Analysis of External Drift Kriging Algorithm with application to precipitation estimation in complex orography.

Bernard Sharon Majani.
January, 2007
Analysis of External drift kriging algorithm with application to precipitation estimation in complex orography.

by

Bernard.S.Majani

Thesis submitted to the International Institute for Geo-information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation, Specialisation: (Geo-informatics)
Dedicated to my late granddad

Mr J.C. Majani

Thanks for believing in me and setting an example as a man of Integrity.

...and to my Loving parents.
Disclaimer

This document describes work undertaken as part of a programme of study at the International Institute for Geo-information Science and Earth Observation. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.
Abstract

Patterns of precipitation are greatly influenced by global geography such as presence of mountains and distance from the sea.

In this research analysing kriging with external drift algorithm and its performance in estimating precipitation in complex orography is worked at to come up with the best model for precipitation estimation in complex orography.

Since rainfall was sparsely sampled we complemented this using secondary attributes that were more densely sampled. The covariates used were elevation, slope aspect, Distance from the coast, vegetation cover factor (from vegetation continuous field) and Topographic Index which is a derived covariate.

The area chosen for this study was Karnataka and Uttaranchal which are States in India. South west and North respectively.

Monthly Rainfall datasets for year 2000 were used for Karnataka and 50 year monthly average rainfall for 1900-1950 was used for Uttaranchal. The covariates were extracted from Satellite images and incorporated as tabular data. Studies were conducted on three months from three seasons: pre-monsoon, monsoon and post-monsoon period. Open source software, R statistics (Gstat) was used for processing and statistical analysis of the data.

The product was six models instead of one universal model that was to be the best precipitation estimator in the two regions for the three different seasons of study. This was attributed to the complexity of the study area, using the data and covariates. It was noted that apart from the covariates used, there are other climatic factors that do play part in contributing to the amount, type and pattern of precipitation received. The “other” climatic factor that stood out was wind, especially during monsoon period. It was also noted that the area received more rainfall during this time and among the covariates used elevation contributes more to the type and amount of rainfall received and its pattern.

Of the six models two had elevation and one a combination of elevation and slope. We expected to find kriging with external drift perform better than ordinary kriging, however it only performed better in one study area (Karnataka).

KEYWORDS: Kriging with external drift, Covariates, Precipitation, Ordinary kriging, Elevation, Slope, Aspect, Distance from coast, Vegetation cover factor, Topographic Index, Geostatistics, Cross validation. RMSE.
Acknowledgements

First and foremost I thank the Almighty God, my Creator and Guide, who gave me the strength and knowledge to go through this rigorous course and execute this work.

I want to express my deep and sincere gratitude to the World Meteorological organization (W.M.O) for giving me chance to study by granting me a scholarship. The Director of Kenya Meteorological Department and Permanent representative of Kenya with W.M.O Dr J.R. Mukabana for his professional advice and support.

I have deep gratitude to Professor, Dr Alfred Stein, Chairman Department of Earth observation Science, Dr V.K Dadhwal, Dean IIRS and Mr C Jeganathan Geo-informatics Course coordinator, for their continued professional advice and guidance throughout my research, thanks for giving me views on the subject and assisting direct my thesis in the right direction to adapt a more scientific approach. I am honoured to have worked with you as my supervisors. What you have invested in me, I promise to use and also pass on to others that may need it.

I am forever grateful to Dr Rossiter who taught me geostatistics and supported me tirelessly to get Geostats codes flowing in ma head and system, and for advice from Edzer Pebesma, and the R-Geostats forum. Dr Kassteele for providing me with his work and offering to advice.

I’m thankful to all the lecturers who taught me during my Msc programme both at IIRS and ITC. My Mum and Dad and bros and sis for being there for me, and believing in me, praying with and for me.

I’ve got lotsa appreciation for all my friends at IIRS and ITC, classmates for the love and encouragement you showed during our time together especially Mr Okhandiar, Kawi, Rajiv, Rafiq, Haille, Gonu, Coco, Ben Mayala, Pankaj, Paco, Devi, Ahmed, Mona, Vinat, Smile, Pook, Lesslie, Lydia, Paul (Ghana) and Hanan. I will always remember the joys and struggles we shared and your willingness to help in times of need.

It was crazy being the only “meero” in Dehradun, life brightened up when I had warm and supportive wananchi wenzangu from mtaani at ITC, Shukran and mob barakas, for making maisha at ITC sasawa.

Not forgetting the Vineyard fellowship of India who were a constant support and encouragement standing in the gap for me when times were hard, thanks and God bless you Big, especially pastor Sunny, Vivek and my home group for being family away from home. The ITC Christian fellowship for offering a warm fellowship that I could also be involved in ministry with.

To all the people I met and made friends with during the course and who supported me in whichever way, I appreciate a lot. The guys we participated with in the FBK games in Hengelo, friends and teammates we participated with in the Batavieren race, especially Mr Wan Bakx for lots of support encouragement and believing in me as a triathlete, to Randolf for times we cycled in the German neighbourhoods and the woods in Enschede.

Friends who called and supported me when I worked to the wee morning hours, Abra Maze God akubless, Dave Mwangi, Keshi n the mum, Nimo, Corrine and Sam thanx for ya prayers and encouragement, Mice, Sues, Lemuel, Clem, CK, n Ngash.

I may not have reached “the end” if you all weren’t by my side, this is but a small product of your investment in my life.- Muchas gracias-
# Table of contents

1. INTRODUCTION .............................................................................................................1
   1.1. INTERPOLATION ........................................................................................................2
   1.2. THE GEOSTATISTICAL APPROACH ...........................................................................2
       1.2.1. KRIGING .............................................................................................................2
   1.3. PRECIPITATION/RAINFALL .....................................................................................3
   1.4. MEASUREMENT OF PRECIPITATION .....................................................................4
   1.5. PROBLEM STATEMENT ............................................................................................4
   1.6. RESEARCH OBJECTIVE ...........................................................................................5
   1.7. RESEARCH QUESTIONS ...........................................................................................5
   1.8. ASSUMPTION AND HYPOTHESES .........................................................................5
       1.8.1. ASSUMPTION ......................................................................................................5
       1.8.2. HYPOTHESES ...................................................................................................6
   1.9. STRUCTURE OF THESIS ..........................................................................................6
2. LITERATURE SURVEY ......................................................................................................7
   2.1. INTRODUCTION ........................................................................................................7
   2.2. OROGRAPHIC PRECIPITATION ...............................................................................7
   2.3. PREVIOUS STUDIES AND THEIR APPLICATIONS .....................................................8
       2.3.1. USE OF GEOSTATISTICAL METHODS ...............................................................8
       2.3.2. USE OF OTHER NON GEO-STATISTICAL METHODS .......................................9
       2.3.3. USE OF COVARIABLES .....................................................................................10
       2.3.4. KRIGING WITH EXTERNAL DRIFT ..................................................................11
       2.3.5. PRECIPITATION ESTIMATION .......................................................................12
       2.3.6. COMPARATIVE ANALYSIS WITH OTHER METHODS ....................................14
3. STUDY AREA ....................................................................................................................18
   3.1.1. CLIMATE .............................................................................................................18
   3.1.2. KARNATAKA .........................................................................................................18
   3.2. UTTARANCHAL .........................................................................................................21
       3.2.1. LOCATION AND BOUNDARIES ......................................................................21
       3.2.2. CLIMATE ..........................................................................................................21
       3.2.3. RAINFALL .........................................................................................................21
       3.2.4. PHYSIOGRAPHY ...............................................................................................22
4. MATERIALS AND METHODS ............................................................................................23
   4.1. MATERIALS ................................................................................................................23
       4.1.1. DATA DESCRIPTION AND PROCESSING ..........................................................23
   4.2. SOFTWARE USED .....................................................................................................24
   4.3. METHODS ..................................................................................................................25
       4.3.1. INTERPOLATION METHODS ..........................................................................25
   4.4. SPATIAL INTERPOLATION .......................................................................................25
   4.5. CLASSIFICATION OF INTERPOLATION PROCEDURES ........................................25
       4.5.1. POINT/AREAL INTERPOLATION ....................................................................25
       4.5.2. AREAL INTERPOLATION .................................................................................25
       4.5.3. GLOBAL/LOCAL INTERPOLATORS .................................................................25
       4.5.4. EXACT/APPROXIMATE INTERPOLATORS .....................................................25
List of figures

Figure 3-1 KARNATAKA LOCATION ................................................................. 19
Figure 3-2 UTTARANCHAL LOCATION .......................................................... 21
Figure 4-1 VARIOGRAM ................................................................................. 27
Figure 4-2 COMMON VARIOGRAM MODELS Source: (Hartkamp, 1999) .......... 28
Figure 4-3 ANISOTROPY ELIPSE .................................................................. 29
Figure 4-4 VARIOGRAM Figure 4-5 MODELLED VARIOGRAM ..................... 40
Figure 4-6 AUTOFIT VARIOGRAM ................................................................. 40
Figure 4-7 OK INTERPOLATION USING EDITED SCRIPT ................................. 41
Figure 4-8 O.K DEFAULT R INTERPOLATION .............................................. 42
Figure 4-9 RESIDUALS HISTOGRAM Figure 4-10 RESIDUALS BUBBLE PLOT .. 42
Figure 4-11 KED INTERPOLATION USING EDITED SCRIPT ......................... 44
Figure 4-12 KED INTERPOLATION USING NORMAL SCRIPT ....................... 44
Figure 5-1 Figure 5-2 ....................................................................................... 47
Figure 5-3 Figure 5-4 ....................................................................................... 47
Figure 5-5 Figure 5-6 ....................................................................................... 47
Figure 5-7 Figure 5-8 ....................................................................................... 48
Figure 5-9 Figure 5-10 ..................................................................................... 48
Figure 5-11 Figure 5-12 .................................................................................. 49
Figure 5-13 Figure 5-14 .................................................................................. 50
Figure 5-15 Figure 5-16 .................................................................................. 50
Figure 5-17 Figure 5-18 .................................................................................. 51
Figure 5-19 Figure 5-20 .................................................................................. 51
Figure 5-21 Figure 5-22 .................................................................................. 52
Figure 5-23 Figure 5-24 .................................................................................. 52
Figure 5-25 Figure 5-26 .................................................................................. 53
Figure 5-27 Figure 5-28 .................................................................................. 53
Figure 5-29 Figure 5-30 .................................................................................. 54
Figure 5-31 Figure 5-32 Figure 5-33 ............................................................... 56
Figure 5-34 Figure 5-35 .................................................................................. 57
Figure 5-36 Figure 5-37 .................................................................................. 58
Figure 5-38 Figure 5-39 .................................................................................. 58
Figure 5-40 Figure 5-41 .................................................................................. 59
Figure 5-42 Figure 5-43 .................................................................................. 59
Figure 5-44 Figure 5-45 .................................................................................. 60
Figure 5-46 Figure 5-47 .................................................................................. 60
Figure 5-48 Figure 5-49 .................................................................................. 61
Figure 5-50 Figure 5-51 .................................................................................. 61
Figure 5-52 Figure 5-53 .................................................................................. 62
Figure 5-54 Figure 5-55 .................................................................................. 62
Figure 5-56 Figure 5-57 .................................................................................. 62
Figure 5-58 Figure 5-59 .................................................................................. 63
Figure 5-60 UTTARANCHAL ELEVATION Figure 5-61 KARNATAKA ELEVATION ... 67
Figure 5-62 WIND IN ASIA OVER SUMMER Figure 5-63 WIND IN ASIA OVER WINTER .... 69
Figure 5-64 WINTER AND SUMMER MONSOON (PRESSURE) ..................................................70
Figure 5-65 DIRECTIONAL VARIOGRAM  Figure 5-66 ANISOTROPY MAP ......................72
List of tables

Table 4-1 OK KED COMPARISON ................................................................. 45
Table 5-1 CORRELATION COEFFICIENT 2KM GRID ........................................ 46
Table 5-2 CORRELATION COEFFICIENT 10KM GRID ........................................ 46
Table 5-3 DETAILS FOR KARNATAKA VARIOMGRAMS ........................................ 49

Table 5-4 ROOT MEAN SQUARED ERROR AND MEAN ERROR FOR THE “BEST” MODELS IN THE THREE MONTHS, APRIL, JULY AND DECEMBER AND COMPARISON OF THEIR PERFORMANCE IN THE OTHER MONTHS ........................................................................ 55

Table 5-5 COMPARISON BETWEEN ORDINARY KRIGING AND EXTERNAL DRIFT KRIGING ROOT MEAN SQUARED ERRORS ........................................................................ 55

Table 5-6 CORRELATION COEFFICIENT FOR UTTARANCHAL .................. 56
Table 5-7 DETAILS FOR VARIOMGRAMS FOR UTTARANCHAL .................. 57

Table 5-8 ROOT MEAN SQUARED ERROR AND MEAN ERROR FOR THE “BEST” MODELS IN THE THREE MONTHS, APRIL, JULY AND DECEMBER AND COMPARISON OF THEIR PERFORMANCE IN THE OTHER MONTHS ........................................................................ 64

Table 5-9 MODEL PERFORMANCE COMPARISON BETWEEN ORDINARY KRIGING AND KRIGING WITH EXTERNAL DRIFT ........................................................................ 64
1. INTRODUCTION

Patterns of precipitation are greatly influenced by global geography such as presence of mountains and distance from the sea. Spatial modelling of a climatic variable is an important issue, because many other environmental variables depend on climate (Marquenez, 2003)

Geo-statistics, which is based on the theory of regionalized variables, is increasingly preferred because it allows one to capitalize on the spatial correlation between neighbouring observations to predict attribute values at unsampled locations. (Phillips et al., 1992). Geo-statistical methods have been shown to be superior to several other estimation methods, such as Thiessen polygon, polynomial interpolation, and inverse distance method (Creutin, 1982; Tabios and Salas, 1985) One of the advantages of geostatistical methods is to use additional information to improve precipitation estimations. (Kravchenko A, 1996)

In Geostatistics, when the variable of interest is sparse or poorly correlated in space, the prediction of this variable over the whole study area may be improved by accounting for secondary information exhaustively sampled over the same study area. The secondary information can be incorporated using kriging with an external drift (Delhomme, 1978; Ahmed and De Marsily, 1987; Galli and Meunier, 1987; Hudson and Wackernagel, 1994; Bourennane et al, 1996, 2000; Gotway and Hartford, 1996). In these studies, only one exhaustive variable was used as an external drift in predicting the variable of interest. However, the use of multi-external drifts in predicting the target variable has been used before in petroleum and hydrology (Renard and Nai-Hsien, 1988; Chile`s, 1991).

A major advantage of Kriging over simpler methods is that sparsely sampled observations of the primary attribute can be complemented by secondary attributes that are more densely sampled. For rainfall, secondary information can take the form of weather-radar observations (Goovaerts, 1999)

Rainfall information is an important input in the hydrological modelling, predicting extreme precipitation events such as droughts and floods, estimating quantity and quality of surface water and groundwater. However, in most cases, the network of the precipitation measuring stations is sparse and available data are insufficient to characterize the highly variable precipitation spatial distribution. This is especially true for mountain areas, where the complexity of the precipitation distribution is combined with the measurement difficulties. Therefore, it is necessary to develop methods to estimate precipitation in areas where precipitation has not been measured, using data from the surrounding weather stations. (Kravchenko A, 1996)

Precipitation information is of interest for a variety of purposes including public safety, water supply, dam design and operation and transport planning. In coastal areas, like Karnataka, orographic precipitation may result from uplift of a moist, but shallow, boundary layer causing increasing precipitation with increasing elevation. Above this layer, the atmosphere becomes drier, causing precipitation to decrease with elevation. When the terrain is complex like in Karnataka then there’s a difference in amount of precipitation received over very short distances. This makes it crucial to have models that can accurately estimate amount of rainfall in space and time.
1.1. INTERPOLATION

During analysis there are some variables which escape our measurement; this is determined directly by the scale and distribution of our data collection, or survey, and its methodology. There are many limitations in data collection and these make the direct measure of continuous spatial data impossible. Probability functions are applied to create an interpolation surface, predicting unmeasured variables at innumerable locations.

A number of ways have been proposed for the interpolation of rainfall data. The simplest approach consists of assigning weights to the more complex ones like kriging (Goovaerts, 2000)

1.2. THE GEOSTATISTICAL APPROACH

Geostatistics started from mineral mining, and is currently applied in many disciplines such as hydrogeology, hydrology, meteorology, and epidemiology. It involves analysis and prediction of spatial or temporal phenomena. The prefix geo is usually associated with geology, owing to it having originated from mining.

Geostatistics addresses the need to make predictions of sampled attributes (i.e., maps) at unsampled locations from sparse, often “expensive” data. To make up for lack of observed data, Geostatistics has concentrated on the development of powerful methods based on stochastic theory.

Geostatistics is associated with a class of techniques for analyzing and predicting values of a variable distributed in space or time. Such values are implicitly assumed to be correlated with each other and the study of such a correlation is usually called a “structural analysis” or “variogram modelling”.

After a structural analysis, predictions at unsampled locations are made using “kriging” or they can be simulated using “conditional simulations”. Briefly the steps in a geostatistical study include:

a) Exploratory data analysis
b) Structural analysis (calculation and modelling of variograms)
c) Making predictions (kriging or simulations)

The main distinction with statistics is that in Geostatistics the variables used are linked to locations. Observations in space are linked to their co-ordinates and each observation has its specific place in space. These are also sometimes referred to as “regionized variables” or “geovariables”.

1.2.1. KRIGING

Kriging is a geostatistical estimation technique which uses a linear combination of sampled values to make predictions. Several authors (Phillips et al., 1992; Tabios and Salas, 1985) have shown that the geostatistical prediction technique (kriging) provides better estimates of rainfall than conventional methods.

To predict we need to know the weights applied to each surrounding sampled data. Kriging allows deriving weights that result in optimal and unbiased estimates. Within a probabilistic framework, kriging attempts to:

a. Minimize the error variance and
b. Systematically set the mean of the prediction errors to zero, so that there are no over – or under-estimates.
There are some cases when other, usually more abundantly sampled data, is used to predict target variable. This data is referred to as secondary data and are assumed to be correlated with the primary data. For example, Kriging is expressed in a “kriging system”, which relates the covariance between the samples, the covariance between each sample to the location to be estimated and the unknown weights. The covariance matrix is inverted to solve for the weights. Kriging allows quantifying the quality of predictions via the kriging variance. The results from kriging are generally of a higher quality and more realistic as compared to the techniques such as inverse distance and triangulation because kriging minimizes the prediction error variance. A measure of the quality of the prediction such as the variance or the standard deviation is also required to assess the reliability of an interpreted map. When predicting in the presence of spatial dependence, observations close to each other are more likely to be similar than those far apart.

1.2.1.1. KRIGING WITH EXTERNAL DRIFT

Kriging with external drift (KED) is a non stationary geostatistical method. Here, we focus on the use of secondary information from a model to obtain better prediction. In the case of KED, predictions at new locations are made by:

\[
Z_{KED}(s_o) = \sum_{i=1}^{n} w_{i}^{KED} * Z(s_i)
\]

For

\[
\sum_{i=1}^{n} w_{i}^{KED} * q_k(s_i) = q_k(s_o); k = 1,...,p
\]

Where \( z \) is the target variable, \( q_k \)'s are the predictor variables i.e. values at a new location \( (s_o) \), \( \delta_0 \) is the vector of KED weights \( (w_{i}^{KED}) \), \( p \) is the number of predictors and \( Z \) is the vector of \( n \) observations at primary locations. (Tomislav and Stein, 2003)

External drift kriging algorithm is explained in-depth in chapter 4 under methods.

1.3. PRECIPITATION/RAINFALL

Precipitation is any product of the condensation of atmospheric water vapour that is deposited on the earth's surface. Precipitation that forms aloft is divided into three categories: Liquid Precipitation; Freezing Precipitation and Frozen Precipitation. It is any form of water that falls from the sky as part of the weather to the ground. This includes snow, rain, sleet, freezing rain, hail, and virga. Precipitation is a major component of the hydrologic cycle, and is responsible for depositing most of the fresh water on the planet. Precipitation begins forming when relatively warm, moist air rises. As the air cools, water vapour begins to condense on condensation nuclei, forming clouds. After the water droplets grow large enough form precipitation. This study addresses orographic precipitation which is the kind of precipitation that occurs on the windward side of mountains. It is caused by the rising air motion of a large-scale flow of moist air across the mountain ridge, resulting in adiabatic cooling and condensation. In parts of the world subjected to relatively consistent winds (for example the trade winds), a wetter climate prevails on the windward side of a mountain than on the leeward (downwind) side as moisture is removed by
orographic precipitation, leaving drier air on the descending (generally warming), leeward side where a rain shadow is observed (Sumner, 1988).

1.4. MEASUREMENT OF PRECIPITATION

Precipitation is mostly measured using rain gauges which are placed on the ground and collect rain etc. It can also be measured or estimated by using radar and satellites. There are some areas which are inaccessible and this makes it difficult to make measurements through the collection of rain in rain gauges. Rain gauges provide information on precipitation, but existing rain-gauge networks, especially in mountainous areas, are generally not dense enough to reveal variability in precipitation over spatial scales of tens of kilometres scales over which topography and precipitation can vary significantly (Anders, 2006).

This calls for different ways of measuring amount of precipitation received in these inaccessible areas and leads to interpolation.

Measured rainfall data are important to many problems in hydrologic analysis and designs. For example the ability of obtaining high resolution estimates of spatial variability in rainfall fields becomes important for identification of locally intense storms which could lead to floods and especially to flash floods. The accurate estimation of the spatial distribution of rainfall requires a very dense network of instruments, which entails large installation and operational costs. Also, vandalism or the failure of the observer to make the necessary visit to the gauge may result in even lower sampling density. It is therefore necessary to estimate point rainfall at unrecorded locations from values at surrounding sites (Goovaerts, 1999).

1.5. PROBLEM STATEMENT

Application of kriging in meteorological studies are rather limited (notable exceptions include (Creutin, 1982) and are often restricted to studies in acid rain (Venkatram, 1988). Measuring at every point where data is needed is prohibited by large associated costs, for this reason spatial interpolation procedures deserve increasing attention (Beek E.G, 1992).

Good quality modelling of precipitation patterns, for example, is a common issue in hydrology, agro-climatology and geomorphology, since the quality of input parameters is a crucial condition to obtain reliable and accurate results over a range of local-to-regional scales.

To estimate precipitation at unsampled locations over a region, a procedure is required to calculate the spatial climatological precipitation mean or the climatological mean at specific locations between the rain gauge stations. However, the precipitation pattern in mountainous areas is not very well known, partly because of the complex topography in these regions and partly because of the sparsity of pluviometric information (Prudhomme and Reed, 1999). Thus, mapping precipitation is critical in mountainous regions with complex rainfall gradients, where the use of radar is problematic and stations are located in valleys with easy access but have low precipitation compared with the surrounding higher terrain (Johansson and Chen, 2003). In complex orography it is hard to use radar information because deriving quantitative precipitation information is hindered by cluttered and blocked regions in mountains.

Several additional spatial climate-forcing factors are most important at scales of less than 1 km, but may also have effects at larger scales. These factors, which include slope and aspect, riparian zones,
and land use/Landover, are typically not accounted for in spatial climate interpolation. Slope and aspect at relatively small scales may play a role in determining the local orographic enhancement of precipitation (Daly, 2006).

The Asian Continent and its contrast with the ocean cause the planetary-scale summer monsoon. Over land, surface properties such as vegetation and soil moisture are considered to be important, but their interaction with precipitation and its role in determining rainfall distributions are complicated and remain not well understood (Shang-Ping-xie, 2005).

To date, (geo)statistical estimates and mapping remain the most commonly used operational and efficient approach, but they also emphasize our current incapacity to account for the atmospheric micro-physics relevant to rainfall (Higgins N A, 2005).

Spatial modelling of a climatic variable is an important issue, because many other environmental variables depend on climate (Kurtzman Daniel, 1999; Marquinez, 2003). Rainfall and several other water-related issues, notably those concerned with water supply, non-point source pollution and groundwater quality impairment, remain of great concern locally and globally (Wilson et al, 2000).

Including spatial dependence explicitly in a predictive model can be an efficient way to improve model accuracy with the available data (Miller, 2005).

The motivation of this investigation is to incorporate parameters, which are mostly not taken into consideration when estimating precipitation in complex terrain.

1.6. RESEARCH OBJECTIVE

This research focuses on an analysis of external drift kriging algorithm. We aim at using Elevation, Slope, Aspect, Distance from the coast, Vegetation Cover Factor and Topographical Index in kriging with an external drift as covariates to increase the prediction accuracy and come out with the best approach for precipitation estimation in complex orography.

1.7. RESEARCH QUESTIONS

The questions that we seek to answer in the proposed research are:

1. Does adding parameters or dropping some parameters alter the resultant interpolation?
2. What are the most relevant factors that contribute to type and amount of precipitation in complex terrain?
3. How does the model behave in the two different topographic situations?
4. Which is the best model for estimating precipitation in complex terrain i.e., one that represents the near true amount and in what sense?

1.8. ASSUMPTION AND HYPOTHESES.

1.8.1. ASSUMPTION

There is good correlation existing between the estimation parameters and rainfall received in the study area, that also aids in better precipitation estimation.

There datasets for the study area are good to carry out statistics on, considering that we were working on complex orography which has a problem of accessibility and few rain gauge stations.
1.8.2. HYPOTHESES

Kriging with external drift will estimate precipitation better in complex orography since it takes more secondary variables that are available all over the region of study into consideration.

Elevation is a major influence type and amount of precipitation in complex orography.

1.9. STRUCTURE OF THESIS

This section briefly describes the structure of the report. It contains six chapters:
Chapter 1. INTRODUCTION.
Gives the Introduction and background of the study, objective and research questions.

Chapter 2. LITERATURE SURVEY.
Looks at various other similar studies that have been carried out before, in what areas and using which methods and how they are related to this study.

Chapter 3. STUDY AREA.
Focuses on the area where the study has been carried out. Here datasets from two places in India have been used, i.e.:
   1. Karnataka – South-west India
   2. Uttaranchal – North India.

Chapter 4. MATERIALS AND METHODS.
This chapter introduces us to concepts of Spatial interpolation, Spatial interpolation methods, Geostatistics and Kriging. It Emphasizes on a geostatistical analysis of external drift Kriging algorithm using datasets from the areas of study mentioned above. It also gives more information on the materials used for the study including datasets and software.

Chapter 5. RESULTS AND DISCUSSION.
Gives detailed discussion of the research findings.

Chapter 6. CONCLUSION AND RECOMMENDATIONS.
Formulates the concluding remarks of this thesis, based on the questions and algorithm model used. It also contains recommendations for the future research work in use of Geostatistics.
2. LITERATURE SURVEY

2.1. INTRODUCTION

Over the past two decades, investigations have been made to elaborate complex statistical models coupling topographic descriptors and climatological rainfall, like for example AURHELY (Benichou and Lebreton, 1987) and PRISM (Daly, 1994), an approach also efficient for extreme rainfall (Prudhomme and Reed, 1999). In the same way, efforts have been undertaken by many authors to address better the problem of orographic effect when carrying out geostatistical techniques for rainfall estimation and to use elevation as an auxiliary variable through multi-variate Geostatistics (Hevesi et al., 1992); (Phillips et al., 1992); (MartinezCob, 1996); (Prudhomme and Reed, 1999); (Goovaerts, 2000). These methodologies have equally been used to map other environmental variables, such as precipitation, temperature or soil properties (Odeh et al., 1995; Holdaway, 1996).

Geographically weighted regression (Ninyerola, 2000) (Marquez, 2003) and a knowledge-based system (Daly, 2002) which combines multiple linear regression and distance weighting, have been used to estimate precipitation in areas where there are no stations nearby. Besides the traditional statistical and geospatial climatology commonly used in a geographic information system (GIS) a variety of geostatistical prediction techniques have been applied to climatic data (Phillips et al., 1992)

In this chapter, we look at some studies mentioned and associate them to our context of this study to derive knowledge of techniques that can aid in improving the quality of this research. These are studies that have been carried out before and have similarities to the research we carried out.

2.2. OROGRAPHIC PRECIPITATION

Precipitation is any product of the condensation of atmospheric water vapour that is deposited on the earth's surface. Precipitation that forms aloft is divided into three categories: Liquid Precipitation; Freezing Precipitation and Frozen Precipitation.

It is any form of water that falls from the sky as part of the weather to the ground. This includes snow, rain, sleet, freezing rain, hail, and virga. Precipitation is a major component of the hydrologic cycle, and is responsible for depositing most of the fresh water on the planet.

Precipitation begins forming when relatively warm, moist air rises. As the air cools, water vapour begins to condense on condensation nuclei, forming clouds. After the water droplets grow large enough form precipitation.

This study addresses orographic precipitation which is the kind of precipitation that occurs on the windward side of mountains. It is caused by the rising air motion of a large-scale flow of moist air across the mountain ridge, resulting in adiabatic cooling and condensation. In parts of the world subjected to relatively consistent winds (for example the trade winds), a wetter climate prevails on the windward side of a mountain than on the leeward (downwind) side as moisture is removed by...
orographic precipitation, leaving drier air on the descending (generally warming), leeward side where a rain shadow is observed.

2.3. PREVIOUS STUDIES AND THEIR APPLICATIONS.

Geostatistics developed from the work carried out by the mining engineer D.G. Krige during the early 1950s in the field of ore reserve estimation. These ideas were formalised and extended by G. Matheron, who was the first to suggest the name ‘Geostatistics’ (Matheron, 1962). As the theoretical development of Geostatistics has continued, it has found practical application in many areas. In addition to the mining industry and earth sciences, the techniques have been used in a wide range of environmental disciplines where there is a correlation in space and/or time between measured values. (Higgins N A, 2005)

The previous studies are categorized according to the kind of study undertaken, but it can be noted that some qualify for more than one category e.g. study by (Clark and A, 2006) where he uses regression but studies precipitation.

2.3.1. USE OF GEOSTATISTICAL METHODS

Burrough (2001) in his study: GIS and Geostatistics: Essential partners for spatial analysis made a review where he demonstrated that GIS statistics and Geostatistics have much to give to each other, particularly when GIS is used for environmental analysis. He reveals that Geostatistics benefits from having standard methods of geographical registration, data storage, retrieval and display, while GIS benefits by being able to incorporate proven methods for testing hypotheses and for handling and understanding errors in data and illustrating their effects on the outcomes of models used for environmental management.

He comments on the Spatial context and the use of external information where he says that the suite of geostatistical methods currently available allow the user to incorporate external information that can be used to modify, and possibly improve, the predictions or simulations required.
2.3.2. **USE OF OTHER NON GEO-STATISTICAL METHODS**

Drogue (2002) carried out a study “statistical-topographic model using an omni-directional parameterization of the relief for mapping orographic rainfall” where he presented an objective, analytical and automatic model of quantification and mapping of orographic rainfall. This was applied to the north-eastern part of France but also applicable in other complex terrain.

The model PLUVIA distributes point measurements of monthly, annual and climatological rainfall to regularly spaced grid cells through a multiple regression analysis of rainfall versus morpho-topographic parameters derived from a digital elevation model. The use of an omni-directional parameterization of the topography is induced by a windowing technique which allows better account to be taken of the synoptic scale weather systems generating the different rainfall quantities of interest and the spatial scale of orographic effects.

He found out that it also provides a more physical interpretation of geographical and topographical parameters selected for spatial estimation.

He then compared the advantages and limitations of the model kriging with external drift and extended collocated co-kriging. The study led to a new method and a derived operational software called PLUVIA.

Christopher Daly (2002) came up with an effective approach that uses the wealth of expert knowledge on the spatial patterns of climate and their relationships with geographic features, to help enhance control and parameterize a statistical technique. This was referred to as PRISM (Parameter-elevation Regressions on Independent Slopes Model)

In operation, PRISM performs the following:
1. uses a digital elevation model (DEM) to estimate the "orographic" elevations of precipitation stations;
2. uses the DEM and a windowing technique to group stations onto individual topographic facets;
3. estimates precipitation at a DEM grid cell through a regression of precipitation vs. DEM elevation developed from stations on the cell's topographic facet; and
4. when possible, calculates a prediction interval for the estimate, which is an approximation of the uncertainty involved.

PRISM was then compared to kriging, detrended kriging, and cokriging in the Willamette River Basin, Oregon. In a jack-knife cross-validation exercise, PRISM exhibited lower overall bias and mean absolute error (Daly, et al, 1993; Phillips, et al, 1992). PRISM was also applied to northern Oregon and to the entire western United States. Detrended kriging and cokriging could not be used in these regions because there was no overall relationship between elevation and precipitation.

PRISM's cross-validation bias and absolute error in northern Oregon increased a small to moderate amount compared to those in the Willamette River Basin; errors in the western United States showed little further increase. PRISM has recently been applied to the entire United States in three separate runs (western, central, and eastern) with excellent results, even in regions where orographic processes do not dominate precipitation patterns.
Goovaerts (1998) in his study “Ordinary Cokriging Revisited” conducted a study on ordinary Cokriging where he related differences between cokriging variants and those of models adopted for the means of primary and secondary variables. The prediction performances of cokriging estimators are assessed using an environmental dataset that includes concentrations of five heavy metals at 359 locations. Analysis of re estimation scores at 100 test locations showed that kriging and cokriging performed equally when the primary and secondary variables are sampled at the same locations. When the secondary information is available at the estimated location, one gains little by retaining other distant secondary data in the estimation.

Boer et al (2002) in their study: “Kriging and thin plate splines for mapping climate variables” studied four forms of kriging and three forms of thin plate splines to predict monthly maximum temperature and monthly mean precipitation in Jalisco State of Mexico. Their main aim was to find an optimal way of including elevation data of the area into the interpolation techniques to increase the prediction accuracy of the climate maps. 7 interpolation techniques were worked upon, 5 including and 2 excluding elevation as additional information. Results showed that techniques using elevation as additional information improve the prediction results considerably. From these techniques, trivariate regression-kriging and trivariate thin plate splines performed best. The results of monthly maximum temperature were much clearer than the results of monthly mean precipitation, because the modelling of precipitation is more troublesome due to higher variability in the data and their non-Gaussian character.

Clark M (2006) undertook a study “Quantitative Precipitation Estimation in Complex Terrain” where he sought to come up with a flexible method that would generate ensemble gridded fields of precipitation in complex terrain. The study sought to interpolate/ extrapolate point measurements of daily precipitation totals to a high-resolution modelling grid (2km). The probabilistic precipitation estimation method (i) explicitly accounts for the effects of topography on the spatial organization of precipitation, (ii) accurately describes errors in high-resolution gridded precipitation estimates, and (iii) can be used to generate conditional ensemble grids of daily precipitation totals for multiple years. The method is based on locally weighted regression, (Rajagopalan and Lall 1998; Loader 1999), in which spatial attributes from station locations are used as explanatory variables to predict spatial variability in precipitation. For each time step, regression models are used to estimate the conditional cumulative distribution function (cdf) of precipitation at each grid cell and ensembles are generated by using realizations from correlated random fields to extract values from the gridded precipitation. The ensemble precipitation grids reproduce the climatological precipitation gradients and observed spatial correlation structure. The probabilistic precipitation estimation method performs well. Probabilistic verification shows that the precipitation estimates are reliable, in the sense that there is close agreement between the frequency of occurrence of specific precipitation events in different probability categories and the probability that is estimated from the ensemble. The largest短coming of this method is its failure to assign high probabilities to extreme precipitation events

2.3.3. USE OF COVARIABLES

Diodato (2005) studied the influence of topographic co-variables on the spatial variability of precipitation over small regions of complex terrain. His objective was to evaluate several spatial interpolation techniques for forecasting the spatial distribution of precipitation in complex topographic
regions. He applied ordinary kriging of precipitation, co-kriging using precipitation and elevation and co-kriging using precipitation and a topographic index variable. He compared the results obtained using alternative methods applied to the same data set. Apart from ordinary kriging examination, two auxiliary variables were added for ordinary co-kriging of annual and seasonal precipitation: terrain elevation data and a topographic index. Cross-validation indicated that the ordinary kriging yielded the largest prediction errors. The smallest prediction errors were produced by a multivariate geostatistical method. However, the results favoured the ordinary co-kriging with inclusion of information on the topographic index.

The application of co-kriging was justified in areas where there were nearby stations and where landform was very complex. He concludes that ordinary co-kriging is a very flexible and robust interpolation method because it may take into account several properties (soft and hard data) of the landscape.

Daly (2006) discusses the relationship between scale and spatial climate-forcing factors, and provides background and advice on assessing the suitability of data sets. He reveals that spatial climate patterns are most affected by terrain and water bodies, primarily through the direct effects of elevation, terrain-induced climate transitions, cold air drainage and inversions, and coastal effects. The importance of these factors is generally lowest at scales of 100 km and greater, and becomes greatest at less than 10 km. He mentions that regions having significant terrain features, and also significant coastal effects, rain shadows, or cold air drainage and inversions are best handled by sophisticated systems that are configured and evaluated by experienced climatologists.

2.3.4. KRIGING WITH EXTERNAL DRIFT

Wackernagel, (1994) investigates the application of kriging with external drift to mapping January mean temperatures in Scotland. In his study “mapping temperature using kriging with external drift: theory and an example from Scotland” He analyses the spatial structure of mean temperature using variograms computed in different directions. From these he reveals that January temperature is second-order stationary in the north-south direction. Hence the variogram exists in that direction and is taken to represent the underlying variogram. He models this variogram and uses it in universal kriging to produce point-kriged estimates on a 5-km square grid. These estimates do not adequately show the variation in temperature between stations and so the correlation with elevation is exploited in universal kriging with elevation as external drift. This method gives a kriged estimate for temperature that reproduces the correlation with elevation at the climate stations. Knowledge of the relationship between temperature and topography guided the choice of elevation as an external drift variable in this study, although the use of elevation alone is a first step to exploring the possibilities of using more than one auxiliary variable for kriging or co-kriging temperatures. Temperatures are measured daily at 149 meteorological stations in Scotland, and January values averaged over a period of 30 years (1951-1980) are used. The stations are all located below 400 m altitude and mainly in agricultural areas. A subset of the elevation at points on a 5-km square grid was used for kriging and as the external drift variable. The kriged estimates using elevation as an external drift variable yielded much more detail than universal kriging using only the temperature data. However, a few unrealistic extrapolations occurred in north-west Scotland with a 16-station neighbourhood. The method of integrating the information about elevation into the mapping of temperature by kriging improves the map of January mean temperatures in Scotland, as compared with kriging based on
temperature data alone. However, some exaggerated extrapolation effects are observed for a region in the Highlands, where stations are scarce and located at low altitudes.

Kassteele (2006) in his study: Statistical Air Quality Mapping, explored the use of external drift kriging with the OPS model output in a reduced monitoring network. He made a comparison with universal kriging (UK) by comparing UK and KED. Parameter estimation was carried out by means of restricted maximum likelihood and Bayesian inference. He worked at a method that aimed at performing a detailed uncertainty assessment of mapping NO2 concentrations in the Netherlands at rural and urban scales using uncertain secondary information from an atmospheric dispersion model. In this study the model-based approach of KED was extended by allowing uncertain secondary information. This new interpolation approach was called error-in-variable KED. It showed a successful creation of concentration maps based on uncertain measurements and uncertain dispersion model output. The Bayesian approach for spatial modelling was extremely useful in this context. He concluded that a combination of observations and a deterministic dispersion model by model-based geostatistical interpolation procedure is successful in reducing uncertainties. The combination led to more accurate and precise spatial interpolation results, in particular, outside the sampling area. If applied as an external drift, the dispersion model output provided more detail in spatial maps than universal kriging. Standard deviations for KED are smaller than those for UK. Furthermore, KED also allows handling of biased additional information. This can be beneficial if the pollution sources are missing or unknown. KED accounts for systematic errors by use of regression parameters.

**2.3.5. PRECIPITATION ESTIMATION**

Kravchenko and Zhang, (1996) Estimation of Mean Annual Precipitation in Wyoming Using Geostatistical Analysis. In her study, she utilized kriging with external drift of topographical information to determine precipitation spatial distributions in Wyoming. Relationships between precipitation and elevation, slope, exposure, latitude, and longitude were used to improve the precipitation estimation. The territory of the state was divided into sub regions with relatively high correlation between precipitation and topographical features, geographical coordinates and geostatistical analyses were carried out for the sub regions. Application of the relationships in the external drift kriging resulted in reduction of kriging variance about 30 to 40% in different parts of the studied area with the maximum reduction up to 70%. The method was especially useful in mountain areas sparsely covered with observations. The external drift kriging allowed to utilize geographical coordinates in the estimation. The addition of geographical coordinates resulted in reducing kriging variance about 54% and increasing the correlation coefficient between observed and estimated values from 0.4 to 0.8 for the precipitation estimation. For all of the sub regions multiple regressions with elevation, slope, exposure as well as latitude and longitude gave the highest correlation coefficients, cross-validation results using the kriging with external drift (KED) were compared with those using the ordinary kriging (OK). Using additional topographical information significantly improved precipitation estimation in all sub regions. Its application in the external drift kriging resulted in much higher estimation precision than the ordinary kriging.

Valesco(2004) worked at a study: “Merging radar and rain gauge data to estimate rainfall fields: an improved geostatistical approach using non-parametric spatial models” They aimed at providing an
improved geostatistical approach able to be applied operationally. They then analysed three geostatistical alternatives to test its efficiency in terms of time computing. The selected approaches were Ordinary Kriging, Kriging with External Drift, and Collocated Cokriging. The study worked at avoiding the typical prior selection of a theoretical correlogram or semivariogram model, and it allows the introduction of spatial information provided by radar and rain gauge measurements. The main idea for this was to transform the experimental (cross-) correlation into density spectrum maps using FFT.

The study was carried out in Cataluña where he tests the performance of these methodologies. Leave-one-out cross validation was used to evaluate, in a statistical way, the quantitative performance of each kriging method. The results showed that rainfall field estimated by KED give the best results in a statistical and qualitative way.

Hosseini (2006) carried out a survey on “Application of Geostatistical Methods for Estimation of Rainfall in Arid and Semi-arid Regions in South West Of Iran”. In their study thin plate smoothing splines, with and without co-variable, weighted moving average, and kriging (ordinary, cokriging and log-kriging) were used to estimate monthly and annual rainfall. They used data of 167 climatic stations in south west of Iran with 22 years records. They divided their study area into sub catchments based on elevation. They had Thin Plate Smoothing Spline with elevation as co-variable the most precise method to estimate annual rainfall.

Atkinson P,(1998) worked on mapping Precipitation in Switzerland with Ordinary and Indicator Kriging. In his study he used ordinary kriging, and indicator kriging to address the problem of estimating values of precipitation at locations from which measurements have not been taken. During the study several problems were raised including: (i) log normality of the data, (ii) non-stationarity of the data and (iii) anisotropy of the spatial continuity. The aim was to compare a variety of different approaches to estimation. Indicator Kriging (informed using directional indicator variogram models) was selected because it is a means to account for log normality and was the method that was unlikely to be used widely within the competition. Accuracy of estimates made using Indicator Kriging were compared with Ordinary Kriging estimates. It was observed that the Ordinary Kriging algorithm, as implemented here, provided more accurate estimates than Indicator Kriging. This was considered to be due, at least in part, to the method used for tail extrapolation and also the small number of data used in estimation (100 data locations). OK was recommended over IK in this instance as OK provided more accurate estimates and was also easier to implement.

Johansson B, (2003) in his study: The influence of wind and topography on precipitation distribution in Sweden: statistical analysis and modelling comparative analysis with other methods. The objective of this paper was to investigate whether the relationship between precipitation, airflow and topography could be described by statistical relationships using data easily available in an operational environment. The purpose was to establish a statistical model to describe basic patterns of precipitation distribution. This model, if successful, can be used to account for the topographical influence in precipitation interpolation schemes. A statistical analysis was carried out to define the most relevant variables, and, based on that analysis, a regression model was established through stepwise regression. Some 15 years of precipitation data from 370 stations in Sweden were used for the analysis. Precipitation data for each station were divided into 48 classes representing different wind directions and wind speeds. Among the variables selected, the single most important one was found to be the location of a station with
respect to a mountain range. On the upwind side, precipitation increased with increasing wind speed. On the leeward side there was less variation in precipitation, and wind speed did not affect the precipitation amounts to the same degree. For ascending air, slope multiplied by wind speed was another important factor. Precipitation amounts were found to be higher for events with easterly and southerly winds than for events with westerly and northerly winds. Possibly, this is an indication of different air mass characteristics. Humidity and stability are known to affect the generation of precipitation.

2.3.6. COMPARATIVE ANALYSIS WITH OTHER METHODS.

Goovaerts P. in his study “Performance comparison of geostatistical algorithms for incorporating elevation into the mapping of precipitation.” Worked at studying three geostatistical algorithms for incorporating a digital elevation model into the spatial prediction of rainfall. The algorithms he used were: simple kriging with varying local means, kriging with an external drift, and collocated cokriging. The techniques were illustrated using annual and monthly rainfall observations at 36 climatic stations in a 5,000 km² region of Portugal. He used cross validation to compare the prediction performances of the three geostatistical interpolation algorithms with the straightforward linear regression of rainfall against elevation and three univariate techniques: Thiessen polygon, inverse square distance, and ordinary kriging.

Larger prediction errors were obtained for inverse square distance and Thiessen polygon which that don’t take into consideration both elevation and rainfall records at surrounding stations. The three multivariate geostatistical algorithms outperformed other interpolators, in particular linear regression, which stresses the importance of accounting for spatially dependent rainfall observations in addition to the collocated elevation. Last, ordinary kriging yielded more accurate predictions than linear regression when the correlation between rainfall and elevation is moderate.

(Kurtzman Daniel, 1999) undertook a study “Mapping of temperature variables in Israel: a comparison of different interpolation methods” where he compared the performance of 2 local interpolation methods, Spline and Inverse Distance Weighting (IDW), with the performance of multiple regression models. These interpolation methods were applied to 4 temperature variables: mean daily temperature of the coldest month, warmest month, lowest mean monthly minimum temperature, and the highest mean monthly maximum temperature. Spline and IDW models with a range of parameter settings were applied to elevation de-trended temperature data. The multiple regression models were based on geographic longitude, latitude and elevation and included terms of first and second order. Accuracy was assessed by a one-left-out cross validation test. Mean daily temperature variables proved more predictable than mean monthly extreme temperature variables. Mean daily temperature variables were predicted more accurately by using a regression model, whereas mean monthly extreme temperature variables were somewhat better predicted by a local interpolation method. The Spline interpolator predicted more accurately than IDW for the 2 summer temperature variables, while IDW performed better for the winter temperature variables. Combining multiple regression and local interpolation methods improved prediction accuracy by about 5% for the extreme temperature variables. He found out that in some instances simple overall regression models can be as effective as sophisticated local interpolation methods, especially when dealing with mean climatic fields.

Bourennane. (2000) Comparison of kriging with external drift and simple linear regression for predicting soil horizon thickness with different sample densities. In their study they examined two
mapping method’s sensitivity to the sampling density of the variable of interest, which is the thickness of a silty-clay–loam. The two methods were simple linear regression and universal kriging with external drift. Their main purposes was to compare the prediction performance of these methods.

Slope was used for thickness of silk loam clay prediction by Simple Linear Regression and by Universal Kriging with External Drift, where slope gradient used as external drift.

The results showed that universal kriging with external drift was more accurate than the Simple linear regression. The improvement of the accuracy of the prediction from SLR to UKEXD was about 38%.

Karamouz, (2006) used multivariate geostatistical algorithms, kriging with varying local means, and kriging with an external drift for estimating spatial variations of precipitation in the western part of Iran. He applied the techniques using annual and monthly precipitation data measured at 54 climatic stations from 1967 to 2000.

He correlated precipitation and the total runoff, leaving the drainage basins to evaluate the accuracy of the estimates. Cross validation was used to compare the accuracy of the techniques. The results showed that kriging estimation variance depends on the framework of the gauging network and it can be used for the optimal design of the rain gauge network. Based on his results, new locations for establishing rain gauges and also recommendation for reducing the number of rain gauges in the region were made.

Goovaerts (1999) incorporated a digital elevation model into the mapping of annual and monthly erosivity values in the Algarve region Portugal. He carried out linear regression of erosivity against elevation, and three geostatistical algorithms: simple kriging with varying local means (SKlm), kriging with an external drift KED, and collocated cokriging.

Cross validation indicated that linear regression, which ignores the information provided by neighbouring climatic stations, yielded the largest prediction errors in most situations. Smaller prediction errors were produced by SKlm and KED that both use elevation to inform on the local mean of erosivity; kriging with an external drift allowed to assess relation between the two variables within each kriging search neighbourhood instead of globally as for simple kriging with varying local means.

The best results were obtained using cokriging that incorporates the secondary information directly into the computation of the erosivity estimate.

The case study showed that the trend coefficients varied substantially across the study area and that elevation tends to influence strongly the KED estimate, especially when the estimated slope of the local trend model is large.

He concluded that digital elevation models are potentially valuable sources of information for the mapping of erosivity values, in particular when climatic stations are sparse. However, there are many different ways to incorporate such exhaustive secondary information, and cross validation results have shown that prediction performances can vary greatly among algorithms.

Yeonsang, (2004) in their study “Inter-comparison of spatial estimation schemes for precipitation and temperature” They undertook a study where they compared the performance of several statistical methods for spatial estimation in two climatologically and hydrologically different basins.

The seven methods assessed were: (1) Simple Average; (2) Inverse Distance Weight Scheme (IDW); (3) Ordinary kriging; (4) Multiple Linear Regression (MLR); (5) PRISM (Parameter-elevation Regressions on Independent Slopes Model) based interpolation; (6) Climatological MLR (CMLR); and (7) Locally Weighted Polynomial Regression (LWP). Regression based methods that used
elevation information showed better performance, in particular, the nonparametric method LWP. LWP is data driven with minimal assumptions and provides an attractive alternative to MLR in situations with high degrees of nonlinearity. For daily time scale, we propose a two step process in which, the precipitation occurrence is first generated via a logistic regression model, and the amount is then estimated using the interpolation schemes. This process generated the precipitation occurrence effectively.

Tsanis, (2004) in her study “Ranking spatial interpolation techniques using a GIS-based dss” developed a GIS based Decision Support System (DSS) to select the appropriate interpolation technique used in studying rainfall spatial variability. The DSS used the Arc-View GIS platform by incorporating its spatial analysis capabilities, the programming language “AVENUE”, and simple statistical methods. A test case from the country of Switzerland is used to demonstrate the applicability of the system. This should aid in better input to hydrological models. This work established an approach by using GIS and historical data to locate the best technique. She recommended data sets with different sizes. Based on the one data set available for this study, it was clear that the Kriging exponential and Kriging Universal models showed consistent performance and provided reliable estimates regardless of the number of gages or the cell size used in the interpolation.

Bourennane and King, (2003) in his study: Using multiple external drifts to estimate a soil variable used kriging with slope gradient and electrical resistivity measurements used as external drifts in universal kriging to predict the depth of a limestone bedrock upper boundary (LUB) in southwest of the Paris. Two indices, (i) the mean error (ME) and (ii) the root mean square error (RMSE), as well as residuals analysis, were computed from the validation sample (observed data) and predicted values. On the 50 test data, the results showed that kriging using two external drifts proved to be less unbiased and more precise compared to kriging using only one external drift variable in predicting the target variable.

Buytaert, (2006) studied the “Spatial and temporal rainfall variability in mountainous areas” focusing on spatial and temporal rainfall patterns were studied. They conducted their study in the south Ecuadorian Andes where they examined rainfall data from 14 rain gauges. A clear intraday pattern was distinguished. Seasonal variation, on the other hand, was found to be low, with a difference of about 100 mm between the driest and the wettest month on an average of about 100 mm month1, and only 20% dry days throughout the year. Rain gauges at a mutual distance of less than 4000 m were strongly correlated, with a Pearson correlation coefficient higher than 0.8. However, even within this perimeter, spatial variability in average rainfall is very high. Significant correlations were found between average daily rainfall and geographical location, as well as the topographical parameters slope, aspect, topography. Spatial interpolation with Thiessen gave good results. Kriging gave better results than Thiessen, and the accuracy of both methods improves when external trends are incorporated. Kriging was more accurate for interpolation of average daily rainfall than Thiessen. The incorporation of external trends improves the accuracy in both methods, but not the error variance. These results suggest that a more detailed assessment of the relation between the topography and the spatial rainfall distribution may be able to improve interpolation results.

Musio(2004) studied different geostatistical methods and uses them to interpolate the spatial distribution of the foliar magnesium content of Silver fir and Norway spruce in the Black Forest. In her study “Predicting magnesium concentration in needles of Silver fir and Norway spruce” she aimed at
identifying the best prediction method that can be useful in the future for cause–effect studies and environmental modelling. At the same time, causal relationships between the response variable and the predictors were investigated. Geostatistical methods with lowest prediction errors which simultaneously provide the highest explanation value were identified. The performance of different methods was measured using cross-validations techniques.

She establish a model for predicting magnesium contents in the needles (a) by using a trend function of soil condition, other site and tree characteristics, and/or (b) by exploiting spatial autocorrelation via geostatistical methods. She compared the prediction performances of different methods currently used in spatial studies, such as a model with independent errors, ordinary kriging, cokriging and kriging with external drift.

Identifying a best prediction method can be useful in the future to predict the magnesium in the needles at unsampled locations. In comparison to OK, the model with independent errors, KED and CK allowed incorporation of explanatory information such as other nutrients, tree characteristics, soil characteristics or locational variables.

KED performed best in terms of prediction; it halved the mean squared prediction error in comparison to OK and the model with independent errors. The prediction error of CK was smaller than for OK. But in comparison to KED, CK did not have the benefit of allowing for a trend function with categorical variables. The model with independent errors and OK yielded similar mean squared prediction errors. The great advantage of OK compared to the other methods was that explanatory information, which might be expensive to measure is not required. The advantage of OK and CK was that fine grid map predictions can be produced for a particular variable without need of further information.

Although she set herself the goal of identifying the best prediction method, she also identified a procedure which allows the investigation of causal relationships at the same time.
3. **STUDY AREA**

To aid the research, two states in India were chosen for the study.
1. Karnataka, which is a state in south West India. It is part of the Western Ghats. It comprises the mountain range that runs along the western coast of India.
2. Uttaranchal state which is in the Northern part of India and part of the Himalayas. These were chosen on the basis of the complexity of their terrain, it was felt that they would be good areas to conduct study of the external drift kriging algorithm in view of its application to complex orography.

### 3.1.1. CLIMATE

Most of the year the monsoons, or seasonal winds, blow from the northeast. In the summer months, they begin to blow from the southwest, absorbing moisture as they cross the Indian Ocean. This warm, moist air creates heavy rains as it rises over the Indian Peninsula and is finally forced up the slopes of the Himalayas. The rains start in early June on a strip of coast lying between the Arabian Sea and the foot of the Western Ghats. A second block of the monsoon starts from the Bay of Bengal in the northeast and gradually extends up the Gangetic Plain. The average annual rainfall for India as a whole is 1,250 mm. The heaviest rainfall occurs along the Western Ghats, often more than 3,175 mm, and on the slopes of the eastern Himalayas and the Khāsi Hills, where the town of Cherrapunji receives 10,900 mm annually. The entire northeast region averages more than 2,000 mm annually. Rain and snow fall in abundance on the entire Himalayan range.

### 3.1.2. KARNATAKA

#### 3.1.2.1. LOCATION AND BOUNDARIES

Karnataka is located in the southern part of India. The it is situated approximately between the latitudes 11.5° and 18.5° North and the longitudes 74° and 78.5° East. It is situated in the Deccan Plateau and is bordered by the Arabian Sea to the west. It is situated at the angle where the Western Ghats and Eastern Ghats of South India converge into the Nilgiri Hills. It extends to about 750 km from North to South and about 400 km from East to West, and covers an area of about 1,91,791 sq.km.
ANALYSIS OF EXTERNAL DRIFT KRIGING ALGORITHM WITH APPLICATION TO PRECIPITATION ESTIMATION IN COMPLEX OROGRAPHY.

3.1.2.2. TOPOGRAPHY

Karnataka has representatives of all types of variations in topography - high mountains, plateaus, residual hills and coastal plains. The State is enclosed by chains of mountains to its west, east and south. It consists mainly of plateau which has higher elevation of 600 to 900 metres above mean sea level. The entire landscape is undulating, broken up by mountains and deep ravines. Plain land of elevation less than 300 metres above mean sea level is to be found only in the narrow coastal belt, facing the Arabian Sea. There are a few high peaks both in Western and Eastern Ghat systems with altitudes more than 1,500 metres. A series of cross-sections drawn from west to east across the Western Ghat generally exhibit, a narrow coastal plain followed to the east by small and short plateaus at different altitudes, then suddenly rising up to great heights. Then follows the gentle east and east-north-west sloping plateau. Among the tallest peaks of Karnataka are the Mullayyana Giri (1,925 m), Bababudangiri (Chandradrona Parvata 1,894 m) and the Kudremukh (1,895 m) and the Pushpagiri (1,908 m) in Kodagu Dt. There are a dozen peaks which rise above the height of 1,500 metres.

3.1.2.3. CLIMATE

Karnataka experiences three main types of climate. For meteorological purposes, the state has been divided into three sub-divisions namely

Tropical Monsoon climate covers the entire coastal belt and adjoining areas. The climate in this region is hot with excessive rainfall during the monsoon season i.e., June to September. The Southern half of the State experiences hot, seasonally dry tropical savannah climate while most of the northern half experiences hot, semi-arid, tropical steppe type of climate.
The winter season from January to February is followed by summer season from March to May. The period from October to December forms the post-monsoon season. The period from October to March has a few spells of rain associated with north-east monsoon which affects the south-eastern parts of the State during October to December.

The months April and May are hot. June experiences high humidity and temperature. The next three months (July, August and September) have reduced day temperature although the humidity continue to be very high.

### 3.1.2.4. TEMPERATURE

Day and night temperatures are more or less uniform over the State, except at the coastal region and high elevated plateau. They decrease south-westwards over the State due to higher elevation and attain lower values at high level stations. April and May are the hottest months. In May, mean maximum temperature shoots up to 40 deg. C over the north-eastern corner of the State, decreasing south-westwards towards the Western Ghat region and the Coastal belt. December and January are the coldest months.

### 3.1.2.5. RAINFALL

Rainfall received is 50cm to over 400 cm. In the districts of Bijapur, Raichur, Bellary and southern half of Gulbarga, the rainfall is lowest varying from 50 to 60 cm. Agumbe in the Sahyadris receives the second heaviest annual rainfall (760cm) in India. It increases significantly in the western part of the State and reaches its maximum over the coastal belt.

The south-west monsoon is the principal rainy season during which the State receives 80% of its rainfall. Rainfall in the winter season (January to February) is less than one per cent of the annual total, in the hot weather season (March to May) about 7% and in the post-monsoon season about 12%. South-West monsoon normally sets in over the extreme southern parts of the State by the beginning of June and covers the entire State by mid June. The rainy months July and August account individually to about 30% and 18% of annual rainfall. There are about 26 rainy days (with daily rainfall of at least 2.5 mm) in the south-west monsoon begins from the northern parts of the State around 2nd week of October and by the 15th October monsoon withdraws from the entire State.

The retreating monsoon current i.e. the north-east monsoon (October to December) effects the eastern parts of South Interior Karnataka and accounts for about 30% of rainfall in this region. Out of the 14 heavy rainfall stations in India, with annual rainfall of more than 500 cm. four stations are situated in Karnataka. They are Agumbe in Tirthahalli taluk of Shimoga district (annual rainfall -828 cm) and Bhagamandala (603 cm), Pullingoth (594 cm) and Makut (505 cm) in Kodagu district.

### 3.1.2.6. FOREST

Karnataka State has a geographical area of 191,791 sq. km of which 38.724 sq.km (20 per cent) is under the control of the Forest Department.
3.2. UTTARANCHAL

3.2.1. LOCATION AND BOUDARIES

Uttaranchal is the 27th State to be formed in India. It has 13 districts. It borders China in the north and Nepal to the east, while its neighbour states are Himachal Pradesh to the west and Uttar Pradesh. It spreads over an area of 55,845 square kilometres and the Himalayas is an integral part of the mountain ecosystem. It falls between Latitude 28°43' N to 31°27' N and Longitude 77°34' E to 81°02' E.

3.2.2. CLIMATE

The Climate of the State is generally temperate. It varies greatly from tropical to severe cold depending on the altitude of the area. Uttaranchal being hilly, temperature variations due to difference in elevation are considerable. In the hilly regions, the summer is pleasant, but in the Doon, the heat is often intense, although not to such degree as in the plains of the adjoining district. The temperature drops below freezing point (0°C) not only at high altitude but even at places like Dehradun during the winters, when the higher peaks are also under snow. The area receives an average annual rainfall of 2073.3 mm. Most of the annual rainfall in the district is received during the months from June to September, July and August being the months that receive the most amount of rainfall.

3.2.3. RAINFALL

Northwest India and the Himalayas are particularly prone to vagaries of severe weather, which claims casualties every year. This region is influenced by extra tropical disturbances that propagate from the west to India during winter, bringing rainfall and chilly weather. These systems severely influence life in the Himalayas by inducing widespread rainfall and, at times, very heavy snowfall associated with squall winds, hail, and severe cold waves. Gale winds and heavy rain/snowfall result in avalanches and landslides (Das et al., 2003).
3.2.4. PHYSIOGRAPHY

Uttaranchal is comprised of two regions, the western half known as Garhwal and eastern half as Kumaon. Most of the northern parts of the Uttarakhand state are part of Greater Himalaya ranges, covered by the high Himalayan peaks and glaciers, while the lower foothills were densely forested but got denuded.

Two of India's mightiest rivers, the Ganga and the Yamuna take birth in the glaciers of Uttarakhand, and are fed by myriad lakes, glacial melts and streams in the region.
4. MATERIALS AND METHODS

4.1. MATERIALS

4.1.1. DATA DESCRIPTION AND PROCESSING

The data used in this study encompasses five recorded variables (Elevation, Slope aspect, Distance from the coast and vegetation cover factor) and one derived variable, Topographic index.

4.1.1.1. RAINFALL DATA ACQUIRED

Daily and monthly rainfall datasets acquired from Drought Monitoring Cell (DMC), Karnataka for the period 1970-2004. Of this dataset, this research utilised data for the year 2000, and the months April, July and December were used. These were chosen because they lie in the different climatic seasons of India which are:

- Cool, dry winter from October to March.
- Hot, dry summer from April to June.
- Southwest monsoon season of warm, torrential rains from mid-June to September.

Uttaranchal Rainfall datasets acquired from Indian Meteorological Department (IMD) Climatology for the period 1900-1950 as an average. The rainfall was measured at 42 meteorological stations, and the values were averaged over the 50 year period.

4.1.1.2. SATELLITE DATA ACQUIRED

ELEVATION

90m DEM. Void-filled seamless Shuttle Radar Topographic Mission (SRTM) data V1, 2004, International Centre for Tropical Agriculture (CIAT), available from the CGIAR-CSI SRTM 90m Database: http://srtm.csi.cgiar.org

This was downloaded in 5 by 5 degree tiles. It was then imported into Erdas imagine 8.7, mosaicked then clipped.

SLOPE

Slope was derived from elevation data above using Arc-GIS. This was degraded to different resolutions, i.e. 2km resolution and 10km resolution.

ASPECT

Aspect was derived from elevation data above using Arc-GIS. This was degraded to different resolutions, i.e. 2km resolution and 10km resolution.

To avoid problems modeling with a circular predictor variable, a transformation used by Roberts and Cooper (1989) was applied to aspect. The transformation formula is as shown below and a table showing the transformation is attached in the appendix.
\[ \text{Aspect} = 0.5 \left( \cos \left( \frac{\text{Asp}}{180} \right) \right) \]

DISTANCE FROM THE COAST
Data acquired from Drought Monitoring cell Karnataka and also used for previous studies. This was also degraded to 2 and 10 kilometres respectively.

VEGETATION CONTINUOUS FIELD
This was acquired from MODIS data at 500meters resolution. It was degraded to 2km resolution. It varies from 1-100 and was converted from this scale to 1-10 to be used in the study. The VCF measures the proportion of natural green vegetation coverage, so that the values vary from zero (no vegetation) to one (abundant vegetation).

TOPOGRAPHIC INDEX
We have also used a new empirical index, called the topographic index (TI) integrating VCF and E variables as a multiplicative function: \( \text{TI} = \text{VCF} \sqrt{\text{E}} \). This was “adopted” from (Diodato, 2005)

4.2. SOFTWARE USED.
In this study, different software was used, depending on the need. The main software used for the research was R statistics and Arc-GIS. Refer to appendix 6 for details on this.
4.3. METHODS

4.3.1. INTERPOLATION METHODS
We worked at coming up with the best method for precipitation estimation in complex orography. The study commenced with study of Ordinary kriging and looking at how good it is at estimating precipitation in complex orography. The dataset used was used for all the methods in the study.

4.4. SPATIAL INTERPOLATION

Spatial interpolation refers to the procedure of estimating the value of properties at unsampled sites within the area covered by existing observations. In almost all cases the property must be interval or ratio scaled. Idea behind spatial interpolation is the observation that points close together in space are more likely to have similar values than points far apart (Tobler's Law of Geography). It’s used as an aid in the spatial decision making process both in physical and human geography and in related disciplines such as mineral prospecting and air quality research, mineral prospecting and precipitation estimation.

4.5. CLASSIFICATION OF INTERPOLATION PROCEDURES

There are several different ways to classify spatial interpolation procedures:

4.5.1. POINT/AREAL INTERPOLATION.
Point interpolation is used for data which can be collected at point locations e.g. weather station readings, spot heights, oil well readings, porosity measurements. Point to point interpolation is the most frequently performed type of spatial interpolation in GIS. Lines to points e.g. contours to elevation grids.

4.5.2. AREAL INTERPOLATION.
Given a set of data mapped on one set of source zones one can determine the values of the data for a different set of target zones.

4.5.3. GLOBAL/LOCAL INTERPOLATORS.
Global interpolators determine a single function which is mapped across the whole region. A change in one input value affects the entire map. Local interpolators apply an algorithm repeatedly to a small portion of the total set of points. A change in an input value only affects the result within the window.

4.5.4. EXACT/APPROXIMATE INTERPOLATORS.
Exact interpolators honor the data points upon which the interpolation is based on and the surface passes through all points whose values are known. This is an important feature in areas like the oil industry. Proximal interpolators, B-splines, and Kriging methods all honor the given data points. Approximate interpolators are used when there is some uncertainty about the given surface values.
4.5.5. **STOCHASTIC / DETERMINISTIC INTERPOLATORS.**

**Stochastic methods** are ones that incorporate randomness. The interpolated surface is conceptualized as one of many that might have been observed, all of which could have produced the known data points. These include trend surface analysis, Fourier analysis, and Kriging procedures such as trend surface analysis allow the statistical significance of the surface and uncertainty of the predicted values to be calculated.

**Deterministic methods**

These don’t use probability theory.

4.5.6. **GRADUAL/ ABRUPT INTERPOLATORS.**

A typical example of a gradual interpolator is the distance weighted moving average which usually produces an interpolated surface with gradual changes. If the number of points used in the moving average is reduced, there would be abrupt changes in the surface.

It may be necessary to include barriers in the interpolation process: semi-permeable, e.g. weather fronts will produce quickly changing but continuous values. Impermeable barriers, e.g. geologic faults will produce abrupt changes.

4.5.7. **COVARIATES.**

Sometimes interpolation can be aided by some additional variable like elevation which helps in interpolating ground temperature or precipitation.

It is well known that, in general, rainfall increases with elevation in the temperate domain with non-uniform gradients (Barry, 1992), and that mountain ranges trigger contrasts between the windward side and the leeward side. However, these 'laws' remain difficult to model and must take into account: the scale of the topography-rainfall interactions, the location and the exposure characteristics of the stations in mountainous basins (Frei and Schar, 1998) and the climate of the study area (Singh and Kumar, 1997).

4.6. **VARIOGRAMS.**

Geostatistics is based on an observation or assumption of spatial correlation; that is, measurements taken at locations that are close together are likely to have more similar outcomes than are those taken further apart. Semi-variograms (often simply referred to as ‘variograms’ in the geostatistical literature, a convention that will be followed in this report) are the principal tool used in geostatistics to quantify this change in correlation.

With increasing distance, the variogram may be defined as half the variance of the increment in the random function, as shown below.

\[ \gamma(h) = \frac{1}{2} \mathbb{E}(y(x) - y(x + h))^2 \]
Variogram analysis consists of the experimental variogram calculated from the data and the variogram model fitted to the data. The experimental variogram is calculated by averaging one half the difference squared of the z-values over all pairs of observations with the specified separation distance and direction. It is plotted as a two-dimensional graph. (Higgins N A, 2005)

From the semi-variogram, various properties of the data are determined:

- The sill, the range, the nugget ($C_0$), the sill/nugget ratio, and the ratio of the sum of square of deviance to the total sum of squares (SSD/SST). The higher the sill value, the higher the prediction variances.
- The nugget is the intercept of the semi-variogram with the vertical axis. It is the non-spatial variability of the variable and is determined when $(h)$ approaches zero. The nugget effect can be caused by variability at very short distances for which no pairs of observations are available, sampling inaccuracy, or inaccuracy in the instruments used for measurement. Observations separated by a distance larger than the range are considered as spatially independent observations.

To obtain an indication of the part of the semi-variogram that shows spatial dependence, the sill-nugget ratio can be determined. If this ratio is close to one, then most of the variability is non-spatial. Normally a “variogram” model is fitted through the empirical semi-variogram values for the distance classes or lag classes. The variogram properties; the sill, range and nugget can provide insights on which model will fit the best (Cressie 1993; Burrough and McDonnell 1998). The most common models are the linear model, the spherical model, the exponential model, and the Gaussian model.

When there is a clear range and sill, a model known as the spherical model often fits the variogram well.

The most common models are as shown in the figure on the next page.
In addition to the problem of stationarity, directionality is likely to be important. Thus, the spatial correlation of a set of measurements is often found to vary with direction, that is, it is anisotropic. Typically, this anisotropy is characterised by an ellipse, as shown in Figure 4.3 above, whose major axis $Y'$ is along the direction of greatest spatial continuity and whose minor axis $X'$ is perpendicular to this.

The direction of the anisotropy is given by the angle $\phi$ (or equivalently $(\phi+180^\circ)$) of the principal axis from North. Deposition from a plume is generally an anisotropic phenomenon, with factors such as the constancy of the wind direction and the terrain influencing the shape of the ellipse.

In practice, it can be difficult to establish anisotropy from the data alone, particularly when there are few measurements. In such cases, supporting information such as wind direction can be used to predict the likely direction of maximum continuity. The anisotropy is expressed quantitatively as the range in the direction of minimum continuity divided by the range in the direction of maximum continuity, where the range is defined as the maximum separation for which values may be considered correlated. Data, which are isotropic, will therefore have an anisotropy parameter equal to one, and a highly anisotropic data set will have a parameter close to zero.
Cross-validation is a convenient tool to check the possibility of the utilisation of any interpolation method or to select the method dependent parameters basing only on the available samples. The "leave-one-out" strategy is the basis of cross-validation: the value of known datum is estimated with the help of all other data where a single location is selected from a dataset and the data at the remaining locations are used to calculate a kriging estimate at that point. The cross-validation error, defined as the difference between the kriging estimate and the measured value, is then calculated for that location.

Then this procedure is repeated for each sample from the data set. The difference between the estimated and known values (cross-validation residuals) gives the "map of errors" and scatter plots of estimated data from the corresponding measured ones. The residuals can be investigated using all possible statistical tools. It is expected for residuals to have the median, the spread and the skewness close to 0 in any region. The mean cross-validation error that results gives an indication of how good the model is. Cross-validation is convenient and powerful tool for comparing different methods and to select the best one.

Once the process is complete, overall error statistics, such as mean absolute error (MAE), bias, and others are calculated (e.g. Willmott et al., 1985a; Legates and McCabe, 1999).

The difference between the estimated value \( Z \) and the corresponding measured value \( Z_I \) is the experimental error

\[
\gamma_i = Z(s) - Z_I(s)
\]

Thus, repeating this estimation for the experimental data size \( n = 51 \), the cross-validation statistics of

Mean error (ME)

\[
ME = \frac{1}{n} \sum_{i=1}^{n} \gamma_i
\]

Root mean Squared error (RMSE)

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \gamma_i^2}
\]
ANALYSIS OF EXTERNAL DRIFT KRIGING ALGORITHM WITH APPLICATION TO PRECIPITATION ESTIMATION IN COMPLEX OROGRAPHY.

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2}
\]

can be calculated, where \( \varepsilon(s_i) \) is the prediction standard error for location \( S_i \).

M.E should be close to zero. RMSE and ASE are indices that represent the goodness of prediction, and RMSSE compares the error variance with a theoretical variance, such as kriging variance. Therefore, it should be close to unity. (Diodato, 2005)

The disadvantage to cross-validation error estimation is that:

- No error information is provided for places where there are no stations.
- The single deletion jack-knife method favours interpolation model parameterizations that heavily smooth the results and reduce local detail, so that deletion of one station is relatively unimportant to the stability of the estimate.
- A related problem is the tendency for jack-knife cross-validation to underestimate errors when stations are located in pairs or clusters; this type of station configuration increases the likelihood that a nearby station will be present to produce a good estimate for one that has been omitted from the data set during the jack-knife deletion process.

Other deletion schemes, such as withholding a stratified sample of data points from the analysis, may sometimes be useful in detecting certain weaknesses in the interpolation. For example, withholding high-elevation stations may help determine how well the system can extrapolate beyond the elevation range of the data. Unlike internally generated model error estimates, cross-validation errors can be compared among interpolation techniques. However, the comparison is valid only when all of the parameters of the interpolation – the domain, input data, grid resolution, etc. – are identical (Phillips et al., 1992) (Daly, 2006)

4.8. KRIGING

Kriging is the generic name for a set of techniques that may be used to estimate the value of a variable at locations where it has not been measured. In its most basic form, a kriging estimate is a weighted linear sum of the measured data. The weights are derived from the variogram or covariance model chosen to represent the phenomenon of interest. Kriging therefore provides an estimate that is more sophisticated than simpler techniques, such as nearest neighbour or inverse distance weighting, because it accounts for the spatial structure of the phenomenon in a way that these techniques cannot. Several variations on the basic kriging algorithm exist which may be appropriate in different situations.

4.8.1. SIMPLE KRIGING.

This, as the most basic form of kriging, assumes that the measured values are realisations of a stationary random function with a constant mean \( m \). The mean must be known but a variant of the procedure will provide an estimate of the mean over the area (which in turn will reduce to the arithmetic mean of the measurements if they are not spatially correlated). The simple kriging estimator for a stationary random function \( Z(x) \) is given by.
ANALYSIS OF EXTERNAL DRIFT KRIGING ALGORITHM WITH APPLICATION TO PRECIPITATION ESTIMATION IN COMPLEX OROGRAPHY.

The simple kriging estimate is calculated for location \( x \) using the measurements at the \( n \) locations \( X \) and their mean \( Z_{sk} \), and the kriging weights \( w_i(x) \).

In most practical situations, a single mean for an entire area is unlikely to provide an adequate approximation and the technique of ordinary kriging is to be preferred.

### 4.8.2. ORDINARY KRIGING

This filters the mean from the simple kriging estimator by imposing the condition that the kriging weights \( w_i(x) \) sum to 1. This results in the bracketed term in the equation above disappearing and creates an estimator that can be used in the more usual practical situation where there are unknown local means, each appropriate to the current local search area. It may be described by the acronym BLUE – Best Linear Unbiased Estimator. It is linear since the estimate is formed from a weighted linear sum of the measured values (within some search radius), and unbiased because it aims to make the average difference between estimated and true values equal to zero, that is, there are no systematic errors which would result in estimates being consistently above or below the true value. In addition to minimising the average error, ordinary kriging also aims to minimise the individual errors for each estimate. It does this by attempting to minimise the variance of the estimation errors, and in this sense, it is the ‘best’ estimate.

The ordinary kriging estimator for the stationary random function \( Z(x) \) is given by the first equation below which is subject to the constraint of the second equation below.

\[
Z_{ok}(x) = \sum_{x \in I} W_i(x)Z(x) \left[ 1 / \sum_{x \in I} W_i(x) \right] m
\]

The optimum size of the local search areas for ordinary kriging is generally dependent on the particular data.

Weights change according to the spatial arrangement of the samples. The linear combination of weights is as follows:

\[
Y_i = \sum_{i \in I} \lambda_i Y_i
\]

where:
- \( Y_i \): estimate of kriging
- \( y_i \): the variables evaluated in the observation locations
- \( \lambda_i \): the kriging weights

The semi-variogram is formulated as follows:
ANALYSIS OF EXTERNAL DRIFT KRIGING ALGORITHM WITH APPLICATION TO PRECIPITATION ESTIMATION IN COMPLEX OROGRAPHY.

\[
\tau(h) = \frac{1}{2} E(y_{(x)} - y_{(x+h)})^2
\]

where:
- \(\tau(h)\): semi-variogram, dependent on lag or distance \(h\).
- \((x, x+h)\): pair of points with distance vector \(h\)
- \(Y(x)-y(x+h)\): difference of the variable at two points separated by \(h\)
- \(E\): mathematical expectation.

Two assumptions are needed to apply kriging namely stationarity and isotropy.
Stationarity for spatial correlation is based on the assumption that the variables are stationary. When \(\gamma(h)\) does not depend on \(x\), where \(x\) is the point location and \(h\) is the distance between the points. So the semi-variogram depends only on the distance between the measurements and not on the location of the measurements.

Isotropy for spatial correlation means \(\gamma(h)\) depends only on \(h\). So the semivariogram depends only on the magnitude of \(h\) and not on its placement.

Usually, stationarity is also necessary for the expectation \(E,y(x)\), to ensure that the expectation doesn’t depend on \(x\) and is constant.

4.8.3 KRIGING WITH A TREND-UNIVERSAL KRIGING.

In both simple and ordinary kriging, the assumption of local stationarity is made; that is, the mean value of the variable over the search area is assumed constant. However, in some circumstances a trend is observed in the data such that the mean varies over the search area and it is therefore no longer locally stationary.
Kriging with a trend, or universal kriging, is a variant of ordinary kriging that can incorporate the effect of a trend on the local mean. It does not require prior knowledge of the mean, but does require a model to be supplied for the trend surface.
The universal kriging algorithm can generate the trend model by fitting a polynomial function to the local data. Alternatively, the trend can be supplied as an external (secondary) variable.
Kriging with a trend uses a random function model that is expressed as the sum of a trend and a residual. The variogram is calculated for the residuals, and used along with the specified trend to Krige the original data.
A requirement for kriging with a trend is the specification of a model variogram of the residuals from that trend.

Search radius

If the density of data is sufficiently high that it is possible to work with a small search radius, there will be little difference between ordinary kriging and kriging with a trend for estimates within the interior of the estimation area (Journel and Rossi, 1989), because ordinary kriging re-estimates the local mean for each search area, and can be considered as kriging with a trend where the trend is the constant local mean. In the interior, the local mean is very similar to the trend value for a small search area. When extrapolating beyond the edge of the dataset, the algorithm has to rely increasingly on the
trend, whether it is given explicitly when using kriging with a trend or calculated by ordinary kriging as the local mean for the data at the edge.

4.8.4. **COKRIGING.**

Cokriging is an extension of the basic kriging algorithm which allows one or more supplementary variables, which are spatially correlated (or assumed to be correlated) with the primary variable of interest, to be included in the estimation process. This is potentially useful if there are few samples of the quantity of interest but a greater number of measurements of the correlated variable. The development of the method is very similar to kriging with a single variable. Cokriging will only improve on the estimates of ordinary kriging when the primary variable is appreciably under sampled.

4.8.5. **KRIGING WITH EXTERNAL DRIFT.**

The external drift method is a particular case of universal kriging. It allows the prediction a variable Z, known only at small set of points of the study area, through another variable s, exhaustively known in the same area. We choose to model Z with a random function \( Z(x) \) and s as a deterministic variable \( s(x) \).

The two quantities are assumed to be linearly related, i.e. it is assumed that \( Z(x) \) is on average equal to \( s(b) \) up to a constant \( a_0 \) and a coefficient \( b_1 \):

\[
E[Z(x)] = a_0 + b_1 s(x)
\]

We examine the case of a second-order stationary random function \( Z(x) \) whose prediction is to be improved by introducing the shape function \( s(x) \) providing detail at a smaller scale than the average sample spacing for \( Z(x) \). The predictor is a linear combination of the sample values at location \( x_i \) (i=1,……,n) with unit sum weight \( w_i \),

\[
Z^*(x_0) = \sum_{i=1}^{n} w_i Z(x_i),
\]

With \( \sum_{i=1}^{n} w_i = 1 \). We look for an unbiased predictor, that is, with a prediction error which is expected to be zero, \( E\{Z^*(x_0)/Z(x_0)\} = 0 \), so that:

\[
E\{Z^*(x_0)\} = E\{Z(x_0)\}
\]

This equality can be developed into:

\[
E\{Z^*(x_0)\} = \sum_{i=1}^{n} w_i E\{Z(x_i)\} = a_0 - b_1 \sum_{i=1}^{n} w_i s(x_i) = a_0 - b_1 s(x_0).
\]

This equation implies that the weights should be on average consistent with an exact interpolation of \( s(x) \):
\[ S(x_i) = \sum_{j=1}^{n} w_j s_j(x_i) \]

The objective function \((O)\) to minimize in this case consists of the prediction variance \(\nu_E^2\) and of two constraints.

\[
0 \leq \nu_E^2 / \sigma_i \left( \sum_{j=1}^{n} w_j / n \right) / \sigma_i \left( \sum_{j=1}^{n} w_j s(x_j) / s(x_i) \right)
\]

where \(\sigma_i\) and \(\sigma_i\) are Lagrange parameters, and \(\nu_E^2\) is the prediction variance which is equal to:

\[
\nu_E^2 = \text{var}[Z^*/Z] = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j C(x_i / x_j) / 2 \sum_{j=1}^{n} w_j C(x_i / x_j) - C(0)
\]

with \(C\) being the covariance function.

The minimum of this quadratic function is found by setting the partial derivatives of the objective function \(O(W, \sigma; \sigma)\) to zero:

\[
\frac{\partial O}{\partial w_i} \bigg|_{0} = \sum_{j=1}^{n} w_j C(x_i / x_j) / \sigma_i s(x_i) \bigg|_{0} - \sigma_i s(x_i) \bigg|_{0} = 0 \quad \text{for} \quad i = 1, 2, \ldots, n
\]

\[
\frac{\partial O}{\partial \sigma_i} \bigg|_{0} = \sum_{j=1}^{n} w_j = 1
\]

\[
\frac{\partial O}{\partial \sigma} \bigg|_{0} = \sum_{j=1}^{n} w_j s(x_j) / s(x_i)
\]

The result of minimization is a system of linear equations called universal kriging equations:

\[
\sum_{j=1}^{n} w_j C(x_i / x_j) / \sigma_i s(x_i) \bigg|_{0} - \sigma_i s(x_i) \bigg|_{0} = 0 \quad \text{for} \quad i = 1, 2, \ldots, n
\]

\[
\sum_{j=1}^{n} w_j = 1
\]

\[
\sum_{j=1}^{n} w_j s(x_j) / s(x_i)
\]

With the minimal prediction variance

\[
\nu_k^2 = C(0) / \sum_{j=1}^{n} w_j C(x_i / x_j) - \sigma_i - \sigma_i s(x_i)
\]

The external drift method thus consists in incorporating into the kriging system additional universality conditions about one or several external drift variables.
S_i(x), i=1, … , M, measured exhaustively in the spatial domain. The functions s_i(x) need to be known at all locations x_i of the samples of Z(x_i), as well as at the nodes of the prediction grid. In this method, we assume a linear relationship between the variable of interest and the auxiliary variable at the observation points of the variable of interest. This assumption is very important in the prediction by the external drift method. Thus, if a non-linear function describes the relation between the two variables, this function should first be used to transform the data of the auxiliary variable. The transformed data could then be used as external drift (Bourennane et al., 2000). Auxiliary variables should be incorporated in the form of an external drift only if they are highly linearly correlated with the variable of interest. Otherwise it is preferable to use the method of co-kriging, which requires the fitting of a model for the cross-variograms between the different variables (Diodato, 2005).

4.8.6. REASON FOR CHOOSING KRIGING WITH EXTERNAL DRIFT.

We preferred the use of KED, because it allows incorporation of many covariates. It also requires a less demanding variogram analysis compared to collocated Kriging which requires a variogram for each of the covariates. Furthermore, comparison studies (Goovaerts, 2000) show KED interpolation to perform better than collocated co-Kriging. (Diodato, 2005) puts it that Cokriging will only improve on the estimates of ordinary Kriging when the primary variable is appreciably under-sampled, yet our primary variable is not very much under sampled.

4.9. THIN-PLATE SMOOTHING SPLINES.

Thin-plate smoothing splines is a related statistical technique (Wahba and Wendelberger, 1980; Cressie, 2003). The software package ANUSPLIN (Hutchinson, 1995) fits thin-plate splines (usually second- or third-order polynomials) through station data in three dimensions: latitude, longitude, and elevation. The main advantages of ANUSPLIN over kriging are that a semivariogram need not be developed (instead, a smoothing term is automatically tuned to minimize the cross-validation error), and the relationship between the climate variable and elevation can vary in space, making the method suitable for large domains. Because a spline is by definition smoothly varying, this approach has difficulty simulating sharply varying climate transitions, which are characteristic of temperature inversions, rain shadows, and coastal effects. Examples of data sets developed using tri-variate thin-plate splines include New et al. (2002) and Hijmans et al.(2006).

4.10. DAYMET

Daymet focuses on the effects of elevation on climate (Thornton et al., 1997). Daymet develops local linear regressions between climate and elevation for each grid cell on a digital elevation model, using data from surrounding stations. Each station is weighted in the regression function by its distance from the target grid cell. This method takes into account the elevational variation of climate, and it’s simple station distance weighting algorithm is computationally efficient. (http://www.daymet.org) Daymet does not have the ability to simulate non-monotonic relationships between climate and elevation, such as temperature inversions, and does not explicitly account for terrain-induced climatic transitions or coastal effects.
4.11. PRISM

PRISM (Parameter-elevation Regressions on Independent Slopes Model) also develops local climate elevation regression functions for each DEM grid cell (Daly, 1994) but calculates station weights on the basis of an extensive spatial climate knowledge base that assesses each station’s physiographic similarity to the target grid cell (Daly, 2002). It uses point data, a digital elevation model (DEM), and other spatial datasets to generate gridded estimates of annual, monthly and event-based climatic parameters.

In operation, PRISM Estimates the elevation of each precipitation station using a smoothed digital elevation model (DEM); and Assigns each DEM grid cell to a topographic facet by assessing slope orientation. PRISM then estimates precipitation at each DEM cell by developing a localized precipitation/DEM-elevation regression function from nearby rainfall stations; and Predicting precipitation at the cell's DEM elevation with this regression function. In the regression function, greater weight is given to stations with location, elevation, and topographic positioning similar to that of the grid cell. Whenever possible, PRISM calculates a prediction interval for the estimate, which is an approximation of the uncertainty involved. (Daly, 1994)

The knowledge base and resulting station weighting functions currently account for spatial variations in climate caused by elevation, terrain orientation, effectiveness of terrain as a barrier to flow, coastal proximity, moisture availability, a two-layer atmosphere (to handle inversions), and topographic position (valley, mid-slope, ridge).

While PRISM explicitly accounts for more spatial climate factors than other methods, it also requires more effort, expertise, and supporting data sets to take advantage of its full capability.

While Daymet and PRISM develop a local regression function between elevation and climate at each DEM grid cell, regional regression techniques develop a single, domain-wide, multivariate regression function between climate and latitude, longitude, and elevation (and sometimes other variables such as wind direction, distance from the coast, etc.). This approach has the advantage of being statistically stable, and often explains a large proportion of the climate variability within the domain. However, a single regression function has difficulty handling spatially varying relationships between one or more explanatory variables and climate across a region, and is therefore usually confined to regional, rather than continental domains that are relatively homogeneous. Examples of the regional regression approach include Ollinger et al. (1995), Goodale et al. (1998), Brown and Comrie (2002), and Johansson and Chen (2005).

4.12. MULTIPLE LINEAR REGRESSION.

The MLR method assumes that a linear relationship between the predictor variables (typically topological variables) and a known response variable (precipitation) can be fitted and used to estimate the response variable at any desired locations. In a simple linear regression model, a single response measurement \( Y \) is related to a single predictor (covariate, regressor) \( X \) for each observation. The critical assumption of the model is that the conditional mean function is linear:

\[
E(Y / X) = \beta_0 + \beta_1 X
\]

In most problems, more than one predictor variable will be available. This leads to the following “multiple regression” mean function:

\[
E(Y / X) = \beta_0 + \beta_1 X + \ldots + \beta_p X_p
\]
$Y_i = \alpha + \delta_i X_{j} + \ldots + \delta_i P_i + \gamma_i$

Where $\alpha$ is the intercept and the $\beta_j$ are called slopes or coefficients. We write $X_{i,j}$ for the $j$th predictor variable measured for the $i$th observation. The main assumptions for the errors $\gamma_i$ is that $E(\gamma_i) = 0$ and $\text{var}(\gamma_i) = \sigma^2$ (all variances are equal). Also the $\gamma_i$ should be independent of each other.

For small sample sizes, it is also important that the $\gamma_i$ approximately have a normal distribution.

These regression coefficients are estimated by minimizing the squared errors. Because of the strong orographic effects on the temperature and precipitation, elevation is included as a predictor variable in most spatial interpolation researches on MLR.

4.12.1. EXPERIMENTAL DESIGN

This part takes you through the different steps we undertook to get the final outcome. It Emphasizes on a geostatistical analysis of external drift Kriging algorithm using datasets from the areas of study. The method that was followed is as shown in the flow chart below.
1. The steps for spatial analysis of precipitation in this study have been as follows:
2. Preparing R Statistics software input files.
3. Analysis of Correlation between precipitation of different months and covariates.
4. Analysis of variograms for selection of model parameters including, Nugget, range(hmax), sill, variogram type
5. Kriging calculations;
6. Calculating average of precipitation in the area;
7. Determining the equations for calculating precipitation according to the data of selected stations in each region;

4.12.1.1. **CONVERSION OF IMAGES.**

We had a DEM which was degraded using neighbourhood statistics in Spatial Analyst Arc map. It was degraded to 2km resolution from 90 Meters.
The Elevation map was then converted to Aspect and Slope using Arc toolbox, Spatial Analyst tools, surface. Slope and Aspect.
We also had a distance from the coast map and vegetation continuous field map.

4.12.1.2. **CLIPPING THE IMAGES.**

The maps were clipped using spatial analyst tools, extraction. Extract by mask. This was done in order to have only the area of study.

4.12.1.3. **EXTRACTION OF DATA.**

We overlaid our rain gauge points over the different images and extracted data from these images.
The points now had values for all the covariates we needed to use.
We used Spatial analyst tools, extract values to points.
After extracting the data, the table was saved as a dbf, then changed to csv.
All this was done in Arc map.
We assessed the csv file and noted there were some points which had “no data” extracted since they were at points where there was no data. These came out as -9999 and we removed them so that we don’t have them having influence on our statistics because this could increase error.
Note: these could not be changed to 0 because 0 is considered a value.
It is while extracting the points in Arc map that we transformed the aspect and Vegetation continuous field and calculated the topographic index using calculate statistics in

4.12.1.4. **INTERPOLATION GRID.**

An interpolation grid was created in R using the expand grid command.
fin_utta_2k <- expand.grid(X=seq(237458.615,558390.928, by=2000),Y=seq(-1895656.696,-1581802.361, by=2000))
This was then saved as a comma separated values file (csv)
>write.table(as.data.frame(fin_utta_2k), file="fin_utta_2k.csv",sep="", row.names=F)

4.12.1.5. **CLIPPING THE GRID AND EXTRACTING COVARIATES DATA.**

The created grid was imported into Arc map and we made a point map using Make XY event layer command.
We then defined the projection into WGS 84 Albers. Noting that all our images were in this projection.
This was then clipped so that we only have grids for our study area remaining. We used Analysis tools, extract, clip in Arc map. We overlaid the grid on the different images in order to extract the various covariates. This was then saved as dbf, changed to csv and removed points that had -9999 from the data table. At this point we had all our tables for the rain gauge data and interpolation grid with all the covariates that we needed to use in them. The csv files were then placed in the directory (workspace) in R and then loaded into R.

### 4.12.1.6. WORKING IN R.

We loaded the csv files rain gauge point data and grid data in R using read.csv command.

```r
> utta <- read.csv("uttarainvcfed.csv")
> uttagrid <- read.csv("uttagridvcfed.csv")
> attach(utta)
```

Then converted them into sp objects.

```r
> coordinates(utta) <- ~ X + Y
> coordinates(uttagrid) <- ~ X + Y
```

### 4.12.1.7. CORRELATION.

First we looked at the correlation of the different independent variables with observed amount of rainfall to help us know whether there is good correlation between the covariates and rainfall. Where there is strong correlation, then the covariate is taken as a good prediction estimator. Good correlation refers to correlation coefficient that lies between -0.5 to -1 and +0.5 to +1.

### 4.12.1.8. MODELING THE VARIOGRAM

The spatial continuity of the data was examined on the basis of our variogram analysis. We made variograms for different months to get the sill, range, partial sill and nugget. The procedure is as explained below using April variogram for Karnataka. The cut-off and width are gotten by using “the eye” fit whereby we adjusted the variogram range and bin size by looking and seeing what the best fit is.

```r
> va <- variogram(Apr.00 ~ 1, ~X+Y, knt, cut-off=270000, width=17000)
Apr.00= object to be estimated, X+Y=coordinates, knt=location of Apr.00, cut-off=range, width=bin size
> plot(va)
```
The variogram was then fitted using the authorized functions, i.e. spherical, Gaussian etc “fitting” the variogram well.

> vma <- vgm(740,"Gau", 135000,850)
740=partial sill, 135=range and 850=nugget
> plot(va, plot=T, model=vma)

We then fitted (va and vma) to get the intrinsic fitted variogram

(vmaf <- fit.variogram(va, vma))

R console output for the fitted variogram

<table>
<thead>
<tr>
<th>model</th>
<th>psill</th>
<th>range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nug</td>
<td>867.8780</td>
<td>0.0</td>
</tr>
<tr>
<td>Gau</td>
<td>942.7247</td>
<td>179331.1</td>
</tr>
</tbody>
</table>

plot(va, model=vmaf,plot=T) shown below.

This was then compared with the one we fitted using our “eye fit” conclusion method, almost similar.

for more details on making a variogram in R run command “?variogram” in R or refer to <http://www.gstat.org/>
4.12.1.9. **APPLICATION OF THE KRIGING MODELS.**

We then carried out ordinary kriging and external drift kriging for the different months, April, July and December for the two states, Uttarakhand and Karnataka, noting that these fall in the three climatic seasons of India and experience different climatic conditions.

**ORDINARY KRIGING**

\[
\text{oka} \leftarrow \text{krige(Apr.00 ~ 1, knt, newdata=kntgrid, model=vmaf)}
\]

R console output:

```
[using ordinary kriging]
source("CODE/ok_plotfnsed.R")
```

the above source command instructs R to source for script that we edited to suit our study, which is in a folder named CODE, ok_plotfnsed is the file. please refer to appendix 3 for details on this script.

To visualize interpolation this we run the command below.

```
plot.kresults(oka, "var1", knt, "Apr.00")
```

OUTPUT:

![Prediction Map](image1)

![Std deviation of the Prediction error](image2)

**Figure 4-7 OK INTERPOLATION USING EDITED SCRIPT**

If the R normal script is used:

```
levelplot(var1.pred~X+Y, as.data.frame(oka), aspect="iso")
```

OUTPUT: on next page…
we then cross validated the model using the following script.

```
okacv <- krig.cv(Apr.00 ~ 1, knt, model=vmaf)
```

R console output:

```
[using ordinary kriging]
[using ordinary kriging]
```

The cross validation was done 41 times for Uttaranchal and 168 times for the number of times

We commanded R to put the residuals of OK into “okares” so as to run statistics on them.

```
okares <- as.data.frame(okacv)$residual
```

A histogram of the residuals was plotted. This helps in looking at how good the model is whereby for a good model, the residuals in the histogram should follow a normally distribution curve. (bell shaped).

A bubble plot of the residuals was also plotted, this revealed to us the points which were under and overestimated by the model.

```
hist(okares,breaks=20,col="lightblue",border="red", main="Residuals for April")
bubble(okacv, "residual")
```
We ran the following command to reveal to us the Root Mean Squared Error and mean error.

\[
> \text{sqrt(mean(okares^2))} \quad \text{mean(okares)}\\
[1] 28.53745 \quad [1] -0.09243677
\]

We cross validated the models that were used in ordinary kriging and compared their root mean squared error and mean error to help find out the model that was a good predictor, i.e. one with RMSE close to zero.

We then proceeded to kriging with external drift.

**EXTERNAL DRIFT KRIGING.**

In external drift kriging, we have five covariates for Uttaranchal and six for Karnataka. These were Elevation, aspect, slope, vegetation cover factor, distance from the sea and topographic index. Topographic index is a derived covariate as compared to the rest of the covariate.

For Uttaranchal we only had five covariates, i.e. we didn’t use distance from the coast here because it would not be a good covariate noting its distance from the sea and it being so much inland. More so data for this was not available.

We started by kriging using one covariate and proceeded to use combination of two, three, up to six different covariates to see which combination gave us the best estimation of precipitation in our area of study.

In total we had 47 different models for one month of one state.

These were then compared using cross validation to come out with the best model in the study area for that month.

The procedure involved.

Here we will use an example of using slope as a covariate.

Computing and modeling the variogram of the residuals by keying in details of the variogram we had calculated earlier i.e. the experimental variogram and display it.

\[
> \text{Avrs5<- variogram(Apr.00~slope,loc=knt);plot(Avrs5, plot.numbers=T)}\\
>(\text{Amrs5<-fit.variogram(Avrs5,vgm(740,"Gau", 135000,850))})
\]

R console output:

<table>
<thead>
<tr>
<th>model</th>
<th>psill</th>
<th>range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Nug  955.1275</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>Gau 8937.6335 791368.5</td>
<td></td>
</tr>
</tbody>
</table>

We then run the external drift kriging model.

\[
> \text{Akrs5<-krige(Apr.00~slope,loc=knt,newdata=kntgrid,model=Amrs5)}
\]

[using universal kriging]

**NOTE:** R displays external drift kriging as universal kriging, in that it runs almost the same model as universal kriging.

source("CODE/ked_plotfnserrms.R")

the above source command instructs r to source for script that we edited to suit our study, which is in a folder named CODE, ok_plotfnsed is the file. please refer to appendix 4 for details on this script.

To visualize interpolation this we run the command below.

plot.kederresults(Akrs5, "var1", knt, "Apr.00")
ANALYSIS OF EXTERNAL DRIFT KRIGING ALGORITHM WITH APPLICATION TO PRECIPITATION ESTIMATION IN COMPLEX OROGRAPHY.

Figure 4-11 KED INTERPOLATION USING EDITED SCRIPT.

Run the script below for the normal external drift output for prediction and prediction error maps.

```r
> levelplot(var1.pred~X+Y, as.data.frame(Akrs5),aspect="iso", main="KED Prediction Map")
> levelplot(var1.var~X+Y, as.data.frame(Akrs5),aspect="iso", main="KED Std deviation Prediction error")
```

Figure 4-12 KED INTERPOLATION USING NORMAL SCRIPT.

To have a look at the mean, median, min and max statistics of our predictions and variations, we run the following command.

```r
> summary(Akrs5)…for this output refer to appendix 5.
> Aslcv <- krig.cv(Apr.00~slope, knt, model = vma)
[using universal kriging]
[using universal kriging]
> Astsl <- as.data.frame(Aslcv)$residual
```
We then compared our ordinary kriging and external drift kriging cross validation RMSE and ME results as shown below. This was done for all the possible combinations of covariates which came to 67 different models, for Karnataka and 31 for Uttaranchal.

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>MEAN ERROR</th>
</tr>
</thead>
<tbody>
<tr>
<td>OK</td>
<td>28.53745</td>
<td>-0.09243677</td>
</tr>
<tr>
<td>KED</td>
<td>28.36692</td>
<td>-0.1037706</td>
</tr>
</tbody>
</table>

Table 4-1 OK KED COMPARISON

The same was applied to all the three months, for the two study areas and we picked out the model which has the least RMSE in each month for the two study areas.

Among the things noted was that different combinations gave different RMSE results and good performance of one model in a given month or a given study area, didn’t necessarily lead to the same outcome for a different month or study area. This is discussed in the results section.

After noticing some “best” models that still had errors, we sought to look at anisotropy. Refer to section 5.7.
5. RESULTS AND DISCUSSION

5.1. KARNATAKA STATE.

This chapter seeks to reveal the outcome of this research basing on the steps followed in the methodology and the research questions that we seek to answer and the objective of the study. It starts with an explorative analysis of the data. To assist in understanding the best ways to go about handling the data and make it easier in analysis too. We then proceed with estimating precipitation using ordinary kriging for both areas Uttaranchal and Karnataka, then external drift kriging with covariates follows.

This is followed by analysis of the results of estimations using the different models (ordinary kriging and kriging with external drift) then analysis of kriging with external drift using different covariates, different seasons then the performance of these models in the two study areas.

5.1.1. EXPLORATIVE DATA ANALYSIS.

Shown below are tables of correlation coefficients of rainfall received and different variables. If a covariate has good correlation then there is a likelihood of it being a good predictor, “holding other climatic influences constant”. This is because apart from a covariate, there are other elements that also come into play that would reduce the ability of a covariate being a good predictor.

The tables are for both 2kilometer and 10kilometer grid.

<table>
<thead>
<tr>
<th></th>
<th>ELEVATION</th>
<th>SLOPE</th>
<th>T-ASPECT</th>
<th>DIST COAST</th>
<th>V.C.F</th>
<th>T.I</th>
</tr>
</thead>
<tbody>
<tr>
<td>APRIL</td>
<td>0.325</td>
<td>-0.021</td>
<td>0.056</td>
<td>-0.422</td>
<td>0.100</td>
<td>0.135</td>
</tr>
<tr>
<td>JULY</td>
<td>-0.338</td>
<td>0.157</td>
<td>-0.018</td>
<td>-0.324</td>
<td>0.231</td>
<td>0.101</td>
</tr>
<tr>
<td>DECEMBER</td>
<td>0.253</td>
<td>0.029</td>
<td>-0.047</td>
<td>-0.303</td>
<td>0.060</td>
<td>0.095</td>
</tr>
</tbody>
</table>

Degrees of freedom = 167

Table 5-1 CORRELATION COEFFICIENT 2KM GRID

In the table above the “best” correlation is between distance to the coast and rainfall amount in April, followed by Elevation and rainfall amount recorded in July. However, they are not good correlations since they are not above 0.5 or below -0.5. It is observed that the correlation between rainfall in the different months and the covariates is poor.

<table>
<thead>
<tr>
<th></th>
<th>ELEVATION</th>
<th>SLOPE</th>
<th>T-ASPECT</th>
<th>DIST COAST</th>
<th>V.C.F</th>
<th>T.I</th>
</tr>
</thead>
<tbody>
<tr>
<td>APRIL</td>
<td>0.307</td>
<td>0.056</td>
<td>-0.072</td>
<td>-0.415</td>
<td>0.213</td>
<td>0.240</td>
</tr>
<tr>
<td>JULY</td>
<td>-0.359</td>
<td>0.253</td>
<td>-0.042</td>
<td>-0.448</td>
<td>0.170</td>
<td>0.056</td>
</tr>
<tr>
<td>DECEMBER</td>
<td>0.250</td>
<td>0.031</td>
<td>-0.016</td>
<td>-0.295</td>
<td>0.187</td>
<td>0.215</td>
</tr>
</tbody>
</table>

Degrees of freedom = 167

Table 5-2 CORRELATION COEFFICIENT 10KM GRID

In the table above, the “best” correlations are between rainfall for April and Distance to the coast followed by July rainfall and elevation.
5.1.1.1. HISTOGRAMS

Histograms for the Observed and normalized rainfall amount for April, July and December.

A good dataset’s histogram should have a normal curve, i.e. bell shaped. If there are breaks and differing amounts then that could be an indicator of clusters in a dataset. In the histograms for Karnataka they are skewed. To do statistics on such a dataset, one can normalize it by for example taking a lognormal of the data e.g. for the different months. This is shown next to each histogram. In the study we did not use the lognormal because of the many 0 values in the data.
5.1.2. **AVERAGE RAINFALL AMOUNT OBSERVED AND ACCOMPANYING VARIOGRAM OF THE MONTH.**

Point maps of amount of rainfall observed in the different months, April, July and August, respectively.

There is more amount of rainfall in the southern part of Karnataka which also has higher elevation.

In July more rainfall is observed in the area close to the sea, and just before the western ghat after which there is very little rainfall observed. This is because of the western ghat influence.
As shown in the point maps, more rainfall is observed in the higher elevation areas of the state, i.e. the western Ghat area (south west) and the hilly area in the south.

<table>
<thead>
<tr>
<th></th>
<th>APRIL</th>
<th>JULY</th>
<th>DECEMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>PARTIAL SILL</td>
<td>942.725</td>
<td>95699.07</td>
<td>1603.968</td>
</tr>
<tr>
<td>RANGE</td>
<td>179331.1</td>
<td>238409.4</td>
<td>323547.6</td>
</tr>
<tr>
<td>NUGGET</td>
<td>867.870</td>
<td>26873.11</td>
<td>171.842</td>
</tr>
<tr>
<td>BINSIZE</td>
<td>17000</td>
<td>21000</td>
<td>14000</td>
</tr>
<tr>
<td>MODEL</td>
<td>Gaussian</td>
<td>Linear</td>
<td>Spherical</td>
</tr>
</tbody>
</table>

Table 5-3 DETAILS FOR KARNATAKA VARIOGRAMS

Variograms help reveal the spatial dependence of the rainfall points, i.e., it shows the maximum distance where point influence others, for different stations. The figures shown above on the right side of the point maps are for the fitted variograms for each month that we use for our study. Usually we can observe that the dissimilarity between values increases on average when the spacing (lag, bin size) between the pairs of sample points is increased. There is good fitting observed in the variograms for April and July as compared to the variogram for December. However the spatial dependence is so poor, just like it was depicted by low correlation.
APRIL
Ordinary kriging for the month of April, the models performance is poor

Kriging with External Drift
The model that performed the best in April for Karnataka state using kriging with external drift was one facilitated by a combination of elevation and slope as covariates, the model performance was based on the value for root mean squared error, where the model with the least RMSE was chosen as the best.

The external drift kriging model trend coefficients were:
(Intercept) = -5.62338, elevation = 0.06444, slope = -1.00650 on 166 degrees of freedom
JULY
ORDINARY KRIGING

KRIGING WITH EXTERNAL DRIFT
In July the best covariate is elevation. Elevation stands out as the best covariate because it is aided by the wind that blows from the Indian ocean, south west monsoon has a great influence and the western ghats assists in aiding of formation of orographic rainfall.

The external drift kriging model trend coefficients were:
(Intercept)= 529.5744 , elevation = -0.4827 on 167 degrees of freedom
DECEMBER ORDINARY KRIGING

KRIGING WITH EXTERNAL DRIFT
December has little rainfall estimated and a very good prediction error map, which reveals that the models performance is quite good. The best predictor covariate was Topographic index.

The external drift kriging model trend coefficients were:
$$(\text{Intercept}) = 7.343167, TI = 0.003174 \text{ on 167 degrees of freedom}$$
5.1.3. **Bubble Plots for the Residuals of the Various Months.**

Bubble plots of the residuals for April, July and December respectively, showing over and underestimation. Over-estimation is indicated by the red bubbles and underestimation is indicated by the green bubbles.

From the bubble plot, it is seen that there is high over and underestimation in the area with complex orography for both ordinary kriging and external drift kriging, despite kriging with external drift having a lower RSME than ordinary kriging algorithm for the same month and same area.

**April**

<table>
<thead>
<tr>
<th>Residuals</th>
<th>Ordinary Kriging</th>
<th>External Drift Kriging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 5-25</td>
<td><img src="image1" alt="Figure 5-25" /></td>
<td><img src="image2" alt="Figure 5-26" /></td>
</tr>
</tbody>
</table>

It is observed that there are so many red and green points, as compared to interpolating the month of December. As it can be observed, performance of the model is not good, especially in the complex region of the state. This is based on the over and underestimation.

**July**

<table>
<thead>
<tr>
<th>Residuals</th>
<th>Ordinary Kriging</th>
<th>External Drift Kriging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 5-27</td>
<td><img src="image3" alt="Figure 5-27" /></td>
<td><img src="image4" alt="Figure 5-28" /></td>
</tr>
</tbody>
</table>
With reference to the histogram and the residuals bubble plot, there is more error in the southern part of Karnataka, which is the area with complex terrain and is the area that receives more rainfall too.

**DECEMBER**

Excluding the histogram of the standardized residuals. These histograms should give an indication of the types of prediction errors obtained by the model. If several outliers exist, it is an indication that the model is not fitting the data well. If so then the spatial model may be inappropriate.

The standardized residuals follow a symmetric shape which reveals that kriging model assumptions may be correct.

Another method useful for evaluating the residuals spatially is to plot the residuals against their longitude or latitude separately. These plots can indicate whether a pattern exists along a one-dimensional line. As before, any pattern other than random noise might indicate an invalid model.

The residuals of the models show that the they are good models. This is based on the nature of the histogram which is normally distributed.

**5.1.5. COMPARE ANALYSIS FOR MODELS IN KARNATAKA.**

For comparing the models, cross validation was used. In cross validation we look at the Root mean squared error which shows the precision of the model and Mean Error which shows the bias of the model. A perfect case, RMSE should be =0.

The errors by ordinary kriging model are quite high especially for the month of July, and less for the other months, April and December. This region receives a lot of rainfall in July owing to its proximity to the ocean and the moisture laden winds (South west monsoon) that blow from the south west.
Although surfaces interpolated by kriging are smooth, we also studied the error map. Noting that all forms of kriging yield estimates of the estimation uncertainty or kriging error. Such values can be mapped to provide error surfaces which are be combined with other information. (Burrough, 2001)

There is higher error in the area with the range of mountain, and more rainfall recorded too (shown by the grey dots in the July prediction map. Basing on this we can say that ordinary kriging doesn’t perform as well in complex terrain.

<table>
<thead>
<tr>
<th>MODEL</th>
<th>APRIL</th>
<th>JULY</th>
<th>DECEMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>ME</td>
<td>%RMSE</td>
</tr>
<tr>
<td>Elev</td>
<td>27.487</td>
<td>-0.029</td>
<td>79.512</td>
</tr>
<tr>
<td>TI</td>
<td>28.892</td>
<td>-0.534</td>
<td>83.575</td>
</tr>
<tr>
<td>elsl</td>
<td>27.475</td>
<td>0.059</td>
<td>79.478</td>
</tr>
</tbody>
</table>

Table 5-4 ROOT MEAN SQUARED ERROR AND MEAN ERROR FOR THE BEST MODELS IN THE THREE MONTHS, APRIL, JULY AND DECEMBER AND COMPARISON OF THEIR PERFORMANCE IN THE OTHER MONTHS.

Ti=Topographic Index
Elsl= combination of elevation and slope.

Table showing comparison between ordinary kriging and external drift kriging root mean squared errors.

<table>
<thead>
<tr>
<th>MONTH</th>
<th>ORDINARY KRYING</th>
<th>EXTERNAL DRIFT KRYING</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>ME</td>
</tr>
<tr>
<td>APRIL</td>
<td>28.71</td>
<td>-0.108</td>
</tr>
<tr>
<td>JULY</td>
<td>179.605</td>
<td>3.302</td>
</tr>
<tr>
<td>DECEMBER</td>
<td>11.696</td>
<td>-0.432</td>
</tr>
</tbody>
</table>

Table 5-5 COMPARISON BETWEEN ORDINARY KRIGING AND EXTERNAL DRIFT KRIGING ROOT MEAN SQUARED ERRORS.
5.2. **UTTARANCHAL STATE.**

5.2.1. **EXPLORATIVE DATA ANALYSIS.**

It was important to do an analysis of the data set by exploring different things like histograms in order to gain some basic understanding. We started by looking at the correlation coefficient of the different variables with Rainfall for the different months. This assisted in determining the strength of correlation.

If a covariate has good correlation then there is a likelihood of it being a good predictor, “holding other climatic influences constant”. This is because apart from a covariate, there are other elements that also come into play that would reduce the ability of a covariate being a good predictor.

<table>
<thead>
<tr>
<th></th>
<th>ELEVATION</th>
<th>SLOPE</th>
<th>T-ASPECT</th>
<th>V.C.F</th>
<th>T.I</th>
</tr>
</thead>
<tbody>
<tr>
<td>APRIL</td>
<td>0.820</td>
<td>0.590</td>
<td>0.011</td>
<td>0.255</td>
<td>0.557</td>
</tr>
<tr>
<td>JULY</td>
<td>0.107</td>
<td>0.173</td>
<td>0.024</td>
<td>0.018</td>
<td>0.085</td>
</tr>
<tr>
<td>DECEMBER</td>
<td>0.671</td>
<td>0.450</td>
<td>0.014</td>
<td>0.116</td>
<td>0.431</td>
</tr>
</tbody>
</table>

**Table 5-6 CORRELATION COEFFICIENT FOR UTTARANCHAL**

The underlined figures indicate where there is good correlation.

**HISTOGRAMS**

Histograms for the Observed rainfall amount for April, July and December.

A good dataset’s rainfall histogram should resemble a normal curve, i.e. bell shaped. If there are breaks and differing amounts then that could be an indicator of clusters in a dataset.
5.2.1.1. AVERAGE RAINFALL AMOUNT OBSERVED and VARIOGRAMS.

Variograms help reveal the spatial dependence of the rainfall points. I.e. It reveals the maximum distance where one point influences another, for different stations. Different aspects of the variograms fitted are shown in the table below.

<table>
<thead>
<tr>
<th></th>
<th>APRIL</th>
<th>JULY</th>
<th>DECEMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>PARTIAL SILL</td>
<td>270</td>
<td>25077.83</td>
<td>32.285</td>
</tr>
<tr>
<td>RANGE</td>
<td>190000</td>
<td>33935.87</td>
<td>56880.39</td>
</tr>
<tr>
<td>NUGGET</td>
<td>38</td>
<td>11339.62</td>
<td>15.801</td>
</tr>
<tr>
<td>BINSIZE</td>
<td>8000</td>
<td>13500</td>
<td>14000</td>
</tr>
<tr>
<td>MODEL</td>
<td>Linear</td>
<td>Gaussian</td>
<td>Gaussian</td>
</tr>
</tbody>
</table>

Table 5-7 DETAILS FOR VARIOGRAMS FOR UTTARANCHAL

The described Variograms are as shown below alongside the amount of precipitation recorded for the months of study. April, July and December.

From the table we can see that the range for the variogram is very high especially for April as compared to July and December which have low range.

To visualize the amount of rainfall recorded for the different months, we used point maps of amount of rainfall observed as shown below. The point maps are for the months of April, July and August respectively.

As it can be seen, more rainfall is observed in the higher elevation areas of the state, i.e. the Himalayas.

APRIL
JULY

Usually we can observe that the dissimilarity between values increases on average when the spacing (lag, bin size) between the pairs of sample points is increased. The behavior at very small scales, near the origin of the variogram, is critical as it indicates the type of continuity of the regionalized variable: differentiable, continuous but not differentiable, or discontinuous. In this last case, when the variogram is discontinuous at the origin, this is a symptom of a 'nugget-effect', which means that the values of the variable change abruptly at a very small scale.

Reason for having a problematic variogram for July can be caused by local clusters of rainfall. As indicated in the variogram above, there are lots of outliers.

OBSERVED RAINFALL AMOUNT FOR JULY

VARIOGRAM FOR JULY

Figure 5-36

Figure 5-37

OBSERVED RAINFALL AMOUNT FOR DECEMBER

VARIOGRAM FOR DECEMBER

Figure 5-38

Figure 5-39
APRIL
ORDINARY KRIGING

EXTERNAL DRIFT KRIGING
Estimation using elevation as drift. The external drift kriging model trend coefficients were: 
(Intercept) = 7.842392 , elevation = 0.014503 on 38 degrees of freedom
JULY
ORDINARY KRIGING

Prediction Map

Std Deviation Prediction error

Figure 5-44

Figure 5-45

EXTERNAL DRIFT KRIGING

Estimation using combination of slope-aspect-vegetation cover factor and topographic index.

The external drift kriging model trend coefficients were:

(Intercept) 408.7039, slope = 11.6917, taspect = 39.2266, vcf = -12.7104, TI = 0.3089

on 35 degrees of freedom

Prediction Map

Std Deviation of prediction error

Figure 5-46

Figure 5-47
**DECEMBER ORDINARY KRIGING**

The external drift kriging model trend coefficients were:

(Intercept) = 18.1799, vcf = 0.7303 on 38 degrees of freedom

---

**EXTERNAL DRIFT KRIGING**

ESTIMATION USING VEGETATION COVER FACTOR (VCF) AS DRIFT.
RESIDUALS

Bubble plots of the residuals for April, July and December respectively, showing over and underestimation. Overestimation is indicated by the red bubbles and underestimation is indicated by the green bubbles.
The geostatistical approach is based on the assumption that sample values are related to one another in a way which is dependent on their distance. Then a primary aim of Geostatistics is to estimate the spatial relationship between sample values. This estimate is used to make spatial prediction of unobserved values from neighbouring samples and to give an estimate of the variance of the prediction error (Musio, 2004).

The spatial distribution of precipitation in July is likely to have been driven by wind orientation and relief (with rain occurring most on the windward side of slopes and least on the leeward side). However, there is little observable relation which persisted across the whole region. The modelling (fitting) of the July variogram is problematical but although the relationship may have been non-linear no obvious association is observed.

Although the precipitation measurements involve simple instrumentation and procedures, they are prone to significant systematic errors that are mostly due to wind affecting gauge catching efficiency (Diodato and Ceccarelli, 2005).

In a study carried out by (Hosseini, 2006) in south west Iran, he found that the Regression coefficient between rainfall and elevation was low in one of the sub catchments (R=0.12) and in two others was better (0.72 and 0.82). In general, for monthly data, this coefficient was less than 0.5. These results showed that the regression coefficient between rainfall and elevation must be greater than 0.6 in order to improve interpolation accuracy using elevation as co variable.

Daly in his study discusses the relationship between scale and spatial climate-forcing factors, and provides background and advice on assessing the suitability of data sets. Spatial climate patterns are most affected by terrain and water bodies, primarily through the direct effects of elevation, terrain-induced climate transitions, cold air drainage and inversions, and coastal effects. The importance of these factors is generally lowest at scales of 100 km and greater, and becomes greatest at less than 10 km. Except in densely populated regions of developed countries (Daly, 2006). We see this phenomena in Karnataka and this are some of the things that lead to precipitation estimation in such terrain difficult, because of the different factors that influence, and possibly even their interaction.

Another method useful for evaluating the residuals spatially is to plot the residuals against their longitude or latitude separately. These plots can indicate whether a pattern exists along a one-dimensional line. As before, any pattern other than random noise might indicate an invalid model.
Root mean squared error shows the precision of the model while mean error shows the bias (accuracy) of the model. A perfect case, RMSE should be =0.

Examining the histogram of the standardized residuals. This histogram should give an indication of the types of prediction errors obtained by the model. If several outliers exist, it is an indication that the model is not fitting the data well. If so then the spatial model may be inappropriate. The standardized residuals follow a symmetric shape which reveals that kriging model assumptions is correct.

5.2.2. CROSS VALIDATION UTTARANCHAL

Root mean squared error was used to summarize the cross validation performance of the models. It shows the precision of the model while mean error shows the bias (accuracy) of the model. A perfect case, RMSE should be =0.

Examining the histogram of the standardized residuals. This histogram should give an indication of the types of prediction errors obtained by the model. If several outliers exist, it is an indication that the model is not fitting the data well. If so then the spatial model may be inappropriate. The standardized residuals follow a symmetric shape which reveals that kriging model assumptions are correct.

Table 5-8 ROOT MEAN SQUARED ERROR AND MEAN ERROR FOR THE “BEST” MODELS IN THE THREE MONTHS, APRIL, JULY AND DECEMBER AND COMPARISON OF THEIR PERFORMANCE IN THE OTHER MONTHS.

tvcf=vegetation cover factor
SLTAVCFTI= COMBINATION OF Slope + Aspect + Vegetation Cover Factor + Topographic Index.

Table 5-9 MODEL PERFORMANCE COMPARISON BETWEEN ORDINARY KRIGING AND KRIGING WITH EXTERNAL DRIFT.
Underlined figures show the models with the least RMSE which reveals the combination of secondary variables that are better precipitation estimators in that region.

The histogram of the model which uses elevation only as a secondary variable, i.e., for April, has its residuals normally distributed. The model which uses vegetation cover factor as a secondary variable, i.e., the best method for the month of December in Uttaranchal has its residual histogram slightly skewed however close to being normally distributed. The model that comes out as the best performer in Uttaranchal for December i.e., one that uses Slope, aspect, vegetation cover factor ad topographic index as covariates for precipitation estimation has its residual histogram not normally distributed. This indicates that it is not a very perfect model. One of the things this can be attributed to is the variogram that was fitted, which was troublesome to fit owing to its nature. The way the residuals are distributed, it reveals that there is presence of clusters.

5.3. PERFORMANCE OF VARIOUS MODELS.

This was done using cross validation. Something that was observed was that using different number of stations to cross validate gives different RMSE, and we settled for using all the rain gauge points available in the state because it was observed that using a specified number of points say 40 point stations gives us a different output every time we run the cross validation model. This is attributed to the “cluster” problem and the fact that specifying 40 point stations, the algorithm searches for 40 points around the one left out, and every time this is done different points are chosen which leads to a different RMSE output. Having noted this we settled for using all the points available. (Daly, 2006) in his study mentions that there is no one satisfactory method for quantitatively estimating errors in spatial climate data sets, because the field that is being estimated is unknown between data points. Perhaps the best overall way to assess errors is to use a combination of approaches, involve data that are as independent from those used in the analysis as possible, and use common sense in the interpretation of results.

One thing that is noted in using leave one out cross validation is that it is susceptible to clustering. whether performed with a single variogram or multiple variograms, and this can be misleading when the data are clustered. In such a case, removing a single observation has little effect on prediction performance at the location of the removed observation since nearby observations provide most of the information for prediction. As a result, leave-one-out cross validation may produce residuals that are overly optimistic and fail to diagnose model problems (Isaaks and Srivastava, 1989).

The accuracy of different methods is regionally dependent. Methods that work well in regions dominated by large-scale frontal systems may not work well in regions dominated by sporadic thunderstorm activity. (Yeonsang, 2004)

5.4. RELATIONSHIP BETWEEN PRECIPITATION AND COVARIATES.

The relationship between elevation and precipitation is complex and highly variable in space, but in general, precipitation generally increases with elevation, owing to forced uplift and cooling of moisture-bearing winds by terrain barriers (Oke, 1978; Barry and Chorley, 1987). Exceptions are when the terrain rises above the height of a moist boundary layer or trade wind inversion, resulting in an increase in precipitation with elevation on lower slopes, and a rapid drying and a decrease of precipitation with elevation on the upper slopes (Mendonca and Iwaoka, 1969).
The elevation of maximum precipitation in such situations is variable, and depends on factors such as the depth of the moist boundary layer, wind speed and direction, terrain profile, and others. In Karnataka, the western Ghat creates a barrier for the moist bearing wind that comes from the Indian ocean, thus rising up to lead to rainfall. This area experiences katabatic winds, which when this wind goes over the barrier to the other side, leeward side, it has already lost its moisture and therefore it is dry and doesn’t lead to formation of clouds and rain on the leeward side. The result of this is having the leeward side of the western Ghat dry or with very little rain observed, as much as there is plenty of rain just a few kilometres away.

The process of blocking and uplifting of moisture-bearing winds amplifies precipitation on windward slopes, and can sharply decrease it on leeward slopes downwind (Daly, 1994). The resulting ‘rain shadow’ effect is common to many mountain ranges worldwide, and is especially noticeable near major moisture sources, such as oceans. These effects are most important at grid cell resolutions of less than 100 km, the scale below which major terrain features can be resolved. (Daly, 2006) (Drogue et al., 2002) also clearly points out that in general, rainfall increases with elevation in the temperate domain with non-uniform gradients and mountain ranges trigger contrasts between the windward side and the leeward side. However, these ‘laws’ remain difficult to model and must take into account: the scale of the topography-rainfall interactions, the location and the exposure characteristics of the stations in mountainous basins and the climate of the study area (Drogue et al., 2002). This is a scenario that we see in these study areas (Uttaranchal and Karnataka).

This is also the reason for the difficulty in fitting the variogram for Karnataka, July data.

5.5. MODEL PERFORMANCE WITH REFERENCE TO COVARIATES AND STUDY AREA

In the objective of the research, we seek to come up with the best method for precipitation estimation in complex orography. In our study scenario we found out that there was no universal method that could be applied to different times (climatic seasons) and different areas like Karnataka and Uttaranchal and give you accurate and precise estimation results. This is owed to the uniqueness of the study area.

Reference is made to Karnataka and Uttaranchal whereby, Karnataka has the western ghat which is a major influence of rainfall especially on the side facing the ocean (windward side). In different seasons, this area experiences different amounts of rainfall (refer to figure 5.19 and 5.23) this shows that in July and in December, the type and amount of rainfall is different and is influenced by different climatic elements. The type of elements that lead to more rainfall in July may are not necessarily the same ones that lead to rainfall in December. In July Karnataka experiences south west monsoon winds blowing from the ocean. This is a major contribution to precipitation in this region and how it is experienced in the region because of the complexity of the topography. One area can receive so much rainfall while another area a few kilometres away may not experience any rain at all.

If we use wind as a covariate in July, it may help us estimate rainfall better, but it may not perform as well in December because of absence of south west monsoon, instead we have a block of wind blowing from the Bengal area that doesn’t lead to same amount and type of precipitation received in July.

Different seasons have different covariates influencing the models and rainfall differently. This comes out clearly in the research, whereby in Karnataka for the month of April, rainfall is influenced more by elevation and assists estimate precipitation more accurately. This is not the case for July and
December. In December we have vegetation cover factor as the best covariate input into the model for best precipitation estimation output. This leads to having different models for this two months. Adding more covariates doesn’t necessarily lead to better estimation. This is also revealed by the different combinations of covariates that have been used to estimate precipitation for the different months and the two study areas.

We would therefore say that the covariates used are not the optimal precipitation estimators in the two regions. There are other climatic elements that influence precipitation that we didn’t use in the study. Basing on literature, pressure and wind would come in handy as other covariates that can be used. This is because pressure influences wind patterns which in-turn influence formation of clouds and where the rainfall will finally occur.

5.5.1. COMPARATIVE ANALYSIS OF KARNATAKA AND UTTARANCHAL

It is noted that different covariates behave different in the two states. Secondary variables that perform well in external drift kriging in Karnataka don’t necessarily perform well in helping estimate precipitation in the Uttaranchal.

Shown are elevation maps of the two states, Karnataka and Uttaranchal.

![UTTARANCHAL ELEVATION](image1)

![KARNATAKA ELEVATION](image2)

**Figure 5-60 UTTARANCHAL ELEVATION**

From the maps one can see the areas with high elevation, in dark blue. From the prediction maps its noted that these are the areas that mostly receive more amount of rainfall. A case in point is the month of April.

Karnataka state experiences a huge amount of rainfall in the month of July which is the monsoon period in India. There is sporadic rainfall, i.e. rainfall occurring in clusters where at one point there is so much rainfall yet at another point not so far, there is very little or no rainfall. This makes it hard for
a “normal” kriging with external drift to estimate precipitation accurately despite inclusion of continuous secondary variables. This can be attributed to the nature of topography in the area. Some rain gauges face the windward side and some leeward side, the rain gauges that face the windward side experience more rainfall than the ones on the leeward side.

The hills in the region are at some places and lack in others and some areas have no hills. These hills increase the complexity of the rainfall in the area, almost creating localized kind of rainfall patterns especially inside the state (Karnataka) as compared to the area close to the sea which mostly experiences convectional rainfall.

Modeling, or estimating rainfall with this kind of topography is hard, since it is hard to fit variograms, and one way would be to divide the study area into regions such that variograms are created for the different regions, then one may be able to get some better correlation between rainfall and the secondary variables like elevation, slope aspect.

From the variograms for Karnataka we see that the range is very large, this indicates the spatial dependence of the rain gauges is high too. See (table 5.6)

The spatial distribution of precipitation is likely to have been driven by wind orientation and relief (with rain occurring most on the windward side of slopes in Karnataka and least on the leeward side). (Drogue et al., 2002) clearly points out that in general, rainfall increases with elevation in the temperate domain with non-uniform gradients and mountain ranges trigger contrasts between the windward side and the leeward side. However, these 'laws' remain difficult to model and must take into account: the scale of the topography-rainfall interactions, the location and the exposure characteristics of the stations in mountainous basins and the climate of the study area (Drogue et al., 2002). This is a scenario that we see in these study areas (Uttarakhand and Karnataka).

Noting that we have two study areas, the models that perform well in Karnataka are not the same ones that are best performers in Uttarakhand, despite them being for the same month. Referring to (table 5.7) for July Karnataka, the best model only has (elevation +slope) as the best covariate and for July Uttarakhand, the best model has elevation only as the best covariate. For Karnataka the best December model has topographic index as the best covariate and for Uttarakhand December the best model has vegetation cover factor as the best covariate.

Basing on this we feel that, the topography of an area influences not just the kind and amount of precipitation experienced, but also the kind of model that stands out as the best performer in terms of accuracy in precipitation estimation. This is owed to the uniqueness of features in an area. Different regions have differing vegetation cover, arrangement of topography etc and this influences rainfall and the type of model that leads to precise and accurate precipitation estimation. In a paper written by (Yuh-Lang-lin, 2006) he mentions that quantitative forecasting of orographic precipitation remains a challenging problem for meteorologists due to complicated dynamical and microphysical processes and their interaction. In a study by (Kastelec, 2002), where she worked on Spatial Interpolation of Mean Yearly Precipitation using Universal Kriging, the analysis of residuals calculated by cross-validation, showed that the results for the eastern part of Slovenia were better than for the western mountainous part. This was attributed due to the lack of observational stations in the mountainous part of Slovenia where the spatial variability of MYP is high.

In our study, the spatial variability if high too and we have different results for the two states. Uttarakhand also experiences local convectional wind patterns that lead to “localized” rainfall. This would be hard to detect using the models,
5.5.2. **RAINFALL: WHAT DETERMINES IT AND HOW IT MOVES IN THE AREA.**

During the summer monsoon, from May to November, moisture is transported into India from the Arabian Sea and transported out to the Bay of Bengal. The transport is reversed for the rest of the year, with low activity in all segments between February and May, as expected. The total moisture advected from oceans is in phase with the total rainfall integrated over land. During the peaks of summer monsoon, the moisture from the ocean exceeds the precipitation, suggesting that moisture may move north over land. Transport of moisture out of the eastern coastline, however, occurs earlier than the transport in from the western coastal line. (Liu, 2005)

Winds blow because of differences in atmospheric pressure. Pressure gradients may develop on a local to a global scale because of differences in the heating and cooling of the Earth's surface. Heating and cooling cycles that develop daily or annually can create several common local or regional thermal wind systems. In our study area, there is difference in pressure during the two seasons and this leads to different wind systems blowing which bring about different rainfall patterns.

These wind systems create different influences on the type of precipitation received at these times in the two study areas. In Karnataka it experiences some little influence of the westerly winds (western disturbances) though there is also some convectional rainfall, as compared to Uttaranchal whose rainfall in winter is influenced totally by the westerly winds. During monsoon, the winds influence rainfall in Karnataka more, and aided by the western ghat leads to much rainfall amounts.

Source: (Sumner, 1988)

**Figure 5-62 WIND IN ASIA OVER SUMMER**

The images above show the wind patterns in Asian continent in winter and summer. In summer wind blows from the Indian ocean and from the west (Europe) during Winter.
The figures above (5-62, 5-63 and 5-64) show times when there’s low and high pressure in the Asian continent and how the wind moves. This in-turn affects precipitation in the area.

The Asiatic monsoon results from complex climatic interaction between distribution of land and water, topography, and tropical and mid-latitudinal circulation. In summer, a low pressure centre forms over northern India and northern Southeast Asia because of higher levels of received solar insolation. Warm moist air is drawn into the thermal lows from air masses over the Indian Ocean. Summer heating also causes the development of a strong latitudinal pressure gradient and the development of an easterly jet stream at an altitude of about 15 kilometres and a latitude of 25° North. The jet stream enhances rainfall in Southeast Asia, in the Arabian Sea. When autumn returns to Asia the thermal extremes between land and ocean decrease and the westerly’s of the mid-latitudes move in. The easterly jet stream is replaced with strong westerly winds in the upper atmosphere. Subsidence from an upper atmosphere cold low above the Himalayas produces outflow that creates a surface high pressure system that dominates the weather in India and Southeast Asia. (Pidwirny, 2006)

Orographic precipitation in a limited region of the central Himalaya has been documented with a dense gauge network and a variety of remote sensing techniques to investigate processes controlling precipitation distribution on the scale of 10–20 km during storm events and monsoon seasons (Barros et al., 2000; Lang and Barros, 2002, 2004; Barros and Lang, 2003). These studies have revealed large gradients in seasonal precipitation totals over short (~10 km) spatial scales that are not simply related to elevation (Barros and Lang, 2003). Additionally, they have documented differences in monsoon season (June–September) diurnal precipitation patterns at high (>2000 m) and low elevations that relate to daytime upslope winds switching to weak night-time down slope winds (Barros et al., 2000; Barros and Lang, 2003).

According to latest studies, it is not only elevation, wind vegetation cover that leads to rainfall in this area but also presence of aerosols in the air. As reported by NASA.

According to a new NASA study, a very different kind of dust, made up of small particles called aerosols, blows in from desert regions, collects in the atmosphere against the slopes of South Asia’s Tibetan Plateau during the region’s monsoon season, and helps trigger rainfall. NASA research scientist William Lau(2006) of NASA’s Goddard Space Flight Centre, Greenbelt, Md., and his team studied the aerosols. They recently found that aerosols in the form of airborne dust lofted from the
desert surface combined with black carbon from industrial emissions, bio-fuel burning, and forest fires can heat the air by absorbing the sun's radiation, which can alter the Asian monsoon cycle. Lau found that these heat-absorbing aerosols, when spun together with warm air currents and moisture, cause a heating effect in the air that triggers the rainy period earlier than usual, lengthening the monsoon season in Asia. Traditionally, aerosols have been seen as only a local environmental problem. Aerosols have not been viewed as an intervening presence in the atmosphere that could affect monsoon rains. Their study is the first to link dust aerosols to monsoon rainfall changes and to claim a specific physical mechanism in the atmosphere, whereby the tiny dust particles interact with the monsoon heat and moisture.

The mechanism operates like an "elevated heat pump," according to Lau. Increased dust or aerosols blowing in from western China, Afghanistan, Pakistan and the Middle East coupled with black carbon emissions from northern India accumulate before the monsoon in late spring against the northern and southern slopes of the Tibetan Plateau.

When the dust absorbs the sun's radiation, it heats the surface air hovering above the mountainous slopes of the region. The heated air rises, and draws warm, moist air in to northern India from the Indian Ocean, like a "heat pump." The warm moist air helps to create more rainfall. The rising motion or air associated with more rain in turn brings in more warm, moist air. The "heat pump" effect actually brings on the monsoon season prematurely in northern India, leading to a longer rainy season. The rising motion associated with the "elevated heat pump" effect will shift the monsoon's path toward the foothills of the Himalayas, meaning that more rain will fall earlier in the season (in May) in northern India as a result, and less over the Indian Ocean to the south.

All these clearly indicate to us that our area of study has more than two variables that determine rainfall amount and pattern and they are complex due to the complex nature of the mountain. These are hard to be determined by the models we came up with in our study and as seen in the results, ordinary kriging performs equally well as kriging with external drift in Uttaranchal despite using covariates in kriging with external drift.

### 5.5.3. Model Performance with Reference to Amount of Rainfall Received.

From the research we note that the month with most amount of rainfall is July, followed by April then December. Basing on the average amount of rainfall in the three months (refer to [appendix 9](#)). The month with higher error (% RMSE) (refer to table 5.8) in Uttarakhand is April and December for Karnataka.

From this we can say that the models aren’t good at estimating precipitation in months with less amount of rainfall.

### 5.5.4. Comparative Analysis of Ordinary Kriging and Kriging with External Drift.

Basing on the results of the study kriging with external drift comes out as a better algorithm for precipitation estimation in complex orography. Referring to the results for Karnataka (table 5.5) it shows that kriging with external drift models performed better by having better RMSE as compared to ordinary kriging. However these results are not satisfactory as per our expectation. We hoped to have a much bigger difference between ordinary kriging and kriging with external drift but as is with the results the difference is very small.

This is attributed to the very poor correlation between rainfall and the covariates used to come up with the model. For ordinary kriging the basic requirement is that there needs to be very good correlation.
between the covariates and the element being estimated (precipitation in our study) as indicated by (Diodato, 2005) in his study that auxiliary variables should be incorporated in the form of an external drift only if they are highly linearly correlated with the variable of interest. Otherwise it is preferable to use the method of co-kriging, which requires the fitting of a model for the cross-variograms between the different variables. (Diodato, 2005)

Looking at our correlation table (table 5.6) we see that there is very poor correlation for most of the covariates apart from (correlation for amount of rainfall for April and Elevation=0.820, April and Slope=0.590 and April and Topographic Index=0.557 and amount of rainfall for December and Elevation) in Uttaranchal.

In the table for Karnataka (table 5.1) correlation that stands out is between distance to the coast and rainfall amount in April, followed by Elevation and rainfall amount recorded in July. However, they are not good correlations since they are not above 0.5 or below -0.5. It is observed that the correlation between rainfall in the different months and the covariates is poor.

Another reason for having a small difference is because of the Search radius. If the density of data is sufficiently high when using ordinary kriging, it is possible to work with a small search radius, and this leads to a small difference between ordinary kriging and kriging with a trend for estimates within the interior of the estimation area (Journel and Rossi, 1989). This is because ordinary kriging re-estimates the local mean for each search area, and can be considered as kriging with a trend where the trend is the constant local mean. In the interior, the local mean is very similar to the trend value for a small search area. When extrapolating beyond the edge of the dataset, the algorithm has to rely increasingly on the trend, whether it is given explicitly when using kriging with a trend or calculated by ordinary kriging as the local mean for the data at the edge.

5.6. ANISOTROPY

After seeing the results for July with very high RMSE, one of the things this was attributed to is the variogram which was difficult to fit. We then came up with directional variograms to check out for anisotropy. The idea was borrowed from a study by (Wackernagel, 1994) where he analysed spatial structure of the mean temperature using directional variograms before estimating temperature using universal kriging.

5.6.1. RESULTS OF ANISOTROPY FOR JULY

DIRECTIONAL VARIOGRAM

Figure 5-65 DIRECTIONAL VARIOGRAM

ANISOTROPY MAP

Figure 5-66 ANISOTROPY MAP
From the above results we can say that, working on data from the area of study is not as easy, it would be good if one can try and carry out various things to the study area in order to carry out estimation of rainfall.

The options are for example dividing the area into regions depending on:
- The elevation.
- The aspect

After which one can then carry out estimation. Dividing the area into regions could give one a better correlation between the covariates and amount of rainfall. Where there is good strong correlation between covariates and the dependent variable one is likely to get good results. Example from the study is the correlation for April rainfall and Elevation which is 0.820 and after precipitation estimation, the error of the model for using elevation as a covariate for the month of April is 7.324, which stands out as the best for the month of April.

5.7. COMPARISON TO FINDINGS OF PREVIOUS RESEARCHES

With reference to the following Studies carried out by different researchers, we can positively say that our results are in line with the findings of other researchers.

AURHELY (Benichou and Lebreton, 1987). statistical-topographic method AURHELY model produced more erroneous results for rainy months as compared to PLUVIA in a study carried out by G drogue (2002) where he came up with a model called PLUVIA. (Refer to literature review) (Valesco, 2004) worked at a study: “merging radar and rain gauge data to estimate rainfall fields: an improved geostatistical approach using non-parametric spatial models” and external drift came out to be the best among ordinary kriging and collocated cokriging.

Wackernagel investigates the application of kriging with external drift to mapping January mean temperatures in Scotland. In his study kriged estimates using elevation as an external drift variable yielded much more detail than universal kriging using only the temperature data. (Bourennane and King, 2003) results from his study Using multiple external drifts to estimate a soil variable showed that kriging using two external drifts proved to be less unbiased and more precise compared to kriging using only one external drift variable in predicting the target variable.

5.8. APPROPRIATENESS/ USEFULLNES OF SELECTED MODELS.

The selected method was appropriate BUT as much as it performs better than ordinary kriging, it still is not “perfect” because for perfection it should estimate precipitation and have an error that is as small as possible.

In our study areas there is presence of many features, citing an example of Karnataka, it is close to the coast, and has the western ghat and presence of mountains. Basing on this we agree with (Daly, 2006) where he says regions having significant terrain features, and also significant coastal effects, rain shadows, or cold air drainage and inversions are best handled by sophisticated systems that are configured and evaluated by experienced climatologists (Daly, 2006)

Among the things this can be attributed to is very poor correlation between our covariates and the rainfall amount recorded. As (Burrough, 2001) says, kriging errors depend on the form of the variogram and the disposition of observations-the more data surrounding an unsampled location and the stronger the autocorrelation structure. The lower the estimation variance.

In our study it shows that the models are useful when:
- There is good correlation between the element being estimated and the covariate.
- When there is moderate amount of rainfall.
In a study carried out by Ebrahim Hosseini et al. they concluded that the regression coefficient between rainfall and elevation must be greater than 0.6 in order to improve interpolation accuracy using elevation as co variable, basing on this would also support noting that our correlations were very low, and having these higher may have improved our estimation accuracy. This leads to us having a very small difference between precipitation estimation in the region using ordinary kriging and using kriging with external drift.

5.9. APPLICATION OF RESULTS

The results are more widely applicable since Kriging with external drift is not tied to precipitation estimation only but can be used to estimate other elements like temperature Wackernagel (1994), air quality (Kassteele, 2006), soil, (Bourennane and King, 2003)
These results can contribute to Geostatistics related problems by guiding researches in knowing the covariates that are better estimators of precipitation (and other climatic elements like temperature, air) in complex terrain and giving ideas of which other areas are explorable and likely to give more in-depth revelations or improving techniques for precipitation estimation in complex orography to perfection.
The results shown in this paper will help guide the selection of appropriate sites for placing rain gauges so that it can aid in reducing errors during precipitation estimation in complex terrain.
6. CONCLUSION AND RECOMMENDATION

This chapter winds up the study by referring to the objective, research questions and the hypothesis that we had to answer using the study.

6.1. EFFECT OF ADDING OR DROPPING COVARIATES ON INTERPOLATION.

Adding or dropping covariates leads to under and overestimation of precipitation. The covariate that increases performance of models when added is elevation. Performance of models is based on RMSE. Refer to table in appendix 7.

In Karnataka and Uttarakhal the optimal number of estimators in April was 4, July 2 and December 2 respectively. After the optimal number increasing the number of parameters leads to higher error.

6.2. THE MOST RELEVANT FACTORS

Based on the research we felt that the most relevant factors that contribute to amount of precipitation in complex terrain were: (These are in order of performance basing on RMSE).

1. Elevation.
2. Slope.
3. Vegetation Cover
4. Topographic index.

Referring to tables in appendix 7 and 8, the study reveals that the covariates that perform better are as listed above in order of better performance. It should also be noted that topographic index is a derived covariate. It is derived from elevation and vegetation cover, and these two are still rated as good precipitation estimators.

Elevation does influence the type and amount of precipitation in complex orography as shown in the two models with best performance in Karnataka. (this is detailed in chapter 5). However it’s not always the case because of other topographic and climatic elements like wind that influence precipitation too. Referring to a study by (Yuh-lang Lin 2006.) he reveals that an orographic precipitating system may form in various ways depending upon ambient flow speed, vertical stability and wind structures, mountain height, horizontal scale and geometry, synoptic and mesoscale environments, etc.

In our study it is found that, these covariates change depending on the season and study area. Elevation helps estimate precipitation better in April and July in Karnataka and not in December and it helps estimate precipitation better in April in Uttarakhal. In Uttarakhal for July and December vegetation cover and elevation influence precipitation more.

This is based on the results that were found. Referring to the table 5.7 with their RMSE and ME. In this table it shows that elevation is a good covariate in July but not in December.
6.3. **BEST MODEL FOR ESTIMATING PRECIPITATION IN COMPLEX TERRAIN.**

In the study we expected to come up with one model that would perform well in terms of accurate precipitation estimation in the two study areas, however through the study we come up with six different models depending on the climatic season and the area.

External drift kriging algorithm models have better performance in Karnataka. We used elevation as a covariate in Karnataka for the month of July where the model is the best but the same model doesn’t come out as the best for July in Uttaranchal instead it performs better in April.

A detailed discussion can be found in chapter 5 and reference to table in appendix 7 and 8. In our study areas there is presence of many features, citing an example of Karnataka, it is close to the coast, and has the western ghat and presence of mountains. Basing on this we agree with (Daly, 2006) where he says Regions having significant terrain features, and also significant coastal effects, rain shadows, or cold air drainage and inversions are best handled by sophisticated systems that are configured and evaluated by experienced climatologists (Daly, 2006)

6.4. **HOW DOES THE MODEL BEHAVE IN A DIFFERENT AREA.**

The models perform differently in the two study areas. The model which performs best in April for Karnataka doesn’t perform best in Uttaranchal. Basing on the percentage RMSE model with elevation that performs best in Karnataka July with 79.94% RMSE has 28% RMSE but is not the best for July instead model with a combination of slope, aspect, vegetation cover factor and Topographic Index emerges as the best with 26.3% RMSE.

6.5. **HYPOTHESIS:**

Kriging with external drift performs better in one study area (Karnataka)
Based on RMSE for Karnataka, we can say that kriging with external drift performs better in this study area. The RMSE figures were normalised and the percentage shows the percentage RMSE.

**PERFORMANCE COMPARISON OF ORDINARY AND EXTERNAL DRIFT KRIGING.**

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>MEAN ERROR</th>
<th>%</th>
<th>RMSE</th>
<th>MEAN ERROR</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>APRIL</td>
<td>28.71</td>
<td>-0.108</td>
<td>83.0488863</td>
<td>27.475</td>
<td>0.059</td>
<td>79.4764</td>
</tr>
<tr>
<td>JULY</td>
<td>179.605</td>
<td>3.302</td>
<td>80.0022272</td>
<td>179.472</td>
<td>0.546</td>
<td>79.943</td>
</tr>
<tr>
<td>DECEMBER</td>
<td>11.696</td>
<td>-0.432</td>
<td>135.213873</td>
<td>10.834</td>
<td>0.08</td>
<td>125.249</td>
</tr>
</tbody>
</table>

**TABLE 6.1**

Kriging errors depend on the form of the variogram and the disposition of observations-the more data surrounding an unsampled location, and the stronger the autocorrelation structure. The lower the estimation variance. (Burrough, 2001)
6.6. LIMITATIONS OF THE STUDY.

During the study it was very hard to get good and consistent daily or monthly data for Uttaranchal, and this led to the use of a 50 year average rainfall.

Early in the study we acquired Wind, Pressure and Temperature data from NCEP re-analysis however this was not used due to very poor(low) resolution of 260km pixel size. Having discovered that wind is a crucial climatic element in determination of precipitation amount and pattern in the study areas, would have been beneficial if it were included in the study.

6.7. RECOMMENDATIONS.

The spatial distribution of precipitation is likely to have been driven by wind orientation and relief (with rain occurring most on the windward side of slopes and least on the leeward side).

(Drogue et al., 2002) also clearly points out that in general, rainfall increases with elevation in the temperate domain with non-uniform gradients and mountain ranges trigger contrasts between the windward side and the leeward side. However, these 'laws' remain difficult to model and must take into account: the scale of the topography-rainfall interactions, the location and the exposure characteristics of the stations in mountainous basins and the climate of the study area. This is a scenario that we see in these study areas (Uttaranchal and Karnataka).

It would be more beneficial in terms of emerging with more accurate models if a study can be carried on with climatic variables (covariates) like wind and pressure for this study area, noting that wind is a big contributor to type, amount and occurrence of rainfall in the two study areas, especially for the months of July and December.

Other options would also be to divide the area depending on height or aspect then be able to study these different regions separately, especially if the individual regions would have better correlation between the covariates and precipitation.

This brings the need to have a model that takes all these climatic variables into consideration. As (Miller, 2005) mentions in her study that including spatial dependence explicitly in a predictive model can be an efficient way to improve model accuracy with the available data.

Bayesian geostatistics are described by (Handcock and Stein, 1993) and can be used in the geo-r package of R statistics. This can be one way of extending the research.

We do agree with (Bahadur, 1999) that there is need to establish a project for studying the Himalayas and have more meteorological data over the Himalayas for better understanding of the weather systems in the area.
7. REFERENCES


ONLINE REFERENCES

http://www.ias.sdms.edu/Orville%20Symposium/Abstracts/Y_Lin.pdf
  Date of access 8th July 2006
  Date of access 22nd December 2006
  Date of access 22nd December 2006
http://www.physicalgeography.net/fundamentals/7o.html
  Date of access 22nd December 2006
http://www.physicalgeography.net/fundamentals/2f.html
  Date of access 22nd December 2006
http://www.nasa.gov/vision/earth/environment/asian_rain.html
  Date of access 22nd December 2006
  Date of access 22nd December 2006
  Date of access 22nd December 2006
http://ugamp.nerc.ac.uk/promise/research/conferences/trieste/mayabstract.pdf
  Date of access 22nd December 2006
http://www.geovista.psu.edu/sites/geocomp99/Gc99/023/ge_023.htm
  Date of access 2nd March Oct 2006
http://www.geog.ucsb.edu/%7Egood/176b/a15.html
  Date of access 3rd October 2006
http://www.personality-project.org/r/
  Date of access 18th July 2006
  Date of access 1st September 2006
8. GLOSSARY

**Lag distance** is a distance \( h \) within which any two samples (a pair) is taken for variogram calculation.

**Lag tolerance** \( \Delta \) is assigned to the lag for setting a search interval for pairs: \( \pm \Delta \). So the pairs are searched within \( h \pm \Delta \) distance. Often, default lag tolerance is set to half lag distance \( \Delta = h/2 \).

**Kriging** is a linear regression prediction using data of the same attribute as that being estimated.

**Ordinary kriging** (OK) is a type of kriging, which is performed with unknown mean and requires that the sum of kriging weights equals to one.

**Simple kriging** (SK) is a type of kriging with a constant known mean. Stationary simple kriging has a constant mean \( m \). Non-stationary simple kriging allows a prior determination of locally varying mean \( m(u) \).

**Cokriging** is a linear regression that uses data defined on different attributes. Simple cokriging is cokriging with no constraint on weights and requiring means for primary and secondary variables.

**Standardized ordinary cokriging** is a type of cokriging consisting of creating new secondary variables with the same mean as the primary variable. Then all the weights are constrained to one.

**Traditional ordinary cokriging** is a type of cokriging with the sum of weights applied to the primary variable set to one, and the sum of weights applied to the secondary variables set to zero. The last condition tends to limit severely the influence of the secondary variables.

**Kriging with trend** (Universal kriging, KT) is a type of kriging with a prior trend model used. A trend model is defined as a smoothly varying deterministic function.

**Kriging with an external drift** is an extension of kriging with trend. The trend model is limited to two terms \( m(u)=a_0+a_1 f_1(u) \) where \( f_1(u) \) is set to the secondary (external drift) variable.
9. APPENDIX

9.1. APPENDIX 1 ASPECT TRANSFORMATION TABLE.
9.2. **APPENDIX 2 - SCRIPT FOR VARIOGRAMS**

```r
> va <- variogram(Apr ~ 1, ~X+Y, utta, cutoff=180000, width=8000)
> plot(va)
8000 is bin size,
180000 is the range,
>vma <- vgm(270,"Lin", 190000,38)
> plot(va, vma)
270 is partial sill, 190000 is range 38 is nugget
> plot(va, plot=T, model=vma)

Fitted variogram
>(vmaf <- fit.variogram(va, vma))
  model psill range
1 Nug 38 0
2 Lin 270 190000

> plot(va, plot=T, model=vmaf)
> vmaf <- fit.variogram(va, vma)
> plot(va, model=vmaf)
> plot(va, model=vmaf, plot=T)
```

9.3. **APPENDIX 3 - SCRIPT FOR ORDINARY KRIGING**

```r
> okd <- krige(Dec ~ 1, utta, newdata=utta.grid, model=vmdf)
> source("CODE/ok_plotfns.R")
> plot.kresults(okd, "var1", utta, "Dec")

> okdcv <- krige.cv(Dec ~ 1, utta, model=vmdf)
> str(okcv)
> bubble(okdcv, "residual")

> okdres <- as.data.frame(okdcv)$residual
> hist(okdres,breaks=20,col="lightblue",border="red", main="Residuals for December")
> sqrt(mean(okdres^2))
> mean(okdres)
```

```r
### R source code, file ok_plotfns.R
### author: DG Rossiter, ITC Enschede (NL) rossiter@itc.nl
### load to workspace with command > source("ok_plotfns.R")
### plot Kriging results: predictions and errors
## with postplot of sample points on predictions
## colour: bpy, grey
## and locations of sample points on errors
```
## Arguments

- `kr.o`: SpatialPointsDataFrame from kriging
- `coordinates` must be named `x`, `y`
- `var1`: prefix for field names `*.pred`, `*.var`, e.g. `"var1"`
- `samp.pts`: SpatialPointsDataFrame sample points
- `f`: field name (quoted) or number for point data values
- `title`: plot.kresults <- function(kr.o, var1, uttagrid, 
f=1, title="") {
  to.eval <- paste("plot.kr <- levelplot(" ,
  paste(var1,"pred",sep=".", " ~ X+Y, as.data.frame(kr.o),
  aspect='iso',
  col.regions=bpy.colors(64), cut=32,
  main=paste(title, 'Prediction'),
  panel = function(x, ...) {
    panel.levelplot(x, ...);
    panel.points(coordinates(uttagrid), col='grey', pch=20, 
    # log scale, but still show minimum
    cex=1.6 * (log10(uttagrid[[f]]) - 
    0.9 * min(log10(uttagrid[[f]]))));
    panel.grid(h=-1, v=-1, col='darkgrey')
  })" );
  eval(parse(text=to.eval));
  to.eval <- paste("plot.kr.e <- levelplot(" ,
  paste(var1,"var",sep=".", " ~ X+Y, as.data.frame(kr.o),
  aspect='iso',
  col.regions=cm.colors(64), cut=32,
  main=paste(title, 'Prediction errors'),
  panel = function(x, ...) {
    panel.levelplot(x, ...);
    panel.points(coordinates(uttagrid), col='green', pch=20,
    cex=0.6);
    panel.grid(h=-1, v=-1, col='darkgrey')
  })" );
  eval(parse(text=to.eval));
  print(plot.kr, split = c(1,1,2,1), more=T);
  print(plot.kr.e, split = c(2,1,2,1), more=F)
}

### Appendix 4 - Script for Kriging with External Drift

```r
>Avre<- variogram(Apr~elevation,loc=utta);plot(Avre, plot.numbers=T)
>(Amre<-fit.variogram(Avre,vgm(270, "Lin", 190000,38)))
```
ANALYSIS OF EXTERNAL DRIFT KRIGING ALGORITHM WITH APPLICATION TO PRECIPITATION ESTIMATION IN COMPLEX OROGRAPHY.

model psill range
1 Nug 38 0
2 Lin 270 190000

> Akre<-krige(Apr~elevation,loc=utta,newdata=uttagrid,model=Amre)
[using universal kriging]
> levelplot(var1.pred~X+Y, as.data.frame(Akre),aspect="iso")
> source("CODE/ked_plotfserr.R")
> plot.kederresults(Akre, "var1", utta, "Apr")

### R source code, file ked_plotfserr.R
### author: DG Rossiter, ITC Enschede (NL) rossiter@itc.nl
### load to workspace with command > source("ked_plotfserr.R")
### plot Kriging results: predictions and errors
### with postplot of sample points on predictions
### colour: bpy, grey
### and locations of sample points on errors
### colour: cm, green
### arguments
### Akre SpatialPointsDataFrame from kriging
### coordinates must be named x, y
### var1 prefix for field names *.pred, *.var, e.g. "var1"
### samp.pts SpatialPointsDataFrame sample points
### f field name (quoted) or number for point data values
### title
plot.kederresults <- function(Akre, var1, uttagrid,
f=1, title="") {
  to.eval <- paste("plot.ked <- levelplot(",
paste(var1,"pred",sep="."),
  " ~ X+Y, as.data.frame(Akre),
  aspect='iso',
  col.regions=bpy.colors(64), cut=32,
  main=paste(title, 'Prediction'),
  panel = function(x, ...) {
    panel.levelplot(x, ...);
    panel.points(coordinates(uttagrid), col='grey', pch=20,
      # log scale, but still show minimum
      cex=1.6 * (log10(uttagrid[[f]]) -
      0.9 * min(log10(uttagrid[[f]]))),)
    panel.grid(h=-1, v=-1, col='darkgrey')
  })
  eval(parse(text=to.eval));
to.eval <- paste("plot.keder <- levelplot(",
paste(var1,"var",sep="."),
  " ~ X+Y, as.data.frame(Akre),
  aspect='iso',
  col.regions=cm.colors(64), cut=32,
ANALYSIS OF EXTERNAL DRIFT KRIGING ALGORITHM WITH APPLICATION TO PRECIPITATION ESTIMATION IN COMPLEX OROGRAPHY.

```r
main=paste(title, 'Prediction errors'),
panel = function(x, ...) {
  panel.levelplot(x, ...);
  panel.points(coordinates(uttagrid), col='green', pch=20,
  cex=.6);
  panel.grid(h=-1, v=-1, col='darkgrey')
}

eval(parse(text=to.eval));
print(plot.ked, split = c(1,1,2,1), more=T);
print(plot.keder, split = c(2,1,2,1), more=F)
```

9.5. APPENDIX 5

```r
>summary(Akrs5)
Object of class SpatialPointsDataFrame
Coordinates:
  min       max
X  -96985.19  399014.8
Y  -3796215.00 -3060215.0
Is projected: NA
proj4string: [NA]
Number of points: 47576
Data attributes:
  var1.pred    var1.var
Min. : -12.670 Min. : 968.6
1st Qu.:  8.636 1st Qu.: 974.2
Median : 19.704 Median : 983.0
Mean : 30.291 Mean : 1001.4
3rd Qu.: 47.400 3rd Qu.: 1002.5
Max. : 121.875 Max. : 3946.7
```

9.6. APPENDIX 6 SOFTWARE USED

9.6.1. R 2.4.0

9.6.1.1. Functionality

Functions can be defined easily and then stored for later use.
- Packages of functions are written by “all of us” to solve particular problems
- e.g. gstat, geo-R, lattice graphics
- Packages can be developed and tailored for a specific lab or problem area, or can be shared through the CRAN repository of (currently) > 400 packages
9.6.1.2. Graphics capability

Doesn’t have graphic capability as compared to Arc-GIS but has some impressive graphs too. Sample graphics taken from personal sites of very impressive graphs done using R
http://addictedtor.free.fr/graphiques/

9.6.1.3. Strengths

Graphics for data exploration and interpretation
• Data manipulation including statistics as data
• Statistical analysis
• Standard univariate and multivariate generalizations of the linear model
• Multivariate-structural extensions (Revelle, 2005) www.personality-project.org/r/

The open source gstat extension package makes the S environment (the R or S plus programs) a very powerful environment for (multivariable) geostatistics.

The package offers several methods for handling one or more exhaustive grids of secondary information for prediction or simulation of primary variable. Secondary variable can be treated as explanatory or predictor variables, leading to linear regression or universal kriging prediction (sometimes referred to as external drift kriging) (Pebesma, 2004)

9.6.1.4. Weakness

A user must decide on the sequence of analyses and execute them step-by-step. However, it is easy to create scripts with all the steps in an analysis, and run the script from the command line or menus

The user must learn a new way of thinking about data, as data frames and objects each with its class, which in turn supports a set of methods. This has the advantage common to object-oriented languages that you can only operate on an object according to methods that make sense and methods can adapt to the type of object. (Rossiter, 2006) http://www.itc.nl/personal/rossiter

9.6.1.5. Availability

R is a project which is attempting to provide a modern piece of statistical software for the GNU suite of software. It’s an international collaboration where there are many contributors to different packages that can be loaded into R.

• R is open source software, public domain version of S+
• R: written by different Statisticians but Initially written by Robert Gentleman and Ross Ihaka of the Statistics Department of the University of Auckland.

R was inspired by the S environment which has been principally developed by John Chambers, with substantial input from Douglas Bates, Rick Becker, Bill Cleveland, Trevor Hastie, Daryl Pregibon and Allan Wilks.

The current version of R is 2.4.0 (2006-10-03) Copyright (C) 2006 The R Foundation for Statistical Computing ISBN 3-900051-07-0.

In R, the package Gstat (Pebesma, E.J., 2004) was used for processing of the data.

9.6.2. ARC - GIS 9.1

Arc-GIS Desktop is a collection of software products that runs on standard desktop computers. It is used to create, import, edit, query, map, analyze, and publish geographic information. There are four products
in the Arc-GIS Desktop collection; each adds a higher level of functionality: Arc-Reader; Arc-View; Arc-Editor; and Arc-Info.

9.6.2.1. Functionality

All Arc-GIS Desktop products share a common architecture, so users working with any of these GIS desktops can share their work with others. Maps, data, symbology, map layers, geo-processing models, custom tools and interfaces, reports, metadata, and so on, can be accessed interchangeably.

9.6.2.2. Graphics capability

The graphics generated by the Arc-GIS packages are relatively straightforward to produce and are of high quality.

9.6.2.3. Strengths

Advanced spatial analysis, data manipulation, and high-end cartography tools. Professional GIS users use Arc-Info for all aspects of data building, modelling, analysis, and map display. The Arc-GIS products are relatively easy to use because of their menu/dialog-box driven nature. Users are essentially walked through the process of performing a spatial interpolation.

9.6.2.4. Weaknesses

Users are limited to the functionality provided within the dialog boxes that drive the various features. As much as it can be customised using Arc-objects, some of the algorithms embedded cannot be manipulated to certain extents to perform non standard analysis as seen in R Statistics software.

9.6.2.5. Statistics capability

It does handle statistics but because of their menu/dialog-box driven nature. Users are essentially walked through the process of performing a spatial interpolation and are not able to manipulate a number of things in the algorithms to suit different elements in customised studies.

9.6.2.6. Availability

It is not open source and is available as commercial software.

9.7. APPENDIX 7 TABLE OF MODELS FOR UTTARANCHAL.

<table>
<thead>
<tr>
<th>MODEL</th>
<th>APRIL RMSE</th>
<th>ME</th>
<th>%</th>
<th>JULY RMSE</th>
<th>ME</th>
<th>%</th>
<th>DECEMBER RMSE</th>
<th>ME</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevation</td>
<td>7.324</td>
<td>0.075</td>
<td>32.236</td>
<td>134.514</td>
<td>-1.473</td>
<td>28.451</td>
<td>5.502</td>
<td>-0.382</td>
<td>27.156</td>
</tr>
<tr>
<td>Slope</td>
<td>8.040</td>
<td>-0.288</td>
<td>35.388</td>
<td>126.747</td>
<td>-1.759</td>
<td>26.808</td>
<td>5.312</td>
<td>-0.484</td>
<td>26.220</td>
</tr>
<tr>
<td>Aspect</td>
<td>7.658</td>
<td>-0.273</td>
<td>33.706</td>
<td>133.215</td>
<td>-3.042</td>
<td>28.176</td>
<td>4.980</td>
<td>-0.468</td>
<td>24.581</td>
</tr>
<tr>
<td>vcf</td>
<td>7.672</td>
<td>-0.379</td>
<td>33.768</td>
<td>131.290</td>
<td>-3.195</td>
<td>27.769</td>
<td>4.589</td>
<td>-0.577</td>
<td>22.648</td>
</tr>
<tr>
<td>TI</td>
<td>7.898</td>
<td>-0.400</td>
<td>34.763</td>
<td>133.113</td>
<td>-2.184</td>
<td>28.154</td>
<td>4.960</td>
<td>-0.681</td>
<td>24.481</td>
</tr>
<tr>
<td>elsl</td>
<td>7.628</td>
<td>0.115</td>
<td>33.572</td>
<td>127.121</td>
<td>-2.009</td>
<td>26.887</td>
<td>5.977</td>
<td>-0.563</td>
<td>29.499</td>
</tr>
<tr>
<td>elta</td>
<td>7.479</td>
<td>0.132</td>
<td>32.919</td>
<td>137.432</td>
<td>-2.357</td>
<td>29.068</td>
<td>5.688</td>
<td>-0.339</td>
<td>28.074</td>
</tr>
<tr>
<td>elvcf</td>
<td>7.433</td>
<td>0.064</td>
<td>32.714</td>
<td>136.011</td>
<td>-2.611</td>
<td>28.767</td>
<td>5.333</td>
<td>-0.428</td>
<td>26.324</td>
</tr>
</tbody>
</table>
### Analysis of External Drift Kriging Algorithm with Application to Precipitation Estimation in Complex Orography

<table>
<thead>
<tr>
<th>Model</th>
<th>April RMSE</th>
<th>April ME</th>
<th>April %</th>
<th>July RMSE</th>
<th>July ME</th>
<th>July %</th>
<th>December RMSE</th>
<th>December ME</th>
<th>December %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevation</td>
<td>27.487</td>
<td>-0.029</td>
<td>79.512</td>
<td>179.473</td>
<td>0.546</td>
<td>79.943</td>
<td>10.882</td>
<td>-0.046</td>
<td>125.799</td>
</tr>
<tr>
<td>Slope</td>
<td>28.367</td>
<td>-0.104</td>
<td>82.056</td>
<td>180.776</td>
<td>-0.121</td>
<td>80.524</td>
<td>10.998</td>
<td>-0.117</td>
<td>127.150</td>
</tr>
<tr>
<td>Aspect</td>
<td>28.246</td>
<td>-0.104</td>
<td>81.706</td>
<td>180.945</td>
<td>0.133</td>
<td>80.599</td>
<td>10.920</td>
<td>-0.064</td>
<td>126.243</td>
</tr>
<tr>
<td>vcf</td>
<td>28.190</td>
<td>-0.202</td>
<td>81.544</td>
<td>195.416</td>
<td>-4.970</td>
<td>87.045</td>
<td>10.887</td>
<td>0.002</td>
<td>125.856</td>
</tr>
<tr>
<td>TI</td>
<td>28.892</td>
<td>-0.534</td>
<td>83.575</td>
<td>190.118</td>
<td>-3.976</td>
<td>84.685</td>
<td>10.834</td>
<td>-0.080</td>
<td>125.245</td>
</tr>
<tr>
<td>dc</td>
<td>28.252</td>
<td>-0.097</td>
<td>81.724</td>
<td>181.264</td>
<td>-0.083</td>
<td>80.741</td>
<td>10.869</td>
<td>-0.059</td>
<td>125.657</td>
</tr>
<tr>
<td>eI</td>
<td>27.475</td>
<td>0.059</td>
<td>79.478</td>
<td>181.341</td>
<td>0.278</td>
<td>80.776</td>
<td>10.982</td>
<td>-0.102</td>
<td>126.961</td>
</tr>
</tbody>
</table>

el=elevation
sl=slope
ta=transformed aspect
vcf=vegetation cover factor
TI=Topographic index.

### 9.8. Appendix 8 Table of Models for Karnataka

<table>
<thead>
<tr>
<th>Model</th>
<th>April RMSE</th>
<th>April ME</th>
<th>April %</th>
<th>July RMSE</th>
<th>July ME</th>
<th>July %</th>
<th>December RMSE</th>
<th>December ME</th>
<th>December %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevation</td>
<td>27.487</td>
<td>-0.029</td>
<td>79.512</td>
<td>179.473</td>
<td>0.546</td>
<td>79.943</td>
<td>10.882</td>
<td>-0.046</td>
<td>125.799</td>
</tr>
<tr>
<td>Slope</td>
<td>28.367</td>
<td>-0.104</td>
<td>82.056</td>
<td>180.776</td>
<td>-0.121</td>
<td>80.524</td>
<td>10.998</td>
<td>-0.117</td>
<td>127.150</td>
</tr>
<tr>
<td>Aspect</td>
<td>28.246</td>
<td>-0.104</td>
<td>81.706</td>
<td>180.945</td>
<td>0.133</td>
<td>80.599</td>
<td>10.920</td>
<td>-0.064</td>
<td>126.243</td>
</tr>
<tr>
<td>vcf</td>
<td>28.190</td>
<td>-0.202</td>
<td>81.544</td>
<td>195.416</td>
<td>-4.970</td>
<td>87.045</td>
<td>10.887</td>
<td>0.002</td>
<td>125.856</td>
</tr>
<tr>
<td>TI</td>
<td>28.892</td>
<td>-0.534</td>
<td>83.575</td>
<td>190.118</td>
<td>-3.976</td>
<td>84.685</td>
<td>10.834</td>
<td>-0.080</td>
<td>125.245</td>
</tr>
<tr>
<td>dc</td>
<td>28.252</td>
<td>-0.097</td>
<td>81.724</td>
<td>181.264</td>
<td>-0.083</td>
<td>80.741</td>
<td>10.869</td>
<td>-0.059</td>
<td>125.657</td>
</tr>
<tr>
<td>eI</td>
<td>27.475</td>
<td>0.059</td>
<td>79.478</td>
<td>181.341</td>
<td>0.278</td>
<td>80.776</td>
<td>10.982</td>
<td>-0.102</td>
<td>126.961</td>
</tr>
</tbody>
</table>
### Analysis of External Drift Kriging Algorithm with Application to Precipitation Estimation in Complex Topography

<table>
<thead>
<tr>
<th>eltavf</th>
<th>27.644</th>
<th>-0.165</th>
<th>79.964</th>
<th>196.868</th>
<th>-4.268</th>
<th>87.700</th>
<th>10.975</th>
<th>0.000</th>
<th>126.882</th>
</tr>
</thead>
<tbody>
<tr>
<td>eltaT</td>
<td>27.630</td>
<td>-0.146</td>
<td>79.924</td>
<td>194.660</td>
<td>-3.935</td>
<td>86.708</td>
<td>10.965</td>
<td>-0.005</td>
<td>126.769</td>
</tr>
<tr>
<td>eltadc</td>
<td>27.617</td>
<td>-0.037</td>
<td>79.888</td>
<td>182.647</td>
<td>0.517</td>
<td>81.357</td>
<td>10.969</td>
<td>-0.049</td>
<td>126.809</td>
</tr>
<tr>
<td>elvcfT</td>
<td>27.599</td>
<td>-0.172</td>
<td>79.834</td>
<td>188.664</td>
<td>-3.950</td>
<td>84.037</td>
<td>10.903</td>
<td>-0.020</td>
<td>126.048</td>
</tr>
<tr>
<td>elvcf</td>
<td>27.487</td>
<td>-0.036</td>
<td>79.512</td>
<td>195.540</td>
<td>-4.364</td>
<td>87.100</td>
<td>10.994</td>
<td>0.056</td>
<td>127.097</td>
</tr>
<tr>
<td>sltaT</td>
<td>27.552</td>
<td>-0.261</td>
<td>82.592</td>
<td>197.408</td>
<td>-4.812</td>
<td>87.932</td>
<td>11.124</td>
<td>-0.058</td>
<td>128.603</td>
</tr>
<tr>
<td>sltaT</td>
<td>27.592</td>
<td>-0.657</td>
<td>85.601</td>
<td>192.195</td>
<td>-3.777</td>
<td>85.610</td>
<td>10.603</td>
<td>-0.147</td>
<td>127.897</td>
</tr>
<tr>
<td>sltaT</td>
<td>28.583</td>
<td>-0.116</td>
<td>82.683</td>
<td>184.720</td>
<td>-0.105</td>
<td>82.280</td>
<td>11.103</td>
<td>-0.123</td>
<td>128.358</td>
</tr>
<tr>
<td>sltaT</td>
<td>28.163</td>
<td>-0.252</td>
<td>81.466</td>
<td>198.067</td>
<td>-4.089</td>
<td>88.226</td>
<td>10.927</td>
<td>-0.020</td>
<td>126.319</td>
</tr>
<tr>
<td>sltaT</td>
<td>28.365</td>
<td>-0.126</td>
<td>82.050</td>
<td>197.926</td>
<td>-4.787</td>
<td>88.163</td>
<td>11.077</td>
<td>0.052</td>
<td>128.054</td>
</tr>
<tr>
<td>sltaT</td>
<td>29.010</td>
<td>-0.515</td>
<td>83.916</td>
<td>193.984</td>
<td>-4.018</td>
<td>86.407</td>
<td>10.919</td>
<td>-0.059</td>
<td>126.237</td>
</tr>
<tr>
<td>sltaT</td>
<td>27.969</td>
<td>-0.134</td>
<td>80.904</td>
<td>197.633</td>
<td>-4.284</td>
<td>88.033</td>
<td>10.903</td>
<td>0.033</td>
<td>126.051</td>
</tr>
<tr>
<td>elsitavf</td>
<td>27.734</td>
<td>-0.175</td>
<td>80.226</td>
<td>197.794</td>
<td>-4.379</td>
<td>88.104</td>
<td>11.072</td>
<td>-0.067</td>
<td>127.997</td>
</tr>
<tr>
<td>elsitavf</td>
<td>27.731</td>
<td>-0.172</td>
<td>80.218</td>
<td>195.407</td>
<td>-3.995</td>
<td>87.041</td>
<td>11.060</td>
<td>-0.074</td>
<td>127.856</td>
</tr>
<tr>
<td>elsitavf</td>
<td>27.602</td>
<td>0.055</td>
<td>79.844</td>
<td>184.448</td>
<td>0.296</td>
<td>82.160</td>
<td>11.072</td>
<td>-0.106</td>
<td>127.994</td>
</tr>
<tr>
<td>elsitavf</td>
<td>27.720</td>
<td>-0.197</td>
<td>80.184</td>
<td>190.202</td>
<td>-3.782</td>
<td>84.722</td>
<td>10.979</td>
<td>-0.020</td>
<td>126.928</td>
</tr>
<tr>
<td>elsitavf</td>
<td>27.685</td>
<td>-0.212</td>
<td>80.083</td>
<td>189.497</td>
<td>-4.187</td>
<td>84.408</td>
<td>10.995</td>
<td>-0.094</td>
<td>127.115</td>
</tr>
<tr>
<td>elsitavf</td>
<td>28.451</td>
<td>-0.303</td>
<td>82.298</td>
<td>199.151</td>
<td>-4.363</td>
<td>88.709</td>
<td>11.030</td>
<td>-0.093</td>
<td>127.518</td>
</tr>
<tr>
<td>elsitavf</td>
<td>28.603</td>
<td>-0.175</td>
<td>82.740</td>
<td>198.521</td>
<td>-4.758</td>
<td>88.428</td>
<td>11.252</td>
<td>-0.006</td>
<td>130.085</td>
</tr>
<tr>
<td>elsitavf</td>
<td>29.570</td>
<td>-0.630</td>
<td>85.537</td>
<td>194.568</td>
<td>-3.955</td>
<td>86.667</td>
<td>11.083</td>
<td>-0.123</td>
<td>128.129</td>
</tr>
<tr>
<td>elsitavf</td>
<td>28.076</td>
<td>-0.153</td>
<td>81.215</td>
<td>199.161</td>
<td>-4.113</td>
<td>88.713</td>
<td>10.979</td>
<td>0.033</td>
<td>126.928</td>
</tr>
<tr>
<td>elsitavf</td>
<td>28.814</td>
<td>-0.241</td>
<td>80.458</td>
<td>191.128</td>
<td>-4.017</td>
<td>85.135</td>
<td>11.073</td>
<td>-0.095</td>
<td>128.009</td>
</tr>
<tr>
<td>elsitavf</td>
<td>27.711</td>
<td>-0.119</td>
<td>80.159</td>
<td>190.976</td>
<td>-3.958</td>
<td>85.067</td>
<td>11.104</td>
<td>-0.035</td>
<td>128.365</td>
</tr>
<tr>
<td>elsitavf</td>
<td>27.610</td>
<td>-0.064</td>
<td>79.868</td>
<td>195.499</td>
<td>-3.875</td>
<td>87.082</td>
<td>11.122</td>
<td>-0.027</td>
<td>128.579</td>
</tr>
<tr>
<td>elsitavf</td>
<td>27.598</td>
<td>-0.056</td>
<td>79.833</td>
<td>197.904</td>
<td>-4.293</td>
<td>88.153</td>
<td>11.160</td>
<td>-0.011</td>
<td>129.018</td>
</tr>
<tr>
<td>elsitavf</td>
<td>28.275</td>
<td>-0.194</td>
<td>81.791</td>
<td>200.171</td>
<td>-4.366</td>
<td>89.163</td>
<td>11.061</td>
<td>-0.036</td>
<td>127.870</td>
</tr>
<tr>
<td>elsitavf</td>
<td>27.588</td>
<td>-0.089</td>
<td>79.803</td>
<td>189.346</td>
<td>-4.134</td>
<td>84.341</td>
<td>11.023</td>
<td>-0.034</td>
<td>127.435</td>
</tr>
</tbody>
</table>
ANALYSIS OF EXTERNAL DRIFT KRIGING ALGORITHM WITH APPLICATION TO PRECIPITATION ESTIMATION IN COMPLEX OROGRAPHY.

<table>
<thead>
<tr>
<th></th>
<th>esltavcfTI</th>
<th>eltadvcvFTI</th>
<th>sladcvcf</th>
<th>sldcTI</th>
<th>elsladcvcfTI</th>
<th>eltadcTI</th>
<th>eltavcf</th>
<th>eltaTI</th>
<th>elsladtadvcvFTI</th>
<th>eltadvfcfTI</th>
</tr>
</thead>
<tbody>
<tr>
<td>sl</td>
<td>-0.241</td>
<td>-0.083</td>
<td>-0.161</td>
<td>-0.616</td>
<td>-0.282</td>
<td>-0.616</td>
<td>-0.165</td>
<td>-0.146</td>
<td>-0.303</td>
<td>-0.051</td>
</tr>
<tr>
<td>ta</td>
<td>80.458</td>
<td>80.134</td>
<td>82.478</td>
<td>85.198</td>
<td>81.963</td>
<td>85.279</td>
<td>79.964</td>
<td>79.924</td>
<td>82.298</td>
<td>79.839</td>
</tr>
<tr>
<td>vcf</td>
<td>191.128</td>
<td>190.098</td>
<td>196.977</td>
<td>192.946</td>
<td>197.561</td>
<td>194.748</td>
<td>196.886</td>
<td>194.660</td>
<td>199.151</td>
<td>197.015</td>
</tr>
<tr>
<td>TI</td>
<td>85.135</td>
<td>84.676</td>
<td>87.740</td>
<td>85.945</td>
<td>88.000</td>
<td>86.747</td>
<td>87.700</td>
<td>86.708</td>
<td>88.449</td>
<td>87.757</td>
</tr>
<tr>
<td></td>
<td>-0.095</td>
<td>0.036</td>
<td>-0.004</td>
<td>-1.120</td>
<td>-0.091</td>
<td>0.035</td>
<td>0.000</td>
<td>-0.005</td>
<td>0.035</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>128.009</td>
<td>127.497</td>
<td>129.155</td>
<td>127.266</td>
<td>128.022</td>
<td>128.022</td>
<td>126.882</td>
<td>126.769</td>
<td>127.518</td>
<td>127.609</td>
</tr>
</tbody>
</table>

el=elevation
sl=slope
ta=transformed aspect
vcf=vegetation cover factor
dc=distance from the coast.
TI=Topographic index.

Elsltavcf= el+sl+ta+vcf=elevation+slope+transformed aspect+vegetation cover factor.

9.9. APPENDIX 9. SUMMARY OF OBSERVED RAINFALL STATISTICS.

KARNATAKA

> summary(Apr.00)

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.00</td>
<td>3.00</td>
<td>18.00</td>
<td>34.57</td>
<td>52.00</td>
<td>184.00</td>
</tr>
</tbody>
</table>

> summary(Jul.00)

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10.0</td>
<td>51.0</td>
<td>93.0</td>
<td>224.5</td>
<td>219.0</td>
<td>1416.0</td>
</tr>
</tbody>
</table>

> summary(Dec.00)

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>8.651</td>
<td>8.000</td>
<td>118.00</td>
</tr>
</tbody>
</table>

UTTARANCHAL

summary(Apr)

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6.30</td>
<td>11.35</td>
<td>20.35</td>
<td>22.72</td>
<td>32.00</td>
<td>54.40</td>
</tr>
</tbody>
</table>

summary(Jul)

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>176.3</td>
<td>330.1</td>
<td>443.2</td>
<td>472.8</td>
<td>624.9</td>
<td>968.0</td>
</tr>
</tbody>
</table>

summary(Dec)

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8.90</td>
<td>14.80</td>
<td>21.80</td>
<td>20.26</td>
<td>25.40</td>
<td>36.10</td>
</tr>
</tbody>
</table>