

Temporal granularity in land surface parameters and gravity anomalies and their relation with land-use/ land-cover and climate over India

Thesis submitted to the Andhra University, Visakhapatnam in partial fulfilment of the
requirement for the award of *Master of Technology in Remote Sensing and GIS*



Submitted By
Prasun Kumar Gupta

Supervised By
Dr. S. K. Srivastav
Scientist/Engineer-‘SG’ and
Group Head, Remote Sensing and Geoinformatics Group



**Indian Institute of Remote Sensing
ISRO, Dept. of Space, Govt. of India
Dehradun - 248001
Uttarakhand, India**

July, 2015

Acknowledgements

It is my privilege to acknowledge the inspiration of all my well-wishers and my critics who have propelled me to continue striving for excellence in all the endeavours that I have undertaken. I am also profoundly grateful to all those who have been associated with this research directly or indirectly.

With a deep sense of gratitude, first and foremost, I would like to thank my supervisor Dr. S.K. Srivastav, Scientist/Engineer ‘SG’ at IIRS, Dehradun for his continuous guidance, patient advice and encouragement at all stages of this work. I am also deeply grateful to former and current Dean (Academics), Dr. S.K. Saha and Dr. S.P.S. Kushwaha for their comments and suggestions during various presentations of the research work.

I wish to express my sincere gratitude to Shri P.L.N. Raju, Group Head, PPEG and Ms. Shefali Agarwal, Course Director and Head, PRSD, without whose support this research would not have seen the light of this day. I am grateful to former and current Directors of IIRS, Dr. P.S. Roy, Dr. Y.V.N. Krishnamurthy and Dr. A. Senthil Kumar for providing a conducive environment and facilities for research.

I would also like to express my gratitude to the Andhra University for affording me the opportunity to carry out this research work. I also wish to thank my colleagues and students for putting up with me. My gratitude to all the scientists of IIRS for their continual guidance during the course. The positive attitude and ever eagerness of the non-teaching staff in offering a helping hand needs a special mention in facilitating my research efforts.

A word of appreciation and gratitude to the great guys at UMD, UMB, NOAA, NASA, USGS, JPL, GFZ, IMD, ISRO etc. for sharing data publicly for the greater good of academics and society.

And finally my profuse thanks to my family and friends for being there for me.

(Prasun K. Gupta)

Date: July 03, 2015

Place: Dehradun

ABSTRACT

Availability of time-series data from satellite has enabled scientists to do trend analysis and prediction modelling for satellite imagery derived land surface parameters. The spatial and temporal resolution inconsistency of these datasets have given rise to problems such as *modifiable areal unit problem* (MAUP) and *modifiable temporal unit problem* (MTUP). Also, spatial and temporal scale for the study of environmental processes are important because models established for any variable at one scale may be unimportant or inoperative at another. This study attempts to understand the effects of granularity in space-time domain in the context of time-series analysis.

Six major parameters, namely, *normalized difference vegetation index* (NDVI), *albedo*, *temperature* (minimum and maximum), *rainfall* and *gravity anomaly* have been used in this study. All except climatic parameters have been derived from satellite-based platforms, whereas the climatic data are ground observed values. The datasets belong to various spatial resolutions (from 1 km to 1 degree) and temporal resolutions (15 day to monthly). Various popular (such as *ordinary least squares*, *Mann-Kendall*) and advanced (such as *seasonal decomposition of time-series using locally weighted regression* (STL/LOESS) and *empirical orthogonal functions* (EOF)) statistical techniques have been applied. A number of open source packages and computer programming techniques have also been implemented.

The study was able to inter-calibrate two large time-series vegetation indices with a goodness-of-fit of 0.72. The seasonal decomposition of all parameters revealed that the datasets were mostly dominated by seasonal components, followed by residual (irregular) components (sometimes as high as 10%). The NDVI time-series data trends showed positive trends in most cropping belts in the country, at annual as well as seasonal time scales. Albedo and NDVI correlation analysis showed positive correlation only in lower Himalayas and Deccan plateau. The study also re-confirmed the findings of IPCC reports that the frequency of extreme events such as anomalies, spells and quantitative measures of temperature and rainfall are increasing in parts of the country, including Northern J&K, Rajasthan, Eastern & Western Ghats and North Eastern states. Both parametric and non-parametric tests gave identical results for all the experimental setups.

The correlation of different parameters and the model built to predict gravity anomalies also gave satisfactory results. The model achieved a stability and maturity of 92% and 84% respectively. Minimum temperature was seen to be the most sensitive parameter in the model, however, the maximum likelihood is that the increase of Rainfall and NDVI values will lead to increase in the gravity anomaly values. The single and joint EOF analysis on the six fields revealed that the assumptions made by the correlation model were correct.

The study is able to show spatial & temporal patterns, correlation and variability in several datasets which have different resolutions and time scales. The models & aggregation techniques used, however, may not hold true at other scales. Hence more detailed investigation is required at various spatio-temporal scales and using other auxiliary datasets such as soil moisture and humidity.

Key words: Temporal granularity, Land surface parameters, Gravity anomaly, Climatic forcing, Trend analysis, Spatio-temporal aggregation, Python

Table of Contents

List of Chapters

Title	Page No.
CHAPTER 1 - INTRODUCTION	1
1.1 Background	1
1.2 Problem Statement & Motivation	1
1.3 Research Identification	2
1.3.1 Research Objective	2
1.3.2 Research Questions	2
1.4 Structure of Thesis	2
CHAPTER 2 - LITERATURE REVIEW	4
2.1 Effects of Spatial and Temporal Aggregation: MAUP and MTUP	4
2.2 Inferences from satellite derived NDVI and correlation with land surface parameters	5
2.3 Methods for Spatial Aggregation	5
2.4 Statistical methods for Trend Analysis	6
2.4.1 Linear Regression	6
2.4.2 Mann Kendall	7
2.4.3 Sen Slope Estimator	7
2.4.4 Eureka	8
2.4.5 EOF	8
2.5 Inferences from trend analysis of Gravity Anomaly	8
CHAPTER 3 - STUDY AREA AND DATA USED	9
3.1 Study Area	9
3.1.1 General Introduction	9
3.1.2 Geography	9
3.1.3 Climate	10
3.1.4 Biodiversity	10
3.1.5 Agriculture and Agro-Ecological Zones	10
3.1.6 Water Resources	11
3.2 Data Used	12
3.2.1 GIMMS NDVI Dataset	12
3.2.2 MODIS NDVI Dataset	12
3.2.3 Gravity Anomaly	13
3.2.4 Albedo Dataset	13
3.2.5 Climatic Forcing (Rainfall Data)	13
3.2.6 Climatic Forcing (Temperature Data)	14
3.2.7 Open Source packages and Programming for Geoprocessing	14
CHAPTER 4 - METHODOLOGY	16
4.1 Methodology for Sub-objective 1	16

4.1.1 Data Preparation	16
4.1.2 Inter-Calibration	18
4.2 Methodology for Sub-objective 2	19
4.2.1 Data Preparation	20
4.1.3 Temporal Aggregation	22
4.1.4 Trend Analysis	23
4.3 Methodology for Sub-objective 3	25
4.3.1 Data Preparation	25
4.3.2 Correlation	26
CHAPTER 5 - RESULTS AND DISCUSSION	29
5.1 Combined Time-Series NDVI	29
5.2 Trends in time-series datasets	31
5.2.1 NDVI trends	31
5.2.2 Albedo trends	34
5.2.3 Temperature trends	41
5.2.4 Precipitation trends	50
5.3 Correlation among datasets and gravity anomaly	54
5.3.1 Results obtained from Eureka	54
5.3.2 Results obtained from EOF	57
CHAPTER 6 - CONCLUSIONS	62
References	65

List of Figures

Title	Page No.
Figure 3.1 Study Area - India	9
Figure 3.2 Agro-ecological zone map of India	11
Figure 4.1 Flowchart for first sub-objective relating to creation of combined NDVI dataset for 1981-2013.	16
Figure 4.2 Flowchart for extracting single dataset from HDF files and mosaic using multi-processing	17
Figure 4.3 Flowchart to synchronize MODIS and NOAA datasets using multi-processing	18
Figure 4.4 Regression of time-series NDVI data from MODIS and NOAA satellites over forested and non-forested random points in Indian sub-continent (2002-2006)	19
Figure 4.5a Flowchart for second sub-objective relating to NDVI and Albedo trends	20
Figure 4.5b Flowchart for second sub-objective relating to climatic trends	20
Figure 4.6 Graph and histogram of 5 day temperature for 63 years data indicating 10th & 90th percentile boundaries	22
Figure 4.7 Trend decomposition using LOESS (a) original data for 768 data points (in time domain), (b) seasonal trend (anomaly), (c) long term trend, and (d) residual component	24
Figure 4.8 Histogram of residuals obtained from LOESS	24
Figure 4.9 Example of clipping of Sen Slope estimates at 95% confidence interval ($\alpha < 0.05$)	25
Figure 4.10 Flowchart for sub-objective 3	26
Figure 4.11 Performance parameters for Eureka correlation run	28
Figure 5.1 NDVI plots before and after calibration for a forested pixel	29
Figure 5.2 NDVI plots before and after calibration for a non-forested pixel	30
Figure 5.3 Scatter plot of GIMSS NDVI against MODIS NDVI (after calibration) for 160 points for time duration July 2002 - December 2006 (fortnightly data)	30
Figure 5.4 Results of STL/LOESS on NDVI. Variance of NDVI time-series (a); relative proportions of NDVI variance contributed by seasonal, trend and residual (irregular) components are shown in (b–d) respectively; and STL components for representative pixel are shown in (e). (July 1981-December 2013)	32
Figure 5.5 Trends in NDVI derived from STL/LOESS results (July 1981-December 2013)	33
Figure 5.6 Significant slopes derived from OLS & M-K / Sen's estimator on Annual Temporal Aggregates of NDVI (1982-2013)	33
Figure 5.7 Significant slopes derived from OLS & M-K / Sen's estimator on Cropping cycle Temporal Aggregates of NDVI (1982-2013)	35
Figure 5.8 Results of STL/LOESS on Albedo (January 2002-December 2013). Variance of albedo time-series (a); relative proportions of albedo variance contributed by seasonal, trend and residual (irregular) components are shown in (b–d), respectively.	36
Figure 5.9 Sample variances and proportions of variances explained by various components of the albedo time-series data along two transects as shown in Fig. 5.8a (January 2002-December 2013)	37
Figure 5.10 Trends in Albedo (derived from STL/LOESS results) (January 2002-December 2013)	38
Figure 5.11 Significant slopes derived from OLS & M-K / Sen's estimator on Annual Temporal Aggregates of Albedo (2002-2013)	38
Figure 5.12 Significant slopes derived from OLS & M-K / Sen's estimator on Cropping cycle Temporal Aggregates of Albedo (2002-2013)	40

Figure 5.13 Correlation statistics between albedo and NDVI datasets (2002-2013)	41
Figure 5.14 Results of STL/LOESS on maximum temperature anomaly data. Variance of temperature time-series (a); relative proportions of temperature variance contributed by seasonal, trend and irregular components are shown in (b–d), respectively (1951-2013)	42
Figure 5.15 Results of STL/LOESS on minimum temperature anomaly data. Variance of temperature time-series (a); relative proportions of temperature variance contributed by seasonal, trend and irregular components are shown in (b–d), respectively (1951-2013)	43
Figure 5.16 Trends in max. & min. temperature anomaly (derived from STL/LOESS results) (1951-2013)	43
Figure 5.17 Results of STL/LOESS on ETCCDI Index ‘6’ monthly maximum Tmax (1951-2013)	44
Figure 5.18 Results of STL/LOESS on ETCCDI Index ‘9’ monthly minimum Tmin (1951-2013)	45
Figure 5.19 Trends in ETCCDI Index 6 and 9 on T_{\max} and T_{\min} (derived from STL/LOESS) (1951-2013)	45
Figure 5.20 Significant slopes derived from OLS & M-K / Sen’s estimator on Annual Temporal Aggregates of Maximum Temperature (1951-2013)	46
Figure 5.21 Significant slopes derived from OLS & M-K / Sen’s estimator on Annual Temporal Aggregates of Minimum Temperature (1951-2013)	46
Figure 5.22 Significant slopes derived from OLS & M-K / Sen’s estimator on ETCCDI (Non seasonal) Indices 2 & 13 (on Tmax data) (1951-2013)	47
Figure 5.23 Significant slopes derived from OLS & M-K / Sen’s estimator on ETCCDI (Non seasonal) Indices 4 & 10 (on Tmin data) (1951-2013)	48
Figure 5.24 Trends in monthly difference of Tmax and Tmin (1951-2013)	49
Figure 5.25 Results of STL/LOESS on Rainfall anomaly data (1901-2011)	50
Figure 5.26 Results of trends in Rainfall anomaly (derived from STL/LOESS) (1901-2011)	51
Figure 5.27 Results of STL/LOESS on Seasonal ETCCDI Index 17 for Rainfall (1901-2011)	51
Figure 5.28 Results of STL/LOESS on Seasonal ETCCDI Index 18 for Rainfall	52
Figure 5.29 Trends in ETCCDI Indices 17 & 18 (as derived from STL/LOESS) (1901-2011)	52
Figure 5.30 Significant slopes derived from OLS & M-K / Sen’s estimator on JJAS temporal aggregated Rainfall data (1901-2011)	53
Figure 5.31 Significant slopes derived from OLS & M-K / Sen’s estimator on ETCCDI Index 23 (1901-2011)	53
Figure 5.32 Significant slopes derived from OLS & M-K / Sen’s estimator on ETCCDI Index 24 (1901-2011)	54
Figure 5.33 Number of Occurrences of each variable (across all models)	55
Figure 5.33a Observed vs. Predicted	56
Figure 5.34 Monthly plots for time duration 2004-2010 of observed and predicted (prefixed with ‘ <i>p</i> ’) gravity anomaly. Y-axis represents gravity anomaly and X-axis represents months.	56
Figure 5.35 Spatial distribution of r and r^2 of predicted vs observed gravity anomalies	57
Figure 5.36 Spatial distribution of normalized RMSE between predicted vs observed gravity anomalies	57
Figure 5.37 Spatial variation of first two eigenvectors of monthly mean albedo data (2004-2010)	58

Figure 5.38 Spatial variation of first two eigenvectors of monthly mean gravity anomaly data (2004-2010)	58
Figure 5.39 Spatial variation of first two eigenvectors of monthly mean NDVI data (2004-2010)	59
Figures 5.40 Spatial variation of first two eigenvectors of monthly mean Rainfall data (2004-2010)	59
Figures 5.41 Spatial variation of first two eigenvectors of monthly mean maximum temperature data (2004-2010)	59
Figure 5.42 Spatial variation of first two eigenvectors of monthly mean minimum temperature data (2004-2010)	60
Figure 5.43 Joint spatial variation of first eigenvector in 6 land-surface parameters (2004-2010).	61
Figure 5.44 Joint spatial variation of second eigenvector in 6 land-surface parameters (2004-2010)	61

List of Tables

Title	Page No.
Table 3.1 Data Used	12
Table 4.1: Difference in MODIS and NOAA NDVI datasets	17
Table 4.2 Coefficients for inter-calibration of NDVI datasets	18
Table 4.3 List of ETCCDI indices chosen for current study	21
Table 4.4 Definition of ETCCDI Indices 13 and 10	22
Table 5.1 Correlation between albedo and NDVI - interpretation key	39
Table 5.2 Relative impact within the model that the input variables (tn, r, n, tx) had on the target variable (g)	55

Abbreviations Used

AEZ	Agro-ecological Zone
AVHRR	Advanced Very High Resolution Radiometer
DOS-DBT	Department of Space - Department of Biotechnology
EOF	Empirical Orthogonal Function
ETCCDI	Expert Team on Climate Change Detection and Indices
FDA	Functional Data Analysis
FOSS	Free and Open Source Software
GDAL	Geospatial Data Abstraction Library
GeoTIFF	Geographic Tagged Image File Format
GFZ	German Aerospace Center
GIMMS	Global Inventory Modeling and Mapping Studies
GLCF	Global Land Cover Facility
GRACE	Gravity Recovery and Climate Experiment
HDF	Hierarchical Data Format
IDL	Interactive Data Language
IIRS	Indian Institute of Remote Sensing
IMD	India Meteorological Department
IPCC	Inter-governmental Panel on Climate Change
J&K	Jammu and Kashmir
JPL	Jet Propulsion Laboratory
LULC	Land Use / Land Cover
LWE	Liquid Water Equivalent
MAUP	Modifiable Areal Unit Problem
M-K	Mann-Kendall
MODIS	Moderate Resolution Imaging Spectroradiometer
MTUP	Modifiable Temporal Unit Problem
NASA	National Aeronautics and Space Administration
NBSSLUP	National Bureau of Soil Survey and Land Use Planning
NDC	National Data Centre
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
NOAA	National Oceanic and Atmospheric Administration
N-S	North-South
OLS	Ordinary Least Squares
PC	Principal Component
PCA	Principal Component Analysis
RMSE	Root Mean Square Error
SMA	Simple Moving Average
SPOT	<i>Satellite Pour l'Observation de la Terre</i>
STL/LOESS	Seasonal Decomposition of Time-Series using Locally Weighted Regression
SWIR	Shortwave Infrared
TMAX	Maximum Temperature
TMIN	Minimum Temperature
UMB	University of Massachusetts Boston
UMD	University of Maryland
URL	Uniform Resource Locator
VI	Vegetation Index
W-E	West-East

CHAPTER 1 - INTRODUCTION

1.1 Background

Time-series satellite observations are now available in chronological order with worldwide coverage. This dataset enables monitoring of ecological conditions a matter of detecting and decoding changes within these datasets. Time series data with high sampling interval often show statistical properties which hinder change detection. Extensive literature is available citing most frequently used approaches for detecting temporal trends, i.e. fitting linear regressions against time, but care is needed in order to avoid spurious trends. The detected slope (or gain) coefficient can be used to calculate the amount of change, but it is not always tested for significant deviation from zero, nor are standard statistical assumptions always respected. Phenological cycles are an important cause for the data to violate assumptions like homogeneous variation and absence of serial correlation in the residuals (de Jong and de Bruin, 2012). Linear models may be fitted directly to seasonal data by remedying for seasonality using temporal aggregation, where the aggregation window (or bin size) corresponds to the length of the season. The resulting bins can be considered as temporal units, which, like spatial units, are modifiable. In case of spatial units, it has been demonstrated that the size may influence the model results, which is known as the Modifiable Area Unit Problem (MAUP) (Openshaw and Taylor, 1979) in (de Jong and de Bruin, 2012). This problem may affect a myriad of spatial studies in geography and remote sensing. Similarly, there is a Modifiable Temporal Unit Problem (MTUP) that is as problematic as the MAUP. This is essentially a question of scale in the temporal dimension (Çöltekin, *et al.*, 2011). In examination of time-series of satellite vegetation indices this problem is easily overlooked, although it may result in detection of incorrect temporal trends in the data. Aspects of the problem include the starting phase of a time-series, its extent and the level of temporal aggregation. The aim of this study is to demonstrate possible MTUP effects in analysis of time-series of satellite data to provide, in this sense, a framework for time-series regression.

1.2 Problem Statement & Motivation

Trends of variation and inter-linkages in vegetation indices, gravity anomalies, precipitation, and temperature leading to land use / land-cover (LULC) change have to be studied in order to understand the Earth's processes and human impacts. Variations in biomass and terrestrial water storage result in variations in the gravitational measures. Similarly, variations in biomass and terrestrial water storage may be linked with climate. These variations may ultimately lead to LULC change. The temporal variations of these important geo-physical parameters are loosely understood. Although these variations can be statistically analysed using various trend analysis techniques, the variation of the spatial and temporal resolution affect the results of trend analysis. The MTUP problem thus assumes significance while analysing the trends in time-series datasets.

The purpose of the proposed research is to study the factors of MTUP that affect results of the trend analysis to pave the way for creation of better trend analytical methods that could reduce the effects of such problems. The space-based normalised difference vegetation index (NDVI), albedo, gravity anomalies (derived from GRACE - Gravity Recovery and Climate Experiment mission) along with precipitation and temperature have been considered in the present analysis.

1.3 Research Identification

1.3.1 Research Objective

The main objective of this research is to study the effect of modifiable temporal unit problem (MTUP) on statistical trend analysis for large, temporal satellite datasets.

1.3.1.1 Sub-objectives

1. To study the compatibility between MODIS-NDVI and NOAA-AVHRR-NDVI and to inter-calibrate them to generate a combined time-series NDVI.
2. To study long-term trends in temporal aggregates of NDVI, albedo, precipitation and temperature.
3. To study the correlation in the trends in different datasets including gravity anomaly for the past decade.

1.3.2 Research Questions

1. Does MODIS-NDVI and NOAA-AVHRR-NDVI complement each other? Can a combined time-series NDVI be generated?
2. Can temporal granularity help in understanding the trend in NDVI, albedo, precipitation and temperature? How do the trends vary in different statistical methods?
3. Is there a correlation between the trends in climatic forcing, satellite derived NDVI, albedo and gravity data?

1.4 Structure of Thesis

The thesis has been structured into following six chapters:

Introduction: The introduction throws light on the study background and explains the basic concepts of MAUP and MTUP for land surface parameters and further deals with the problem statement, research objectives and research questions.

Review of Literature: This chapter discusses the literature reviewed and referred for accomplishing this study. It briefly explains the concepts and techniques which are employed and why they are used in this study.

Study Area and Data used: The detailed description of the study area and its significance is illustrated in this chapter. It provides a very informative overview to the geography, climate as well as biodiversity in a concise manner. The other section of this chapter provides a detailed overview of the data sets used and the Open Source packages and Programming used for Geoprocessing in this study.

Methodology: This chapter explains the systematic way followed in this research to answer the research questions. This includes the various algorithms implemented and the methods involved in the study.

Results and Discussion: The research outcomes are discussed extensively in this chapter.

Conclusions: Based on the outcomes and the discussions, conclusions and recommendations are provided to the scientific community.

CHAPTER 2 - 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.

2.1 Effects of Spatial and Temporal Aggregation: MAUP and MTUP

The major obstacle in doing time-series analysis of spatial data is in combining spatially and temporally unmatched data sets (Gotway and Young, 2002). Many statistical issues are also associated with combining such data for modelling and inferences (Shellman, 2004). The choice of an appropriate spatial and temporal scale for the study of climatic processes has been extremely important because mechanisms essential to the spatial and temporal dynamics for any variable at one scale may be unimportant or inoperative at another (de Jong *et al.*, 2011). These facts are proven true in the study of human, animal, and plant populations and has led many researchers in sociology, agriculture, ecology, geography, statistics, and environmental sciences to consider scale issues in detail (Kendall and Yule, 1950).

In many cases, spatial aggregation is necessary to create consequential units for analysis. Openshaw and Taylor (1979) first coined the term modifiable areal unit problem, referred as the MAUP, which is a direct outcome arising out of spatial aggregation. Many studies have illustrated the MAUP as two interconnected problems. The first problem referred to as the scale effect or aggregation effect. It arises as different inferences are obtained when the same set of data is grouped into increasingly larger areal units and vice-versa. The second problem, often termed as the grouping effect or the zoning effect. It arises due to unusual formations of the areal units leading to differences in unit and shape at the similar scales and hence results in the variations in results (Jelinski and Wu, 1996). Both these issues can be, and often are, present in a single analysis.

It has also been proved by many studies that the choice of temporal resolution is critical as it defines the time units for observation. Similar to spatial aggregation, there are consequences of temporal aggregation in time-series models (Buishand *et al.*, 2004). In general, the data is aggregated over fixed number of units may be in years or months or days. Aggregating the data like this will create temporal units which are modifiable. These units influence the analysis and also affect the amount of change detected. This issue has been termed as Modifiable Temporal unit Problem (Çöltekin *et al.*, 2011). MTUP is analogous to MAUP yet it has not been formally developed to address issues that can be applied as a standard. The issue of Modifiable Temporal Unit Problem (MTUP) can be remediated three important aspects are considered for any time-series analysis - duration (how long), temporal resolution (how often), and the point in time (when).

Data aggregation is done to simplify the large data sets by summarizing groups of data elements. Meaningful patterns can often achieve by repeated cycles of aggregation process. But, the consequences of this process are MAUP and MTUP as already explained. The availability of related work which involves issues of MTUP is relatively less because this problem has recently been

addressed (Çöltekin et al., 2011). A related study has been done to study linear trends in seasonal vegetation time-series and the modifiable temporal unit problem over part of Australia (De Jong and De Bruin, 2012). Their results show that linear regression can be used to quantify trends in cyclic data using Ordinary least squares (OLS). They have shown how the temporal unit affects the estimation of model parameters and how the amount of absolute change that was attributed to MTUP has been estimated.

2.2 Inferences from satellite derived NDVI and correlation with land surface parameters

NDVI is one of the most important and commonly used vegetation indices, defined as equation $NDVI = (NIR - RED) / (NIR + RED)$, where RED is the reflectance in the red channel and NIR is the reflectance in the near-infrared channel. The RED and NIR bands contain more than 90% of vegetation information. (Baret, *et al*, 1989). The characteristics of NDVI time-series can be disaggregated into a set of quantitative metrics that may be used to derive information about vegetation phenology and land cover (Hill and Donald, 2003)

Time-series of NDVI are frequently used to map the spatio-temporal behaviour of vegetation cover. NDVI can, for example, be calculated from spectral radiance measurements of Advanced Very High Resolution Radiometer (AVHRR). NDVI is a measure of the fractional vegetated ground cover. To Use NDVI as a correct representation of spatio-temporal vegetation cover behaviour, relations between NDVI and measured vegetation cover have been established by researchers (Maignan *et al.*, 2008). Changes in the vegetation cover are often related to variations in seasonal weather conditions and the moisture availability in the subsoil. Therefore, relation between phenological behaviour, climate and land-use/land-cover is studied by several researchers (Anyamba and Tucker, 2005, Herrmann *et al.*, 2008).

NDVI trends have been used for many purposes, including assessment of ecological response to global warming, phenological change, crop status, land cover change and desertification. For example, systematic greening found in the Sahel is most likely due to climatic variations and recovery from severe droughts (Herrmann *et al.*, 2008). The effects of human-induced land degradation are highlighted by some studies and disputed by others. At the global-scale, combined NDVI trends with rain-use efficiency has been used as a proxy of degradation (Bai *et al.*, 2008). Most analyses established trends by linear regression of NDVI, integrated annually or seasonally but it is not always clear whether a fitted slope coefficient differs significantly from zero or what may be the effect of integration by calendar year in the southern hemisphere where growing seasons straddle the year end (de Jong and de Bruin, 2011).

2.3 Methods for Spatial Aggregation

Spatial aggregation is widely used in the studies, such as land use/cover monitoring, ecological resource management, that is carried out at regional, national and global level. Spatial aggregation divides the input grid of fine resolution raster image into blocks and the value for each block is determined to generate coarse resolution aggregated grid (Raj *et al.*, 2013).

Mainly two types of approaches are used for aggregating fine resolution remote sensing data. One approach is based on numerical aggregation in which the mathematical method such as mean,

median central pixel resampling (Bian and Butler, 1999) is applied on the input grid DN values and the result is assigned to the output aggregated grid. Another approach is based on the categorical aggregation that assigns the class label to output grid by logical processing on input grid categories such as selection of frequently occurring class, random selection of class from input grid (He *et al.*, 2002).

For the purpose of aggregation of continuous images (such as NDVI, Albedo etc.) from finer to coarser resolution, the first approach i.e., numerical aggregation does not affect the spatial patterns in the data, causes a smoothening effect and provides sufficient explanation when related with other variables (Raj *et al.*, 2013).

2.4 Statistical methods for Trend Analysis

Many trend analysis methods exist, however the selection of the most appropriate statistical method will depend on the following characteristics:

- Normal Distribution
- Abrupt Changes
- Cycles
- Outliers
- Missing Values
- Censored Data
- Serial Correlation

There are two main categories for statistical trend analysis methods: (1) parametric approaches - statistics for normally distributed data (such as regression etc.), and (2) non-parametric approaches - statistics that are not as dependent on assumptions about data distribution (such as Mann-Kendall etc.).

The most common method to detect changes in cyclic time-series is the use of a linear model obtained from OLS regression. The slope coefficient α , or gain, is used to calculate the change in Y as β times the number of bins. This number is determined by the aggregation level (or by number of observations per bin), which is equivalent to the sample interval.

2.4.1 Linear Regression

The equation of analyzing the trend using linear regression is given by equation 2.1.

$$Y_t = \alpha + \beta \cdot \omega_t \quad \text{Equation 2.1}$$

where ω_t , the residual, is ideally independent and identically distributed, i.e. white noise. The dependent variable Y can be any kind of VI or cyclical environmental parameter in general. The most common spectral vegetation indices are based on the rapid change in reflectance of chlorophyll between the red and near-infrared (NIR) ranges. In the present study, NDVI, which is a commonly used proxy for terrestrial photosynthetic activity, has been used. (de Jong, *et al.*, 2011)

2.4.2 Mann Kendall

The Mann-Kendall test statistic has been extensively used for detecting linear trends in bio-physical parameters. This test is less sensitive to the outliers as it is based on the order statistics. Like other trend tests, the Mann-Kendall test assumes observations to be independent and identically distributed. The test statistic in the Mann-Kendall test follows a standard normal distribution. In Mann-Kendall test, the null hypothesis is that the data are independent and randomly ordered. Therefore, the significance of trends at a desired significance level can be evaluated by comparing its value with standard normal variate. The impact of serial correlation on the Mann-Kendall test has been observed to results in increase or decrease in the rejection rate of the null hypothesis. It has been observed that this factor considerably reduces the power of the test.

The Mann Kendall statistic S is given as

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(x_j - x_k) \quad \text{Equation 2.2}$$

where,

$$\text{sgn}(x_j - x_k) = \begin{cases} +1, & (x_j - x_k) > 0 \\ 0, & (x_j - x_k) = 0 \\ -1, & (x_j - x_k) < 0 \end{cases} \quad \text{Equation 2.3}$$

This test has been used in number of studies and helped in trend detection for various regions across the globe (Sonali and Kumar, 2012).

2.4.3 Sen Slope Estimator

If a linear trend is present in a time-series, then the true slope (change per unit time) can be estimated by using a simple non-parametric procedure developed by Sen (1968). The slope estimates of N pairs of data are first computed by

$$Q_i = \frac{x_j - x_k}{j - k} \text{ for } i=1, \dots, N \quad \text{Equation 2.4}$$

where x_j and x_k are data values at times j and k ($j > k$), respectively. The median of these N values of Q_i is Sen's estimator of slope. If N is odd, then Sen's estimator is computed by

$$Q_{med} = Q_{(N+1)/2} \quad \text{Equation 2.5}$$

and if N is even, then Sen's estimator is computed by

$$Q_{med} = \left[\left(Q_{N/2} + Q_{(N+2)/2} \right) / 2 \right] \quad \text{Equation 2.5}$$

Finally, Q_{med} is tested by a two-sided test at the $100(1 - \alpha)\%$ confidence interval and the true slope may be obtained by the non-parametric test (Partal and Kahya, 2006).

2.4.4 Eureqa

The detection of free-form natural laws from various physical systems is a nontrivial problem which has haunted scientists for a long time. A robust approach developed by the software “Eureqa” attempts to find correlation among datasets which may be as complex as Hamiltonians and Lagrangians (Schmidt and Lipson, 2009). The approach is to first remove outliers, smoothen the data and then perform symbolical regression on chosen algebraic/ trigonometric/ exponential etc. operators to find the optimum solution by minimizing various error metrics.

2.4.5 EOF

EOF or Empirical Orthogonal Functions is also known as Principal Component Analysis (PCA). When choosing a statistical model, one of the big challenges is to pick as few predictors as possible which are good in explaining much of the past data (lowest mean squared error). EOF analysis provides a rational method for doing this. Also, EOF is one of the most popular methods for analyzing the variability of a single field, i.e., a field of only one scalar variable. The method finds the spatial patterns of variability, their time variation, and gives a measure of the “importance” of each pattern (Bretherton *et al.*, 1992). The “modes of variability” provided by EOF analysis are primarily *data modes*, and not necessarily *physical modes*; and represent standing oscillations in real valued fields, and not propagating patterns.

2.5 Inferences from trend analysis of Gravity Anomaly

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 study the variation in the gravitational potential due to the changes in the surface of Earth. Many researchers have concluded that there exists a positive trend in the gravity anomaly in the southern Indian region (Rodell *et al.*, 2009). The Indo-Gangetic plain has been witnessing a negative trend in the gravity anomaly (Rodell *et al.*, 2009). The trends in the gravitational anomaly have been result of the increased ground water depletion in the region (Tiwari *et al.*, 2009). 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, and Uttar Pradesh along with the union territory of Delhi have been witnessing significant drop in the water table because of irrigational practice.

Gravity anomalies have therefore been studied for changes in total water storage in form of groundwater, surface water and glaciers. There are recent studies in which vegetation is also being considered as a factor influencing gravity anomaly (Yang *et al.*, 2014).

CHAPTER 3 - STUDY AREA AND DATA USED

3.1 Study Area

3.1.1 General Introduction

The study area, India, is a South Asian country located within 8° 8' N to 37° 4' N and 68° 9' E to 97° 23' E (Figure 3.1). It occupies a geographical area of about 3,287,263 km² (1,269,219 sq. mi) of which 90.44% is covered by land and 9.56 % is covered by water.

It is the seventh-largest country by geographical area. The Indian Ocean, the Arabian Sea and the Bay of Bengal bounds it on the south, south-west and the south-east. India shares land borders with Pakistan, China, Nepal, Bhutan, Burma and Bangladesh. In the Indian Ocean, India is in the proximity of Sri Lanka and the Maldives; in addition, India's Andaman and Nicobar Islands share a maritime border with Thailand and Indonesia. The current study does not take into account the area covered by the Andaman and Nicobar Islands and the Laccadive islands.

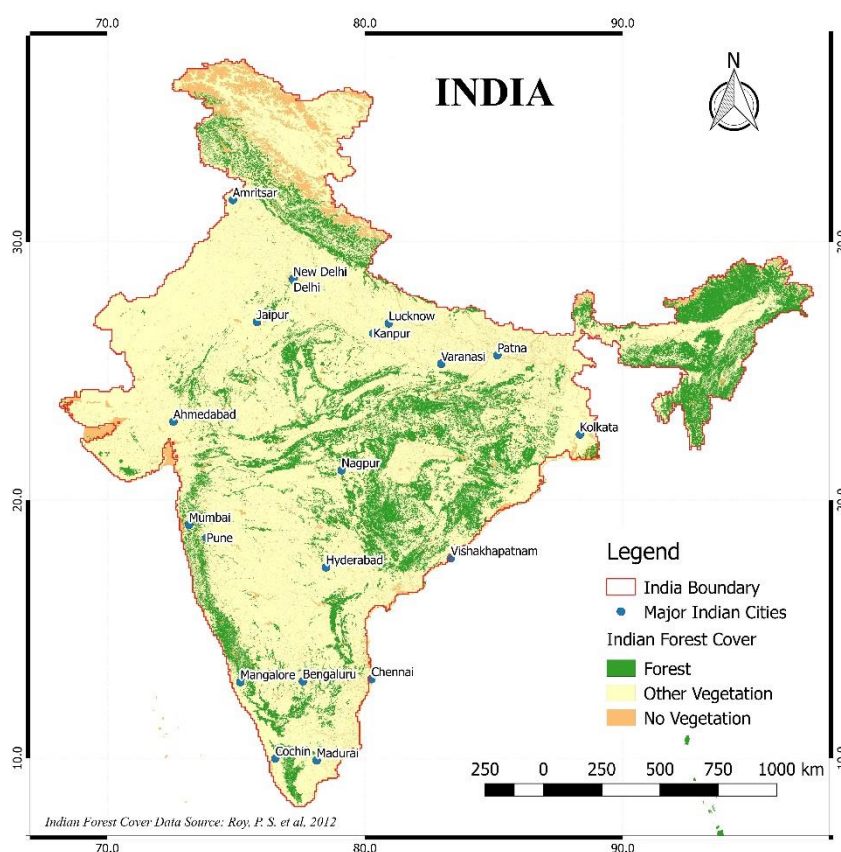


Figure 3.1 Study Area - India (Source: Roy, P. S. *et al.*, 2012)

3.1.2 Geography

India lies on top of the minor Indian tectonic plate, which in turn belongs to the Indo-Australian Plate. India's defining geological processes started 75 million years ago when the Indian subcontinent, then part of the southern supercontinent Gondwana, began a north-eastward drift

across the then-unformed Indian Ocean that lasted 50 million years. The subcontinent's subsequent collision with, and subduction under, the Eurasian Plate pushed up the planet's highest mountains, the Himalaya. They border India in the north and the north-east. In the former seabed immediately south of the emerging Himalaya, plate movement created a vast trough that has gradually filled in with river-borne sediments; it now forms the Indo-Gangetic Plain. To the west lies the Thar Desert, this is cut off by the Aravalli Range in the east.

3.1.3 Climate

The Indian climate is strongly influenced by the Himalaya and the Thar Desert, both of which influence the economically and culturally pivotal summer and winter monsoons. The Himalaya prevent cold Central Asian katabatic winds from blowing in, keeping the bulk of the Indian subcontinent warmer than most locations at similar latitudes. . The Thar Desert plays a crucial role in exerting a pull on the moisture-laden south-west summer monsoon winds that, between June and October, provide the major chunk of India's rainfall. The summer and winter seasons are depended on the axial tilt of Earth as it oscillated from summer solstice to winter solstice in June and December every year, for the Northern Hemisphere. India can be sub-divided into four major climatic groups: tropical wet, tropical dry, subtropical humid, and montane.

3.1.4 Biodiversity

India lies within the Indomalaya eco-zone and contains three biodiversity hotspots. One of the 17 megadiverse countries, it hosts 7.6% of all mammalian, 12.6% of all avian, 6.2% of all reptilian, 4.4% of all amphibian, 11.7% of all piscine, and 6.0% of all flowering plant species. Endemism is high among plants in eco-regions such as the Shola forests. Habitat ranges from the tropical rainforest of the Andaman Islands, Western Ghats, and North-East India to the coniferous forest of the Himalaya. Between these extremes lie the moist deciduous Sal forest of Eastern India; the dry deciduous Teak forest of central and southern India; and the babul-dominated thorn forest of the central Deccan and western Gangetic plain. Less than 12% of India's landmass bears thick jungle.

3.1.5 Agriculture and Agro-Ecological Zones

The major seasonal aspects in case of agriculture are the crop seasons defined as:

- Rabi (December- February),
- Kharif (July-September), and
- Zaid (March-June)

Figure 3.2 shows different Agro-ecological zones in India. Agro-Ecological Zone (AEZ) is a systematic assessment of the soil and climatic resources which is a pre-requisite for formulating efficient land use plan for various regions in India. Mapping of the various agro-ecological regions will help in identifying appropriate cropping patterns for a particular region. To assess yield potentialities of different crops, crop combinations in different agro-ecological regions/zones are delineated. National bureau of soil survey and land use planning- ICAR have differentiated twenty-one such zones on the basis of different ecosystems, physiographic, soil type, climate and growth period for vegetation.

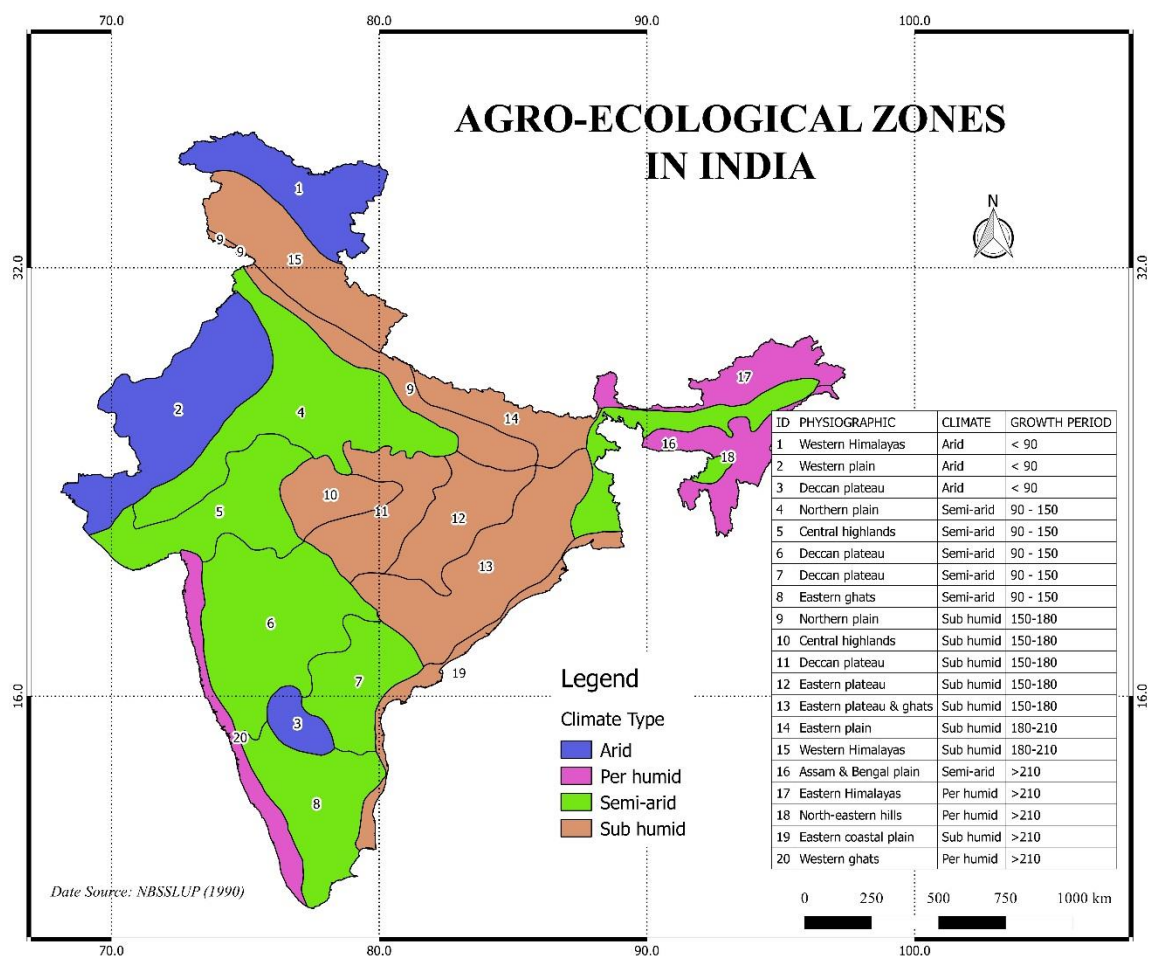


Figure 3.2 Agro-ecological zone map of India (Source: NBSSLUP, 1990)

3.1.6 Water Resources

The normal yearly precipitation India receives is 4,000 km³, out of which 700 km³ is immediately lost to the atmosphere, 2,150 km³ and 1,150 km³ flows into the ground and as streams respectively. The aggregate water resources in the country have been estimated at 1,953 km³. About 62% or 1,202 km³ of the total water resources is available in the Ganga-Brahmaputra-Meghna basin. The remaining 23 basins account for 751 km³ of the total water resources.

The yearly water availability as far as utilizable water resources in India is 1,122 km³. Other than this, the quantity of 123 km³ to 169 km³ additional return flow will also be available from increased use from irrigation, domestic and industrial purposes by the year 2050. The per capita availability of utilizable water, which was about 3,000 m³ in the year 1951, has been diminished to 1,100 m³ in 1998 and is expected to be 687 m³ by the year 2050.

3.2 Data Used

The study is data intensive and large spatio-temporal data sets were handled. Summary of the resolutions and availability of the datasets are listed in Table 3.1. However, only select years for each product were used for calibration, long term analysis and correlation objectives.

Data	Product Name	Spatial Resolution	Temporal Resolution	Availability	Data Web Reference
NDVI (NOAA-AVHRR)	GIMMS	8 km	15 days	Jul 1981-Dec 2006	[1]
NDVI (MODIS)	MYD13A2	1 km	16 days	Jul 2002-Dec 2013	[2]
Albedo - Combined (SWIR)	MCD43B3 Gap filled	1 km	16 days	Jan 2002-Dec 2013	[3]
Temperature - (Tmax, Tmin) °C	IMD 3/2008	1° x 1°	Daily	Jan 1951-Dec 2013	[4]
Rainfall - (per day) mm	IMD4 6/2014	0.25° x 0.25°	Daily	Jan 1901-Dec 2011	[5]
GRACE (LWE thickness) cm	JPL-RL05.1	1° x 1°	Monthly	Apr 2002-Nov 2011	[6]
Data Web References (Accessed on June 23, 2015): [1] http://glcf.umd.edu/data/ [2] http://earthexplorer.usgs.gov/ [3] ftp://rsftp.eeos.umb.edu/data02/Gapfilled/2013/ [4] http://www.imdpune.gov.in/publication/pub_index.html [5] ftp://podaac-ftp.jpl.nasa.gov/allData/tellus/L3/land_mass/RL05/geotiff					

Table 3.1 Data Used

3.2.1 GIMMS NDVI Dataset

The Global Inventory Modeling and Mapping Studies (GIMMS) NDVI data product are available for a 25 year period from 1981 to 2006. This data set is derived from imagery obtained from the AVHRR instrument onboard the (National Oceanic and Atmospheric Administration) NOAA satellite series 7, 9, 11, 14, 16 and 17. This is an NDVI dataset that has been corrected for calibration, view geometry, volcanic aerosols, and other effects not related to vegetation change (<http://glcf.umd.edu/data/gimms/>).

This dataset has a spatial resolution of 8 km and a temporal resolution of 15 days. There are two 15-day composites per month, the first for day 1-15, and the second for day 16 to the end of the month.

3.2.2 MODIS NDVI Dataset

The moderate-resolution imaging and spectrometer (MODIS) NDVI data product was planned to maintain continuity with the NOAA AVHRR-NDVI data. It has specific required characteristics which reduce noise and uncertainty associated with the instrument and external sources (Justice *et al.*, 1998).

The 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 NDVI 16-day composite products (MYD13A2) at 1 km spatial resolution generated from the Aqua satellite were used for the study. The dataset is available in the MODIS sinusoidal projection and HDF file format.

Tiles used - h23v05, h24v05, h24v06, h24v07, h25v05, h25v06, h25v07, h25v08, h26v06 and h26v07 (10 tiles, covering the Indian sub-continent).

3.2.3 Gravity Anomaly

The Gravity Recovery and Climate Experiment (GRACE), a joint mission of National Aeronautical and Space Administration (NASA) and the German Aerospace Center (GFZ), has been making detailed measurements of Earth's gravity field since its launch in March 2002 (Tapley *et al.*, 2004).

The data is in the Liquid Water Equivalent (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. Studies suggest that this data can correlate to total water storage, biomass and other factors influencing mass.

The data used in this study has a temporal resolution of 1 month and a spatial resolution of $1^\circ \times 1^\circ$ (a 300 km wide Gaussian filter has also been applied to the data). The latest dataset, reprocessed by Jet Propulsion Laboratory (JPL), called RL05.1, have been used in the present study. The data has been provided starting from April 2002 till November 2011.

Due to data being not available for the months of June 2002, July 2002, June 2003, January 2011 and June 2011, the time duration from January 2004 till December 2010 has been taken for correlation analysis as explained in Section 4.3.

3.2.4 Albedo Dataset

The MODIS *albedo* product (MCD43B3) provides 1-kilometer data portraying both directional hemispherical reflectance (black-sky albedo) at local solar noon and bihemispherical reflectance (white-sky albedo) (Justice *et al.*, 1998). This MCD43B3 albedo magnitude is produced from the 16-day anisotropy models provided in MCD43B1 and represent mean of the underlying 500m values. The MCD43B3 albedo quantities are provided as a level-3 gridded product in the Sinusoidal projection.

The study attempts to use the best available representation of time-series albedo available. The term “MCD”, meaning “Combined”, product utilizes both Terra and Aqua data, hence providing the highest probability for quality input data. Moody *et al.*, 2004 have developed an algorithm to create a spatially complete, gap filled dataset through the following major steps: (1) the use of the high quality data for that time periods; (2) a phenological temporal curve fitting exercise for pixels without high quality retrievals; (3) a spatial fitting exercise; and as a last resort (4) a spatial smoothing effort.

The white sky albedo products in shortwave region of the electromagnetic spectrum have been used in this study.

3.2.5 Climatic Forcing (Rainfall Data)

From the India Meteorological Department (IMD), a high spatial resolution $0.25^\circ \times 0.25^\circ$ gridded, daily rainfall data, called IMD4, for 1901-2010 over the Indian subcontinent region have been used in this study (Pai *et al.*, 2014). In comparison with previous gridded datasets, due to its higher (better) spatial resolution and higher density of rainfall stations, the spatial rainfall distribution -

like heavy rainfall areas in the orographic regions of the west coast and over northeast, low rainfall in the leeward side of the Western Ghats etc. - are more realistically and better presented.

This data product is the fourth version of IMD gridded daily rainfall data developed in 2008 based on 6995 quality controlled rain-gauge stations over India. The data are arranged in 135 x 129 grid points which were created by interpolating the station data. The starting point of the grid is 6.50° North and 66.50° East.

A well tested interpolation method, named Shepard's method, was used to interpolate the station data into regular grids of 0.25 degrees. IMD rainfall data were made available for research work after several comparisons and quality control evaluations and the utility of this dataset is already proven in many meteorological studies.

3.2.6 Climatic Forcing (Temperature Data)

Another data set from IMD, a coarse resolution 1 degree gridded (approximately 110 km), daily temperature dataset was used for this study (Srivastava *et al.*, 2009). The daily temperature (Minimum, Maximum and Mean) data were collected from 395 quality controlled stations over India. IMD has prepared this dataset for the period 1969-2009 for research purposes. This data are arranged in 32x35 grid points. The unit of temperature is in degrees Celsius. For the leap year, data for 366 days were created. The data are available in binary and text format for every year.

IMD presently maintains around 550 geographically distributed surface observatories in the entire country, where daily surface air temperature observations (maximum and minimum) are recorded. These data are compiled, digitized, quality controlled and archived at the National Data Centre (NDC) of IMD. They have used a modified version of the Shepard's angular distance weighting Algorithm for the data preparation. Daily temperature data at a coarse resolution are useful for analysing extreme climate changes. For environmental modelling applications and validation of climate model simulations, the IMD temperature datasets have shown their efficiency in various studies.

3.2.7 Open Source packages and Programming for Geoprocessing

There is a plethora of packages available in the free and open source software (FOSS) domain for all types of applications. This study makes an attempt to utilize some for their agility and user-friendliness. The packages used in this study include (in alphabetical order):

EUREQA

Eureqa is a package (earlier available free to cost) which derives relationships between "natural" variables using user-specified arithmetic operators (Schmidt and Lipson, 2013).

GDAL

Geospatial Data Abstraction Library or GDAL is the de-facto standard for reading and writing spatial (raster & vector) data (GDAL, 2014). It was used standalone as well as an extension to the Python programming language and R programming language (R-GDAL (Hijmans, 2014)).

IDL

Interactive Data Language or IDL is a blazing fast array intensive programming language (Fanning, 2003).

MATPLOTLIB

The plotting and graphing engine for Python programming language (Hunter, 2007).

NUMPY

The numerical extension to Python programming language which is very useful for matrix operations (Dubois *et al.*, 1996).

PYTHON

A multi-purpose high level language which is the language of choice these days for geoprocessing (Downey, 2012).

R

R-programming language is widely used globally for statistical operations (R Development Core Team, 2008).

R-FDA

The functional data analysis extension to R-programming language was also used (Ramsay *et al.*, 2014).

SCIPY

The scientific extension to Python programming language which is like a Swiss knife for all statistical & scientific jobs (Jones *et al.*, 2001).

CHAPTER 4 - METHODOLOGY

This chapter describes the research methodology followed for pre-processing the land surface parameters, climatic forcing and gravity anomalies, conducting exploratory analysis, removing seasonality by creating temporal granules, and carrying out trend analysis and correlation studies to achieve the objectives of this study.

The three research sub-objectives require the study to have three distinct workflows.

4.1 Methodology for Sub-objective 1

The flowchart shown in Figure 4.1 aims to address the first research sub-objective which studies the compatibility between MODIS-NDVI and NOAA-AVHRR-NDVI and to inter-calibrate them to generate a combined time-series NDVI.

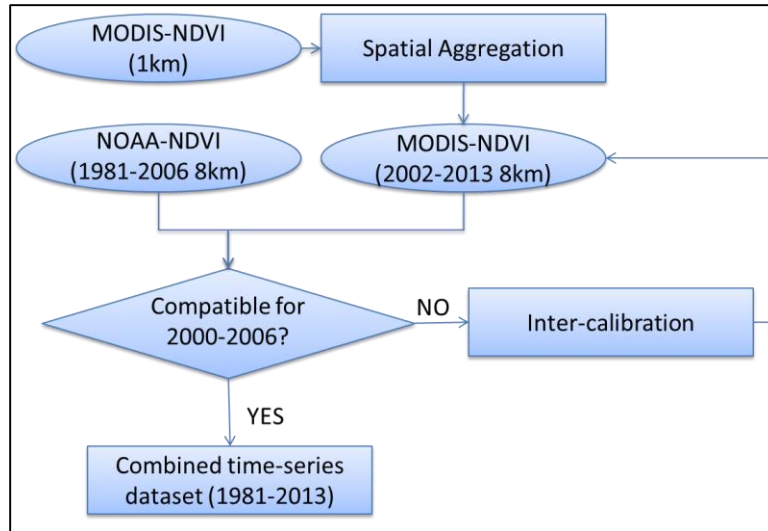


Figure 4.1 Flowchart for first sub-objective relating to creation of combined NDVI dataset for 1981-2013.

4.1.1 Data Preparation

To achieve the first objective of the proposed research, the MODIS NDVI (MYD13A2) and NOAA-AVHRR NDVI (GIMMS) datasets were downloaded and pre-processed. This was achieved in the following steps:

1. Identification of MODIS Tiles

The facilities REVERB¹ and GLOVIS² of NASA and USGS, were made use of to identify all MODIS tiles lying within the Indian sub-continent. India shape file from Survey of India was generalized to fit within 50 nodes. The tiles identified were h23v05, h24v05, h24v06, h24v07, h25v05, h25v06, h25v07, h25v08, h26v06 and h26v07 (10 tiles in all).

2. Downloading

¹<http://reverb.echo.nasa.gov/>

²<http://glovis.usgs.gov/>

A script in Python language was made to compile the web-address/links for identified files to be downloaded. The URL (<http://e4ftl01.cr.usgs.gov/>) stores all the files separated in different folders based on date of acquisition. A code to extract all URLs, match with tiles required and store them in a separate file was written and executed. A bulk downloading software was later used to read all the links and download those (5500+ files, ~69.6 Gbs). The workflow is shown in Figure 4.2.

3. Extraction to Geotiff file format

The downloaded Hierarchical Data File (HDF) format files are compressed files having multiple layers. The dataset required was at “Layer 0” and of data type “16 bit signed integer”. The Arcpy Python library’s function “ExtractSubDataset_management” was used in an iterative manner to extract same layer from all the files.

4. Mosaic

The 10 Geotiff files for each date needed to be mosaiced to form a single tile. The Arcpy Python library’s function “MosaicToNewRaster_management” was used to do this. Sequential processing of the data files required large computing time. Hence, a parallel computing approach was approached. The Arcpy library is not “thread-safe”, hence a “multiprocessing” based approach was used. The built-in “multiprocessing” Python library was used to create a pool and assign tasks were executed asynchronously.

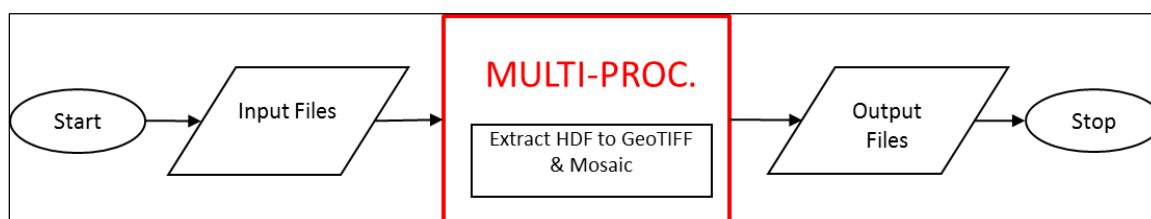


Figure 4.2 Flowchart for extracting single dataset from HDF files and mosaic using multi-processing

5. Pre-processing of GIMSS NDVI dataset

GIMSS dataset was downloaded from the University of Maryland Global Land Cover Facility Data distribution centre website. The files were available in Geotiff file format for global coverage (except Greenland and Antarctica). However, MODIS and NOAA datasets had many different data parameters as given in Table 4.1.

	MODIS	NOAA
Projection	Sinusoidal	Lat-Lon
Missing data values	-3000	-10000
File format	HDF	TIF
Resolution	1km	8km
Extent	Not compatible	

Table 4.1: Difference in MODIS and NOAA NDVI datasets

The Geospatial Data Abstraction Library (GDAL) utility was extremely functional to merge the two datasets into same parameters. The workflow shown in Figure 4.3 was adopted to synchronize the datasets. An example syntax of using the “gdalwarp” utility from GDAL in command line is given below:

```
gdalwarp.exe -t_srs EPSG:4326 -srcnodata -3000 -dstnodata -10000 -ts 403 417 -cutline
boundary.shp -te 68.0770 6.8229 97.3861 37.1502 -r average -of GTiff input.tif
resample/output.tif
```

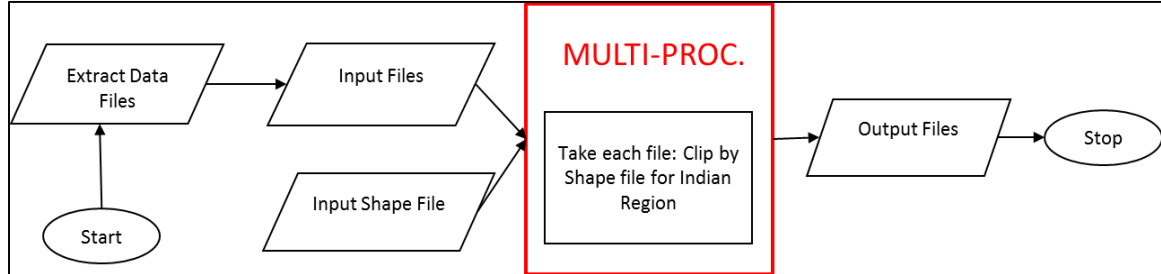


Figure 4.3 Flowchart to synchronize MODIS and NOAA datasets using multi-processing

4.1.2 Inter-Calibration

Once the MODIS and NOAA datasets were downloaded and preliminary data scraping was done, inter-calibration was carried out to generate a single combined long-term time-series NDVI dataset. To perform inter-calibration, the approach suggested by (Steven *et al.*, 2003) was implemented. The approach suggested a linear model (linear regression approach), as given in equation 4.1, to perform inter-calibration between MODIS and NOAA datasets.

$$NDVI_{reference} = a + b \cdot NDVI_{measured} \dots\dots\dots \text{Equation 4.1}$$

Steven *et al.*, 2003 have reported the coefficients ‘a’ and ‘b’ for many NDVI datasets. The coefficients for MODIS and NOAA inter-calibration reported by them are given in Table 4.2.

Independent NDVI	Dependent NDVI	
	AVHRR	MODIS
AVHRR		a=0.004, b=1.103
MODIS	a=0.003, b=0.904	

Table 4.2 Coefficients for inter-calibration of NDVI datasets (as reported by Steven *et al.*, 2003)

An attempt was made to experimentally derive the coefficients for the Indian sub-continent. The MODIS NDVI product used was MYD13A2 which is a 1 km resolution product. This was resampled to 8 km *i.e.*, to that of GIMMS NDVI resolution. Since some amount of averaging has taken place, the GIMMS NDVI is kept as the reference dataset and MODIS NDVI as the measured dataset.

To derive the coefficients ‘a’ and ‘b’, 100 random point samples were taken each in forested and non-forested areas in the Indian region. The forest and non-forest areas were delimited using the DOS-DBT project data (Roy *et al.*, 2012). The data from overlapping period for GIMMS NDVI and MODIS NDVI *i.e.*, July 2002 - December 2006, was extracted for the 200 points (100 forest + 100 non-forest) and regressed. As can be seen in the results, shown in Figure 4.4, the spread of the data points is large and no consensus can be derived from the coefficient of determination values.

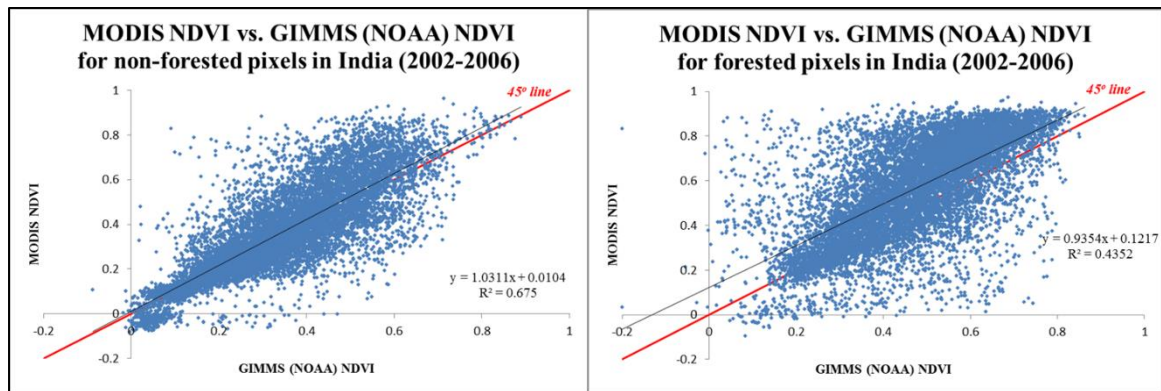


Figure 4.4 Regression of time-series NDVI data from MODIS and NOAA satellites over forested and non-forested random points in Indian sub-continent (2002-2006)

Steven *et al.*, 2003 carried out the same experiments in a controlled environment using calibrated test sites and has proposed coefficients values as in Table 4.2. Hence, in the current study, equation 4.2 below is used to make a long term NDVI data set.

$$NDVI_{reference} = 0.003 + 0.904 \cdot NDVI_{measured} \dots \dots \dots \text{Equation 4.2}$$

Another GDAL utility ‘gdal_calc.py’ utility was used to perform the 'raster calculator' operation, to bring the measured NDVI (MODIS in this case) up to the reference NDVI (GIMMS in this case). Example syntax to run ‘gdal_calc’ on all files in a folder in windows command line is as given below:

```
Resample_files_folder> for %i in (*.tif) do gdal_calc -A %i --outfile=
Output_files_folder\%i --calc="0.003+0.904*A" --NoDataValue=-10000
```

GIMMS NDVI dataset from July 1981 till December 2006 was joined with the modified MODIS NDVI from January 2007 till December 2013.

4.2 Methodology for Sub-objective 2

The flowchart shown in Figure 4.5a and 4.5b aim to address the second research sub-objective which attempts to study the long-term trends in temporal aggregates of NDVI, albedo, temperature and precipitation.

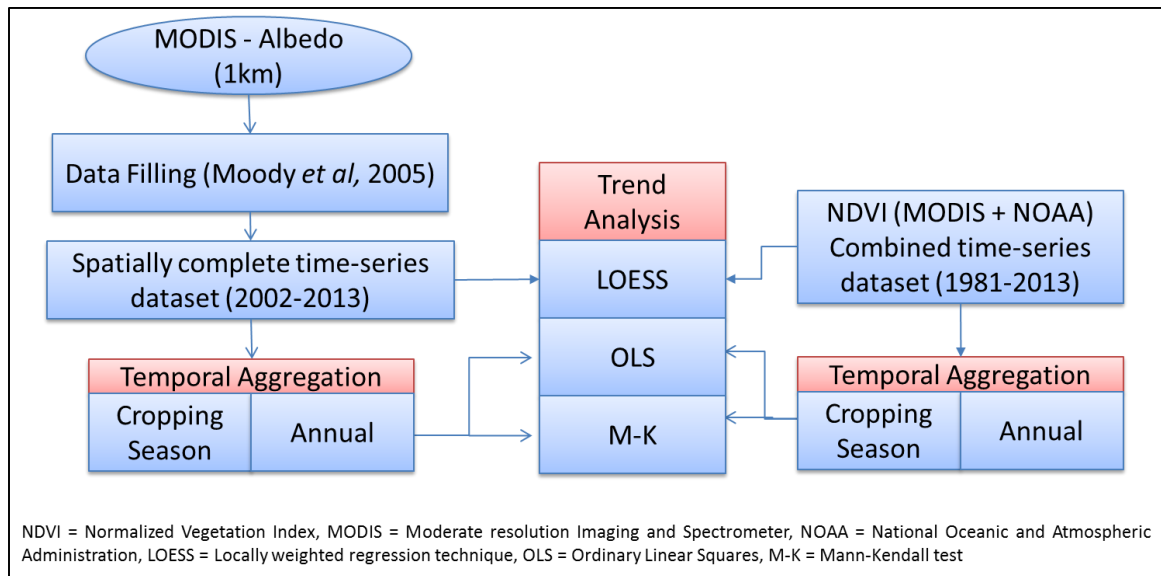


Figure 4.5a Flowchart for second sub-objective relating to NDVI and Albedo trends

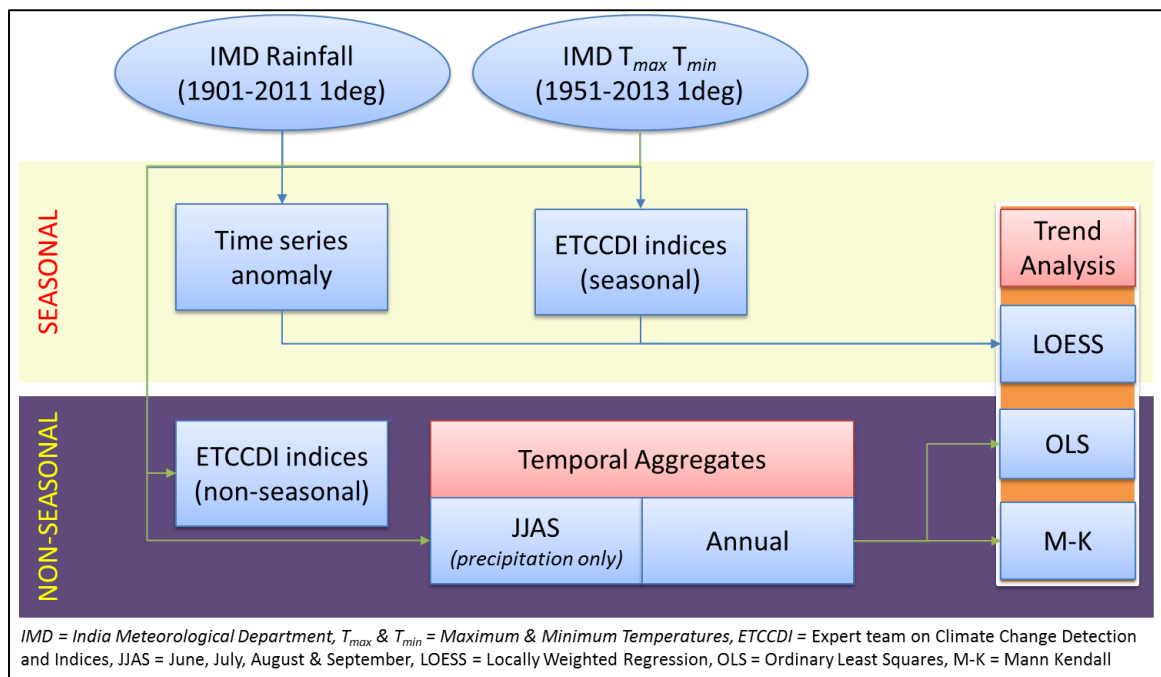


Figure 4.5b Flowchart for second sub-objective relating to climatic trends

4.2.1 Data Preparation

Albedo

The process for pre-processing MODIS NDVI data (as described in Section 4.1.1) was also used to prepare the Albedo dataset. The MODIS product “MCD43B3” was used. The “Layer 17” and “Layer 19” were extracted and mosaiced. They are the Visible and Shortwave white sky albedo products. GDAL was used to do preliminary data pre-processing, an example for which is shown below:

```
Raw_Albedo_files_folder>for /r %i in (*.hdf) do gdalwarp.exe -t_srs EPSG:4326 -
srcnodata 32767 -dstnodata 32767 -cutline boundary.shp -of GTiff -te 68.14156741
8.03336612 97.32488907 37.14168781 %i "%~ni.tif"
```

Albedo data is prone to quality issues due to presence of clouds and atmospheric interferences at the time of acquisition. Hence, to use the best available representation of time-series albedo, the dataset developed using Moody *et al.*, 2004 were used. The algorithm creates a spatially complete, gap filled through the following major steps: (1) the use of the high quality data for that time periods; (2) a phenological temporal curve fitting exercise for pixels without high quality retrievals; (3) a spatial fitting exercise; and as a last resort (4) a spatial smoothing effort.

Climatic data (Rainfall and Precipitation)

The climatic forcing data obtained from the India Meteorological Department (IMD) available in GRD (binary) file format was used. IDL codes were written to; (1) convert them to spatially referenced TIFF file format, (2) find average rainfall & temperature (min and max) (ignoring pixels having missing values), and (3) creating anomaly data for entire time-series of daily rainfall and temperature (min and max) datasets. The anomaly data were created by subtracting daily climatic data from climatic normal generated by averaging daily data for the period.

To better understand the climatic variables, apart from the time-series anomaly data, a few indices given by Expert team on Climate Change Detection and Indices (ETCCDI) were also prepared. Both types of indices, seasonal and non-seasonal, were chosen for both rainfall and temperature data, as given in Table 4.3. The index numbers given are as per (etccdi.pacificclimate.org).

	Seasonal	Non-Seasonal
Tmax	6. TXx, Monthly maximum value of daily maximum temperature	2. SU, Number of summer days
		13. TX90p, Percentage of days when TX > 90 th percentile
Tmin	9. TNn, Monthly minimum value of daily minimum temperature	4. TR, Number of tropical nights
		10. TN10p, Percentage of days when TN < 10 th percentile
Rainfall	17. Rx1day, Monthly maximum 1-day precipitation	23. CDD. Maximum length of dry spell
	18. Rx5day, Monthly maximum consecutive 5-day precipitation	24. CWD. Maximum length of wet spell

Table 4.3 List of ETCCDI indices chosen for current study

Python programs were written to derive these indices from IMD dataset. The name and description of seasonal indices indicate clearly the data created. Index 2 (SU) is the annual count of days when TX (daily maximum temperature) is >25°C. Similarly, Index 4 (TR) is the annual count of days

when TN (daily minimum temperature) is $>20^{\circ}\text{C}$. Indices 23 and 24 (CDD and CWD) are count of consecutive days (maximum length) of dry ($< 1\text{mm}$) and wet ($\geq 1\text{mm}$) daily rainfall.

Derivation of Indices 13 and 10 (TX90p and TN10p) is a two-step process. In first step, the 10th and 90th percentile temperature values are derived from minimum and maximum temperature datasets respectively, the definition for which is given in Table 4.4. In this study, the base period has been selected as the entire duration for which the temperature data are available i.e., 1951-2013.

TX90p, Percentage of days when $\text{TX} > 90^{\text{th}}$ percentile:

Let TX_{ij} be the daily maximum temperature on day i in period j and let TX_{in90} be the calendar day 90th percentile centred on a 5-day window for a base period 1990. The percentage of time for the base period is determined where:

$\text{TX}_{ij} > \text{TX}_{in90}$

TN10p, Percentage of days when $\text{TN} < 10^{\text{th}}$ percentile:

Let TN_{ij} be the daily minimum temperature on day i in period j and let TN_{in10} be the calendar day 10th percentile centred on a 5-day window for a base period. The percentage of time for the base period is determined where:

$\text{TN}_{ij} < \text{TN}_{in10}$

Table 4.4 Definition of ETCCDI Indices 13 and 10

The second step is finding the number of days having temperatures greater than or lesser than the percentile value. This can be best explained as shown in Figure 4.6 which shows 315 data points in the temporal domain for a single pixel (24°N, 81°E) for 5 consecutive dates (January 01 to January 05) for the years 1951 till 2013, the 10th & 90th percentile lines and the histogram of all the values.

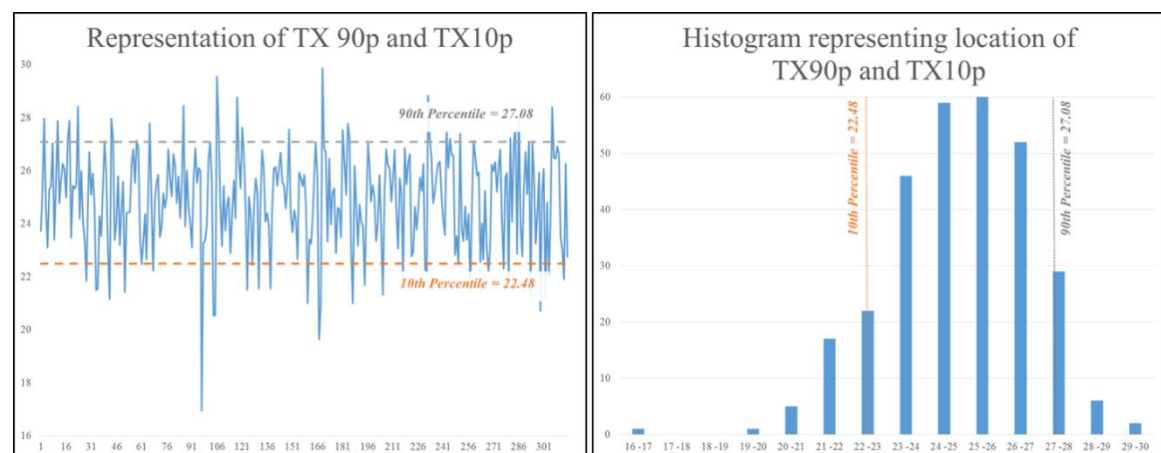


Figure 4.6 Graph and histogram of 5 day temperature for 63 years data indicating 10th & 90th percentile boundaries

4.1.3 Temporal Aggregation

Programs in IDL language were written to derive the temporal aggregates of consolidated NDVI and Albedo time-series data. As the NDVI data are available fortnightly, two files from each month are taken for the time period under consideration, and averaged - ignoring missing values. The following aggregates were prepared:

1. Annual (24 files per year)

2. Cropping season:

- a. Rabi (December - February) (6 files per year)
- b. Kharif (July - September)(6 files per year)
- c. Zaid (March - June) (8 files per year)

The MODIS acquisition dates were taken into consideration for the temporal aggregation of Albedo data. For the Climatic datasets, annual aggregates were prepared for temperature and rainfall. Additionally, monsoon period (June, July, August and September or JJAS), rainfall data was also aggregated.

These consolidated NDVI, Albedo, Climatic anomaly, ETCCDI indices time-series datasets and their temporal aggregates were used for trend analysis as explained in section 4.1.4.

4.1.4 Trend Analysis

One of the primary objectives of MTUP problem is to study the effects of differently sized time granules on regression patterns obtained from them. The time domain can be modified suitably such as to create time granules which are ‘deseasonalized’, i.e., the effect of repetitive seasonal variations are minimized. The consolidated time-series data has seasonal cycles hence not suitable for direct regression techniques, but is an ideal candidate for Locally Weighted Regression (LOESS) technique described next. The temporal aggregates based on annual and cropping season cycles are suitable candidates for direct regression techniques such as OLS and Mann-Kendall test with Sen’s slope estimator as explained later in this section.

LOESS

Seasonal decomposition of time-series (STL) using LOESS (Cleveland *et al.*, 1990) is used to decompose the seasonal time-series (NDVI, Albedo, Climatic anomaly, Seasonal ETCCDI) as given in the equation 4.3.

$$Y_t = T_t + S_t + R_t \dots \dots \dots \text{Equation 4.3}$$

Where Y_t is the data value at time t , T_t is the trend component, S_t is the seasonal component and R_t is the residual component.

The LOESS technique was simplified to take 2 main arguments - (1) s.window - the span of the LOESS window for seasonal extraction, and (2) t.window - the span of the LOESS window for trend extraction. A ballpark cropping season length of 2.5 months gives s.window value as 5 spans (at 2 spans in 1 month) and less than 5 years for trend extraction gives t.window as 100.

Figure 4.7(a) shows the actual NDVI data points for a pixel (row 200, column 150) for the entire duration from 1982 till 2013 (32 years * 24 bi-weekly datasets = 768 data points). Using the technique described above, the seasonal, residual and long term trends were derived. The mean of the seasonal trend was subtracted from itself and plotted in (b). Figure 4.7(c) shows the long-term trend component derived using 100 as t.window. Figure 4.7(d) shows the residual derived from subtracting smoothed seasonal data from the original data points.

The histogram (Figure 4.8) of the residuals shows that there is even distribution and hence no bias or auto-correlation exists after smoothing.

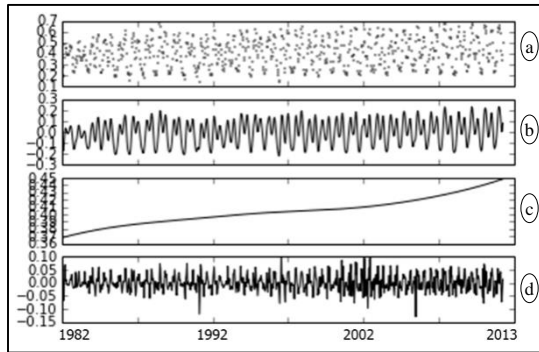


Figure 4.7 Trend decomposition using LOESS (a) original data for 768 data points (in time domain), (b) seasonal trend (anomaly), (c) long term trend, and (d) residual component

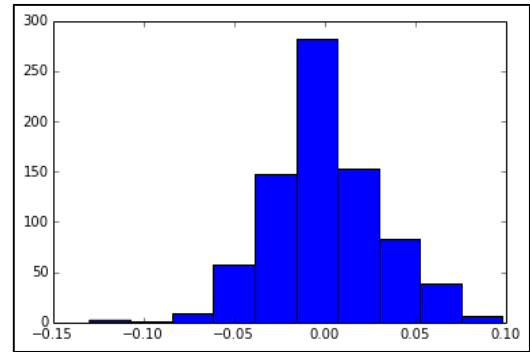


Figure 4.8 Histogram of residuals obtained from LOESS

```

3450 - done - 144.71 s
3451 - done - 147.53 s
3452 - done - 145.56 s
3453 - done - 145.27 s
3454 - done - 147.45 s
3455 - done - 119.70 s
3456 - done - 115.31 s
3457 - done - 116.32 s
3458 - done - 116.20 s
3459 - done - 116.62 s
3460 - done - 121.06 s
3461 - done - 120.11 s
3462 - done - 115.35 s
3463 - done - 116.05 s
3464 - done - 84.78 s
3465 - done - 89.77 s
3466 - done - 86.48 s
3467 - done - 84.44 s
3468 - done - 84.19 s
3469 - done - 85.16 s
3470 - done - 84.21 s
3471 - done - 83.90 s
3472 - done - 61.61 s
3473 - done - 61.05 s
3474 - done - 62.80 s
3475 - done - 61.72 s
3476 - done - 61.85 s
3477 - done - 60.83 s
3478 - done - 62.73 s
3479 - done - 62.92 s
3480 - done - 61.95 s
3481 - done - 0.03 s
3482 - done - 0.00 s
3483 - done - 0.00 s
3484 - done - 0.02 s
3485 - done - 0.00 s
3486 - done - 0.00 s
3487 - done - 0.00 s
3488 - done - 0.00 s
3489 - done - 0.00 s
3490 - done - 0.00 s
3491 - done - 0.00 s
3492 - done - 0.00 s
Time 4853203.28 s
>>>

```

The STL/LOESS algorithm is computationally expensive and both Python and IDL programs were written and time of execution compared for both on the Albedo dataset (552 files; dimensions 3502 x 3493; 16 bit integer data). STL on Albedo when performed using an improvised code in IDL programming language gave results in 1 day; however, sampling variance for some datasets could not be determined. The Python code took 56 days (4,853,203.28 seconds) to run (update: 26 December, 2014) as shown in the screenshot on the left.

OLS

LINREGRESS routine from SCIPY.STATS module in Python programming language has been used to perform trend analysis by fitting the paired data(x_i, y_i) to the linear model, $Y = mX + c$. A two tailed p-test with 95% confidence interval (p value less than 0.05) was also calculated and multiplied with the derived slope to return only significant slope values.

Mann-Kendall Test and Sen's Slope Estimator

First and Second lag auto-correlation tests were done, and then presence of trend was determined using the Mann-Kendall Test. A Python program with some optimization and parallelization could achieve the desired objectives, including distributed Sen's Slope estimates for the entire Indian sub-continent. An image showing only the Sen Slope values at pixels having p-value (derived from Mann-Kendall test) less than 0.05 (95% confidence interval) is also shown, as given in the example shown in Figure 9. Figure 9a&b shows the spatial variation of Sen Slope estimate and p-value as derived from Mann-Kendall test. A mask is created based on p-values less than 0.05 and applied to the Sen Slope estimates to generate the significant Sen Slope estimate map, as shown in Figure 4.9.

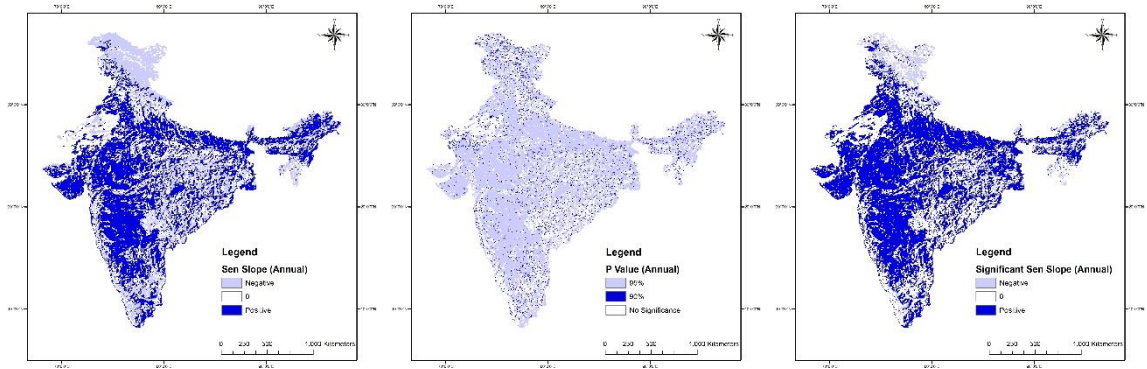


Figure 4.9 Example of clipping of Sen Slope estimates at 95% confidence interval ($\alpha < 0.05$)

4.3 Methodology for Sub-objective 3

The flowchart shown in Figure 4.10 aims to address the third research sub-objective which studies the correlation in the trends in different datasets including gravity anomaly for the past decade.

4.3.1 Data Preparation

The GRACE gravity anomaly data came as global files in 1 degree resolution. Clipping them to India extent and interpolation them (using nearest neighbour) to half a degree resolution was done using GDAL, samples of which are shown in Sections 4.1.1, 4.1.2 and 4.2.1.

Albedo, NDVI and Climatic datasets were resampled to half a degree spatial resolution, averaged in temporal domain to create monthly aggregates (to match that of gravity anomaly data) and masked using a common mask using GDAL commands, Python and IDL programming.

'Normals' for all datasets were derived and subtracted from the entire time-series to generate the anomaly time-series datasets. This anomaly dataset for all six datasets was used further for correlation.

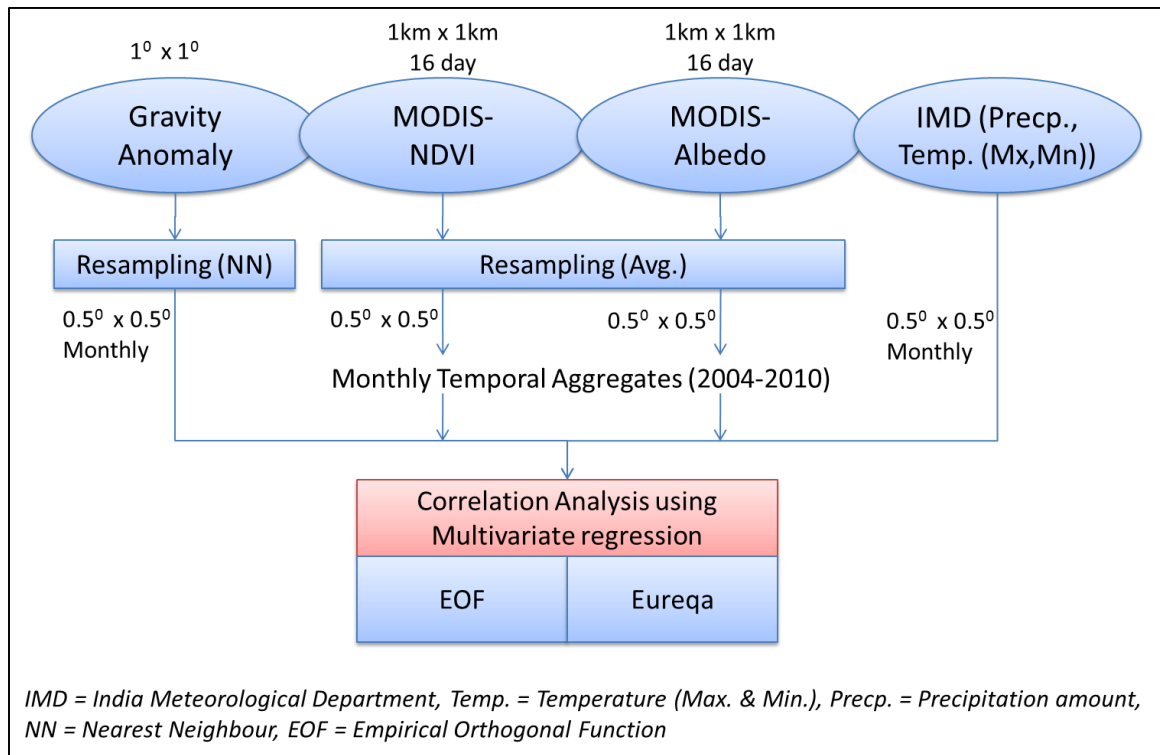


Figure 4.10 Flowchart for sub-objective 3

4.3.2 Correlation

The correlation among six variables was attempted using the EOF technique and the Eureqa package. The basic premise behind this part of the study is to understand the dependence of gravity anomalies on other physical processes. The terrestrial water component, which comprises of water stored as surface water, soil moisture, moisture at the vadose zone and groundwater, have been reported to be the major component effecting gravity. Negative impact due to surface irrigation and ground water extraction also impacts gravity (Tiwari *et al.*, 2009). Attempts have been made to proxy the effect of these factors by using NDVI and Rainfall datasets.

EOF

EOF or PCA has generated immense interest in research community for its capability to generate statistically derived “data” modes which may represent “standing oscillations” or “propagating signals.” This multivariate statistical technique aims to “reduce a dataset containing a large number of variables to a dataset containing fewer (hopefully many fewer) new variables.” These new signals when applied to many fields (or variables) may help “uncover meaningful joint relationships among fields in an exploratory setting, where clues to possibly unknown underlying physical mechanisms may be hidden in the complex relationships among several fields” (Wilks, 2011).

Hence, the primary usefulness of EOF is (1) to identify a small subspace that contains most of the dynamics of the observed natural system, and (2) to identify modes of variability.

The EOFs are found by computing the eigenvalues and eigenvectors of a spatially weighted anomaly covariance matrix of a field. Hence to perform EOF for a single field, the following steps are carried out:

1. The input variables are converted to anomaly data by subtracting the mean of entire dataset with each layer. In the current study all 6 datasets were resampled spatially and temporally to $0.5^\circ \times 0.5^\circ$ grids and 1 month respectively. They were then averaged for the entire time duration 2004-2010 and subtracted from each month layer.
2. Most commonly, the spatial weights are the cosine of the latitudes or, better for EOF analysis, the square root of cosine of latitude values. These are calculated and multiplied by each layer.
3. Covariance matrix of $k \times k$ dimensions is calculated where k is the number of grids for which the spatial EOF is being run.
4. The eigenvectors and eigenvalues are calculated. The derived eigenvalues provide a measure of the percent variance explained by each mode. The eigenvectors represent the pattern of variance in the spatial scale. The units are totally arbitrary. The only importance of the values is whether they are positive or negative.
5. To interpret the patterns, the principal components are calculated, which are the spatially weighted anomaly data projected onto the eigenvectors. The principal component attached to the corresponding eigenvector spatial pattern provides the sign and overall amplitude of the pattern as a function of time.

The steps above will help explain the spatial variability in single fields. The process is modified to explain variability in a set of L fields. The anomaly data is stacked together such that the covariance matrix is $kL \times kL$. The modes derived by such modification explains the joint variability in the datasets.

The EOF technique was used with codes written in both R and IDL programming language. Multiple libraries from the R programming language were used such as raster, fda, plyr, rgdal, sp, zoo, base, cluster, tcltk, stringr.

Eureqa

The Eureqa software provides a user friendly interface and is designed to work on noisy data. The results of the symbolic regression are presented in an intuitive way. Symbolic regression, as implemented in Eureqa, creates an initial subset of functional forms from user-specified building blocks, which are stored as the operator and terminal sets. These building blocks can consist of a range of operators, including arithmetic, trigonometric, exponential, etc. Using genetic programming methods, Eureqa generates a number of solutions, each offering a potential function that explains the target data by input data.

Eureqa automatically splits the input data into training and validation datasets. The training set is used to generate and optimize solutions, and the validation set is used to test how well those models generalize to new data. Eureqa randomly shuffles the data and splits it into training and validation data sets based on the total size of your data. Training data is taken from the start of dataset and validation data will be taken from the end (after shuffling). For very large data sets (like in current study) it uses a smaller fraction of data for training ($\leq 10,000$ data points) and remaining data are used for validation.

Data from all pixels were reformatted into ASCII and fed into Eureqa for the entire time-series. Handling missing data and remove outliers (Threshold: 2.00 multiple of IQR) were the data preparation instructions fed into the software.

The basic building blocks for the target expression (Equation 4.4 and 4.5) was specified as below:

- Basic (+, -, *, /), Trigonometric (sin, cos), Exponential (exp, log, ^) and History (Simple moving average)

The process was attempted first with 2 independent variables NDVI (n) and Rainfall (r) and 1 dependent variable Gravity anomaly (g) for the target expression given in Equation 4.4.

$$g = f(n, r) \dots \dots \dots \text{Equation 4.4}$$

The model failed to converge beyond 70% even after running for over 6 hours on an Intel® i7-3.60Ghz 8 core CPU, 64 bit machine. The model achieved was only 6% stable & 55% mature and gave low coefficient of correlation and r^2 (Goodness of Fit) values.

The process was attempted next with all 5 independent variables and 1 dependent variable for the target expression given in Equation 4.5.

$$g = f(a, n, r, tx, tn) \dots \dots \dots \text{Equation 4.5}$$

Where g, a, n, r, tx and tn denote Gravity, Albedo, NDVI, Rainfall, Maximum temperature and Minimum temperatures respectively. The process was run for just over 3 hours on an Intel® i7-3.60Ghz 8 core CPU, 64 bit machine. The model was stopped when it reached 100% convergence and model was 90%+ stable, as shown in Figure 4. 11.

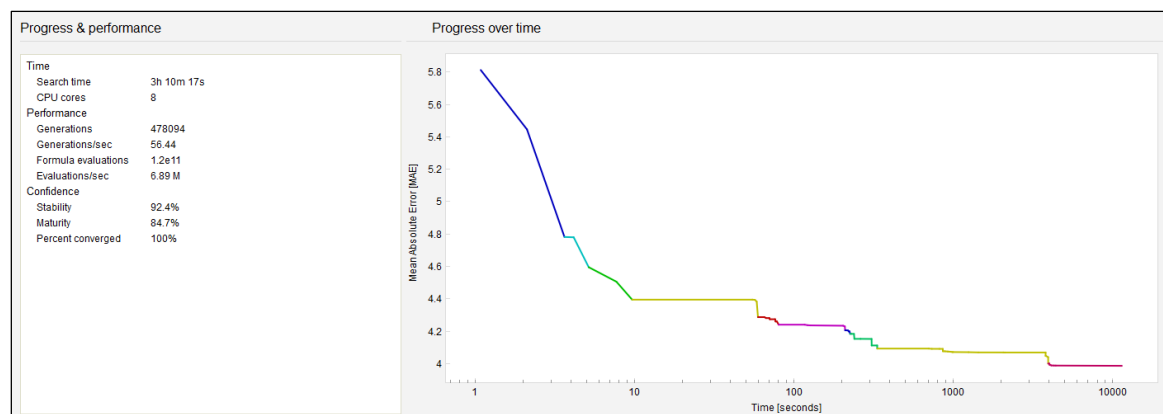


Figure 4.11 Performance parameters for Eureka correlation run

The results obtained are discussed in next chapter. The coefficients obtained for Equation 4.4 were applied to the original independent variables and regressed GRACE anomaly was also compared with the original.

CHAPTER 5 - RESULTS AND DISCUSSION

The study attempts to understand the effect of temporal granularity i.e., the effect of selecting differently sized temporal aggregates to do statistical trend analysis for spatio-temporal datasets over Indian sub-continent. The results are categorised to answer the research sub-objectives.

5.1 Combined Time-Series NDVI

Section 3.2.1 and 3.2.2 discussed about the two NDVI datasets used (NOAA-GIMMS and MODIS-MYD13A2). Section 4.1 discussed about the process followed to inter-calibrate the two datasets.

After spatial aggregation (averaging algorithm), of MODIS NDVI dataset (1 km spatial resolution) to GIMMS NDVI dataset (8 km spatial resolution), the calibration model was applied on MODIS NDVI and resultant dataset was merged with GIMMS NDVI, to provide the final consolidated time-series NDVI dataset. Total 780 files were obtained (2 fortnights per month * 12 months per year * 32 ½ years (July 1981 - December 2013)) in this dataset.

Plots between GIMMS NDVI, MODIS NDVI before and after calibration, for a forested pixel (Figure 5.1) and a non-forested pixel (Figure 5.2), for the overlapping period (i.e., July 2002 - December 2006), were drawn. As can be clearly seen, the calibration model attempts to bring the NDVI values closer.

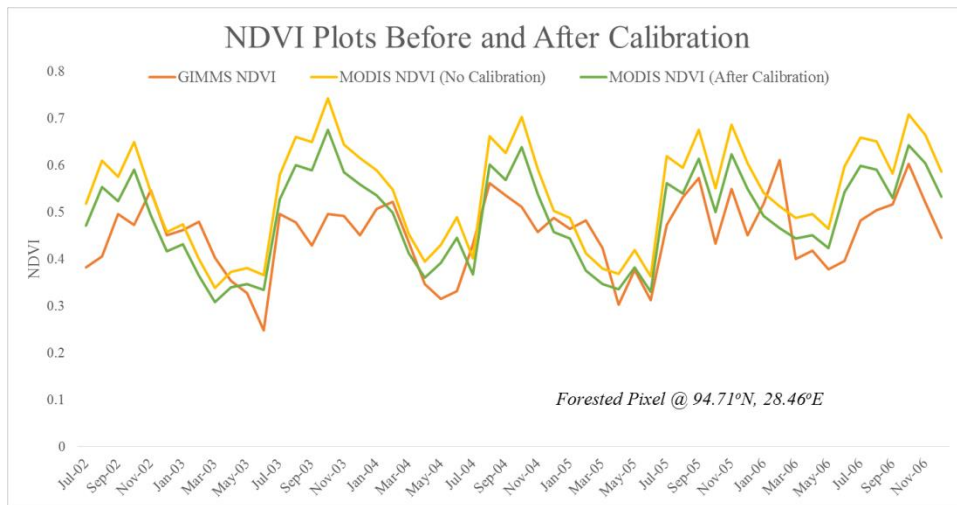


Figure 5.1 NDVI plots before and after calibration for a forested pixel

Random 160 points in forested and non-forested areas were used to test the goodness of fit. The r-squared was found to be 0.72 with a majority of the points lying on the 45° line as shown in Figure 5.3.

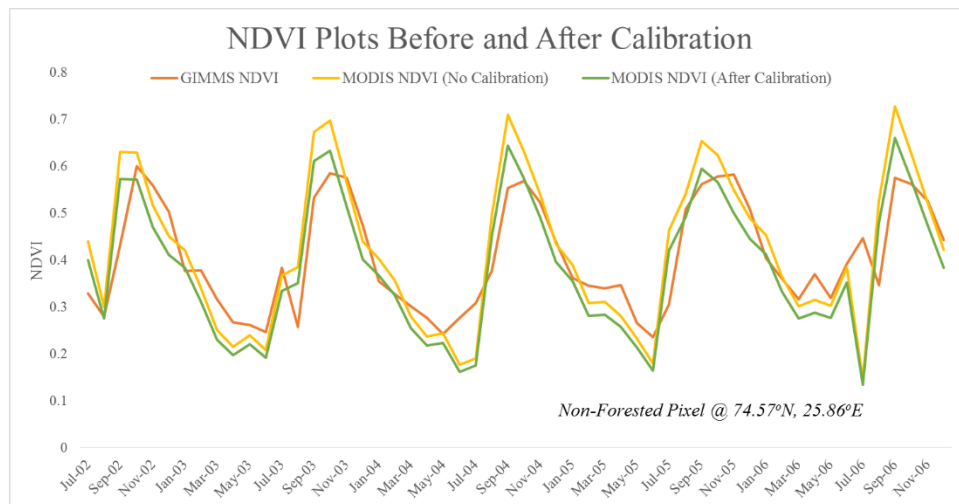


Figure 5.2 NDVI plots before and after calibration for a non-forested pixel

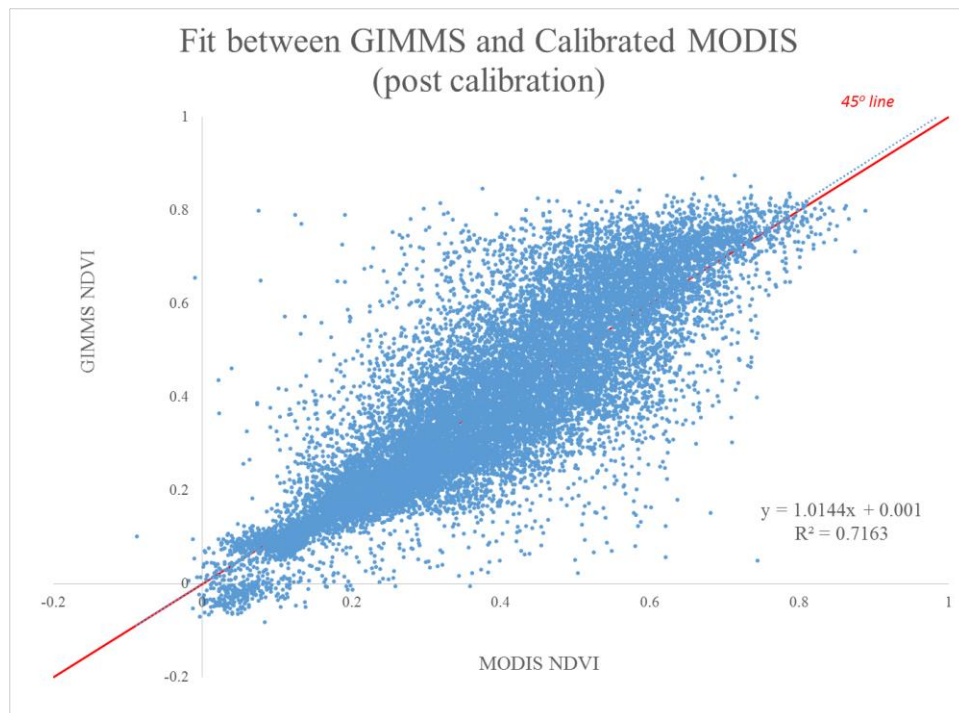


Figure 5.3: Scatter plot of GIMSS NDVI against MODIS NDVI (after calibration) for 160 points for time duration July 2002 - December 2006 (fortnightly data)

5.2 Trends in time-series datasets

Section 3.2.1, 3.2.2, 3.2.4, 3.2.5 and 3.2.6 discussed about the NDVI, Albedo, Precipitation and Temperature datasets. Section 4.2 discussed about the process followed to find long-term trends in temporal aggregates of NDVI, albedo, temperature and precipitation.

5.2.1 NDVI trends

STL/LOESS on NDVI

As discussed in Section 4.1.4, STL/LOESS was applied on the 32.5 years (15 day temporal resolution) NDVI data from July 1981 till December 2013. To obtain insight into the NDVI dynamics, the magnitudes of the trend, seasonal and residual (irregular) components were compared. To do this, the sample variance of each component is calculated as a percentage of the variance of the original NDVI time-series data.

Figure 5.4a maps the variances in the original series. Higher variances in NDVI (>0.04) are observed in Punjab, Haryana, Arunachal Pradesh and the Western Ghats, among other sporadic occurrences throughout the country. Low variances (<0.01) are observed in the Western Himalaya, Western Plain and Kutch Peninsula and parts of Eastern Ghats.

Figure 5.4b shows the contribution of the STL seasonal components to the overall variance across the country. Seasonality primarily governs the decomposition of variances with values ranging from 60-100%. Lesser impact of seasonality is found in Gangetic plains, parts of Bengal and Odisha Plain and Western Ghats.

Figure 5.4c shows the contribution of the STL trend components to the overall variance. Trend is not a major component of variance. Higher percentages of trend components are seen in the Western Himalaya, Kutch and Kathiawar peninsula, Deccan plateau and Assam plains.

The median percentage of variance due to irregular component in NDVI values is about 7% (Fig. 5.4d). Higher percentages of irregular components in NDVI values are observed in entire North East India (except Assam plains), Eastern and Western coastal plains.

For the purpose of visualizing the regional structure in STL-derived trends, it is convenient to reduce each one to a single number. Here, for each pixel an index of overall annual change has been defined as $(T_n - T_1)/n$, where $n = 32.5$ is the number of years of record in the analysis period 1981-2013 and T_1 & T_n is the value of the STL trend component in the 1st & last year of the record as defined above.

Figure 5.5 shows a map of the long-term trends calculated in this way. Positive trends dominate the Indian land mass however negative trends are observed in Northern Himalayas, upper reaches of North Eastern Hills, Rajasthan, Western Ghats and parts of Central plateau. Positive trends in NDVI could be because of increase in agriculture propelled by irrigation development. This is in agreement with the increase in net irrigated area reported in Lee *et al.*, 2009.

Overall, the pattern is very similar to that derived in Figure 5.6a using linear regression. The STL trends are, however, more realistic than trends estimated by linear regression.

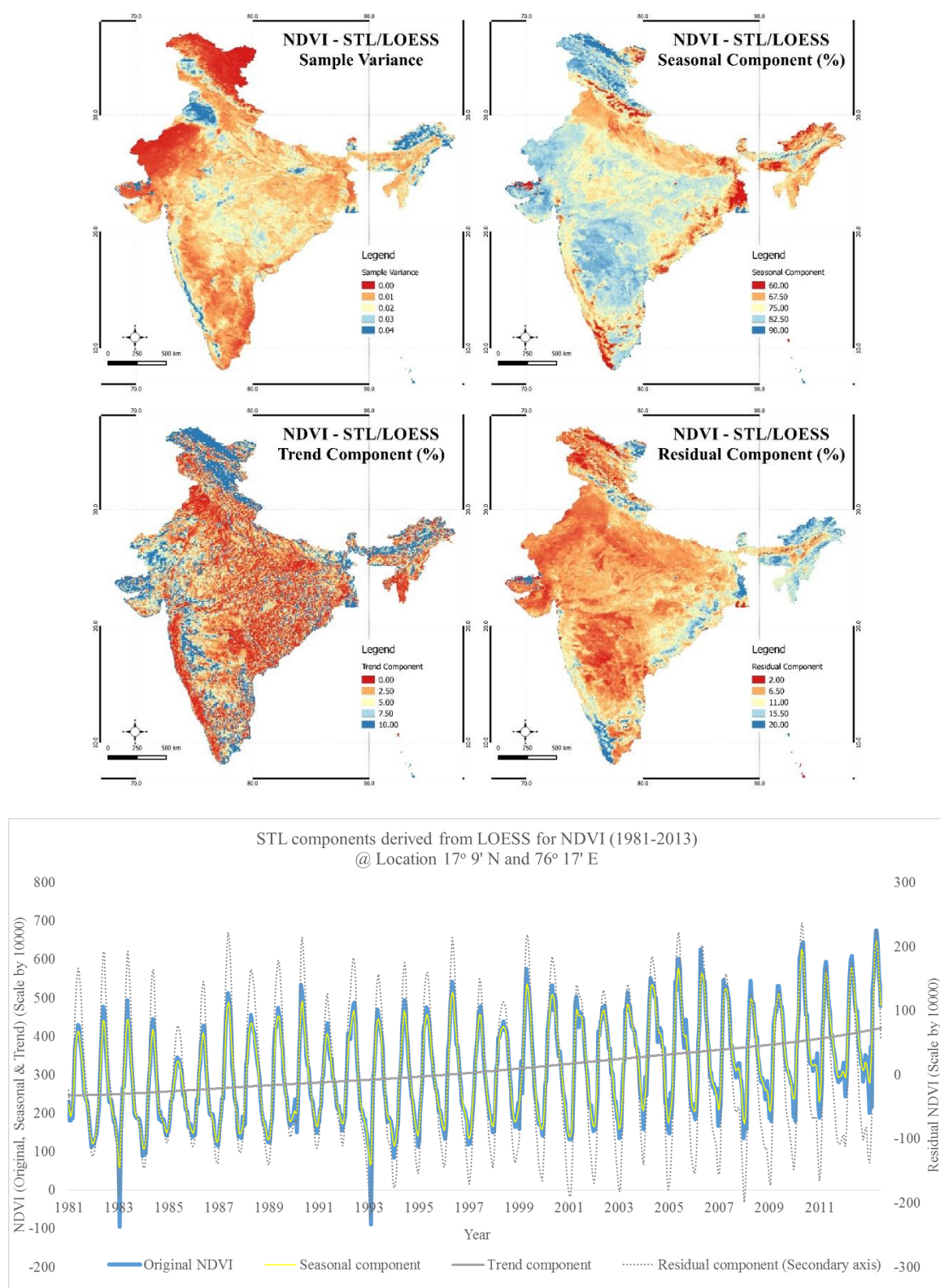


Figure 5.4 Results of STL/LOESS on NDVI. Variance of NDVI time-series (a); relative proportions of NDVI variance contributed by seasonal, trend and residual (irregular) components are shown in (b–d) respectively; and STL components for representative pixel are shown in (e). (July 1981–December 2013)

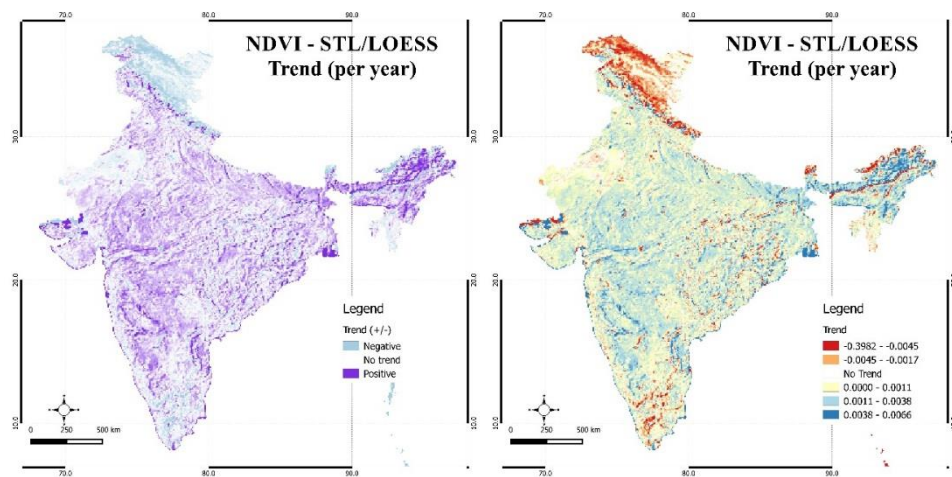


Figure 5.5 Trends in NDVI derived from STL/LOESS results (July 1981-December 2013)

OLS & M-K on Annual Temporal Aggregates of NDVI

Annual temporal aggregates were prepared as discussed in Section 4.1.3 and OLS & M-K test were applied as discussed in Section 4.1.4. Figure 5.6 shows the results of OLS and M-K techniques on annual temporal aggregates of combined NDVI dataset (1982-2013). The results have been masked at 95% confidence interval (p-value less than 0.05).

Figure 5.6a represents the slope values of the OLS technique in 3 main categories, i.e., Negative trend (values less than 0.0), Zero or No trend (values equal to 0.0) and Positive trend (values greater than 0.0). The majority of the area in the India sub-continent shows positive trend which also resembles the area under cropping, such as Gangetic plains, Central highlands & Kathiawar peninsula, Deccan plateau and Assam & Bengal plains. Portions of the Eastern plateau and Eastern Ghats and large tracts of Western Himalaya show negative trends.

Figure 5.6b represents the slope values the slope from the Sen's slope estimator. The results are similar to that obtained from OLS technique. It is observed that the area under negative trend has reduced and has been replaced by no significant trend.

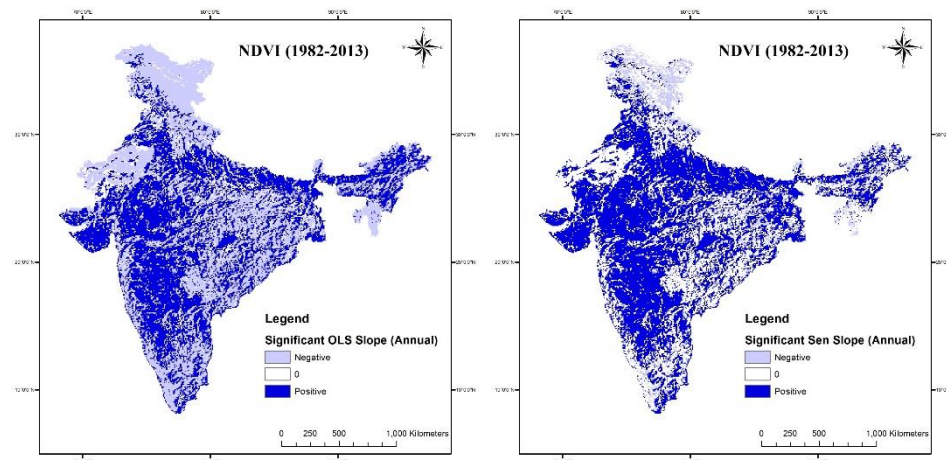


Figure 5.6 Significant slopes derived from OLS & M-K / Sen's estimator on Annual Temporal Aggregates of NDVI (1982-2013)

OLS & M-K on Cropping Cycle Temporal Aggregates of NDVI

The temporal aggregates for cropping cycles (prepared as described in Section 4.1.3) were subjected to the OLS and Sen's slope estimators for derivation of trends in the datasets. The trends, as shown in Figure 5.7, were masked for significant values (at 95% confidence interval). The cropping seasons used, namely Kharif, Rabi and Zaid are described in 3.1.5.

The Kharif or monsoon crop aggregated data shows positive trends in majority land mass area of the country and mimics the annual temporal aggregate trends (Figure 5.7a&b). The slope values or trends in OLS and Sen's estimator are similar to each other; except that the negative trends in OLS were replaced by no significant trend in Sen's estimator plus M-K test outputs. The negative trends are more pronounced in upper reaches of Northern Himalayas.

Figure 5.7b-e shows the slope trends in NDVI aggregated for Rabi (December-February) and Zaid (March-June) cropping cycle. NDVI trends for Rabi or winter season crop reflects the areas under Rabi wheat (Punjab, Haryana, parts of Uttar Pradesh, Madhya Pradesh and Rajasthan) and rice (parts of Gangetic Assam and Bihar plains) crops. Areas in J&K, Eastern plateau (Chhattisgarh and Odisha) and parts of Mizoram and Tripura show negative NDVI trends. Zaid cropping cycle trends show positive values in Gangetic plains, North Eastern hills and Deccan plateau & negative trends in Punjab & Haryana.

5.2.2 Albedo trends

STL/LOESS on Albedo

As discussed in Section 4.1.4, STL/LOESS was applied on the 12 years (8 day temporal resolution) albedo data from January 2002 till December 2013. To obtain insight into the albedo dynamics, the magnitudes of the trend, seasonal and residual (irregular) components were compared. To do this, the sample variance of each component is calculated as a percentage of the variance of the original albedo time-series data.

Figure 5.8a maps the variances in the original series. Higher variances in albedo (90-100%) are observed in the Western and Eastern Himalaya and in parts of Kutch peninsula. Medium variances (50-89%) are observed in Northern plains, Central Highlands, some parts of Deccan plateau and in parts of North Eastern hills. Low variances are observed in rest of India.

Figure 5.8b shows the contribution of the STL seasonal components to the overall variance across the country. Generally, seasonality is the primary component of variance in albedo except in Western and Southern regions of the country. Albedo values that are dominated by the seasonal component (>80% of variance) occur mainly in the Gangetic plains, Western Ghats and North Eastern hills. Seasonality is the least important component in the Western India (Rajasthan, Punjab, and Haryana), Southern Karnataka and extreme reaches of Jammu and Kashmir.

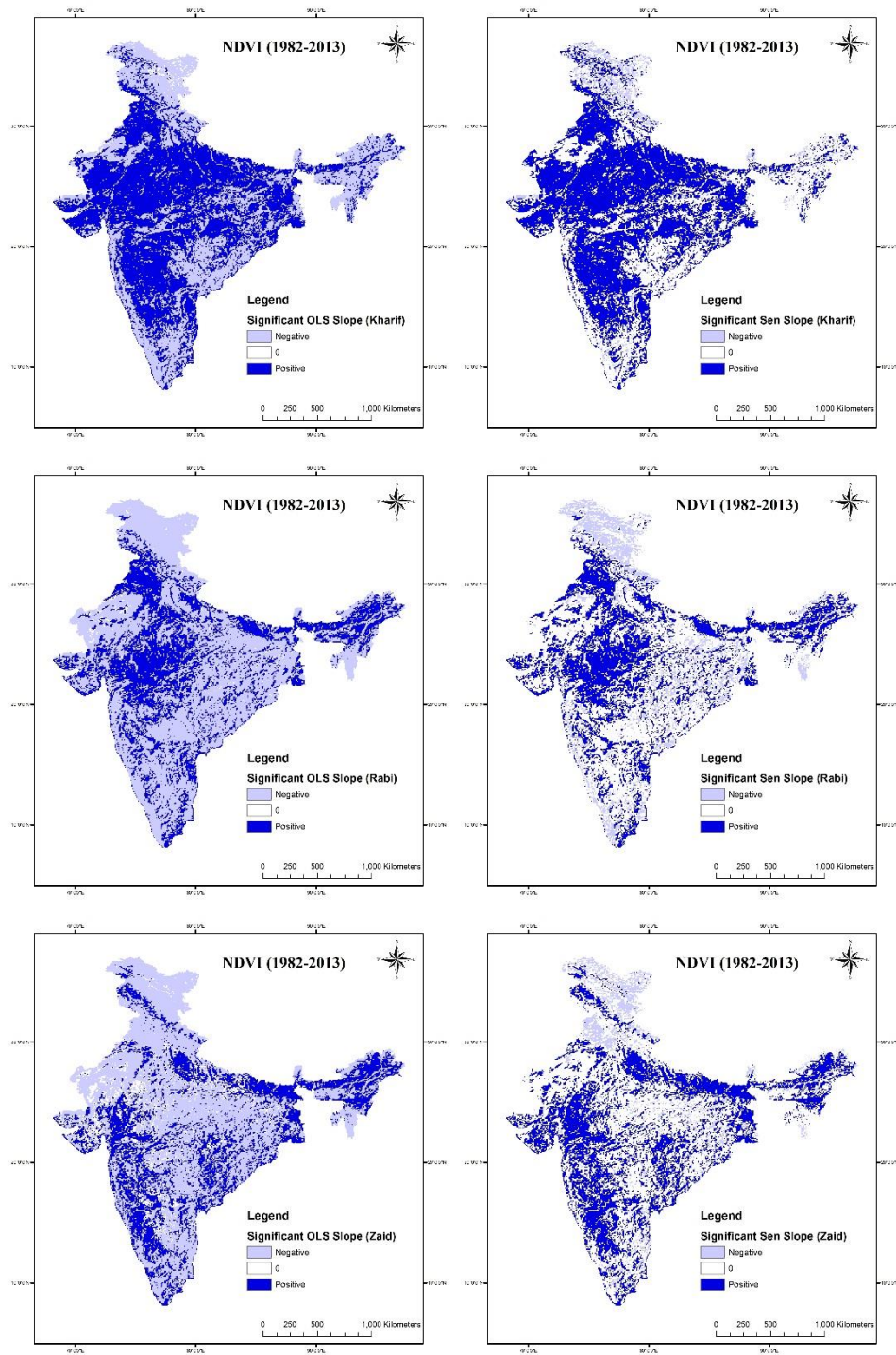


Figure 5.7 Significant slopes derived from OLS & M-K / Sen's estimator on Cropping cycle Temporal Aggregates of NDVI (1982-2013)

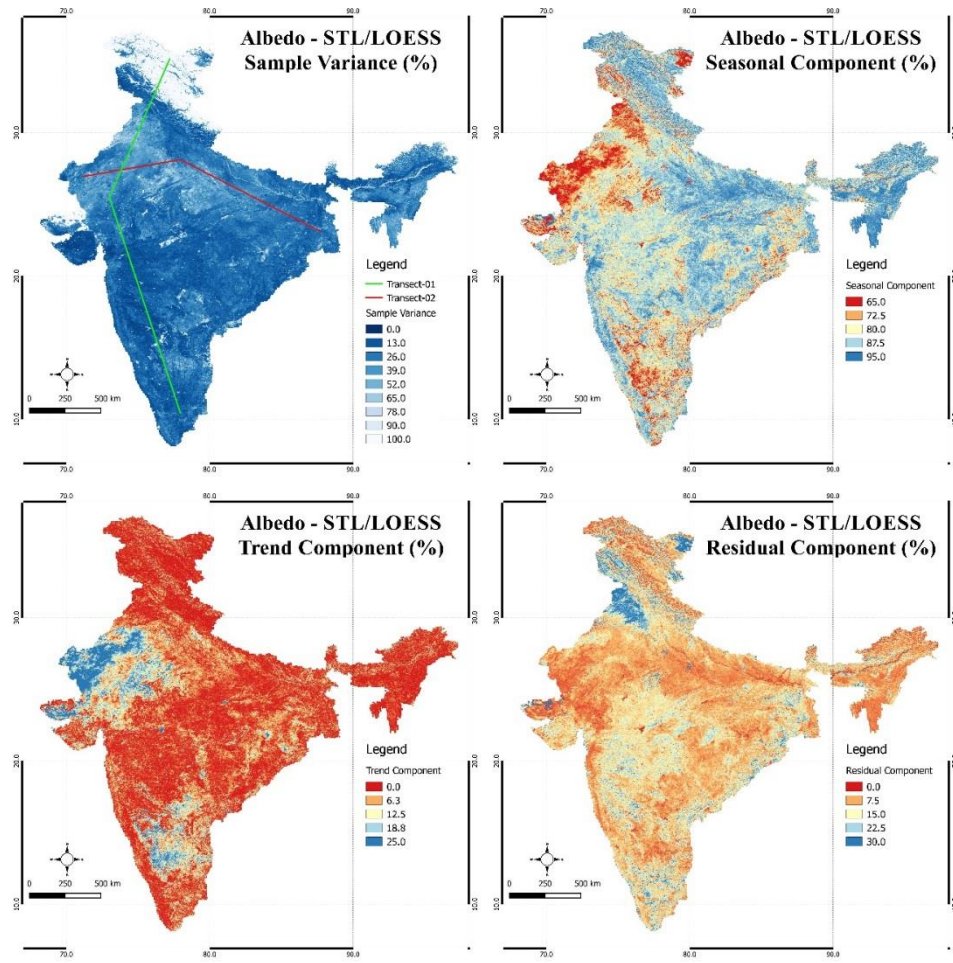


Figure 5.8 Results of STL/LOESS on Albedo (January 2002-December 2013). Variance of albedo time-series (a); relative proportions of albedo variance contributed by seasonal, trend and residual (irregular) components are shown in (b–d), respectively. Relative proportion of variance in albedo explained by all time-series components along two profiles lines (N-S and W-E) on Fig. 5.8a are shown in Fig. 5.9.

Figure 5.8c shows the contribution of the STL trend components to the overall variance. Trend is the major component of variance in few locations. Higher percentages of trend components are concentrated in the Western Plain, Kutch Peninsula and Deccan Plateau. Although trends are detected in a large area of Western India (Rajasthan), total variances of albedo are rather small (10–30%) here. In contrast, areas with elevated trend variances are also observed in the Kutch peninsula where mean albedo fluctuations are high (>80%). The median percentage of variance due to residual (irregular) component in albedo values is about 10% (Fig. 5.4d). Higher percentages of irregular components in albedo values are observed in Punjab, Haryana and parts of Jammu and Kashmir.

Spatial variations in albedo values variance and its decomposition into various time-series components are represented in two regional transects (Figs. 5.8a, and 5.9). The first transect (N-S) shows variations in seasonality, trend and irregular components from Western Himalaya down to the Deccan plateau, through Northern plains & Central highlands; and the second transect (W-E) shows the variations from the Western plain & Kutch peninsula to Eastern plateau & Eastern Ghats.

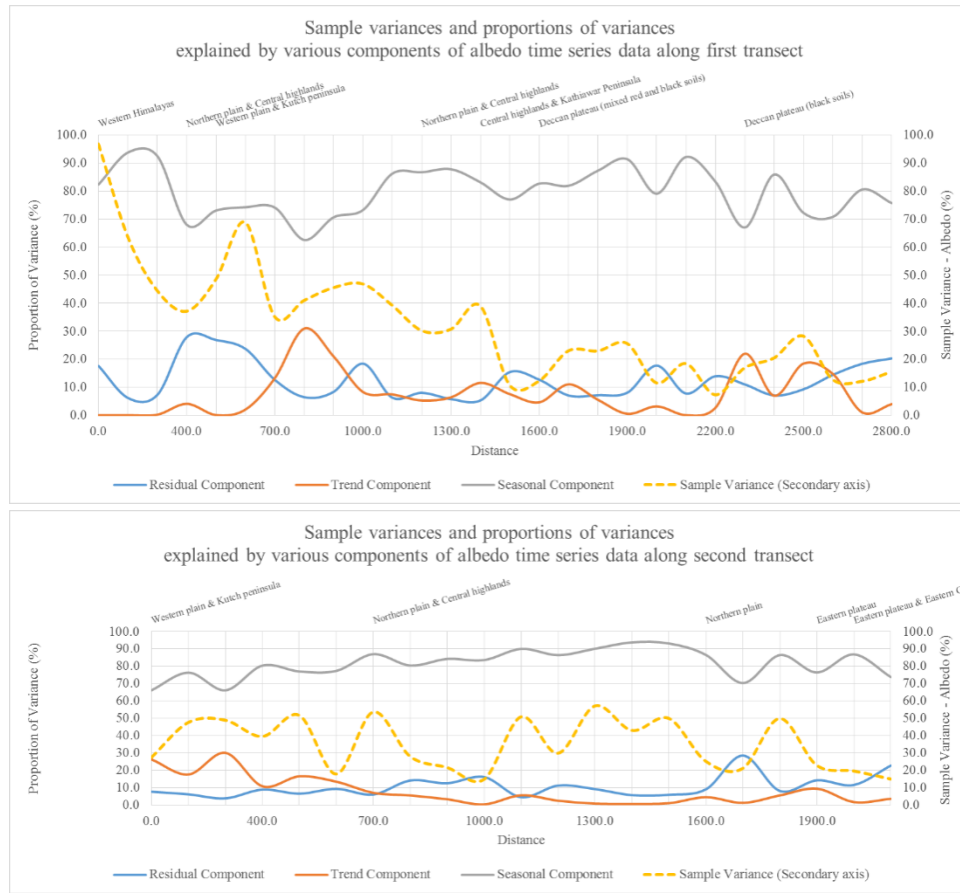


Figure 5.9 Sample variances and proportions of variances explained by various components of the albedo time-series data along two transects as shown in Fig. 5.8a (January 2002-December 2013)

Figure 5.9 shows that along both transects, seasonality is the major component of albedo variance; however, trends dominate in the Western plain & Kutch Peninsula and around Deccan plateau with black soils. The irregular component explains approximately 10% of the variance in albedo throughout the country. Irregular components are higher in the high productive irrigated agricultural areas of Punjab and Haryana.

The regional structure in STL-derived trends has been defined and visualized in Figure 5.10 using the formulation similar to Section 5.2.1; however for albedo dataset, ‘n’ is the number of years of record in the analysis period 2002-2013, has been kept at 12. The albedo trends show a mixed distribution of negative and positive values in the country. The lower and middle Himalayas, parts of Gangetic plains in Bihar, Bengal & Assam and the Central highlands show positive trends. Higher reaches of Northern Himalayas, Rajasthan, Kutch peninsula, Deccan and Eastern plateau show negative trends.

The spatial pattern emerging from the albedo trend derived from STL/LOESS is very similar to that derived from OLS & Sen’s slope / M-K test as shown in Figure 5.11, however, they are more realistic than trends estimated by linear regression.

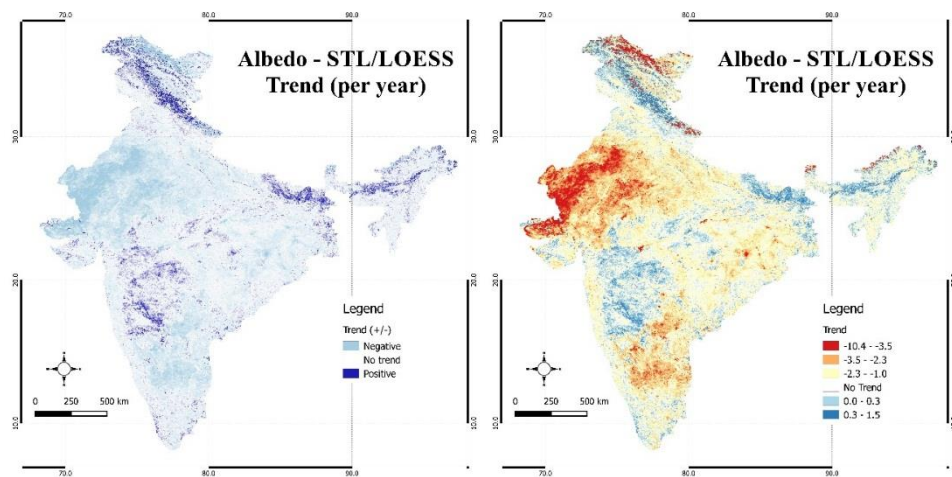


Figure 5.10 Trends in Albedo (derived from STL/LOESS results) (January 2002–December 2013)

OLS & M-K on Albedo Temporal Aggregates

Temporal aggregates was done for Annual and cropping cycle (Section 4.1.3) and OLS & Sen's slope estimator / M-K test techniques were applied (Section 4.1.4). Figure 5.11 shows the areas under positive and negative trend as derived from application of OLS & M-K algorithm on the temporal aggregates.

Figure 5.11a maps the trend of slopes derived using OLS method for albedo in annual temporal aggregates for 12 years data (January 2002 till December 2013). Large tracts of Rajasthan, Eastern plateau, Eastern Ghats and Deccan plateau show negative trends. Very few areas (small patches in North Karnataka and Assam plains) show positive trends.

The same is replicated in the results from Sen's slope estimates masked by significance at 95% confidence interval, as shown in Figure 5.11b.

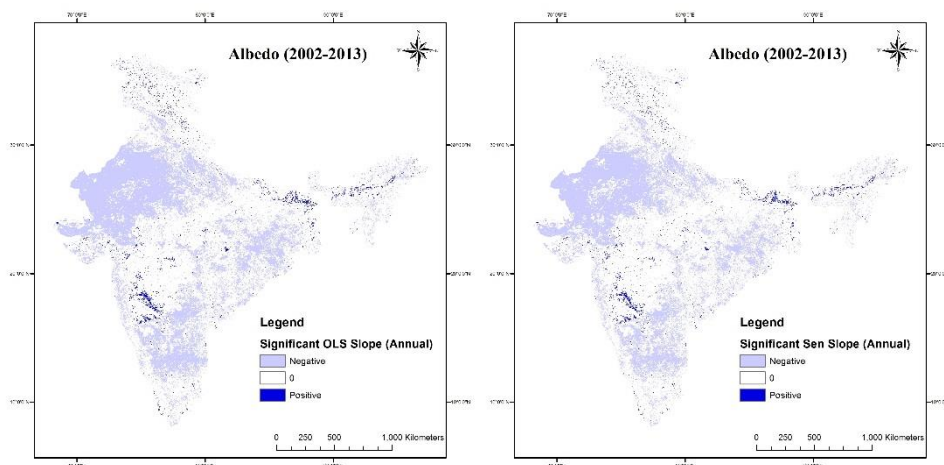


Figure 5.11 Significant slopes derived from OLS & M-K / Sen's estimator on Annual Temporal Aggregates of Albedo (2002–2013)

Figures 5.12a-e shows trends in albedo aggregated over different cropping seasons. As seen for NDVI data trends in Figures 5.7a-e, the significant slopes derived from OLS and Sen's slope / M-K test are nearly identical for albedo data trends.

Kharif season data shows negative trends in Rajasthan and in scattered locations in South and Central India. Positive trends are observed in very few places in North Bengal, Madhya Pradesh and Maharashtra.

Rabi season trends indicate negative trends in Rajasthan, Odisha, Mizoram, entire Western Ghats and South India. Positive trends are observed in lower and middle Himalayas, Gangetic plains and Assam plains.

Negative trends in Rajasthan, Western Uttar Pradesh, Western Madhya Pradesh, Karnataka, Southern Andhra Pradesh and Odisha is observed in Zaid season trends. Positive trends are scattered over Northern Himalayas, Bihar & Assam plains and parts of Maharashtra.

Albedo and NDVI Correlation Analysis

The second objective of this study focusses on the long term trends in different datasets. An attempt has been made prepare and use the datasets for the highest resolution and longest time duration available at the time of data processing. The time periods or length of dataset (in number of years) used in NDVI and Albedo trend analysis are different. However, the study by de Jong and de Bruin, 2012 suggests that the “slope coefficient is inversely related to the extent of the time series”.

To interpret the long term trends in albedo and NDVI, an interpretation key has been shown in Table 5.1. For different land use and land cover classes, NDVI and albedo are either positively or negatively correlated. To compare the trends or slopes (derived from STL/LOESS) for NDVI (Figure 5.5) and albedo (Figure 5.10), the keys given in Table 5.1 can be used.

























		NDVI	Albedo	Correlation
Vegetation				Negative
				Negative
Urban areas / Bare land				Negative
				Negative
Water body / Soil moisture / Irrigated area				Positive
				Positive
Snow				Negative
				Negative

Table 5.1 Correlation between albedo and NDVI - interpretation key

Areas receiving snowfall, such as Himalayas and higher reaches of North Eastern Hills, show positive trends in albedo and negative trends in NDVI (negative correlation). The cropping belts show increasing trends in NDVI and decreasing trends in albedo (negative correlation). Some areas under irrigation / higher soil moisture, show positive correlation, i.e., negative NDVI and negative albedo trends.

Quantitative correlation was performed by selecting MODIS albedo and NDVI data for the time period (June 2002 - December 2013) and statistically significant (at 95% confidence interval) Pearson’s correlation coefficient (r) and coefficient of determination (r^2) was calculated, pixel-wise, as shown in Figure 5.13a-b. The two datasets were resampled spatially and temporally to 8 km and 16 day resolutions.

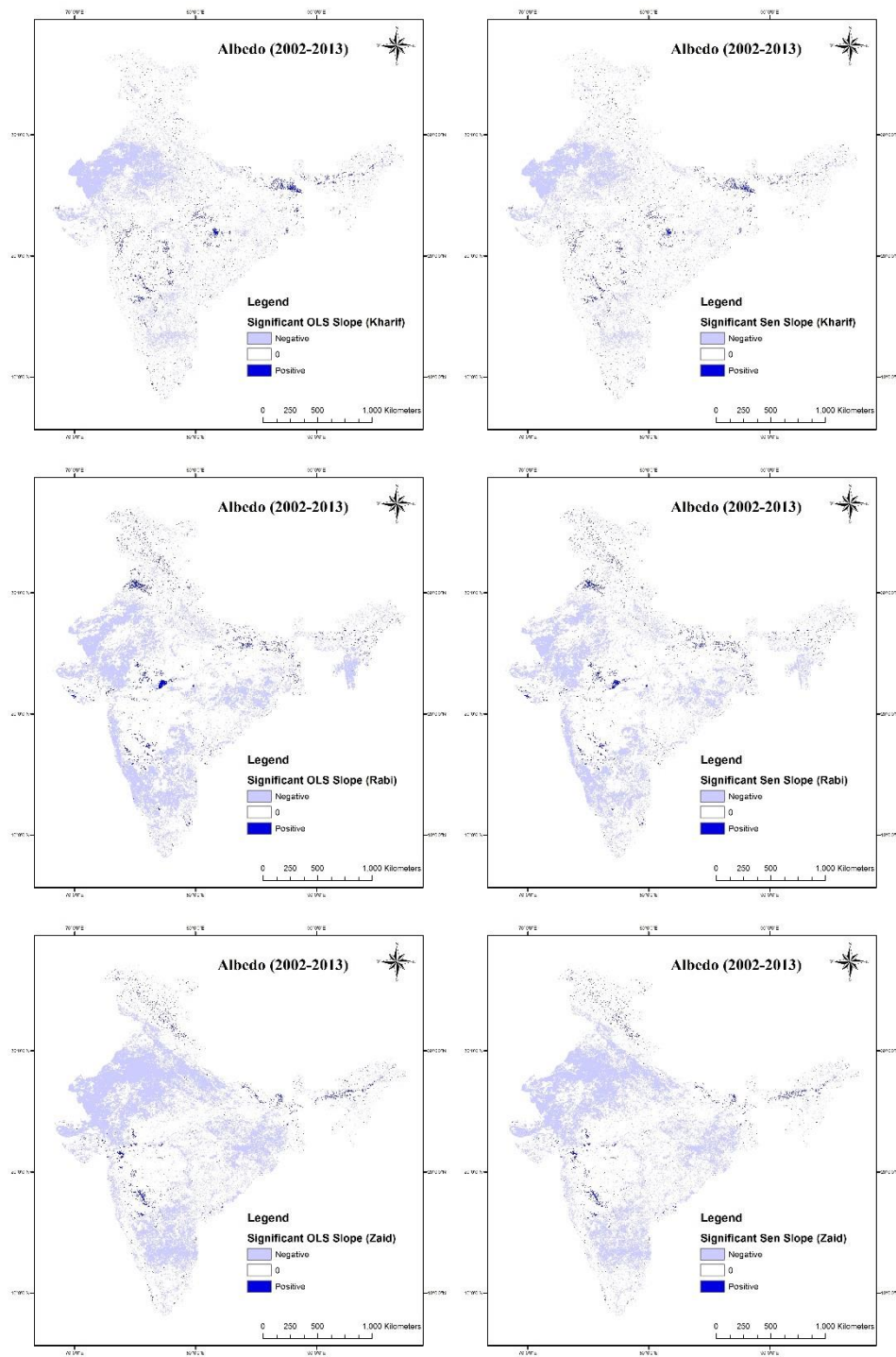


Figure 5.12 Significant slopes derived from OLS & M-K / Sen's estimator on Cropping cycle Temporal Aggregates of Albedo (2002-2013)

The results are similar to as discussed above, negative ‘r’ values were observed in Himalayan and cropping belts and positive ‘r’ values in irrigated areas and areas with high soil moisture. R^2 values or goodness-of-fit statistic, mostly, shows higher values in areas with negative correlation.

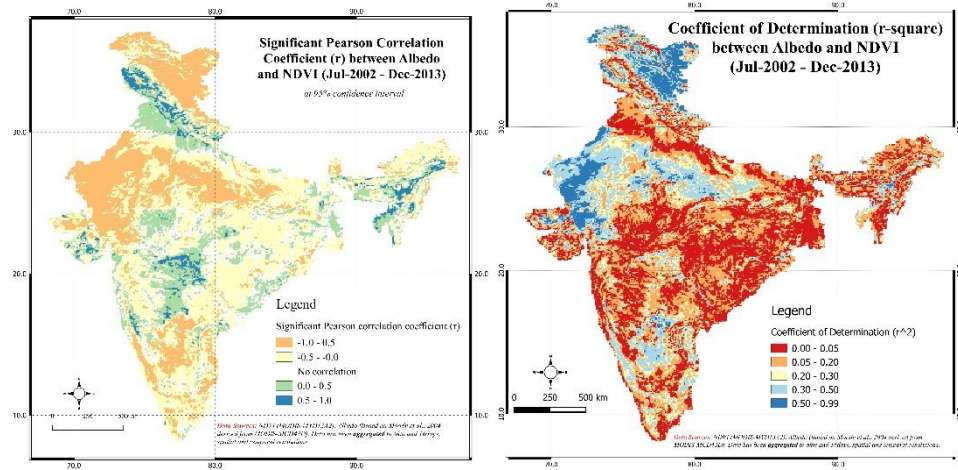


Figure 5.13 Correlation statistics between albedo and NDVI datasets (2002-2013)

The results of this analysis can be further improved by studying spatial variability of soil moisture and ground water extraction patterns.

5.2.3 Temperature trends

STL/LOESS on Temperature Time-Series Anomaly

As discussed in Section 4.1.4, STL/LOESS was applied on the 63 years (1 day temporal resolution) maximum and minimum temperature anomaly data from January 1951 till December 2013. This anomaly dataset was prepared by subtracting the temperature ‘normals’ derived by averaging all the datasets and then subtracting from the original series.

Figure 5.14a & 5.15a maps the variances in the original series. Higher variances in temperatures ($>40^\circ$) are observed in J&K, Punjab and Haryana. Lower variances are observed while going South of India, towards the equator. Least variances ($<10^\circ$) were observed in Kerala and Tamil Nadu.

A ‘banding’ effect is seen in the results derived from STL/LOESS components (Figure 5.14b-d & 5.15b-d). Northern India is found to have higher contribution from seasonal components as compared to Southern India. Similarly lower contribution from trend components are found from Northern India as compared to Southern India.

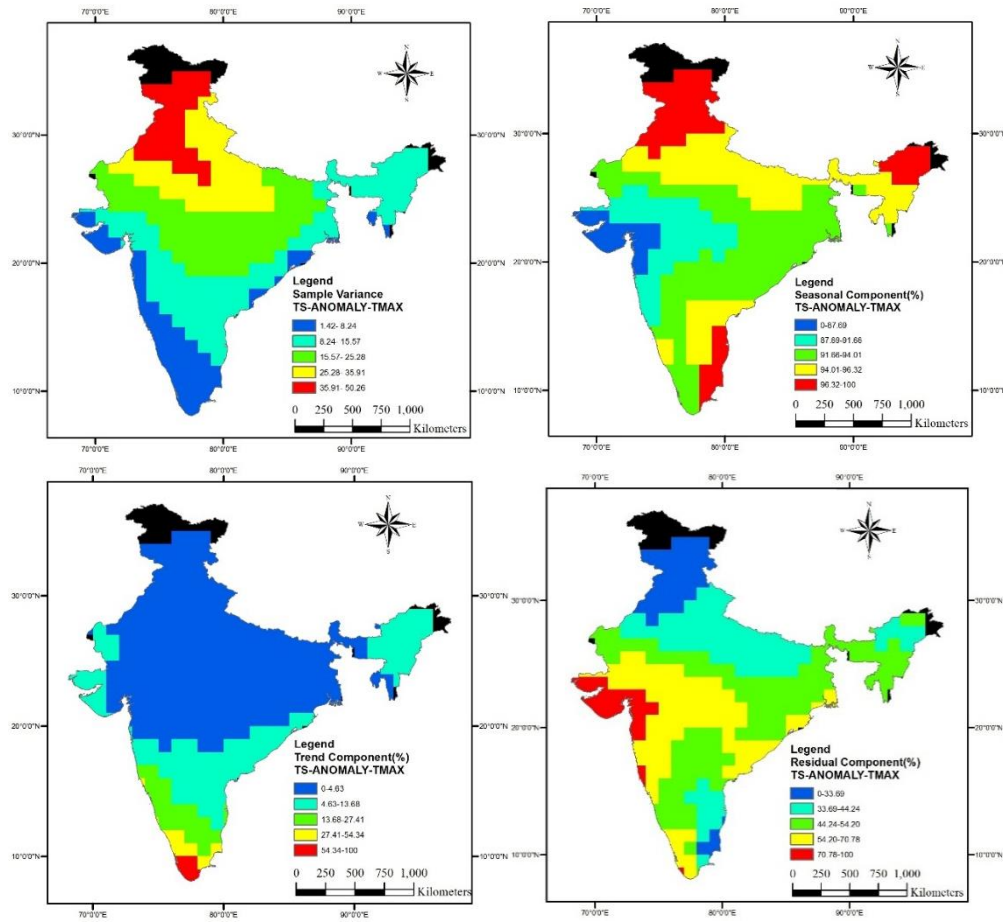


Figure 5.14 Results of STL/LOESS on maximum temperature anomaly data. Variance of temperature time-series (a); relative proportions of temperature variance contributed by seasonal, trend and irregular components are shown in (b–d), respectively (1951-2013)

Figure 5.16 shows the trends derived from STL/LOESS for maximum and minimum temperature anomaly data. Maximum temperature reported increasing trend in Southern peninsula and decreasing trend in Gangetic plain and Central Highlands. Minimum temperature reported increasing trend in Rajasthan and decreasing trend in Bihar, West Bengal and parts of Odisha.

STL/LOESS on Temperature ETCCDI Indices

As discussed in Section 4.2.1 (Climatic data), ETCCDI Indices 6 and 9 (monthly maximum value of daily maximum temperature and monthly minimum value of daily minimum temperature) were prepared and STL/LOESS was applied on the monthly maximum and minimum temperature aggregates.

Figure 5.17a & 5.18a maps the variances in the original series of ETCCDI Indices 6 and 9. Higher variances were observed in Northern India as compared to Southern India.

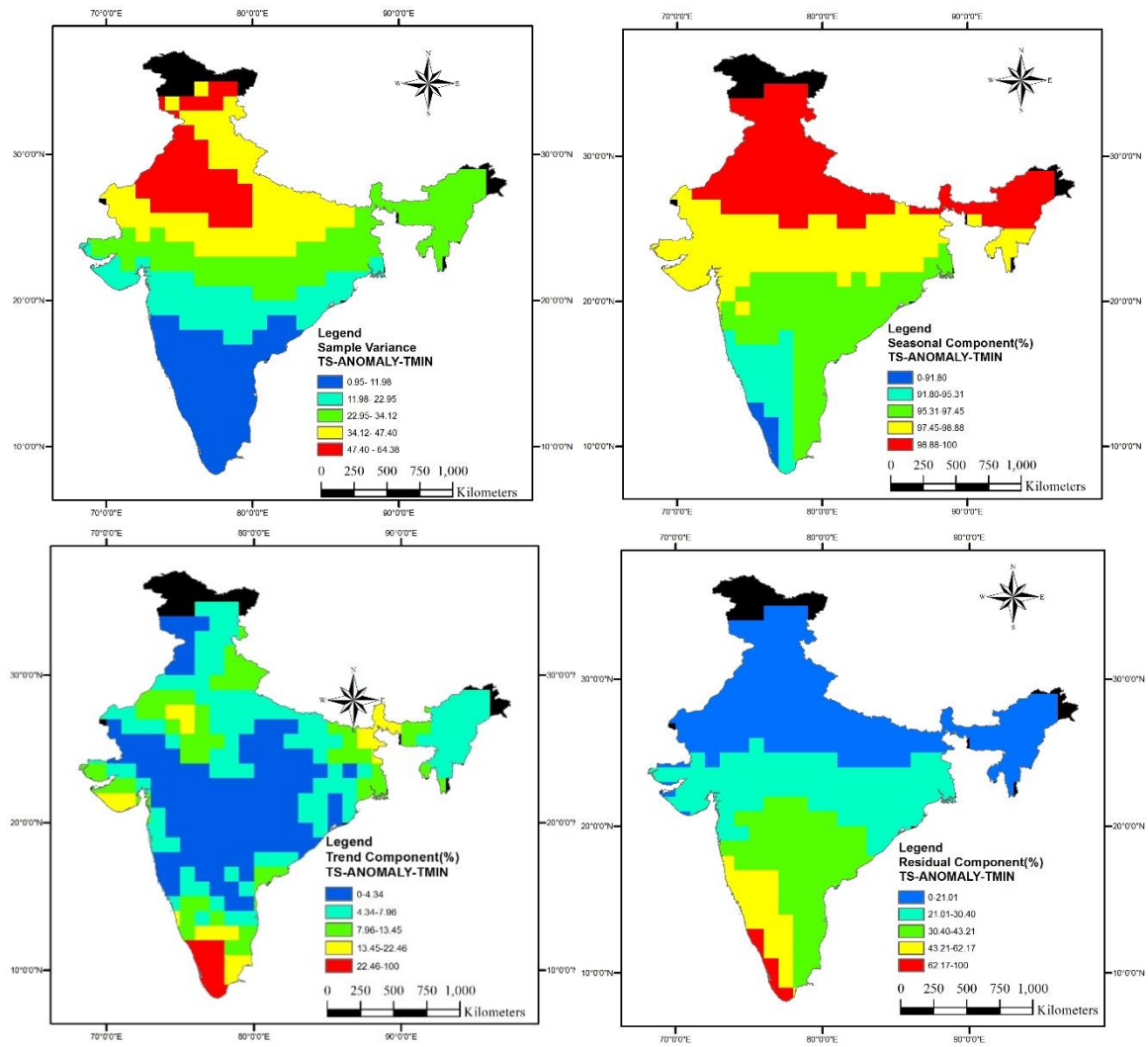


Figure 5.15 Results of STL/LOESS on minimum temperature anomaly data. Variance of temperature time-series (a); relative proportions of temperature variance contributed by seasonal, trend and irregular components are shown in (b–d), respectively (1951-2013)

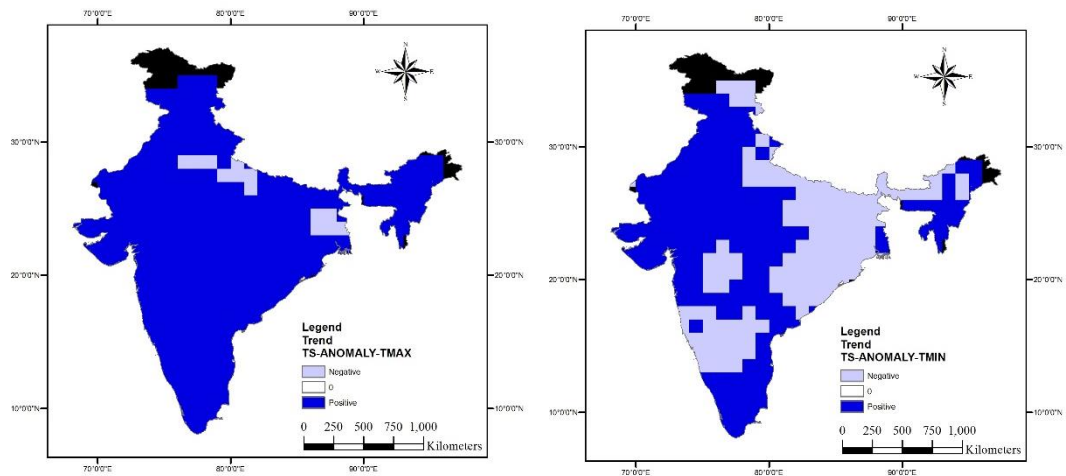


Figure 5.16 Trends in max. & min. temperature anomaly (derived from STL/LOESS results) (1951-2013)

Monthly maximum temperatures show complete effect of seasonality and very low impact of trend as seen in Figures 5.16b-c. Monthly minimum temperatures show the banding effect as discussed in previous paragraph.

Figures 5.19a-b show the trends obtained from STL/LOESS on Index 6 and 9. A general increasing trend is observed in the entire country. A maximum increase of about 0.01°C per year is reported in monthly maximum temperatures for North Eastern India. The western half of the country shows positive trends in monthly minimum temperatures and eastern half shows negative trends. These graphs are also similar to Figure 5.16 showing trends in maximum and minimum temperature anomaly (derived from STL/LOESS results) for the same time period i.e., 1951 to 2013.

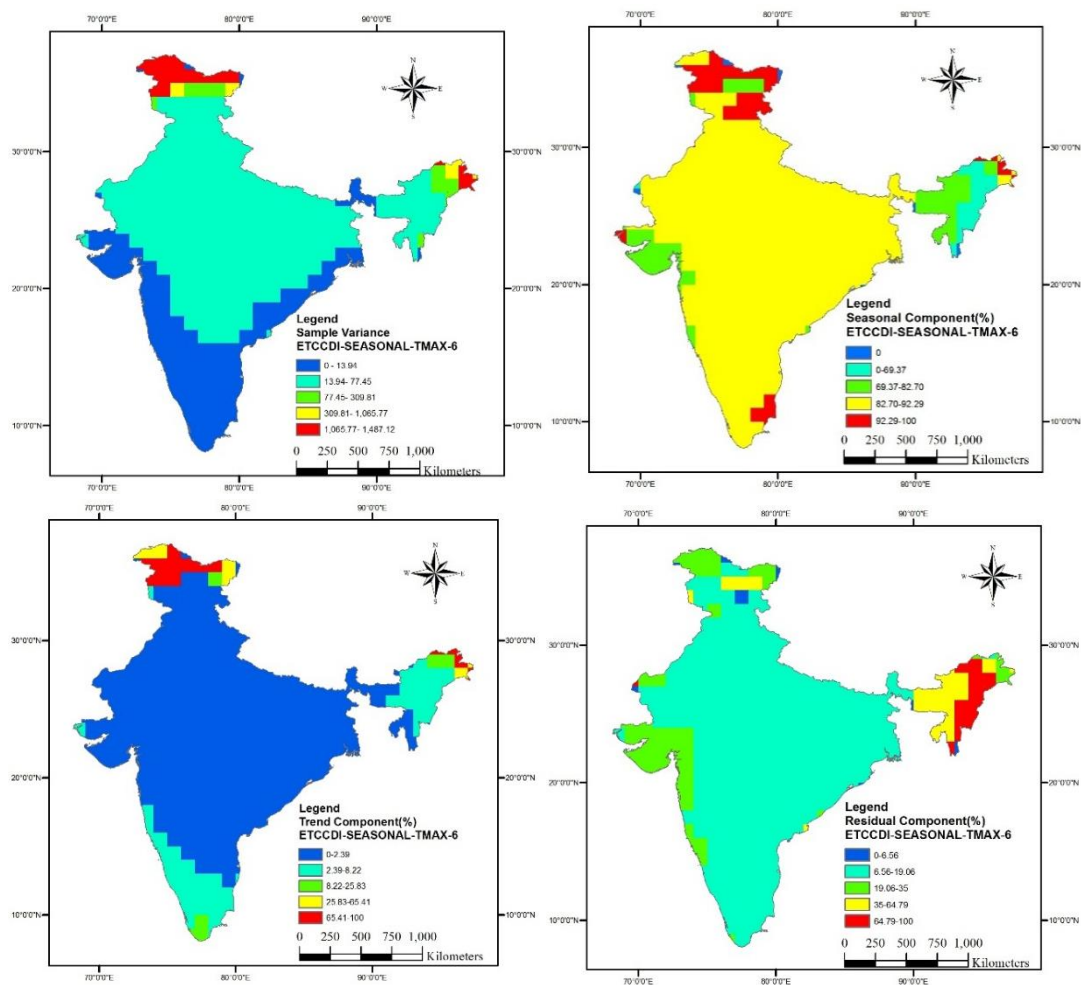


Figure 5.17 Results of STL/LOESS on ETCCDI Index '6' monthly maximum T_{max} (1951-2013)

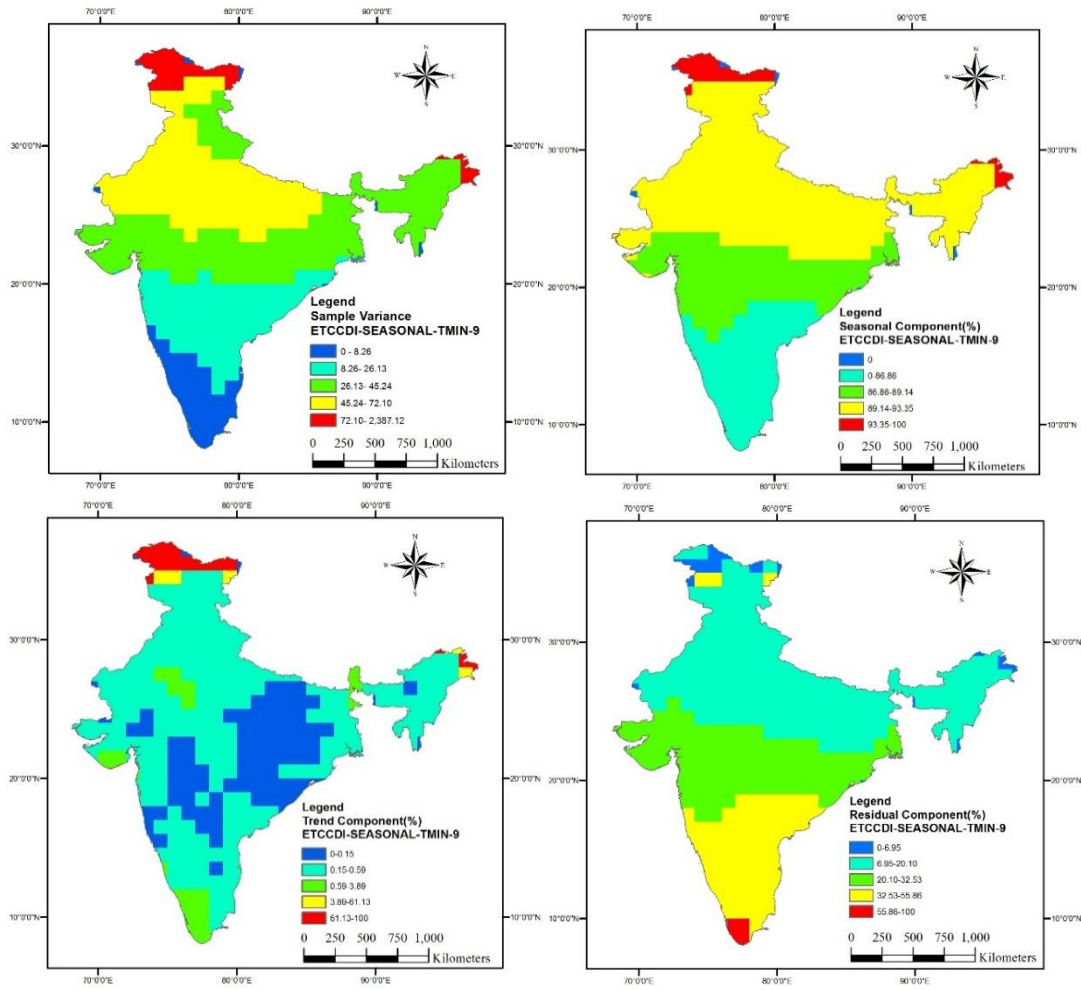


Figure 5.18 Results of STL/LOESS on ETCCDI Index '9' monthly minimum T_{min} (1951-2013)

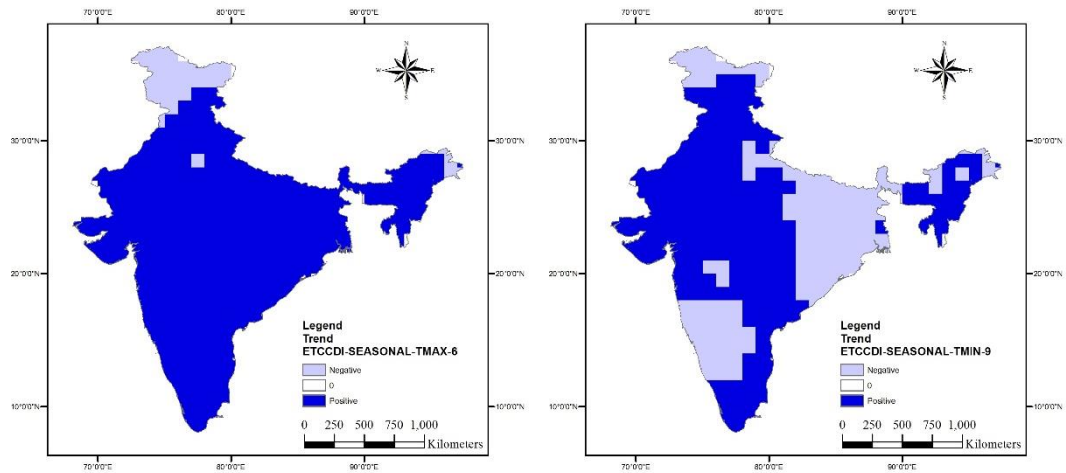


Figure 5.19 Trends in ETCCDI Index 6 and 9 on T_{max} and T_{min} (derived from STL/LOESS) (1951-2013)

OLS & M-K on Temperature Annual Temporal Aggregates

Figures 5.20a&b and 5.21a&b shows the results obtained on application of OLS and Sen's slope estimator algorithms on annually aggregated maximum and minimum temperature data. Both the techniques give identical results at such coarse resolution ($1^\circ \times 1^\circ$).

Most of the south, west and eastern India show increasing trends in maximum temperatures. Parts of Punjab, Haryana, Rajasthan, Gujarat, Tamil Nadu, Andhra Pradesh and North Eastern states show positive trends in the minimum temperatures data.

Although some negative trends were observed in the Sen's slope estimates, they were clipped out after masking for 95% confidence interval ($\alpha < 0.05$, derived from M-K test).

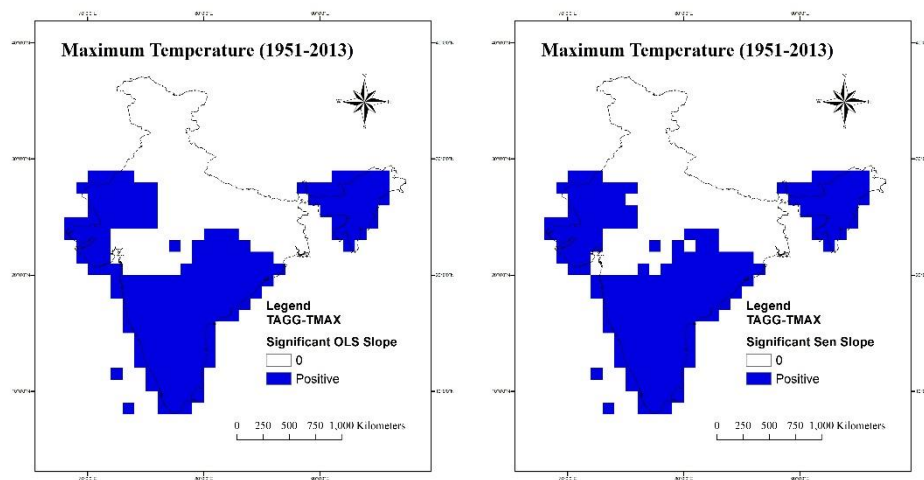


Figure 5.20 Significant slopes derived from OLS & M-K / Sen's estimator on Annual Temporal Aggregates of Maximum Temperature (1951-2013)

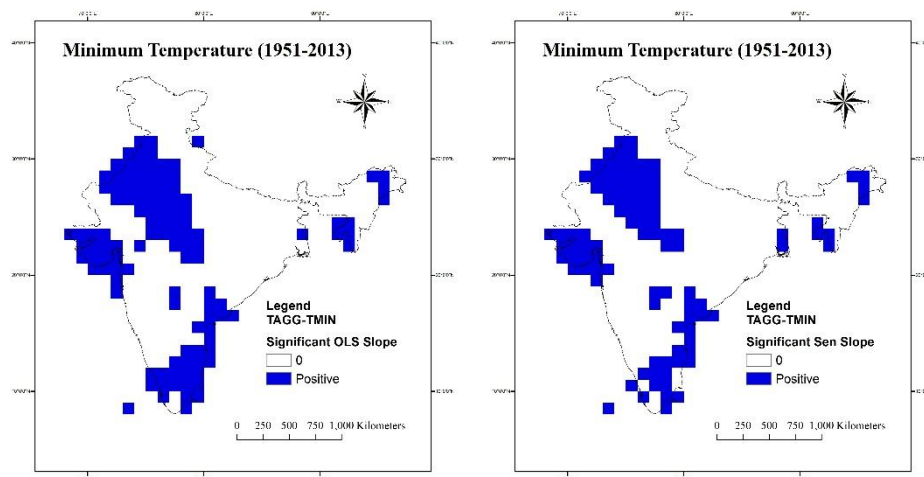


Figure 5.21 Significant slopes derived from OLS & M-K / Sen's estimator on Annual Temporal Aggregates of Minimum Temperature (1951-2013)

OLS & M-K on Maximum Temperature ETCCDI Temporal Aggregates

Figure 5.22a-d shows significant slopes derived from OLS & M-K / Sen's estimator on non-seasonal ETCCDI Indices 2 & 13 on maximum temperature datasets. Index 2 represents the number of summer days (SU) and Index 13 represents the percentage of days when maximum temperatures are greater than the 90th percentile of values centred at a 5-day window (TX90p).

The results obtained from both parametric and non-parametric test are identical. Number of summer days i.e., days with daily maximum temperature is greater than 25°C shows an increasing trend in Northern Himalayas, North East states and parts of Western Ghats.

The TX90p results are similar to those of trends from annual temporally aggregated maximum temperature data (Figure 5.20). Except J&K and the Gangetic plains, the rest of India shows increasing maximum temperatures for this ETCCDI Index.

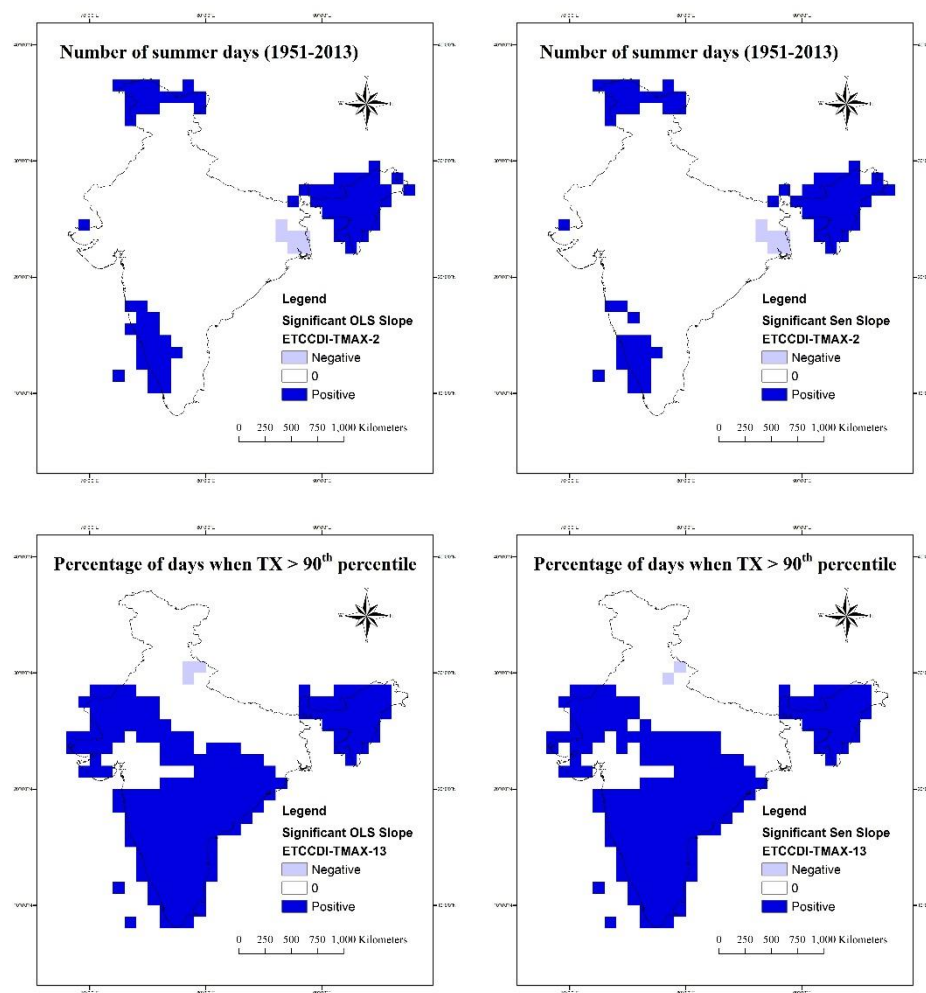


Figure 5.22 Significant slopes derived from OLS & M-K / Sen's estimator on ETCCDI (Non seasonal) Indices 2 & 13 (on T_{max} data) (1951-2013)

OLS & M-K on Minimum Temperature ETCCDI Temporal Aggregates

The non-seasonal ETCCDI indices 4 and 10 i.e., number of tropical nights and percentage of days when minimum temperatures is less than 10th percentile of a 5-day window period (TN10p) - were subjected to the OLS and Sen's slope estimator methods. The slope values at 95% confidence interval are shown in Figures 5.23a-d, from both the methods. As seen in the above 2 examples, the OLS and Sens' slope estimator values with M-K test are identical in nature.

Parts of J&K, Rajasthan, Kutch peninsula, Kerala, Tamil Nadu, West Bengal and Arunachal Pradesh show positive trends in number of tropical nights (i.e., days with minimum temperatures are greater than 20°C). Nearly the same areas, with larger footprint, show decreasing trends in the TN10p data results. Only exception was the parts of J&K where positive trends were observed in both the results.

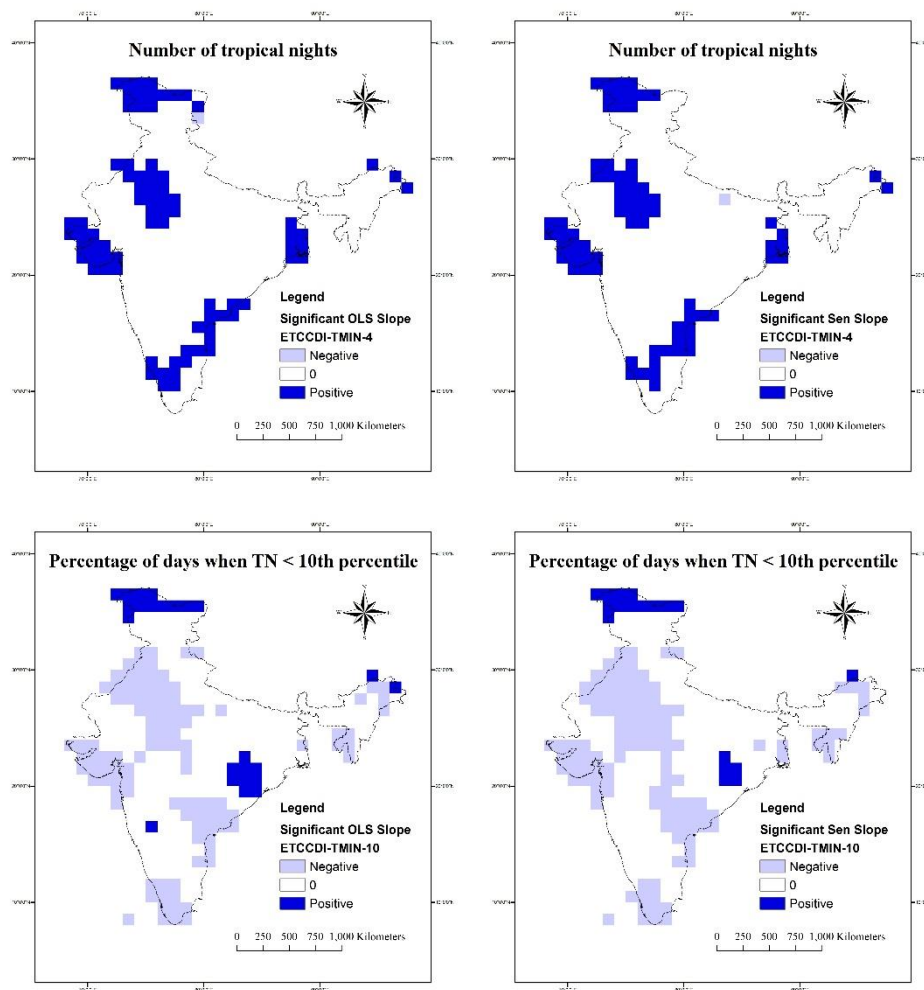


Figure 5.23 Significant slopes derived from OLS & M-K / Sen's estimator on ETCCDI (Non seasonal) Indices 4 & 10 (on T_{min} data) (1951-2013)

Trend analysis on difference of monthly maximum and minimum temperature data (ΔT)

Maximum and minimum temperature values and its associated ETCCDI indices were studied using the statistical methods described in Section 4.1.4. However an important ETCCDI Index 16 (DTR) which suggests to study trends in daily temperature range i.e, trends in monthly mean difference between maximum and minimum temperatures was also carried out.

For ease of calculation, instead of taking an average of daily differences, the difference of monthly average of maximum and minimum temperatures was calculated for all 12 months for 63 year time period from 1951 till 2013. The spatial trend in these ($12 \times 63 =$) 756 months were calculated using both the OLS and Sen's slope estimator techniques. After masking with 95% confidence interval ($\alpha < 0.05$) calculated again for each pixel, identical results were obtained from both techniques. The results from the Sen's slope estimator is shown in Figure 5.24.

The result clearly shows parts of Haryana, Punjab and adjoining areas having a marked decreasing trend in the monthly difference of maximum and minimum temperatures. This is in line with results reported by Lee *et al.*, 2009. Also increasing trends are reported in South India, Odisha and parts of Rajasthan and North Eastern states. The increase and decrease of monthly range of temperatures is in the order of 0.001°C .

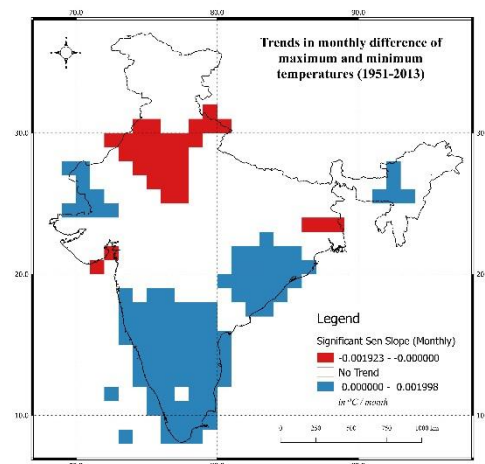


Figure 5.24 Trends in monthly difference of T_{\max} and T_{\min} (1951-2013)

5.2.4 Precipitation trends

STL/LOESS on Precipitation Time-Series Anomaly

As discussed in Section 4.1.4, STL/LOESS was applied on the 111 years (1 day temporal resolution) rainfall anomaly data from January 1901 till December 2011. This anomaly dataset was prepared by subtracting the temperature ‘normals’ derived by averaging all the datasets and then subtracting from the original series.

Figure 5.25a shows the variance in rainfall anomaly values (mm) which is ~50mm throughout the country except the Western Ghats and north east hills where it is reported to go up to 500mm in extreme cases. This variance is mainly (>70%) explained by seasonal component throughout the country (Figure 5.25b).

The trend component (Figure 5.25c) shows a handful of pixels having trend in rainfall data. The residual component (Figure 5.25d) due to irregular rainfall anomaly show irregular patterns in Southern and Western parts of the country. The median of the residual component is ~6% only.

Figure 5.26 shows the trends in rainfall anomaly as derived from STL/LOESS technique. The Western Himalayas and Western Ghats report an increasing trend in rainfall anomaly of about 0.08 mm / year, whereas North Eastern Hills report a decreasing trend in rainfall anomaly of about 0.04 mm / year.

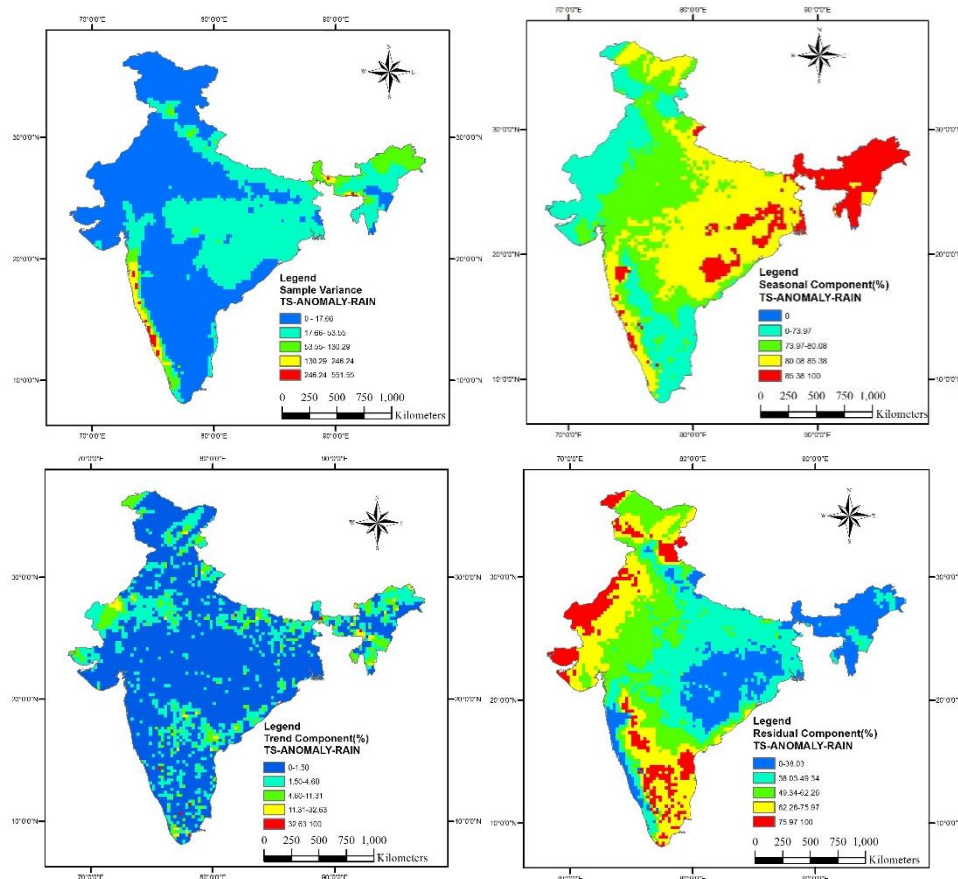


Figure 5.25 Results of STL/LOESS on Rainfall anomaly data (1901-2011)

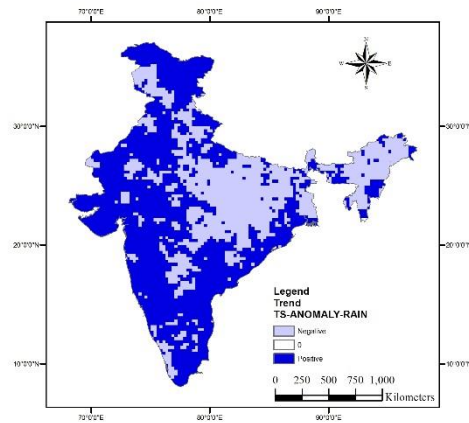


Figure 5.26 Results of trends in Rainfall anomaly (derived from STL/LOESS) (1901-2011)

STL/LOESS on Precipitation ETCCDI Indices

Figure 5.27 and 5.28 show the results of STL/LOESS on seasonal ETCCDI indices 17 & 18 for rainfall. Both sets of results show that rainfall (1 day maximum and 5 day maximum) are governed completely by seasonality and irregularities. The irregularities seem to be concentrated in North, North West, West and South India. The median of the both the residual components were found to be ~8%.

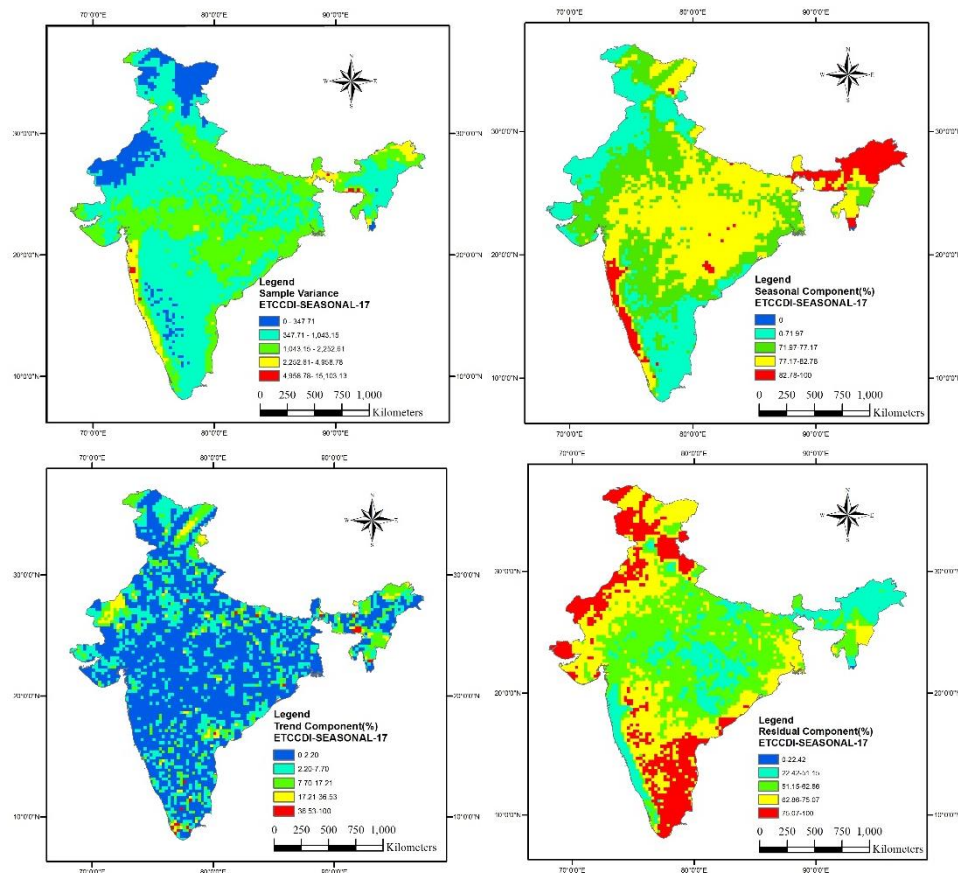


Figure 5.27 Results of STL/LOESS on Seasonal ETCCDI Index 17 for Rainfall (1901-2011)

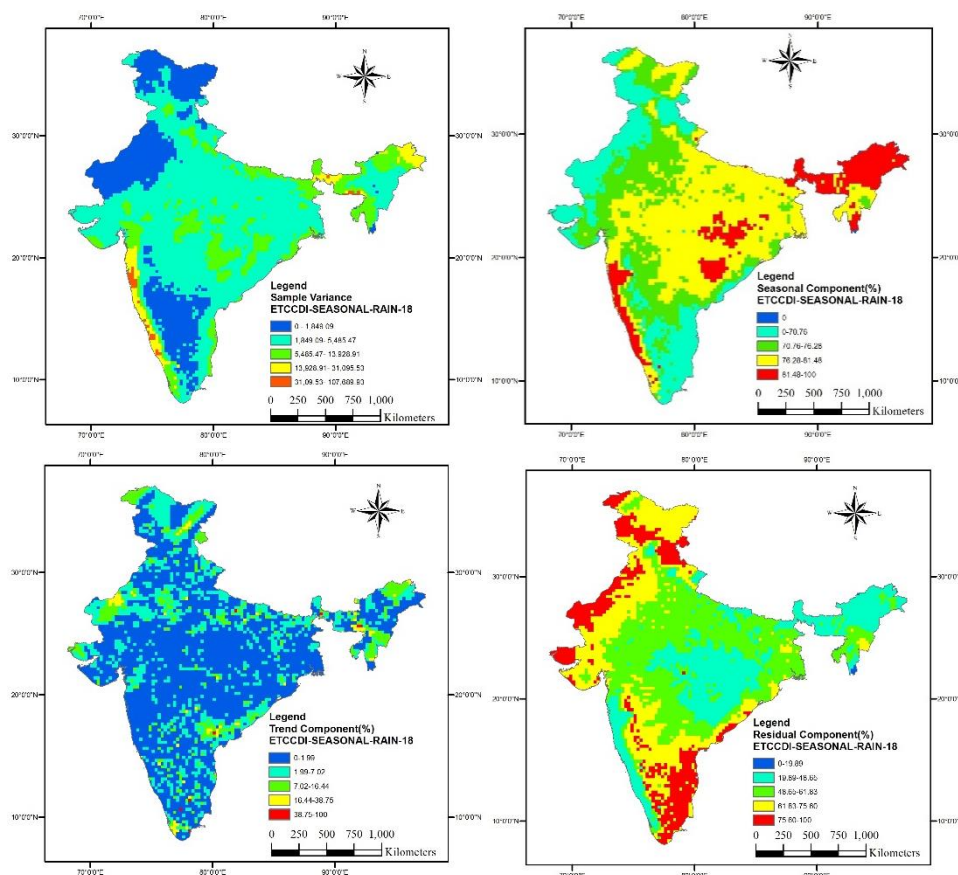


Figure 5.28 Results of STL/LOESS on Seasonal ETCCDI Index 18 for Rainfall

Figure 5.29 show the trends as derived from STL/LOESS results on these 2 seasonal ETCCDI indices. Both are identical in nature with large areas in South, West and North India showing increasing trends. Decreasing trends are observed in parts of Central plateau, Gangetic plains, and North Eastern states.

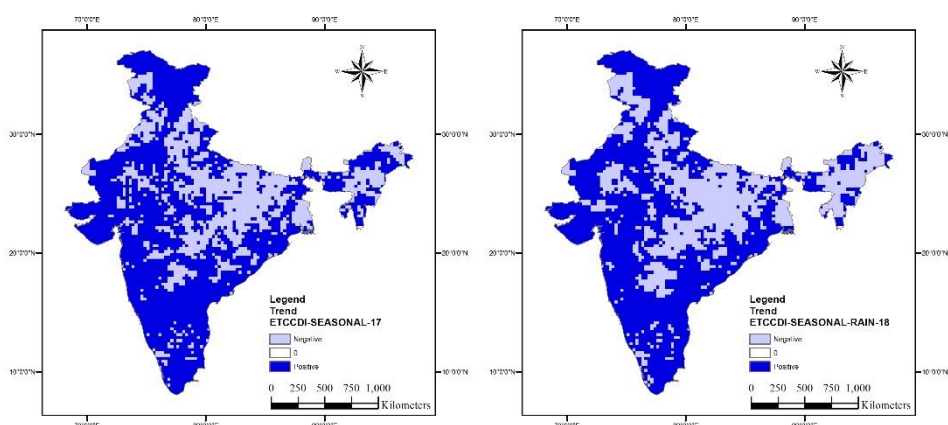


Figure 5.29 Trends in ETCCDI Indices 17 & 18 (as derived from STL/LOESS) (1901-2011)

OLS & M-K on Precipitation JJAS Temporal Aggregates

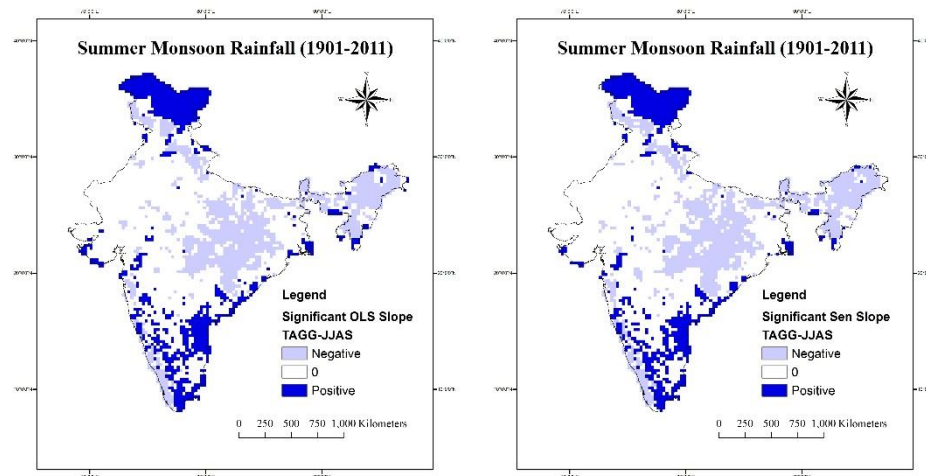


Figure 5.30 Significant slopes derived from OLS & M-K / Sen's estimator on JJAS temporal aggregated Rainfall data (1901-2011)

Figure 5.30a shows significant trends (obtained from OLS) in rainfall data over the Indian summer monsoon rainfall season. Positive trends are obtained in parts of J&K, South India and along the Eastern coast. Negative trends are observed in Bihar, Chhattisgarh, parts of Madhya Pradesh and North Eastern states. Figure 5.30b shows trends obtained from Sen's slope estimator with M-K test and show identical results as that of the OLS method.

OLS & M-K on Precipitation ETCCDI Temporal Aggregates

Figure 5.31 and 5.32 shows results of application of OLS and Sen's slope estimator (at 95% confidence interval) on ETCCDI Index 23 and 24 data. Results of Index 23 (Figure 5.31) which shows trends in maximum length of dry spell shows significant negative trends in Northern Himalaya and parts of Rajasthan. Positive trends are seen in Kutch & Kathiawar regions, parts of South India and Central highlands & plateaus.

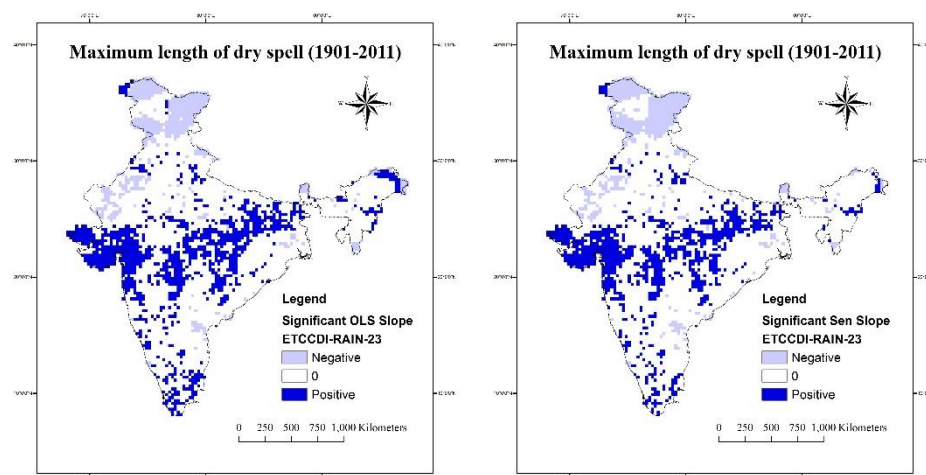


Figure 5.31 Significant slopes derived from OLS & M-K / Sen's estimator on ETCCDI Index 23 (1901-2011)

Results of Index 24 (Figure 5.32) for maximum length of wet spell shows identical results for OLS and Sen's slope estimator; just as seen for Index 23. Positive trends are seen in parts of J&K and scattered places in Eastern & Western Coasts and North East states. However, negative trends are predominant over the rest of India showing shortening of maximum length of wet spells across India.

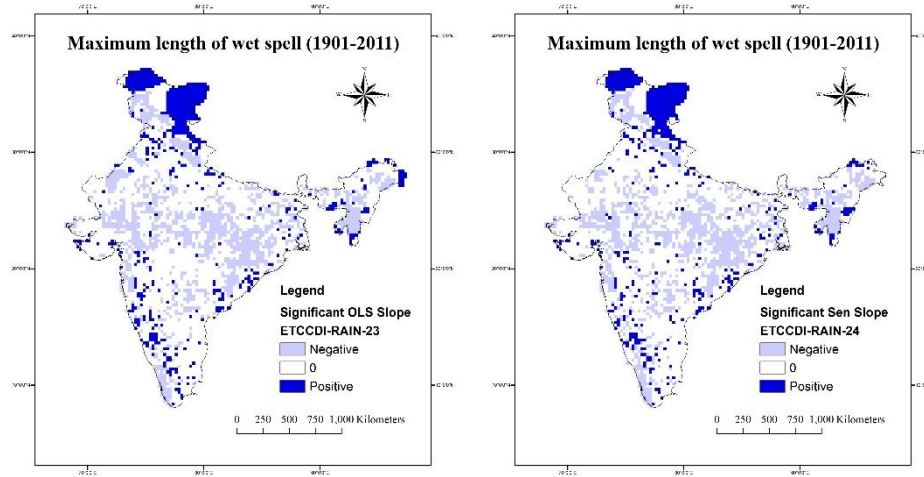


Figure 5.32 Significant slopes derived from OLS & M-K / Sen's estimator on ETCCDI Index 24 (1901-2011)

5.3 Correlation among datasets and gravity anomaly

Section 3.2.3 discussed about the Gravity anomaly datasets followed by Section 4.3 which discusses about the process followed to find correlation among different datasets.

5.3.1 Results obtained from Eureqa

The technique as discussed in Section 4.3.2 (Eureqa) was implemented and results obtained. The inputs were 'anomaly values' at monthly temporal aggregation of input variables (NDVI, Albedo, Rainfall, Maximum and Minimum Temperatures) and output variable (Gravity) at $0.5^\circ \times 0.5^\circ$ spatial resolution for the time period 2004-2010.

As discussed, the model was tried to be setup with 2 independent variables NDVI (n) and Rainfall (r) and 1 dependent variable Gravity anomaly (g). However, this attempt failed as the model failed to converge beyond 70% even after running for over 6 hours and could achieve was only 6% stability & 55% maturity.

The second attempt was on 5 independent variables and 1 dependent variable and the model was stopped, it has achieved 100% convergence and was 90%+ stable. The final model fitting equation derived is given in Equation 5.1.

Model:

$$g = b \cdot \text{sma}(r, 4) + c \cdot n + d \cdot \cos(\text{sma}(tn, 15)) + e \cdot tx \cdot \text{sma}(r, 7) + f \cdot n \cdot \text{sma}(tx, 4) \dots \text{Equation 5.1}$$

where g , r , n , tn , tx , were model variables representing gravity anomaly, rainfall anomaly, NDVI anomaly, minimum and maximum temperature anomalies respectively; and coefficients b , c , d , e ,

f were assigned values 1.797511, 0.002937, 2.534784, 0.171359, 0.000493. ' $sma(u, v)$ ' in the equation represents simple moving average of the ' u ' variable by ' v ' time steps and ' cos ' is cosine.

18 different models with varying complexities were attempted by the software before stagnating on the current one (19th model). The number of occurrences of each variable (across all models) is shown in Figure 5.33 and the variable sensitivity in Table 5.2.

Variable	Sensitivity	% Positive	Positive Magnitude	% Negative	Negative Magnitude
tn	1.0763	0.57	1.1535	0.43	0.97252
r	0.60941	1	0.60981	0	0.00268
n	0.47551	0.99	0.48159	0.01	0.051982
tx	0.35182	0.37	0.30178	0.63	0.38146

Table 5.2 Relative impact within the model that the input variables (tn, r, n, tx) had on the target variable (g)

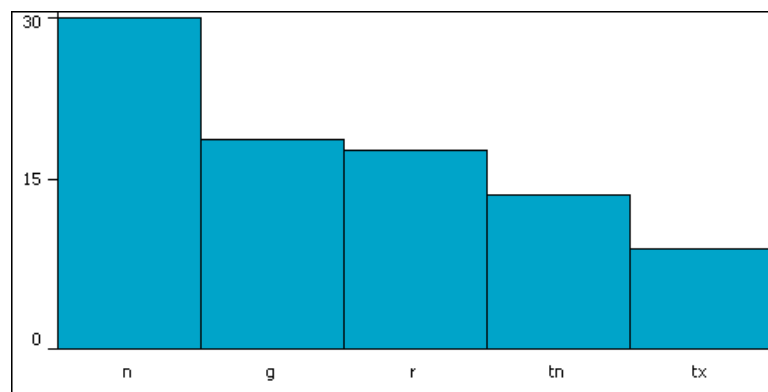


Figure 5.33 Number of Occurrences of each variable (across all models)

Noticeably, the software ignored the input variable ' a ' (for albedo) as it proxies ' n ' (NDVI), in some manner, and that it would increase the model complexity. The attributes of Table 5.2 are explained below.

% Positive: The likelihood that increasing this variable will increase the target variable. If % positive = 70%, then 70% of the time increases in this variable lead to increases in the target variable (but the remaining 30% of the time it either decreases it or has no impact). If % positive = 0%, increases in this variable will not increase the target variable.

Positive Magnitude: When increases in this variable lead to increases in the target variable, this is generally how big the positive impact is.

% Negative: The likelihood that increasing this variable will decrease the target variable. If % negative = 60%, then 60% of the time increases in this variable lead to decreases in the target variable (but the remaining 40% of the time it either increases it or has no impact). If % negative = 0%, increases in this variable will not decrease the target variable.

Negative Magnitude: When increases in this variable lead to decreases in the target variable, this is generally how big the negative impact is.

The model output was quantified by calculating various statistical parameters and spatially distributed comparisons. The observed versus predicted scatter plot is shown in Figure 5.33a. **Overall an r^2 value (Goodness of Fit) of 0.657 and Correlation Coefficient (r) of 0.811 was obtained.**

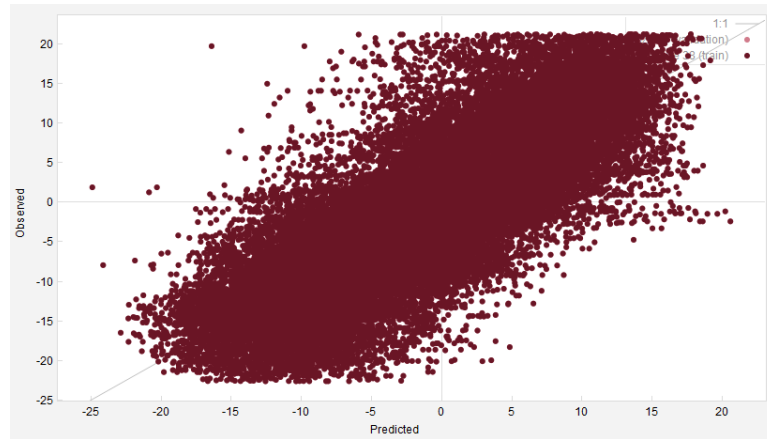
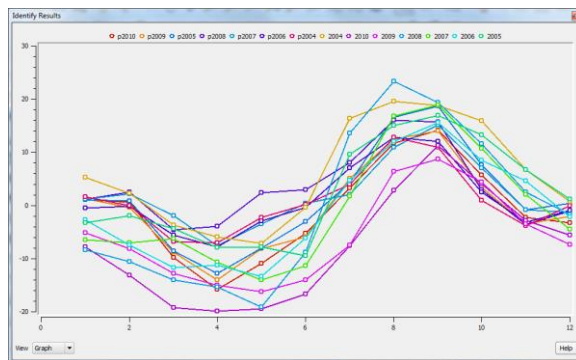
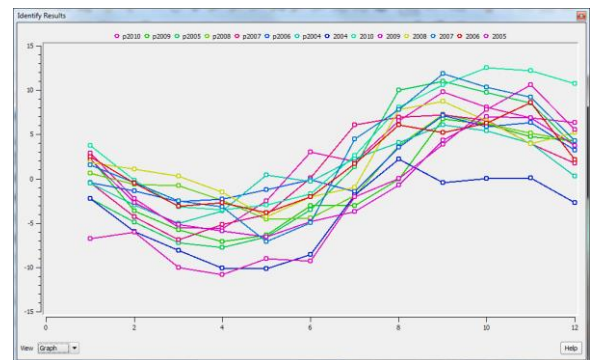


Figure 5.33a Observed vs. Predicted

The derived model (in Equation 5.1) was applied on observed variables and the gravity anomaly was predicted and predicted gravity anomaly obtained. The results at 2 pixels, first at 83.25°N 26.75°E and second at 79.25°N 15.75°E was plotted in Figure 5.34a & b respectively. The model is found to be correctly predicting the trajectory of gravity anomaly each year.



(a) Location = 83.25°N 26.75°E



(a) Location = 79.25°N 15.75°E

Figure 5.34 Monthly plots for time duration 2004-2010 of observed and predicted (prefixed with 'p') gravity anomaly. Y-axis represents gravity anomaly and X-axis represents months.

To understand the spatial correctness of the derived model, 3 spatially distributed parameters were derived namely Pearson's correlation coefficient (r), Coefficient of Determination (r^2) and normalized root mean square error (RMSE) in Figures 5.35 and 5.36. The results are encouraging as they show high correlation coefficient (near 1.0) for most of the Indian land mass. The goodness of fit measure which ranges from 0.25-0.75 for most of the land mass, depicts a good fitting of expected values with observed values.

The normalised RMSE summarizes that most of the area has low discrepancy between predicted and observed values. Higher RMSE is observed in Western Ghats and parts of J&K. A normalized RMSE of less than 0.30 is observed in the rest of India.

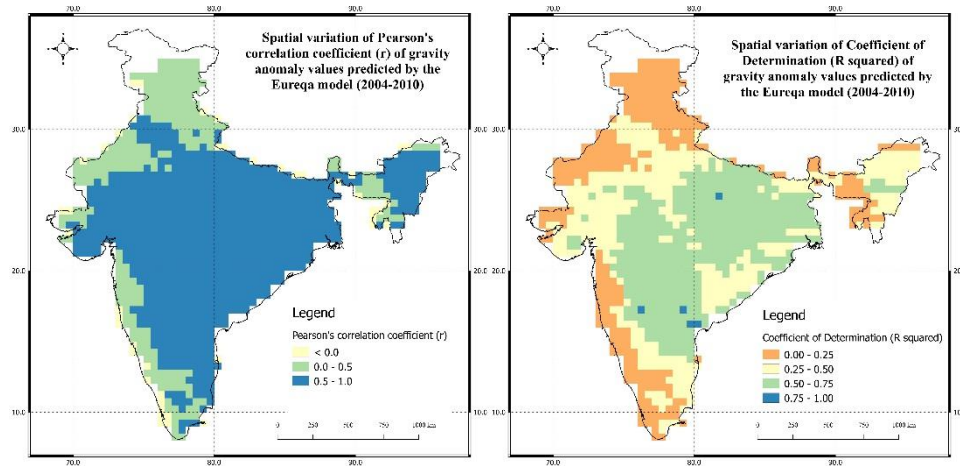


Figure 5.35 Spatial distribution of r and r^2 of predicted vs observed gravity anomalies

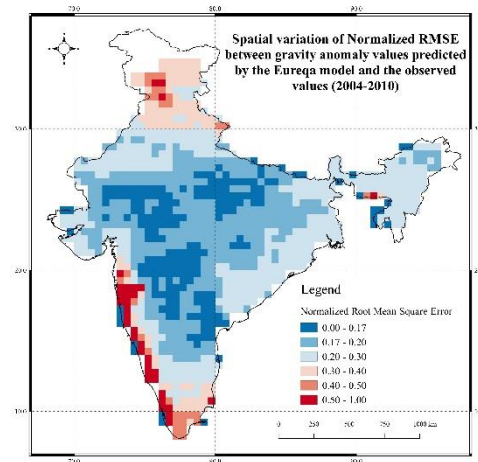


Figure 5.36 Spatial distribution of normalized RMSE between predicted vs observed gravity anomalies

5.3.2 Results obtained from EOF

Variability of the 6 land surface parameters namely albedo, gravity anomaly, NDVI, rainfall, maximum temperature and minimum temperature is calculated as described in Section 4.3.2. Figures 5.37 shows the first 2 modes of variability (or EOF-1 and EOF-2) which effectively represents 46.36% and 26.39% of the entire variability in the monthly albedo data from 2004 till 2010.

The first mode shows negative variances in Maharashtra, Assam and their adjoining areas. This result is in agreement with trends derived from time-series albedo data using STL/LOESS shown in Figure 5.10. The patches of positive trend in Maharashtra and Assam could explain the negative variances in these areas.

The spatial pattern in positive variance in the second mode of albedo are in agreement with the spatial pattern of the seasonal component of albedo time-series as derived using STL/LOESS (Figure 5.8b).

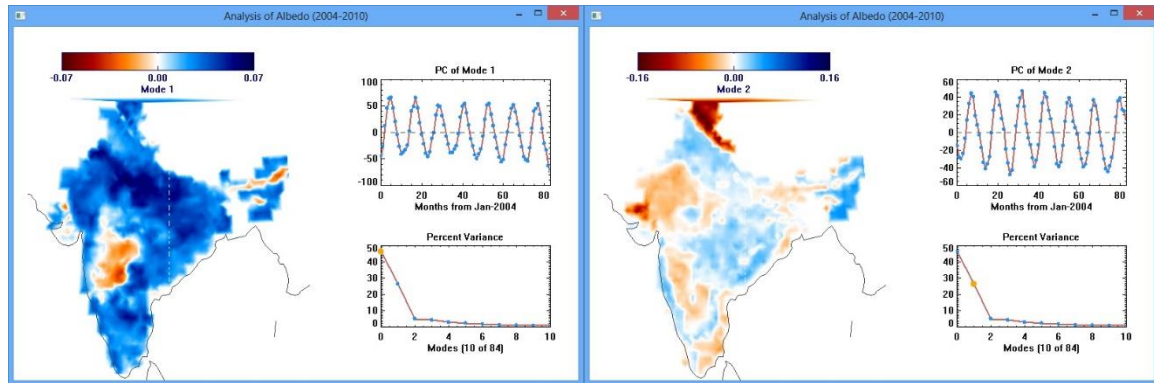


Figure 5.37 Spatial variation of first two eigenvectors of monthly mean albedo data (2004-2010)

Figures 5.38 shows EOF-1 and EOF-2 of the monthly gravity anomaly data from 2004 till 2010, which represents 83.20% and 8.49% of the entire variability in the time-series dataset. Although negative variances are seen completely in EOF-1, the Eastern and Central regions of the country explain greater negative variance in the dataset as compared to the North and South regions. The second mode, EOF-2, shows a distinct line dividing the country in 2 halves showing negative and positive variances.

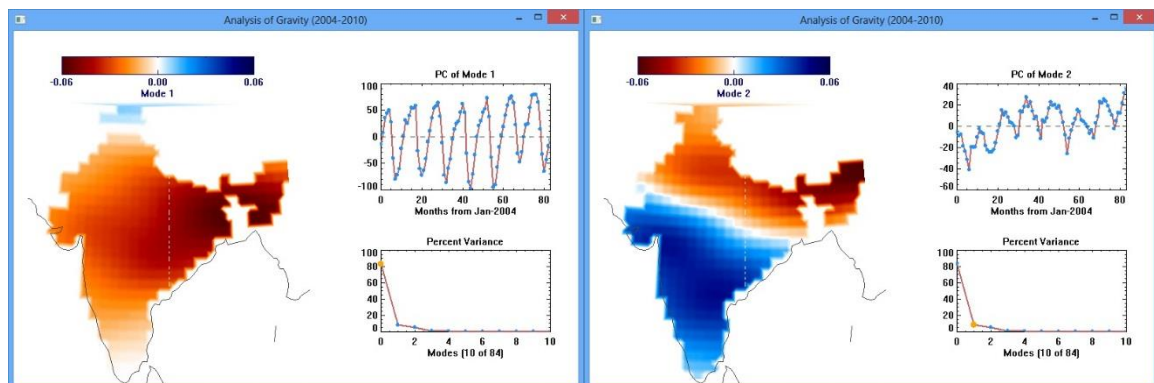


Figure 5.38 Spatial variation of first two eigenvectors of monthly mean gravity anomaly data (2004-2010)

Figures 5.39 shows the EOF-1 and EOF-2 of the monthly mean NDVI data. The EOF-1 shows negative variance covering the country. However, the spatial patterns in high negative variances match with the high seasonal component contribution as derived from STL/LOESS (Figure 5.4b).

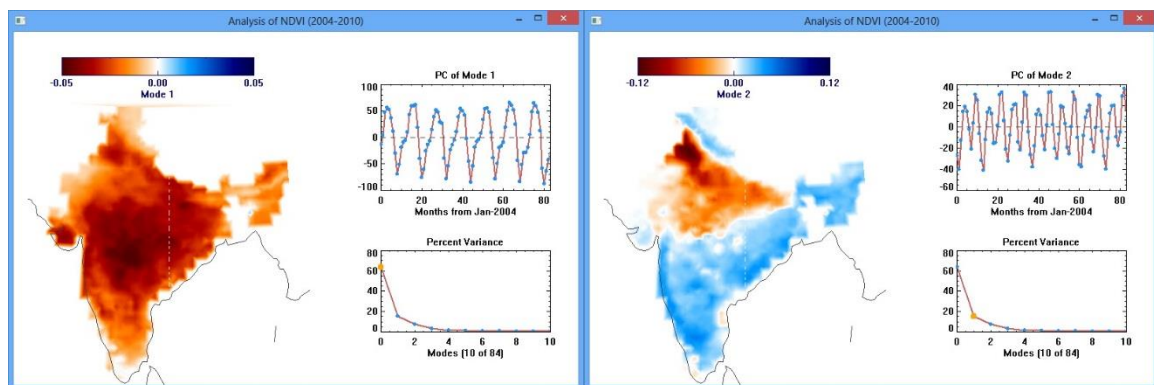
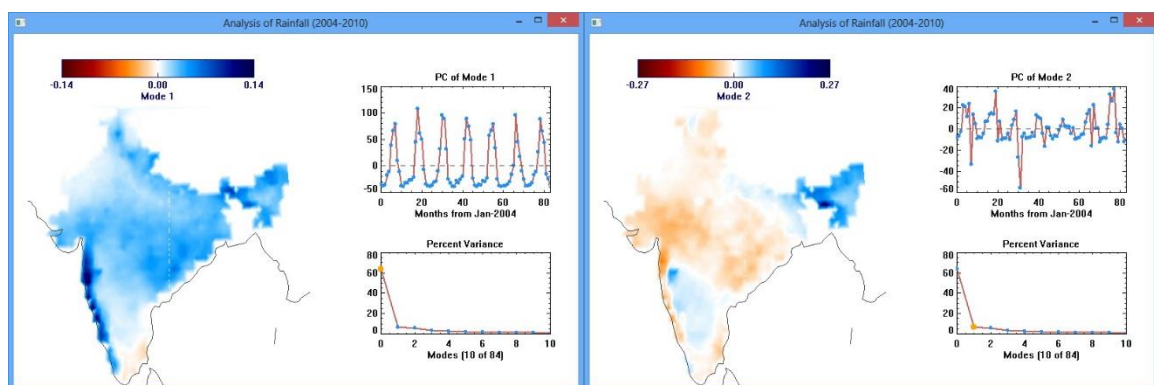


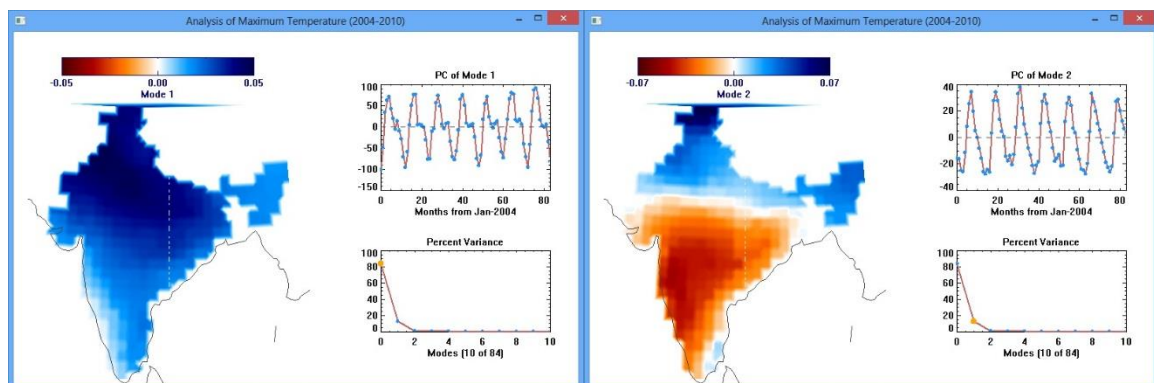
Figure 5.39 Spatial variation of first two eigenvectors of monthly mean NDVI data (2004-2010)

Figures 5.40 show spatial variability as explained by EOF-1 and EOF-2 for monthly mean rainfall data during 2004-2010. Mode 1 shows general positive variance in the country except in parts of Tamil Nadu. High positive variance is observed in Western Ghats and parts of North East which receive above average rainfall annually. EOF-2 shows negative variability in most parts of the country except Deccan and North Eastern states.



Figures 5.40 Spatial variation of first two eigenvectors of monthly mean Rainfall data (2004-2010)

The variances in monthly mean maximum temperature is shown in Figures 5.41. The first mode shows positive variability throughout the country with higher values towards North India. The second mode shows the “banding” effect as seen in STL/LOESS derived trends (Section 5.2.3).



Figures 5.41 Spatial variation of first two eigenvectors of monthly mean maximum temperature data (2004-2010)

Figures 5.42 shows the EOF-1 and EOF-2 for monthly mean minimum temperature, which explains 96.59% and 1.99% of the entire time-series dataset. The results are similar to that of minimum temperature and mimic the “banding” effect shown in Section 5.2.3.

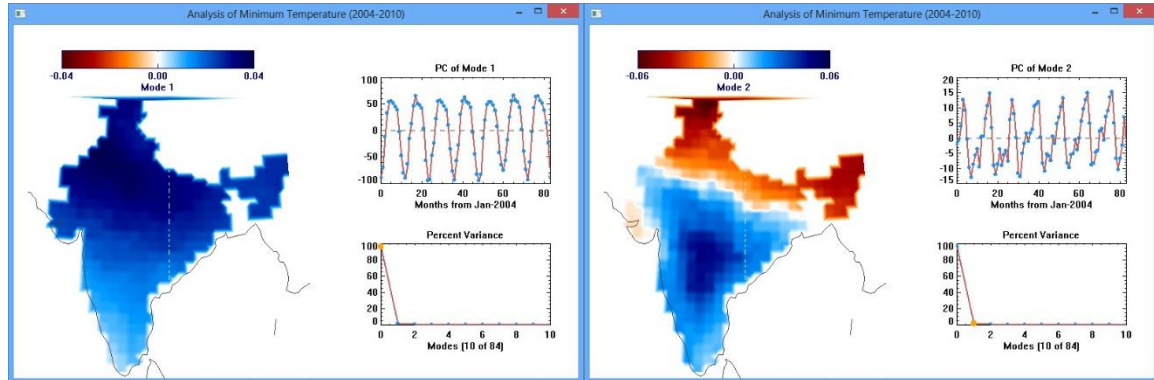


Figure 5.42 Spatial variation of first two eigenvectors of monthly mean minimum temperature data (2004-2010)

The gravity anomaly spatial variability is explained better by temperature (minimum and maximum) and NDVI as compared to Rainfall and Albedo. This data supports the results obtained from Eureka which developed a model giving more weightage of temperature values, NDVI & rainfall, and completely eliminated albedo from its model setup.

The results from simultaneous PCA performed on all 6 fields are shown in Figures 5.43 and 5.44. Figure 5.43 shows the first mode of all the 6 variables which explains the joint variances of 42.63%.

The results, although similar to single field PCA, explains the joint variance in the time-series datasets for all 6 fields. The high negative joint variance in the NDVI dataset matches spatially to the variance explained by high positive variances in albedo (Figure 5.37). These are reconfirmed by the NDVI-albedo correlation analysis done at the end of Section 5.2.2.

Figure 5.44 shows the second mode of joint variability and explains about 32.98% of the total joint variance.

The analysis was repeated by projecting the anomaly data on rotated eigenvectors. The spatial patterns and time series obtained for the rotated EOFs were found to be essentially the same as for the unrotated case. This indicates that the EOF decomposition is robust.

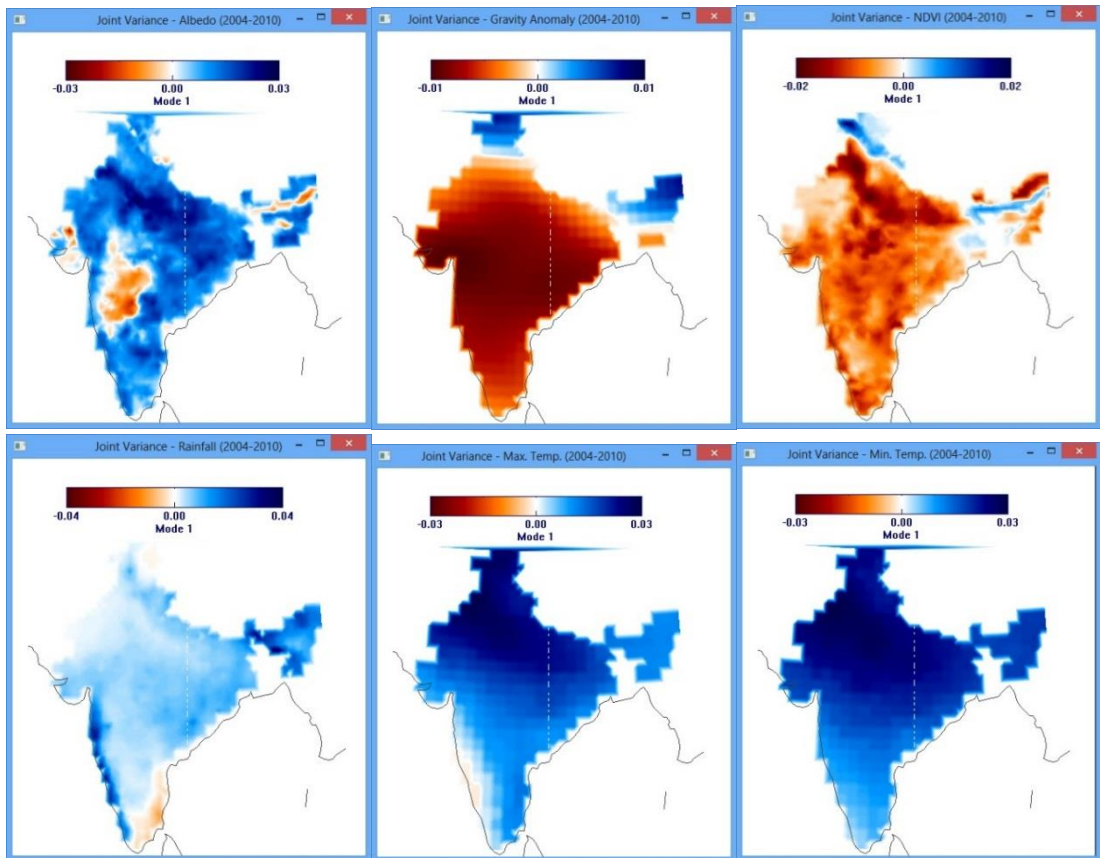


Figure 5.43 Joint spatial variation of first eigenvector in 6 land-surface parameters (2004-2010).

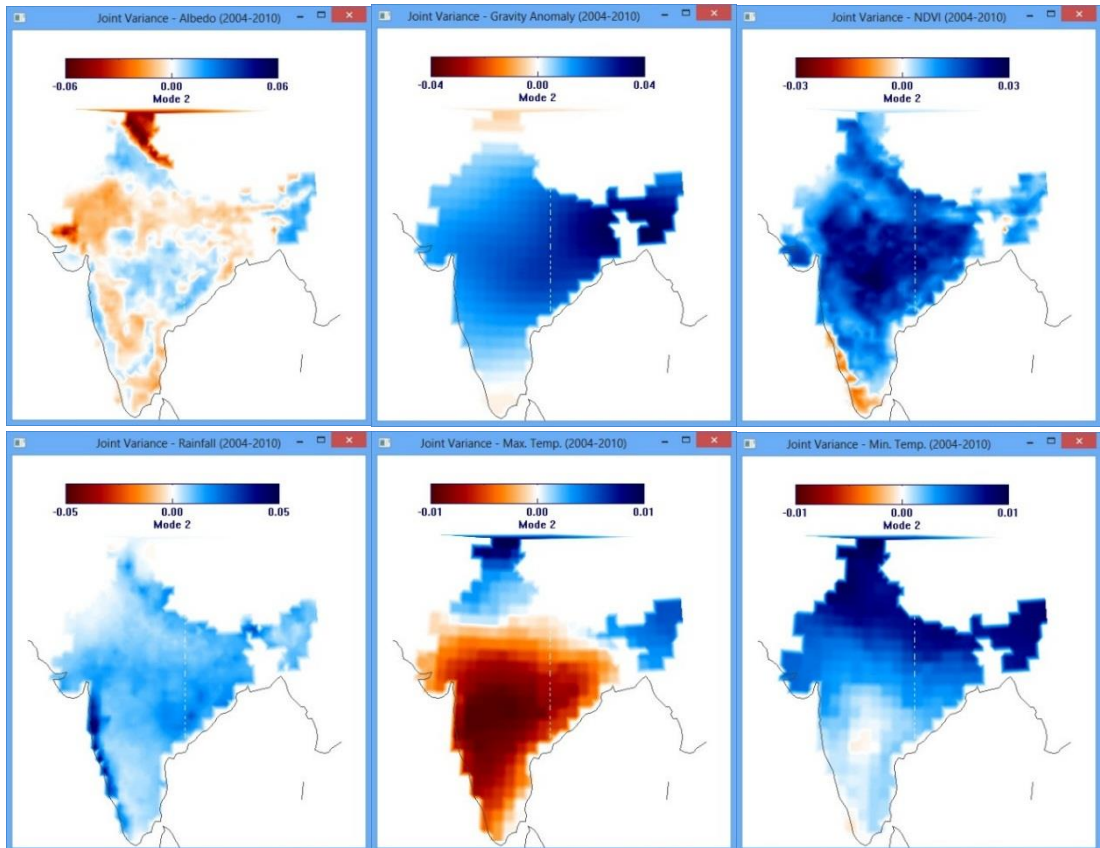


Figure 5.44 Joint spatial variation of second eigenvector in 6 land-surface parameters (2004-2010)

CHAPTER 6 - CONCLUSIONS

The study aimed to understand the effect of temporal granularity in land surface parameters and gravity anomalies and their relation with land-use/ land-cover and climate over India. To achieve this, the study was sub-divided into 3 major component / research questions, the answer to which are given below.

Does MODIS-NDVI and NOAA-AVHRR-NDVI complement each other? Can a combined time-series NDVI be generated?

NDVI plays an important role to understand vegetation phenology. Both MODIS-NDVI and GIMMS-NDVI were found to complement each other on modelling an inter-relationship between them.

A least square regression based model was used to understand the inter-dependency of MODIS-NDVI and GIMMS-NDVI. It was found that model parameters derived from linear relationships did not give acceptable statistical fitness parameters. The result did not change even when applied separately to forested (permanent vegetation cover) and non-forested (temporary vegetation cover) pixels.

Model parameters derived from strictly controlled & calibrated testing environment gave good statistical fitness between the two datasets. This enabled to create a combined time-series dataset which was used for long term studies.

Can temporal granularity help in understanding the trend in NDVI, albedo, precipitation and temperature? How do the trends vary in different statistical methods?

To understand the effect of temporal granularity, an attempt was made to deseasonalize the datasets by creating temporal aggregates. Season lengths and season start dates for different parameters are dissimilar in various parts of the country. Selection of time granules is a critical procedure which needs to be carried out carefully in order to avoid false results. Coarser/Higher aggregation levels tend to overestimate the amount of trends detected in such analysis. It may also result in higher variations in terms of significance of observed trends as well. Sufficient temporal sampling is critically required to obtain significant trends but different temporal bins shows variation in the detected significant trend.

For NDVI and Albedo, aggregates were created by averaging for annual and cropping cycle basis. For temperature and precipitation, apart from annual and monsoon season aggregates, an attempt was made to aggregate climatic datasets based on WMO recommended ETCCDI indices. Sets of four and six, seasonal and non-seasonal indices were selected from multitude of ETCCDI indices to understand the trends from them.

For any trend analysis, the choice of statistical method for evaluation and quantification of trends is very critical. The STL/LOESS technique has been applied pixel wise on all datasets without aggregation, and the technique decomposed the patterns into individual components explaining contribution of season, long-term trend and irregularities; making it less sensitive to the outliers.

It was seen that the STL/LOESS method improves on the previous linear regression analyses both by allowing a more flexible representation of the underlying trend, and by considering all aspects of the time-series simultaneously.

The M-K test assumes observations to be independent and identically distributed. These properties make M-K test a strong statistical test which is among the most widely used and extensive methods for detecting trends. Therefore, M-K test was preferred and considered to be the most reliable statistical estimators for detecting statistical significance, in this study.

The trends in albedo and NDVI were similar as derived from OLS, Sen's slope estimator and STL/LOESS. The application of these techniques on different time scale (annual, JJAS, cropping cycle etc.) based temporal aggregates revealed new trends. These results obtained from different trend analysis methods showed identical spatial patterns, even after masking for 95% confidence interval.

The application of these algorithms on climatic data reconfirmed the findings of IPCC reports that the frequency of extreme events such as anomalies (in North, West and South India), spells (in Northern J&K, Rajasthan, Eastern & Western Ghats and North Eastern states) and measures of temperature and rainfall (Northern J&K and South India) are increasing.

However, datasets and trends derived from them, for Northern J&K, Western Ghats and North Eastern states need to be further investigated for spurious trend detection.

With respect to statistical estimators, OLS and Sen's slope gave similar results for high resolution datasets (albedo and NDVI) and nearly identical results for coarse resolution datasets (temperature and rainfall).

Is there a correlation between the trends in climatic forcing, satellite derived NDVI, albedo and gravity data? Do these correlations reflect LULC and climate change scenarios?

The study was able to establish a working model based on building blocks and natural progression to understand the impact of land-surface parameters on gravity anomalies. Minimum temperature was seen to be the most sensitive parameter in the model, which could emphasize the role of condensation of water in liquid or solid form to gravity anomaly. However, the maximum likelihood is that increasing the NDVI and rainfall values will increase the gravity anomaly values. NDVI, which also stood out to be the third most sensitive parameter, shows the impact of biomass on gravity anomalies.

The EOF results also established correlation between the trends in the variables used. The spatial variability explained by the fields were found to have a spatial match with trends and outputs of other methods applied on the fields individually. Joint EOF reconfirmed Eureka output by spatially matching the variances of important fields and that there exists homogenous correlation among the variables.

General conclusions

Positive trends in NDVI could indicate effects of green revolution (plant breeding, agrochemicals, mechanized farming), irrigation development, advent of double or more cropping, afforestation and even climate change (increasing temperatures lead to longer growing season lengths).

Negative trends in NDVI could be because of deforestation, conversion of land-use from cropping to others leading to urbanisation or wastelands, conversion of double or more cropping to single crops, decline in irrigation, climate change (higher temperatures and erratic rainfall lead to low crop yield) and other socio-economic causes.

On similar lines, albedo trends could be because of increasing snow line, increase in wastelands / decrease in vegetation etc. Trends in climatic patterns are effect by many regional and global factors such as *El Niño* and *La Niña*, subtropical Indian Ocean dipole and global trade winds.

Spatio-temporal studies require immense computational resources. Modern day programming language have banked upon techniques such as compiler optimization, multi-threading and multicore processing to bring a better user experience in terms to runtime and responsiveness. Multi-purpose languages such as Python have enabled GIS specialists and computer programmers to attempt what was virtually impossible till about a decade back.

Limitations and Further recommendations

A genuine scientific research gives rise to more questions than it could possibly answer. This work has also raised questions for further research. The problem in combining distinct data at different spatial and temporal resolution will always give rise to MAUP and MTUP. The current study has a major limitation in combining datasets of different resolutions (spatial and temporal) and time scales. More emphasis is required in the right approach to study such wide array of datasets.

During this study, few problems related to data used were also encountered. Missing data, incompatible GIS data formats and humungous data size did pose some technological problems. Use of HPC and Cloud computing needs to be explored along-with newer technologies such as Machine learning and HADOOP framework, for spatial data processing.

References

- Anyamba, A., & Tucker, C. J. (2005). Analysis of Sahelian vegetation dynamics using NOAA-AVHRR NDVI data from 1981 -2003. *Journal of Arid Environments*, 63(3), 596-614.
- Bai, Z. G., Dent, D. L., Olsson, L., & Schaepman, M. E. (2008). Proxy global assessment of land degradation. *Soil use and management*, 24(3), 223-234.
- Baret, F., Guyot, G., & Major, D. J. (1989). TSAVI: a vegetation index which minimizes soil brightness effects on LAI and APAR estimation. In *International Geoscience and Remote Sensing Symposium, 1989. IGARSS'89. 12th Canadian Symposium on Remote Sensing, 1989,3*, 1355-1358. IEEE.
- Bian, L., & Butler, R. (1999). Comparing effects of aggregation methods on statistical and spatial properties of simulated spatial data. *Photogrammetric Engineering and Remote Sensing*, 65, 73-84.
- Bretherton, C. S., Smith, C., & Wallace, J. M. (1992). An intercomparison of methods for finding coupled patterns in climate data. *Journal of climate*, 5(6), 541-560.
- Buishand, T. A., Shabalova, M. V., & Brandsma, T. (2004). On the choice of the temporal aggregation level for statistical downscaling of precipitation. *Journal of Climate*, 17(9), 1816-1827.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess, *Journal of Official Statistics*, 6(1), 3-73.
- Çöltekin, A., De Sabbata, S., Willi, C., Vontobel, I., Pfister, S., Kuhn, M., & Lacayo, M. (2011). Modifiable temporal unit problem. In *ISPRS/ICA workshop "Persistent problems in geographic visualization" (ICC2011), Paris, France* (Vol. 2).
- de Jong, R., & de Bruin, S. (2012). Linear trends in seasonal vegetation time-series and the modifiable temporal unit problem. *Biogeosciences*, 9(1), 71-77.
- de Jong, R., de Bruin, S., de Wit, A., Schaepman, M. E., & Dent, D. L. (2011). Analysis of monotonic greening and browning trends from global NDVI time-series. *Remote Sensing of Environment*, 115(2), 692-702.
- de Jong, R., & de Bruin, S. (2011). Time-series of vegetation indices and the modifiable temporal unit problem. *Biogeosciences Discussions*, 8(4), 8545-8561.
- Downey, A. (2012). *Think Python: How to Think Like a Computer Scientist*. Green Tea Press.
- Dubois, P. F., Hinsen, K., & Hugunin, J. (1996). Numerical Python. *Computers in Physics*, 10, 262-267.
- Fanning, D. J. (2003). IDL programming techniques--. *IDL programming techniques--2nd ed./David W. Fanning. Fort Collins, Co. Fanning Software Consulting*.
- GDAL. 2014. GDAL - Geospatial Data Abstraction Library: Version 1.11.1, Open Source Geospatial Foundation, <http://gdal.osgeo.org>
- Gotway, C. A., & Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, 97(458), 632-648.

- He, H. S., Ventura, S. J., & Mladenoff, D. J. (2002). Effects of spatial aggregation approaches on classified satellite imagery. *International Journal of Geographical Information Science*, 16(1), 93-109.
- Herrmann, S. M., Anyamba, A., & Tucker, C. J. (2005). Recent trends in vegetation dynamics in the African Sahel and their relationship to climate. *Global Environmental Change*, 15(4), 394-404.
- Hijmans, R. J. (2014). raster: raster: Geographic data analysis and modeling. R package version 2.3-12. <http://CRAN.R-project.org/package=raster>
- Hill, M. J., & Donald, G. E. (2003). Estimating spatio-temporal patterns of agricultural productivity in fragmented landscapes using AVHRR NDVI time-series. *Remote Sensing of Environment*, 84(3), 367-384.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 0090-95.
- Jelinski, D. E., & Wu, J. (1996). The modifiable areal unit problem and implications for landscape ecology. *Landscape ecology*, 11(3), 129-140.
- Jones, E., Oliphant, T., & Peterson, P. (2001). SciPy: Open source scientific tools for Python. <http://www.scipy.org/>.
- Justice, C. O., Vermote, E., Townshend, J. R. G., Defries, R., Roy, D. P., Hall, D. K., Salomonson, V. V., Privette, J. L., Riggs, G., Strahler, A., Lucht, W., Myneni, R. B., Knyazikhin, Y., Running, S. W., Nemani, R. R., Wan, Z. M., Huete, A. R., Van Leeuwen, W., Wolfe, R. E., Giglio, L., Muller, J. P., Lewis, P., & Barnsley, M. J. (1998). The Moderate Resolution Imaging Spectroradiometer (MODIS): Land Remote Sensing for Global Change Research. *IEEE Transactions on Geoscience and Remote Sensing*, 36(4), pp 1228 -1249.
- Kendall, M. G., & Yule, G. U. (1950). *An introduction to the theory of statistics*. Charles Griffin & Company.
- Lee, E., Chase, T. N., Rajagopalan, B., Barry, R. G., Biggs, T. W., & Lawrence, P. J. (2009). Effects of irrigation and vegetation activity on early Indian summer monsoon variability. *International Journal of Climatology*, 29(4), 573-581.
- Maignan, F., Bréon, F. M., Bacour, C., Demarty, J., & Poirson, A. (2008). Interannual vegetation phenology estimates from global AVHRR measurements: Comparison with in situ data and applications. *Remote Sensing of Environment*, 112(2), 496-505.
- Moody, E. G., King, M. D., Platnick, S., Schaaf, C. B., & Gao, F. (2005). Spatially complete global spectral surface albedos: Value-added datasets derived from Terra MODIS land products. *IEEE Transactions on Geoscience and Remote Sensing*, 43(1), 144-158.
- Openshaw, S., & Taylor, P. J. (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. *Statistical applications in the spatial sciences*, 21, 127-144.
- Pai, D. S., Sridhar, L., Rajeevan, M., Sreejith, O. P., Satbhai, N. S., & Mukhopadhyay, B. (2014). Development of a new high spatial resolution (0.25°×0.25°) long period (1901 -2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *Mausam*, 65(1), 1-18.

Partal, T., & Kahya, E. (2006). Trend analysis in Turkish precipitation data. *Hydrological processes*, 20, 2011-2026.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

Raj, R., Hamm, N. A., & Kant, Y. (2013). Analysing the effect of different aggregation approaches on remotely sensed data. *International Journal of Remote Sensing*, 34(14), 4900-4916.

Ramsay, J. O., Wickham, H., Graves, S. & Hooker, G. (2014). fda: Functional Data Analysis. R package version 2.4.4. <http://CRAN.R-project.org/package=fda>

Roy, P. S., Kushwaha, S. P. S., Murthy, M. S. R., Roy, A., Kushwaha, D., Reddy, C. S., & Behra, M. D. *Biodiversity characterisation at landscape level: national assessment. Indian Institute of Remote Sensing, Dehradun, India*. ISBN 81-901418-8-0, (2012).

Schmidt, M., & Lipson, H. (2013). Eureqa (Version 1.10). Available from <http://www.eureqa.com/>

Schmidt, M., & Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science*, 324(5923), 81-85.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63(324), 1379-1389.

Shellman, S. M. (2004). Time-series intervals and statistical inference: The effects of temporal aggregation on event data analysis. *Political Analysis*, 12(1), 97-104.

Sonali, P., & Kumar, D. N. (2012). Review of trend detection methods and their applications to detect temperature changes in India. *Journal of Hydrology*, 476, 212-227.

Srivastava, A. K., Rajeevan, M., & Kshirsagar, S. R. (2009). Development of a high resolution daily gridded temperature data set (1969 -2005) for the Indian region. *Atmospheric Science Letters*, 10(4), 249-254.

Steven, M. D., Malthus, T. J., Baret, F., Xu, H., & Chopping, M. J. (2003). Intercalibration of vegetation indices from different sensor systems. *Remote Sensing of Environment*, 88(4), 412-422.

Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C. (2004). The gravity recovery and climate experiment: Mission overview and early results. *Geophysical Research Letters*, 31(9).

Tiwari, V. M., Wahr, J., & Swenson, S. (2009). Dwindling groundwater resources in northern India, from satellite gravity observations. *Geophysical Research Letters*, 36(18).

Tucker, C.J., Pinzon, J. E., Brown, M. E., Slayback, D., Pak, E. W., Mahoney, R., Vermote, E., & El Saleous, N. (2005). An Extended AVHRR 8-km NDVI Data Set Compatible with MODIS and SPOT Vegetation NDVI Data. *International Journal of Remote Sensing*, 26(20), 4485-5598.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 91 in the International Geophysical Series). Academic press.

Yang, Y., Long, D., Guan, H., Scanlon, B. R., Simmons, C. T., Jiang, L., & Xu, X. (2014). GRACE satellite observed hydrological controls on interannual and seasonal variability in surface greenness over mainland Australia. *Journal of Geophysical Research: Biogeosciences*, 119(12), 2245-2260.