GeoComputational Methods for Surface and Field Data Interpolation

Sara Zandi

A thesis submitted to

Auckland University of Technology

in Partial fulfilment of the requirements for the degree of

Master of Philosophy

School of Computing and Mathematical Sciences

November, 2013

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a University or other institution of higher learning, except where due acknowledgment is made in the acknowledgments.

Signed 

Date 30/04/2014

Acknowledgements

I would like to express my deepest gratitude to my supervisors, Prof. Philip Sallis and Prof. Stephen MacDonell for their excellent guidance, caring, patience and providing me with great atmosphere for doing research.

I would also like to thank my mother Gohar, sister Maryam and brother Mostafa. They were always supporting me and encouraging me with their best wishes.

Finally, I would like to thank my husband, Akbar Ghobakhlou who has always been there for me, overseeing my thesis, giving the best suggestions and standing by me through this.

List of Figures

List of Tables

List of Abbreviations

AR	Actual Rainfall
APIs	Application Programming Interface
APIs	Application Programming Interfaces
AI	Artificial Intelligence
AUT	Auckland University of Technology
A-E	Auto-Estimator
CGIS	Canada Geographic Information System
CS	Catalogue Service
CA	Cellular Automaton
CV	Coefficient Variation
CI	Computational Intelligence
CAD	Computer-Assisted Design
CRUD	Create, Read, Update and Delete
SDI	Data Infrastructure Initiatives
DBMS	Database Management System
DEMs	Digital Elevation Models
ESRI	Environmental System Research Institute
XML	Extensible Markup Language
GIS	Geographical Information System
GOES	Geostationary Operational Environmental Satellites
GPS	Global Positioning Systems
G	Goodness-of-Prediction Estimate
HANTS	Harmonic Analysis of Time Series
H-E	Hydro-Estimator

ITs	Information Technologies
IFOV	Instantaneous-Field-Of-View
IDS	Inverse Distance Squared
IDW	Inverse Distance Weighting
LIST	Line Intersect Sampling Tool
ME	Mean Error
MPB	Mountain Pine Beetle
NIWA	National Institute of Water and Atmospheric Research
NOAA	National Oceanic and Atmospheric Administration
NN	Nearest Neighbour
NWP	Numerical Weather Prediction
OGC	Open Geospatial Consortium
OSS	Open-Source Software
OK	Ordinary Kriging
PPGIS	Public Participatory Geographical Information Systems
RBF	Radial Basis Functions
RDBMS	Relational Database Management System
RMSE	Root Mean Square Error
RMSSE	Root Mean Square Standardized Error
RS-GKDD	Rough Set Geographic Knowledge Discovery in Databases
STAR	Satellite Applications and Research
SR	Satellite Rainfall
SOM	Self-Organizing neural Maps
SFSQL	Simple Features for SQL
SDI	Spatial Data Infrastructures
SDTS	Spatial Data Transfer Standard

STNs	Spatial-Temporal Neighbourhoods
SQL	Structured Query Language
SPOT	Systeme Probatoire dObservation de la Terre
TINs	Triangulated Irregular Networks
USGS	United States Geological Survey
UTALCA	Universidad de Talca
WCS	Web Coverage Service
WFS	Web Feature Service
WMS	Web Map Service
WWW	World Wide Web

Abstract

GeoComputation is a relatively recent and emergent discipline containing numerous methods and techniques for fields such as electrical engineering, computer science, geography, biology. This thesis set out to describe the development of the research domain for GeoComputation and learning area so, to compare and evaluate the methods conventionally used for problem definition and investigation by those working in the discipline. . In recent years, this field has extended beyond geospatial data processing, analysis and depiction to incorporate refined methods of mathematical modelling (particularly for spatio-temporal data point estimation of discrete and continuous event/instance values) and the anticipation (and prediction) of events in Nature. In some cases that are referenced in this thesis, successful solutions to scientific problems have been generated by combining techniques from geographical information systems with those from emerging computer science research areas such as neuro-computing, data mining (heuristic searching for example) and cellular automata. The methods used for data analysis are examined as they are applied for two case studies and conclusions regarding their worth are outlined.

Table of Content

Chapter One

1. Introduction

The academic discipline of Geodetic Science and its professional practice for land surveying and other aspects of Geodesy have evolved over time as technology has advanced. An obvious example of this is the progression from ancient distance measurement techniques using standard-length based rods, through to the use of optical instruments such as the theodolite, to laser resonance equipment and now terrestrial satellite housed remote sensing devices. The use of computational techniques to process geo-spatial data became more possible with the advent of relational databases and software for three-dimensional screen and print visualization, together with the rapid growth in faster and cheaper computing capabilities. Because of the combination of science, engineering and related technologies, geographic and geodetic issues and questions have become the domain of study by researchers from numerous disciplines and a multi (or trans) disciplinary field has emerged, which since the mid-1990's has become known as *GeoComputation*.

GeoComputation is an emerging field of research, which advocates the use of computational-intensive techniques such as neural networks, heuristic search and cellular automata for spatial data analysis. Individual Geo-Computational methods have become accepted as an effective solution to particular spatial analysis problems and they have emerged as imbedded function elements in more commonly used software. Examples of this phenomenon include the extensive use of remote sensing data analysis, geo-simulation techniques and visualization tools in software packages and systems where location data is included as related to the primary application focus. Real Estate information is one example of this and trip planning, hotel accommodation, holiday destination and land use are others.

The nature of GeoComputation is also discussed in this thesis. This new and emerging field is briefly reviewed in terms of its salient characteristics with reference to published research and some common applications are described. In doing this, an explanation of the key techniques of GeoComputation and in particular spatial analysis are summarized and evaluated in terms of their relevance to the research outlined in this thesis.

1.1 **Motivations for the Research**

The first motivation to undertake this work has been has been driven by the expanding interest of spatial prediction in the literature and the apparent limitations that past experience in other research domains brings to the field of GeoComputation. In order to investigate this, different techniques for spatial data interpolation and regression are explained, compared and evaluated in light of the need for new ways to predict location-event data point values over time.

The second motivation relates more particularly to data mining and analysis techniques for geospatial data processing. To date this work as reflected in the literature has mostly been driven by what might be regarded as traditional statistics, ranging from descriptive and inferential techniques for data depiction purposes through to Bayesian and Markovian methods for trend and prediction modelling scenarios. When supplemented by spatial interpolation and other time-place event estimation methods, a growing element of GeoComputation is now being referred to as GeoStatistics. How much this development is part of an evolving GeoComputational research domain reality characteristic is considered in this thesis as a primary impact factor when defining the field itself. It is apparent that the research domain and practice of GeoComputation is for example, providing a powerful and appropriate set of tools for spatial analysis in the agricultural, earth and environmental sciences. Examples of this are given later in the thesis.

1.2 **Thesis Structure**

In this section a brief overview of each chapter of the thesis is given as a guide to what is later presented.

Chapter 2:

In this chapter, the published (principal) definitions of GeoComputation that have been described by individual authors are given. These include some from abstracts of papers and some from GeoComputation conferences. This is considered as a comprehensive set of commonly accepted definitions of GeoComputation. A literature review undertaken for this research is outlined with expository reflections on each article to make obvious the distinctions evident between empirical research results and conceptual models behavioural observation descriptions and sometimes only descriptive representations. The published corpus of literature relating directly to GeoComputation is not large but it does consist of journal articles, papers in conference proceedings and some monographs, mostly textbooks on which, this thesis has drawn.

Chapter 3:

In this chapter the functional elements of a geospatial database are defined, indicating the inter-relationship (and inter-dependence) of source data and the types of data incorporated in this integrative entity. The software used to build and process these databases is also described with examples of some products and their functions. The stages of geographic database design and the key techniques for structuring geographic information are described and evaluated in terms of their potential points of failure and their distinctive operational features.

Chapter 4:

This chapter introduces different types of data models and describes the process of modelling a geographical reality in GIS. Data models are very vital to GIS because of having a major impact on the type of analytical operations that can be preformed and controlling the way that data are stored. Several spatial interpolation approaches involved in modelling surfaces are also discussed in this chapter.

Chapter 5:

Having described the field of GeoComputation and some computational and analytical methods that enact it for practical application in such fields of endeavor as agriculture and the influence of climate and atmosphere on it, this chapter aims to demonstrate some of this science and technology by use of two case studies. In the first case study three common interpolation methods are used to study the spatial distributions of soil pH in a vineyard. Interpolation techniques were used to estimate the pH measurement in un-sampled points and create a *continuous* dataset (time interval-event defined rather than discrete time-data point observed) that can be represented over a geospatial plane…a map of the entire study area observed over time. The accuracy and efficiency of the generated maps are examined and the most fitting technique for the soil pH in the study area is identified. This case study is an example of processing *ground truth* data using GeoComputational methods.

In the second case study, the aim is to demonstrate how GeoComputational methods are used to process *environmental influence* data, rather than only *ground truth* data. In this second case study the distribution of monthly average rainfall as recorded by satellite remote sensing instruments is compared with terrestrial telemetry instruments in specific locations. These two data sources provide a further comparison of wide area rainfall distribution (macro climate) with specific geo-referenced location measurements for micro-climate influence. Samples are observed over 24 hour periods and the source data is compared for variation. When other factors such as temperature and wind velocity as included in a model with this rainfall data, time-condition varying trends can be established and estimates made for input to weather prediction models that can be useful as information for agricultural (especially crop-related) management decision making purposes. There are significant errors in the rainfall estimations obtained from satellite products due to factors such as measurement techniques, retrieval algorithms wind, cloud, temperature, humidity and infrequent satellite overpasses. The chapter concludes with a suggestion for future work as this field evolves and develops such that it can be performance evaluated over time.

Chapter 6:

After describing the case studies in the previous chapter, their results are summarised here and conclusions drawn regarding their usefulness in exposing the field of Geo-Computation and in particular, that aspect of this discipline that relates to research and practice in methods for surface and field data interpolation. The thesis set out describes the field of GeoComputation and to compare and evaluate the principal data processing and analytical elements of it.

Chapter Two

2. **GeoComputation-An Overview**

2.1 **Introduction**

The literature sources and material read, described and evaluated in this chapter is arranged in chronological sequence according to when the publications appeared.

Part of the review undertaken was expository and related to distinctions made between empirical research results and conceptual, behavioral and sometimes only descriptive published articles. The corpus of literature relating directly to GeoComputation is not large but it does consist of journal articles, papers in conference proceedings and some monographs, mostly textbooks. The articles referred to here provide a comprehensive but only a representative sample of publications in this area of research and practical application.

There are debates among researchers over the inception of GeoComputation as an emerging field in the computational science community. Some writers believe that GeoComputation has been around as long as the computer was used in geography, whereas others seem to think that GeoComputation is a newer and specifically, a distinctive concept.

Stan Openshaw is generally recognized as the 'father' of GeoComputation. As he stated in his book entitled *GeoComputation* (Openshaw & Abrahat, 2000), the word GeoComputation first "appeared" in his spell checker dictionary in 1996, after a coffee time discussion relating to a computational geography conference in Leeds. The word 'computational geography' used at the time was too restricted and did not include other related disciplines within the context of "geo" studies. Soon after, Tavi Murray, an ice geographer, changed the words 'computational geography' into a more general and meaningful way and coined the term GeoComputation with a capital C preceding the word Computation. The subsequent use of an uppercase C in GeoComputation is however, generally attributed to Bob Abrahart. Again in his book (Openshaw & Abrahat, 2000), Openshaw says that the C in capital format emphasizes the importance of Computation component whereas in the Preface to GeoDynamic (Atkinson P. M., 2004), Atkinson notes that "the uppercase C in GeoComputation was retained to emphasize that each part (geography, computation) was equally important." As is so often the case it seems, apparently minor details such as this can have deep meaning.

2.2 What is GeoComputation?

GeoComputation has been described and defined many times in the past. It remains only one term used for various research activities and rather than being an over-arching discipline descriptor, it is often subsumed with other names for entities such as The Geoinformatics Research Centre at the Auckland University of Technology (AUT, 2013) or the Centre for Geomatics at the University of Talca in Chile (UTALCA, 2013) . Early references to each of these terms (Geoinformatics and Geomatics) such as in a Software Engineering Journal of the time by Sallis and Benwell (1993) can be read now as attempts at defining an emerging discipline that was more-or-less GeoComputation; a term first coined in 1996. For the purposes of this thesis after considering the contemporary literature on the topic, GeoComputation is regarded as the extant umbrella term.

In the Preface to GeoComputation (Openshaw & Abrahat, 2000), Stan Openshaw and Bob Abrahart talk about GeoComputation as a follow-on revolution occurring after Geographic Information System (GIS). They list three challenges arising from the basic structure of GeoComputation:

1) Better and fuller use of GIS and non-GIS generated data

2) Utilizing the data beyond the data gathering

3) Further expanding the concept of Geoinformatic sciences toward real world problem solving.

According to Openshaw, there are three inter-related components for GeoComputation. First is geographical or environmental data. Second, the modern computational technologies and the third component is a high performance computing hardware. In his paper (Openshaw, 2000), two aspects that make GeoComputation special were identified. The first is emphasis on "geo" subjects, which has been de-emphasized until recent years. Second is the distinctive power of the computation, which provides scientists with new or better solutions for insoluble problems.

The subsequent use of GeoComputation research "…arose from the application of computer power to solve geographical problems." According to Ress and Turton (1998) GeoComputation is about solving geographical problems using the computational methods. They see GeoComputation as "the progress of applying computing technology to geographical problems" which is different from doing geography with the computer. Helen Couclelis (1998) also believes "Novelty" is not an accurate explanation for GeoComputation as both notion of GeoComputation and its techniques were mentioned by Dobson's (1983) in 'Automated Geography" article. Her working definition for GeoComputation is the use of computational methods and techniques "to portray spatial properties", to solve geographical problems and to describe the concept of geography. She agrees that "the universe of computational techniques applicable to spatial problems" is a true characterization of the current state of art. Couclelis believes GeoComputation does not have much contribution in science society and its lack of epistemological definition makes it more like a "grab-bags" of tools powered by Artificial Intelligence (AI) and Computational Intelligence (CI) fields (p18,27). She notes that GeoComputation lacks the theoretical concept and it needs to"... formulate a model based on the computational notion of machine "and "… develop a coherent perspective on geographical space" that justify the 'geo' prefix." (p. 25, 28)

Unlike Helen Couclelis' pessimistic attitude, Mark Gahegan (1999) has a promising practical view toward GeoComputation. He sees GeoComputation as an enabling technology, one needed to fill the "…gap in knowledge between the abstract function- ing of these tools…and their successful deployment to the complex applications and data sets that are commonplace in geography." (p. 206). He defines GeoComputation concern as "…to enrich geography with a toolbox of methods to model and analyze a range of highly complex, often non-deterministic problems." (p. 204) Bill Macmillan (1998) in the Epilogue of GeoComputation, A Primer, does not agree with Couclelis either. He argues that GeoComputation includes the latest forms of computational geography and that it is not an incremental development (p.257). He says Stan Open- shaw's latest offer the "Geographical Correlates Explanation Machines", answers Couclelis criticism of lack of a major demonstration projects (p .261). Macmillan accepts that sound theory is needed, but he believes that it has already been provided by Openshaw in form of inductivisim.

Longley (2001) believes that the GeoComputation community needs to more emphases on association between GeoComputation and GIS. He describes GeoComputation as a computational approach to human and geography problem and "In some important respects, the term GeoComputation is synonymous with geographic information science….although it has often put greater emphasis upon the use of high-performance computers.” (p. 140)

In 1999 Steve Harding studied the word frequency and association analysis of the conference abstracts and prepared a poster contacting these words and phrases. The larger text symbolized a broader concept and smaller text represented a particular concept. Each of the four basic concepts: space, time, data, and computing were represented by different font styles and colors (Figure.2.1) (Harding S. , 1999).

The poster was a motivating tool to visualize the nature of GeoComputation; however it was subjective to its authors' observation and experiences (Ehlen & Douglas, 2002). GeoComputation is about empowering geography with a range of computational methods in order to analysis, model and solve currently unsolvable or even unknown problems. It is a rich technology representing both geography and computer sciences. It challenges computer scientists to provide geographers and researches with enabling computational techniques, algorithms, and paradigms.

2.3 **List of citations in the field of GeoComputation**

2.3.1 1996-2000

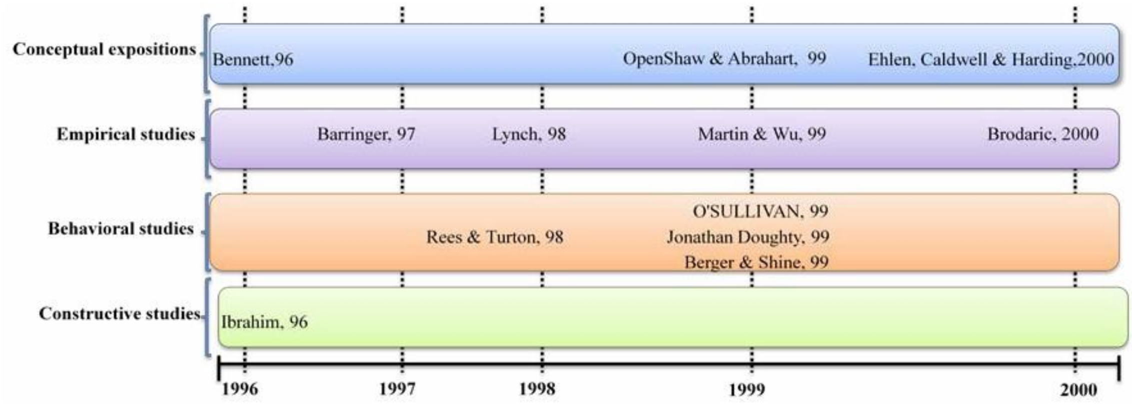


Figure 2.2: Years 1996-2000

2.3.1.1 Conceptual expositions

The principal milestone and publications that have defined the GeoComputation descriptive are listed.

2.3.1.1.1 Spatial Reasoning for Geographic Information Systems (Bennett, 1996)

This paper argues that Geographic Information Systems (GIS) do not support spatio-temporal analysis with the kinds of data objects that they model. It proposes an alternative method to provide GIS with a multi-dimensional reasoning capability. It also explains a mechanism whereby complex compound GIS queries can be expressed and resolved. It is well argued and the method described supports the author's proposition. It is a workable solution to the functional limitations of GIS software.

2.3.1.1.2 GeoComputation research agendas and futures (Openshaw & Abrahart, 1999)

The aim of this article is to devise a GeoComputation research agenda and identifies its desirable key topics. It also encourages researches to write down their ideas about the potentials of GeoComputation, possible research agendas, interesting speculation in order to increase and improve inter-researcher dialogue in this emerging field. Five thematic topics of the GeoComputation named in this paper are: Theoretical developments, Empirical analysis, Modeling and simulation, Social aspects and Automation of analysis and modeling functions. It is a good contribution to the research activity in this research domain.

2.3.1.1.3 The Semantics of GeoComputation (Ehlen, Caldwell, & Harding, 2000)

This paper is a bottom-up approach looking at GeoComputation in terms of what GeoComputation researchers say they do in their data analysis, software development and implementation. Phrase analysis software investigates the scope of GeoComputation by examining the body of research and abstracts presented at the five conferences between 1996 and 2000. This is a very useful review resource for research in this field.

2.3.1.2 Empirical studies

Three publications reflecting empirical developments in GeoComputation are reviewed in this study.

2.3.1.2.1 Meso-scale mapping of soil temperatures in the Mackenzie Basin, New Zealand. (Barringer, 1997)

In this paper, Barringer notes that data collected at climate sites in New Zealand do not represent much of the surrounding area and they contain significant errors, which makes it difficult to interpolate local climate to other adjacent areas. This paper explains the development and testing of an empirical model for soil temperature prediction. The model provides the patterns of spatial variation in soil temperature.

2.3.1.2.2 Converting Point Estimates of Daily Rainfall onto a Rectangular Grid (Lynch, 1998)

This paper describes the Arc/Info GIS techniques that assist researchers in converting point estimates of daily rainfall to a raster plane which can be used to describe the daily rainfall in a real environment. At discussion section, the author recommends that either the Schäfer daily rainfall estimation method or the inverse distance weighting method can be used to convert point daily rainfall measurements at un-gauged positions.

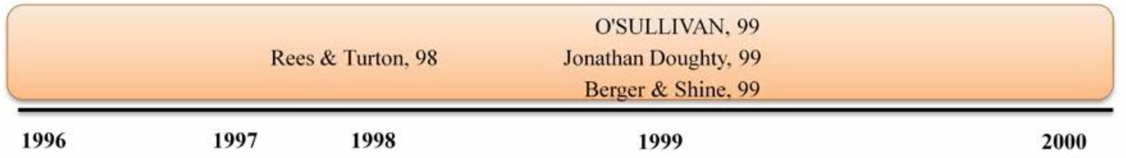
2.3.1.2.3 Empirical CA simulation from a high resolution population surface (Martin & Wu, 1999)

In this paper Cellular Automaton (CA) simulation is applied to simulate urban growth using empirical data of south-east region of the UK. It examines the empirical and theoretical concerns when implementing CA simulations. A population surface modeling technique is used to create a land use model. It implements two simulation models using different growth constraint, based on a range of empirical data sources. The paper also includes some urban growth scenarios and explains that it is essential to link the application of simulation to GIS order to interpretation of the CA results in a real world context.

2.3.1.2.4 GeoComputing with Geological Field Data: Is there a 'ghost in the machine'? (Brodaric, 2000)

This study explores the degree of correlation between field data and their generalized geological classes. Data and analysis from some geologists are compared and correlated using unsupervised and supervised classification techniques with the Self-Organizing neural Maps (SOM). This paper also reports correlation results, challenges in organizing largely qualitative data for the SOM and inductive and deductive reasoning processes.

2.3.1.3 Behavioral studies



2.3.1.3.1 Investigation of the Effects of Input Uncertainty on Population Forecasting (Rees & Turton, 1998)

The paper investigates the effects of improbability in the model inputs on projection outcomes using a model for 12 European Community countries and 71 regions. The projection is based on historical data and professional's opinion. Scenarios are driven for fertility, mortality, extra-European Community migration, inter-country and inter-region migration.

2.3.1.3.2 Exploring the structure of space: towards geo-computational theory (O'SULLIVAN, 1999)

This paper describes ongoing research that seeks to answer Couclelis' (1998a, 1998b) call for a specifically GeoComputational theory in her Challenges for GeoComputation, through a phenomenological investigation of a new class of models of spatial processes.

It proposes a method by which the properties of such models might be usefully explored, and concludes with some initial results from such an exploration. Model structure is characterized by using some of the graph structural measures that have been proposed. A brief survey of some graph measures is presented as well.

2.3.1.3.3 Interoperable geospatial objects (Jonathan Doughty, 1999)

This paper describes a research prototype that creates an Extensible Markup Language (XML) representation of geospatial metadata and interoperability characteristics for transmission between systems and manipulates them via Java-based components for validation.

2.3.1.3.4 Estimating sub-pixel geospatial features (Berger & Shine, 1999)

The method represented in this paper is used to estimate the sub-pixel geospatial features in special situations. It also explains solutions to estimate the mini-images that might have been seen within the Instantaneous-Field-Of-View (IFOV) which leads to tagging of clusters of nondescript pixels. The study is also intended to reduce the ambiguity in the correspondence problem in standard photogrammetry.

2.3.1.4 Constructive studies

2.3.1.4.1 GPS Based Public Transport Route Information System (Ibrahim, 1996)

This paper describes a public transport live passenger route information system based on a Global Positioning Systems (GPS).The system provides on-board passengers with real-time location information. The aim of this study is to provide information to passengers and increase the use of public transport which will help to overcome the rising traffic congestion problems in big cities.

The system receives the vehicle's spatial information from a GPS receiver. GPS data is transferred to a computer for processing which is pre-loaded with the co-ordinates of the bus-stops on the route of a bus. A program calculates the distance between the bus-stops and the current location of the bus in real-time and then provides the bus routes in the area of interest as well as estimated arrival time.

2.3.2 2001-2005

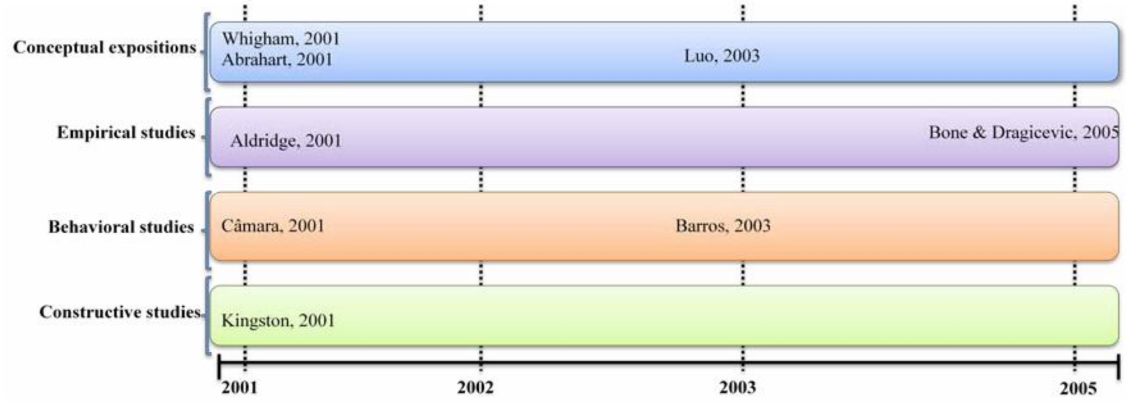
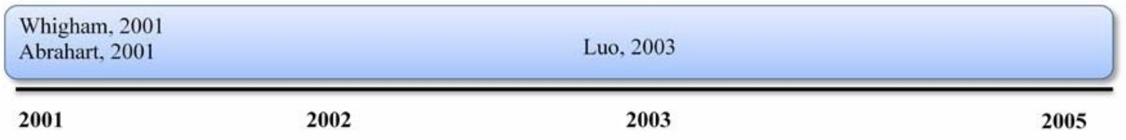


Figure 2.3: Years 2001-2005

2.3.2.1 Conceptual expositions



2.3.2.1.1 Spatial Information: Problems, Challenges and Directions (Whigham, 2001)

The paper highlights challenges and the diversity of fields such as spatio-temporal data, data management, data models and representation, intelligent spatial systems, mobile delivery and the World Wide Web (WWW) in spatial information science and research areas.

2.3.2.1.2 Why are computational processes important? (Abrahart, 2001)

This paper argues that developments in computer sciences must be a part of the geo-graphical research agenda and there is a need for investigation on the principal influences of computational processes on digital geographical researches.

2.3.2.1.3 GeoComputation: A Coordination-Oriented Approach (Luo, 2003)

This paper attempts to set up a methodological foundation for knowledge sharing of geospatial problem solving, integration of tools and simplifying the uses of GeoComputational techniques for domain users.

2.3.2.2 Empirical studies

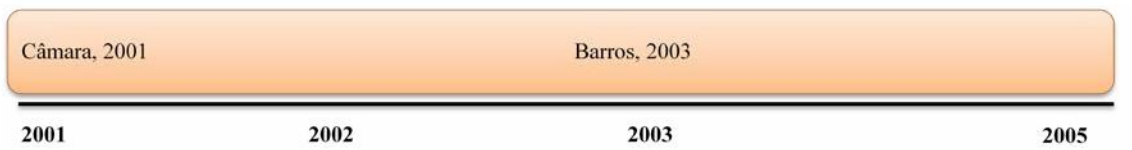
2.3.2.2.1 A Rough Set Based Methodology for Geographic Knowledge Discovery (Aldridge, 2001)

This paper explains the Rough Set Geographic Knowledge Discovery in Databases (RS-GKDD) methodology. The proposed methodology is applied to ecological data relating the spatial distribution of greater glider possums to the quality of their environment. The result of analysis by other researchers using several knowledge induction methods are used to compare the results of the present work.

2.3.2.2.2 Sensitivity of a Fuzzy-Constrained Cellular Automata Model of Forest Insect Infestation (Bone & Dragicevic, 2005)

This paper focused on a case study of mortality patterns on native pine (Pinus contorta) caused by infestations of Mountain Pine Beetle (MPB). This study intended to integrate fuzzy set theory with GIS based CA modelling to model tree mortality patterns and examine the sensitivity of the model to other spatial properties such as the size of the cells in the raster grid, the number of cell states, the types of transition rules or the number of iterations for which the model is performed.

2.3.2.3 Behavioral studies



2.3.2.3.1 GeoComputation techniques for spatial analysis: are they relevant to health data? (Câmara, 2001)

This paper is a brief survey of GeoComputational techniques and examined some of the main branches of research in GeoComputation. The paper attempts to provide a unified perspective of this new research field and draws the attention of the public health community to the new analytical possibilities offered by GeoComputational techniques.

2.3.2.3.2 Simulating Urban Dynamics in Latin American Cities (Barros, 2003)

The present paper focuses on a specific kind of urban growth that happens in Third World cities, called 'peripherisation', which is characterized by the construction of low-income residential areas in the marginal ring of the city and a perpetuation of a dynamic core-periphery spatial pattern. This study presents two different simulation exercises that explore different kind of urban growth in Latin American cities.

2.3.2.4 Constructive studies

2.3.2.4.1 MedAction: An Internet Tool for Forecasting Land Use Change and Land Degradation in the Mediterranean Region (Kingston, 2001)

MedAction is a web-based mapping tool that provides planners, decision-makers and citizens with a modeling system which predicts land use change and land degradation risk. The proposed application is a web-based modeling system which is both an information service and a scenario based analysis tool. It alerts climate change scenarios and visualizes the effects on land use change and land degradation. The model also integrates predictions of the climatic, the physical and the socio-economic environment to create scenario-based forecasts of agricultural land-use and land degradation at a one decimal minute resolution.

2.3.3 2006-2010

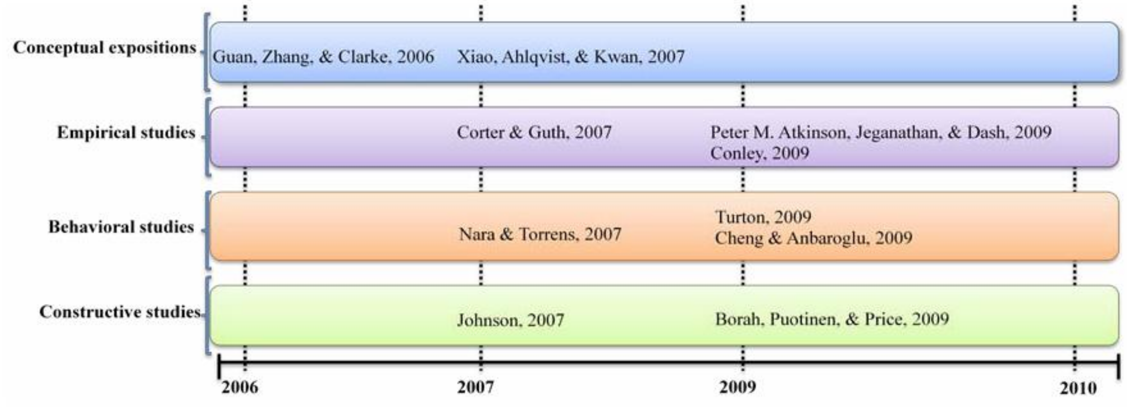
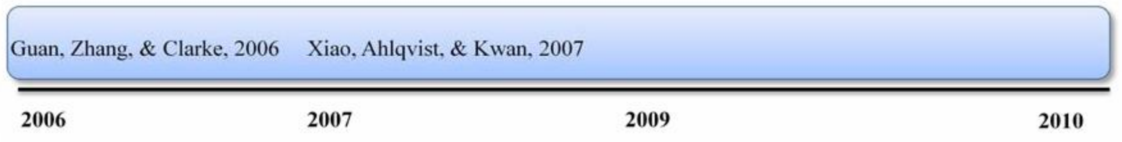


Figure 2.4: Years 2006-2010

2.3.3.1 Conceptual expositions



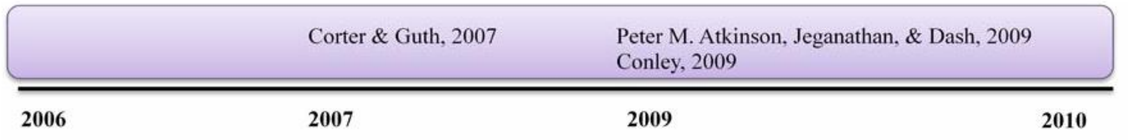
2.3.3.1.1 GeoComputation in the Grid Computing Age (Guan, Zhang, & Clarke, 2006)

This paper discusses challenges such as usability, feasibility, applicability that current GeoComputation faces. A Grid-based geospatial problem-solving architecture proposed to provide an easy-to-use geospatial problem-solving environment that integrates multiple complicated GeoComputational processes at an acceptable cost. The model composed of four tires: presentation tier, service tier, model tier and grid tier.

2.3.3.1.2 Public Participation GIS for the General Public? (Xiao, Ahlqvist, & Kwan, 2007)

The complex features of available GIS packages require technical tuition and financial resources which causes the lack of public in Public Participatory Geographical Information Systems (PPGIS). The paper discusses the conceptual framework of PPGIS that can be made available to the general public. The three primary requirements for the proposed model are interaction, sharing and computation.

2.3.3.2 Empirical studies



2.3.3.2.1 Unsupervised Classification of Submarine Landslides (Corter & Guth, 2007)

The paper investigates the differences between landforms in the two lakes (Crater Lake in Oregon and Lake Tahoe located in California) and the surrounding environment. It also examines the use of a collection of geomorphometric environmental parameters for unsupervised terrain classification.

The classification methods demonstrate marginal success in entirely separating the landslide from other terrain. Greater than 90% of the landslide points were sited together in the same cluster in both lakes. In both cases, the method also placed several points not in the landslide in the same cluster as the landslide points.

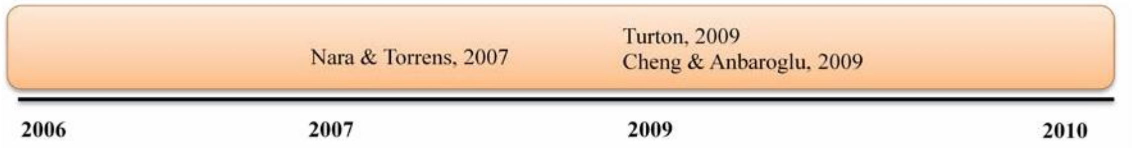
2.3.3.2.2 Analyzing the effect of different GeoComputational techniques on estimating phenology in India (Peter M. Atkinson, Jeganathan, & Dash, 2009)

The study compares and analysis the capability of some of frequently used methods in smoothing time-series satellite sensor data over multiple landscapes in India. The five smoothing techniques are Harmonic Analysis of Time Series (HANTS), asymmetric Gaussian, double logistic, Savitzky-Golay and discrete Fourier based approach. It was found that parameters from HANTS and TIMESAT do not provide an adequate fit to large landscapes. The current study recommends the Fourier based approach as it was fitting for all natural vegetation and agricultural types in India.

2.3.3.2.3 An Evaluation of the Extended Power of Cluster Detection Methods under Sampling Strategies (Conley, 2009)

This paper aims to detect a cluster's extent or shape under different sampling strategies by applying spatial cluster detection methods. Detecting cluster's shape using the traditional power statistic is not accurate as it may not match the shape of the original cluster. This paper studies this inequality by applying power statistic to datasets generated with different sampling methods. As a result, the complete decrease in power for SaTScan is unexpectedly large and FleXScan maintained its power the best of all the methods. The ability of cluster detection methods to accurately detect the shape of a cluster decreases more than the standard power under sampling conditions.

2.3.3.3 Behavioral studies



2.3.3.3.1 Fractal Analysis of Pedestrian Egress Behavior and Efficiency (Nara & Torrens, 2007)

The objective of this research is to provide a spatial analysis of pedestrian egress behaviours and efficiency in different crowd environments and explores the capability of using fractal analysis for quantifying tortuosity of movement paths.

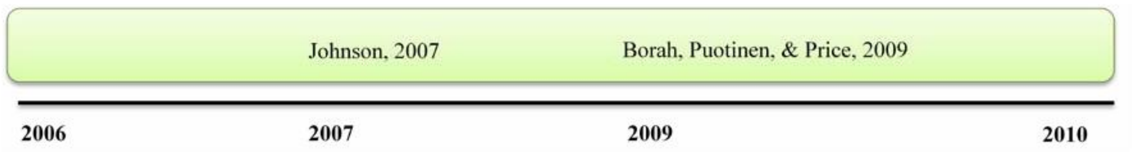
2.3.3.3.2 Experiences Teaching GIS with Open Source Software (Turton, 2009)

This paper provides a data point for geographic teaching by recounting the experiences of the author in teaching a masters level courses in the Pennsylvania State University's e-education program. Student responses even though the course is considered as one of the hardest in the programme, they all have learned something beyond button pushing and the standard software packages.

2.3.3.3.3 Methods on Defining Spatial-Temporal Neighborhoods (Cheng & Anbaroglu, 2009)

Since there is no formal definition on integrated Spatial-Temporal Neighbourhoods (STNs), this paper intends to provide a quantitative approach to reflect the relationship between time and space in the data precisely and reviews available methods on defining STNs.

2.3.3.4 Constructive studies



2.3.3.4.1 Health Atlas Ireland (Johnson, 2007)

This paper introduces Health Atlas Ireland, Open-Source Software (OSS) which integrates GIS, database and statistical components. The intention of the system is to help answer questions related to health events, emergency response, health services and demographics, at first in the Republic of Ireland and eventually worldwide as related to Irish Health Services.

2.3.3.4.2 Line Intersect Sampling Tool (LIST): A GIS-based tool for Spread Analysis (Borah, Puotinen, & Price, 2009)

Current commercial and open source GIS software provide options for random point sampling of spatial data. However, random sampling of a distributed trend requires a line-sampling or area-sampling technique. Line intercept sampling has not been implemented for spread analysis and no GIS tools allow random directional sampling of disturbance progressions maps. In this paper a new GIS-based tool called the Line Intersect Sampling Tool (LIST) is introduced to address this issue. LIST is an ArcGIS tool developed in Python for creating random line sample transects for progression datasets.

2.4 **Research Methods**

2.4.1.1 Meso-scale mapping of soil temperatures in Mackenzie Basin, New Zealand. (Barringer, 1997)

This paper proposes an empirical model to predict the spatial variability of soil temperature within the central South Island high country of New Zealand.

Two types of data (site attribute data and dependent variable) are collected from 43 sites in the Grampians. The sites elevation ranged from 600 to 1800 meters and the four primary aspects (i.e., north, east, south, and west) were considered. Aspect is recorded into degrees south of north. Soil temperature data are collected every four month using a hand held digital thermometer sealed by a protrude aluminum.

Multiple linear regression approach is used to model how the value of the soil temperature varies when any one of the site attribute data (slope, altitude and aspect) changes. The calculated correlation coefficients from the seasonal regression analyses are high (0.83 to 0.96). The small regression interaction terms between elevation and location (0.001) and aspect and location (-0.003 to 0.001) shows that the regression is applicable throughout the area in excess of 15000 km^2.

2.4.1.2 Sensitivity of a Fuzzy-Constrained Cellular Automata Model of Forest Insect Infestation (Bone & Dragicevic, 2005)

The aim of this study is to simulate tree mortality patterns caused by integrating fuzzy set theory with GIS-based CA modelling. It also examines the model's responsiveness to different spatial properties of a fuzzy-constrained CA. The study addresses the issue of applying fuzzy sets to modelling dynamic ecological processes. It uses lodge pole pine mortality patterns caused by swarm of MPBs are as a case study. The results from the analysis of the Tree Mortality Model indicate that the model is significantly sensi- tive to the choice of the size of both the neighbourhood and study area.

It shows that fewer trees were attacked in smaller neighbourhoods during the first two cycles.

The sensitivity analysis also shows the smaller study area received considerably less attacked trees for the central area than the model using the entire study area.

In addition, the CA model that uses larger neighbourhood and study area sizes also ensure that MPB attacked less susceptible trees once plenty amount of high susceptible trees were attacked and MPB population increased.

2.5 Unsupervised Classification of Submarine Landslides (Corter & Guth, 2007)

This paper aims to explore differences between landforms in the two lakes and the surrounding terrain, and to study the use of a suite of geo-morphometric terrain parameters for unsupervised terrain classification. The two m and 1/3 arc second Digital Elevation Models (DEMs) for Crater Lake was compared to test the effect of re-interpolating the DEM on environment statistics.

The preformed classification method showed marginal success in completely separating the landslide from other terrain. More than 90% of the landslide points were placed together in the same cluster for both lacks. In both cases, the method also located lots of points not in the landslide in the same cluster as the landslide points. Unsupervised classification is reliable for categorizing terrain and finding the steep slopes, gentle lake bottoms, and the transitional areas between them. Since the landslide shares many likenesses with other flat areas, unsupervised classification is not accurate in separating the landslides.

2.6 **Conclusions**

GeoComputation is a fresh field of endeavour yet reflects a broad spectrum of research methods, which are applied appropriately to the sub fields of this problem domain. GeoComputation is a cutting age research within the field of GIS and geospatial analysis and it is strongly influenced by latest programming development, data processing and interface design. Individual GeoComputational methods become accepted as an effective solution to particular spatial analysis problem and they start to emerge in more common software with extensive usage such as remote sensing data analysis, geo-simulation techniques and visualization tools.

GeoStatistics is perhaps considered a separate field in its own right but the investigation conducted for this thesis suggests that its methods (a mix of standard statistics and some contemporary data clustering and visualisation methods) are directly related to the problems addressed within the field of GeoComputation and therefore, it is rightly an intrinsic part of this research domain, especially where empirical methods are being used. The association of GeoStatistics, GIS, statistics and GeoComputation provides a powerful and compatible set of tools for spatial analysis in the agricultural, earth and environmental sciences.

System usability and some other aspects of GeoComputation where conceptual design and human opinion are quantified may best be described using behavioural research methods. In short, GeoComputation is essentially similar to other areas of research activity in Computer Science and Information Systems but due to the discipline boundaries it crosses, unique methodologies arise for individual problem solving and can at times cause concern for researchers because they do not strictly conform as one approach to research that may in undertaken in any of the contributing discipline areas. Notwithstanding, a corpus of literature and a body of knowledge exists as is evolving for GeoComputation and therefore, it can be considered as serious science in its own right.

Chapter Three

3. **Geodetic Data and Computer Sciences**

3.1 **Geographical Information system (GIS)**

A Geographical Information System (GIS) is defined as a set of tools for spatial data collection, storing, retrieval, transformation and visualisation from the real world for a particular set of reasons (Burrough, 1986). In their design and implementation they are intrinsically interactive. The first GIS were developed by Canada Geographic Information System (CGIS) (Peuquet 1977 in Star and Estes 1990) to deal with environmental impact issues and practical problems. The aim of CGIS was to analyse Canadian land inventory data, to find marginal lands. The GIS was first implemented in 1964 (Deuker 1979 in Star and Estes 1990). It was first commercialised in 1970 during operation of image processing and remote sensing. Several institutes and organisations such as the Environmental System Research Institute (ESRI) in California were interested in using GIS and began investing in their development. ESRI used special GIS software such as Arc View, Arc Map, Map Info, IDRISI etc. (Star and Estes 1990). Geographic Information System could be understood in two parts "Geography and Information system". First "Geography" is study of spatial relationship between man and environment using map as a key tool. Secondly "Information System" is a continuous sequence of data collection, storage, analysis, and use the derived information in decision-making (Calkins and Tomlinson 1977 in Star and Estes 1990).

GIS interprets and visualizes data in many ways and reveals relationships, patterns, and trends in the form of maps, globes, reports, and charts. A GIS is capable of using different set of operation to work with spatially referenced geo-data. Manual data elements such as maps, aerial and ground photograph, statistical report are used in GIS. GIS can be applied in almost every field for measuring, mapping, monitoring and modelling environmental features GIS is continuous process of data actuation, pre-processing, data management, manipulation and analysis and product generation (Star and Estes 1990; Congalton and Green 1995; Sivertun 1993). GIS processes relevant information in a standard form and generates computerized maps and further used for integrated analysis. It manipulates each datum in its entire spatial and temporal context and provides the decision makers with analysis to work with. Table 3.1 summarises definitions of GIS from its audience's point of view.

Table 3.1: Definitions of GIS, and groups who find them useful (Longley, Goodchild, Maguire, & Rhind, 2011)

GIS definition	Groups who find it helpful
A container of maps in digital form	The general public
A computerised tool for solving geographic problems	Decision makers, community groups, planners
A spatial decision support system	Management scientists, operational researchers
A mechanized inventory of geographically distributed features and facilities	Utility managers, transportation officials, resource managers
A tool for revealing what is otherwise in-visible in geographic information	Scientists, investigators
A tool for performing operations on geo-graphic data that are too tedious or expensive or inaccurate in performed by hand	Resource managers, planners

There is a growing awareness of the economic and strategic worth of GIS in almost every industry. GIS implementations can optimize operational expenses through reduction in fuel consumption and staff time, better customer service, and more efficient scheduling. Making proper a decision is critical to the success of an organization. GIS is the go-to technology for making better decisions about location. For example, during typhoons, disaster response commanders need to know all the different situations occurring at each point in time and find disaster information from older periods and monitor current data in real time (Figure 3.1).

Figure 3.1: This GIS-based disaster decision support system helps Taiwan plan for and respond to typhoons (www.esri.com, 2013).

GIS-based maps and visualizations provide advanced communication between public and professional fields and departments. GIS provides a powerful support for managing and maintaining reliable records about the geographical status and changes (Esri, 2009). Figure 3.2 shows a GIS-based state-wide cadastral database used to manage all the geographic data associated with the huge territory of Montana state in the United States.

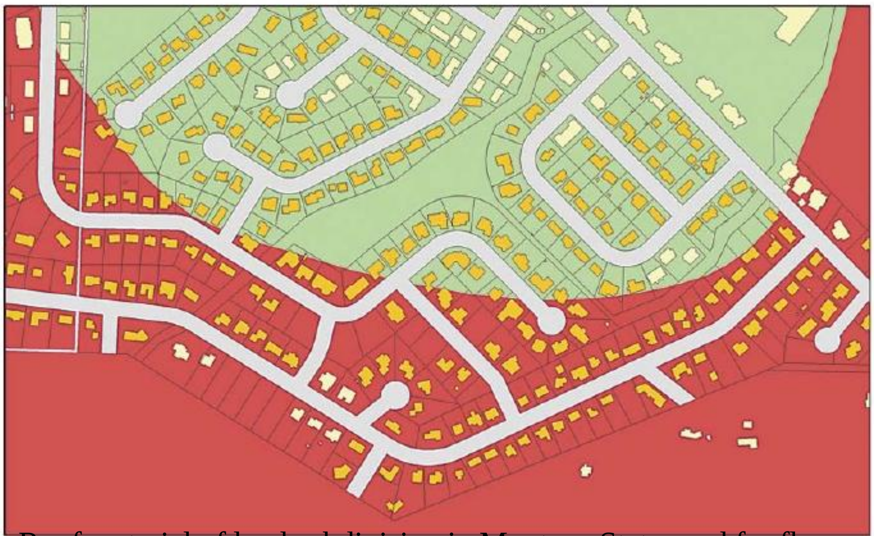


Figure 3.2: Roof material of land subdivision in Montana State used for flammability classifications (www.esri.com, 2013).

The disadvantage of GIS method is that the procedure of data collection and the software system are expensive. In addition, using GIS requires skills in a variety of geographic, computer science and system engineering field.

3.2 **Data Collection**

One of the key characteristics of GIS is their ability to contain diverse geographic data types and integrate data about location from different sources. In this section, data collection methods are categorized into data capture and data transfer classes.

3.2.1 Data Capture

Data capture can be data input from either primary data sources (data collected in digital format specifically for GIS use) or secondary sources (digital and analogue datasets that require to be fitted into digital format in a GIS project) (Longley et al, 2011) (Table 3.2). Characteristics of raster and vector data models are described in chapter four.

Primary GIS sources examples include the French Systeme Probatoire dObservation de la Terre (SPOT) and Quikbird Earth satellite images and vector building survey measurements captured using a survey station. Digital satellite remote-sensing images are one of the primary data capture sources for GIS use. Remote sensing is the science of obtaining information about an object or area through the analysis of measurements made at a distance from the object (Eastman, 2001).

Remote sensing includes the set of sensors, platforms. Data processing methods are used to receive information about the earth surface features and atmosphere properties through sensors. The field of remote sensing has grown from the camera and film in aerial photography to electro-optical sensors in optical or microwave regions of electromagnetic spectrum bands (Sahoo, 2012).

Secondary GIS sources include aerial photographs or paper maps that can be scanned and vectorised. It is not always easy to determine the difference between primary and secondary, and raster and vector sources. For example, digital satellite remote sensing data may be thought as secondary as they are not stored straight into a GIS database (Longley et al, 2011). However, according to (Longley, Goodchild, Maguire, & Rhind,

2011) these data are considered as primary because they have gone through minimum transformation and the data characteristics are appropriate for direct use in GIS projects.

Table 3.2: Geographic Data Collection Techniques

	Raster	**Vector**
Primary	Digital remote sensing images Digital aerial photo graphs	G P S measurements Survey measurements
Secondary	Scanned maps/photographs/ DEMs frommaps	Topographic surveys Toponymy database

Data collection has been always a major project task, and is considered as a time consuming, costly and tedious task. The costs of data capture account for up to eighty five percent of the cost of a GIS. Data collection involves a series of sequential stages. Developing a project plan and establishing the user's needs is an important stage to any project. Preparation in data collection is key stage. It includes data collection, redrafting low quality map resources, editing and removing noises from images and preparing GIS hardware and software systems that are compatible with data. Next stage is to conversion and transformation of data to digital format and involves the majority of the effort. Data validation and improvement are performed in editing and improvement stage. The final phase is to identify the project's success and failures in both qualitative and quantitative ways. Figure 3.3 illustrates five stages in data collection projects (Longley, Goodchild, Maguire, & Rhind, 2011).

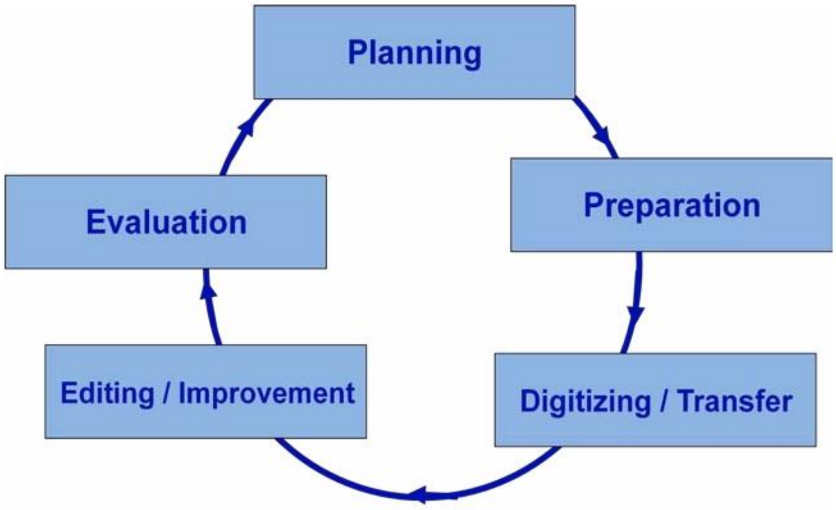


Figure 3.3: Stages in data collection projects

3.2.2 Data Transfer

Data transfer is a method of importing data into GIS from external sources. The availability of datasets is constantly changing and many types of the geographic data are sold mostly and increasingly from web sites. According to (Longley, Goodchild, Maguire, & Rhind, 2011) the best way to find geographic data is to search the Internet and use resources such as geo-libraries and geo-portals provided by national or global Spatial Data Infrastructure Initiatives (SDI). However, there is a major challenge to evaluate the fitness of data when using data obtained from the Web. These data need to be projected on top of a map with an adequate accuracy and independent verification of the geometric properties of a sample of objects. Another challenge with data obtained from external resources is that they are usually encoded in many different formats. This variety of data formats are developed in response to diverse user requirements.

To decrease the high cost of geo-databases development, many software and open Application Programming Interfaces (APIs) have been developed to transfer and reused data. Many image formats, AutoCAD DWG and ESRI Shape files can be read directly by GIS. There are complex formats, such as Spatial Data Transfer Standard (SDTS) are designed for exchange reasons and involve advance processing before they can be displayed in GIS application. Third party translation system software can be used to convert data to a format that is compatible with the target GIS software (Longley, Goodchild, Maguire, & Rhind, 2011). Figure 3.4 shows transformation of data by direct read and translation intermediate.

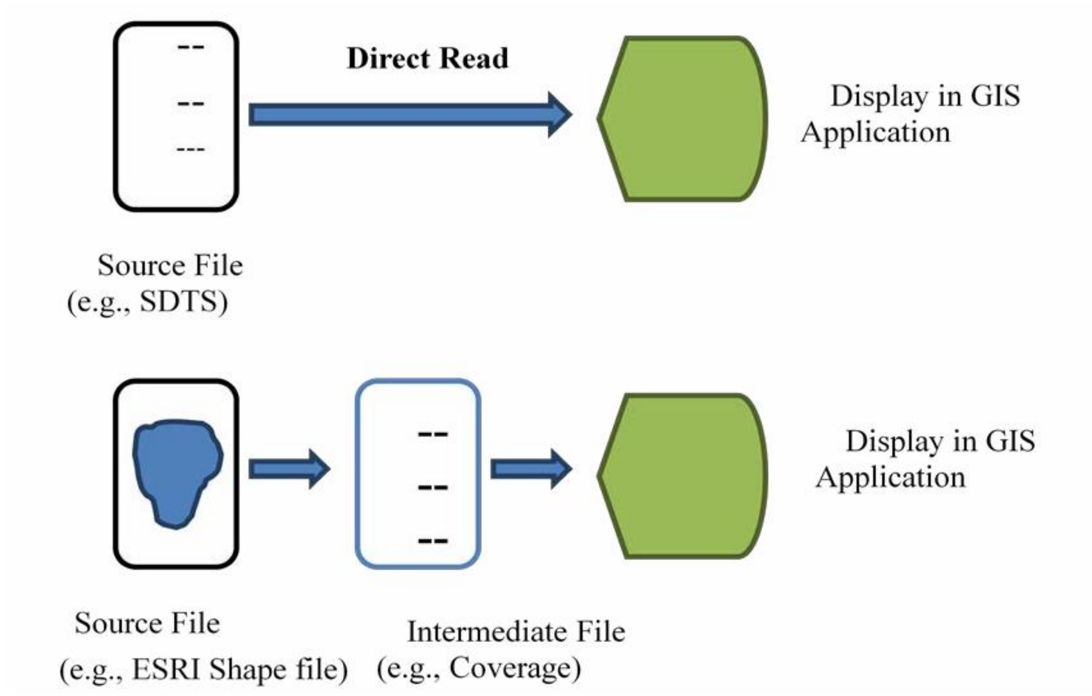


Figure 3.4: Comparison of data access by direct read and translation

The Feature Manipulation Engine from Safe Software supports over 300 geographic data formats. There are two translation issues that need to be addressed by geographic data translation software. First is syntactic translation which deals with translating digital symbols such as letters and numbers. Second is semantic translation which is much more difficult process and involves converting meaning inherent in geographic information (Safe Software, 2013).

3.3 **Creating and Maintaining Spatial Databases**

Spatial data is spatially referenced data that act as a model of reality. Spatial data represent the geographical location of features for example points, lines, area etc. Spatial data typically include various kinds of maps, ground survey data and remotely sensed imagery and can be represented by points, lines or polygons. According to (Anselin, 1992) the integration between GIS and spatial data analysis has always been an important demand in research community since Goodchild (1987). Several calls for development of GIS into a research tool to explore and analyze spatial relationships have been planed (e.g., Openshaw, 1990). An increasing number of review articles, conceptual outlines, and guides for practical implementation of the linkage were also performed (Anselin and Getis, 1992), (Bailey, 1992), (Fischer and Nijkamp, 1992), (Goodchild et al.,1992), and (Anselin, Dodson and Hudak, 1992).

A spatial database can be thought of any set of data describing the semantic and spatial properties of real world phenomena. Spatial databases used to be applied in a GIS and Computer-Assisted Design (CAD) system was joined with a Database Management System (DBMS). Application Programming Interface (API) can access to spatial engines to support the spatial database applications. Spatial database can be applied in a universal (object-relational) server with spatial extension or in a web server with spatial viewer. These spatial databases can use flat file, hierarchical, network, relational, object-oriented, multi-dimensional or hybrid structures. In addition, they can be organised in different architectures such as stand-alone GIS, client-server solutions, intranets or spatial data warehouses In fact, this evolution of the GIS market follows the general trends of mainstream ITs (Bédard, Merret, & Jiawei, 2001).

A *geodatabase* is considered as a critical part of a GIS both because of the cost of data collection and maintenance and because of its impacts on all the modelling, analysis and decision making activities. In recent years geographic database have become increasingly large and complex. Table 3.3 represents potential GIS database volumes for some typical applications. The size of a U.S National Image Mosaic will be over 25 terabytes and the Britain's MasterMap has approximately 450 million vector features in its database collected by the Great Britain Ordnance Survey (Longley P. A., Goodchild, Maguire, & Rhind, 2011). Both spatial and attribute data allows the database to be exploited in more ways than a conventional database allows, as GIS provides all the functionality of the DBMS and adds spatial functionality.

Table 3.3: Potential Volumes. (Longley P. A., Goodchild, Maguire, & Rhind, 2011) GIS database.

Database Volumes	
1 megabyte	Single dataset in a small project database
1 gigabyte	Entire street network of a large city or small country
1 terabyte	Elevation of entire Earth surface recorded at 30 m intervals Satellite
1 petabyte	image of entire Earth surface at 1 m resolution
1 exabyte	A future 3D representation of entire Earth at 10 m resolution

According to (Bédard, 1999) in a spatial database, analysis is the process of understanding and describing the users need. Analysis results in a formal and detailed database requirements specification. The action of defining and describing how the analysis result is called designing. Practical topics such as limitations of technology for supervising the spatial database, the desired performance and flexibility and the implementation of security needs are addressed in designing stage. Figure 3.5 shows the relationship and roles between GIS and DBMS software.

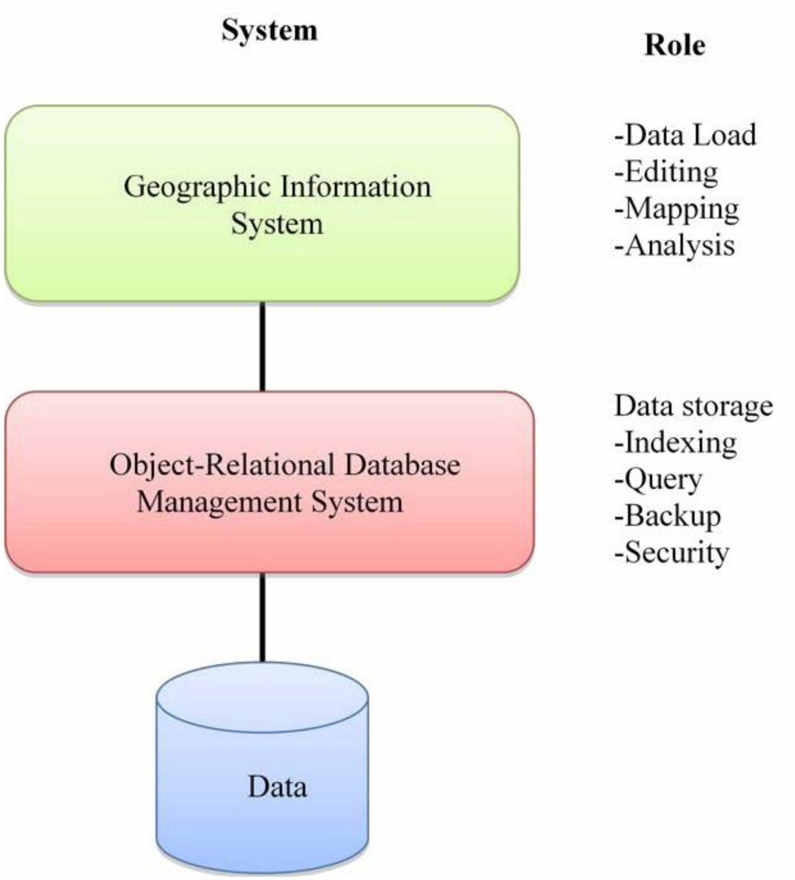


Figure 3.5: The tasks of GIS and DBMS

All operational GIS are founded upon a geographic database. Each GIS database is organised as a series of thematic layer that can be integrated using a geographic location and represent and answer questions about a particular problem. Designing a GIS database is based on identifying the thematic layers to be used and specifying the contents and representations of each thematic layer. The process of designing a GIS database uses key GIS and DBMS design methods.Each geodatabase design starts with identifying the geographic representations for each dataset. Individual geographic entities can be represented as feature classes (collection of points, lines, and polygons), imagery and raster, continuous surfaces that can be represented using features (such as contours), DEMs, or Triangulated Irregular Networks (TINs) and attribute tables for descriptive data (Esri, 2008). Figure 3.6 illustrates the more common geodatabase elements that can be employed in a geodatabase design.

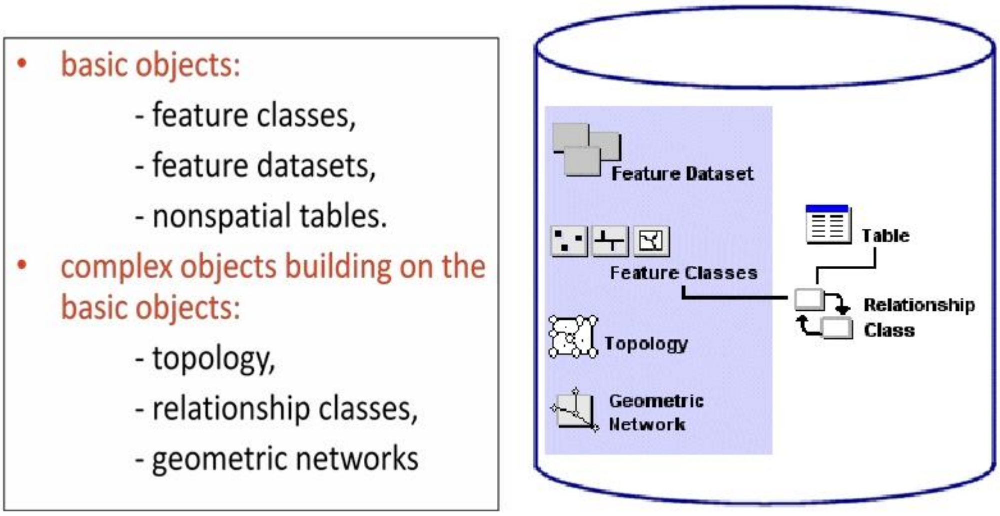


Figure 3.6: Geodatabase Objects

All of the elements of non-graphic geographic data are stored in geographic databases as alphanumeric characters. Graphic data in spatial databases includes points, lines, polygons, and other map features such as projections, coordinate systems, and cartographic symbols. Graphic data is commonly stored in the database in the form of coordinates, symbols, rules, or pixels. Each Geodatabase represents both the ordered collections of simple features (e.g., feature classes, feature dataset, non-spatial tables), as well as complex object (e.g., topology, relationship classes and geometric networks) properties that add rich GIS behaviour.

Every geodatabase will contain feature classes. A feature class is a collection of geographic features with the same geometry type (e.g., point, line, or polygon), a common set of attribute columns and the same coordinate system. However, most Geodatabases contain comprehensive data model that adds advanced geodatabase elements. Next step is organising spatially related feature classes into a common feature dataset. A feature dataset is a collection of spatially or thematically related feature classes that share a common coordinate system. Feature datasets are used to hold feature classes that participate in a shared topology, a network dataset, a geometric network, or a terrain (Esri, 2008).

Geodatabase analysis depends on accurately modeling spatial relationships thus creating topology is essential. For example if analysis involves calculating the total area for different types of land cover, gaps between land cover polygons will result in inaccurate totals. There are three types of topology available in the geodatabase: geodatabase topology (over 20 topology rules), map topology, and geometric network topology.

Each type of topology is created from feature classes that are stored within a feature dataset. A feature class can participate in only one topology at a time (Arcture & Zeiler, 2004).

There are four steps introduced by (Arcture & Zeiler, 2004) in designing a geodatabase:

- **Conceptual Design**
    - Identify the information products that will be produced by your GIS.
    - Identify the key thematic layers based on your information.
    - Specify the scale ranges and spatial representation for each thematic layer.
    - Group representations into datasets.
- **Logical Design**
- Define the database structure and behaviour for descriptive attributes.
    - Identify attribute fields.
    - Specify valid values and ranges.
    - Apply subtypes to control behaviour.
    - Model relationships.
- Define the spatial properties of the datasets.
- Propose a geodatabase design.
    - Define the set of geodatabase elements for each data theme.
    - Study existing designs for ideas and approaches that work.
- **Physical Design**
- Implement, prototype, review and refine your design.
- Design work flows for building and maintaining each layer.
- Document your design using appropriate methods.

The recent evolution of the software industry indicates that spatial databases are becoming mainstream solutions seamlessly integrated with non-spatial corporate data. This is happening more and more, with new categories of tools outside the traditional GIS packages. In comparison to non-spatial databases, these solutions offer a higher level of diversity both within and across categories. The complexity of spatial objects is also inherently higher; issues such as geometry, spatial reference systems, movement, spatial precision, spatial integration, metadata management, database versioning, data quality analysis, and so on may have a tremendous impact when designing a database for spatial querying, spatial analysis, spatial data exchange, and system interoperability (Oosterom, 1993).

3.4 Computation and Queries

Storing spatial data in a spatial database provides a simple Structured Query Language (SQL) for Create, Read, Update and Delete (CRUD) operations as well as analysis operations. Spatial databases provide the ability to use geography as a query parameter and operations parameter.

3.4.1 Spatial Column Types

A non-spatial database supports data types as strings, numbers and dates format. A spatial database includes one or more types for indicating geographic features. The Open Geospatial Consortium (OGC) defined basic model as Simple Features for SQL (SFSQL) for all the particular spatial database implementations. The basic geographic types are in spatial database include geometry, point which is a single coordinate with usually but not necessarily two dimensions, line-string which is a set of two or more coordinates with a linear version of the path between the coordinates. Polygons are another type supported in spatial database and include a set of one or more closed to linearrings, one exterior ring that defines a bounded area, and a set of interior rings that define exceptions such as holes to the surrounded areas. Multipoints, Multilinestrings, Multipolygons and Geometrycollections (a heterogeneous set of geometries) are additional data types in spatial database. The SFSQL specifies the particular rules for building valid geometries, the legal representations of geometries in both ASCII and binary form, and a set of basic functions for constructing, inspecting, measuring and manipulating geometries (Leslie, 2009).

3.4.2 Spatial Indexes

Spatial databases are usually large and geographic queries are computationally expensive. Geographic queries can take a very long time to complete. Topological structuring helps to speed up adjutancy and network queries. Another solution to assist queries is the use of index. In DBMS jarsgon, creating an index helps to avoid expensive full-table scans. Indexing for standard types (numbers, strings, and dates) is usually done with b-tree indexes. B-trees partition the data using the natural sort order to put the data into quickly findable buckets (Longley P. A., Goodchild, Maguire, & Rhind, 2011). The natural sort order of numbers, strings, and dates is pretty easy to determine - every value less than, greater than or equal to every other value. However, the natural sort orders of polygons are different. Since polygons can overlap, be contained in one another, and are arrayed in a multi-dimensional cartesian space, a B-tree cannot be used to efficiently index them. Spatial index gives fast answers to the queries looking for objects that are within a bounding box. A bounding box is the minimum rectangle capable of containing any given feature. Indexes have to perform quickly. So instead of providing exact results, as b-trees do, spatial indexes provide approximate results. A common geodatabase query looks up for lines inside a polygon. However, the index can efficiently find lines which have bounding boxes and are contained in this polygon's bounding box. There are verities of actual spatial indexes implemented by various databases. Quad trees and grid-based indexes can be implemented in shipping spatial databases, but the most common implementation is the R-tree. Figure 3.7 represents an R-tree index showing how each shape is proxied by a bounding box, and how the bounding boxes are built into a searchable tree structure (Leslie, 2009).

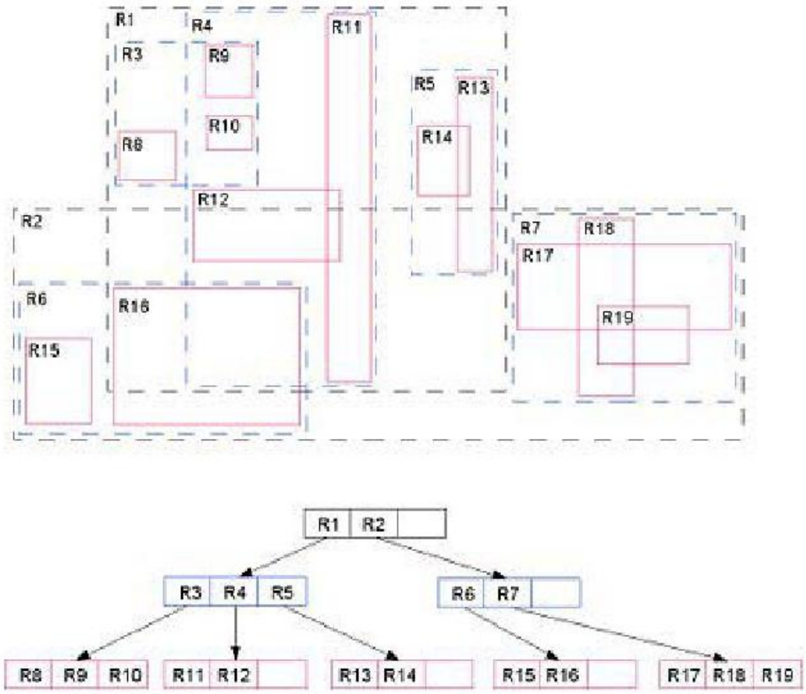


Figure 3.7: R-tree Index (Leslie, 2009).

Spatial indexes only provide the approximation. A functional spatial database includes functions capable of testing the relationships between geometries along with indexes. Implementing functions in geodatabase provides answers that are mathematically correct and computationally efficient (Arcture & Zeiler, 2004).

3.4.3 Spatial Functions

A fully functional spatial database provides a complete set of functions for analyzing the components of geometries, the relationships between geometries, and manipulating geometries. Common sets of functions are defined by the OGC SFSQL and implemented by all the spatial databases (Ryden, 2005).

According to (OpenGeo, 2013) these functions are categorised as below:

- Construction is used to build geometries from text and binary representations
- Serialization is used for exporting geometries into various text and binary representations (e.g., KML, GML, JSON and SVG)
- Predicates tests relationships between geometries and returning Boolean answers.
- Analysis and measurement are used to provide results in numerical summaries (areas, lengths, and distances) about geometries.
- Accessors strips out parts from geometries (e.g., rings from polygons, points from line-strings).
- Builders take geometry inputs and build new altered outputs.
- Aggregates capture geometry sets and return single resultants. Union is the most common aggregate function.

3.5 Web Services for Geospatial Applications

The recent years, evolution of GIS application enables the end users to apply GIS to a problem by accessing the data and the software remotely. In the last years, several Spatial Data Infrastructures (SDI) providing Geospatial web services have been built up. GIS are evolving from traditional client-server architecture to web service architec- ture to the response of demands for geospatial interoperability and adoption of the open standards. In the web service architecture the web is used for delivering not only data but also geo-processing functionality that can be wrapped in interoperable web services (Anderson & Moreno, 2003). In fact, web service is a kind of self-contained and self-described software components that can be discovered and invoked by other software components through the web. In the web services view, every different system or component offers some services for others, and every system does its job by just calling or combining suitable services over Internet (Cömert, 2004). Geospatial web services and Web services technologies are different from each other. Web services are composed of particular set of technologies and protocols but Geospatial web services are comprised of defined set of interface implementation specifications which can be implemented with diverse technologies (Volter, Kricher , & Zdun, 2005). Web services provide standard information interoperability. There are different definitions for web services. A web service is a self contained, modular application which is accessible through standard interfaces over the network. (Fisher, 2002) describes web services as an "applications accessible to other applications over the Web". The fundamental web service technology is the standard message exchange between the applications in their life cycle (Akinci, 2004).

Open Geospatial web services are online services that allow others to access data and maps hosted by another group using interoperable technology. The specifications for serving these services are defined by the OGC. Web Feature Service (WFS), Web Map Service (WMS) and Web Coverage Service (WCS) are the most fundamental Geospatial web services which are introduced by OGC There are several types of services and documents defined by OGC. WMS Services are more commonly used to serve maps and map layers, and perform basic queries about these layers. WFS allows for data features (as GML) to be accessed directly. Web context documents are standard documents that describe a particular map derived from WMS Layers. Catalogue Service (CS) provides catalogues for OGC web services and supports the ability to publish and search collections of descriptive information (metadata) for data, services, and related information objects (Zhang, 2005) (Figure 3.8).

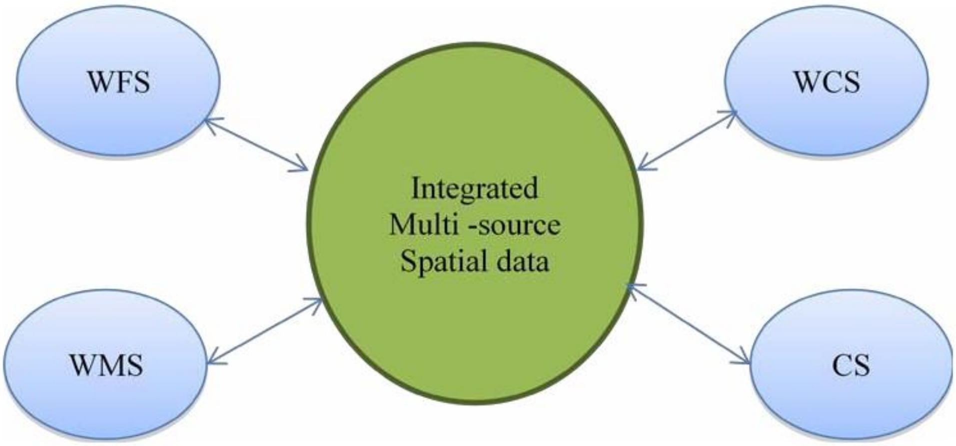


Figure 3.8: OGC Web Services

According to (Volter, Kricher, & Zdun, 2005) the main goal of Geospatial web services development was to share geospatial data and services among heterogeneous geospatial processing systems. At the same time, in Information Technology (IT) world, using Web services technologies are the best solution to provide interoperability among heterogeneous systems in scattered and decentralized environments (Figure 3.9).

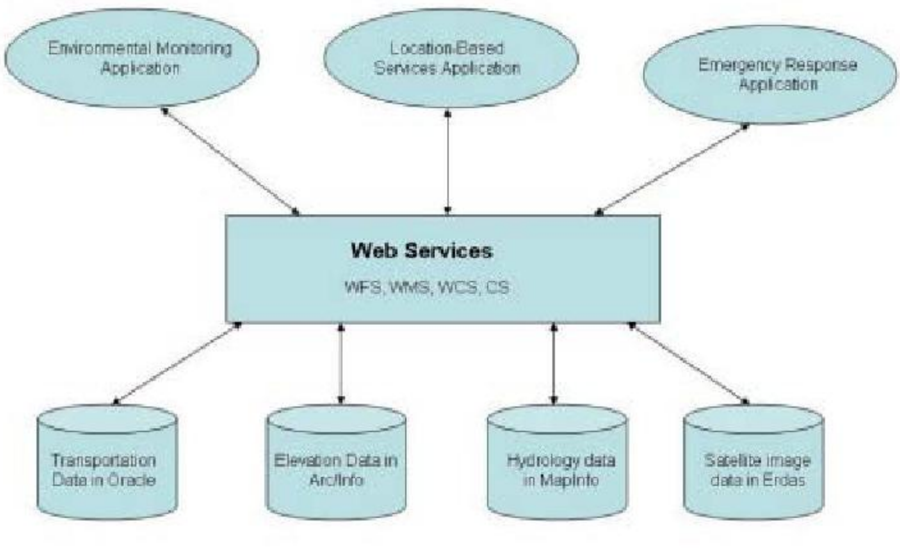


Figure 3.9: Sharing heterogeneous spatial data through web services (Zhang, 2005)

There has been an ongoing effort to produce standards-based specifications and approaches for detection, assessment, access, visualization and management of geospatial information. Geospatial web services are enabling information exchange approach with a support by OGC, and the adoption of these standards by mapping and surveying organizations. One of the major advantages of the Web services approach to geospatial data is the consumption of data as determined by user requirements (Kralidis, 2005).

There are at least three geospatial web based services with a focus on geospatial information. First, data discovery web services supports search and discovery to geospatial data and services. Second, data visualization web services that offer visualization images of the actual geospatial data. Finally, data access web services which are responsible for access to the actual geospatial data (FGDC, 2010) (Figure 3.10). Geospatial web services offer better control for data access, providing the user with the data on demand. They enable user to access information with filters allowing for data subsetting. With a hybrid finer-grained access to data, users can download data in specific area of interest with specific attribution criteria. Geospatial web services are based on standards based approaches and specifications from the OGC. OGC approaches are based on commonly adopted approaches in the broader IT / IM standards bodies (W3C, ISO, OASIS), these components represent, at a technical level, nothing more than another set of tools and applications which can be used by various IT developer / application communities (Kralidis, 2005).

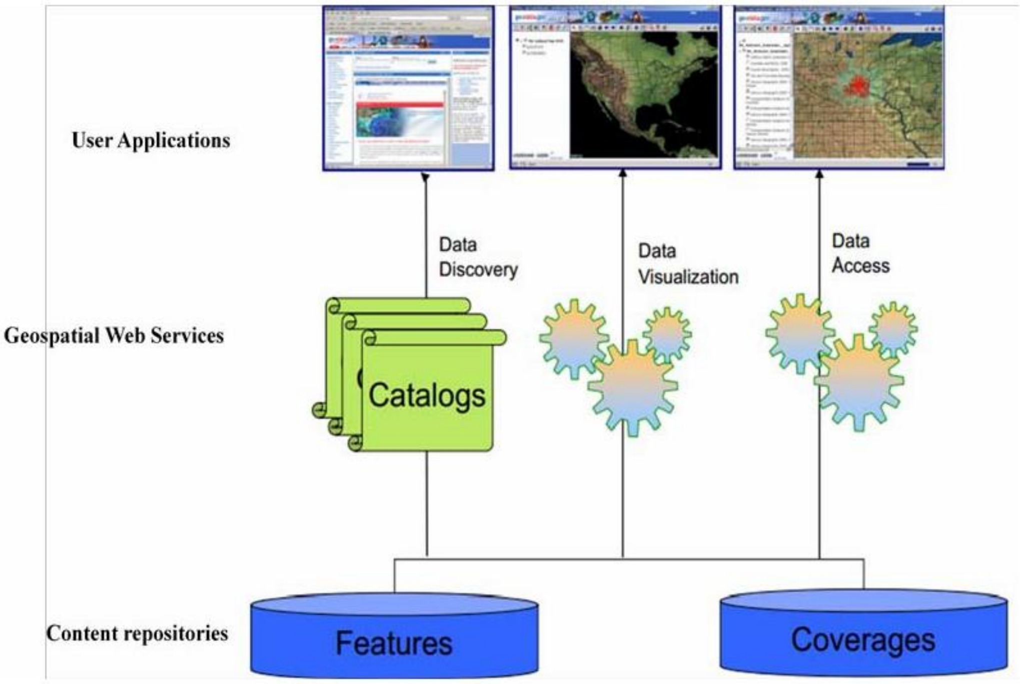


Figure 3.10: Types of Geospatial Web Services

3.6 **Conclusions**

This chapter has presented a selection of GIS application areas and specific instances within each of them. GIS is being implemented in a wide range of situations from day-to-day problem solving to fundamental curiosity- driven science. The staged of data collection methods have been discussed throughout the need of planning and understanding user's demand; the need to prepare data, so that the software and data are compatible; the need to digitize and improve data; and the need for data evaluations. A GIS software system includes an integrated collection of computer programs that implement geographic storage, processing, and visualisation function. GIS provide the necessary tools to load, store, edit, query, analyse and display spatial data. The cost of creation and maintenance of spatial database as well as its impact on all analysis, modelling and decision making, makes it a vital part of an operation GIS. Major commercial vendors have released spatial database extensions to their standard DBMS products. Although these systems are using different technologies, scope and features, they all store, mange, and query geographic objects. Spatial database design involves the creation of conceptual, logical, and physical design with practical steps. Spatial databases are usually large and make geographic queries computationally expensive. Spatial indexing is used to speed up and support efficient query and analysis in spatial databases. The rapid development of spatial technologies has enabled users to benefit from distributed GIS. The Geospatial web services offer a vision of a future in which geographic location is central to a wide range of geospatial applications and information technologies.

Chapter Four

4. **Surfaces and Field Analysis**

This section introduces different types of data models and focuses on how geographical reality is modelled in GIS. By way of introduction, there is a difference between the terms models and representation. According to (Longley, Goodchild, Maguire, & Rhind, 2011) model is used in practical and database situations, whereas representation is typically used in conceptual discussions.

4.1 **Modeling Surfaces**

Surface data are meteorological data that are measured at the earth's surface (technically, somewhere between the ground level and 10m). For every location in the set *{x,y}* there is a single scalar value *z* representing the value of measured parameter at that point. In some cases a number of z-values belong to each (*x,y*) location. The term surface or scalar field refers to information that is recorded in a single interval or ratio-scaled value. For example, spatial datasets such as atmospheric pressure at a particular altitude recorded at a specific time interval. Another example is in case of multi-spectral band sensing data, where separate spectral bands are coded for each pixel of an image and stored as separate layers within an image folder. The term vector field refers to the fields that have multiple values for the same variable at a single time and location (e.g. geological borehole data) and/or have directional factor (e.g. wind speed and direction) (Smith, Goodchild, & Longley, 2007).

Collection and storage of geospatial information become a fundamental pre-requisite to process of analysing and interpreting such data. Data that describes the physical world are obtained from National Mapping Agencies (NMA) and sample surveys (Smith, Goodchild, & Longley, 2007).

At the national level, the NMAs are playing a key role in the development of national Spatial Data Infrastructure (SDI) (LAND, 2003) and provide topographic maps and geographic information derived from aerial, satellite and terrestrial surveys (Smith, Goodchild, & Longley, 2007). These data are analysed and stored in paper map or digital formats. The elevation of the terrain over a specified area, usually at a fixed grid interval over the surface of the earth is stored in DEMS (Barringer & Lilburne, 1997). The closer together the grid points are located, the more detailed the information will be in the file. However, since these values are generated using computer programs there is tendency in appearance of the artefacts even in most accurate national datasets such as GB Ordnance Survey and United States Geological Survey (USGS) datasets (Smith, Goodchild, & Longley, 2007). Most DEMs are interpolated from the most commonly available source of topographic data -digital contours which in turn have been generated photogrammetrically from aerial photographs. In many cases a quantitative assessment of DEM accuracy is not addressed, and to secondary parameters (e.g., slope and aspect) error propagation is not considered (Fryer, 1994).

Sample survey is one of the principal sources of filed data. A limited set of sample values are collected for a set of distinct locations and can be used to calculate the values at the infinite number of points across the surface. Values at unsampled location - defined as a fine grid- are assigned to the surface by performing interpolation and prediction methods on measured data. Grid data may be subject to estimated or un- known errors of modelling.

4.1.1 Data Model Overview

Specific details of related aspects of the real world can be fitted in a framework known as a data model. Data models are the bases of any GIS which represents objects in the digital environment of the computer (Maguire & Grisé, 2001). A data model also can be described as a template for data with a set of assumption about the nature of data.

There is no single type of all-encompassing GIS data model that is best for all situations as the GIS is being used for different purposes and the characteristic of phenomenon that end users study are different.

"…Human perception of space is frequently not the most efficient way to structure a computer database and does not account for the physical requirements of storing and repeatedly using digital information. Computers for handling geographical data therefore need to be programmed to represent the phenomenological structures in an appropriate manner…" (Burrough & McDonnell, 1998). This pertinent quotation reminds that as well as defining the entities the computer require unambiguous instruc- tions to turn specific entity data into digital representations. According to Longley, Goodchild, Maguire and Rhind (2011) for a successful design of spatial data models within a GIS it is helpful to consider the real world in terms of four levels of abstrac- tions. A data model provides system developers and users with a common understanding and reference point.

Users and system developers participate in a data modelling process that successively engages with each of following levels:

- Reality which is a collection of real world events such as building and people.
- Conceptual model which is human-oriented model of particular objects that are related to problem domain.
- Logical model which is represented as lists and diagrams.
- Physical model that corresponds to the actual implementation in GIS and includes tables and data files and the relationship between objects types stored in database

Figure 4.1 shows four levels of GIS data model abstraction (Longley, F.Goodchild, J. Maguire, & W. Rhind, 2011).

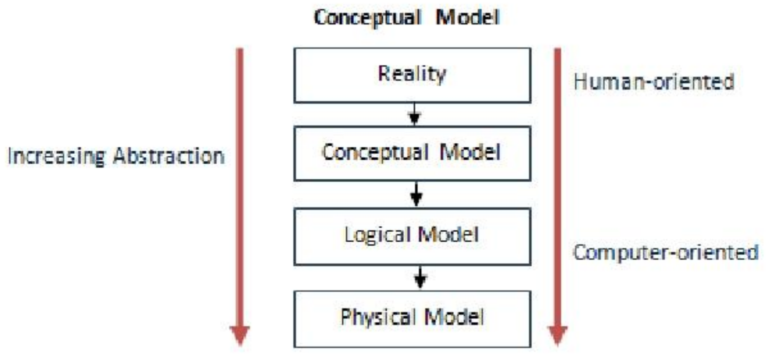


Figure 4.1: Levels of GIS data model abstraction.

Data model enables developers to represent an application domain in such a way that could be translated into a design and implementation of a system. For users, it describes the structure of the system, independent of specific items of data or details of the particular application.

4.1.2 GIS Data Models

Data model is a collection of construct to represent real world objects in the digital environment of the computer. A data model is the basis of any GIS and choosing the right type of data model are critical to the success of a GIS project and have a direct influence on the analysis results. Operational GIS are used by GIS users for creating maps, querying database or performing analysis. Figure 4.2 illustrates the role of data model in GIS.

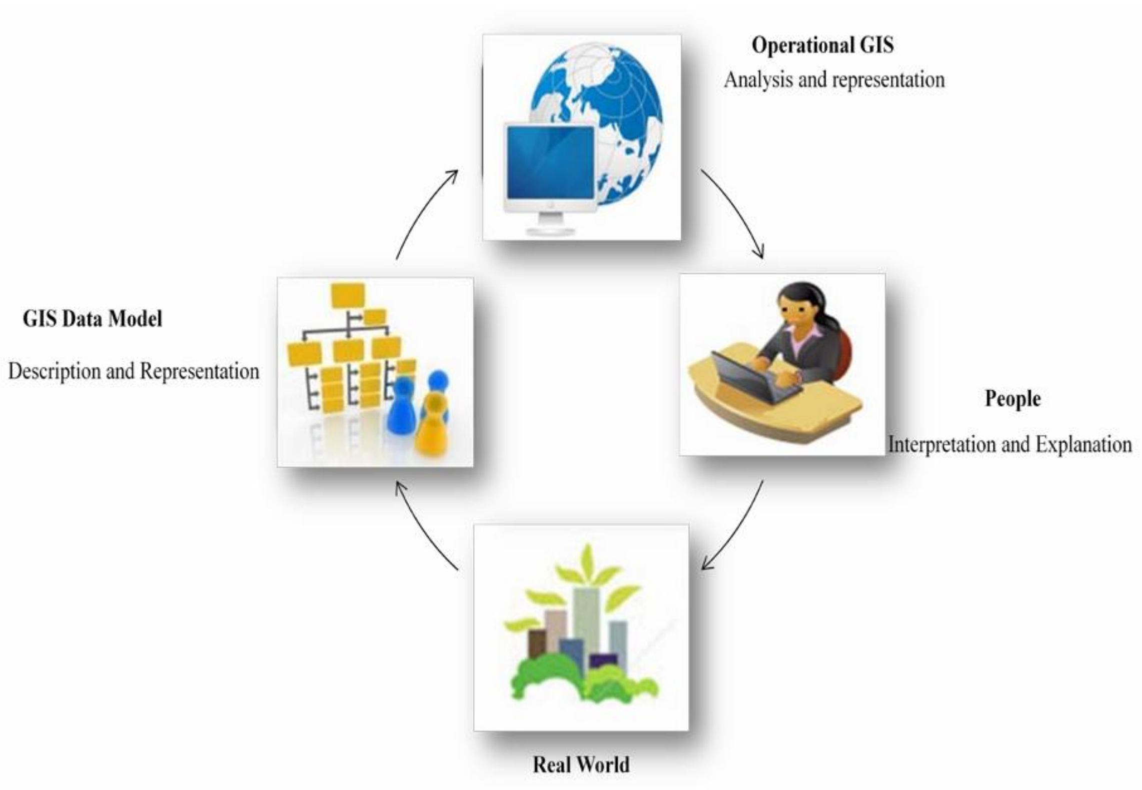


Figure 4.2: The role of data models in GIS.

The key types of geographic data models are raster/grid, vector/geo-relational topological, network, TIN and object data models. Surfaces can be represented using vector models such as contour or iso-lines and TINs or raster models. The most surface analysis in GIS is done on raster or TIN data. Entities of the same geometric types group together (dimensionally) are called Class or layer. Layers are also a general term for datasets in GIS. Different layers of different geometric types can be combined into one grouped layer which makes it easier to implement validation and editing rules and building relationship between entities (Longley, F.Goodchild, J. Maguire, & W. Rhind, 2011).

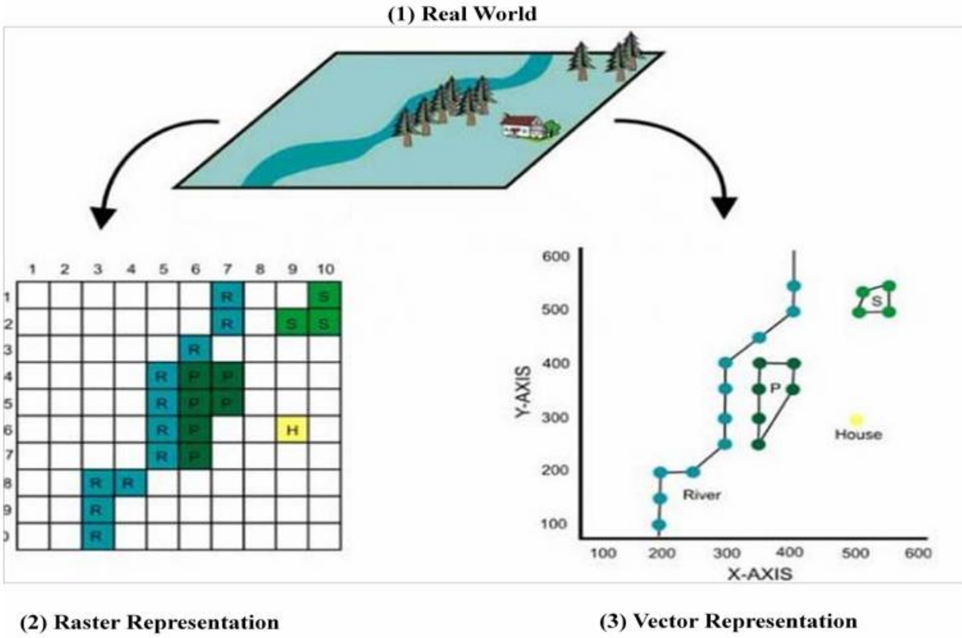


Figure 4.3: Comparison of the Raster and Vector Models. The landscape in 1 is shown in a raster representation (2) and in a vector representation (3). The pine forest stand (P) and spruce forest stand (S) are features. The river is a line feature, and the house is a feature point (GIS Cookbook, 2007)

4.1.2.1 Raster Data Models

According to (Bernhardse, 2002) raster data are applied in at least four ways:

1. Models describing the real world
2. Digital scans of existing maps
3. Compiling digital satellite and image data
4. Automatic drawing driven by raster output devices

A raster data model is composed of rectangular arrays of regularly spaced square cells. Raster data model can represent real world objects using a network of square cells laid over the landscape holding the value of attributes with metadata in each cell. Each row in attribute table is either a cell or a pixel class and each column is an attribute (UW, 2010). Because squares or rectangle are often used and a graphical view of them resembles classic grids of squares, it is sometimes called a grid model. Metadata include geographic coordination of the upper-left corner of the grid, row and column numbers, size of cell and projection. The data arrays are usually saved as compressed file or a record in a Relational Database Management System (RDBMS) (Bernhardse, 2002). While any type of geographic data can be stored in raster format, raster datasets are especially suited to the representation of continuous, rather than discrete, data. Figure 4.4 demonstrate a raster data representation. All commonly used programming languages support array handling therefore raster data can be managed easily in computers. However, one of the main operational problems with a raster is the inefficiency in terms of computer storage space. To improve the storage efficiency many compression technique have been developed.

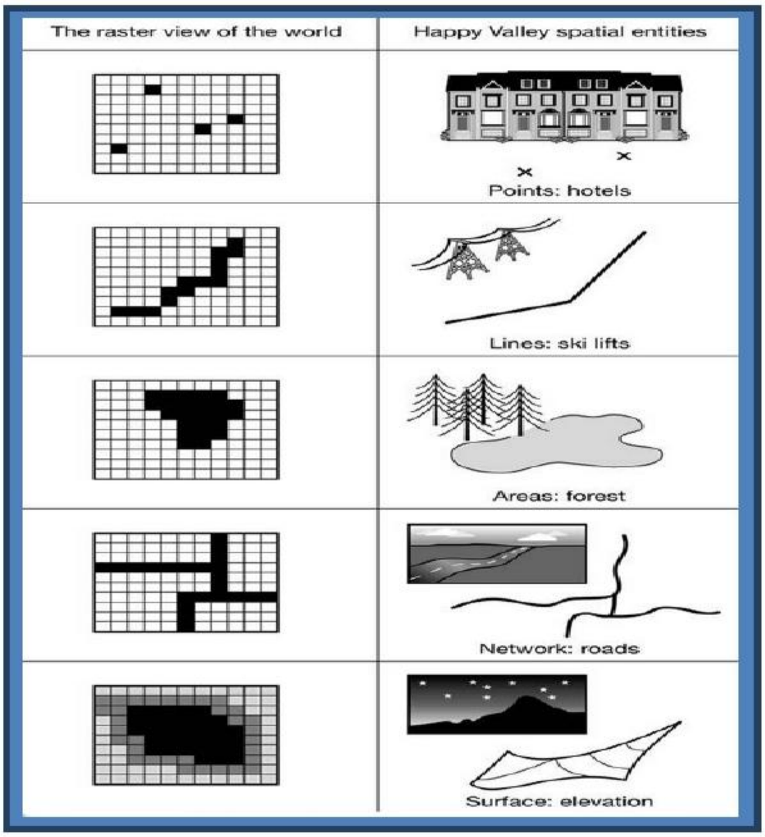


Figure 4.4: Raster Data Model

4.1.3 Compression techniques- an Overview:

Compression methods are most effective on raster data with large homogeneous areas. The effect of method is rather small if the raster have continues changes in cell values (Bernhardse, 2002).

4.1.3.1 Run-length coding

Geographical data tends to be "spatially auto-correlated", meaning that objects which are close to each other tend to have similar attributes. "All things are related, but nearby things are more related than distant things" (Tobler 1970). Because of this principle, neighbouring pixels are expected to have similar values. Therefore, instead of repeating pixel values, the raster can be coded as pairs of numbers - (run length, value). The run-length coding is a simple compression method and is widely used. The primary data elements are pairs of values, involving a pixel value and a repetition count which specifies the number of pixels in the run. Data are built by reading successively row by row through the raster, creating new values every time the pixel value changes or the end of the row is reached (GITTA, 2010).

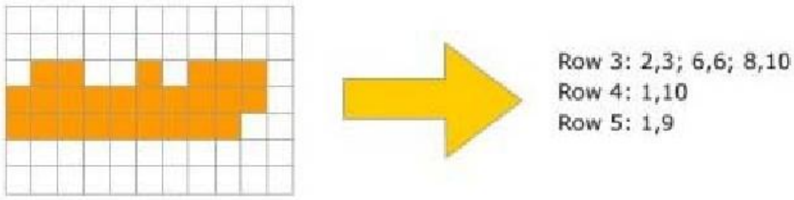


Figure 4.5: Interior of an area by run-lengths

If a run is not required to break at the end of each line data compression can be applied even further. (Figure 4.5, Codes -III)

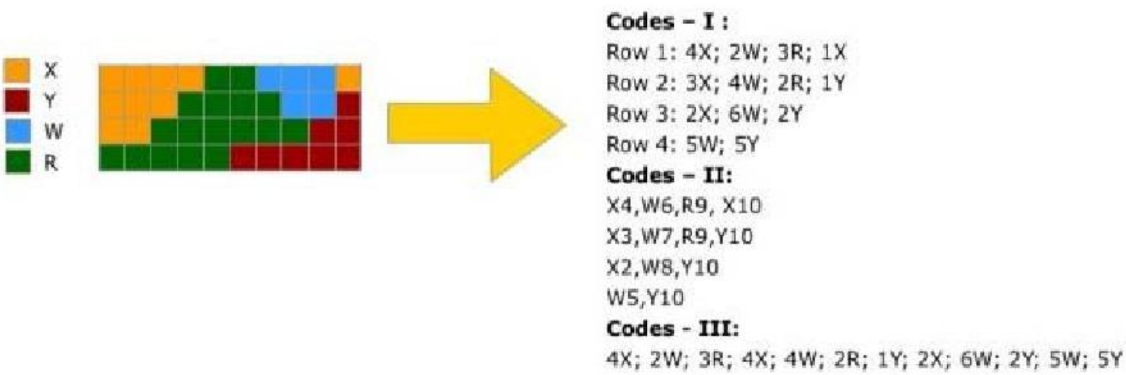


Figure 4.6: Applying run-length in a multiple attribute case

Run-length coding on DEM data or any other type of data where neighbouring pixels mostly have different values is not very effective.

4.1.3.2 Block Wise coding (lossless)

This method is a two-dimensional version of run-length encoding. The homogeneous quadratic areas are stored as one block. For each block the position, the size and, the contents of the pixels are stored. Column and rows number are used to show the position of block's centre or bottom left corner. The radius from the blocks centre to the cells at the corners of the block indicates the size of the block. The data volume increases with the storage of a single cell value, but the volume will be reduced with squares of four or more cells (Bernhardse, 2002). According to Green and Rapek (1990) the quadrants that contain part of the region subdivided, which adds another level to the tree. A black square represents a quadrant that is filled completely by part of the region (Figure 4.7).

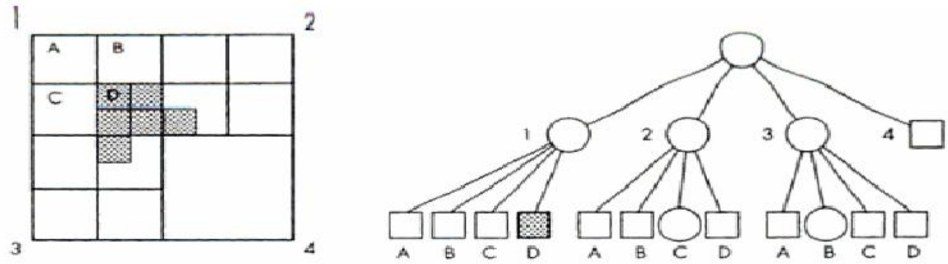


Figure 4.7: Black Square (Adapted from Green and Rapek 1990)

A larger square could comprise several raster cells with similar values. Homogeneous region that are not square or does not match with the pattern of squares in use can be divided into homogeneous squares (Bernhardse, 2002). There are four ways to organize your raster data (raster data models): the raster dataset, the mosaic dataset, the raster catalogue, and raster that are attributes of a feature.

4.1.3.3 Quad-tree coding

The quad-tree compression technique is the fastest compression method and is very widely used. Quad-tree coding stores the information by subdividing a square region into quadrants, each of which may be further subdivided in squares until the contents of the cells have the same values (GITTA, 2010). Bernhardse (2002) describes quad-tree structure as an inverted tree whose leafs is the indicators to the attributes of the homogenous squares and branches are pointers to smaller values. Subdivision of a tree is continued until each leaf is represented either by a black or a white square (Figure 4.8). (Adapted from Green and Rapek 1990)

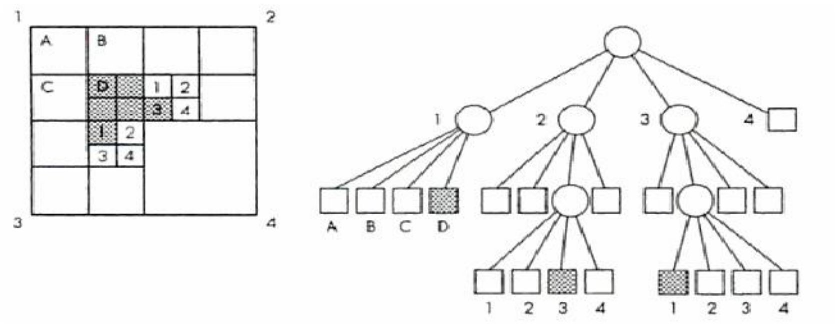


Figure 4.8: Only 19 leaves are needed to store this region into a quad-tree square

The advantages and disadvantages of quad-tree methods are listed by (Sahoo, 2012) as below:

The advantages of the quad-tree model are:

- Rapid data manipulation, because homogeneous areas are not divided into the smallest cells used.
- Rapid search, because larger homogeneous areas are located higher up in the point structure
- Compact storage because homogeneous squares are stored as units
- efficient storage structure for certain operation, including searching for neighbouring squares or for squares containing a specific point

The disadvantages of the quad-tree model are:

- The structure set up involves significant processing time.
- expanded processing may involve alternations and updating
- Data must be relatively homogenous
- Complex data may require more storage capacity than ordinary raster storage.

4.1.3.4 Comparing raster data storage models

One of the major design decisions in managing raster data is whether to store all the data in a single dataset or as a catalogued collection of many datasets. According to (Esri, 2008), there are four possible storage models for multiple raster datasets: Firstly, store each raster dataset individually, secondly mosaic them into one large raster dataset, thirdly store them as members of a mosaic dataset and finally store them as members of a raster catalog.

Datasets which are not neighbouring to each other or are rarely used on the same project are usually stored in the individual raster datasets. Mosaicking inputs together to form one large, single extent of raster data is suitable for many applications;

4.1.3.4.1 Raster datasets

A Raster dataset is a single picture of an object or a seamless image of a spatially continuous area. The image can be a single original image or the result of many images mosaicked together. Data in raster datasets are homogeneous data with a single format, data type. Metadata are stored once and will be applied to the complete dataset. Raster datasets are preferred for fast display of large quantities of raster data at any scale. They are also used when the overlaps between mosaicked images do not need to be taken. Raster datasets can be used as a data source in many geo-processing and analysis tools as well as image Analysis window. However, raster dataset is slower to update for file and personal geo-database because the entire file has to be rewritten (Esri, 2008).

4.1.3.4.2 Raster catalogs

A collection of raster datasets are displayed as a single layer in Raster catalog. Raster catalog can be in different coordinate systems and can have different data types. Data stored in Raster catalog can have several formats, file sizes and data types. Data are stored as attribute columns for each raster dataset item in the raster catalog. Raster catalog datasets can manage multi-row raster tables for many purposes. Raster catalog are recommended for massive image repositories and preserving overlaps between datasets. They can be served for managing time series data, and when differences among adjoining images prevent mosaicking. One of the main operational problems associated with raster catalog datasets is that displaying many raster dataset items from a raster catalog can be time consuming. They also cannot be used in the Image Analysis window. File-based raster catalogs with different data types may not also render well (Esri, 2008).

4.1.3.4.3 Mosaic dataset

Mosaic dataset is a collection of raster datasets stored as a catalog and provides more storage, management, browsing, and querying collections of raster data. It is presented as a mosaic but each raster dataset in the collection are accessible.

Data in Mosaic datasets are heterogeneous data which may have multiple formats, data types, file sizes, and coordinate systems. Metadata can be stored within the raster record as well as attributes in the raster catalog table. Many geo-processing and analysis tools use Mosaic dataset as a data source. Mosaic dataset manages large collections of raster data. Mo- saiced images can be generated with no loss of data by taking advantage of the mosaic on-the-fly processing methods. Mosaic datasets are fast to display at any scale and can be used in the Image Analysis window (Esri, 2008). According to (ArcGIS, 2012) it is recommended to use a mosaic dataset for multidimensional data, querying, storing metadata, and overlapping data, and it provides a good hybrid solution.

According to (Smith, Goodchild, & Longley, 2007) datasets encoded using the raster data models look like conventional map and communicate information fast which make them practically useful as a backdrop map. They are commonly associated with the field conceptual data model and are widely used for analytical applications such as surface-water flow analysis and disease modelling. In overall, Mosaic dataset and raster catalog are good ways to manage a large number of raster datasets. Mosaic dataset or unmanaged raster catalog can maintain a directory or listing of data holdings. However, mosaic datasets are the recommended choice as it is a catalog of raster data that can be viewed as a single raster dataset (mosaicked image).Also, a referenced mosaic dataset can be stored outside a geo-database as an Administrative Template (.amd) file.

4.1.4 Vector data models

The precise nature of vector data model representation methods and its efficiency in data storage as well as the quality of its cartographic output make it widely used in GIS. The vector data model is associated with the discrete object view and each object in real world is classified into geometric types: point is recorded as single coordinate pairs, line as a set of ordered coordinates, or polygon as one or more line segments that form a polygon area. The coordinates may have two (x, y: latitude and longitude), three (x, y, z: latitude, longitude and elevation) or four (x, y, z, m: latitude, longitude, elevation and time) dimensions (Longley P. , 2005). According to Bernhardse (2002) lines can include certain lines or part of lines (e.g., road curves) that can be defined by mathematical functions.

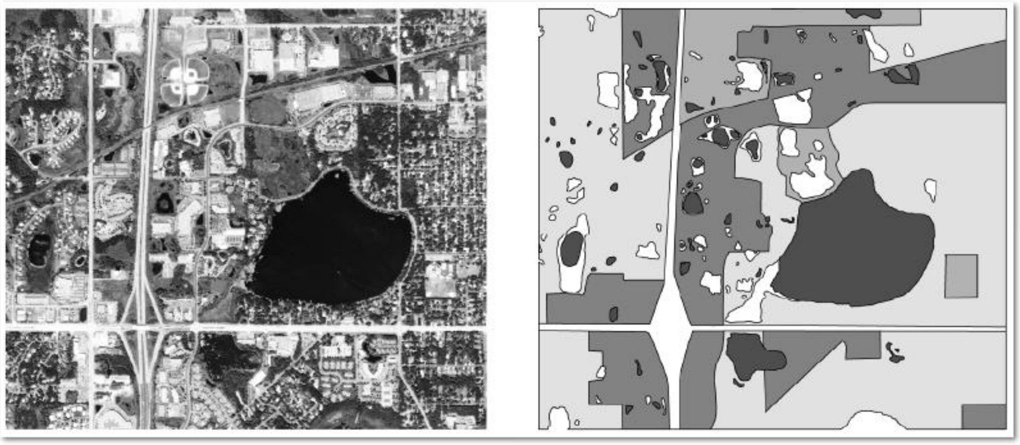


Figure 4.9: Vector representation of hydrography from an aerial photo

4.1.4.1 Features

Features are discrete geographic phenomenon in coverage with its vector data model. Features of the same geometric type stored in a geographical database are called feature classes. There are two types of features: simple and topological. The Open Geospatial According to (Neumann, Freimark, & Wehrle, 2010) consortium (OGC) defines a standard that offers the digital storage of geographical data with spatial attributes as well as non-spatial attributes. This specification, called Simple Features also includes spatial operators that can be used to generate new geometries based on existing geometries. Simple feature datasets are useful in GIS because they are easy to create and store and fast to retrieve. However, simple feature datasets are short of advanced data structure such as topology and do not support operations such as shortest-path network analysis and polygon adjacencies (Longley, Goodchild, Maguire, & Rhind, 2011).

"Topology is the science and mathematics of relationships used to validate the geometry of vector entities Topological features is simple features with advanced data structures using topological rules." Topological rules empower GIS in data validation, editing, modelling incorporated feature behaviour and query optimization (Longley, F.Goodchild, J. Maguire, & W. Rhind, 2011). Topological or topology-based data are useful for detecting and correcting digitising errors (e.g. two lines in a roads vector layer that do not meet perfectly at an intersection). Topology is necessary for applying spatial analysis such as network analysis. More complex definitions of points, lines, and polygons can be used to capture the internal structure of an entity; these definitions may be functional or descriptive. Topological links can be used to indicate how lines are linked into polygons or linked networks, respectively. In recent years more highly structured ways of encapsulating entity data have been possible through the object-oriented approach (Burrough & McDonnell, 1998).

Topology utility in GIS can be grouped into two broad areas. The first of these is to support spatial analyses. These can include using connectivity for network analysis; area definition to determine containment; and contiguity for neighbourhood analyse (Bansal V. K., 2011). Testing the topological integrity of a dataset is a useful approach to validate the geometric quality of the data and to inspect if it is appropriated for geographic analysis (Topology and Topological Rules, 2008). Four useful data validation topology tests are provided by Longley (2005). First is Network connectivity which connects all network elements to create a graph and "snappes" all network elements at junctions together. Second is line intersection which tests if there are junctions at intersecting polylines, but not at crossing polylines (e.g., at a bridge or underpass).Third one is overlap validation test which checks if adjacent polygons overlap as it is important to build databases free from overlaps and gaps (e.g., land ownership). The other usage of topology utility in GIS is to support the development of spatial database. Knowledge of the topological circumstances in a data set is useful to discover structural problems with the feature database (e.g., polygons which are not closed or are overlapping). It can also be used to automatically create features and ensure feature integration (Topology and Topological Rules, 2008).

According to (Longley, Goodchild, Maguire, & Rhind, 2011) topology also improves the productivity of editors in several ways. It enables manipulation of common, shared polylines and nodes as single geometry objects and ensures that the common geometries are alike. Rubberbanding is another way of improvement of editors' productivity. Rubberbanding procedure moves the node, polyline, or polygon boundary and receives interactive response about the location of all topologically connected geometry. The productivity of editor can be improved by "snapping" technique which speed up editing and keeps the quality data in high standard. Tracing a network analysis method mostly used in utility applications and empower the user is to ensure the connectivity of liners. Many GIS queries can be optimised by computing and storing the information in advance. Optimisation of queries can be used if network tracing, polygon adjacency, containment and intersections (Maguire, Goodchild, & Rhind, 1991).

4.1.5 Network

The network data model is a special type of the topological feature model. There are two primary types of networks: radial (e.g., stream and storm drainage) and looped (e.g., water distribution networks) .Flows always has an upstream or downstream direction in radial or tree networks whereas in looped networks self-intersections are common occurrences (Longley, Goodchild, Maguire, & Rhind, 2011). In GIS, network topological relationships define how lines are connected together at points. Defining the direction of the flow through the network is useful for the purpose of network analysis. The rate of the flow is modelled as weights on the nodes and lines. According to (Longley, 2005) in geo-relational implementation of topological network feature model, the information of geometry and topology is typically stored in computer files and the attribute are held in a linked database.

4.1.6 Triangulated Irregular Network (TIN)

As surface structures have dimensional properties between 2-D and 3-D, the term 2.5-D is used to describe them. A TIN is a digital data structure used in a GIS to create and represent a surface. A TIN is a vector-based representation of the physical surface, made up of continues no overlapping triangular and irregularly distributed nodes and lines in a three dimensional coordinates (x, y, and z). TINs are often derived from the elevation data of a rasterized DEM (Longley, Goodchild, Maguire, & Rhind, 2011). Tessellation or tiling of the surface is a collection of plane figures that fills the surface with no overlaps and no gaps. Generalizations to higher dimensions are also possible. Three-dimensional visualizations are created by reproduction of the triangular surface. In regions where there is little variation in surface height, the points may be widely spaced whereas in areas of more intense variation in height the point density is increased (EMU, 2008).

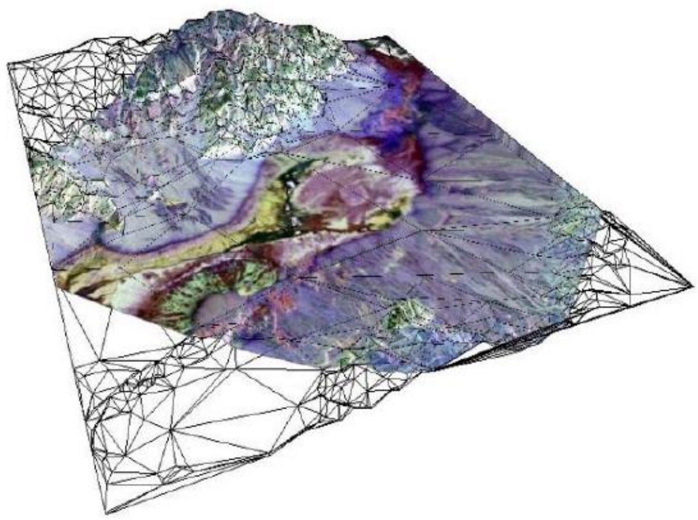


Figure 4.10: TIN surface of Death Valley, California

4.2 Interpolations Methods and Creating Surface

Interpolation is the process of predicting the values of a certain variable of interest at unsampled locations based on measured values at points within the area of interest (Burrough and McDonnell 1998). According to (Sahoo, 2012) spatial interpolation is a process of using points with known values to estimate values at other points forming the surface. The motivation behind spatial interpolation is based on Tobler's first law of Geography according to which "Everything in space is related to every other thing but points close together are more likely to be similar than the points which are far apart". In general, two study points of a few meters apart are more likely to have the same altitude than points on two hills some kilometres apart.

In meteorology and climatology, spatial information is almost entirely founded on measurements at certain points. Selecting a suitable spatial interpolation method that transforms points to surfaces plays a major role in meteorology and climatology (Luo, Taylor, & Parker, 2013). These methods have been applied to various disciplines. It was used by Journel and Huijbregts (1978) in mining engineering and Goovaerts (1997), Burrough and McDonnell (1998) and Webster and Oliver (2001) applied these methods in environmental sciences.

Since spatial interpolation is based on statistics, there are inevitably a certain assumptions and optimizations. There are several factors affecting the performance of spatial interpolation methods. The errors are mainly generated from sample data density (Stahl, Moore, Floyer, Asplin & McKendry, 2006), grid size or resolution (Hengl, 2007), sample spatial distribution (Collins & Bolstad, 1996), data variance (Schloeder, Zimmerman, & Jacobs, 2001), surface types (Zimmerman, Pavlik, & Ruggles, 1999). It is difficult to select an appropriate interpolation method for a given input dataset when there are no consistent findings about the effect of these factors on spatial interpolator's performance (Li & Heap, 2011). Errors generated by spatial interpolation and their propagation in analysis models will certainly influence the quality of any decision-making supported by spatial data.

4.2.1 Feature of Spatial Interpolation Methods

This section is concerned with the philosophy and several approaches to spatial interpolation.

4.2.1.1 **Deterministic versus Stochastic**

Interpolation techniques can be classified into deterministic and Geo-statistical methods. Deterministic interpolation methods use mathematical functions to calculate the values at unknown points based on either the extent of similarity (e.g. Inverse Distance Weighting (IDW)) or the degree of smoothing (e.g. Radial Basis Functions (RBF)) in relation with neighbouring data points (Peralvo M. , 2004). Deterministic methods do not use any probability theory. They explain spatial phenomena by physical means. In practice deterministic methods are preferable because they are conceptual and less abstract. This quality makes the deterministic approach especially important (Luo, Taylor, & Parker, 2013).

Stochastic methods include the idea of uncertainty meaning that the interpolated surface is conceptualised as one of many that might have been observed and all of which could have produced the known data points. They often calculate the computation of statistical importance of the surface and uncertainty of the predicted values. Stochastic methods for spatial problems are often referred to as Geo-statistics (Luo, Taylor, & Parker, 2013). A defining concept of Geo-statistics is that spatial phenomena are spatially autocorrelated. In other words, locations that are close together tend to be more similar than locations that are far apart from each other (Peralvo M., 2004). Geo-statistical techniques use both, mathematical and statistical methods to predict values at unmeaseured points. These methods provide probabilistic estimates of the quality of the interpolation based on the spatial autocorrelation among data points and an estimate the accuracy of the predictions by using the statistical properties of the measured points (Burrough & McDonnell 1998, Esri, 2008).

4.2.1.2 Exactness

Interpolators can be either exact or inexact. Result values in exact interpolators are identical to the measurements at sampled point; wherase the predicted value of inexact interpolators are different from the observed values at given points. Exact interpolators honour the data points upon which the interpolation is based. The surface passes through all points whose values are known (Luo, Taylor, & Parker, 2013). Sharp peeks are avoided in inexact interpolators which results in a smoother and more reasonable output surface. The statistics of the differences between predicted and measured values are commonly used as estimators of model quality (Burroughs & McDonnell 1998). According to (Luo, Taylor, & Parker, 2013) inexact interpolators are used when there is an uncertainty about the given values. In many data sets there are global trends, which vary slowly, overlaid by local fluctuations, which vary rapidly and produce an error in the recorded values. The effect of smoothing will reduce the effects of error on the resulting surface.

4.2.1.3 Global versus Local

In addition, interpolation methods are generally divided into two groups of local and global methods. Local methods operate within a small zone around the position of the predicted point (neighborhood). IDW, local polynomial, and RBF are exapmles of local deterministic methods. Global interpolation methods operate on the whole extent of the study area and use all the available measured points to create predictions for the whole area of interest (Peralvo M. , 2004). These methods are mostly used for evaluation and removing global variations caused by physical trends in the data.

4.2.1.4 Gradual versus Abrupt

Another distinguishing characteristic of spatial interpolators is the smoothness of the predicted surface that is produced. Some methods produce a discontinuous surface with sharp and abrupt changes, while gradual interpolators (e.g., distance-based weighted averages) produce a relatively smooth and gradual surface. The proximal Nearest Neighbour (NN) method, sets unknown points equal to the nearest measured point and produces an abrupt surface. The criteria used in the selection of the weight values in relation to the distance effects the degree of surface smoothness. Criteria include simple distance relations (e.g., IDW), minimisation of variance (e.g., Ordinary Kriging), and minimisation of curvature and enforcement of smoothness (e.g., Splines) (Li & Heap, 2008).

4.2.1.5 Inclusion of Covariates

A final characteristic is the ability of an interpolator to include additional variables other than distance between points in the interpolation process. In some cases inclusion of other variables in the spatial interpolation model can provide more accurate predicted surface. These additional information identifies changes associated with the spatial process that otherwise would be missed. Including additional variables helps to reduce the uncertainty of the interpolation model's output if, in fact, the covariates explain some of the spatial variability in the process of interest. Elevation, topography, and land use are examples of Geographic variables of use in explaining spatial variation. In air pollution modelling, other covariates that might improve predictions include other correlated or meteorology information. For example, if ozone concentrations are correlated with temperature, adding temperature to an ozone spatial interpolation model should improve the accuracy of the predictions (EPA, 2004).

4.2.2 Deterministic Interpolation Methods

Local interpolation methods use a small area around the interpolated point and estimates oh unknown values are made with data from the location in the immediate neighbourhood to get the best fitting. Spline, Inverse Distance Weighting IDW and Thiessen polygons are the most common local interpolation methods. A cross validation procedure can assist in the choice of model and parameters.

4.2.2.1 Inverse Distance Weighting (IDW)

IDW is a type of deterministic method for multivariate interpolation with a known scattered set of points. The assumption is that each input point has a local influence that reduces with distance. The IDW technique calculates a value for each node by examining surrounding data points that lie within a inter-point distance. The value of the unvisited point is the distance-weighted average of data points occurring within a neighbourhood or area surrounding the unvisited point (Smith, Goodchild, & Longley, 2007).

As the distance increases, weight reduce by a factor of p. Weighting of the sampled locations depends on the power parameter p. In other word, when distance increases the weight decreases too (Sahoo, 2012).

There are very few decisions to make regarding model parameters. The main factor affecting the accuracy of IDW is the value of the power parameter (Isaaks & Srivastava, 1989). The choice of power parameter and neighbourhood size is arbitrary (Webster & Oliver, 2001). The most popular choice of p is two and the resulting method is often called Inverse Distance Squared (IDS). The power parameter can also be chosen on the basis of error measurement (e.g., minimum mean absolute error, result in the optimal IDW) (Collins & Bolstad, 1996). The incensement in power parameter results in smooter surface, and it was found that the estimated results become less accurate when p is one and two compared with p value of four (Ripley, 1981).

IDW calculation is fast and simple and it is good way to take a first look at the interpolated surface. IDW gives reasonable results for many types of data. However, there are a few drawback s with the IDW method. First, the choice of the weighting power may misrepresent the data as too smooth or too rolling if little is known about the underlying surface. Weights reduce as the distance increases, especially when the value of the power parameter increases, so nearby samples have a heavier weight and have more influence on the estimation, and the resultant spatial interpolation is local (Isaaks & Srivastava, 1989). Second, IDW interpolations are easily affected by uneven distributions of observational data points since an equal weight will be assigned to each of the points even if it is in a cluster. Third, the map quality need to be assessed by taking extra observations as there is no built-in method of testing for the quality of predictions. Finally, the interpolated values of any point within the dataset are bounded by the maximum and minimum of the observed data points as IDW is a smoothing technique (Luo, Taylor, & Parker, 2013). This is considered to be an important weakness because, in order to be useful, an interpolated surface should predict accurately certain important features of the original surface. For example, the locations and magnitudes of maxima and minima, even when they are not included as original sample points (Lam, 1983).

4.2.2.2 Radial Basis Functions (RBF)

At early days, cartographers used rules called splines to fit smooth curves through a number of fixed points. Spline functions are mathematical equivalents of theses flexible rulers (Burrough & Mcdonnell, 1998).

Splines are repeated approximation of a third order trend surface to a limited numbers of known points. The splines consist of polynomials with each local polynomial of degree p. The polynomials describe pieces of a line or surface and are fitted together so that they join smoothly (Burrough & McDonnell, 1998; Webster & Oliver, 2001). The places where the pieces join are called knots. The choice of knots is arbitrary and may have a dramatic impact on the estimation (Burrough & McDonnell, 1998). For degree p = 1, 2, or 3, a spline is called linear, quadratic or cubic respectively. Typically the splines are of degree three and they are cubic splines (Webster & Oliver, 2001).

RBF is the name given to a family of five deterministic exact interpolation techniques: thin-plate spline, spline with tension, completely regularized spline, multi-quadratic function and inverse multi-quadratic function. In an exact interpolation the predicted values are identical with those measured at the same point and the generated surface requires passing through each measured points. The predicted values can vary above the maximum or below the minimum of the measured values (Nikolova and Vassilev, 2006). The estimated values of the methods are based on a mathematical function that minimises total curvature of the surface, generating quite smooth surfaces.The only assumption in RBF is that the input data are linear (Smith, Goodchild, & Longley, 2007). The predict values might be above the maximum or below the minimum of the measured values. A smoothing parameter controls the smoothness of the RBF's resulting surface. The estimated values of the methods are based on a mathematical function that minimizes overall surface curvature, generating quite smooth surfaces (Sahoo, 2012).

According to (Luo, Taylor, & Parker, 2013) the general definition of RBF is given by a linear grouping of the basic functions.

$$Z(s_0) = p(x) + \sum_{i=1}^{n} w_i \emptyset\left(\|s_i - s_0\|\right) \quad (1)$$

the term $p(x)$ represents a polynomial of degree at most k, and the second term represents its 'proximity' or 'fidelity' to the data. Here, Ø(r) is a radial basis function, $r = \|s_i - s_0\|$ is Euclidean distance between the prediction location s_0 and each data location s_i, and $\{w_i : i = 1, 2, \ldots, n\}$ are estimated weights.

There are a few differences for the basic function such as Thin-plate spline, Multi-quadric function and Inverse multi-quadric function. Each of these basis functions has a different shape and creates slightly different interpolation surface. For smoothing, thin-plate splines are commonly used to interpolate elevation to create digital elevation and interpolate large areas fast and efficiently.

The interpolating values can be calculated fast as they use quite few points at a time. The measurement errors associated with the data are small; hence the predictions are very close to the values being interpolated. They are analytic, flexible and unlike the TSA, RBF save small-scale features (Buhmann, 2003). The RBF methods are able to predict values above the maximum and below the minimum measured values as in the cross-section of the data points. The smoothness of RBF means that mathematical derivatives can easily be calculated for direct analysis of surface geometry and topology (Luo, Taylor, & Parker, 2013).

In other hand; RBF performs poorly on data sets with sudden changes in a small distance.
The most significant drawback of RBF method is that a view of reality generated by some of the functions (i.e. thin-plate spline) is unrealistically smooth (Luo, Taylor, & Parker, 2013). Other disadvantage is that there are no direct estimates of the errors associated with RBF interpolation, though these may be obtained by a recursive technique known as 'jack-knifing' (Burrough & Mcdonnell, 1998).

4.3 Geo-statistic Interpolation Methods

Point pattern analysis involves the ability to describe patterns of locations of point event and test whether there is a significant occurrence of clustering of points in a particular area. In the recent years, many different fields have started to use point pattern analysis. Climatologists conduct "Hot Spot" analysis in order to understand location of frost. This information can be used to model scenarios for desired optimal cultivation management outcomes. The results can be used to modify and optimize both growing and production methods.

4.3.1 Geo-statistics

Geo-statistics is a branch of applied statistics developed by George Matheron of the centre de Morophologie Mathematicque in Fontainebleau, France. The initial reason of introducing Geo-statistics was to calculate ore grade changes within a mine. However, the principles have been applied in geology and other scientific disciplines (McKillup & Darby Dyar, 2010). A biblio- graphic research (Zhou, Huai-Cheng, Yun-Shan, & Chao-Zhong, 2007) found that the top ten area that use Geo-statistics are geosciences, water resources, environmental sciences, agriculture or soil sciences, mathematics, statistics and probability, ecology, civil engineering, petroleum engineering, and limnology.

Geo-statistics was defined by (Smith, Goodchild, & Longley, 2007) as "... models and methods for data observed at a discrete set of locations, such that the observed value, z_i, is either a direct measurement of, or is statically related to, the value of an underlying continues spatial phenomenon, F(x,y),at the corresponding sampled location(x_i,y_i) within some spatial region A." A unique aspect of Geo-statistics is the use of regionalized variable where variables are a mixture of random and completely deterministic variables. The phenomenon demonstrates spatial connections; however, it is not always possible to sample the entire study area. Therefore, unknown values must be estimated from data collected from particular locations that can be sampled. The size, shape, orientation, and spatial arrangement of the sample area control the ability of predicting the unknown points (Sahoo, 2012).

Variograms

Spatial autocorrelation between sample points can be analyzed using correlograms, covariance functions and variograms (semivariograms) which can provide the information for optimising interpolation weights and search radii. In this section variograms are discussed as one of the most important elements of regionalized variable theory. Regionalized variable theory uses a geo-statistical measure called the semivariance to express the degree of spatial dependence between points on a surface. Semivariance is used for descriptive analysis where the spatial structure of the data is investigated using the semivariogram. The semivariogram is the first step towards a quantitative description of the regionalized variation by (Smith, Goodchild, & Longley, 2007).

The semivariance is half the variance of the difference among possible points which are distributed in a constant distance (Luo, Taylor, & Parker, 2013). A semivariogram bases these predictions by the level of spatial autocorrelation, that is, dependence between sample data values which decrease as the distance between observations increase (Lam, 1983). Hence, the semivariogram acts as dissimilarity function as the variance of the variation increases with distance.

There are several characteristic that are associated with semivariogram function. Sill is the semivariogram value at the height that it levels off. It is composed of the nugget effect and the partial sill. The nugget effect is discontinuity at the beginning and represents the non-spatial variation due to measurement error and micro-scale variation. The distance at which the semivariogram reaches to the sill is called the range Figure (4.12).

In empirical semivariogram a gradual increase beyond the global variance value is a sign of significant spatial trends in the variable. Spatial trends result in a negative correlation between variable values separated by large lags(Smith, Goodchild, & Longley, 2007).

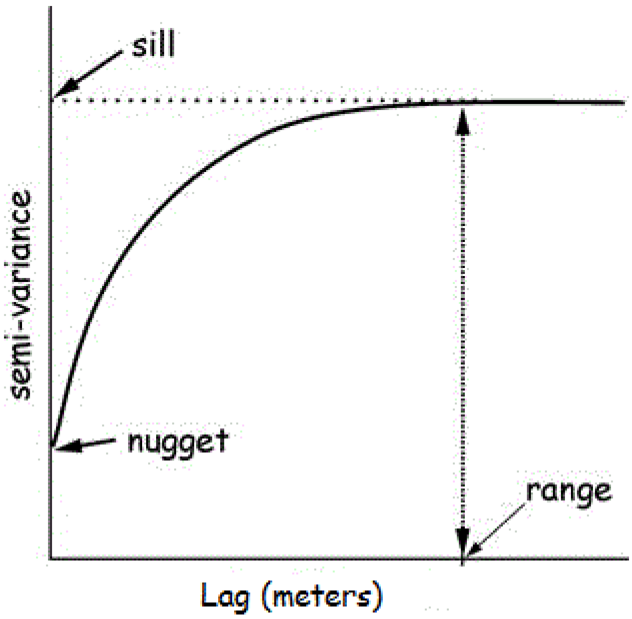


Figure 4.11: Sill, Range and Nugget

Calculation of a semivariogram model starts by creating an experimental variogram using a series of available data. The semivariogram provides information about a regionalized variable, including the size of the zone of influence around a sample, the isotropic or anisotropic character of the variable, and the consistency of the variable through space (Cressie, 1993, Collins, 1995). However, in stochastic simulation the empirical semivariogram is replaced with an acceptable semivariogram model. Therefore, geo-statisticians choose from a palette of acceptable or licit semivariogram models such as nugget model, spherical model, exponential and Gaussian model (Otero, 2007).

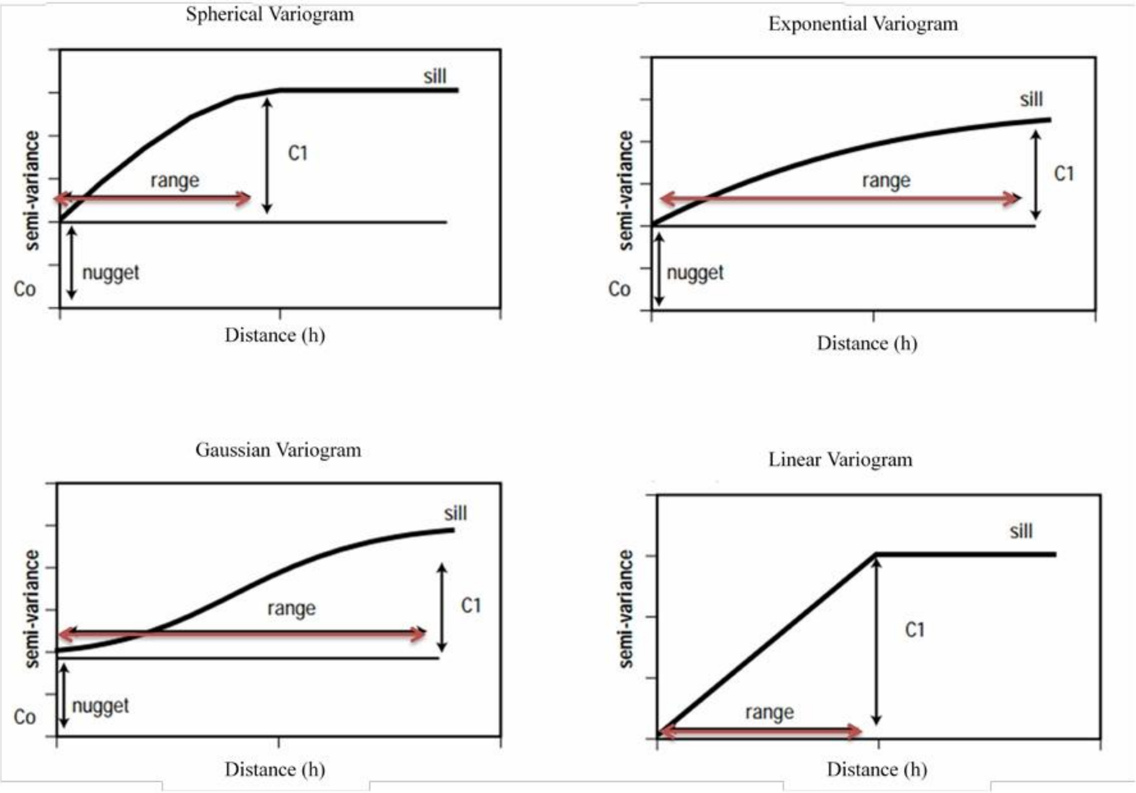


Figure 4.12: Most commonly used variogram models spherical, exponential, Gaussian and Linear

The nugget model represents the discontinuity at the origin due to small-scale variation. The spherical model actually reaches the specified sill value at a specified range. The exponential and Gaussian approach the sill asymptotically, with a representing the practical range, the distance at which the semivariance reaches 95% of the sill value. Estimation of variogram and modeling is central part for interpolation and structural analysis. A weighted least squares method is usually used to find the best fitting of variogram models. Even so, variogram fitting is an interactive process requiring considerable judgement and skill. Since empirical semivariograms are often quite noisy, quite a bit of subjective judgment goes into selecting a good model (Cressie, 1985) and (Bohling, 2005).

4.3.2 Kriging

Kriging is a very distinctive interpolation method and was invented by the aid of a South African mining engineer named D.G. Krige. Krige used empirical observations of weighting to estimate the ore content of mined rock by comparing known values sampled from early mining explorations (Cressie, 1993, Collins, 1995, Lam, 1983). Kriging is a stochastic method that uses a linear combination of weights at sampled points to estimate the value at an unknown point (Collins, 1995). Burrought and Mac-Donnell (1998) define kriging as a multi-stage procedure of interpolation that builds upon calculation and modelling of the experimental variogram. Kriging assumes that the spatial variation of an attribute such as elevations of the land surface is neither totally random nor deterministic. There are three components used in spatial variation: a spatially related component to represent the variation of the regionalized variable; a structure to express a trend and a random error expression.

Different assumptions about the regionalized variables have led to development of varieties of kriging such as ordinary, universal, simple and indicator. In this study, Ordinary Kriging is discussed. Ordinary kriging assumes that mean is constant and unknown and focuses on the spatial parts. In this method only data points in the local neighborhood are used for the estimation.

4.3.3 Ordinary Kriging (OK)

Ordinary kriging is by far the most common type of kriging in practice. It uses dimensionless points to estimate other dimensionless points (i.e. soil pH in first case study). Unlike IDW, kriging method is not deterministic but extends the proximity weighting approach of IDW to incorporate random components where the exact point is unknown. In ordinary kriging, the regionalized variable is assumed to be stationary and no drift exists. This assumption allows for an estimate of an unknown value (Cressie, 1988).

From a theoretical point of view, kriging employs the theory of regionalized variables to produce the drawing of statistical inference. There is advantage in using kriging over polynomial interpolation and other interpolation methods, since the degree of interdependence of the sample points has been taken into account. The kriging also moderate the impact of clustering because the weight assigned to a sample is lowered to the degree that its information is duplicated by nearby, highly correlated samples. Kriging provides an estimate of the error of the unknown points. This error information reveals the density of control points and the spatial correlation degree within the surface. The error maps analysis can help with planning the future sampling. However; gaining an optimal estimation is computationally complex. Computer processing times will be much longer if a map is generated by kriging because large set of synchronized equations must be solved for every unknown grid. As there are properties that not independent and the system is underdetermined, trial-and-error trialling is necessary to determine the best combination. For this reason, to warrant the best possible estimates of the surface, kriging probably should be applied in those instances where the best possible estimates of the surface are essential, the data are of reasonably good quality and estimates of the error are needed (Sheimy, 1999).

4.4 **Conclusions**

This chapter has presented a review of spatial interpolation methods, which are relevant for GIS applications. The spatial interpolation methods are developed for specific types of environmental data and variables. Each method has its own specific data requirements, assumption and properties. Many factors have an effect on the performance of spatial interpolation methods. Some examples of these factors are sampling density, sample spatial distribution and clustering, surface type, data variance, data normality, quality of secondary information, and grid resolution or size. In some cases the interactions among different factors may affect the accuracy of the result. However, there are no consistent findings about how these factors affect the performance of the spatial interpolators.

IDW method can be used on vast variety of data. Comparing with IDW method, most of RBF methods are aesthetically user friendly and generates a clear outline of the data fast. In other hand, RBF can be misleading in estimation of attribute values for numerical models and generate over smoothed surfaces. The kriging estimate is an optimal unbiased estimate. Kriging method provides an estimate of the error of the unknown points, an asset not provided by other interpolation procedures. The error map provided by kriging is very useful in analyzing the reliability of each feature in the kriged map. However; kriging is a time consuming procedure as data must be studied in advance to consider the stationarity, determination of the semivariograms, adjustment of the neighborhood size and selecting the proper order of the trend if there is any. Generating a kriging map is also taking longer comparing to RBF and IDW methods.

Chapter Five

5. **Case Studies**

This chapter presents two case studies that apply and evaluate various methods discussed in Chapter Four. The objective of the first case study is twofold: first to examine spatial variability in soil pH in two different horizons. Second objective is to generate maps of soil pH using three common interpolation methods naming OK, RBF. The accuracy and efficiency of the generated maps are examined and the most fitting technique for the soil properties in the study area is also identified. The second case study uses regression analysis and kernel estimation method to model errors' contributed from infrequent satellite overpasses of estimated rainfalls.

5.1 **Case study one:** Evaluation of Spatial Interpolation Techniques for Mapping Soil ph and Moisture in Different Depths

5.1.1 Introduction

Adjusting soil acidity or alkalinity improves soil nutrition without adding extra fertilizers. Soil nutrients needed by plants in the largest amount are referred to as macronutrients. In addition to macronutrients, plants also need trace nutrients and both macro and trace nutrient availability is controlled by soil pH. Understanding of spatial variability of soil properties is important in site-specific management. Analysis of spatial variation of soil properties is fundamental to sustainable agricultural and rural development. The special variability of soil property is often measured using various interpolation methods resulting in map generation. Selecting a proper spatial interpolation method is crucial in surface analysis, since different methods of interpolation can lead to different surface results. Managing spatial variability, which is popularly known as precision farming is essential for serving dual purpose of enhancing productivity and reducing ecological degradation.

IDW and its modifications are the most often-applied deterministic interpolation method (Nalder & Wein, 1998). OK, IDW, and RBF are three well-known spatial interpolation techniques commonly used for characterizing the spatial variability and interpolating between sampled points and generating the prediction maps. The variety of available interpolation methods has led to questions about which is most appropriate in different contexts and has stimulated several comparative studies of relative accuracy.

In this study, three common interpolation methods are used to study the spatial distributions of soil pH in a vineyard. Interpolation techniques were used to estimate the pH measurement in unsampled points and create a continuous dataset that could be represented over a map of the entire study area. The method investigated includes; IDW, RBF and OK. The performance of conventional statistics showed that soil pH had a law variation in this study.

5.1.2 Related Work

Site-specific management has received considerable attention due to the three main potential benefits of increasing input efficiency, improving the economic margins of crop production and reducing environmental risks. Uniform management of crops grown under spatially variable conditions can result in less than optimum yields due to nutrient deficiencies as well as excessive fertilizer application that may potentially reduce environmental quality (Redulla *et al*., 1996). Improvement of soil productivity, quality and capacity of soil also known as soil restorative is the basis of a sustainable agricultural system. Soil pH has an influential affect on plant nutrient availability by controlling the chemical forms of the nutrient. Knowledge about spatial variation of soil properties is considered a key variable when implementing "good farming rules" towards sustainable rural development (Karydas *et al*., 2009). The sustainable soil properties for efficient crop production depend heavily on the structural properties and the concentration of the soil solution. Spatial variability of soil properties is somewhat inherent in nature because of variations in soil parent materials and microclimate (Zhao et al. 2007). Knowledge of soil spatial variability and the relationships among soil properties is important for evaluating agricultural land management practices (Huang et al. 1999). Geostatistics can be used for studying and predicting the spatial structure of georeferenced variables, generating soil properties map and understanding the distribution of these (Krasilnikov et al. 2008).

Soil variability in thefield is generally defined with classic statistical methods and is assumed to have a random variability (Cemek et al. 2007). Soil variability occurs as a result of the effect and interaction of various processes in the soil profile (Parkin 1993). According to Webster (1985) soil characteristics generally show spatial dependence. Samples close to each other have similar properties than those far away from each other. However, the classical statistics, assuming the measured data to be independent, is not capable of analyzing the spatial dependency of the variables (Vieira et al. 1983).

Managing spatial variability, which is popularly known as precision farming is essential for serving dual purpose of enhancing productivity and reducing ecological degradation (Rabi N, 2003). Among statistical methods, geo-statistical kriging-based techniques have been often used for spatial analysis (Deutsch, 2002). IDW and its modifications are the most often-applied deterministic interpolation method (Nalder and Wein, 1998). OK, IDW, and RBF are three well-known spatial interpolation techniques commonly used for characterizing the spatial variability and interpolating between sampled points and generating the prediction maps. The variety of available interpolation methods has led to questions about which is most appropriate in different contexts and has stimulated several comparative studies of relative accuracy. Geostatistical analysis is used to dene the spatial dependency of soil properties, both isotropically fi and anisotropically (Burgess and Webster 1980; McBratney and Webster 1983). Proper geostatistical analysis will provide valuable information for the spatial distribution of soil properties in agricultural field (Liu and Yang 2008).

5.1.3 Materials and Methods

A. Study area and data sampling

The data were collected from the Kumeu vineyard located in Kumeu region in Auckland, New Zealand in May 2011 (Figure 5.1). Fifty four soil samples were collected as part of an ongoing research project in the Geoinformatics Research Centre (Scannavino et al., 2011). The size of the study area was approximately 400 x 150 m. Soil samples were collected from two depths: 5-25cm, 25-45cm. Soil pH was measured three times on the field using Field Scout pH 110 Meter data logger. Soil moisture and temperature were collected along with the geo-coordination for each sampling point.

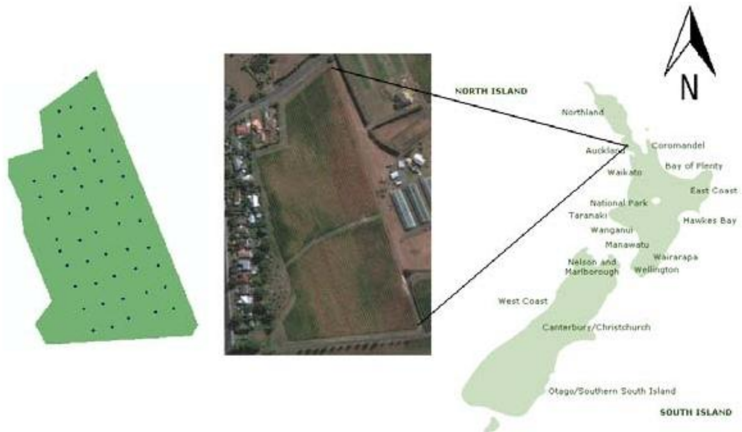


Figure 5.1: Location of study area and sampling patterns

B. Methods

Data analysis was performed in three stages: (1) Conventional statistics such as mean, maximum, minimum, median, Standard Deviation (S.D), Coefficient of Variation (CV), and skewness were performed using Minitab 16 Statistical Software; (2) Anderson Darling normality test was applied to analyse frequency distribution and normality of data; (3) Both deterministic and geostatistical interpolation methods was used to estimate soil pH values at unsampled points and generate their distribution maps. The methods investigated in this case study include; IDW, RBF and OK. Deterministic interpolation techniques create surfaces from measured points, based on either the extent of similarity (IDW) or the degree of smoothing (RBF). Geostatistical interpolation techniques (OK) use the statistical properties of the observed points. Geostatistical techniques quantify the spatial autocorrelation among measured points and account for the spatial configuration of the sample points around the prediction location (Esri, 2008).

ArcGIS 10 (ESRI 2010) was used to examine the spatial coloration between measured sample points through semivariograms functions. Experimental anisotropy semivariogram were examined to model the spatial relationship in the dataset and to find out the best fit model that passes through the points in the semivariogram. In this study, Spherical, Exponential, Gaussian and K-Bessel functions were inspected to determine the best fitted model in OK interpolation method. Model selection for semivariograms was done by comparing the deviation of estimates from the measured data and performing a cross-validation test over the entire dataset. The values of Mean Error (ME), Root Mean Square Error (RMSE), Mean Standardized Error (MSE) and Root Mean Square Standardized Error (RMSSE) were estimated to prove the performance of the best fitted theoretical models. The performance of interpolation techniques, in terms of the accuracy of predictions, was based on the comparison of the measure of accuracy and on one measure of effectiveness, namely the Goodness-of-Prediction Estimate (G). The G measure gives an indication of how effective a prediction might be. A G value equal to 100% indicates a perfect prediction.

Geostatistical parameters such as range, nugget and nugget ratio values were calculated for pH. Sill is the lag distance between measurements at which one value of dataset does not weight its neighbouring values. Range is the distance at which the variogram reaches the sill value (Zhao et al. 2007; Baalousha 2010). Theoretically, the semivariogram value at zero separation distance (i.e., lag = 0) is zero. However, at an infinitely small separation distance, the semivariogram often shows a nugget effect, which is some value greater than zero. The spatial dependency of soil properties was graded based on the nugget variance effect. The properties was considered strongly dependent if the rate was equal or lower than 25%, fairly dependent if it was between 25 and 75%, and weakly dependent if it was 75% or more (Cambardella et al. 1994).

5.1.4 Results

The results of performing the conventional statistics on available dataset and the summery statistic are shown in Table 5.1. Coefficient Variation (CV) is the most discriminating factor for describing variability of soil properties than the other parameters such as SD, mean, median, etc. (Xing-Yi et al.2007). In this study, soil pH had CV of 7.24% and 6.24% in depths A and B respectively. According to the grading of CV for soil property variations provided by Warrick (1998), there was a law variation (CV <15%) of soil pH in both depths in this study area.

Figure 5.2 shows the performance of Anderson-Darling normality test on soil pH values for both depths. The calculated P-Value for soil pH at both depths was greater than 0.05 and the data points almost follow a straight line. A graphical summary of descriptive statistic is presented in Figure 5.3.

5.1: Statistical analysis of soil ph in sampling area

Depth	Min	Max	Mean	Median	SD	CV	Skewness	Kurtosis
(A): 5-25	4.75	6.82	5.86	5.84	0.42	7.25	0.11	0.26
(B): 25-45	5.13	6.82	6.06	6.05	0.39	6.40	-0.04	-0.32
Avg	4.94	6.82	5.96	5.94	0.41	6.83	0.04	-0.03

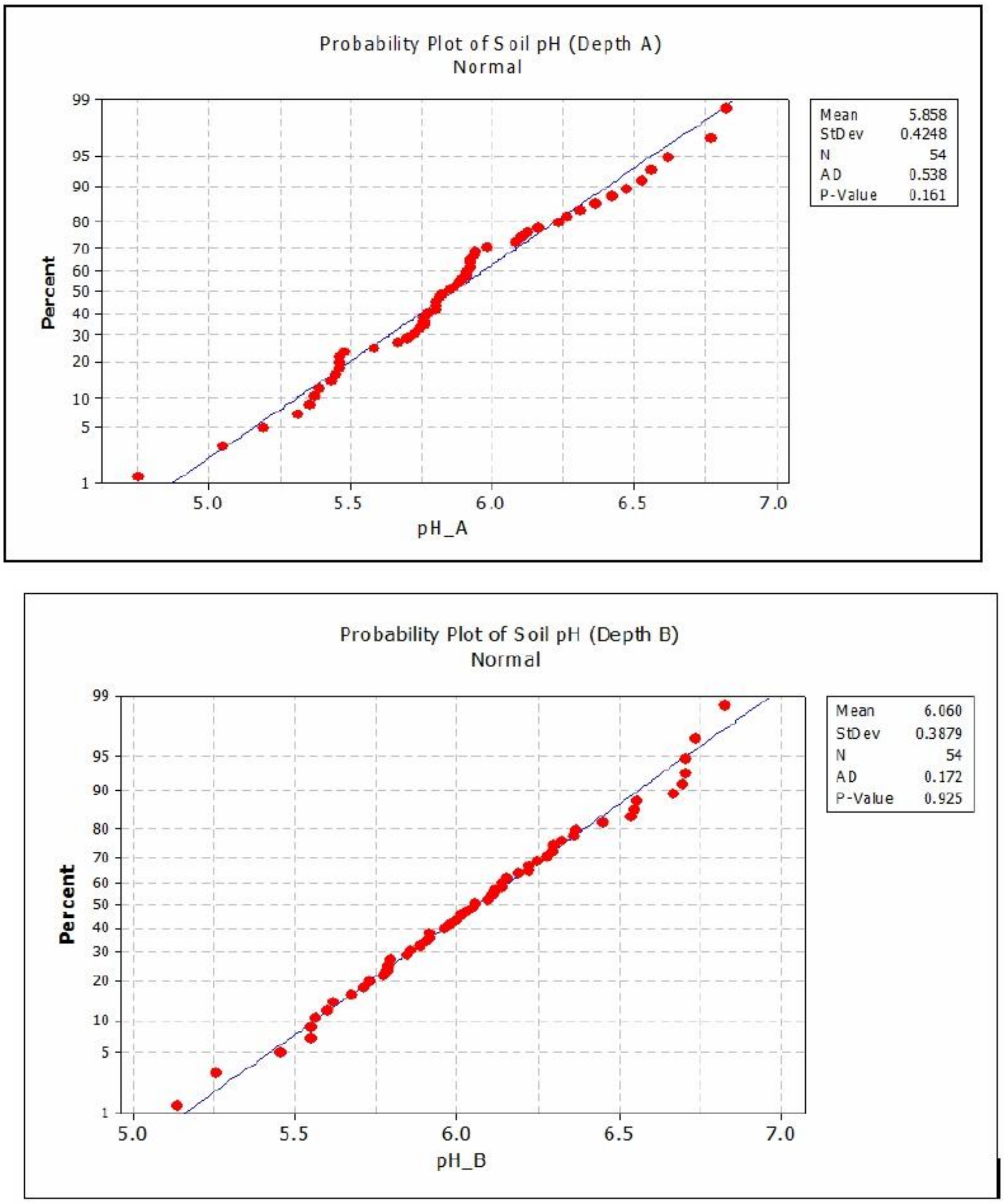


Figure 5.2: Anderson-Darling Normality Test on soil pH: (a) Depth A; (b) Depth B

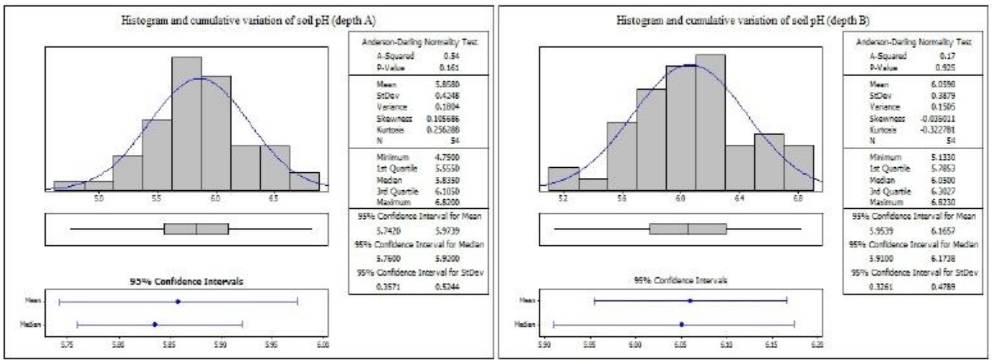


Figure 5.3: Soil pH graphical summary, histogram of pH with an overlaid normal curve, boxplot, 95% confidence intervals for mean, and 95% confidence intervals for the median: (a) depth A; (b) depth B

Trend analysis on soil samples showed that pH at both depths had a trend that can be attributed to the geographic characteristics of the study area (Figure 5.4). However trend was too weak in depth A whereas for depth B both North-South and East-West trend lines exhibited a clear curve. Experimental anisotropic semivariograms were calculated to identify the possible spatial structure of the soil pH. To determine the predictability of the theoretical model, prediction error statistics were calculated using cross validation test for all models and best models were identified (Table 5.2). The best model was selected based on four criteria: the standardized mean nearest zero, the smallest RMSE, the average standard error nearest the root-mean-squared prediction error and the standardized root-mean-squared prediction error nearest one. In this study, Exponential and Gaussian models were selected as the best fitted model using the cross validation test (Table 5.3). These models were fitted to experimental semivariograms for soil pH in depth A and B respectively (Fig 5.5a, b). The parameters of selected semivariograms models of soil pH are represented in Table 5.3

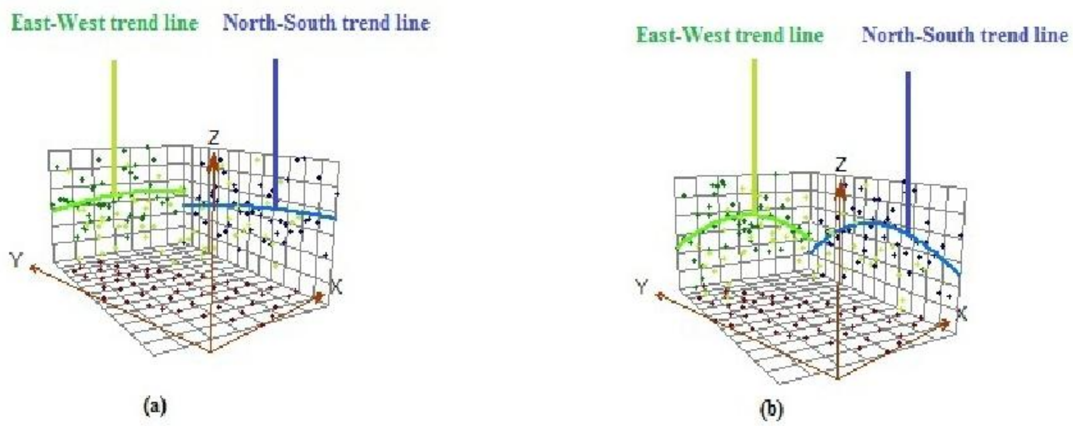


Figure 5.4: Trends of soil pH: (a) depth A; (b) depth B

Trangmar et al. (1985) and Goovaerts (1997) indicated that sampling intervals influence the semivariogram range. The variogram range indicates the horizontal extent of autocorrelation or spatial dependence of the data. The range of the variogram can be used to guide further soil sampling in existing fields or fields with a similar degree of soil spatial variation. According to the classification of Chien et al. (1997) by the nugget-to-sill ratio soil pH of both depths had a moderate spatial dependency (Table 5.3).

Table 5.2: Cross Validation Result

	Model	Mean	RMS	MS	RMSS	ASE
A	**Spherical**	0.013558	0.446315	0.025356	1.061026	0.417128
	Exponential	0.013765	0.455569	0.026825	1.036655	0.436969
	Gaussian	0.011454	0.438971	0.018373	1.135542	0.382375
	K-Bessel	0.013967	0.450624	0.026225	1.041815	0.429506
B	**Spherical**	0.006877	0.324916	0.018045	0.931229	0.351262
	Exponential	0.006697	0.33348	0.017283	0.931822	0.359504
	Gaussian	0.004754	0.319076	0.012831	0.957702	0.336557
	K-Bessel	0.004954	0.319041	0.012935	0.945738	0.340724

Table 5.3: The parameters of selected semivariograms models of soil pH variation.

Depth	Model	Nugget (Co)	Sill (Co+Cs)	Nugget/Sill (%)	Spatial dependency
A	Exponential	0.06	0.14	31.66	M
B	Gaussian	0.07	0.07	46.83	M

% nugget= (nugget semivariance/ total semivariance)*100, S strong spatial dependence (% nugget <25), M moderate spatial dependence (25<% nugget <75) and weak spatial dependence (% nugget >75)

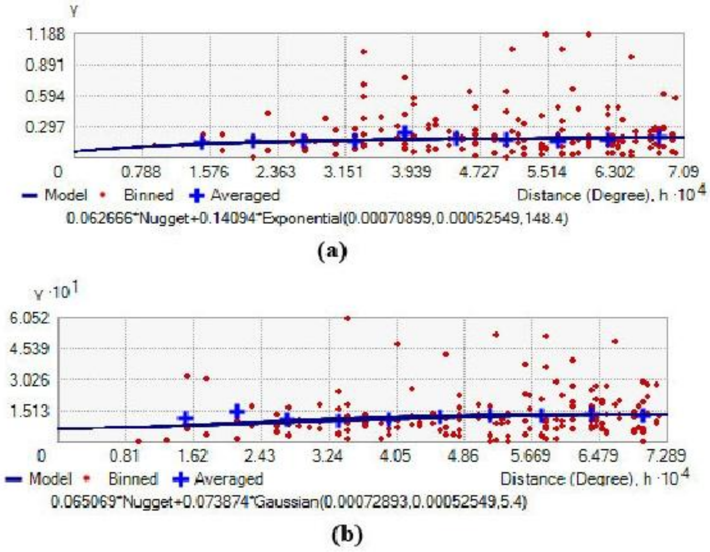


Figure 5.5: Best fitted semivariograms for soil pH: (a) depth Exponential; (b) depth B Gaussian

After modelling the variograms, three different techniques OK, RBF and IDW were used to predict the spatial distribution of soil pH. The Mean-Squared Error (MSE) and the Goodness of prediction (G %) were calculated as measures for accuracy and effectiveness of soil pH at both depths. The comparison of MSE values between the methods shows that the smallest errors are achieved by the RBF and OK methods in depth A and B respectively (Table 5.4).

The Goodness of prediction results shows positive values for sampled soil pH at both depths, signifying that the use of interpolation techniques was appropriate for mapping soil pH. Finally, maps of spatial distribution of studied soil pH were generated in GIS environment. The mapping result of IDW, RBF and OK are shown in Figure 5.7. There were different spatial patterns of pH (Figures 5.5-7) for both soil depths. Resulted maps of soil pH also showed different spatial variation with depth. Higher pH value was found at the south part of the study area for depth A and east and southeast parts for depth B.

Table 5.4: Result of Mean Error (ME) and Mean Square Error (MSE)

	Method	G	ME	MSE
Depth A	**RBF**	99.43	0.0095	0.4411
	IDW	99.43	0.0230	0.4420
	OK	99.45	0.0148	0.4468
Depth B	**RBF**	99.71	0.0039	0.3236
	IDW	99.67	0.0183	0.3464
	OK	99.72	0.0048	0.3191

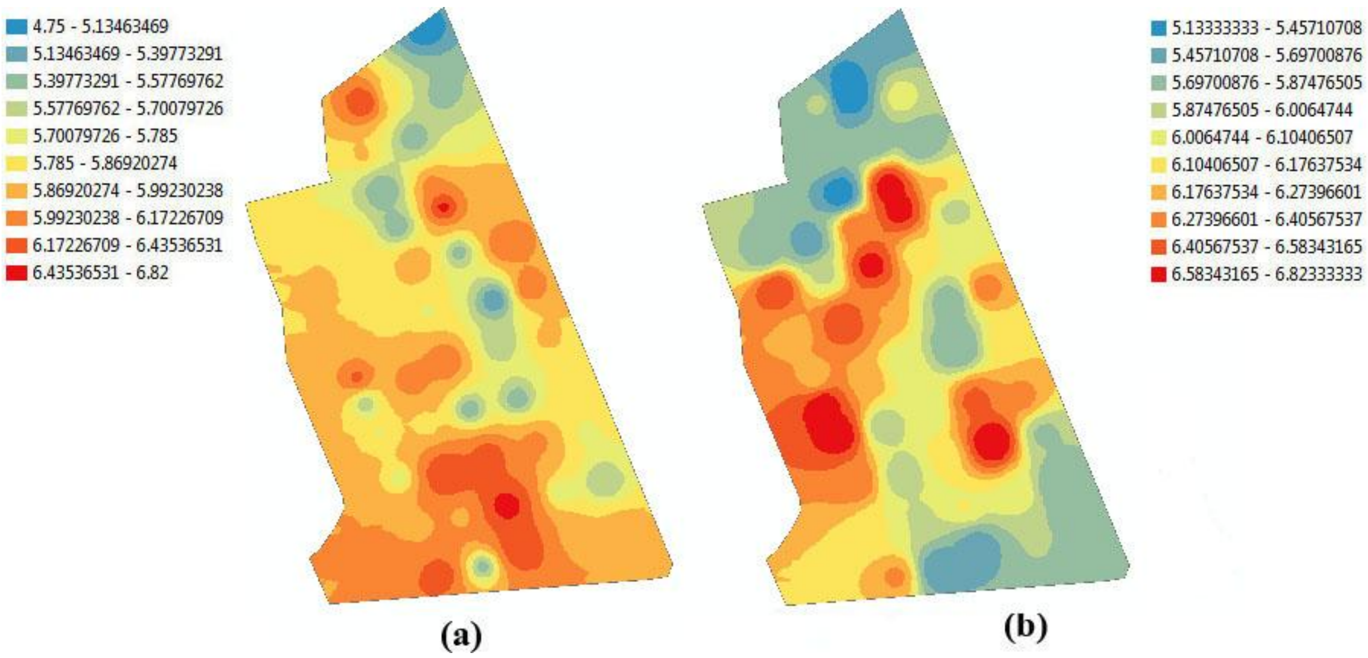


Figure 5.6: Prediction map of soil pH generated by IDW: (a) Depth A; (b) Depth B

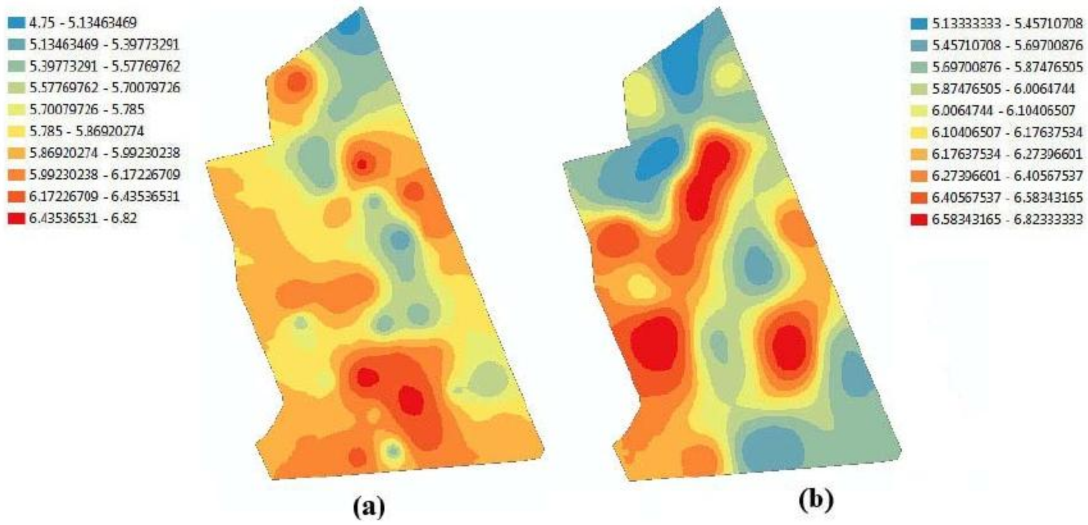


Figure 5.7: Prediction map of soil pH generated by RBF: (a) Depth A; (b) Depth B

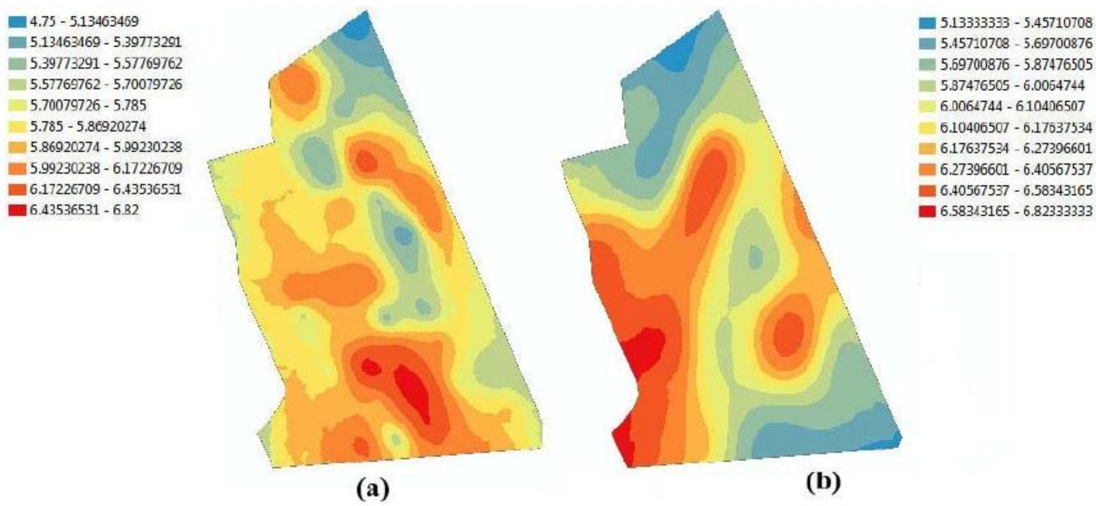


Figure 5.8: Prediction map of soil pH generated by OK: (a) Depth A; (b) Depth B

5.1.5 Discussions

This study evaluated the performance of three commonly used interpolation methods for soil pH. OK, RBF and IDW techniques were used to explore and assess the spatial variation of soil pH at two different depths within a vineyard. Results from both statistical approaches indicated that moderate spatial variability existed across the field for soil pH considered in this study. The results revealed that RBF was considered to be an accurate and adequate method for spatial interpolation and evaluation of soil properties for depth one. The result also showed that OK is suitable method for prediction and mapping the spatial distribution of soil pH for depth two. The results of the goodness of prediction measurements confirmed that OK was considered as the most effective method for prediction of unsampled soil pH values. Interpolation maps generated by a combination of geostatistical techniques and digital data in a GIS environment were accurate enough to improve the identification of soil properties such as pH, which is the first step for site-specific management. Comparison of generated soil pH maps of different depths provides an effective way to quantify soil map unit purity.

To improve predictions of sparse information from soil surveys it could be very useful to use more intensive, densely sampled data. spare data, may not have sufficient information to describe the spatial structure if it does exist, and therefore different methods may not perform dramatically different. Hence, it is crucial to make sure that the spatial sampling intensity is adequate to capture the possible spatial structure of data. Density of the samples is phenomenon and geography dependent. The selection of the interpolation method also needs to be made with care so that the interpolation method can fully utilize the spatial information captured by the samples.

This case study has demonstrated a typical application of GeoComputational methodology utilising specific geospatial data attributes for variable identification and algorithms for processing and depicting the results of the study. The blend of database and data processing software, together with the tolls for result visualisation is regarded here to have completeness in terms of the aim of the thesis, which is to demonstrate conformity of these methods for surface and field data interpolation as being consistent with what characterizes GeoComputation as a domain specific multi disciplinary area of research.

5.2 **Case study two**: Estimation of Actual Rainfall from Satellite Rainfall

5.2.1 Introduction

Rainfall is the major driving force of the hydrologic cycle, thus dictating the evolution of key land surface states that determine the critical fluxes. Satellite-based rainfall monitoring provides a method of producing rainfall estimates for the entire continent, without the need for extensive real time surface observations. Satellite Rainfall (SR) estimates have their own limitations in terms of estimation accuracy and many operational SR products lack any estimate of their uncertainty. This leads to a situation in which the SR products generally accepted to come with significant errors, but they have no quantitative information about the distribution of these errors and are not able to take this uncertainties into account at the decision making stage. This fact has led to a series of validation studies aimed at quantifying the errors in SR estimations. A possible solution to this major problem is to construct a model that characterizes the conditional distribution of Actual Rainfall (AR) rate given SR estimates. In this case study SR estimates were validated against actual rain gauge data in the Auckland region where there is a dense network of gauges available.

5.2.2 Materials and Methods

Pairs of AR and SR data are needed to quantify the accuracy of SR product. In this study, it was assumed that averages of high-resolution ground-based radar rainfall estimates are sufficiently accurate approximations of AR. The commonly used method to examine the effect of infrequent satellite overpasses is to take the averages of the ground-based radar rainfall estimates as AR, and to simulate SR as the average of data sampled from a high-resolution rainfall field according to a specified satellite sampling pattern (e.g., (Oki and Sumi, 1994), (Steiner, 1996), (Steiner et al., 2003) and (Gebremichael and Krajewski, 2004)). As a result, the SR and AR values represent monthly rainfall averaged over the Auckland region. The SR values represent estimates from a satellite that samples the hourly rainfall over a given area.

5.2.2.1 Actual Rain Data (AR) Collection

Actual Rain gauge data has been obtained from the National Institute of Water and Atmospheric Research (NIWA) database. Average daily rainfall data was gathered form thirty eight stations throughout Auckland City, New Zealand. The temporal resolution of the provided rain gauge data was daily and in measuring the rain gauges with recorded data from two seasons. It was decided to use data collected in two different months of the year 2013. February 2013 was chosen as dry season and May 2013 was selected as wet season. To avoid any inconsistency or missing data during these time periods, it was ensured that all thirty eight stations were operating and collecting data.

5.2.2.2 Satellite Rain data (SR) Collection using Hydro-Estimator Digital Global Data

Satellite estimation of rainfall was obtained from Hydro-Estimator provided by the National Oceanic and Atmospheric Administration (NOAA) website (NOAA, 2013).

Satellite estimation of rainfall value was stored on a latitude/longitude grid with 8000 column and 311 rows. The corresponding latitude/longitude points were stored on a separate world coordination file. Each line in coordination file contained the latitude and longitude, respectively of the corresponding data point in the SR data file. The first line of the world coordination file contained the latitude and longitude of the northwest corner of the first data point in SR data files. Oddly, positive values in the coordination file signified degrees north (latitude) and west (longitude) which was counter to the usual convention. In this study, negative longitude values in the range of -167 to -178 and latitude values of -36 to -37 were selected for New Zealand, Auckland region. The longitude values were later corrected for map projection to show the spatial distribution of these SR points.

SR data files contain 1-hour Hydro-Estimator rainfall accumulations holding values ranging from 0 to 256. These values were converted to rainfall accumulation in millimetre (mm) using bellow equation (STAR Satellite Rainfall Estimates, 2013):

$$R = (value-2) * 0.30 \quad (2)$$

Where a value of 0 indicates missing rainfall and a value of 2 means no rainfall.

Due to large size of both hourly rain data files and the world coordination file the process of extracting the hourly rainfall estimation amounts was time consuming. There were also other issues regarding the data files provided by NOAA's website which made the process of data collection even slower. First, some data files were not available due access permission problems. Secondly, after extracting SR values of the desired region and converting these values into (mm) using provided equation, it mostly resulted in negative rain values. The author has communicated several times with one of the meteorologist in Centre for Satellite Applications and Research (STAR) to obtain corrected data. It appeared that there was some error in some of the past rainfall estimates that was producing the negative value which should never occur. This issue was resolved later and the whole process of downloading files and data extraction and conversion was performed again.

Figure 5.9 illustrates the study region and spatial distribution of the AR and SR locations used in this study.

Figure 5.9: Spatial Distribution of AR and SR in Auckland Region

5.2.2.2.1 Hydro-Estimator (H-E): An Overview

Satellite-based estimates of rainfall have been used operationally at NOAA / ESDIS since the late 1970's, starting with the largely manual Interactive Flash Flood Analyzer (IFFA; Scofield 1987), and then progressing to the fully automated Auto-Estimator (Vicente et al. 1998). The original Auto-Estimator (A-E) algorithm was developed for deep, moist convective systems. Over time, enhancements and improvements to the program led to a completely new product, called the Hydro Estimator (H-E).

The H-E (Scofield and Kuligowski 2003) is the current-generation operational algorithm at NESDIS and has been used since 2002. H-E is currently used by the National Weather Service for monitoring potential flash flood events. Precipitation rates are primarily based on the cloud top temperature obtained from GOES 12 and GOES 10 (10.7 micron) (CONUS, 2013). Estimates of rainfall from satellites can provide critical rainfall information in regions where data from gauges or radar are unavailable or unreliable, such as over oceans or sparsely populated regions.

The H-E estimates rainfall rates by using infrared data from NOAA's Geostationary Operational Environmental Satellites (GOES). Estimates are calculated every 15 minutes for the continental United State. Estimates for the rest of the world are produced using available geostationary data over Europe, Africa, and western Asia (METEOSAT), and eastern Asia (MTSAT) (STAR Satellite Rainfall Estimates, 2013).

According to Scofield and Kuligowski (2003) the hydro estimation starts by determining a pixel of interest temperature compared to its surroundings using the variable Z.

$$Z = -(T - \mu_T)/\partial_T \qquad (3)$$

where T is the pixel brightness temperature,

μ_T is the mean value of T in the surrounding cloud,

∂_T is the standard deviation of T in the same region.

It is assumed that cloud's pixels with a positive value of Z are above updrafts and are colder than their surroundings. These pixels have non zero rainfall rates and other pixels are assumed not to be producing rainfall.

Once raining pixels have been identified, rainfall rate is based on a number of factors, including T (higher rain rates for lower values of T, based on an exponential curve fit), Numerical Weather Prediction (NWP) model perceptible water PW (higher rain rates for higher values of PW), and Z (higher rain rates for higher values of Z). The value of T is adjusted downward in regions where the NWP-derived convective equilibrium temperature is above 213 K, in regions where T values significantly below 213 K would not be expected from thermodynamic considerations. Such regions can still contain strong updrafts and heavy rainfall, but will not exhibit the very cold cloud signatures typically associated with such rainfall.

Adjustments to these rain rate estimates are then made based on NWP-derived mean surface-to-700 hPa relative humidity (reduction of rainfall rates in dry regions), and on updrafts or downdrafts induced by topography (enhancement in regions where the horizontal wind incident on surface topography produces upward vertical motion; reduction in regions where downward motion results). This algorithm has shown some skill at fine-scale rainfall estimation. The highest resolution of this data is quarter-hourly at a 4km spatial resolution between latitude 60N and 60S (Scofield and Kuligowski 2003).

5.2.3 Statistical Analysis

The aim of this study is to calculate the probability distribution of all the possible AR values using the value of SR, assuming *Y* as SR estimation and *X* as AR estimation. The probability distribution is calculated by an estimate of $f(y|x)$, the conditional density at $Y = y$ given $X = x$, for a range of possible x values, X. The conditional density estimation used in this study is based on a non-parametric Kernel estimator.

Figure 5.10 illustrates histogram and Kernel density for AR and SR estimation plus the mean and standard deviation measurements for Auckland region in February and May 2013. An evident discrepancy between AR and SR estimation is depicted in Figure 5.10: Histogram and Kernel Density of AR and SR Estimation for Auckland Region in February and May 2013.The means for both AR and SR estimation is far apart while the standard deviation of AR is much smaller than SR estimation. This results in a much wider histogram for SR estimation.

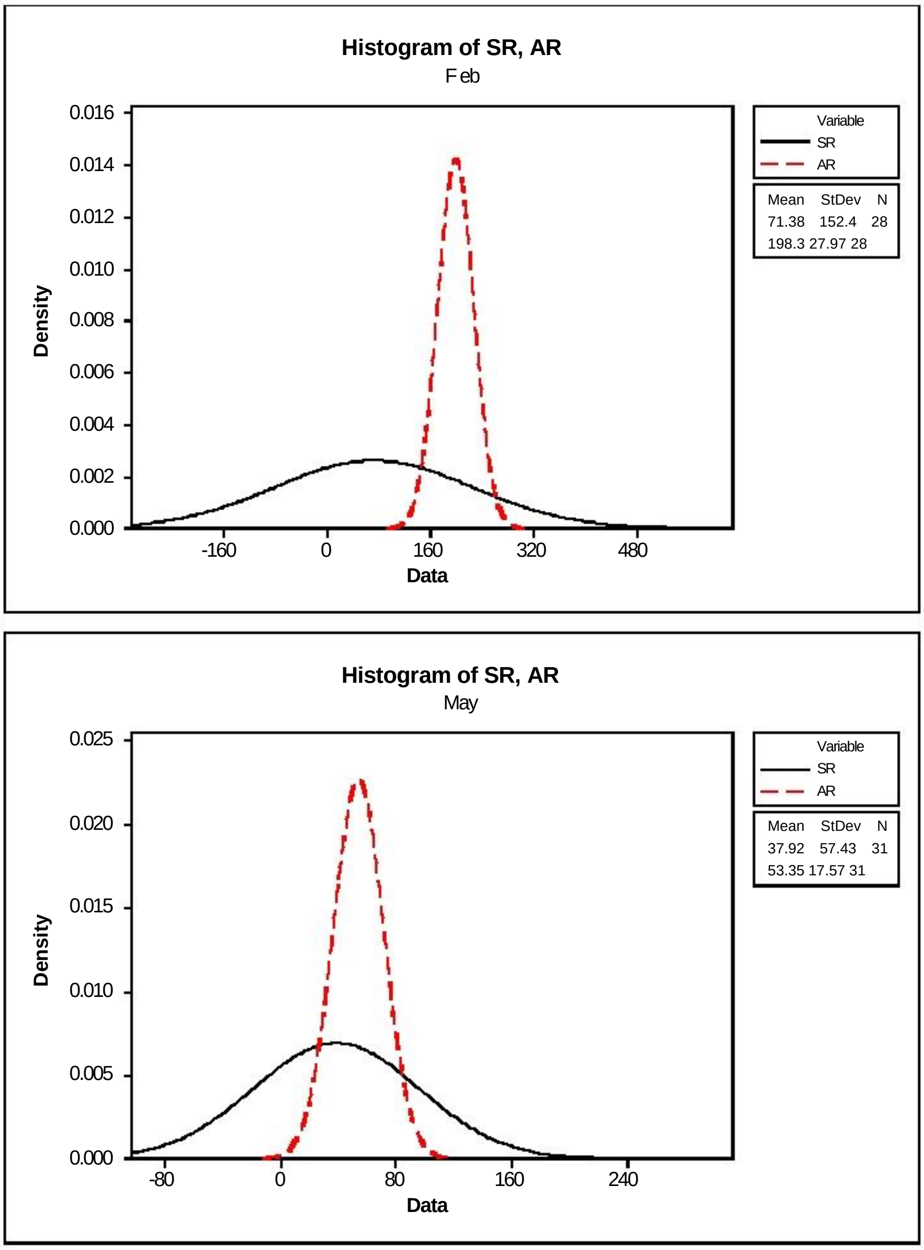


Figure 5.10: Histogram and Kernel Density of AR and SR Estimation for Auckland Region in February and May 2013.

Table 5.5: Regression Analysis of AR and SR for February 2013

The Regression Equation Result				
AR = 52.2 + 0.0298 SR				
Predictor	**Coef**	**SE Coef**	**T**	**P**
Constant	52.214	3.848	13.57	0.000
SR	0.02982	0.05655	0.53	**0.602**
S = 17.7895 **R-Sq = 0.9%** R-Sq(adj) = 0.0%				

Table 5.6: Variance Analysis of AR and SR for February 2013

Analysis of Variance						
Source	DF	SS	MF	F	P	
Regression	1	88.0	88.0	0.28	0.602	
Residual Error	29	9177.5	316.5			
Total	30	9265.5				
Unusual Observations						
Obs	SR	AR	Fit	SE Fit	St	Resid
1	0	93.00	52.21	3.85	40.79	2.35R
30	275	36.20	60.41	13.77	-24.21	-2.15RX

Table 5.7: Analysis of AR and SR for February and May 2013

The Regression Equation Result				
AR = 203 - 0.213 SR				
Predictor	Coef	SE Coef	T	P
Constant	202.74	5.565	36.43	0.000
SR	-0.2135	0.1125	-1.90	0.069
S = 26.7152 R-Sq = 12.2% R-Sq(adj) = 8.8%				

Table 5.8: Variance Analysis of AR and SR for May 2013

Analysis of Variance					
Source	DF	SS	MF	F	P
Regression	1	2570.5	2570.5	3.60	0.069
Residual Error	26	18556.3	713.7		
Total	27	21126.7			

Unusual Observations						
Obs	SR	AR	Fit	SE Fit	St	Resid
4	158	104.100	169.09	16.20	-64.99	-3.06RX
9	174	220.50	165.57	17.97	54.93	2.78RX

5.2.4 Results and Discussion

Conditional distribution of AR given SR estimation was modelled using a non-parametric Kernel estimator and a regression method. Regression analysis was performed on data collected in February and May to see whether the SR estimation explains a significant amount of variance in AR. This will indicate the accuracy of using SR estimation as an alternative for AR in locations where there is no AR available (e.g. middle of Ocean or direst). The p-values calculated in the analysis tables (see Table 5.6: Variance Analysis of AR and SR for February 2013Table 5.8: Variance Analysis of AR and SR for May 2013) of variance indicate that the relationship between AR and SR estimation is not statistically significant at a-level of 0.05. This is also shown by the p-value for the estimated coefficient of SR, which is 0.602 in May and 0.069 in February.

The value of coefficient of determination (R^2) shown in Table 5.5: Regression Analysis of AR and SR for February 2013 Table 5.7: Analysis of AR and SR for February and May 2013 indicate that SR estimation explains only 0.9 and 12.2% % of the variance in AR for May and February respectively. This means that the current model does not fit the data well. Unusual observations for day 1 and 30 in May and day 4 and 9 of February were identified as shown in Table 5.6: Variance Analysis of AR and SR for February 2013Table 5.8: Variance Analysis of AR and SR for May 2013. This could indicate that these observations are outlier due to large standardized residual.

Although the results and analysis of data in this case study indicates that the SR estimation is far from being accurate, it does not undermine the objective of this study. Further study is required to evaluate the accuracy of SR estimation obtained from various estimators.

Chapter 6

6. Conclusions

GeoComputation is a fresh field of endeavor yet reflects a broad spectrum of research methods, which are applied appropriately to the sub fields of this problem domain. Individual GeoComputational methods become accepted as an effective solution to particular spatial analysis problem and they start to emerge in more common software with extensive usage such as remote sensing data analysis, geo-simulation techniques and visualization tools. System usability and some other aspects of GeoComputation where conceptual design and human opinion and other qualitative data are quantified may best be described using behavioural research methods. In short, GeoComputation is essentially similar to other areas of research activity in Computer Science and Information Systems.

GeoComputation is a cutting age research within the field of GIS and geospatial analysis and it is strongly influenced by latest programming development, data processing and interface design. Most important part of GIS is the data. A GIS thus consists of an extensive database of geographic information involving both positional data about land features and descriptive data and sets of programmes of applications, which enable the data to be input, assessed, manipulated, analyzed and reported. A direct measurement captures the primary geographic data sources especially for GIS. Secondary sources are obtained from other systems or reused from earlier studies. Data captures accounts for over half of the total cost of a GIS project. Direct input of digital data into a geodatabase minimizes the time consumption and the possible errors. However; close coupling of data collection devices and GIS database is not always possible. Raster and vector data capture can be used to overcome this issue.

Remote sensing is most popular real time method for gathering and displaying the information about the features in raster format. Remote sensing data provides synoptic viewing, data comparability, repeat coverage and historical record logging also helpful for generation of thematic maps. Remote sensing is a valuable source of input for spatial databases. Primary vector data capture is major source of geographic data gathering using surveying and GPS. Ground survey was found a very time consuming and expensive practice but still the best way to obtain highly accurate point locations.

The spatial interpolation methods have been developed for and applied to various disciplines. Many sampling factors and data characteristics affect the estimations of the methods but there are no consistent findings about how these factors affect the performance of the spatial interpolators. This makes it hard to select a fitting spatial interpolation method for a given input dataset.

This thesis provided a review of three spatial interpolation (IDW, RBF and Kriging) methods and evaluates them by comparing their performance on environmental data.

IDW interpolation uses the proximity factor adopted by Thiessen polygons and the slow changes of a trend surface. The assumption is that objects that are close to one another are more similar than those that are farther apart. In overall, IDW is a fast and simple method. There are few decisions to make regarding model parameters. IDW gives reasonable results for most types of data. The knowledge about the underlying surface is needed when choosing the weighting power as it may misrepresent the data as too smooth or too uneven. Uneven distributions of observational data points can affect the IDW interpolations result since an equal weight will be assigned to each of the points even if it is in a cluster. According to (Lam, 1983) IDW interpolations usually have a bird-eyes prototype around solitary data points that have values that differ greatly from their neighbourhoods, however this can be modified to a certain extent by altering the search criteria for the data points to account for anisotropy.

RBF is a special class of such as Gaussian or elliptical function. RBF response changes monotonically with distance from a central point. These radial functions may be fitted to points and merge together to minimize curvature. The interpolating values can be quickly calculated since RBF use relatively few points at a time. RBF are analytic, flexible, and test data for smooth surfaces and predictions are very close to the values being interpolated. Unlike IDW method, the RBF can predict values above the maximum and below the minimum measured values in the cross-section of the data points. Mathematical derivatives can easily be calculated for direct analysis of surface geometry and topology using the smoothness of RBF. However, RBF is poor for data sets which exhibit abrupt changes in small distance. According to (Burrough & Mcdonnell, 1998) there are no direct estimates of the errors associated with RBF interpolation, though these may be obtained by a recursive technique known as 'jack-knifing'. Thin-plate spline functions may provide a misleading result in the estimation of attribute values for numerical models.

Ordinary kriging is by far the most common type of kriging in practice. It uses dimensionless points to estimate other dimensionless points (i.e. wind speed surface in this project). In ordinary kriging, the regionalized variable is assumed to be stationary and no drift exists. This assumption allows for an estimate of an unknown value to be calculated using a weighted average of the known values or control points.

Kriging provides an estimate of the error of the unknown points, an asset not provided by other interpolation procedures. In other hand, a large set of simultaneous equations must be solved for every unknown grid estimated by kriging. Therefore, computer run times will be significantly longer if a map is produced by kriging rather than by another method. In addition, an extensive prior study of the data must be made to test for stationarity, determining the form of the semivariogram, setting the neighbourhood size and selecting the proper order of the drift if it exists.

In first case study, OK, RBF and IDW techniques were used to explore and asses the spatial variation of soil pH at two different depths within a vineyard. The results revealed that RBF was considered to be an accurate and adequate method for spatial interpolation and evaluation of soil properties for depth one. OK method was found suitable method for prediction and mapping the spatial distribution of soil pH for depth two. Results from both statistical approaches indicated that moderate spatial variability existed across the field for soil pH considered in this study. The results of the goodness of prediction measurements confirmed that OK method was considered as the most effective method for prediction of unsampled soil pH values. Interpolation maps generated by a combination of geostatistical techniques and digital data in a GIS environment were accurate enough to improve the identification of soil properties such as pH, which is the first step for site-specific management. Comparison of generated soil pH maps of different depths provides an effective way to quantify soil map unit purity.

Spatial interpolation methods have been applied to many disciplines. There has been substantial development over the past decade from the accuracy, multivariate frameworks and robustness points of view. However, many factors affect the performance of the methods and there are no consistent findings about their effects. As Burrough (1986) concluded: '*It is unwise to throw one's data into the first available interpolation technique without carefully considering how the results will be affected by the assumptions inherent in the method. A good GIS should include a range of interpolation techniques that allow the user to choose the most appropriate method for the job at hand*.'

In second case study, the conditional distribution of AR given measures from SR products was modelled. The model was applied for monthly rainfall averaged over an area in the Auckland City, New Zealand. The results and analysis of data in this case study indicate that the SR estimation is far from being accurate, which could be due to the data resolution and missing satellite data. Further study is required to study AR and SR data in other areas and seasons. Introducing a paradigm shift towards probabilistic estimations instead of the typical deterministic ones can be more beneficial to the users of the rainfall products in decision making process.

References

Abrahart, R. J. (2001). Why Are Computational Processes Important? Proceedings Of The 6th International Conference On Geocomputation. Brisbane, Australia.

Akinci, H. (2004). Geospatial Web Services For E-Municipality. Paper Presented At The Xxth ISPRS Congress, , 12-23 July , Istanbul, Turkey.

Aldridge, C. H. (2001). A Rough Set Based Methodology For Geographic. Proceed ings Of The 6th International Conference On Geocomputation. Brisbane, Australia.

Anderson Darling Test. (N.D.). Retrieved 07 11, 2011, From NIST/SEMATECH E-Handbook Of Statistical Methods: Http://Www.Itl.Nist.Gov/Div898/Handbook/

Anderson, G., & Moreno, S. (2003). Building Web-Based Spatial Information Soluions Around Open Specifications And Open Source Software. Transactions In GIS , 447-466.

Anselin, L. (1992). Spatial Data Analysis With Gis: An Introduction To Application In The Social Sciences. Santa Barbara: National Center For Geographic Information And Analysis.

Anselin, Luc And Arthur Getis (1992). Spatial Statistical Analysis And Geographic Information Systems, The Annals Ofregional Science 26, 19-33.

Anselin, Luc, Rustin Dodson And Sheri Hudak (1992). Linking GIS And Spatial Data Analysis In Practice. Paper Presented At The 32nd European Congress Of The Regional Science Association International, Brussels, Belgium.

Arcgis. (2012). Raster Data Organization. Esri.

Arctur, D., & Zeiler, M. (2004). Designing Geodatabases: Case Studies In GIS Data Modeling. ESRI Press.

Atkinson, P., Foody, G., Darby, S., & Wu, F. (2004). Geodynamics. Boca Raton: CRC Press.

AUT. (2013, 10 11). AUT. Retrieved 10 11, 2013, From Auckland University Of Technology: Http://Www.Aut.Ac.Nz/

Baalousha H (2010) Assessment Of A Groundwater Quality Monitoring Network Using Vulnerability Mapping And Geostatistics: A Case Study From Heretaunga Plains, New Zealand. Agric Water Manage 97:240-246. Doi:10.1016/J.Agwat.2009.09.013

Bailey, Trevor (1992). Statistical Analysis And Geographic Information Systems: A Review Of The Potential And Progress In The State Of The Art. EGIS 92, Proceedings Of The Third European Conference On Geographic Information Systems, Pp. 186-203 (Utrwht, EGIS Foundation).

Bansal, V. K. (2011). Use Of GIS And Topology In The Identification And Resolution Of Space Conflicts. Ascelibrary.Org , 159-171.

Bansal, V. K. (2011). Use Of GIS And Topology In The Identification And Resolution Of Space Conflicts. Journal Of Computing In Civil Engineering , 159-171.

Barringer, J. R. (1997). Meso-Scale Mapping Of Soil Temperatures In The. Presented At The Second Annual Conference Of Geocomputation. University Of Otago, New Zealand.

Barringer, J. R., & Lilburne, L. (1997). An Evaluation Of Digital Elevation Models For Upgrading New Zealand Land Resource Inventory Slope Data. Second Annual Conference Of Geocomputation. New Zealand: University Of Otago, New Zealand.

Barros, J. (2003). Simulating Urban Dynamics In Latin American Cities. Proceedings Of The 7th International Conference On Geocomputation. United Kingdom.

Bédard, Y. (1999). Principles Of Spatial Database Analysis And Design. John Wiley & Sons, Inc.

Bédard, Y., Merrett, T., & Han, J. (2001). Fundamentals Of Spatial Data Warehousing For Geographic Knowledge Discovery. London: Taylor & Francis.

Bennett, B. (96). The Application Of Qualitative Spatial Reasoning To GIS . Extended Abstracts From The 1st International Conference On Geocomputation. UK.

Berger, H., & Shine, J. (1999). Estimating Subpixel Geospatial Features. Proceedings Of The 4th International Conference On Geocomputation. Virginia,USA.

Bernhardse, T. (2002). Geographic Information Systems: An Introduction. In T. Bernhardse, Basic Data Models (P. 73). New York: John Wiley & Sons.

Bohling, G. (2005, 10 17). Introduction To Geostatistics And Variogram Analysis. Retrieved 03 01, 2013, From GIS Myanmar Community: Http://Gismyanmar.Org/Geofocus/Wp-Content/Uploads/2013/01/Variograms.Pdf

Bone, C., & Dragicevic, S. (2005). Sensitivity Of A Fuzzy-Constrained Cellular Automata Model Of. Proceedings Of The 8th International Conference On Geo-computation. United States Of America.

Borah, R., Puotinen, M., & Price, O. (2009). Line Intersect Sampling Tool (LIST):A GIS-Based Tool For Spread Analysis. Proceedings Of The 10th International Conference On Geocomputation. Australia.

Brodaric, B. G. (2000). Geocomputing With Geological Field Data: Is There A Ghost In The Machine? Proceedings Of The 5th International Conference On Geocomputation. Chatham.

Buhmann, M. (2003). Radial Based Functions:Theory And Implemenation. Cambridge : Cambridge University Press.

Burgess TM, Webster R (1980) Optimal Interpolation And Isorithmic Mapping Of Soil Properties. I.

Burrough P A & R A Mcdonnell, 1998. Principles Of Geographic Information Systems (Oxford University Press) 356 Pp.

Burrough, P. (1986). Principles Of Geographical Information Systems For Land Resources Assessment. Oxford University Press.

Burrough, P. A., & Mcdonnell, R. A. (1998). Principles Of Geographical Information Systems. (P. 98). Oxford University Press.

Calkins, H. W. & Tomlinson. R.F. (1977). Geographic Information Systems: Methods And Equipment For Land Use Planning. International Geographic Union

Câmara, G. (Sept./Oct. 2001). Geocomputation Techniques For Spatial Analysis: Are They Relevant To Health Data? Cad. Saúde Pública Vol.17 No.5 Rio De Janeiro .

Cambardella CA, Moorman TB, Novak JM, Parkin TB, Karlen DL, Turco RF, Konopka AE (1994) Field-Scale Variability Of Soil Properties In Central Iowa Soils. Soil Sci Soc Am J 58:1501-1511

Cameron, C., & Trivedi, P. (2005). Density Estimation And Nonparametric Regression.

Cemek B, Gu¨Ler M, Kilic¸ K, Demir Y, Arslan H (2007) Assessment Of Spatial Variability In Some Soil Properties As Related To Soil Salinity And Alkalinity In Bafra Plain In Northern Turkey. Environ Monit Assess 124:223-234. Doi:10.1007/S10661-006-9220-Y

Cheng, T., & Anbaroglu, B. (2009). Methods On Defining Spatial-Temporal Neighborhoods.

Chien YJ, Lee DY, Guo HY, Houng KH (1997). Geostatistical Analysis Of Soil Properties Of Mid-West Taiwan Soils. Soil Sci., 162: 291-297. Cliffs, Prentice Hall: New Jersey

Collins Jr., Fred. C. (1995).A Comparison Of Spatial Interpolation Techniques In Tempera- ture Estimation". Doctoral Dissertation. Virginia Polytechnic Institute And State University, Blacksburg, VA.

Collins, F.C. , Bolstad, P.V. (1996). A Comparison Of Spatial Interpolation Techniques In Temperature Estimation, Proceedings, Third International Conference/Workshop On Integrating GIS And Environmental Modeling, Santa Fe, NM. Santa Barbara, CA: National Center For Geographic Information And Analysis, Santa Barbara.

Cömert, C. (2004). Web Services And National Spatial Data Infrastructure (NSDI). Proceedings Of Geo-Imagery Bridging Continents, Xxth ISPRS Congress. Istanbul, Turkey: Commission 4.

Congalton, R. G. & Green, K. (1995). The Abcs Of GIS: An Introduction To Geographic Information Systems. In Wetland And Environmental Applications Of GIS (Eds. J.G. Lyons & J. Mccarthy) Pp. 9-24. Lewis Publishers, Boca Baton, Florida.

Conley, J. F. (2009). An Evaluation Of The Extended Power Of Cluster. Proceedings Of The 10th International Conference On Geocomputation.

CONUS. (2013, 07 09). Retrieved 09 21, 2013, From NOAA Star: Http://Www.Star.Nesdis.Noaa.Gov/Smcd/Emb/Ff/CONUS.Php

Corter, J., & Guth, P. (2007). Unsupervised Classification Of Submarine Landslides. Proceedings Of The 9th International Conference On Geocomputation. National University Of Ireland, Maynooth.

Couclelis, H. Geocomputation In Context, In: P.A. Longley, S.M. Brooks, R. Mcdonnell, B. Macmillan (Eds.), Geocomputation: A Primer, John Wiley, Chichester, 1998, Pp. 17-29.

Cressie, N. (1985), "Fitting Variogram Models By Weighted Least Squares," Mathematical Geology, 17(5), 563-570.

Cressie, Noel. A. C. Statistics For Spatial Data. (1993) Wiley Series In Probability And Mathematical Statistics, Iowa State University. New York, NY.

Davis, J. C. (2002). Statistics And Data Analysis In Geology. John Wiley & Sons.

Deutsch CV (2002) Geostatistical Reservoir Modeling, 1st Edn. Oxford University Press, New York

Dobson, J.E (1983),"Automated Geography", The Professional Geographer 35, Pp.135-134

Dueker, K J. & Kjerne, D. (1979) Multipurpose Cadastre: Terms And Definitions. Falls Church VA, American Society For Photogrammetry And Remote Sensing And American Congress On Surveying And Mapping

Eastman, J. R. (2001). Introduction To Remote Sensing And Image Processing. Worcester: Clark University.

Ehlen, J., & Douglas, R. L. (2002). Geocomputation: What Is It? Elsevier Science Ltd.

Ehlen, J., Caldwell, D., & Harding, S. (2000). The Semantics Of Geocomputation. Proceedings Of The 5th International Conference On Geocomputation.

EMU. (2008). Eastern Mediterranean University. Retrieved 3 1, 2013, From Triangulated Irregular Network: Http://Www.Emu.Edu.Tr/

EPA. (2004). Developing Spatially Interpolated Surfaces And Estimating Uncertainty. Research Triangle Park.

Esri. (2008). Deterministic Methods For Spatial Interpolation. Retrieved August 12, 2012, From Esri:

Http://Help.Arcgis.Com/En/Arcgisdesktop/10.0/Help/Index.Html#//003100000023000000.Htm

Esri. (2008). Understanding Raster Data.

Esri. (2009). ESRI. Retrieved 02 14, 2013, From ESRI International User Conference: Http://Events.Esri.Com/Uc/Qanda/Index.Cfm?Fuseaction=Answer&Conferenceid=2a8e2713-1422-2418-7f20bb7c186b5b83&Questionid=2441

FGDC. (2010, 06 04). Federal Geographic Data Committee. Retrieved 05 12, 2012, From Federal Geographic Data Committee: Http://Www.Fgdc.Gov/Training/Nsdi-Training-Program/Materials/Geowebservices_Intro_20100604.Pdf

Fischer, Manfred M. And Peter Nijkamp (1992). Geographic Information Systems And Spatial Analysis, The Annals Of Regional Science 26, 3-17.

Fisher, M. and Jendorck, E. Chapter 1: Introduction to Web Services, Java Web Services Tutorial. Sum Microsystem, 2002.

Fryer, J. G., Chandler, J.H. & Cooper, M. A. R. (1994). On The Accuracy Of Heighting From Aerial Photographs And Maps: Implications To Process Models. Earth Surface Processes And Landforms 19:577-583.

Gahegan, M. (1999). What Is Geocomputation? Transactions In GIS, 3, 203-206.

Gardner, J.V., And Dartnell, P. (2001, 2000). Multibeam Sonar Survey Of Crater Lake, Oregon: Data, GIS, Images, And Movies: U.S. Geological Survey Digital Data Series DDS-72

GIS Cookbook. (2007, 10 25). Retrieved 07 11, 2013, From GIS Cookbook For Lgus: Http://Www.Cookbook.Hlurb.Gov.Ph/Book/Export/Html/203

GITTA. (2010, 06 18). Raster Data Compression. Retrieved 1 1, 2013, From Geographic Information Technology Training Alliance: Http://Www.Gitta.Info/Datacompress/En/Html/Rastercomp_Chain.Html

Goodchild, M. (1987). A Spatial Analytical Perspective On Geographicalinformation Systems. International Journal Of Geographical Information Science , 327-334.

Goodchild, M.F., R. Haining And S. Wise (1992). Integrating GIS And Spatial Data Analysis: Problems And Possibilities, International Journal Of Geographic Information Systems 6(5), pp. 407-423.

Goovaerts P (2000) Geostatistical Approaches For Incorporating Elevation Into The Spatial Interpolation Of Rainfall. J Hydrol 228:113-129

Goovaerts, P.(1997). Geostatistics For Natural Resources Evaluation. Oxford University Press, New York, 483 Pp.

Guan, Q., Zhang, T., & Clarke, K. C. (2006). Geocomputation In The Grid Computing Age. Springer Berlin / Heidelberg.

Harding, S. (2002). Geocomputation: What Is It? Elsevier Science Ltd.

Hecht, L., Jr. (2002). Web Services Are The Future Of Geoprocessing.

Hengl, T.(2007). A Practical Guide To Geostatistical Mapping Of Environmental Variables, Luxembourg: Office for Official Publication of The European Communities

Huang X, Skidmore EL, Tibke G (1999) Spatial Variability Of Soil Properties Along A Transect Of CRP And Continuously Cropped Land. In: 10th International Soil Conservation Organization Meeting Held, Purdue University And The USDA-ARS National Soil Erosion Research Laboratory, Pp 641-647

Ibrahim, D. (1996). GPS Based Public Transport Route Information System. Extended Abstracts From The 1st International Conference On Geocomputation. United Kingdom.

Isaaks, E. H. And R. M. Srivastava. (1989). An Introduction To Applied Geostatistics. Oxford Univ. Press, New York, Oxford.

Isaaks, E.H., Srivastava, R.M. (1989). Applied Geostatistics. Oxford University Press, New York, 561 Pp.

Jin Li & Andrew D. Heap. (2011). A Review Of Comparative Studies Of Spatial Interpolation Methods In Environmental Sciences: Performance And Impact Factors. Ecological Informatics 6: 228-241

Johnson, H. (2007). Health Atlas Ireland. Geocomputation 2007. Ireland,.

Jonathan Doughty, P. J. (1999). Interoperable Geospatial Objects. Proceedings Of The 4th International Conference On Geocomputation.

Journel, A.G. And Huijbregts, C.J., (1978). Mining Geostatistics. Academic Press, London, 600 Pp.

Karydas, C., Gitas, I., & Koutsogiannaki, E. L. (2009). Evaluation Of Spatial Interpolation Techniques For Mapping Agricultural Topsoilproperties In Crete. Earsel Eproc. , 8, 26-39.

Kingston, R. (2001). Medaction: An Internet Tool For Forecasting Land Use Change And Land Degradation In The Mediterranean Region. Proceedings Of The 6th International Conference On Geocomputation. Brisbane, Australia.

Krasilnikov P, Carre´ F, Montanarella L. (2008). Soil Geography And Geostatistics, Concepts And Applications. European Commission, Joint Research Centre, Institute For Environment And Sustainability, Luxembourg

Kravchenko A & D Bullock. (1999). A Comparative Study Of Interpolation Methods For Mapping Soil Properties. Agronomy Journal, 91: 393-400

Kravchenko A N. (2003) . Influence Of Spatial Structure On Accuracy Of Interpolation Methods. Soil Science Society Of America Journal, 67: 1564-1571

Krivoruchko K & C A Gotay. (2003) . Using Spatial Statistics In GIS. In: International Congress On Modelling And Simulation, Edited By D A Post, 713-736

Lam, N. (1983). Spatial Interpolation Methods: A Review. The American Cartographer. 10(2): 129-149.

LAND, N. (2003). The Role Of The National Mapping Agencies In Building. FIG Working Week, (P. PS2.1). Paris.

Leslie, M. (2009, 10 14). Introduction To Postgis. Retrieved 01 15, 2012, From Http://Revenant.Ca/Www/Postgis/Workshop/Introduction.Html

Li, J., & Heap, A. D. (2008). A Review Of Spatial Interpolation Methods For Environmental Scientists. Geoscience Australia, Record 2008/23.

Li, J., & Heap, A. D. (2011). A Review Of Comparative Studies Of Spatial Interpolation Methods In Environmental Sciences: Performance And Impact Factors. Ecological Informatics , 228-241.

Liu G, Yang X. (2008). Spatial Variability Analysis Of Soil Properties Within A Field. In: IFIP International Federation For Information Processing. Vol 259: Computer And Computing Technologies In Agriculture. Vol 2; Daoliang Li; Springer Boston Pp 1341-1344

Liu, X., Hu, J., & Ma, J. (2011). Quantitative Evaluation Of Spatial Interpolation Models Based On A Data-Independent Method. In D. Chen, Advances In Data, Methods, Models And Their Applications In Geoscience (Pp. 53-70). Croatia: Intech.

Longley, P. (2005). Geographic Data Modeling. In P. Longley, Geographic Information Systems And Science (P. 211). John Wiley & Sons.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2011). Geographic Information Systems & Sciences. United States Of America: John Wiley &Sons.

Lou, W., Taylor, M., & Parker, S. (N.D.). Spatial Interpolation For Wind Data In England And Wales. Retrieved 08 10, 2011, From Intermet: Http://Intermet.Csl.Gov.Uk/Wind.Pdf

Luo, J. A. (2003). Geocomputation: A Coordination-Oriented Approach. Proceedings Of The 7th International Conference On Geocomputation. UK.

Luo, W., Taylor, M., & Parker, S. (2013). Spatial Interpolation For Wind Data In England And Wales. UK: Intermet .

Lynch, S. D. (1998). Converting Point Estimates Of Daily Rainfall Onto A Rectangular Grid. Proceedings Of The 3rd International Conference On Geocomputation. United Kingdom.

Macmillan, B. (1998).Epilogue. In P.A.Longley, S.M.Brooks, R.Mcdonn Ell, & B.Macmillan (Eds.), Geocomputation, A Primer. Chichester: John Wiley And Sons.

Maguire, A. J., & Grisé, S. (2001). Data Models for Object-Component Technology. GITA 2001. California.

Maguire, D., Goodchild, M., & Rhind, D. (1991). Geographical Information Systems: Principles. Harlow, UK: Longman.

Marsaglia, G., & Marsaglia, J. (2004). Evaluating The Anderson-Darling Distribution. Journal Ofstatistical Software 9. Journal Ofstatistical Software 9.

Martin, D., & Wu, F. (1999). Empirical CA Simulation From A High Resolution Population Surface. Proceedings Of The 4th International Conference On Geocomputation. Virginia,USA.

Mcbratney AB, Webster R. (1983). How Many Observations Are Needed For Regional Estimation Of Soil Properties? Soil Sci 135(3):177-183

Mueller T G, N B Pusuluri, K K Mathias, P L Cornelius, R I Barnhisel & S A Shearer. (2004). Map Quality For Ordinary Kriging And Inverse Distance Weighted Interpolation. Soil Science Society Of America Journal, 68: 2042-2047

Nalder, I.A. And Wein, R.W. (1998). Spatial Interpolation Of Climatic Normals: Test Of A New Method In The Canadian Boreal Forest. Agricultural And Forest Meteorology, 92: 211-225.

Nara, A., & Torrens, P. M. (2007). Fractal Analysis Of Pedestrian Egress Behavior And Efficiency. Geocomputation Proceedings 2007,National University Of Ireland,. Maynooth, Eire.

Neumann, A., Freimark, H., & Wehrle, A. (2010). Geodata Structures And Data Models. Retrieved 02 04, 2013, From Institute Of Cartography: Https://Geodata.Ethz.Ch

Nikolova N, Vassilev S (2006). Mapping Precipitation Variability Using Different Interpolation Methods. Proceedings Of The Conference On Water Observation And Information System For Decision Support (BALWOIS).S May 2006, Ohrid, Macedonia, 25-29

Oosterom,P (1993). Reactive Data Structures For Geographic Information Systems. Oxford, Oxford University Press

Openshaw, S. & Abrahart, R. J. (2000). Preface. In S. Openshaw, & R. J. Abrahart (Eds.), Geocomputation. London: Taylor and Francis.

Openshaw, Stan (1990). Spatial Analysis And Geographical Information Systems: A Review of Progress and Possibilities. In H. Scholten And J. Stillwell (Eds.), Geographical Information Systems For Urban And Regional Planning, Pp. 153-63 (Dordrecht, Kluwer Academic).

Openshaw,S. & Abrahart, R . (1999). Research Agendas And Futures In Geocomputation (Pp. 379-381).

O'SULLIVAN, D. (1999). Exploring the Structure Of Space: Towards Geo-Computational Theory. Proceedings of the 4th International Conference On Geocomputation.

Otero, J. K. (2007). Development Of A Rational Design Approach (Lrfd Phi) For Drilled Shafts Considering Redundancy, Spatial Variability, And Testing Cost. UNIVERSITY OF FLORIDA.

Peralvo, M. (2004). Influence Of DEM Interpolation Methods In Drainage Analysis. GIS Hydro 04, University Of Texas, Austin, Texas.

Peter M. Atkinson, P. M., Jeganathan, C., & Dash, J. (2009). Proceedings Of The 10th International Conference on Geocomputation. Proceedings Of the 10th International Conference On Geocomputation. Sydney, Australia.

Rabi N, S. (2003). Geoestatistic In Geoinformatics For Managing Spatial Variability. Retrieved 09 09, 2011, From IASRI: Http://Www.Iasri.Res.In/Ebook/EB_SMAR/EBook_Pdf%20files/Manual%20IV/6-Geostatistics.Pdf

Redulla, C.A., J.L. Havlin, G.J. Kluitenberg, N. Zhang And M.D. Schrock. (1996). Variable Nitrogen Management For Improving Groundwater Quality. Proceedings Of The 3rd International Conference On Precision Agriculture, Jun. 23-26, Minneapolis, MN., Pp: 1101-1110.

Rees, P., & Turton, I. (1998). Investigation of The Effects Of Input Uncertainty On Population Forecasting. Proceedings of The 3rd International Conference On Geocomputation. UK.

Ress, P. And Turton, I. (1998) "Geocomputation. Solving Geographical Problems With Computing Power", Environment and Planning A, 30, Pp.1835-1838

Ripley, B.D. (1981). Spatial Statistics. John Wiley & Sons, New York, 252 Pp.

Ryden, K. (Ed.): Opengis Implementation Specification for Geographic Information - Simple Feature Access - Part 1: Common Architecture. Open Geospatial Consortium. Http://Www.Opengeospatial.Org/Standards/Sfa (2013) (Accessed 17 January 2013)

Safe Software. (2013). Retrieved 05 27, 2013, From Safe Software: Http://Www.Safe.Com/

Saffari, M., Yasrebi, F., Sarikhani, R., & Gazni, M. (2009). Evaluation Of Artificial Neural Network Models For Prediction Of Spatial Variability Of Some Soilchemical Properties. Research Journal Of Biological Sciences. , 4.

Sahoo, P. M. (2012). Data Models In GIS. In P. M. Sahoo, Basics of GIS And GPS. New Delhi : Indian Agricultural Statistics Research Institute.

Sallis, PJ And Benwell, GL (1993) Geomatics: The Influence Of Informatics On Spatial Information Processing, IN Barta,BZ, Hung,SL And Cox, KR, Software Engineering Education (A-40), Elsevier Science B.V North-Holland, Pp 199-208

Scannavino Jr. F., Ghobakhlou, A., Perez-Kuroki, A., And Sallis, P., And Cruvinel P.(2011). Spatial Variability Of Soil Ph: A Case Study In Vineyards. In Proceedings Of IEEE Workshop On Instrumentation And Measurement Society New Zealand Chapter Workshop on Smart Sensors, Measurement And Instrumentation (Abstract), Auckland University of Technology (AUT), New Zealand, 10-11 March 2011. P 24.

Schloeder, C.A., Zimmerman, N.E. And Jacobs, M.J. (2001). Comparison Of Methods For Interpolating Soil Properties Using Limited Data. Soil Science Society Of American Journal, 65: 470-479

Scofield, R. A., And R. J. Kuligowski (2003). Status And Outlook Of Operational Satellite Precipitation Algorithms for Extreme-Precipitation Events. Mon. Wea. Rev., 18, 1037-1051.

Sivertun, A. (1993). Geographical Information Systems (GIS) As A Tool For Analysis And Communication Of Multidimensional Data (Geographical Reports). Dept. Of Geography, University Of Umea.

Smith, M. J., Goodchild, M. F., & Longley, P. A. (2007). Geospatial Analysis. Leicester: Matador.

Soil Ph And Nutrient Availability. (2011, 06 14). Retrieved September 5, 2011, From Extension : Http://Www.Extension.Org/Pages/9875/Soil-Ph-And-Nutrient-Availability

Stahl, K., Moore, R.D., Floyer, J.A., Asplin, M.G. And Mckendry, I.G., 2006. Comparison of Approaches for Spatial Interpolation of Daily Air Temperature in a Large Region with Complex Topography and highly variable station density

STAR Satellite Rainfall Estimates. (2013, 07 09). Retrieved 09 01, 2013, From NOAA Star: Http://Www.Star.Nesdis.Noaa.Gov/Smcd/Emb/Ff/Hydroest.Php

Star, J. & Estes, J. (1990). Geographic Information System: An Introduction, Englewoods

TOBLER, W.R. (1970). A Computer Movie Simulating Urban Growth In The Detroit Region. Economic Geography, 46, 234-240.

Topology And Topological Rules. (2008). Retrieved 1 10, 2013, From The School Of Natural Resources: Http://Www.Srnr.Arizona.Edu/Rnr/Rnr417/Topology.Pdf

Trangmar BB, Yost RS, Uehara G (1985) Application Of Geostatistics To Spatial Studies Of Soil Properties. Adv Agron 38:45-94

Turton, I. (2009). Experiences Teaching GIS With Open Source Software. Proceedings Of The 10th International Conference On Geocomputation. Sydney, Australia.

Tutorial, August 2002, Available From Http://Java.Sun.Com/Webservices/Docs/1.0/Tutorial/

UW. (2010, 09 30). The GIS Spatial Data Model. Retrieved 05 27, 2012, From Http://Courses.Washington.Edu/Gis250/Lessons/Introduction_Gis/Spatial_Data_Model.Html#Raster

Vieira SR, Hatfield JL, Nielsen DR, Biggar JW (1983) Geostatistical Theory And Application To Some Agronomic Properties. Hilgardia 51:1-75

Volter, M., Kricher, M., & Zdun, U. (2005). Remoting Patterns: Fundamental Of Enterprise, Internet And Realtime Distributed Object Middleware . New Jersey, USA: John Wiley and Sons, Inc.

Volter, M., Kricher, M., & Zdun, U. (2005). Remoting Patterns: Fundamental Of Enterprise, Internet And Realtime Distributed Object Middleware . New Jersey, USA: John Wiley and Sons, Inc.

Webster, R. (1985) Quantitative Spatial Analysis Of Soil In The Field. Adv Soil Sci 3:1-70

Webster, R. And M.A. Oliver. (1992). Sampling Adequately To Estimate Variograms of Soil Properties. J. Soil Sci., 43: 177-192

Webster, R., Oliver, M., (2001). Geostatistics for Environmental Scientists. John Wiley & Sons, Ltd, Chichester, 271 Pp.

Whigham, P. A. (2001). Spatial Information: Problems, Challenges and Directions. Proceedings of the 6th International Conference on Geocomputation. Brisbane, Australia.

Xiao, N., Ahlqvist, O., & Kwan, M.-P. (2007). Public Participation GIS For The General Public? Geocomputation 2007. NUI Maynooth, Ireland.

Xing-Yi Z, Yue-Yu S, Xu-Dong Z, Kai M, Herbert SJ (2007) Spatial Variability Of Nutrient Properties in Black Soil of Northeast China. Pedosphere 17(1):19-29

Zhang, C. (2005). The Roles of Web Feature And Web Map Services In Real-Time Geospatial Data Sharing For Time- Critical Applications. Department Of Geography And Geology, University Of Wisconsin.

Zhao Y, Peth S, Kru¨Mmelbein J, Horn R, Wang Z, Steffens M, Hoffmann C, Peng X (2007) Spatial Variability of Soil Properties Affected By Grazing Intensity In Inner Mongolia Grassland. Ecol Model 205:241-254

Zhou, F., Huai-Cheng, G., Yun-Shan, H., & Chao-Zhong, H. (2007). Scientometric Analysis of Geostatistics. Scientometrics. Volume, 73.

Zimmerman, D., Pavlik, C., & Ruggles, A. (1999). An Experimental Comparison Of Ordinary And Universal Kriging And Inverse Distance Weighting. Mathematical Geology , 375-390.