volcanic Eruptions 3 and 4



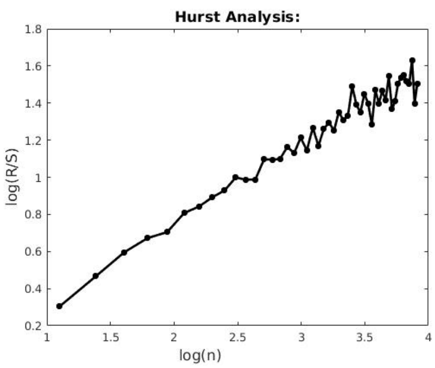
3

(a) Volcanic Eruption 5

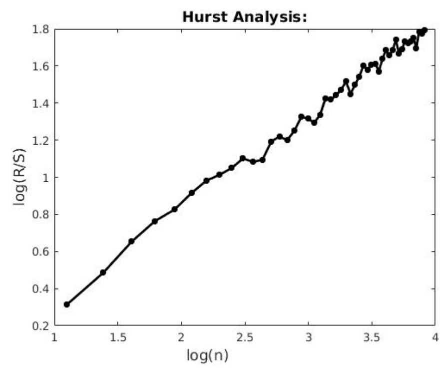


(b) Volcanic Eruption 6

Figure 8: R/S for Volcanic Eruptions 5 and 6



(a) Volcanic Eruption 7

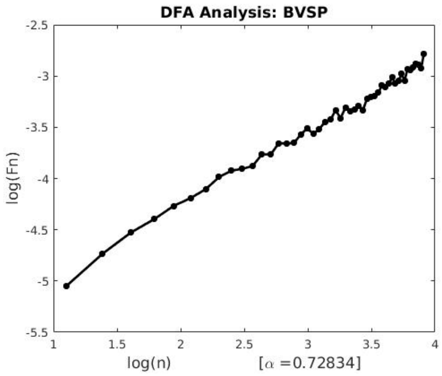


(b) Volcanic Eruption 8

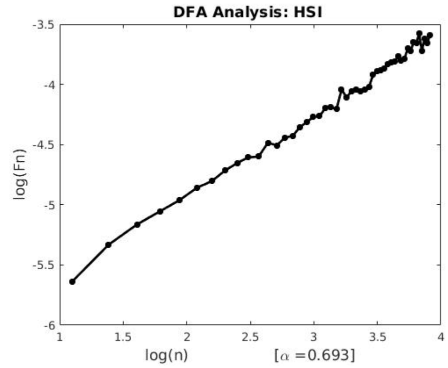
Figure 9: R/S for Volcanic Eruptions 7 and 8

Appendix B

This appendix presents the figures obtained from our numerical proceedure using the Detrended fluctuation analysis (DFA). Figures 10 - 14 are the plots obtained from the (DFA) applied to the various financial markets and figure 15-18 are the plots obtained from the (DFA) applied to the various volcanic eruptions recorded at different times [33].

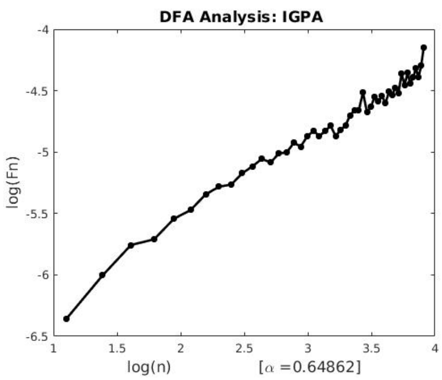


(a) BVSP

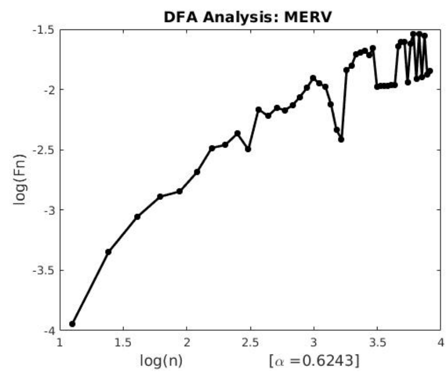


(b) HSI

Figure 10: DFA for BVSP and HSI

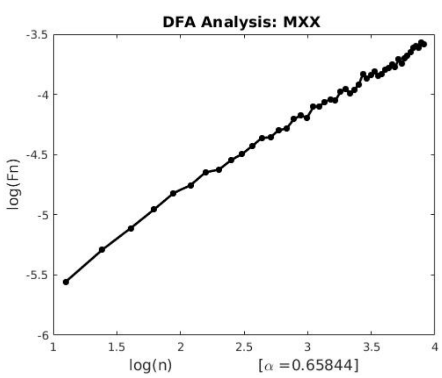


(a) IGPA

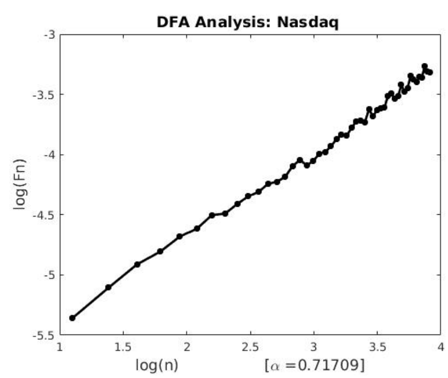


(b) MERV

Figure 11: DFA for IGPA and MERV

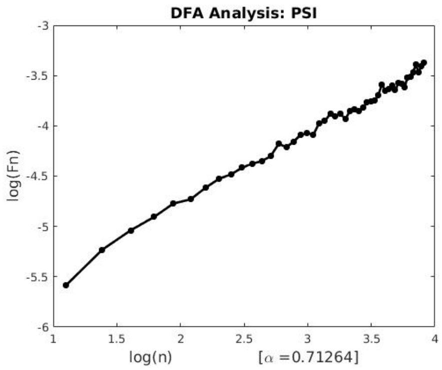


(a) MXX

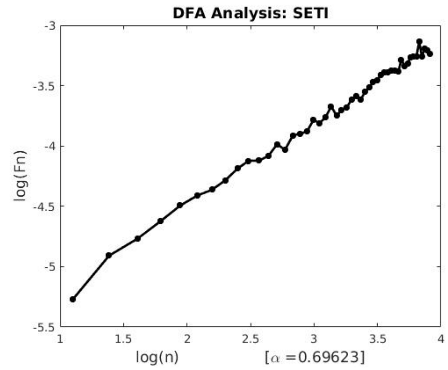


(b) Nasdaq

Figure 12: DFA for MXX and Nasdaq

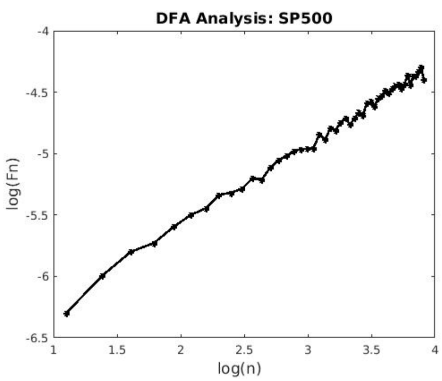


(a) PSI

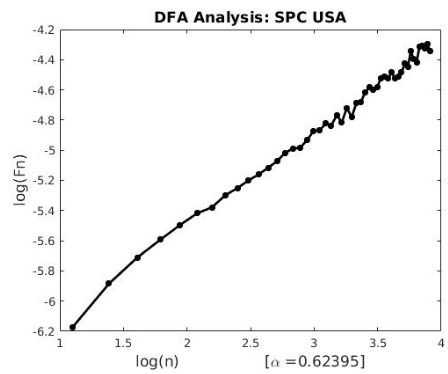


(b) SETI

Figure 13: DFA for PSI and SETI

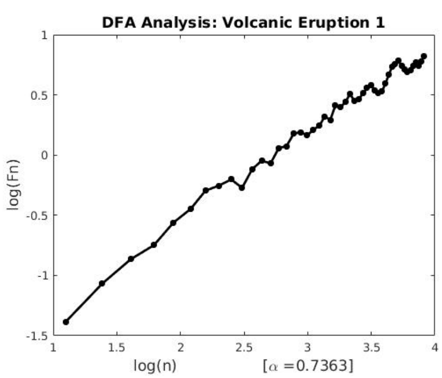


(a) SP500

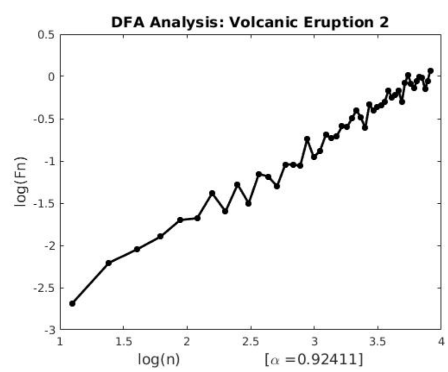


(b) SPC USA

Figure 14: DFA for SP500 and SPC USA

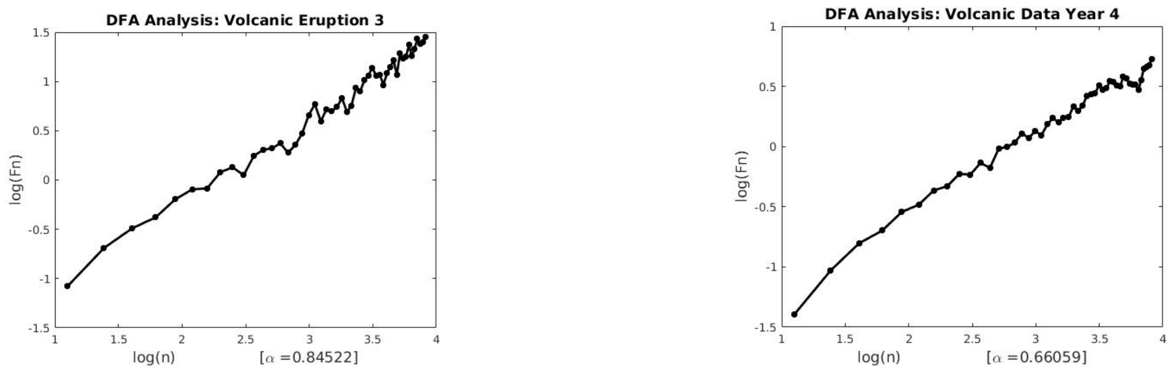


(a) Volcanic Eruption 1



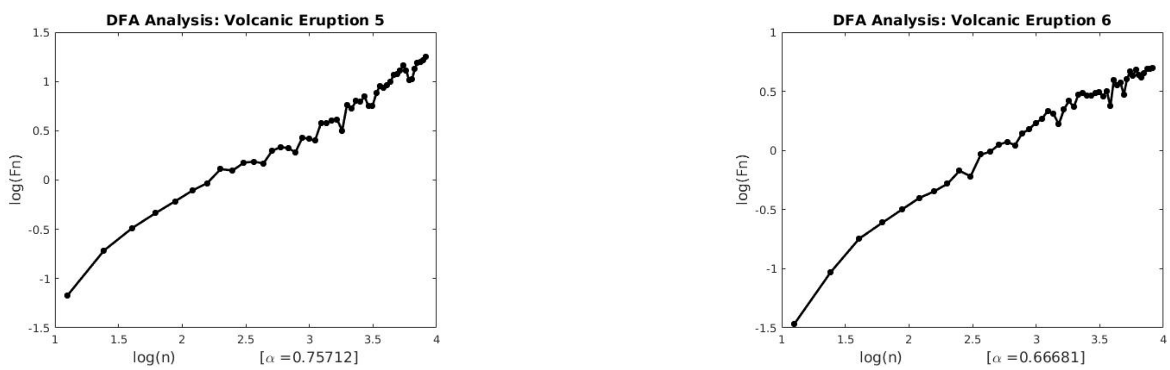
(b) Volcanic Eruption 2

Figure 15: DFA for Volcanic Eruptions 1 and 2



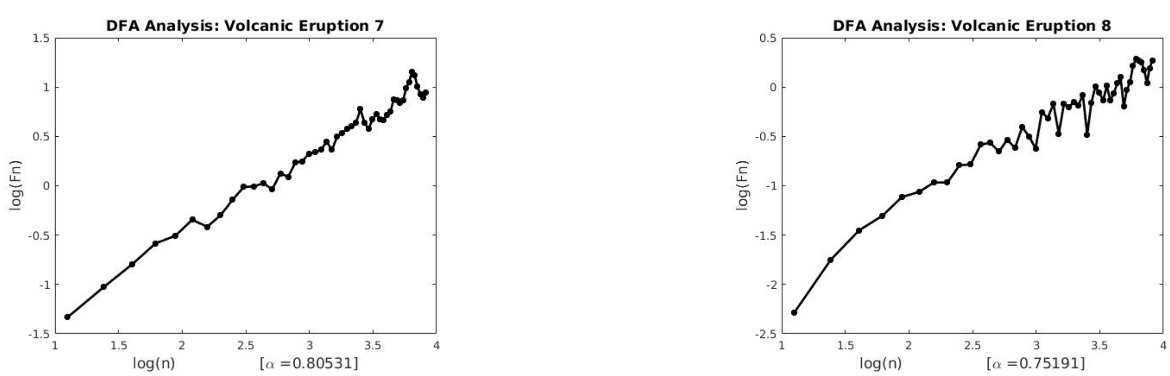
(a) Volcanic Eruption 3 (b) Volcanic Eruption 4

Figure 16: DFA for Volcanic Eruptions 3 and 4



(a) Volcanic Eruption 5 (b) Volcanic Eruption 6

Figure 17: DFA for Volcanic Eruptions 5 and 6

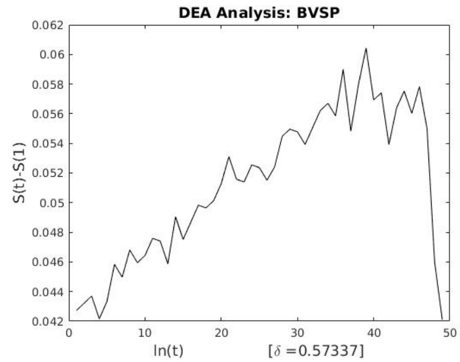


(a) Volcanic Eruption 7 (b) Volcanic Eruption 8

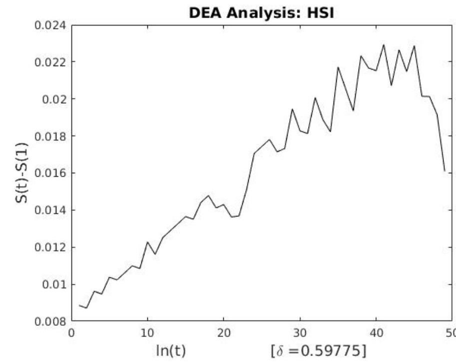
Figure 18: DFA for Volcanic Eruptions 7 and 8

Appendix C

This appendix presents the figures obtained from our numerical proceedure using the Diffusion entropy analysis (DEA). Figures 19 - 23 are the plots obtained from the (DEA) applied to the various financial markets and figure 24-27 are the plots obtained from the (R/S) applied to the various volcanic eruptions recorded at different times [33].

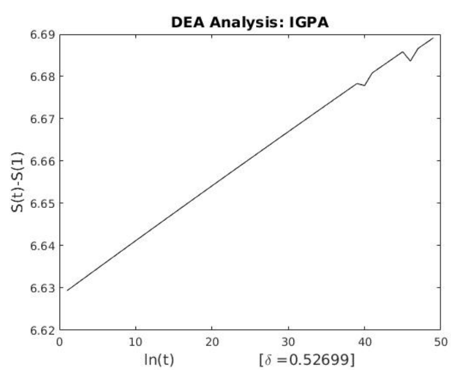


(a) BVSP

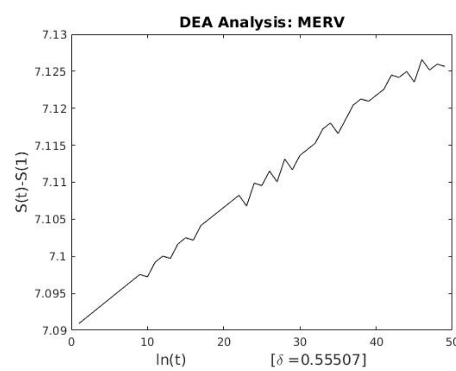


(b) HSI

Figure 19: DEA for BVSP and HSI

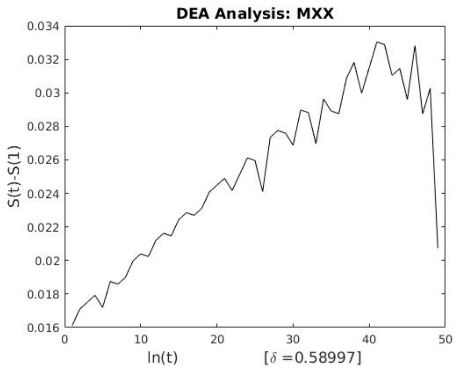


(a) IGPA

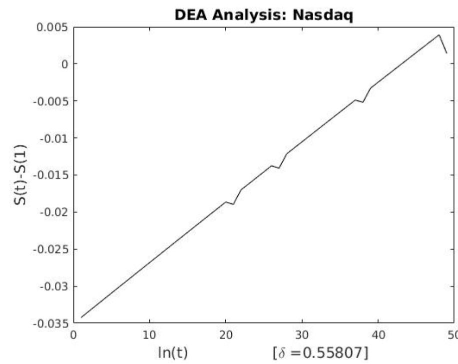


(b) MERV

Figure 20: DEA for IGPA and MERV

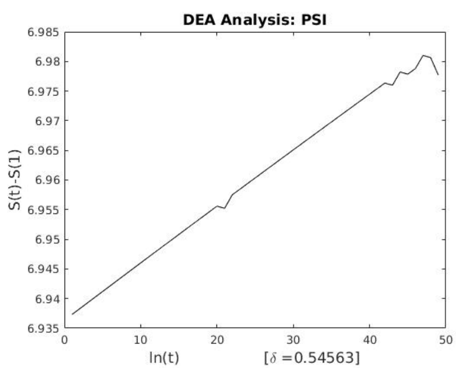


(a) MXX

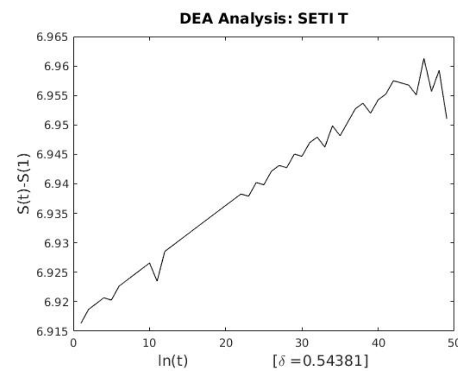


(b) Nasdaq

Figure 21: DEA for MXX and Nasdaq

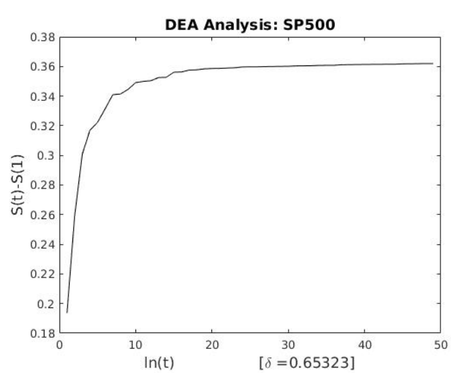


(a) PSI

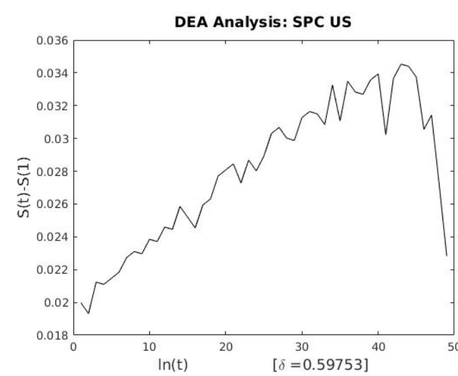


(b) SETI

Figure 22: DEA for PSI and SETI

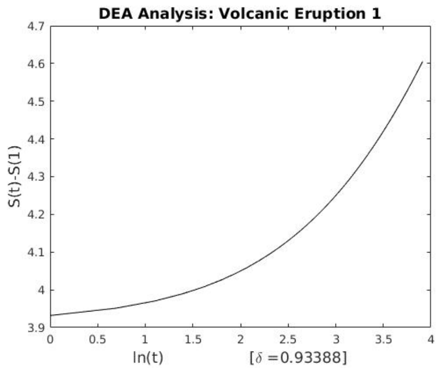


(a) SP500

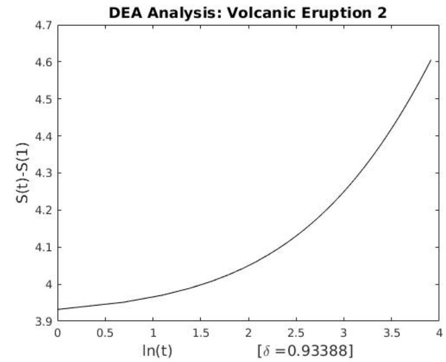


(b) SPC USA

Figure 23: DEA for SP500 and SPC USA

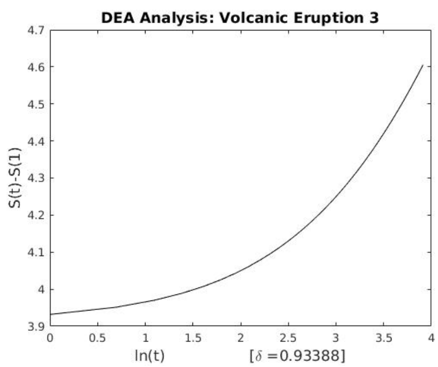


(a) Volcanic Eruption 1

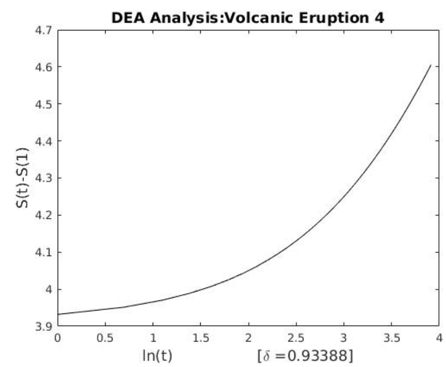


(b) Volcanic Eruption 2

Figure 24: DEA for Volcanic Eruptions 1 and 2

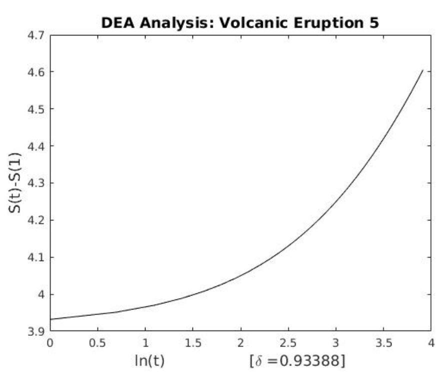


(a) Volcanic Eruption 3

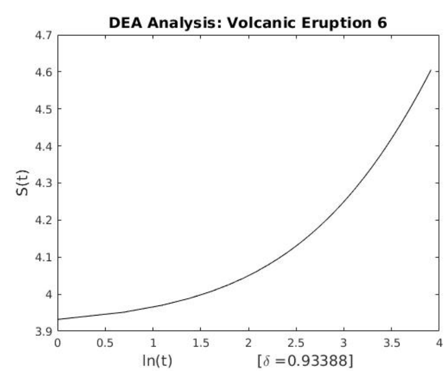


(b) Volcanic Eruption 4

Figure 25: DEA for Volcanic Eruptions 3 and 4

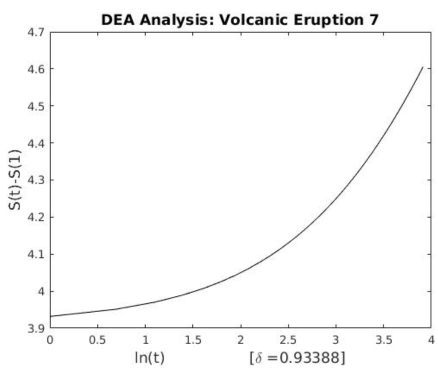


(a) Volcanic Eruption 5

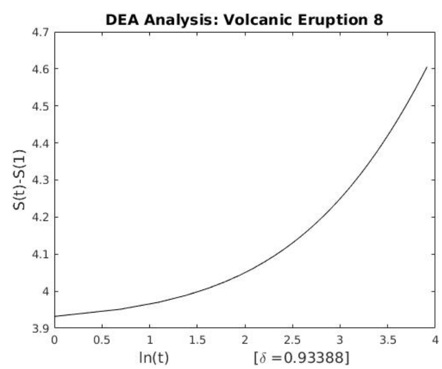


(b) Volcanic Eruption 6

Figure 26: DEA for Volcanic Eruptions 5 and 6



(a) Volcanic Eruption 7



(b) Volcanic Eruptions 8

Figure 27: DEA for Volcanic Eruptions 7 and 8

Curriculum Vita

Peter K Asante was born on May 21, 1990. Currently, he is pursuing a Ph.D. degree in Computational Science at The University of Texas at El Paso (UTEP) under the supervision of Prof. Maria C. Mariani. He had his Masters degree in the field of Mathematical Sciences at Kansas State University, Manhattan,KS, where his focus was on applied Mathematics and acquired a certificate in applied Mathematics in addition to his Master's degree. Prior to that, he had his Bachelor's degree in Mathematics at the Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. His experience with research as an undergraduate student fuelled his interest in pursuing graduate studies in mathematics. During his master's program at Kansas State he realized he was more interested in applying mathematical concepts in a practical way to different fields and was impressed with how mathematical models could provide a way to understand complex systems that occured both in nature and outside of nature. This led him to pursue his doctoral degree in Computational Science where he seeks to apply and develop mathematical and stochastic models to help understand naturally occuring systems as well as stochastic systems.

Currently his work involves the application of data analytics on Financial & Geophysical time series, with particular interest in scaling models for time series, systems of stochastic differential equation and entropy measures in detecting long-memory effects of time series data. He began his Ph.D. program in the Spring of 2018. He worked as a Teaching Assistant at KNUST after his Bachelor's degree, at Kansas State University during his Masters degree program and currently works as a PhD teaching assistant at the University of Texas at El Paso.He has one publication as a co-author, one conference paper as a co-author and has presented at two conferences. He is also a member of the Society for Industrial and Applied Mathematics (SIAM) and Computational Science Students Association (CPSSA) where he serves as the RSO.