SPATIO-TEMPORAL REGRESSION KRIGING FOR PREDICTING RAINFALL FROM SPARSE PRECIPITATION DATA IN GHANA

JACOB AZUMAANA ADIGI
February, 2019

SUPERVISORS:
Dr. Frank B. Osei
Dr. Mariana Belgiu

SPATIO-TEMPORAL REGRESSION KRIGING FOR PREDICTING RAINFALL FROM SPARSE PRECIPITATION DATA IN GHANA

JACOB AZUMAANA ADIGI
Enschede, The Netherlands, February, 2019

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

SUPERVISORS:
Dr. Frank B. Osei
Dr. Mariana Belgiu

THESIS ASSESSMENT BOARD:
Prof. Dr. ir. A. Stein (Chair)
Dr. ir. G. B. M. Heuvelink (External Examiner, Wageningen University & Research, Laboratory of Geo-Information Science and Remote Sensing)

ABSTRACT

Reliable and spatially exhaustive surfaces that provide accurate spatial and temporal distribution of rainfall are key requirements for making climate-related informed decisions such as management of water resources and ecological modeling. Different approaches to predict rainfall from sparsely available data is a notable subject of research in spatial statistics. In this study, we carried out a spatio-temporal regression kriging to predict rainfall in Ghana by applying a model-based approach using maximum likelihood method to estimate the model parameters. Mean monthly, mean annual and cumulative annual data was computed from daily measurements of 26 rain gauge stations from 2001 to 2010 distributed over 238,540 km^2 area of Ghana. The spatial coordinates, elevation derived from digital elevation model (DEM) and one-month time series Normalized Difference Vegetation Index (NDVI) images of the Moderate Resolution Imaging Spectroradiometer (MODIS) were used as predictors. Due to multicollinearity between predictors, three linear regression models were formulated and used to carry out mean monthly, mean annual and cumulative annual predictions of rainfall at $1km^2$ grid. Their performance was evaluated by leave-one-out cross-validation using the root mean square error and coefficient of correlation metrics. The second regression kriging model with the spatial coordinates especially the northing as predictors performed best in five months for the mean monthly predictions, six years for cumulative annual predictions and the mean annual predictions. The third model with NDVI as predictor also performed in five months and three years. The third model with a subset of all predictors performed in two months and one year. The results uncovered the spatial and temporal distribution of rainfall in the country. The south-western part records high rainfall and the northern part less rainfall. June is the wettest month and January the driest month in the country. There was an increasing trend in rainfall from 2001 to 2010.

ACKNOWLEDGEMENTS

My heartfelt gratitude goes to the Almighty God for the strength and wisdom to go through the daily demands of this research. I would like to acknowledge my country Ghana for the sponsorship to undertake this course and Ghana Meteorological Agency for the data provided for this research.

I would also like to extend to my supervisors Dr. Frank B. Osei and Dr. Mariana Belgiu, my sincere gratitude for their guidance, suggestions, and supervision to make this research a great success.

My warm gratitude also goes to my wife Winifred Adigi and family, my family, my Uncle Denis Amenga and the family, my pastor Rev Dr. Stephen Wengam and the church for their love, sacrifice, support, and prayers. Am also highly indebted to Brother Kwabena Adu-Boahene and the family for being a great blessing.

And to all my classmates, friends in ICF and all who have contributed in one way or the other to the success of this work I would like to say thank you.

I dedicate this work to my lovely wife
Winifred Adigi and my daughter Glory Adigi

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Motivation and Problem Statement

Climate variables such as precipitation data are of prime importance and its spatial distribution is required for water resources management, hydrologic and ecologic modeling, recharge assessment and irrigation scheduling (Mair & Fares, 2011; Di Piazza, Conti, Noto, Viola, & Loggia, 2011). Reliable and exhaustive rainfall information is a critical requirement for the successful modeling and assessment of these processes.

Direct and accurate measurements of precipitation data at a fine resolution will require a dense network of meteorological stations (Goovaerts, 2000). Ghana as a third world country lacks the capacity to afford the installation of these stations at higher density. A network of meteorological stations which are the direct source of reliable precipitation data are therefore sparsely located making it difficult to characterize this highly variable phenomenon and its spatial and temporal distribution (Keblouti, Ouerdachi, & Boutaghane, 2012). The level of sparsity becomes more pronounced as a result of missing data due to the malfunctioning of some rain gauge stations. This affects the continuity of available data from the sparsely available stations.

A practical indirect alternative to providing spatially exhaustive precipitation information is the use of ground-based meteorological RADARs and satellite platforms with mounted remote sensing devices. The accuracy and resolution of the estimates provided by these indirect methods are often insufficient and unreliable. As a result, these methods still require calibration and validation using historical data from direct and reliable rain gauge measurements (Bostan, Heuvelink, & Akyurek, 2012; Lanza, Ramírez, & Todini, 2001). This calls for alternative methods that provide the means to accurately estimate precipitation data at unsampled locations.

Interpolation methods that have been proposed for estimating precipitation data at unsampled locations include geostatistical interpolation and deterministic techniques. According to Goovaerts (1999), geostatistical methods provide the best results in the estimation of precipitation since they take into consideration spatial dependences which are usually observed for precipitation. Geostatistics is (Goovaerts (2000) as quoted by Mendez & Calvo-Valverde (2016)) "based on the theory of regionalized variables and provides a set of statistical tools for incorporating the spatial correlation of observations in data processing". The use of secondary variables such as elevation, radar imagery or land use in combination with precipitation as established by previous studies results in more accurate estimation than using only precipitation measurements (Hofierka, Parajka, Mitasova, & Mitas, 2002).

Despite the robustness of different spatial interpolation techniques, there is always an element of uncertainty as to which method is most applicable for a given set of data. According to Luo, Taylor, & Parker (2008), accuracies vary significantly among spatial interpolation methods depending on the spatial attributes of the data at hand. Burrough, McDonnell, & Lloyd (1998) indicated that in the abundance of data, most interpolation techniques produce results that are similar. This becomes completely different when data are sparsely located and the choice of interpolation method to estimate data at unsampled locations becomes a great concern. In this study, we investigate and model the spatial structural

dependence in sparse rainfall data by applying the Maximum Likelihood method to estimate the drift and covariance model parameters.
Through spatial exploratory analysis, the presence of trend was evident in the data, especially with the latitude. This necessitated the use of non-stationary geostatistical methods also called the 'hybrid' techniques by Bishop et al. (2000). According to Bishop et al. (2000), the use of ordinary univariate kriging is inappropriate whenever trend exists in the data. Ordinary kriging was therefore not undertaken in this research.

These hybrid algorithms which assume a spatially varying mean by including a trend surface model (Deutsch & Journel, 1998), are used to estimate the mean monthly, mean annual and cumulative annual precipitation of Ghana using data from 2001 to 2010 from 26 rain gauge stations distributed over the country.

1.2. Research Identification

Mapping of the spatial distribution of precipitation is important for many applications in ecological studies, environmental sciences, and epidemiology of infectious diseases. Meteorological stations serve as the source for accurate precipitation data but are often sparsely located. This becomes more evident in a third world country like Ghana that lacks the resources to establish a denser network of these stations. This project sought to create exhaustive precipitation information from fine-scale covariates using regression kriging in monthly and annual time steps. The Normalized Difference Vegetation Index (NDVI) from the Moderate Resolution Imaging Spectroradiometer (MODIS), elevation derived from digital elevation model (DEM) and the spatial coordinates are the secondary variables incorporated in this study. The results of this research will be useful for applications in ecological and epidemiological studies in Ghana.

1.3. Research Objectives

The main objective of this research is to predict the mean monthly and annual spatial and temporal distribution of rainfall in Ghana by applying a model-based approach.
The specific objectives are:

a. To evaluate the influence of secondary variables (time series NDVI and elevation) on the spatial and temporal distribution of precipitation.

b. To evaluate different covariance functions to infer the spatial structure of precipitation in the study area for a given time period.

c. To evaluate different regression models to determine the model that provides the best prediction results for a given time period.

1.4. Research Questions

Specific objective 1:

i. What is the relationship between elevation and the spatial distribution of precipitation in the study area for a given time period?

ii. Can time series of NDVI, a remote-sensed derived covariate from MODIS be used to infer the spatial and temporal distribution of precipitation in the study area?

Specific objective 2:

i. Which covariance function is appropriate for capturing the spatial structure of precipitation in the study area for a given time period?

ii. What are the optimal model parameters of the chosen covariance function?

Specific objective 3:

i. Which regression model is most appropriate for the prediction of precipitation in the study area for a given time period?

1.5. Innovation Aimed At

The innovation of this research is aimed at producing spatially exhaustive rainfall information from sparsely available rainfall point data by applying a model-based maximum likelihood method in the estimation of the model parameters. There are scarcely published research studies that have applied model-based approaches to estimate precipitation from sparse data, especially in Ghana.

According to Hofierka et al. (2002), the use of secondary variables such as elevation, radar imagery and land use in combination with precipitation results in a more accurate estimation of precipitation than using only precipitation measurements. The incorporation of time series Normalized Difference Vegetation Index (NDVI), a remote-sensed derived secondary information and DEM in the above-mentioned method to map the spatial and temporal distribution of rainfall also adds to the novelty of this research. The influence of NDVI to accuracy in mean monthly and annual rainfall estimation is evaluated.

1.6. Thesis Outline

This section outlines the structure of the thesis. Chapter 1 captures a description of the motivation and problem statement, the research objectives, questions, and innovation. A literature review of related studies is provided in Chapter 2. Chapter 3 provides information regarding the study area, the datasets and software used. The methodology utilized is provided in Chapter 4. Chapter 5 provides the results and analysis. The discussions and limitations of the research are provided in Chapter 6. Chapter 7 captures the conclusion and recommendations.

2. LITERATURE REVIEW

2.1. Introduction

In this section, we undertake a literature review on geostatistics and previous related research projects that have been carried out taking into consideration the spatial interpolation methods used, covariance parameter estimation methods applied and the incorporation of auxiliary variables in mapping climatic variables especially precipitation.

2.2. The Concept of Geostatistics

Geostatistics, also known as spatial statistics is concerned with the analysis and prediction of phenomena that vary in space and time. Phenomena that are of interest to the geostatistician are often very expensive and are therefore sparsely located. The need to obtain exhaustive information in the spatial and/or temporal distribution of a given phenomenon within a study area by carrying out predictions at unsampled locations is the subject matter of geostatistics (Richard Webster And Margaret A. Oliver, 2007).

There are several spatial interpolation methods that are used for spatial predictions at unsampled locations. These can be grouped into two categories namely, deterministic and geostatistical methods. Examples of deterministic methods include Splines, Thiessen polygons and Inverse Distance Weighting (Hartkamp, De Beurs, Stein, & White, 1999; Keblouti et al., 2012). Examples of geostatistical methods include Ordinary kriging, regression kriging, multiple linear regression, geographically weighted regression, universal kriging, co-kriging and kriging with external drift. The following are advantages associated with geostatistical methods over deterministic methods (Goovaerts, 1999, 2000);

- Geostatistical methods allow one to make use of the spatial correlation between neighboring observations to estimate values at unvisited locations.
- They provide estimates of the prediction error (kriging variance) at unsampled locations.
- They allow the integration of the primary attribute with secondary attributes that are sampled at higher density.

2.3. Related works

The applications of rainfall information in gridded format are numerous and varied. Exhaustive rainfall information serves as key inputs for basin management, hydrological and water quality applications (Ly, Charles, & Degré, 2011). The successful running of hydrological models, research into agriculture, the planning and management of water resources all require exhaustive rainfall information (Basistha, Arya, & Goel, 2008). Extensive research has been carried out with the application of different methods to produce high resolution rainfall information.

Bostan et al. (2012) compared multiple linear regression (MLR), geographically weighted regression (GWR), ordinary kriging (OK), regression kriging (RK) and universal kriging (UK) in mapping the average annual precipitation over Turkey. Elevation map, surface roughness, distance to the nearest coast, river density, aspect, land cover, and eco-region were used as covariates. In an R-squared, root mean square error (RMSE) and standardized mean square error (SMSE) performance assessment method, UK was considered the most accurate for the spatial interpolation of the precipitation distribution of Turkey.

Harris et al. (2010) made a comparison between MLR, GWR, OK, UK and geographically weighted regression kriging (GWRK) models using simulated data sets. The authors concluded that UK was the best performing model. Bajat et al. (2013) also carried out a study to map the average annual precipitation in Serbia using regression kriging. Digital elevation model and spatial coordinates were used as covariates. By comparative analysis with the multiple linear regression method, the authors concluded that regression kriging performed better by cross-validation measures.

Cantet (2017) carried out a comparison study to map the mean monthly and mean annual precipitation in a small island called Martinique which is located in the Lesser Antilles using data from 35 meteorological stations. Spatial interpolation methods such as regression kriging and external drift kriging were compared through a cross-validation procedure. The different regression kriging methods that were applied include multiple linear regression kriging, principal component regression kriging, and partial least squares regression kriging. In his performance assessment by cross-validation, external drift kriging outperformed the regression kriging methods and was considered the best method for mapping precipitation in the island.

Buytaert et al. (2006) investigated the spatial and temporal rainfall variability in the western mountain range of the Ecuadorian Andes using rainfall data from 14 stations. They compared kriging with Thiessen polygon technique and reported kriging produced accurate interpolation of rainfall than the Thiessen polygon. They also indicated the fact that the accuracy of both methods can be improved by the incorporation of external trends.

Ly, Charles, & Degré (2011) in their paper "Geostatistical interpolation of daily rainfall at catchment scale" in Belgium, compared geostatistical and deterministic approaches to interpolate rainfall using 30 years daily rainfall data from 70 rain gauge stations. In a cross-validation performance assessment, the geostatistical and inverse distance weighting algorithms which take into account the spatial dependence between neighboring observations performed better than the Thiessen polygon algorithm. According to the authors, the Thiessen polygon technique failed to depict the true spatial variation of rainfall in the study area.

Masson & Frei (2014) carried out a study on the spatial analysis of precipitation in a high mountain region of the European Alps. Kriging with external drift (KED) using local topographic height as the only predictor was reported to produce smaller interpolation errors as compared to linear regression. According to the authors, the incorporation of more predictors only resulted in a marginal improvement of the kriging algorithm. They also underscored the fact that the use of a single predictor field in KED improves interpolation accuracy as compared to ordinary kriging.

Lark (2000) compared the method of moments and the maximum likelihood method to estimate variograms from simulated and real data. The maximum likelihood method was observed to be efficient than the method of moments in different sampling intensities of the simulated data. However, the method of moments was efficient where there were small nugget variance and large correlation range of the data. Both methods were reported to be susceptible to positively skewed simulated data. Todini & Ferraresi (1996) advocated the following reasons for using the Maximum Likelihood method for estimating model parameters:

- The need for an estimation algorithm that is objective.
- The need to have parameter values estimated by minimizing (or maximizing) the objective function in the kriged variables' space with available actual observations and meaningful residuals but not in the variogram space without available observations.
- The possibility of having the variance-covariance matrix of the parameters estimated.

Pardo-Igúzquiza (1998) also identified the following advantages for the application of Maximum likelihood inference in geostatistics:

- The parameters that are of interest are the only ones estimated.
- It is easy to assess the uncertainty of the estimates.
- Model selection may be done by using the ML function.
- As compared to other methods, it is more efficient in terms of mean square error.

Yoon, Kim, & Park (2015) estimated monthly precipitation in South Korea by comparing different statistical linear interpolation models using data from 441 stations. The linear models compared were the general linear model, the generalized additive model, the spatial linear model, and the Bayesian spatial regression model. The secondary variables that were incorporated in the study include the longitude, latitude, elevation, topographic aspect and coastal proximity. The Bayesian spatial model was reported to outperform the other models based on the root mean square error, mean absolute error and correlation coefficient indexes.

Goovaerts (2000) integrated a digital elevation model in the estimation of rainfall in Portugal using three multivariate geostatistical algorithms such as simple kriging with varying local means, kriging with external drift and colocated cokriging. A cross-validation comparison was made to evaluate the prediction performance of the multivariate algorithms against univariate techniques such as Thiessen polygon, inverse square distance, and ordinary kriging. The Thiessen polygon and inverse square distance algorithms which ignore both the elevation and rainfall observations at neighboring stations reported larger prediction errors. The multivariate techniques were reported to outperform the other interpolators, especially in a linear regression, where the rainfall observations and the colocated elevation is taken into account. Ordinary kriging was also reported to outperform linear regression when there is a moderate correlation between rainfall and elevation.

Time-series of remote sensing products have much to offer in geostatistics as far as the contribution of auxiliary variables to the improvement of spatial prediction accuracy is concerned. There has been a quest in recent times to improve the accuracy of climatic mapping by increasing the scope of covariates to time series of remote-sensing based variables (Hengl et al., 2012). Hengl et al. (2012) incorporated time-series of MODIS land surface temperature (LST) images to map daily temperature in Croatia. According to the authors, the use of the time series product led to a significant improvement in accuracy.

Hu, Shu, Hu, & Xu (2017) undertook a study on spatiotemporal regression kriging to predict mean monthly precipitation in Xinjiang using time-series MODIS data and digital elevation model. According to the authors, spatiotemporal regression kriging performed better by a leave-one-out cross-validation method when compared with spatiotemporal multi-linear regression and spatiotemporal kriging. The authors also indicated that the normalized difference vegetation index is one of the optimal covariates for mean monthly precipitation prediction.

The focus of this project differs from the related works in the light of the interpolation methods to be compared, the covariance structural modeling approach and the covariates to be considered in estimating precipitation from sparse data. This project sought to assess the performance of different regression models of the hybrid geostatistical algorithm regression kriging in estimating mean monthly and annual precipitation by incorporating DEM and NDVI, a remote sensed derived time-series data from MODIS, as predictors. The performance assessment of these models was carried out by the leave-one-out cross-validation measures.

3. STUDY AREA AND DATA DESCRIPTION

3.1. Study Area

The study area for this project covers the entire country of Ghana as shown in Figure 2. Ghana is a West African country located along the Atlantic Ocean and the Gulf of Guinea. Ghana lies approximately between latitude 4^{o} and 12^{o}N and longitude 4^{o}W and 2^{o}E. It has a total area of 238,540 km^2. The rainfall dynamics in Ghana shows considerable variation between the northern and the southern parts. The northern part experiences rainfall between the months of May and November and the southern part experiences a bimodal wet season with the major season from March to July and the minor season from September to November.

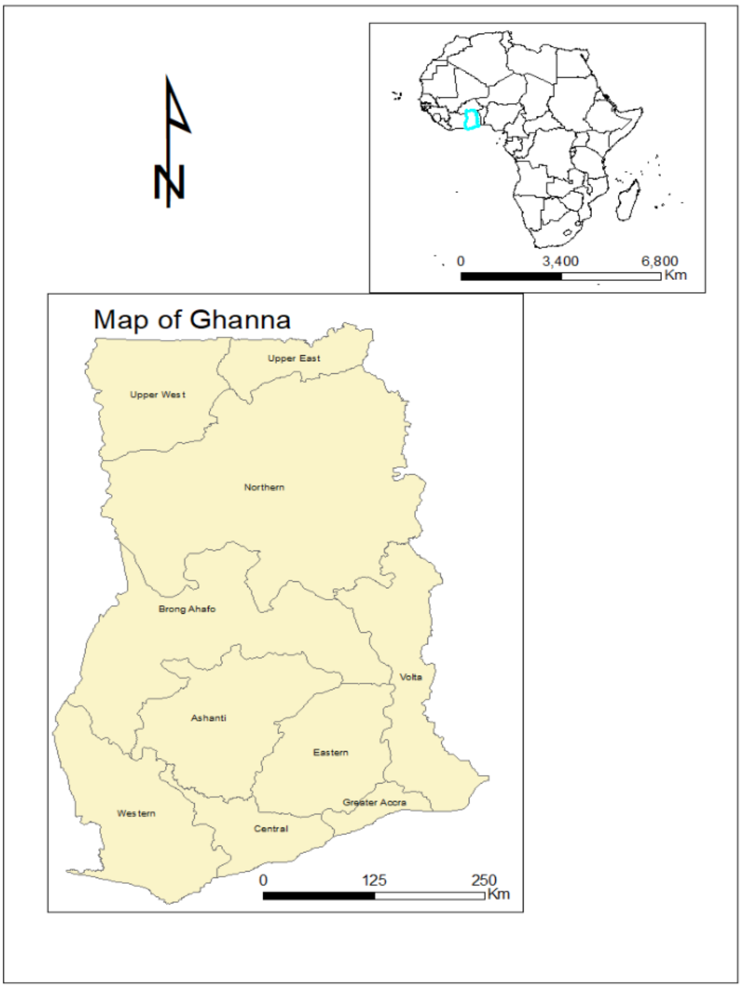


Figure 3.1. Map of study area

3.2. Datasets

3.2.1. Precipitation Data

Rain gauge measurements from 26 meteorological stations from 2001 to 2010 are used in this study. The choice for this period is based on the use of the NDVI time-series data from MODIS which is available from 2001 to date.

The original precipitation data received from the Ghana Meteorological Agency consist of daily precipitation measurements for thirty-two stations. Due to data gaps, only twenty stations with complete data from 2001 to 2010 are used for analysis.

3.2.2. Digital Elevation Model (DEM) Data

The Shuttle Radar Topography Mission (SRTM), a 90m spatial resolution digital elevation model data was downloaded from the website of the U.S Geological Survey https://gdex.cr.usgs.gov/gdex/. This data which is a global public dataset was downloaded and clipped to the study area. It was resampled from the 90m spatial resolution to 1km spatial resolution using the nearest neighbour algorithm. The figure below shows the spatial location of the rain gauge stations used in this study.

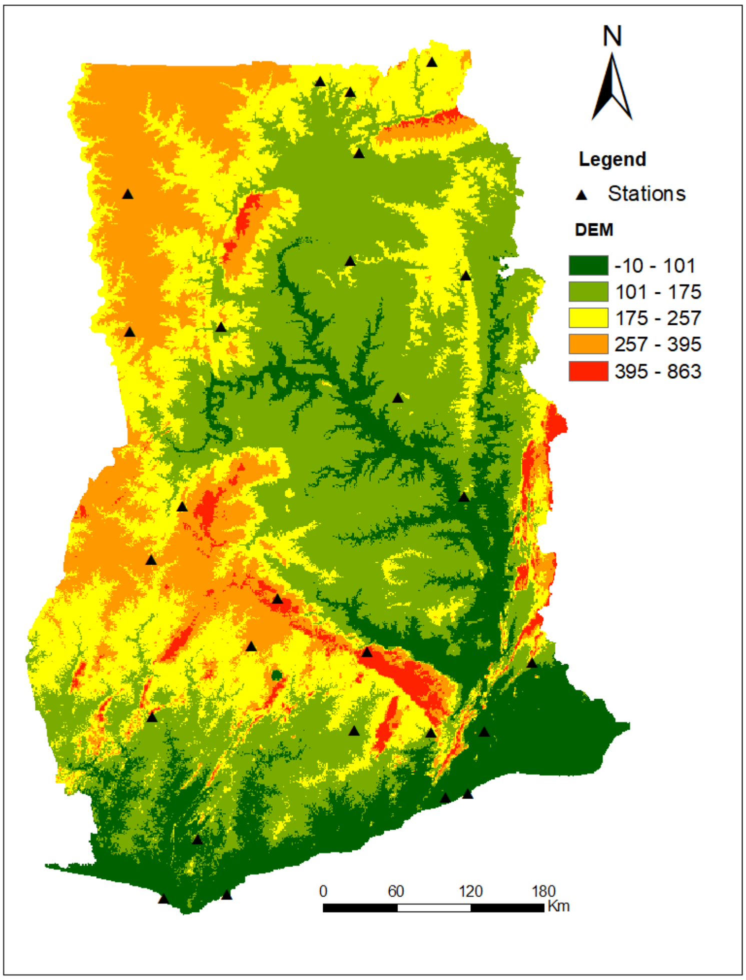


Figure 3.2: Digital Elevation Model of Ghana showing the spatial locations of the rain gauge stations

3.2.3. Time Series MODIS NDVI Data

Monthly time series images of the Moderate Resolution Imaging Spectroradiometer (MODIS) Normalized Difference Vegetation Index (NDVI) at 1km resolution were obtained from the USGS website. These images were downloaded using the MODIStsp (Busetto et al., 2018) and mapedit packages in the R software.

This readily available NDVI product is calculated by the National Aeronautics and Space Administration (NASA) as the ratio of the difference between the near-infrared radiation and the visible radiation to the sum of the near-infrared radiation and the visible radiation. NDVI values range between -1 and +1 with a value close to +1 indicating high density of green leaves. This can be written mathematically as

$$NDVI = \frac{(NIR - VIS)}{(NIR + VIS)}$$

Where NIR represents reflectance in the near infrared channel and VIS represents reflectance in the visible channel.

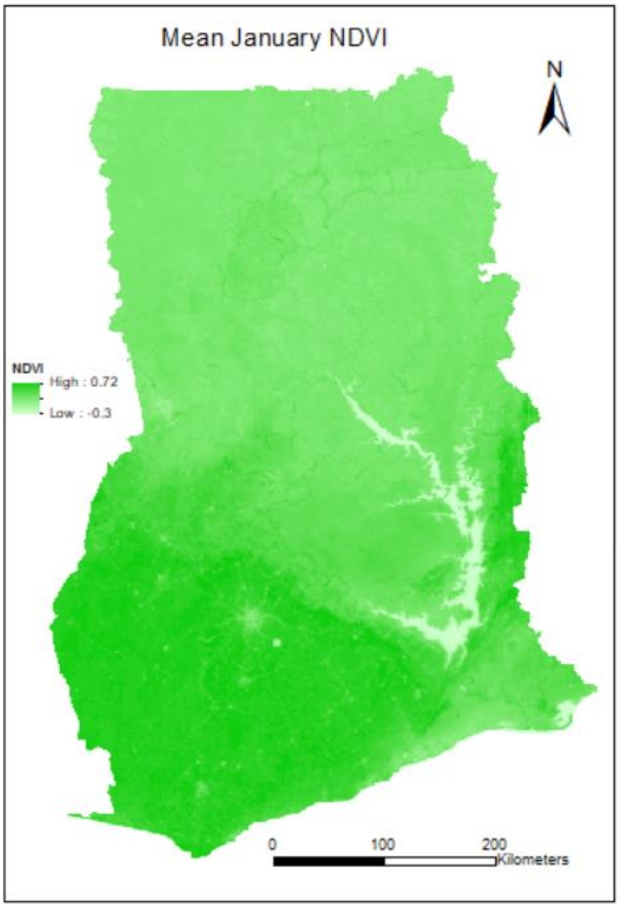


Figure 3.3: Map of Ghana showing NDVI for mean January

3.3. Software Used

Different software is used in this study to accomplish different tasks. These include;

- ArcGIS. The Extract by Mask tool in Spatial Analyst was used to extract the DEM to the boundaries of the study area.
- R software. Different packages of the R software (R Core Team., 2015) were used to accomplish different task including exploratory data analysis, modeling, and prediction.

4. METHODOLOGY

The methodology undertaken to carry out this research is outlined in Figure 4.1.

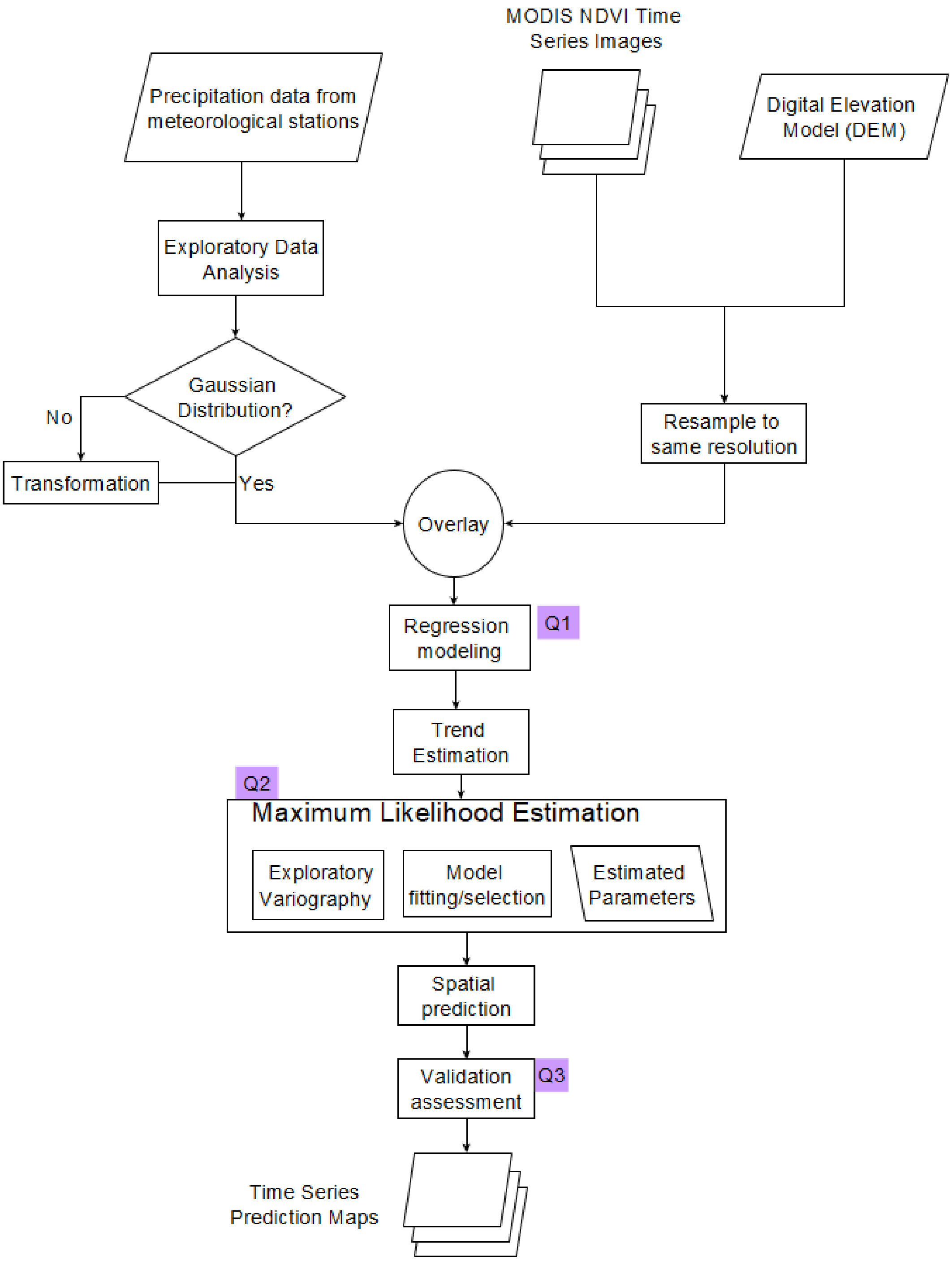


Figure 4.1: Flowchart of the methodology

4.1. Data Processing

4.1.1. Primary Data

The data received from the Ghana Meteorological Agency was the daily precipitation records for thirty-two stations from January 2001 to December 2017. Since the inclusion of missing data to the computation of average and cumulative rainfall for stations will lead to unrepresentative and wrong estimations, stations with missing data for more than 10 days in a month (Hartkamp et al., 1999) were eliminated from the analysis. The period from 2001 to 2010 gave the maximum number of stations with continuous data and was therefore considered for this analysis. Mean monthly, mean annual and cumulative annual rainfall was computed from the daily records of twenty-six (26) stations. The table below shows the descriptive statistics of the mean monthly and annual precipitation data for the 26 stations.

Table 4.1: Descriptive statistics of the mean monthly and annual precipitation (mm) data

	Mean	Median	SD	Max	Min	Skew	Kurt
January	14.60	12.39	12.92	49.49	0	0.76	0.06
February	29.72	22.12	25.53	90.77	2.40	0.69	-0.78
March	69.50	54.51	45.30	148.18	4.76	0.13	-1.44
April	121.28	122.89	39.67	191.66	38.95	-0.17	-0.70
May	147.31	139.98	49.05	296.61	81.85	1.11	1.21
June	189.4	183.9	65.62	464.1	129.10	2.72	8.98
July	147.94	144.14	49.37	240.16	65.47	0.11	-1.1
August	129.74	89.02	88.17	292.43	17.40	0.33	-1.52
September	167.08	180.16	59.28	296.96	51.59	-0.32	-0.23
October	136.59	147.07	57.11	233.75	59.11	0.02	-1.47
November	49.57	41.79	47.19	157.91	1.42	1.1	0.36
December	20.62	14.06	21.26	70.8	0	0.93	-0.07
Annual	1227.4	1272.2	275.56	1954.9	728.2	0.67	0.39

4.1.2. Secondary Data

The auxiliary variables incorporated in this study include the Moderate Resolution Imaging Spectroradiometer (MODIS) Normalized Difference Vegetation Index (NDVI) images of the Tera Earth Observation systems platform and the Digital Elevation Model (DEM) of the Shuttle Radar Topography Mission (SRTM). Mean monthly NDVI was calculated for each month by stacking the monthly 1km time series NDVI images according to the months of the year using the stack functions of the "raster" package (Hijmans et al., 2018) in R software.

4.2. Data Integration

Both the primary data and the secondary data were recorded in their respective coordinate reference system and at different resolutions, especially for the secondary data. It is therefore necessary to ensure all data are in the same coordinate reference system to enhance compatibility. Both the primary and the secondary data were transformed from the WGS84 geographic coordinate system to the WGS84 UTM Zone 30N coordinate reference system. The DEM was resampled from 90m resolution to the same resolution as the MODIS NDVI and both images were stacked together and overlaid with the point data to extract the corresponding covariates at the point data locations. This was done using the stack and extract functions of the raster package (Hijmans et al., 2018) in the R software.

4.3. Exploratory Data Analysis

Spatial and non-spatial exploratory data analysis are key statistical operations which are necessary prior to the formulation of models in geostatistics.

4.3.1. Non-Spatial Exploratory Analysis

Exploratory analysis was carried out to investigate the non-spatial structure of the primary data. This was done by plotting histograms and Q-Q plots of the monthly precipitation data. Details of these plots are shown in chapter 5. This makes it possible to investigate the distribution of the data as well as detect and eliminate possible outliers. The data for the mean monthly, mean annual and cumulative annual precipitation was positively skewed. We carried out a log-transformation of the data before model calibration and prediction. After predictions, the results were back-transformed to the original scale of the data (Hengl, Heuvelink, & Rossiter, 2007).

4.3.2. Spatial Exploratory Analysis

Spatial exploratory data analysis includes circle plots of the response variable with respect to the spatial locations of the data. Scatter plots to investigate the relationship between the response variable and the auxiliary information used were produced (Diggle & Ribeiro, 2007; Richard Webster And Margaret A. Oliver, 2007). This enabled us to gain insight with regards to the presence of a trend in the data. A discovered trend in the data indicated a spatially varying mean and suggested the inclusion of a trend surface model. Possible outliers that were also discovered through this exploratory analysis were eliminated.

4.4. Regression Modelling

We carried out three different linear regression models between the log-transformed precipitation and the predictors in this study (Neter, J., Kutner, M. H., Nachtsheim, C. J., Wasserman, 2005). The predictors include MODIS NDVI, elevation, the northing (X) and easting (Y) coordinates. It became evident that the MODIS NDVI was correlated with the northing coordinates. Because of the multicollinearity, we decided to formulate three exploratory models to choose significant predictors for each model. In the first model, we regressed log-transformed precipitation on all the predictors and a stepwise regression was carried out to select statistically significant predictors for the model. NDVI was still maintained in the first model because the multicollinearity ceased in the months with higher rainfall. In the second model, we regressed log-transformed precipitation on the spatial coordinates and the statistically significant predictor was chosen for the model or both are used if significant. In the third model, we regressed log-transformed precipitation on the ancillary variables MODIS NDVI and elevation and the statistically significant predictor selected for the model or both are used if significant.

The statistically significant predictors were chosen as the ones with their p-value less than a significance level of 0.05. These predictors were used as the trend in modeling the covariance structure. In few cases especially for Model 1, where all predictors were insignificant at 0.05 significance level, we considered the predictors provided by the stepwise regression for the model.

The three formulated models are shown in equation (1) to (3).

Model 1: $\log(Precipitation) = \beta_0 + \beta_1 NDVI + \beta_2 ELEV + \beta_3 X + \beta_4 Y$ (1)

Model 2: $\log(Precipitation) = \beta_0 + \beta_1 X + \beta_2 Y$ (2)

Model 3: $\log(Precipitation) = \beta_0 + \beta_1 NDVI + \beta_2 ELEV$ (3)

where β_0 is the intercept, β_1, β_2, β_3 and β_4 are the regression coefficients of the predictors and ELEV represents elevation. ELEV, X and Y are predictors which are temporally constant and NDVI is temporally dynamic and is provided for each period of model fitting.

4.5. Model Definition

The stationary Gaussian model has the following assumptions (Diggle & Ribeiro, 2007, pg. 29);

- The Gaussian process $\{S(x): x \in \mathbb{R}^2\}$ for locations $x_1, \ldots, x_n$ has mean $\boldsymbol{\mu}$, variance $\sigma^2 = \mathrm{Var}\{S(x)\}$ and correlation function $\rho(h) = \mathrm{Corr}\{S(x), S(x')\}$ such that distance $h = \|x - x'\|$;
- Given $\{S(x): x \in \mathbb{R}^2\}$, the measured value y_i of a geostatistical data at location x_i, are realisations of mutually independent random variables Y_i, distributed normally with mean $E[Y_i|S(.)] = S(x_i)$ and conditional variance τ^2

The model can be defined as

$Y_i = S(x_i) + Z_i$ (4)

where $\{S(x): x \in \mathbb{R}^2\}$ represents the first assumption and Z_i are mutually independent N(0, τ^2) random variables.

The stationary Gaussian model can be extended for a spatially varying mean by including a linear regression model in place of the stationary mean.

The Gaussian random field model for a spatially varying mean is given as (Paulo Ribeiro Jr, Diggle, & Paulo Ribeiro Jr, 2018):

$Y(x) = \mu(x) + S(x) + e$ (5)

where

- $Y(x)$ is the observed variable at location x in a planar Euclidian coordinates.
- $\mu(x) = X\boldsymbol{\beta}$ is the mean component of the model (trend).
- $S(x)$ is a stationary Gaussian process with variance σ^2 (partial sill) and correlation function ϕ (the range parameter).
- e is the nugget variance or measurement error.

4.6. Parameter Estimation by Maximum Likelihood Method

A robust alternative to estimate the model parameters of the linear effects and the covariance structure is the Maximum Likelihood method proposed by Mardia & Marshall (1984). To apply the Maximum Likelihood method, the variable of interest is assumed to be a realization from a random Gaussian process (Diggle & Ribeiro, 2007, pg. 112).

The spatial trend $\mu(x)$ in equation (5) is either a function of the spatial coordinates or spatially referenced covariates such that $\mu(x) = \mathrm{D}\beta$,

$Y \sim N(D\beta, \sigma^2 R(\varphi) + \tau^2 I)$ (6)

where D is the $n \; x \; p$ matrix of covariates, β is the corresponding vector of regression parameters, and R depends on a scalar or vector valued correlation function parameter(s) φ.

The likelihood function is given as:

$$L(\beta, \tau^2, \sigma^2, \varphi) = -0.5\{nlog(2\pi) + log\{|(\sigma^2 R(\varphi) + \tau^2 I)|\} + (y - D\beta)^T(\sigma^2 R(\varphi) + \tau^2 I)^{-1}(y - D\beta)\}, \quad (7)$$

The values of the regression coefficients defining the mean process and the covariance parameters that maximize the log-likelihood function for a given set of data yields the maximum likelihood estimate of the parameters. Considering the spatial dependence (nugget:sill ratio) described by the variogram as

$$v^2 = \frac{\tau^2}{\sigma^2}, \quad (8)$$

Then the matrix

$$V = R(\varphi) + v^2 I. \quad (9)$$

Given V, the log-likelihood function is maximized at

$$\hat{\beta}(V) = (D^T V^{-1} D)^{-1} D^T V^{-1} y \quad (10)$$

and

$$\hat{\sigma}^2(V) = \frac{\{y - D\hat{\beta}(V)\}^T V^{-1}\{y - D\hat{\beta}(V)\}}{n} \quad (11)$$

$\hat{\beta}(V)$ becomes the generalized least squares estimate provided V is known.

A substitution of $\hat{\beta}(V)$ and $\hat{\sigma}^2(V)$ in equation (10) and (11) respectively into the log-likelihood function gives a concentrated log-likelihood function in equation (12).

$$L_0(v^2, \varphi) = -0.5\{n\log(2\pi) + n\log\hat{\sigma}^2(V) + \log|V| + n\} \quad (12)$$

Numerical optimization of equation (12) with respect to φ and v is carried out followed by a back substitution to obtain $\hat{\sigma}^2$ and $\hat{\beta}$. Details of the maximum likelihood estimation method can be found in (Diggle & Ribeiro, 2007).

4.7. Model Selection

In the method of moments, the plausible variogram model is selected as the one with the minimum sum of square errors or root mean square error. The plausible covariance function in the maximum likelihood method is selected as the one with the maximum log-likelihood function or the one with the minimum negative log-likelihood function (Pardo-Igúzquiza, 1998).

Plausible variogram models (Oliver & Webster, 2014) used in this study include the following;

The Spherical function which is given as:

$$\gamma(h) = c_0 + c\left\{\frac{3h}{2r} - \frac{1}{2}\left(\frac{h}{r}\right)^3\right\} \text{ for } 0 < h \le r \quad (13)$$

$$= c_0 + c \text{ for } h > r$$

$$= 0 \text{ for } h > 0$$

where h is the distance between pairs of points, c_0 is the nugget variance which represents small variability at distances less than the minimum sampling distance or the measurement error, and the range r is the distance beyond which spatial dependence ceases to exist. The total variance known as the sill is given as $c_0 + c$.

The Exponential function which is given as:

$$\gamma(h) = c_0 + c\left\{1 - \exp\left(-\frac{h}{a}\right)\right\} \text{ for } 0 < h \quad (14)$$

$$= 0 \text{ for } h = 0$$

The parameters have the same explanation as in equation (14) whiles *a* is a distance parameter. The exponential function approaches the sill asymptotically at an effective range usually given as $r' = 3a$.

The Gaussian covariance function is given as:

$$\gamma(h) = c_0 + c\left\{1 - \exp\left(-\frac{h^2}{a^2}\right)\right\} \text{ for } 0 < h \qquad (15)$$

$$= 0 \text{ for } h = 0$$

The parameters have the same meaning as in the exponential model. The effective range is given as $r' = \sqrt{3}a$

4.8. Spatial Predictions

We carried out spatial predictions at 1km square grid using the prediction equation (Diggle & Ribeiro, 2007, pg. 37):

$$\hat{S}(x) = \mu(x) + \sum_{i=1}^{n} w_i(x)(y_i - \mu(x)) \qquad (16)$$

where $\hat{S}(x)$ is the regression kriging predictor at location x , $\mu(x)$ is the mean or trend, $w_i(x)$ is a function of the covariance parameters and $(y_i - \mu(x))$ is the interpolated residuals.

4.9. Evaluation of Interpolation Methods

Due to the sparse nature of the data available, a separate dataset was not created for validation. The leave-one-out cross-validation comparison is undertaken to evaluate the prediction performance of the methods. The cross-validation procedure is carried out by holding one data point using the remaining dataset to predict the value at that point(Hengl, 2007). This procedure is repeated for all the data points. The mean error and the root mean square error are obtained using the difference between the observed and the predicted values. Performance assessment is done by comparing the mean error (ME), root mean square error (RMSE) and the coefficient of multiple determination (R-square). The R-square indicates the amount of variability explained by the model. The best performing method is chosen as the one with the minimum RMSE (Bostan et al., 2012) and maximum R-squared. The mean error which is a measure of the prediction bias is expected to be close to zero for unbiased methods. The root mean square error which is also a measure of the prediction precision is expected to be small (Odeha, McBratney, & Chittleborough, 1994).

The ME and RMSE are calculated using the following formulas (Bishop et al., 2000):

$$ME = \frac{1}{n}\sum_{1}^{n}\{y(x_i) - \hat{y}(x_i)\} \qquad (17)$$

$$RMSE = \sqrt{\left[\frac{1}{n}\sum_{i=1}^{n}\{y(x_i) - \hat{y}(x_i)\}^2\right]} \qquad (18)$$

The R-square is also calculated using the equation below:

$$R - square = 1 - \frac{SSerr}{SStot} \quad SSerr = \sum_{i=1}^{n}(y(x_i) - \hat{y}(x_i))^2 \text{ and}$$

$$SStot = \sum_{i=1}^{n}(y(x_i) - \bar{y})^2 \qquad (19)$$

where $y(x_i)$ is the observed precipitation value, $\hat{y}(x_i)$ is the predicted precipitation value, $SSerr$ is sum of squares of the residuals, $SStot$ is the total sum of squares, $\bar{y}$ is the mean of the observations and *n* is the total number of data points.

5. RESULTS AND ANALYSIS

In this chapter, we outline the findings of the methodology carried out in chapter 4. We commence with the result of both the spatial and non-spatial exploratory data analysis, followed by regression analysis to select significant predictors for modeling and prediction. We conclude with a cross-validation assessment of the methods applied in this study.

5.1. Non-Spatial Exploratory Analysis

The distribution of the data was investigated by carrying out exploratory data analysis. The data for most of the months were not normally distributed and a log transformation was undertaken to achieve approximate normal distribution. For months with zero values of precipitation, we added 1 to all the values before the transformation since the log transformation of zeros is not possible. After prediction and back transformation to the original scale of the data, we subtracted one from the values to obtain the original data. Table 5.1 shows the descriptive statistics of the log-transformed values of the mean monthly and annual precipitation data. The transformation did not achieve a complete improvement in the normality of the distribution of the data for all months. We decided to use the transformed data for analysis as the scope of the appropriateness of the Gaussian model can be broadened by assuming that the model is still valid when the response variable is transformed (Diggle & Ribeiro, 2007, pg. 60). There is quite some similarity between the mean and the median values, and a reduced level of skewness after the transformation as shown in the table below.

Table 5.1: Descriptive statistics of the log-transformed mean monthly and annual precipitation data

	Mean	**Median**	**SD**	**Max**	**Min**	**Skew**	**Kurt**
January	2.20	2.59	1.27	3.92	0	-0.65	-1.08
February	3.01	3.12	1.0	4.52	1.22	-0.19	-1.42
March	3.94	4.02	0.95	5.01	1.75	-0.88	-0.36
April	4.74	4.82	0.38	5.26	3.69	-1.01	0.56
May	4.95	4.95	0.31	5.70	4.42	0.32	-0.37
June	5.21	5.22	0.27	6.14	4.87	1.37	2.98
July	4.95	4.98	0.36	5.49	4.20	-0.41	-0.86
August	4.59	4.5	0.83	5.68	2.91	-0.42	-1.05
September	5.04	5.20	0.45	5.70	3.96	-1.18	0.45
October	4.83	4.99	0.46	5.46	4.10	-0.34	-1.47
November	3.37	3.76	1.22	5.07	0.88	-0.42	-1.15
December	2.31	2.71	1.51	4.27	0	-0.44	-1.40
Annual	7.09	7.15	0.22	7.58	6.55	0.03	-0.11

Graphical exploration of the data was carried out using histograms and Q-Q plots to study the structure of the data and to detect possible outliers. Figure 5.1 and Figure 5.2 below show the histograms and Q-Q plots of the original data and the log-transformed data respectively. Log-transformation of the annual data was also carried out as shown in Figure 5.3 and Figure 5.4.

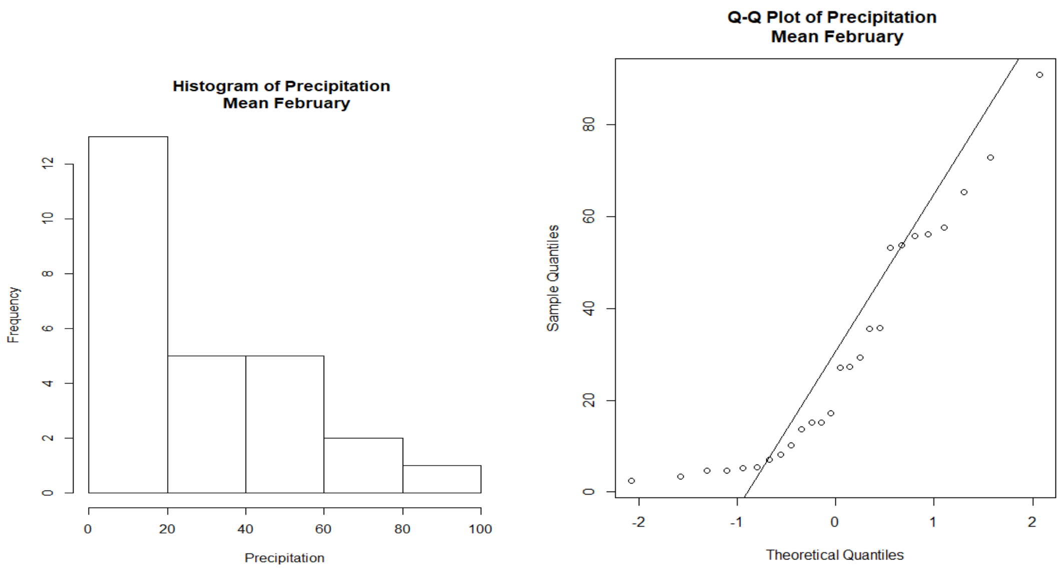


Figure 5.1: Histogram and Q-Q plot of the original precipitation data for the month of February

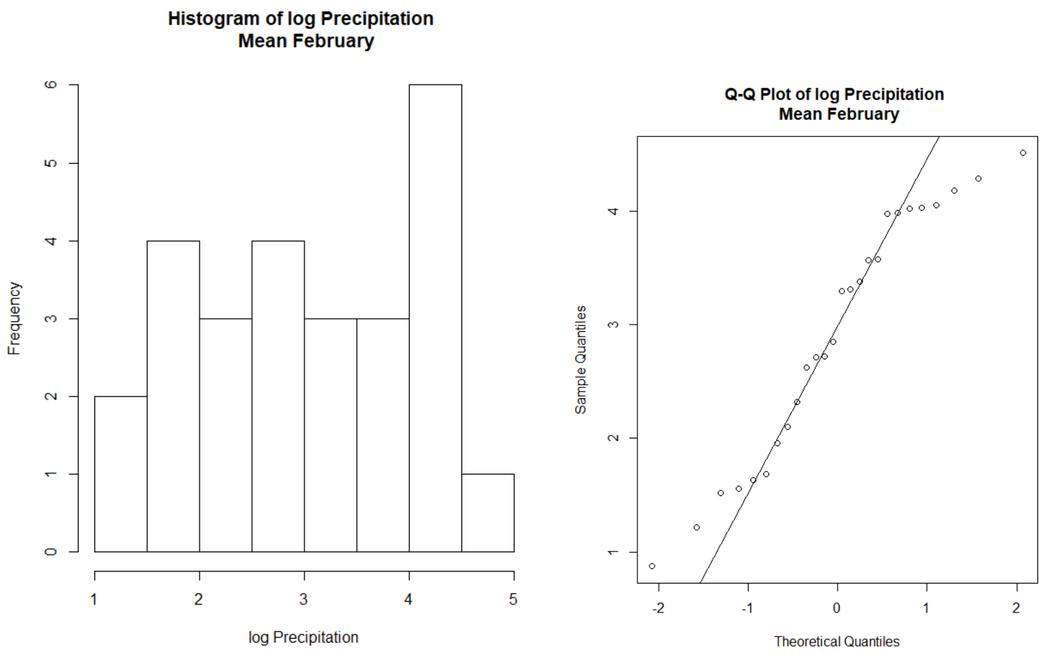


Figure 5.2: Histogram and Q-Q plot of the log-transformed precipitation data for the month of February

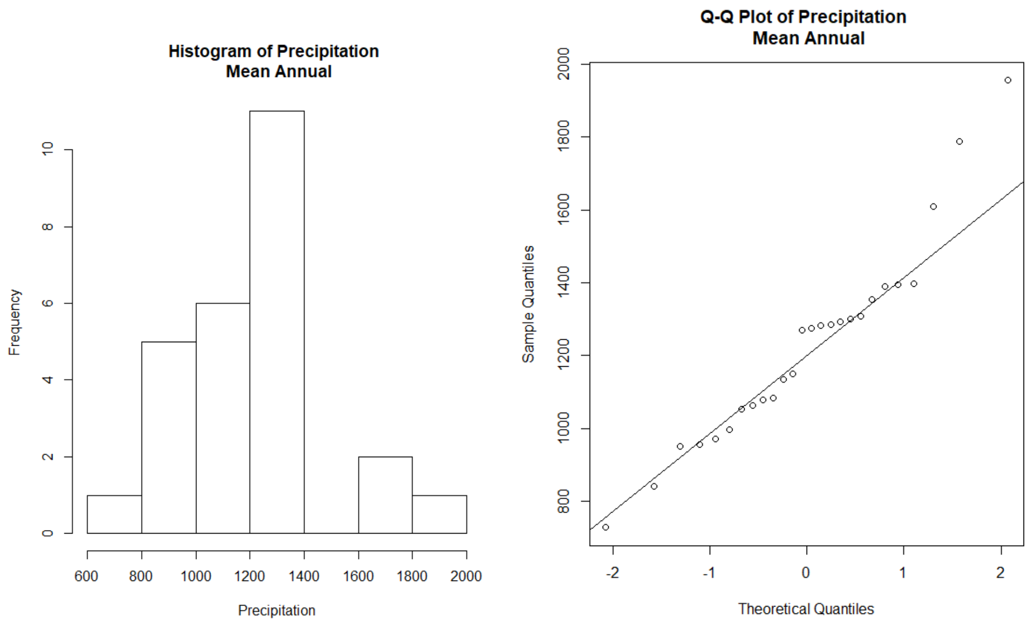


Figure 5.3: Histogram and Q-Q plot of mean Annual precipitation data

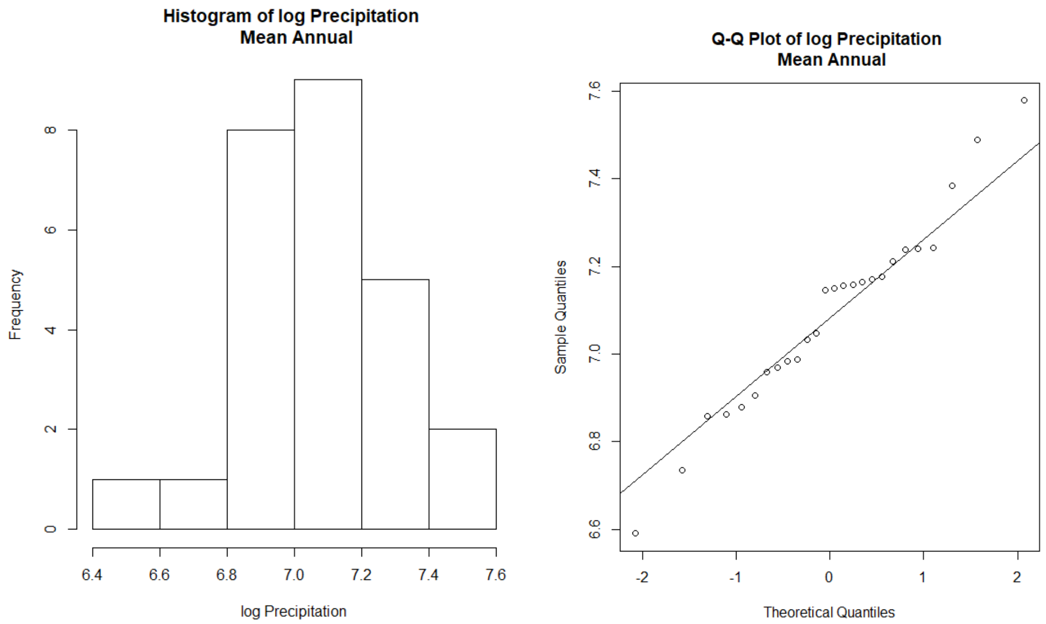
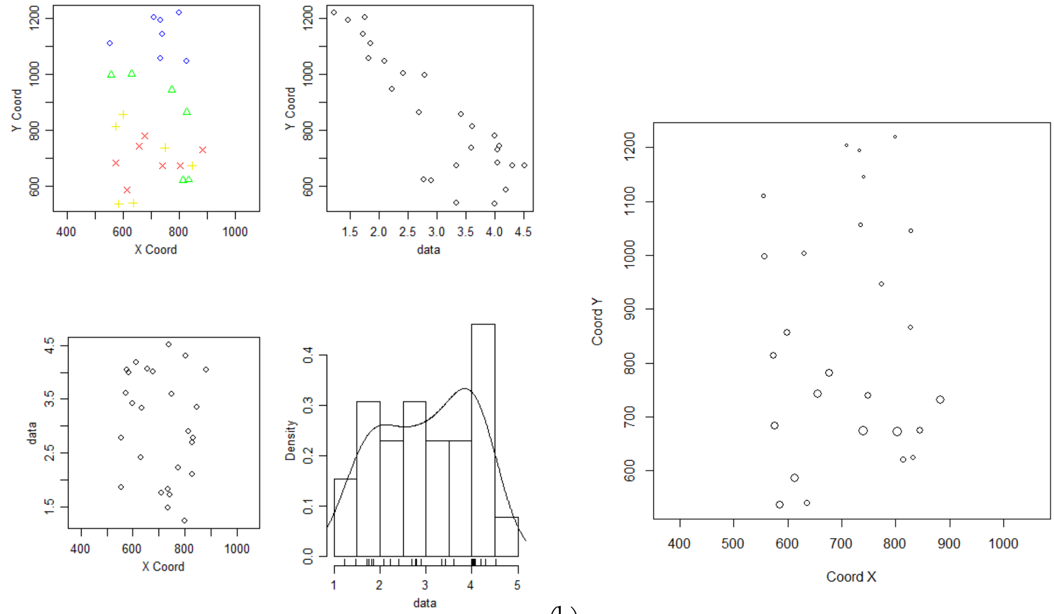


Figure 5.4: Histogram and Q-Q plot of the log-transformed annual precipitation data

5.2. Spatial Exploratory Analysis

The graphical plots in Figure 5.5 (a) show a plot of the spatial locations of the log-transformed precipitation data, their correlation with the Y coordinate (Northing), their correlation with the X coordinates (Easting) and a histogram of the data respectively. This spatial exploratory analysis of the data for the month of February shows a high negative correlation of the data with the Northing. This suggested the need for the inclusion of a trend surface model in this analysis. Figure 5.5 (b) also shows the plot of the data in proportion to the data values indicating the decreasing trend of precipitation from the south to the north. Figure 5.6 (a) and (b) also show the spatial exploratory plots of the log-transformed annual precipitation with NDVI and the Easting as a trend. Though there is temporal variation in precipitation which is captured by the monthly analysis, the spatial exploratory analysis of the annual precipitation also shows the decreasing trend of precipitation from the south-west to north-east.



(a) (b)

Figure 5.5: Graphical plots of spatial exploratory data analysis for log-transformed mean February precipitation

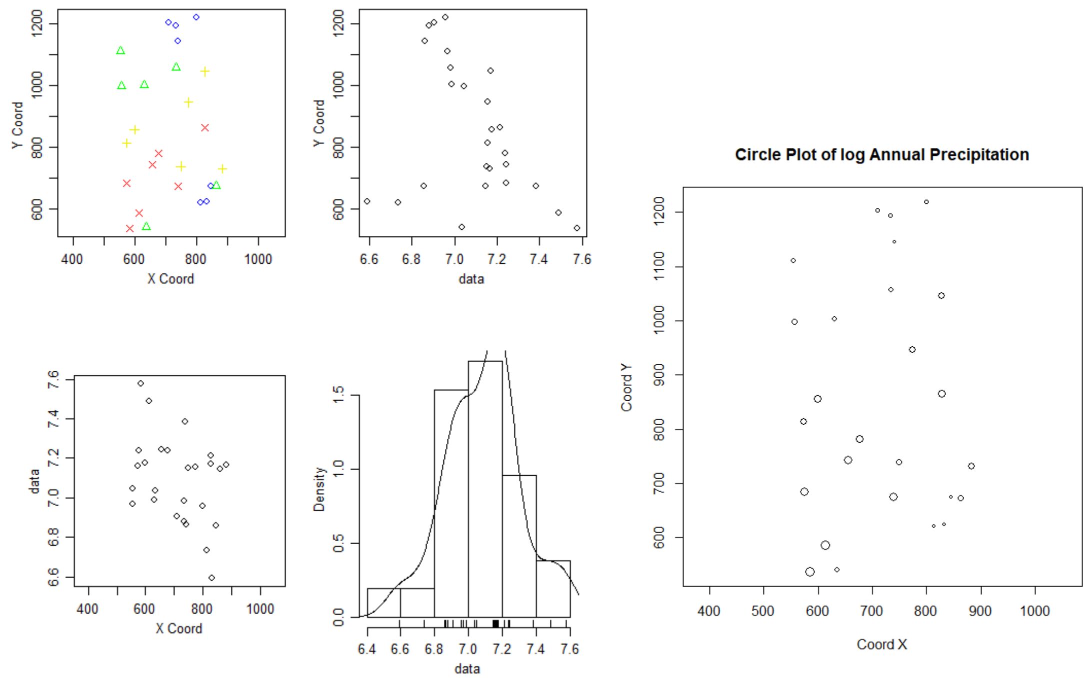


Figure 5.6: Graphical plots of spatial exploratory data analysis for the log-transformed mean annual precipitation

5.3. Regression Analysis

Spatial exploratory analysis of the data revealed the presence of a trend signifying a spatially varying mean. The spatial coordinates, elevation, and NDVI were used as predictors or trend to model the mean component of the kriging models. Regression analysis was carried out between log-transformed precipitation and the predictors to identify significant predictors for the different regression models as indicated in Section 4.4.

The correlation between the log-transformed precipitation and each of the predictors for the monthly and annual data was also analyzed as shown in table 5.2.

Table 5.2: Correlation coefficients of log-transformed precipitation with covariates

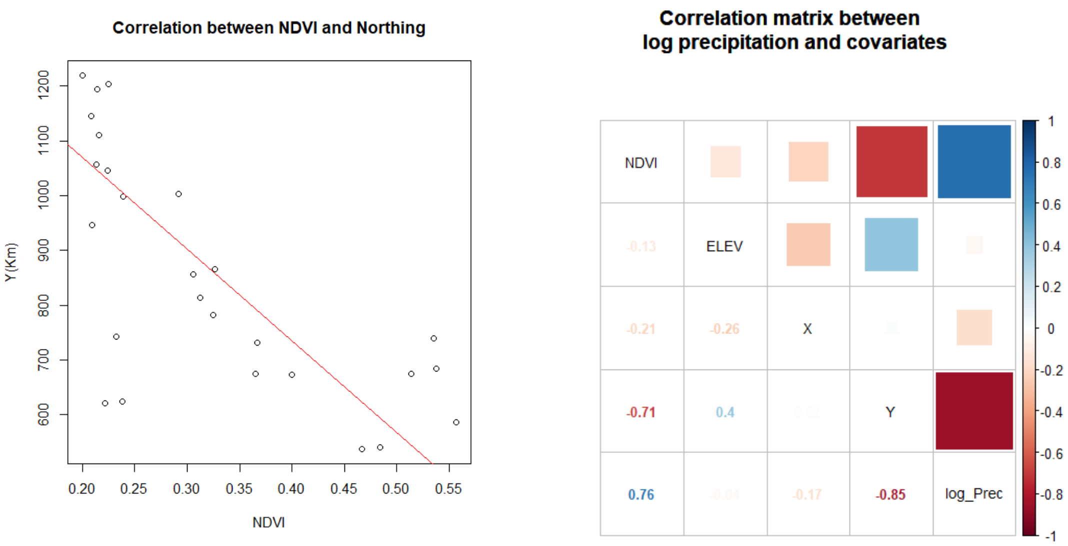
	NDVI	ELEV	X	Y
January	0.66	-0.12	-0.03	-0.89
February	0.78	-0.04	-0.17	-0.85
March	0.74	0.01	-0.15	-0.85
April	0.68	0.11	-0.42	-0.63
May	0.64	-0.27	-0.39	-0.86
June	0.34	-0.33	-0.17	-0.70
July	-0.004	-0.03	-0.20	0.56
August	0.50	0.47	-0.10	0.87
September	0.41	0.65	-0.05	0.55
October	0.40	0.13	-0.35	-0.61
November	0.67	-0.18	-0.19	-0.92
December	0.74	-0.21	-0.06	-0.93
Annual	0.60	0.13	-0.38	-0.36

There is a moderate positive correlation between log-transformed precipitation and NDVI for January, March, April, May, August, November, and December ranging from 0.50 to 0.74 and a strong correlation for February. There is a weak correlation for the months of September and October and a very weak correlation in June and July. The correlation was weak for months with higher rainfall and stronger for months with less rainfall. There is a very strong correlation between log-transformed precipitation and the northing for all months with correlation coefficients greater than 0.5 or -0.5. Generally, there was a very weak correlation for the elevation and Easting covariates with precipitation except in September for elevation and in April for the Easting. Notwithstanding the weak correlation of elevation with precipitation, elevation showed significance in the regression model for January, February, March, April, September, and October with p-value from 0.037 (*p<0.05) in October to 0.0008 (*p<0.001) in March.

There was multicollinearity between NDVI and the Northing which affected the significance of NDVI as a predictor in the regression Model 1 for most of the months. NDVI only showed significance for the months where its correlation with the Northing was weak and moderate. However, NDVI was very significant as a predictor when precipitation was regressed on NDVI only in Model 3 except for the months of June and July where the correlation with precipitation was very weak. Table 5.3 below shows the coefficient of correlation between NDVI and the Northing. Figure 5.7 (a) shows the correlation plot between NDVI and the northing and (b) the correlation matrix between log precipitation and all covariates for the month of December.

Table 5.3: Coefficient of correlation between NDVI and the Northing

Period	Correlation
January	-0.65
February	-0.71
March	-0.71
April	-0.68
May	-0.61
June	-0.55
July	-0.29
August	0.40
September	0.25
October	-0.05
November	-0.54
December	-0.69
Annual	-0.27



(a) **(b)**

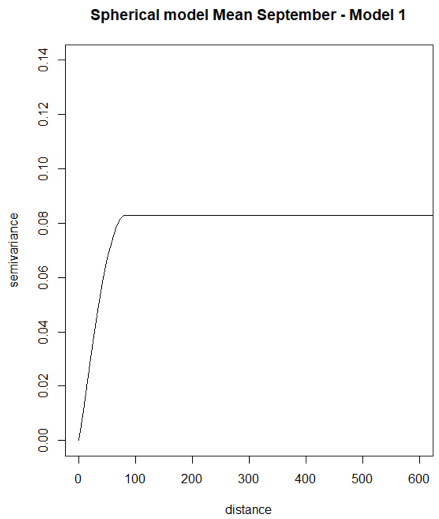
Figure 5.7: (a) Scatter plot of the correlation between NDVI and the Northing (b) correlation matrix between log precipitation and all covariates for the month of February

5.4. Parameter Estimation by Maximum Likelihood

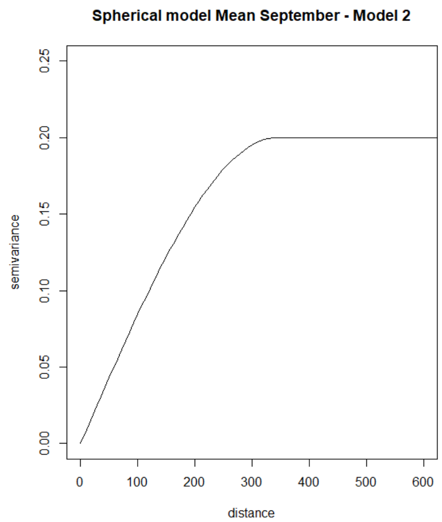
5.4.1. Exploratory Variographic Analysis

The spatial structure of precipitation for the mean monthly and annual precipitation was analyzed using different covariance models. The covariance models used in this analysis include the spherical model, the exponential model, and the Gaussian model. Since our approach is purely model-based using the maximum likelihood method of parameter estimation, the empirical variogram was not used as the basis for inference but only as an exploratory tool to select plausible parametric models for the covariance structure and also to identify initial covariance parameters (Diggle & Ribeiro, 2007). The model with the maximum log-likelihood function was selected for spatial prediction.

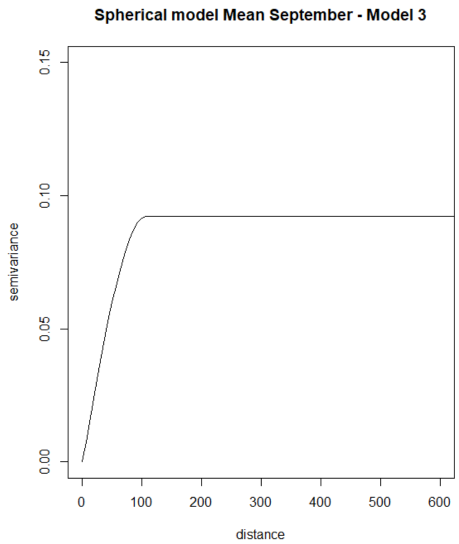
Figure 5.8 (a), (b) and (c) below show the empirical variogram and a fitted spherical model for mean September and mean annual log precipitation after removing the trend. These were used for exploratory purpose to identify initial parameters for covariance parameter estimation and prediction.



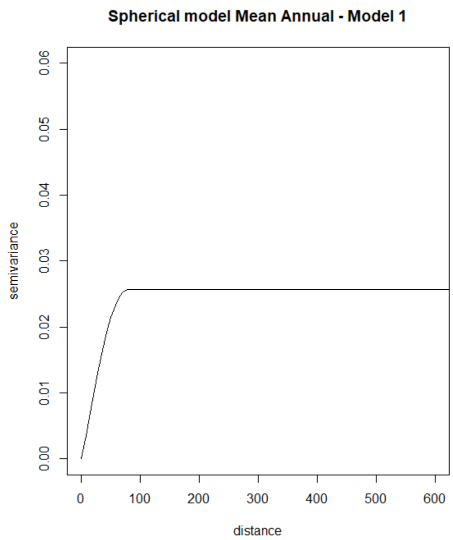
(a)



(b)



(c)



(d)

Figure 5.8: Empirical variogram and a fitted spherical model for mean September and Mean Annual precipitation when the trend in (a) elevation and Northing, (b) Northing (c) NDVI and elevation (d) NDVI and the Easting were removed

5.4.2. Model Selection

The covariance parameters needed for the accurate estimation of precipitation at unsampled locations is largely dependent on the type of covariance function that best describes the spatial structure of precipitation in the study area. For each of the kriging methods employed, the appropriate covariance function was selected based on the maximum value (Myung, 2003) of the maximized log-likelihood function. Table 5.4, 5.5 and 5.6 provide the details of the preferred covariance function for each period in Model 1, Model 2 and Model 3 respectively.

Table 5.4: Maximized log-likelihood values for covariance functions in Model 1

Period	**Maximized Log-likelihood**			**Preferred function**
	Exponential	Spherical	Gaussian	
January	-19.07	-18.58	-18.69	Spherical
February	-13.06	-11.52	-13.06	Spherical
March	-9.48	-9.57	-9.64	Exponential
April	3.246	3.676	3.744	Gaussian
May	22.6	22.6	22.6	Spherical
June	7.521	7.763	6.602	Spherical
July	0.1891	-0.00085	0.6254	Gaussian
August	-10.44	-10.16	-9.786	Gaussian
September	-4.3	-3.889	-5.555	Spherical
October	-0.5967	0.0086	0.8181	Gaussian
November	-13.26	-12.99	-13.02	Spherical
December	-18	-18	-18	Spherical
Annual	10.91	11.29	11.62	Gaussian

Table 5.5: Maximized log-likelihood values for covariance functions in Model 2

Period	**Maximized Log-likelihood**			**Preferred function**
	Exponential	Spherical	Gaussian	
January	-21.01	-20.37	-20.42	Spherical
February	-16.79	-16.68	-16.05	Gaussian
March	-14.34	-14.22	-14.09	Gaussian
April	2.449	3.115	3.248	Gaussian
May	22.2	22.27	22.2	Spherical
June	7.253	7.034	7.002	Exponential
July	0.179	0.7822	0.6254	Spherical
August	-11.66	-11.49	-11.05	Gaussian
September	-6.084	-5.897	-6.82	Spherical
October	-3.108	-2.986	-1.803	Gaussian
November	-14.73	-14.49	-15.28	Spherical
December	-21.05	-20.91	-21.05	Spherical
Annual	9.095	9.634	10.1	Gaussian

Table 5.6: Maximized log-likelihood values for covariance functions in Model 3

Period	Maximized Log-likelihood			Preferred function
	Exponential	Spherical	Gaussian	
January	-25.13	-25.9	-24.42	Gaussian
February	-18.41	-17.68	-17.45	Gaussian
March	-14.82	-13.84	-12.08	Gaussian
April	1.164	1.862	2.16	Gaussian
May	11.8	11.72	13.96	Gaussian
June	3.031	3.47	0.9348	Spherical
July	-0.9911	-0.4444	-0.1757	Gaussian
August	-16.99	-16.46	-16.56	Spherical
September	-4.683	-4.592	-4.617	Spherical
October	-3.568	-3.069	-4.209	Spherical
November	-20.03	-19.27	-16.84	Gaussian
December	-27.97	-27.25	-25.31	Gaussian
Annual	9.095	9.634	10.1	Gaussian

5.4.3. Model Parameters

The covariance models that are selected based on the criteria in Section 5.4.2 after the optimisation of the likelihood function provide the parameters of the linear effect and covariance structure. It is important to note that in the maximum likelihood estimation method, both the trend coefficients and the covariance parameters are estimated concurrently. An optimisation criterion using the "*likefit*" function of the geoR package gives the maximum likelihood estimates of the parameters in each of the models. Table 5.7, 5.8, and 5.9 show the estimated parameters of the mean, the signal variance, the nugget variance and the range in each of the kriging models applied for each period. The abbreviations SPH = Spherical, EXP = Exponential and GAU = Gaussian.

Table 5.7: Selected covariance functions and model parameters in Model 1

Period	Coefficients					Psill	Rang (Km)	Nug.	Loglik	Model
	β_0	β_1	β_2	β_3	β_4					
January	6.4869		0.0021		-5.473	0.176	139.3	0.0871	-18.58	SPH
February	6.4487		0.0028		-4.655	0.1459	65.13	0	-11.52	SPH
March	7.0543		0.0031		-4.3484	0.1223	16.67	0	-9.479	EXP
April	6.5322		0.0005	-1.335	-1.0690	0.0547	73.87	0.0039	3.744	GAU
May	6.6173	0.1477		-1.078	-1.143	0.0009	76.62	0.0094	22.6	SPH
June	6.2099	-0.257			-1.0156	0.0342	86.06	0	7.763	SPH
July	4.2933	-0.001			0.8141	0.0799	205	0.0265	0.6254	GAU
August	1.7500	0.8574			2.9379	0.1295	38.38	0	-9.786	GAU
September	4.1622		0.0016		0.7151	0.0829	80.43	0	-3.889	SPH
October	5.5872	0.6900	0.0012		-1.5741	0.0654	53.12	0	0.8181	GAU
November	6.1269	1.8472			-4.3421	0.0556	361.6	0.1164	-12.99	SPH
December	7.7265		0.0023		-6.8966	0.0038	50.07	0.2301	-18	GAU
Annual	6.9769	0.9094		-0.431		0.0259	39.44	0	11.62	GAU

Table 5.8: Selected covariance functions and model parameters in Model 2

Period	Coefficient			Psill	Range (Km)	Nugget	Loglik	Model
	β_0	β_1	β_2					
January	6.403		-4.887	0.2569	158.3	0.065	-20.37	SPH
February	6.5351		-4.111	0.235	50.85	0	-16.05	GAU
March	6.5309		-3.180	0.211	157.3	0.0726	-14.09	GAU
April	6.5849	-1.5040	-0.8732	0.0648	78.94	0.0021	3.248	GAU
May	6.7617	-1.1034	-1.2065	0.0027	74.36	0.0078	22.27	SPH
June	5.9992		-0.9231	0.0349	31.11	0	7.253	EXP
July	4.2806		0.8229	0.0676	311.7	0.015	0.7822	SPH
August	1.9277		3.2351	0.1504	47.59	0	-11.05	GAU
September	3.4801		1.7216	0.2001	344.7	0	-5.897	SPH
October	6.6084	-0.9154	-1.2790	0.0885	63.74	0	-1.803	GAU
November	9.1208	-1.7897	-5.2613	0.1837	67	0	-14.49	SPH
December	7.8075		-6.4601	0.169	59.67	0.1252	-20.91	SPH
Annual	7.8412	-0.5147	-0.4258	0.0303	45.87	0	10.1	GAU

Table 5.9: Selected covariance functions and model parameters in Model 3

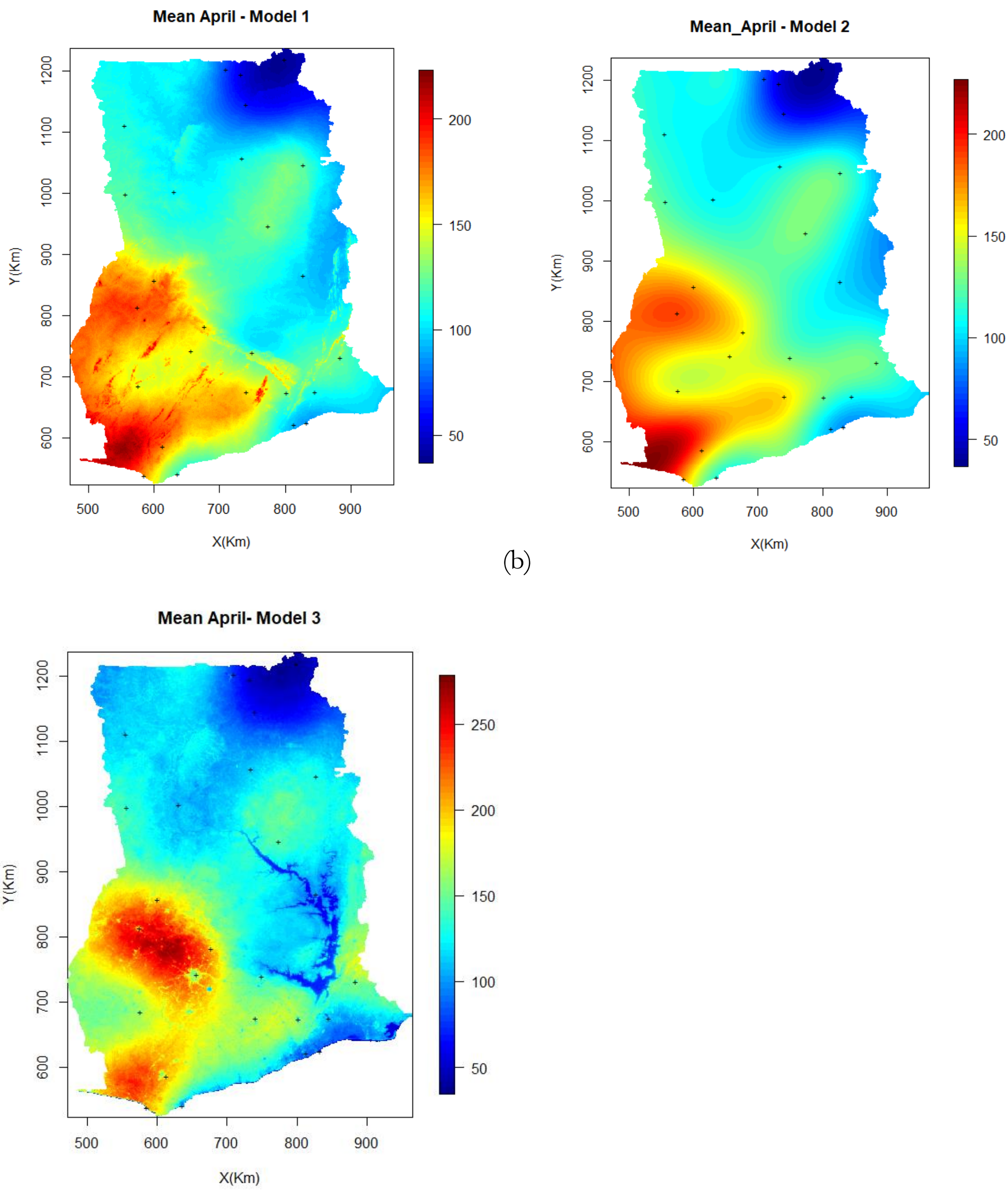
Period	Coefficients			Psill	Range (Km)	Nugget	Loglik	Model
	β_0	β_1	β_2					
January	1.627	1.4589		0.8872	152.1	0.0744	-24.42	GAU
February	1.8697	2.5109		0.4662	358.9	0.1298	-17.45	GAU
March	2.3183	1.3942		1.831	620	0.0706	-12.08	GAU
April	4.3616	0.9369		0.0803	81.32	0	2.16	GAU
May	5.0923	0.2366		0.5246	1134	0.0107	13.96	GAU
June	5.3512	-0.118		0.0709	786.6	0.0205	3.47	SPH
July	5.0259	-0.028		0.1349	249	0.0261	-0.1757	GAU
August	4.291	0.5432		0.6673	944	0.046	-16.46	SPH
September	4.249	0.9162	0.0018	0.0922	106.6	0	-4.592	SPH
October	4.4456	0.5287		0.2262	511.3	0	-3.069	SPH
November	2.0746	1.5535		1.41	560.1	0.1133	-16.84	GAU
December	1.1201	1.8042		1.924	563.8	0.2368	-25.31	GAU
Annual	6.66	0.9376		0.028	43.45	0	10.9	GAU

5.5. Spatial Prediction

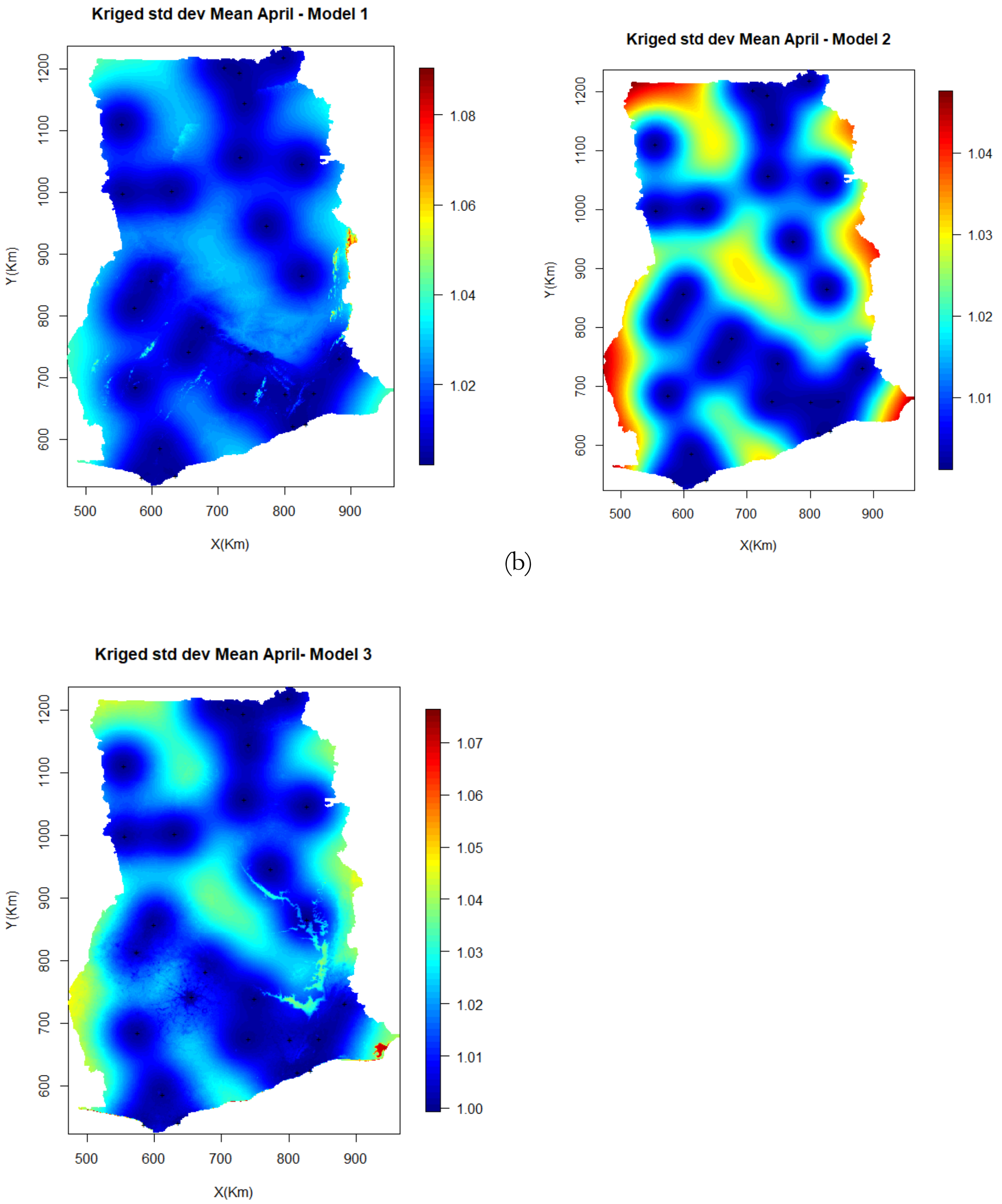
Spatial predictions were carried out for mean monthly, mean annual and cumulative annual rainfall for the period under analysis using the three kriging models.

5.5.1. Mean Monthly Prediction

Model 2 outperformed the other methods in five months (February, April, May, September, and October). Model 3 also performed better than the others in five months (March, July, August, November, and December). Model 1 was the least performing model and performed better in January and June. Figure 5.9 (a), (b) and (c) show the predictions for the month of April using Model 1, Model 2, and Model 3 respectively. The kriging standard deviation for the three models are shown in Figure 5.10. Figure 5.11 shows the time series mean monthly predictions for all the months using the preferred kriging model in each month.



(c)
Figure 5.9: Predicted precipitation maps for mean April using (a) Model 1 (b) Model 2 and (c) Model 3



(c)
Figure 5.10: Kriging standard deviation using (a) Model 1 (b) Model 2 and (c) Model 3

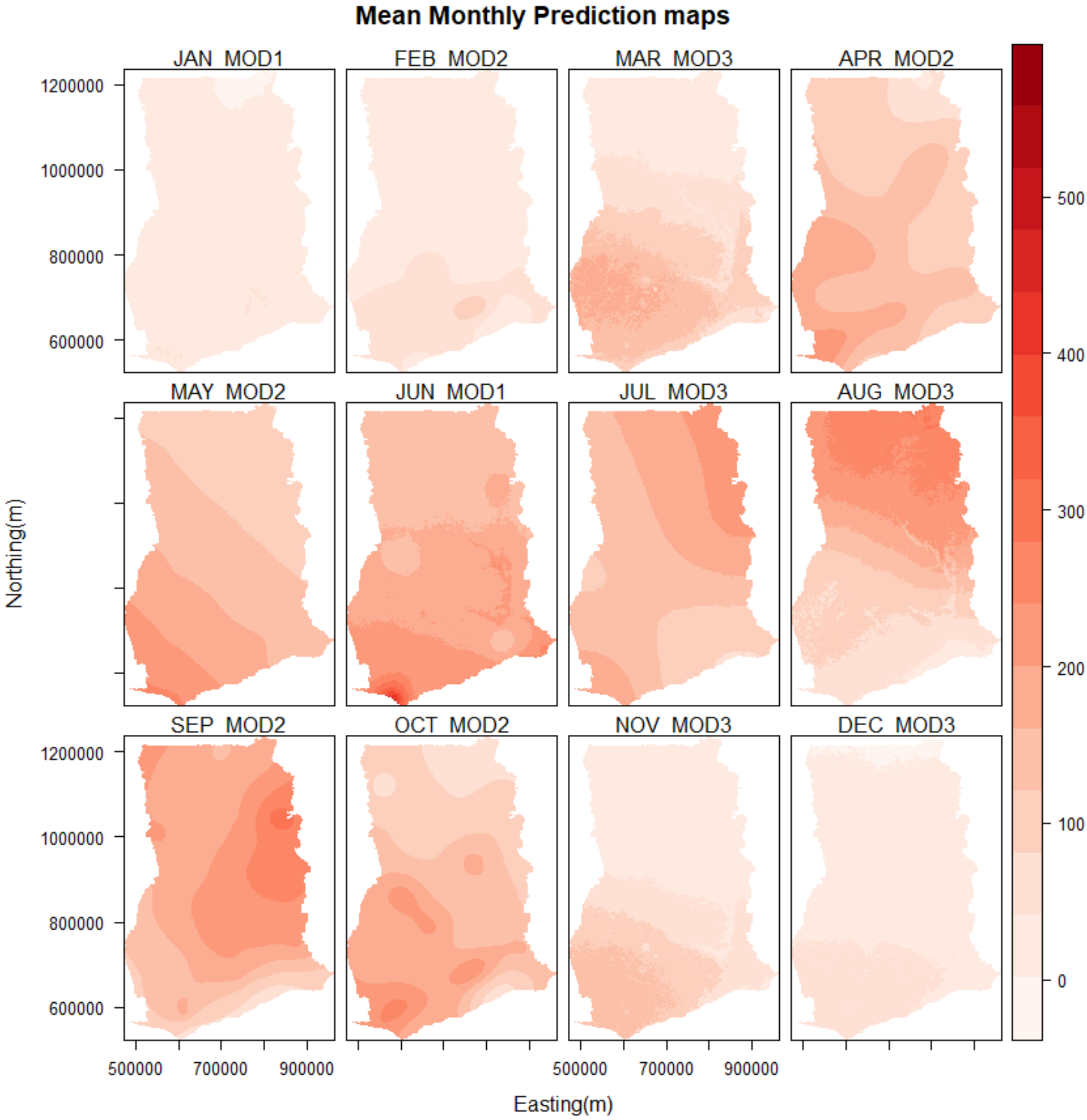
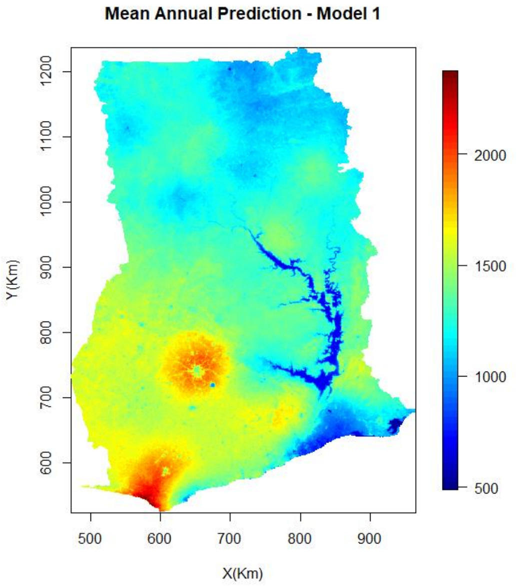


Figure 5.11: Predicted mean monthly precipitation maps of Ghana using Model 1, Model 2 and Model 3.

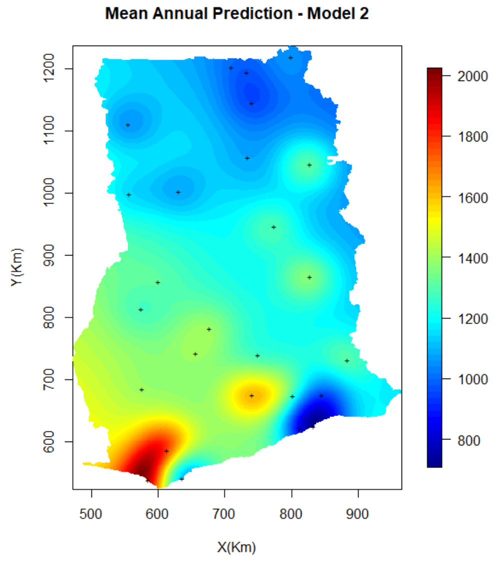
MOD1, MOD2 and MOD3 represent Model 1, Model 2 and Model 3 respectively.

5.5.2. Mean Annual Prediction

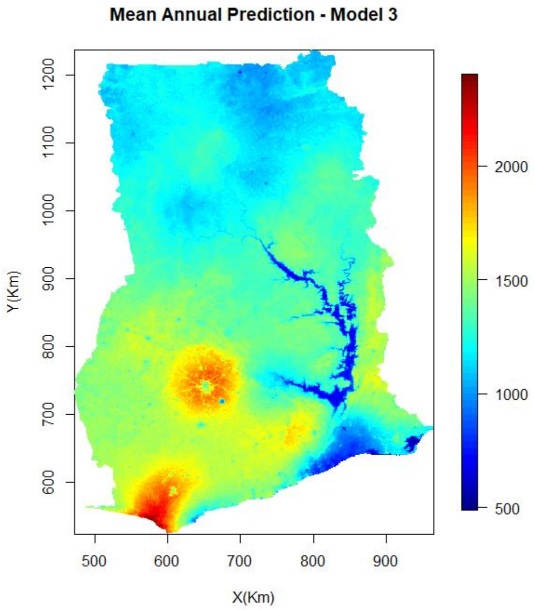
Model 2 outperformed Model 1 and Model 3 in the mean annual predictions. Though the spatial coordinates especially the northing were strongly correlated with the log-transformed precipitation for the monthly data which accounted for the performance of Model 2 in five months, their correlation with the log-transformed mean annual precipitation data was weak. Notwithstanding the weak correlation, Model 2 still outperformed Model 1 and Model 3. Figure 5.12 (a), (b) and (c) show the mean annual prediction maps using Model 1, Model 2 and Model 3 respectively.



(a)



(b)



(c)

Figure 5.12: Predicted Mean Annual precipitation maps using (a) Model 1 (b) Model 2 and (c) Model 3

5.5.3. Cumulative Annual Prediction

In order to investigate the cumulative dynamics of precipitation in Ghana, we carried out cumulative annual predictions from 2001 to 2010 using the three models and selected the best model for each year through cross-validation. Figure 5.13 shows the spatial and temporal distribution of cumulative annual rainfall in the country for the ten-year period.

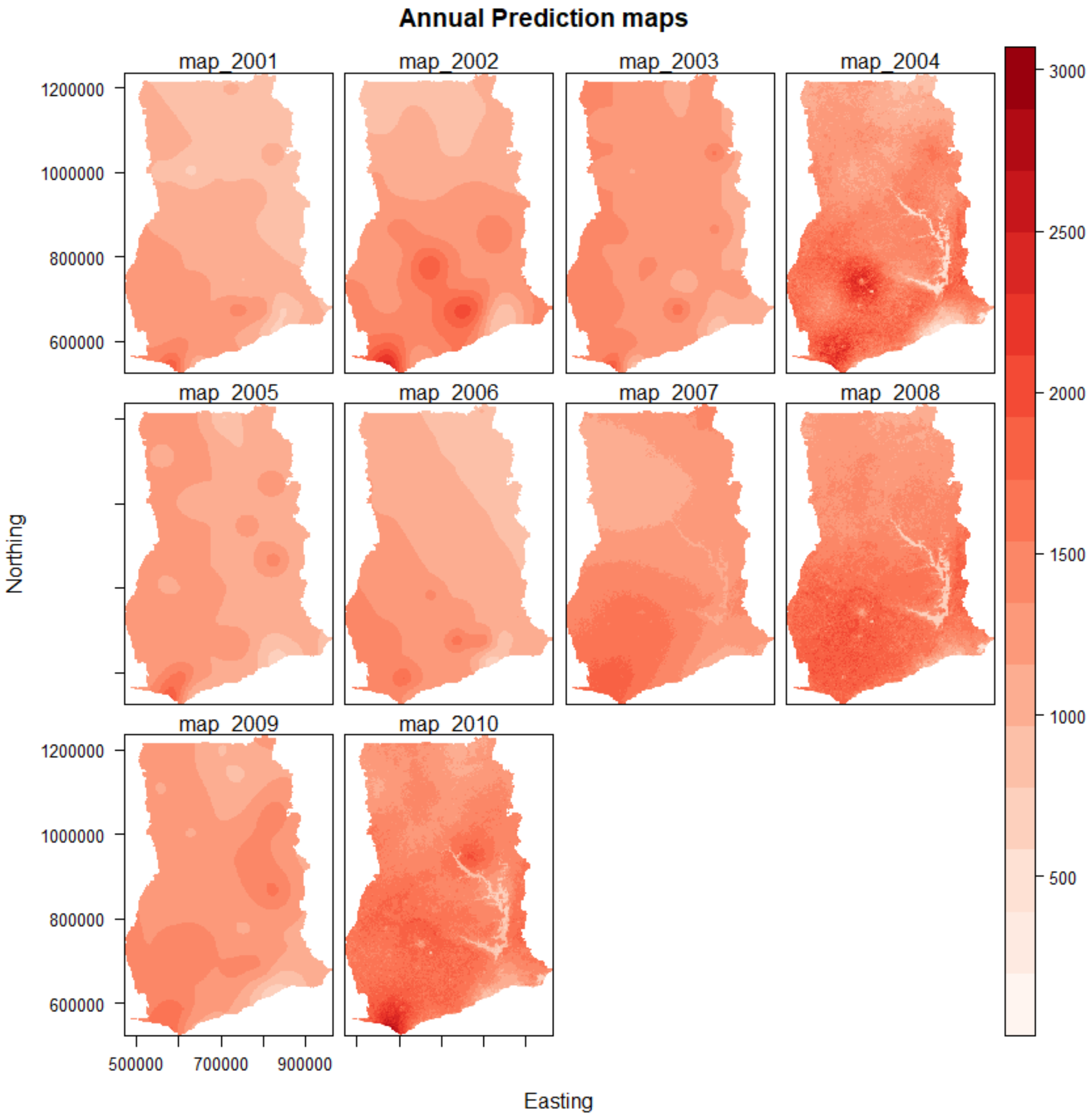


Figure 5.13: Cumulative annual prediction maps using Model 1, Model 2 and Model 3.

5.6. Performance Assessment of Models

Table 5.10 and Table 5.11 show the leave-one-out cross-validation statistics carried out to evaluate the performance of the three models in mean monthly and cumulative annual predictions. For mean monthly prediction, Model 2 performed better in five months (February, April, May, September, and October) with the minimum root mean square error and maximum coefficient of determination. Model 3 equally performed better in five months (March, July, August, November, and December) and Model 1 for two months (January and June) as shown in Table 5.10. For cumulative annual prediction, Model 2 performed better in 2001to 2004, 2007 and 2009. Model 3 performed in 2005, 2008 and 2010 and Model 1 for 2006.

Table 5.10: Cross-Validation statistics for the three Models in mean monthly and annual predictions

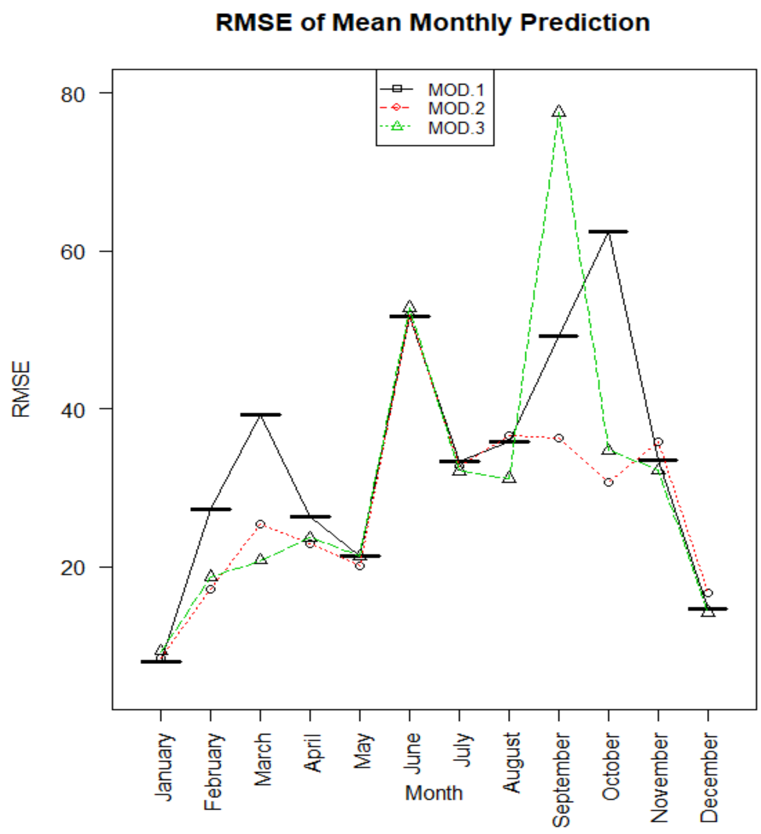
	Model 1			Model 2			Model 3			Preferred Method
	ME	RMSE	R sq.	ME	RMSE	R sq.	ME	RMSE	R sq.	
Jan	0.748	8.005	0.6041	1.155	8.5417	0.5536	0.392	9.3147	0.4602	Model 1
Fe	0.1991	27.321	0.4113	4.024	17.148	0.5566	1.856	18.758	0.4440	Model 2
Mar	-2.736	39.229	0.2238	2.862	25.453	0.6558	1.773	20.840	0.7815	Model 3
Apr	1.937	26.361	0.5433	2.802	22.973	0.6565	1.744	23.673	0.6317	Model 2
May	0.937	21.40	0.8024	1.423	20.083	0.8265	1.32	21.334	0.8040	Model 2
Jun	4.579	51.685	0.3598	4.492	51.835	0.3559	5.205	52.845	0.3320	Model 1
Jul	3.589	33.32	0.5317	4.02	32.688	0.5509	3.548	32.146	0.5643	Model 3
Aug	3.857	35.841	0.8304	4.898	36.610	0.8242	6.796	31.242	0.8758	Model 3
Sep	4.868	49.297	0.2878	3.542	36.334	0.6130	1.142	77.595	0.385	Model 2
Oct	-0.846	62.56	0.2480	6.699	30.670	0.7144	2.142	34.735	0.6168	Model 2
Nov	1.974	33.571	0.4756	1.853	35.799	0.4032	2.888	32.269	0.5177	Model 3
Dec	0.696	14.755	0.5003	0.598	16.66	0.2330	1.538	14.259	0.5377	Model 3
Ann	23.22	225.85	0.3088	35.49	209.96	0.4135	24.16	221.71	0.3347	Model 2

The abbreviations Jan = January, Feb = February etc and Ann = Annual

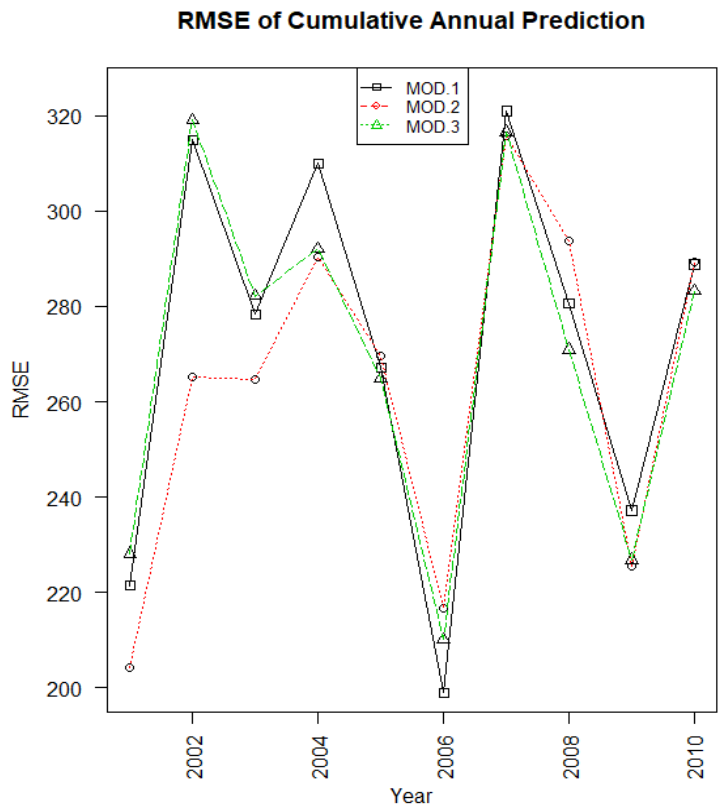
Table 5.11: Cross-Validation statistics for the three Models in cumulative annual predictions

	Model 1			Model 2			Model 2			Preferred Method
	ME	RMSE	R sq.	ME	RMSE	R sq.	ME	RMSE	R sq.	
2001	18.62	221.43	0.2098	24.13	204.26	0.3323	17.66	228.09	0.1606	Model 2
2002	20.52	314.96	0.2726	41.60	265.27	0.4946	27.88	319.04	0.2562	Model 2
2003	27.58	278.39	0.0342	35.94	264.68	0.0913	26.23	282.15	0.0230	Model 2
2004	34.72	309.96	0.2733	59.10	290.36	0.3810	32.31	292.15	0.3542	Model 2
2005	31.43	267.23	0.1621	41.90	269.64	0.1559	31.19	264.91	0.1766	Model 3
2006	20.51	198.95	0.51	30.30	216.79	0.4209	22.25	210.10	0.4515	Model 1
2007	40.65	320.96	0.1477	34.71	315.67	0.1723	30.63	316.63	0.1649	Model 2
2008	20.33	280.63	0.2117	32.31	293.82	0.1418	20.32	270.97	0.2653	Model 3
2009	33.97	237.18	0.2286	37.80	225.44	0.3085	33.12	226.86	0.2948	Model 2
2010	25.24	288.81	0.3295	40.32	289.30	0.3352	29.80	283.33	0.3569	Model 3

Figure 5.14 (a) and (b) show plots of the RMSE of prediction using Model 1, Model 2, and Model 3 for the mean monthly and cumulative annual prediction. The months with less precipitation recorded lower RMSE and the months with more precipitation recorded higher RMSE in each of the models. Though Model 2 performed better than the other models for five months, it was more bias than the other models for those months except September. Generally, all models showed different levels of underprediction with mean errors above zero. Model 1 also showed overpredictions with MEs below zero as a result of the incorporation of elevation in accounting for the trend. Figure 5.15 (a) and (b) also show plots of the mean error or bias in prediction for mean monthly and cumulative annual predictions.

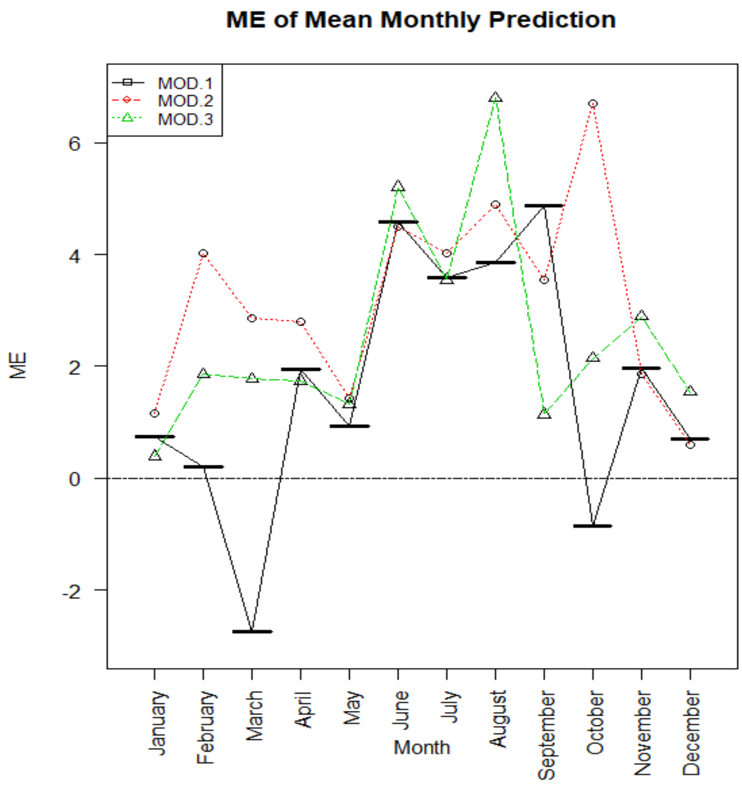


(a)

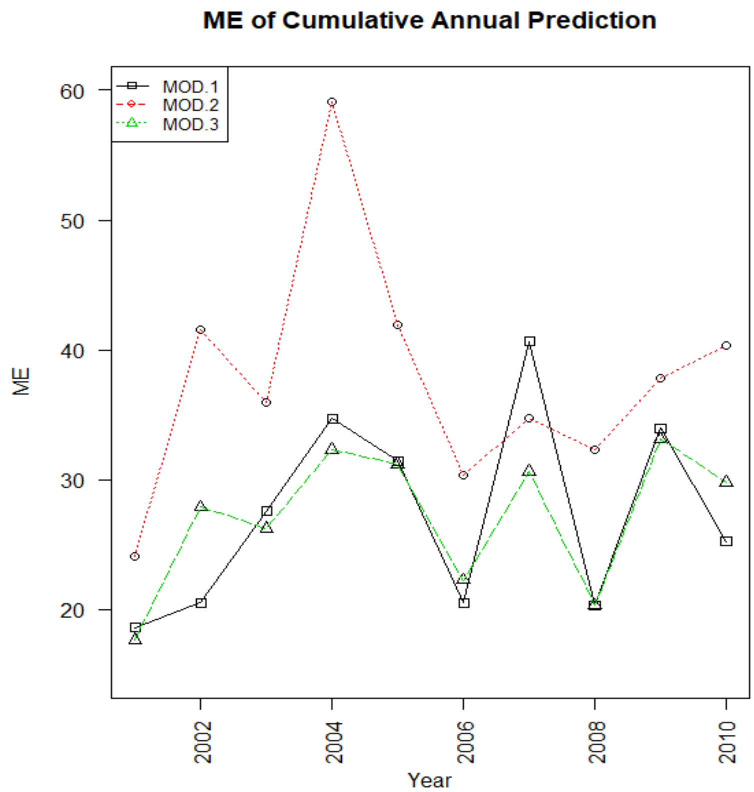


(b)

Figure 5.14: RMSE of prediction using Model 1, Model 2 and Model 3 for (a) mean monthly and (b) cumulative annual precipitation estimation



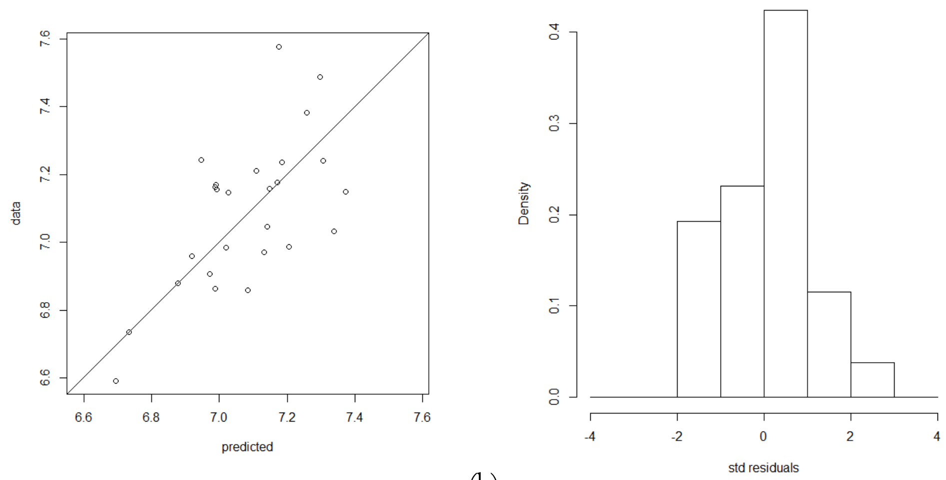
(a)



(b)

Figure 5.15: ME of prediction using Model 1, Model 2 and Model 3 for (a) mean monthly and (b) cumulative annual precipitation estimation

Figure 5.16 shows the cross-validation plots of the log-transformed mean annual precipitation. The scatter plot of the observed and the predicted shows a reasonable fit with some slight deviations as showing in Figure 5.16(a). The histogram of Figure 5.16(b) shows the residuals of the log-transformed mean annual precipitation. For a good model, the residuals are expected to be normally distributed.



(a) (b)

Figure 5.16: Cross-validation (a) scatter plot (b) histogram of residuals for log-transformed mean annual precipitation

5.7. Spatial Dependence

Figure 5.17 shows the monthly variation of the estimated sill and nugget effect for each of the interpolation models. The largest sill and nugget effect is observed mostly in the months with less precipitation as shown in Figure 5.17 (a), (b) and (c). This is as a result of the higher variation of precipitation in those months. The months with more precipitation have less sill and nugget effect.
Figure 5.18 also shows the nugget to sill ratio for each of the interpolation models in each month. According to Cambardella et al. (1994), a nugget to sill ration of 0 to 25% correspond to strong spatial dependence, 25% to 75% corresponds to moderate spatial dependence and above 75% correspond to weak spatial dependence.
Figure 5.19 shows the range of spatial dependence described by the three models. Model 3 with NDVI as a predictor, shows a higher range of spatial dependence beyond the scale of the study area especially in the months with more precipitation. This shows the existence of the trend beyond the scale of the study area.

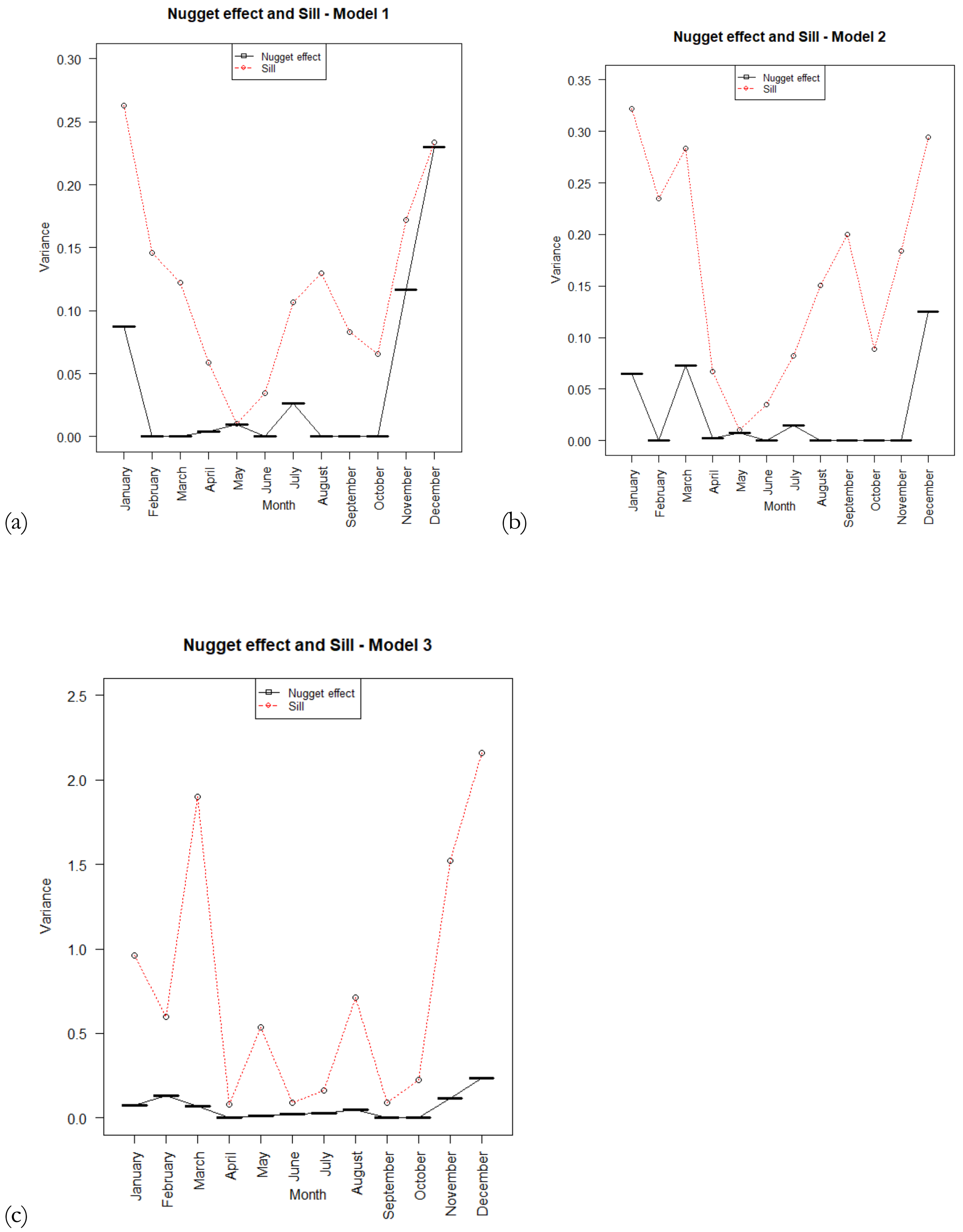


Figure 5.17: Nugget effect and total sill in (a)Model 1, (b) Model 2, and (c) Model 3 for each month

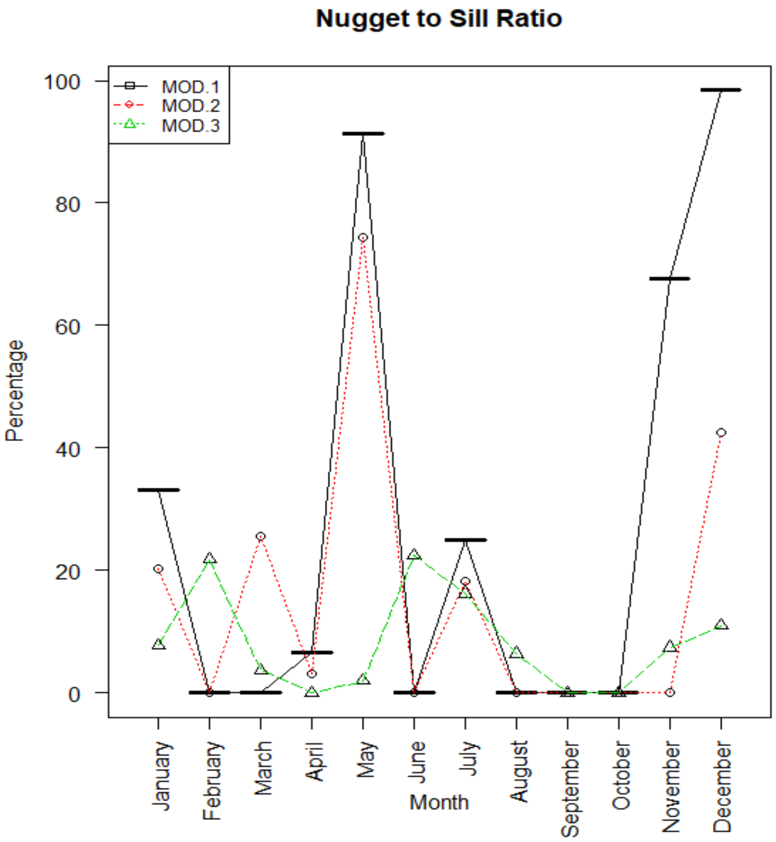


Figure 5.18: A graph of Nugget to Sill ratio in each month for Model 1, Model 2 and Model 3

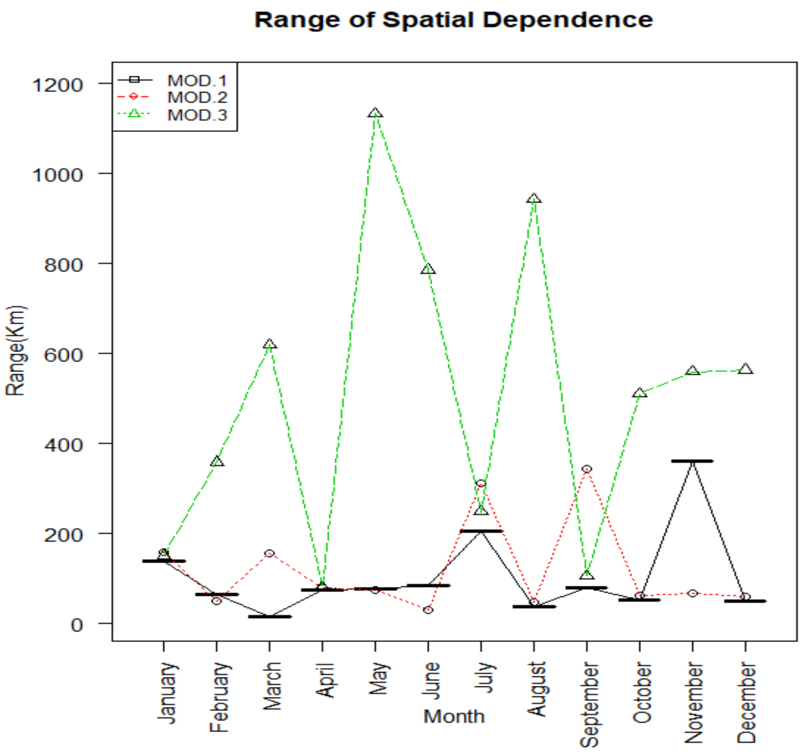


Figure 5.19: Range of spatial dependence in each month for Model 1, Model 2 and Model 3

6. DISCUSSION

In this chapter, we present a discussion on the results obtained through the methodology we applied to predict rainfall from sparse precipitation data.

6.1. Assessment of Predictors

The inclusion of elevation as a trend component in regression kriging did not contribute much in the prediction results for both monthly and annual precipitation. Although weakly correlated with precipitation, elevation showed significance with a p-value less than a significance level of 0.05 in January, February, March, April, September, October, and December in the multiple linear regression model. Model 1 with elevation as a predictor only performed better than the other models in January. There was an extreme over-prediction of precipitation in most of the months for which elevation showed significance and was included as a predictor. The over predictions were mainly observed in higher elevations as compared to lower elevations. The significance of elevation in the regression models for the abovementioned months could possibly be due to local influences. This was not a surprising discovery judging from the weak correlation between log-transformed precipitation and elevation presented in Table 5.2. The correlation of elevation with precipitation ranged from 0.01 to 0.65. Our findings are consistent with Goovaerts (2000), who found that ordinary kriging performed better than linear regression with elevation when the correlation between elevation and rainfall was moderate (ρ<0.75). A stronger correlation between elevation and rainfall is therefore necessary for elevation to bring improvement in spatial prediction of rainfall.

There was a moderate and strong correlation between NDVI and log-transformed precipitation ranging from 0.5 to 0.78 in most months except June and July as shown in Table 5.2. The weak correlation of NDVI with precipitation in June and July could be as a result of the fact that it rains everywhere in the country within that period. NDVI was simultaneously correlated with the northing coordinates making it less significant in Model 1 except the months with more rain. Model 3 outperformed Model 2 and Model 1 in March, July, August, November, and December when NDVI was incorporated as the drift. Though NDVI was weakly correlated with log-transformed mean July precipitation, Model 3 still performed better in July. Remotely-sensed derived covariate NDVI is, therefore, an optimal covariate for the accurate prediction of rainfall in the above-mentioned months.

The spatial coordinates especially the northings were highly correlated with log-transformed mean monthly precipitation and this accounted for the performance of Model 2 in five months. Their correlation with the log-transformed mean annual precipitation was weak. Though weakly correlated, Model 2 performed better in the mean annual prediction using the spatial coordinates as predictors. The spatial coordinates are therefore optimal covariates for mean annual prediction of rainfall in Ghana.

It could be observed in Figure 5.14(a) that all models performed similarly in May, June and July which are the months with more rainfall as shown in Figure 5.11. All models also performed similarly in cumulative annual predictions as shown in Figure 5.14(b). It is important to note that Model 2 with the spatial coordinates as predictors performed closely with Model 3 in the months that model 3 performed better using NDVI as a predictor. Though NDVI improved predictions in these months, the spatial coordinates could be used for overall monthly predictions in the absence of NDVI data.

6.2. Spatial Structure of Precipitation

The choice of a variogram model for spatial predictions depends largely on the covariance structure of the input data. It is therefore important to evaluate different models to infer the spatial structure of the input data for a given period. The spherical, exponential and Gaussian models were evaluated in this study and the appropriate model was selected for parameter estimation and prediction. The spherical and Gaussian models best captured the spatial structure of precipitation in the monthly and annual data. There was significant variation in the spatial dependence of precipitation from one month to the other resulting in the variation of variogram parameters. The months with less precipitation exhibited higher nugget effect as compared to months with more precipitation except for May with more precipitation and higher nugget effect. We may attribute this to the fact that it rains everywhere in the country during the month of May making it difficult to capture variability less than the minimum sampling distance. The nugget to sill ratio which describes the strength of spatial dependence (Cambardella et al., 1994) was high in the months of May and December as shown in Figure 5.18. The higher this ratio, the weaker the spatial dependence which is as a result of the high nugget effect. There was strong to moderate spatial dependence in most of the months especially months with high rainfall.

6.3. Spatial and Temporal Distribution of Precipitation

Precipitation varies both in space and time and this was clearly depicted in the monthly analysis. It was observed that precipitation varies spatially with decreasing trend from the south-west to the north-east. There is more precipitation in the south-western part of the country from January to June and from October to December than in the northern part. Axim located in the south-western part records the highest rainfall in the country and Bawku the lowest in the north-eastern within this period. The north-eastern part also depicted higher precipitation than the south in July and August whiles the easting part recorded higher precipitation in September as shown in Figure 5.11. This is in line with the rainfall pattern of Ghana with the southern part exhibiting a bimodal wet season from March to July (major season) and September to November (minor season) whiles the northern part experiences a single wet season from May to October as shown in Figure 5.11.

Precipitation shows an increasing trend from January to June at a maximum and declines steadily to August. It rises again to a maximum in September and finally declines to December. June is the month with the highest rainfall and January the lowest in the interpolation results. August is the transition between the major and the minor seasons in the southern part resulting in low precipitation in the south as compared to the north. This temporal variation of precipitation in the country among other factors could possibly be due to the influence of the two major air masses in the country. The south-west monsoon wind that hits the country beginning at Axim could explain why it records the highest rainfall in the country for most of the months. It could also be argued that the high rainfall record of Axim is as a result of its location along the coast. This argument would be flawed by the fact that towns like Tema and Accra are coastal towns yet are among the areas recording low rainfall in the country. Therefore, proximity to the coast has less or no fundamental effect on the rainfall variability in Ghana and a predictor like distance to the coast may not provide an improvement in rainfall predictions. The dry north-east trade wind commonly known as harmattan that blows from the Sahara Desert towards the north-eastern part of Ghana in December to March could possibly be the reason for the low precipitation rate in the north.

The result of the cumulative annual predictions indicates the spatial and temporal distribution of rainfall from 2001 to 2010. The results show that there was less rainfall in 2001 and more rainfall in 2010. This shows an increasing trend of rainfall in the country for that decade. This provides reliable information for drought analysis since drought-prone areas can easily be spotted.

7. CONCLUSION AND RECOMMENDATION

7.1. Conclusion

The main objective of our study was to predict rainfall from sparse precipitation data using regression kriging by applying model-based maximum likelihood method in estimating the drift and covariance parameters. We carried out a comparative analysis to evaluate the performance of three regression kriging models in estimating the mean monthly, mean annual and cumulative annual rainfall of Ghana through cross-validation measures using the RMSE and R-square metrics. Monthly and annual averages were computed from daily observations of precipitation from 2001 to 2010 using data from 26 rain gauge stations to carry out this analysis. The secondary variables used in this study include the spatial coordinates, MODIS NDVI, and elevation derived from DEM. The first model with significant predictors selected from a multiple linear regression of all the predictors performed better in two months (January and June) for mean monthly predictions and in 2006 for cumulative annual predictions. There was a strong correlation between mean monthly precipitation and the northings which accounted for the performance of the second model in five months (February, April, May, September, and October) for the mean monthly predictions, in six years (2001-2004, 2007 and 2009) for cumulative annual predictions and in the mean annual predictions. The third model with NDVI as predictor performed better in five months (March, July, August, November, and December) for mean monthly predictions and in three years (2005, 2008 and 2010) for cumulative annual predictions.
There was a slight difference between the RMSEs of the Model 3 with the spatial coordinates as predictors and Model 2 with NDVI as a predictor. We therefore consider MODIS NDVI as an valuable covariate for mean monthly rainfall estimation in Ghana. The RK model with elevation as a predictor only performed in January. There were overpredictions for most of the months where elevation was included as a predictor in the multiple linear regression. Elevation was not significant in the mean annual and cumulative annual predictions.

Answers to Research Questions

Specific Objective 1:

1. **What is the relationship between elevation and the spatial distribution of precipitation in the study area for a given time period?**

   The was a very weak correlation between precipitation and elevation for most of the months. Elevation was only moderately correlated with precipitation in September as shown in Table 5.2. Though weakly correlated with precipitation, elevation contributed to the improvement of rainfall predictions in only January when it was used as a predictor in a multiple linear regression with the northing coordinates.
   January is the driest month in Ghana and rainfall in this month can be partly due to orographic effects.

2. **Can time series of NDVI, a remote-sensed derived covariate from MODIS be used to infer the spatial and temporal distribution of precipitation in the study area?**

The one-month time series of MODIS NDVI was simultaneously correlated with the northing coordinates and was not significant for most of the months when it was used in a multiple linear regression with all predictors for Model 1.
NDVI became very significant for most months and years when it was used as a predictor in Model 3 except the months with more rainfall such as June and July.
The significance of NDVI led to the improvement of prediction in five months (March, July, August, November, and December) for the mean monthly predictions and three years (2005, 2008 and 2010) for the cumulative annual predictions.

Specific Objective 2:

1. **Which covariance function is appropriate for capturing the spatial structure of precipitation in the study area for a given time period?**

   The appropriateness of three covariance functions was evaluated using the maximum likelihood method of parameter estimation. These included the spherical function, the exponential function, and the Gaussian function. The preferred covariance function was selected for each period of model fitting by considering the function with the maximum value of the maximized log-likelihood.

2. **What are the optimal model parameters of the chosen covariance function?**

   For each period of prediction, the preferred covariance function was selected and the estimated parameters used for prediction. Details of the functions used for each period and the optimal model parameters are shown in Section 5.4.3.

Specific Objective 3:

1. **Which regression model is most appropriate for the prediction of precipitation in the study area for a given time period?**

   Three different regression models were formulated due to multicollinearity between NDVI and the northing coordinates as indicated in Section 4.4.
   Their performance was evaluated by leave-one-out cross-validation as indicated in Section 5.6.
   The first model performed better in two months (January and June) for the mean monthly predictions and 2006 for the cumulative annual predictions.

   The second model performed better in five months (February, April, May, September, and October) for the mean monthly predictions, in six years (2001-2004, 2007 and 2009) for cumulative annual predictions and in the mean annual predictions.

   The third model performed better in five months (March, July, August, November, and December) for mean monthly predictions and in three years (2005, 2008 and 2010) for cumulative annual predictions.

7.2. Recommendation

The following recommendations are suggested for future research:

- We recommend further research to evaluate the influence of remotely sensed rainfall estimates to the accuracy of mean monthly and annual precipitation estimation.

- An extension of this work with more data to space-time kriging with parameters estimated by the maximum likelihood method

LIST OF REFERENCES

Bajat, B., Pejović, M., Luković, J., Manojlović, P., Ducić, V., & Mustafić, S. (2013). Mapping average annual precipitation in Serbia (1961–1990) by using regression kriging. *Theoretical and Applied Climatology*, *112*(1–2), 1–13. https://doi.org/10.1007/s00704-012-0702-2

Basistha, A., Arya, D. S., & Goel, N. K. (2008). Spatial distribution of rainfall in Indian Himalayas - A case study of Uttarakhand Region. *Water Resources Management*, *22*(10), 1325–1346. https://doi.org/10.1007/s11269-007-9228-2

Bishop, T. F. a. A., Dunbar, M. S., Shatar, T. M., McBratney, A. B., & Odeh, I. O. a. A. (2000). *An overview of pedometric techniques for use in soil survey*. *Geoderma* (Vol. 97). https://doi.org/10.1016/S0016-7061(00)00043-4

Bostan, P. A., Heuvelink, G. B. M., & Akyurek, S. Z. (2012). Comparison of regression and kriging techniques for mapping the average annual precipitation of Turkey. *International Journal of Applied Earth Observation and Geoinformation*, *19*(1), 115–126. https://doi.org/10.1016/j.jag.2012.04.010

Burrough, P. A., McDonnell, R., & Lloyd, C. D. (1998). *Principles of geographical information systems* (Third Edit). New York: Oxford University Press.

Buytaert, W., Celleri, R., Willems, P., Bièvre, B. De, & Wyseure, G. (2006). Spatial and temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes. *Journal of Hydrology*, *329*(3–4), 413–421. https://doi.org/10.1016/j.jhydrol.2006.02.031

Cambardella, C. A., Moorman, T. B., Parkin, T. B., Karlen, D. L., Novak, J. M., Turco, R. F., & Konopka, A. E. (1994). Field-Scale Variability of Soil Properties in Central Iowa Soils. *Soil Science Society of America Journal*, *58*(5), 1501. https://doi.org/10.2136/sssaj1994.03615995005800050033x

Cantet, P. (2017). Mapping the mean monthly precipitation of a small island using kriging with external drifts. *Theoretical and Applied Climatology*, *127*(1–2), 31–44. https://doi.org/10.1007/s00704-015-1610-z

Deutsch, C., & Journel, A. (1998). *GSLib. Geostatistical* …. https://doi.org/10.1016/0098-3004(94)90041-8

Diggle, P., & Ribeiro, P. J. (2007). *Model Based Geostatistics*. *Learning* (Vol. 26). New York. https://doi.org/10.1007/978-0-387-98135-2

Goovaerts, P. (1999). Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma*, *89*(1–2), 1–45. https://doi.org/10.1016/S0016-7061(98)00078-0

Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, *228*(1–2), 113–129. https://doi.org/10.1016/S0022-1694(00)00144-X

Harris, P., Fotheringham, A. S., Crespo, R., & Charlton, M. (2010). The Use of Geographically Weighted Regression for Spatial Prediction: An Evaluation of Models Using Simulated Data Sets. *Mathematical Geosciences*, *42*(6), 657–680. https://doi.org/10.1007/s11004-010-9284-7

Hartkamp, A. D., De Beurs, K., Stein, A., & White, J. W. (1999). Interpolation Techniques for Climate Variables. *NRG-GIS Series 99-01*, 26. https://doi.org/10.1016/j.apsusc.2010.08.104

Hengl, T. (2007). *A Practical Guide to Geostatistical Mapping of Environmental Variables*. Retrieved from http://europa.eu

Hengl, T., Heuvelink, G. B. M., & Rossiter, D. G. (2007). About regression-kriging: From equations to case studies $. *Computers & Geosciences*, *33*, 1301–1315. https://doi.org/10.1016/j.cageo.2007.05.001

Hengl, T., Heuvelink, G. B. M., Tadić, M. P., & Pebesma, E. J. (2012). Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. *Theoretical and Applied Climatology*, *107*(1–2), 265–277. https://doi.org/10.1007/s00704-011-0464-2

Hijmans, R. J., Van Etten, J., Cheng, J., Sumner, M., Mattiuzzi, M., Greenberg, J. A., … Golicher, D. (2018). *raster: Geographic Data Analysis and Modeling*. Retrieved from http://cran.r-project.org/package=raster

Hofierka, J., Parajka, J., Mitasova, H., & Mitas, L. (2002). Multivariate Interpolation of Precipitation Using Regularized Spline with Tension. *Transactions in GIS*, *6*(2), 135–150. https://doi.org/10.1111/1467-9671.00101

Hu, D., Shu, H., Hu, H., & Xu, · Jianhui. (2017). Spatiotemporal regression Kriging to predict precipitation using time-series MODIS data, *20*, 347–357. https://doi.org/10.1007/s10586-016-0708-0

Keblouti, M., Ouerdachi, L., & Boutaghane, H. (2012). Spatial Interpolation of Annual Precipitation in

Annaba-Algeria - Comparison and Evaluation of Methods. *Energy Procedia*, *18*, 468–475. https://doi.org/10.1016/J.EGYPRO.2012.05.058

Lanza, L. G., Ramírez, J. A., & Todini, E. (2001). Introduction Stochastic rainfall interpolation and downscaling. *Hydrology and Earth System Sciences*, *5*(2), 139–143. Retrieved from https://www.hydrol-earth-syst-sci.net/5/139/2001/hess-5-139-2001.pdf

Lark, R. M. (2000). Estimating variograms of soil properties by the method-of-moments and maximum likelihood. *European Journal of Soil Science*, *51*(4), 717–728. https://doi.org/10.1046/j.1365-2389.2000.00345.x

Lorenzo Busetto, Luigi Ranghett, Leah Wasser, J. H. (2018). *Package "MODIStsp" Title A Tool for Automating Download and Preprocessing of MODIS Land Products Data*. Retrieved from https://cran.r-project.org/web/packages/MODIStsp/MODIStsp.pdf

Luo, W., Taylor, M. C., & Parker, S. R. (2008). A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. *INTERNATIONAL JOURNAL OF CLIMATOLOGY Int. J. Climatol*, *28*, 947–959. https://doi.org/10.1002/joc.1583

Ly, S., Charles, C., & Degré, A. (2011). Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrol. Earth Syst. Sci*, *15*, 2259–2274. https://doi.org/10.5194/hess-15-2259-2011

Mair, A., & Fares, A. (2011). Comparison of Rainfall Interpolation Methods in a Mountainous Region of a Tropical Island. *Journal of Hydrologic Engineering*, *16*(4), 371–383. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000330

Mardia, K. V., & Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, *71*(1), 135–146. https://doi.org/10.1093/biomet/71.1.135

Masson, D., & Frei, C. (2014). Spatial analysis of precipitation in a high-mountain region: exploring methods with multi-scale topographic predictors and circulation types. *Hydrology and Earth System Sciences*, *18*(11), 4543–4563. https://doi.org/10.5194/hess-18-4543-2014

Mendez, M., & Calvo-Valverde, L. (2016). ScienceDirect Assessing the performance of several rainfall interpolation methods as evaluated by a conceptual hydrological model. *Procedia Engineering*, *154*, 1050–1057. https://doi.org/10.1016/j.proeng.2016.07.595

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*, 90–100. https://doi.org/10.1016/S0022-2496(02)00028-7

Neter, J., Kutner, M. H., Nachtsheim, C. J., Wasserman, W. (2005). *Applied Linear Statistical Models* (5th Editio). McGraw-Hill.

Odeha, I. O. A., McBratney, A. B., & Chittleborough, D. J. (1994). Spatial prediction of soil properties from landform attributes derived from a digital elevation model. *Geoderma*, *63*(3–4), 197–214. https://doi.org/10.1016/0016-7061(94)90063-9

Oliver, M. A., & Webster, R. (2014). A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena*, *113*, 56–69. https://doi.org/10.1016/j.catena.2013.09.006

Pardo-Igúzquiza, E. (1998). Maximum Likelihood Estimation of Spatial Covariance Parameters. *Mathematical Geology*, *30*(1), 95–108. https://doi.org/10.1023/A:1021765405952

Paulo Ribeiro Jr, A. J., Diggle, P. J., & Paulo Ribeiro Jr, M. J. (2018). *Package "geoR" A package for geostatistical analysis*. Retrieved from http://www.leg.ufpr.br/geoR

R Core Team. (2015). R: A Language and Environment for Statistical Computing,R Foundation for Statistical Computing, Vienna, Austria. Retrieved February 6, 2019, from https://www.r-project.org/

Richard Webster And Margaret A. Oliver. (2007). *Geostatistics for Environmental Scientists* (2nd ed.). Chichester: John Wiley & Sons Ltd.

Todini, E., & Ferraresi, M. (1996). Influence of parameter estimation uncertainty in Kriging. *Journal of Hydrology*, *175*(1–4), 555–566. https://doi.org/10.1016/S0022-1694(96)80024-2

Yoon, S., Kim, M. K., & Park, J. S. (2015). Comparison of statistical linear interpolation models for monthly precipitation in South Korea. *Stochastic Environmental Research and Risk Assessment*, *29*(5), 1371–1382. https://doi.org/10.1007/s00477-015-1031-7