2019-06

Modeling of Climatic Index on Infectious Diarrheal Disease

Ahasan, Md. Nazmul

University of Rajshahi

Modeling of Climatic Index on Infectious Diarrheal Disease

A dissertation submitted to the University of Rajshahi in fulfillment of the requirements for the degree of ***Doctor of Philosophy (PhD)***

Submitted by

Md. Nazmul Ahasan

Under the Supervision of

Dr. Md. Mesbahul Alam
Professor, Department of Statistics
University of Rajshahi
&
Dr. Dinesh Mondal
Scientist, Laboratory Sciences Division
ICDDR,B Dhaka

Institute of Environmental Science
University of Rajshahi
Rajshahi-6205, Bangladesh

June, 2019

DECLARATION

I do hereby declare that the dissertation entitled ***Modeling of Climatic Index on Infectious Diarrheal Disease***, submitted to the Institute of Environmental Sciences (IES), University of Rajshahi, Rajshahi-6205, Bangladesh, for the Degree of **Doctor of Philosophy** is a unique, completely new and is the researcher's own work, and has not been submitted elsewhere for any degree or diploma.

Author

Institute of Environmental Science
University of Rajshahi
Rajshahi-6205, Bangladesh
June, 2019

CERTIFICATE

This is to certify that the dissertation entitled ***Modeling of Climatic Index on Infectious Diarrheal Disease*** is an original work done by **Mr. Md. Nazmul Ahasan**, PhD Fellow, Institute of Environmental Sciences, University of Rajshahi, Rajshahi-6205, Bangladesh and Assistant Professor in Statistics, Department of Statistics, Rajshahi Government College, Rajshahi for the degree of ***Doctor of Philosophy*** under my supervision.

I certify that any part of this dissertation in any form or by any means completely or partially has not been submitted elsewhere for any other degree or diploma.

(**Dr. Dinesh Mondal**)
Co-Supervisor
Scientist, Laboratory Sciences Division
ICDDR,B Dhaka

(**Prof. Dr. Md. Mesbahul Alam**)
Principal Supervisor
Department of Statistics
University of Rajshahi

Md. Mesbahul Alam,PhD
Professor
Department of Statistics
University of Rajshahi
Rajshahi-6205,Bangladesh

Dedicated to

My Well Wishers

ACKNOWLEDGMENTS

First and foremost, I would like to deepest gratitude to the Almighty Allah who has given me physical and mental strength, patience and opportunity for being able to complete this research task successfully.

I express my sincere gratitude to my honorable supervisor ***Professor Dr. Md. Mesbahul Alam****, Department of Statistics, University of Rajshahi, Rajshahi - 6205, Bangladesh, for his brilliant guidance, valuable suggestions, all out cooperation and critical comments throughout the research. I am also greatly indebted to my research co-supervisor Dr. Dinesh Mondal, ICDDR,B, Dhaka, for helpful ideas on this research and providing necessary data.*

I am pleased to express my special thanks to my dearest friend and faculty Dr. Md. Abdul Khalek, Associate Professor, Department of Statistics, University of Rajshahi, for judicious, valuable and thoughtful suggestions to complete the research work. I am also grateful to Professor Dr. Sayedur Rahman and Professor Dr. Md. Golam Hossain, Department of Statistics, University of Rajshahi, Bangladesh for inspiration and motivation in environmental and epidemiological research.

I acknowledge my gratefulness to the Institute of Environmental Science (IES), University of Rajshahi for offering me all kinds of academic supports and facilities for the study. My deep and sincere gratitude to Professor Dr. Golam Sabbir Sattar, Director,Institute of Environmental Science (IES) and Professor Dr. Md. Sultan-Ul-Islam (Ex-Director, IES),University of Rajshahi for their co-operation in a number of ways. My debt of gratitude goes to my teachers of IES, Professor Dr. Md. Golam Mostafa, Professor Dr. Md. Abul Kalam Azad, Professor Dr. Md. Redwanur Rahman, Dr. S M Shafiuzzaman and Zakiya Yasmin. Throughout my Research period, their encouragement, sound advice, good teaching, and lots of good ideas has enriched my study. Special thanks to Mr. Md. Sultan Ali, Secretary, IES for helping me much at IES.

I am highly grateful to Ministry of Education and University Grants Commission (UGC), Bangladesh for granting me administrative and financial support for this study.

Finally, I fall short of words to express my sincere thanks to my family members for their sacrifice that relaxed me during painstaking days, which lead me to all the success in this study.

Md. Nazmul Ahasan

June, 2019 Md. Nazmul Ahasan

Abbreviations and Acronyms

ACF	:	Autocorrelation Function
ADF	:	Augmented Dicky-Fuller
AIC	:	Akaike Information Criteria
AR(p)	:	Autoregressive of order p
ARCH(p)	:	Autoregressive Conditional Heteroscedasticity of order p
ARIMA(p,d,q)	:	Autoregressive Integrated Moving Average of order p, d, q
ARMA	:	Autoregressive Moving Average
BBS	:	Bangladesh Bureau of Statistics
BIC	:	Bayesian Information Criteria
CDC	:	Climate Diagnostic Centre
CFE	:	Cumulative Forecast Error
CIRES	:	Cooperative Institute for Research in Environmental Science
DCC	:	Dhaka City Corporation
DLNM	:	Distributed Lag Nonlinear Model
DMI	:	Dipole Mode Index
DOAJ	:	Directory of Open Access Journal
ENSO	:	El-Nino Southern Oscillation
ESRL	:	Earth System Research Laboratory
GAM	:	Generalized Additive Model
GARCH(p,q)	:	Generalized Autoregressive Conditional Heteroscedasticity of order p,q
GLM	:	Generalized Linear Model
HDWT	:	Haar Discrete Wavelet Transformation
ICDDR,B	:	International Centre for Diarrheal Disease and Research, Bangladesh
ICOADS	:	International Comprehensive Ocean-Atmosphere Data Set
IDD	:	Infectious Diarrheal Disease
I.I.D	:	Independently and Identically Distributed
INGARCH(p,q)	:	Integer Valued Generalized Autoregressive Conditional Heteroscedasticity of order p,q for Poisson
INGARCH-NB(p,q)	:	Integer Valued Generalized Autoregressive Conditional

		Heteroscedasticity of order p,q for Negative Binomial
IOD	:	Indian Ocean Dipole
LM	:	Lagrange Multiplier
MA(q)	:	Moving Average of order q
MAD	:	Mean Absolute Deviation
MAE	:	Mean Absolute Error
MAPE	:	Mean Absolute Percentage Error
MSE	:	Mean Squared Error
NAO	:	North Atlantic Oscillation
NOAA	:	National Oceanic and Atmospheric Administration
PACF	:	Partial Autocorrelation Function
RMSE	:	Root Mean Squared Error
SLP	:	Sea Level Pressure
SOI	:	Southern Oscillation Index
SST	:	Sea Surface Temperature
W-GARCH	:	Wavelet-Generalized Autoregressive Conditional Heteroscedasticity
WHO	:	World Health Organization
W-INGARCH	:	Wavelet-Integer Valued Generalized Autoregressive Conditional Heteroscedasticity
WMO	:	World Meteorological Organization

CONTENTS

List of Tables

List of Figures

Abstract

Ever-increasing demand of quality health care services, service delivery institutions and service providers struggle to meet up excess demands, particularly associated with peak events like diarrhea. Front-line health delivery services and providers are not usually adequately pre-informed and do not have adequate resources to meet the needs of health care. The main motivation of this study is to model the infectious diarrheal patients hospitalized data considering its variations over time as the dynamics of the disease depends on various climate factors. The monthly data were used the number of infected diarrheal patients who were hospitalized for diarrhea from January, 1993 to December, 2017 as target and IOD & MEI as explanatory.

The selected models were those with appropriate distribution for count data, flexible and that allows insertion of explanatory variables experimented. In literature, it is found that Poisson and Negative Binomial distributions have been widely used in GLM with recurrence, but its performance is weak in comparison to the Wavelet-INGARCH model. This research finds suitable forecasting models for time series counts of infectious diarrheal patients, and demonstrates expected results for decision-maker related to diarrheal disease. The practical consequence behind the research was to incorporate wavelet decomposition whether the data are either stationary or non-stationary or linear or non-linear or seasonal or having trend. The W-INGARCH-NB(1, 1) model with MEI as covariate has been found outstanding performance in forecasting number of infectious diarrheal patients.

List of Publications

1) **Ahasan, Md. Nazmul**, Md. Abdul Khalek and Md. Mesbahul Alam (2019). Modeling via Wavelet GARCH Algorithm on Multivariate ENSO Index, *International Journal of Scientific and Research Publications (IJSRP)*. **9(7).** (Accepted, Impact Factor:3.27)
2) **Ahasan, Md. Nazmul**, Md. Abdul Khalek and Md. Mesbahul Alam (2019). MODELING OF INDIAN OCEAN DIPOLE INDEX VIA W-GARCH METHODS, International *Journal of Civil Engineering and Technology (IJCIET).***10(6)** (Accepted, Impact Factor:10.7810)

3) Khalek, Md. Abdul, Md. Nazmul Ahasan and Md. Ayub Ali (2018). Groundwater Table Volatility Forecasting using Hybrid Wavelet-GARCH Model in the Northwest Bangladesh, J. *Hydrology Science and Technology*. (Impact Factor: 1.82)

4) Ahasan, Md. Nazmul, Md. Abdul Khalek and Md. Mesbahul Alam (2017). *Univariate Modeling and Forecasting of ICOADS Data on Tropical Indian Ocean: A GARCH Application*, "International Conference on Bioinformatics and Biostatistics for Agriculture, Health and Environment" during 20-23 January, 2017, Venue: Rajshai University Campus, Rajshah, Bangladesh.

CHAPTER ONE
Overview of the Study

Chapter One
Overview of the Study

1.1. Introduction

One of the most important factors that influence diarrheal disease is climatic variations. Many human infectious diseases oscillation are affected by climate variability (Kovats, Bouma, Hajat, Worrall, & Haines, 2003). McMichael et al. (2000) pointed out that field-based epidemiological research of the climatic influences on disease requires sufficient information enabling to realize the effects of climatic factors. Another important phenomenon is the use of quantitative model. It is well known to use of statistical models in providing guidelines to decision makers for making effective decisions within the state of the time series data is still a crucial task. Making decisions on quantitative measures is highest type of efficiency that can utilize existing material to the best advantage for future as well. As such, modeling of climatic index and forecasting number of infectious diarrheal patients are the focal fact of this study. This research reflects time series data which are typically consisted of a single subject's sequence of observations obtained as time progresses. Therefore, the present study is a typical time series modeling of diarrheal patients count data incorporating global climatic factors.

Ocean atmospheric climate phenomena on inter-annual time scales, affects not only the climate system of the earth and is also associated with disease, natural disasters and social conflicts. Climatic conditions influence the culture and economy of societies and the performance of economies as well. Specifically, El-Nino as an extreme climate event is known to have notable effects on health, agriculture, industry, and related areas (Fan, Meng, Ashkenazy, Havlin, & Schellnhuber, 2017).

MEI, IOD, NAO, PDO etc. are the strongest climate atmospheric index which measure our climate and weather projects of the earth. Some of the typical measures (events) are in regular cycle, while some are not. When they recur in the form of regular cycles of fluctuations in climate factors, they are called climate oscillations. The term oscillation is used because such fluctuations are not perfectly periodic. For example, El-Nino returns every four and half years. But actually it may or may not return. Or it may return too early or too late. So, ocean atmospheric phenomena(s) are quasi-periodic (Greenland, 2003).

The knowledge of climatic pattern impacts on human illness is centuries old. The Rom aristocracy took their summers at higher cooler elevations in part to avoid seasonal outbreaks of malaria. The food processes of salting, smoking, and currying of meats began in temperate climates due to an understanding of the association between food spoilage and diarrheal disease. So the earth has a disruption of normal global weather patterns resulting in a variety of weather conditions. These phenomena have both direct and indirect effects on human infectious disease (Kontra, 2017). Variations in the amplitude, duration, and distribution of ocean atmospheric phenomena can lead to major, sometimes massive and disruptive, heat wave-drought-flood cycles. More extreme variations in weather conditions, including severe storms as well as more frequent non-seasonal hurricanes and typhoons, are consequences of global warming as well. The resulting effects on health and disease, especially in impoverished regions with poor water, sanitation, and health care infrastructure can be devastating (Hales, Edwards, & Kovats, 2003).

One the other hand, global climate is a complex system, the dynamics of which are still poorly understood. Understanding climate pattern (variability) as a determinant of infectious disease is increasingly seen as a keystone of climate change preparedness and an urgent area of need in developing and less developing countries (Murray, Vos, Lozano, AlMazroa, & Memish, 2014). Therefore, modeling and forecasting of climatic index time series data can be a crucial tool of diarrheal disease study. Aside from awareness and treatment of people, forecasts can help people and governments anticipate extreme catastrophe, which may cause illness to death.

Historical data provides a rich opportunity to evaluate climate-health interactions. However, long-term time series data are often not available and confounders over a

similar study period are infrequently identified at the same spatial and temporal scale (Schulze, 2000). Variation(s) in climatic factors over time and the complexity of diarrheal disease causation complicates the use of such limited data sets to derive climate-health relationships for a particular region. With few exceptions, longer term data sets are identified from high income countries, focusing research in areas of the world with greater economic potential to mitigate and adapt to the effects of climate change both socially and environmentally (McMichael et al., 2006). Understanding the potential health impacts of climate change in low-income countries will be essential in developing mitigation and adaptive strategies designed to protect these vulnerable populations expected to be impacted the hardest and likely the least able to adapt (Mendelsohn, Dinar & Williams, 2006).

1.2 Necessary Terminologies

Necessary climatic and infectious diarrheal disease terminologies related to this dissertation are briefly discussed in the following sub-sections.

1.2.1 *Climatic Index*

Indices are diagnostic tools used to describe the state of a climate system. Climate indices are most often represented with a time series; each point in time corresponds to one index value. A climate index is any calculated value that can describe the state of the climate as well as changes in a climate system. Climate indices use air temperature, air pressure, sea surface temperature, and precipitation data in their calculations. The idea behind an index is that a particular set of values, measured from a particular set of locations and time, can be combined to create a general picture of overall climate for a region or even the entire planet. Like ocean atmospheric indices, analysis of long term trends in climate indices can help to make predictions about the future state of Earth's weather. In our study, two types of climatic indices are used, namely multivariate ENSO index (MEI) and Indian Ocean dipole (IOD) index.

A. Multivariate ENSO Index (MEI)

Multivariate ENSO index originates from the term El-Nino (which is Spanish for the Christ child) was originally used by fishermen along the coasts of Ecuador and Peru to refer to warm pacific ocean current that typically appears around the Christmas

time and lasts for several duration (Room, Forecasts & Portal, 2000). Figure 1.1 indicates the tropical Pacific Ocean that is extended from 23.26^0 N to 23.26^0 S and from 80^0W to 105^0E encompassing MEI.

The El-Nino affects both weather towards climate and marine life around the globe (Wikarmpapraharn & Kositsakulchai, 2010). The relationship between El-Nino activity and resulting climatic effects has been well established. During 1920's Sir Gilbert Walker, a British scientist discovered a relationship between barometer readings at stations on the Eastern and Western sides of the pacific. This phenomenon was called Southern Oscillation.

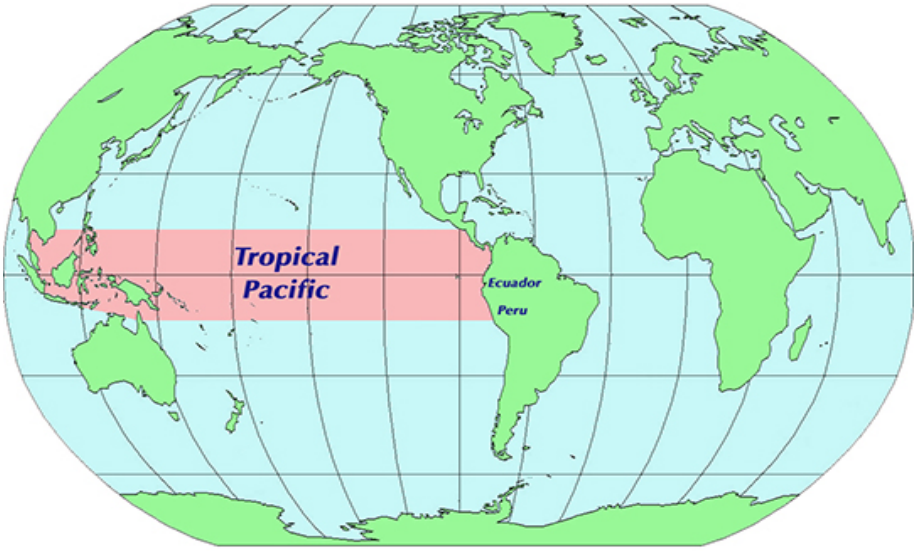


Figure 1.1: Tropical Pacific Ocean for MEI

In order to capture the relationship between El-Nino and Southern Oscillation, scientist at NOAA-CIRES Climate Diagnostic Centre (CDC) at the University of Colorado calculated an index called multivariate ENSO index, abbreviated as MEI, is a method used to characterize the intensity of an El-Nino Southern Oscillation (ENSO) event. Given that ENSO arises from a complex interaction of a variety of climate systems, MEI is regarded as the most comprehensive index for monitoring ENSO since it combines analysis of multiple meteorological and oceanographic components (Mazzarella, Giuliacci & Scafetta, 2013). MEI is determined as the first principal component of six different parameters: sea level pressure, zonal and meridional components of the surface wind, sea surface temperature, surface air temperature and cloudiness using data from the International Comprehensive Ocean-

Atmosphere Data Set, ICOADS (Wolter & Timlin, 2011). The MEI is derived from multiple climate parameters and has been shown to reflect the nature of the coupled ocean-atmosphere system better than either the SOI or SST-based indices. This is because the MEI integrates more information than other indices and is less vulnerable to non-ENSO related variability in a single variable. MEI is calculated twelve times per year for each "sliding bi-monthly season", characterized as January–February, February–March, March–April, and so on. Large positive MEI values indicate the occurrence of El-Nino conditions, while large negative MEI values indicate the occurrence of La Niña conditions.

B. Indian Ocean Dipole (IOD) Index

The IOD, also known as the Indian Nino, is an irregular oscillation of sea-surface temperatures in which the Western Indian Ocean becomes alternately warmer and then colder than the eastern part of the ocean. Monsoon in India is generally affected by the temperature between Bay of Bengal in the east and the Arabian Sea in the west. Both of these poles are situated within the equatorial belt of the Indian Ocean (i.e., the IOD index is defined as the difference between SST anomalies of the western (50–70°E, 10°S–10°N) and eastern (90–110°E, 10°S–10^0N) tropical Indian Ocean. Figure 1.2 shows the Eastern and Western pole of tropical Indian Ocean for DMI. The IOD involves an aperiodic oscillation of sea-surface temperatures (SST), between "positive", "neutral" and "negative" phases. A positive phase sees greater-than-average sea-surface temperatures and greater precipitation in the western Indian Ocean region, with a corresponding cooling of waters in the eastern Indian Ocean—which tends to cause droughts in adjacent land areas of Indonesia and Australia. The negative phase of the IOD brings about the opposite conditions, with warmer water and greater precipitation in the eastern Indian Ocean, and cooler and drier conditions in the west. The IOD phenomenon was first identified by climate researchers in 1999 (Saji, Goswami, Vinayachandran & Yamagata, 1999; Webster, Moore, Loschnigg, & Leben, 1999).

The IOD also affects the strength of monsoons over the Indian subcontinent. A significant positive IOD occurred in 1997–98, with another in 2006. The IOD is one aspect of the general cycle of global climate, interacting with similar phenomena like the El Niño-Southern Oscillation (ENSO) in the Pacific Ocean.

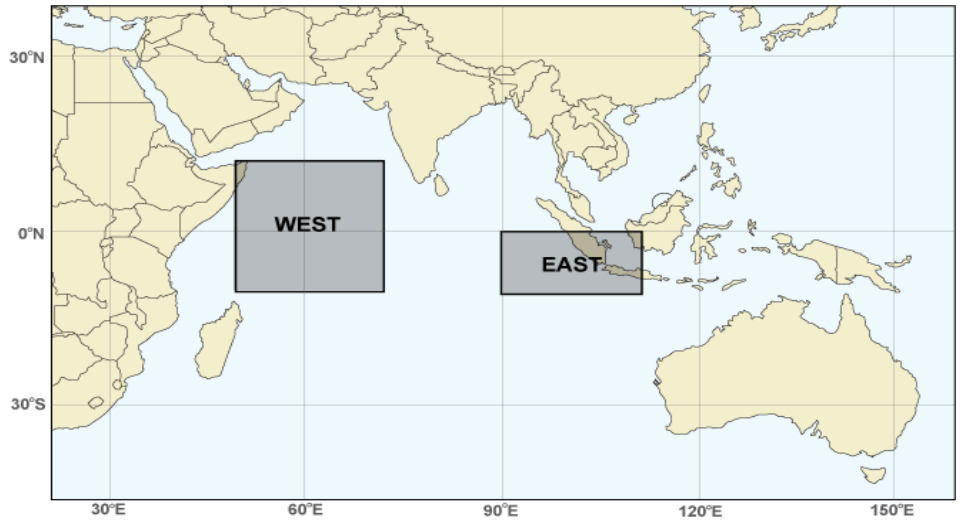


Figure 1.2: Tropical Indian Ocean for IOD

1.2.2 *Link between MEI and IOD Index*

The question is whether IOD event in the Indian Ocean is independent of ENSO events in the Pacific. This is a natural answer, when we have two oscillators in ocean(s) somehow related to each other under the global atmospheric bridge? (Yamagata et al., 2002). Owing to the dependency between the two data series, the simultaneous correlation coefficient between IOD and MEI indices is somehow related. Allan et al. (2001) found the correlation coefficient increases in the seasonally stratified data, reaching up to 0.54 in September-November. Therefore, we are favour to conclude that IOD events occur as a part of ENSO events. The remote forcing of the IOD by ENSO is an important source of long-lead predictability of the IOD (Cai, Sullivan & Cowan, 2009; Lim, Hendon, Zhao & Yin, 2017)

Development of the IOD link with the El Nino-Southern Oscillation (ENSO) in the Pacific has been explored by a number of studies as well (Hong, Lu, & Kanamitsu, 2008; Lim et al., 2017; Shinoda, Hendon & Alexander, 2004). In general, the (+)ve IOD tends to occur with El-Nino and the (−)ve IOD with La-Nina, although exceptions exist (Cai et al., 2009). This relation between the IOD and ENSO peaks in boreal autumn when remotely forced wind anomalies over the Indian Ocean, as a result of the alteration of the Walker Circulation in response to ENSO in the Pacific, are reinforced by strong local positive air-sea feedbacks (Hendon, 2003; Hendon, Lim & Liu, 2012). The remote forcing of the IOD by ENSO is an important source of long-lead predictability of the IOD (Cai et al., 2009; Lim et al., 2017). The IOD may

also act to reinforce or weaken concurrent development of ENSO (Behera et al., 2006; Luo et al., 2010) and subsequent development of ENSO in the subsequent year (Izumo et al., 2010).

As IOD and ENSO indices are non-orthogonal, however, the vast IOD influences have not been appreciated so far. In a series of literature, Saji and Yamagata (2002), using a partial correlation analysis, have demonstrated that the enhancement of the East African rain is dominated by the positive IOD rather than El-Nino. Following the above studies, the IOD appears as the second dominant mode of SST anomalies, as compared to the gravest basin-wide mode related to ENSO events. IOD may evolve without the external forcing from the Pacific ENSO but it interacts with the Pacific phenomenon in some occasions possibly through the atmospheric bridge (Behera and Yamagata, 2002).

1.2.3 *Diarrheal Disease*

Diarrheal diseases affect people of all ages irrespective of their socio-economic status and are particularly prevalent among poor people. Diarrheal diseases are a major threat to human health and still represent a leading cause of morbidity and mortality worldwide (Sarker et al., 2018). Although the burden of the diarrheal diseases is much lower in developed countries, it is a significant public health problem in low and middle-income countries like Bangladesh. Though diarrhea is preventable and managed with low-cost interventions, it is still the leading cause of morbidity according to the patient who sought care from hospitals in Bangladesh indicating that significant resources are consumed in treating those patients. According to the latest global burden of disease study in 2015, about 2.39 billion of diarrheal cases occurred globally and approximately 0.53 million of under five children died every year (Kassebaum et al., 2017; Liu et al., 2016). Specifically, incidence and case-fatality ratios are much higher in lower and middle income countries (Mashoto, Malebo, Msisiri, & Peter, 2014). In Bangladesh, diarrhea diseases are still very common to all, especially among children under 5 years old (Sarker et al., 2018). The diseases are highly sensitive to climate, showing seasonal variations in numerous sites (Drasar, Tomkins, & Feacham, 1978). Relative humidity and temperature influence the rate of replication of different types of pathogens such as bacteria and protozoa, and also the survival of entroviruses in the environment which is another cause of diarrheal diseases (Oliveira, Gabbay, Ishak, & Linhares, 1999). Differentiation between infectious and non-infectious diarrheal disease are given below.

A. Infectious Diarrheal Disease

Diarrhoea is usually a symptom of an infection in the intestinal tract, which can be caused by a variety of bacterial, viral and parasitic organisms. Rotavirus and Escherichia coli, are the two most common etiological agents of moderate-to-severe diarrhea. An overview of the causative agents in Diarrhea can be observed in the Table 1.1.

The incidence of pathogens causing diarrhea is known as infectious diarrhea that varies between developed and developing world setting. In developed countries about 70% of diarrheal cases are of viral (40% rotavirus), 10% - 20% of bacterial and <10% of protozoal origin (Cheng, McDonald & Thielman, 2005; Cunliffe et al., 1998; Wilson, 2005). In developing countries 50% - 60% cases are of bacterial (Entropathogenic Esherichia coli 25%, Campylobacter jejuni 10% - 18%, Salmonella spp. and Shigella spp. 5% each), 35% of viral (15% - 25% rotavirus) origin, and in many the cause is unidentified or mixed (Naghipour, Nakagomi, & Nakagomi, 2008).

Table 1.1: A few of Causative agents for Infectious Diarrheal Disease

Bacteria	Viruses	Parasites
Diarrheagenic Escherichia coli	Rotavirus	Cryptosporidium parvum
Campylobacter jejuni	Norovirus (calicivirus)	Giardia intestinalis
Vibrio cholerae O1,	Adenovirus	Microsporida
V. cholerae O139	Astrovirus	Entamoeba histolytica
Shigella species	Cytomegalovirus	Isospora belli
V. parahaemolyticus		Cyclospora cayetanensis
Bacteroides fragilis		Dientamoeba fragilis
C. coli		Blastocystis hominis
C. upsaliensis		
Nontyphoidal		
Salmonellae		
Clostridium difficile		
Yersinia enterocolitica		
Y. pseudotuberculosis		

Source: *Journal of clinical gastroenterology*, *47*(1), 12-20.

Diarrhea, including that of parasitic origin, remains one of the most common illnesses among children and, as reported by the WHO, is one of the major causes of infant and childhood mortality in developing countries (Boschi-Pinto, Velebit & Shibuya, 2008). Intestinal opportunistic parasitic infections are important causes of diarrhea which is a serious health problem in tropical regions. Giardia spp. and Cryptosporidium spp. are common parasitic causes of human diarrhea with the prevalence rate of 1% - 3% in the industrialized world and 4% - 17% in developing countries (Hörman, Korpela, Sutinen, Wedel & Hänninen, 2004). Causes of infectious diarrhea are multi agents.

B. Non Infectious Diarrheal Disease

Diarrhea that is not infected by a certain agents is known as non-infectious. Globally, an estimated 2 billion cases of diarrheal disease occur each year, and 1.9 million children under the age of 5 years, mostly in developing countries, die from diarrhea. Diarrhea is characterized by abnormally loose or watery stools. Some people frequently pass stools, but they are of normal consistency. This is not diarrhea. Similarly, breastfed babies often pass loose, pasty stools. This is normal. These are non-infectious diarrhea. Non-infectious diarrhea can be caused by toxins (e.g., certain types of food poisoning), chronic diseases (e.g., cystic fibrosis) or antibiotics (e.g., ampicillin). Non-infectious diarrhea does not spread from person-to-person.

1.2.4 *Link of Global Climatic Index and Infectious Diarrhea*

According to the estimates from global burden of disease 2015, diarrhea was responsible for more than 1.31 million deaths around the world (Wang et al., 2016). Although the burden of acute infectious diarrhea is greatest in low-income countries, it is a common cause of outpatient visits and hospital admissions in high-income countries as well (Troeger et al., 2017). The Asia Pacific region is home to more than half of the world’s population and is regarded as one of the most vulnerable areas under the influence of global climatic factors(Woodward, Hales, & Weinstein 1998). As the impact of global climatic events (either heat waves or cold waves) becomes more intense, we expect meteorological factors might affect the timing and intensity of infectious diarrhea in these areas (Hashim & Hashim, 2016). Local weather factors like temperature (Chou et al., 2010; D'souza, Hall, & Becker, 2008), relative humidity and rainfall (Mellor, Kumpel, Ercumen, & Zimmerman, 2016) to be linked to diarrhea-associated morbidity. Global climatic factors like the Indian Ocean Dipole

(IOD) and El-Nino Southern Oscillation (ENSO) are also considered to influence the causes of infectious gastroenteritis (Onozuka, 2014).

Figure 1.3 shows he World Health Organisation posited a number of potential ENSO, or more specifically, El Niño-related health impacts, based either on known health outcomes arising from past ENSO events or conceptual relationships of climate and health, given that ENSO produces discernible variations in health sensitive climate. That is the most prominent source of inter annual global climate variability which affects weather conditions, such as temperature, rainfall, wind speed and direction, and storm tracking throughout the world. Hashizume, Terao & Minakawa (2009) and Shaman & Lipsitch (2013) shows these severe effects of fluctuation and differentiation from region to region. The IOD is another global climate phenomenon that arises from ocean-atmosphere interactions which affect climate patterns in the tropical Indian Ocean (Hashizume, Chaves & Minakawa, 2012; Hashizume et al., 2011; Hashizume et al., 2009). Moreover, the WHO in 2003 quantified the impact of global warming on diarrhea, and reported that warming by 1^0C was associated with a 5% increase in diarrhea. Although regional differences and contrasting effects of temperature on different kinds of diarrhea are evident (Kolstad & Johansson, 2011), few studies have examined the non-stationary associations of infectious gastroenteritis and global climate variability.

Figure 1.3: Health consequences of El-Nino
Source: https://www.who.int/features/2016/el-nino-3-630x420-2.jpg

The IOD and ENSO indices support the proposal that large-scale climate indices might be more useful for forecasting sensitive health events than local weather variables (Hallett et al., 2004; Stenseth et al., 2003). Climate can affect the transmission of infectious diseases through several linear and non-linear pathways in a host population (Onozuka, 2014). Climate can also affect several biological traits of the organisms involved in the life cycle of parasites, from individual life histories to population dynamics (Hallett et al., 2004). The WHO and the Intergovernmental Panel on Climate Change (IPCC) have identified changes in the incidence of diarrhea as one of the most important future health effects of climate change (Walker et al., 2013). While local weather most likely only affects the biological components of disease transmission, large-scale climate patterns might also influence contextual components of disease transmission, such as population susceptibility (Adriano Mazzarella, Giuliacci & Pregliasco, 2011).

In this sense making a relationship and interpretation between climatic factors and infectious diarrhea is important for disease control and prevention. An early warning system based on large scale climate indices forecast could be implemented.

1.3 Essentials of Modeling and Forecasting

Modeling and forecasting of future events is based on a foreknowledge acquired through a systematic process or intuition (Armstrong, 2001; Lewis, McGrath & Seidel, 2009). It is the application of science and technology to forecast the conditions of event(s) for a given location and time. Forecasting has advanced over time and has gained the solution of many specialized areas, including the fields of health (Marno et al., 2010; McInnes et al., 2017; Rogers et al., 2010), economics and commerce (Armstrong, 2001; Makridakis & Taleb, 2009), environment , including meteorology, (WMO, 2009; Patz et al., 2003), technology and politics (Fildes, 2006; Manton, Soldo & Vierck, 1984; Rogge, 1982).

Human beings have attempted to guess the weather informally for millennia and formally since the 19th century. Climate forecasts are made by collecting quantitative data about the past state of the atmosphere at a given place and using meteorology to project how the atmosphere will change. Once a human-only endeavor based mainly

upon changes in barometric pressure, current weather conditions, and sky condition or cloud cover, but climate forecasting now relies on integrated discipline(s) with computer and statistical based models that take many atmospheric factors into account (Dirmeyer, Schlosser, & Brubaker, 2009). Human input is still required to pick the best possible forecast model to base the forecast upon, which involves pattern recognition skills, Tele-connections, knowledge of model performance, and knowledge of model biases.

The inaccuracy of forecasting is due to the chaotic nature of the atmosphere, the massive computational power required to solve the equations that describe the atmosphere, the error involved in measuring the initial conditions, and an incomplete understanding of atmospheric processes. Hence, forecasts become less accurate as the difference between current time and the time for which the forecast is being made (the *range* of the forecast) increases. The uses of model consensus help narrow the error and pick the most likely outcome in the present study.

Climate forecasts are important because they are used to protect life and property. Forecasts based on temperature and precipitation are important to agriculture, and therefore to traders within commodity markets. Temperature forecasts are used by utility companies to estimate demand over coming days. On an everyday basis, people use weather forecasts to determine what to wear on a given day. Since outdoor activities are severely curtailed by heavy rain, snow and wind chill, forecasts can be used to plan activities around these events, and to plan ahead and survive them. In 2009, the US spent $5.1 billion on weather forecasting (Ubaldi, 2013) . Along with climatic forecasts, Health forecasting is a novel and valuable tool for predicting future health events or situations such as demands for health services and healthcare needs. It facilitates preventive medicine and health care intervention strategies, by pre-informing health service providers to take appropriate mitigating actions to minimize risks and manage demand (Soyiri & Reidpath, 2013).

Health service(s) is (are) the most important component of any health system. The WHO reports that effective health service delivery requires some key resources including information, finance, equipment, drugs and well-motivated staff (Frieden, 2010). Given the ever increasing demand for both the coverage and quality of health care services, health service delivery institutions and service providers struggle to

tackle situations of excess demand particularly associated with peak events (Bradley, 2005; Derlet, 2002). This is because front-line health delivery services and providers are not usually adequately pre-informed and do not have adequate resources to meet the needs of a ‘‘higher than normal’’ demand for health care. Hence, improving the access, coverage and quality of health services depends on the ways these services are pre-informed, organized and managed.

Modeling and forecasts is the most valuable when they provide sufficient warning for timely, remedial action to be taken. Providers make critical decisions and resource allocations to meet the potential demand. Meanwhile, being able to meet the demand for a forecast that provides abundant time for preparatory activities often requires the use of a good forecasting technique and sufficient reliable data. It also comes with an additional compromise as to the precision and accuracy of the forecast (Myers, Rogers, Cox, Flahault, & Hay, 2000). Hence, finding a fine line between what is predictable and the demand for specific forecast is a key challenge in modeling and forecasting. Hence, forecasts ought to be probabilistic in nature, taking the form of probability distributions over future quantities and events (Dawid, 1984).

1.4. Motivation of the Study

Any effects of climatic factors probably operate on the groups of factors in different ways and will likely be at least nonlinear and time dependent. Most models developed for climate-sensitive health determinants and outcomes provide global or large regional estimates of changes in risks associated with climate change for short period using causal relationship. But links between the disease and climate is multifactorial, given the multivariate nature of climate change under time series and nonlinear pathways in both disease and climate processes. In most of the cases, the integer valued temporal data of interest is multivariate and these multivariate temporal processes often have a complicated dependency structure. Hence an efficient multivariate temporal modeling and forecasting strategy under our target is a very challenging task.

1.5 Objectives of the Study

To establish a new efficient integer valued distributional temporal model of the disease volatility behavior in incidence of diarrheal disease agent(s) with large scale climatic factors.

- To depict an exploratory study of the disease (Diarrhea) with respect to global climatic indices under statistical investigation;
- To design a new model based on volatility features of global climatic indices;
- To establish a new model based on volatility features of count time series data such as infectious diarrheal diseases (IDD);
- Forecasting and checking model stability of the proposed model.

1.6 Organization of the Dissertation

The study has been organized into eight chapters. This section briefly discussed the structure of the dissertation.

The **first** chapter named overview of the study will discuss more about the research topic, presenting the study problem, the fundamentals of its choice and the objectives that guide the study and its purpose.

The **second** chapter is dedicated to brief review of previous works that the main basis of the research, as well as the fundamental concepts

The **third** chapter is introducing methodologies, materials and tools used in the different stages of the research

The **fourth** chapter named environmental settings of the study and will be discussed different environmental factors related to peripheral communities of the area under study.

The **Fifth** chapter will contain the exploratory results of the investigation and a preliminary development of a sustainability framework for the study.

The **sixth** chapter integrates the results of an alternative approach, for climatic indices considered for the study. This chapter also selected best model for the sustainable

detection of the climate variability. The chapter is all about choosing the best model in terms of feasibility and sustainability criteria.

The **seventh** chapter integrates the results of final model. It is all about choosing the best model in terms of feasibility and sustainability criteria for infectious diarrheal disease using climatic indices.

The **eighth** chapter devoted to the useful findings, general recommendations based on issues identified on the development of the study as well as limitations of the study

CHAPTER TWO
Review of Literatures

Chapter Two
Review of Literatures

2.1. Introduction

Review of literature(s) is a center for research in every research field. A literature review compiles and evaluates the research available on a certain topic or issue that is studied and published. It ensures that researchers do not duplicate work that has already been done, clues as to where future research is heading or recommend areas on which to focus, identifies inconsistencies, gaps and contradictions in the literature. Therefore, it is the backbone of research.

The ongoing research work develops a new statistical model based on volatility (conditional variance) technique in a hybrid statistical framework for more accurate and computationally efficient with large scale climate sensitive disease domain. The modeling method is applied to a real world surveillance data on infectious diarrheal disease. Related literatures are reviewed and presented in following areas:

- Modeling and forecasting of global climatic index(s), i.e. ENSO and IOD under mean, volatility, wavelet de-noising and hybrid model.
- Modeling and forecasting of discrete time series, i.e. infectious diarrheal disease under mean, volatility, wavelet de ’noising and hybrid model.

2.2. Review of Literatures

Hashizume et al., (2013) performed a cross wavelet coherency analysis to examine patterns of association between monthly cholera cases in the hospitals in Dhaka and Matlab (1983–2008) and indices for both IOD and ENSO. They found that the strength of both the IOD and ENSO associations with cholera hospitalizations changed across time scales in the study period. In Dhaka, 4-year long coherent cycles were observed between cholera and the index of IOD in 1988–1997. In rural Matlab, the effect of ENSO was more dominant while there was no evidence for an IOD effect on cholera hospitalizations. They called for the consideration of non-stationary, possibly non-linear, patterns of association between cholera hospitalizations and climatic factors in cholera epidemic early warning systems.

In 2011 Hashizume M, et al. investigated a negative binomial generalized linear model (GLMs) to examine whether a change in the DMI for a given month was associated with a change in monthly number of cholera cases. They discussed that temporal associations between climate and disease can be confounded by temporal trends and seasonal patterns. To account for seasonality in the incidence of cholera that is not directly linked with the IOD, indicator variables for months were included in the model. Indicator variables for the years of the study were also incorporated into the model to allow for long-term trends and other variation between the years. To allow for autocorrelations, an autoregressive term at order 1 was incorporated into the models. Both negative and positive dipole events are associated with an increased incidence of cholera in Bangladesh with varying time lags.

In another study cross-wavelet coherency analysis investigated to measure the pattern of associations between indices Indian Ocean Dipole (IOD) and El-Nino Southern Oscillation (ENSO) (Onozuka, 2014). Infectious gastroenteritis cases were non-stationary and significantly associated with the IOD and ENSO for a period of approximately 1 to 2 years. This association was non-stationary and appeared to have a major influence on the oscillation of infectious gastroenteritis transmission. Results suggest that non-stationary patterns of association between global climate factors and incidence of infectious gastroenteritis should be considered when developing early warning systems for epidemics of infectious gastroenteritis.

A systematic review article conducted by Demissie and Mengisitie (2017) in January 2017 using electronic databases Google Scholar, PubMed, and Directory of Open Access Journals (DOAJ) shows, there is need to predict an increasing global burden of diarrheal disease with climate variability as a risk factor due to inability to safely store health data. The review was limited to studies reporting the impacts of El-Nino on diarrheal diseases or studying associations between climate change associated with El-Nino and diarrheal diseases outcomes. Over 2600 scholarly papers and potential published articles identified in the initial electronic search, of which 30 fulfilled the inclusion criteria. Evidence for an association between disease risk and ENSO is more robust when analysis uses a long time-series that incorporate more than one event and when there is an appropriate geographical aggregation of data. The result of this systematic review confirmed that most of the studies noted a significant association between diarrheal disease and El-Nino. However, research on the impact of El-Nino on diarrheal disease is limited. Longitudinal studies over extended periods of time that investigate the link between El-Nino/climate change and diarrheal disease are needed. There is a need for studies to be expanded to include more countries in the region and to include other environmental, social and economic factors that might affect the incidence of diarrheal disease.

Chen et al., (2018) examined the association between weather variable(s) and acute diarrhea. A distributed lag non-linear model (DLNM) was used for the non-linearity and lagged effect of temperature used for 12 regions of Hong Kong, Taiwan, and Japan during 2012–2016. The study called for the integration of temperature-based early warning systems into existing action plans might facilitate timely interventions of AD during cold seasons. The model is an exploratory study and they are not consider the effect of heterogeneity for count time series features.

Huq et al. (2005) reported a time series analysis using clinical and various environmental data (water temperature, air temperature, water depths, rainfall, pH, salinity, water conductivity, dissolved O_2, count of bacteria and copepod counts in water samples, etc.) collected at biweekly intervals in four rural locations in Bangladesh for four years. They found a significant negative association of rainfall with cholera incidence in some locations with different lags (0−8 weeks). However, they did not control seasonal effects in the models and a linear relationship was

assumed. Furthermore, so many environmental variables with various lags were examined by stepwise regression without taking into account a priori decision of confounders and intermediates. This may cause biased estimates of the results.

In 2017, Nafissatou et al. examined prevalence of gastroenteritis viruses in association with meteorological variables in Ouagadougou, Burkina Faso. The relationship between monthly enteric virus prevalence and climatic variable was evaluated using parametric Pearson correlation coefficient test. The results can provide valuable information necessary to alert health care providers when a period of infection in the community is likely to occur.

Chou et al., (2010) investigated and quantified the relationship between climate variations and diarrhea-associated morbidity in subtropical Taiwan. Specifically, this study analyzed the local climatic variables and the number of diarrhea-associated infection cases from 1996 to 2007. This study applied a climate variation-guided Poisson regression model to predict the dynamics of diarrhea-associated morbidity. The proposed model allows for climate factors (relative humidity, maximum temperature and the numbers of extreme rainfall), auto regression, long-term trends and seasonality, and a lag-time effect. The authors emphasized to develop a model based on the climatic variation information for disease control management.

Bennett et al. (2012) studied surveillance data from two cohorts in a sub-urban shanty town in Lima, Peru, 1995 through 1998 to understand climate variability and diarrheal disease at the community level and informed predictions for future climate change scenarios, whether the El Nino climate pattern is associated with increased rates of diarrhea among Peruvian children. Diarrheal incidence through El Nino was modeled between-subject heterogeneity with random effects Poisson models. El-Nino associated climate variability affects community rates of diarrhea, particularly during the cold season and among older children. They recommended public health officials should develop preventive strategies for future El Nino episodes to mitigate the increased risk of diarrheal disease in vulnerable communities.

Carlton et al., (2013) quantified to account for the extent to which social and bio-physical factors modify climate-disease relationships, focusing on rainfall. The study tested the hypothesis that community drinking water treatment, improved sanitation, hygiene practices, social cohesion, and long-term rainfall patterns modify the

relationship between heavy rainfall events and diarrhea incidence. For each analysis, they used random-effects Poisson regression with the number of incident diarrhea cases in a given village as the outcome. The study found, heavy rainfall events were associated with increased diarrhea incidence following dry periods and decreased diarrhea incidence following wet periods.

Kolstad & Johansson (2011) investigated to assess a range of plausible health impacts of climate change and quantified the impacts of projected regional warming on diarrhea disease for Peru. A range of linear regression coefficients was combined to compute projections of future climate change -induced increases in diarrhea using the results from five empirical studies and a 19-member climate model ensemble for which future greenhouse gas emissions were prescribed. The model is applied in six geographical regions and finds uncertainties associated with future projections of diarrhea and climate change. These uncertainties can be attributed primarily to the sparsely of empirical climate–health data. The study highlighted, future climate change may bring disastrous increases in diarrhea. Most important is that the uncertainties associated with these increases are unacceptably large. More accurate empirical data for the relationships between climate and health are clearly needed.

To measure the effects of *El Niño* on hospital admissions for diarrhea Checkley et al., (2000) considered the daily data on hospital admissions from the Oral Rehydration Unit, and meteorological data from the Peruvian Weather Service, and used time-series model through correlation structure between adjacent values of the dependent variable. The analytical approach introduced the use of smooth curves (regression splines) to model changes in the pattern of admissions after the onset of El-Nino. The method was found useful for analyzing meteorological effects on health outcomes like diarrhea, acute respiratory infections, or malaria, and can control for confounding variables in the regression model. El Nino had an effect on hospital admissions greater than that explained by the regular seasonal variability in ambient temperature. If global climate change occurs as forecasted, the observed adverse health effects of El Niño may be an analogue for future changes in environmental health.

To measure the effects of *El Niño* on hospital admissions for diarrhea Hashizume et al., (2007) used weekly rainfall, temperature and number of hospital visits for non-cholera diarrhea and conducted time-series regression. Relationship of the number of

weekly non-cholera cases with rainfall and temperature using generalized linear Poisson regression models allowing for over dispersion. An autoregressive term at order one was incorporated into the models. The number of cases increased with higher temperature, particularly in those individuals at a lower socio-economic and sanitation status.

In a longitudinal cohort study De Magny et al. (2007) examined the relationship of cholera and climate was explored, by analyzing monthly 20-year cholera time series for five coastal areas in Africa's. Wavelet methods were used as tools to provide information on the evolution of the periodic component over time and allow quantification of non-stationary associations between time series. Results of the study suggest that global and regional scale climate variability influence both the temporal dynamics and the spatial synchrony of cholera epidemics in human populations. They observed synchrony is also coherent with both the local and global climate variability quantified by the Indian Oscillation Index.

Rodó et al. (2002) investigated a quantitative evidence for an increased role of global climate variability on the temporal dynamics of an infectious disease like cholera. The evidence is based on time-series analyses of the relationship between El Nino Southern Oscillation (ENSO) and cholera prevalence in Bangladesh during two different time periods. Firstly singular spectrum analysis (SSA) was applied to isolate the main inter annual variability in the data. The original time series are then analyzed to quantify the strength of the association between cholera and ENSO to identify transient couplings. In the study, it is noticed that there is a significant association and this climate phenomenon accounts for over 70% of disease variance.

In a retrospective, observational and exploratory study shows the relationships with environmental temperature, presence and severity of El Niño on the number of adult diarrheal patients attending the Hospital (Lama et al., 2004). A multiple regression analysis was performed to predict the burden of acute diarrhea in adults per month. The number of acute diarrhea cases was the response variable. The mean monthly environmental temperature (in centigrade degrees), presence of cholera cases during the observation period, presence of a weak or strong El Niño phenomenon, and different interactions between these variables were entered into the model as independent variables. The presence of cholera and the severity of the El Niño event

(weak or strong) were considered as dummy variables. A stepwise backward elimination procedure (p value entrance and removal criteria: 0.05 and 0.1 respectively) was used for fitting the final model. The study explored, during 1991-1996, an increased number of visits to the hospital due to acute diarrhea in the warmer month(s) were observed. As an early warning system to predict epidemics of diarrhea in adults, the model have a tremendous impact on healthcare strategies and management of health services in general.

Musengimana et al. (2016) examined the relationships between climatic parameters (minimum and maximum temperature) and aggregated surveillance diarrhea incidence in a community. A Poisson regression model and a lagged Poisson model with autocorrelation were investigated to test the relationship between climatic parameters (minimum and maximum temperature) and incidence of diarrhea. The study finds an association between an increase in minimum and maximum temperature and incidence diarrhea.

Ardkaew & Tongkumchum (2009) in a statistical modeling and group study identified the patterns of diarrhea incidence in the north eastern border provinces of Thailand. The study used statistical models that combined linear regression and generalized estimating equations (GEE). According to the study, it was found, a seasonal pattern higher in January to March and April to June.

Cash et al., (2010) finds a link between the incidence of cholera, a paradigmatic waterborne bacterial illness endemic to Bangladesh, and the ENSO. Nonparametric statistical analysis was applied to identify regions of SST anomalies associated with variations in Bangladesh rainfall, the Nino-3 and Nino-4 indices forecasting for Cholera Risk in an ensemble of pacemaker simulations using atmospheric model. The study reflects Bangladesh summer rainfall is enhanced following winter El Nino events, providing a plausible physical link between El Nino and cholera incidence.

Venegas-Pérez et.al. (2016) in a case study identified the impact of El Nino and La Nina phenomena on the Acute Diarrheal Disease as well as other possible environmental factors during the period 2000–2010 in the population of the state of Aguascalientes, Mexico. Descriptive statistics based on the calculation of measures of central tendency and the dispersion were performed. To correlate the phenomena under study with the rate of acute diarrheal disease, the Pearson correlation

coefficients were obtained. Meteorological and epidemiological data showed a correlation between El Niño and Acute Diarrheal Disease, showing a morbidity rate increased by the increase in the number of days with temperatures above 30^0C, which occurred under conditions of El Niño phenomenon.

Kale, Hinde & Nobre (2004) in a time series modeling recognized the temporal pattern of diarrhea disease in children less than 5 years of age to provide support for decisions about prevention and control of the disease. Generalized linear model (GLM) was derived using variables related to weather and month. An exploratory analysis of the diarrhea disease time series using a generalized additive model (Poisson family) with a spline smoother of time was performed. General pattern for both time series was found by graphical inspection and fitting of appropriate GLMs. The study provides some additional evidence that severe cases of diarrhea disease may be attributed to rotavirus.

Martinez et al., (2017) evaluated the predictability of cholera dynamics for the city in recent times based on El Nino lessons. A mechanistic temporal model and spatio-temporal model was carried out for methodology that incorporates both epidemiological processes and the effect of ENSO, as well as a previously published statistical model. The models produced accurate performance for the low-cholera years preceding 2016 but failed to predict the lack of an outbreak in response to the strong 2015±2016 El Niño.

Martinez et al., (2004) discussed the effects of the meteorological variation on the prevalence of diarrhea disease. Statistical investigation has been made for the effect of the climate on the epidemic diseases using the time series of meteorological elements and the number of patients. Authors used the correlation and the EOF analysis to the time series of diarrheal patients count and meteorological elements in Bangladesh. Total diarrhea diseases and diarrheal cholera was found common anomaly pattern but rotavirus seems different.

Worku and Sahile, (2018) reviewed the impact of ENSO climate changes especially temperature, rainfall and sunshine hours in Ethiopia. ENSO was found the harmful property of climate change, this impact on agriculture production such as socio-economic, cultural, political, biological, ecological and institutional that shapes the human environment interactions.

Nur'utami and Hidayat (2016) examined relative influences of IOD and ENSO on Indonesian rainfall. Composite analysis was used for investigating ocean-atmosphere physics over Indo-Pacific when the phenomena of IOD, ENSO, and combinations of both are occurred. The study highlights the atmosphere–ocean interaction in Indo-Pacific sector which plays an important role on Indonesian rainfall variability.

Other study done by Fisman et al. (2016) observed relative influences of IOD and ENSO on Indonesian rainfall are investigated. Relative Risk, temporal trends, and confidence intervals for incidence were estimated using Poisson regression models, with weighted monthly case counts as the outcome. The impact of ENSO suggests that warmer temperatures and extreme variation in precipitation events influence risks of vector-borne and enteric disease.

Azmoodehfar and Azarmsa (2013) showed that El Nino and La Nina should be added to the weather forecast variables because of their consequences on different tele-connected phenomenon. Ocean atmospheric variables have great uncertainty in a nonlinear pattern. A probabilistic approach has been chosen to show the effect of El Nino and La Nina on the monthly maximum temperature and found that the El Nino and La Nina should be added to the weather forecast variables because of their major effects.

In a time series study during 1971–2006 by Paz (2009) analyzed the possible association between the cholera rates in south-eastern Africa and the annual variability of air temperature and sea surface temperature (SST) at regional and hemispheric scales. Using Poisson regression time series analysis, associations have been found between the annual increase of the air temperature in south eastern Africa and the cholera incidence increase in the same area. A significant exponential increase of cholera rates in humans during the study period has been confirmed.

Bhandari et al. (2012) assessed the relationship between climatic variables, and diarrhea and to find out the range of non-climatic factors that can confound the relationship of climate change and human health. They conducted trend analysis based on correlation. Time Series analysis was also conducted. Statistically significant correlation between occurrence of diarrheal cases, temperature and rainfall were observed. However, climate variables were not the significant predictors of diarrheal occurrence.

Another study by Bhandari et al. (2012) in time series study investigated the question "is Climate Change likely to worsen the public health threat of diarrheal disease in Botswana?". A forward-backward stepwise variable selection approach was used to minimize the Bayesian information criterion for model selection. There was a strong positive autocorrelation ($p < 0.001$) in the number of reported diarrhea cases at the one-month lag level. Climatic variables (rainfall, minimum temperature, and vapour pressure) predicted seasonal diarrheal with a one-month lag in variables ($p < 0.001$).

In a time series study by Alexander et al. (2013) found both of IOD and ENSO have strong effects on the climate of Asian region. The impact of natural disasters in recent past is increasing and particularly in south Asian region.

A study conducted by Cazelles et al. (2005) used wavelet approach and found that association between ENSO events and incidence of dengue is non-stationary from 1986 to 1992 and appears to have a major influence on the synchrony of dengue epidemics in Thailand.

Abdussalam (2016) used generalized additive model (GAM) and revealed the growing concern of the potential impact of climate change on the dynamic of infectious diseases in the future. The study finds that temperature, rainfall and humidity may facilitate both the transmission and the development of invasive cholera. Identifying, quantifying and improving disease surveillance will likely enhance our understanding and ability to predict and reduce diarrheal incidence.

Socio-demographic and climate studies done by Budyanra (2017) found that all explanatory variables: age, nutritional status, maternal education level, maternal age, maternal behavior of washing hands, access to sources of drinking water, access to sanitation facilities and household density significantly affect the incidence of children diarrhea.

McGregor and Ebi (2018) showed the degree of relationship between an ENSO tele-connection index and a time series of incidence data for a specific climate-sensitive disease. The study highlighted to move beyond a purely statistical treatment of climatic-health variability associations to one where diagnostic analyses are undertaken to identify the underlying climate mechanisms that form the cascade of processes that link ocean-atmosphere interactions with health.

Modarres and Ouarda (2013) investigated ENSO factors and revealed that ENSO is the most energetic climate signal. Any change in ENSO characteristics will have serious consequences for the global climate system. They observed that ENSO was more dynamic and uncertain in recent decades. In addition, the non-linearity and non-stationary of the SOI volatility increases in recent decades.

Walker et al. (2012) in a review study remarked that diarrhea incidence rates might be declining slightly, but yet study and funds are needed to improve both prevention and treatment practices in low- and middle-income countries.

In an empirical study on diarrhea and climate based on modeling method, Kolstad and Johansson (2010) observed that uncertainties are associated with future projections of diarrhea and climate change. They noticed that these uncertainties can be attributed primarily to the sparsely of empirical climate–health data. They also highlighted the need for empirical data in the cross section between climate and human health.

2.3. Research Gap

Knowledge sharing with the present study based on review of literature(s), some important motivations is given below:

- **First:** Most of the existing studies are based on the impact of local climatic factors on diarrhea, mostly used the temperature and rainfall data collected from limited ground monitors. Although a very few studies based on global climatic driver factors are examined but they are limited to normally distributed linear association based study. Therefore, none of the studies are noticed where infectious diarrheal pattern have been modeled with respect to ocean-atmospheric multivariate climatic index(s);
- **Second:** In terms of statistical methodology, most of the models used in the previous literatures are mean model for continuous time series data. But links between disease and climatic factors (Global/local) are non-linear, non-normal and volatility (Conditional variance) based. No study forecasts the future burden of diarrhea in relation to climatic driver factors under count time series data features of volatility in statistical distribution format.

2.4. Conclusion

The conventional statistical study was carried out by many researchers throughout the world. But the conventional method with shrinkage tools in integer valued distributional model in multivariate arena is still not found as an efficient technique. Effort is required to move seasonal health forecasting beyond the proof of concept phase through establishing a model of ocean atmospheric impacts in temporal probabilistic frameworks.

CHAPTER THREE
Materials and Methods

Chapter Three
Materials and Methods

3.1 Introduction

This chapter introduces some basic concepts of modeling that will assist to choose appropriate methodology and make guidelines for material collection. Finding and testing efficient method(s) are very important towards any research study and so included the chapter. This chapter will explore research process in the context of research questions and objectives.

The analysis of experimental data that have been observed at different points in time leads to new and unique problems in statistical modeling and inference. The obvious correlation introduced by the sampling of adjacent points in time can severely restrict the applicability of the many conventional statistical methods traditionally dependent on the assumption that these adjacent observations are independent and identically distributed. The systematic approach by which one goes about answering the mathematical and statistical questions posed by these time correlations is commonly referred to as time series analysis.

Modeling and forecasts plays an important part in both decimal and integer valued time series cases. Integer valued data for the study are commonly found in different areas including biomedicine, public health and finance, the serious development of count time series models only seen recently, see for example, Xu et al. (2012) and Weiss (2018). It is, however, preferable to work with integer valued time series data model with simpler theory but easier to be implemented.

Integer valued time series analysis and its application have become increasingly important in various fields of research. An important example from epidemiology is the monthly number of registered infections by certain agents, which is routinely collected by public health authorities. Important objectives of such data analysis are the prediction of future values for adequate planning of resources, the detection of unusual values pointing at some epidemics or the proper description of e.g. seasonal patterns for better understanding and interpretation of data generating mechanisms. Examples from other than patient counts are the number of stock market transactions per minute, from finance, or the hourly number of defect items, from industrial quality control. It is no doubt as an interesting and very important area of study that one can be overwhelmed easily by its scope, theories, application and achievements. Thus before delving into the subject area of the research, we will firstly refer to the defensive tools behind the methods adopted for the study.

3.2 Methods

3.2.1 *Time series*

A time series is a collection of observations measured sequentially through time. These measurements may be made continuously through time or to be taken in a discrete set of time points (Cryer and Kellet, 1991). Mathematically it may be defined by the following functional relationship $U_t = f(t)$, where U_t is the value of the study variable under consideration at time t. There are four types of time series data. They are as follows:

- Continuous time series
- Count time series
- Sampled time series and
- Accumulated time series.

Time series modeling is a dynamic research area which has been attracted by the research community over last few decades. The main aim of time series modeling is to carefully collect and rigorously study the past observations of a time series to develop an appropriate model which describes the inherent structure of the series.

This model is then used to generate future values for the series, i.e. to make forecasts. Time series forecasting thus can be termed as the act of predicting the future by understanding the past (Raicharoen, et al., 2003). The main goals of time series analysis are as follows:

- Descriptive: identify patterns in correlated data trends and seasonal variation.
- Explanation: understanding and modeling the data.
- Forecasting: prediction of short-term trends from previous patterns.
- Quality control: deviations of a specified size indicate a problem (Gottman, 1981).

3.2.2 *Count Time Series*

The word “count” is typically used as a verb meaning to enumerate units, items, or events. We might count the number of road killings observed on a stretch of highway, how many patients died/admitted at a particular hospital within hours/day/week/ month of having a disease/problem, or how many separate sunspots were observed in March 2013. “Count Data,” on the other hand, is a plural noun referring to observations made about events or items that are enumerated.

In statistics, count data refer to observations that have only nonnegative integer values ranging from zero to some greater undetermined value. Theoretically, counts can range from zero to infinity, but they are always limited to some lesser distinct value.

When the data being modeled consist of a large number of distinct values, even if they are positive integers, many statisticians prefer to model the counts as if they were continuous data.

Generally, there are four types of count variables. They are:

- A count or enumeration of events
- A count of items or events occurring within a period of time or over a number of periods
- A count of items or events occurring in a given geographical or spatial area or over various defined areas
- A count of the number of people having a particular disease, adjusted by the size of the population at risk of contracting the disease (Harris et al., 2014).

3.2.3 *Stochastic Process*

The term “stochastic” comes from the Greek word “stokhos” which means a target or bull’s eye. A random or stochastic process is a collection of random variables $\{X_t\}$ indexed by a set $T, i.e. t \in T$. If T consists of the integers (or a subset), the process is called a Discrete Time Stochastic Process and denoted by Y_t. If T consists of the real numbers (or a subset), the process is called Continuous Time Stochastic Process. If $T \in R^2$, then the process is called a Random Field. Similarly, these processes may take on values which are real or restricted to the integers and are called continuous state space or discrete state space accordingly.

3.2.4 *Unit Root Tests for Stationary*

In statistics, a unit root test whether a time series variable is non-stationary using an autoregressive model. A well-known test that is valid in large samples is the augmented Dickey-Fuller test. The optimal finite sample tests for a unit root in autoregressive models were developed by Jhon Denis Sargan and Alok Bhargava. Another test is Philips-Perron test. These tests use the existence of a unit root as the null hypothesis.

Time series can be characterized in many ways. First, we want to focus on the presence of trends in the time series. There are two types of trends, such as:

- Deterministic trend and
- Stochastic trends

A stochastic trend is a random walk, which may or may not contain deterministic or stochastic drift. A time series that contain a random walk process is termed as a unit root process.

A. Dickey-Fuller (*DF*) *Test*

We analyze the case where a unit root test is based on a Dickey-Fuller regression the only deterministic term of which is fixed intercept. Suppose, however, as could well be a case, that the actual data-generating process includes a broken linear trend. It is shown theoretically, and verified empirically, that under the I(1) null and I(0) alternative hypothesis the DF test can display a wide range of different characteristics depending on the nature and location of the break.

In statistics, the DF test is whether a unit roots is present in autoregressive model. It is named after the statisticians Dickey and Fuller (1970), who developed the test. The easiest way to introduce this test is to consider the following model:

$$Y_t = Y_{t-1} + \varepsilon_t \tag{1}$$

where ε_t is the stochastic error term that follows the classical assumptions, namely, it has zero mean, constant variance σ^2, and is non autocorrelated. Such an error is also known as white noise error term, the equation (1) is a first order or AR(1), regression in that we regress the value of y at time *t* on its value at time (*t* - 1). Now if the coefficient of Y_{t-1} is in fact equal to 1, we face the problem that is known as the unit root problem *i.e.*, a non-stationary situation.

Therefore we run the following regression

$$Y_t = \rho Y_{t-1} + \varepsilon_t \tag{2}$$

And actually find that $\rho = 1$, then we say that the stochastic variable Y_t has a unit root. In (time series) econometrics a time series that has a unit root is known as random walk model. And random walk is an example of non-stationary time series. Equation (2) is often expressed in an alternative form as

$$\nabla Y_t = (\rho - 1)Y_{t-1} + \varepsilon_t$$

$$\Longrightarrow \nabla Y_t = \delta Y_{t-1} + \varepsilon_t \tag{3}$$

where, $\delta = (\rho - 1)$ and ∇ is known as first difference operator. Note that

$$\nabla Y_t = Y_t - Y_{t-1}$$

Making use the definition, we can easily see that (2) and (3) are the same. However, now the null hypothesis is that $\delta = 0$. If δ is in fact to 0, we can write (3) as

$$\nabla Y_t = (Y_t - Y_{t-1}) = \varepsilon_t \tag{4}$$

where (4) says that the first difference of a random walk time series (ε_t) are a stationary time series because by assumption ε_t is purely random. To find out if a time series is non-stationary, run the regression (2) and find out if $\hat{\rho} = 1$, or equivalently, estimate (3) and find out if $\hat{\delta} = 0$ on the basis of the say, the t-statistics. Unfortunately, the t value thus obtained does not follow student's t distribution even in large samples.

Under the null hypothesis that $\rho = 1$, the conventionally computed t statistics is known as the $\tau(tau)$ statistic whose critical values have been tabulated by Dickey and Fuller on the basis of Monte Carlo simulations. In the literature tau (τ) test is known as the DF test. If the null hypothesis $\rho = 1$ is rejected (i.e., time series is stationary), we can use the usual student t test.

If the computed absolute value of the τ statistic (i.e., $|\tau|$) exceeds the DF absolute critical τ values then we do not reject the null hypothesis that the given time series is stationary. For the theoretical and practical reasons, the DF test is applied to regression in the following form:

$$\nabla Y_t = \delta Y_{t-1} + \varepsilon_t \tag{5}$$

$$\nabla Y_t = \beta_1 + \delta Y_{t-1} + \varepsilon_t \tag{6}$$

$$\nabla Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \varepsilon_t \tag{7}$$

where *t* is the time or trend variable. In each case the null hypothesis is that $\delta = 0$ that is there is unit root. The difference between (5) and the other two regressions lies (equation 6 and 7) in the conclusion of the constant (intercept) and the trend term.

B. Augmented Dicky-Fuller (*ADF*) *Test*

The most widely used test for non-stationary is the augmented ADF unit root test developed by Dicky and Fuller (1979 and 1981). They consider three different regression equations that can be used to test the presence of unit root Y_t:

$$\nabla Y_t = \delta Y_{t-1} + \sum_{j=1}^{J} \varphi_j \nabla Y_{t-j} + \mu + \beta t + \varepsilon_t \tag{8}$$

$$\nabla Y_t = \delta Y_{t-1} + \sum_{j=1}^{J} \varphi_j \nabla Y_{t-j} + \mu + \varepsilon_t \tag{9}$$

$$\nabla Y_t = \delta Y_{t-1} + \sum_{j=1}^{J} \varphi_j \nabla Y_{t-j} + \varepsilon_t \tag{10}$$

In equation 8, 9, and 10 the dependent variable is ∇Y_t. This implies that Y_t has a unit root if $\delta = 0$. Hence, the null hypothesis of the test equation 8, 9 and 10 states non-stationary of $Y_t: H_0: \delta = 0$ $(Y_t \ has\ a\ unit\ root)$. The three equations differ in the deterministic regressor. The choice between the three equations is an important issue in unit root testing. One problem is that the additional estimated parameters reduce the degree of freedom and the power of the test. Reduced power means that the researcher

may conclude that the process contains a unit root where it is not the case. The second problem is that an appropriate statistic for testing $\delta = 0$ depends on which regressors are included in the equation. For example, if the data generating process includes a deterministic trend, omitting the term β_t give an upward bias in the estimated value of additional regressor(s), however, increasing the absolute value of the critical values so that the researcher may fail to reject the null of a unit root.

The test is implemented through the usual *t* statistic of $\widehat{\delta}$. The *t* statistic of the three models are denoted as t_τ, t_μ and *t* respectively. Alternatively, Dickey and Fuller (Dickey & Fuller, 1979) suggested F statistics to test the joint hypothesis $\delta = \beta = \mu = 0$ and $\delta = \beta = 0$ in equation (8) and the joint hypothesis $\delta = \mu = 0$ in equation (9), denoted as (φ_1). Under the null hypothesis of non-stationarity the *t* statistic and t_τ, t_μ and F-statistic do not have the standard t and F-distributions, but are functions of Brownian motions. Critical values of the asymptotic distributions of these t-statistics are provided by Fuller (1976) and have recently been improved by MacKinnon (1991) through larger sets of replications. Dickey and Fuller (1981) list critical values for the F-statistics of $\varphi_1, \varphi_2 \; and \; \varphi_3$. Dolado et al. (1990) develop a systematic testing strategy between the alternative equations in the unit root testing procedure consists of the following steps:

***Step*-1:** In the most unrestricted equation (8) the null hypothesis of stationary is tested with t_τ. If the null hypothesis is rejected, variable Y_t is trend stationary and there no need to proceed any further.

***Step*-2:** If the null hypothesis is not rejected, we test for the significance of the deterministic trend under the null hypothesis $\delta = \beta = 0$ nusing the statistic φ_3. If it is significant, the present of the unit root can be tested again, nothing that t-statistic follows now a standard t-distribution.

***Step*-3:** If δ and β are jointly insignificant in equation (8), we estimate the equation without the deterministic trend, (9) and test for the unit root using t_μ and it's critical values. If the null hypothesis is rejected, we may stop again and conclude that variable Y_t is stationary.

***Step*-4:** If the null hypothesis is not rejected, we test for the significance of the constant term under the null hypothesis $\delta = \mu = 0$ using φ_1. If the constant term is significant, we test for the unit root using the standard normal distribution.

***Step*-5:** If δ and β are jointly insignificant in equation (9), we estimate equation (10) and test for the presence of a unit root. The process ends either with the result that variable Y_t is stationary or that Y_t contain a unit root.

If we can't reject the null hypothesis in any of the steps of the strategy, we conclude that Y_t is non-stationary and needs to be differenced at least one to become stationary. To detect the order of integration *d*, of the series Y_t we proceed by testing the difference series until the unit root hypothesis is rejected. So, if Y_t is found to be non-stationary and ∇Y_t is found to be stationary then Y_t is called 'integrated of order 1 (denoted as $I(1)$). If we can only reject the null of a unit root after differencing *d* times, we conclude that the series is integrated of order *d*. Stochastic trends in marketing are often linear and sometimes quadratic, so *d* rarely exceeds 2 (Horváth et al., 2002). The number of lags (L) in equations (8), (9) and (10) is often determined by the AIC, BIC or by a selection procedure advocated by (Perron, 1989). The ADF test assumes that the variable under considerations is continuous and can take any real values.

C. Phillips-Perron (*PP*) *Test*

In statistics, the Phillips–Perron (1988) test is a unit root test which is used in time series analysis to test the null hypothesis that a time series is integrated of order 1. It builds on the DF test of the null hypothesis $\rho = 0$ in $\nabla Y_t = \rho Y_{t-1} + \varepsilon_t$, where ∇ is the first difference operator. Like the ADF test, the PP test addresses the issue that the process generating data for Y_t might have a higher order of autocorrelation than is admitted in the test equation - making Y_{t-1} endogenous and thus invalidating the DF t-test. Whilst the augmented DF test addresses this issue by introducing lags of ∇Y_t as regressors in the test equation, the PP test makes a non-parametric correction to the t-test statistic. The test is robust with respect to unspecified autocorrelation and heteroskedasticity in the disturbance process of the test equation. Davidson and MacKinnon (2004) report that the PP test performs worse in finite samples than the ADF test (Davidson & MacKinnon, 2004).

D. Kwiatkowski-Phillips-Schmidt-Shin (*KPSS*) *Test*

The KPSS (1992) test differs from the other unit root tests described here in the series Y_t is assumed be stationary under the null hypothesis. The KPSS statistic based on the residuals from the OLS regression of y_t on the exogeneous variable x_t:

$$y_t = x_t\delta + u_t$$

The Lagrange multiplier (LM) statistic is defined as:

$$LM = \sum_t \frac{\left(s(t)\right)^2}{T^2 f_0}$$

where f_0 is an estimator of the residual spectrum at frequency zero and $s(t)$ is a cumulative residual function : $S(t) = \sum_{r=1}^{t} \hat{u}_t$

Based on the residual $\hat{u}_t = y_t - x_t\,\hat{\delta}(0)$. We point out that the estimator of δ used by GLS detrending since it is based on a regression involving the original data, and not on the quasi-differenced data. To specify the KPSS test, we must specify the set of exogenous regressors x_t and a method for estimating f_0. The reported critical values for the LM test statistic are based upon the asymptotic results presented in KPSS.

3.2.5 *Test of Normality*

The checking of normality of a residual is very important because a standard assumption is that the random shocks are normality distributed. This permits us to perform approximate *t*-tests on coefficient significance at the estimation stage. One way to check the normality is to examine the histogram of the residuals. Another is to plot residuals in normal probability plot (Cook & Weisberg, 1983). Both of the procedure provide a helpful graphical presentation of the data, they do not, however, provide any formal test of normality. There are various formal tests for normality. The most recent and efficient test for normality of the fitted residuals is Jarque-Bera test for normality.

A. Jarque-Bera (JB) Test

Firstly we assume normality of the error term of the model. In finite samples the hypothesis testing and confidence interval estimation of the parameters of the model equation relied on the normality assumption. The assumption of normality of errors

can be tested by JB test (1982). JB test is a two degrees of freedom χ^2 test based on the skewness and kurtosis. For a normal distribution skewness=0 and kurtosis=3.

From sample data of one can compute these as follows:

Let x_1, x_2, …, x_T be a random sample and its mean $\bar{x}$ and compute the following moments

$$\mu_2 = \frac{1}{T}\sum(x_i - \bar{x})^2, \mu_3 = \frac{1}{T}\sum(x_i - \bar{x})^3, \mu_4 = \frac{1}{T}\sum(x_i - \bar{x})^4 .$$

Skewness = μ_3^2 / μ_2^3 and Kurtosis = μ_4 / μ_2^2

To test the following hypothesis

H_0: Skewness = 0 and Kurtosis = 3

The JB statistics is

$$JB = T\left(\frac{Skewness}{6} + \frac{(Kurtosis - 3)^2}{24}\right) \sim \chi_2^2$$

If JB > $\chi^2_{(\alpha,2)}$, then the decision rejects the null hypothesis meant that data don't follow normal distribution at 100(1-α)% level of significance with 2 df.

3.2.6 *Stationary Process Diagnostics*

When an AR(p) process is represented as $\varepsilon_t = \varphi(L)y_t$, then $\varphi(L) = 0$ is known as the characteristic equation for the process. It is proved by Box, Jenkins, Reinsel, & Ljung (2015) that a necessary and sufficient condition for the AR(p) process to be stationary is that all the roots of the characteristic equation must fall outside the unit circle. Hipel and McLeod (1994) mentioned another simple algorithm by Joseph and Pagano (1981) for determining stationary features of an AR process. For example as shown AR(1) model $y_t = c + \varphi_1 y_{t-1} + \varepsilon_t$ is stationary when $|\varphi| < 1$, with constant mean $\mu = \frac{c}{1-\varphi_1}$ and constant variance $\gamma_0 = \frac{\sigma^2}{1-\varphi_1^2}$.

An MA(q) process is always stationary, regardless of the values the MA parameters (Hipel & McLeod, 1994). The conditions regarding stationary and invertible of AR and MA processes also hold for an ARMA process. An ARMA(p, q) process is stationary if all the roots of the characteristic equation $\varphi(L) = 0$ lie outside the unit

circle. Similarly, if all the roots of the lag equation $\theta(L) = 0$ lie outside the unit circle, then the ARMA(p, q) process is invertible and can be expressed as a pure AR process. Widely used few diagnostic tools are presented below -

A. Autocorrelation (ACF) and Partial Autocorrelation Functions (PACF)

To determine a proper model for a given time series data, it is necessary to carry out the ACF and PACF analysis. These statistical measures reflect how the observations in a time series are related to each other. For modeling and forecasting purpose it is often useful to plot the ACF and PACF against consecutive time lags. These plots help in determining the order of AR and MA terms.

For a time series the auto-covariance at lag *k* is defined as

$$\gamma_k = cov(x_t, x_{t+k}) = E[(x_t - \mu)(x_{t+k} - \mu)]$$

The Autocorrelation Coefficient at lag k is defined as:

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

Here μ is the mean of the time series, i.e. $\mu = E[x_t]$. The auto-covariance at lag zero i.e. γ_0 is the variance of the time series. From the definition it is clear that the autocorrelation coefficient ρ_k is dimensionless and so is independent of the scale of measurement. Also, clearly $-1 < \rho_k < 1$. Box and Jenkins (Box et al., 2015) termed ρ_k as the theoretical Autocorrelation Function (ACF).

Another measure, known as the Partial Autocorrelation Function (PACF) is used to measure the correlation between an observation *k* period ago and the current observation, after controlling for observations at intermediate lags (i.e. at lags < k). At lag 1, PACF(1) is same as ACF(1). The detailed formulae for calculating PACF are given in (Box et al., 2015; Hipel & McLeod, 1994).

Normally, the stochastic process governing a time series is unknown and so it is not possible to determine the actual or theoretical ACF and PACF values. Rather these values are to be estimated from the training data, i.e. the known time series at hand. The estimated ACF and PACF values from the training data are respectively termed as sample ACF and PACF. As given in (Box et al., 2015), the most appropriate sample estimate for the ACF at lag k is $\gamma_k = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \mu)(x_{t+k} - \mu)$

3.2.7 *Mean Model*

In general models for time series data can have many forms and represent different stochastic processes. There are two widely used linear time series models in literature, viz. Autoregressive (AR) and Moving Average (MA) models. For seasonal time series forecasting, a variation of ARIMA, viz. the Seasonal Autoregressive Integrated Moving Average (SARIMA) (Hamzaçebi, 2008) model is used. ARIMA model and its different variations are based on the famous Box-Jenkins principle and so these are also broadly known as the Box-Jenkins models (Box et al., 2015; Hipel & McLeod, 1994; Lee; G. P. Zhang, 2003). Non-linear models are appropriate for predicting volatility changes in time series (Parrelli, 2001). Many famous non-linear models have been suggested in literatures. Autoregressive Conditional Heteroskedasticity (ARCH) model and its variations like Generalized ARCH (GARCH), Exponential Generalized ARCH (EGARCH) etc. are widely used. The Threshold Autoregressive (TAR) model (Park, 1999), the Non-linear Autoregressive (NAR) (Zhang, 2007) model, the Nonlinear Moving Average (NMA) (Perrelli, 2001) model, etc. are also important for modeling.

A. Autoregressive Moving Average (*ARMA*) *Model*

An ARMA(p, q) model is a combination of AR(p) and MA(q) models and is suitable for univariate time series modeling. In an AR(p) model the future value of a variable is assumed to be a linear combination of p past observations and a random error together with a constant term. Mathematically the AR(p) model can be expressed as (Hipel & McLeod, 1994).

$$y_t = c + \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t$$

Here y_t and ε_t are respectively the actual value and random error (or random shock) at time period t, $\varphi_i (i = 1, 2, \cdots, p)$ are model parameters and c is a constant. The integer constant p is known as the order of the model. Sometimes the constant term is omitted for simplicity. Usually for estimating parameters of an AR process using the given time series, the Yule-Walker equations (Hipel & McLeod, 1994) are used. Just as an AR(p) model regress against past values of the series, an MA(q) model uses past errors as the explanatory variables. The MA(q) model is given by (Cochrane, 2005; Hipel & McLeod, 1994):

$$y_t = \mu + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t.$$

Here μ is the mean of the series, φ_j $(i = 1, 2, \cdots, q)$ are model parameters q is the order of the model. The random shocks are assumed to be white noise process, i.e. a sequence of independent and identically distributed random variables with zero mean and a constant variance σ^2. Generally, the random shocks are assumed to follow the typical normal distribution. Thus conceptually a moving average model is a linear regression of the current observation of the time series against the random shocks of one or more prior observations. Fitting an MA model to a time series is more complicated than fitting an AR model because in the former one the random error terms are not fore-seeable.

AR and MA models can be effectively combined together to form a general and useful class of time series models, known as the ARMA models. Mathematically an ARMA(p, q) model is represented as :

$$y_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}.$$

Here the model orders p, q refer to p autoregressive and q moving average terms. Usually ARMA models are manipulated using the lag operator notation. The lag or backshift operator is defined as $Ly_t = y_{t-1}$. Polynomials of lag operator or lag polynomials are used to represent ARMA models as follows:

AR(p) model : $\varepsilon_t = \varphi(L) y_t$.

MA(q) model : $y_t = \theta(L)\varepsilon_t$.

ARMA(p, q) model : $\varphi(L) y_t = \theta(L)\varepsilon_t$.

Here $\varphi(L) = 1 - \sum_{i=1}^{p} \varphi_i L^i$ and $\theta(L) = 1 + \sum_{j=1}^{q} \theta_j L_j$.

It is shown in (Hipel & McLeod, 1994) that an important property of AR(p) process is invertibility, i.e. an AR(p) process can always be written in terms of an MA(∞) process. Whereas for an MA(q) process to be invertible, all the roots of the equation $\theta(L) = 0$ must lie outside the unit circle. This condition is known as the invertible condition for an MA process.

B. Autoregressive Integrated Moving Average (*ARIMA*) *Model*

The ARMA models, described above can only be used for stationary time series data. However in practice many time series such as those related to socio-economic and business show non-stationary behavior. Time series, which contain trend and seasonal patterns, are also non-stationary in nature (Faraway & Chatfield, 1995; Flaherty & Lombardo, 2000; Hamzaçebi, 2008). Thus, from application view point ARMA models are inadequate to properly describe non-stationary time series, which are frequently encountered in practice. For this reason the ARIMA model (Flaherty & Lombardo, 2000) is proposed, which is a generalization of an ARMA model to include the case of non-stationary as well.

In ARIMA models a non-stationary time series is made stationary by applying finite differencing of the data points. The mathematical formulation of the $ARIMA_{(p,d,q)}$ model using lag polynomials is given below :

$$\varphi(L)(1-L)^d y_t = \theta(L)\varepsilon_t$$

$$i.e.\left(1-\sum_{i=1}^{p}\varphi_i L^i\right)(1-L)^d y_t = \left(1+\sum_{j=1}^{q}\theta_j L^j\right)\varepsilon_t.$$

Here, p, d and q are integers greater than or equal to zero and refer to the order of the autoregressive, integrated, and moving average parts of the model respectively.

- The integer d controls the level of differencing. Generally d=1 is enough in most cases. When d=0, then it reduces to an ARMA(p,q) model.
- An ARIMA(*p*,0,0) is nothing but the AR(*p*) model and ARIMA(0,0,*q)* is the MA(*q*) model.
- ARIMA (0,1,0), i.e. $y_t = y_{t-1} + \varepsilon_t$ is a special one and known as the Random Walk model. It is widely used for non-stationary data.

C. Seasonal Autoregressive Integrated Moving Average (*SARIMA*) *Model*

Box and Jenkins have generalized the ARIMA model to deal with seasonality and proposed model is known as the Seasonal ARIMA (SARIMA) model. In this model seasonal differencing of appropriate order is used to remove non-stationary from the series. A first order seasonal difference is the difference between an observation and the corresponding observation from the previous year and is calculated as $z_t = y_t - y_{t-s}$, where z_t is the seasonally differenced series. For monthly time series s=12 and

for quarterly time series s=4. This model is termed as $SARIMA(p,d,q) \times (P,D,Q)[s]$ model.

3.2.8 *Volatility*

Volatility is a measure of the intensity of unpredictable changes in time series and it is commonly time varying dependent as recognized by Baillie (1997) among others and can be think as a random variable that follows a stochastic process. The task of any volatility model is to describe the historical pattern of volatility and possibly use this to forecast future volatility. An important characteristic of time series is that the periods of high volatility tend to be more persistent than periods of lower volatilities. Another stylized effect is that the data series exhibit non-normality and excess of kurtosis Engle and Ng (1993). The volatility has some important features:

- The volatility changes over time and that jumps in the volatility are unusual.
- The volatility is calculated by the assumption that it has a Geometric Brownian motion.
- The volatility appears in clusters.
- The volatility does not grow to infinity.
- The volatility reacts different on a drop in the prices of the underlying asset then it does for an increase in the price of the underlying asset (Grek, 2014).

In climatology, volatility is the degree of changeability or variance in ambient climatic conditions over time. High volatility in climate means that climate conditions can change dramatically over a short period of time, whereas lower volatility means that climate conditions do not oscillate dramatically in the short-term. Uncertainty and risk are higher during more volatile climatic intervals. From an adaptive standpoint, gradual changes in mean conditions over a century (e.g. from wet to dry conditions) present a very different set of challenges than dramatic inter-annual or decadal changes between extreme wet and dry conditions. Our primary argument is that highly volatile climate conditions make it difficult for individuals or institutions to determine the costs and benefits of one strategy over another.

To evaluate the impact of climatic volatility on the rise and fall, we examine published climate records. We measure climatic volatility through time by using the empirical dispersion index,

$$F = (Q_3 - Q_1) / (Q_3 + Q_1)$$

where Q_1 is the first and Q_3 is the third quartile of any data vector .The dispersion index calculated is also called the Quartile dispersion coefficient. Higher F-values indicate greater climatic volatility.

Conditional volatility features in time series do not comply with the assumptions of the least squares method. One of the essential propositions of least squares models is homoscedasticity that assumes that the standard deviation of error terms is constant. This implies that the expected value of all error terms, when squared, is the same at any point. Data, in which the variance of error is changing, where the error terms may change for some ranges of data, are heteroskedastic by definition. When dealing with heteroskedasticity using the Ordinary Least Squares (OLS) regression, the estimated standard errors and confidence intervals tend to be too thin. In other words, a sequence or a vector of a random variable is defined as heteroskedastic if there are different variances for the random variables. One of the most powerful tools to deal with heteroskedasticity is using the idea of autoregressive heteroskedastic process or ARCH-type models (GARCH-family models), which fill the gap in both academic and practical literature.

3.2.9 *Volatility Model*

Nonlinear models are considered for better time series analysis and forecasting. Campbell, Lo and McKinley (1997) made important contributions towards this direction. According to them almost all non-linear time series can be divided into two branches: one includes models nonlinear in mean and other includes models non-linear in variance. As an illustrative example, here we present two nonlinear time series models.

- Non-linear Moving Average (NMA) Model: $y_t = \varepsilon_t + \alpha\varepsilon_{t-1}^2$. This model is non-linear in mean but not in variance.
- Eagle's (1982) ARCH Model: $y_t = \varepsilon_t + \alpha\sqrt{\varepsilon_{t-1}^2}$ is heteroskedastic. This model has several other variations, like GARCH, EGARCH etc.

A. Autoregressive Conditional Heteroscedasticity (*ARCH*) *Model*

Autoregressive Conditional Heteroscedasticity model denoted by ARCH is the first model that provides a systematic frame work for volatility modeling, the basic idea of ARCH model(s) is that:

- the residual $\{\varepsilon_t\}$ is serially uncorrelated, but dependent.
- the dependence of $\{\varepsilon_t\}$ can be described by a simple quadratic function of its lagged values.

In ARCH models the conditional variance has a structure of the conditional expectation in an AR model. ARCH(q) model is one for which the variance at time *t* is conditional on observations at the previous q times.

Let us assume $\{r_t\}$ be the yield on an index series is

$$r_t = \mu + \sigma_t \varepsilon_t$$

where ε_t ~ i.i.d N(0,1). The residual yield at time *t*, $r_t - \mu$ is

$$\alpha_t = \sigma_t \varepsilon_t.$$

In an ARCH(1) model, first developed by Engle (1982), $\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2$, where $\alpha_0 > 0$ and $\alpha_1 \geq 0$ to positive variance and $\alpha_1 < 1$ for stationary. Under an ARCH(1) model, if the residuals yield, α_t is large in magnitude, our forecast for next period's conditional volatility, σ_{t+1} will be large. This model returns are conditionally normal. Also, the yields, r_t are uncorrelated but are not i.i.d. For time varying σ^2 will lead to longer tails, relative to a normal distribution, in the unconditional distribution of α_t (Campbell & Andrew, 1997). The ARCH(1) is a special case of ARCH(q) and therefore what applies for ARCH(q) also applies for ARCH(1).

B. Generalized Autoregressive Conditional Heteroscedasticity (*GARCH*) *Model*

The GARCH is an extension of the ARCH model similar to the extension of an AR to ARMA process. An ARMA model may have fewer parameters than AR model, similarly a GARCH model may contain fewer parameters as compared to an ARCH model, and thus a GARCH model may be preferred to an ARCH model.

In real world problem, the conditional variance or volatility changes with time and this is typical for many types of time series. Such variations are known as

heteroscedastic. It makes harms in predicting the future volatility pattern whether it is increasing or decreasing per time period. Therefore, Engle (1982) introduced an ARCH model namely ARCH(p) originally to describe U.K. inflationary uncertainty as a function of past errors. It is called autoregressive since the model allows an AR (p) structure for the conditional heteroscedasticity with $\varepsilon_t \sim N(0, \sigma_t^2)$. The model is defined as

$$r_t = \varepsilon_t \sigma_t; \forall t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2. \quad (11)$$

where $\alpha_0 > 0,\ \alpha_i \geq 0$ for $i = 1, 2, \dots, p$ and σ_t^2 is the variance of the process.

Followed by ARCH, Generalized Autoregressive Conditional Heteroscedasticity GARCH(p,q)) was first introduced by Bollerslev (1986) to solve the problems encountered in ARCH (p) model with higher order by allowing more flexible lag structure. It is sometimes useful to consider the ARCH(∞) representation of a GARCH(p, q) process or one can say that the GARCH model is an infinite order ARCH model and often provides a highly parsimonious lag shape. Using the same approach corresponding to ARMA models, this extension of ARCH process is used to reduce the infinite number of estimated parameter. The GARCH(p, q) process uses values of the past squared observations and past variances to model the variance at time *t,* where the process $\{r_t\}$ conditioned on the past value, the σ-field generated from $r_{t-1}, r_{t-2}, \dots, r_1$ with h_{t-1} follows normal distribution, mean zero and variance, h_t. It can be defined as

$$\left(r_t \middle| h_{t-1}^{\varepsilon_t}\right) \sim N(0, \sigma_t) \quad (12)$$

$$h_t = \gamma_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j} \quad (13)$$

Since the mid-1980s, these models have become actively discussed and studied among both academics researchers and practitioners and lead to several extensions of GARCH (p,q) being developed. For example, integrated GARCH model (IGARCH) model was suggested by Engle & Bollerslev (1986) whereby the model is defined to be integrated in variance. Such model is argued to be both theoretically important for the asset pricing models and empirically relevant. On the other hand, Nelson (Nelson, 1991) proposed the exponential GARCH model known as EGARCH in order to

capture the leverage effect of the stock market. Sentana (1995) introduced quadratic GARCH model (QGARCH) to cope with asymmetric effects of shocks on volatility. In addition, the parameters in the GARCH (p, q) model change according to the sign or the size of shock ε_i which lead the model being interpreted as a regime-switching model known as Markov-Switching GARCH (MSW-GARCH) model. This model developed by Gray (1996) assumes that the regime is determined by an unobservable Markov-process.

GARCH(1, 1) Model

In an ARCH(1) model, next period's variance only depends on last period's squared residual so a crisis that caused a large residual would not have the sort of persistence that we observe after actual crises. This has led to an extension of the ARCH model to a GARCH, or Generalized ARCH model, first developed by (Bollerslev, 1986) which is similar in spirit to an ARMA model. In a GARCH(1, 1) model,

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 h_{t-1}$$

where $\alpha_0 > 0, \alpha_1 > 0$, $\beta_1 > 0$, and $\alpha_1 + \beta_1 < 1$, so that our next period forecast of variance is a blend of our last period forecast and last period's squared return. We can see that just as an ARCH(1) model is an AR(1) model on squared residuals, an ARCH(1,1) model is an ARMA(1,1) model on squared residuals by making the same substitutions as before, $h_t = r_t^2 - v_t$

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 h_{t-1}\ \sigma_t^2 - v_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 (r_{t-1}^2 - v_{t-1})\sigma_t^2$$
$$= \alpha_0 + (\alpha_1 + \beta_1) r_{t-1}^2 + v_t - \beta_1 v_{t-1}$$

which is an ARMA(1, 1) on the squared residuals. The unconditional variance of r_t is

$$var(r_t) = \mathrm{E}(r_t^2) - (\mathrm{E}[r_t])^2 = \mathrm{E}(r_t^2) = E(\sigma_t^2 \varepsilon_t^2) = E(\sigma_t^2)$$
$$= \alpha_0 + \alpha_1 E[r_{t-1}^2] + \beta_1 \sigma_{t-1}^2 = \alpha_0 + (\alpha_1 + \beta_1) E\lceil \sigma_{t-1}^2 \rceil$$

And since r_t is a stationary process,

$$var(r_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

And since, $r_t = \sigma_t \varepsilon_t$, the unconditional variance of returns, $E[\sigma^2] = E[r_t^2]$, is also $\alpha_0/(1 - \alpha_1 - \beta_1)$. Just as an ARMA(1, 1) can be written as an AR(∞), a GARCH(1, 1) can be written as an ARCH(∞),

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \sigma_t^2$$
$$= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1(\alpha_0 + \alpha_1 \varepsilon_{t-2}^2 + \beta_1 \sigma_{t-2}^2) \quad = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$$
$$+ \alpha_0\beta_1 + \alpha_1\beta_1 \varepsilon_{t-2}^2 + \beta_1^2 \sigma_{t-2}^2 \quad = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_0\beta_1$$
$$+ \alpha_1\beta_1 \varepsilon_{t-2}^2 + \beta_1^2(\alpha_0 + \alpha_1 \varepsilon_{t-3}^2 + \beta_1 \sigma_{t-3}^2)$$

and so on, we get

$$\sigma_t^2 = \frac{\alpha_0}{1-\beta_1} + \alpha_1 \sum_{i=0}^{\infty} \varepsilon_{t-1-i}^2 \beta_1^i$$

So that the conditional variance at time t is the weighted sum of past squared residuals and the weights decrease as you go further back in time. Since the unconditional variance of returns is $\mathrm{E}[r_t^2] = \alpha_0/(1 - \alpha_1 - \beta_1)$

It is often useful not only to forecast next period's variance of returns, but also to make an l-step ahead forecast, especially if our goal is to price an option with l steps to expiration using our volatility model. Again starting from the GARCH(1, 1) equation for σ_t^2, we can derive our forecast for next period's variance, $\hat{\sigma}_{t+1}^2$

$$\hat{\sigma}_{t+1}^2 = \alpha_0 + \alpha_1 E[(\varepsilon_{t-1}^2 | I_{t-1})] + \beta_1 \sigma_t^2$$

According to the above equation(s), it is clear that $\hat{\sigma}_{t+1}^2 \to \sigma^2$ as $t \to \infty$ so as the forecast horizon goes to infinity, the variance forecast approaches the unconditional variance of ε_t. From the 1-step ahead variance forecast, we can see that $(\alpha_t + \beta_t)$ determines how quickly the variance forecast converges to unconditional variance. If the variance spikes up during a crisis, the number of periods *k*, until it is halfway between the first forecast and the unconditional variance is $(\alpha_1 + \beta_1)k = 0.5$, so the half-life is given by $k = ln(0.5)/\, ln(\alpha_1 + \beta_1)$.

3.2.10 *Wavelet Method*

Wavelet transforms allow to decompose time series into low frequency and high frequency data, there by exposing trends, break-down points, and discontinuities in the data that other signal analysis methods might miss (Adamowski et al., 2009; Nalley et al,. 2013). Another advantage of wavelet transformations is the flexibility of choice in selecting a mother wavelet according to the properties of the time series (Conraria & Soares 2011; Pingale et al., 2014).

Wavelet analysis is a popular tool due to its capability to reveal evidence within the indication in both the time and scale domains (Nourani et al., 2008). It is appropriate for non-stationary data. Most climatic time-series data are non-stationary; therefore, wavelet transformations are used for these kinds of data.

Continuous wavelet analysis (CWT) represents the sum over all time of the signal, multiplied by scale and shifted versions of the mother wavelet analysis ψ [Kim and Valdes, 2003] given by

$$\psi(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} x(t)\psi^* \left(\frac{t-\tau}{s}\right) dt \tag{14}$$

where, *s* is the scale parameter, *t* is time, and τ is the shift parameter. Each scale corresponds to the width of the wavelet. While a CWT is useful in processing various images and signals, it is seldom used for prediction as its calculations are intricate and lengthy in terms of time. As an alternative, in prediction applications, the discrete wavelet transform (DWT) is applied, due to its simplicity and shorter calculation time. The scales and shifts of the DWT are usually based on powers of two (dyadic scales and shifts). This is obtained by altering the wavelet transform:

$$\psi(m)_{j,m} = \frac{1}{\sqrt{|s_0^j|}} \sum_k \psi \left(\frac{k - m\tau_0 s_0^j}{s_0^j}\right) x(k) \tag{15}$$

where, *j* and *m* are integers which control the scale and shift, respectively, $s_0 > 1$ is a fixed expansion step, and τ_0 is a shift parameter based on the aforementioned expansion step. The impact of discretizing the wavelet transform is that the time-

space scale is sampled at discrete levels. The DWT functions like a pair of high-pass and low-pass filters. The time series is decomposed into one comprising its trend and one comprising the high frequencies and the fast events (Adamowski & Sun, 2010). In the present research, the detail coefficients and approximation sub-time series were obtained using Eq. (14 & 15).

3.2.11 *Link of GARCH and Wavelet*

General GARCH models can only describe the overall volatility features of the time series, but cannot describe partial volatility features of the series and multi-scale information. Ignoring partial volatility features and multi-scale information of the time series will affect the precision of the model. To take advantage of partial volatility features and multi-scale information of variables when forecasting, a hybrid model namely, W-GARCH has been proposed by wavelet translating and thereafter applying GARCH.

In this study, the influence of noise on the generalization power of GARCH processes with regard to the volatility function in oceanographic data index, namely MEI and IOD are investigated.

3.2.12 *Wavelet De-Noising*

Wavelet smoothing operators have been used behind threshold features, which consists to eliminate small wavelet coefficients which are not distinguishable from pure noise. The remained coefficients are used for a better reconstruction of the signal. The wavelet shrinkage principle (Kerkyacharian & Picard, 1993) is perfect for the scope and applies a threshold de-noising procedure to the data at hand by shrinking the wavelet coefficients toward zero so that only a limited number will be considered for reconstructing the original signal. Removing noise from the signal should allow for a better reconstruction, which might be crucial for climatic time series with unobservable volatility structure. Still, from the perspective of statistical inference, we are clearly employing a non-parametric procedure, which does not rely on particular assumptions about the nature of the function $f(t)$ and adopts a criterion similar to a locally adaptive band width in kernel estimation.

In summary, the following smoothing procedure is implemented:

- HDWT algorithm is applied to the data to get the empirical wavelet detail and smooth coefficients;
- Wavelet coefficients are shrunken toward zero by threshold features;
- Inverse HDWT is applied to the threshold wavelet coefficients to reconstruct the signal.

3.2.13 *Count Time series Model*

A. Poisson model

Observations of dependent counts can in many cases be modeled successfully through the Poisson distribution. Let $\{y_t\}$ denote a time series of counts taking non-negative integers values, where y_t is the response process. According to the conditional law, $\{y_t\}$ is specified by assuming that the conditional density of the response given the past is Poisson with mean λ

$$f(y_t; \lambda|F_{t-1}) = \frac{e^{-\lambda}\lambda^{y_t}}{y_t!} \quad ; \qquad y_t = 0, 1, 2, \dots\dots\dots$$

where λ is the mean and variance of the process.

However, in practice the variance of the errors could be larger than the mean (although it can also be smaller). When the variance is larger than the mean, two other extensions of the Poisson model are more suitable. In the over-dispersed Poisson model, an extra parameter is included to estimate how much larger the variance is than the mean. This estimated parameter is then used to correct the effects of the larger variance. The negative binomial distribution can be used as an alternative to the Poisson distribution.

B. Negative Binomial Model

The negative binomial distribution is a form of the Poisson distribution in which the distribution's parameter is itself considered as a random variable. The variation of this parameter can account for the variance of the data that is higher than the mean. In order to deal appropriately with over-dispersed Poisson count data, the link used for the negative binomial needs to be the same as that of the Poisson model, namely, the log link (Hilbe, 2011).

It is especially useful for discrete data over time. Since the negative binomial distribution has one more parameter than the Poisson, the second parameter can be used to adjust the variance independently of the mean and the variance is given by

$$v(y_t) = \mu_t + \frac{\mu_t^2}{\emptyset}$$

The density function written as

$$f(y_t; \mu_t, \emptyset) = \frac{gamma(\emptyset + y_t)}{gamma(y_t + 1) \times gamma(\emptyset)} \left(\frac{\mu_t}{\mu_t + \emptyset}\right)^{y_t} \left(\frac{\emptyset}{\mu_t + \emptyset}\right)^{\emptyset}, y_t = 1, 2, \dots$$

is a distribution of exponential family, and the random component vector y_t and the systematic component is given by

$$\eta_t = g(\mu_t) = \sum_{t=1}^{p} Z_t \beta$$

where $\eta_t = (\eta_1, \eta_2, \dots, \eta_n)^T$ is the linear predictor, $g(.)$ is the link function, which in this case is canonical. $Z_t = (z_{t_1}, z_{t_2}, \dots, z_{t_p})'$ is a vector that represents the explanatory variables and $\beta = (\beta_1, \beta_2, \dots, \beta_t)'$ the vector of regression parameters, usually estimated by the maximum likelihood method.

C. INGARCH(p, q) Model

The standard model for analyzing the conditional mean is INAR (Integer-valued AR) model using 'binomial thinning operator' (Hwang and Basawa, 2011). In addition, there are conditional Poisson, conditional binomial model, conditional magnetic logistic model (Fokianos, 2011).

The ARMA model or the GARCH model, which is a representative time series analysis model, is a model that deals with real-valued time series observed over time. It is not an approximate model when applied to the time series data having only integer value. It is injudicious to do. Thus, there is a need for an appropriate model for the volatility of the count time series with non-negative integer values, such as the number of infected individuals. Since the conditional variance changes over time in the time series data taking an integer value, it is often inappropriate to assume (conditional) assumption of equi-distribution that is not really sound. Ferland et al. (2006) used an integer-valued GARCH model with the Poisson distribution as the

conditional distribution of the time series, reflecting the need for a conditionally heteroskedasticity model for these time series data.

Many tend to over-dispersion (Zhu, 2012). Analysis of volatility of coefficient time series In order to deal with this and the scatter problem in the field, Zhu (2011) proposed a conditional distribution, Negative Binomial INGARCH (INGARCH-NB). In this research various GARCH models with various coefficients are introduced and applied them to the data on the number of infectious diarrheal patients (count time series).

The standard GARCH model, which is a model for time series volatility analysis, is defined as follows.

$$(\varepsilon_t|F_{t-1})\sim N(0,\sigma_t^2),$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

Here, the data is $\{\varepsilon_t\}$ and F_{t-1} is the information set up to (t – 1). Similar to the above GARCH model Ferland et al. (2006) used mathematical properties to manipulate coefficient time series data with integer values. The proposed integer-valued GARCH, that is, the INGARCH model, is as follows:

$$(\varepsilon_t|F_{t-1})\sim Poisson(\lambda_t)$$

$$\lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \lambda_{t-j}$$

where $\alpha_0 > 0;\ \alpha_i \geq 0;\ i = 1, 2, \cdots, p;\ \beta_j \geq 0;\ j = 1, 2, \cdots, q;\ p \geq 1, q \geq 0.$

In the above model when $q = 0$, the GARCH model becomes the INARCH (p), which assumes a normal distribution as a conditional distribution. The conditional distribution of y_t assumes Poisson distribution. In the INGARCH (p, q) model, the variance depends on past values in the observed time series and on their past values. In this model, mean and variance are equal to the conditional distribution of the Poisson distribution with characteristics, and the over-dispersion (Zhu, 2012). To overcome these shortcomings Zhu (2011) proposed the following negative binomial INGARCH, denoted by INGARCH-NB (p, q).

$$(\varepsilon_t|F_t) \sim NB(r, p_t)$$

$$\lambda_t = \frac{1-p_t}{p_t} = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \lambda_{t-j}$$

where NB represents the negative binomial distribution and r is a positive integer $\alpha_0 > 0;\ \alpha_i \geq 0;\ i = 1, 2, \cdots, p;\ \beta_j \geq 0;\ j = 1, 2, \cdots, q;\ p \geq 1, q \geq 0.$

The INGARCH-NB(p, q) model unlike the INGARCH (p, q) model, the conditional distribution of ε_t assumes negative binomial distribution instead of Poisson distribution. The model uses the INGARCH(p, q) model and the extreme observations, which are common in the coefficient time series data (Zhu, 2011, 2012). In the above model, when q = 0, it becomes INARCH-NB(p) model and a brief description of the model is

D. INGARCH(1, 1) Model

The INGARCH(1, 1) model, the primary model of INGARCH (p, q) as given by-

$$(\varepsilon_t|F_{t-1}) \sim Poisson(\lambda_t)$$

$$\lambda_t = \alpha_0 + \alpha_1 X_{t-1} + \beta_1 \lambda_{t-1}\ ; \text{where } \alpha_1 < 1.$$

E. INGARCH-NB(1, 1) Model

The INGARCH-NB(1, 1) model, which is the first model of INGARCH-NB(p, q), is as follows.

$$(\varepsilon_t|F_t) \sim NB(r, p_t)$$

$$\lambda_t = \frac{1-p_t}{p_t} = \alpha_0 + \alpha_1 \varepsilon_{t-1} + \beta_1 \lambda_{t-1}$$

where $\alpha_0 > 0; \alpha_1 \,\&\, \beta_1 < 1$ and r is a positive integer. The normality condition of this model is $|(r\alpha_1 + \beta_1)^2 + r\alpha_1^2| < 1$ (Zhu, 2011). The conditional mean and conditional variance are derived from the characteristics of the negative binomial distribution as follows

$$E(\varepsilon_t|F_{t-1}) = r\lambda_t$$

$$Var(\varepsilon_t|F_{t-1}) = r\lambda_t(1 + \lambda_t)$$

The above equation shows that the conditional variance has a larger value than the conditional average. Using the property, the problem of over-dispersion, is difficult to explain in the Poisson distribution. However, there is a disadvantage that the parameter *r* must be additionally considered. Specifically, INARCH-NB(1, 0), where q = 0, the model is as follows:

$$(\varepsilon_t|F_{t-1}) \sim NB(r, p_t)$$

$$\lambda_t = \frac{1-p_t}{p_t} = \alpha_0 + \alpha_1 \varepsilon_{t-1}\ ;\ \alpha_1 < 1\ .$$

F. INGARCH model without Covariate

According to Ferland et al. (2006) the integer valued GARCH(p, q) with Poisson distribution is given by

$$(\varepsilon_t|F_{t-1}) \sim Poisson(\lambda_t(\theta))$$

$$\Pr(\varepsilon_t = y_t|F_{t-1}) = \frac{\exp(-\lambda_t)\,{\lambda_t}^{y_t}}{y_t!}\ \ ; y_t = 0, 1, 2, \ldots$$

where $\{\varepsilon_t\}$ be a time series of counts and F_{t-1} to be the history of the process up to time (t-1) with

$$\lambda_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \lambda_{t-1}.$$

According to Zhu (2011) the integer valued GARCH(p,q) with negative binomial distribution is given by

$$\lambda_t = \frac{1-p_t}{p_t} = \lambda_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \lambda_{t-1}.$$

G. INGARCH Model with Covariate(s)

According to Fokianos and Fried (2010) the integer valued GARCH(p,q) with covariates for poisson distribution is given by

$$(\varepsilon_t = y_t|G_{t-1}) \sim Poisson(w_t)$$

$$w_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j=1}^{q} \beta_j w_{t-j} + \sum_{k=1}^{s} \gamma_j X_{t,k}$$

Where $\{y_t\}$ be a time series of counts and G_{t-1} to be the history of the process up to time (t-1) with covariates X_t.

According to Zhu (2011), Christou and Fokianos (2014) and Liboschik (2015) the integer valued GARCH(p, q) with negative binomial distribution is given by

$$(\varepsilon_t = y_t|G_{t-1}) \sim NB(r, \vartheta_t)$$

$$w_t = \frac{1-\vartheta_t}{\vartheta_t} = w_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{q} \beta_j \vartheta_{t-j} + \sum_{k=1}^{s} \gamma_j X_{t,k}$$

3.2.14 *Model Selection and Validation Criteria*

There are many kinds of algorithm used to select best model and the general method is to choose what measure of forecast error is most appropriate. For this research, root mean squared error (RMSE), mean absolute error (MAE), Mean absolute deviation (MAD), Mean square error (MSE) will be used.

The mathematical formula of RMSE and MAE are as follows:

(a) $\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}$

(b) $MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|$

(c) $MAD = \frac{1}{n_F}\sum_{t=1}^{n_F}|y_t - y_F|$

(d) $MSE = \frac{1}{n_F}\sum_{t=1}^{n_F}(y_t - y_F)^2$

(e) $MAPE = \left(\frac{1}{n_F}\sum_{t=1}^{n_F}\frac{y_t - y_F}{y_t}\right) \times 100\%$

Akaike Information Criteria (AIC)

AIC is one of the most important criteria for checking adequacy as well as the lag order of a model (Sakamoto, Ishiguro, & Kitagawa, 1986).

AIC is defined as,

$$AIC = \log(\frac{\sum\hat{\varepsilon}_i^{\,2}}{n}) + \frac{2k}{n}$$

where $\sum\hat{\varepsilon}_i^{\,2}$ is the sum of squared residuals. In principle, one could select a lag structure by increasing the number of lags up to the point where the AIC reaches a minimum value.

Bayesian Information Criteria (BIC)

$$BIC = n \ln(\hat{\sigma}^2) + k \ln(n)$$

where n is the dimensionality of the model and k is the number of parameters to be estimated and $\hat{\sigma}^2$ is error variance.

3.3 Data Source

A. Infectious Diarrheal Disease (*IDD*) *Patients:*

For the research IDD considered as the regressor variable, is the monthly count of hospitalized patients with diarrhea who attended the International Centre for Diarrheal Disease Research, Bangladesh (ICDDR,B) in Dhaka. The ICDDR,B hospital serves a large urban population within the city of Dhaka and adjacent areas of Dhaka city. It provides free treatment for more than 100,000 cases of diarrhea each year. A surveillance system was established at the ICDDR,B in 1979 to systematically sample children and adults with diarrheal illnesses (Stoll et al., 1982).

For each diarrheal patient of the ICDDR,B hospitals in Dhaka the stool was microbiologically examined to identify common enteric pathogens. A patient is enrolled in the study when either bacterial, viral and parasitic agents are identified in the stool specimen. From the surveillance data that had been collected over a 24-year period January 1993 to December 2017. From these data, the monthly infectious diarrheal cases in Dhaka are used for this research.

B. Multivariate ENSO Index (MEI)

The MEI is the first principal component of six variables over the tropical Pacific such as sea surface temperature, sea level pressure, zonal and meridian components of the surface wind, air temperature and total cloudiness fraction of the sky from ICOADS dataset developed by NOAA. MEI values are computed for every month based on the two preceding calendar months.

MEI does a better job than other indices for the overall monitoring of the ENSO phenomenon, including worldwide correlations with surface temperatures and rainfall.

MEI data that are used for this research for the period January, 1950 to December, 2017 and collected from the following website,

https://www.ncdc.noaa.gov/teleconnections/enso/indicators/mei/

C. Indian Ocean Dipole (IOD) Index

IOD index is the difference between the sea surface temperature anomaly of the West (50°E to 70°E and 10°S to 10°N) and Eastern (90°E to 110°E and 10°S to 0°N) of the Indian ocean where the intensity of the IOD is represented and its index is calculated.

This climatic phenomenon was discovered in 1999 by Dr. Toshio Yamagata (Director of climate variations research program, University of Tokyo), Dr. Saji Hameed and other staff and named the IOD (Indian Ocean Dipole).

The index of the IOD is positive when the Indian Ocean sea surface temperature is warmer than the normal to the West and below than the normal in the East. The index of IOD is negative when the Indian Ocean sea surface temperature is colder than normal in the West and warmer than normal in the East. IOD time series are taken for the period January, 1870 to December, 2017 and detail of the data is given in the following website,

https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/dmi.long.data

3.4 Conclusion

Mean model, volatility model and wavelet volatility model have been described in methods and materials chapter sequentially mathematically and philosophically with their respective validation rule that evaluates overall validity and reliability. Finally, INGARCH models have been used in count data application with climatic time series. However the methods of the study have been presented in the study for finding a better model to forecast number of infected diarrheal patients.

CHAPTER FOUR

Environmental Settings of the Study

Chapter Four
Environmental Settings of the Study

4.1 Introduction

Environmental settings of the study are the identification of the structure, dimensions and components (fields) that constitute the contents of a study in relation to different environmental characteristics. Environmental settings address those aspects that are resulted in the factors from the environment. It is actually focused on overall general scenario of real ground and target settings. The present research belongs to a special environmental settings and features. As the study deals with both environment and human health, this chapter could be helpful in making decisions of dimensions.

4.2 Geophysical and Socio Economic Features of Bangladesh

The geo situation of Bangladesh lies between 20.40^0N and 26.38^0N latitude and between 88.01^0 E and 92.41^0E longitude. It is bounded by India to the north, west and east, Myanmar to the southeast and the Bay of Bengal to the south. The total area of Bangladesh is 147,570 km^2 of which 119,624 km^2 (81.0%) is effective land area, 8,236 km^2 (5.6%) is occupied by river networks and the forest area consists of 19,710 km^2 (13.4 %) (BBS, 2011a).

Bangladesh can be divided into three major physiographic regions: hills, verandas and floodplains. Floodplains constitute the lion's share of the country (80%) and have been formed by the deposition of alluvium by three mighty rivers (the Ganges,

Brahmaputra and the Meghna, collectively known as GBM) and their tributaries and distributaries; the northern and southern hilly areas of Tertiary and Quaternary age occupy about 12 % of the country, whilst only 8% of country is made up of terrace areas (Brammer, 2002). The country has a humid subtropical climate with wide seasonal variations in rainfall, generally warm temperatures and high humidity (Rashid, 1991). There are three distinct seasons, being summer (March–May), the monsoon (June–October) and the winter or dry season (November–February). However, from a meteorological point of view, the year has been categorized into four seasons. They are pre-monsoon (March–May), monsoon (June–September), post-monsoon (October–November) and winter, which starts in December and ends in February. The pre-monsoon season is also known as summer and is the hottest part of the year, characterized by high evaporation rates. The mean temperature ranges from 27^0 to 31^0C during this period. Rainfall varies spatially with the frequent occurrence of thunderstorms which provide about 10–15 % of the annual total. Abundant rainfall occurs in the monsoon with the influence of southerly or south-westerly winds. Temperature drops with the onset of the summer monsoon and approximately 75–80 % of total rainfall occurs during the rainy season. In contrast, the winter is cool and dry with little rainfall. The mean winter temperature is 18^0C with the lowest being 3–5^0C in the northwest part of the country, and only 5% of annual rainfall occurs during this season. This is the sunniest period in Bangladesh with high solar radiation. The hydrographic network of the country is very dense and more than 250 rivers crisscross the country (Chowdhury and Salehin, 1987). These rivers act as lifelines in some of the areas and have supported agriculture, navigation, groundwater recharge, fish habitats and land-building activities since time immemorial.

The economy of Bangladesh is mostly agrarian with 48% of the labor force engaged in agriculture and related activities (BBS, 2010a). A provisional estimate shows that the agricultural sector, together with fishing and forestry, contributed 18.43 % to the Gross Domestic Product (GDP) in 2010–2011 at current market prices, whilst the contribution from the manufacturing sector including large, medium and small scale industries was 18.17 % (BBS, 2012a). Currently, the total GDP of the country is 110.6 billion US$ with the GDP per capita being 770 US$ (World Bank, 2012). Based on the most recent Household Income and Expenditure Survey (HIES), the incidence

of poverty is declining with time. For example, the rate of poverty was 40 % in 2005 and decreased to about 31.5 % in 2010 at national level. Similar results can be found for the rural and urban poverty rates (BBS, 2011b). According to the most recent population and housing census (BBS, 2012b), the literacy rate for males and females is 54.1 and 49.4 %, respectively, at the national level.

The ever-growing population in Bangladesh is currently placing enormous pressure on natural resources. Despite the fact that the annual growth rate has been reduced significantly, the size of population is still large as compared to the size of the country. As a result, both landlessness and environmental degradation have become pervasive in recent times (Choudhury, 2008). The available statistics show that the population of Bangladesh was 71.4 million in 1974 with an annual growth rate of 3.5% (BBS, 1984) which increased to 87.1 million in 1981, 111.4 million in 1991 and 123.9 million in 2001 (BBS, 1994 and 2004). According to the 2011 population and housing census, the total population of the country is 149 million with an annual growth rate of 1.4% (BBS, 2012b). This shows that the total population of the country has increased by 72 million between 1974 and 2011. The density of population has also increased from 590/km^2 in 1981 to 976/km^2 in 2011. If the current trend continues, the population of Bangladesh is expected to reach about 194 million in 2050 (UN, 2012). Understandably, this population exerts tremendous pressure on a limited resource base of the country. For instance, the total proportion of land holdings being farmed has been reduced from 72.7 % in 1983 to about 58.6 % in 2008 (BBS, 2010a).

Along with national population growth, the percentage of people living in urban areas has also been rising since the country's independence from Pakistan in 1971 (Islam, 2005a). For example, 1.8 million people were living in urban areas in 1951 which increased to 13.5 million in 1981, 22.5 million in 1991, 31 million in 2001 and 33.5 million in 2011 (BBS, 2008 and 2012a,b). This overwhelming growth in urban population is largely attributed to rural–urban migration, natural growth and the redefinition of urban areas (BBS, 2008). However, it is asserted by many that rural-urban migration is the most dominant factor (Rouf and Jahan, 2007; Islam 2005a). Because of such growth in urban population and associated infrastructure, every year Bangladesh loses 0.3 % of its cultivated land (BBS, 2010a). Furthermore, the country

faces enormous challenges in coping with the infrastructures and service requirements for its rapidly growing urban population (Rashid, 2008). As a result, widespread environmental degradation has become a common feature in the urban areas of the country, particularly in the capital city – Dhaka (Dewan et al., 2012; Siddiqui et al., 2010; Islam, 2005a).

Apart from rapid population growth, the probable effects of climate change have now become a matter of serious concern in Bangladesh in spite of its tiny contribution to the global carbon dioxide emission, at 0.3 metric ton per capita per annum (World Bank, 2012). Bangladesh's vulnerability to climate change is further aggravated by its geographical location, weak economy and low capacity to address the devastating impacts (Shahid, 2010; Mirza, 2009). Studies have demonstrated that the annual average rainfall of Bangladesh is increasing at a rate of 6.58 mm/year (Shahid, 2011a) which has major implication for future flooding in the country. A 23–29 % increase in flooded areas is predicted for Bangladesh if global temperature rises by 2^0C (Mirza et al., 2003). In another study, Shahid (2012) predicted an average increase in temperature of 1.4^0C in 2050 and 2.4^0C in 2100. Rising temperature would seriously affect current agricultural practices (Shahid, 2011b) and could severely disrupt food production in the region (Douglas, 2009). One metre rise in sea level could inundate as much as 16% of Bangladesh (Ahmed, 2009) and might displace nearly 30 million people (BBS, 2009 in Mirza, 2009). Furthermore, global warming may increase the frequency of tropical cyclones and storm surges in the coast. A study by Islam and Peterson (2009) indicated that the rate of tropical storms making landfall has increased by 1.18% per year since 1950 which may be linked with global warming.

4.3 Geophysical and Socio Economic Features of Dhaka City

Dhaka, the capital of the People's Republic of Bangladesh, first attracted attention when it became the provincial capital in 1905 (Rizvi, 1969). After the partition of the Indian subcontinent in 1947 into India and Pakistan, Dhaka became the capital of the then East Pakistan. Bangladesh emerged as an independent state in 1971 by liberation war, and the country retained Dhaka as its capital. Since then, Dhaka has been the major focus of administrative, social, educational and cultural activities. Some demographic characteristics are shown in Table 4.1.

Table 4.1: Demographic characteristics of Dhaka City Corporation (DCC) area

Year	Total Households	Population	Density (km^2)	Sex Ratio	Household size (Average)	Literacy Rate	Growth Rate (%)
1961	61,983	368,575	10,840	145	5.9	----	–
1974	197,361	1,403,259	38,979	140	7.1	----	–
1981	367,513	2,475,710	18,755	141	6.1	----	11.15
1991	574,807	3,612,850	23,475	131	5.6	48.1	5.22
2001	1,109,514	5,327,306	34,629	131	4.7	57	6.55
2011	1,580,672[a]	7,033,075	49,182	131	4.5	65.1	4.08

(Source: BBS, 1998, 2008, 2012b); 'a' includes general institutional households

The climate of the megacity can be classified as tropical monsoon and is characterized by three distinct seasons: monsoon, summer and winter. The average annual rainfall is about 2,000 mm, of which 80% occurs during the monsoon. The temperature during the summer months ranges between 28^0C and 34^0C. In winter, the temperature ranges between 10^0C and 21^0C with little rainfall. There are five major river systems flowing across the megacity, namely, the Buriganga–Dhaleshwari system to the south, Bansi–Dhaleshwari system to the west, Turag–Lubundha to the north and the Lakhya–Balu system to the east and southeast. The megacity and its adjoining areas are composed of alluvial terraces of the southern part of the Pleistocene 'uplands' known as the Madhupur tract and low-lying areas around the junction of the Meghna and Lakhya rivers.

With the increasing dependency on groundwater extraction, the water table is dropping rapidly. The rate of decline has reached as high as 2.5 m per year in the vicinity of the Central Business District (CBD) (Hoque et al., 2007).

Sources of drinking water in the megacity vary according to the locality, although groundwater remains the major source (Ahmed et al., 2005). Fifty-three percent of households have access to supply water, whilst 43.1% rely on tube wells for their potable water sources (BBS, 2008). The remaining households depend on deep tube wells, ponds and other sources such as rivers.

Due to increasing anthropogenic activities, the quality of surface water has been deteriorating in the rivers and other water bodies such as lakes and canals (Dewan et al., 2012). As Dhaka hosts more than 40% of the country's industries (BBS, 2005) and 75% of the export-oriented garment industry (Islam, 2005a), industrial effluents along with domestic wastes are directly discharged into the rivers, and are the primary cause of water contamination (Hossain and Rahman 2012; BBS, 2010b; Rahman and Hossain, 2008; Sohel et al., 2003; Karn and Harada, 2001). Water quality parameters, such as dissolved oxygen (DO) have been found to be completely depleted in the Buriganga river during the dry season whilst biological oxygen demand (BOD) increases at the same time (Dewan et al., 2012; Sohel et al., 2003), indicating the existence of substantial amounts of inorganic and non-biodegradable components in the rivers (BBS, 2010b; Karn and Harada, 2001). Slum dwellers typically build open latrines on the roadside or construct hanging latrines over the water bodies. When natural and storm water run-off transports human waste to the surrounding water bodies, it results in severe water pollution (BBS, 2010b; Hossain, 2006; Hasan and Mulamoottil, 1994).

Other environmental hazards in the megacity include severe local storms such as tornados and nor'westers. Time series analysis of nor'wester occurrences (1954–2000) in Bangladesh revealed that Dhaka Megacity is highly exposed to local damaging storms, and it had the highest incidence of such events (on a weekly or fortnightly basis) resulting in severe property damage (Dewan and Peterson, 2003).

4.4 Health Settings in Bangladesh

The health system of Bangladesh is publicly financed by a top-down approach from national to community level. There is no health insurance or community financing for health care expenditure. In most cases, the poor population pays out of pocket for their health care. This study also assessed availability and accessibility of health system at community level for adaptation. However, the knowledge and perception portion of the study would be a remarkable submission.

Bangladesh's population estimated to be 162.9 million in mid-2016 and will be 186.5 million in mid-2030; and 202.2 million in mid-2050 (Population Reference Bureau,

2016). In 1973, when the country launched its First Five-Year Plan (1973-78), population was estimated to be 74.0 million and the rate of population growth was then 3.0% per annum. In 1975, contraceptive prevalence rate (CPR) was reported to be 8.5% (BFS, 1975) as against the present estimate of 52 per cent, (DHS, 1999) showing an average increase of 1.8% per annum since then. In 1989, total fertility rate (TFR) and CPR were estimated at 4.9% and 32.0% respectively (BFS and CPS, 1989). The corresponding figures in 1999 are 3.3% and 52% respectively. Bangladesh has achieved this progress against the backdrop of low literacy rate (54%), low status of women and low income per capita US $ 350 and so on. Now women (15-49 years) using modern methods of contraception are about 54% and all methods (both modern and traditional) are 62% (BDHS, 2014). According to Bangladesh Sample Vital Statistics 2016, the estimated crude death rate was 5.1 per 1,000 people. The rate was 5.7 in rural areas and 4.2 in urban areas. This rate has shown a decline from 5.3 in 2012 to 5.1 in 2016. The crude birth rate, which is the simplest measure of fertility, has been estimated at 18.7 per 1,000 in 2016 compared to 18.8 in 2015. The child birth rate (CBR) fell from 18.9 in 2012 to 18.7 in 2016. The rural CBR, as expected, is higher than the urban CBR: 20.9 versus 16.1. The general fertility rate worked out to 69 per 1,000 women with a much higher rate (79) in rural areas compared to 57 in urban areas. Male children show a somewhat greater decline in the infant mortality rate (IMR) (20.6%) than their female counterparts (12.5%). The decline in the IMR is more pronounced (17.6%) in rural areas than in the urban ones (9.7%). In conformity to with this decline in the IMR, the neo-natal mortality rate also fell from 21 deaths per 1,000 live births in 2012 to 19 deaths per 1,000 live births in 2016 without revealing any significant sex differentials. The area of residence did not influence the neo-natal mortality rate. The post-neo-natal mortality rate, which nearly remained static over the last three years, was around nine deaths per 1,000 live births. Child mortality was estimated to be 1.8 deaths per 1,000 children in 2016, which is marginally lower than the previous year's rate. The rate, however, fell from 2.3 in 2012 to 1.8 in 2016, registering an almost 22% decline in five years. Life expectancy in Bangladesh has gone up by 2.2 years during the past five years, attaining an average of 71.6 years in 2016 from 69.4 years in 2012. A similar decline was noted in

the infant mortality rate, from 33 per 1,000 live births in 2012 to 28 in 2016 (BSVS, 2016).

4.5 Relation between Global Climate and Health

Global climatic factors do not affect all regions of the world with same intensity. The effects depend upon the geographical setting, whereas the intensity of impacts of the events depends upon both the geographical and the socio-economic setting of a country. To build a world where people have sustainable livelihood, food security, health care services and green environment knowledge of global climate factors and its impacts are important in respective setting areas.

The warming of sea surface temperatures in the tropical Pacific causes changes in Pacific trade winds and ocean currents, setting off a chain reaction of weather disturbances worldwide. Scientists believe that Bangladesh is one of the most vulnerable countries in the world to El Nino or La Nina effects (Choudhury, 1994). The country experienced very adverse El Nino effects in 1997-98, which included severe drought in various places and severe La Nina effects in 1998, which included the most devastating floods. It is thought that the country was also affected by the 1972-73, 1976, 1982, 1986 and 1994 El Nino events. The country's geographical location and climatic characteristics made it vulnerable to such warm events.

El Nino tele-connection(s) influences the global climate in different regions in different ways. In Bangladesh, El Nino is generally associated with drought, whereas La Nina, which corresponds to the positive value of the Southern Oscillation Index (SOI), results in increased rainfall leading mostly to torrential rainfall (Syed Anwar, 1999).

As climate change is an issue of great importance to Bangladesh and other low-income countries. Our study in this area and generate evidence to support national, regional and global policy-making.

Links between global climate factors and spread of infectious diseases, particularly diarrhea is well documented. In recent years govt. has been building our expertise in fields such as water engineering and environmental science. Scientists are working on several studies of ecosystems and environmental drivers of poverty, land loss, migration to cities, salt contamination of drinking water, and related issues. We are

providing social and economic data on livelihoods at household level in vulnerable areas, to support modeling of climate change impacts during the current century. We have extensive experience of health and population research in Bangladesh, so are well placed to shape and inform discussions on responses to climate change. We will provide the Government of Bangladesh with regular updates on climate change in the country. We will also ensure that discussions are relevant to other countries facing similar challenges.

Still considerable challenges remain in the forefronts in the efforts to improve the health status of the population, reduce health inequalities, improve the quality of care and public satisfaction with healthcare, and to increase the efficiency and sustainability of service-delivery agencies. These challenges point to the growing need for appropriate and applied research to enhance the knowledge about factors affecting the governance, provision, organization, financing and use of healthcare and health services as well as at the role of key multispectral players within the healthcare system. Where resources are scare, it is vital that health system be strengthened so that every decision is the best decision. Health systems research can support that decision-making.

4.6 Conclusion

Global climate factors and its impacts is becoming an emerging research arena in recent decades. Geography and climate have made the country vulnerable to different meteorological, hydrological and geological hazards especially to global climate factors. They (factors) also provide information about the social actions and economic measures that societies take to avoid or mitigate different impacts and maintain the capacity to provide the services that are essential for life and human well-being in health. Lack of knowledge from information of global climate factors could go against all our efforts and may slow down our pace to the development.

CHAPTER FIVE
Exploratory Data Analysis

Chapter Five
Exploratory Data Analysis

5.1 Introduction

Exploratory Data Analysis (EDA) is a fundamental early step after data collection and pre-processing, where the data is simply visualized, plotted, manipulated, without any assumptions, in order to help assessing the quality of the data and building models. It extracts the information enfolded in the data and gain the main idea behind the objective of the study. It is considered to be a crucial step in any data science project. EDA is essential for a well-defined and structured data science step and it should be performed before any statistical analysis phase. EDA reveals the behavior of the data with some minimum prior assumptions and thus allows the data to guide the choice of appropriate models (Cox, 2017; Behrens & Yu, 2003; Hoaglin et al., 2011). It is a robust technique, using the graphical inspection of the data series, where one can decide how data modeling and building can be forwarded. It helps to ensure whether a few extraordinary data values either influence or not the result. Data will be explored by some descriptive measures, graphs and plots to realize the evidences of study guided by time series philosophy. This chapter has included some descriptive and exploratory presentations of the study variables.

5.2 Nature of the Data Series

5.2.1 *Time Series Plot*

As a first step of any research study, the visual inspection of actual data is highly recommended. Simple time series plot with the time on x-axis and concentration on y-axis with suitable scale, in which individual measurements are represented as points, could provide a valuable information on the linearity, normality, seasonality, volatility etc. and often also on the long-term trend. A simple visualization can also serve as a good base for the decision, whether it is necessary to exclude outliers or made some

other pre-processing; moreover, advanced analysts can suggest data transformation as well.

Three run sequence plots are shown in time series format. Top of the figure is explored for monthly data index of a global climate series called the Indian Ocean Dipole (IOD) (Fig. 5.1(a)) of 1776 months ranging over the years 1870–2018. The IOD measures changes as the difference between SST anomalies of the western (50^0E–70^0E, 10^0S–10^0N) and eastern (90^0E–110^0E, 10^0S–10^0N) tropical Indian Ocean. The series tend to exhibit repetitive behavior, with regularly repeating cycles that are easily visible.

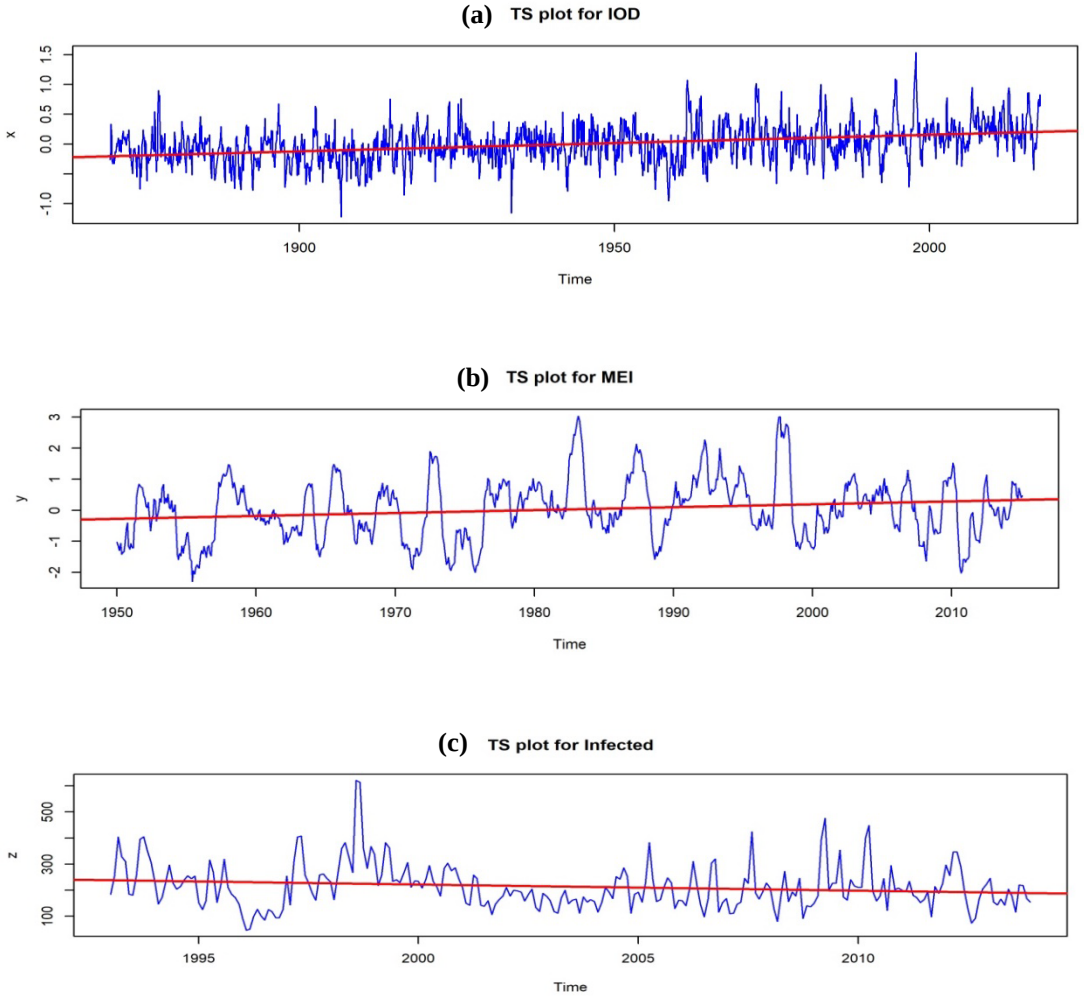


Figure 5.1: Time series plot of IOD (top), MEI (middle) and IDD (bottom) for actual dataset.

This behavior is of interest because underlying processes may be irregular and the rate or frequency of oscillation characterizing the behavior of the underlying series would help to identify modeling. The figure 5.1(b) suggests a non-stationary behavior.

Accordingly, a time series plot of IDD counts over time is presented in Figure 5.1 (c). It is clear from the graph that the IDD cases have somehow a seasonal component with regular occurring almost every year. Besides seasonality, the graph shows a very slight trend of declining over time.

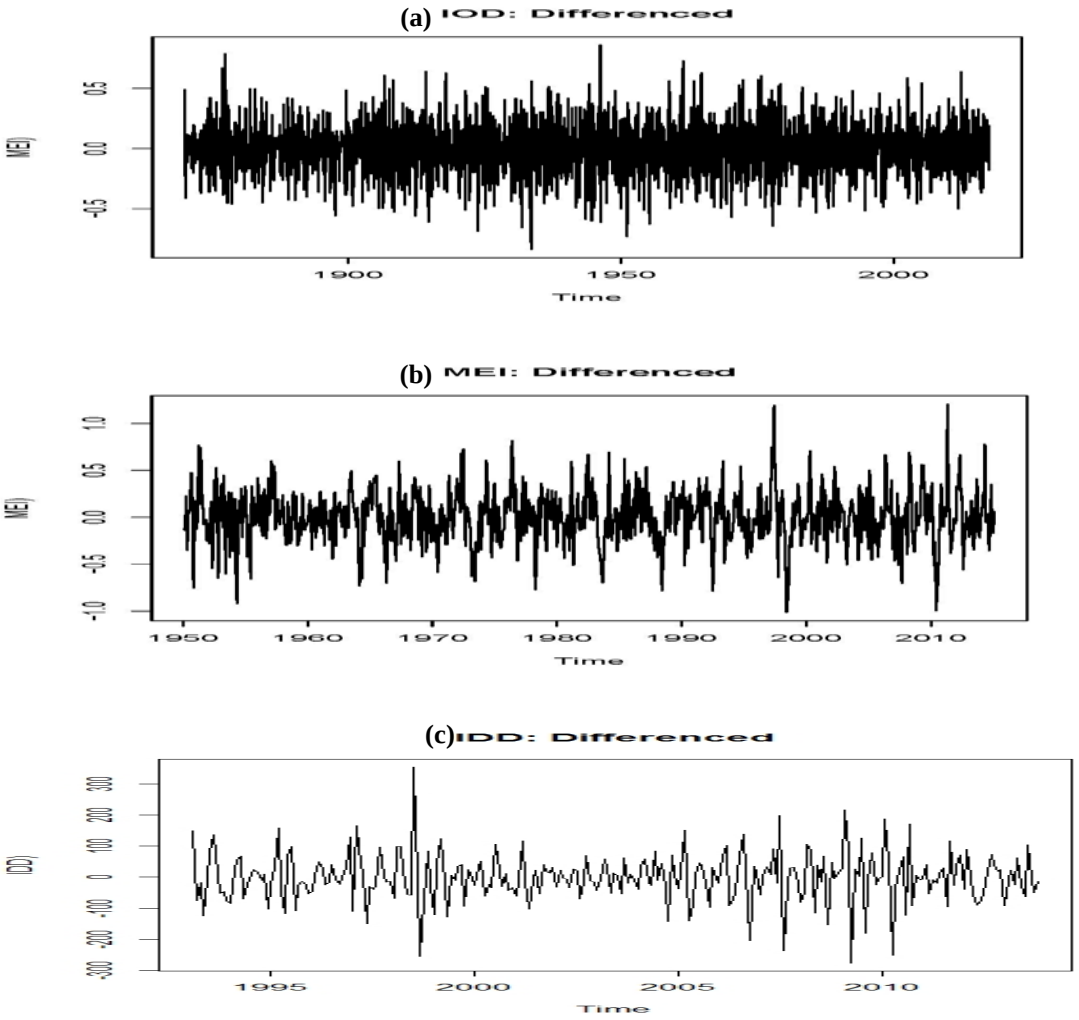


Figure 5.2: Time series plot of IOD (top), MEI (middle) and IDD (bottom) for difference dataset.

Figure 5.2(a), 5.2(b) and 5.2(c) are the exploration of monthly values for difference dataset of IOD, MEI and IDD respectively. Global climatic time series called the Multivariate Enso Index (MEI) for a period of 816 months ranging over the years 1950–1918 is given in middle of the figure. MEI series in Figure 5.2(b) tends to exhibit repetitive behavior, with regularly repeating events that are easily visible. This irregular behavior is of interest because underlying processes may be regular and the rate or frequency of oscillation characterizing the behavior of the underlying series would help to identify them. It can also be noted that the cycles of the MEI are repeating at an extra ordinary order from others even for difference data.

5.2.2 *Box-and-Whisker Plot*

In descriptive statistics, a box-and-whisker plot (also known as a box-plot) is a convenient way of graphically depicting groups of numerical data through their five-number summaries (the smallest observation, lower quartile (Q_1), median, upper quartile (Q_3), and largest observation). A box-plot may also indicate which observations, if any, might be considered outliers. Box-plot can be useful to display differences between populations without making any assumptions of the underlying statistical distribution. The spacing between the different parts of the box-plot indicates the degree of dispersion (spread) and skewness in the data, and identifies outliers. Figure 5.3 gives the box-plot of the data under consideration, and it is observed that all the data series contain outliers. All most every month has a clear differentiation among the descriptive statistics provided. Outliers can have a disproportionate effect on time series models and produce misleading results.

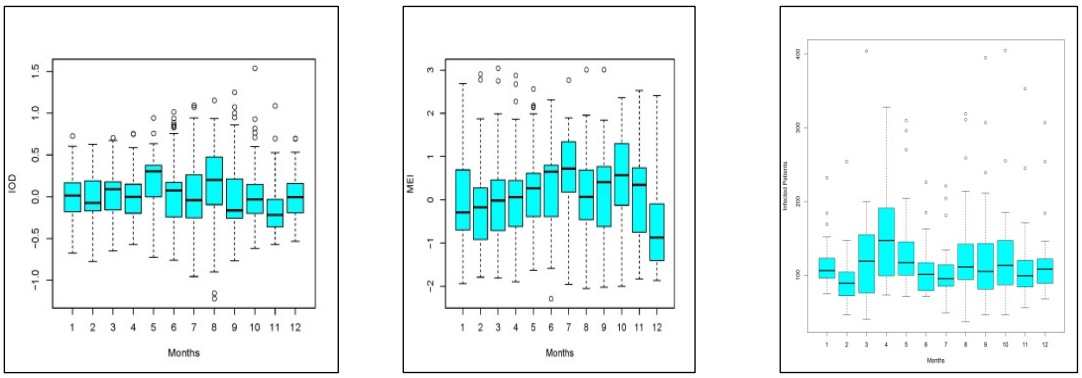


Figure 5.3: Box plot of IDD (right), MEI (middle) and IOD (left) for actual dataset.

5.2.3 Volatility Nature of the Dataset:

Figures 5.4 to 5.6 explore plots of monthly standard deviation. The plots give evidence about the volatility nature of the data. This exploration suggests volatility model. Figure 5.4 gives the monthly variation of the IOD data under consideration. It shows the vacillation of the series in the average and variance and it is observed that all the data series oscillates to every month has a clear differentiation among the descriptive statistics provided.

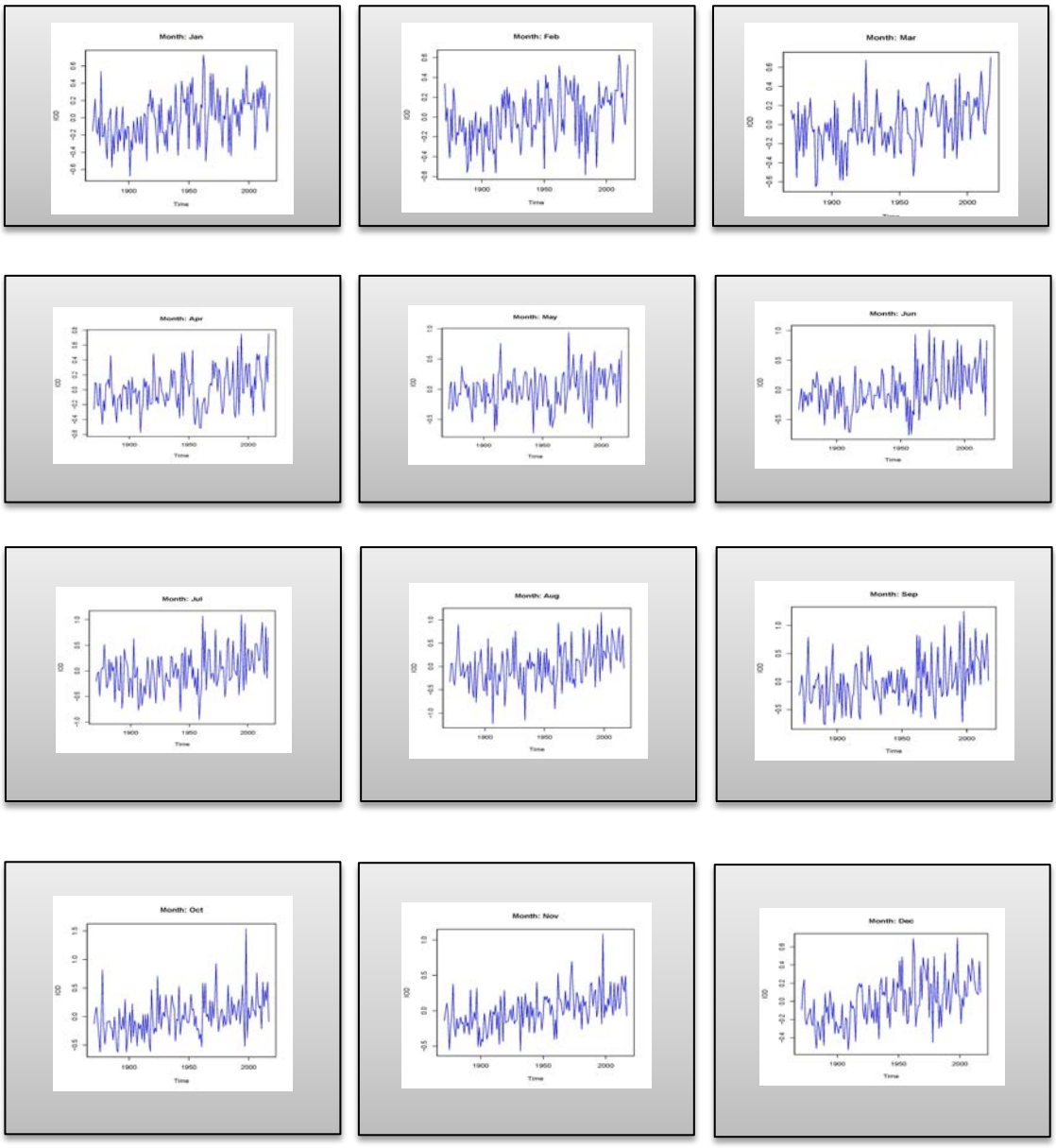


Figure 5.4: Monthly variation plot of IOD for actual.

Figure 5.5 gives the monthly variation of the MEI data under consideration. It also shows the vacillation of the series in the average and standard deviation and it is observed that all the data series mutability to every month has a clear differentiation among the descriptive statistics provided.

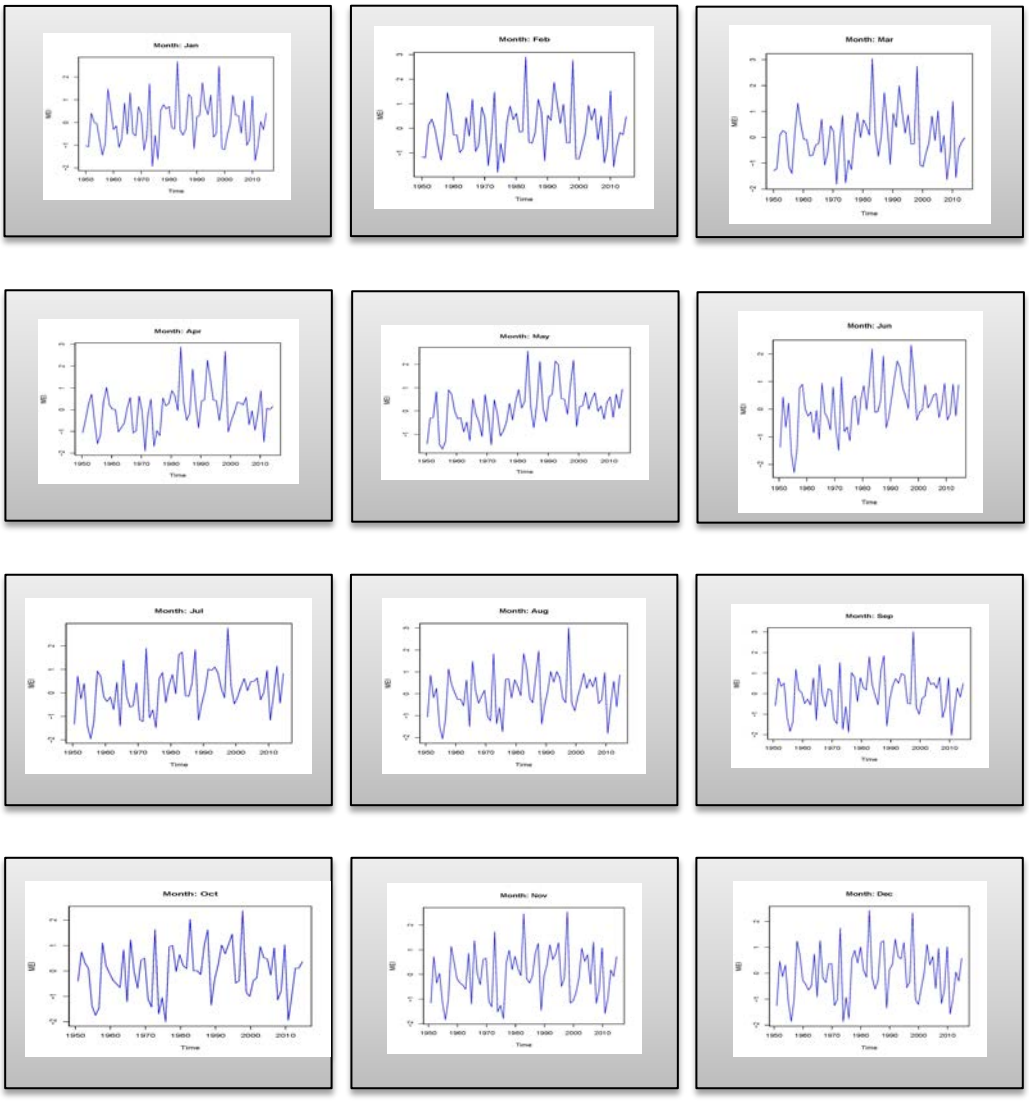
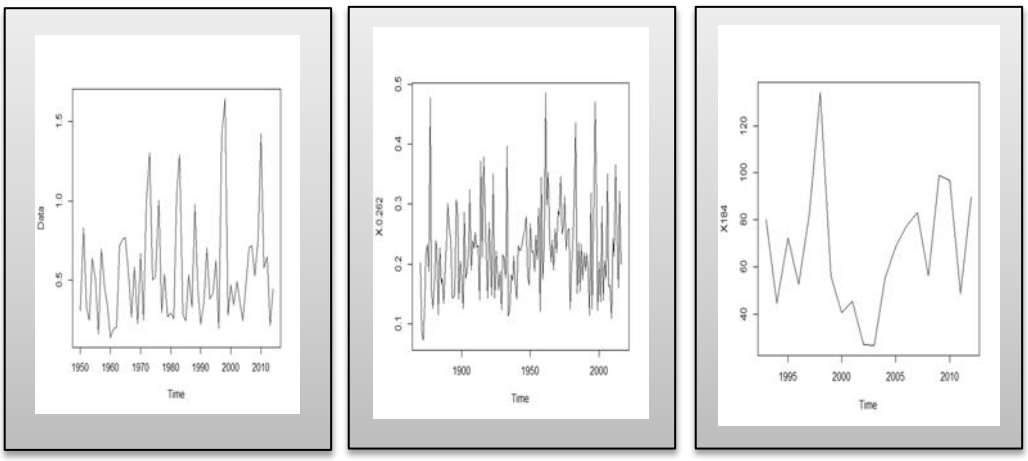


Figure 5.5: Monthly variation plot of MEI for actual.

Figure 5.6 gives the monthly variation of the IOD, MEI and IDD data under consideration. It also shows the fluctuation of the series in the average and standard deviation and it is observed that all the data series have lack of mutability to every month and has clear differentiation among the descriptive statistics provided.



MEI IOD IDD

Figure 5.6: Monthly SD plot of MEI, IOD and IDD for actual

5.2.4 *Correlogram from Autocorrelation*

Seasonal patterns of time series can be examined via correlograms. The correlogram displays graphically and numerically the autocorrelation function (ACF), that is, serial correlation coefficients (and their standard errors) for consecutive lags in a specified range of lags. Ranges of two standard errors for each lag are usually marked in correlograms but typically the size of auto correlation is of more interest than its reliability because we are usually interested only in very strong autocorrelations. While examining correlograms, we should keep in mind that autocorrelations for consecutive lags are formally dependent. If the first element is closely related to the second, and the second to the third, then the first element must also be somewhat related to the third one, etc. This implies that the pattern of serial dependencies can change considerably after removing the first order auto correlation (i.e., after differencing the series with a lag of 1).

5.2.5 *Correlogram from Partial autocorrelation*

Another useful method to examine serial dependencies is to examine PACF, where the dependence on the intermediate elements (those within the lag) is removed. If a lag of 1 is specified then the partial autocorrelation is equivalent to auto correlation.

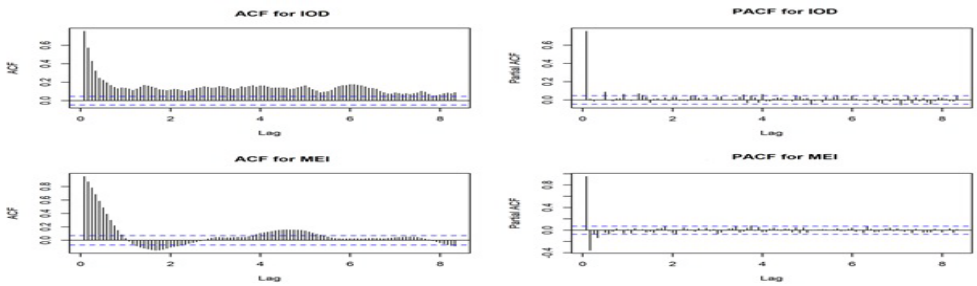


Figure 5.7: ACF and PACF plot for Multivariate ENSO and IOD.

From figure 5.7, we find that the sample ACF values exceed the rough critical values at lags in many points. Decision about ACF and PACF plots suggest that the dataset are correlated.

5.3 Tests for Stationarity

To investigate whether the monthly data are stationary or not, the ADF test has been applied for data. The results are reported in Table 5.1.

As explained the null of the ADF test is that the series is non-stationary. The 5% critical (rejection) region for the test that you used is $-2.928 < t < 2.928$. Thus as the measured t is not in the critical region the hypothesis is accepted and the time series may be taken as non-stationary. This is confirmed by the Mackinnon approximate p-value for z(t): 0.9622 which implies that assuming a unit root there is a probability of 96% of finding an ADF statistic less than 0.047. Then we use DF-GLS variation of the ADF statistic as this is more powerful than the standard ADF statistic which satisfies our expectation.

5.4 Unit Root Tests

These findings were challenged by Perron (1989), who argues that in the presence of a structural break, the standard ADF tests are biased towards the acceptance of the null hypothesis. Perron argues that most macroeconomic series are not characterized by a unit root but rather that persistence arises only from large and infrequent shocks, and that the economy returns to deterministic trend after small and frequent shocks.

Applying the procedure for testing the unit root hypothesis, which allows for the possible presence of the structural break, has at least two advantages. First, it prevents yielding a test result, which is biased towards non-rejection, as suspected by Perron (1989). Second, since this procedure can identify when the possible presence of structural break occurred, then it would provide valuable information for analyzing whether a structural break on a certain variable is associated with a particular

government policy, economic crises, war, regime shifts or other factors.

Firstly we apply DF and ADF test to the data for actual and then for de-noised. Models for these series consist of constant but no time trend. We want test the hypothesis

H_0: The process is non-stationary

H_1: The process is stationary.

Table 5.1: Table of ADF test statistic and critical values of IOD, MEI and IDD

Level	**Factors**	**Actual**		**First Differenced**	
		t-statistic	**p-value**	**t-statistic**	**p-value**
Actual	IOD	-3.418	0.888	-5.352	0.481
	MEI	-1.878	0.154	-5.540	0.054
	IDD	-2.541	0.121	-13.098	0.021
Denoised	IOD	-4.881	0.010	-7.689	0.010
	MEI	-9.949	0.030	-9.041	0.000
	IDD	-15.470	0.000	-14.770	0.000

To go insight of the data, different unit root tests were applied to detect the pattern of stationarity and out of them only ADF test successfully detect the stationarity pattern. Results of ADF test are shown in Table 5.1. Observing p-value for actual data have different significant level and for denoised data all the studied variables are found statistically significantly stationary at desired level of significance.

5.5 Conclusion

Initial investigations on data under study are performed so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of statistics and graphical representations. It is observed from ACF, PACF plots, all the studied climatic and disease variables IOD, MEI and IDD are non-stationary at actual and stationary at denoised case. But it is very interesting to note that the results revealed all studies variables are significantly stationary at first difference for both actual and denoised cases. Same findings are observed by ADF test.

CHAPTER SIX
Model Building: Climatic Indices

Chapter Six
Model Building: Climatic Indices

6.1 Introduction

Modeling and forecasts of global oceanographic index and related climatic variables considering as key factors for giving better health care services of infectious diarrheal disease patients. An effective management on climate anomalies impact depends on the performance towards better accuracy and forecasts. This section introduces W-GARCH algorithm to improve the forecasting performance in global climate index time series data. MEI for the period January, 1950 to December, 2017 and IOD for the period from January, 1870 to December, 2017 of dipole mode index on tropical Indian Ocean are used to compare the GARCH model and proposed W-GARCH(1, 1) model. The data are divided into two groups: *Training data*, from January, 1870 to December, 2008 and January, 2009 to December, 2017 as *Test data*. The goodness of performance has been investigated via the AIC, AIC_c, BIC, SC, HQ, RMSE, QQ plot and McLeod-Li test.

In this chapter, approximation series has been used since the series behave as the main component of the transform, while the detail series provides “small” adjustments. The procedure explained in this chapter is as follows:

- Decompose MEI and IOD through the Haar wavelet transformation.
- Use a specific GARCH model fitted to each one of the approximation series to make the forecasting behaviour of the constructed model successful.

- The results of GARCH and W-GARCH are compared using the standard criteria as discussed in chapter three.

6.2 Model Building: Multivariate ENSO Index (MEI)

MEI, influenced by oceanographic tele-connected factors over different time horizons that range from minutes to years, in the past, present, and future have become a topic of interests to the meteorological experts as well as data analysts for long days. It is a strong climate-dominant mode in the tropical Pacific, which has major effects on the global climate system and ecosystem as well as significant socio-economic consequences around the globe. Operational and strategic decision-making based on ENSO has been taken into account not just the realized effects of climate variability, but also potential effects in the decision making in different areas. Thus, a better accuracy of performance and effective decision is vital via modeling and forecasts of MEI data series. ARIMA has been applied in modeling and forecasting for MEI.

Following the ARIMA tools, best model is selected from a set of models by checking performance accuracy. In chapter four, it is observed from EDA that volatility features in MEI and IOD time series data are present and hence, GARCH is applied. But GARCH fails to capture highly irregular phenomena, especially in climatic oscillation and other highly unanticipated events that can lead to significant change in the respective arena. Contrary to that, wavelet transform, a non-linear component, has been applied to overcome the pitfalls of GARCH. But the results were not enough good with respect to accuracy of performance.

Following ARIMA, GARCH and wavelet to gain a better performance in modeling and forecasting MEI data a hybrid W-GARCH model has been proposed.

In Table 6.1, different GARCH orders have tested for MEI data set from 1 to 3 over its various combinations. GARCH(1, 2) has the smallest (AIC = 0.189, BIC = 0.180 and AICc = 0.152) therefore, this model is selected to be the best fit of all models fitted. The other models have greater AIC and AICc, they were provided only for comparison purposes.

Table 6.1: Comparison of GARCH and W-GARCH models for MEI dataset

GARCH(p, q)	AIC		BIC		AIC_c	
	Actual	De-noised	Actual	De-noised	Actual	De-noised
(1,1)	0.262	**-4.303**	0.337	**-4.268**	0.282	**-4.290**
(1,2)	**0.188**	0.100	**0.180**	0.170	**0.152**	0.122
(1,3)	0.279	0.279	0.395	0.209	0.299	0.209
(2,1)	0.259	0.260	0.359	0.259	0.359	0.259
(2,2)	0.211	0.209	0.391	0.311	0.377	0.277
(2,3)	0.896	0.846	0.291	0.191	0.393	0.323
(3,1)	0.899	0.849	0.291	0.191	0.334	0.234
(3,2)	0.810	0.760	0.330	0.230	0.324	0.324
(3,3)	0.861	0.808	0.413	0.213	0.334	0.333

6.2.1 *Performance checking of final model: MEI*

Table 6.2 gives performance of different model using the criteria AIC, SC and HQ with respect to normal, student-t and generalized error distribution (GED).

Table 6.2: Final Model selection for actual and denoised

Criteria	Distribution					
	Normal		Student's t		GED	
	GARCH (1, 2)	W-GARCH (1, 1)	GARCH (1, 2)	W-GARCH (1, 1)	GARCH (1, 2)	W-GARCH (1, 1)
AIC	0.188	**-4.303**	0.209	-1.426	0.259	---
SC	0.180	-4.268	0.307	-1.385	0.307	---
HQ	0.152	-4.290	0.278	-1.410	0.278	---

From Table 6.2, the minimum value of AIC, SC and HQ is considered to select the best model for this research. All GARCH models for data are included in this test between (0, 1) and (2, 2). Some of the combination measures based on degree of parameters are complex roots that were not considered in the estimation procedure. The results obtained based on real roots are presented in this Table.

The GARCH model for the actual MEI data is found as GARCH(1, 2) with AIC about Gaussian distribution to 0.259 as presented in Table-6.2, while the GARCH model for the transform data is selected as W-GARCH (1, 1) with AIC equal to -4.303 as presented in same table as well. Although the fitted GARCH model for the transform data using Haar wavelet transform is selected as W-GARCH(1, 1) based on AIC about Gaussian distribution equal to -4.303, shows some other criteria of various distributions about the result. Considering all these criteria W-GARCH(1, 1) model is found to be better than the GARCH(1, 2) model.

Moreover, the Haar wavelet transform gives more sufficient result and better than Daubechies wavelet transform in the forecasting that is not cited here. However, in some statistical literature, Daubechies wavelet transform is better than Haar wavelet in the decomposition, but in this research a different result has been observed, the reason behind this related to the data set since two wavelet transformation used different approximation for comparison.

Following the results and discussion it can be noted that global climate anomalies impact depends on the performance towards GARCH and Wavelet are two widely used forecasting models captured better accuracy. Although wavelet and GARCH both are suitable for an irregular low and high frequency signals in global climatic time series but their hybrid effect gives a better results for forecasting. Finally, for this study of MEI data a hybrid forecast model, W-GARCH(1, 1) is the best fitted.

6.2.2 *Residual analysis of MEI for model adequacy and diagnostic*

Analytical results are explored in two ways: graphical method and estimated numerical facts. Both are supportive to each other.

Figure 6.1 presents residuals for ACF and PACF of MEI for model diagnostics and it is evident that W-GARCH is superior to ARIMA and GARCH as in run sequence plot residuals are decreasing sequentially over the fitted models respectively. Accordingly in ACF and PACF, residuals are within the limit other than two.

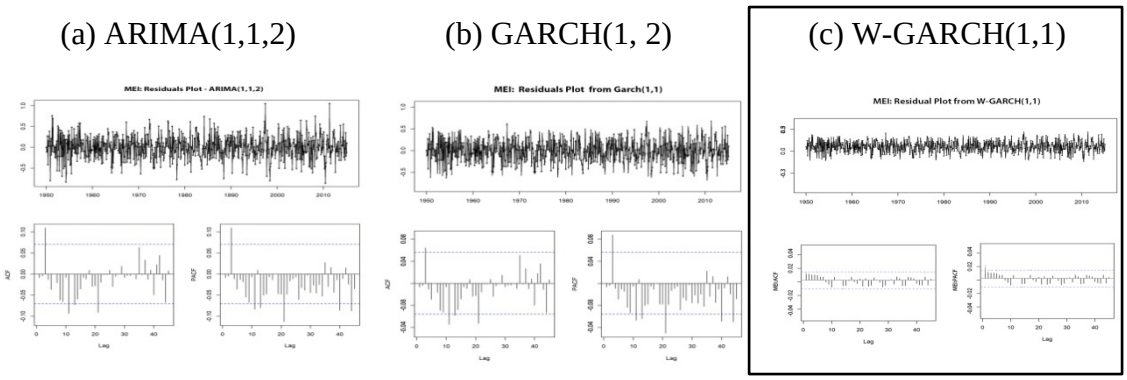


Figure 6.1: Residuals, ACF and PACF of MEI for model diagnostics

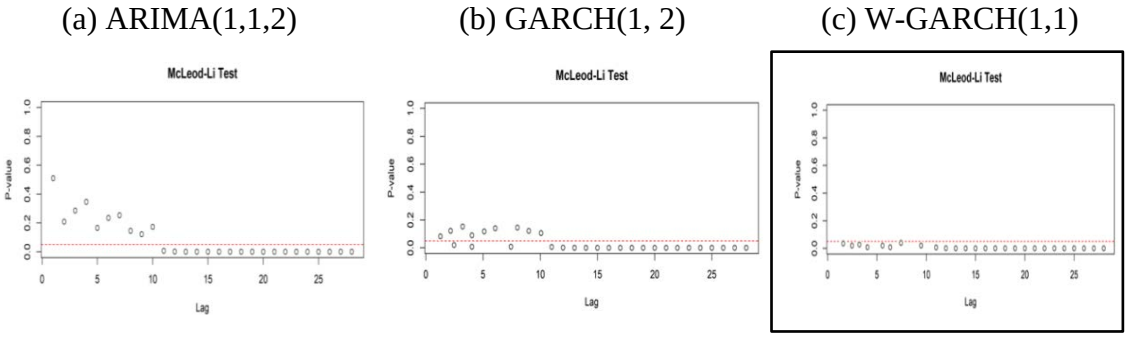


Figure 6.2: McLeod-Li test for model stability

Figure 6.2 describes McLeod-Li test for model stability and W-GARCH(1, 1) model is found to be stable in both training and test data.

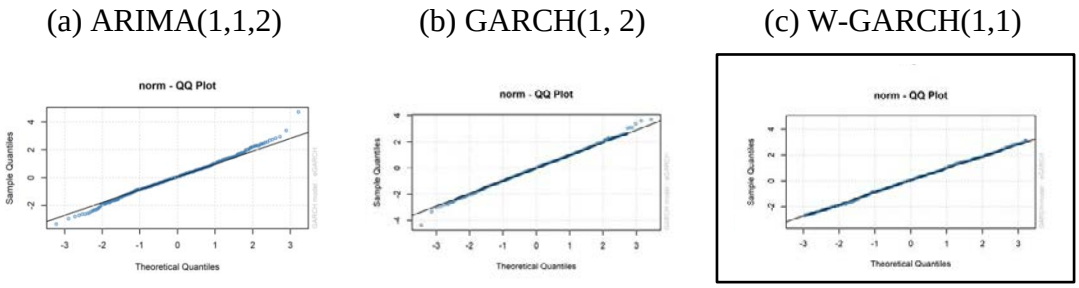


Figure 6.3: QQ Plots for normality test

In Figure 6.3 the standardized residuals from the three models are presented. The QQ plot for the W-GARCH is the only plot showing residuals that can be considered to follow a normal distribution, while for the ARIMA and GARCH this assumption does not hold. In concluding remarks, if the Wavelet transform is used for forecasts in MEI data series, then the result of the W-GARCH (1, 1) model attained better forecast and forecasting accuracy is more stable than the actual MEI data.

6.3 Model Building: Indian Ocean Dipole Index (IOD)

The IOD is a coupled ocean-atmosphere phenomenon in the tropical Indian Ocean, which is characterized by anomalous cooling of SST in the south eastern equatorial Indian Ocean and anomalous warming of SST in the western equatorial Indian Ocean. Although, this type of climatic time series analysis is critical to determine realized effects via modeling and forecasts strategy but it could help understand and forecast the system's future behavior on variability. Subsequently results of the study can be beneficial to realized field.

The AR, MA and ARIMA are some widely recognized statistical forecast models which predict future observations of a time series on the basis of some linear function of past values and white noise terms. As such, these models impose the inherent constraint of linearity on the data generating function with constant variance.

In this sense as a key indicator of climate variability measure in longer time domain, this section employs time series GARCH model, with wavelet transformation techniques to develop a more efficient model with applications to IOD data like MEI modeling. After the data, IOD, being modeled through different modeling tools, goodness of performance is checked. Results of Table 6.3 showed to select best model about GARCH with respect to actual and de-noised.

Table 6.3: Model selection of GARCH for IOD dataset

Order	Selection Criteria					
	AIC		BIC		AIC_c	
	Actual	De-noised	Actual	De-noised	Actual	De-noised
(1, 1)	0.570	**-12.873**	-0.270	**-12.877**	0.288	**-12.867**
(1, 2)	-0.266	-0.366	-0.288	-0.243	0.570	0.501
(1, 3)	-0.292	-0.262	-0.170	-0.170	0.570	0.570
(2, 1)	-0.270	-0.270	-0.288	-0.288	0.570	0.570
(2, 2)	**-1.203**	-1.003	**-1.195**	-1.095	**0.274**	0.054
(2, 3)	-0.273	-0.273	-0.291	-0.291	0.570	0.570
(3, 1)	0.570	0.570	-0.520	-0.520	0.266	0.206
(3, 2)	-0.269	-0.209	-0.265	-0.265	0.308	0.271
(3, 3)	-0.291	-0.291	-0.287	-0.290	0.282	0.232

In Table 6.3, different GARCH order have tested to model IOD data set from 1 to 3 over its various combination. GARCH (2, 2) has the smallest (AIC = -1.203, BIC = -1.295 and AICc = 0.274) therefore, this model is selected to be the best fit among all fitted models. The other models have greater AIC and AICc, they were provided only for comparison purposes.

6.3.1 *Performance checking of final model: IOD*

Table 6.4 gives different model performance criteria using the criteria AIC, AICc, and BIC with respect to normal, student-t and generalized error distribution.

Table 6.4: Final model selection for IOD

Criteria	GARCH(1,1)	GARCH(1,2)	GARCH(2,2)	W-GARCH(1,1)
Normal Distribution				
AIC	-0.290	-0.288	-1.203	**-12.873**
BIC	-0.273	-0.270	-1.195	-12.855
AICc	-0.283	-0.282	0.274	-12.867
Student's t Distribution				
AIC	-0.292	-0.292	-0.291	**-28.107**
BIC	-0.274	-0.270	-0.266	-28.902
AICc	-0.287	-0.284	-0.282	-25.043
Generalized Error Distribution				
AIC	-0.291	-0.291	-0.290	-20.750
BIC	-0.273	-0.260	-0.265	-20.728
AICc	-0.284	-0.283	-0.281	-20.742

In this section, the minimum value of AIC, BIC and AICc is considered to select the best model selection criteria. All choices of GARCH models for data are included in the test between (0, 1) and (2, 2). Some of the combination measures based on degree of parameters are complex roots that should not be satisfied the estimation procedure.

The GARCH model for the actual IOD data series is measured from the combination between (0, 1) and (2, 2) with AIC, BIC and AICc with respect to Gaussian, student-t and generalized error distribution as presented in Table 6.3. While the GARCH model for the transform data by using wavelet transform is selected as W-GARCH (1, 1)

with AIC equal to -28.107 for Student-t distribution accordingly. But the fit of GARCH model is selected as W-GARCH (1, 1) based on AIC about Gaussian distribution equal to -12.855. As student-t distribution is a sampling distribution derived from Normal distribution and it's accommodation is a trial and error composition, the fitted model W-GARCH(1, 1) following Gaussian distribution could be preferred.

Furthermore, the standard errors of model parameters which measure the variation between index data after wavelet transforms are also small. All of these criteria explain that GARCH with the Haar wavelet transform gives more sufficient result better than GARCH in the forecasts.

Although wavelet and GARCH both are suitable for an irregular low and high frequency signals in climatic time series but their hybrid effect gives a better results of modeling. As such, in this section, we have fitted a hybrid forecast method that applies GARCH and Wavelet.

Following the results and discussion it can be noted that global climate anomalies impact depends on the performance towards GARCH and Wavelet are two widely used forecasting models captured better accuracy. Although wavelet and GARCH both are suitable for an irregular low and high frequency signals in global climatic time series but their hybrid effect gives a better results of forecasts. As such, the study for IOD data a hybrid forecast model W-GARCH(1, 1) is successfully fitted.

6.3.2 *Residual analysis for model adequacy and diagnostic: IOD*

Analytical results are explored in two ways: graphical method and estimated numerical facts. Both are supportive to each other.

ARIMA(2, 1, 2) GARCH(2, 2) W-GARCH(1, 1)

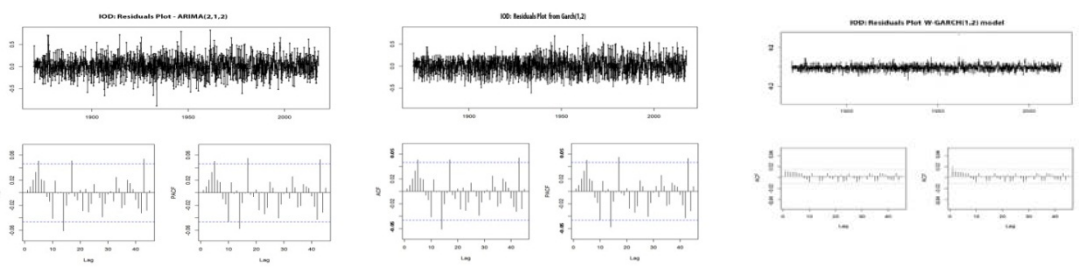


Figure 6.4: Residuals Plot for IOD

From Figure 6.4, it is evident that W-GARCH is superior to that of ARIMA and GARCH as in run sequence plot residuals are decreasing sequentially over the fitted models respectively. Accordingly in ACF and PACF, residuals are within the limit other than two.

ARIMA(2, 1, 2) GARCH(2, 2) W-GARCH(1, 1)

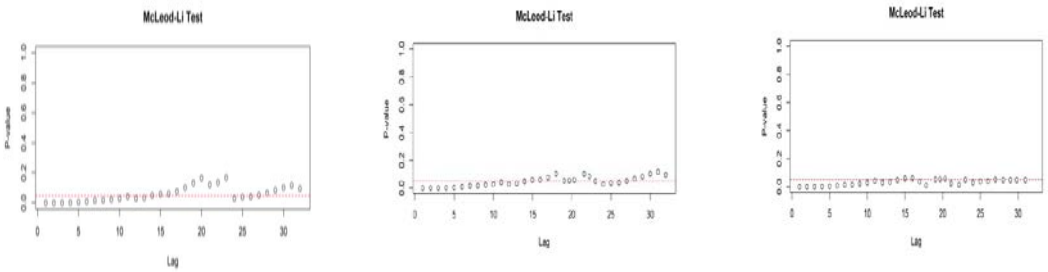


Figure 6.5: McLeod-Li test for model stability of IOD

In figure 6.5, according to McLeod-Li test W-GARCH(1, 1) model is found to be more stable.

ARIMA(2, 1, 2) GARCH(2, 2) W-GARCH(1, 1)

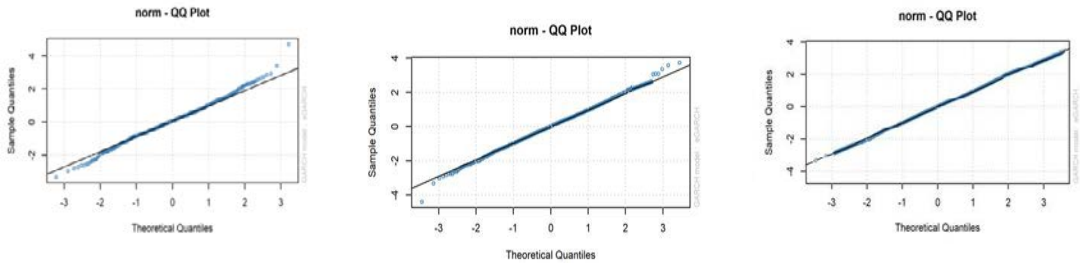


Figure 6.6: QQ Plots for normality test for IOD

In Figure 6.6 the standardized residuals from the three models are presented. The QQ plot for the W-GARCH(1, 1) is the only plot showing residuals that can be considered to follow a straight line, while for the ARIMA and GARCH this assumption is not achieved.

6.4 Conclusion

Modeling of MEI and IOD data series is a complex process for its oscillation characteristics. Three different methods- mean model, volatility model and smoothed volatility techniques for modeling and prediction future values for the monthly global climate data index are applied. First we made fitting of mean model (ARIMA), model

identification, model selection, parameter estimation and forecasting with the selected model maintaining the proper validation rule and then we made fitting of GARCH model, model identification, testing for ARCH effects, model selection, parameter estimation, diagnostic checking of the model and forecasting with the GARCH on different orders of p, q. After then, in final consideration we made fitting of W-GARCH(1, 1) model.

At last we made a comparison between the results of the best GARCH and the best W-GARCH models in order to determine which is better to use in similar situation. Excellent modeling performances were confirmed by the W-GARCH(1, 1) model with Gaussian as error distribution for MEI and W-GARCH(1, 1) model with student-t as error distribution for IOD.

CHAPTER SEVEN

Model Building with Covariate(s)

Chapter Seven
Model Building with Covariates

7.1 Introduction

The fight against infectious diseases does not only require treating patients and setting up measures for prevention but also demands the timely recognition in order to take necessary actions. Effective modeling and forecasts for the type of data is essential inputs for policy-making, planning in health care.

Number of hospitalized patients for infectious diarrhea is count time series data. In literature, most of the modeling and forecasting number of hospitalized patients was used time series regression model with Gaussian distribution. But monthly hospitalized patients for infectious diarrheal data are not continuous. In particular, this type of data is greatly influenced by external factors, which can be represented by including covariate(s) into a model. But studies about the effects of covariates and volatility nature for this type of data have largely fallen behind due to the fact that the underlying process, whose behavior determines the dynamics of the observed process, is not observed. That’s why this study is concerned with a successful model for count time series data IDD with external factors climatic indices MEI and IOD as covariates.

In this chapter, an integer-valued GARCH(1, 1) model is considered which is a popular count time series model, and combine the model with wavelet transformation followed by the Poisson or negative binomial process with GARCH structured parameter.

7.2 Model Building: Infectious Diarrheal Disease (IDD)

Model building approach for this chapter consists of monthly data on number of IDD patients in ICDDR'B from January 1993 to December 2017. An appropriate model will be selected to fit IDD and forecast values would be compared. Before starting the model building process, the condition that should be satisfied is that the series must be generated by a stationary process. In chapter four, stationary condition is checked by using DF, ADF and KPSS test and found that the series is non-stationary.

Following the model building steps, sequentially models- ARIMA, ARCH and GARCH were fitted initially over different orders and distribution for both climatic indices MEI and IOD in Chapter 6. Also, decomposed the time series by wavelet transformation into time-frequency space, W-GARCH(1, 1) model is found to be more appropriate. After that, as the traditional model is misleading and inappropriate for integer valued time series data, initially three models - ARMA, INARCH, INGARCH and INGARCH-NB are fitted for different orders and the models are experimented with and without covariates.

Twelve types of models were studied for modeling monthly number of IDD as target variable and MEI & IOD are considered as covariates along with Poisson and Negative Binomial distribution.

7.2.1 *Model Identification*

A time series exhibiting conditional heteroscedasticity were tested by the Engle (1982) ARCH test which is a Lagrange multiplier (ARCH-LM) test as shown in Table 7.1. The ARCH-LM test shows that there exist ARCH effects in all three factors since the p-values of those main keywords are smaller than 5% level of significance. The Ljung-Box test shows that serial dependence in the factors since the p-values of those main keywords are also smaller than 5% level of significance. Therefore, identified the validity is to consider the covariate IOD and MEI for forecasting target variable IDD. In Figure-7.1, target variable shows a very short decreasing trend over the years. It means that the time plots of IDD are non-stationary even if the ACF plots appears several lags autoregressive model, but the PACF plots display the moving average model with order 2 or 3.

Also from Figure 7.1, it is evident that IDD are not stationary and hence ARIMA, ARCH cannot be applied in general. For this reason, an attempt has be made to deal with INGARCH model with covariates for further analysis.

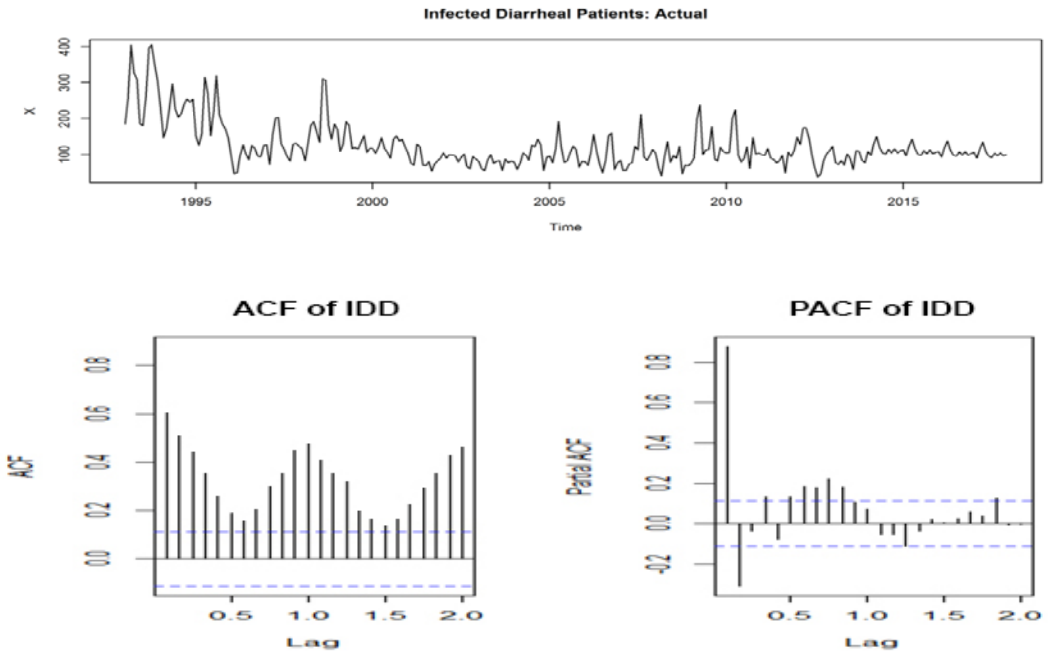


Figure7.1: Time series plot for IDD

Table 7.1: ARCH-LM test for actual and de-noised IDD, MEI and IOD

Test statistic	Actual			De-noised		
	IDD	IOD	MEI	IDD	IOD	MEI
χ^2	20.42	8.417	15.387	17.130	8.210	10.416
p-value	0.005	0.052	0.021	0.000	0.068	0.080
Decision	Present	Present	Present	Present	Present	Present

Table 7.2: Ljung-Box test for actual and de-noised IDD, MEI and IOD

Test statistic	Actual			De-noised		
	IDD	IOD	MEI	IDD	IOD	MEI
χ^2	53.350	8.115	11.972	62.080	8.292	12.742
p-value	0.050	0.706	0.050	0.000	0.002	0.008
Decision	(+)ve	(+)ve	(+)ve	(+)ve	(+)ve	(+)ve

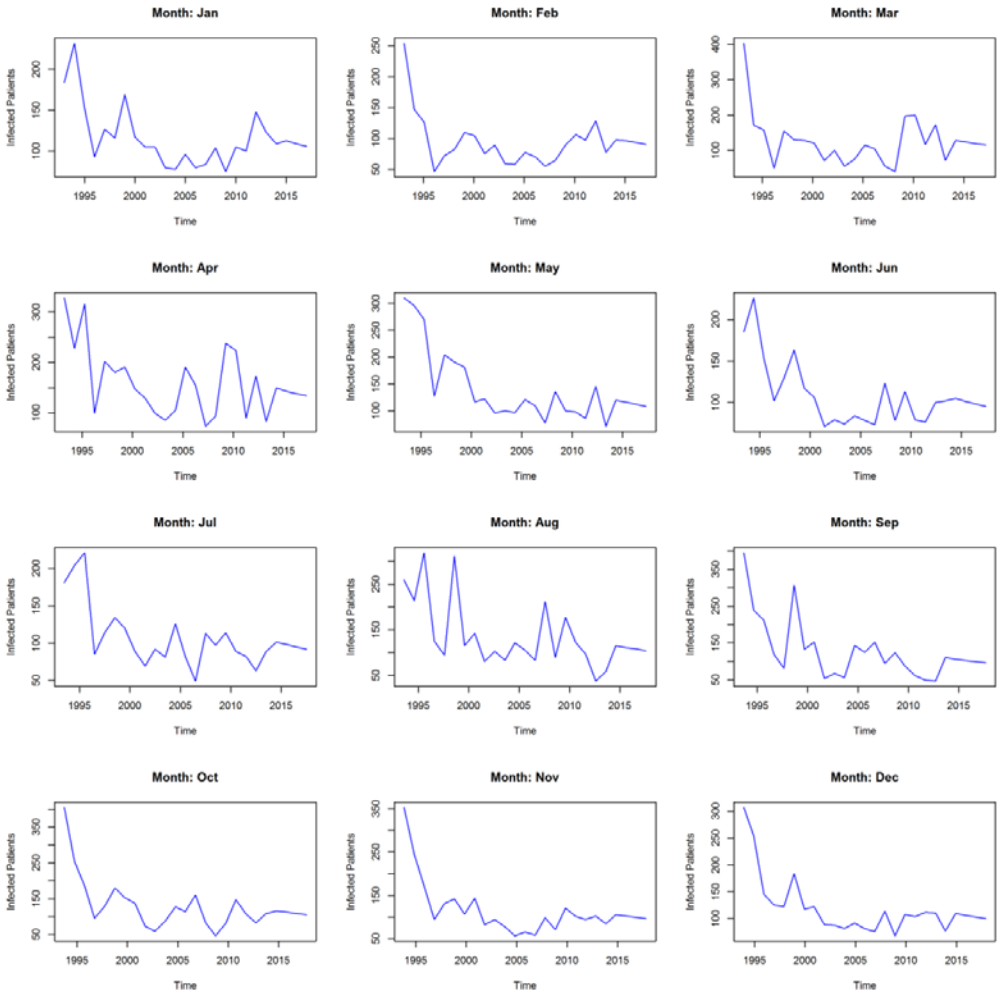


Figure7.2: Monthly variation of IDD for the period 1993 to 2017

Figure 7.2 gives the monthly variation of the IDD data under consideration. It shows monthly variation exists in the data, implies mean is not constant over time. It is also observed that all the data series oscillates to every month has a clear differentiation among the descriptive statistics provided. Also Figure 7.1 shows the vacillation of the series about ACF and PACF as well. INARCH/ INGARCH structure for fitting data series could be better choice for prediction.

7.2.2 *Model Performance*

Based on the results of Table 7.1 and Figures 7.1 - 7.2, the main count time series models are INGARCH(1, 1) and INGARCH(1, 1) for wavelet with Poisson and negative binomial distributions. So these models were applied to twelve models of IDD and covariate(s) with IOD, MEI and {IOD, MEI}.

Table 7.3 Results of different Univariate and Multivariate INGARCH model with covariates for actual and Wavelet translated data

Model	With covariates		Without covariates	
	AIC	BIC	AIC	BIC
Model: IDD on IOD				
INGARCH(1, 1) (Actual)	0.825	0.936	1.738	1.474
INGARCH(1, 1) (Wavelet)	-4.884	-4.332	-3.233	-3.534
INGARCH-NB(1, 1) (Actual)	-1.630	-1.800	1.436	2.172
INGARCH-NB(1, 1) (Wavelet)	-7.721	-7.457	-6.754	-5.056
Model(s): IDD on MEI				
INGARCH(1, 1) (Actual)	-0.252	-2.390	1.738	1.474
INGARCH(1, 1) (Wavelet)	-7.832	-0.849	-3.233	-3.534
INGARCH-NB(1, 1) (Actual)	-0.016	-18.840	1.436	2.172
INGARCH-NB(1, 1) (Wavelet)	-12.597	-1.242	-6.754	-6.754
Model(s): IDD on {MEI, IOD}				
INGARCH(1, 1) (Actual)	-0.252	-2.390	1.738	1.474
INGARCH(1, 1) (Wavelet)	-5.832	-0.641	-3.233	-3.534
INGARCH-NB(1, 1) (Actual)	0.015	58.840	1.436	2.172
INGARCH-NB(1, 1) (Wavelet)	-15.597	-1.542	-6.754	-6.754

According to chapter 4, IDD data are found to be non-stationary and should be analyzed by count time series models. Table 7.1 shows the model comparison of IDD with and without covariates. From Table 7.3, the fact that INGARCH-NB (1, 1) for wavelet with covariate MEI fitted IDD data well than other eleven count time series models including INGARCH-NB (1,1) with covariates because the values of AIC and BIC are smallest among the experimented twelve models. In fact, we did apply the cases in Table 7.3 to the model by Fokianos and Fried (2012), but the values of AIC and BIC by Poisson autoregression model are much bigger than the ones by the above

suggested eleven different count time series models. So better prediction could be executed by the selected model. This means that the model with covariate is superior to the models without covariate. The same goes for the remaining cases.

Table7.4: Estimates of the Parameters of all experimental models

Model	**With covariates**					**Without Covariates**		
	α_0	α_1	β_1	γ_1	γ_2	α_0	α_1	β_1
Model: =IDD on IOD								
INGARCH-Pois(1,1) (Actual)	0.60	0.03	0.03	1.93	n/a	5.05	0.99	0.00
INGARCH-Pois(1,1) (Wavelet)	0.60	0.00	0.03	1.93	n/a	8.21	0.97	0.99
INGARCH-NB(1,1) (Actual)	0.56	0.03	0.97	1.93	n/a	5.05	0.99	0.00
INGARCH-NB(1,1) (Wavelet)	0.56	0.09	1.62	1.93	n/a	4.21	1.97	0.04
Model: IDD on MEI								
INGARCH-Pois(1,1) (Actual)	2.40	0.28	0.01	1.96	n/a	5.05	0.99	0.00
INGARCH-Pois(1,1) (Wavelet)	2.40	0.28	0.01	1.96	n/a	8.21	0.97	0.99
INGARCH-NB(1,1) (Actual)	2.40	0.28	0.11	1.96	n/a	5.05	0.99	0.00
INGARCH-NB(1,1) (Wavelet)	2.40	0.28	0.05	1.96	n/a	4.21	1.97	0.04
Model: IDD on {IOD, MEI}								
INGARCH-Pois(1,1) (Actual)	1.26	20.48	1.47	0.71*	1.68	5.05	0.99	0.00
INGARCH-Pois(1,1) (Wavelet)	1.21	12.25	1.33	5.05*	1.97	8.21	0.97	0.99
INGARCH-NB(1,1) (Actual)	1.26	20.48	6.03	8.21*	1.24	5.05	0.99	0.00
INGARCH-NB(1,1) (Wavelet)	1.21	12.25	5.40	8.71*	0.99	4.21	1.97	0.04

From Table 7.4, it can be concluded that out of the twelve models, some estimates are significant and some are insignificant but for the model INGARCH-NB (1,1) for

wavelet with MEI's expresses significant estimates. Interestingly for the models with covariates {IOD, MEI}, the estimates with IOD are insignificant in four models.

7.2.3 Squared Residual Analysis

In time series modeling, the selection of the best model fit to the data is directly related to whether residual analysis is performed well. The plots of the standardized residuals, the sample ACF of the residuals, normal Q-Q plot, PIT histogram are examined. The horizontal dashed line at 5% helps judge the size of the p-values. Whether all the spikes are within the significance limits or some of them are outsides. The Q-Q plot of standard residuals which explore the distributional shape, suggests that the distribution may have a tail thicker than that of a normal distribution and may be somewhat skewed to the right and show some non-normality on the tails of the distribution and a nearly straight line suggesting that the residuals follow approximately normal distribution, and the model seems to fit quite well. Therefore, to proceed by using the model to forecast future values of the given data series would be interesting.

INGARCH(1,1) INGARCH-NB(1,1)

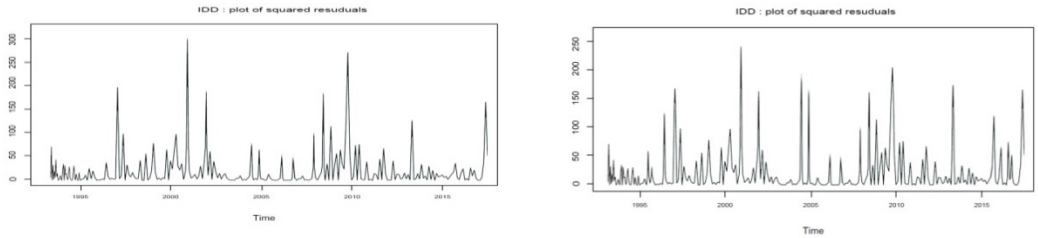


Figure 7.3: Squared residuals plot of INGARCH(1, 1) and INGARCH-NB(1, 1) for IDD without covariate

Figure 7.3 shows that the residual plots of the INGARCH(1, 1) and INGARCH-NB(1,1) models without covariate INGARCH-NB(1,1) has the smaller residuals than other. Figure 7.4 represents the residual plots for models under IDD and MEI as covariate. In the results from Figure 7.4, it is observed that the residuals of the model INGARCH-NB(1,1) for wavelet has gained the smallest residuals. Accordingly, in the results from Figure 7.5, it is revealed that the residuals of the multivariate model INGARCH-NB(1,1) for actual data gained the smallest residuals. Therefore, in the results from Figures 7.3 to 7.6 and Table 7.5, it is observed that the residuals with covariate (MEI) of the model INGARCH-NB(1,1) for wavelet has gained smallest

than any models whether it is multivariate or univariate. As compared to multivariate models for forecasting, univariate model under MEI is doing better. There are several reasons behind that. IOD is a descriptive measure index between two poles and considers only one factor whereas MEI is the first PC of six factors. In addition to that modeling of IOD did not follow normal distribution but student's–t distribution that is a trial and error composition.

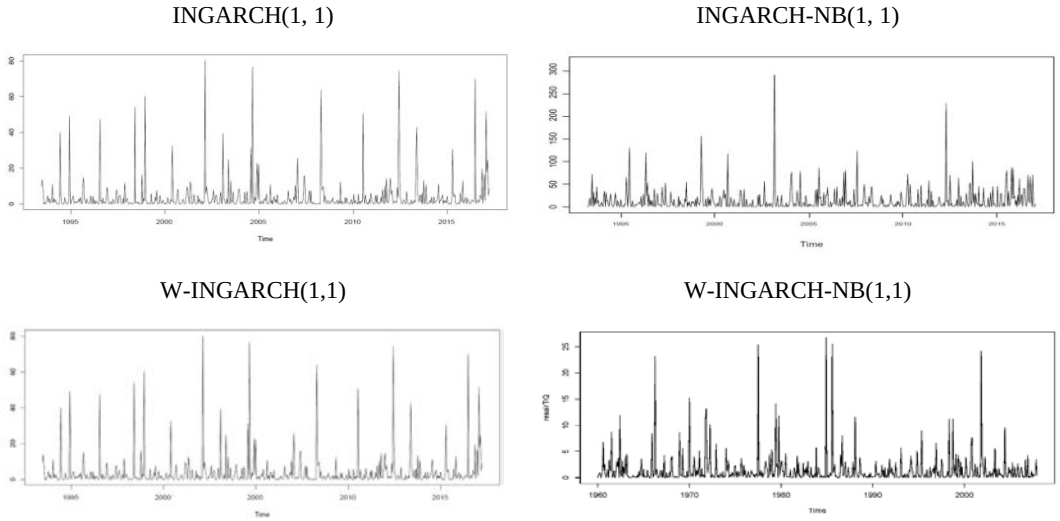


Figure 7.4: Squared residuals plot of INGARCH (1, 1) and INGARCH-NB(1, 1) for actual and de-noised IDD with covariate MEI

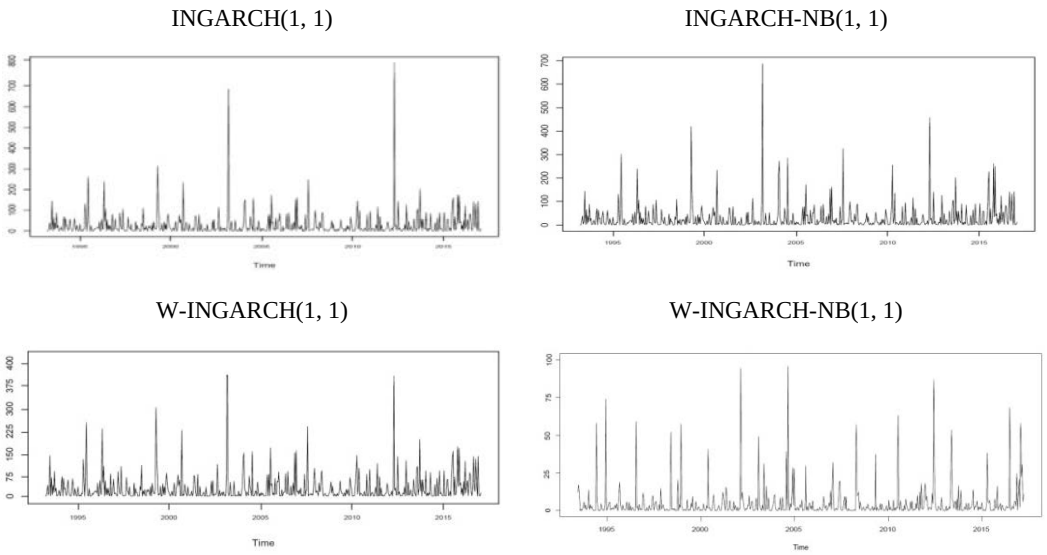


Figure 7.5: Squared residuals plot of INGARCH(1, 1) and INGARCH-NB(1, 1) for actual and de-noised IDD with covariate IOD

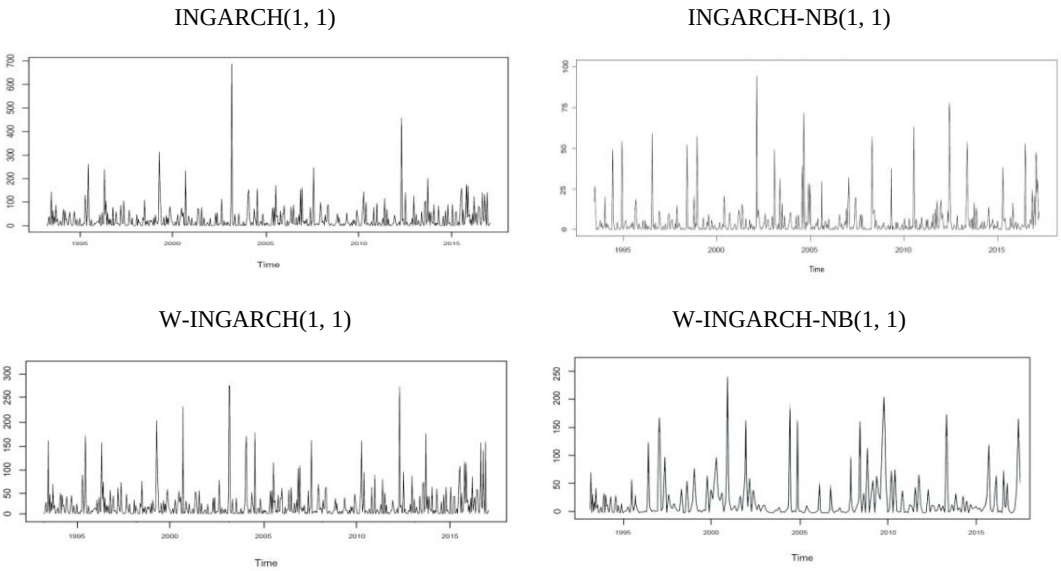


Figure 7.6: Squared residuals plot of INGARCH(1, 1) and INGARCH-NB(1, 1) for actual and de-noised IDD with covariates {IOD, MEI}

Table7.5: Comparison of Univariate and Multivariate models using maximum lebel of squared residuals

Model	Covariates			Without covariates
	MEI	IOD	{IOD,MEI}	
INGARCH(1, 1)	700	800	700	300
W-INGARCH(1, 1)	80	400	300	300
INGARCH-NB(1, 1)	300	700	100	250
W-INGARCH-NB(1, 1)	**25**	100	250	250

7.2.4 *Residual Analysis: PIT, ACF and PACF*

Figures 7.7 to 7.10 represent the time-series behavior of the fitted values, PIT histogram and the correlograms. The distribution of data is asymmetric given the fact that the right has a lot of zeros and the left have outliers. In Figure 7.7, 7.8 and 7.9 the u-shape indicates under dispersion of the predictive distribution. It can be noted in the ACF the autocorrelation, in the negative binomial and Poisson feature points outside the confidence interval. But INGARCH-NB(1, 1) for wavelet with covariate MEI is within the confidence limit.

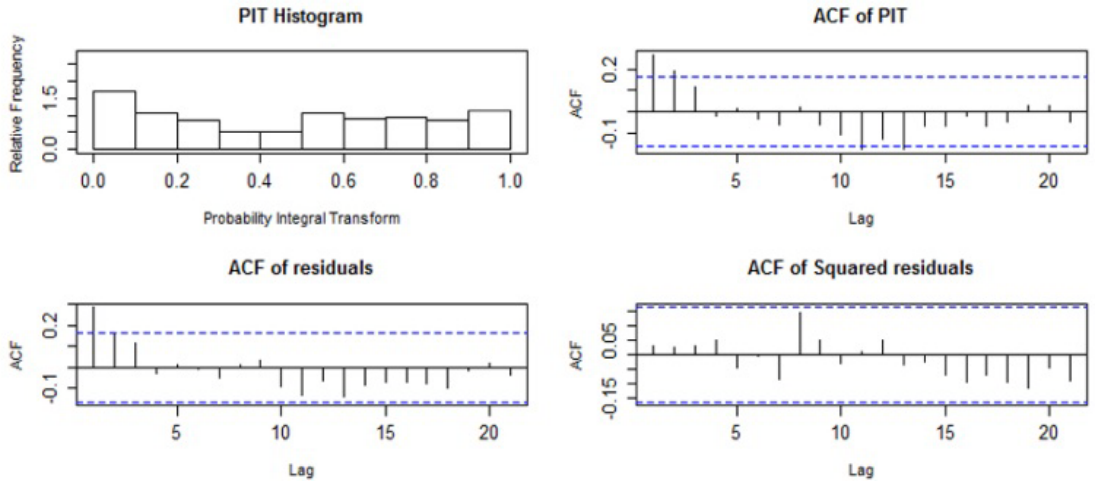


Figure 7.7: PIT, ACF and PACF of W-INGARCH (1, 1) with covariate MEI.

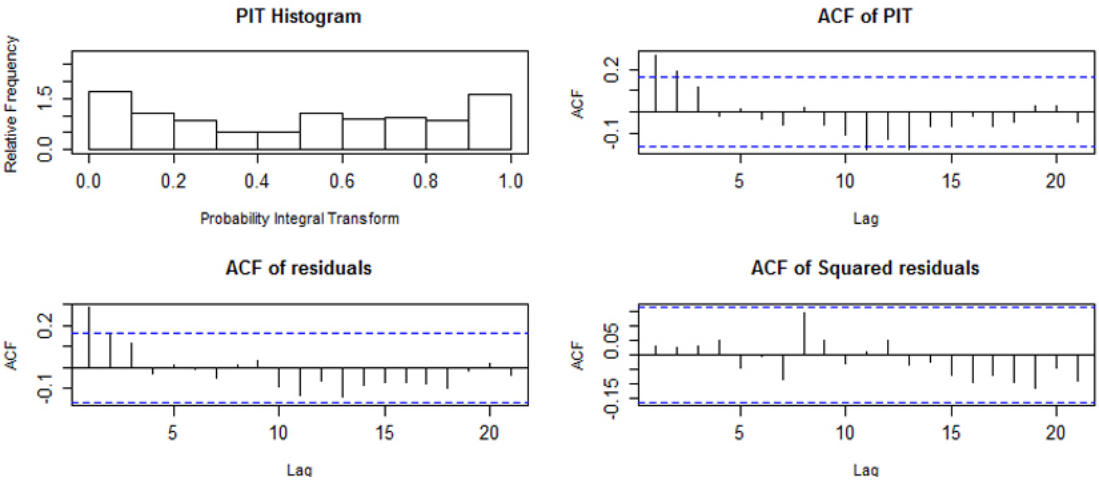
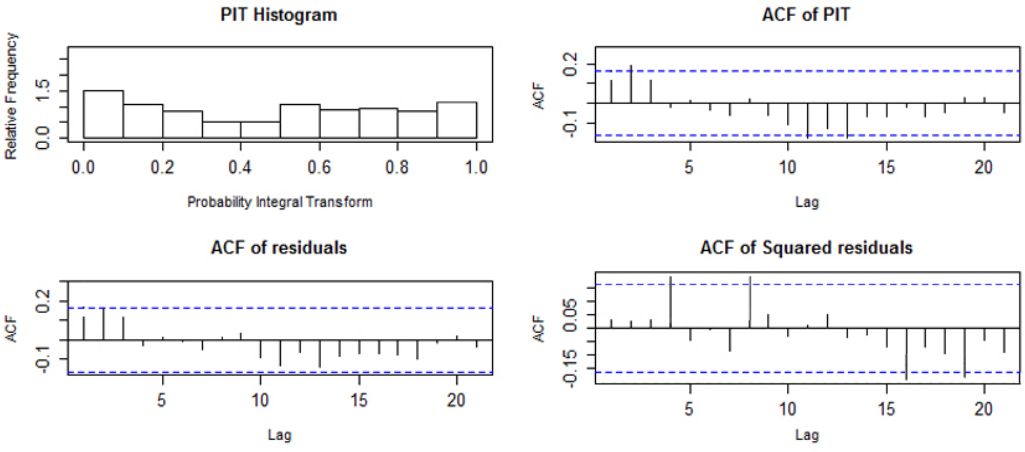


Figure 7.8: PIT, ACF and PACF of W-INGARCH-NB(1, 1) with covariate {IOD,MEI}.



F

7.2.5 *Forecast accuracy*

Results from figure 7.10-7.12 the standardized residuals for three models are presented. The QQ-plot for the INGARCH-NB(1,1)with wavelet covariate MEI is the only plot showing residuals that can be considered to follow a normal distribution, while for the INGARCH(1,1) with wavelet covariate MEI and INGARCH-NB(1,1) with covariates {IOD, MEI} for wavelet is not.

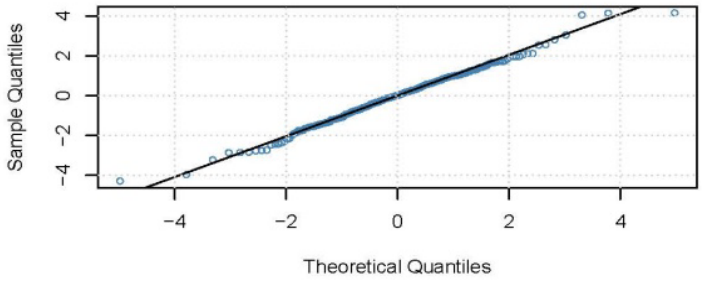


Figure 7.10: QQ Plot for normality test of W-INGARCH(1,1) with covariate MEI.

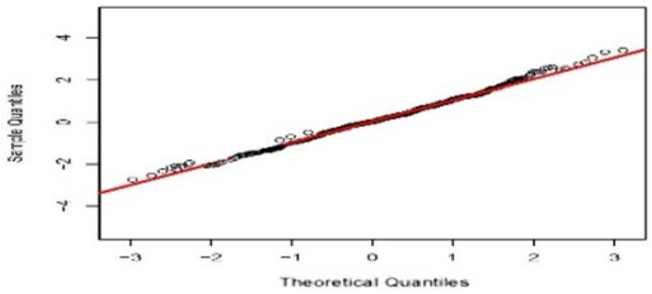


Figure 7.11: QQ Plot for normality test of W-INGARCH-NB(1, 1) with covariates {IOD, MEI}.

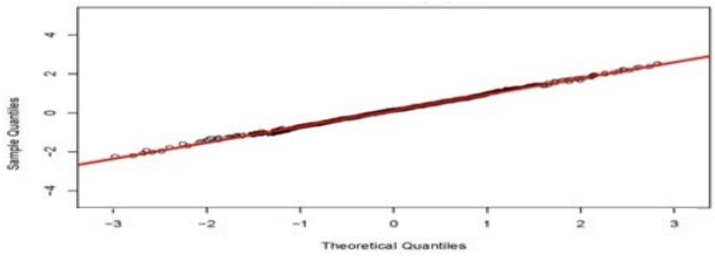


Figure 7.12: QQ Plot for normality test of W-INGARCH-NB(1, 1) with covariate MEI.

7.3 Conclusion

The aim of this study was to investigate models that are applicable to time series for count IDD. This research brought out the advantages of using models developed for time series counts in addition to the conventional techniques used in count data

modeling, and demonstrated that actually these models has best characteristics and flexible for the case in this study.

In this chapter, IDD was target variable and MEI and IOD were used as covariates. From the results it is revealed that W-GARCH-NB(1,1) model with covariate MEI is superior to the models without covariate(s).

CHAPTER EIGHT
Summary and Conclusion

Chapter Eight
Summary and Conclusion

8.1 Introduction

This chapter provides a short summary of the main results obtained in this research and scope of future study.

8.2 Summary

Chapter one summarizes the key concepts to justify the title of the study and have tried to answer what and why the study is relevant to modeling and forecasting based on time varying parameters. The vital question and objective of the research was- how much effective global scale climatic index helps quantifying the incidence of disease via temporal distributional formation? Accordingly, the main objective was to establish a new efficient count valued distributional temporal model of the volatility behavior in the incidence of diarrheal disease agent(s) with and without global scale climatic factors. At the end of the chapter the motivation remarks concluded that

(1) Quantitative modeling and forecasts about health events could be a quality of task(s) in research with respect to global scale climate factors
(2) A true forecast of hospitalized patients is one of the key challenges for health care policy makers to better allocate medical resources and service providers. However, the justification demonstrated that further improvements are needed for modeling and forecasting approaches to reliably useful for policy making purposes.

In Chapter two, the research gap is mainly explored based on literature review. This chapter also raised different issues of climate data analysis and limitations of statistical methods to solve these issues. Most of the models used in previous research related to the study are mean model for continuous time series data. But links between disease and climatic factors (global/local) are non-linear, non-normal and volatility based are found in literatures. Although a few studies forecasted the future burden of diarrhea in relation to global climate driver factors, MEI or IOD, under count time series data modeling through GLM. But the shrinkage tools, wavelet, with volatility model in count time series distributional model for univariate or multivariate is still not found as an efficient technique like the proposed model. Finally, the chapter is ended up with several research gaps to answer - can large scale climatic index helps forecasting the count infectious diarrheal disease via wavelet GARCH temporal distributional formation?

Chapters three, four and five considered methodology and exploratory data learning part that grounded the base of analytical methods to forward modeling for non-linear, non-normal and volatility features. In these chapters data processing, presentation and some descriptive measures are taken that make the base to final modeling. The run sequence plot of time series (in level, in first seasonal difference, monthly and yearly), boxplot, ACF and PACF have been used to inspect and resulted in:

Data series(s) tend to exhibit repetitive behavior, with regularly repeating cycles. Monthly oscillated behavior is found in the series MEI, IOD and IDD and irregular rate or frequency of oscillation was characterized to go forward for volatility modeling. In the series it also found that the cycles of the data series are repeating at an extra ordinary order from others implied that outlier is contained. Boxplot shows the variations of data series under study with outlier, non-linear and non-normal features. After that, stationary condition is checked. The ADF and KPSS tests show that the series is non-stationary.

In Chapter six, three different methods of modeling namely mean model, volatility model and smoothed volatility model are used for analyzing the data series to predict future values for the monthly global climate data index. Observing ACF and PACF several ARIMA models have been fitted. The results of diagnostic checking, model

comparison, model adequacy, model selection and forecasting performance of the models are summarized as follows -

- ARIMA orders had tested over its different order and ARIMA(1, 1, 2) for MEI, ARIMA(2, 1, 2) for IOD and INARMA(2, 2) for IDD had been found appropriate. The correlogram of the estimated residual for the respective models reveal the estimated residuals obtained from selected models are not white noise.
- QQ plot shows that the errors are roughly normal for all selected models. But the McLeo-Li test shows that they may be non-normal for all the cases.
- Having checked the arch effects and positive serial dependency in MEI and IOD, GARCH orders had tested over its different order for actual and de-noised. GARCH(1, 2) and GARCH(1,1) for actual MEI and de-noised MEI respectively are selected as the best model. One the other hand, GARCH(2, 2) and GARCH(1,1) for actual IOD and de-noised IOD respectively are selected as the best model.
- Finally, comparison between the results of the best GARCH and the best W-GARCH were performed with distributions normal, student-t and GED to determine which performs better. Excellent modeling performances were confirmed by the W-GARCH(1, 1) model in normal distribution for MEI and W-GARCH(1, 1) model in student-t distribution for IOD.

In Chapter seven, proposed non-linear count time series hybrid model W-INGARCH-NB(1, 1) for IDD patients' count data with covariate MEI are investigated. Performance of the proposed model showed best results as AIC, BIC, squared residuals, Pit histogram has been quantified the lowest. On the other hand, the W-INGARCH-NB(1, 1) model with covariates {IOD, MEI} also gains lowest AIC and BIC than the proposed model but the later model's coefficient of IOD in estimation procedure showed insignificant.

8.3 Recommendations

Burden of infectious diarrhea is still much in low-income countries; it is a common cause of outpatient visits and hospital admissions in high-income countries as well.

Global climatic events (either heat waves or cold waves) are becoming more intense. As a result, the ever-increasing demand for both the coverage and quality of health care services, health service delivery institutions and service providers have to tackle situations of excess demand. But the front-line health delivery services and providers are not usually adequately prior informed and do not have adequate resources to meet the needs for health care. Therefore, improving the access, coverage and quality of health services depends on the modeling and forecasting, where service providers can be pre-informed, organized and managed. The proposed model will be beneficial for that. The specific recommendation is that every health service provider should build an organizational real time database for long range planning, prediction and logistic support. In addition to that a zonal socio-demographic index should be designed based on visiting patients.

8.4 Further Research Scope

In this study, the methodology applied to only one hospital, but the approach is expected better performance to analyze trends of other similar infectious diseases that is sensitive to climatic factors. Building a general model for forecasting trends of disease based on several hospitals with advanced integer-valued time series models in spatial method would be an interesting work. Robustification of the proposed model would be another interesting challenge to the researcher.

References

References

Abdussalam, A. F. (2016). Modelling the Climatic Drivers of Cholera Dynamics in Northern Nigeria Using Generalised Additive Models. *Intl. J of Geography and Environmental Management*, ISSN 2504-8821, **2(1):** 84-97.

Adamowski, J., and K. Sun. (2010). Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology*, 390(1-2), 85-91.

Ahmed KM, M. Nishigaki, M Dewan. (2005) Constraints and issues on sustainable groundwater exploitation in Bangladesh. *J Groundw Hydrol*, **47(2):** 163–179.

Ahmed MF (2009) Environment impacts of sea-level rise on the coastal areas of Bangladesh: Climate change and the tasks for Bangladesh. *Bangladesh Poribesh Andolon and Bangladesh Environment Network*, Dhaka, pp 88–100

Ahn, S., G. Lee, and J. Jeon. (2000). Analysis of the M/D/1-type queue based on an integer-valued first-order autoregressive process. *Operations Research Letters*, **27(5),** 235-241.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, **19(6),** 716-723.

Al Wadia, M., and M. Tahir Ismail. (2011). Selecting wavelet transforms model in forecasting financial time series data based on ARIMA model. *Applied Mathematical Sciences*, **5(7),** 315-326.

Alexander, K., M. Carzolio, D. Goodin, and E. Vance. (2013). Climate change is likely to worsen the public health threat of diarrheal disease in Botswana. *International journal of environmental research and public health*, **10(4),** 1202-1230.

Allan, R. J., N. Nicholls, P. D. Jones, and I. J. Butterworth. (1991). A further extension of the Tahiti-Darwin SOI, early ENSO events and Darwin pressure. *Journal of Climate*, **4(7),** 743-749.

Allan, R., D. Chambers, W. Drosdowsky, H. Hendon, M. Latif, N. Nicholls, and Y. Tourre. (2001). Is there an Indian Ocean dipole and is it independent of the El Niño-Southern Oscillation. *CLIVAR exchanges*, **21,** 18-22.

Alosh, M. (2009). The impact of missing data in a generalized integer-valued autoregression model for count data. *Journal of biopharmaceutical statistics*, **19(6),** 1039-1054.

Ardkaew, J., and P. Tongkumchum. (2009). Statistical modelling of childhood diarrhea in northeastern Thailand. Southeast Asian *Journal of Tropical Medicine and Public Health*, **40(4),** 807.

Arino, J., C. Bauch, F. Brauer, S. M. Driedger, A. L. Greer, S. M. Moghadas, and P. Van Den Driessche. (2011). Pandemic influenza: modelling and public health perspectives. *Mathematical Biosciences & Engineering*, **8(1)**, 1-20.

Arino, J., F. Brauer, P. Van Den Driessche, J. Watmough, and J. Wu. (2008). A model for influenza with vaccination and antiviral treatment. *Journal of theoretical biology*, **253(1)**, 118-130.

Armstrong, J. S. (2001). Principles of forecasting: a handbook for researchers and practitioners (Vol. 30): *Springer Science & Business Media*.

Auffhammer, M., S. M. Hsiang, W. Schlenker, and A. Sobel. 2013). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy*, **7(2),** 181-198.

Azmoodehfar, M. H., and S. A. Azarmsa. (2013). Assessment the effect of ENSO on weather temperature changes using fuzzy analysis (case study: Chabahar). *APCBEE procedia*, **5,** 508-513.

Baillie, R. (1997). Time Dependent Conditional Heteroskedasticity. *Paper presented at the manuscript, Workshop of Time Series Analysis, Arrábida, Portugal*.

Bangladesh Bureau of Statistics (BBS)(2012a).
http://www.bbs.gov.bd/WebTestApplication/ userfiles/Image/BBS/GDP_2011.pdf.
Accessed 15 Sept 2012

Bangladesh Bureau of Statistics (BBS) (1984) Bangladesh population census 1981, National series, analytical findings and national tables. Bangladesh Bureau of Statistics, Ministry of Planning, Dhaka

Bangladesh Bureau of Statistics (BBS) (1994) Bangladesh population census 1991, vol 1, Analyti- cal findings. Bangladesh Bureau of Statistics, Ministry of Planning, Dhaka

Bangladesh Bureau of Statistics (BBS) (2004) Bangladesh population census 2001, National series. Bangladesh Bureau of Statistics, Ministry of Planning, Dhaka

Bangladesh Bureau of Statistics (BBS) (2008) Population census – 2001, vol 3 National series, Urban area report, Bangladesh Bureau of Statistics, Ministry of Planning, Dhaka

Bangladesh Bureau of Statistics (BBS) (2010a) Census of agriculture 2008. Bangladesh Bureau of Statistics, Ministry of Planning, Dhaka

Bangladesh Bureau of Statistics (BBS) (2011a) Monthly statistical bulletin – February 2011. Bangladesh Bureau of Statistics, Ministry of Planning, Dhaka.

Bangladesh Bureau of Statistics (BBS) (2011b) Household Income and Expenditure Survey (HIES) – 2010. Bangladesh Bureau of Statistics, Ministry of Planning, Dhaka

Bangladesh Bureau of Statistics (BBS) (2012b) Population and housing census 2011, Community series – Dhaka, Gazipur, Narayanganj. Bangladesh Bureau of Statistics, Ministry of Planning, Dhaka

Behera, S. K., J. J. Luo, S. Masson, S. A. Rao, H. Sakuma, and T. Yamagata. (2006). A CGCM study on the interaction between IOD and ENSO. *Journal of Climate*, **19(9),** 1688-1705.

Behrens, J. T., and C. H. Yu. (2003). Exploratory data analysis. *Handbook of psychology*.

Bennett, A., L. D. Epstein, R. H. Gilman, V. Cama, C. Bern, L. Cabrera and. C. R. Sterling. (2012). Effects of the 1997–1998 El Niño episode on community rates of diarrhea. *American journal of public health*, **102(7),** e63-e69.

Bhandari, G., S. Gurung, M. Dhimal and C. Bhusal. Climate change and occurrence of diarrheal diseases: Evolving facts from Nepal. *J Nepal Health Res Counc,* **10(22),** 181-186.

Black, R. and C. Lanata. (1995). Epidemiology of diarrheal diseases in developing countries (pp. 13-36): *Raven Press, New York.*

Blundell, R., Griffith, R. and F. Windmeijer. (2002). Individual effects and dynamics in count data models. *Journal of econometrics,* **108(1),** 113-131.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, **31(3),** 307-327.

Boschi-Pinto, C., L. Velebit and K. Shibuya. (2008). Estimating child mortality due to diarrhoea in developing countries. *Bulletin of the World Health Organization,* **86**, 710-717.

Bowerman, B. L., R. T. O'Connell and A.B. Koehler. (2005). Forecasting, time series, and regression: an applied approach: *Thomson Brooks/Cole*.

Box, G. and G. Jenkins. (1976). Time Series Analysis, Forecasting and Control, Holden Day, San Francisco.

Box, G. E., G. M. Jenkins, G. C. Reinsel and G. M. Ljung. (2015). Time series analysis: forecasting and control: John Wiley & Sons.

Boyle, J., M. Jessup, J. Crilly, D. Green, J. Lind., M. Wallis and G. Fitzgerald (2012). Predicting emergency department admissions. *Emerg Med J,* **29(5),** 358-365.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika,* **52(3),** 345-370.

Bradley, V. M. (2005). Placing emergency department crowding on the decision agenda. *Nursing Economics*. **23(1),** 14.

Brammer H (2002) Land use and land use planning in Bangladesh. *The University Press Limited, Dhaka.*

Brännäs, K., J. Hellström and Nordström (2002). A new approach to modelling and forecasting monthly guest nights in hotels. *International Journal of Forecasting*, **18(1),** 19-30.

Bretherton, F. P. (1982). Ocean climate modeling. *Progress in Oceanography*, **11(2),** 93-129.

Brockwell, P. J., and R. A. Davis (2016). Introduction to time series and forecasting: *springer*.

Butterton, J., and S. Calderwood. (2005). Acute infectious diarrheal diseases and bacterial food poisoning. *HARRISONS PRINCIPLES OF INTERNAL MEDICINE*, **16(1),** 754.

C, K. C. (2016). Integer-valued ARCH and GARCH Models.

Cai, W., A. Sullivan, and T. Cowan (2009). How rare are the 2006–2008 positive Indian Ocean dipole events? *An IPCC AR4 climate model perspective. Geophysical Research Letters*, **36(8).**

Cameron, A. C., and P.K. Trivedi (2013). Regression analysis of count data **(Vol. 53)**: *Cambridge university press.*

Cardinal, M., R. Roy, and J. Lambert (1999). On the application of integer valued time series models for the analysis of disease incidence. *Statistics in Medicine*, **18(15)**, 2025-2039.

Carlton, E. J., J. N. Eisenberg, J. Goldstick, W. Cevallos, J. Trostle, and K. Levy (2013). Heavy rainfall events and diarrhea incidence: the role of social and environmental factors. *American journal of epidemiology*, **179(3)**, 344-352.

Casals, M., M. Girabent-Farres, and J. L. Carrasco (2014). Methodological quality and reporting of generalized linear mixed models in clinical medicine (2000–2012): a systematic review. *PloS one*, **9(11),** e112653.

Cash, B. A., X. Rodó, J. L. Kinter III and M. Yunus (2010). Disentangling the impact of ENSO and Indian ocean variability on the regional climate of Bangladesh: Implications for cholera risk. *Journal of Climate*, **23(10)**, 2817-2831.

Cazelles, B., M. Chavez, A. J. McMichael & S. Hales (2005). Nonstationary influence of El Nino on the synchronous dengue epidemics in Thailand. *PLoS medicine*, **2(4),** e106.

Chaovalit, P., A. Gangopadhyay, G. Karabatis and Z. Chen (2011). Discrete wavelet transform-based time series analysis and mining. ACM Computing Surveys *(CSUR)*, **43(2),** 6.

Chatfield, C. (2016). The analysis of time series: an introduction: Chapman and Hall/CRC.

Checkley, W., L.D. Epstein, R. H. Gilman, D. Figueroa, R. I. Cama, J. A. Patz and R. E. Black (2000). Effects of EI Niño and ambient temperature on hospital admissions for diarrhoeal diseases in Peruvian children. *The Lancet*, *355(9202)*, 442-450.

Chen, C. S., C. H. Liu and H. C. Su (2008). A nonlinear time series analysis using two stage genetic algorithms for stream flow forecasting. *Hydrological Processes*: *An International Journal*, **22(18),** 3697-3711.

Chen, C., J. Hu, Q. Meng, and Y. Zhang (2011). Short-time traffic flow prediction with ARIMA-GARCH model. *Paper presented at the Intelligent Vehicles Symposium* (IV), 2011 *IEEE.*

Chen, C. C., B. C. Lin, L. Yap, P. H. Chiang and T. C. Chan (2018). The Association between Ambient Temperature and Acute Diarrhea Incidence in Hong Kong, Taiwan, and Japan. *Sustainability*, **10(5),** 1417.

Chen, C. C., B. McCarl, and H. Hill(2002). Agricultural value of ENSO information under alternative phase definition. *Climatic Change*, **54(3),** 305-325.

Cheng, A. C., J.R. McDonald and N. M. Thielman (2005). Infectious diarrhea in developed and developing countries. *Journal of clinical gastroenterology*, **39(9),** 757-773.

Chiann, C., and P.A. Morettin (1998). A wavelet analysis for time series. *Journal of Nonparametric Statistics*, **10(1),** 1-46.

Chou, W. C., J. L. Wu, Y. C. Wang, H. Huang, F. C. Sung and C. Y. Chuang (2010). Modeling the impact of climate variability on diarrhea-associated diseases in Taiwan (1996–2007). *Science of the Total Environment*, **409(1),** 43-51.

Choudhury AKMK (2008) Land use planning in Bangladesh. A H Development Publishing House, Dhaka

Chowdhury JU and M. Salehin (1987) Floods and their processes. *In: Proceedings of the international*

Clements, M. P., P. H. Franses and N.R. Swanson (2004). Forecasting economic and financial time-series with non-linear models. *International Journal of Forecasting*, **20(2),** 169-183.

Colwell, R. R. and J. A. Patz (1998). Climate, infectious disease and health: an interdisciplinary perspective.

Cook, R. D., and S. Weisberg (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*, 70(1), 1-10.

Cox, V. (2017). Exploratory data analysis. In Translating Statistics to Make Decisions (pp. 47-74). *Apress, Berkeley, CA*.

Cryer, J. D., & Kellet, N. (1991). Time series analysis: Springer.

Cui, Y., and R. Lund (2009). A new look at time series of counts. *Biometrika*, **96(4),** 781-792.

Darda, M. A. (2014). On Modelling Temperature Volatility Dynamics. *Jahangirnagar university journal of information technology (jujit)* vol. 3, issue. 1, september 2014, vol. 3,(issue. 1,september 2014).

Data, C. (2009). Guidelines on Analysis of extremes in a changing climate in support of informed decisions for adaptation. *World Meteorological Organization*.

Daubechies, I. (1992). Ten lectures on wavelets: *SIAM*.

Derlet, R. W. (2002). Overcrowding in emergency departments: increased demand and decreased capacity. *Annals of emergency medicine*, **39(4),** 430-432.

Dewan AM, MH Kabir, K. Nahar, MZ Rahman (2012) Urbanization and environmental degradation in Dhaka Metropolitan Area of Bangladesh. *Int J Environ Sustain Dev* **11(2):**118–146

Dickey, D. A. and W.A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74(366a),** 427-431.

Dong, Z., X. Guo, J. Zheng and L. Xu (2001). Calculation of noise resistance by use of the discrete wavelets transform. *Electrochemistry communications*, **3(10),** 561-565.

Douglas I (2009) Climate change, flooding and food security in South Asia. *Food Secur* **1 (2):**127–136

Drasar, B., A Tomkins and R. Feacham (1978). Seasonal aspects of diarrheal disease. Seas. Dimens. to Rural poverty.(Report to UK Overseas Dev. Assoc. London Sch. Hyg. Trop. Med. London. London, UK.

D'souza, R., G. Hall and N. Becker (2008). Climatic factors associated with hospitalizations for rotavirus diarrhea in children under 5 years of age. *Epidemiology & Infection*, **136(1),** 56-64.

Earnest, A., S. Tan and A. Wilder-Smith (2012). Meteorological factors and El Nino Southern Oscillation are independently associated with dengue infections. *Epidemiology & Infection*, **140(7),** 1244-1251.

Estrada, F., and P. Perron (2014). Detection and attribution of climate change through econometric methods. *Boletín de la Sociedad Matemática Mexicana*, **20(1),** 107-136.

Fahrmeir, L., and G. Tutz (2013). Multivariate statistical modelling based on generalized linear models: Springer Science & Business Media.

Fan, J., J. Meng, Y. Ashkenazy, S. Havlin and H.J. Schellnhuber (2017). Network analysis reveals strongly localized impacts of El Niño. *Proceedings of the National Academy of Sciences*, **114(29),** 7543-7548.

Ferland, R., A. Latour and D. Oraichi (2006). Integer valued GARCH process. *Journal of Time Series Analysis*, **27(6),** 923-942.

Fisman, D. N., A. R. Tuite and K.A. Brown (2016). Impact of El Niño Southern Oscillation on infectious disease hospitalization risk in the United States. *Proceedings of the National Academy of Sciences*, **113(51),** 14589-14594.

Fokianos, K. (2015). Statistical analysis of count time series models: A GLM perspective. Handbook of discrete-valued time series, Handbooks of Modern Statistical Methods, 3-28.

Fokianos, K., and R. Fried (2010). Interventions in INGARCH processes. *Journal of Time Series Analysis*, **31(3),** 210-225.

Fokianos, K., A. Rahbek and D. Tjøstheim (2009). Poisson autoregression. *Journal of the American Statistical Association*, **104(488**), 1430-1439.

Franke, J., and T. Seligmann (1993). Conditional maximum likelihood estimates for INAR (1) processes and their application to modelling epileptic seizure counts. *Developments in time series analysis*, 310-330.

Freeman, E., S. D. Woodruff, S. J. Worley, S. J. Lubker, E. C. Kent, W. E. Angel and L. Gates (2016). ICOADS Release 3.0: a major update to the historical marine climate record. *International Journal of Climatology*.

Gauthier, G., and A. Latour(1994). Convergence forte des estimateurs des paramètres d'un processus GENAR (p). Paper presented at the Annales des Sciences Mathématiques du Québec.

Ghil, M., M. Allen, M. Dettinger, K. Ide, D. Kondrashov, M. Mann and F. Varadi (2002). Advanced spectral methods for climatic time series. Reviews of geophysics, 40(1).

Gluhovsky, A., and E. Agee (2007). On the Analysis of Atmospheric and Climatic Time Series. *Journal of Applied Meteorology and Climatology*, **46(7),** 1125-1129. doi: 10.1175/jam2512.1

Gneiting, T., F. Balabdaoui and A. E. Raftery (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series* B *(Statistical Methodology)*, **69(2)**, 243-268.

Gottman, J. M. (1981). Time-series analysis comprehensive introduction for social scientists.

Gouriéroux, C. (2012). ARCH models and financial applications: Springer Science & Business Media.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, **42(1),** 27-62.

Greenland, D. (2003). An LTER Network overview and introduction to El Niño-Southern Oscillation (ENSO) Climatic signal and response: *Oxford University Press: New York.*

Grek, Å. (2014). Forecasting accuracy for ARCH models and GARCH (1, 1) family: Which model does best capture the volatility of the Swedish stock market?

Hales, S., S. Edwards and R Kovats, R. (2003). Impacts on health of climate extremes. *Climate change and human health: Risks and responses*, 79-96.

Hallett, T., T. Coulson, J. Pilkington, T. Clutton-Brock, J. Pemberton, and B Grenfell (2004). Why large-scale climate indices seem to predict ecological processes better than local weather. *Nature*, **430(6995),** 71.

Hamilton, J. D. (1994). Time series analysis (Vol. 2): Princeton university press Princeton, NJ.

Hang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, **50**, 159-175.

Hannan, E. J., and B. G.Quinn (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)*, **41(2),** 190-195.

Hardin, J. W., Hardin, J. W., Hilbe, J. M., & Hilbe, J. (2007). Generalized linear models and extensions: Stata press.

Harris, T., J. M. Hilbe and J. W. Hardin (2014). Modeling count data with generalized distributions. **Stata Journal**, **14(3),** 562-579.

Hashim, J. H., and Z. Hashim (2016). Climate change, extreme weather events, and human health implications in the Asia Pacific region. *Asia Pacific Journal of Public Health*, **28(2_suppl),** 8S-14S.

Hashizume, M., B. Armstrong, S. Hajat, Y. Wagatsuma, A. S. Faruque, T. Hayashi and D. A. Sack (2007). Association between climate variability and hospital visits for non-cholera diarrhoea in Bangladesh: effects and vulnerable groups. *International journal of epidemiology*, **36(5),** 1030-1037.

Hashizume, M., B. Armstrong, Y. Wagatsuma, A. Faruque, T. Hayashi, and D. A. Sack (2008). Rotavirus infections and climate variability in Dhaka, Bangladesh: a time-series analysis. *Epidemiology & Infection*, **136(9),** 1281-1289.

Hashizume, M., L. F. Chaves and N. Minakawa (2012). Indian Ocean Dipole drives malaria resurgence in East African highlands. *Scientific reports*, **2**, 269.

Hashizume, M., L.F. Chaves, A. Faruque, M. Yunus, K. Streatfield, and K. Moji (2013). A differential effect of Indian ocean dipole and El Niño on cholera dynamics in Bangladesh. *PloS one*, **8(3),** e60001.

Hashizume, M., A. Faruque, T. Terao, M. Yunus, K. Streatfield, T. Yamamoto, and K. Moji (2011). The Indian Ocean dipole and cholera incidence in Bangladesh: a time-series analysis. *Environmental health perspectives*, **119(2),** 239.

Hashizume, M., T. Terao, and N. Minakawa (2009). The Indian Ocean Dipole and malaria risk in the highlands of western Kenya. *Proceedings of the National Academy of Sciences, pnas.* 0806544106.

Heinen, A. (2003). Modelling time series count data: an autoregressive conditional Poisson model. Available at *SSRN* 1117187.

Hendon, H. H., E. P. Lim, and G. Liu (2012). The role of air–sea interaction for prediction of Australian summer monsoon rainfall. *Journal of Climate***, 25(4),** 1278-1290.

Hermanianto, A. B., A. Buono and K. K. Nisa. Length of Rainy Season Prediction Based on Southern Oscillation Index and Dipole Mode Index Using Support Vector Regression.

Allen, M., M. Dettinger, K. Ide, D. Kondrashov, M. Mann and F. Varadi (2002). Advanced spectral methods for climatic time series. *Reviews of geophysics*, **40(1).**

Hilbe, J. M. (2011). Modeling count data: Springer.

Hipel, K. W. and A. I. McLeod (1994). Time series modelling of water resources and environmental systems (Vol. 45): *Elsevier*.

Hoaglin, D. C., F. Mosteller and J.W. Tukey (2009). Fundamentals of exploratory analysis of variance (Vol. 367): John Wiley & Sons.

Hurvich, C. M., and C. L. Tsai (1989). Regression and time series model selection in small samples. *Biometrika,* **76(2),** 297-307.

Ikarmpapraharn, C. and E. Kositsakulchai (2010). Relationship between ENSO and rainfall in the Central Plain of Thailand. Kasetsart *Journal (Natural Science),* **44,** 744-755.

Islam N (2005a) Dhaka now: contemporary urban development, Golden jubilee series no. 2. Bangladesh Geographical Society (BGS), Dhaka

Islam T. and RE Peterson (2009) Climatology of landfalling tropical cyclones in Bangladesh 1877–2003. *Nat Hazards* **48(1):**115–135

Izumo, T., J. Vialard, M. Lengaigne, C. de Boyer Montegut, S.K. Behera, J.J. Luo and T. Yamagata (2010). Influence of the state of the Indian Ocean Dipole on the following year's El Niño. *Nature Geoscience*, **3(3),** 168.

Jevrejeva, S., J. Moore and A. Grinsted (2003). Influence of the Arctic Oscillation and El Niño Southern Oscillation (ENSO) on ice conditions in the Baltic Sea: The wavelet approach. *Journal of Geophysical Research: Atmospheres*, **108(D21).**

Johansson, M. A., D. A. Cummings and G. E. Glass (2009). Multiyear climate variability and dengue—El Nino southern oscillation, weather, and dengue incidence in Puerto Rico, Mexico, and Thailand: a longitudinal data analysis. *PLoS medicine*, **6(11),** e1000168.

Jung, R. C. and A. Tremayne (2011). Useful models for time series of counts or simply wrong ones? *AStA Advances in Statistical Analysis*, **95(1),** 59-91.

Kale, P. L., J. P. Hinde and F. F. Nobre (2004). Modeling diarrhea disease in children less than 5 years old. *Annals of epidemiology*, **14(6),** 371-377.

Karim, S. A. A., M. T. Ismail, M. K. Hasa and J. Sulaiman (2013). Denoising the temperature data using wavelet transform. *Applied Mathematical Sciences*, **7(117),** 5821-5830.

Kassebaum, N., H. H. Kyu, L. Zoeckler, H.E. Olsen, K. Thomas, C. Pinho, and T. T. Ghiwot (2017). Child and adolescent health from 1990 to 2015: findings from the global burden of diseases, injuries, and risk factors 2015 study. *JAMA pediatrics*, **171(6),** 573-592.

Kedem, B., and K. Fokianos (2005). Regression models for time series analysis (Vol. 488): John Wiley & Sons.

Koelle, K., X. Rodó, M. Pascual, M. Yunus, and G. Mostafa (2005). Refractory periods and climate forcing in cholera dynamics. *Nature*, **436(7051),** 696.

Kolstad, E. W., and K. A. Johansson (2010). Uncertainties associated with quantifying climate change impacts on human health: a case study for diarrhea. *Environmental health perspectives*, **119(3),** 299-305.

Konno, T., H. Suzuki, N. Katsushima, A. Imai, F. Tazawa, T. Kutsuzawa, and N. Ishida (1983). Influence of temperature and relative humidity on human rotavirus infection in Japan. *Journal of infectious diseases*, **147(1),** 125-128.

Kovats, R. S., M. J. Bouma, S. Hajat, E. Worrall and A. Haines (2003). El Niño and health. *The Lancet*, **362(9394),** 1481-1489.

Kriechbaumer, T., A. A. Angus, D. Parsons and M. R. Casado (2014). An improved wavelet–ARIMA approach for forecasting metal prices. *Resources Policy*, **39,** 32-41.

Kruijshaar, M. E., J. J. Barendregt and N. Hoeymans (2002). The use of models in the estimation of disease epidemiology. Bulletin of the World Health Organization, 80, 622-628.

Labat, D. (2010). Cross wavelet analyses of annual continental freshwater discharge and selected climate indices. *Journal of Hydrology*, **385(1-4),** 269-278.

Lama, J. R., C. R. Seas, R. León-Barúa E. Gotuzzo and R. B. Sack (2004). Environmental temperature, cholera, and acute diarrhoea in adults in Lima, Peru. *Journal of Health, Population and Nutrition*, 399-403.

Lee, J. Univariate time series modeling and forecasting (Box-Jenkins Method). Econ 413, lecture 4.

Lee, T., T. Ouarda and S. Ousmane (2011). Predictability of climate indices with time series models. Stochastic Hydrology of the Great Lakes–A Systemic Analysis, 45.

Lewis, J. B., R. J. McGrath and L. F. Seidel (2009). Essentials of applied quantitative methods for health services: Jones & Bartlett Learning.

Lim, E.-P., H. H. Hendon, M. Zhao, and Y. Yin (2017). Inter-decadal variations in the linkages between ENSO, the IOD and south-eastern Australian springtime rainfall in the past 30 years. *Climate Dynamics*, **49(1-2),** 97-112.

Lim, G. h., Y. C. Suh, Y. and B. M. Kim (2006). On the origin of the tropical Atlantic decadal oscillation based on the analysis of the ICOADS. *Quarterly Journal of the Royal Meteorological Society*, **132(617),** 1139-1152.

Liu, H., J. Shi and X. Qu (2013). Empirical investigation on using wind speed volatility to estimate the operation probability and power output of wind turbines. *Energy Conversion and Management*, **67**, 8-17.

Lobitz, B., L Beck, A. Huq, B. Wood, G. Fuchs, A. Faruque and R. Colwell (2000). Climate and infectious disease: use of remote sensing for detection of Vibrio cholerae by indirect measurement. *Proceedings of the National Academy of Sciences*, **97(4),** 1438-1443.

Luo, J.-J., R. Zhang, S. K. Behera, Y. Masumoto, F. F. Jin, R. Lukas and T. Yamagata (2010). Interaction between El Nino and extreme Indian ocean dipole. *Journal of Climate,* **23(3),** 726-742.

MacDonald, G. M., and R. A. Case (2005). Variations in the Pacific Decadal Oscillation over the past millennium. *Geophysical Research Letters*, **32(8).**

MacDonald, I. L., and W. Zucchini (1997). Hidden Markov and other models for discrete-valued time series (Vol. 110): CRC Press.

Maniruzzaman KM (eds) Urbanization in Bangladesh: patterns, issues and approaches to planning. *The Bangladesh Institute of Planners (BIP),* Dhaka, pp 1–24

Marno, P., M. Chalder, T. Laing-Morton, M. Levy, P. Sachon and D. Halpin (2010). Can a health forecasting service offer COPD patients a novel way to manage their condition? *Journal of health services research & policy*, **15(3)**, 150-155.

Mazzarella, A., A. Giuliacci and I. Liritzis (2010). On the 60-month cycle of multivariate ENSO index. *Theoretical and applied climatology*, **100(1-2),** 23-27.

Mazzarella, A., A. Giuliacci and F. Pregliasco (2011). Hypothesis on a possible role of El Niño in the occurrence of influenza pandemics. *Theoretical and applied climatology*, **105(1-2),** 65-69.

McGregor, G., and K. Ebi (2018). El Niño Southern Oscillation (ENSO) and Health: An Overview for Climate and Health Researchers. *Atmosphere*, **9(7),** 282.

McKenzie, E. (2003). Ch. 16. Discrete variate time series. Handbook of statistics, 21, 573-606.

McMichael, A. J., R. E. Woodruff and S. Hales (2006). Climate change and human health: present and future risks. *The Lancet*, **367(9513**), 859-869.

Mellor, J., E. Kumpel, A. Ercumen and J. Zimmerman (2016). Systems Approach to Climate, Water, and Diarrhea in Hubli-Dharwad, India. *Environmental science & technology,* **50(23),** 13042-13051.

Meyer, Y. Wavelets: Algorithms and Applications (SIAM, Philadelphia, 1993). MATH Google Scholar.

Milidiu, R. L., R. J. Machado and R. P. Rentería (1999). Time-series forecasting through wavelets transformation and a mixture of expert models. *Neurocomputing*, **28(1),** 145-156.

Mirza MMQ (2009) Climate change risks in Bangladesh. In: Bhattacharya D, Ahmed MF, Islam N, Matin MD (eds) Climate change and the tasks for Bangladesh. Bangladesh Paribesh Andolon and Bangladesh Environment Network, Dhaka, pp 69–87

Misiti, M., Y. Misiti, G. Oppenheim and J. M. Poggi (2008). Wavelet ToolboxTM 4 User's Guide. The MathWorks, Inc, 3, 1-47.

Modarres, R., and T. Ouarda (2014). Modeling the relationship between climate oscillations and drought by a multivariate GARCH model. **Water Resources Research, 50(1),** 601-618.

Modarres, R., and T. B. Ouarda (2013). Generalized autoregressive conditional heteroscedasticity modelling of hydrologic time series. **Hydrological Processes, 27(22),** 3174-3191.

Modarres, R., and T. B. Ouarda (2013). Testing and modelling the volatility change in ENSO. *Atmosphere-Ocean*, **51(5),** 561-570.

Modarres, R., and T. B. Ouarda (2014). Modeling the relationship between climate oscillations and drought by a multivariate GARCH model. *Water Resources Research*, **50(1),** 601-618.

Murray, C., T. Vos, R. Lozano, M. A. AlMazroa and Z. A. Memish (2014). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010 (vol 380, pg 2197, 2012).

Musengimana, G., F. K. Mukinda, R. Machekano and H. Mahomed (2016). Temperature variability and occurrence of diarrhoea in children under five-years-old in Cape Town metropolitan sub-districts. *International journal of environmental research and public health*, **13(9),** 859.

Myers, M. F., D. Rogers, J. Cox, A. A. Flahault and S. Hay (2000). Forecasting disease risk for increased epidemic preparedness in public health. *Advances in Parasitology*, **47**, 309.

Nason, G. P., and B. W. Silverman (1995). The stationary wavelet transform and some statistical applications Wavelets and statistics (pp. 281-299): Springer.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica: *Journal of the Econometric Society*, 347-370.

Neyman, J., and E. S. Pearson (1933). The testing of statistical hypotheses in relation to probabilities a priori. *Paper presented at the Mathematical Proceedings of the Cambridge Philosophical Society*.

Nur'utami, M. N., and R. Hidayat (2016). Influences of IOD and ENSO to Indonesian rainfall variability: role of atmosphere-ocean interaction in the Indo-Pacific sector. *Procedia Environmental Sciences*, **33,** 196-203.

Oliveira, C. S. d., Y. B. Gabbay, R. Ishak and A. D. C. Linhares (1999). Gastroenterites virais.

Onozuka, D. (2014). Effect of non-stationary climate on infectious gastroenteritis transmission in Japan. *Scientific reports*, **4,** 5157.

Onozuka, D., and M. Hashizume (2011). Weather variability and paediatric infectious gastroenteritis. *Epidemiology & Infection*, **139(9),** 1369-1378.

Onozuka, D., M. Hashizume and A. Hagihara (2010). Effects of weather variability on infectious gastroenteritis. *Epidemiology & Infection*, **138(2),** 236-243.

Organization, W. H. (2003). Climate change and human health: risks and responses: summary.

orku, M., and S. Sahile (2018). *Journal of Climatology & Weather Forecasting*.

Ouedraogo, N., S. M. T. Ngangas, I. J. O. Bonkoungou, A. B. Tiendrebeogo, K. A. Traore, I. Sanou and N. Barro(2017). Temporal distribution of gastroenteritis viruses in Ouagadougou, Burkina Faso: seasonality of rotavirus. *BMC Public Health*, **17(1),** 274.

Pabón, J. D., and R. Martínez (2016). Preface: Third ENSO Conference: findings and key messages. *Advances in Geosciences*, **42**, 91.

Patz, J., A. Githeko, J. McCarty, S. Hussein, U. Confalonieri and N. De Wet (2003). Climate change and infectious diseases. *Climate change and human health: Risks and responses*, **6**, 103-137.

Perrelli, R. (2001). Introduction to ARCH & GARCH models. University of Illinois Optional TA Handout, 1-7.

Rahman, M. J., and M. A. M. Hasan (2014). Performance of wavelet transform on models in forecasting climatic variables *Computational Intelligence Techniques in Earth and Environmental Sciences* (pp. 141-154): Springer.

Rashid H (1991) Geography of Bangladesh, 2nd edn. The University Press Limited, Dhaka.

Rashid KBS (2008) Bangladesh: resource and environmental profile. A H Development Publishing House, Dhaka

Reiner, R. C., A. A. King, M. Emch, M. Yunus, A. Faruque and M Pascual (2012). Highly localized sensitivity to climate forcing drives endemic cholera in a megacity. *Proceedings of the National Academy of Sciences*, **109(6),** 2033-2036.

Reis, A. R., and A. A. Da Silva (2005). Feature extraction via multiresolution analysis for short-term load forecasting. *IEEE Transactions on power systems*, **20(1),** 189-198.

Rizvi SNH (1969) East Pakistan district gazetteers: Dacca. East Pakistan Government Press, Dacca

Rodó, X., M. Pascual, G. Fuchs, and A. Faruque (2002). ENSO and cholera: a nonstationary link related to climate change? *Proceedings of the National Academy of Sciences*, **99(20),** 12901-12906.

Rogers, D. P., M. A. Shapiro, G. Brunet, J. C. Cohen, S.J. Connor, A. A. Diallo and D. Hemming (2010). Health and climate–opportunities. *Procedia Environmental Sciences*, **1**, 37-54.

Romilly, P. (2005). Time series modelling of global mean temperature for managerial decision-making. *Journal of environmental management*, **76(1),** 61-70.

Room, M., S. T. Forecasts and C. C. W.Portal (2000). Multivariate enSO index (Mei).

Rouf MA and S. Jahan (2007) Spatial and temporal patterns of urbanization in Bangladesh. In: Jahan S,

Saji, N., B. Goswami, P. Vinayachandran and T. Yamagata (1999). A dipole mode in the tropical Indian Ocean. *Nature*, **401(6751),** 360.

Sang, Y. F. (2013). A review on the applications of wavelet transform in hydrology time series analysis. *Atmospheric research*, **122**, 8-15.

Sanusi, R. A., F. Adebola and N. Adegoke (2016). Cases of Road Traffic Accident in Nigeria: A Time Series Approach. *Mediterranean journal of social sciences*, **7(2 S1**), 542.

Sarker, A. R., M. Sultana, R. A. Mahumud, N. Ali, T. M. Huda, S. Haider and R. Van Der Meer (2018). Economic costs of hospitalized diarrheal disease in Bangladesh: a societal perspective. *Global health research and policy*, **3(1),** 1.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, **6(2),** 461-464.

Shabbar, A., and M. Khandekar (1996). The impact of el Nino Southern oscillation on the temperature field over Canada: Research note. *Atmosphere-Ocean*, **34(2)**, 401-416.

Shahid S (2010) Recent trends in the climate of Bangladesh. *Clim Res* **42(3):**185–193

Shahid, S.(2010). Probable impacts of climate change on public health in Bangladesh. *Asia Pacific Journal of Public Health*, **22(3),** 310-319.

Shahid S (2012) Vulnerability of the power sector of Bangladesh to climate change and extreme weather events. *Reg Environ Change* **12(3**):595–606

Shaman, J., and M. Lipsitch (2013). The El Niño–Southern Oscillation (ENSO)–pandemic Influenza connection: Coincident or causal? *Proceedings of the National Academy of Sciences*, **110(Supplement 1),** 3689-3691.

Siddiqui K, Ahmed J, Siddiqui K, Huq S, Hossain A, Nazimud-Doula S, Rezawana N (2010) Social formation in Dhaka, 1985–2005: a longitudinal study of society in a third world megacity. Ashgate Publishing Group, Surrey

Silva, I. (2012). Analysis of discrete-valued time series.

Silva, I. (2012). Analysis of discrete-valued time series.

Singh, R. B., Hales, S., De Wet, N., Raj, R., Hearnden, M., & Weinstein, P. (2001). The influence of climate variation and change on diarrheal disease in the Pacific Islands. Environmental health perspectives, 109(2), 155.

Society, R., & Follett, S. B. (2002). Infectious diseases in livestock: summary and main recommendations: Royal Society.

Soltani, S. (2002). On the use of the wavelet decomposition for time series prediction. Neurocomputing, 48(1), 267-277.

Soyiri, I. N., & Reidpath, D. D. (2013). An overview of health forecasting. Environmental health and preventive medicine, 18(1), 1.

Stenseth, N. C., Ottersen, G., Hurrell, J. W., Mysterud, A., Lima, M., Chan, K. S., . . . Ådlandsvik, B. (2003). Studying climate effects on ecology through the use of climate indices: the North Atlantic Oscillation, El Nino Southern Oscillation and beyond. Proceedings of the Royal Society of London B: Biological Sciences, 270(1529), 2087-2096.

Stoll, B. J., Glass, R. I., Huq, M. I., Khan, M., Holt, J. E., & Banu, H. (1982). Surveillance of patients attending a diarrhoeal disease hospital in Bangladesh. Br Med J (Clin Res Ed), 285(6349), 1185-1188.

Szolgayová, E., Arlt, J., Blöschl, G., & Szolgay, J. (2014). Wavelet based deseasonalization for modelling and forecasting of daily discharge series considering long range dependence. Journal of Hydrology and Hydromechanics, 62(1), 24-32.

Taylor, J. W., & Buizza, R. (2004). A comparison of temperature density forecasts from GARCH and atmospheric models. Journal of Forecasting, 23(5), 337-355.

Teshima, A., Yamada, M., Hayashi, T., Wagatsuma, Y., & Terao, T. (2004). Climate impact on seasonal patterns of diarrhea diseases in Tropical area. University of Kyoto, Japan.

Thai, K. T., Cazelles, B., Van Nguyen, N., Vo, L. T., Boni, M. F., Farrar, J., . . . de Vries, P. J. (2010). Dengue dynamics in Binh Thuan province, southern Vietnam: periodicity, synchronicity and climate variability. PLoS neglected tropical diseases, 4(7), e747.

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

Torrence, C., & Webster, P. J. (1999). Interdecadal changes in the ENSO–monsoon system. Journal of Climate, 12(8), 2679-2690.

Troeger, C., Forouzanfar, M., Rao, P. C., Khalil, I., Brown, A., Reiner Jr, R. C., . . . Ahmed, M. (2017). Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet Infectious Diseases, 17(9), 909-948.

United Nations (UN) (2012) World urbanization prospects – the 2011 revision. UN, New York

Venegas-Pérez, M. E., Ramírez-López, E. M., López-Santos, A., Magaña-Rueda, V. O., & Avelar-González, F. J. (2016). The impact of phenomena El Niño and La Niña and other environmental factors on episodes of acute diarrhoea disease in the population of Aguascalientes, Mexico: a case study. Advances in Geosciences, 42, 15-21.

Walker, C. L. F., Perin, J., Aryee, M. J., Boschi-Pinto, C., & Black, R. E. (2012). Diarrhea incidence in low-and middle-income countries in 1990 and 2010: a systematic review. BMC Public Health, 12(1), 220.

Wang, H., Naghavi, M., Allen, C., Barber, R. M., Bhutta, Z. A., Carter, A., . . . Coates, M. M. (2016). Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet, 388(10053), 1459-1544.

Wang, W., Van Gelder, P., Vrijling, J., & Ma, J. (2005). Testing and modelling autoregressive conditional heteroskedasticity of streamflow processes. Nonlinear processes in Geophysics, 12(1), 55-66.

Wargon, M., Guidet, B., Hoang, T., & Hejblum, G. (2009). A systematic review of models for forecasting the number of emergency department visits. Emergency Medicine Journal, 26(6), 395-399.

Webster, P. J., Moore, A. M., Loschnigg, J. P., & Leben, R. R. (1999). Coupled ocean–atmosphere dynamics in the Indian Ocean during 1997–98. Nature, 401(6751), 356.

Weiss, C. H. (2018). An Introduction to Discrete-valued Time Series: John Wiley & Sons.

Wolter, K. (1993). Monitoring ENSO in COADS with a seasonally adjusted principal component index. Paper presented at the Proc. of the 17th Climate Diagnostics Workshop, 1993.

Wolter, K., & Timlin, M. S. (2011). El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI. ext). International Journal of Climatology, 31(7), 1074-1087. Woodland, L., Burgner, D., Paxton, G., & Zwi, K. (2010). Health service delivery for newly arrived refugee children: a framework for good practice. Journal of paediatrics and child health, 46(10), 560-567.

World Bank (2012) http://data.worldbank.org/country/bangladesh#cp_cc. Accessed 17 Dec 2012

World Bank (2012) http://data.worldbank.org/country/bangladesh#cp_cc. Accessed 17 Dec 2012

Xu, H.-Y., Xie, M., Goh, T. N., & Fu, X. (2012). A model for integer-valued time series with conditional overdispersion. Computational Statistics & Data Analysis, 56(12), 4229-4242.

Yamagata, T., Behera, S. K., Luo, J. J., Masson, S., Jury, M. R., & Rao, S. A. (2004). Coupled ocean atmosphere variability in the tropical Indian Ocean. Earth's Climate, 189-211.

Yamagata, T., Behera, S. K., Rao, S. A., Guan, Z., Ashok, K., & Saji, H. N. (2002). The Indian Ocean dipole: A physical entity. CLIVAR exchanges, 24(7), 2.

Yao, S., Song, Y., Zhang, L., & Cheng, X. (2000). Wavelet transform and neural networks for short-term electrical load forecasting. Energy Conversion and Management, 41(18), 1975-1988.

Ye, L., Yang, G., Van Ranst, E., & Tang, H. (2013). Time-series modeling and prediction of global monthly absolute temperature for environmental decision making. Advances in Atmospheric Sciences, 30(2), 382.

Zeger, S. L., Liang, K.-Y., & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. Biometrics, 1049-1060.

Zhou, B., He, D., & Sun, Z. (2006). Traffic predictability based on ARIMA/GARCH model. Paper presented at the Next Generation Internet Design and Engineering, 2006. NGI'06. 2006 2nd Conference on.

Zhou, G., Minakawa, N., Githeko, A. K., & Yan, G. (2004). Association between climate variability and malaria epidemics in the East African highlands. Proceedings of the National Academy of Sciences, 101(8), 2375-2380.

Zhu, F. (2011). A negative binomial integer valued GARCH model. Journal of Time Series Analysis, 32(1), 54-67.

Zhu, F. (2012). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. Journal of Mathematical Analysis and Applications, 389(1), 58-71.

Zhu, F. (2012). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. Journal of Mathematical Analysis and Applications, 389(1), 58-71.

Appendices

Appendix

R Code

```
#*********************************************************************#
#Load Packages
#-------------------------------------------------------------------------------------------------
library(forecast);library(tseries);library(reshape2);library(TSA);library(ggplot2);
library(TTR); library(scatterplot);library(quantreg) library(fGarch);library(e1071);
library(fracdiff);library(FinTS);library(foreign);library(TSPred);library(rugarch);
library(rmgarch); ibrary(neuralnet);library(datasets);library(wmtsa);library(wavelets);
library(waveslim);library(timsac);library(expsmooth);library(astsa);library(vars);
library(fpp);library(caret);library(dtw);library(fArma); ibrary(splus2R);ibrary(xtable);
library(bayesGARCH);library(betategarch);library(dplyr);library(tidyr);
library(tscount);library(showtext);library(quantmod);library(wavethresh)
library(thief); library(nnfor); library(tibble); library(lazyeval);library(curl);
library(rlang); library(WaveletArima)
packages<-c("forecast","tseries","reshape2","TSA","ggplot2", "TSA", "scatterplot3d",
"quantreg","fGarch","e1071", "fracdiff", "FinTS", "foreign", "TSPred", "rugarch",
"rmgarch","httpuv","neuralnet","datasets","wmtsa", "wavelets", "waveslim", "timsac",
"expsmooth","qualV","GEVStableGarch","astsa","vars","fpp", "caret", "purrr", "dtw",
"fArma","splus2R","xtable", "nnfor", "bayesGARCH", "betategarch", "dplyr", "tidyr",
```

```
"showtext", "quantmod", "wavethresh", "glue","thief", "nnfor", "tibble", "lazyeval",
"curl", "rlang", "WaveletArima", "tscount", “acp”)

install.packages(packages, dep=T)

################################################

getwd()

setwd("E:")

xx=read.csv("data.csv", header=T)

head(xx)

x = ts(xx, start=c(1950, 1), frequency=12)

ts.plot(x, col=4)

abline(h=mean(x), col=2, lwd=2)

abline(h=mean(x)-2, col=3, lwd=2, lty=2)

abline(h=mean(x)+2, col=3, lwd=2, lty=2)

seasonplot(x, col=rainbow(length(x)/12), year.labels=T)

x1 = ts(abs(xx), start=c(1950, 1), end=c(1970,12), frequency=12)

x = ts(xx, start=c(1950, 1), frequency=12); x

tsp(xx)

boxplot(xx)

boxplot(y ~ cycle(x), xlab="Months", ylab= "IOD", col=5)

summary(x)

ts.plot(x, col=4, xlab="Year", ylab="IDD", main="Infected Diarrheal Patients:
        Actual")

seasonplot(x, col=rainbow(length(x)/12+1), year.labels=T)

seasonplot(x^2, col=rainbow(length(x)/12+1), year.labels=T)

acf(x, lag.max = 100)

pacf(x, lag.max = 100)
```

```
ndiffs(x)
dx=diff(x, d=1)
ndiffs(dx)
par(mfrow=c(3,1))
ts.plot(x, main="Actual")
ts.plot(x^2, main="Squared: Actual")
ts.plot(dx, main="1st Diff")
par(mfrow=c(1,1))
boxplot(x)
boxplot(y ~ cycle(y), xlab="Months", ylab= "IOD", col=5);summary(y)
ts.plot(z,col=4,xlab="Year", ylab="IDD", main="Infected Diarrheal Patients: Actual")
```

############### **Function for Wavelet decomposition**###################

```
main<- function(t, trans=0, omega=6, j=0){
        dial<-2*2^(j*.125)
        sqrt((1/dial))*pi^(-1/4)*exp(1i*omega*((t-trans)/dial))*exp(-((t-
        trans)/dial)^2/2) }
main
 (J<-length)
(wt<- matrix(rep(NA, (length(meanDev))*(J+1)), ncol=(J+1)))
for(jj in 0:J){
        for(kk in 1:length(meanDev)){
                wt[kk, jj+1]<- main(t=1:(length(meanDev)), j=jj,
        trans=kk)%*%meanDev
        }
}
```

```
################## Wavelet smoothing: S8######################
 (MEI_W = wavMODWT(MEI, wavelet = "s8", n.levels = 10))
(IOD_W = wavMODWT(IOD, wavelet = "s8", n.levels = 10))
(IDD_W = wavMODWT(IDD, wavelet = "s8", n.levels = 10))
 ########Extraction of the real part for the reconstruction for wavelet######
################## Reconstruct as in formula:#######################
(dial<- 2*2^(0:J*.125))
(xrec<- rep(NA,(length(meanDev))))
for(ll in 1:(length(meanDev))){
        xrec[ll]<- 0.2144548*sum(re.wt[ll,]/sqrt(dial))
}
(xrec<-xrec+xMean)
######################## Save the data as De-noised######################
write.csv(xrec, E:/Sumon_Thesis/PhD_Thesis_Paper/DataResult/data_Denoised.csv")
################## Plot for Actual and de-noised data######################
tiff("E:/Sumon_Thesis/PhD_Thesis_Paper/DataResult/FigureAll/Fig01.tiff",
        width=3200, height=3200, units="px", res=400)
plot(xx, type="l", col="green"); lines(xrec, col="red")
legend("topright", lty=1, col=c("green", "red"), c("Actual Data", "Wavelet
        Transformed"))
dev.off()
tiff("E:/Sumon_Thesis/PhD_Thesis_Paper/DataResult/FigureAll/Fig02.tiff",
        width=3200, height=3200, units="px", res=300)
ts.plot(xA, col="green3", xlab="Year", ylab="Infected Patients", main="Infected
        Diarrheal Patients: Actual & Denoised")
lines(xD, col="red", lty=2)
```

```
legend("topright", lty=c(1,2), col=c("green3", "red"), c("Actual", "Denoised"))
dev.off()
plot = function(X, title="", lag.max=NULL) {
        layout(matrix(1:3, 3, 1))
        plot.ts(X, main=title)
        acf(X, main=title, lag.max=lag.max)
        pacf(X, main=title, lag.max=lag.max)        }
tiff("E:/Sumon_Thesis/PhD_Thesis_Paper/DataResult/FigureAll/Fig03.tiff",
        width=3200, height=3200, units="px", res=400)
makplot(xA, title="Infected Diarrheal Patients: Actual")
dev.off()
#
tiff("E:/Sumon_Thesis/PhD_Thesis_Paper/DataResult/FigureAll/Fig04.tiff",
        width=3200, height=3200, units="px", res=400)
makplot(xD, title="Infected Diarrheal Patients: Denoised")
dev.off()
###
adf.test(xA, alternative = "stationary")
adf.test(xD, alternative = "stationary")

dxA = diff(xA, differences = 1)
dxD = diff(xD, differences = 1)
plot(dxA, col=4)
plot(dxD, col=4)
#
###
```

```
tiff("E:/Sumon_Thesis/PhD_Thesis_Paper/DataResult/FigureAll/Fig05.tiff",
        width=3200, height=3200, units="px", res=400)
makplot(dxA, title="Infected Diarrheal Patients: Stationarity of Actual")
dev.off()
#
tiff("E:/Sumon_Thesis/PhD_Thesis_Paper/DataResult/FigureAll/Fig06.tiff",
        width=3200, height=3200, units="px", res=400)
makplot(dxD, title="Infected Diarrheal Patients: Stationarity of Denoised")
dev.off()
###
tiff("E:/Sumon_Thesis/PhD_Thesis_Paper/DataResult/FigureAll/Fig07.tiff",
        width=3200, height=3200, units="px", res=200)
par(mfrow=c(2, 1))
seasonplot(xA, col=rainbow(length(xx)/12), ylab="Infected Diarrheal Patients",
        xlab="Year", main="Seasonal plot: Actual", year.labels=T)
seasonplot(xD, col=rainbow(length(xx)/12), ylab="Infected Diarrheal Patients",
        xlab="Year", main="Seasonal plot: Denoised", year.labels=T)
par(mfrow=c(1, 1))
dev.off()
#
tiff("E:/Sumon_Thesis/PhD_Thesis_Paper/DataResult/FigureAll/Fig07b.tiff",
        width=3200, height=3200, units="px", res=350)
seasonplot(xA, col=rainbow(length(xx)/12), ylab="Infected Diarrheal Patients",
        xlab="Year", main="Seasonal plot: Actual", year.labels=T)
dev.off()
# Monthly TS Plot
```

```
tiff("E:/Sumon_Thesis/PhD_Thesis_Paper/DataResult/FigureAll/Fig09.tiff",
        width=3200, height=3200, units="px", res=350)
boxplot(xx ~ cycle(xx), xlab="Months", ylab= "Infected Patients", col=5)
dev.off()
 (annualx <- aggregate(xx)/12)
(annualxsd <- aggregate(xx, FUN = sd))
#
tiff("E:/Sumon_Thesis/PhD_Thesis_Paper/DataResult/FigureAll/Fig10.tiff",
        width=3200, height=3200, units="px", res=400)
op=par(mfrow=c(3,1))
plot(xx, xlab="Year", ylab="Number", main="Infected Patients: Actual", col="blue")
plot(annualx, xlab="Year", ylab="Infected Patients", , main="Infected Patients:
        Yearly Average", col="blue")
plot(annualxsd, xlab="Year", ylab="Infected Patients", , main="Infected Patients:
        SD",col="blue")
par(op);dev.off()
#
(JanX <- window(xx, start = c(1993, 1), freq = TRUE))
(FebX <- window(xx, start = c(1993, 2), freq = TRUE))
(MarX <- window(xx, start = c(1993, 3), freq = TRUE))
(AprX <- window(xx, start = c(1993, 4), freq = TRUE))
(MayX <- window(xx, start = c(1993, 5), freq = TRUE))
(JunX <- window(xx, start = c(1993, 6), freq = TRUE))
(JulX <- window(xx, start = c(1993, 7), freq = TRUE))
(AugX <- window(xx, start = c(1993, 8), freq = TRUE))
(SepX <- window(xx, start = c(1993, 9), freq = TRUE))
```

```
(OctX <- window(xx, start = c(1993, 10), freq = TRUE))
(NovX <- window(xx, start = c(1993, 11), freq = TRUE))
(DecX <- window(xx, start = c(1993, 12), freq = TRUE))
tiff("E:/Sumon_Thesis/PhD_Thesis_Paper/DataResult/FigureAll/Fig11.tiff",
     width=3200, height=3200, units="px", res=300)
op=par(mfrow=c(4,3))
plot(JanX, ylab="Infected Patients", col=4, main="Month: Jan")
plot(FebX, ylab="Infected Patients", col=4, main="Month: Feb")
plot(MarX, ylab="Infected Patients", col=4, main="Month: Mar")
plot(AprX, ylab="Infected Patients", col=4, main="Month: Apr")
plot(MayX, ylab="Infected Patients", col=4, main="Month: May")
plot(JunX, ylab="Infected Patients", col=4, main="Month: Jun")
plot(JulX, ylab="Infected Patients", col=4, main="Month: Jul")
plot(AugX, ylab="Infected Patients", col=4, main="Month: Aug")
plot(SepX, ylab="Infected Patients", col=4, main="Month: Sep")
plot(OctX, ylab="Infected Patients", col=4, main="Month: Oct")
plot(NovX, ylab="Infected Patients", col=4, main="Month: Nov")
plot(DecX, ylab="Infected Patients", col=4, main="Month: Dec")
par(op);dev.off()
####################Fitting of Arima############################
a=auto.arima(x); a
summary(a)
m<-arima(x, order=c(p,d,q), season=list(order=c(P,D,Q),freq=12)) #p,q=(1 to 3),d=1
f<-fitted(m)
fc<-forecast(m, h=6)
```

```
plot(fc); lines(f, col=2)

adf.test(xA, alternative = "stationary")

adf.test(xD, alternative = "stationary")

dxA = diff(xA, differences = 1)

dxD = diff(xD, differences = 1)

plot(dxA, col=4);plot(dxD, col=4)

seasonplot(xA, col=rainbow(length(xx)/12), ylab="Infected Diarrheal Patients",
    xlab="Year", main="Seasonal plot: Actual", year.labels=T)

seasonplot(xD, col=rainbow(length(xx)/12), ylab="Infected Diarrheal Patients",
    xlab="Year", main="Seasonal plot: Denoised", year.labels=T)

par(mfrow=c(1, 1));dev.off()

seasonplot(xA, col=rainbow(length(xx)/12), ylab="Infected Diarrheal Patients",
    xlab="Year", main="Seasonal plot: Actual", year.labels=T)

dev.off()

plot(xx, xlab="Year", ylab="Number", main="Infected Patients: Actual", col="blue")

plot(annualx, xlab="Year", ylab="Infected Patients", , main="Infected Patients:
    Yearly Average", col="blue")

plot(annualxsd, xlab="Year", ylab="Infected Patients", , main="Infected Patients:
    SD",col="blue")

par(op);dev.off()

#

(JanX <- window(y, start = c(1870, 1), freq = TRUE))

(FebX <- window(y, start = c(1870, 2), freq = TRUE))

(MarX <- window(y, start = c(1870, 3), freq = TRUE))

(AprX <- window(y, start = c(1870, 4), freq = TRUE))

(MayX <- window(y, start = c(1870, 5), freq = TRUE))
```

```
(JunX <- window(y, start = c(1870, 6), freq = TRUE))
(JulX <- window(y, start = c(1870, 7), freq = TRUE))
(AugX <- window(y, start = c(1870, 8), freq = TRUE))
(SepX <- window(y, start = c(1870, 9), freq = TRUE))
(OctX <- window(y, start = c(1870, 10), freq = TRUE))
(NovX <- window(y, start = c(1870, 11), freq = TRUE))
(DecX <- window(y, start = c(1870, 12), freq = TRUE))
#
std <- function(x) sd(x)/sqrt(length(x))
std(JanX);std(FebX);std(MarX);std(AprX);std(MayX);std(JunX);std(JulX);std(AugX;
std(SepX);std(OctX);std(NovX);std(DecX)
tiff("E:/Today/Sumon_Thesis/Fig31.tiff", width=3200, height=3200, units="px",
     res=300)
op=par(mfrow=c(4,3))
plot(JanX, ylab="IOD", col=4, main="Month: Jan")
plot(FebX, ylab="IOD", col=4, main="Month: Feb")
plot(MarX, ylab="IOD", col=4, main="Month: Mar")
plot(AprX, ylab="IOD", col=4, main="Month: Apr")
plot(MayX, ylab="IOD", col=4, main="Month: May")
plot(JunX, ylab="IOD", col=4, main="Month: Jun")
plot(JulX, ylab="IOD", col=4, main="Month: Jul")
plot(AugX, ylab="IOD", col=4, main="Month: Aug")
plot(SepX, ylab="IOD", col=4, main="Month: Sep")
plot(OctX, ylab="IOD", col=4, main="Month: Oct")
plot(NovX, ylab="IOD", col=4, main="Month: Nov")
```

```
plot(DecX, ylab="IOD", col=4, main="Month: Dec")
par(op)
dev.off()
#################################################################
Box.test(z^2, type="Ljung-Box", lag = 12)    ## use Box.test from stats package
ArchTest(z)              #Use ArchTest() function from FinTS package Engle's LM test
 ########################Fitting of GARCH and W-GARCH##############
(gg1<-garchFit(~garch(1,1), data=xx, include.mean=F,
          cond.dist= "norm", trace=FALSE))
(gg2<-garchFit(~arma(1,2)+garch(1,1), data=xx, include.mean=F,
          cond.dist= "norm", trace=FALSE))
summary(gg1)
summary(gg2)
plot(xx[10:50], type="l")
which(xx>10)
plot(gg2@data, type="l", col=4, main="**********")
plot(gg2@h.t, type="l", col=4, main="Estimated Volatility")
stdRes<-gg2@residuals/sqrt(gg2@h.t)
hist(stdRes, breaks=20, prob=T,main="Histogram of standardized residuals \n from
          ARIMA-GARCH for *****")
# lets add a normal density curve for sample mean and variance
curve(dnorm(xx, mean=mean(stdRes), sd=sd(stdRes)), col=2, lwd=2, add=TRUE)
# Jarque-Bera test
jarque.bera.test(stdRes)
ArchTest(stdRes, lags=5)
```

```
durbinWatsonTest(lm(stdRes~1), max.lag=5)
spec<- ugarchspec(variance.model=list(model="eGARCH",garchOrder=c(1,1)),
mean.model=list(armaOrder=c(1,2),include.mean=F),distribution.model="norm")
eg<- ugarchfit(spec=spec, data=xx); eg
plot(eg, which=2)
plot(eg, which="all")
(gFit<- fGarch::garchFit(~ garch(1,1), data = xx, trace = FALSE))
predict(gFit, n.ahead = 1, doplot=F)[3]
predict(gFit, n.ahead = 10, doplot=T)
predict(gFit, plot=TRUE)
(armaGarch= garchFit(formula=~arma(2,4)+garch(1,1), data=xx, trace=FALSE))
plot(armaGarch, which=2)
#################
garchFit(x = "dem2gbp", cond.dist = "sged",
        shape = 1, include.shape = FALSE, algorithm = "nlminb+nm")
garchFit(formula.mean = ~arma(0,0), formula.var = ~garch(1,1),  init.rec = c("mci",
"uev"), algorithms = c("sqp", "nlminb", "bfgs"))
(gg<- garchFit(formula=~arma(0,0)+garch(1,1), data=xx, algorithms = "sqp"))
(gg<- garchFit(formula=~arma(1,2)+garch(1,1), data=xx, algorithms = "sqp"))
(gg<- garchFit(formula=~arma(1,2)+garch(1,1), data=xx, algorithms = "sqp"))
(gg<-  garchFit(formula=~arma(0,1)+garch(1,1),  data=xx,  init.rec="mci",
        algorithms="nlminb"))
gg@fit$coef
g=garchFit(xx~arma(0,1)+garch(1,2), cond.dist = "std"); g
garchFit(xx~aparch(1,1), delta = 2, include.delta = FALSE)
g@fit$ics
```

```
g@fit$cvar
g@fit$hessian
g@fit$params
g@fit$series$order
g@fit$series$init.rec
g@fit$params$params
g@fit$params$cond.dist
g@fit$params$control
ug_var<- gg@fit$var
ug_res2<- (g@fit$residuals)^2
plot(ug_res2, type = "l")
lines(ug_var, col = "green")
ugfore <- ugarchforecast(g, n.ahead = 10)
ug_f <- ugfore@forecast$sigmaFor
plot(ug_f, type = "l")



egSpec<- ugarchspec(variance.model=list(model="eGARCH",garchOrder=c(1,1)),
          mean.model=list(armaOrder=c(1,2)))
egFit<- ugarchfit(egSpec, xx)
egFit
coef(egFit)
FutureForecast<-ugarchforecast(egFit, n.ahead=20)
FutureForecast
library(fBasics);
```

```
jarqueberaTest(xx)

str(egFit)

volatility<-egFit@fit$fitted.values;

volatility<-volatility[,1];

par(mfrow=c(3,1))

plot.ts(xTr)

plot.ts(xx)

plot.ts(volatility)

par(mfrow=c(1,1))


adf.test(xx, alternative = "stationary")

acf(xx, main='')

pacf(xx, main='')

dx = diff(xx, differences = 1)

plot(dx)

adf.test(dx, alternative = "stationary")

acf(dx, main='')

pacf(dx, main='')

kpss<- ur.kpss(xx, type="mu", lags="short"))

(adf<- ur.df(xx, type="trend", lags=20, selectlags="BIC"))

summary(kpss)

summary(adf)

spec.pgram(xx)

(m<- arima(xx, order=c(2,1,1), seasonal=list(order=c(1,1,2), period=12)))

(mb<- Arima(xx, order=c(2,1,1), seasonal=list(order=c(1,1,2), period=12)))
```

```
tsdiag(m)
(mFcast=forecast(mb, 36, level=95))
plot(mFcast)
(res=residuals(m))
resQ=res^2
par(mfcol=c(3,1))
plot(resQ)
acfQ<-acf(resQ, lag.max=100, main="ACF Squared Residuals")
pacfQ<-pacf(resQ, lag.max=100, main="PACF Squared Residuals")
par(mfcol=c(1,1))
shapiro.test(res)
hist(res); lines(density(res), col=2)
qqnorm(res); qqline(res, col=2, lwd=2)
(g<- garchFit(xx~arma(1,1)+garch(1,1),data=xx, trace=F))
summary(g)
(gg<- garchFit(xx ~arma(2,1)+ aparch(1,1), data=xx, trace=F))
(tg<- garchFit(xx ~arma(2,1)+ garch(1,1), data=xx, delta=1, trace=F))
(gjr<- garchFit(xx ~arma(2,1)+ garch(1,1), data=xx, delta=2, leverage = TRUE,
    trace=F))
# cond.dist = c("norm", "snorm", "ged", "sged", "std", "sstd", "snig", "QMLE")
(g1<- garchFit(xx ~arma(2,1)+ garch(1,1), data=xx, cond.dist = "sstd", trace=F))
(g2<- garchFit(xx ~arma(2,1)+ garch(1,1), data=xx, cond.dist = "ged", trace=F))
plot(g2, which=3)
#
g3<- garchFit(formula=~arma(2,1)+garch(1,1), data=xx,
```

```
        cond.dist="norm", include.mean=T, trace=F)
summary(g3)
(arch<- garch(res, order=c(0,8), trace=F))
(loglik=logLik(arch))
summary(arch)
ht=arch$fit[,1]^2
plot(ht)
mFitted=fitted.values(m)
low=mFitted-1.96*sqrt(ht)
high<-mFitted+1.96*sqrt(ht)
plot(xx, type="l", col=3, main=" ")
lines(low, col=2, lty=2)
lines(high, col=2, lty=2)
legend("topleft", lty=c(1,2), col=c(3,2), c("Actual", "95% CI"))
#
archRes<-res/sqrt(ht)
qqnorm(archRes, main="ARIMA-ARCH residuals")
qqline(archRes, col=2, lwd=2)
#
garchoutput <- garch(res^2, order=c(1,1))
CIC<-AIC(garchoutput)
CIC
(CIC = AIC(garchoutput)/length(res^2))
#
h<-garchFit(~arma(1,1)+garch(2,2),data=xx,cond.dist="std",TRACE=F)
```

```
predict(h,10)


######
(gg1<- garchFit(formula=~arma(0,1)+garch(1,1), data=xx, include.mean=FALSE))
(gg2<- garchFit(formula=~arma(1,1)+garch(1,1), data=xx, include.mean=FALSE))
(gg3<- garchFit(formula=~arma(2,1)+garch(1,1), data=xx, include.mean=FALSE))
(gg4<- garchFit(formula=~arma(2,1,1)+garch(1,1), data=xx, include.mean=FALSE))
(gg5<- garchFit(formula=~arma(2,1,1)+garch(1,1), data=xx, include.mean=F,
        trace=T))
(gg6<- garchFit(formula=~arma(2,1,1)+garch(1,1), data=xx, include.mean=F,
        trace=F))
(gg7<- garchFit(formula=~arma(2,1,1)+garch(1,1), data=xx, include.mean=F,
        cond.dist="std", trace=T))
summary(gg5)
summary(gg7)
plot(gg5, which="all")



gjr<-garchFit(xx~arma(0,1)+aparch(1,1), data=xx, delta = 2, include.delta = FALSE)
tgarch<- garchFit(xx~arma(0,1)+aparch(1,1), data=xx, delta = 1, include.delta =
        FALSE, leverage = FALSE)
gg<-garchFit(xx~arma(1,1)+garch(1,2), data=xx, cond.dist = "sstd")
gg@fit$hessian
gg@fit$ics
gg@fit$params
gg@fit$coef
```

```
gg@fit$series$model
gg@fit$series$order
gg@fit$series$init.rec
gg@fit$params$U
gg@fit$params$V
garch1<-garchFit(~arma(0,1)+garch(1,1), data=xx, trace=FALSE,
    include.mean=TRUE)
predict(garch1, n.ahead=25)
spec<- ugarchspec(variance.model = list(garchOrder=c(1,1)),
    mean.model=list(armaOrder = c(2, 1)), distribution.model = "norm")
garch<- ugarchfit(spec=spec, data=xx, solver.control = list(trace=0))
garch
garch@fit$coef
garch@fit$sigma
garch@fit$z
str(garch)
#
(arch1<-garch(xx, order=c(0,1),trace=FALSE))
(arch2<-garch(xx, order=c(0,2),trace=FALSE))
(arch10<-garch(xx, order=c(0,10),trace=FALSE))
(garch11<-garch(xx, order=c(1,1),trace=FALSE))
options(show.signif.stars=FALSE)
summary(arch11)
###
(vol.estmat<-cbind(
```

```
        arch1$fitted.values[,1],
        arch2$fitted.values[,1],
        arch10$fitted.values[,1],
        garch11$fitted.values[,1]))
#
par(mfcol=c(2,1))
ts.plot(vol.estmat[,1], col=c(1,2,3,4), main="ARCH(1)",ylab="vol")
abline(h=sqrt(var(xx)),col=1, lwd=3)
ts.plot(vol.estmat[,2],col=c(2,2,3,4), main="ARCH(2)",ylab="vol")
abline(h=sqrt(var(xx)),col=1, lwd=3)
par(mfcol=c(1,1))
#
par(mfcol=c(2,1))
ts.plot(vol.estmat[,3], col=c(3,2,3,4), main="ARCH(10)",ylab="vol")
abline(h=sqrt(var(xx)),col=1, lwd=3)
ts.plot(vol.estmat[,4],col=c(4,2,3,4), main="GARCH(1,1)",ylab="vol")
abline(h=sqrt(var(xx)),col=1, lwd=3)
par(mfcol=c(1,1))
```

####################Fitting of the INGARCH and W-INGARCH############

```
data <- tsglm(ts=data, model=list(past_obs=1, past_mean=c(7,13)))
summary(data)
plot(data)
ddistr(x, meanvalue, distr=c("poisson", "nbinom"), distrcoefs)
pdistr(q, meanvalue, distr=c("poisson", "nbinom"), distrcoefs)
qdistr(p, meanvalue, distr=c("poisson", "nbinom"), distrcoefs)
```

```
rdistr(n, meanvalue, distr=c("poisson", "nbinom"), distrcoefs)

sddistr(meanvalue, distr=c("poisson", "nbinom"), distrcoefs)

ardistr(response, meanvalue, distr=c("poisson", "nbinom"), distrcoefs)

checkdistr(distr=c("poisson", "nbinom"), distrcoefs)

ingarch.mean(intercept, past_obs=NULL, past_mean=NULL)

ingarch.var(intercept, past_obs=NULL, past_mean=NULL)

ingarch.acf(intercept, past_obs=NULL, past_mean=NULL, lag.max=10,type=c("acf",
        "pacf", "acvf"), plot=TRUE)

datafit <- tsglm(ts=data, model=list(past_obs=1, past_mean=c(7,13)))

data_intervdetect <- interv_detect(fit=datafit, taus=80:120, delta=1)

plot(datafit_intervdetect)

marcal(datafit)

pit(datafit)

predict(datafit, n.ahead=1) #prediction interval using conditional distribution

predict(datafit,n.ahead=1,newobs=NULL,newxreg=NULL, evel=0.95,global=FALSE,
type=c("quantiles", "shortest", "onesided"), method=c("conddistr", "bootstrap"),
B=1000, estim=c("ignore", "bootstrap", "normapprox", "given"), B_estim=B)

QIC(datafit)

AIC(datafit)

datafit_resid <- residuals(datafit, type="pearson")

plot(datafit_resid)

acf(datafit_resid)

scoring(datafit)

datafit_pois <- tsglm(data, model = list(past_obs = 1, past_mean = 13), xreg =
            interventions, distr = "poisson")

datafit_nbin <- tsglm(campy, model = list(past_obs = 1, past_mean = 13), xreg =
```

```
                interventions, distr = "nbinom")
acf(residuals(campyfit_pois), main = "ACF of response residuals")
marcal(campyfit_pois, main = "Marginal calibration")
lines(marcal(campyfit_nbin, plot = FALSE), lty = "dashed")
legend("bottomright",legend = c("Pois", "NegBin"), lwd =1,lty = c("solid", "dashed"))
pit(datafit_pois, ylim = c(0, 1.5), main = "PIT Poisson")
pit(datafit_nbin, ylim = c(0, 1.5), main = "PIT Negative Binomial")
rbind(Poisson = scoring(datafit_pois), NegBin = scoring(datafit_nbin))
Target_data <- IDD[, "VanKilled"]
regressors <- cbind(MEI = IDD[, c("MEI")],linearTrend = seq(along = timeseries)/12)
Target_data_Tr <- window(IDD, end = 2012 + 11/12)
MEI_Tr <- window(regressors, end = 2012 + 11/12)
Target_datafit <- tsglm(Target_data_Tr, model = list(past_obs = c(1, 12)), link =
                 "log", distr = "poisson", + xreg = MEI_Tr)
summary(IDD, B = 500)
IDD_Test <- window(IDD, start = 2013, end = 2017 + 11/12)
MEI_Tr <- window(MEI, start = 2012, end = 2017 + 11/12)
predict(IDD, n.ahead <- 12, level = 0.9, global = TRUE, B = 2000, newxreg =
                    IDD_Test)$pred
IDD_alldata <- tsglm(IDD, link = "log", model = list(past_obs = c(1, 12)), xreg =
                MEI, distr = "poisson")
IDD_alldata <- tsglm(IDD, link = "log", model = list(past_obs = c(1, 12)), xreg =
                MEI, distr = "nbin")
Accuracy(IDD_alldata)
##############################END#################################
```