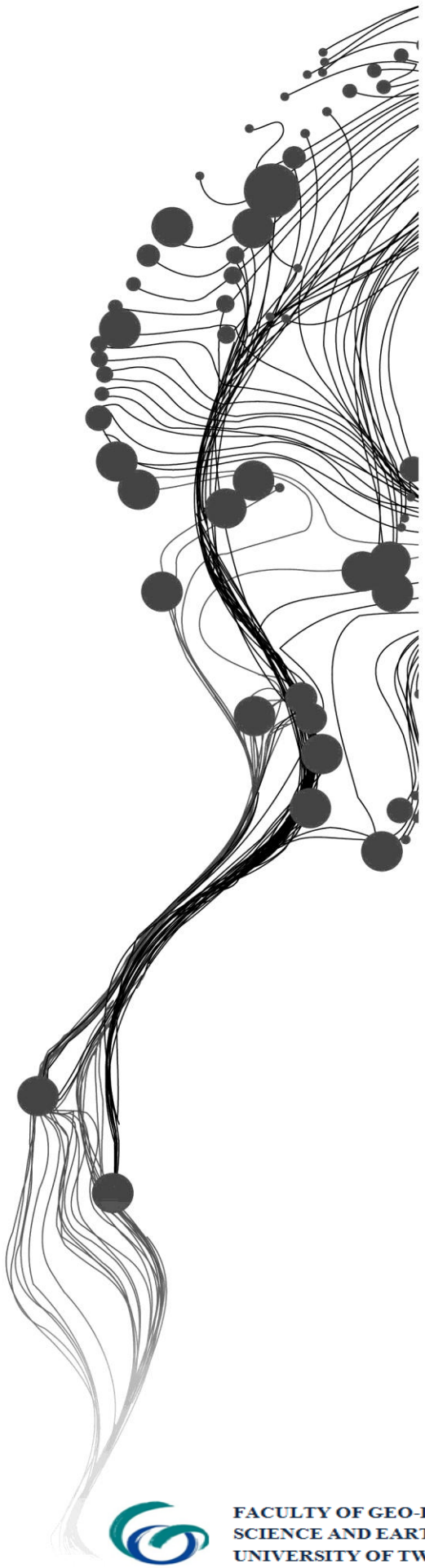
Studying the spatio-temporal trends in space based gravity observations and exploring the relation with precipitation and biophysical parameters over India

RAJ BHAGAT PALANICHAMY
March, 2014

ITC SUPERVISOR
Dr. Raul Zurita Milla

IIRS SUPERVISORS
Mr. Prasun K. Gupta
Dr. S.K. Srivastav



Studying the spatio-temporal trends in space based gravity observations and exploring the relation with precipitation and biophysical parameters over India

RAJ BHAGAT PALANICHAMY
Enschede, the Netherlands [March, 2014]

Thesis submitted to the Faculty of Geo-information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geoinformatics

THESIS ASSESSMENT BOARD:

Chairperson : Prof. dr. ir. Alfred Stein
External Examiner : Dr. S.K Bartarya, Senior (Wadia Inst of Himalayan Geology)
ITC Supervisor : Dr. Raul Zurita Milla
IIRS Supervisor : Mr. Prasun Kumar Gupta
IIRS Supervisor : Dr. S.K. Srivastav

OBSERVERS:

ITC Observer : Dr. Nicholas Hamm
IIRS Observer : Dr. S. K. Srivastav

FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION, UNIVERSITY OF TWENTE, ENSCHEDE, THE NETHERLANDS

INDIAN INSTITUTE OF REMOTE SENSING
Indian Space Research Organisation
Department of Space, Government of India

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, The Netherlands. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

Dedicated to my parents and friends....

ABSTRACT

Space based Gravity observations of GRACE mission vary based on the mass changes on the surface of Earth. The mass changes on the surface are caused due to spatio-temporal variations in vegetation, surface and subsurface water, urbanization, snow amount, and plate tectonics. The studying of the spatial and temporal patterns in these parameters is crucial, as it can help in understanding the surface mass changes and also the anthropogenic effects on gravitational measurements. Considering the space-time variation of these parameters, the choice of the spatial and temporal scale of these datasets is vital because the trends and patterns could change with different spatial and temporal scales.

This research intends to study the Modifiable Spatio-Temporal Unit Problem (MSTUP) which arises due to spatial and temporal aggregation of datasets. Trend detection at various spatial and temporal scales could help in the understanding of the effects. The present study was done to understand the trends in GRACE LWE Thickness, vegetation, precipitation, urbanization, and snow amount over India and also to find if various parameters such as vegetation, precipitation, earthquakes, urbanization, and snow amount affect the GRACE LWE Thickness anomalies. The datasets used for the study are GRACE level-3 LWE Thickness land data, MODIS Enhanced Vegetation Index, TRMM Precipitation Rate, AMSR-E Snow Water Equivalent Thickness. Using the GPWv3 Population Density dataset and the DMSP "Night lights" OLS data, a proxy for urbanization is created. An attempt to mathematically model the relationship between the various parameters and gravity measurements using an Artificial Neural Network (ANN) is done in the research. The model uses precipitation, vegetation, snow water equivalent thickness, urbanization factor as explanatory variables for explaining the variation in GRACE LWE Thickness.

The results from the trend analysis indicate that for the time period between 2003 and 2010, GRACE LWE Thickness has a negative trend over North India and a positive trend in some parts of the Deccan and Tibetan plateaus. The correlation studies indicate that the vegetation, precipitation correlate positively with GRACE LWE Thickness. A negative correlation between urbanization and GRACE LWE thickness was found indicating that ground water depletion is faster in urban regions than in rural regions. Earthquakes between 2003 and 2010 don't have any effect on GRACE level-3 land data over the Indian Subcontinent. An ANN model was created with a RMS Error of 2.89 cm of LWE Thickness. Spatial and temporal resolutions of the datasets are found to affect the trend analysis results and the correlation results.

Keywords: *GRACE Gravity LWE Thickness anomaly, Modifiable Spatio-Temporal Unit Problem,Artificial Neural Network, Precipitation, Vegetation, Urbanization Index.*

ACKNOWLEDGEMENTS

It is a great opportunity for me to get an admission in such a wonderful course of MSc, IIRS-ITC JEP. I would like to express my gratitude to all the people who has supported me during my entire MSc research work.

On the completion of my MSc thesis, I owe my deepest gratitude to my both IIRS supervisors Mr. Prasun Kumar Gupta and Dr. S.K. Srivastav for their continuous support, guidance, motivation and extraordinary scientific perception. Thank you sir, for your precious time, support throughout my academics at IIRS. They helped me whenever I had approached them, even during my hard time. I don't have enough words to express my gratitude for the faith you had in me.

I would like to thank Dr. Raul Zurita Milla, my ITC supervisor, for his valuable guidance and valuable suggestions for this research work. Dr. Raul had huge influence on me since the first day of ITC class. The first sentence that I heard from him was "Correlation doesn't imply causation" which did have huge impact in my research and my ideas.

I am grateful to Dr. Y.V.N. Krishna Murthy, the Director, IIRS and Dr. P.S. Roy, Former Director, IIRS for providing excellent research environment and infrastructure to carry out this research work. I would like to show my extreme gratitude to Mr. P.L.N. Raju, Group Head, RSGG, IIRS for his constant support and providing critical inputs for making this MSc program an invaluable experience. I am also thankful to Dr. Nicholas Hamm, ITC for being there for us at all times and for providing valuable advices and especially for making our stay at ITC an easy and memorable one. I would like to show my gratitude to Dr. David Rosseiter for his advices and innovative inputs. The Data Analysis module he taught me cannot be forgotten as I helped me understand logics, possibilities, and improbabilities.

I am also honoured by receiving lectures from Dr. Alexei Voinov and Dr. Alfred Stein, Dr. Kraak ITC in the field of modelling, statistics and Visualization.

I would like to thank my family friends who had been with me during the toughest times. Words cannot express my affection for you guys. I would thank all my friends Ponraj , Abhishek, Sahiti, Girija, Mustafa, Krithiga, Surya, Guru,Gunamani, Aravind, Ashutosh, Ram, Raja, Aniket, Sanjay, Sanjeev,K.D,Neeraj, Yeshu, Joyson, Abishek.A, Saif, Meenakshi, Vedika, Chandana, Sagarika, Srikanth, Ridhika and Anukesh. Thank you for keeping the wonderful times you had with me and also for the encouragement that you guys gave.

I offer my greatest appreciation to my Dad, Mom, Anna, Anni, Rhea and Divya for their infinite support. I would have to specially thank Aloys Divya for being with me during my tough times.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION:

1.1. Background

Understanding the gravity variation in Earth would be very helpful in studying the geodynamical process like tectonic movements, mass variation in surface of Earth, and anthropological influence on the geophysical processes. Greater insight on regional to global-scale water cycle variability is a pre-requisite to design developmental and adaptation strategies for climate change. These necessitate a research on the variation of gravity and water cycle. Gravity Recovery And Climate Experiment (GRACE) mission has been launched in 2002 for studying gravity on the surface of Earth("GRACE Tellus" 2013). The data from the mission could be utilized to study these processes.

Mass changes on the surface are recorded by the GRACE mission. These mass changes are caused due to various parameters on the surface of Earth that change spatially and temporally. Precipitation due to the two monsoon seasons (summer and winter) cause the recharging of water in surface and subsurface reservoirs. The precipitation in monsoon thus causes one of the greatest mass changes over the Indian subcontinent(Wahr, Molenaar, and Bryan 1998). This mass change would vary based on the amount of rainfall and also the time period of the rainfall.

Both natural and anthropogenic factors cause change in vegetation. The change in biomass over short term and medium term could affect the mass on surface of Earth. Natural vegetation tend to increase during the monsoon seasons and decrease during the rest of the seasons. Even though agricultural practices is dependent on the monsoons, large scale irrigation projects in the Indian subcontinent initiated since the 20th and 21st centuries(Spate and Learmonth 1972) have made that the time of the change of vegetation could also be dependent on the irrigational practice and water availability in the region. One of the major modes of irrigation in Northern Indian plains is the usage of ground water(Rodell, Velicogna, and Famiglietti 2009). The aquifers over the Northern Indian region have been found to deplete faster due to the increasing amount of water required for irrigation and drinking water purposes(Rodell, Velicogna, and Famiglietti 2009). These depletions in the sub surface water cause a huge mass change.

Since the liberalization in 1990's India has been growing faster and is one of the fastest growing countries in the world(Spate and Learmonth 1972). Due to the growth, the economy of India has changed from a predominantly agricultural economy to an economy dependent on manufacturing and services. These have contributed to the fast urbanization of the country. Urbanization has been very rapid that the population of Delhi has risen from 9.4 million in the 1990s to 16.4 million in 2011("Census of India: Provisional Population Totals Paper 1 of 2011: NCT of Delhi" 2014). Urbanization has been uneven in India and the surrounding regions. Southern India has urbanized very fast and the southern state of Tamil Nadu has half of population in the urban regions of the state. Whereas the north eastern state of Arunachal Pradesh has less than 25 percent of its population living in urban regions("Census of India: Population Enumeration Data (Final Population)" 2014). The increase in the urban sprawl is a variation of the mass in the surface of Earth. Most of urban structures in India are built with brick, mortar, and concrete. The change in the mass of the region could be caused due to the new buildings that have been

constructed on the surface. There have been so many skyscrapers built in urban regions in the last two decades with the use of steel columns(Morel et al. 2001). This could affect the mass in the surface. Also the population increase in these regions has also increased in the demand of water for drinking and other domestic purposes. Most of the population of India are without proper water supply to their houses and workplaces("Census of India: Population Enumeration Data (Final Population)" 2014). These problems have caused the population to be dependent on bore-wells for the water requirement. These have caused depletion in the ground water storage. The depletion is known to affect the mass in the surface of Earth. Even though rainwater harvesting schemes have been implemented in few states such as Tamil Nadu("Rain Water Harvesting" 2014) and Rajasthan, there has been little effort in other states and regions. Water depletion and shortage is one of the key issues facing the subcontinent. This has huge impact on the population sustainability, agriculture, and economic development of the region. The study of GRACE data could be very helpful in understanding the ground water depletion in the country.

Surface snow and ice are found in the Polar Regions and also in the mountainous regions across the globe. The amount of snow and ice in these regions are affected by seasons and also due to global warming(Bajracharya, Mool, and Shrestha 2007). In the Indian subcontinent seasonal snow fall occurs in the Himalayan region, Hindu Kush region and also in Tibet. These regions also host permanent snow and glacier regions. The glacier melt in the summer season is one of the major sources of water for the entire South, South East and East Asian countries(Hasnain 2002). Major rivers of Asia such as the Indus, Ganges, Brahmaputra, Mekong, Irrawady, Yangtze, and Amu Darya rise in the Himalayan-Karakoram-Hindu Kush (HKH) mountain system(Hasnain 2002). Due to global warming, the number of glacial lakes formed in the region has been increasing(Bolch et al. 2012). Melting of the glaciers has increased in the last few years. This can cause change in the mass in the surface in both long and short term. The long term trends can be due to global warming and glacial retreat. The short term change would be due to the variation of amount of snow and ice due to the change of amount of snow and melt during various seasons.

Earthquakes are caused due to plate tectonics(Scholz 2002). The Indian plate has been moving towards the Asian plate and the Burmese plate for many million years(Stein and Okal 2005). This was the reason for the growth of Himalayas and the formation of various trenches and the mountainous system in the Indonesian region(Scholz 2002). The constant movement of these plates have caused huge strain on the rocks in the region. The sudden release of the strain accumulated in the region is commonly known as Earthquake. The sudden release of strain could affect the structure of the rocks and also could cause permanent effects on the geography of the region. Various regions have known to move or sub duct due to Earthquakes. The Sumatra Earthquake on December 24th 2004 was known to shift the sea floor(Abercrombie, Antolik, and Ekström 2003)(Han et al. 2008). These huge changes could affect the mass on the surface of Earth. Plate tectonics could also cause long term effects on the mass of the region(Han et al. 2008). The movement of Indian plate has been responsible for the increase of the height of the Himalayan mountain range, formation of the Tibetan plateau and the Indo-Gangetic plain(Scholz 2002). Even though the formation of these geographical features took a longer period of time, it can be observed in short term too. These changes could affect the mass in surface of earth in long term.

The mass change in the surface of Earth is thus caused by various geophysical phenomena. This makes it very important to study the trends followed by the various parameters and also the trends followed by the change of mass in the surface of Earth. The trend analyses for such studies are known to be affected by the spatial and temporal scales of the data. The problem that arises due to the variation in spatial scale is known as Modifiable Areal Unit Problem (MAUP). Spatial aggregation of data has been known to affect trend analysis and various other analyses that are done on datasets. Since the GRACE gravity data is available in a very coarse resolution, the effect of the spatial resolution on trend analysis need to be studied properly for better understanding the usefulness of the trend analysis methods. The variation of temporal scale is also known to affect the trend analysis too. The problem caused by the variation of the temporal scale is known as the Modifiable Temporal Unit Problem (MTUP). These effects of both the spatial and temporal scales on various analytical methods gives rise to the problem known as the Modifiable Spatio-Temporal Unit Problem (MSTUP)(Maurya).

1.2. Problem statement and motivation:

The study of the trends in the mass change across the Indian subcontinent is crucial to the study of various geophysical phenomena such as the monsoons, plate tectonics, biomass change, and also the anthropogenic induced effects such as varying irrigation patterns, urbanization, population growth, ground water depletion. The trends in these important parameters and the relationship between the parameters are crucial in understanding the cause and effect of the changes caused to the mass of Earth. These require statistical trend analysis on the parameters and also the modelling of the relationship between them. Hence, the fluctuations caused by parameters such as precipitation, vegetation, snow water equivalent, and urbanizations on the Mass change or gravity anomaly requires to be studied. This could be helpful in understanding the variability in the individual parameters and also their effects in particular regions. Since the anomalies in the gravity values are primarily caused by these parameters, it is an essential requirement to understand the effect of every parameter in the research. This requires a detailed mathematical modelling of the relationship.

The trend analysis and the models that could be used for such an analysis would be affected by the variation in the temporal and spatial scales of the data. Particularly when the data of GRACE observed gravity anomalies are in a very coarse spatial resolution temporal resolution, the trend analysis and the relationship studies could be affected. Hence the selection of spatial and temporal scales for such trend analysis studies becomes very crucial. This necessitates the study of the effects of these scales on the trend analysis and the models.

1.3. Objectives:

1.3.1. Main Objective:

- To analyse the spatial and temporal trends in gravity, precipitation, vegetation, urbanization, and snow cover and the effects of MSTUP on the trend analysis.

1.3.2. Sub-Objectives:

- To model the interaction between gravity, precipitation, vegetation , urbanization, and snow cover using Artificial Neural Network and to understand the error propagation in the model
- To find the effects of earthquake induced variations in the model and trend analysis
- To understand the effects of MSTUP on modelling

1.4. Research Questions:

- Is there any specific trend in the variation of the gravity measures over India?
- Do the variations in the parameters i.e. precipitation, vegetation, urbanization, and snow cover affect the trend in gravity? (If so, how?)
- How does error propagate through the model and which parameters influence the model?
- Do earthquakes affect the trend and models?
- How does the use of various spatial and temporal resolutions of the various parameters affect the trend, correlation (if any), and the model?

1.5. Innovation:

Modifiable Spatio Temporal Unit Problem (MSTUP) is a relatively new problem in the remote sensing and geo-informatics field(Maurya). Even though the effects and problems rising due to Modifiable Areal Unit Problem and Modifiable Temporal Unit Problem have been studied well, the combined effects of both the problems have not been studied very clearly. This study aims to study the effects the MSTUP by studying the effects of various spatial and temporal scales on trend analysis. This could be very useful for interpreting the trend results arrived at in the previous researches and also in the future researches for the same study area and the dataset. The research has also been done to understand the relationship between precipitation, vegetation, snow water equivalent, urbanization, and gravity. The creation of the urbanization dataset for the study of the effect of urbanization using the population density and night lights datasets is a very

new approach. The study of the relationship between urbanization and gravity anomalies has never been done. This study also intends to study the relationship between these various parameters causing the surface mass change and also an attempt on modelling the relationship between these parameters using the Artificial Neural Network algorithm. Such a model is new when considering the parameters involved. This model could explain the variations of gravity anomalies with respect to the anomalies in the parameters such as precipitation, vegetation, snow water equivalent, and urbanization. The study also tries to find the signals caused due to earthquake in the Indian land region which has never been attempted.

2. STUDY AREA AND DATA:

The study area used for the study is the Indian subcontinent. The spatial extent of the study is 4.5°N to 38.5°N in latitudes and from 65.5°E to 98.5°E in longitudes. India is a very diverse country with varying climatic regions. The time period of the study is from January 2003 to December 2010.

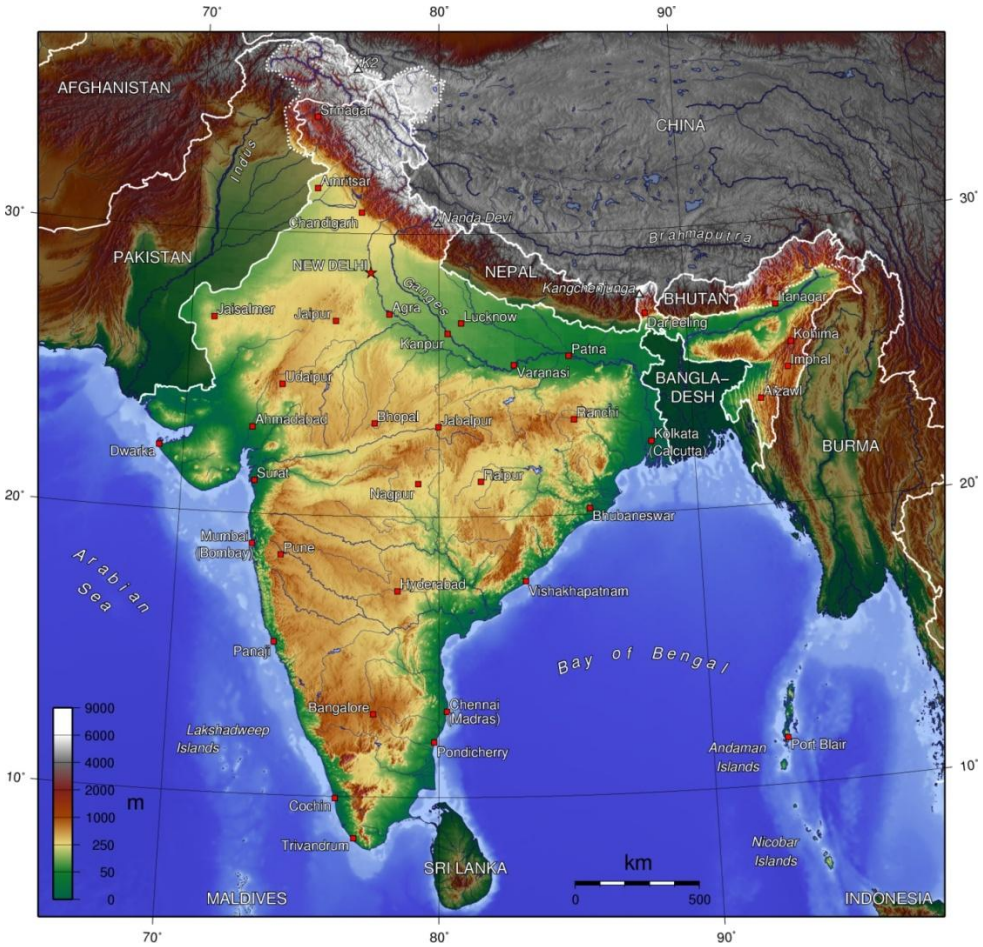


Figure 1 Study Area - India

2.1. Geography and Geology:

The study area primarily consists of two active plates and one plate boundary surrounding the Indian plate. The Indo-Australian plate has been moving towards the Asian plate for 40 million years. The collision of the two plates caused the rise of the Himalayan mountain system, the Tibetan plateau and for the subduction in the Gangetic plains(Khattri and Tyagi 1983). The Himalayan- Karakoram-Hindu Kush mountain system is a very active plate boundary. The region has been affected by various earthquakes. The most devastating of the earthquakes include the dash earthquake of Kashmir earthquake of 2005.Apart from these mountainous regions, mild earthquakes occur in the Western Ghats region in the southern Indian region. The Deccan plateau region was a major volcanic region 65 million years ago (mya) (Bondre, Duraiswami, and Dole 2004). Even though most of the earthquakes in the study are were caused by plate tectonics, there were two devastating intra-plate earthquakes in the region(Schweig et al. 2003). The two earthquakes are the Latur earthquake of 1992 and the Bhuj earthquake of 2002(Schweig et al. 2003).

2.2. Climate and Vegetation:

The precipitation in the region varies from 20cm in the Thar Desert to the mountains of Meghalaya where the precipitation is around 1000cm(Hartmann and Michelsen 1989). The vegetation in the region is also widely varying. The region has dense rainforests in the north eastern Indian states of Arunachal Pradesh, Assam, Meghalaya, Mizoram, Nagaland, and Manipur(Sagreiya 1967)(Puri 1960). Rainforests are also found in the Arakan-Yoma in Myanmar. In the south Rainforests are found along the Western Ghats and Srilanka. The Rainforests in Western Ghats are spread in the states of Karnataka, Kerala, and Tamil Nadu. The North Western part of the study area is typically dry except for the Indus valley which is flooded by the Indus River. The Thar desert is found in the northwestern Indian state of Rajasthan and Bhawalpur state of Pakistan(Singh et al. 1974). The Tibetan plateau is not vegetated and is found in the north-eastern regions of the study area. The Himalayas are covered by alpine forests and meadowlands. Permanent snow occurs in the Hindu Kush-Karakoram-Himalaya Mountain system. There are many glaciers in the region. Some of the glaciers are the source for major rivers such as the Indus, Ganges, and Brahmaputra that flow in the region. The Gangotri glacier which is the origin of the river Ganges is one of the largest glaciers. During the winter months between October and March, Snow fall occurs in the major portions of the mountain system, the Tibetan plateau. The amount of snow is increased during this season. The snow and the glaciers melt during the summer seasons feeding the rivers in the region(Hasnain 2002).

2.3. Population:

The study area has a quarter of the population of the world("Census of India : Population Enumeration Data (Final Population)" 2014). The cities in the region are growing very fast and are some of the fastest growing cities. Delhi and Mumbai are some of the largest cities in the world by population. Apart from them, there are many metropolitan areas that have population more than 5 million. These include the cities of Kolkata, Karachi, Dhaka, Lahore, Chennai,

Bengaluru, and Hyderabad (India). Since the liberalisation of the Indian economy, the cities have been growing very fast and the urban population have been increasing ever since(Jones and Visaria 1997).

2.4. Data:

2.4.1. GRACE gravity dataset:

The gravity dataset used for the study is from Gravity and Climatology Experiment (GRACE) mission. The GRACE mission is a joint mission of National Aeronautical and Space Administration (NASA) and German Aerospace Agency (GFZ)("GRACE Tellus" 2013). The GRACE mission has two satellites which move in sync with the other. The distance between the two are measured by the usage of dash in both the satellites. Their positions are also recorded with the help of the Global Positioning System (GPS) satellites. The satellites' movements are affected very minutely by the change in the gravitational pull from Earth below. The mass change on the surface is the cause for the variation of the pull. The positional change is recorded and using which the variation in the gravitational pull is determined. The gravitational pull which is calculated in milligal (mg) units is then corrected for the atmospheric variations, tidal variations. Thus the second level data is produced. The GRACE second level data is provided as spherical harmonic coefficients and has various errors. In-order to correct the correlating errors in the GRACE data, various filtering techniques is performed on the dataset. The filters include a 200km spatial filter, Gaussian filter (Swenson and Wahr 2006)("GRACE Tellus: GRACE MONTHLY MASS GRIDS - OVERVIEW" 2013). Post these corrections, the data is provided in the Liquid Water Equivalent anomaly (LWE Thickness anomaly) format. The LWE Thickness format is technically defined as the gravitational pull exerted by a water column of a particular thickness and has the units of centimetres of liquid water equivalent thickness. The dataset is provided by Jet Propulsions Laboratory (JPL). Even though the data is provided by German Aerospace Agency, Jet propulsions laboratory and Centre for space research (Texas), the data provided by JPL has the least missing months in the data("GRACE Tellus: GRACE 'Months'" 2013).

2.4.2. Precipitation data:

Precipitation dataset is used from the Tropical Rainfall Measuring Mission (TRMM) data. TRMM is a joint mission by National Aeronautical and Space Administration (NASA) and the Japanese Space Agency (JAXA) (Mission and Kubota 2013). The TRMM - 3B42v7 dataset is used for the research. The 3B42 algorithm used in the dataset produces infrared-merged precipitation and the root mean square of the error estimate(Huffman et al. 2007)("TRMM" 2014). The 3B42 estimates are produced in various steps("TRMM" 2014). The first step involves the use of the microwave sensor in the estimation of precipitation. The second step uses the microwave estimated precipitation to calculate the infrared estimated precipitation. The third step involves

the combination of microwave and infrared based precipitation estimates. These are then scaled to the monthly temporal resolution format. The data is available for the tropical region between 50^o North and 50^o south. It has a monthly temporal resolution and a spatial resolution of 0.25 degree("TRMM" 2014). The TRMM precipitation rate's unit is mm/hr.

2.4.3. Vegetation data

Enhanced Vegetation Index (EVI) is used for the study. The Enhanced Vegetation Index was developed by to overcome the defects of Normalized Vegetation Index (NDVI). Enhanced Vegetation Index is more sensitive to changes in bio mass rather than the reflection on the surface of the vegetation. NDVI is more sensitive to the reflection on the surface of vegetation and it saturates in asymptotically high biomass regions(Huete et al. 2002). Unlike NDVI, EVI also uses the reflection in blue band for calculation.

The formula for Enhanced Vegetation Index (EVI) is given by,

$$Enhanced\ Vegetation\ Index\ (EVI) = G * \frac{NIR - Red}{NIR + C1 * Red - C2 * Blue + L}$$

Here, G is the Gain factor;

L is the canopy background adjustment;

NIR , Red, and Blue are the reflectance in the respective bands;

C1 and C2 are coefficients of aerosol resistance term.

In EVI calculations, the values for L, G, C1, and C2 are given by 1, 6, 7.5 and 2.5 respectively ("MODIS Vegetation Indices (MOD13) C5 User's Guide: Introduction" 2014). The Moderate-resolution Imaging and Spectrometer (MODIS) instrument was launched aboard the terra and aqua satellites by NASA. MODIS provides 36 band data at various spatial and temporal resolutions. The MODIS Vegetation Indices monthly global 0.05 Deg CMG data (MYD13C2) calculated from the aqua satellite data was used for the study. The dataset is available in the MODIS sinusoidal projection and HDF file format. The dataset has a spatial resolution of 0.05 degree and a monthly temporal resolution("MYD13C2 | LP DAAC :: NASA Land Data Products and Services" 2014, 13). The MYD13C2 has both the NDVI and EVI datasets along with error estimates. The EVI dataset from the MYD12C2 was used for the study.

2.4.4. Snow data

Snow Water Equivalent thickness is generally defined as the equal amount of water thickness that could be obtained by melting of a particular amount of snow. It generally describes the amount of snow in the region. The amount of snow can be determined by measuring the microwave brightness temperature from radiation released by the snow present in a snow pack, glacier. Since when the depth of snow increases the scattering caused by the snow crystals increase as there are more crystals(Hall et al. 2001). There are many algorithms that are used for the purpose of deriving the snow thickness from the microwave radiation. The National Snow and Ice Data Centre (NSIDC) provide the AMSR-E/Aqua L3 Global EASE-grids dataset which provides the Snow Water Equivalent (SWE) thickness. The data collected from the NASA Earth Observation Satellite - Aqua for the calculation of SWE("SNOW WATER EQUIVALENT — GES DISC" 2014). The AMSR-E dataset has a 25km resolution and monthly temporal resolution. The dataset

is available as EASE grids (Equal-Area Scalable Earth Grids) with a set of 721 rows and 721 columns("SNOW WATER EQUIVALENT — GES DISC" 2014).

2.4.5. Night lights data

The Defence Meteorological Satellite Program (DMSP)' Operational Line-scan System(OLS) dataset is used for the research. The OLS dataset is produced from a set of 9 Satellites during the time period between 1992 and 2010. The sensors use the optical and near infrared data to produce the night light imagery which comprises of light emission which are both man-made and the natural light sources in the surface of Earth. The sources of these lights could be forest fires, mine fires, street lighting, domestic lighting, industrial lighting, and also record the auroras in the polar region(Small, Pozzi, and Elvidge 2005). The dataset is popularly known as "Night Lights" data and has been used in various studies related to urbanization, economic development, and natural hazards. The Temporal resolution of the dataset is one year. The spatial resolution of the dataset is 30 arc seconds.

2.4.6. Population Density dataset

Population density is useful for the studies related to urbanization. However actual population density data is not available for the Indian region as the Indian ministry of statistics hasn't produced the spatial dataset of the census data. Hence the University of Columbia's World gridded population density dataset (GPWv3) is used for the study. The GPWv3 uses the individual census dataset from the countries across the world and a spatial dataset is created. University of Columbia in collaboration with the Food and Agricultural Organization (FAO) produced the gridded population dataset for the time period between 1990 and 2000("Gridded Population of the World (GPW), v3 | SEDAC" 2013). The population and population density for the years between 1990 and 2000 have been modelled and the data for the years 2005 and 2010 are predicted. The population predicted for the Indian region has a difference of over 5% from the actual population. However the dataset is widely used as it is the only available global spatial population density dataset. The spatial resolution of the dataset is 2.5 arc minutes and the temporal resolution of the dataset is 5 years("Gridded Population of the World (GPW), v3 | SEDAC" 2013).

2.4.7. Earthquake dataset

A list of earthquakes with magnitude above 6 is required for the research for exploration of the effect of earthquakes on the gravity anomaly data and the model that would be subsequently created. The list of earthquakes along with their spatial co-ordinates and time were obtained from the United States Geological Survey (USGS). USGS provides the dataset in the comma separated value file format (CSV) with the latitudes, longitudes, magnitude of the earthquake and also the depth of the earthquake.

Table 1 - Data resolution

Dataset	Spatial resolution	Temporal Resolution	Data Availability
GRACE LWE Thickness Level 3	1 degree	1 month	2003-2010
TRMM Precipitation rate	0.25 degree	1 month	2003-2010
MODIS EVI	0.05 degree	1 month	2003-2010
AMSR-E SWE	25 km	1 month	2003-2010
DMSP – OLS	30 arcseconds	1 year	2003-2010
GPW v3 Population Density data	2.5 arcminutes	5 years	2000-2010

2.5. Structure of the thesis

The Thesis comprises of six chapters. The first chapter explains the background of the research and further deals with the problem statement, motivation, objectives of the research and the research questions. The second chapter explains the datasets used, the time period of study and the study area used and the justification for the same. The third chapter is the researches referred to for the current research. The fourth chapter deals with the methodology which is applied in the research. The fifth chapter deals with the results arrived in the research and the discussions around the results. The sixth chapter deals with the conclusions and recommendations for future work.

3. LITERATURE REVIEW

The following chapter deals with the literature and related work done to this research. Trend analyses have been used in various geophysical sciences and economics for analysing the events that had happened and also for future prediction. Various analytic methods are also used in various researches for analysing the patterns temporally and also for the understanding the variation of the patterns spatially and temporally. Such studies are very useful in understanding the dynamics of various geophysical phenomena in space and time.

3.1. Inferences from trend analysis results for Gravity

Various researches have been done on the trend in the gravity anomalies in the Indian subcontinent. Since the launch of the GRACE mission, global scale gravity anomaly data has

been available which allowed researchers to model the temporal variations geoid, variation in the gravitational potential due to the changes in the surface of Earth. Many researches have concluded that there exists a positive trend in the gravity anomaly in the southern Indian region(Rodell, Velicogna, and Famiglietti 2009). The Indo-Gangetic plain has been witnessing a negative trend in the gravity anomaly(Rodell, Velicogna, and Famiglietti 2009). Rest of the south Asian region don't have a significant trend. The trends in the gravitational anomaly have been result of the increased ground water depletion in the region. The ground water depletion has been acute in the north Indian agricultural region. Ground water has been increasingly used for irrigation in the region. Increasing amount of tube wells, bore wells in the region have been observed. Indian states of Rajasthan, Haryana, Punjab, Uttar Pradesh along with the union territory of Delhi have been witnessing significant drop in the water table because of irrigational practice. However, the reason for the positive trend in southern india has not been conclusively proved. The suspected reasons for the positive trend in the region include tectonic signal and the effect of the Sumatra earthquake on the gravity anomaly along the coastal regions of south eastern India(Tiwari, Wahr, and Swenson 2009). The Sumatra earthquake occurred on December 26, 2004 and affected the southern coast of the region rather than the south eastern region(Han et al. 2008).

3.2. Inferences from trend analysis results for precipitation, vegetation and snow

Long term analysis and forecasting for precipitation requires study of the precipitation in larger temporal time period. A small time period precipitation trend analysis could yield unrealistic or improbable results(Maurya). Over the years climate change has been affecting precipitation in a significant manner over the entire globe. The effect of Global warming has caused variation in the trends. Extreme precipitation trends have been observed in the region(Sen Roy and Balling 2004). Snow and glacier have shown retreat in the last decades due to global warming (Zhao and Moore 2006). The increase in the glacier melt is increasing the flow of the rivers in the region(Hasnain 2002).

3.3. Causes for the variation of gravity anomaly

Gravity anomaly varies in space and time due to many reasons. The gravitational anomaly variation is caused primarily by hydrological factors such as precipitation and ground water("GRACE Tellus: GRACE MONTHLY MASS GRIDS - OVERVIEW" 2013)(Wahr, Molenaar, and Bryan 1998). Snow melting, glacial movement and glacial rebound has also been known to affect gravity("GRACE Tellus: GRACE MONTHLY MASS GRIDS - OVERVIEW" 2013). The variations in the gravitational anomaly have been used for studying the effects of global warming on ice caps in the Polar Regions(Gunter et al. 2009). The breaking of ice and melting of ice in the Antarctic regions have been quantified using the GRACE mission data. Seasonal increase and decrease of snowfall has affected the seasonal variation in gravity anomaly data. Earthquakes have also affected the gravity observations due to GRACE. The Alaska earthquake has affected the gravitational pull in the Alaskan region(Sun and Okubo 2004). The Sumatra Earthquake (boxer day earthquake) on December 26, 2004 affected the gravity measurement in the Indian Ocean region. The Sumatran earthquake caused the movement of the

ocean floor(Han et al. 2008). This movement caused variation in the mass causing a variation in the gravitational pull. Plate tectonics in long term could cause significant variation in gravitational anomaly. The subduction and upliftment of rocks due to plate movement cause significant variation. Tectonic signals have been speculated as one of the reasons for the positive anomaly in the gravitational anomaly in the Deccan region in India(Majumdar and Bhattacharyya 2010).

3.4. Statistical methods for trend analysis

Various trend analysis methods have been used in various researches for the finding the trend in temporal data. Trends could be varying with every trend analysis algorithm. There are many statistical issues with respect to modelling and the inferences from the time series analysis. Non-normal data, correlated errors in time series, correlated datasets, and correlated observational errors are some of the statistical issues(De Beurs and Henebry 2005). The time series analysis is affected by both the observational error and also by the time period taken for the analysis. The data used for temporal analysis of climate change led to the *"hockey stick"* controversy where the time period taken had affected the inferences arrived at(Stern 2007). Among the most populous of the trend analysis methods is the linear regression method. The slope found from the linear regression algorithm is the rate of change of the parameter with respect to time. Linear regression results are however not very accurate as they are likely to be affected by the existence of outlier in the temporal dataset(Muhlbauer, Spichtinger, and Lohmann 2009). The least squares method for estimating the trend in the value is one of the most common algorithms used.

3.5. Statistical methods for analysing relationship

The most commonly used methods for understanding the relationship between the data is the correlation coefficients. Two types of correlation coefficients are normally used for the analysis. The pearson's correlation coefficient and the Spearman's Rho correlation coefficient. The pearson's correlation coefficient explores the linear relationship between two sets of data. The value of the pearson's correlation coefficient varies from -1 to +1. -1 indicates purely negative correlation whereas +1 indicates purely positive correlation(FIELLER, HARTLEY, and PEARSON 1957). However the problem with the pearson's correlation coefficient is that the effect of outliers on the correlation coefficient. It easily gets affected by the values that could be rare or unrealistic. However spearman's rho correlation coefficient is more efficient when it comes to the outliers as it doesn't get affected by the outliers(FIELLER, HARTLEY, and PEARSON 1957). Spearman's Rho correlation coefficient is the second most popular measure for analysis of correlation.

3.6. Modifiable Spatio-Temporal Unit Problem

The data used for scientific research have always been of different spatial and temporal resolutions. Because the resolutions vary spatially and temporally, the analysis on the datasets could yield different results and have been problematic for arriving at inferences(Jelinski and Wu 1996). One of the primary issue is the usage of spatially and temporally un-matching datasets. The spatial and temporal resolutions of the data have a crucial role when the data is used for the

analysis of geophysical phenomena. The usage of some scales could yield obscured or unrealistic analysis results for particular geophysical phenomena(Maurya). Hence scales of the data need to be chosen properly for better understanding of the geophysical phenomena. Modifiable Areal Unit Problem (MAUP) has been observed problematic in various researches(Dark and Bram 2007)(Fotheringham and Wong 1991). There are two problems associated with the MAUP. The first problem is the aggregation and the second problem is the zonation. The Aggregation technique used and the neighbourhood values may cause the differences in the results. Zonation problem is defined as the one caused by the selection of various combinations of neighbourhoods(Maurya). While spatially aggregating the data, there is multiple number of combinations of zones possible and each could generate variations in the results.

Similar to the areal unit problem, Modifiable Temporal Unit Problem (MTUP) deals with the problems rising due to the temporal granularity of the dataset(Cöltekin et al. 2011). The choice of the time unit, granularity and time period of study could affect the outcome of the results. Researches have indicated that the trend analysis on various temporal granularities for vegetation, precipitation could yield different analysis results. The researches show that the cyclic pattern and the linear regression based trend analysis have been affected by the temporal units. Similarly various methods for aggregating the data temporally could cause variations in the analysis of the trend(Maurya).

The combined effect of these two problems gives rise to the Modifiable Spatio-Temporal Unit Problem (MSTUP) which is very recent. The researches done for addressing the MSTUP is very minimal(Maurya).

3.7. Modelling approaches

To understand the variations in the gravitational measures and the ability of other parameters to explain the variations, a modelling has to be done with gravity as the explained variable and various parameters such as precipitation, vegetation, urbanization, snow water equivalent as the explanatory variables. There are various algorithms used for modelling and predicting the values. Simpler linear regression and non-linear regression models are uni-variate in nature as they have one dependent variable and one independent variable. This requires the use of machine learning algorithms for understanding the variation and the classification of the dataset. Some of the commonly used machine algorithms include Artificial Neural Network, Random Forest algorithm, Decision tree algorithm, and naïve Bayesian classifier algorithm. The Naïve Bayesian algorithm however assumes that the variables are independent of each other. The variables used for modelling the gravity values particularly precipitation and vegetation are dependent on each other. This makes the naïve Bayesian classifier unsuitable for the purpose(Tu 1996). The Decision tree algorithm and random forest algorithm are classifiers that are dependent on the initial condition and also overfit to the data for which it is trained to. In an ANN model, the inter dependence of variable (or redundancy) could increase the accuracy of the model however would require longer time for processing and creation of the model. ANN can also model non-linear relationships between the explanatory variables and the explained variable. The drawback of the ANN model is the blackbox nature of it(Tu 1996). The initial assumption for the parameters begins randomly and it autocorrects itself based on the training dataset. ANN based models have

been widely used in various fields for various applications like pattern recognition , forecasting etc.

4. METHODOLOGY:

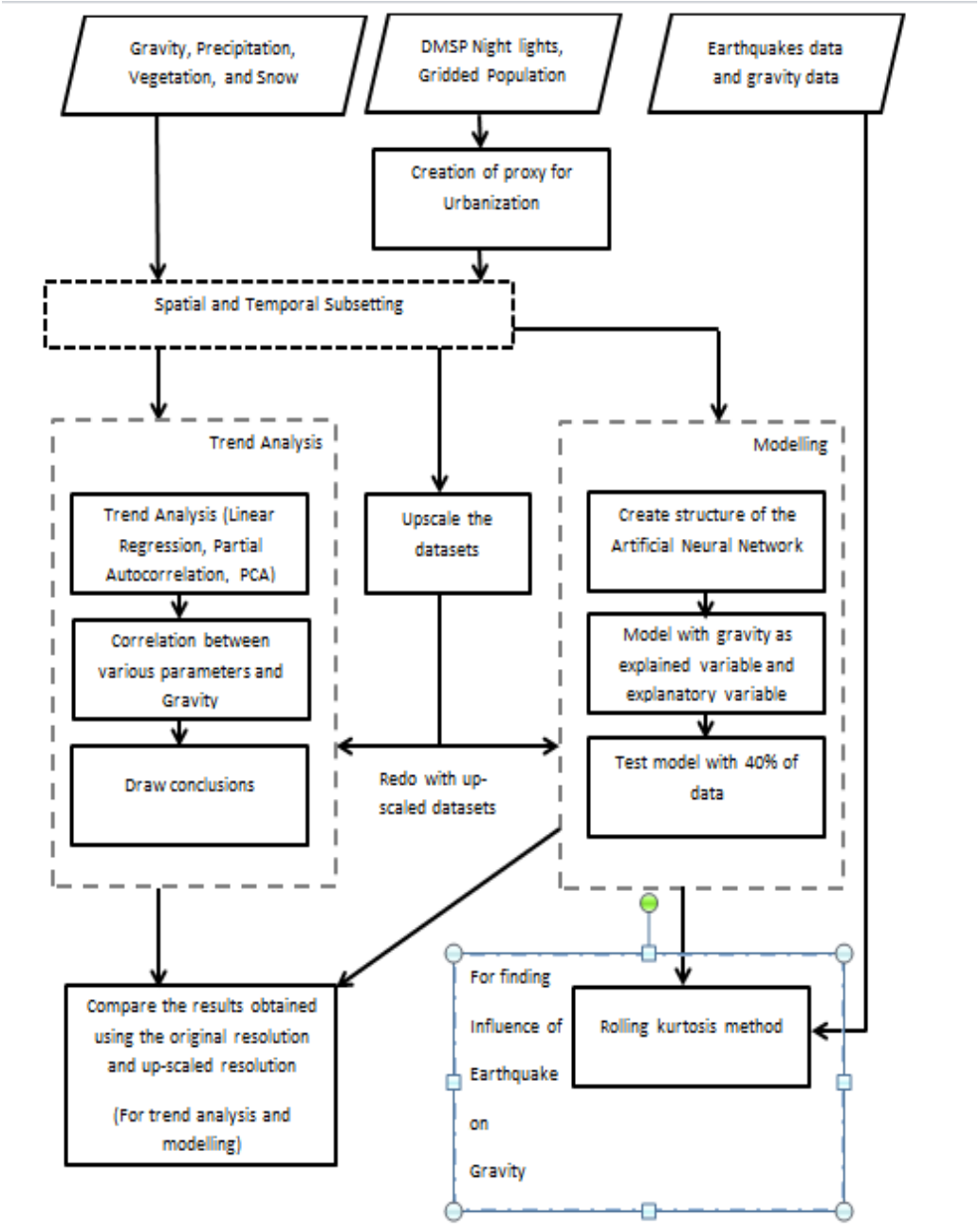


Figure 2 General Methodology chart

4.1. Data preparation:

4.1.1. GRACE Gravity dataset

JPL provides the GRACE Liquid Water Equivalent Thickness data in the NetCDF format("GRACE Tellus" 2013). It is stored as a three dimensional grid dataset. The file contains data for the entire globe between 180°W to 180°E Longitude and 90°S and 90°N Latitude. The temporal extent of the data is from 2003 Jan to 2011 Dec. The dataset for the entire globe is clipped for the bounding box of 4.5°N, 65.5°E, 38.5°N, and 98.5°N. The dataset is also temporally clipped between 2003 January to 2010 December.

4.1.2. TRMM Precipitation dataset

The TRMM Precipitation dataset is provided by NASA in the NetCDF format for the entire globe as a three dimensional grid dataset. The file contains the data for the region between 180°W to 180°E Longitude and 90°S and 90°N Latitude. The temporal extent of the data is from 2003 January to 2010 December("TRMM" 2014). The TRMM precipitation dataset is subsetted spatially for the bounding box of 4.5°N, 65.5°E, 38.5°N, and 98.5°N. However the GRACE LWE dataset is provided as anomaly data based on the average between 2004 January and 2009 December. Hence the Precipitation rate dataset is averaged for the time period between 2004 January and 2009 December. From the actual precipitation data, the averaged data is subtracted to arrive at the Precipitation anomaly data.

4.1.3. Enhanced Vegetation Index dataset:

The MODIS Enhanced Vegetation Index dataset is provided in the HDF format with sinusoidal projection system. Using the MODIS Reprojection Tool (MRT), the dataset is mosaicked, subsetted and converted to GeoTIFF format with a Geographic grid system. The dataset is subsetted spatially for the extent of 4.5°N, 65.5°E, 38.5°N, and 98.5°N. Along with the spatial subset, the dataset is also temporally subsetted for the time period of Januaury 2003 to December 2010. The format of the file is converted from HDF to GeoTIFF using MRT. From the HDF dataset, only the Enhanced vegetation Index data and the enhanced Vegetation index quality datasets are extracted. All the files are further converted into NetCDF format using python and the files are arranged in a 3 dimensional grid dataset similar to the GRACE LWE dataset. Inorder to create the EVI Anomaly dataset, the EVI data is averaged for the time period of 2004 January and 2009 December. From the actual EVI data the averaged data is subtracted to arrive at the EVI anomaly.

4.1.4. Snow Water Equivalent thickness dataset:

The AMSR-E Snow water equivalent thickness dataset is present in the HDF format with a polar projection. Using python, Pyhdf and the Scipy packages, the AMSR-E Snow Water Equivalent Thickness dataset was converted into the NetCDF format. The dataset is spatially subsetted for the bounding box of 4.5°N, 65.5°E, 38.5°N, and 98.5°N. In order to create the SWE Thickness

Anomaly, snow water equivalent thickness for every pixel is averaged for the time period between 2004 January and 2009 December. The Anomaly data is then created by subtracting the actual data and the averaged data.

4.1.5. Urbanization Dataset:

For creating the Urbanization Index, two datasets were used. The two datasets are the DMSP "Night Lights" dataset and the University of Columbia GPWv3 Population Grid dataset. The DMSP "Night lights" data taken by various sensors are provided in the GeoTIFF format("Earth Observation Group - Defense Meteorological Satellite Progam, Boulder | Ngdc.noaa.gov" 2013). The datasets for the years between 2000 and 2010 are taken. However the datasets are from three different sensors. Because of that, the illuminations tend to differ. In-order to avoid the differences that arise due to these, all the datasets are calibrated using the Elvidge coefficients. The usage of Elvidge coefficients for every sensor data is an alternative method for calibrating the datasets instead of using actual sensor calibration which are not available(Elvidge et al. 1997). The dataset is then subsetted spatially for the extent of 4.5°N, 65.5°E, 38.5°N, and 98.5°N. The temporal granularity of the dataset is one year. After applying the Elvidge coefficients for calibration of the DMSP Night lights data, The Maximum Night lights value is 63. This value of 63 is usually present in large metropolitan cites.

The University of Columbia GPWv3 population density dataset is provided in the BIN format("Gridded Population of the World (GPW), v3 | SEDAC" 2013). The file format of the GPWv3 population density dataset is changed to GeoTIFF and is then subsetted for the extent of 4.5°N, 65.5°E, 38.5°N, and 98.5°N. The GPWv3 dataset is available for the years of 2000, 2005, and 2010. Population density growth follows a second order non-linear curve. For every pixel, a second order non-linear regression is fitted and the value for every year in between is predicted. The dataset is resampled to the spatial resolution of DMSP OLS data. Since population density is a ratio of population to area, the resampled data for every pixel is the same as the parent pixel.

To create the Urbanization Index dataset, the population density grid data and the night lights data are used. The following formula is used for calculating the urbanization index.

$$Urbanization\ Index\ (U.I) = \frac{Population\ density * Night\ lights\ value\ in\ the\ pixel}{Maximum\ night\ lights\ value\ in\ the\ image}$$

To create the Urbanization Index Anomaly dataset, the UI values are averaged for every pixel between the time period of 2004 and 2009. The Anomaly is then found out by subtracting the averaged UI value from the actual UI value. After performing the calculation, the dataset is converted to a three dimensional NetCDF format.

4.1.6. Earthquake dataset:

The Earthquake dataset provided by the United States Geological Survey is in the Comma Separated Value (CSV) file format. The dataset is used as provided by the USGS.

4.2. Preliminary Data Analysis:

The preliminary data analysis involves the plotting of the datasets to understand the pattern and the relationship before doing the trend analysis and relationship analysis on the dataset. The GRACE LWE Thickness Anomaly data is plotted in the timescale along with the Precipitation rate anomaly, Enhanced Vegetation Index anomaly, Urbanization Anomaly, Snow Water Equivalent Index Anomaly. The Enhanced Vegetation Index Anomaly is scaled down by a factor of 3000 so that it could be visualized along with the other factors which have lesser anomalous values. This time series would be helpful in identifying the pattern in a temporal manner.

To assist in the finding of the relationship between various parameters of the study, the parameters (Precipitation Rate anomaly, Enhanced Vegetation Index anomaly, Snow Water Equivalent Thickness anomaly and Urbanization Index anomaly) are plotted against the GRACE LWE Thickness anomaly. This helps in understanding the existence of linear or non-linear relationship between various parameters and the GRACE LWE Thickness anomaly.

4.3. Trend Analysis

The major trend analysis used for the research is the linear regression analysis which is a parametric test. Linear regression trend analysis can be done in many software packages. For the research, the linear regression analysis was done using the Scipy package from Python.

4.3.1. Linear Regression

Linear Regression is normally used for the calculation of the analysis of two variables where one is dependent on the other. In order to find the equation of the straight line from the dependent and independent variable values, least squares estimation method is done. The dependent variable y for every set of value is a linear combination of the parameters in the independent value (McCuen 2002). If in a linear regression for modelling, n points are there and one independent variable; two parameters β_0 and β_1 , then the Straight line is given by the following equation.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1,2,3 \dots n$$

Here, ε_i is the error term and "i" is the observation. Based on a random sample of the model, the population parameters are estimated to obtain the sample model.

$$\hat{y} = B_0 + B_1 x_i$$

The residual, $e_i = y_i - \hat{y}$ is the variation between the value of the dependent variable and the predicted model, $\hat{y}$. Ordinary least squares method attempts in reducing this variation and for the model which has the least mean square error is taken as the linear model. The sum of the square residuals is given by the formula,

$$SSE = \sum_{i=1}^{n} e_i$$

The reduction of the error in this results in a linear set of equations of the parameters which are used to solve the get the parameter estimates B_0 and B_1. (Helsel and Hirsch 1992)

The following method is used for finding the slope maps of for the trend. In a linear regression trend analysis for anomaly data, the intercept map would be of less importance. Using the slope

values calculated using the least squares estimator method and the p-value for the fit, the regions of significant slopes are found. The method is shown in the figure

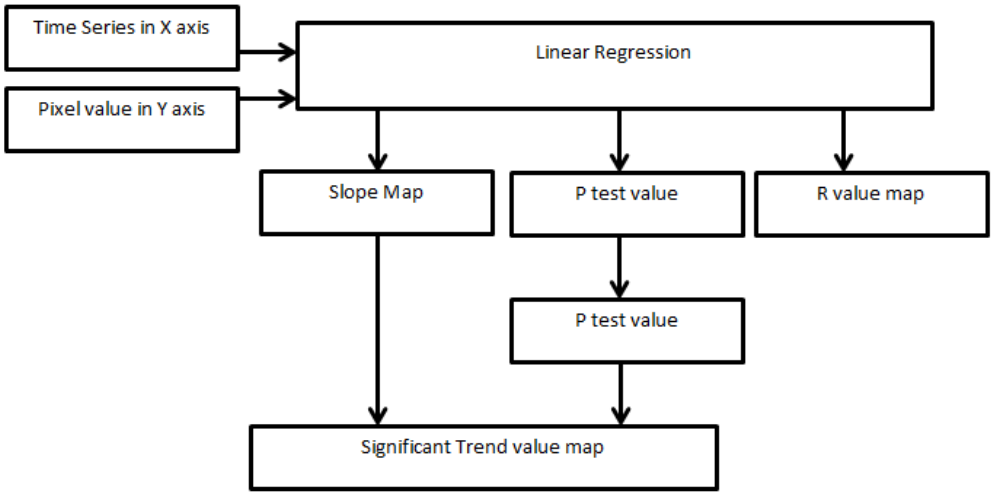


Figure 3 Linear regression methodology

The linear regression for the data is done pixel wise across the time series. Negative slopes indicate a negative trend in the pixel and positive slope indicate positive slope in the respective pixel. The same linear regression analysis is used for finding the significant trend regions in the Indian subcontinent for Gravity, Precipitation, Vegetation, Urbanization, and Snow water equivalent thickness. The slope gives the rate of change the particular value over the unit time period.

4.3.2. Partial Autocorrelation function

Partial Autocorrelation function is done on a time series data to find the repetivity in the data. It helps in understanding the seasonality. The time step in which the data repeats could be found out using the Partial autocorrelation function. The Partial autocorrelation function is done using the package in python. For a perfect sine wave the partial autocorrelation function is also a sine wave. So using the Partial autocorrelation function, the cyclicity in the datasets could be found out. The calculation for partial autocorrelation function is done based on the Box-Jenkins method in time series modelling(Box, Jenkins, and Reinsel 2013). If in a time series Z_t, the partial autocorrelation of a lag k, which is denoted by $\propto (k)$ is the autocorrelation between t and Z_{t+k}. It has a linear dependence of t+1 to Z_{t+k-1}.

$$\propto (1) = Cor\ (z_{t,}, z_{t+1})$$

$$\propto (k) = Cor\ \left(z_{t+k,} - P_{t,k}(z_{t+k}), z_{t+1} - P_{t,k}(z_t)\right), for\ k \geq 2$$

The Box, Jenkins and Reinsel algorithm is commonly used for the autocorrelation function. In python, partial autocorrelation function is implemented using the scipy package.

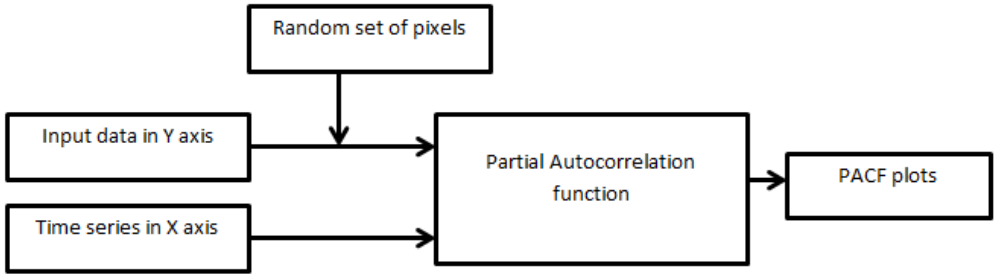


Figure 4 Partial Autocorrelation Function methodology

4.3.3. Deseasonalization

De-seasonalization of the dataset becomes necessity since the temporal long term and medium term trends in the dataset could be affected by the seasonal fluctuations in the dataset. Inorder to avoid the effects caused due to these seasonal fluctuations in the long term trend analysis, deseasonalization is done. The de-seasonalization has been done by using moving average. Based on the cyclic nature of the dataset, the window size of the moving average function is chosen. Usually the datasets tend to have cycles of one year. Hence the moving average function window is chosen to be 12 time steps (since the dataset is monthly dataset). Deseasonalization is not required for the urbanization dataset as it is present in the yearly temporal scale. Linear regression trend analysis is repeated on the deseasonalized dataset. This may or may not have different trends when compared with the trend analysis done on the original anomaly dataset. The linear regression based trend analysis method is discussed in earlier chapter.

4.3.4. Principal component Analysis

Principal Component Analysis (PCA) on a time series is very helpful in understanding the trends and variations from the trends in every pixel. The PCA on time series is similar to the PCA on spatial dataset however that the PCA on time series. The principal component analysis is used to identify the patterns in data. PCA is dimension reduction method using which the number of dimensions is reduced and hence the redundancy is reduced. This however doesn't incur the loss of information. Even though the PCA technique is used widely in Digital Image Processing, it also has application in finding redundancy and patterns in the dataset. When the principal component analysis is applied on a time series, it results in the finding of the long term and short term trends in the data as it is indicative of redundancy and variation from the redundancy. Implementing PCA is done in the following method. The mean of the data across the time period is found out initially. The actual values in all the initial data dimensions are then subtracted by the mean. The resultant data is similar to an anomaly dataset for which the mean is zero. After the subtraction of mean, the covariance matrix is found. The covariance matrix will be of n-dimension. The number of dimensions is determined by the number of time steps for which the dimension reduction needs to be done. Since the data for which the principal component analysis is done has 96 time steps (January 2003 to December 2010), the covariance matrix created will have a dimension of 96. The size of the data hence would be 96x96. The covariance matrix is a square matrix with 96 elements across its diagonal. The Eigen values and Eigen vectors for the

covariance matrix are then calculated. This results in 96x1 shaped Eigen values matrix and 96x96 shaped Eigen vector matrix. The resultant Eigen vector matrix is rearranged based on the eigen values. The largest eigen value has the most significant relationship with the actual data. With the Eigen vectors thus found, a feature vector is formulated by using the following form.

$$Feature\ Vector = (eigen1, eigen2, eigen3, \dots eigen96)$$

The number of eigen vectors used is decided based on significance. Since the last Eigen vectors usually signify noise or unwanted variation, it is usually omitted. The component data is then given by the matrix multiplication between the row feature vector (post selecting the required components) and the transposed matrix of the actual data.

$$Component\ data = Row\ Feacture\ Vector\ x\ Data\ Transposed$$

The component data is the resultant matrix(Smith 2002) which has the data items in columns and the dimensions along the rows. The maximum number of components for the PCA on the time series analysis would hence be the number of time steps. In the study the maximum number of components used is 20. As the later components tend to show the noise and the unwanted variability in the dataset rather than any meaningful variation, the number of components is limited to 20.

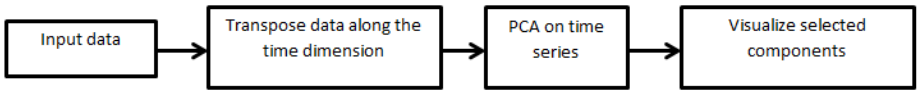


Figure 5 Principal Component Analysis flowchart

4.4. Artificial Neural Network

4.4.1. Creating the model

Artificial Neural Network for mathematically modelling the relationship in the system was created using the pybrain package in python. The pybrain is a very powerful package for machine learning algorithms and for creating supervised and unsupervised learning methods. Using the pybrain package a feedforward network was created. The structure of the neural network is given in the figure below. The neural network algorithm for the relation is a feed-forward neural network. The structure consists of one input layer, one hidden layer and one output layer. There is also a bias module used. Bias modules are used for the shifting of the values for calculations. Bias becomes a necessity when modelling values which have different units. A full connection is used for connecting the various layers. A full layer is one in which all the input nodes are connected to all the hidden layer nodes, which in turn are connected to all the nodes in the output layer. The Bias module is also connected to all the hidden layer nodes and the output layer. There are four nodes for the input layer representing the four parameters which are used as explanatory variables. There are four nodes in the hidden layer which have an activation function of hyperbolic tangent function (tanh(x)). The use of hyperbolic tangent function is that it can be used to scale and rotate the value in both negative and positive regions which the sigmoid function cannot do. Since some of the explanatory variables in the network could have negatively correlating relationship and some of them with positively correlating relationship, a hyperbolic

tangent function becomes a necessity. If no activation functions are used it creates a linear relationship between the input variables and the output variable. The weightages for the connections are determined by using the back propagation training algorithm on the network using the training datasets.

4.4.2. Creating the training datasets and testing datasets

The training datasets are created for training the neural network to determine the weightages of the connections. Every pixel in the GRACE LWE Thickness anomaly dataset has a value in every time step i.e every month. Every pixel in every time step in GRACE LWE Thickness anomaly dataset has a corresponding value in every pixel in every time step for all the explanatory variables (Precipitation, Vegetation, Snow Water Equivalent thickness, and Urbanization). All such instances which don't have NaN (Not a Number) or inf (Information less) values are taken as individual records. Out of the total available individual records, 75% of the records are randomly selected and are arranged as the training dataset and the remaining 25% of the records are used for testing. Thus the training and testing datasets are created for model creation purposes.

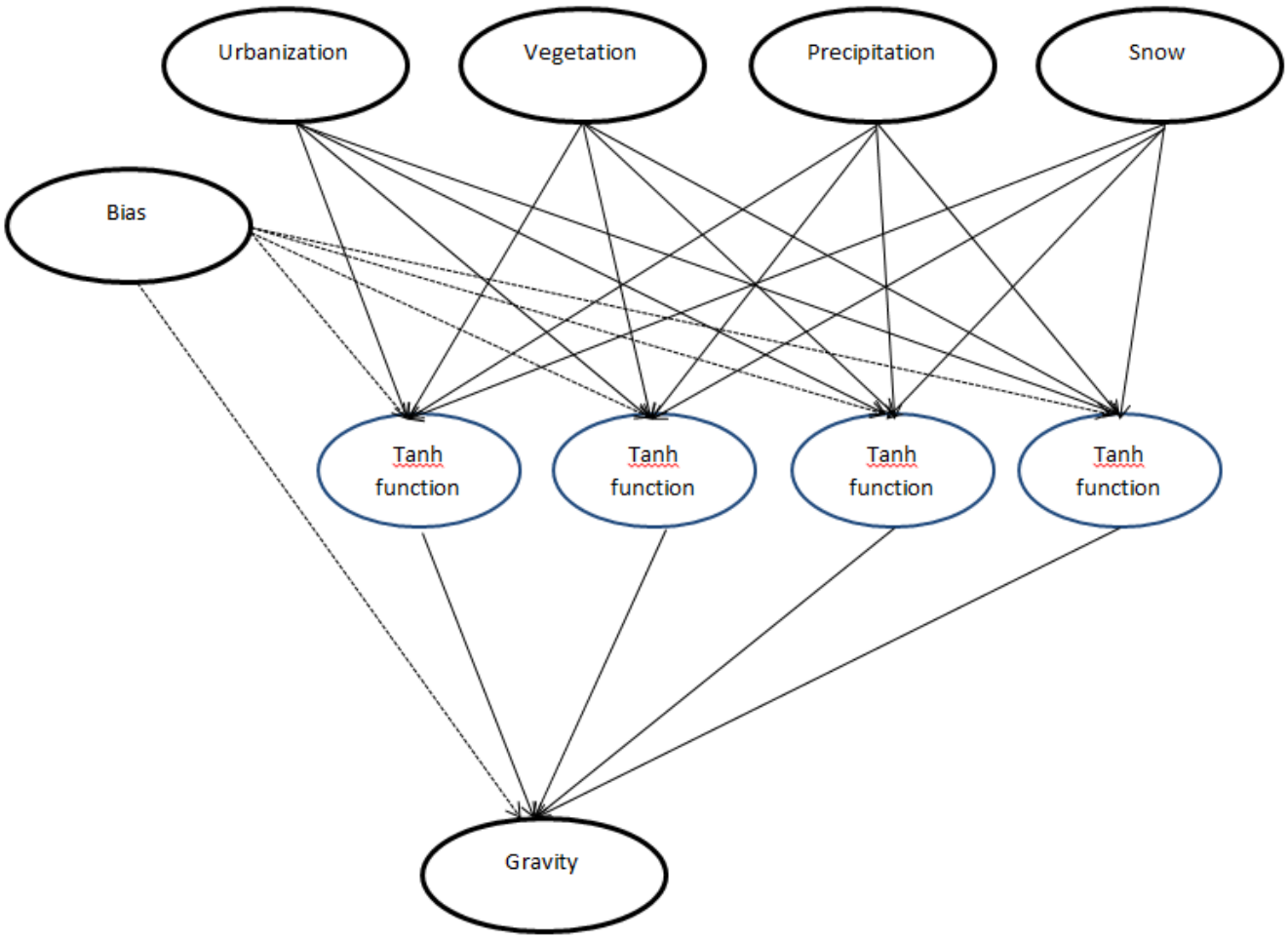


Figure 6 Artificial Neural Network Structure

4.4.3. Training the model

The model created is trained using the back propagation trainer from the pybrain package in Python. Since the Artificial Neural Network is a black box model because of the random origins, every instance of training could yield arrive at different weightages. Hence the set of weightages for which the error in the testing is the least could be used for the model. To find the error, using the weightages created by the back propagation trainer, the values of explanatory variables from the testing dataset is activated. This yields a predicted value for gravity anomaly for the particular instance. These are then checked with the actual gravity value. Based on the difference between the actual gravity anomaly value and the predicted anomaly value, the root mean square is calculated. A new root mean square value is present hence for every time the network is trained using the training dataset. In order to find a suitable set of weightages, the training is repeated for 100 times. The trained set of weightages for which the root mean square value is the least of the 100 iterations is thus realized as the model parameter values. The model structure and the associated parameter values are stored in an XML file for future use and activation.

For finding the effects of MSTUP on the models, the model is created with the temporally and spatially upscaled resolutions of all the datasets. A comparative analysis between the models created at various resolutions is then done.

4.5. Monte-Carlo Simulation Method

Monte-Carlo simulation method is used for understanding how the error in input data influences the model predictions. The enhanced vegetation index, precipitation, and the urbanization index data have their own observational and other statistical error. These errors in the input data might affect the prediction of the model. For analysing it the Monte-Carlo method is used. In a Monte Carlo simulation method 1000 Gaussian distributed random values are generated based on the mean (input data value) and the standard error of the input data at the particular pixel. Numpy package in python is used for this purpose. Once the generation is done, the neural network is activated with thousand combinations of randomly generated input. The output for every combination is then plotted as a distribution. The output plot would signify the effect of the error caused on the model predictions.

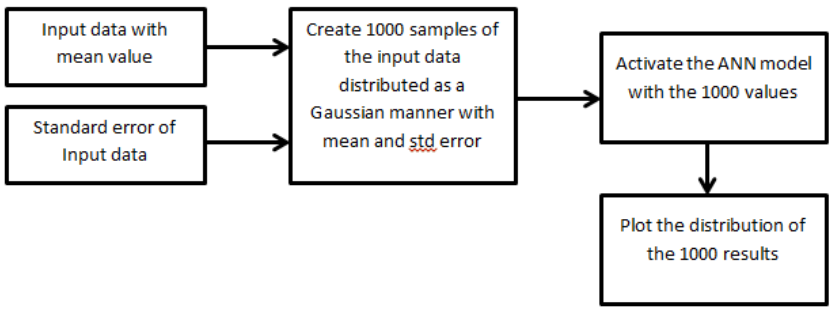


Figure 7 Monte - Carlo method for understanding Error propagation

4.6. Model activation For finding the influence of parameters

For finding which parameter influences gravity the most, the model that has been created is used. The model is activated with all the non-zero values of parameters other than one respective parameter .The root mean square or the deviation from the actual value is found. The method is repeated for every parameter. The parameter with the least root mean square is the one which influences the gravity value most.

4.7. Rolling Kurtosis

Kurtosis is defined as the fourth moment about mean. Kurtosis is calculated for finding the peakedness in the data. It was defined by Pearson in 1906. Kurtosis formula is given by,

$$\text{Kurtosis} = \frac{\mu_4}{\sigma^4} - 3$$

Where, μ_4 is the fourth cumulant and σ^4 is the square variance. Kurtosis is used to find the peakedness of a given distribution of data, rolling kurtosis across a temporal data is used to find the peakedness of the data in a particular time step when compared with the neighbouring time steps. Hence rolling kurtosis would give the sudden rise or fall in the values. Rolling kurtosis is calculated in the same manner of kurtosis except of calculating the average amongst a population, the average in a particular window is used. The rolling kurtosis over the time period of GRACE liquid water equivalent thickness data would be helpful in analysing the impact of earthquake which is a rare but sudden event. In the study, the window for calculating the moving kurtosis is 5. This shows a sudden peakedness of the kurtosis value within 5 time steps. By comparing the change in the value of rolling kurtosis over the pixel where the earthquake occurred and its neighbouring pixels in the time period of the earthquake would show the sudden effects of an earthquake.

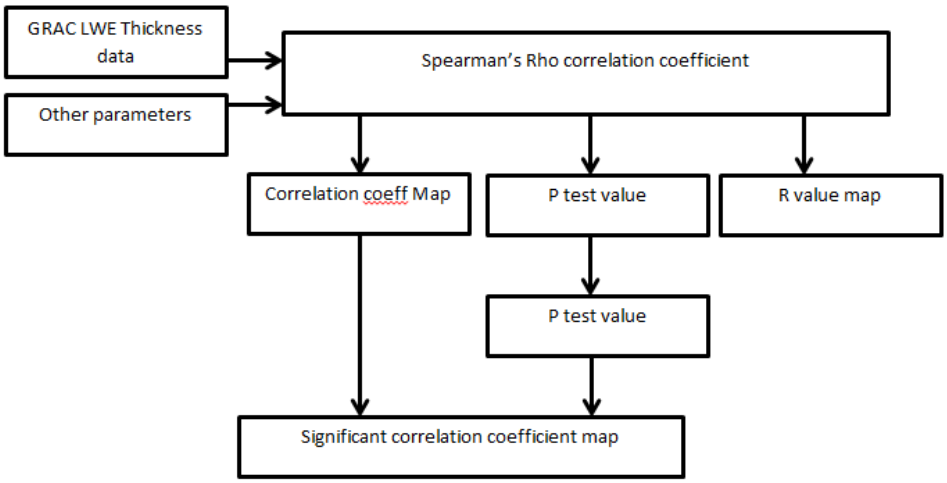
4.8. Relationship Analysis

For understanding the relationship between the parameters, the correlation coefficient between various parameters (EVI Anomaly, PCP Anomaly, UI Anomaly, and SWE Anomaly) and the GRACE LWE Thickness Anomaly is calculated. There are two primary correlation coefficients that could be used for understanding the relationship between two variables. The coefficients are the Pearson's correlation coefficient and the spearman's rho correlation. The pearson's correlation coefficient has tendency to be affected by outlier whereas the spearman's rho coefficient is not affected heavily by the outliers(FIELLER, HARTLEY, and PEARSON 1957). Hence the Spearman's rho coefficient was used for the understanding the relationship.
The spearman's rank correlation coefficient is calculated by the formula,

$$\text{Rho(p)} = \frac{6\sum d^2}{n^2(n+1)}$$

Where d is the difference between paired ranks and n is the number of pairs("Correlation Coefficients" 2014).

For the purpose of understanding the effect of resolutions on the correlation coefficient, the process is done on the data with coarser resolution.



5. RESULTS AND DISCUSSION

5.1. Data preparation

The entire anomaly data were prepared as mentioned in the methodology chapter. The urbanization dataset was created as explained in the chapter. The urbanization dataset manages to single out urban centres from the initial night lights map. The Indo-Pakistan border in Rajasthan, oil rigs offshore of Mumbai which are visible in the night lights map are not visible in the newly created urbanization index map. Urbanization Index dataset hence could be considered reliable.

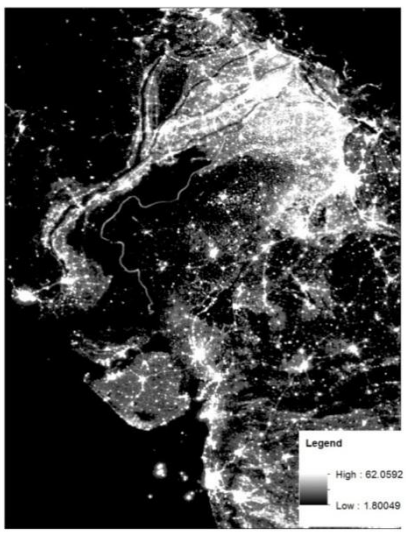


Figure 10 DMSP OLS dataset

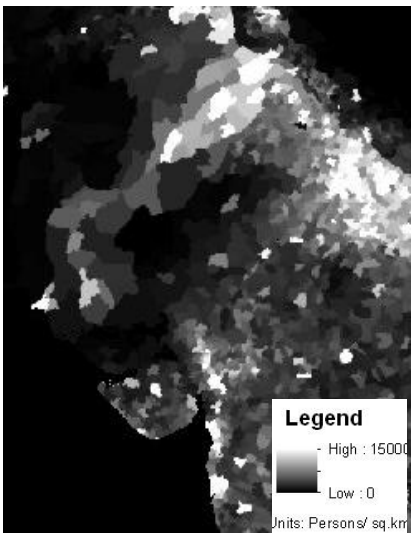


Figure 9 Population Density dataset

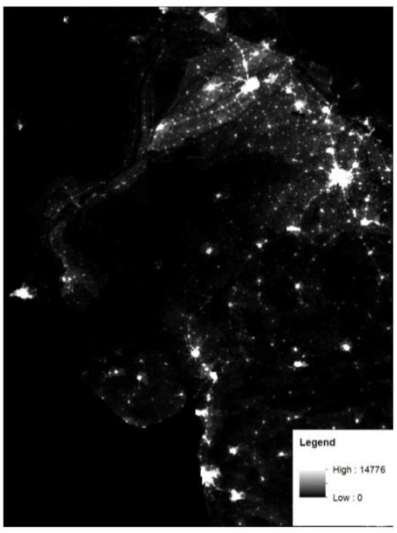


Figure 8 Urbanization Index dataset

5.2. Preliminary analysis

The preliminary time series data plot shows that the seasonal variation of gravity anomaly is synchronized with that of the vegetation anomaly and precipitation anomaly parameters. The seasonal increase and decrease of the gravitational anomaly is related to the monsoon seasons. The time series plot in the figure given below explains the variation in the time steps. The time of the increase of the decrease of the gravity anomaly (LWE Thickness) is varying in various regions because of the variation of the monsoon season and intensity. The gravity anomaly along with vegetation and precipitation show a near sinusoidal wave pattern in the south, east and most parts of north India. In drier regions in the west and in the Tibetan plateau, the wave pattern is absent due to the drier climate and weaker monsoons. Due to the scanty rainfall in the drier region, LWE thickness is not following a sinusoidal pattern in the regions. In some parts of north India, the wave shows a significant decline.

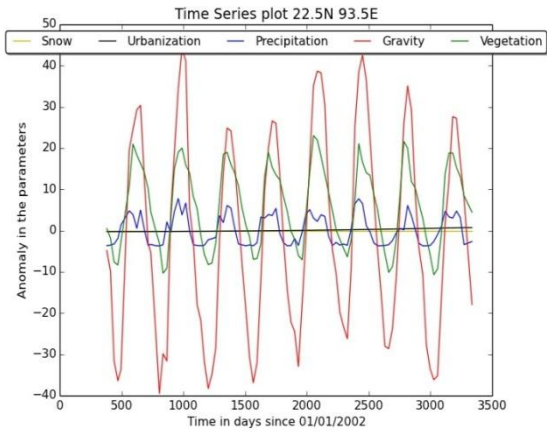


Figure 11 Time series plot

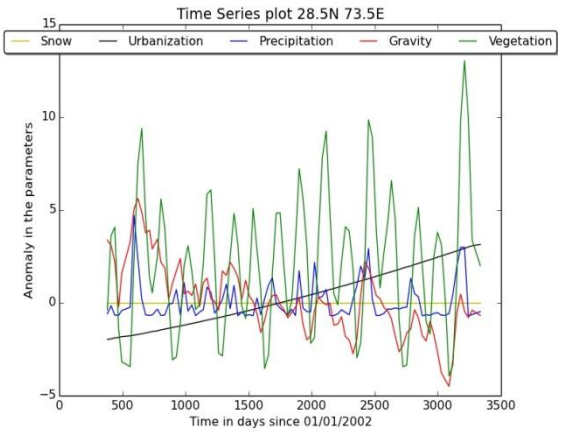


Figure 12 Time Series plot

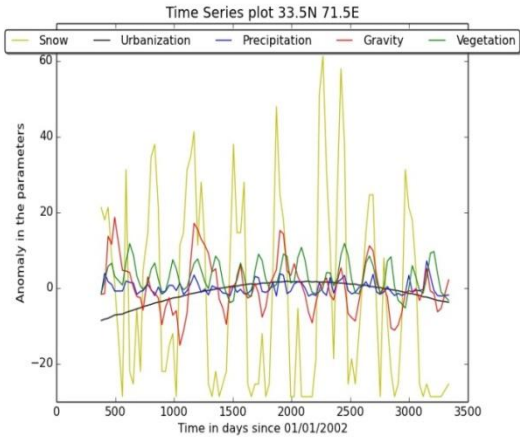


Figure 12 Time series plot

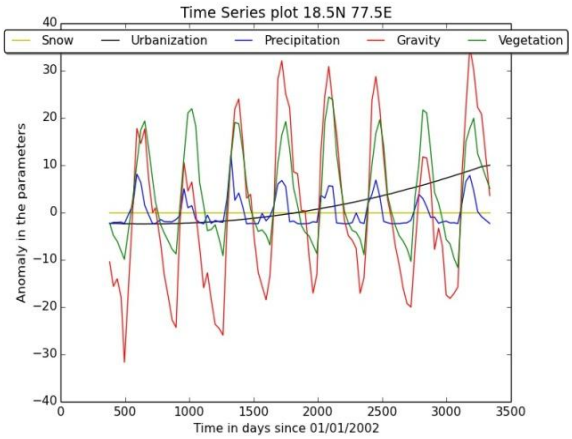


Figure 13 Time series plot

5.3. Trend Analysis on Gravity LWE Thickness

The trend analysis on gravity was done using linear regression, partial autocorrelation function and principal component analysis methods.

5.3.1. Linear regression results

Linear regression on the gravity anomaly data was done from the year January 2003 to December 2010. The observations of the linear regression results are in line with the researches done before on the gravity anomaly data(Rodell, Velicogna, and Famiglietti 2009). Significant trend results are found in the northern southern parts of India. North Indian regions covering the states of Haryana, Delhi, Punjab, Uttar Pradesh, and parts of Rajasthan are showing significant negative trend. One of the causes for the negative trend is due to the decrease in the ground water in the region. The depletion of ground water causes huge mass change in the subsurface of Earth which causes decrease in the gravitational pull in the particular region. The results are in line with the trend analysis researches done before on the north Indian region for studying the depletion of the ground water in the region(Tiwari, Wahr, and Swenson 2009). The major causes for the depletion of ground water in the region are due to the pumping of ground water for irrigation and for domestic uses.

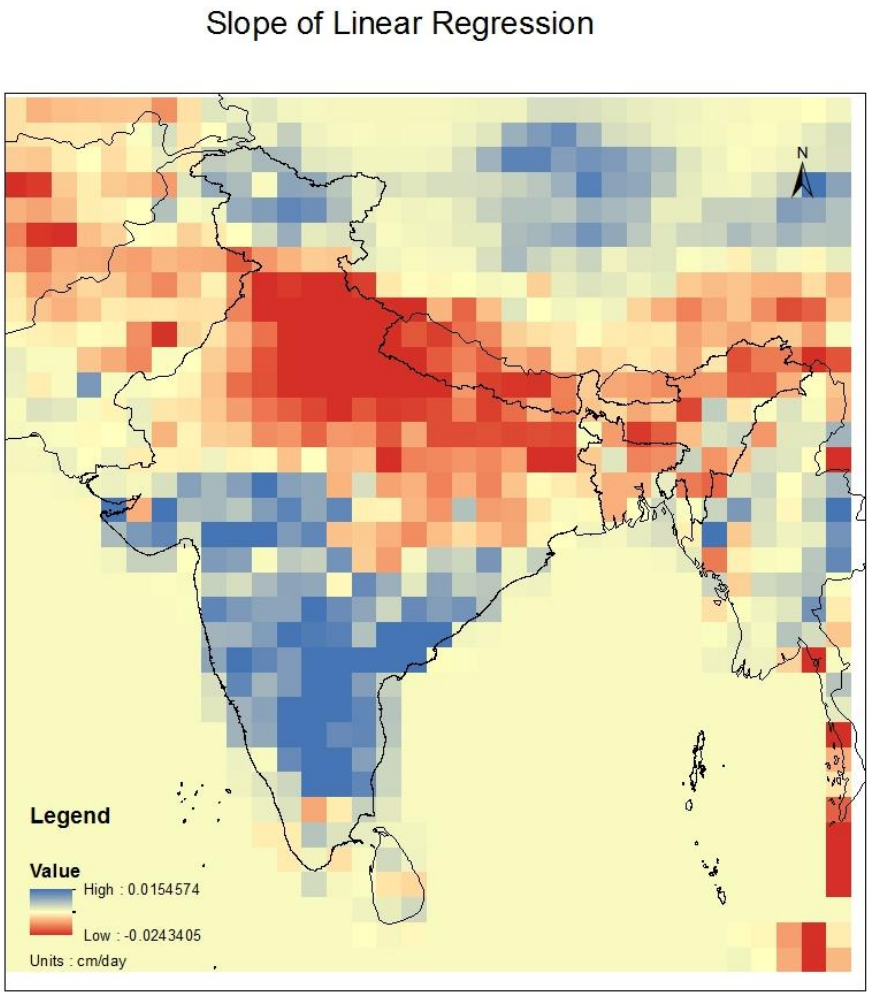


Figure 14 Linear regression results - GRACE gravity anomaly

5.3.2. Partial Autocorrelation Function Results

Partial Autocorrelation function (PACF) results show a wave pattern in southern, eastern, and northern regions of the country. The time period for repetition of the wave is roughly 12 months in these regions. These confirm that the gravity anomaly is primarily affected by the monsoons. However as the time lag increases, the autocorrelation value decreases even in the time of monsoons. This can be seen in the final time steps of the partial autocorrelation function. This decrease shows that there is a long term trend in the particular pixel. Since a perfect sinusoidal wave without long term trend would have a similar sinusoidal wave pattern in the partial auto correlation function values, the decline in the autocorrelation values indicate the presence of a trend in regions. Since the drier regions, gravity anomaly is not affected by monsoons and other seasonal hydrological patterns; the autocorrelation function yields different results in the region. This could be seen in the figure below. The figure 17 shows a typical Indian region where a 12 month cycle is seen along with a trend (either increasing or decreasing). In the drier regions of Western Pakistan, Thar desert, and the Tibetan regions, the autocorrelation function shows that there is no cycle in the region.

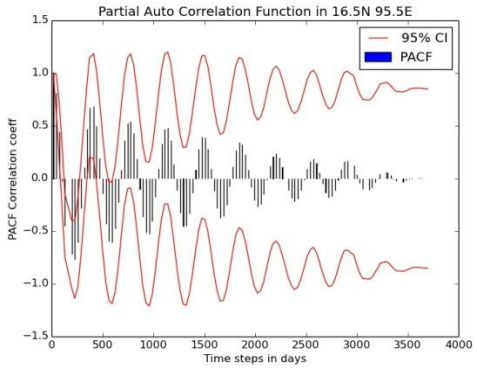


Figure 16 PACF results for Grace LWE

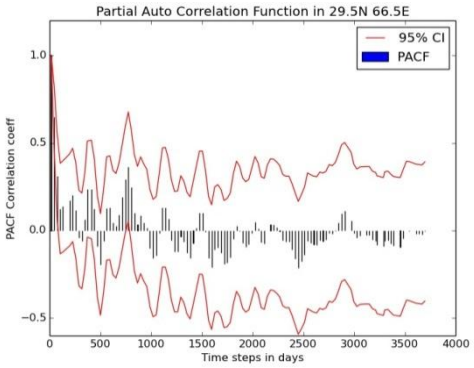


Figure 15 PACF results for Grace LWE

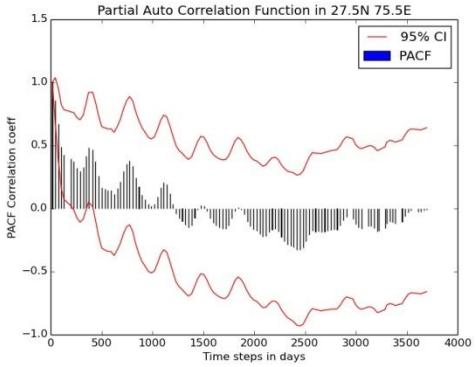


Figure 17 PACF results for Grace LWE

5.3.3. Deseasonalization

Deaseasonalization was done on the gravity dataset and the linear regression was performed on the dataset. The linear regression on the de-seasonalized dataset has similar results to those observed by the linear regression on the original dataset. Little differences have been observed in

the eastern region. These are primarily due to the fluctuations in either 2010 or in 2003 as the seasonal variations in those years could particularly affect the trends.

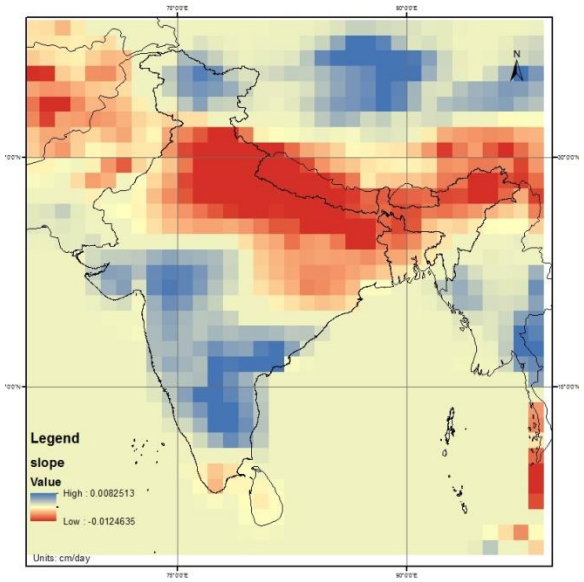


Figure 18 Linear regression results on de-seasonalized gravity data

5.3.4. Principal Component Analysis

The principal component analysis on the time series shows the temporal variation of the gravity. The first component of the PCA result shows the overall LWE thickness. The first component is supposed to show the redundancy present in the time series of the data. High LWE Thickness values are observed in the Eastern parts of India which have high rainfall and also receive water from both the Ganges-Brahmaputra river system. Lower amount of LWE thickness anomaly is observed in the western region of the study area covering the arid parts of Pakistan, Afghanistan and the Rajasthan state of India. The Thar Desert which forms a major part of Rajasthan and eastern parts of Pakistan has lower LWE thickness value. The lower LWE thickness is because of the scanty rainfall and the absence of larger rivers except the Indus valley. The Indus valley shows a slightly higher LWE thickness when compared with the drier regions around it.

The second component of PCA shows the variation of the gravity anomaly from the redundant data. This component hence indicates the long term trend in the gravity dataset. Second component results are similar to the trends observed in the linear regression trend analysis test. The third component of PCA on time series shows the variation of the gravity anomaly in seasonal scale. It signifies the variation observed in every time step from the redundant value. Hence it shows the significant values. The seasonality in the region is seen as lower component loading values in the image. Strong seasonality is observed in the Ganges-Brahmaputra delta region whereas the regions around Delhi, Haryana, Punjab, Western Uttar Pradesh and some parts of Tamil Nadu show a low seasonality. This could be due to the irrigation practiced in the region causing increase or decrease of LWE thickness. The later components tend to show the minor variations and noise in the dataset.

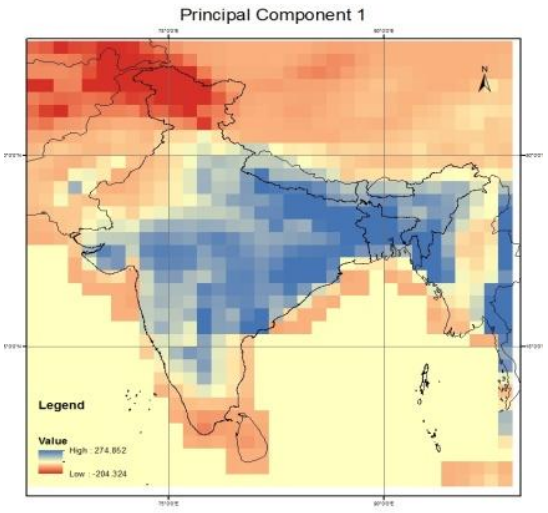


Figure 19 Principal component – gravity

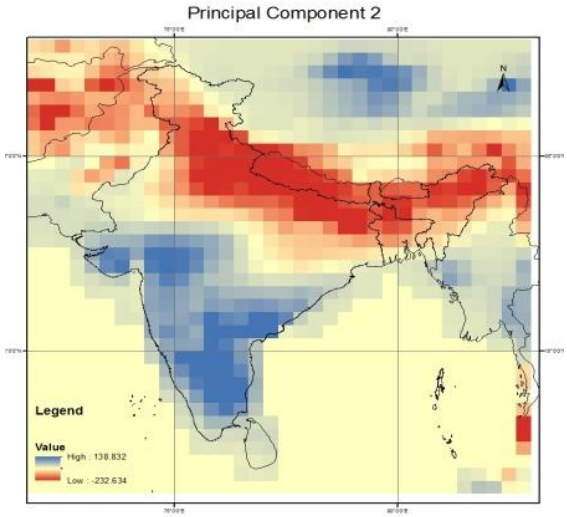


Figure 20 Principal component - gravity

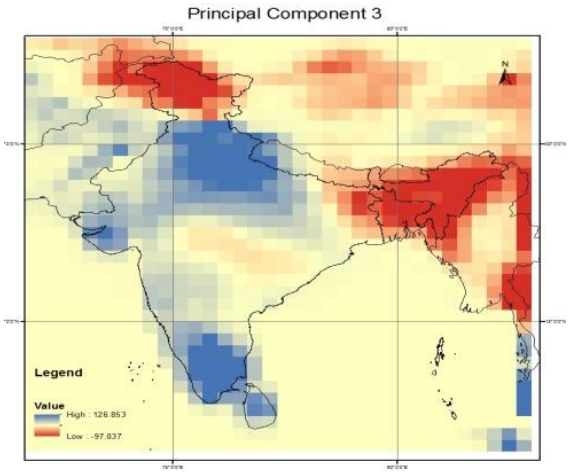


Figure 21 Principal component - gravity

5.4. Trend Analysis on Precipitation

Similar to gravity anomaly data, various trend analyses were performed on TRMM precipitation monthly resolution dataset.

5.4.1. Linear regression results

The linear regression results show that a significant positive trend exists in few parts of interior Karnataka and Bhutan. The linear regression slope indicates an increase of this much cm of rainfall in the region between the time period of 2003-2010. Significant negative trends have been observed in the Himalayan region around the borders of Uttarakhand, Kashmir, Nepal, and Tibet. The negative trend occurs on the northern slopes of the Himalayan mountain region.

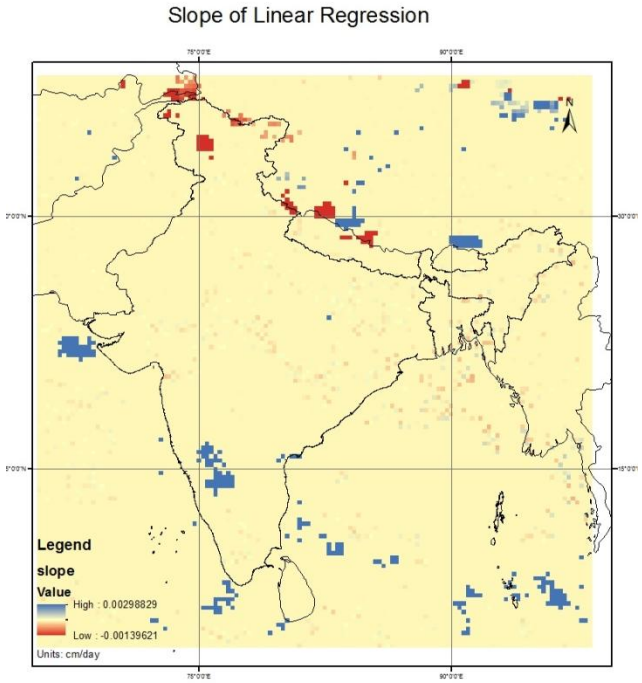


Figure 22 Linear regression on Precipitation

5.4.2. Partial Autocorrelation Function

Partial autocorrelation function on the TRMM Precipitation dataset show sinusoidal wave in most of the Indian subcontinent except the drier regions. The points for which the partial autocorrelation function are the same as the points used for the partial autocorrelation in grace dataset which is shown in figure. The cycle period for precipitation anomaly is similar to that of the LWE thickness which is around 12 months in most of the regions. The PACF results are similar for the regions receiving the south-west monsoon and the regions receiving the north-east monsoon (Tamil Nadu). Western coastal region (Malabar –Konkan belt) which receives rainfall in both the monsoon seasons has a periodic cycle of 12 months too.

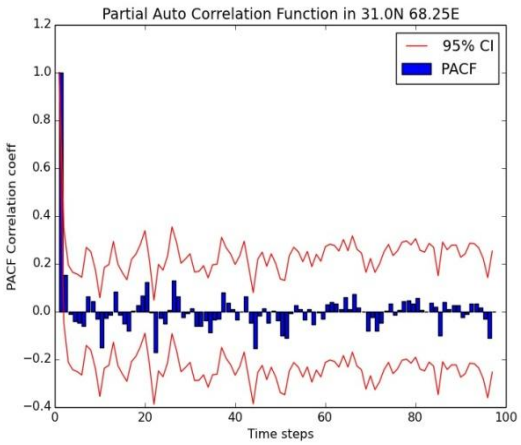


Figure 23 PACF precipitation data

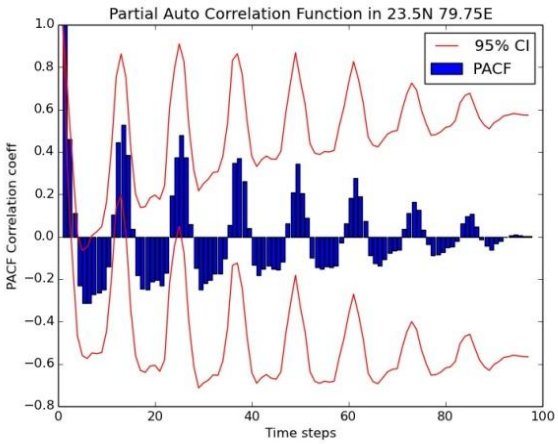


Figure 24 PACF precipitation data

5.4.3. Deseasonalization

The deseasonalized precipitation dataset was created using the rolling average method from the original TRMM precipitation Anomaly dataset. The linear regression done on the deseasonalized precipitation dataset gives a different picture than the linear regression done on the original dataset. Significant positive trends have been observed in precipitation rate in the western coastal regions, Interior Karnataka, and few parts of northern Pakistan and Eastern Tibet. Significant negative trends have been observed in the Himalayan mountain region comprising the state of Uttarakhand, and the countries of Nepal and Tibet. Negative trends have been observed in the regions of central India, Bangladesh and some parts of the Arakan mountain range in Myanmar. These results are different when compared with the linear regression results done with the original dataset. The trends in Pakistan, western coastal regions, eastern Tibet, Arakan region, central India, Gujarat and Bangladesh were not observed in the linear regression tests on the original dataset. These indicate that the seasonal variation in the years of 2003 or 2010 could be affecting the trend results arrived at in the linear regression test using the original precipitation anomaly dataset.

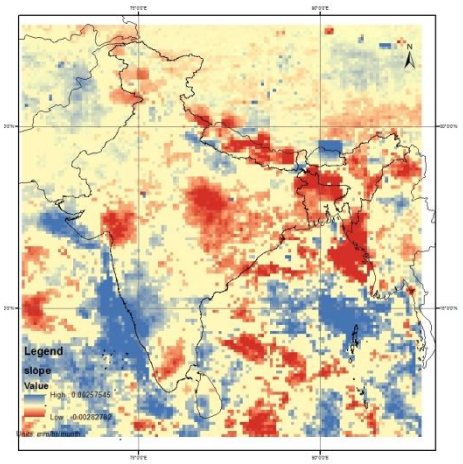


Figure 25 linear regression results of deseasonalized precipitation dataset

5.4.4. Principal Component Analysis

Principal component analysis on the TRMM precipitation anomaly dataset was done. The first component of the PCA shows the redundancy in the region. The first component indicates that Meghalaya, Arakan coast the Tennaserim coast and the Western coastal regions of India are of high rainfall regions. Regions in the southern slope of the Himalayas and the central regions of India and Myanmar have moderate amount of rainfall. The regions of Pakistan, Tibet, Afghanistan, Western states of India, and the Deccan plateau have received lower rainfall.

The second component of the PCA test on TRMM precipitation anomaly dataset shows the long term trend in the region. The second component results are similar to the results arrived at the linear regression results of the deseasonalized dataset of the TRMM precipitation anomaly. The second component results show variations from the component one exist in the regions of

Bangladesh, Arkaan coast, northern India, northern Pakistan, Himalayan region, Gujarat and central India. These indicate that a long term trend existing in these regions. The component three which shows the seasonal variation in the dataset shows that the precipitation rate over Bangladesh and North-Eastern parts of India is strongly seasonal. The southern states of India which receive rainfall in both the seasons (north east monsoon and south west monsoon) have negative loadings. The Images for PCA are annexed in the Appendix IV.

5.5. Trend Analysis on Vegetation Index

5.5.1. Linear regression results

The Enhanced Vegetation Index anomaly dataset that was created was used for the linear regression testing to find the slope. The trend results show significant positive trends in the Deccan plateau which also had a positive trend in precipitation. The increased amount of the vegetation might be due to the increased amount of rainfall. Negative trend in EVI anomaly have been observed in parts of Arakan coast indicating vegetation loss in the region. Some positive and negative trends have been observed in the Indus valley region of Pakistan and over some parts of Rajasthan.

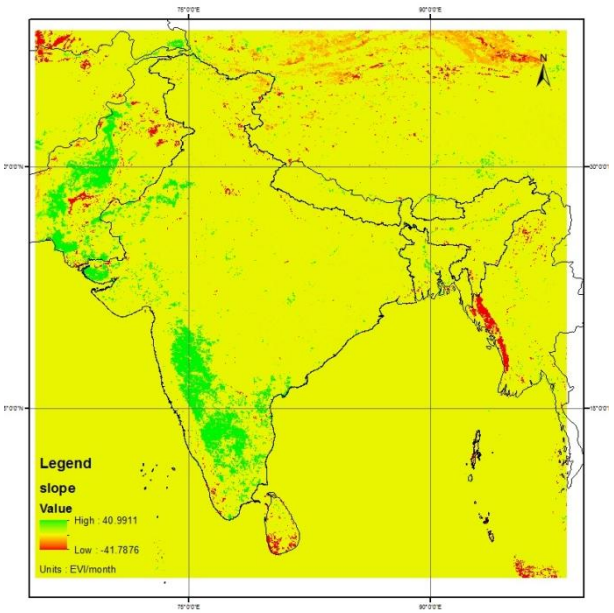


Figure 26 Linear regression results of EVI over India

5.5.2. Partial Autocorrelation Function

Partial autocorrelation function on EVI anomaly index show cyclic pattern in EVI similar to those observed in partial autocorrelation functions done on precipitation anomaly and gravitational anomaly. The EVI values tend to repeat every 12 months in most of India, Bangladesh, Srilanka, and Myanmar. However in the drier regions, EVI doesn't show any cyclical pattern.

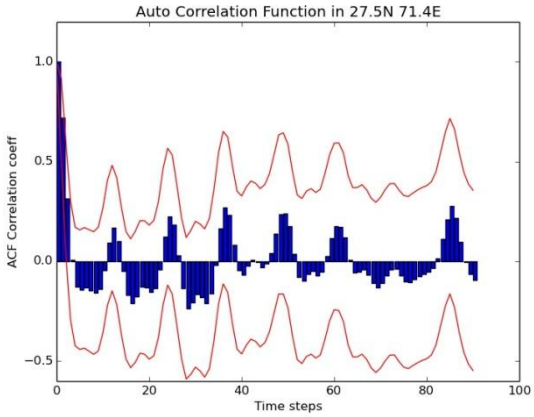


Figure 27 PACF – EVI

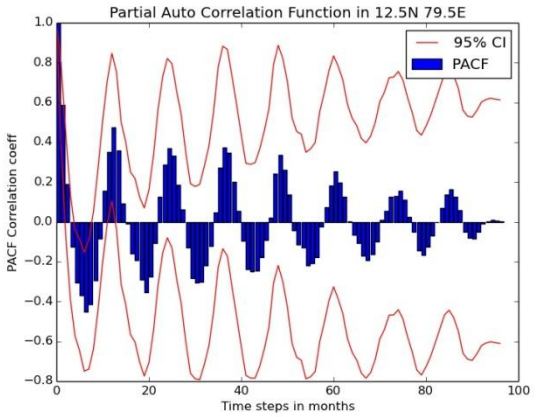


Figure 28 PACF - EVI

5.5.3. Deseasonalization

Linear regression done on the deseasonalized Enhanced Vegetation Anomaly dataset has similar results to the linear regression done on the original dataset. Minor variations have been observed in parts of Northern Pakistan, and Northern India region. These could have been due to the seasonal variations or the short term variations during the years of 2003 and 2010. Significant negative trends have been found in the Arakan coast region of Myanmar and significant positive trends have been observed in the Deccan plateau.

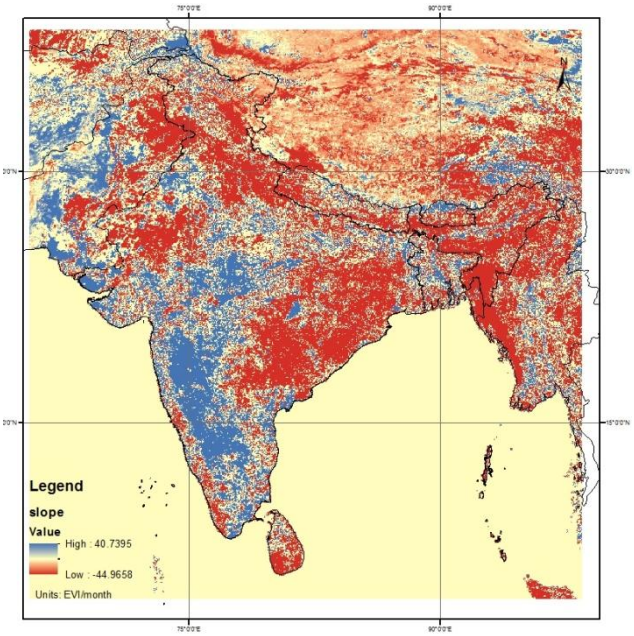


Figure 29 Linear regression on deseasonalized EVI data

5.5.4. Principal Component Analysis

PCA results on the EVI anomaly dataset show the variation of EVI dataset over the time period. The first component which indicates redundancy indicates overall vegetation amount. This indicates the place where the EVI is high all through the year. The Western coastal regions, Central India, North Eastern India and Arakan Coast of Myanmar show high loadings values for the first principal component. This indicates that these regions are greener than the rest of the Indian regions. The rain shadow regions of southern Deccan plateau, Thar Desert, Tibetan plateau, Afghanistan and North western regions of Pakistan which have arid climate are showing lesser amount of vegetation. The second component which indicates the variation from the first component shows the variation that exists from the first component. The second component has high loadings values in the Northern parts of India particularly in the states of Punjab, Haryana, western Uttar Pradesh, and Delhi. High loadings are also observed in Punjab (Pakistan), the Indus valley , the Krishna-Godavari delta regions and Sylhet region of Bangladesh. The regions which have high loadings are also some of the world's well irrigated lands. The irrigation pattern in these regions could have caused the variation. The other components tend to show the local variations and the noise in the temporal dataset. The components are displayed in the Appendix.

5.6. Trend Analysis on Snow data

5.6.1. Linear regression results:

Linear regression was performed on the SWE anomaly dataset. Significant positive trends have been observed in the Himalayan mountain chain in the regions of Himachal Pradesh and Uttarakhand Significant negative trends have been observed over the Eastern Himalayas in the Bhutan region and the Brahmaputra. Negative trends have been observed in the regions of Tien Shan Mountains which border the northern parts of Tibetan plateau. No trends have been observed over the Hindu Kush Mountains and the Karakoram mountains.

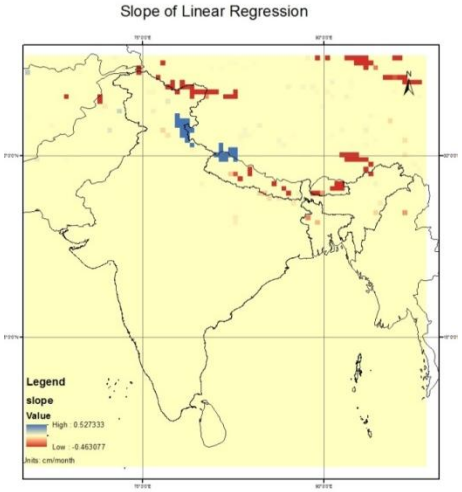


Figure 30 Linear regression results- SWE

5.6.2. Partial Autocorrelation Function

Partial autocorrelation function performed on snow water equivalent indicates the seasonal structure in the snow water equivalent. In occasionally snowing regions in Tibet the autocorrelation function provide sudden peaks which indicate that the snow accumulation in the region has been rare and not annual. However in Himalayan Mountains, Karakoram and the Hindukush regions where it snows every year, a cycle pattern is formed. Roughly every 12 months the snow water equivalent repeats itself and in most of the regions, a long term trend exist causing the fall in the value of the autocorrelation function.

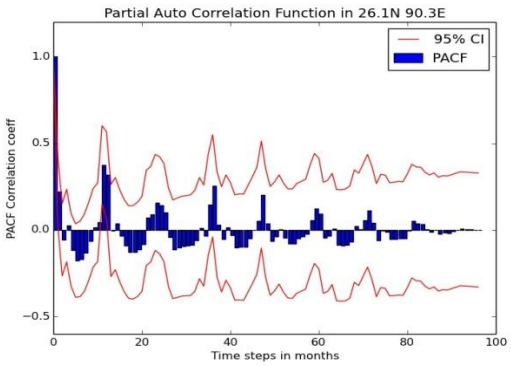


Figure 31 PACF - SWE data

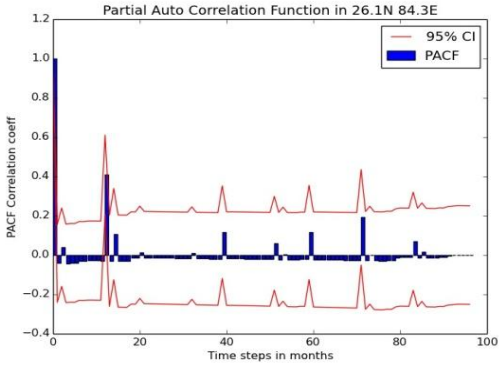


Figure 32 PACF - SWE data

5.6.3. Deseasonalized dataset

The linear regression performed on the deseasonalized snow water equivalent anomaly dataset indicates a significant negative trend in the Hindu Kush-Karakoram region and the Eastern Himalayas along the border between Arunachal Pradesh and Tibet. Significant positive trend has been observed in the Himachal and Uttarkhand-Nepal sections of the Himalayan mountain range. The significant negative trend along the Hindukush-Karakoram region and the Eastern Himalaya section was not observed when the original SWE anomaly dataset was used.

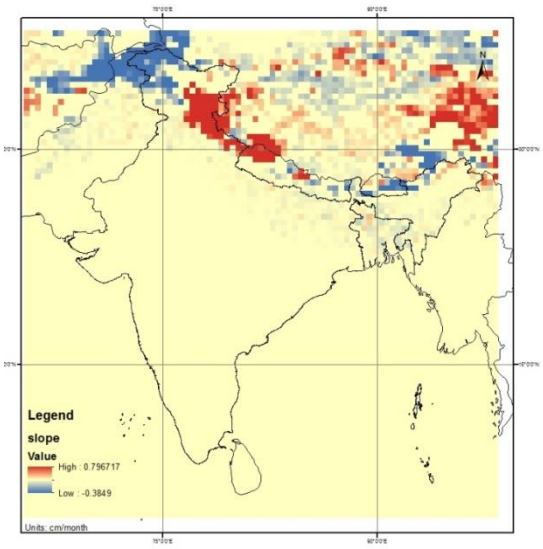


Figure 33 Linear regression results on SWE data

5.6.4. Principal Component Analysis

The Principal component analysis was performed with the SWE Anomaly dataset for trend analysis. The first component gives the overall snow content in the region. It provides the regions which have significant amount of snow in all the years. The first component has high high loading values in the Eastern Himalayas and the Karakoram- Pamir section. It also has high loading values over the Central and the Western Himalayas. Moderate loading values in first component over the Tibetan plateau indicate that the region faces snow cover. The component results are added in the appendix

5.7. Trend Analysis on Urbanization

5.7.1. Linear regression results

Linear regression on the urbanization index anomaly shows that almost all the cities in the region have been urbanizing very fast. The outward growth of cities has been observed. In most of the cities, the central parts of the city are observed of having either a negative trend or no trend at all. The city with an exception for this case is Kabul. This might be due to the fact that the city witnessed lesser growth before the 2001 Afghanistan war and has been growing fast ever since. The negative trend in the city centre might be due to the using Elvidge coefficients rather than the actual sensor calibration coefficients. However the outward growth is clearly observed with the cities tending to grow radially. Some rural regions of Northern India have been observed of having significant negative trend.

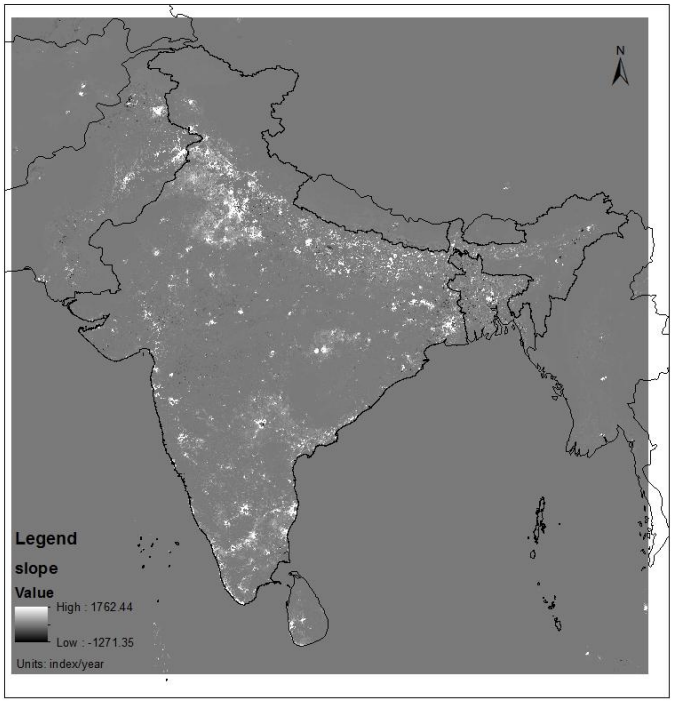


Figure 34 Linear regression results - Urbanization

5.8. Effects of MSTUP on trend analysis

Variation of spatial and temporal resolutions causes varying effects on trend analysis. The linear regression analysis was done on the data for the decreased spatial resolutions and the temporal resolutions of gravity anomaly. The percentage of pixels having positive trend tend to decrease when the temporal resolution is decreased whereas the percentage of pixels having negative trend increases when the temporal resolution is decreased. Similar patterns existed when the resolution is decreased for the other parameters such as precipitation, vegetation, snow water equivalent thickness, and urbanization index anomaly (refer appendix II). The results are effectively indicating that the change in the spatial and temporal resolutions of the datasets affect the trend analysis in varying manner. The reason for the decrease of the significant positive trends and increase of the significant negative trends is unclear.

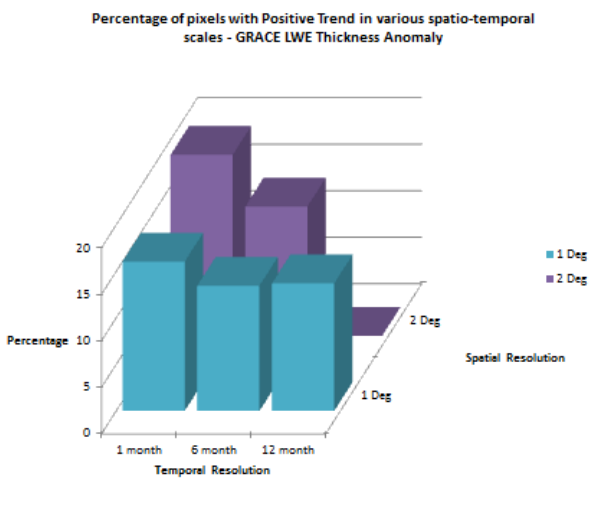


Figure 35 Percentage of pixels having Positive trend - gravity data

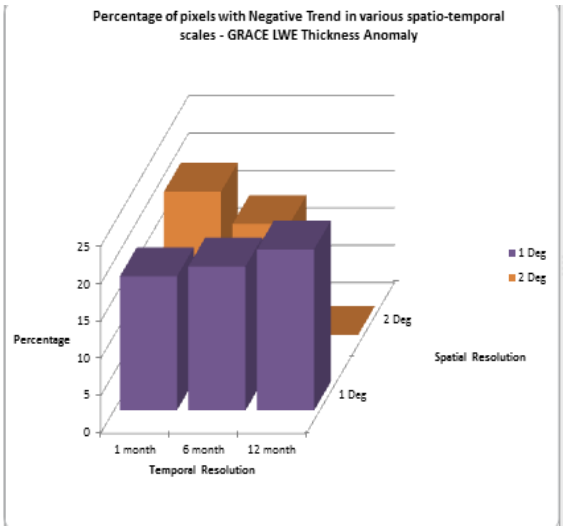


Figure 36Percentage of pixels having Negative trend - gravity data

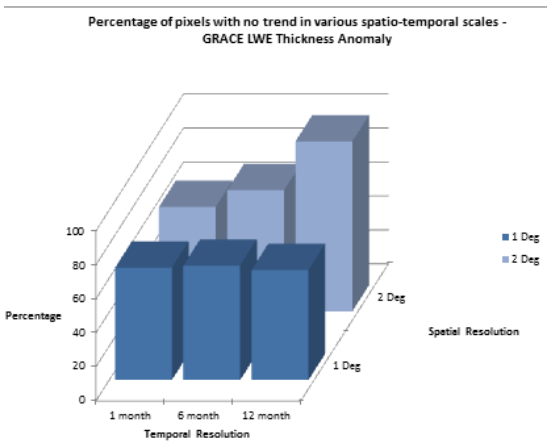


Figure 37 Percentage of pixels having No trend - gravity data

5.9. Relationship analysis

5.9.1. Model

The Artificial Neural Network model was created with the structure as discussed in the chapter earlier. Based on the computational speed available, the number of training iterations was taken as 100. After training for 100 times, the training for which the root mean square was the least was selected. The root mean square values of all the 100 trainings were between 2.8 and 4.5 (refer appendix). The RMS value oscillated between these since the ANN is a black box model where the initial conditions change for a training iteration. The lowest RMS value was approximately 2.89. The trained structure for which the value was 2.89 was selected as the final model structure. The structure of the model was stored in XML format for later use. The stored XML file containing the model parameters, model structure is annexed in appendix I.

Using the model parameters, the model is re activated with all the explanatory variables for predicting the values of LWE Thickness anomaly. The predicted LWE thickness dataset does save the temporal variation and the sinusoidal pattern in many pixels. The LWE Thickness predicted for the year 2008 January is displayed in the figure below. Figure 2 shows the actual LWE thickness anomaly value from GRACE data. The differences observed show that there is no spatial correlation. However when temporally plotted as a time series, the sinusoidal pattern created due to the monsoonal pattern is clearly observed.

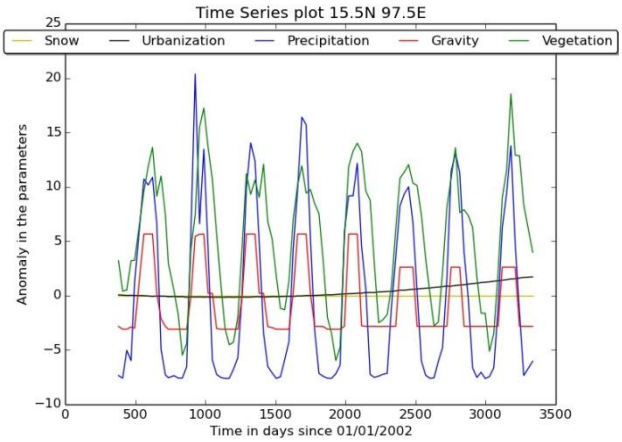


Figure 38 time series plot of the predicted data using the model. red line indicates modelled gravity data

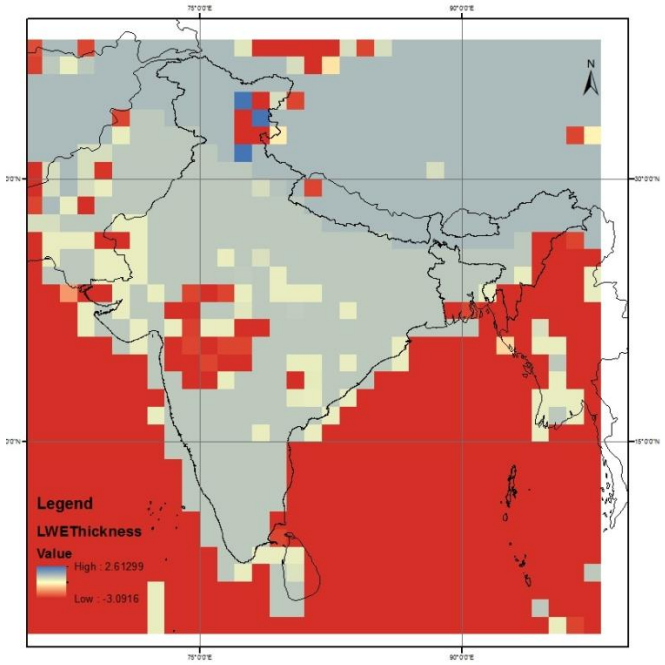


Figure 39 gravity data when predicted using the modelled data

5.9.2. Monte Carlo simulation results for understanding error propagation

The Monte-Carlo simulation method for understanding error propagation is done using the model created earlier. The randomly selected inputs from a Gaussian distribution of the input values based on the mean and the standard deviation of the parameters were used for the Monte-Carlo simulation. 100 randomly selected values were used for understanding the error propagation in the model. Even though the inputs follow a Gaussian distribution, the output values don't have a Gaussian distribution. The output values don't follow any proper distribution and have "islands" of values as shown in figure . This shows that the Artificial Neural Network algorithm behaves similar to the classification algorithm. The minor changes in the input values don't give rise to a newer value as seen in the figure. Because of this nature, the values predicted using the random inputs don't have much difference with the values predicted using the mean of the input data. This effectively concludes that the errors in the input data (explanatory variables) don't have much effect on the modelled data (explained variable).

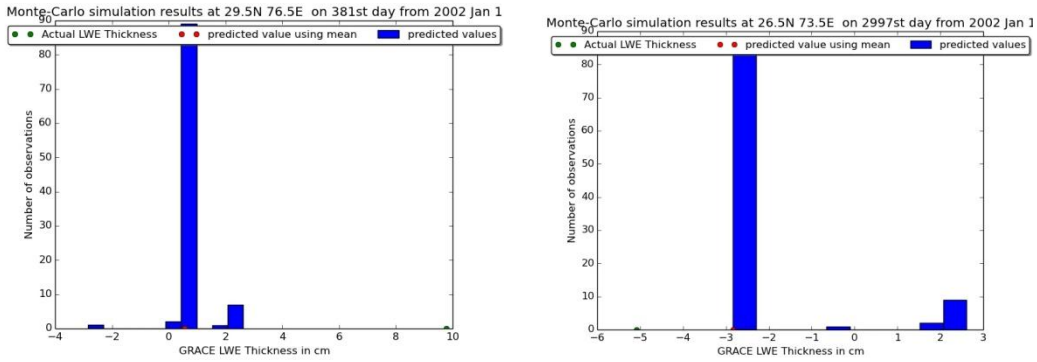


Figure 40 Monte Carlo simulation results

5.9.3. Effect of parameters

Individual effects of parameters on the model were found by activating the model without the respective parameter. When activated without a particular parameter, the RMS value tends to increase more. The model was then activated without urbanization index anomaly. The RMS value was found to be 9.7 cm of LWE Thickness. The model was then activated without the SWE anomaly parameter and the RMS value was observed to be 10.1 cm of LWE Thickness. When the model was activated without the EVI anomaly parameter the RMS value rose to 10.4 cm of LWE Thickness. When the model was activated without the EVI anomaly parameter, the RMS value was 10.34 cm of LWE thickness. The RMS value was found to be greatest when the EVI anomaly parameter was not used for the activation of the ANN model. The model hence is influenced more by EVI anomaly than the other parameters. This can also be seen in the plot shown in figure in appendix, where the synchronization is greater between EVI anomaly and LWE Thickness. This doesn't necessarily indicate the influence of biomass on the LWE thickness, however it seems to correlate. When concerned with the model, EVI anomaly tends to influence the most.

5.9.4. Relationship between parameters

5.9.4.1. Correlation Coefficient results

The spearman rho correlation coefficient was calculated between GRACE LWE Thickness and other parameters such as the EVI anomaly, SWE Anomaly, Precipitation Anomaly, Urbanization Index anomaly.

Precipitation and LWE Thickness anomaly have high positive correlation over the entire study area. There are few drier regions in the Western Pakistan, Thar Desert, and Tibet that don't have significant positive correlation. There are isolated pixels of Negative correlation however these are mostly noise or there are corruption from Sea. In the south, there is no significant positive trend over the state of Kerala. The region has significant amount of backwaters which store lots of water. This could have caused the absence of positive correlation in the region. However over the entire Indian region, the spearman correlation coefficient between LWE Thickness anomaly and the Precipitation Anomaly was calculated.

Enhanced Vegetation Index anomaly and LWE Thickness anomaly have very high positive correlation in most parts of mainland India. However, significant but slight positive correlation in the Gangetic valley and delta regions of few rivers has been observed. In these places the river water, irrigation could have caused the disparity in in the correlation. Significant negative correlation has been observed in the Hindu Kush-Karakoram region and also parts of Tibetan plateau. The negative correlation occurs in places of lower overall vegetation content which could was observed in the first component of principal component analysis done on the vegetation index anomaly data. The regions are also regions of intense snow accumulation as seen in the component one of principal component analysis done on snow water equivalent data. Even though significant positive correlation between vegetation and LWE thickness has been observed, the causation cannot be confirmed as during seasons of high precipitation, the vegetation index also rises. Since the presence of water and vegetation are highly correlated to

each other, vegetation cannot be inferred to cause the trends in LWE Thickness. But it could be inferred with confidence that vegetation positively correlates with LWE thickness.

Snow water equivalent thickness and LWE thickness anomaly correlate have a very significant positive correlation in the Hindukush- Pamir region. These regions are dry and lack in vegetation content. Only Snow content varies in the region, hence they cause significant positive correlation. However in the eastern Himalayas, central Himalayas, and the southern parts of Tibetan plateau, there exists a significant negative correlation between snow content and gravity anomaly. This is probably due to the effect of other parameters such as precipitation, glacial movement. The region is also home to many glaciers and is the source of three major rivers (Indus, Brahmaputra and Ganges) and all of their tributaries. In these regions, precipitation and other parameters also would have affected the correlation between snow and LWE thickness. The time series plot in fig shows such a pixel where both precipitation and snow are present. Snow water equivalent tends to have significant positive correlation coefficient with LWE thickness when the other parameters are absent in the region.

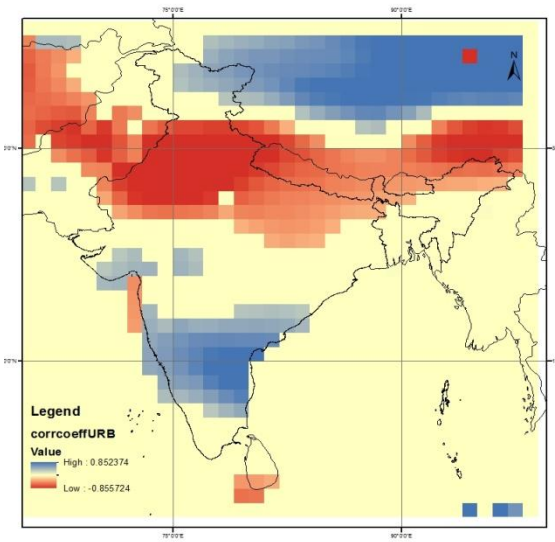


Figure 41 Correlation coefficient - TRMM vs. Gravity

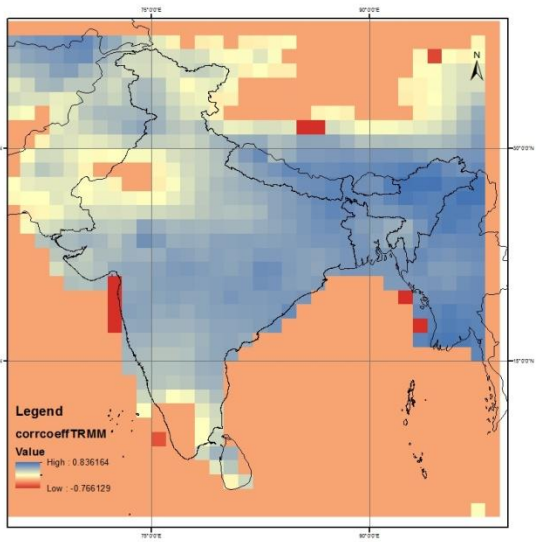


Figure 42 Correlation coefficient - Urbanization Index vs. Gravity

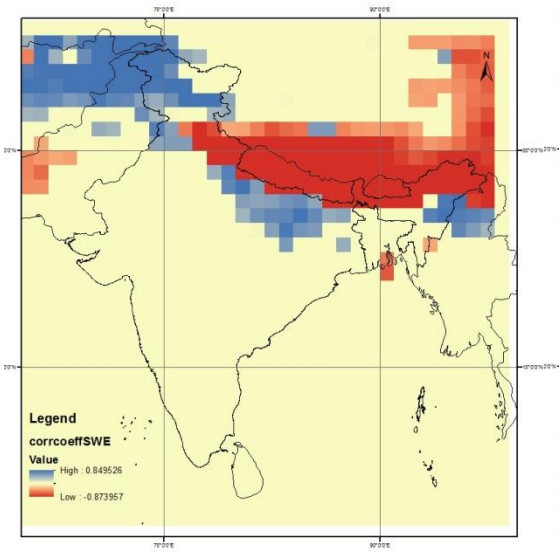


Figure 43 Correlation coefficient - SWE vs. Gravity

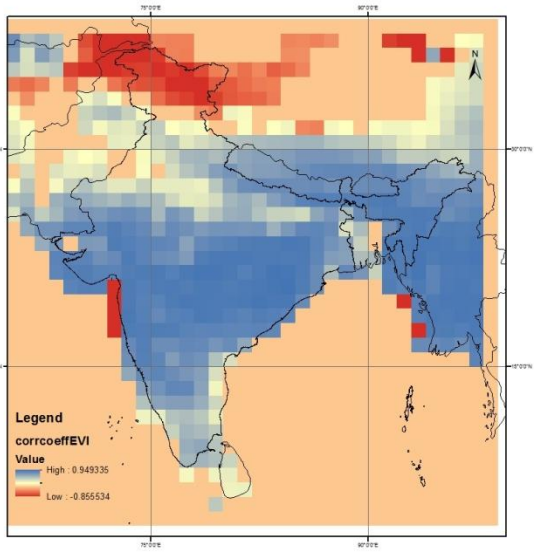
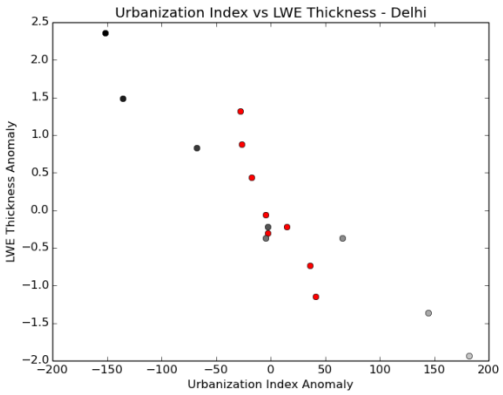


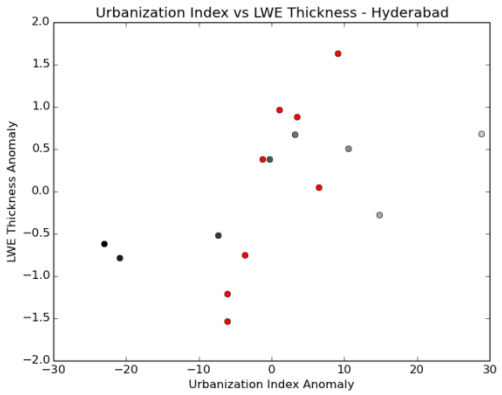
Figure 44 Correlation coefficient - Vegetation vs. Gravity

5.9.4.2. Effect of Urbanization

The spearman correlation coefficient map between urbanization Index and gravity anomaly shows that there are significant negative correlation in Gangetic valley, South Eastern Tibet and Indus valley. There are significant negative correlation in south India and the central and northern Tibetan plateau. Such a correlation exists because in most of the regions urbanization has been having significant positive trend which has been explained in chapter. Because of the presence of the universal positive trend in urbanization anomaly, the correlation coefficient map produces a result similar to the linear trend result of gravity anomaly. Hence the correlation coefficient map for the study between the urbanization index and the LWE thickness has no use. The effect of urbanization can be studied only with the comparative study between pixels which have high urbanization pattern and the pixels around the respective pixel which don't have very high urbanization. When the urbanization index is plotted against the gravity anomaly values in the pixels containing major cities and the rural region surrounding it, the relationship is understood. For example the figure shows the plot between urbanization index anomaly for the pixel containing Delhi city and the mean of the neighbouring pixels. For this plot the arithmetic mean of the surrounding 8 pixels are used. This plot shows that the urbanization index anomaly has been increasing very fast in the pixel containing the Delhi metropolitan area but not so high in the pixels around Delhi. The gravity anomaly in Delhi is observed to be having a more negative slope when compared with the pixels around it. This concludes that the Urbanization is causing decrease in the gravity anomaly. The next figure shows the city of Hyderabad, where positive trend had been observed in gravity anomaly. When LWE thickness anomaly is plotted

against urbanization anomaly, it has been observed that even though both have a positive trend, the rapid urbanization is causing lower positive trend in the city of Hyderabad than the surrounding pixels. Thus it can be clearly said that urbanization is negatively correlating with LWE Thickness anomaly. Even though addition of new buildings in the regions should positively correlate with the LWE Thickness anomaly, negative correlation has been observed. This negative correlation could be due to the pumping of ground water in the cities for domestic purposes. Since the population of these cities have been increasing at a rapid pace, the ground water pumping would have increased causing the negative correlation. Other results are displayed in appendix V. In the figure below, the values of LWE thickness are plotted against the Urbanization anomaly. The black colour indicates the year being 2003 and grey indicates 2010. The red colour values are the values corresponding to the mean of the surrounding pixels whereas the grey shade values are the values corresponding to the central pixel where the urbanization is happening.





5.9.5. Effect of MSTUP on the model

The ANN model was created with the coarser spatial and temporal resolutions of all the parameters. The model gets created and the predictions are susceptible to the change of the resolution. Since the model gets formed with a lesser number of samples, it gets over fit. The root mean square of the models get reduced because of the over fit and lesser number of samples. Hence a comparison study between the model created with the original dataset and the up-scaled dataset becomes impossible. However the models were created with the upscaled data. The parameters calculated are attached in appendix VI.

5.10. Effect of MSTUP on correlation coefficient

Correlation coefficients between the LWE thickness and various parameters are affected by the coarser spatial and temporal resolutions. The decreasing spatial resolution doesn't impact the positive correlation gravity anomaly and the precipitation rate. However, the temporal resolution seems to decrease the percentage of pixels which have the positive correlation. On the other

hand, the percentage of pixels with negative correlation tends to reduce when the spatial resolution is decreased and the percentage of pixels increase when the temporal resolution is decreased. The effect of various spatial and temporal resolutions on the correlation coefficient varies with the parameters. The results are attached in the Appendix III.

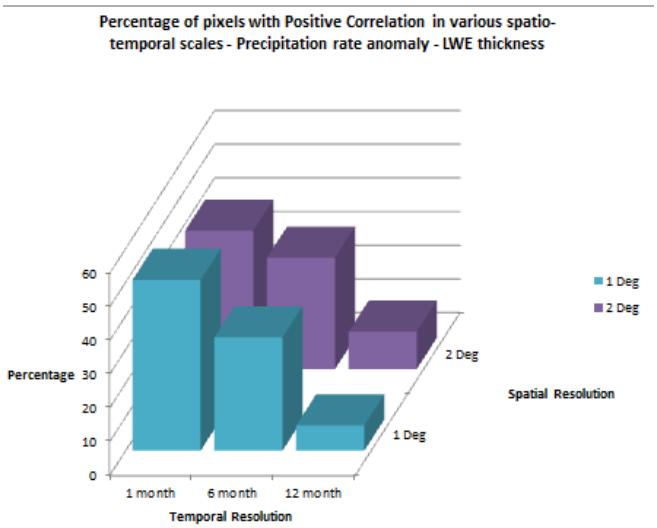


Figure 45 Percentage of pixels with positive correlation in various spatio-temporal scales (Gravity vs Precipitation)

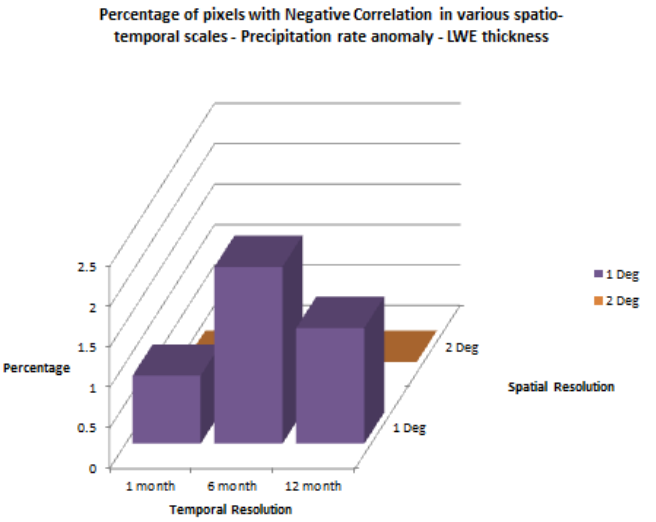


Figure 46 Percentage of pixels with Negative correlation in various spatio-temporal scales

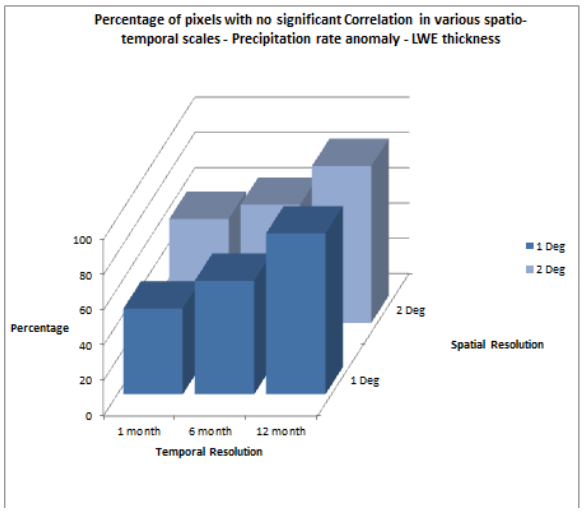
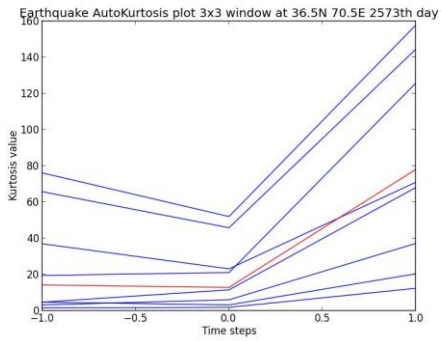
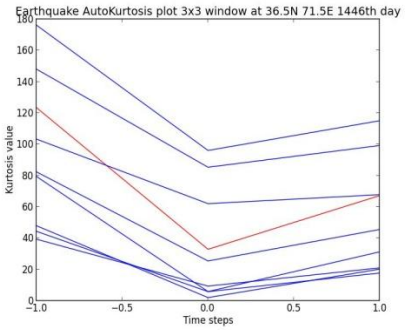


Figure 47 Percentage of pixels with no significant correlation in various spatio-temporal scales

5.11. Earthquake Analysis – Moving Kurtosis results

The moving kurtosis- comparison method shows that there is no significant earthquake signal in the level 3 LWE thickness anomaly data. The plot shows the moving kurtosis values in a pixel where earthquake had occurred along with the moving kurtosis values of the neighbouring pixels. The window size is taken as 3x3. The movement of the kurtosis value which is supposed to indicate sudden peakedness don't have any deviation from pattern when compared with the neighbouring pixels. This result convincingly says that for the earthquakes that occurred in india between 2003 and 2010, there is no signal in the LWE Thickness dataset. The Sumatra earthquake (mag -9.5) which occurred on 2004 December 26 occurred off the coast of Sumatra and had impact on gravity anomaly data. However, in the current research the grace land data has been used which doesn't contain the pixel where the epicentre of the earthquake lies. Also the third level GRACE data used in this research had been subject to various spatial smoothening filters. These spatial smoothening filters could have removed the earthquake signals from the third level GRACE data. The figures displayed below show the rolling kurtosis on two earthquakes. The rest of the results are annexed in appendix.





6. CONCLUSION AND RECOMMENDATIONS

6.1. Conclusion

1) *Is there any specific trend in the variation of the gravity measures over India?*
   Yes. Based on the linear regression test, partial autocorrelation results and the principal component analysis results, it could be concluded that there are specific trends of gravity anomaly over India. Across the northern Indian plains, there exists a negative trend and in the Deccan plateau there is a positive trend. Gravity anomaly varies seasonally in a cyclic manner in most parts of India. However in drier regions of the country, the cyclic pattern is absent.
2) *Do the variations in the parameters i.e. precipitation, vegetation, urbanization, and snow cover affect the trend in gravity? (If so, how?)*
   Yes. The gravitational anomaly pattern follows the same pattern as the precipitation and the vegetation datasets. There is high positive correlation between these parameters and the gravity anomaly. The correlation coefficient maps conclusively say that these parameters correlate a lot. Snow amount has high positive correlation the gravity in regions where the influence of other parameters is lower. Urbanization affects the long term trend of gravity in various regions. The negative correlation pattern could be observed when checked in pixel level.

3) *How does error propagate through the model and which parameters influence the model?*
   An Artificial Neural Network model was created with a root mean square of 2.89 cm of LWE thickness. However the values of error in the input value least affects the outcome of the model. The ANN model behaves similar to a classifier algorithm. Hence the effect of the error from the input data is minimal. The distribution pattern of the monte carlo simulation doesn't yield any proper mathematical distribution. The parameter that has influences the model the most is enhanced vegetation index. Vegetation might be likely to have behaved like a proxy for ground water and surface water storage. Earlier researches have concluded that the effect of bio mass on the gravity is negligible when compared with water level.
4) *Do earthquakes affect the trend and models?*
   No. The there is no significant effect of earthquakes on the GRACE gravity level three anomaly data. Since there is no presence of earthquake effect on the gravity anomaly data, it could be concluded that it doesn't affect the trend or the model. Earlier researches have suggested that earthquakes affect gravity anomaly, however the data used is different. The

third level GRACE gravity data after undergoing spatial and Gaussian filters would have lost the earthquake signal in it.

5) *How does the use of various spatial and temporal resolutions of the various parameters affect the trend, correlation (if any), and the model?*
   The trend analysis results are significantly affected by the spatial and temporal resolutions. Even though a common pattern amongst all the parameters doesn't exist, individual pattern exist in every parameter. In case of GRACE gravity data, as the temporal resolution decreases, the percentage of regions with significant positive trend decreases. However in case of significant negative trend regions, as the temporal resolution decreases the percentage of significant negative trend regions increase however in the lower spatial resolution, the percentage of significant negative trend regions decrease. This shows that the spatial and temporal resolutions create complexity in trend analysis results. In the case of the correlation results between gravity anomaly and the precipitation anomaly, the percentage of positively correlating regions decrease as the temporal resolution is decreased whereas there is no significant decrease as the spatial resolution is decreased. The percentage of significantly negatively correlating regions doesn't have a uniform pattern even though the results arrived at when using different spatial resolutions. The model is least likely to be affected because of the nature of the artificial neural network. The ANN model behaves like a classifier, hence when created with lower spatial and temporal resolutions it still tends to over fit the values. Since the number of samples is lesser for lower spatial and temporal resolutions, root mean square tends to come down.

6.2. Recommendations

Based on the research, it could be said that the trend analysis results, have been significantly affected by the variation of spatial and temporal resolutions of the datasets. Care should be taken before using which combination of spatial and temporal resolution that could be used for the dataset. Since the model was created successfully with precipitation, vegetation, urbanization, and snow cover; further parameters influenceing the values like plate tectonics (Station GPS data can be used), ground water and soil erosion should be considered. The influence of these parameters on the gravity values need to be studied for further understanding of the mass changes over the surface of Earth. For the study of earthquake effects on gravity, second level GRACE gravity anomaly data needs to be used rather than the third level gravity anomaly data. However further study with GRACE gravity data would be significantly helpful in finding the understanding the hydrological cycle as the values of gravity correlate positively with the water cycle.

Modifiable Spatio-Temporal Unit Problem has been found to influence the trends and correlations. Other datasets with various spatial and temporal resolutions should be

studied and also a model should be attempted with it to identify the optimal spatial and temporal resolutions for analysis in geophysical phenomena. The usage of various resolutions of data and higher resolutions of data require better computing algorithms and computing power. Better machine learning algorithms and optimization programs should be devised for quicker and better outputs.

REFERENCES

\cCöltekin, Arzu, Stefano De Sabbata, Corina Willi, Irene Vontobel, Sebastian Pfister, Matthias Kuhn, and Martin Lacayo. 2011. "Modifiable Temporal Unit Problem." In *ISPRS/ICA Workshop "Persistent Problems in Geographic visualization"(ICC2011), Paris, France*. Vol. 2. http://www.geo.unizh.ch/~sdesabba/docs/ModifiableTemporalUnitProblem.pdf.

Abercrombie, Rachel E., Michael Antolik, and Göran Ekström. 2003. "The June 2000 Mw 7.9 Earthquakes South of Sumatra: Deformation in the India–Australia Plate." *Journal of Geophysical Research: Solid Earth (1978–2012)* 108 (B1): ESE 6–1–ESE 6–16.

Bajracharya, Samjwal R., Pradeep Mool, and Basanta Raj Shrestha. 2007. *Impact of Climate Change on Himalayan Glaciers and Glacial Lakes: Case Studies on GLOF and Associated Hazards in Nepal and Bhutan*. International Centre for Integrated Mountain Development Kathmandu.

Bolch, T., A. Kulkarni, A. Kääb, C. Huggel, F. Paul, J. G. Cogley, H. Frey, J. S. Kargel, K. Fujita, and M. Scheel. 2012. "The State and Fate of Himalayan Glaciers." *Science* 336 (6079): 310–14.

Bondre, Ninad R., Raymond A. Duraiswami, and Gauri Dole. 2004. "Morphology and Emplacement of Flows from the Deccan Volcanic Province, India." *Bulletin of Volcanology* 66 (1): 29–45.

Box, George EP, Gwilym M. Jenkins, and Gregory C. Reinsel. 2013. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

"Census of India : Population Enumeration Data (Final Population)." 2014. Accessed March 9. http://www.censusindia.gov.in/2011census/population_enumeration.aspx.

"Census of India : Provisional Population Totals Paper 1 of 2011 : NCT of Delhi." 2014. Accessed March 8. http://www.censusindia.gov.in/2011-prov-results/prov_data_products_delhi.html.

"Correlation Coefficients." 2014. Accessed March 9. http://www.andrews.edu/~calkins/math/edrm611/edrm05.htm.

Dark, Shawna J., and Danielle Bram. 2007. "The Modifiable Areal Unit Problem (MAUP) in Physical Geography." *Progress in Physical Geography* 31 (5): 471–79.

De Beurs, K. M., and G. M. Henebry. 2005. "A Statistical Framework for the Analysis of Long Image Time Series." *International Journal of Remote Sensing* 26 (8): 1551–73.

"Earth Observation Group - Defense Meteorological Satellite Progam, Boulder | Ngdc.noaa.gov." 2013. Accessed August 27. http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html.

Elvidge, Christopher D., Kimberly E. Baugh, Eric A. Kihn, Herbert W. Kroehl, and Ethan R. Davis. 1997. "Mapping City Lights with Nighttime Data from the DMSP Operational Linescan System." *Photogrammetric Engineering and Remote Sensing* 63 (6): 727–34.

FIELLER, Edgar C., Herman O. HARTLEY, and Egon S. PEARSON. 1957. "Tests for Rank Correlation Coefficients. I." *Biometrika* 44 (3-4): 470–81.

Fotheringham, A. Stewart, and David WS Wong. 1991. "The Modifiable Areal Unit Problem in Multivariate Statistical Analysis." *Environment and Planning A* 23 (7): 1025–44.

"GRACE Tellus." 2013. Accessed August 27. http://grace.jpl.nasa.gov/.

"GRACE Tellus: GRACE MONTHLY MASS GRIDS - OVERVIEW." 2013. Accessed August 27. http://grace.jpl.nasa.gov/data/gracemonthlymassgridsoverview/.

"GRACE Tellus: GRACE 'Months.'" 2013. Accessed August 27. http://grace.jpl.nasa.gov/data/GraceMonths/.

"Gridded Population of the World (GPW), v3 | SEDAC." 2013. Accessed August 27. http://sedac.ciesin.columbia.edu/data/collection/gpw-v3.

Gunter, Brian, T. Urban, R. Riva, M. Helsen, R. Harpold, S. Poole, P. Nagel, B. Schutz, and B. Tapley. 2009. "A Comparison of Coincident GRACE and ICESat Data over Antarctica." *Journal of Geodesy* 83 (11): 1051–60.

Hall, Dorothy K., George A. Riggs, Vincent V. Salomonson, J. S. Barton, K. Casey, J. Y. L. Chien, N. E. DiGirolamo, A. G. Klein, H. W. Powell, and A. B. Tait. 2001. "Algorithm Theoretical Basis Document (ATBD) for the MODIS Snow and Sea Ice-Mapping Algorithms." *NASA GSFC, September*. https://eospso.gsfc.nasa.gov/sites/default/files/atbd/atbd_mod10.pdf.

Han, Shin-Chan, Jeanne Sauber, Scott B. Luthcke, Chen Ji, and Fred F. Pollitz. 2008. "Implications of Postseismic Gravity Change Following the Great 2004 Sumatra-Andaman Earthquake from the

Regional Harmonic Analysis of GRACE Intersatellite Tracking Data." *Journal of Geophysical Research* 113 (B11). doi:10.1029/2008JB005705. http://doi.wiley.com/10.1029/2008JB005705.

Hartmann, Dennis L., and Marc L. Michelsen. 1989. "Intraseasonal Periodicities in Indian Rainfall." *Journal of the Atmospheric Sciences* 46 (18): 2838–62.

Hasnain, SYED IQBAL. 2002. "Himalayan Glaciers Meltdown: Impact on South Asian Rivers." *International Association of Hydrological Sciences, Publication* 274: 417–23.

Helsel, Dennis R., and Robert M. Hirsch. 1992. *Statistical Methods in Water Resources*. Vol. 49. Elsevier.

Huete, A., Kamel Didan, Tomoaki Miura, E. Patricia Rodriguez, Xiang Gao, and Laerte G. Ferreira. 2002. "Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices." *Remote Sensing of Environment* 83 (1): 195–213.

Huffman, George J., Robert F. Adler, David T. Bolvin, Guojun Gu, Eric J. Nelkin, Kenneth P. Bowman, Yang Hong, Erich F. Stocker, and David B. Wolff. 2007. "The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales." *Journal of Hydrometeorology* 8 (1).

Jelinski, Dennis E., and Jianguo Wu. 1996. "The Modifiable Areal Unit Problem and Implications for Landscape Ecology." *Landscape Ecology* 11 (3): 129–40.

Jones, Gavin W., and Pravin Visaria. 1997. *Urbanization in Large Developing Countries: China, Indonesia, Brazil, and India*. Oxford University Press.

Khattri, K. M., and A. K. Tyagi. 1983. "Seismicity Patterns in the Himalayan Plate Boundary and Identification of the Areas of High Seismic Potential." *Tectonophysics* 96 (3): 281–97.

Majumdar, T. J., and R. Bhattacharyya. 2010. "Generation and Study of Satellite Gravity over Gujarat, India and Their Possible Correlation with Earthquake Occurrences." *Geocarto International* 25 (4): 269–80.

Maurya, Ravi. "Effects of the Modifiable Temporal Unit Problem on the Results of the Trend Analysis of Climatic Forcing and Satellite Derived Vegetation Index over India". Dehradun: Indian Institute of Remote Sensing.

McCuen, Richard H. 2002. *Modeling Hydrologic Change: Statistical Methods*. CRC press.

Mission, Tropical Rainfall Measuring, and Takuji Kubota. 2013. *TRMM*. Accessed August 27. http://agriculturedefensecoalition.org/sites/default/files/pdfs/33S_2006_NASA_TRMM_Senior_Review_Proposal_2007_Publication.pdf.

"MODIS Vegetation Indices (MOD13) C5 User's Guide: Introduction." 2014. Accessed March 9. http://www.ctahr.hawaii.edu/grem/mod13ug/sect0001.html.

Morel, J. C., A. Mesbah, M. Oggero, and Peter Walker. 2001. "Building Houses with Local Materials: Means to Drastically Reduce the Environmental Impact of Construction." *Building and Environment* 36 (10): 1119–26.

Muhlbauer, Andreas, Peter Spichtinger, and Ulrike Lohmann. 2009. "Application and Comparison of Robust Linear Regression Methods for Trend Estimation." *Journal of Applied Meteorology & Climatology* 48 (9).

"MYD13C2 | LP DAAC :: NASA Land Data Products and Services." 2014. Accessed March 9. https://lpdaac.usgs.gov/products/modis_products_table/myd13c2.

Puri, Gopal Singh. 1960. "Indian Forest Ecology. A Comprehensive Survey of Vegetation and Its Environment in the Indian Subcontinent [Vols. I & II]." *Indian Forest Ecology. A Comprehensive Survey of Vegetation and Its Environment in the Indian Subcontinent [Vols. I & II]*.

"Rain Water Harvesting." 2014. Accessed March 9. http://www.tn.gov.in/dtp/rainwater.htm.

Rodell, Matthew, Isabella Velicogna, and James S. Famiglietti. 2009. "Satellite-Based Estimates of Groundwater Depletion in India." *Nature* 460 (7258): 999–1002. doi:10.1038/nature08238.

Sagreiya, Kamta Prasad. 1967. "Forests and Forestry." *Forests and Forestry*.

Scholz, Christopher H. 2002. *The Mechanics of Earthquakes and Faulting*. Cambridge university press.

Schweig, EUGENE, JOAN Gomberg, MARK Petersen, MICHAEL Ellis, PAUL Bodin, LAUREL Mayrose, and B. K. Rastogi. 2003. "The Mw 7.7 Bhuj Earthquake: Global Lessons for Earthquake Hazard in Intra-Plate Regions." *JOURNAL-GEOLOGICAL SOCIETY OF INDIA* 61 (3): 277–82.

Sen Roy, Shouraseni, and Robert C. Balling. 2004. "Trends in Extreme Daily Precipitation Indices in India." *International Journal of Climatology* 24 (4): 457–66.

Singh, Gurdip, R. D. Joshi, S. K. Chopra, and A. B. Singh. 1974. "Late Quaternary History of Vegetation and Climate of the Rajasthan Desert, India." *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 267 (889): 467–501.

Small, C, F Pozzi, and C Elvidge. 2005. "Spatial Analysis of Global Urban Extent from DMSP-OLS Night Lights." *Remote Sensing of Environment* 96 (3-4): 277–91. doi:10.1016/j.rse.2005.02.002.

Smith, Lindsay I. 2002. "A Tutorial on Principal Components Analysis."

"SNOW WATER EQUIVALENT — GES DISC." 2014. Accessed March 9. http://disc.gsfc.nasa.gov/hydrology/data-holdings/parameters/snow_water_equivalent.shtml.

Spate, Oskar Hermann Khristian, and Andrew Thomas Amos Learmonth. 1972. "India and Pakistan: Land, People and Economy." *India and Pakistan: Land, People and Economy*.

Stein, Seth, and Emile A. Okal. 2005. "Seismology: Speed and Size of the Sumatra Earthquake." *Nature* 434 (7033): 581–82.

Stern, Nicholas. 2007. *The Economics of Climate Change: The Stern Review*. cambridge University press.

Sun, Wenke, and Shuhei Okubo. 2004. "Coseismic Deformations Detectable by Satellite Gravity Missions: A Case Study of Alaska (1964, 2002) and Hokkaido (2003) Earthquakes in the Spectral Domain." *Journal of Geophysical Research: Solid Earth (1978–2012)* 109 (B4).

Swenson, Sean, and John Wahr. 2006. "Post-processing Removal of Correlated Errors in GRACE Data." *Geophysical Research Letters* 33 (8).

Tiwari, V. M., J. Wahr, and S. Swenson. 2009. "Dwindling Groundwater Resources in Northern India, from Satellite Gravity Observations." *Geophysical Research Letters* 36 (18).

"TRMM." 2014. Accessed March 9. http://trmm.gsfc.nasa.gov/3b42.html.

Tu, Jack V. 1996. "Advantages and Disadvantages of Using Artificial Neural Networks versus Logistic Regression for Predicting Medical Outcomes." *Journal of Clinical Epidemiology* 49 (11): 1225–31.

Wahr, John, Mery Molenaar, and Frank Bryan. 1998. "Time Variability of the Earth's Gravity Field: Hydrological and Oceanic Effects and Their Possible Detection Using GRACE." *Journal of Geophysical Research* 103 (B12): 30205–30.

APPENDICES

Appendix - I:

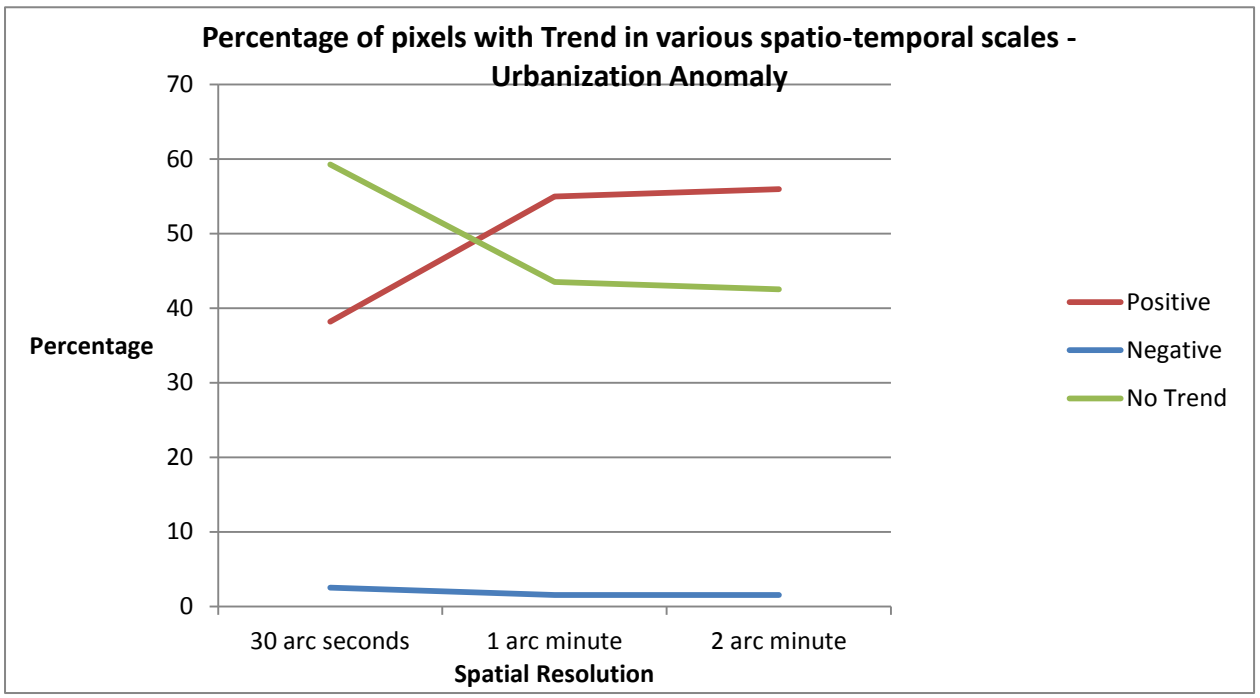
1. **MODEL RESULT**

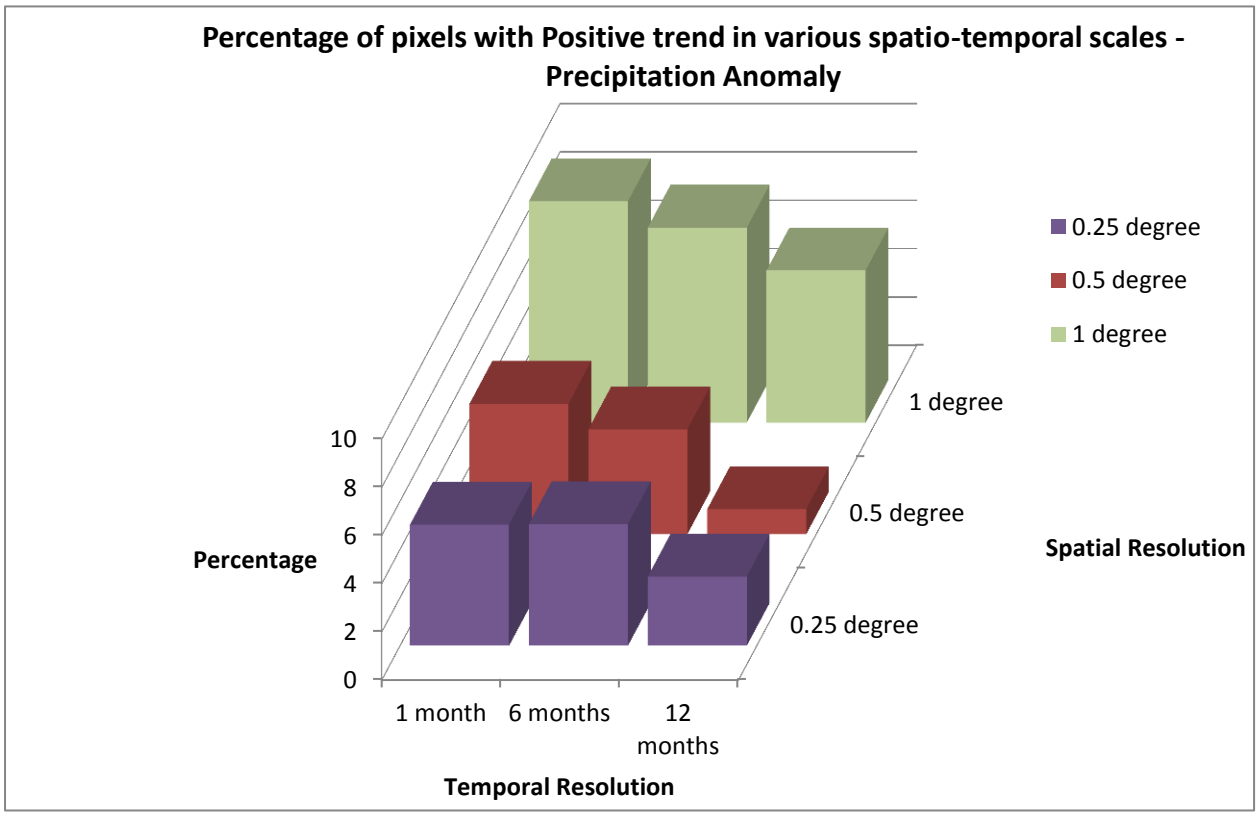
RMS ERROR: 2.89432304624
MODEL ERROR 95% CI: (-6.3544094957988024, 4.0718628925500751)
STD DEVIATION: 2.65981223905
<FullConnection 'FullConnection-13': 'hidden0' -> 'out'>
((0, 0), 1.5215914362593788)
((1, 0), -2.722772038111386)
((2, 0), -0.31033528223307877)
((3, 0), -0.12951972239540424)
<FullConnection 'FullConnection-14': 'in' -> 'hidden0'>
((0, 0), 24.933347179584068)
((1, 0), -131.45277581588365)
((2, 0), 55.372766679538692)
((3, 0), 35.98131000101457)
((0, 1), 18.753599539547992)
((1, 1), -14.123245367706602)
((2, 1), -99.65204764095445)
((3, 1), -17.256692498521531)
((0, 2), -13.403078385929485)
((1, 2), 60.223034020722487)
((2, 2), -35.728745425615777)
((3, 2), 3.6711060893669774)
((0, 3), -17.758276017449724)
((1, 3), -24.842563041668488)
((2, 3), 1.735668991456166)
((3, 3), -11.701768166632892)
<FullConnection 'FullConnection-15': 'bias' -> 'out'>
((0, 0), 1.5926224913418272)
<FullConnection 'FullConnection-16': 'bias' -> 'hidden0'>
((0, 0), -32.854847110505624)
((0, 1), 55.033208240998412)
((0, 2), 50.987869744076193)
((0, 3), 4.2650108267693279)

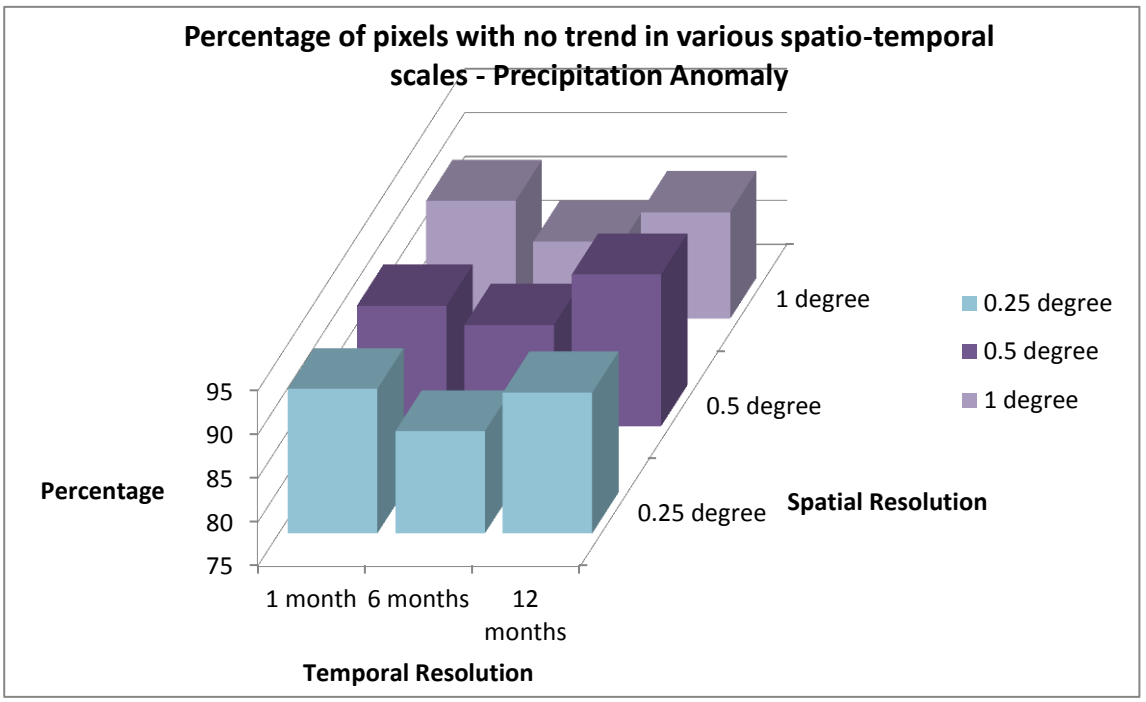
2. XML File

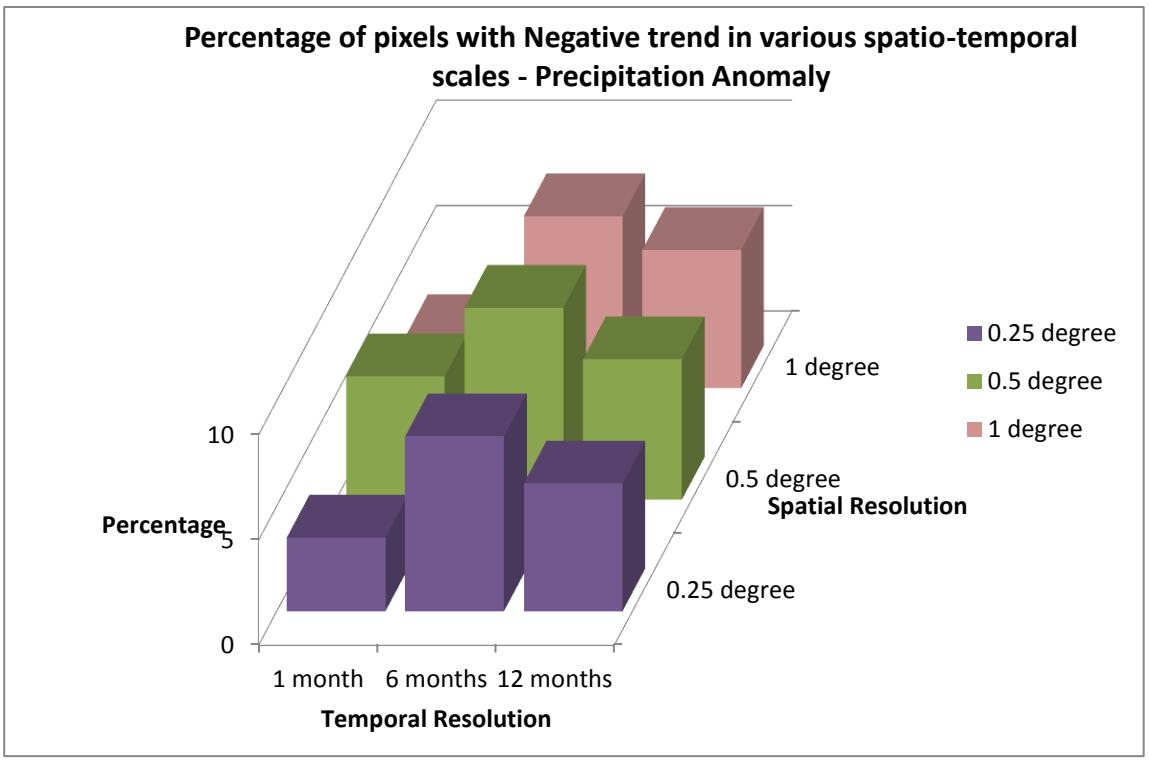

```
<?xml version="1.0" ?>
<PyBrain>
	<Network class="pybrain.structure.networks.feedforward.FeedForwardNetwork" name="FeedForwardNetwork-17">
		<name val="'FeedForwardNetwork-17'"/>
		<Modules>
			<LinearLayerclass="pybrain.structure.modules.linearlayer.LinearLayer"inmodule="True"
name="in">
				<dim val="4"/>
				<name val="'in'"/>
			</LinearLayer>
			<LinearLayer class="pybrain.structure.modules.linearlayer.LinearLayer" name="out"
outmodule="True">
				<dim val="1"/>
				<name val="'out'"/>
			</LinearLayer>
			<BiasUnit class="pybrain.structure.modules.biasunit.BiasUnit" name="bias">
				<name val="'bias'"/>
			</BiasUnit>
			<TanhLayer class="pybrain.structure.modules.tanhlayer.TanhLayer" name="hidden0">
				<dim val="4"/>
				<name val="'hidden0'"/>
			</TanhLayer>
		</Modules>
		<Connections>
			<FullConnection class="pybrain.structure.connections.full.FullConnection" name="FullConnection-
15">
				<inmod val="bias"/>
				<outmod val="out"/>
				<Parameters>[1.5926224913418272]</Parameters>
			</FullConnection>
			<FullConnection class="pybrain.structure.connections.full.FullConnection" name="FullConnection-
16">
				<inmod val="bias"/>
				<outmod val="hidden0"/>
				<Parameters>[-32.854847110505624, 55.033208240998412, 50.987869744076193,
4.2650108267693279]</Parameters>
			</FullConnection>
			<FullConnection class="pybrain.structure.connections.full.FullConnection" name="FullConnection-
14">
				<inmod val="in"/>
				<outmod val="hidden0"/>
				<Parameters>[24.933347179584068, -131.45277581588365, 55.372766679538692,
35.98131000101457, 18.753599539547992, -14.123245367706602, -99.65204764095445, -17.256692498521531, -13.403078385929485,
60.223034020722487, -35.728745425615777, 3.6711060893669774, -17.758276017449724, -24.842563041668488, 1.735668991456166, -
11.701768166632892]</Parameters>
			</FullConnection>
			<FullConnection class="pybrain.structure.connections.full.FullConnection" name="FullConnection-
13">
				<inmod val="hidden0"/>
				<outmod val="out"/>
				<Parameters>[1.5215914362593788, -2.722772038111386, -0.31033528223307877, -
0.12951972239540424]</Parameters>
			</FullConnection>
		</Connections>
	</Network>
</PyBrain>
```

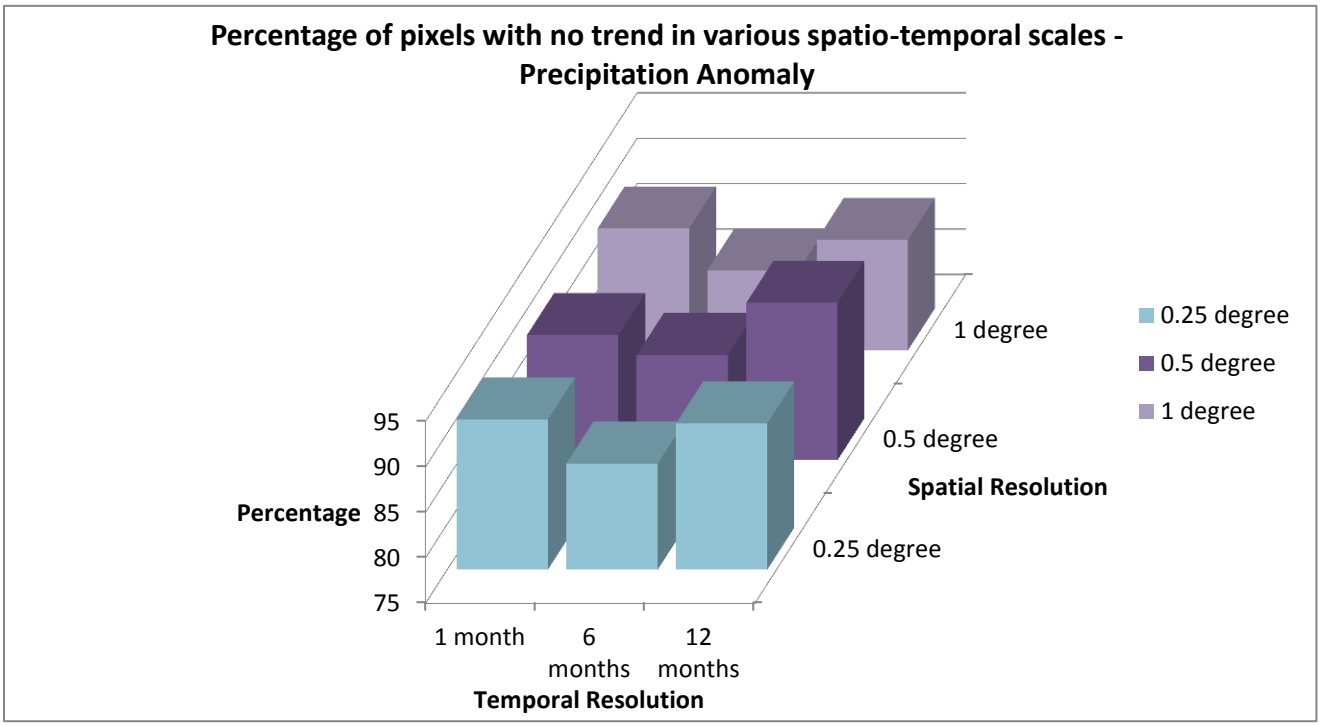
Appendix - II:

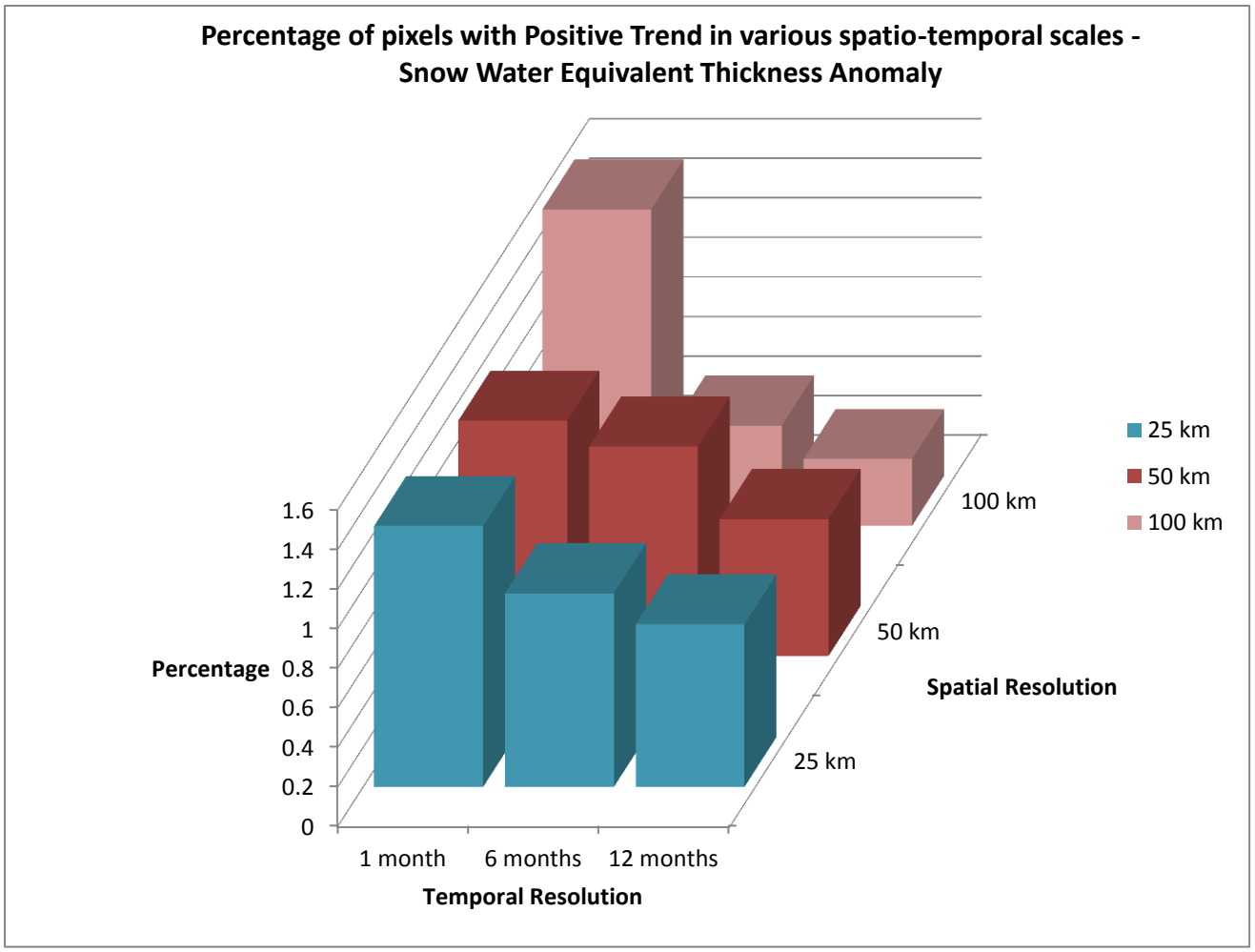

Percentage of pixels with Trend in various spatio-temporal scales - Urbanization Anomaly
70
60
50
40
30
20
10
0
Percentage
30 arc seconds
1 arc minute
2 arc minute
Spatial Resolution
Positive
Negative
No Trend

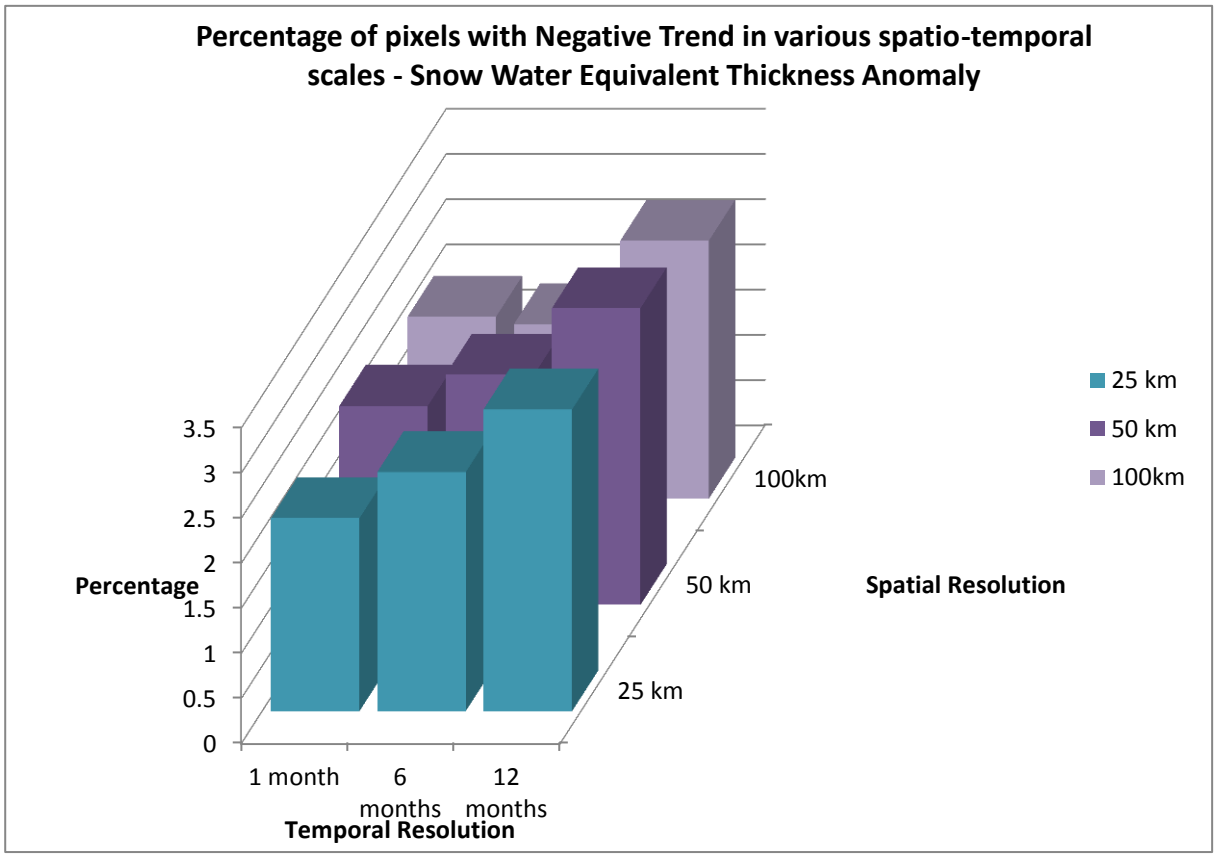

Percentage of pixels with Positive trend in various spatio-temporal scales - Precipitation Anomaly
10
8
6
4
2
0
Percentage
1 month
6 months
12 months
Temporal Resolution
1 degree
0.5 degree
0.25 degree
Spatial Resolution
0.25 degree
0.5 degree
1 degree

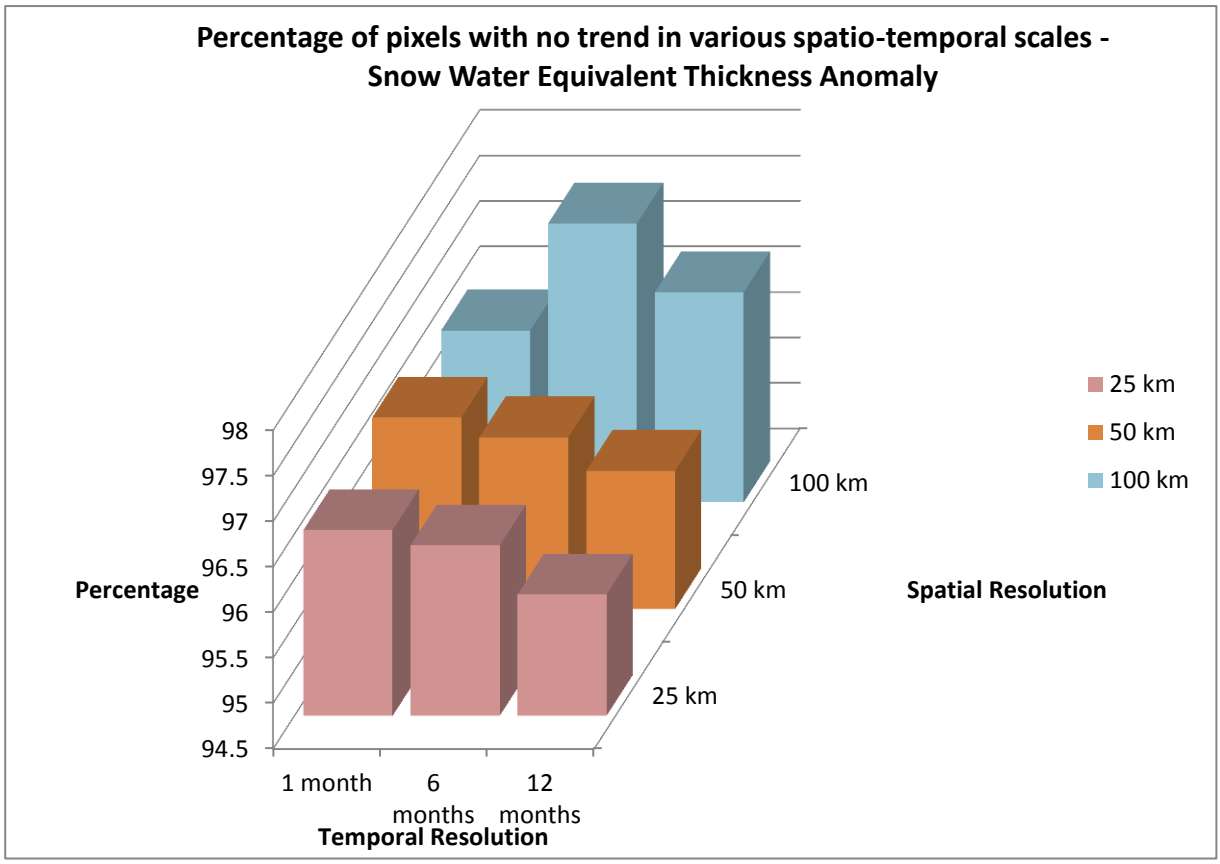

Percentage of pixels with no trend in various spatio-temporal scales - Precipitation Anomaly
95
90
85
80
75
Percentage
1 month
6 months
12 months
Temporal Resolution
1 degree
0.5 degree
0.25 degree
Spatial Resolution
0.25 degree
0.5 degree
1 degree


Percentage of pixels with Negative trend in various spatio-temporal scales - Precipitation Anomaly
10
5
0
Percentage
1 month
6 months
12 months
Temporal Resolution
1 degree
0.5 degree
0.25 degree
Spatial Resolution
0.25 degree
0.5 degree
1 degree

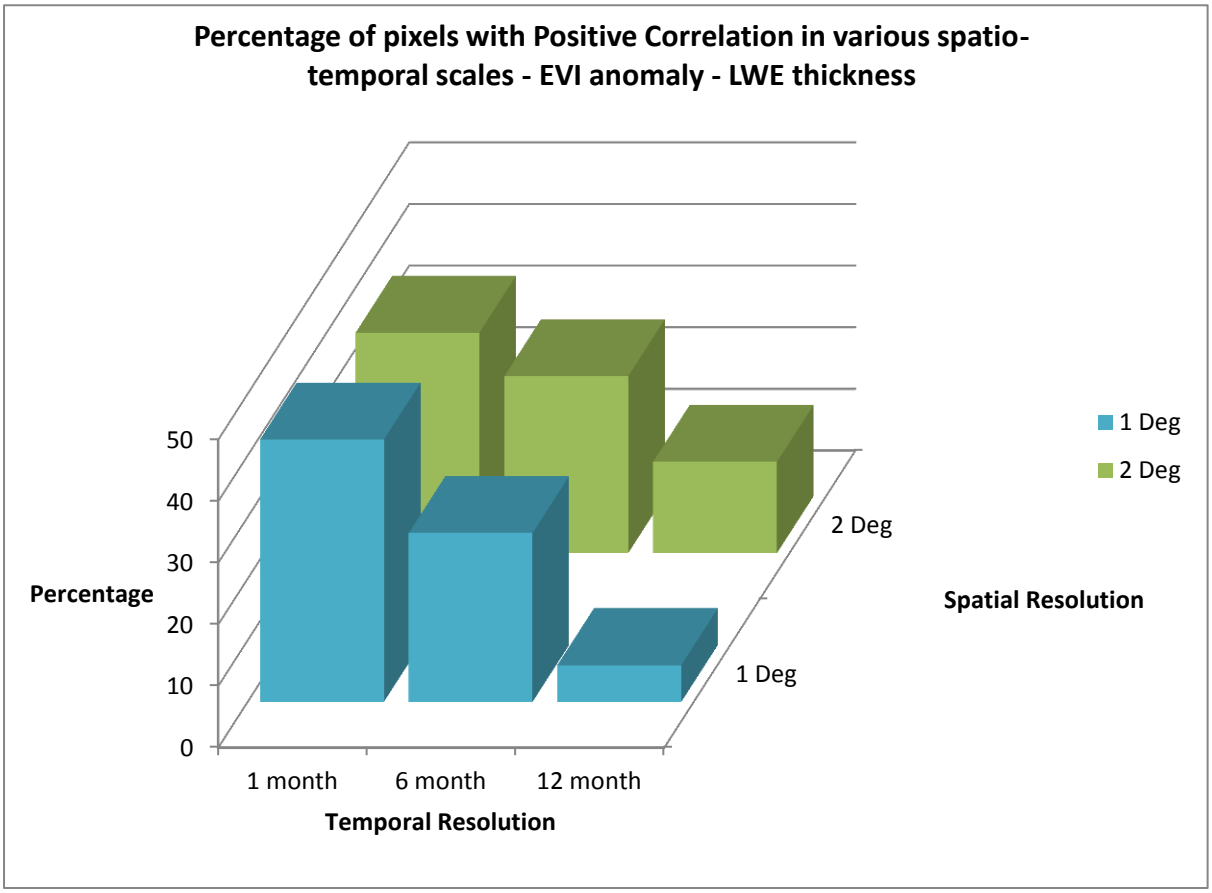

Percentage of pixels with no trend in various spatio-temporal scales - Precipitation Anomaly
Percentage
95
90
85
80
75
1 month
6 months
12 months
Temporal Resolution
1 degree
0.5 degree
0.25 degree
Spatial Resolution
0.25 degree
0.5 degree
1 degree

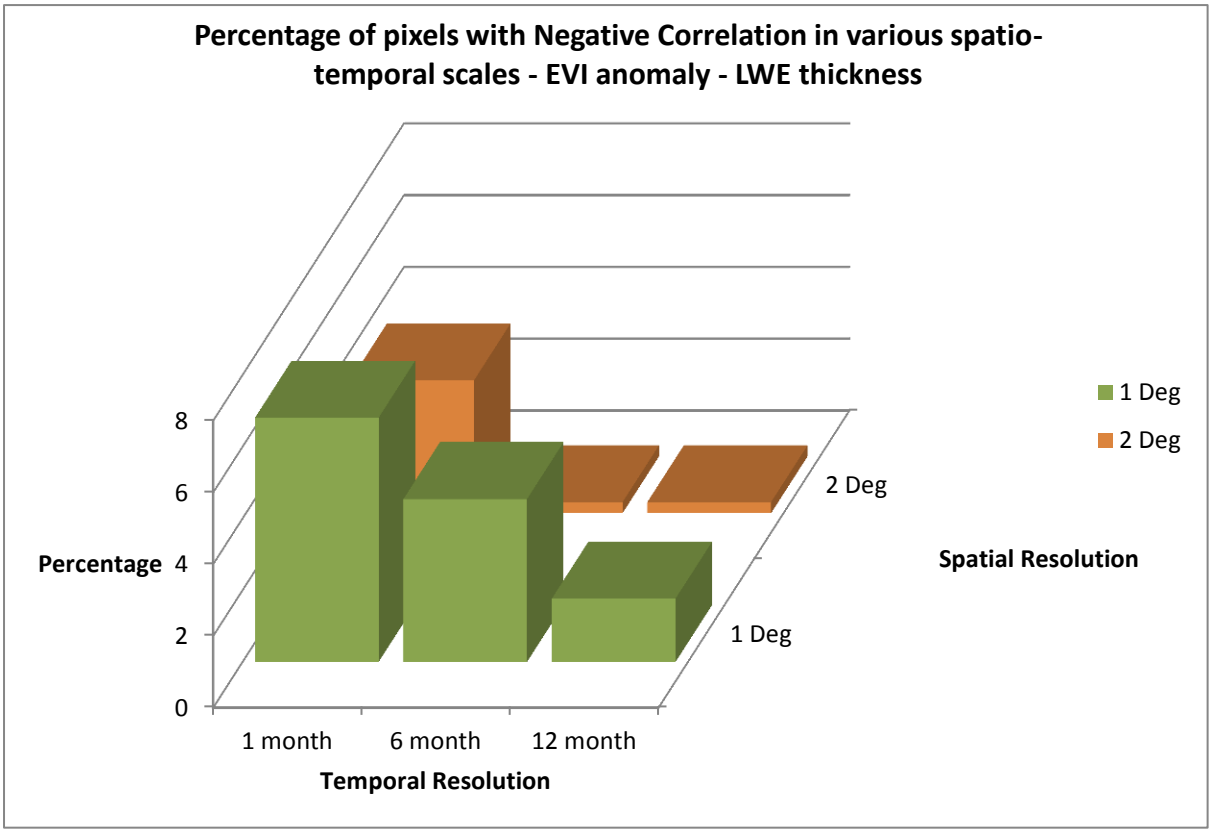

Percentage of pixels with Positive Trend in various spatio-temporal scales - Snow Water Equivalent Thickness Anomaly
Percentage
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1 month
6 months
12 months
Temporal Resolution
100 km
50 km
25 km
Spatial Resolution
25 km
50 km
100 km

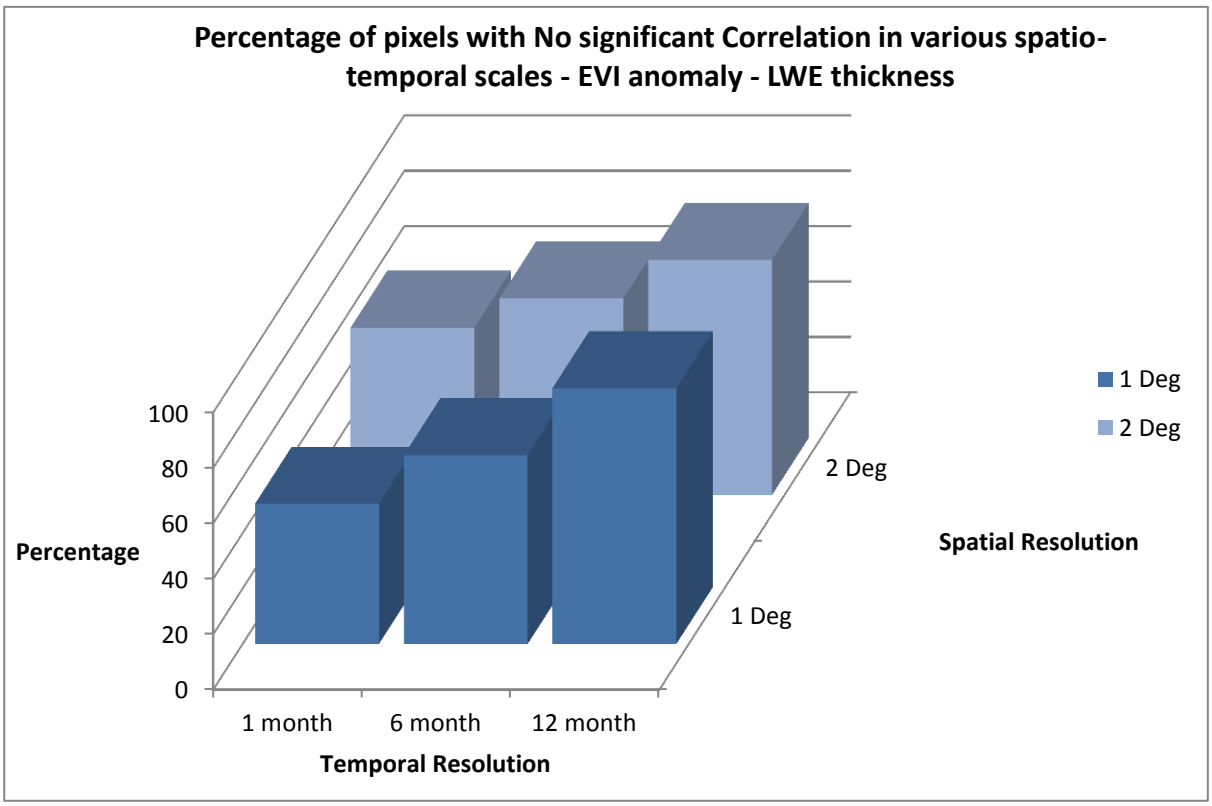

Percentage of pixels with Negative Trend in various spatio-temporal scales - Snow Water Equivalent Thickness Anomaly
Percentage
3.5
3
2.5
2
1.5
1
0.5
0
1 month
6 months
12 months
Temporal Resolution
25 km
50 km
100km
Spatial Resolution
25 km
50 km
100km

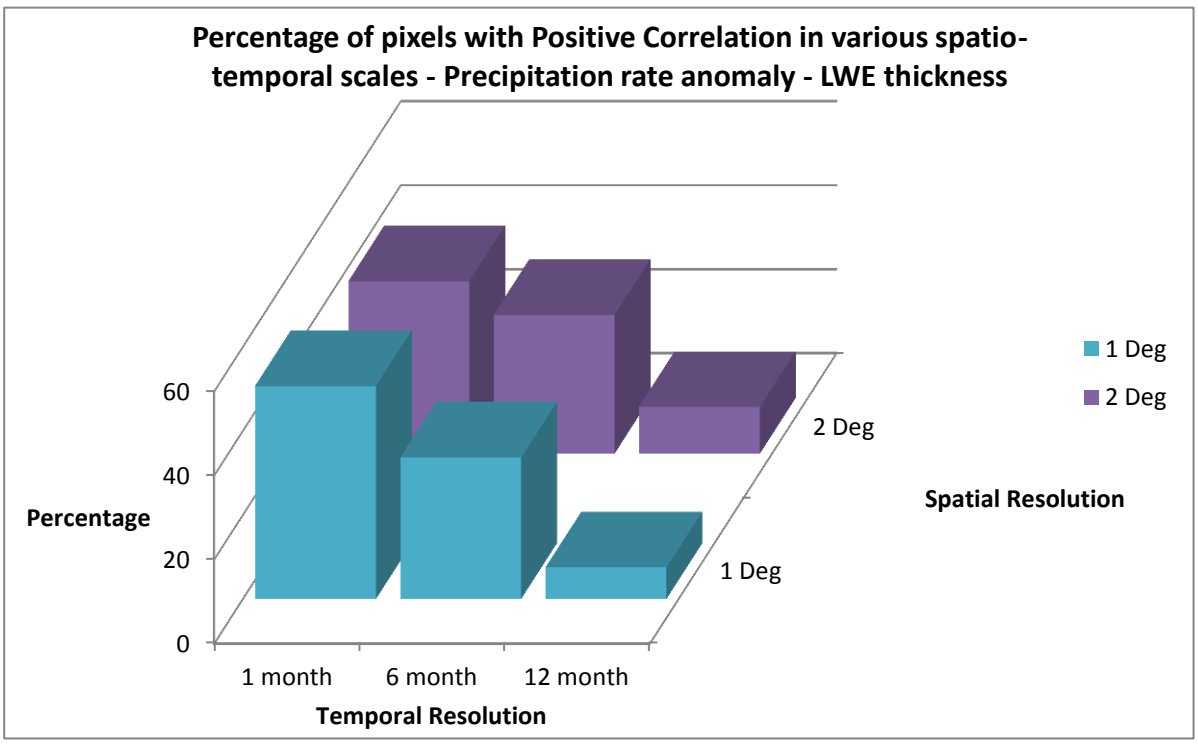

Percentage of pixels with no trend in various spatio-temporal scales - Snow Water Equivalent Thickness Anomaly
Percentage
98
97.5
97
96.5
96
95.5
95
94.5
1 month
6 months
12 months
Temporal Resolution
25 km
50 km
100 km
Spatial Resolution
25 km
50 km
100 km

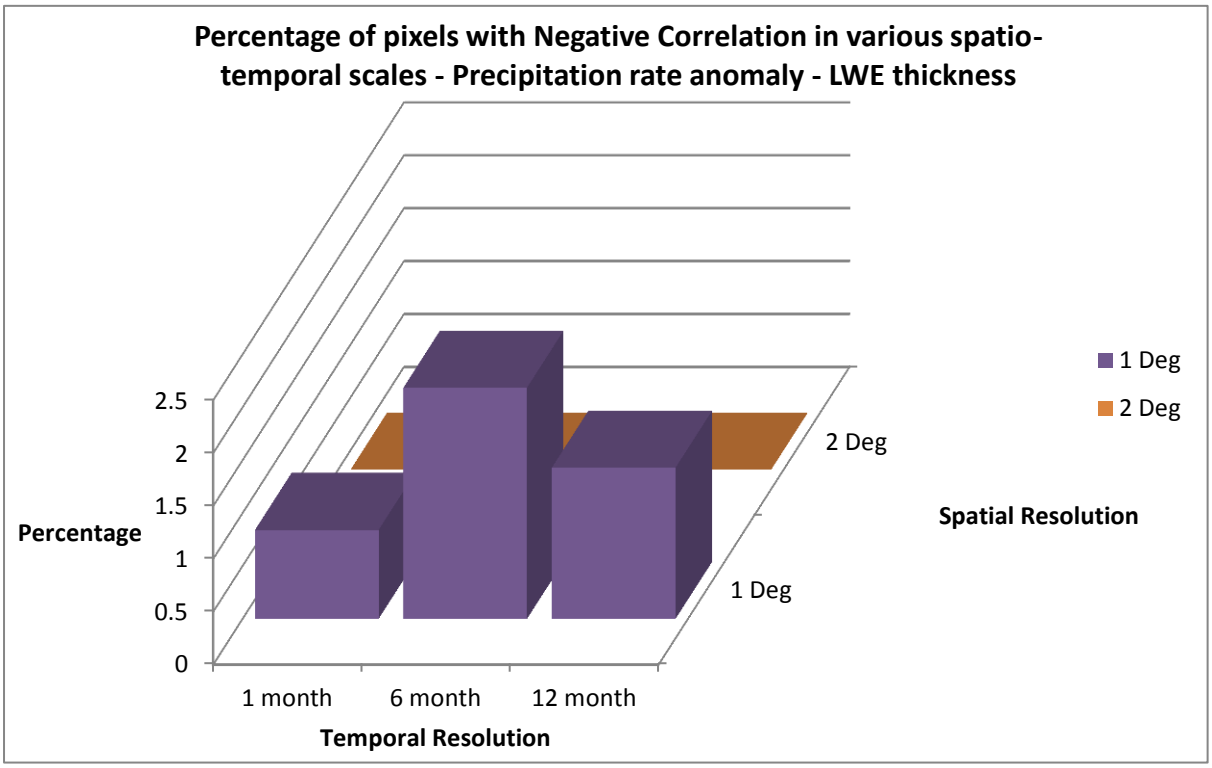
Appendix III:

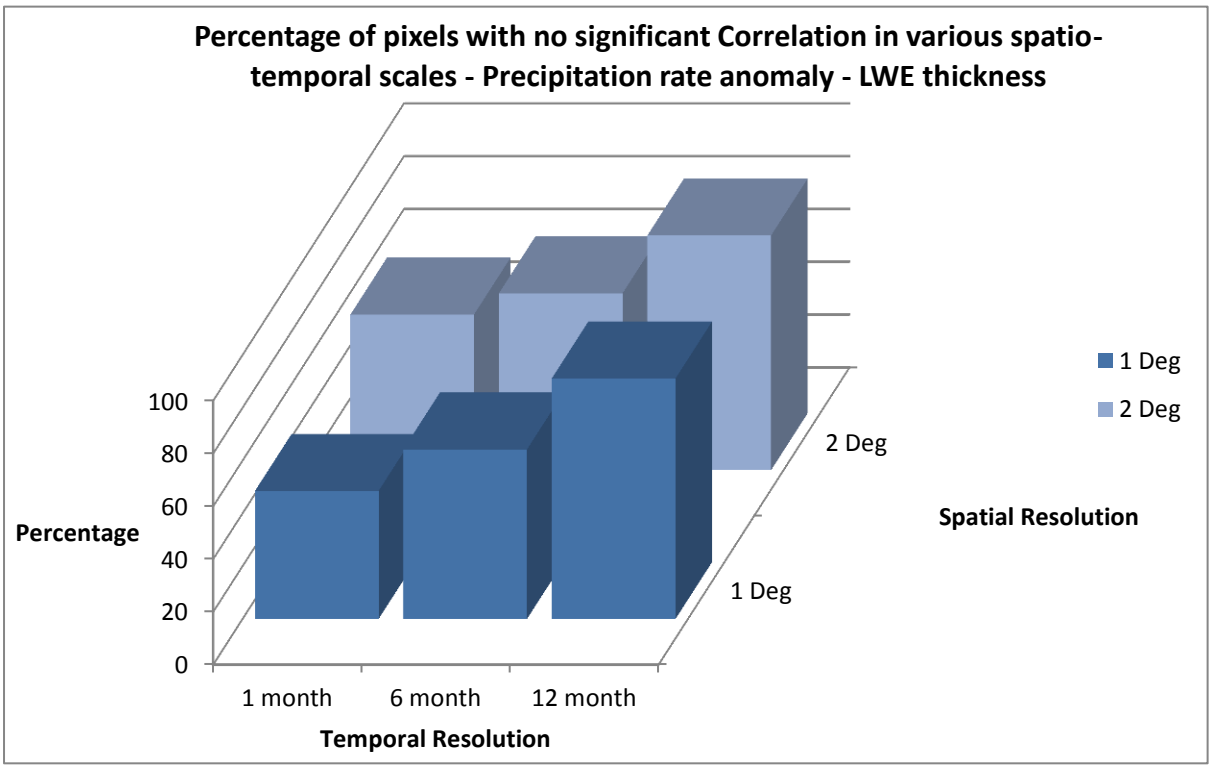


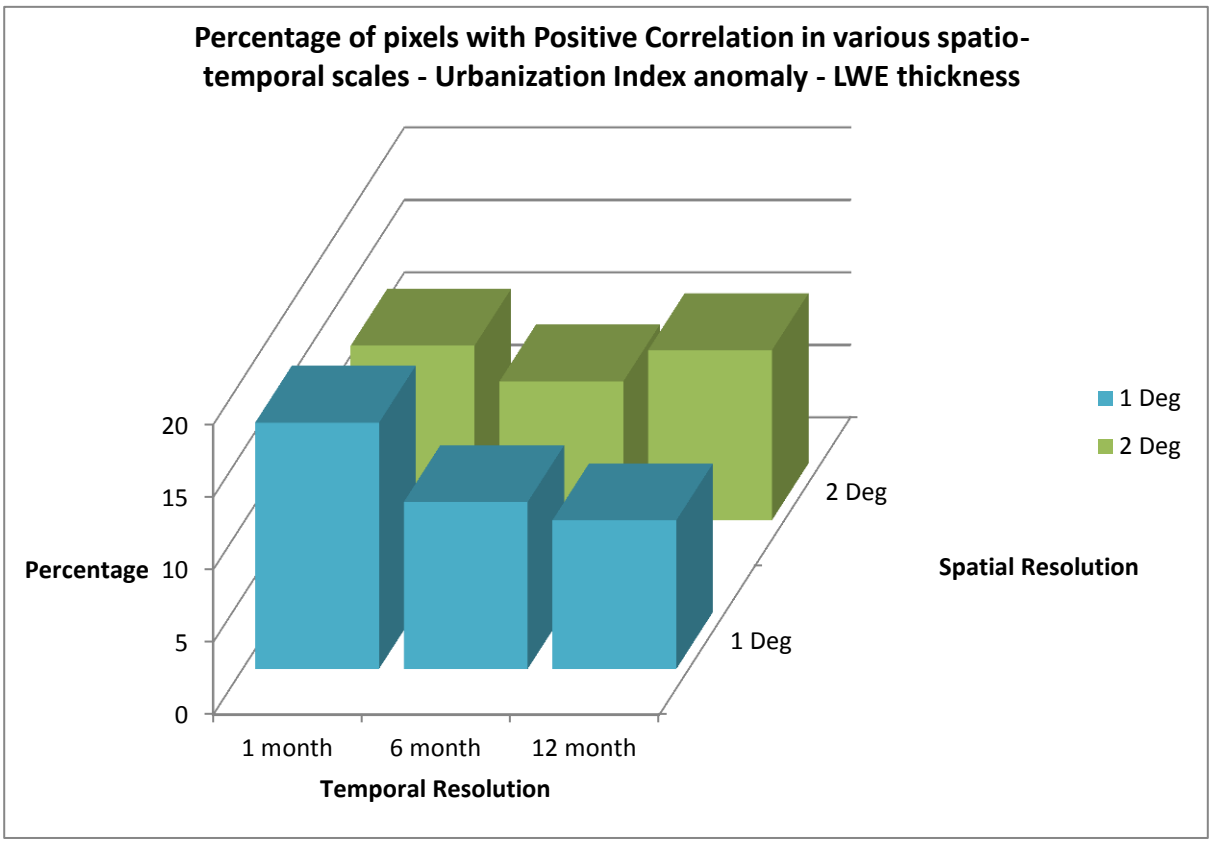


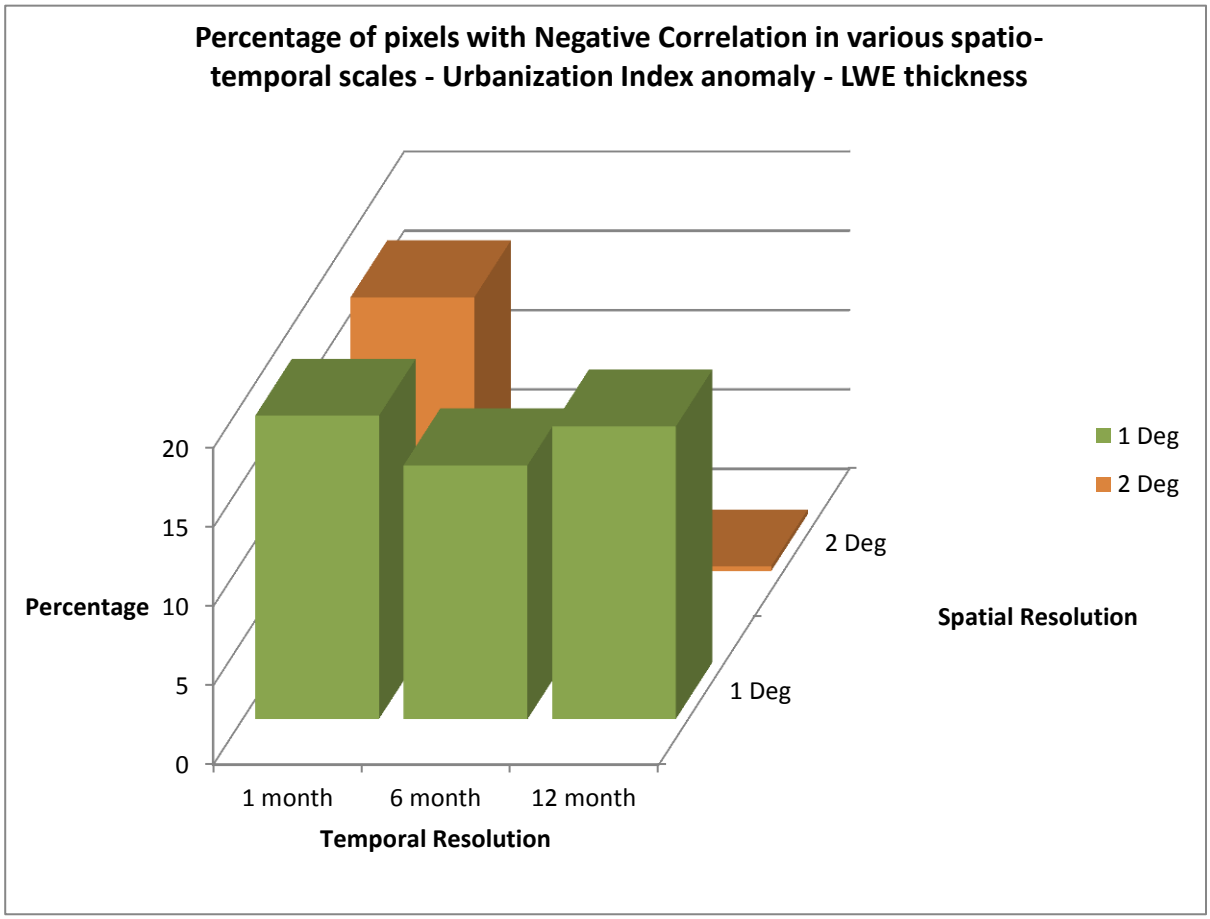

Percentage of pixels with No significant Correlation in various spatio-temporal scales - EVI anomaly - LWE thickness
Percentage
100
80
60
40
20
0
1 month
6 month
12 month
Temporal Resolution
1 Deg
2 Deg
Spatial Resolution
1 Deg
2 Deg

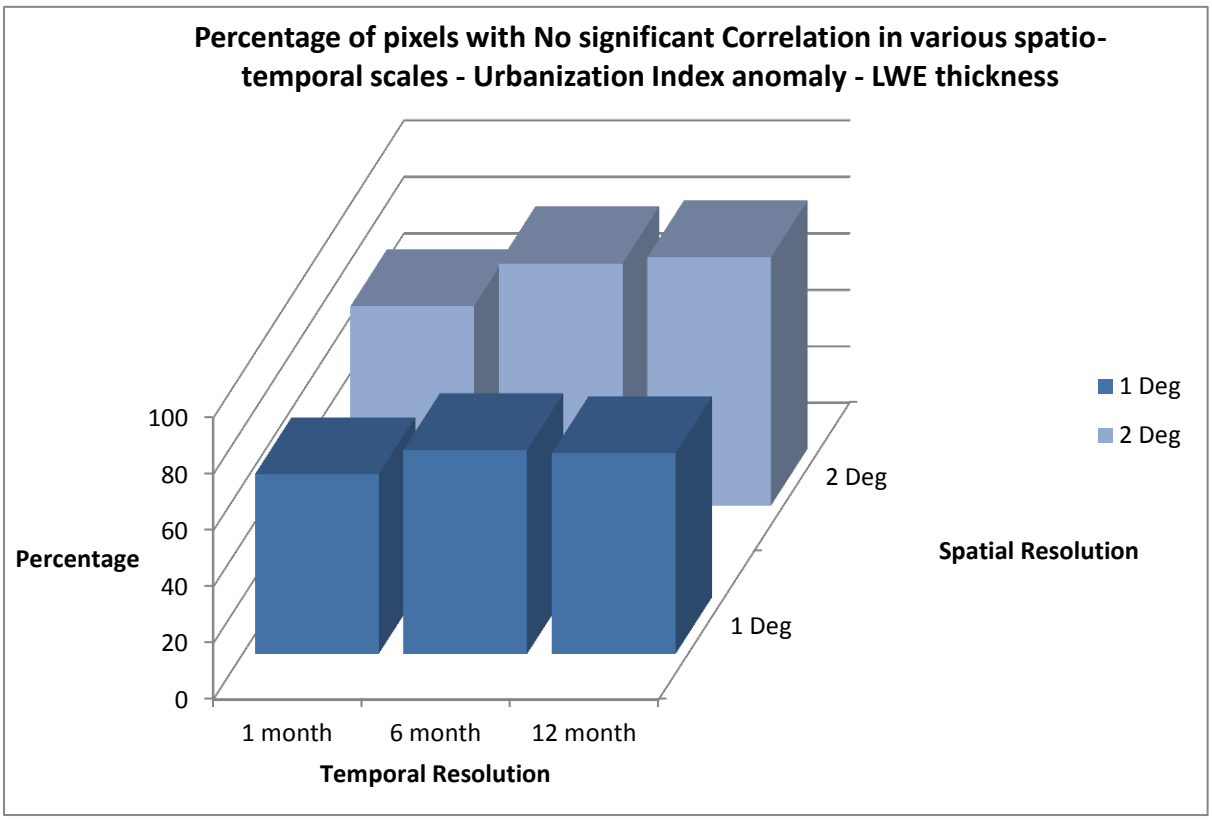

Percentage of pixels with Positive Correlation in various spatio-temporal scales - Precipitation rate anomaly - LWE thickness
Percentage
60
40
20
0
1 month
6 month
12 month
Temporal Resolution
1 Deg
2 Deg
Spatial Resolution
1 Deg
2 Deg

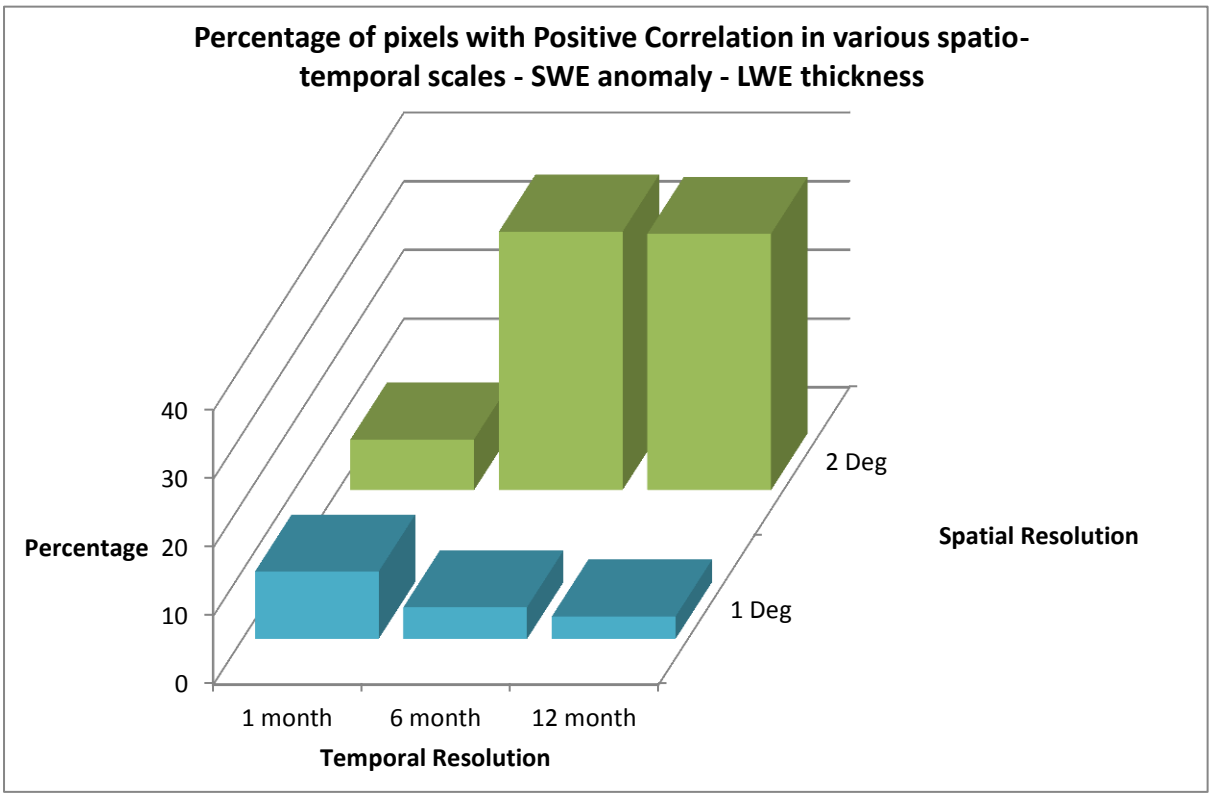

Percentage of pixels with Negative Correlation in various spatio-temporal scales - Precipitation rate anomaly - LWE thickness
Percentage
2.5
2
1.5
1
0.5
0
1 month
6 month
12 month
Temporal Resolution
1 Deg
2 Deg
Spatial Resolution
1 Deg
2 Deg

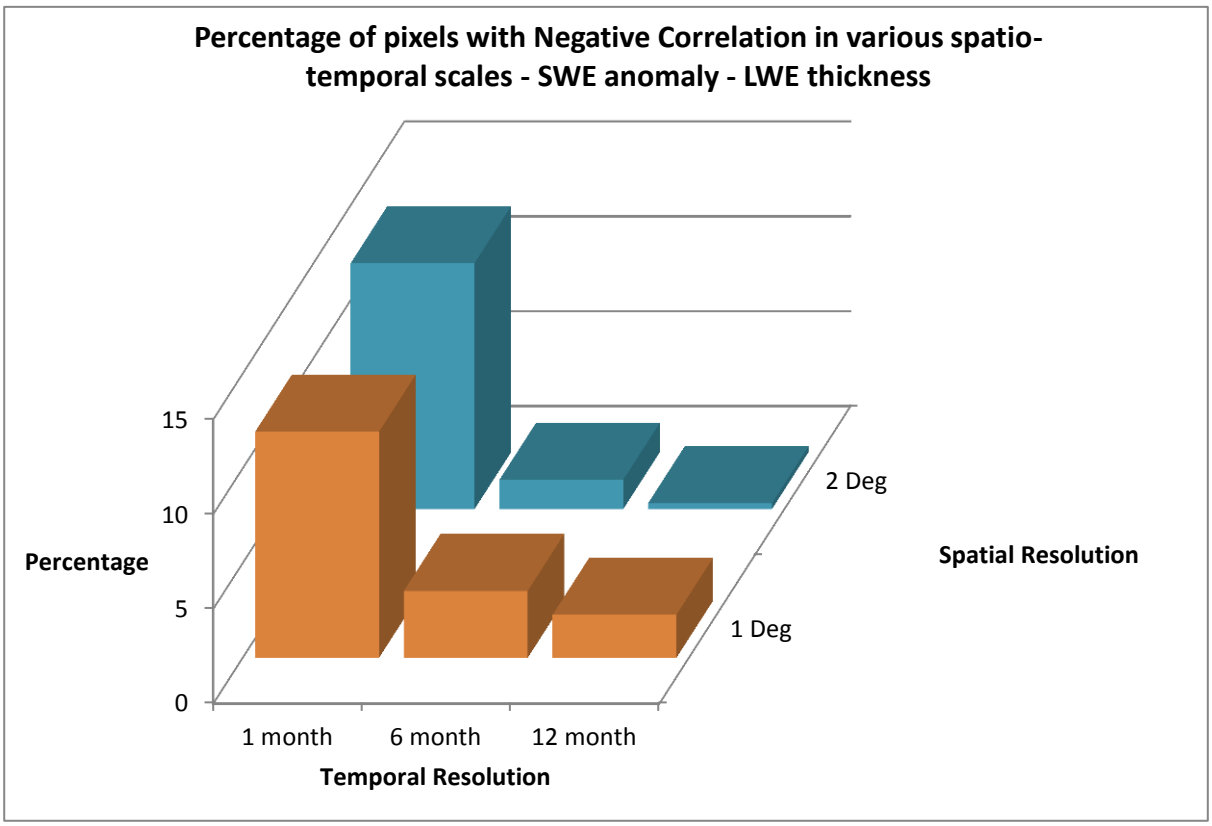

Percentage of pixels with no significant Correlation in various spatio-temporal scales - Precipitation rate anomaly - LWE thickness
Percentage
100
80
60
40
20
0
1 month
6 month
12 month
Temporal Resolution
1 Deg
2 Deg
Spatial Resolution
1 Deg
2 Deg

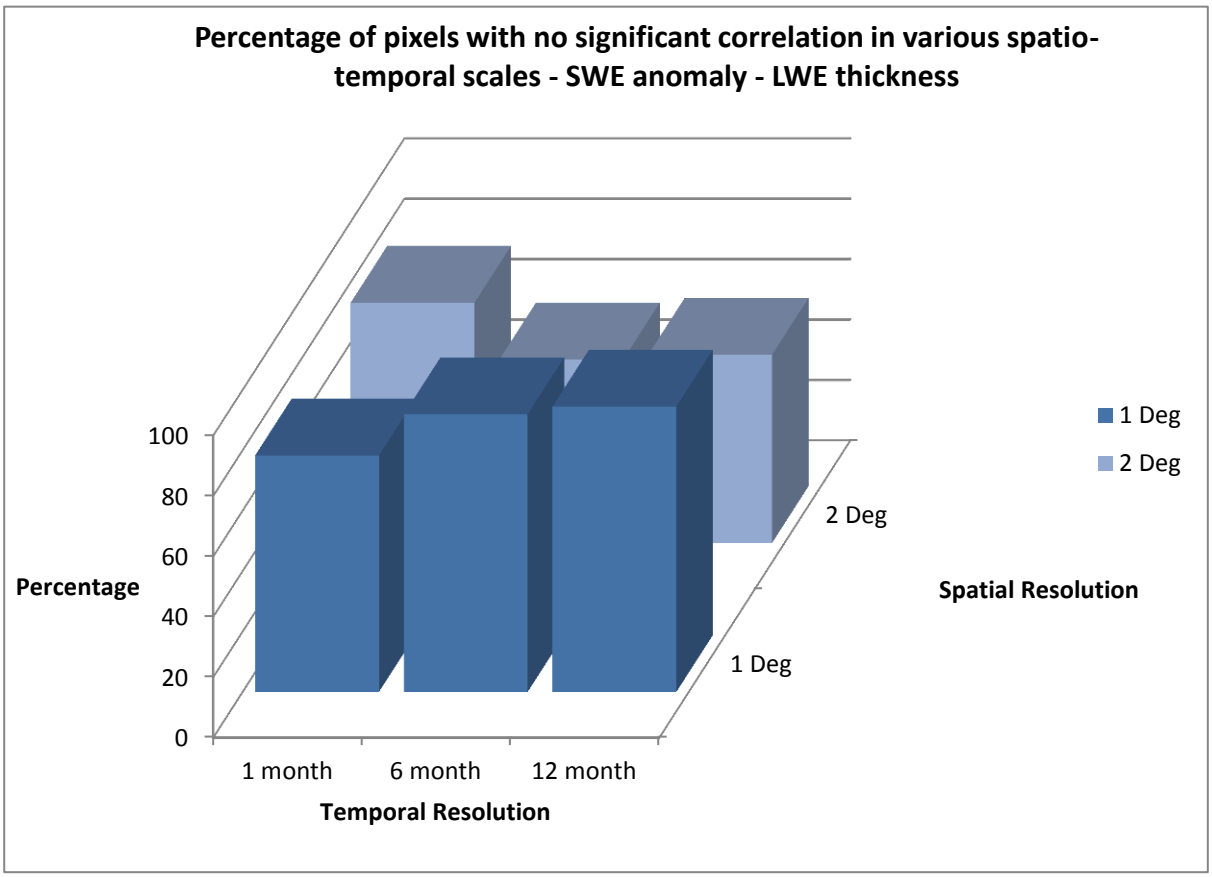

Percentage of pixels with Positive Correlation in various spatio-temporal scales - Urbanization Index anomaly - LWE thickness
Percentage
20
15
10
5
0
1 month
6 month
12 month
Temporal Resolution
1 Deg
2 Deg
Spatial Resolution
1 Deg
2 Deg


Percentage of pixels with Negative Correlation in various spatio-temporal scales - Urbanization Index anomaly - LWE thickness
Percentage
20
15
10
5
0
1 month
6 month
12 month
Temporal Resolution
1 Deg
2 Deg
Spatial Resolution
1 Deg
2 Deg

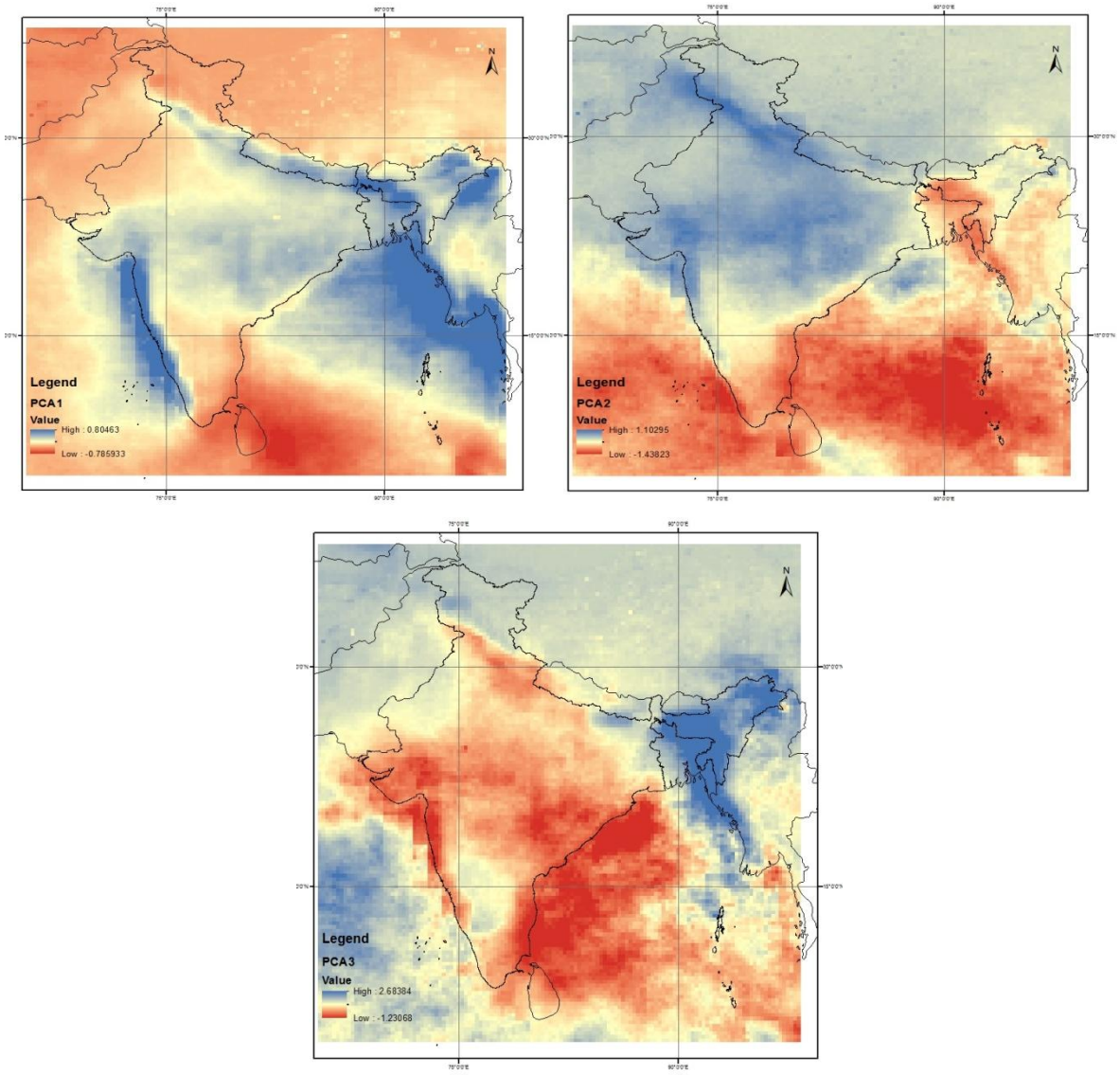

Percentage of pixels with No significant Correlation in various spatio-temporal scales - Urbanization Index anomaly - LWE thickness
100
80
60
40
20
0
Percentage
1 month
6 month
12 month
Temporal Resolution
1 Deg
2 Deg
Spatial Resolution
1 Deg
2 Deg


Percentage of pixels with Positive Correlation in various spatio-temporal scales - SWE anomaly - LWE thickness
40
30
20
10
0
Percentage
1 month
6 month
12 month
Temporal Resolution
2 Deg
1 Deg
Spatial Resolution

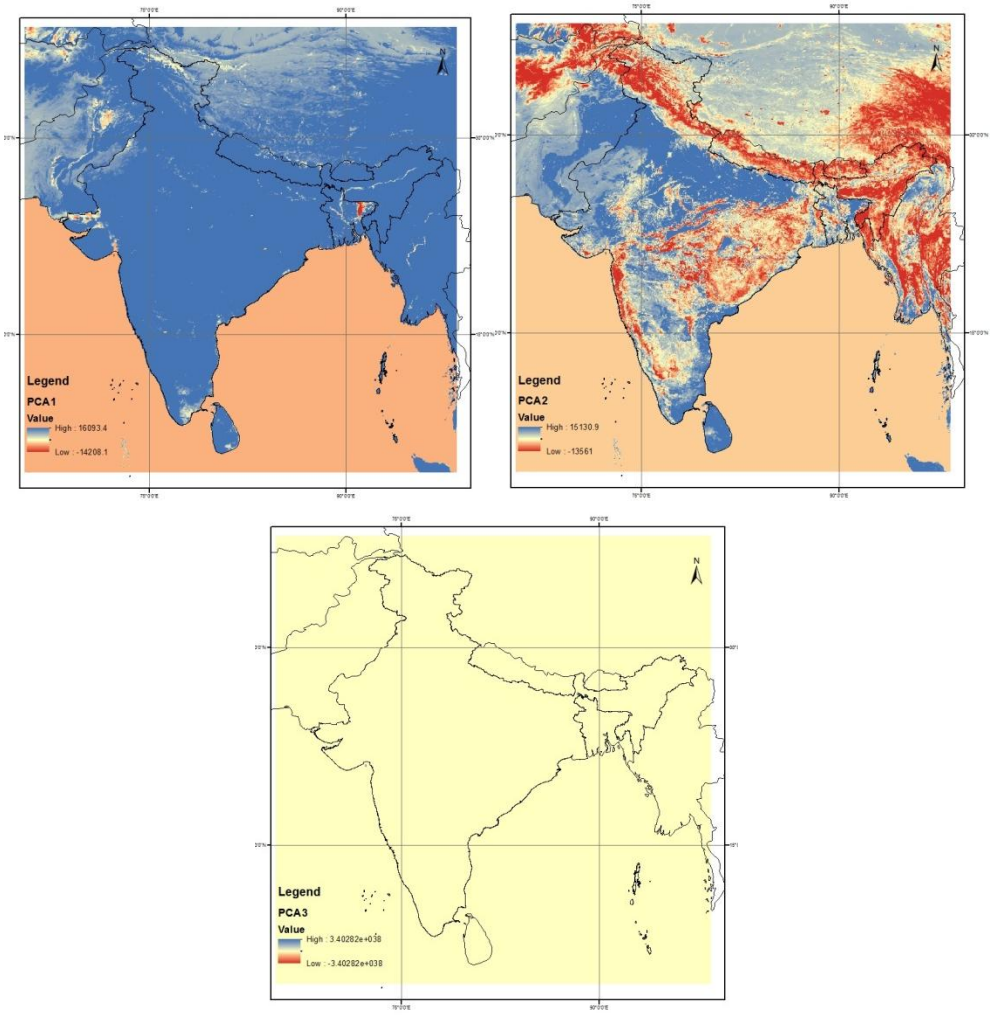

Percentage of pixels with Negative Correlation in various spatio-temporal scales - SWE anomaly - LWE thickness
Percentage
15
10
5
0
1 month
6 month
12 month
Temporal Resolution
2 Deg
1 Deg
Spatial Resolution


Percentage of pixels with no significant correlation in various spatio-temporal scales - SWE anomaly - LWE thickness
Percentage
100
80
60
40
20
0
1 month
6 month
12 month
Temporal Resolution
2 Deg
1 Deg
Spatial Resolution
1 Deg
2 Deg

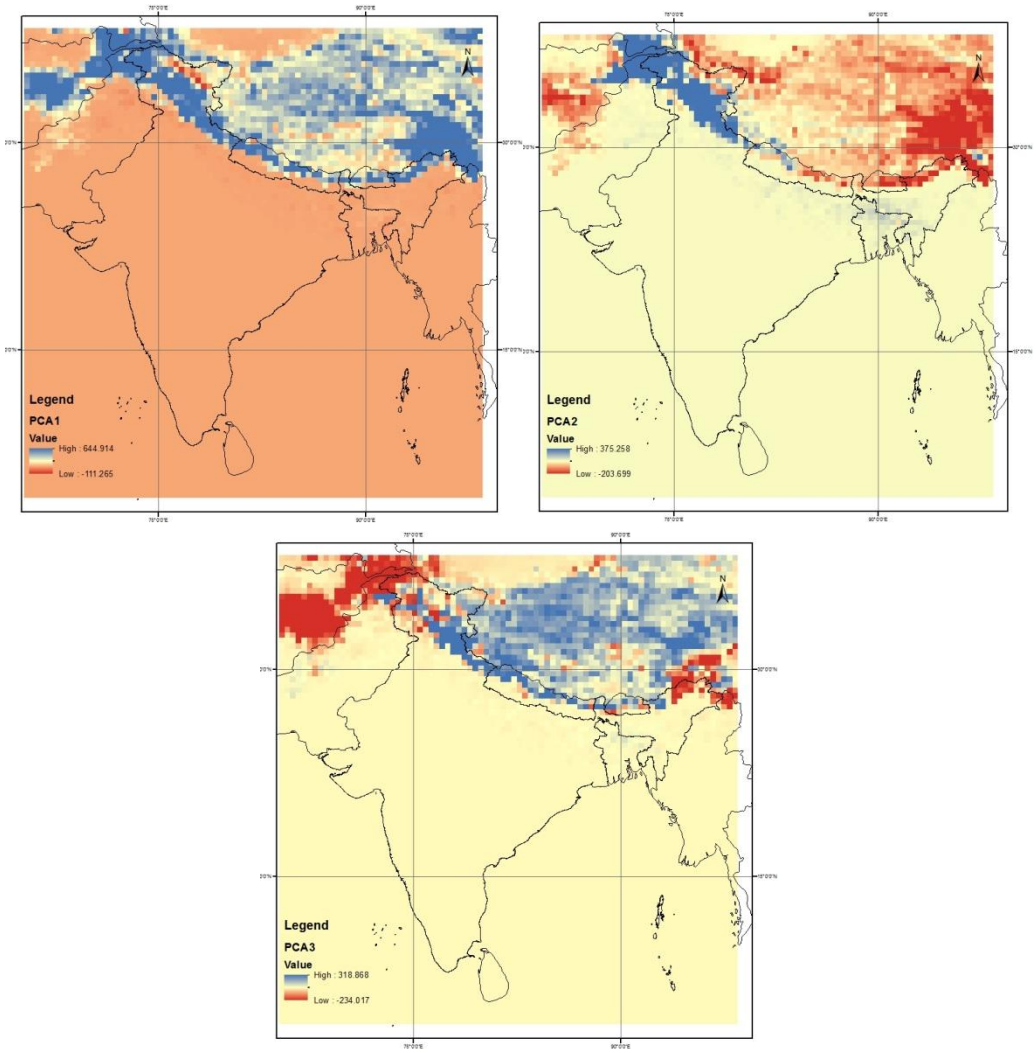
Appendix IV:



PCA Analysis TRMM



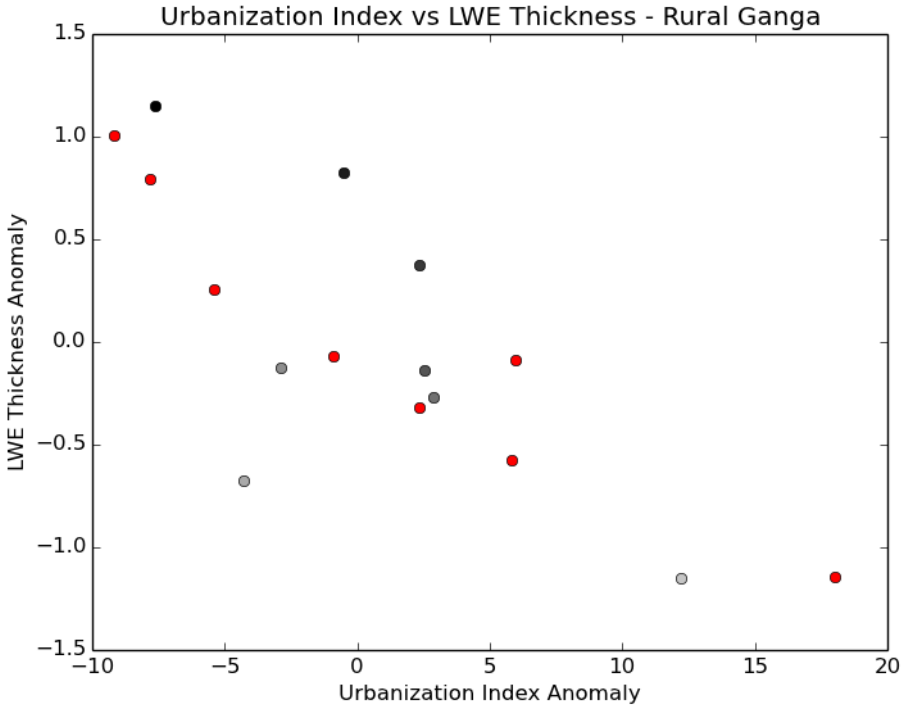
PCA Analysis VIPC1

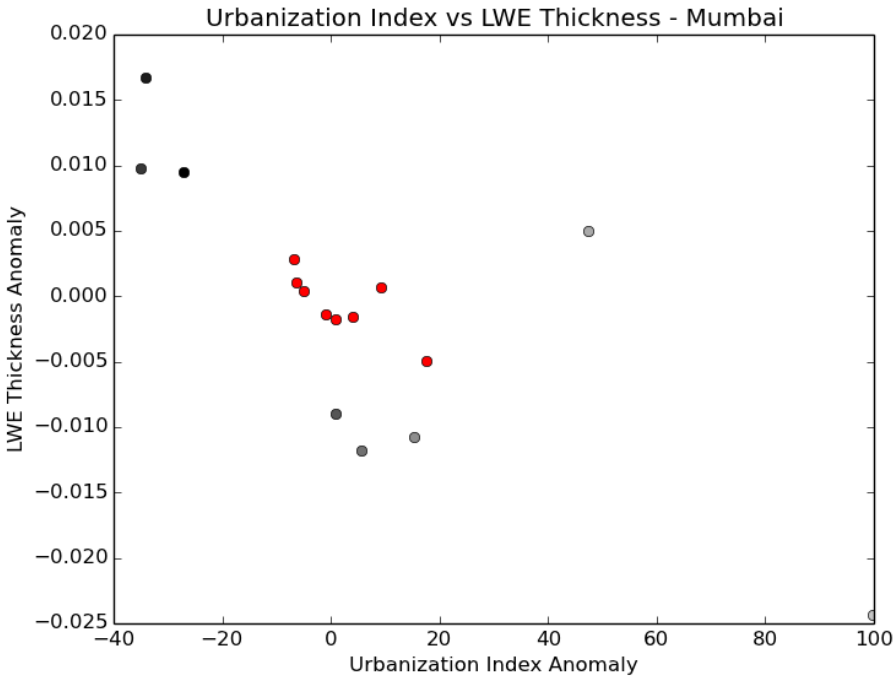


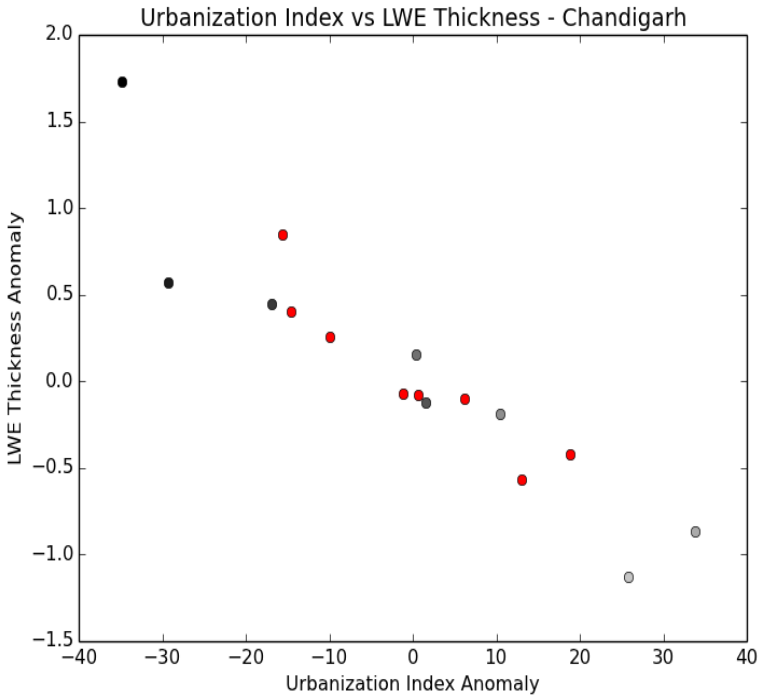
PCA Analysis SWE

APPENDIX – V: Plot of Urbanization index anomaly values when compared with the neighbouring pixels.

Note: Red colour indicates the mean value of the pixels around the central urbanizing pixel. The shades of black indicate the value over the time period. The lighter grey indicates 2010 and black indicates 2003






Urbanization Index vs LWE Thickness - Chandigarh
LWE Thickness Anomaly
Urbanization Index Anomaly
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−40
−30
−20
−10
0
10
20
30
40

APPENDIX VI: Parameters of the models created using the upscaled datasets

1. **2 degree spatial I month temporal resolution dataset**

RMS ERROR: 1.31459665224
MODEL ERROR 95% CI: (-0.97782925074170668, 2.7779881680973659)
STD DEVIATION: 0.958134294422
<FullConnection 'FullConnection-22': 'hidden0' -> 'out'>
((0, 0), 1.0218607476069177)
((1, 0), 1.6584085694114885)
((2, 0), -0.12272851311203202)
((3, 0), 0.20757932783295488)
<FullConnection 'FullConnection-23': 'in' -> 'hidden0'>
((0, 0), 0.68854218561380798)
((1, 0), -1.752591475138872)
((2, 0), -3.1311480180714333)
((3, 0), 1.2087343816680625)
((0, 1), -0.30848476829210936)
((1, 1), 1.8279634387704091)
((2, 1), -0.6993311127833538)
((3, 1), -0.32807566881106753)
((0, 2), 1.4271639143194128)
((1, 2), 0.15716064154735207)
((2, 2), -2.2556549860429285)
((3, 2), 0.93356825448540448)
((0, 3), 1.1401394100063362)
((1, 3), 0.12353380109213487)
((2, 3), -1.1950544236181306)
((3, 3), 1.0005310773067149)
<FullConnection 'FullConnection-24': 'bias' -> 'out'>
((0, 0), -1.838603399477923)
<FullConnection 'FullConnection-25': 'bias' -> 'hidden0'>
((0, 0), 0.89393313230135585)
((0, 1), 1.148279922710371)
((0, 2), -0.68295841820028858)
((0, 3), -1.6081615973354721)

2. **2 degree spatial 6 months temporal resolution dataset**

RMS ERROR: 1.05393995768
MODEL ERROR 95% CI: (-1.4663507187641276, 2.3152097484839262)
STD DEVIATION: 0.964701519282
<FullConnection 'FullConnection-31': 'bias' -> 'out'>
((0, 0), -0.76563481448979387)
<FullConnection 'FullConnection-32': 'bias' -> 'hidden0'>
((0, 0), -4.2983797591456296)

((0, 1), 2.2016172512789622)
((0, 2), -1.4179707321498636)
((0, 3), -2.3069527193343746)
<FullConnection 'FullConnection-33': 'hidden0' -> 'out'>
((0, 0), 0.097354813695590237)
((1, 0), 0.036106369991329497)
((2, 0), 0.24144998699004694)
((3, 0), -1.1516274719858135)
<FullConnection 'FullConnection-34': 'in' -> 'hidden0'>
((0, 0), 2.9707487593717823)
((1, 0), -6.249772073856386)
((2, 0), 0.21804701264027845)
((3, 0), 1.5842239040181592)
((0, 1), 1.7429464548913365)
((1, 1), -0.46203648693661348)
((2, 1), 2.1605583052518567)
((3, 1), -0.61517745759328535)
((0, 2), -1.8823466598586096)
((1, 2), 0.38825675810570659)
((2, 2), 0.027208839624675539)
((3, 2), 2.456389102260792)
((0, 3), -3.2472500654023486)
((1, 3), 4.0058612225190178)
((2, 3), -1.3349566047907966)
((3, 3), 0.3190638452026649)

3. 2 degree spatial 12 month temporal resolution dataset

RMS ERROR: 1.00418228817
MODEL ERROR 95% CI: (-1.6529834949710804, 2.1572438437530121)
STD DEVIATION: 0.97201463108
<FullConnection 'FullConnection-40': 'in' -> 'hidden0'>
((0, 0), 0.54480409995711598)
((1, 0), -0.36369043520626748)
((2, 0), -0.6546439256449107)
((3, 0), 1.0688054180543503)
((0, 1), 0.024369939102743775)
((1, 1), 1.0557974597369728)
((2, 1), 0.51674140415676817)
((3, 1), 0.22511453151543648)
((0, 2), -2.8450892365065656)
((1, 2), -0.76855591130381506)
((2, 2), -0.57301114106617523)
((3, 2), 0.51704908691317752)

((0, 3), 0.084662612660249464)
((1, 3), -1.1564029893525136)
((2, 3), -0.26463757582749475)
((3, 3), -0.13191784927384717)
<FullConnection 'FullConnection-41': 'bias' -> 'out'>
((0, 0), -0.32277770163179947)
<FullConnection 'FullConnection-42': 'bias' -> 'hidden0'>
((0, 0), 1.5968993628286405)
((0, 1), -1.5357983101893609)
((0, 2), -1.9389725920916203)
((0, 3), -1.8458541963390143)
<FullConnection 'FullConnection-43': 'hidden0' -> 'out'>
((0, 0), -0.37447011662810459)
((1, 0), -1.3134740954542894)
((2, 0), 0.72451621871986371)
((3, 0), -0.096715725186141369)

4. **1 degree spatial 6 month temporal resolution dataset**

RMS ERROR: 3.29016449405
MODEL ERROR 95% CI: (-7.0448690225030113, 5.4844911217099614)
STD DEVIATION: 3.19632407612
<FullConnection 'FullConnection-13': 'hidden0' -> 'out'>
((0, 0), 0.77994914496052403)
((1, 0), -1.9306688249956083)
((2, 0), 2.5328200969433059)
((3, 0), 0.006705116533805601)
<FullConnection 'FullConnection-14': 'bias' -> 'out'>
((0, 0), 1.4662950297400352)
<FullConnection 'FullConnection-15': 'bias' -> 'hidden0'>
((0, 0), 9.8612369107907014)
((0, 1), -0.68505855286445616)
((0, 2), -34.761287015268159)
((0, 3), -0.96085192516348972)
<FullConnection 'FullConnection-16': 'in' -> 'hidden0'>
((0, 0), 6.0781325663275005)
((1, 0), -33.115323819651216)
((2, 0), -0.77743280327016573)
((3, 0), -15.968377835572666)
((0, 1), -9.3421048353843066)
((1, 1), -11.604532617437052)
((2, 1), 3.5214743313838484)
((3, 1), 0.62639941516571918)

((0, 2), -18.375425919509112)
((1, 2), -30.882470782414298)
((2, 2), 10.261698653489097)
((3, 2), 55.187894301782777)
((0, 3), 3.8960028769270658)
((1, 3), 12.128083941924441)
((2, 3), -1.9122612712772928)
((3, 3), 13.437804042208469)

5. 1 degree spatial I2 month temporal resolution dataset

RMS ERROR: 1.81901634986
MODEL ERROR 95% CI: (-3.7092908766788106, 3.3943543357264607)
STD DEVIATION: 1.81218769029
<FullConnection 'FullConnection-4': 'in' -> 'hidden0'>
((0, 0), 6.856471054590723)
((1, 0), -8.9331293115252581)
((2, 0), -0.32225786244236659)
((3, 0), 2.213229736758767)
((0, 1), 0.83446657149718584)
((1, 1), 12.455183948584555)
((2, 1), -0.83097788361483971)
((3, 1), 0.85562165455029981)
((0, 2), 6.4763572656422408)
((1, 2), 7.6626826924754381)
((2, 2), -1.4760622596831456)
((3, 2), 2.5133446178421504)
((0, 3), -6.0748825124476378)
((1, 3), -7.123047110268482)
((2, 3), 1.6215877082399361)
((3, 3), 0.011306232404230662)
<FullConnection 'FullConnection-5': 'hidden0' -> 'out'>
((0, 0), -0.10910344036809989)
((1, 0), 0.53813279826316474)
((2, 0), 0.35003376329821295)
((3, 0), 1.2425836051930719)
<FullConnection 'FullConnection-6': 'bias' -> 'out'>
((0, 0), 1.0032232102456775)
<FullConnection 'FullConnection-7': 'bias' -> 'hidden0'>
((0, 0), -4.1136569022397955)
((0, 1), 3.7005645955338262)
((0, 2), 0.082470801535659832)
((0, 3), -3.1496453372295656)