Advances in Geographical and Environmental Sciences

Manish Kumar · R. B. Singh · Anju Singh · Ram Pravesh · Syed Irtiza Majid · Akash Tiwari

Geographic Information Systems in Urban Planning and Management

IGU UGI

Springer

Advances in Geographical and Environmental Sciences

Series Editors

Yukio Himiyama, Hokkaido University of Education, Asahikawa, Hokkaido, Japan

Subhash Anand, Department of Geography, University of Delhi, Delhi, India

Advances in Geographical and Environmental Sciences synthesizes series diagnostigation and prognostication of earth environment, incorporating challenging interactive areas within ecological envelope of geosphere, biosphere, hydrosphere, atmosphere and cryosphere. It deals with land use land cover change (LUCC), urbanization, energy flux, land-ocean fluxes, climate, food security, ecohydrology, biodiversity, natural hazards and disasters, human health and their mutual interaction and feedback mechanism in order to contribute towards sustainable future. The geosciences methods range from traditional field techniques and conventional data collection, use of remote sensing and geographical information system, computer aided technique to advance geostatistical and dynamic modeling.

The series integrate past, present and future of geospheric attributes incorporating biophysical and human dimensions in spatio-temporal perspectives. The geosciences, encompassing land-ocean-atmosphere interaction is considered as a vital component in the context of environmental issues, especially in observation and prediction of air and water pollution, global warming and urban heat islands. It is important to communicate the advances in geosciences to increase resilience of society through capacity building for mitigating the impact of natural hazards and disasters. Sustainability of human society depends strongly on the earth environment, and thus the development of geosciences is critical for a better understanding of our living environment, and its sustainable development.

Geoscience also has the responsibility to not confine itself to addressing current problems but it is also developing a framework to address future issues. In order to build a 'Future Earth Model' for understanding and predicting the functioning of the whole climatic system, collaboration of experts in the traditional earth disciplines as well as in ecology, information technology, instrumentation and complex system is essential, through initiatives from human geoscientists. Thus human geosceince is emerging as key policy science for contributing towards sustainability/survivality science together with future earth initiative.

Advances in Geographical and Environmental Sciences series publishes books that contain novel approaches in tackling issues of human geoscience in its broadest sense — books in the series should focus on true progress in a particular area or region. The series includes monographs and edited volumes without any limitations in the page numbers.

Manish Kumar · R. B. Singh · Anju Singh ·
Ram Pravesh · Syed Irtiza Majid · Akash Tiwari

Geographic Information Systems in Urban Planning and Management

Springer

Manish Kumar
Department of Geography
Central University of Haryana
Mahendergarh, Haryana, India

R. B. Singh
Department of Geography
Delhi School of Economics
University of Delhi
New Delhi, Delhi, India

Anju Singh
Department of Geography
Aditi Mahavidyalaya
University of Delhi
New Delhi, Delhi, India

Ram Pravesh
MapmyIndia
New Delhi, Delhi, India

Syed Irtiza Majid
Department of Geography
Central University of Haryana
Mahendergarh, Haryana, India

Akash Tiwari
Department of Geography
Central University of Haryana
Mahendergarh, Haryana, India

ISSN 2198-3542 ISSN 2198-3550 (electronic)
Advances in Geographical and Environmental Sciences
ISBN 978-981-19-7854-8 ISBN 978-981-19-7855-5 (eBook)
https://doi.org/10.1007/978-981-19-7855-5

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

In the late twentieth century, the term sustainable development came to represent an ideal outcome in the sum of all planning and management objectives. As advocated by the United Nations-sponsored World Commission on Environment and Development in *Our Common Future* (1987), *sustainability* is defined as 'development that meets the needs of the present without compromising the ability of future generations to meet their own needs'. According to *World Urbanization Prospects: 2005 Revision* prepared by the United Nations Population Division, the population increase at the rate of 1.8% per year is expected to raise world urban population to 60% by 2030. It was only 30% of the world population in 1950, 47% in 2000, and in 2008, for the first time, half the world's population was living in towns and cities. As such the rising urban density necessitates that suitable urban planning and management practices be executed by the urban administration for better quality of life. Accordingly, urban planning management aims to take care of the residents' housing, employment, recreation, trade and business, sanitation, mobility and communication besides preserving the natural and built heritage of the place.

Urban planning and management are mutually supportive and highly complicated activity. Most of the urban issues are related to land use activities as land is a scarce and highly valuable commodity. Additionally, the administrative bodies and other stakeholders (like urban local bodies or ULBs in India) are not equipped with the professional and technical capacity to effective plan and execute the urban planning and management. Conventional methods of surveying and maintaining land records cannot successfully manage the urban scenario today, as the pace of popular attitudes and technology has increased ahead of time. Geographic Information System (GIS) plays a pivotal role in the field of urban planning and management and provides better solutions for numerous urban problems. It offers the extraordinary vistas to better understand existing requirements of a city and designs to fulfil its specific needs.

This book titled *Geographic Information Systems in Urban Planning and Management* achieves the objective of developing scientific knowledge base on geospatial technologies among planners, researchers, scientist, professionals, students and laymen and offers them with better understandings on urban planning and management at various levels of planning. It exhibits the importance of GIS in better understanding current urban challenges and provides new insights on how to apply GIS in urban planning.

The book is organized into two major parts. Part I discourses on the fundamentals of Geographic Information System in eight chapters, and it also deals with the advance techniques of spatial planning. Chapter 1 deliberates on the historical, definitive and introductory aspects of Geographic Information System. It also discusses linkages between the component geospatial technologies and related integration models and provides a general account applied GIS in the real world. Chapter 2 explains the concepts of map projection, coordinate system and georeferencing of various geospatial data products. It also offers a detailed account of several coordinate systems. Chapter 3 on GIS data models is a discussion on structure, components and specifications of raster and vector data models. Important raster data products are detailed out in the chapter along with the deliberation on vectorization and rasterization. The chapter also explains raster data encoding techniques and compression of the raster data. Spaghetti and topological vector data models, with a special focus on the fundamental precepts of topology, are also discussed. Chapter 4 enumerates the sources of geospatial data and explains the methods of inputting spatial and non-spatial data into GIS. A discussion on errors encountered in the geospatial data input process and spatial data editing (topological and non-topological) is also provided. Chapter 5 explains the process of geovisualization and deliberates on the standards of cartographic representation of several data types. Chapter 6 explains the spatial data analysis for vector data including buffering, overlay, distance measurement, and minor operations like dissolve, clip, append, select, erase, eliminate, split, etc. And for raster data local operations, focal operations and zonal operations are discussed. This chapter also explains the requirement and potential of spatial data analysis in GIS. Chapter 7 is a discussion on management of non-spatial data. It involves the discussions on non-spatial data types and databases, attribute tables, database management, relational data model, methods of linking spatial and non-spatial data in GIS and its manipulation. Chapter 8 deliberates on the applications of GIS in urban planning and management with special reference to microlevel planning, hydrological management, sustainable development, agricultural resource management and sustainable tourism development and sustainable tourism development.

Part II showcases the applications of geospatial data products and GIS-based methodologies and techniques in the studies of urban planning and management through seven case studies (Chaps. 9–15). It demonstrates the addressal of real-world case studies using a bunch of GIS applications. Case Study 1 depicts the monitoring and modelling of land use change in Mirzapur District of Uttar Pradesh. For this study, data were downloaded from USGS Earth Explorer and utilized on the GIS Environment ArcGIS and ERDAS Imagine to deduce the information on land use-land cover change over the period 2001–2020. Case Study 2 is based on the future land use-land cover simulation. This case study was performed for Bangalore city. Land use-land cover map of the region was prepared for three-time intervals (2001, 2011 and 2016). Cellular automata-Markov chain model was used in this case study to predict the future Land use-land cover map for the year 2025. After forecasting the future land use map, change was analysed using Land Change Modeller. Case Study 3 shows the effectiveness of Multi-criteria Evaluation Techniques for the selection of potential sites for housing development. This case study was performed for the Hawalbagh Block of Almora District in Uttarakhand. Analytical Hierarchy Process (AHP) was selected as a Multi-criteria Evaluation Techniques which consider several heterogeneous criteria such as slope, road proximity, land use, distance from developed land, landslide, lithology, drainage proximity, lineament, and aspect as a major criterion in site selection. Finally, suitability map was classified into five classes, viz. very less, less, medium, high and very high. Case Study 4 on analysis of urban green space and selection of potential site for green space expansion in NCT Delhi involves the identification of green space using the remote sensing data, estimation of per capita green space, and weighted buffering analysis for potential site selection. Case Study 5 shows the importance of applicability of fuzzy TOPSIS as Multi-criteria Decision-Making Technique for the selection of the alternative landfill site selection of Varanasi District in Uttar Pradesh. To select the best alternate landfill sites in the Varanasi, distance from waste production centre, distance from roads, depth to ground water, distance from water bodies, distance from settlement land, types of soils and slope were considered. These criterions were classified into beneficial and non-beneficial criterions on the basis of expert knowledge. There were five alternative landfill sites considered for the selection of best sites. Case Study 6 is a study of modelling the urban flood susceptibility of Srinagar, Jammu and Kashmir, India using novel fuzzy multi-layer perceptron neural network (fuzzy MLPNN). It discusses the creation of conditioning factor databases, inventory databases and modelling process associated with urban flood modelling and validation of resultant cartographic outputs. Case Study 7 assesses, maps and predicts the urban heat island scenarios in the Srinagar City region using the land surface temperature (LST) as means of study, and the methodologies of mono-window algorithm for LST retrieval, LST thresholding for urban heat island identification, cellular automata-Markov chain integrated model (CA-Markov) for future prediction and multi-layer perceptron neural network (MLPNN) for estimating future mean temperature.

This book has used data from several data sources to carry out the work of case studies including that from Indian Space Research Organisation (ISRO), National Remote Sensing Centre (NRSC), Survey of India (SOI), Office of the Registrar General and Census Commissioner, India, US Geological Survey, European Space Agency, etc. among others. We extend our gratitude to all of these data providers for availing us the information and enabling us to accomplish the substantial work of this book. Additionally, all the authors of the numerous books, research papers, and various websites that have been cited in this work are duly acknowledged, and we extend out heartful gratitude to them as well.

Last but not least, we are indebted to Late Prof. R. B. Singh Sir who enthusiastically encouraged us to take up this challenging task of writing the book. Their contribution and inspiration in shaping this book is laudable, and we are grateful to him from the core of our hearts.

Mahendergarh, India Manish Kumar
New Delhi, India R. B. Singh
New Delhi, India Anju Singh
New Delhi, India Ram Pravesh
Mahendergarh, India Syed Irtiza Majid
Mahendergarh, India Akash Tiwari

Contents

About the Author

The responsible series editor of this book is Dr. R. B. Singh.

Part I
Fundamentals of Geographic Information Systems

Chapter 1
Introduction of Geographic Information System

Abstract A geographic information system (GIS) is a computer system that is used to store, manage, analyse and display geospatial data. GIS has been used by professionals in natural resource management, land use planning, natural hazards, transportation, health care, public services, market area analysis and urban planning since the 1970s. It has also become a crucial tool for normal operations for government organizations at all levels. GIS integration with the Internet, GPS, wireless technology and web services has recently found applications in location-based services, web mapping, in-vehicle navigation systems, collaborative web mapping and donated geographic information. You should be able to understand the following after reading this chapter:

- Basic concepts of the GIS
- Historical background of the GIS
- Interrelation between GIS, Remote Sensing and Global Positioning System (GPS)
- Capabilities and briefly understanding about the GIS

Keywords GIS · Remote sensing · Global positioning system · Integration model

1.1 Introduction

We are currently living in a technological era. We have sophisticatedly accepted a technology-driven lifestyle. Most phenomena in this era may be effectively described and interpreted by vividly visualizing the phenomena. Geographic Information System (GIS) is one such computer-based approach that is commonly used for storing, interpreting and analysing phenomena occurring on the earth's surface. Geographic Information Systems (GIS) investigates the spatial linkages, patterns and trends in geography. GIS is a system for creating, managing, analysing and mapping several kinds of data. By integrating the location data with that the other forms of the descriptive data, GIS is helpful to connect data to the map. It made crucial to perform groundwork to map and analysis which is being employed in science and practically

M. Kumar et al., *Geographic Information Systems in Urban Planning and Management*, Advances in Geographical and Environmental Sciences,
https://doi.org/10.1007/978-981-19-7855-5_1

each sector. GIS is helpful for the detailed comprehension, relationship and analysis of geographical data. It not only improves the efficiency but also management, planning and conservation strategies could be very well implemented using GIS. In this chapter we'll understand the historical evolution of the GIS. A well-defined definition of the GIS will also be understood in this chapter. In this interesting journey to start the visit of GIS world, its linkages with Remote Sensing (RS) and Global Positioning System (GPS) will be highlighted in this very first chapter. To develop wide understanding about the GIS it's crucial to acknowledge with the basic components of the GIS. So, for this aspect major basic component of the GIS is discussed in this chapter. Finally, we'll understand the importance and capability of the GIS through daily life examples and applications.

1.2 Evolution of Geographical Information System

Nigel Waters of The University of Calgary, Canada has very well discussed the historical evolution of GIS in his research paper entitled "**GIS: History**". Under this section we'll discuss the development of GIS from antecedents to contemporary world. Before introducing the Geographical Information System (GIS) to the world of academic geography there were huge lack of knowledge about the spatial discreteness and its linkages. Growth of quantitative and theoretical approaches to geography in the 1950s and 1960s posed a challenge to the uniqueness of place. Era of the supremacy of spatio-scientific temperament was promulgated with the evolution of GIS during 1960s. Those students who had completed their graduation and have some basic idea about the use and building of new software tools were getting on the basis of GIS based knowledge during late 1980s. Meanwhile, there are many students to praising the spatial uniqueness and qualitative geography while many also are against to it since the invention of GIS. In this section we are intended to discuss the scope and extension of the GIS during last two decades and will try to elaborate that how the individual GIS working personnel recognizes spatial uniqueness and commonalities.

1.2.1 Looking Behind the Geographic Information Systems

Information regarding organization, depiction and analysis of the spatial data is briefly provided by the individuals who have basic idea of GIS operation. During the "Battle of Yorktown", the French geographer Berthier made map overlays depicting military movements, which are widely referenced in these talks. Dr. John Snow who was a British physician tracked the outbreak of Cholera in London in 1854 and concluded that it is water-borne disease rather than air-borne disease (Waters

1998). They both (Berthier and Snow) tried to map the data so that it could be understood the basic concept behind its generation and outbreak. Besides these developments in the mapping GIS specialists were still not using the combined spatial data and overlay process. This became the important part of mapping approximately a century later. This type of mapping in planning was explored and originated by Steinetz, Parker and Jordan during 1976. Warren Manning performed research on the town of Billerica, USA in 1912. His work was published in the journal "Landscape Architecture Quarterly" in 1913. He used overlay analysis in his research. On the same track Steinetz, Parker and Jordan also performed overlay analysis in 1912 for designing a plan for Düsseldorf city in Germany. McHarg (1969) immediately popularized this technique also. He is considered as having extraordinary inspiration on GIS development because of his book named "Design with Nature". But this book, "Design with Nature" is not able to describe the method for the computerization of the overlaying process. This type of methodology is also mentioned by Steinetz et al. (1976). This is crucial to point that various breakthroughs happened around the notified military zone and its vicinity prior to the introduction of software and hardware systems that were legally acknowledged as "Geographic Information Systems. A detailed assessment was provided by Clarke and Cloud (2000) on the importance of US CORONA programme for the investigation satellites under the Central Vigilance Agency: 1959–1972 and SAGE programme for the purpose of image processing and analysis.

1.2.2 Development in Geographic Information Systems

Coppock et al. (1991) classified the development of GIS within four periods:

(a) The Pioneer Period: Mid 1950s–Early 1970s
(b) The Period of Government Supported Experiments: Mid 1970s–Early 1980s
(c) The Commercial Period: Early 1980s–Early 1990s
(d) Era of the User Dominance: Starting from 1990s

Coppock and Rhind tried to present a chronological development, innovations, improvement in software and hardware over the predescribed period. Coppock and Rhind's paradigm is followed in this discussion. An optional way of perspective on the early days of GIS was provided by the Foresman (1998). Much operational and pioneer chapters on the origin of GIS were presented in the edited book of the Foresman.

Timeline which is defined by Foresman in the development of GIS has several superimposing time periods. They are listed here:

(a) The Pioneer Age: Mid-1950s–Early 1970s
(b) Research and Development Age: Early 1970s–Mid-1980s
(c) Implementation and Vendor Age: Early 1980s–Mid-1990s
(d) Client Application Age: Early 1990s–Early 2000s

(e) Local and Global Network Age: Late 1990s–Early 2000s

In the field of academic geography, computation and environmental awareness these advancements could be then interlinked.

1.2.2.1 The Pioneer Period: The Mid-1950s–Early 1970s

The previously stated overlay or layer-cake approach for geographical data arrangement and Brian Berry's popularization of the geographical data matrix were conceptual advancements during this time period. The data matrix was originally arranged with columns indicating locales and rows reflecting their properties. When the concept of Berry was executed into early GIS programs to create geographical matrix much interoperable having standard database technology, this arrangement was placed. Canada Geographic Information System (CGIS) was developed as the first completely functional vector-based Geographic Information System during 1960s. The CGIS has developed a variety of wonders innovation which include hardware for the purpose of laser scanning and software for the vectorization of the result images and to store the layers of raster effectively with Morton ordering. This development was done in collaborations with Spartan Air Services (company of Roger Tomlinson) and Government of Canada Land Inventory. Roger Tomlinson is commonly referred regarded as the "Father of GIS" due to his extraordinary work in the evolution and development of GIS. Persons who were working in the field of the geospatial technologies were not able to access the applicability of computer science because at that time computer science was not able to provide the all-necessary information in the field of GIS development. With the passage of computing development from mainframe computing GIS would be supposed to benefited. External advancements in database technology were also important to the advancement of GIS. During this pivotal period, Howard Fisher's "Harvard Laboratory for Computer Graphics and Spatial Analysis", founded in 1965, set the groundwork for following breakthroughs. As a result, this laboratory contributed a lot by developing the widely used mapping software such that as SYMAP, CALFORM, SYMVU, GRID, etc. North American Institutions were widely used the programmes, viz. SYMAP, CALFORM, SYMVU for the creation of line printer maps and 2D and 3D plots. In 1969, William Warntz was selected as the Director of the Harvard Laboratory. He did significant contribution for example he assumed that surface's critical features can be utilized to generate Triangulated Irregular Network (TIN) models in order to produce the much-compacted storage of the surface features. Jack Dangermond, who is related to leading GIS company of the world ESRI, was also associated with the renowned Harvard Laboratory. ESRI is responsible for the development of world class GIS application named ArcGIS. Chrisman provides a comprehensive history of the Harvard Laboratory's early years (2006). In 1967, the Experimental Cartography Unit (ECU) was formed at the UK by David Bickmore. The ECU worked on same path as Harvard Laboratory worked tremendously in the field of computer-based GIS analysis and mapping. Very first geocoded census was released by US Census Bureau in 1970. The concepts which are thus shared through the organizations is helping the field of academic developments, governmental planning and industrial development etc. One of the earliest such organizations is the Urban and Regional Information Systems Association (URISA). URISA conducted its first conference in 1963 and has had yearly conferences since then. The proceedings of the URISA conferences has been always great source of knowledge regarding the growth and development of GIS.

1.2.2.2 Government-Funded Experimental Research Period: Mid-1970s–Early 1980s

During this time, concepts and software were being developed for the academic developments, governmental planning and industrial development. Such that, methods created in the mid-1960s to tackle location-allocation problems were now accessible as separate programs. These programmes were being tried to integrate within software system (Waters 1998). Development of the Swedish Road Data Bank and various other digital databases was result of the government-sponsored research programmes in the Europe. Development and growth in the field of computer science is being mirrored by the advancements and development in the GIS. GIS software are progressively migrated to new computing platforms by allowing the time-sharing capacity of mainframe computer systems to mini computers and desktop microcomputers.

The Minnesota Land Management Information System, (MLMIS), was one of the most famous parts of the government funded experimental research mainframe of this era. Near the end of this period, it was observed that, the data accuracy problem, challenges related to data access and interface of the nonuser-friendly command line were rapidly going to be obsolete because of very high maintenance cost. Huge mainframe systems were sequentially phased out by the development of strong workstations in the early 1980s.

1.2.2.3 The Commercial Period: Early 1980s–Late 1980s

Tomlinson, 1987 gave the ‘state-of-the-art’ review on the GIS, newly developed subdiscipline in geography in the very first volume of that which would be the flagship publication in the field of the geography. It was observed by him that GIS software use in government and commercial organizations has made tremendous progress, particularly for the transportation and facilities development and planning, local level planning, agricultural and environmental development and planning, and engineering and technologies sectors. He predicted that, rather than the private sector, new breakthroughs would emerge from university and government. This was only partially correct, as Esri published ARC/INFO, the first commercial GIS, in 1982. The CGIS model of managing spatial and attribute data independently was modified by ARC/INFO. The first employed Esri’s topological ARC structure, whereas the second used the INFO relational database. Due of the popularity of the IBM PC desktop computer, Esri developed PC ARC/INFO in 1986. Tomlinson (1987) correctly described that the lack of educational capabilities would act as a huge barrier in the spread of the knowledge of GIS. In order to encourage the fundamental research in the field of GIS development, a grant was awarded to the consortium of universities in 1988 by the US National Science Foundation (NSF). Department of geography of 'University of Santa Barbara'; state university of New York; Surveying and Engineering department of 'University of Maine' was included in the consortium of universities. Besides other developments, National Center for Geographic Information and Analysis (NCGIS) established a fundamental GIS prospectus framework which consisted 75 lectures fragmented into 3 semesters. Thus, each semester consists of 25 lectures. The lectures originally comprised additional laboratory material and were written by 35 different authors. They were a huge success, with over 1300 copies delivered to over 70 countries by January 1995. This fundamental curriculum was translated into various other international languages. This core curriculum was framed to provide basic knowledge which will permit the faculty who will teach the GIS to explain the broadened scope of the GIS to the students so that they could have idea about applicability of the GIS in the daily life. Four Regional Research Laboratories (RRLs) were created by the Economic and Social Research Council of the UK in February 1987 at London, Edinburgh, Cardiff and Newcastle. They all were aimed to data management, software development, spatial analysis, research training and development.

1.2.2.4 User Dominance Period: Late 1980s–Mid-1990s

Coppock and Rhind highlighted the noticeable competition among the GIS software development companies since the fourth periods because of extreme user dominance. This resulted in the reduction of the broad market to a limited number of providers. This vast competitiveness tendency was aided by the several GIS conferences and annual 'GIS World Sourcebook' publications (Waters 1998). As a result of the increasing competition, various companies have emerged in the field of GIS software development. In this period of time, service providers switched themselves from the Command Line Interface complexity towards graphical user interfaces (GUIs). TIGER file system was released by the 'US Census Bureau' in 1990. Meanwhile, other countries like UK and Canada made available the census data on a cost recovery basis. Under the 'Aarhus Convention', which turned into the law in October, 2001, many governments have allowed increasing open access to geospatial data in recent years. However, the pact was only signed by European and Asian countries. GIS software for niche markets was provided by a number of companies. For example, TransCAD is developed by Caliper Corporation focused for the transportation and planning; IDRISI is developed by Clark University focused to operate several geospatial and modelling-related tasks. During this time, the push in the USA to incorporate GIS instruction into the K-12 curriculum gained traction. These all efforts were reviewed, and resource material and suggestions for educating the GIS at elementary and secondary level schools were presented during the 1st national conference on the educational applications of the Geographic Information Systems which was held in 1995 under the sponsorship of Technical Educational Research Centres (TERC).

1.2.3 Geographic Information Science (GIScience, GISc) Era

The academic subject of GIS underwent a major transformation in the mid-1990s. Goodchild made a major theoretical contribution in 1992. He stated forcefully that the disciplines should be move away from its general characteristics of questions and answer. GIS as a discipline has very huge scope of exploration (Goodchild 1992). In the introduction part of his paper, he claimed that majority of the initial period of the GIS was driven by the technology. Major concern was to obtain the geographical data to be processed within the information system. Various government agencies, remote sensing companies are developing new technologies for the acquisition of the data to be processed within the information system. Goodchild mentioned on the importance of the handling and the exploitation of the data which is stored in the GIS database. Way ahead was put forward during the recently developed conferences series, Spatial Data Handling Symposia under the International Geographical Union. In 1984, for the first time such type of meeting was held in Zurich. Nascent discipline of the GIScience was devised into series of the topics by the Goodchild which included:

(a) Handling of the Spatial Data, (b) Data Collection and Its Management, (c) Data Capturing, (d) Spatial Statistics and Analysis of the Data, (e) Modelling and Theories of the Spatial Data, (f) Structures of the Data, (g) Algorithms and the Processing of the Data, (h) Visualization of the Data, (i) Analytical Tools and Operations, (j) Issues Related to Institutions, Management and Ethics.

Whereas few of the advancements benefited from computational advancements (e.g. data visualization in a scientific way, usually referred as second computer science revolution), the majority of them initiated novel subdisciplines in the field of geography. ‘Spatial Analysis’ and ‘GeoStatistical Analysis’ could be considered as some new fields which has now become the integral part of the department in the form of ‘Information Sciences’ or ‘Geomatics Engineering’, etc. This new change in the geographical horizon could be observed by the alteration of the names of the renowned journals and emphasis more on the science than the system.

For example, International Journal of Geographic Information Systems change its name from “______System” to “______Science” in 1997. This new name was detailed described and explained by then editor, Peter Fisher. He deeply acknowledged to Goodchild for providing a rationale to new focus on the journal and the novel discipline. Cartography and Geographic Information Systems was renamed by an American cartographer in 1990. This was continued till 1999 when ‘system’ was renamed to 'science'. This nomenclature system was kept unchanged during 2nd edition of the Geographical Information Systems. They all mentioned the both terms, viz. "Geographic Information Systems" and "Geographic Information Science" in the very first edition of the textbook which got published after two years (Longley et al. 2001).

"The Annals of the Association of American Geographers" accept the submission under four major categories. Major of them are methods, models and GIScience. Table of contents of this journal contained few more ambiguous acronyms on GIS. Novel educational and research-related resources are available at the 'Digital Library of the Earth System', 'Centre for Spatially Integrated Social Science at NCGIA', Virtual Campus of the ESRI, etc. (Waters 2013). A detailed review was given by Goodchild in 2010 on the progress in the field of GIScience after 20 years the concept was discussed. This period was also marked by new educational initiatives. The NCGIA board of directors recommended in 1990 for the formation of new organization for assisting researchers in the field of GIScience. According to the brief history of the University Consortium for Geographic Information Science (UCGIS), an ad hoc steering committee was formally developed in 1991. This committee consisted 16 members which come from seven various disciplines. But it was unfortunate that none of the member of committee were listed in its historical note. Annual meetings and symposia have been promoted by the UCGIS. It is noticeable fact that a 'Body of Knowledge' (BoK) has been developed by UCGIS for the teaching purposes under the collaboration of the Association of American Geographers. This was recognized as a natural progression from NCGIA's fundamental framework. On the other hand, the BoK had been conceptualized in a quite different way. This newly developed BoK was divided into 10 areas; these 10 areas were further subdivided within 73 units. Again these 73 units were further divided into 330 topics. Deficiency of the explicit structure for the continuous updating is one of the major weaknesses of the core curricula of the BoK. Meanwhile it becomes important to notice that the process of the updating in the curricula in computer science is performed every six years. GIScience has yet to accomplish this. This could not be ignored that these GIS software have expanded tremendously all over the world. ESRI's former education specialist, Mike Phoenix, highlighted that his software had been extended up to more than 60 different types of the universities till the end of 1990s (Waters 1998). In present time evolution and development of GIS has not only impacted the geographical disciplines but also several other interdisciplines are hugely impacted due to it. The implementation of the geospatial tools and techniques is growing day by day.

1.3 Meaning and Definition of Geographic Information System

According to the previous sections, GIS can be defined as a computer system capable of acquiring, storing, analysing and displaying geographically referenced information; that is, data identified by place. A GIS is also defined by practitioners as the procedures, operating staff and spatial data that go into the system.

Many scholars defined the GIS in many ways, out of them few important are cited here.

1. According to Chang (2019), A geographic information system (GIS) is a type of database containing geographic data (that is, descriptions of phenomena for which location is relevant), combined with software tools for managing, analyzing, and visualizing those data.
2. According to Ali (2020), A Geographic Information System (GIS) is a system of computer software, hardware and data, personnel that make it possible to enter, manipulate, analyze, and present data, and the information that is tied to a location on the earth's surface.
3. According to National Geographic Society, A geographic information system (GIS) is a computer system for capturing, storing, checking, and displaying data related to positions on Earth's surface. GIS can show many different kinds of data on one map, such as streets, buildings, and vegetation. This enables people to more easily see, analyze, and understand patterns and relationships.
4. According to Burrough (1986), "GIS is defined as a powerful set of tools for collecting, storing, retrieving at will, transforming, and displaying spatial data from the real world."
5. According to Goodchild (1985), "GIS is a system that uses a spatial database to provide answers to queries of a geographical nature."

Further, except these all there are also a variety of definitions. GIS is treated as a spatial data handling system by Marble and Peuquet (1983). Cowen (1990) treats GIS as an information system which handles geographically or spatially referenced data. A different view was provided by Berry (1986), by pointing out the GIS as an internally referenced, automated, spatial information system. Few authors like Devine and Field (1986) explained GIS simply as a device to expand the use of maps. Parker (1988) defined GIS as an information technology which stores, analyses and displays both spatial and non-spatial data.

1.4 Linkages Between GIS, Remote Sensing and Global Positioning System

Before discussing the linkages of Global Positioning System (GPS), Remote Sensing (RS) and Geographical Information System (GIS), their definition must be cleared primarily. Since we have already discussed about GIS, it is not necessary to elaborate again it here. Here, only we'll focus on the definition of the rest two concepts, and then will discuss the linkage/integration among these segments.

There are wide variety of definitions available to define Remote Sensing (RS). Lillesand and Kiefer (1994) defined it as the science and art of obtaining information about an object, area, or phenomenon through the analysis of data acquired by a device that is not in direct contact with the object, area, or phenomenon under investigation. Colwell (1983) defined it as the gathering and processing of information about the Earth's environment, particularly its natural and cultural resources, through the use of photographic and related data acquired from an aircraft or satellite. Campbell (1996) explained the RS as the science of deriving information about the Earth's land and water areas from images acquired at a distance. It usually relies upon measurement of electromagnetic energy reflected or emitted from the features of interest.

On the other hand, GPS, because to its brief history, does not have not too much definition. Wooden (1985) explained GPS as an all-weather, space-based navigation system under development by the Department of Defense to satisfy the requirements for the military forces to accurately determine their position, velocity, and time in a common reference system, anywhere on or the earth on a continuous basis. Bock (1996) justified the work of GPS as to provide a global absolute positioning capability with respect to a consistent terrestrial reference frame.

1.4.1 Need of Linkages Between GPS with Remote Sensing and Global Positioning System

GIS database could consist two types of data. It could be either geographical type of data or it could be the thematic type of data. In ancient times, geographical data and a few thematic data associated to it was digitized through the available topographic maps or land use-land cover maps. These maps represent the secondary characters, and due to over map generalization, all the necessary features could not be expressed by such type of the maps. With the development of the geospatial technology and remote sensing these constraints could be addressed very well. With the passage of sequential development, aerial pictures and satellites images could be considered as unique and more real-time based data for the preparation of the maps and deducing the valuable information. All the major definitions of the remote sensing are highlighting the ability to acquire data efficiently. Original aerial photographs and satellite images, whether digital or analogue, are georeferenced to their respective coordinate systems. Suppose, one wants to examine the phenomena or area from various frames of photographs or whether these data need to be spatially layered from other sources, they all must to be georeferenced for sharing a matching coordinate system. Georeferencing necessitates some form of ground control. Topographic maps have traditionally been used to generate geometric control for remote sensing imagery because they are widely available and mathematically dependable.

However, topographic maps are entirely useless in locations where there are no recognizable landmarks or where the scene has changed dramatically since the maps were published. In this situation, GPS can be a cost-effective solution for acquisition of data having automatically referenced data to universal coordinate system. Real-time or near-real-time access to point, linear, or even areal data is possible. To georeference remotely sensed images, point-based data at specified landmarks, also known as ground control points (GCPs), can be used. Because of its inability to obtain fully 2D areal data, GPS data, even if areal in nature, can never replace aerial pictures or satellite images.

The preceding description demonstrates how remote sensing, GIS and GPS are inherently complimentary in their primary purposes. Each technology has its own set of constraints. When used separately, each technology may be difficult or impossible to use in some applications. Their strengths can only be fully utilized through integration. Integration will not only make their applications in resource management and environmental monitoring (e.g. wildfire fighting) easier, but will also increase their scope of application (e.g. real-time emergency response).

1.4.2 Integration Models

Integration of Remote Sensing, GPS and GIS could be conceptualized and characterized by four models (Gao 2002): linear, interactive, hierarchical and complicated. They are discussed further below.

1.4.2.1 Linear Model

In the linear paradigm, data flows linearly from GPS to remote sensing and finally to a GIS (Fig. 1.1). The distinctive strength of each component is maximized in this model. In other words, GPS is used for the collection of the geometric control for the aerial photographs and satellite imageries. Photographs that have been rectified could then further imported into a GIS database. The model's linear nature implies that the three components are not equally essential. The end goal of the integration, a GIS, plays a dominant role. Rectified photographs are stored in the GIS database. The GIS is used for all spatial studies and modelling. In comparison, GPS's role is secondary that the point-based data for the remote sensing is being provided by it. Because of deficiency of the direct connection between GPS and GIS, data derived through GPS platform is not possible to access directly in the final GIS database. For bridging the gap between remote sensing and other data in GIS, GPS data could be used. These could be used for the standardization of the satellite images and aerial photographs for the georeferencing of the system which uses other type of GIS data.

Fig. 1.1 Linear model of integration

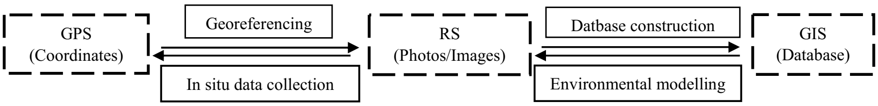


Fig. 1.2 Interactive model of integration

1.4.2.2 Interactive Model

The structure of the interactive model is very similar to that of the linear model (Fig. 1.2). Closer inspection reveals that not only the data flows between GPS and Remote Sensing, but also between Remote Sensing and GIS. As a result, Remote Sensing could no longer be viewed as merely supplying data to a GIS. Although data can move from a GIS to RS in this approach, integration of the 'left-to-right' type is slightly much typical than that of data movement in the reverse way, as indicated in Fig. 1.2.

The interactive aspect of GIS and Remote Sensing makes judging their relative importance challenging, even if GPS as a data collecting tool is seen as less important. Unlike the linear model, GPS data can be overlaid with remote-sensing-derived results to map elements like roads that are unseen on satellite imagery due to their coarse spatial resolution (Treitz et al. 1992). After integration, GPS data is directly viewable in the GIS database.

1.4.2.3 Hierarchical Model

The hierarchical model has two levels of integration (Fig. 1.3). Level one of integration is happened between GPS and Remote Sensing data. Aside from the previously mentioned, remotely sensed images are corrected by incorporating the GPS coordinates. These rectified and accurate images are utilized for characterizing the in-situ samples on the basis of their locations. Relationship between the variables to be investigated and its image attributes could be aided by overlaying the in situ samples with that of digital photographs or satellite imageries (Usery et al. 1995; Gao and O'Leary 1997). A spatial analysis tool such as S-Plus can be used to construct a statistical relationship between the two variables.

The second level of integration (modelling) incorporates remotely sensed data, other GIS data and/or mathematical models. Depending on the data format, spatial modelling can be carried out via remote sensing or a GIS. Modelling in raster format is as easy to do with a digital image analysis system as it is with a GIS. No direct

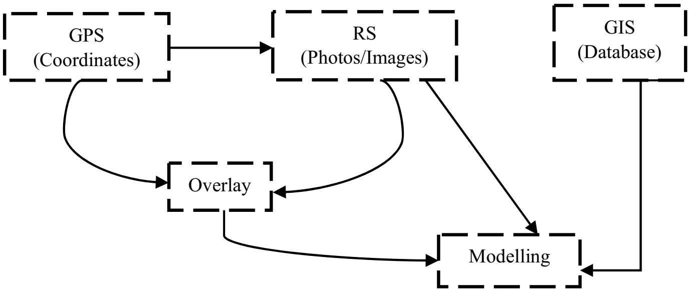


Fig. 1.3 Hierarchical model of integration

relationship between the GIS and Remote Sensing is observed in this integration approach.

1.4.2.4 Complex Model

This model represents the all-optimum linkages which could be possible between any of the two nodes. This model, shown in Fig. 1.4, depicts the ultimate or the total integration among these three nodes (GIS, GPS and Remote Sensing). An additional interaction is observed between the GPS and GIS in this model than the previous three models. In this situation, GPS data can be easily transferred to a GIS database to be updated or created from scratch (Bor 1994). It could be in the form of point, linear or areal data. Before integration, geometric properties should be changed for matching the data which is already contained within GIS database. Integration of these all nodes have found to be extremely useful to increase the precision of the acquired data. GPS receiver is primarily used for the measurement of coordinates which are associated with precision-farming variables and a GIS is utilized for the integration, storage and analysis of the data (Swindell 1995; Lachapelle et al. 1996). The integration of a GIS and GPS is comparable to the integration of Remote Sensing and GPS. Such type of relationship is formed by validation of the GIS modelling in the field or when more ground data at positions determined by the modelled results is gathered in the field.

1.5 Basic Components of Geographic Information System

There are five major components in Geographic Information System:

(a) Hardware
(b) Software
(c) Data
(d) People

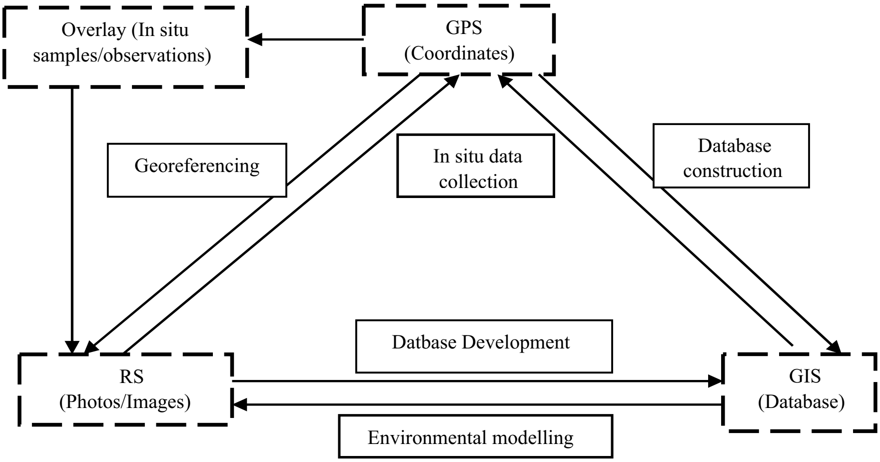


Fig. 1.4 Complex model of integration

(e) Methods (Fig. 1.5)

(a) Hardware

The computer on which GIS software operates is referred to as the hardware. Nowadays, computers are in different shape and size, and they can be either desktop or server-based. It should be noted that ArcGIS is considered as server-based computer

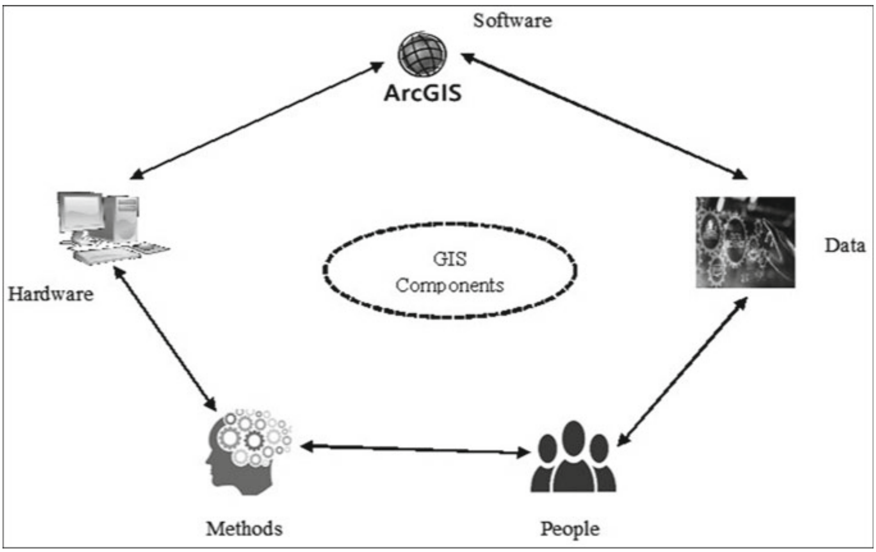


Fig. 1.5 Components of GIS

which is able to process the GIS-related operations on network or in cloud. To increase the efficiency of the processing computer should be highly configured. Some examples of hardware components include: a motherboard, a hard drive, a processor, a graphics card, a printer and so on. All of these components work together to ensure that a GIS software runs smoothly.

Some major hardware components are:

(i) *Motherboard:* A board on which major hardware components are mounted or a location where all components are connected.
(ii) *Hard Drive:* It is also known as a hard disc, which is a storage device for data.
(iii) *Processor:* The processor is the most important component of a computer; it executes calculations. It is known as the Central Processing Unit (CPU).
(iv) *RAM:* RAM is the temporary storage location for all running programmes.
(v) *Printer:* It is an output device that is used to print an image, a map or a document. There are different types of printers on the market.
(vi) *Monitor:* It is a screen that displays output data. There are several types of monitors available today, including cathode ray tube (CRT), liquid crystal display (LCD), light emitting diodes (LEDs) and others.

(b) Software

The next component is GIS software. Software is essential for the processing of spatial data and to deduce the valuable information from it. Software help for the purpose of querying, editing displaying, projecting and so many things. These data are stored by using the Relational Database Management System (RDBMS). Few popular GIS software are ArcGIS, QGIS, SAGA GIS, etc.

Some major software components are:

(i) GIS tools: The following are the essential tools for exploring GIS data.
(ii) RDBMS: GIS data is stored in a relational database management system. GIS software may retrieve data from or insert data into RDBMS.
(iii) Query tools: Tools for querying, inserting, deleting and other SQL operations that operate with database management systems (Standard Query Language).
(iv) Layout: A good layout window for designing a map.

(c) Data

Data is considered as the fuel of the GIS. Data is the most important and expensive component of the GIS. Data could be either in the form of graphical or in the form of tabular format. Broadly this data could be considered as the vector and raster-based data. These both types of data could be processed on the GIS environment to deduce the basic information. By performing the digitization, digital GIS data could be extracted from the analogue or paper format data. Ground control points (GCPs) or known coordinate are used to spatially register the raster image throughout the process of the digitization. This is commonly referred to as rubber sheeting or geo referencing. Digitizing a raster image yields polygons, lines and points. The raster image can be registered with coordinates, which is commonly referred to as correcting the image. The majority of registered images are exported in TIFF format. GIS data, as previously stated, can be raster or vector.

There are two types of GIS data:

(i) Raster

Raster data are in the form of cells. These cells are also known as pixels. Raster data could be obtained from the aerial photograph or satellite-based imageries. To deal with the continuous data, such type of data could be used.

(ii) Vector

On the contrary to raster data, vector data is used to deal with discrete type of data. In vector format data is stored in the form of x–y coordinate format. There are three types of vector data:

(a) Point
(b) Line
(c) Polygon.

(d) People

People used the GIS software and take its charge. Dramatic improvement in the field of the hardware and the software made it easy and compatible to process the GIS applications. Increased efficiency of the computers has increased the potential of the people to use it and deduce the information and analysis for their research. Thus, people could be considered as the major segments for the successful implementation of the GIS.

(e) Methods

A well-designed plan and business operation standards are essential for successful GIS operation. Methods can differ among organizations. Any organization has documented their GIS operating procedure plan. These documents address a variety of GIS-related issues, including the amount of GIS experts necessary, GIS software and hardware, the process for storing data, the type of database management system (DBMS) and more. A well-organized plan would be able to cope up with all of such types of issues.

1.6 Capabilities of Geographic Information System

Geographical Information System (GIS) is a broad term that refers to any computer-based capability for modifying geographical data. Data input, data storage, data management (data manipulation, updating, changing and exchange) and data reporting are all hardware and software aspects of GIS (retrieval, presentation, analysis, combination, etc.).

All of these actions and operations are used by GIS as a tool to create its database. Geospatial Information Systems (GIS) enable the capture, management, analysis and visualization of spatial data. GIS has several applications across a wide range of industries and areas. Spatial information can play a critical role in project development or just provide a splash of colour to technical papers. Integrate sustainability's GIS specialists are also environmental professionals, giving them a unique viewpoint on how to use GIS to present, solve and manage spatial data in the context of a site's natural environmental assets. Integrate sustainability is able to create competitive geospatial solutions by experimenting with and utilizing open-source tools.

Major abilities of the GIS could be understood within following heads:

- Generation of the maps
- Spatial data management and their organizations
- Digitizing and generation of the spatial data
- Analysis of the spatial data
- Sharing of knowledge, tips and tricks from using open-source software.

1.7 Basic Application of Geographic Information System in Recent World

Geographic Information Systems (GIS) have various important applications, and technological advancements have significantly enhanced GIS data, specifically how it can be used and what can be achieved as a result. Geographic Information Systems (GIS) are effective decision-making tools for any business or industry because they allow for the analysis of environmental, demographic and topographic data. Data intelligence derived by GIS applications assists businesses, industries and consumers in making educated decisions.

1.7.1 Application in Cartographic Mapping

GIS can be used to provide a visual representation of data. Google Maps is a great example of a web-based GIS mapping service that people use for navigation. Smart mapping technology, on the other hand, has improved dramatically and is employed in products such as Nobel's GeoViewer, which provides cities, municipalities and private industry with an in-depth look into electric and water district assets in the field.

1.7.2 Application in Telecom and Network Services

Geographic data can be included into sophisticated network design, optimization, planning and maintenance tasks. This information improves telecom procedures by allowing for improved customer relationship management and location services.

1.7.3 Application in Accident Analysis and Hot Spot Analysis

GIS data aids in the identification of accident areas, and data intelligence may be used to optimize road networks. This intelligence aids in the improvement of road safety measures as well as better traffic management.

1.7.4 Application in Urban Planning

GIS data is used to examine urban growth and the direction of expansion. When used correctly, it can uncover new sites for further development while taking into account a variety of elements that are required for effective construction.

1.7.5 Application in Transportation Planning

GIS data is frequently employed in the management of transportation challenges. Companies can prepare for a new road or rail route by incorporating environmental and topical data into a GIS platform.

1.7.6 Application in Environmental Impact Analysis

The information obtained by GIS applications is critical for conserving natural resources and protecting the environment. The magnitude of human impact on the environment is assessed in impact statements, which GIS integration aids in indicating.

1.7.7 Application in Agriculture

GIS data aids in the development of more efficient farming techniques, as well as the enhanced analysis of soil data. This has the potential to enhance food production in various places of the world.

1.7.8 Application in Disaster Management and Mitigation

Efficient GIS systems safeguard the environment and aid in risk and catastrophe management.

1.7.9 Application in Navigation

GIS is very much used for the development of navigation maps which are extremely helpful for the civilians. Continuous updation of the web maps on the basis of GIS data has proven its applicability among public in their everyday life.

1.7.10 Application in Natural Resources Management

GIS is very much able in the properly maintenance and management of the forest data. It is crucial for the allocation and spatial distribution of water through the region.

1.7.11 Application in Banking

Banking has grown into a market-driven industry, and success of the bank is determined by the ability of the bank to provide the customer-centric services. GIS data is critical in the banking business for planning, organizing and making decisions.

1.7.12 Application in Planning and Community Development

GIS data is very much helpful in assisting, understanding and responding to the major global concerns. With the passage of time, growing field of GIS is expanding its scope made it readily available for planners. Such type of geospatial tools has ability to revolutionize the planning phenomenon by providing the accurate and precise real-time data.

1.7.13 Application in Irrigation Water Management

Availability of the water shows a direct impact on the crop yield for a specific region. It is crucial to implement the geospatial analysis for the selection of suitable crops which could give their potential yield in that particular type of region.

1.8 Conclusion

On completion of this chapter, we are not only able to understand the basic concepts of GIS but also able to develop a chronological understanding about historical evolution of GIS, which would help us to understand and established a link with upcoming future changes and modifications. Knowledge of the interaction of GIS with Remote Sensing and GPS will increase the scope of applicability in various field with increased accuracy and efficiency. We understood the application of GIS in various sectors which proved the huge applicability of GIS. Further, it should be noted that there is huge scope of the applicability of GIS in several fields. It is us who can explore in the ocean of GIS as much as we can and make our lives much easier.

Questions

(1) Discuss the history of GIS.
(2) What are the different models in GIS? Explain them in order of their importance.
(3) Briefly describe the different components of GIS.
(4) Explain the importance of remote sensing and GIS in contemporary world. Do you identify any limitation in Remote Sensing-based survey and assessment?

References

Ali E (2020) Geographic information system (GIS): definition, development, applications and components. Department of Geography, Ananda Chandra College, India

Berry JK (1986) Learning computer-assisted map analysis. J Forestry 39–43

Bock Y (1996) Introduction. In: Kleusberg A, Teunissen PJG (eds) GPS for Geodesy. Springer-Verlag, Berlin, Germany, pp 3–36

Bor WT (1994) The use of GPS in GIS applications. J Geogr Sci 17:77–85
Burrough PA (1986) Principles of geographical information systems for land resources assessment. Oxford University Press, N.Y., p 193
Campbell JB (1996) Introduction to remote sensing, 2nd edn. Taylor & Francis, London, U.K., p 622
Chang KT (2019) Introduction to geographic information systems, vol 9. McGraw-Hill, Boston
Clarke KC, Cloud JG (2000) On the origins of analytical cartography. Cartogr Geogr Inf Sci 27(3):195–204
Colwell RN (Ed) (1983) Manual of remote sensing (2nd edn). American Society of Photogrammetry, Falls Church, Virginia, two volumes, 2440
Coppock JT, Rhind DW (1991) The History of GIS. In: Maguire DJ, Goodchild MF, Rhind DW (eds) Geographical information systems. Longman, London, pp 21–43
Cowen DJ (1990) GIS versus CAD versus DBMS: what are the differences? In: Introductory readings in geographic information systems. CRC Press, pp 70–80
Devine HA, Field RC (1986) The gist of GIS. J Forestry 17–22
Foresman TW (ed) (1998) The history of geographic information systems: perspectives from the pioneers. Prentice Hall, Upper Saddle River, NJ
Gao J (2002) Integration of GPS with remote sensing and GIS: reality and prospect. Photogramm Eng Remote Sens 68(5):447–454
Gao J, Leary SMO (1997) Estimation of suspended solids from aerial photographs in a GIS. Int J Remote Sens 18(10):2073–2085
Goodchild ME (1985) Geographic information systems in undergraduate geography: a contemporary dilemma. Oper Geogr 8:34–38
Goodchild MF (1992) Geographical information science. Int J Geogr Inf Syst 6(1):31–45. https://doi.org/10.1080/02693799208901893
Lachapelle G, Cannon ME, Penney DC, Goddard T (1996) GIS/ GPS facilitates precision farming. GIS World 9(7):54–56
Lillesand TM, Kiefer RW (1994) Remote sensing and linage interpretation, 3rd edn. Wiley & Sons, New York, N.Y., p 750
Longley PA, Goodchild MF, Maguire DJ, Rhind DW (2001) Geographic information systems and science. John Wiley & Sons, Chichester, UK
Marble DF, Peuquet DJ (1983) Geographic information systems and remote sensing. In: Manual of remote sensing (2nd ed). American Society for Photogrammetry and Remote Sensing, Falls Church, Virginia.
Mark DM, Chrisman N, Frank AU et al (1997) The GIS history project. UCGIS Summer Assembly, Bar Harbor, ME. http://www.ibrarian.net/navon/paper/The_GIS_History_Project.pdf?paperid=1965613. Accessed on 1 April 2016
McHarg IL (1969) Design with nature. Doubleday, New York
Parker, H.D. (1988) The unique qualities of a geographic information system: a commentary. Photogramm Eng Remote Sens 54(11):1547–1549
Steinitz C, Parker P, Jordan L (1976) Hand-drawn overlays: their history and prospective uses. Landscape Archit 66(5):444–455
Swindell J (1995) A rich harvest: integrating GPS and GIS on the farm. Mapp Awareness 9(1):32–35
Tomlinson RF (1987) Current and potential uses of geographical information systems: the North American experience. Int J Geogr Inf Syst 1(3):203–218
Treitz PM, Howarth PJ, Gong P (1992) Application of satellite and GIS technologies for land-cover and land-use mapping at the rural-urban fringe: a case study. Photogramm Eng Remote Sens 58(4):439–448
Usery EL, Pocknee S, Boydell B (1995) Precision farming data management using geographic information systems. Photogramm Eng Remote Sens 61(11):1383–1391

Waters N (1998) Geographic information systems. Encycl Libr Inf Sci 63:98–125

Waters N (2013) The geographic information science body of knowledge 2.0: toward a new federation of GIS knowledge. In: Arnold O, Spickermann W, Spyratos N, Tanaka Y (eds) Webble technology. Communications in Computer and Information Science 372. Springer, Heidelberg, pp 129–142

Wooden WH (1985) Navstar global positioning system: 1985. In: Proceedings of the first international symposium on precise positioning with the global positioning system, 15–19 April (vol 1). Rockville, Maryland, pp 23–32

Chapter 2
Referencing and Coordinate Systems in GIS

Abstract The fundamental requirement of Geographic Information Systems (GIS) is the use of spatially aligned map layers. If they do not, obvious errors can occur. GIS users are accustomed to working with map features on a plane surface. The map elements correspond to spatial features on the surface of earth. Map features are located using a plane coordinate system represented by *x*, *y*-coordinates, while spatial features are located using geographic coordinate system expressed by longitudes and latitudes. A map projection acts as the bridge between the two coordinate systems. Map projection converts the curved surface of the earth into a plane, yielding a map projection suitable for use with a projected coordinate system. Georeferencing is the process of connecting the internal coordinate system of a digital map or aerial photograph to a system of actual geographic coordinates on the ground. After completion of this chapter, you would be able to understand the following topics:

- Map projection and their types
- Coordinate system
- Concept of geographic coordinate system
- Concept of projected coordinate systems
- Some widely used projections
- Concept of georeferencing and its application in raster and vector data.

Keywords Projection · Coordinate system · Datum · UTM · UPS · SPC · Georeferencing

2.1 Introduction

Map projection attempts to convert the curved surface of earth, or a section of it, into a flat sheet (paper, cloth, etc.) or computer screen. Map projection, in simple terms, converts globe from spherical (3D) form to a planar (2D) form. GIS users are accustomed to working with map features on a plane surface. The map elements correspond to spatial features on the surface of earth. Map features are located using a plane coordinate system represented by *x*, *y*-coordinates, while the spatial features are located using geographic coordinate system expressed by longitudes and latitudes. A map projection acts as the bridge between the two coordinate systems. It converts

M. Kumar et al., *Geographic Information Systems in Urban Planning and Management*, Advances in Geographical and Environmental Sciences,
https://doi.org/10.1007/978-981-19-7855-5_2

the curved surface of the earth into a plane, yielding a map projection suitable for use with a projected coordinate system. The Coordinate Reference System (CRS) then determines the manner in which the two-dimensional, projected map in GIS corresponds to real-world locations. The choice of map projection and CRS to employ is determined by the area extents of the study region, the type of research to be performed, and, in many cases on the data availability.

We often download vector and raster datasets from the Internet for GIS applications. Some digital datasets are referenced in longitude and latitude values, while others are referenced in various projected coordinate systems. It is appropriate to analyse these datasets at beginning of the project so as to avoid incongruities in coordinate system of these datasets. In this context, processing refers to projection and reprojection. Datasets are projected from geographic coordinate system to projected coordinate system, and datasets are reprojected from one projected coordinate system to other. Projection and reprojection are characteristically the initial tasks in a GIS project.

2.2 Map Projection

Globes are a traditional way of modelling of the earth. However, there are certain problems with this strategy. Although these retain most of earth's shape and demonstrate the shape of large geographical features like continents, one cannot conveniently carry a globe. Also, globe is useful only at very small scales (e.g. 1:100 million).

Most of the thematic data utilized in GIS has much bigger scale. Its files typically have a scale of 1:250,000 or higher, with regard to the quantity of information. Globes of such a scale would be difficult as well as costly to manufacture, and considerably difficult to handle and transport. Consequently, cartographers have devised a set of procedures known as map projections that are aimed to depict the spherical world in two dimensions with reasonable accuracy.

When viewed up close, the world appears pretty flat. However, as observed from space, the world seems to be roughly round. Maps are representations of reality, as is explained in the forthcoming topic on map production. In addition to the characteristics of earth's surface features, maps are intended to represent shape and arrangement of features as well. Map projections have their respective benefits and drawbacks. The optimum projection for a map is determined by its scale and the purposes for which it will be used. For example, if used to map the entire continent of Africa, a projection can have intolerable distortions, yet it can be an outstanding option for a large-scale map of a country. Some map design elements may be influenced by the attributes of a map projection. Some projection methods are better to map tiny areas, some to map an area with extensive longitudinal range and some to map an area with a large latitudinal extent.

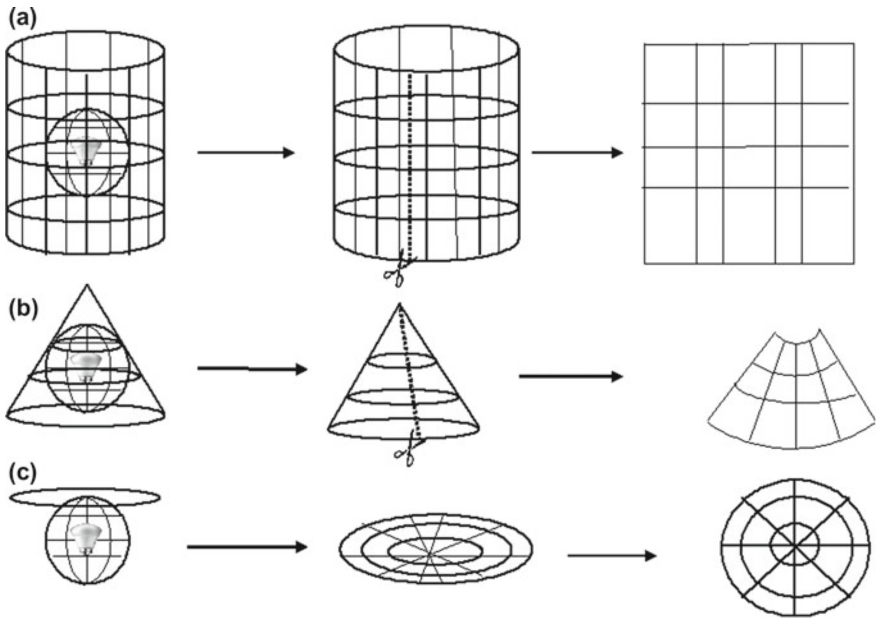


Fig. 2.1 **a** Cylindrical projection; **b** conical projection; **c** planar projection

2.2.1 Types of Projection

Projection techniques could be classified on the basis of following properties.

2.2.1.1 On the Basis of Developable Surface

Map projection is best demonstrated by placing light source inside a transparent globe containing opaque earth features. The feature outlines should then be projected onto a two-dimensional flat sheet of paper. By encircling the globe with developable surface of cylindrical shape, cone shape or planar surface, cylindrical, conical or planer projections can be produced. These are known as map projection families. As a result, there is a cylindrical projection family, a conical projection family and a planar projection family (Fig. 2.1).

2.2.1.2 On the Basis of Drawing Techniques

Projections are divided into three types based on their construction method. These include perspective, non-perspective, and conventional or mathematical projections. Perspective projection is created with the help of a light source which projects image of the globe's network of parallels and meridians onto a suitable developable surface.

Non-perspective projection is created independent of light or the casting of shadows on flattened surfaces. Mathematical or conventional projection is the one that is produced through mathematical formulae and have least relationship to the projected image.

2.2.1.3 On the Basis of Preserved Properties

We know that the precision of shape, area, distance and direction are the main global properties to be preserved during the projection process. However, no projection has the characteristic to preserve all of four properties at the same time. As a result, a projection can be created based on the specific need in order to maintain the desired quality. Thus, projections may be classified as equal-area, orthomorphic, azimuthal and equidistant projections based on the global properties they preserve. Equal-area projection is also known as homolographic projection which preserves the global property of area of the feature that it represents. Orthomorphic or true shape projection accurately depicts the shape of earth's surface features. Shape is largely preserved at the expense of area accuracy. Azimuthal or true bearing projection correctly represent the direction of all points from the centre. Equidistant or true scale projection is one in which the scale is correctly maintained and distances can be calculated accurately. However, no projection can maintain the consistency of the scale throughout. It can only be kept in proper working order along a few specific parallels and meridians.

2.2.1.4 On the Basis of Source of Light

Projections can be categorized as gnomonic, stereographic or orthographic based where the light source is placed in a perspective projection. Gnomonic projection is achieved by placing the light in the globe's centre. If the light source is positioned at a point diametrically opposite to the point at which developable surface tangentially contacts the globe, a stereographic projection is created. When the light source is positioned at infinite distance from the globe, in the direction opposite to the point where developable surface tangentially contacts it, an orthographic projection is drawn. Figure 2.2 is representing it.

2.3 Coordinate System

A coordinate system is a scheme of identifying the position of a point or geometric piece inside a geographic framework with the help of numbers or coordinates. This aids in the integration of geographic datasets by means of common places. Within this framework, a coordinate system is also utilized to denote the location of geographic features, observational sites, GPS points and imaging. Cartesian coordinates are the simplest coordinate system, consisting of perpendicularly oriented coordinate axes.

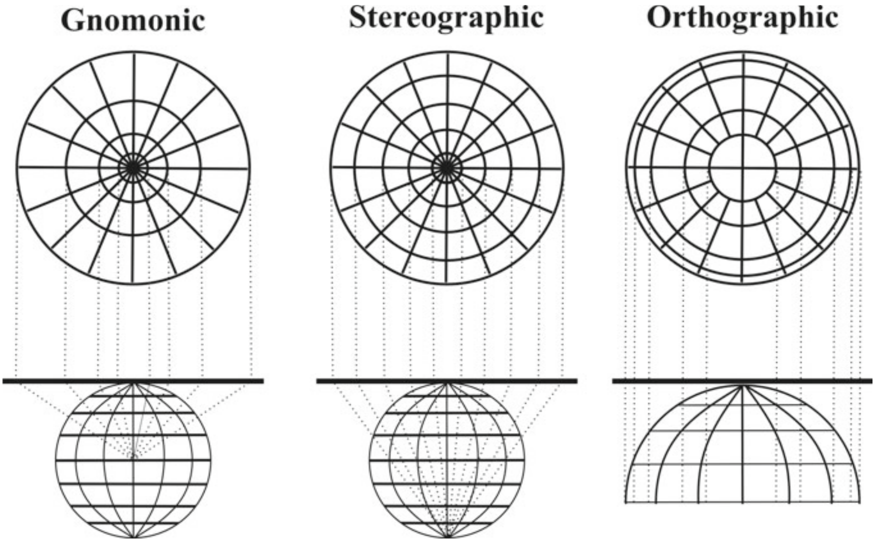


Fig. 2.2 Projection on the basis of light source

A single sort of coordinate system may be better to utilize than others depending on the type of problem under examination.

In mathematics, coordinate system is a technique of assigning an n-tuple of numbers (scalars) to every point in an n-dimensional space. The theory of manifolds includes this concept (Morita 2001). In many contexts, "scalars" refers to real numbers, although it can also refer to complex numbers or elements of another commutative ring, depending on the context. It is practically impossible to give a single consistent coordinate system for entire space in complex spaces. For example, a set of coordinate systems known as graphs is combined to build an atlas that covers the entire space. The surface of the earth is a simple example (which motivates the nomenclature).

Singularities are places where one or more of the coordinates are not well defined in informal usage of coordinate systems. The origin in polar coordinate system (r), on the plane, for example, is singular, while the radial coordinate takes well-defined value ($r = 0$) at origin; it may have any angular value and therefore cannot be a well-defined function.

2.4 Geographic Coordinate System

A geographic coordinate system (GCS) is a referencing system that uses latitudes and longitudes to determine locations on a spheroid or sphere for designation of locations on the earth's three-dimensional surface (Kennedy and Kopp 2000). GCS is frequently referred to just as datum; however, a datum is only one component of GCS; as in addition to datum (spheroid-based), it consists of prime meridian and angular unit of measurement.

To put it in other words, the geographic coordinate system is a standard to find spatial features on the surface of earth in which longitudes and latitudes are used. The longitudes and latitudes are angular measurements as longitudes measure angle eastward or westward of the prime meridian, while latitudes measure angle northward or southward of the equator. The longitude of location X in Fig. 2.3, for example, is angle $a°$ *West*, while the latitude of location Y is the angle $b°$ *North* (Chang 2019).

Major segments of the GCS are discussed below:

(a) Parallels

Parallels are the lines that trace horizontal path on globe from east to west and have fixed latitudinal values. They are equidistant and parallel, forming concentrical rings about the globe. Equator is the largest parallel on globe, dividing it in two halves. It is equidistant from each pole, and its latitudinal value is 0°. In the northern hemisphere, parallels assume positive latitudinal values ranging from 0° to +90°, and in southern hemisphere they assume negative latitudinal values ranging from 0° to −90°.

(b) Meridians

Meridians are the lines that transverse vertically from north to south and have fixed longitudinal values. They form semi-circles round the globe, which connect at poles. Prime meridian (0°) is the longitudinal line that depicts the beginning of longitudes. It is the line that runs via Greenwich, England. Other longitudinal lines including the ones passing through Bern, Bogota and Paris were also designated as prime meridians. Positive longitudinal values ranging from 0° to + 180° are on the east of prime meridian up to the antipodal meridian (180°). On the west of prime meridian, negative longitudinal values ranging from 0° to −180° degrees are found (Fig. 2.4).

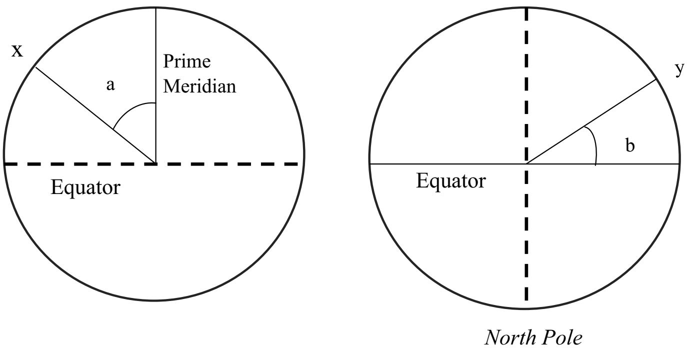


Fig. 2.3 Longitude of location X is '$a°$ West' (Left), and latitude of location Y is '$b°$ North' (Right)

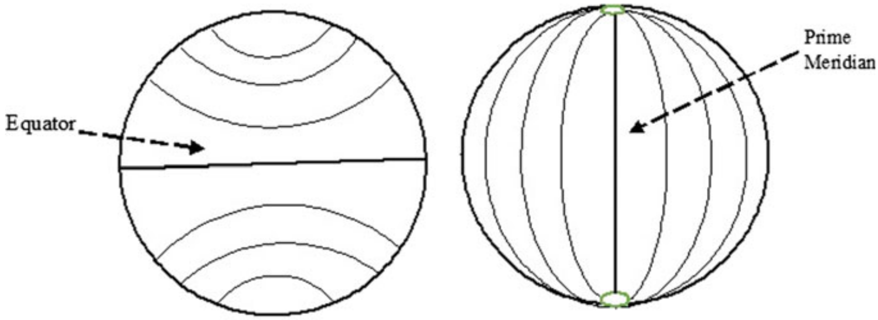


Fig. 2.4 Latitude and longitude lines

(c) Graticule

The latitudes and longitudes are drawn around the earth's surface to generate a grid of lines known as the graticule. Its origin is (0,0), the point of intersection of 0° latitudinal line and 0° longitudinal line. Equator is the sole line of the graticule on which the linear distance equivalent of one latitudinal degree equals to linear distance equivalent of one longitudinal degree. Because longitudinal lines converge at poles, the stretch between adjacent longitudes varies at each parallel. Consequently, as near the poles, the distance equivalent of one latitudinal degree becomes larger than the distance equivalent of one longitudinal degree.

The graticule makes it tough to find out the length of the latitudinal lines which are concentric rings that reduce in radius as one moves from equator to the poles. Latitudes at poles become single points which coincide with the location of intersection of the longitudinal lines. At the equator, distance equivalent of one longitudinal degree is roughly 111.321 kms, but at 60° latitude, distance equivalent of one longitudinal degree is just 55.802 kms (as per Clarke, 1866 spheroid). Consequently, since the distance equivalents of longitudes are not uniform across the latitudes, the distance between points cannot be accurately measured with the help of angular units of measurement.

Figure 2.5 is representing the graticule of the earth:

(d). Sphere and Spheroid Approximations

Coordinate systems are generally described by the sphere or the spheroid approximations of the shape of the planet. Since earth does not have a perfect round shape, spheroid is used to preserve map accuracy depending on where you are located on the surface of the planet. Spheroids are ellipsoids with an ellipse base, whereas a sphere has a circle base.

Two radii determine the shape of the ellipse. The longer radius is referred to as the semi-major axis, while the shorter radius is referred to as the semi-minor axis. A three-dimensional shape created by rotating the ellipse about one of its axes is known as an ellipsoid.

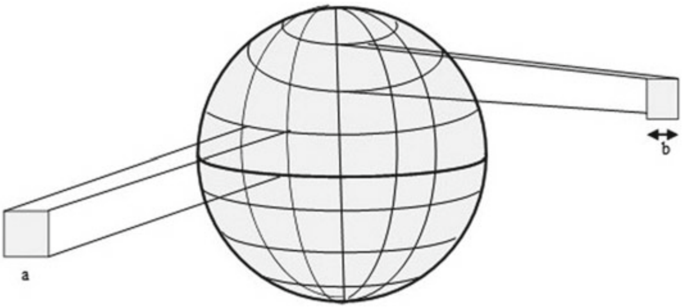


Fig. 2.5 Graticule of the earth **a** distance between one degree of longitude at equator = 111.321 km; **b** distance between longitudes at 60 degree of latitudes = 55.802 km

The sphere, spheroid approximation of the earth, the major axis and minor axis are depicted in Fig. 2.6.

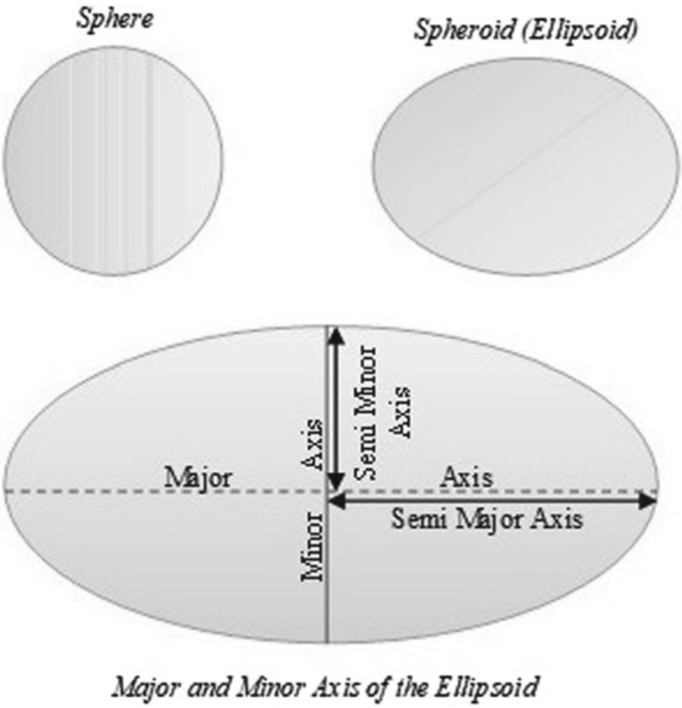


Fig. 2.6 Sphere and spheroid approximation of earth

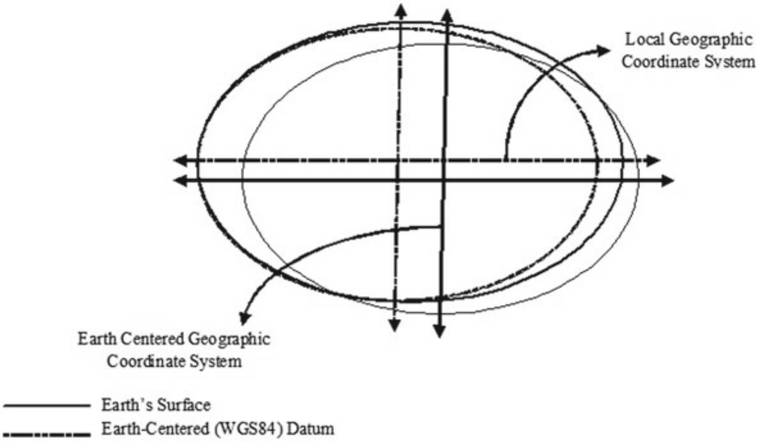


Fig. 2.7 Alignments of datum

(e). Datum

A datum is a set of values that define the spheroid's situation relative to the earth's centre. The datum defines the origin and orientation of latitude and longitude lines and serves as a frame of reference for measuring locations. Some data are global in nature, with the goal of providing decent overall precision all over the world. A local datum's spheroid is aligned to closely appropriate the earth's surface at local level. As a result, the coordinate system measurements will be inaccurate when used with an area other than the one for which they were designed.

Figure 2.7 demonstrates the manner in which various data line up with the surface of earth. At this location, the local datum, NAD27, aligns thoroughly with the surface as compared to the earth-centred datum, WGS84. When the datum is changed, the geographic coordinate system and coordinate values change.

2.5 Projected Coordinate System

A map projection provides a framework for projected coordinate system. Map projections and projected coordinate systems are often used interchangeably. The Lambert conformal conic, for example, a map projection, may denote a coordinate system as well. Nevertheless, projected coordinate system is intended to determine accurate calculation and positioning and is commonly used in map projects of larger scales (1:24,000 or higher).

The accuracy of locating and positioning a feature with respect to other features is thus a critical factor to consider in designing the projected coordinate system. Projected coordinate systems are frequently divided into several zones to maintain the desirable levels of measurement accuracy. The zones are distinguished by a distinct projection centre. Furthermore, projected coordinate systems are defined not merely by parameters of their fundamental map projection, for the parameters of geographic coordinate system (e.g. datum) out of which the map projection is obtained are also important.

The Universal Transverse Mercator (UTM) grid system, the Universal Polar Stereographic (UPS) grid system and the State Plane Coordinate (SPC) system are the three most common coordinate systems.

2.5.1 *The Universal Transverse Mercator (UTM) Grid System*

The UTM grid system, which is used all over the world, splits the surface of earth into 60 zones and is limited in the latitudinal extensions of 84° N to 80° S. Each zone has longitudinal extent of 6° and is designated by sequential nominal digits, with 1st zone starting from 180°W to 174°W. UTM zones are also subdivided into north zone and south zone depending on the two hemispheres of the earth. As a result, the designations of UTM zones include nominal numbers along with the letters denoting the position above or below the equator. Figure 2.8 is representing the UTM grid system.

In UTM grid system, zones are represented by secant case of transverse Mercator, having scale factor of 0.9996 at central meridian, and equator as the origin latitude. The standard meridians are located 180 kms east and west of the central meridian (Fig. 2.9). The use of separate projection for each zone of UTM grid system intends to uphold the accuracy of at least one part in 2500 (i.e. distances measured for 2500 m

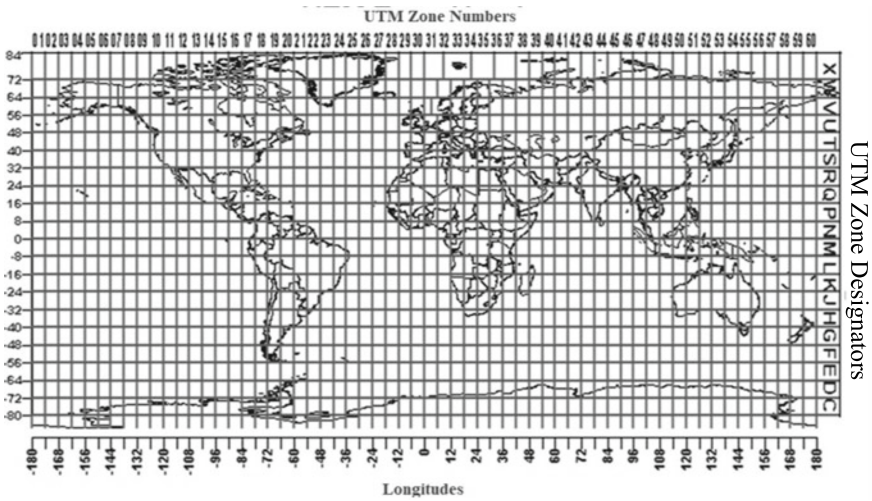


Fig. 2.8 UTM grid system

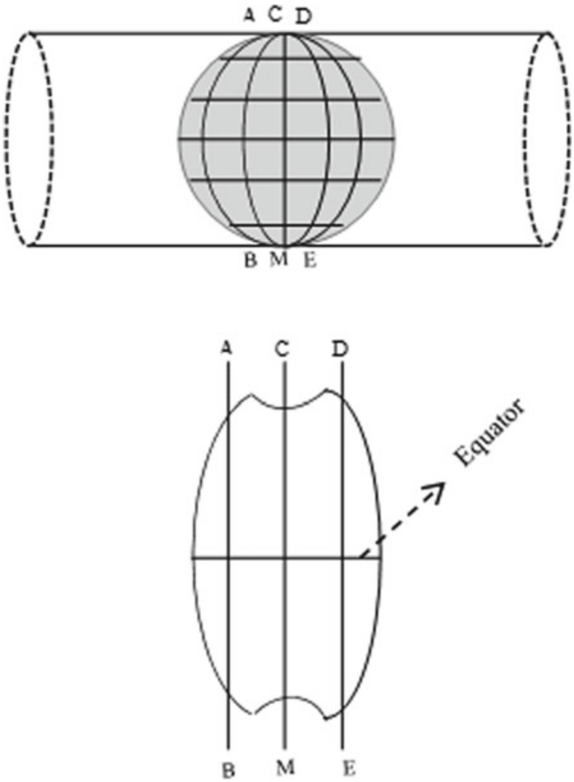


Fig. 2.9 A secant case transverse Mercator projection is represented by a UTM zone. The central meridian is CM, and the standard meridians are AB and DE. The standard meridians are 180 kms to the west and east of the central meridian. Each UTM zone extends from 84° N to 80° S and covers 6° of longitude. For illustration purposes, the size and shape of the UTM zone have been exaggerated

stretch in UTM grid will be accurate within a metre of the true measure) (Kimerling et al. 2011).

UTM coordinates in the northern hemisphere are determined using a false origin at equator and 500,000 m west of central meridian of the UTM zone. Similarly in the southern hemisphere, UTM coordinates are determined using a false origin 10,000,000 m south of equator and 500,000 m west of central meridian of the UTM zone.

Because of the usage of false origin points, UTM coordinates are always positive, but these might be bulky numbers. To maintain data precision when performing coordinate computations, we may employ x-shift and y-shift values to all coordinate readings and lessen the digits in bulky numbers. Major limitation of the UTM coordinate system is that it is not good for small-scale maps and does not extend to North and South Poles.

2.5.2 The Universal Polar Stereographic (UPS) Grid System

To locate position(s) on surface of earth, the universal polar stereographic (UPS) coordinate system is made use of in combination with universal transverse Mercator (UTM) system. UPS, like UTM, employs the metric-based Cartesian grid on a conformally projected surface. It covers Polar Regions of earth, specifically those that lie north of 84° N and south of 80° S, as well as an additional stretch of 30 min of latitudinal extension into the UTM grid.

Directions can become complicated in the Polar Regions, because all lines of graticule converge at the poles. As a result, the angular inclinations between north of the UPS grid and true north may vary up to 180° in some places. As such north of UPS grid becomes true south, and/or vice versa. North of UPS grid is subjectively defined as with respect to the prime meridian in Antarctica and the 180th meridian in the Arctic Ocean; thus, east and west on the grids are defined with respect to 90° E and 90° W meridians when one moves away from the poles.

2.5.3 The State Plane Coordinate (SPC) Grid System

The State Plane Coordinate System (SPCS) is a collection of 124 geographic zones or coordinate systems developed for precise geographic areas of USA. Every state has a single or multiple plane zone(s), boundaries of which typically correspond to county lines. There are 110 zones in the mainland USA, plus 10 in Alaska, 5 in Hawaii, 1 in Puerto Rico and 1 in US Virgin Islands. State and local governments rely heavily on the system for geographic data. Its popularity can be attributed to two main factors. To begin with, it specifies locations using a simple Cartesian coordinate system instead of complex spherical coordinate systems (the GCS of latitudes and longitudes). Plane surveying methods can be executed easily in simple *x, y*-coordinates of Cartesian system, which speeds up and simplifies calculations. Secondly, this system is extremely precise within each and every zone (error is less than 1:10,000). Outside of a specific state zone, the system's precision rapidly drops, rendering it unsuitable for regional or national mapping.

The transverse Mercator projection or the Lambert conformal conic projection is used for a majority of state plane zones. The choice among these two map projection systems is determined by the state's morphology and zones. States with extensive east–west extensions are usually subdivided into two zones which can be covered by standard meridians of conical projection. Because the projection cone touches the globe along two latitudinal lines, the Lambert conformal conic projection is suitable for upholding precision along an east–west stretch. As the circumference of the projection cylinder is aligned along the longitudinal meridian, zones with long north–south extensions may be suitably covered by transverse Mercator projection. The Alaska panhandle uses an oblique Mercator projection, which reduces the cumulative error along the *x, y* axes because its major dimensions are on the diagonal.

2.6 Widely Used Projections

Some most commonly projections are listed below:

2.6.1 Azimuthal Projection—Stereographic

Ptolemy, around 150 AD, is the first documented account of this projection. However, it is thought that this projection was generally known much earlier, possibly as early as the second century BC.

It is undoubtedly one of the most common azimuthal projections, especially for Polar Regions, although it may also be employed for small-scale mapping of continents like Australia. The projection's main draw is that the earth looks as if seen from space. Shapes are well retained across the map with this conformal projection, while significant distortions occur towards the map's edge. The map is not equal area, but the directions are true from centre of the map (contact point of our hypothetical sheet of paper).

A straight line that passes via the centre point is a Great Circle, which is an intriguing characteristic of the stereographic projection. The benefit of this is that for a point of interest (e.g. Canberra, Australia's capital city), actual distances to other points of interest may be determined using a map that employs the Stereographic projection and is centred on that point of interest (e.g. Canberra to Wellington, New Zealand). Figure 2.10 is representing it.

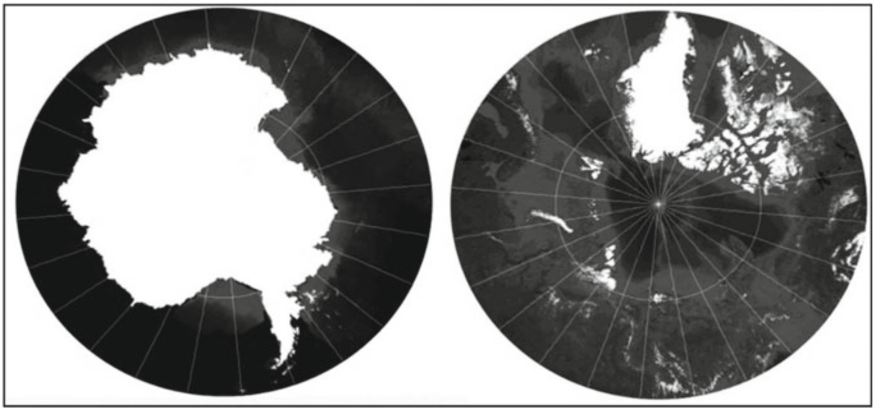


Fig. 2.10 Stereographic projection

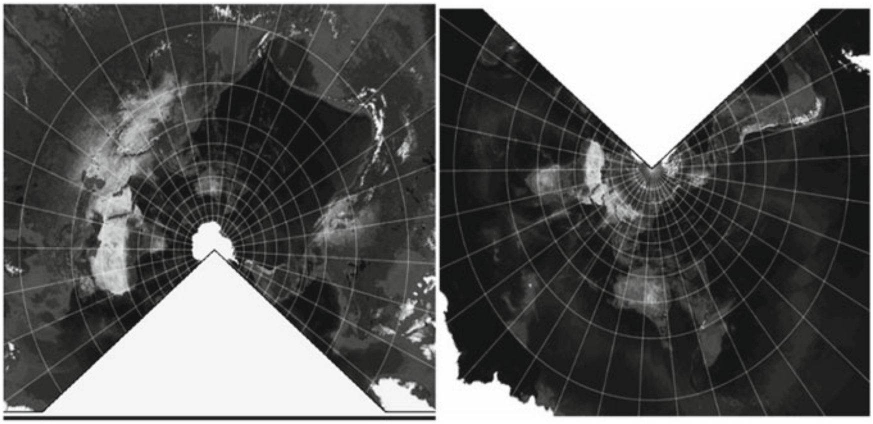


Fig. 2.11 Conic projection

2.6.2 Conic Projection—Lambert Conformal Conic

Johann Heinrich Lambert, mathematician and physicist, was born in Germany and lived in France, and his mathematics was innovative at the time. It continues to be significant now. Both his conformal conic and transverse Mercator projections were published in 1772.

Today, the Lambert conformal conic projection is used to map huge areas (on a smaller scale) in the mid-latitudes, such as USA, Europe, Australia, etc. Aeronautical charts, like the 1:100,000 scale World Aeronautical Charts map series, have grown particularly popular.

Two Standard Parallels were often employed in this projection (latitudinal lines which are concentric circles).

Shapes are retained for significant distances near Standard Parallels, indicating that the projection is conformal. The shapes of world maps are wildly warped away from Standard Parallels. This is why regional maps of mid-latitude regions (about 20° to 60° N/S) are so popular. Distances are true only along Standard Parallels in this projection. Directions are mostly correct throughout the map. This is represented in Fig. 2.11.

2.6.3 Cylindrical Projection—Mercator

The Mercator projection, invented in 1569 by a Flemish cartographer and geographer named Geradus Mercator, is one of the most well-known map projections. Because it has the potential to depict lines of continuously true directions, it turns out to be the standard projection for nautical expeditions. A continuously true direction is one in which the straight line joining any two points on the map is same direction in the compass. This was a critical aspect of this projection, when marine expeditions and navigation relied solely on direction.

The Equator is always the Standard Parallel in the Mercator projection. A globe map is always rectangular for the longitudinal and latitudinal lines intersect at right angles to each other in its formation. In addition, the longitude lines are evenly spaced. However, as one moves away from the Equator, the space between the lines of latitude grows. This relationship is what allows for a consistent true direction between any two spots on the map.

While this relationship between latitude and longitude lines preserves accurate orientation, it allows for distortion of regions, forms and distances. There is negligible distortion at the Equator. Distances are always correct along the Equator, but not anywhere else on the map. The areas and shapes are well conserved between 15° N and S. The area and the shape are reasonably well-preserved further outside to nearly 50° N and S. As a result, the Mercator projection is only advised for usage in the Equatorial zone for purposes other than sea navigation. Despite these distortions, the Mercator projection is typically considered to be conformal. Because shapes are essentially true within small areas, this is the case. Figure 2.12 is representing the Mercator cylindrical projection.

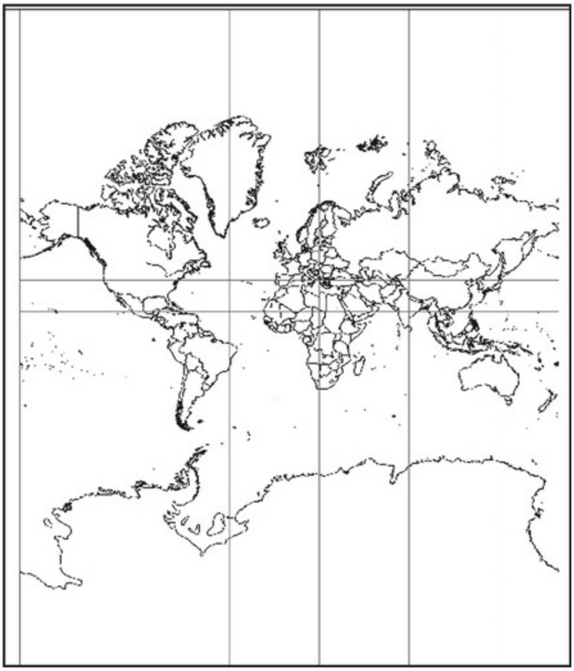


Fig. 2.12 Cylindrical Mercator projection

2.6.4 Cylindrical Projection—Robinson

Arthur H. Robinson, a Wisconsin geography professor, invented a projection for global maps in the 1960s that has grown far more widespread as compared to Mercator projection. It was designed because modern cartographers were dissatisfied with the Mercator projection's distortions and desired a global projection that resembled reality. Robinson projection took over from Mercator projection as the favourite projection for world maps at the time. Rand McNally and National Geographic are two major publishing organizations that employ Robinson projection. The Equator is its Standard Parallel for it is a pseudo-cylindrical projection. It, however, suffers from the same distortion problems as Mercator projection.

The area and the shape are well-preserved for 0° to 15° N/S extents. But the permissible distortion range has been increased from 15° N/S to 45° N/S. In addition, the Polar zones have less distortion. The lines of height and longitude in the Robinson projection are evenly distributed across the map, unlike the Mercator projection. Also, major distinction from Mercator is that just the longitudinal line in the map's centre (Central Meridian) is straight, while the rest are bent, with the curvature increasing as the distance from the Central Meridian rises.

The Robinson projection trades distortions for a more pleasant appearance. It is not conformal, equal area, equidistant or azimuthal. This is represented in Fig. 2.13.

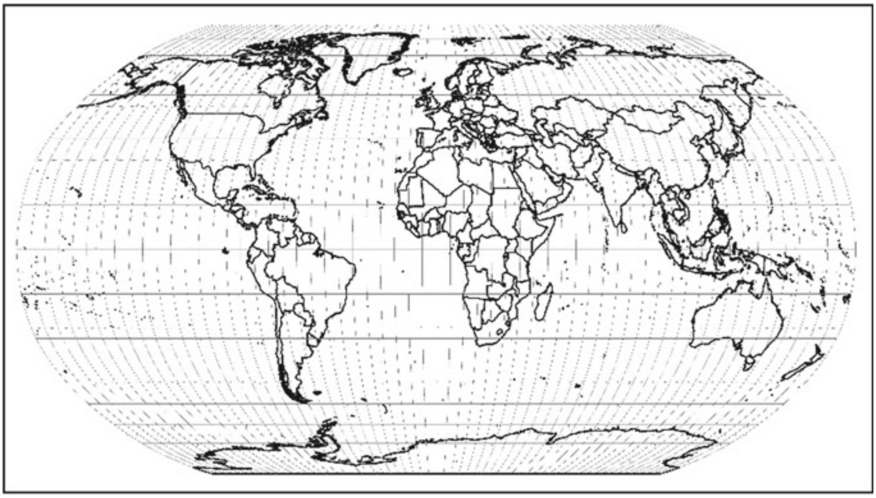


Fig. 2.13 Robinson projection

2.6.5 Cylindrical Projection—Transverse Mercator

Johann Heinrich Lambert, mathematician and physicist, was born in Germany and lived in France, and his mathematics was innovative at the time. It continues to be significant now. Both his conformal conic and transverse Mercator projections were published in 1772.

The Transverse Mercator projection is developed from Mercator projection as base, which has been extremely successful. The Mercator projection's key strength is that its extreme precision at the Equator (the tangential region for our imaginary piece of paper—also known as the Standard Parallel), whereas the projection's chief flaw is the amplification of distortion away from the Equator. Due to these advantages and disadvantages, Mercator projection is well suited to mapping regions with an east–west extents near the Equator, and not for mapping regions with larger north–south extension (e.g. South America).

Instead of contacting at the Equator, Lambert's stroke of inspiration was to have the fictitious piece of paper touch any longitudinal line. On map, it is known as the Central Meridian. This allowed for the creation of accurate maps of sites with a north–south orientation. The mapmaker just had to choose a Central Meridian that ran through the map's centre. Following figure is depicting this projection. This projection is represented in Fig. 2.14.

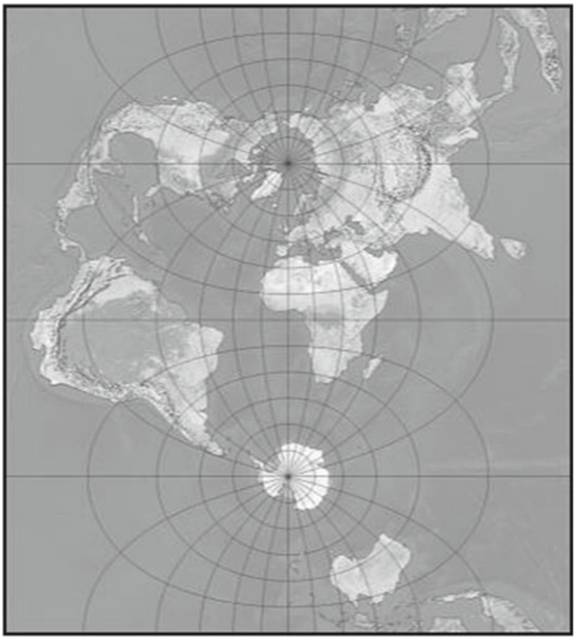


Fig. 2.14 Transverse Mercator projection

2.7 Georeferencing

The coordinates of a digital map or aerial photograph can be registered to the on-ground geographic coordinates. This is called georeferencing. On-ground location of every feature of the georeferenced map or aerial photograph can be determined by linking it to a predefined geographical coordinate system.

Georeferencing is a common term for a variety of procedures involved in identification of unique geographical objects. Broadly speaking, the phrase geographical object denotes any object or structure that can be reasonably linked a location in geographic space, as in points of object (POI), road, places, bridge, building or agricultural areas. Geographic location indicates a point in the space. Geographical places may be demarcated in several spatial dimensions, like zero-dimensional locations, one-dimensional linear features, two-dimensional area features and rare three-dimensional bodies. POI, for instance, may be mapped in dimensionless point location, while roads may be mapped to linear stretch of locations. A building is generally registered in two-dimensional space, independent of their portrayal as three-dimensional features.

Hill (2009) differentiates among informal and formal ways of specifying places. Informal georeferencing includes domain of referencing precise location in science and technology, such as employing spatial reference systems, whereas formal georeferencing involves the domain of colloquial references to spatial objects as in name of place. We shall investigate formal georeferencing approaches throughout this work.

The term georeferencing can be described in a variety of ways. Sommer and Wade, for example, define georeferencing as "the process of aligning geographic data to a known coordinate system so that it can be displayed, queried, and analyzed with other geographic data" (Sommer and Wade 2006). The examination of ways for identifying spatial objects in general (Zheng et al. 2011) and the domains of georeferencing multimedia (Zheng et al. 2011) are examples of coarser viewpoints.

In general, there are three forms of information that can be made use of in referencing spatial objects.

(1) Geometrical information defines an object's geometric qualities, such as a road segment's layout and shape.
(2) In georeferencing, the graphical structure produced by network of particular spatial features, like roads, rivers, etc. is of key relevance for topological inquiries.
(3) Semantic information refers to several semantic features that can be linked to a geographic location, like a place name or road name. For exclusively identifying a spatial object, common georeferencing procedures contemplate at least one category, but generally a mixture of these forms is considered.

Matching is the process of identifying spatial items and assigning these to certain locations. While creating a map from ground up, the mapmaker identifies spatial items, locations and relationships among them. Georeferencing is commonly done with the geodetic reference systems like WGS-84 (National Imagery and Mapping

Agency 2000) that consist of standard earth coordinate frame, reference ellipsoid and geoid that can establish a nominal sea-level. Additional referencing systems, dependent on their utility, may be used. Navigation devices may be made use of in data collection process to some extent. GPS tracker provides WGS-84 geodetic reference system coordinates that may be linked to specific geographic locations on the map. However, assigning things to places is frequently a time-consuming operation. From raster datasets like aerial photographs or satellite images to entities on pre-authored vector datasets, there are many different sorts of geographical items to match on a map. Various procedures and algorithms have grown as a function of the type of matching needed.

An assignment between geographical objects and geographical places is the result of a matching process. Frequently, the object referred to on a map comes from a different map having its own reference system. In such a situation, matching yields correspondence between the two and expresses similarities as well as differences, i.e. the information about which items may be detected on both maps and which could not (or only partially).

Conflation is the process of compounding map data from numerous sources to create a fresh map. Geographic Information Systems, Longley et al. (2005) defines conflation as the act of merging geographic information from overlapping sources so as to retain accurate data, remove redundancy and reconcile data clashes. Conflation is done by finding correspondence between maps with the help of matching algorithm(s) for discrete structures (nodes, segments, edges, etc.) to correlate and combine the data into single fresh map. Yuan and Tao (1999) suggested a classification system that divides conflation into horizontal (uniting neighboring areas) and vertical conflation (uniting multiple maps of same region). Horizontal conflation, on one hand, is most commonly executed on vector maps, vertical conflation, on the other hand, can be executed using raster and vector maps (Zhang 2009).

2.7.1 Georeferencing of Raster Images

Assigning specific objects of raster images to location references is a frequent problem in the georeferencing. It can be accomplished by combining procedures and techniques from a variety of disciplines, including remote sensing, photogrammetry, as well as computer visualization, image processing, pattern recognition, etc.

Raster images of earth's surface are frequently acquired using satellites, aerial vehicles or ground-based platforms. Raw image is preprocessed and ortho-rectified (geometric correction) for removing distortions caused by sensor tilt, topography or lens' optical properties. The image is then divided into numerous segments using segmentation algorithm(s) like edge detection (Lindeberg 2001), region growing (Adams and Bischof 1994), split and merge (Douglas and Peucker 1973), etc. Identified segments are processed further using feature extraction, which employs dimensionality reduction to aid in the segregation of spatial features of the image

(Guyon and Elisseeff 2003). Pixels can thence be identified and classified as roads, agriculture, etc. using a suitable classifier for the particular features.

Raster images are georeferenced by connecting them to digital maps using ground control points (GCPs) (De Leeuw et al. 1988). GCPs are recognized ground locations, chosen for easy identification in satellite imagery or aerial photographs. These can be road crossings or man-made target features. Because ortho-rectified images have an unfluctuating scale, it is enough to associate at least 3 GCPs with pixels in an image that has been classified appropriately. Rest of the pixels, as well as additional classified objects, can then be georeferenced to positions on the map. GCPs can be pre- or post-marked; premarking involves spotting GCPs with targets prior to flight, whereas post-marking involves selecting GCPs after the flight using natural features of image that are deemed suitable for the purpose (Army Corps of Engineers of the United States, 2002). Distinctive non-natural or natural target features are appropriate, but it is to be taken care of the fact that whether a GCP is used for horizontal or vertical referencing. Premarking provides a broad method for identifying locations in comparatively monotonous geographical space, whereas post-marking allows a GCP to be placed in the best possible location.

2.7.2 Georeferencing of Vector Images

Georeferencing vector data involves determination of association between the vector and the corresponding element over a digital map. In its most fundamental form, a pair of coordinates must be allocated to a specific map. This is called as positioning or map-matching. The given coordinate pairs are generally subject to varying degrees of imprecision, e.g. when obtained from consumer-grade GPS instrumentation. In order to match coordinates on a map, certain filtering, rectification and validation must be implemented. A map-matcher working on the navigation device used to navigate streets, for example, must adhere to the constraints that a valid location can only be located over road segments.

Location referencing refers to the process of georeferencing multipart vector data that can also have a spatial extent, as in linear features, area features, road networks, to a vector map. To illustrate this, the terminology point location stands for POI such as bus-stop, whereas linear feature stands for a series of connected point locations like that comprising a route.

The referencing of locations could be done by static or dynamic means (Schneebauer and Wartenberg 2007). Static means use predefined location codes or lookup tables to exclusively recognize locations on a map, like European Traffic Message Channel standard (TMC) (ISO 14819-1 2003) or the Japanese Vehicle Information and Communication System standard (VICS) (Yamada 1996). Dynamic location referencing techniques, on the other hand, such as ISO 17572-3/AGORA-C (Wevers and Hendriks 2005), Xi's improved AGORA-C variant (Xi et al. 2008), OpenLR (TomTom International B.V 2012), TPEG-LOC (CEN ISO/TS 2006), MEILIN (Wartenberg 2008) or TPEG-LOC2/ULR (Fraunhofer). Thus, locations encoded using dynamic means can be decoded and referenced on a map that is topologically and geometrically similar to the map that was made use of in encoding process.

2.8 Conclusion

We discussed the concept of map projection and how it is important. Understanding of coordinate systems would lead to create a conceptual horizon to grasp up various important coordinate systems. We understood the concept of latitudes, longitudes, datum and graticule under geographical coordinate systems. We also understood the concept of projected coordinate systems and three important types of the projected coordinate systems, i.e. Universal Transverse Mercator (UTM), Universal Polar Stereographic (UPS) and Universal Polar Stereographic (UPS). We also discussed the concept of georeferencing and how it is done for raster and vector data.

Detailed understanding of these all concepts would help us to create the foundation for the better understanding of upcoming chapters and working of GIS.

Questions

1. Define the Map Projection and their different types. Discuss the importance of Map Projection also.
2. Differentiate between geographic coordinate system and projected coordinate system.
3. What do you understand by "Datum"? Explain with suitable example.
4. Elucidate the fundamental difference between conical and cylindrical projection.
5. What do you understand with 'Georeferencing'? Why 'Georeferencing' is important in GIS?

References

Adams R, Bischof L (1994) Seeded region growing. IEEE Trans Pattern Anal Mach Intell 16(6):641–647

Chang KT (2019) Introduction to geographic information systems, vol 4. McGraw-Hill, Boston

De Leeuw AJ, Veugen LMM, Van Stokkom HTC (1988) Geometric correction of remotely-sensed imagery using ground control points and orthogonal polynomials. Int J Remote Sens 9(10–11):1751–1759

Douglas DH, Peucker TK (1973) Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartographica Int J Geogr Inf Geovisualization 10(2):112–122

Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3(Mar):1157–1182

Hill LL (2009) Georeferencing: the geographic associations of information. Mit Press

http://ecoursesonline.iasri.res.in/mod/page/view.php?id=2075

http://wiki.gis.com/wiki/index.php/Coordinate_system#cite_ref-1

https://desktop.arcgis.com/en/arcmap/latest/map/projections/transverse-mercator.htm

https://mathworld.wolfram.com/CoordinateSystem.html
https://ncert.nic.in/textbook/pdf/kegy304.pdf
https://www.icsm.gov.au/education/fundamentals-mapping/projections/commonly-used-map-projections
https://www.tandfonline.com/doi/full/https://doi.org/10.1080/19475683.2013.868826
Kennedy M, Kopp S (2000) Understanding map projections. Esri, Redlands, CA
Kimerling AJ, Muehrcke JO, Buckley AR, Muehrcke PC (2011) Map use: reading and analysis. Esri Press
Lindeberg T (2001) Edge detection. In: Hazewinkel M (ed) Encyclopedia of mathematics. Springer, Berlin
Longley PA, Goodchild MF, Maguire DJ, Rhind DW (2005) Geographic information systems and science. John Wiley & Sons
Morita S (2001) Geometry of differential forms (No. 201). American Mathematical Soc
Schneebauer C, Wartenberg M (2007) On-the-fly location referencing methods for establishing traffic information services. IEEE Aerosp Electron Syst Mag 22(2):14–21
Sommer S, Wade T (2006) A to Z GIS: an illustrated dictionary of geographic information systems (2nd ed, vol 89). Esri Press, Redlands
Wartenberg M (2008) Mathematical Methods for Location Referencing. Shaker
Wevers K, Hendriks T (2005) AGORA-C on-the-fly location referencing. In: 12th world congress on intelligent transport systems. San Francisco, CA
Xi D, Weifeng L, Tongyu Z (2008) Improved dynamic location reference method agora-C based on rule optimization. In: International conference on computer science and software engineering, Wuhan, December 12–14
Yamada S (1996) The strategy and deployment plan for VICS. IEEE Commun Mag 34(10):94–97
Yuan S, Tao C (1999) Development of conflation components. Proc Geoinformatics 99:1–13
Zhang M (2009) Methods and implementations of road-network matching. Doctoral dissertation, Technische Universität München
Zheng YT, Zha ZJ, Chua TS (2011) Research and applications on georeferenced multimedia: a survey. Multimedia Tools Appl 51(1):77–98

Chapter 3
GIS Data Models

Abstract This chapter is concerned with the understanding of the raster and the vector data models of GIS and their data structure. After reading this chapter, you will understand the following:

- The concept and structure of the raster data model.
- Raster data encoding techniques and compression of the raster data.
- A detailed account of the Landsat imagery.
- The concept and structure of the vector data model.
- Spaghetti and topological vector data models, with a special focus on the fundamental percepts of topology.
- Advantages and disadvantages of the two GIS data models.

Keywords Raster data model · Vector data model · Data structure · Vectorization and rasterization

3.1 Introduction

A data model is a collection of guidelines, rules and/or constructs that are used to describe, convert and represent real-world geographical phenomena (field and objects) in the digital world of computers. It is a conceptual understanding of how geospatial data and associated attributes are represented as digitally and logically linked spatial objects consisting of geometry and attributes. In general, the geometry is represented by a geometric-topological structure, while the attribute data is maintained by a thematic or semantic structure. In GIS, the geometric data is represented by two major model types, i.e. the *raster data model* and *the vector data model*. Figure 3.1 shows the representation of the spatial data in a raster (*grey* pixels in the grid of cells) and vector (*black* points, lines and polygon) formats.

M. Kumar et al., *Geographic Information Systems in Urban Planning and Management*, Advances in Geographical and Environmental Sciences,
https://doi.org/10.1007/978-981-19-7855-5_3

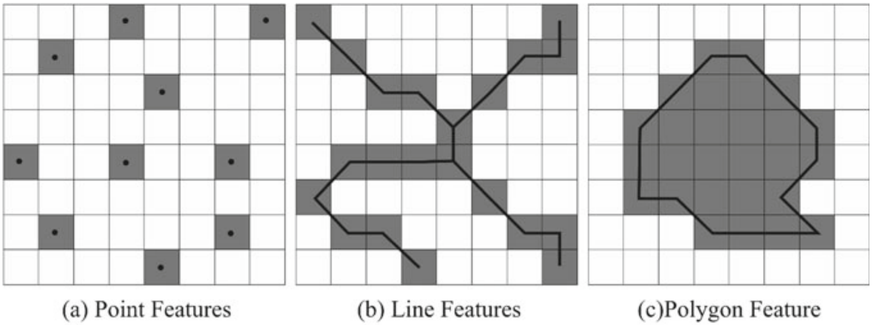


Fig. 3.1 Spatial data representation in GIS

3.2 Raster Data Model

A raster is a collection of contagious cells that are regularly spaced and have associated cell values. The raster data model employs a raster grid to cover the space in which each grid cell represents a particular aspect of a spatial phenomenon at the location of the grid cell. The changes in the values of cells depict the spatial variation of the geographic phenomena represented in the raster data model (Chang 2018). Since it is useful for representing continuous geographical phenomena or features like altitude, rainfall, slope, etc., the raster data model is also named as field data model.

Raster data is the most dominant data type in the GIS. Commonly used raster data products comprise the digital elevation model (DEM), digital surface model (DSM), digital terrain model (DTM), satellite imagery, digital orthophotos and other graphics.

3.2.1 Components of Raster Data Model

Raster, often called an image or grid, structurally consists of cells that represent continuous surface phenomena or features. The geographical feature representation in the form of pixels is such that a point feature is denoted by a single pixel, a line feature has a sequence of adjacent pixels, and an area feature is depicted by a grid of contagious pixels. The cells are called pixels in geoinformatics. For analysis and storage purposes, a raster is befittingly supposed to have rows and columns of pixels, and each pixel in the raster is demarcated by its row and column number, as rows function as *y-coordinate* and columns function as *x-coordinate.*

The *pixel values* in the raster data may be categorical (as in the land use raster) or numerical (as in the rainfall raster). In a categorical raster, the pixel values do not have decimal values and contain only integer values. On the other hand, a numerical raster contains decimal values. Accordingly, a categorical raster is called an *integer* *raster*, while a numerical raster is called a *floating-point raster*. The *pixel size* of a raster is defined as the extent of the actual/on-ground area covered by one pixel of the image. It represents the spatial resolution of image. To illustrate, the ALOS PALSAR digital elevation model shown in Fig. 3.2a has a pixel size of 12.5 m^2, which indicates that each pixel in this raster has a length and breadth of 12.5 m, and the ground area depicted by one pixel in this raster is 156.25 m^2. The satellite image in Fig. 3.2b, on the other hand, is a Landsat 5 image with a cell size of 30 m^2, i.e. each pixel represents 900 m^2 of ground area. The digital elevation model has fine spatial resolution while the satellite image has a coarse spatial resolution. *Pixel depth* is the number of bits that are required to store the pixel values. Since the *bit* is the smallest data unit in the binary language system of computers, it is convenient to understand pixels according to this very unit. The raster can have a *2-bit*, *8-bit*, *16-bit* and so on, pixel depth. The stretch of values that an *n-bit* raster pixel can store is calculated as the *nth* power of *2*. Illustratively, a *2-bit* raster pixel can store $2^2 = 4$ different pixel values only, *an 8-bit* raster pixel can store $2^8 = 256$ different pixel values, and a 16-bit raster pixel can store $2^{16} = 65{,}536$ different pixel values. The higher the pixel depth, the wider the stretch of values that a pixel can store, and the lower the pixel depth, the pixel can store a narrower stretch of values only. The greater the range of values that a pixel can store, the greater the ability of the raster data product to represent speckled information. The raster images may be composed of *single* or *multiple bands*. In case of single band rasters (like elevation), one pixel of the image has only one cell value, while in case of multi-band rasters (like satellite imagery), one pixel of the image has numerous cell values associated with it. The raster images must have a proper spatial reference for processing and analysis in the geographic information system.

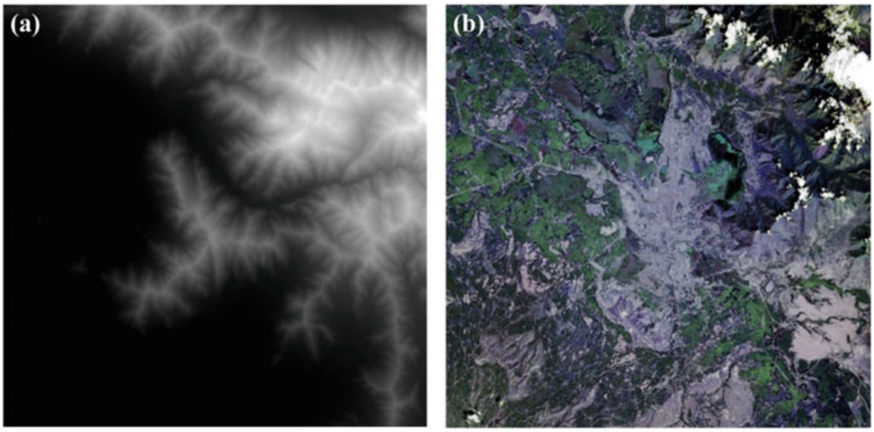


Fig. 3.2 Raster data products: **a** digital elevation model (DEM) and **b** satellite image in natural colour

3.2.2 Raster Data Structure and Data Compression

The data encoding scheme used to store raster data in a computer system is termed as the raster data structure. Cell-by-cell encoding, run-length encoding and quadtree are the most popular schemes of raster data structure. In *cell-by-cell encoding*, the raster is stored in the form of a matrix, cell values of which are coded into a file by rows and columns as shown in Fig. 3.3a. It is the most straightforward raster data structure. When the pixel values in the raster change abruptly, cell-by-cell encoding is preferable. This is applicable in the case of DEMs in which the elevation values change between adjacent pixels continuously. Single-band satellite imagery also uses cell-by-cell encoding. However, the multiple band satellite imagery requires additional handling in band sequential (.bsq) format, band interleaved by line (.bil) format and band interleaved by pixel (.bip) format. The problem with this approach is that it becomes inefficient when used to store a raster with recurring pixel values, such as a geology raster.

The *run-length encoding* approach groups neighbouring pixels with the same pixel value and stores them in the file as rows. A *run* is a collection of pixels having identical values in a single row. Alternatively, a *run* can be defined as the distance between the beginning and finishing pixel of the same value in a row. This is illustrated in Fig. 3.3b. The benefit of run-length encoding is that it allows for the storage of raster data with repetitive pixel values in a relatively minimal amount of storage space. The *quadtree approach* is a more efficacious way of structuring raster data, in which a hierarchy of quadrants is developed by constantly dividing the raster until each quadrant has only one distinct pixel value. A quadtree contains *nodes* and *branches*. Nodes symbolize the quadrants. Depending upon the pixel value(s) inside the quadrant, a node shall be described as *leaf node* (if all pixels inside it have the same value) or a *non-leaf node* (if pixels inside it contain different values). While the non-leaf node can be divided further, the leaf node is the endpoint and must be codified with the homogenized value of the quadrant. Codification is undertaken after the raster subdivisions are completed. Quadtree and spatial indexing techniques are used to carry out the codification. This is illustrated in Fig. 3.4. The raster shown on the left is divided into quadrants until the leaf nodes are reached in all the quadrants. Then quadtree codification of the raster is done starting from the NW and ending at the NE quadrant. At the first level, the raster is divided into four non-leaf nodes. At the second level, the quadtree diagram finds three grey leaf nodes, three white leaf nodes and ten non-leaf nodes. At the third level (last in this case), 16 grey nodes and 24 white nodes are deduced.

Before the raster data (DEM or satellite image) can be imported and used, GIS software would first receive statistics about the raster, such as its data structure, area extent, cell size, band count and no-data values. Usually, this information is stored as a *header file*. Additionally, the raster dataset may contain other supporting files.

A reduction in the volume of data is called *data compression*. Data compression is related to the raster data structure. Data compression is very important for data exchanging, web mapping, etc. which requires data to be of limited operable size.

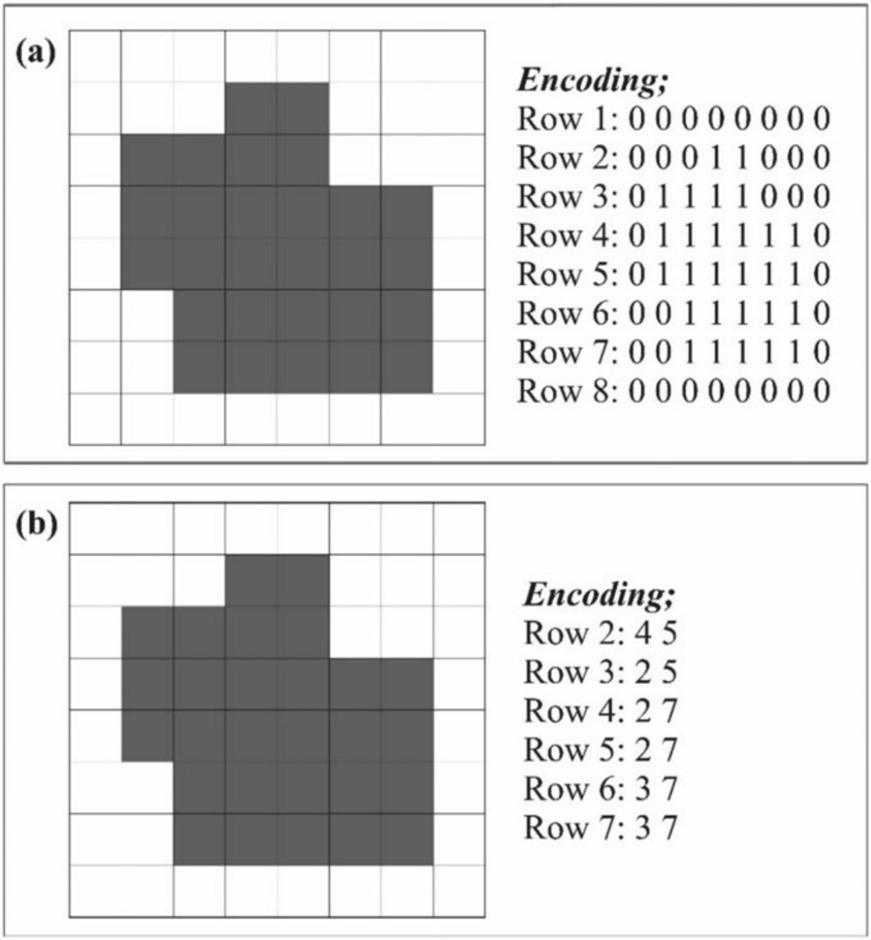


Fig. 3.3 Raster data structure: **a** cell-by-cell encoding and **b** run-length encoding

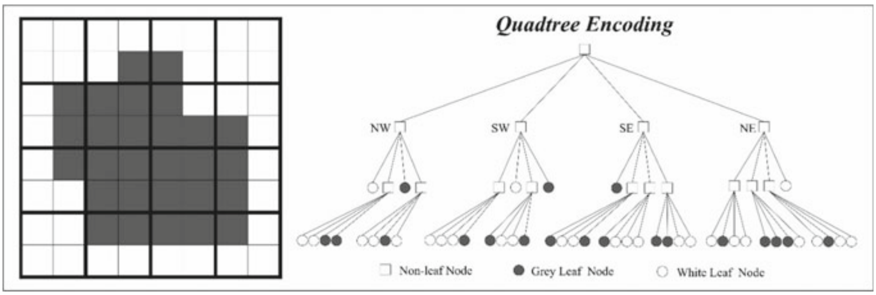


Fig. 3.4 Quadtree raster data structure

The *run-length* and the *quadtree* encoding are also understood as data compression techniques, in addition to raster data structuring approaches. A variety of other data compression techniques are used. These can be classified as *lossless compression* (when the pixel values are preserved such that the original raster can be reconstructed if needed) and *lossy compression* (when pixel values get modified and the initial raster cannot be reconstructed completely). The lossy compression yields a higher compression ratio than that of the lossless compression. The lossless compression is suitable for the raster data to be analysed, while the lossy compression can be adopted for compressing background images that are not part of the actual analysis. Examples of lossless compression are *run-length* encoding and *Lempel–Ziv–Welch.*

3.2.3 Important Raster Data Products

Raster data is the most often utilized type of data in geoinformatics. The digital elevation model, satellite imagery, digital orthophotos, bi-level scanned files, digital raster graphics, land cover data and other graphic files are forms of raster data products. In this section, we shall limit our discussion to the definition and construction of digital elevation models and satellite imagery (Landsat) only.

The ***digital elevation model (DEM)*** is a representation of elevation of the topographic surface of bare earth. It is an arrangement of pixels that contain the elevation values of the area they represent. Traditionally, contour line interpolation and stereo-plotting of aerial photographs, etc. were used to create the DEMs. Of late the development of DEMs from remote sensing data has gained more attention and acceptability. DEM can be developed from a pair of, or more, optical sensor images of a similar area taken from distinct directions. ASTER DEM (*Res. 30 m*) and SPOT 5 (*Res. 20 m*) DEM are examples of optical sensor-derived DEM, which is constructed from the Terra ASTER images and SPOT 5 images, respectively.

InSAR creates the digital elevation models from a pair of, or more, synthetic aperture radar (SAR) images. It builds upon the reflective surfaces like vegetation, bare-ground and land use features to generate the DEM. Shuttle Radar Topography Mission (SRTM) DEM is produced using satellite-based SAR data, and it covers nearly 80–85% earth's land surface. InSAR DEMs of different resolutions are available, and lately, the spatial resolution of these DEMs has fined out to up to 1 m (derived from TerraSAR-X and distributed by Airbus Defence and Space). Advanced Land Observing Satellite-Phased Array type L-band Synthetic Aperture Radar (ALOS-PALSAR) is an L-band Synthetic Aperture Radar (SAR) that generates the digital elevation models in all weather and operates day and night. Its DEM (*Res. 12.5 m*), generally referred to as Radiometric Terrain Correction (RTC) products, is distributed by the Alaska Satellite Facility (ASF), a Distributed Active Archive Center (DAAC) of NASA.

LiDAR, by its principle, measures the distance between the active sensor and the target as a function of the time interval taken by the laser pulse in the two-way travelling. The GPS and Inertial Measurement Unit (IMU) of LiDAR determine the location and orientation of the platform, respectively. This can advantageously be used to generate high-resolution digital elevation models (*Res. 0.5–2 m*), forest heights, etc.

Satellite imagery, as the name suggests, refers to the remote sensing data (digital images) of the earth's surface that is captured using various sensors (active or passive), deployed on space-based earth-orbiting platforms. Remote sensing sensors collect data on electromagnetic radiation emitted, reflected and reflected from objects. Light is a wave with a wavelength and frequency that make up the electromagnetic spectrum. The longer the wavelength, the less energy it contains, and vice versa. The electromagnetic spectrum is broad, and not all wavelengths are equally effective to interact significantly with remote sensing targets.

Imagery, in general, is a very effective visual aid and satellite imagery as a source of derivative information, such as planimetric analysis, and classification schemes, which can be used to deduce information about land use, land cover, and vegetation, etc.

The popularity of satellite image processing for geographical studies has expanded. Satellite imagery provides numerous benefits to scientific users, including global views, multi-scale observations, frequent observations, direct and non-destructive observations, comprehensive coverage and even non-visible spectral features. As a result of these substantial benefits, a diverse range of applications of satellite imagery has emerged. Moreover, the declining costs, increasing picture resolutions and expanded accessibility (e.g. online availability and free access) of satellite photos for the public at large, all contribute to their increasing acceptance in research. Over the last several decades, satellites have been employed to acquire a tonne of information on the earth's surface. Change detection, image classification, risk assessment, agriculture, natural resource management, coastal and marine, urban geography, monitoring and understanding of global weather patterns, tectonic activity studies, cryosphere research and emergency management and response are just a few of the scientific applications of satellite images. Additionally, it is beneficial to conduct analytical research of a spatial nature.

The satellite imagery is qualified by the different types of resolutions. *Spatial resolution* refers to on-ground extent of area covered by a single pixel in the imagery, and it is commonly described in raster data as pixel size. The *spectral resolution* of a satellite sensor is defined by the number and width of the spectral bands it observes. *Radiometric resolution* is synonymous with the raster data structure as it depicts the stretch of values that a raster image can store. Satellites in orbit take periodic observations of the same point on the earth, referred to as *temporal resolution.* These timeframes vary according to orbit, altitude, location and swath, among other factors, and are critical for temporal analysis and change detection.

Many of the satellite images we get are in different colours, as compared to what we perceive with our own eyes. They are generally monochromatic (black-and-white). GIS enables these black-and-white images to be assigned false colours. Red, green and blue are the three fundamental colours of light. By using a distinct primary colour for each band, GIS can display a raster image in three distinct bands simultaneously. When these three photos are combined, we obtain the colour composite image. When data allocated to red, green and blue is collected outside the visible zone, a *false colour composite (FCC)* is generated. When data allocated to red, green and blue is collected at the corresponding wavelengths, a *natural colour composite (NCC)* is produced. Different satellites have different band structures, and natural and false-colour composites are created accordingly.

In the world of remote sensing and geographic information systems, a multitude of satellite imagery products is accessible. Among others, the US Landsat satellite imagery, the French SPOT satellite imagery and the ESA's Sentinel images are prominent. Here, we will limit our discussion to Landsat imagery only. The *Landsat* programme formally began planning and development in 1967. NASA and the U.S. Geological Survey (USGS) are two federal partners in the Landsat programme. Now, more than four decades after the launch of Landsat 1 in 1972, the Landsat programme continues to ensure the supply of high-quality data for research, innovation and problem resolution, including earth surface phenomena, acquisition planning, data storage and maintenance, and provision of data products suitable directly for analytical studies. The availability of archival and fresh imagery has culminated in a multitude of novel applications and scientific considerations. International programmes and conventions (e.g. forest monitoring, climate change) benefit from access to consistently acquired and calibrated data with the expectation of future continuity (Wulder et al. 2019). Landsat imagery has revolutionized the quality of analytical and descriptive outputs in a wide range of research areas, including cryosphere, aquatic ecosystems, surface water mapping, vegetation phenology, albedo, surface temperature, water use and crop yield (evapo-transpiration), forest monitoring and climate change.

Landsat 1, 2 and 3 obtained images with a spatial resolution of 57 m $\times$ 79 m and radiometric resolution of 6-bits, using the *multispectral scanner (MSS).* MSS data was gathered using the red, green and two near-infrared bands. Its data does not equate to the later Landsat sensors. Consequently, implementing effective MSS atmospheric correction and cloud masking is complex. This serious trade, however, becomes irrelevant when the research involves Landsat data gathered prior to 1985 as MSS is the individual source of worldwide data for studies that require moderate resolution digital imagery for 13 years (1972–1984).

The Thematic Mapper (TM) scanner aboard Landsat 4 acquired images in 1982 with a spatial resolution of 30 m and 7 spectral bands including blue, green, red, near-infrared, mid-infrared I, mid-infrared II and thermal infrared). Thematic Mapper was able to utilize a wider range of wavelengths of the electromagnetic spectrum and obtained a more comprehensive image of the earth's surface. In 1984, a second TM was launched on Landsat 5. It was designed and built concurrently with Landsat 4 and carried similar payloads.

In 1993, the launch of the Landsat 6 mission failed to reach orbit. It carried an Enhanced Thematic Mapper (ETM) scanner, which could not be operationalized, as such. Following this Landsat 7 was successfully launched in 1999, that carried Enhanced Thematic Mapper Plus (ETM+), that imitates the competences of the highly successful Thematic Mapper. ETM+ has more features than its predecessors, making it a more adaptable and resourceful tool for change research, land cover supervision and large area mapping. On-board panchromatic band with 15 m spatial resolution, a thermal IR channel with 60 m spatial resolution and an on-board data recorder are among these features. Since the introduction of Landsat 8 in 2013, Landsat science and services have grown in breadth and depth, with significant advancements in data quality and depth. The Landsat 8 satellite payload includes two scientific instruments: the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). OLI delivers seasonal scans of the global landmasses with spatial resolutions of 30 m (visible, NIR, SWIR); and 15 m (panchromatic). A new deep blue band (0.43–0.45 m) and a new shortwave infrared band (1.36–1.38 m) are added to this instrument which further refines its spectral resolution. TIRS collects 100 m (thermal) data for two extra narrow spectral bands in the thermal zone formerly covered by a single wide spectral band on Landsat 4–7. In 2021, Landsat 9 was launched, and it carries two instruments that are essentially like those on Landsat 8, i.e. the Operational Land Imager 2 (OLI-2) and the Thermal Infrared Sensor 2 (TIRS-2) (TIRS-2). OLI-2 and TIRS-2 sensors detect light at eleven different wavelengths, including visible, near-infrared, shortwave infrared and thermal infrared, as it is reflected or emitted from earth's surface (Fig. 3.5).

3.3 Vector Data Model

The vector data model, alternatively termed the discrete object model, represents geographical information on the earth's surface using discrete objects. It is predicated on the idea that the earth's surface is made up of discrete features that can be represented using well-defined point, line and polygon boundaries. The *x, y* coordinates of a feature define its location in the real world. A single *x, y* coordinate pair defines a point. Lines have two or more *x, y* coordinates. Lines that form polygon boundaries define polygons. These *x, y* coordinates refer to the latitude and longitude of the feature.

As mentioned earlier, there are three fundamental vector features, i.e. points, lines and polygons in geographic information systems. A point feature possesses simply the property of being in a particular location. A point is characterized by a pair of *x, y* coordinates. It is suitable for representing singular and/or discrete geographical features with specified *x, y* coordinates, such as wells, trees, kilns, fire lookout towers, gas activity sites and weather stations. The node and the vertex are other point features in GIS. The point is a self-contained entity, but the *node* is a topological intersection that characterizes a common *x, y* coordinate pair shared by intersecting lines and/or polygons. A *vertex,* on the other hand, is any bend across a line or polygon feature that is not an intersection of lines and/or polygons. *Lines* are considered to be one-dimensional features. A line has two endpoints and may include extra points that indicate its form. Its shape can be determined by connecting straight line segments or by generating a smooth curves using mathematical functions. A line feature in GIS is composed of a single line or a collection of lines (polyline feature). Lines are suitable for depicting linear elements including highways, streams and

Band Number	Band Name	Wavelength	Spatial Resolution
		Landsat 9 (OLI-2/TIRS-2)	
Band 1	Visible	(0.43 - 0.45 µm)	30-m
Band 2	Visible	(0.450 - 0.51 µm)	30-m
Band 3	Visible	(0.53 - 0.59 µm)	30-m
Band 4	Red	(0.64 - 0.67 µm)	30-m
Band 5	Near-Infrared	(0.85 - 0.88 µm)	30-m
Band 6	SWIR1	(1.57 - 1.65 µm)	30-m
Band 7	SWIR2	(2.11 - 2.29 µm)	30-m
Band 8	Panchromatic	(0.50 - 0.68 µm)	15-m
Band 9	Cirrus	(1.36 - 1.38 µm)	30-m
Band 10	TIRS1	(10.6 - 11.19 µm)	100-m
Band 11	TIRS2	(11.5 - 12.51 µm)	100-m
		Landsat 8 (OLI/TIRS)	
Band 1	Coastal Aerosol	(0.43 - 0.45 µm)	30 m
Band 2	Blue	(0.450 - 0.51 µm)	30 m
Band 3	Green	(0.53 - 0.59 µm)	30 m
Band 4	Red	(0.64 - 0.67 µm)	30 m
Band 5	Near-Infrared	(0.85 - 0.88 µm)	30 m
Band 6	SWIR1	(1.57 - 1.65 µm)	30 m
Band 7	SWIR2	(2.11 - 2.29 µm)	30 m
Band 8	Panchromatic	(0.50 - 0.68 µm)	15 m
Band 9	Cirrus	(1.36 - 1.38 µm)	30 m
Band 10	TIRS1	(10.6 - 11.19 µm)	100 m
Band 11	TIRS2	(11.5 - 12.51 µm)	100 m
		Landsat 7 (ETM+)	
Band 1	Blue	(0.45 - 0.52 µm)	30 m
Band 2	Green	(0.52 - 0.60 µm)	30 m
Band 3	Red	(0.63 - 0.69 µm)	30 m
Band 4	Near-Infrared	(0.77 - 0.90 µm)	30 m
Band 5	SWIR	(1.55 - 1.75 µm)	30 m
Band 6	Thermal	(10.40 - 12.50 µm)	60 m
Band 7	Mid-Infrared	(2.08 - 2.35 µm)	30 m
Band 8	Panchromatic	(0.52 - 0.90 µm)	15 m
		Landsat 4-5 (TM)	
Band 1	Visible	(0.45 - 0.52 µm)	30 m
Band 2	Visible	(0.52 - 0.60 µm)	30 m
Band 3	Visible	(0.63 - 0.69 µm)	30 m
Band 4	Near-Infrared	(0.76 - 0.90 µm)	30 m
Band 5	Near-Infrared	(1.55 - 1.75 µm)	30 m
Band 6	Thermal	(10.40 - 12.50 µm)	120 m
Band 7	Mid-Infrared (IR)	(2.08 - 2.35 µm)	30 m
		Landsat 1-5 (MSS)	
Band 4	Visible	(0.5 to 0.6 µm)	79 m
Band 5	Visible	(0.6 to 0.7 µm)	79 m
Band 6	Near-Infrared	(0.7 to 0.8 µm)	79 m
Band 7	Near-Infrared	(0.8 to 1.1 µm)	79 m

Fig. 3.5 Landsat imagery specifications

boundaries. In addition to the location, lines have the attribute of length associated with them. A line is characterized by a series of *x, y* coordinates. A *polygon* is a two-dimensional shape that possesses the attributes of perimeter and area in addition to the location. A polygon is characterized by a series of *x, y* coordinates. It is constructed by looping back many lines to draw a closed feature. On the first line segment, the first coordinate pair (point) is the identical to the last coordinate pair of the last line segment. The perimeter or boundary of a polygon is defined by its connected, closed, non-intersecting lines. A polygon feature is made up of a single polygon or a collection of polygons that may exist independently or in conjunction with each other. Additionally, within its extents, a polygon may include a hole, resulting in an outer and an inner boundary. Polygons are suitable for representing municipal boundaries, geological formations, lakes, soils, vegetation zones, and urban regions, among other things.

Certain GIS implementations mandate the representation of a single geographical feature in multiple ways. This is referred to as a *multi-representation system.* As such, the geographical feature should be seen as a point feature at times and as a polygon feature at others. A good illustration of this is the development of maps at a variety of scales. While villages can be depicted as point features on a small-scale map of the national road network, thematic mapping of a certain district requires villages to be represented as polygons.

Vector data is prepared in three important steps. The first stage categorizes geographical features as points, lines and polygons across space and describes their locations and shapes using points and their *x, y* coordinate pairs. The second stage establishes a conceptual foundation for the attributes and spatial interactions of geometric objects. The third stage encodes and saves vector data in digital data that can be reretrieved, analysed and processed.

3.3.1 Vector Data Structure

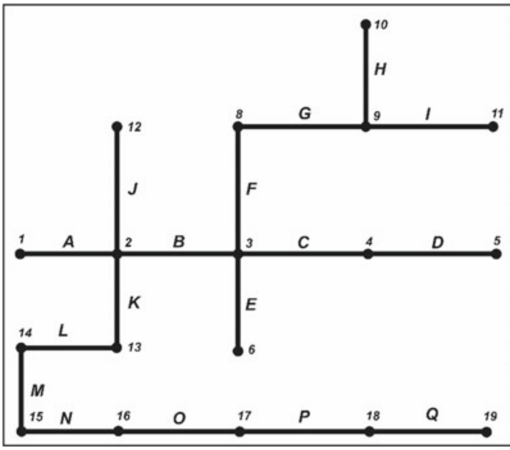
Vector data structure refers to how the vector representations of geographical features are generated and stored in computer systems. In GIS, different structures of the vector data model are used to generate, store and manage vector information. Each of them has different advantages and disadvantages. Here, we shall elaborate on two of the most popular data structures, i.e. the spaghetti data model and the topological data model. Vector data in GIS can be topological or non-topological, depending on whether or not topology is incorporated into the data.

The spaghetti data model is the simplest vector data structure. Features in the spaghetti model are depicted as a string of *x, y* coordinates (or as a single *x, y* coordinate pair for a vector containing a only one point) with no inherent structure. Each line in this model can be thought of as a single strand of spaghetti which is transformed into complex shapes by adding more strands of spaghetti (Campbell and Shin 2012). In spaghetti structure, adjacent polygons must be composed of their own stands of spaghetti. That is to say, each polygon must have a unique set of *x, y* coordinate pairs, even if adjoining polygons share some strands. This creates redundancy in the data model, hence reducing its efficiency. Even though each strand of spaghetti has a locational identifier, spatial associations are not recorded in the spaghetti model. This leads to a deficiency of topological information, which is challenging when performing measurements or analyses. As a result, the computational requirements are quite high when modern analytical techniques are applied to vector data constructed in this manner. Nonetheless, the spaghetti data model's basic structure enables quickly reproduce graphics and maps for topological information is not required in the process.

Unlike the spaghetti data model, the *topological data model* incorporates topological information into the datasets, as the term suggests. Topology refers to a bunch of principles that describes the relationship between adjacent points, lines and polygons and the manner in which their geometries are related. Consider two nearby polygon features as an example. The spaghetti model defines the common border of the adjacent features as distinct, undistinguishable lines. With topology included in the data model, only one line can be used to depict this common border. Additionally, topology is associated with maintaining spatial features when shapes are subject to geometric transformations. It enables more effective projection and reprojection of map data.

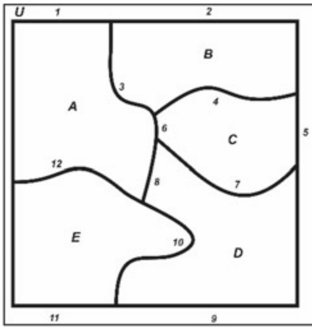
Three fundamental topological percepts need to be discussed to comprehend the topological data model. To begin, the connectivity of the feature dataset is depicted by *arc-node topology* percept. As previously established, nodes are the intersection sites of two or more arcs in topological data model. In case of arc-node topology, arcs have a first node and an ending node. Additionally, connecting each pair of nodes is a line segment, occasionally referred to as a link, that has its unique identification number and refers to both the starting and ending node. Arcs *A, B, J* and *K* all overlap in Fig. 3.6 because they share node *2*. As a result, it is possible to move along arc *A* and turn onto arc *K*, but not from arc *A* to arc *D*, as they lack a common node. The second topological percept is *polygon-arc topology*, which defines the concept of area. According to the area definition, arcs that join to encircle an area create a polygon. In polygon-arc topology, arcs are utilized to form polygons, and each arc is recorded just once. This reduces the quantity of data stored and entails that the edges of adjacent polygons do not overlap. The polygon-arc topology depicted in Fig. 3.7 demonstrates that polygon *C* is composed of arcs *4, 5, 7* and *6.*

The third fundamental topological percept, *polygon topology*, determines contiguity of the area features. This percept is predicated on the premise that the polygons sharing a boundary are considered contiguous. It necessitates that arcs of a polygon have their particular orientation, which enables the determination of adjacency information, i.e. which polygon is on what side. Because polygons that share an arc are considered contiguous, the left and right sides of each arc can be determined. This left and right information is plainly contained within the topological data model's attribute information. The universe polygon is a critical element of polygon topology since it depicts the area surrounding the study area. As illustrated in Fig. 3.8, arc 6 is enclosed on the left by B and on the right by C. The universe polygon, polygon A, is to the left of arcs 1, 2 and 3.



Arc	Start Node	End Node
A	1	2
B	2	3
C	3	4
D	4	5
E	3	6
F	7	8
G	8	9
H	9	10
I	9	11
J	2	12
K	2	13
L	13	14
M	14	15
N	15	16
O	16	17
P	17	18
Q	18	19

Fig. 3.6 Arc-node topology percept



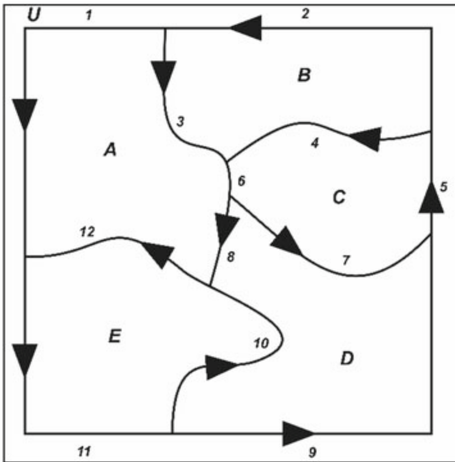
Polygon	Constituent Arcs
A	*1, 3, 6, 8, 12*
B	*2, 4, 3*
C	*4, 5, 7, 6*
D	*7, 9, 10, 8*
E	*10, 11, 12*

Fig. 3.7 Polygon-arc topology percept

3.4 Vectorization and Rasterization

GIS software can display raster and vector data concurrently and quickly convert between the two. Raster and vector data complement one another in numerous ways. Thus, the conversion of these two forms of data has turned out to be a standard and desirable practice in GIS projects.

Vectorization also called digitization is the process of converting raster data to vector data. It is composed of three fundamental aspects, i.e. thinning, extraction and topological reconstruction. The lines have a length but no breadth. Vectorization is the process of converting raster data to vector data. It is composed of three fundamental aspects, i.e. thinning, extraction and topological reconstruction. The lines have a length but no breadth. A subsequent line smoothing technique can assist in removing such raster data abnormalities.



Arc	Left Polygon	Right Polygon
1	A	U
2	B	U
3	B	A
4	C	U
5	C	U
6	C	A
7	C	D
8	D	A
9	D	U
10	E	D
11	E	U
12	E	A

Fig. 3.8 Polygon topology percept

Rasterization is the process of converting vector data to raster data. This is accomplished by assigning point, line and polygon attribute values to raster cells that overlap with the corresponding features on the vector base map. Rasterization involves three basic stages. The first stage creates a raster with a given cell size that spans the area covered by vector data and initializes all cell values to zero. The size of the raster cells should be carefully chosen to meet the study's needs for geometric resolution, storage space and so on. The second step modifies the values of cells such that the points, lines and polygon boundaries are separately coded. The third step fills the polygon interiors with the polygon cell values. As such, rasterization is a backward step, as raster boundaries are an approximation of the original vector boundaries and the original objects lose their topological properties.

3.5 Conclusion

This chapter discusses the definitive and structural aspects of GIS data models. In GIS, data plays the most important role in providing the inputs for the analysis and interpretation of spatial objects/phenomena. As such the data may be spatial data of geometric form or the non-spatial attribute data that qualifies the spatial data. Raster and vector data models are used for the representation of spatial data in the GIS. The actual geographical space is graphically represented with the help of these data models. Although the major aspects of the data models have been discussed, we aim to conclude this chapter with a brief discourse on the advantages and disadvantages of the two GIS data models.

The raster data model is the most dominant GIS data model that represents the geospatial information using the grid of pixels. This model benefits from the low cost and widespread availability of the technology required to make raster images. Raster images are structurally simple. Each grid cell in the raster image corresponds to a only one value, and raster overlay analysis is quite straightforward. However, raster files are often quite huge. Particularly for raster images created using the cell-by-cell encoding approach, the total number of values recorded for a particular dataset results in extremely large files. Also, a raster file that spans a huge area and has high resolution will consume hundreds of gigabytes of storage space. Data compression techniques are available for handling such a problem. Also, depending on how far one zooms into a raster image, the details and coherence of the raster are gradually lost in a sea of pixels. Geometric adjustments that occur during map reprojection also cause raster graphics problems. Changing map projections alters the size and form of the original input layer, frequently resulting in loss or gain of pixels. The raster data model becomes bothersome when the spatial analysis attempts to overlay and analyse several raster graphics created at varying scales and pixel resolutions.

Vector data model that uses points, lines and polygons to represent spatial information tends to be more accurate and precise representation of reality as compared to raster data model. Additionally, vector data enables a greater degree of control over the scale of observation and analysis. For the fact that coordinate pairs associated with a point, line or polygon features represent tiny precise locations, zooming deeply into a vector image does not alter the view of a vector in the same way that it alters the view of a raster. Vector data also has a more compact data structure, which results in substantially reduced file sizes than their raster equivalents. The vector model is topological in nature. When a vector model is used, the topological information simplifies spatial analysis (e.g. network analysis, proximity analysis, error detection and spatial transformation). On the other hand, the vector data model has two significant disadvantages. Firstly, the data structure is typically more complex as compared to the simple raster data model. Because the location of vertices must be explicit in the model, there are no shortcuts for saving storage space like the run-length and quadtree encoding approaches offer in the raster data model. Secondly, spatial analysis implementation could be somewhat complex due to slight variations in the accuracy and precision of input datasets. Likewise, vector data manipulation and analysis methods are sophisticated, which can result in high processing needs, particularly when dealing with larger datasets.

Questions

1. What is a data model in GIS?
2. What components make up raster data models?
3. What are the various raster data structuring schemes?
4. How does the spaghetti vector data model vary from the topological vector data model?
5. How many fundamental topological percepts are there? Discuss.
6. What is the significance of data conversion in GIS environment?

References

Campbell J, Shin M (2012) Essentials of geographic information systems. Saylor Academy.
Chang KT (2018) Introduction to geographic information systems-ninth edition. McGraw-Hill Education, 2 Penn Plaza, New York, NY 10121. ISBN: 978-1-259-92964-9
Wulder MA, Loveland TR, Roy DP, Crawford CJ, Masek JG, Woodcock CE, Zhu Z et al (2019) Current status of landsat program, science, and applications. Remote Sens Environ 225:127–147. https://doi.org/10.1016/j.rse.2019.02.015

Chapter 4
Data Input in GIS

Abstract This chapter discourses on the understanding of the process of data input in the GIS. It deliberates upon all the phases involved in the process from data collection to data preparation. After reading this chapter, you will understand the following:

- The different sources of geospatial data.
- Various methods of spatial data input in GIS.
- Errors encountered in the geospatial data input process.
- Spatial data editing by topological, non-topological methods among others.
- Non-spatial data entry and attribute data manipulation in GIS.

Keywords Spatial data input · Geospatial data sources · Spatial data editing · Non-spatial data input · Attribute data manipulation

4.1 Introduction

A Geographic Information System (GIS) is a computer-based information system that accepts and stores data in a digital format. In GIS, data is the prerequisite for mapping, analysis and modelling (GIS). GIS integrates the data from a multitude of sources into a common format that can be compared and analysed. The process of encoding data into a digital format and writing it to the GIS database is called data input. The process of importing data into a GIS requires encoding both the spatial data (the location of features on a map) and non-spatial data (descriptive or numeric information about features) into a GIS. Coordinates on a particular Cartesian coordinate system are used to encode the locational data. The projections and scales of the source maps may differ, and multiple data transformation steps may be required to unify all data to a single coordinate system. The spatial and non-spatial data that act as the fuel in the GIS analyses are input, edited, analysed and stored in GIS as databases. The spatial database is built to represent the spatial data in the raster and vector formats. These models have been discussed definitively and structurally in the previous chapter. The non-spatial database, which consists of the attributes related to the spatial data, has different representations in the georelational and the object-based vector data models, which are discussed in the later section of this chapter.

M. Kumar et al., *Geographic Information Systems in Urban Planning and Management*, Advances in Geographical and Environmental Sciences,
https://doi.org/10.1007/978-981-19-7855-5_4

The definition of data input highlights that there must be suitable sources of geographical data that could be encoded into digital format for GIS analyses. The production of GIS data historically obligated the painstaking efforts of digitizing paper maps. However, as a result of technical improvements, the situation has now altered. Fresh GIS data can be developed from a wide variety of data sources. In addition to paper maps, we can employ primary data with *x, y* coordinates as the data source. These primary data sources range from manually collected field data to advanced satellite imagery. The definition of data input also stresses the methods by which the geographic data is encoded into the digital format. Spatial data can be entered into GIS by digitization, data transfer, etc. These techniques are explored in detail in the sections that follow.

4.2 Sources of Geospatial Data

Geospatial data is information on objects, events or other features that have a physical location on or near the earth's surface. It combines location data (primarily earth coordinates) and attributes data (the characteristics of the object, event or phenomena concerned). Existing geospatial data can be input into GIS directly from the variety of sources available in the digital world and accessible through the Internet. Government agencies, non-profit organisations and private companies provide, offer or sell the geospatial data in formats suitable for input and analyses in the GIS. In general, the governments have now been building and maintaining the geospatial data in the administrative system as Spatial Data Infrastructure (SDI). The National Aeronautics and Space Administration (NASA), the U.S. Geological Survey (USGS), the European Space Agency (ESA) and the Indian Space Research Organization (ISRO) among others, provide the geospatial data at global as well as respective national levels. This geospatial data including (satellite imagery, DEMs, etc.) is georeferenced and available in digital formats suitable for direct input to GIS (Fig. 4.1).

Apart from the existing spatial data, new geospatial data can be created from various sources. These include analogue maps, remote sensing data, GPS data, field surveys, etc. *Analogue maps* are tangible maps that can be viewed directly on a piece of paper or fabric. Traditional drawn or printed productions, map-like outputs of

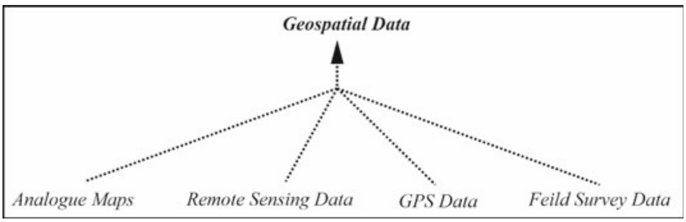


Fig. 4.1 Sources of geospatial data

aerial photography and satellite imagery can be included in this category. Analogue maps must be obtained from credible and reliable sources like the Survey of India, the Census of India, the National Remote Sensing Centre, the National Bureau of Soil Survey and Land Use Planning, the Geological Survey of India, etc. *Remote sensing data* gathered via space-based platforms, aerial vehicles or ground-based platforms can be analysed digitally to generate a wide variety of geospatial data. The quality of remote sensing data has improved significantly in terms of spatial resolution, to the point that it can be used efficiently to extract rivers, roads, paths, buildings, forests, rivers, etc. As such, it serves directly as the basic framework for a variety of GIS-based analyses, including land use-land cover mapping, vegetation studies, crop health assessment, soil erosion, mineral exploration, water body monitoring and glacial dynamics. Satellite images can provide useful data since they are collected at regular intervals and so allow for change detection studies in the terrestrial and aquatic ecosystems. Digital remote sensing data are freely available from a variety of sources, including the Bhuvan Indian Geoportal, the United States Geological Survey's EarthExplorer, the European Space Agency's Sentinel, the Global Land Cover Facility (GLCF), the Japan Aerospace Exploration Agency (JAXA), etc. In India, National Remote Sensing Centre, Hyderabad is the nodal centre for the distribution of remote sensing data products. *Field survey data* is a collection of locational data and associated information gathered during a survey of a field region. It can be used to verify and cross-check spatial data built out of the other sources, especially the remote sensing data. Along with locating physical and manmade structures and determining quantities (precipitation, humidity, mineral potential and so on) across space, field survey data may include distances, directions and altitudes. Measurements of distances can be made with a tape measure or an electronic distance-measuring tool. A transit, theodolite or total station can be used to determine the azimuth or bearing of a line. An azimuth is a counter clockwise angle measured from the north end of a meridian to a line. Azimuths are measured in degrees and range from 0° to 360°. A bearing is an acute angle formed by the intersection of a line and a meridian. The bearing angle is always preceded by letters indicating the quadrant (NE, SE, SW or NW) in which the line is located. Levels and rods can be used to determine the elevation difference between two points in feet or metres. Point data obtained during field surveys can be imported into a GIS and extrapolated or interpolated to determine the spatial distribution of the geographical phenomena. It is possible to generate it from a text file containing x, y coordinates. Each pair of x, y coordinates results in the creation of a point. As a result, we may generate spatial data from a file containing the locations of meteorological stations, epicentres, etc.

The Global Positioning System (GPS), a constellation of 24 satellites designed and operated by the US military, transmits data continually, allowing for the accurate identification of a position on earth by measuring its distance from the satellites. Along with geographic coordinates, GPS gives altitude and time data in text format (ASCII) suitable for GIS input. The GPS receiver determines its distance (range) from a constellation satellite by calculating the travelling time and speed of the signals received from the satellite. With three satellites available concurrently, the receiver estimates its position in space (x, y, z) with respect to the earth's centre. A fourth satellite compensates for timing errors in order to identify the precise location. After that, the receiver's position in space is translated to latitude, longitude and altitude in WGS-84 coordinates. GPS users have at least four satellites visible from any location on the earth's surface thanks to the Navigation Satellite Timing and Ranging (NAVSTAR) satellite constellation. A critical part of using GPS for spatial data entry is rectifying GPS data for inaccuracies. There can be noise errors, including ephemeris (positional) errors, clock errors (orbital errors between monitoring times), atmospheric delay errors and multipath errors (signals bouncing off obstructions before reaching the receiver). Differential GPS (DGPS) is an augmentation technique, which can significantly reduce noise errors with the aid of reference or base stations. Reference stations are located at precisely measured places and are operated by commercial enterprises and public agencies. Using its known position, the reference receiver can calculate what the travel time of the GPS signals should be. Thus, the correction factor is defined as the difference between the expected and actual journey times. The reference receiver computes error correction factors for all visible satellites. These correction factors are then made available to all GPS receivers within the reference station's coverage area.

Apart from GPS, there are similar systems of global or regional importance across the world, such as the GLONASS (Russia), the Galileo (Europe) and Beidou (China). India's space agency, the Indian Space Research Organization (ISRO), and its commercial arm, ANTRIX, developed the Indian Regional Navigation Satellite System (IRNSS), which was approved by the Indian government in 2006 and later renamed Navigation Indian Constellation (NavIC) in 2016, for operational purposes. It is an independently built regional navigation satellite system by India. NavIC is composed of eight satellites (three in geostationary orbit and five in inclined geosynchronous orbit) orbiting roughly 36,000 kms above the earth's surface. The NavIC operates in a manner very similar to the Global Positioning System (GPS) of the USA. It provides accurate position data to users in India as well as users in an area extending up to 1500 kms outside Indian borders. It offers two types of services: standard positioning services (SPS) and restricted positioning services (RS) (RS).

GPS and other alike positioning systems can simply provide location-specific information (point data) as discussed above. A collection of GPS readings along a line can be used to determine the existence of a line feature as shown in Fig. 4.2, and a series of GPS-measured lines can be used to determine the existence of an area feature as shown in Fig. 4.3. This is why GPS has become a valuable tool for geographic data collection, validation and monitoring.

It is also convenient to distinguish between primary and secondary raster and vector geographic data. *Primary data sources* are digital data sources obtained particularly for a GIS project. Examples of typical GIS inputs are SPOT and IKONOS Earth satellite pictures and building-survey measures taken using a total station, etc. *Secondary sources* are digital and analogue datasets that were originally collected for another reason and require digitization for usage in a GIS project. Typical secondary sources include scanned colour aerial pictures of metropolitan areas and paper maps

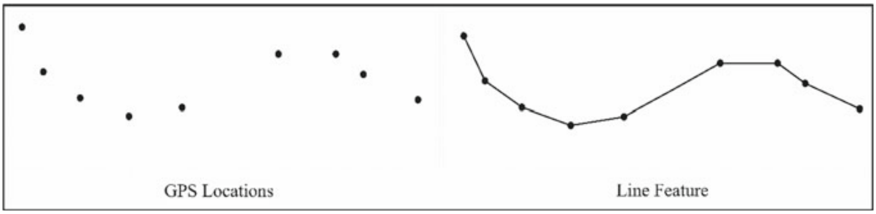


Fig. 4.2 Line feature construction using GPS data

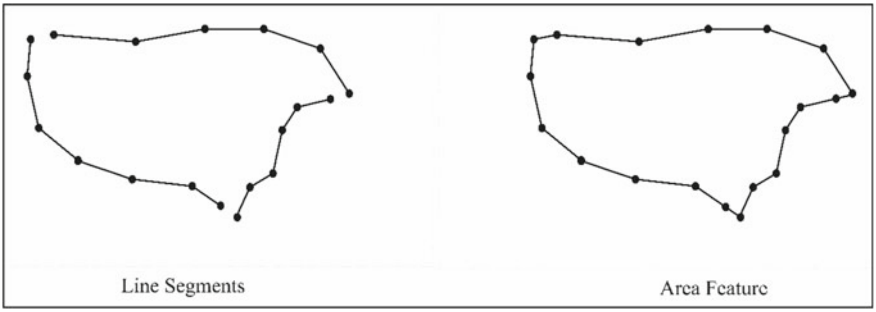


Fig. 4.3 Area feature construction using GPS data

that may be scanned and vectorized from the United States Geological Survey or Institut Geographique National, France (Longley et al. 2001).

4.3 Spatial Data Input in GIS

The data input process is a significant choke point in GIS technology. It usually absorbs 80% or more of the project's budget and time. The procedures are generally time-consuming, tedious and prone to error. It is crucial to establish solutions to save costs while improving the reliability of GIS Data Input. Additionally, it is necessary to automate as much of the input process as possible, while ensuring that automated input does not result in major editing problems later on. To input data into a GIS, we require tools for converting geographical data of various forms to a digital format. This can be accomplished through the use of the various methods as shown in Fig. 4.4.

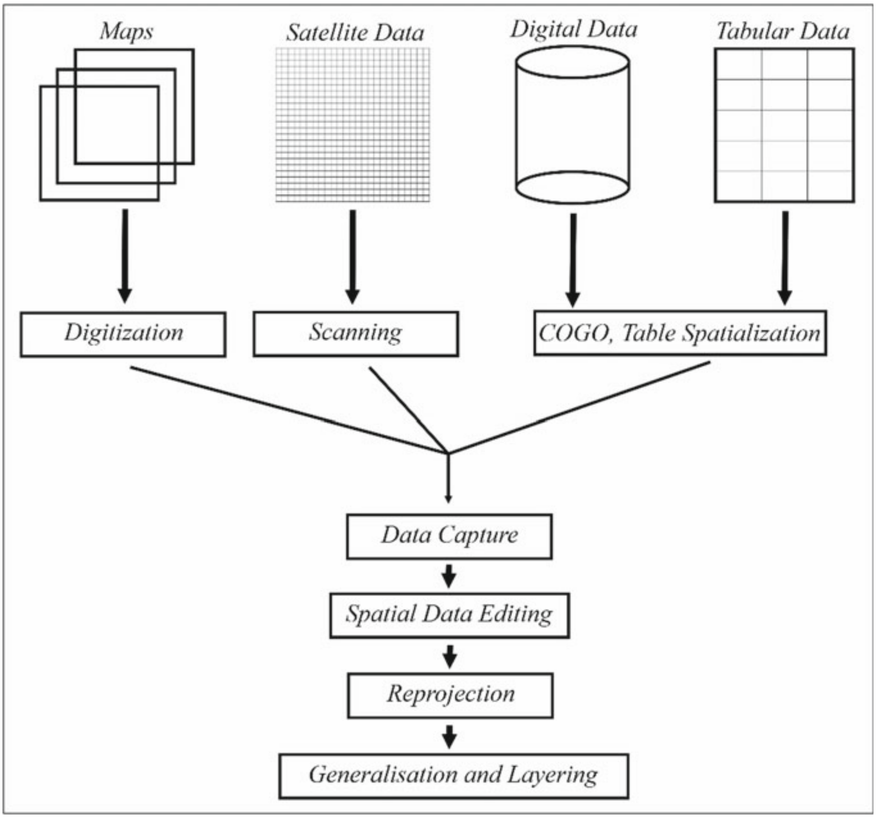


Fig. 4.4 Methods of data input

4.3.1 Scanning

Scanning is the most frequently used form of automatic digitization and is the best way of data encoding when extensive raster datasets are required. It is a process that converts an analogue map to a raster-format scanned file, which is subsequently transformed into vector format through tracing. The monochrome and multicolour maps accessible in the form of analogue maps can be scanned into a digital format using the appropriate scanner types. The quality of the raster image is determined by the inherent resolution of the scanner.

After scanning, the scanned raster is often vectorized by a technique called tracing. Tracing entails line thinning, feature extraction and reconstruction. Tracing can be performed manually, in which the user specifies the line to be traced from the start to the end, or semi-automatically, in which the user specifies only the starting node and the computer traces all the lines connecting to it.

There are numerous scanning technologies available for the automated capture of geographical data. All of these methods have the advantage of rapidly capturing spatial data from a map. The scanning issue is that scanners are frequently fairly costly. They rarely distinguish between geometry, text and symbols. Additionally, scanned data requires extensive human editing. While automatic scanning can be extremely fast, it cannot be guaranteed to be error-free.

4.3.2 Digitization

Digitization is the conversion of analogue data to a digital format. It is the process of converting the spatial information on a map into a digital format. The point, line and areal features of the map are translated to *x, y* coordinates. A point is represented by a single pair of coordinates, a line by a string of coordinate pairs, and an area (polygon) is designated when several lines are combined with a label point inside an outline. Digitization, per se, can be defined as the capturing of a series of points and lines. Digitizing can be performed on a digitizing tablet (Fig. 4.5) or a computer screen (Fig. 4.6).

A *digitizing tablet* consists of an electronic mesh of wires that can detect the position of the cursor. The computer detects and interprets the position of an indicator as it moves across the surface of the digitization tablet. The cursor contains control buttons that enable system control without diverting attention away from the digitizing tablet and towards a computer interface. The operator just clicks a button on the cursor after matching the cursor's crosshair with the point to transfer the point *x, y* coordinates of the point to the computer system with which it is connected. The absolute precision of large-size digitizing tablets is approximately 0.001 in.

To carry out digitization using the digitizing tablet, a map is placed and affixed to the table. The coordinate system is specified in the connected computer system. Digitization often begins with the identification of a set of three or more control

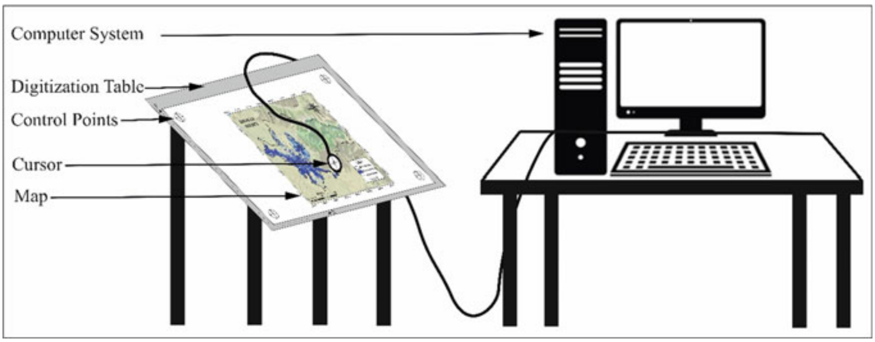


Fig. 4.5 Digitization tablet system

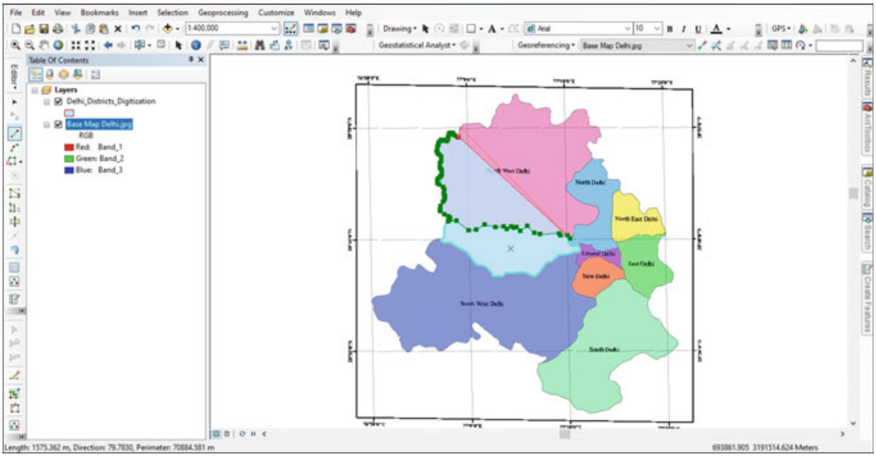


Fig. 4.6 Heads-up digitization

points, which are later used to georegister the digital map with real-world coordinates. Digitizing point features is straightforward: a single click records the location of each point. Line features can be digitized in either point or stream mode. In point mode, the operator chooses points to capture. Lines are digitized in stream mode at a predetermined time or distance interval. For instance, lines can be automatically digitized at intervals of 0.05 in. If the digitized features contain a large number of straight-line segments, the point mode is chosen. According to the vector data model's interpretation of a polygon as a collection of lines, digitizing polygon features is equivalent to digitizing line features. Additionally, a polygon may include a label, which is treated as an internal point.

Heads-up digitization is another term for digitizing on a computer screen. It is particularly valuable for modest digitization initiatives like watershed delineation. On-screen digitization begins with the scanning of the base map to make a digital copy that can be loaded onto the computer system. After that, the scanned map is georeferenced using an appropriate coordinate system. On the computer screen, the georeferenced map is deployed as the base map for tracing out the features using a graphical user interface device (mouse, digital pencil, etc.).

The difficulties encountered during the process of converting maps to digital format via the digitization process differ according to the analyst. It is dependent on the analyst's expertise and skill, as well as the density of the map's points, lines, and polygons. Errors occur due to the instability of the hard copies of the base maps. If the map has been stretched or shrunk in the meantime, freshly digitized points will be slightly out of alignment with previously digitized points. Errors on these maps are entered into the GIS database. The GIS database error level is directly proportional to the source map error level. In the heads-up digitization, it is important to note that the accuracy of scanned output data depends on the scanner, the software used to process it, and the source document. Scanned data quality and amount depend on scanner resolution. A GIS user should look for splines, which appear black on the scanned output, after scanning the paper map. This can be removed by thinning.

4.3.3 Coordinate Geometry

Coordinate geometry (COGO) is a technique for spatial data input that involves the calculation and entry of geometric coordinates. It generates geographical data in the form of points, lines and polygons from survey data. This approach is advantageous for developing extremely exact cartographic definitions of property and is thus more suited to land record management at the cadastral or municipal level.

4.3.4 Table Spatialization

Geospatial data can be generated from a text file that has intact *x*, *y* coordinates, either geographic (in decimal degrees) or projected. There is a point associated with every *x*, *y* pair of coordinates. It is therefore possible to build a spatial information dataset/layer using the locations of weather stations, storm epicentres, the path of a hurricane, etc. As with rainfall data from meteorological stations, the point dataset generated using this method can be spatially interpolated or extrapolated in the GIS using a variety of methods (Inverse Distance Weighting, Kriging, etc.). This creates an area expression of the phenomenon under investigation.

4.3.5 Data Entry Errors and Spatial Data Editing in GIS

A fundamental requirement for Geographic Information System (GIS) applications is efficient and good quality spatial data. However, the process of spatial data entry is prone to a variety of errors. We rely on spatial data editing to meet the standards of good quality analysis in a GIS environment. As the raster data is represented by a regular grid and fixed cells, spatial data editing does not apply to it. Vector data, on the other hand, may have editable errors. For example, newly digitized layers are likely to contain digitizing errors, while existing spatial layers may be out of date and so require revision. Locational and topological errors are two major types of errors in the Vector data.

Location errors are geometric flaws in scanned features that vary according to the data source utilized to digitise them. In primary geospatial data, spatial data accuracy is dependent on instrument parameters such as the sensor's spatial resolution in remote sensing data products, etc. Satellite images may have a spatial resolution of less than one metre to one kilometre. Similarly, GPS point data precision might range from several millimetres to ten metres. On the other hand, three frequent situations can lead to errors between secondary geospatial data and the source map. The first is attributed to human error during manual digitization. The second situation involves scanning and tracing issues. A tracing method typically encounters complications when raster lines are too close together, too wide, or too thin and broken. Collapsed lines, malformed lines and additional lines are all examples of digitizing problems caused by scanning and tracing issues. The third possibility is that errors occurred during the conversion of the digital map to real-world coordinates. The conversion can result in differences between digitized and source lines if the control points are incorrect.

Topological errors violate the topological relationships in a vector data model. The connectivity, area definition and contiguity topological linkages may be lost in newly digitized geospatial data. Also, about more than thirty topology rules control the spatial interactions between point, line, and polygon features in the GIS environment. These rules apply to features within a feature class or to two or more participating feature classes (Chang 2018). There are several typical topological errors while working with spatial features, which can be classed as polygon-specific and line-specific topological errors. A polygon is defined by the presence of a closed perimeter. In the process of digitization, they may overlap, have gaps between them or have unclosed boundaries if their boundaries are not precisely scanned. A line that has a beginning and an end may incorporate dangling nodes and pseudo-nodes as two common topological errors. When line features do not meet exactly at their endpoints, *dangling nodes* occur. If a gap exists between lines, it becomes an *undershoot dangle*, and if a line is overextended, it becomes an *overshoot dangle*. The *pseudo-node* separates a line feature unnecessarily into multiple line segments.

Spatial data editing can be used to remove locational and topological errors from spatial data vectors. To edit spatial data, we must utilise a GIS that is capable of detecting and exposing errors and providing tools for fixing them. Spatial data editing entails topological and non-topological operations, as well as edge matching, line generalization, and line smoothing. *Topological editing* operations include **cluster processing** in which *XY tolerance* is specified and used to snap the vertices falling in the square area of tolerance. It is useful for handling dangling node errors. If the parts of features in a dataset (in a single layer or multiple layers) are supposed to be coincident, a temporary **map topology** can be built to avoid or remove the topological errors. Also, the topology rules of the GIS package can be used to avoid and remove the topological errors. *Line Extending, Line Trimming, Moving Features, Deleting Features, Integrating Feature/Feature boundaries in a specific X, Y Tolerance, Feature Reshaping, Line Splitting, Polygon Splitting,* etc., are the common non-topological editing operations. Non-topological editing also affords the creation of new features from out of the existing features using operations like *Merge, Buffer, Union and Intersection.*

Edge-matching connects lines along an edge of a layer to lines on a neighbouring layer, ensuring that the lines are continuous across the border between the layers. The term **line generalization** refers to the act of reducing the complexity of a line by eliminating some of its points. Lines with an excessive number of points may not always improve the analysis outcome, but they do demand additional processing time. When a large number of points are digitized by scanning or stream-mode digitization, or when a source layer is to be shown at a smaller scale, line generalization is required. **Line smoothing** is a procedure that involves altering lines using mathematical functions as with splines. Perhaps the most critical aspect of line smoothing is its application to data visualization. Computer-generated lines, such as isolines on a precipitation map, are frequently rough and unattractive. These lines can be smoothed for the purpose of displaying data.

4.4 Non-spatial Data Input in GIS

The non-spatial data also called the attribute data describes the characteristics of the spatial features. The non-spatial data of raster and vector data models are different from one another. Each pixel in the raster image has its unique corresponding data value. The attribute table lists the pixel values as per their frequencies (Table 4.1).

In the vector models, the features are qualified (described) by their attributes (single or multiple). The framework of attribute data entry and storage is different in the georelational and the object-based vector data models. The georelational data model (e.g. shapefiles) stores the attribute data, separately from the spatial data, in a relational database. It uses the feature identification numbers (ID) to link the attribute data to the spatial features as shown in Table 4.2. On the other hand, the object-based data model (e.g. geodatabase) combines both geometries and attributes in a single system. Each spatial feature has a unique object ID and an attribute to store its geometry as shown in Table 4.3. Although the two data models handle the storage of spatial data differently, both operate in the same relational database environment.

Table 4.1 Raster data attribute table

Object ID	Value	Count
0	1501	1252
1	1502	2831
2	1503	1503
3	1504	3618

Table 4.2 Vector data attribute table (georelational data model)

Record	Geology-ID	Area	Perimeter
1	1	156.02	624.98
2	2	368.58	1472.62
3	3	1002.22	4108.83
4	4	539.96	2212.39

Table 4.3 Vector data attribute table (object-based data model)

Object ID	Shape	Shape_Area	Shape_Length
1	Polygon	156.02	624.98
2	Polygon	368.58	1472.62
3	Polygon	1002.22	4108.83
4	Polygon	539.96	2212.39

In GIS, attribute data is organized in tables by rows (records) and columns (fields), with each row corresponding to a particular spatial feature and each column corresponding to a particular characteristic of spatial features. Attribute tables are classified into two categories. The feature attribute table and the non-spatial attribute table are distinct in that the former with access to the feature geometry has default fields that store the features' geometry, and geometric attributes such as area, perimeter and length, while the latter does not have direct access to the geometry of the features but can be linked to the spatial data via the common field when necessary.

The procedure of entering attribute data into a GIS is carried out in steps. To begin, the field is established in the attribute table concerning the data characteristics. It entails determining the field's name, length, data type and decimal places. The specifications of the field become the properties of the field. As a result, it is critical to consider the field's intended use before specifying it. The data is then manually entered into the attribute table if the amount of data to be entered is not large, or if the data is not available in the digital file formats. However, it is preferable to acquire the digital data files (from an authentic source) and integrate the data into GIS via the joins or relate options. GIS software packages can import files in delimited text, dBASE and Excel formats. The common field in the GIS attribute table of the spatial data file and the attribute data table is used to match the records and integrate the data tables.

In GIS, attribute data manipulation is possible and several data manipulation procedures are employed. These include *adding data fields* (for classification or computation of the data), *deleting data fields* (to reduce confusion and increase the operational efficiency of the system by removing the unwanted data fields from the geospatial data files, imported directly into GIS), *a new classification* of the existing attribute data and *computation* of the attribute data (using mathematical formulations over the existing attribute data).

4.5 Conclusion

Data is the prerequisite of analysis in the Geographic Information System (GIS). It is also a significant bottleneck in the implementation of GIS technology, and due to the associated expenses, considerable research has been conducted to develop suitable input techniques. Sharing digital data is one method of circumventing the input bottleneck. Geospatial data sources are proliferating and optimizing. Digital data is becoming increasingly available. Free access to such data is gaining popularity throughout the world, with the hope that civilian application of the data can result in effective problem resolution and sustainable growth. The methods for entering data into Geographic Information Systems (GIS) are evolving and becoming more refined over time. Among the data entry methods outlined in this chapter, each has distinct advantages and disadvantages that clearly distinguish it from the others. The method chosen is determined by the volume of data to be digitized, the data format, and the project economics, among other factors. Regardless of the manner of data entry, errors are common, and GIS systems, such as ArcGIS, conveniently assist in identifying, highlighting, and rectifying errors through the use of appropriate tools and algorithms. Thus, geospatial analyses are desired in the modern era for being refined and exact, error-free and pertinent to the issues at hand.

Questions

1. How does spatial data vary from non-spatial data?
2. Define data inputting in GIS?
3. What are the primary sources and various types of geospatial data?
4. How can GPS data be utilized to develop area features?
5. How is digitization performed with the digitization tablet system?
6. Why is the editing of spatial data required?

References

Chang KT (2018) Introduction to geographic information systems-ninth edition. McGraw-Hill Education, 2 Penn Plaza, New York, NY 10121. ISBN: 978-1-259-92964-9

Longley PA, Goodchild MF, Maguire DJ, Rhind DW (2001) Geographic information systems and science. Wiley, London

Chapter 5
Data Visualization and Output

Abstract This chapter is primarily concerned with comprehending the process of geovisualization as a transitional and final stage in the execution of GIS projects. You will understand the following after reading this chapter:

- The concept and process of geovisualization.
- Outputs of GIS data.
- GIS project results that are cartographic in nature.
- Aesthetics inherent in the production of maps based on qualitative and quantitative data.
- Terrain feature mapping.
- Elements of time series mapping.

Keywords Geovisualization · GIS data output · Cartographic output · Vectorization and rasterization

5.1 Introduction

Visualization translates directly as '*visual representation*'. It is the process of creating the visual illustration of data. The advantage of the visual form of data is that it is straightforward for everyone to understand as compared to the numerical, tabular and textual forms of data. In GIS, the geographical database is turned into a visual form, i.e. a map, for abstracting the complicated reality of the earth's surface for comprehending a certain feature/phenomenon, easily. The process of converting the geographic database into maps for understanding through the visualization process is called *map visualization* in GIS or simply *geovisualization.* It is a collection of tools and strategies that assist in the comprehension of spatial data. Geovisualization is most commonly used to refer to the practice of interactively visualizing geographic data at any stage of spatial analysis in GIS. It can also indicate the creation of the final visual output (e.g. charts, maps or a combination of these), as well as the associated methodologies. Interactive on-screen visualization or the final output maps can help deal with questions and answers related to the spatiotemporal analysis of geographic objects/phenomena. Maps reveal spatial relations and patterns. Besides assisting in the identification of geographical features/phenomena, they also

M. Kumar et al., *Geographic Information Systems in Urban Planning and Management*, Advances in Geographical and Environmental Sciences,
https://doi.org/10.1007/978-981-19-7855-5_5

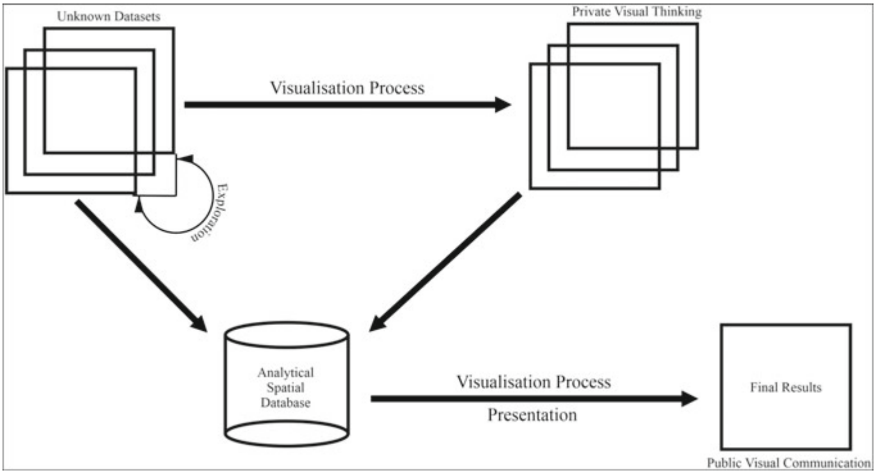


Fig. 5.1 Role of geo-visualisation

demonstrate their thematic attributes and evolutionary aspects. On-screen maps are interactive and offer the option of working out complex queries (Fig. 5.1).

5.2 Geovisualization Process

Geovisualization also refers to the process of developing scientific visualizations in order to build models for map-based scientific visualization. It is primarily concerned with the presentation and exploration of data. While *presentation* is merely the visual transfer of facts to the public, *exploration* is the process of private visual thinking aimed at evaluating the quality, checking correctness through intermediate checks and improving the map's quality, among other things.

Visualization is largely determined by the level at which geographic data must be graphically represented. Furthermore, the purpose for which spatial data is to be graphically visualized governs the geovisualization process. We know that, geovisualization can be and are created at any stage of data handling and spatial analysis. It can be that the user requires to take intermediate visual views of the spatiotemporal database (or a part of it) at various stages in his project. It can be useful in checking the database structure and the consistency in the data acquisition process. Interim visualization processes demonstrate the need for integrated approach to geoinformatics. Apart from their primary function of communicating geographic information, maps have begun to play an important role in the visual thinking process. Nowadays, users of visualization products are interested in *dynamic presentation* and *user interaction*, as well as immediate and *real-time data access*. As a result, the definition of geovisualization has evolved as the methods for presenting data in a visual format have expanded in range.

Additionally, the visualization process serves as a means of communication between the GIS professional and the map audience. The process is also directed by the question "*how do I say what to whom, and is it effective?*" where '*how*' refers to the visualization process, '*I*' refers to the GIS professional, '*what*' refers to the information (spatial data and its attributes), and '*whom*' refers to the intended audience (Fig. 5.2) and the later clause of the sentence suggests that the efficiency of the visual output must be tested by the GIS professional. As such, the visualization process necessitates the creation of complete topographic map sheets, newspaper maps, sketch maps, animated interactive process presenting maps, three-dimensional views of metropolitan systems like traffic situations, among other things. In contrast to the previously held belief that maps must be accepted once they have been drafted by the cartographer, the visualization process allows for the incorporation of expert knowledge and feedback from the map audience in order to improve the output maps' effectiveness.

During the geovisualization process (Fig. 5.3), the cartographic methods and techniques are applied to translate the spatiotemporal database into visual illustrations like maps or on-screen interactive graphics. The scale of the map decides the level of generalization to be incorporated into the map. The type of data (thematic, topographic, etc.) and the nature of data (qualitative or quantitative) has a bearing on the designing of the output. Modern GIS professionals have a plethora of tools at their disposal for visualizing data and modifying its visual appearance as needed. These tools consist of the functions (data classification algorithms, line smoothing functions, etc.), rules and cartographic conventions. Geospatial data visualization uses longitude, latitude and altitude to present two-dimensional and three-dimensional data. With time series data, it becomes a four-dimensional geospatial visualization. Mostly geospatial data visualizations are two-dimensional, such as a standard map or a thematic map. Another method of geovisualization is simulation, which creates

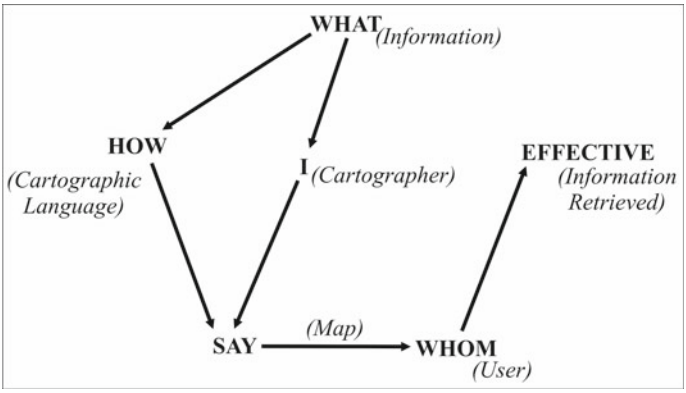


Fig. 5.2 Cartographic communication process

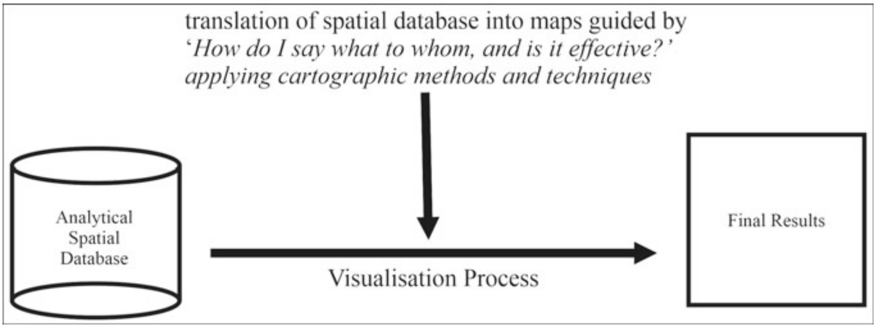


Fig. 5.3 Process of geovisualization

motion by displaying still photographs in sequence. This enables dynamic and interactive visualization of geographical data, such as simulations of continental drift, plate movement and flight across topography. Simulation maps take advantage of time to give the display an additional visual dimension. The interactive user interface is the key technique for geovisualization. This enables the authentication protocol to gather extra information from the user as the authentication session progresses. For instance, a user who is interacting with Google Earth data may rediscover the proper route between two points by drawing the route on a map or discovering other relationships in the data. This notion was also employed in conventional cartographic maps by analysing data using a coloured pencil or ribbon, but it was less effective when analysing multiple datasets at the same time. In this case, GIS enables the analysis of several maps from multiple sources on a single platform using multi-criteria methodologies (Singh and Kumar 2013). For instance, we can close or add new layers for analysis purposes, such as buildings, roads, boundaries and market layers. Thus, geovisualization is a dynamic process in which a person sorts, highlights, filters and transforms data in order to discover patterns and linkages.

5.3 GIS Data Output

The outputs of GIS analysis can be displayed as maps, charts, reports or a combination of these. These are all examples of GIS output. The GIS output is categorized as cartographic and non-cartographic. Cartography is the best way to represent geographic features/phenomena. It's useful to use non-cartographic outputs for summarising tabular data and documenting any value that was determined by geography. The GIS can generate either a hard-copy or a soft-copy map. Additionally, it serves as an input for other GIS projects, the formats of which may be easily exported to ensure compatibility with other GIS environments, such as *.kml, *.kmz, *.dxf, *.shp, *.mdb and *.lyr.

Non-cartographic outputs are frequently required to present the characteristics and facts of geographical data. GIS generates a variety of non-cartographic outputs such as tables, charts and reports. Tables and charts contain geographical and non-spatial information about the features that are occasionally required to interpret the maps. Charts complement maps by visually summarising the data in the tables. With the use of line graphs, bar graphs, pie charts, area charts and scatter plots, it is possible to visualize the distribution, trend, pattern and relationship between statistical data.

Cartographic outputs are the major products of the GIS analysis projects. They aid in the precise determination of the answer to the questions involving *where*. Also, the maps are highly useful for showing the spatial distribution and shape of geographic features/phenomena. As such, it answers the questions involving *what*. Maps are also capable of representing the temporal dynamics of a place, a geographic features/phenomenon, etc., and as such answer the questions involving *when.* A good map can help the analyst make better inferences and draw correct conclusions. It is necessary to conduct cartographic data analysis in order to develop cartographic outputs (maps) in GIS. The data to be visualized must be described in order for the interpretation to be clear. It could be *qualitative* (nominal/categorical) or *quantitative* (interval and ratio). Ordinal data falls in the middle of the two categories of data. Table 5.1 lists the characteristics of these data types.

The cartographic depiction of geographical data for the intended audience requires the use of standard and fundamental map elements. They include but are not limited to, point symbols, line symbols, area symbols and text. Additionally, there are defined colours in cartography that must be used appropriately, such as blue for water and green for vegetation. Furthermore, colour-symbol fusions such as the blue line depicting a water channel (river) and alike must be considered.

Table 5.1 Characteristics of different data types used in the process of geovisualization

Data type	Measurement scale	Characteristics
Qualitative data	Nominal/Categorical	• Numerical labelling of data or categories of data • The data can only be categorized
	Ordinal	• Ranking of the elements in a particular order, without any numerical deductions possible • The data can be categorized and ranked
Quantitative data	Interval	• A numerical scale where the order of the variables is known as well as the difference between these variables can be deduced. The deductions are however based on the arbitrary zero value • The data can be categorized and ranked, and evenly spaced
	Ratio	• A variable measurement scale that not only produces the order of variables but also makes the difference between variables known along with information on the value of true zero • The data can be categorized, ranked, evenly spaced and has a natural zero

Bertin (1967) describes six kinds of *visual variables*. These are illustrated in Fig. 5.4. These variables can be used to distinguish between the fundamental symbols. Cartographers differentiate geographical objects using visual variables while adhering to the rules of cartographic grammar. These variables also have an effect on the audience's perception of the map understanding. As shown in Table 5.2, there is a significant relationship between the visual variables and the nature of the mapped data.

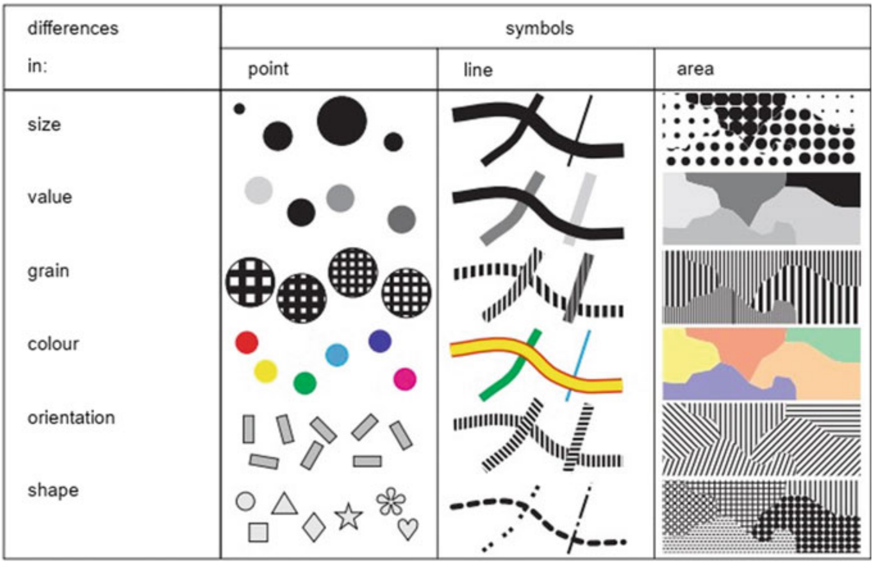


Fig. 5.4 Bertin's classification of visual variables. *Source* Plate 11 in Kraak and Ormeling (1996)

Table 5.2 Visual variables and their perception properties

Perception properties	Visual variables	Measurement scales			
		Nominal	Ordinal	Interval	Ratio
	Dimension of the plane	X	X	X	X
Order and quantities	Size		X	X	X
Order	Grey value Texture		X X	X X	
Equal importance	Colour and Hue Orientation Shape	X X X			

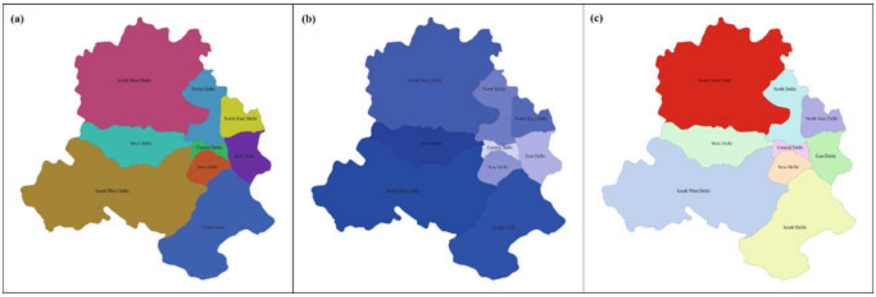


Fig. 5.5 Mapping qualitative data (Districts of NCT Delhi): **a** proper symbology, **b** misuse of tints of blue and **c** misuse of bright red

5.3.1 Cartographic Representation of the Qualitative Data

When it comes to depicting qualitative data in a cartographic output format, it is crucial to remember that each unit of qualitative data should receive the same level of attention. An illustration would be the creation of a map showing several watersheds in a specific area. Each watershed is expected to be equally stressed in this situation, and none of the watersheds should stand out. When random colours of comparable intensity are used to denote geographic units (administrative units like districts), as seen in Fig. 5.5a, geographic units can be discerned without undue emphasis on any single unit. On the other hand, depicting qualitative data in monochrome, as illustrated in Fig. 5.5b, or with colours of dissimilar intensity, as illustrated in Fig. 5.5c, results in an erroneous portrayal of quantitative data.

5.3.2 Cartographic Representation of the Quantitative Data

When it comes to the cartographic depiction of quantitative data, the map is intended to portray both the actual quantity/amount of a certain parameter and its geographical distribution. As such, the symbols that will be employed must possess quantitative perception qualities. To illustrate this, Fig. 5.6a depicts the population of different districts of Uttarakhand. The number of people is represented by dot symbols of varied sizes, and the map also depicts the population distribution across the state. If this map were instead shown in monochrome, as illustrated in Fig. 5.6b, the reader would be unable to discern the geographic distribution of the population or realize the fact that the population is not necessarily homogeneous within the geographic unit. On the other hand, a multi-colour scheme, as illustrated in Fig. 5.6c, influences how readers interpret the map because it does not define any absolute differences between a group of geographical units (districts) represented by the same colour.

When dealing with quantitative data that is relative, such as population density, it is important to understand that the numbers have a direct relationship to the area

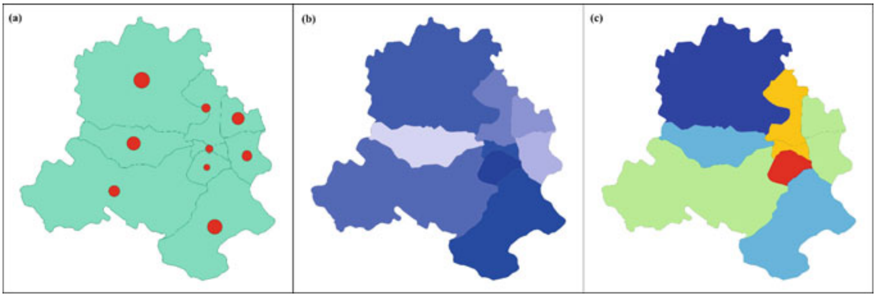


Fig. 5.6 Mapping absolute quantitative data (population of NCT Delhi): **a** proper symbology, **b** misuse of tints of blue and **c** misuse of colours

they represent. And presenting such data in the form of a map involves the use of sequential values of a certain colour to illustrate the hierarchy or order of the quantities depicted. The population density map of Uttarakhand depicted in Fig. 5.7 is a simple and effective demonstration of this. The map reader will naturally link lighter colour values with low population density and vice versa. However, if this type of data is represented in a multicolour scheme or with non-sequential colour values, the map reader will have difficulty associating perceptions and depictions.

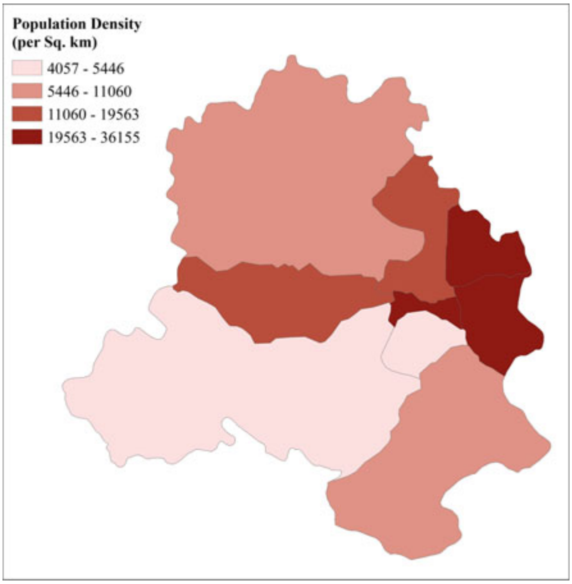


Fig. 5.7 Mapping relative quantitative data (population density of NCT Delhi)

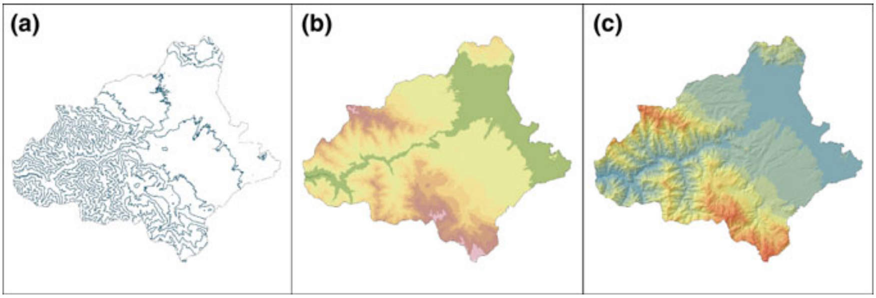


Fig. 5.8 Mapping terrain elevation (Baramulla, J&K, IN): **a** contour map, **b** layer tints map and **c** shaded-relief map

5.3.3 Mapping Terrain Elevation

Terrain maps can be created using digital elevation models or point data with *x, y* coordinates and height information (including peaks and other characteristic features of the terrain). These maps can take on a variety of forms. Contour maps are isoline maps in which each line connects points in the region with similar elevation values. It is a clear and simple method of representing elevation in traditional cartography and is the most frequently used. To visually enhance the contour map's information content, the space between adjacent contours can be filled with colours and values consistent with the conventions (green for low elevation and brown for high-elevation areas). This is referred to as *hypsometric* or *layer tinting* (Fig. 5.8).

The addition of shaded relief enhances the three-dimensional visual characteristics of the terrain maps in the advanced computer-based mapping process. Shaded relief creates shading effects in the virtual world by utilizing complete three-dimensional information. Because certain low-elevation features become obscured by high-elevation features in the three-dimensional view, it is critical that the user has access to interactive functions (panning, zooming, rotating, scaling, etc.) when dealing with such terrain data outputs.

5.3.4 Cartographic Representation of the Time Series Data

The growing availability of data collected over time and the emphasis of GIS professionals on studying changes induced by real-world processes have resulted in the research of geographical features/phenomena along the fourth dimension of time. To visualize temporal analysis on the data, it is critical that the user has access to a composite output set of multi-temporal maps in suitable print formats or via computer-based interactive graphics, simulations or similar options.

Temporal analysis is concerned with charting the changes that have occurred in a region across time. This could be modifications to the geometries (for spatial datasets) or the attributes (for non-spatial data), or both. To illustrate, one may simply visualize the administrative boundary dynamics. Additionally, urban regions have grown in size, demonstrating that urban expansion is a spatiotemporal process. Changes in land ownership also occur across time and are an example of non-spatial data change. On a broad level, there are three temporal data representing cartographic techniques, as shown in Fig. 5.9. *Single static map* uses specific graphic variables (visual variables like value) and symbols to indicate the change or specify an event. Maps depicting urban expansion might employ dark tints to depict older built-up areas and lighter tints to depict newly developed areas. *Series of static maps* represents the changes over the time in the form of temporal snapshots. These maps are put in a sequence, and the change is perceived by going through the individual maps in succession. *Simulation* (or *animated maps*) help perceive the changes in the region by displaying the sequence of the snapshots as successive frames in a videographic format. To view the results of simulations or animations, interactive tools must be available. Viewing the animation does not immediately clarify the situation for the viewer. As a result, the software packages used to display the simulations include features such as 'pause' (to look at the particular frame), '(fast-)forward' and '(fast-)backward', 'step-by-step display'.

Time series data mapping needs the use of symbols that indicate the changes in the data. For instance, arrows are used to illustrate movement, and the size of the arrows might indicate the amount of the shift. Additionally, size modifications can be

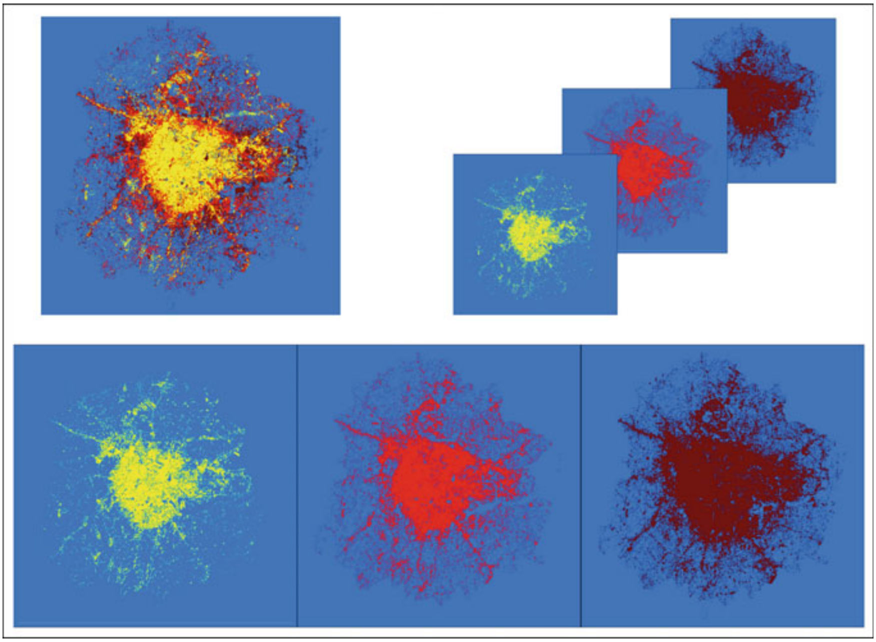


Fig. 5.9 Mapping time series data

made to point and line symbols to indicate growth or decline through time. Historic maps may feature special point symbols to show dynamic events, such as *lightning* for ‘riots’, *crossed swords* for ‘battles’ and so forth.

5.4 Conclusion

Geovisualization is a process for visualizing the entire process of making decisions about the earth’s resources, features and phenomena. It involves all processes, from the formulation of the question to the final analysis and presentation of the results, through the exploration, analysis and processing of data, as well as the synthesis and presentation of facts. Geovisualization is not restricted to the generation of final maps or non-cartographic outputs from GIS projects, but is also a useful component of the project’s intermediate inspections. Two- and three-dimensional graphics can be used to visualize spatial data. These graphical representations can be created using both cartographic and non-cartographic outputs. Cartographic representations must incorporate cartographic principles and language, as well as a range of map elements, such as scale, title, subtitle, legend, data sources and map layout. Aesthetics of the map must be considered during the process. The map layout, visual hierarchy, and usage of visual variables, among other elements, all contribute to the map’s ease of comprehension and ability to transmit the necessary information.

Questions

1. Define geovisualization and explain its relevance at different phases of a GIS project.
2. What is non-cartographic data output?
3. What six types of visual variables did Bertin classify in 1967?
4. Which mapping technique is appropriate for visualizing several watersheds of a certain region?
5. What approaches are used to represent the terrain features?

References

Bertin J (1967) Semiology graphique. Mouton, Den Haag, The Netherlands, pp 460–466

Kraak MJ, Ormeling FJ (1996) Cartography: visualization of spatial data. Addison-Wesley Longman, London, U.K.

Singh RB, Kumar D (2013) Land-use planning for sustainable development using geoinformatics: policy implications for drylands. In: Developments in soil classification, land use planning and policy implications. Springer, Dordrecht, pp 563–575

Chapter 6
Spatial Data Analysis

Abstract In the preceding chapters, we have explained the GIS data models, data types, data entry and editing in the Geographic Information System. In addition, digital data can be exploited for mapping and extracting information through spatial analysis for effective management of geographic objects/features/phenomena and identifying the new associations, thereby enhancing our knowledge of the real world. After reading this chapter you will be able to appreciate the following:

- The necessity for spatial analysis
- Analytical capabilities of GIS
- Methods of spatial analysis using vector data
- Methods of spatial analysis using raster data
- The measurement from the datasets of several feature types.

Keywords Spatial analysis · Analytical capabilities of GIS · Vector data analysis · Raster data analysis

6.1 Introduction

Spatial analysis can be defined as the process of extracting new information from existing stored data that provides new insights. In addition to map creation, spatial analysis is one of the chief capabilities of the Geographic Information System (GIS). It is a fundamental component of a GIS that enables an in-depth examination of the topological and geometric properties of one or more datasets. GIS is centred on spatial analysis in numerous ways (Longley et al. 2001). The spatial analysis comprises all the transformations, manipulations and methods that can be executed over the geographic data in order to augment its value, facilitate decision-making, and discover patterns and irregularities that are not usually discernible. In other words, spatial analysis refers to transformation of raw data into valuable information in pursuit of scientific analyses or improved decision-making. The analytical capabilities of GIS use spatial and non-spatial (attribute) data to address the problems of spatial relevance. Analysis is the process of resolving and separating the geographic system into its constituent units to elucidate their essence and structural interconnections and to ascertain the underlying functional principles. It relies on logical

M. Kumar et al., *Geographic Information Systems in Urban Planning and Management*, Advances in Geographical and Environmental Sciences,
https://doi.org/10.1007/978-981-19-7855-5_6

Fig. 6.1 Rationale and process of spatial analysis

connections between geographic features and their attributes, as well as operational procedures based on the spatial relationships between the features (Fig. 6.1).

Typically, the spatial analysis consists of five steps: defining the objective, preparing the data, selecting the appropriate tools and techniques, conducting the research and estimating the results. The real world offers numerous questions regarding the physical, socioeconomic and environmental processes/phenomena. There are countless questions about our actual system. To answer these questions, it is pertinent to create spatial simulations of real-world systems using spatial analyses based on various types of dataset relationships. These simulations reveal the primary patterns in the geographic data, thereby making new information accessible. To resolve various spatial queries, numerous interconnected factors and standards are required; the models can be crafted as the result of a single- or multilayer spatial analysis. In addition to attribute query and spatial query, spatial analysis in GIS also involves the generating new datasets from the primary databases. On the basis of vector or raster databases, or blend of both, spatial analysis can be conducted. In this chapter, various spatial analysis techniques, such as single/multilayer operations/overlays, spatial modelling, geometric modelling, point pattern analysis, network analysis, surface analysis and raster/grid analysis, are discoursed in depth. In addition, this chapter focuses on analytical functions which can serve as application model building blocks. Hopefully, it will be explicit to the student that these operations can be mingled in a variety of ways to perform elaborate analysis.

6.2 Analytical Capabilities of GIS

There are numerous ways to categorize the analytical capabilities of a GIS. According to Aronoff (1989), numerous analytical functions can be performed in GIS, including classification, retrieval, measurement, overlay, neighbourhood and connectivity functions (Fig. 6.2a). *Classification, retrieval and measurement functions* are performed on a single (vector or raster) data layer, frequently utilizing the associated attribute data. Based on attribute values or attribute ranges, classification permits the interpreting of features to a class (defining data patterns). Pixels can be classified as representing different crops based on the reflectance characteristics found in a raster. The retrieval functions enable a selective search for data. Thus, we could retrieve all agricultural fields where a specific crop is cultivated. Measuring functions permit the computation of distances, lengths and areas. In GIS, the most frequently used functions are *overlay functions*. They enable the grouping of two (or more) spatial data layers by juxtaposing their position and treating overlapping and non-overlapping regions differently. Numerous GIS packages offer overlays via algebraic language, where overlay functions are expressed as a formula with data layers as their arguments. Using overlay functions, operators such as intersection, union, difference and complement can be implemented. Unlike overlays, *neighbourhood functions* analyse the attributes of the region adjacent to the location of a feature. The neighbourhood functions scan the vicinity of particular features and compute value for the feature location. Search capabilities enable the extraction of features that lie within a specified search window. Search window may take the shape of a rectangle, a circle or a polygon. The generation of buffer zones is one of the most well-known neighbourhood functions of GIS. It establishes a two-dimensional buffer around particular feature(s). The width of the buffer can be fixed or adjustable. Interpolation functions estimate unidentified values based on identified values at a few neighbourhood locations. This characteristically is suitable for continuous fields, such as elevation, when the stored data cannot yield a direct answer for the desired locations. Topographical functions decipher the characteristics of an area by considering its nearby surroundings as well. Examples include slope calculations on digital terrain models (continuous field data). *Connectivity functions* rely on networks, such as road networks, waterways in coastal areas and communication lines in telecommunication. These network systems illustrate the connections between geographic features. Contiguity assesses a property of a collection of adjacent geographical elements. A satellite image can be analysed to find a forest of a particular shape and size using the contiguity functions. In GIS, *network analyses* can be used to compute the parameters of connected linear features. The network can include roadways, public transportation routes, electricity lines and other means of transportation. For routing purposes, the analyses of these network systems may involve calculating the shortest paths between two-point locations in the network. Further it may include finding all locations that lie within a specified distance or time with respect to a starting location, for allocation purposes, and determining the network's capacity for transport between a specified source and sink location. Visibility functions are also included in this list, as they

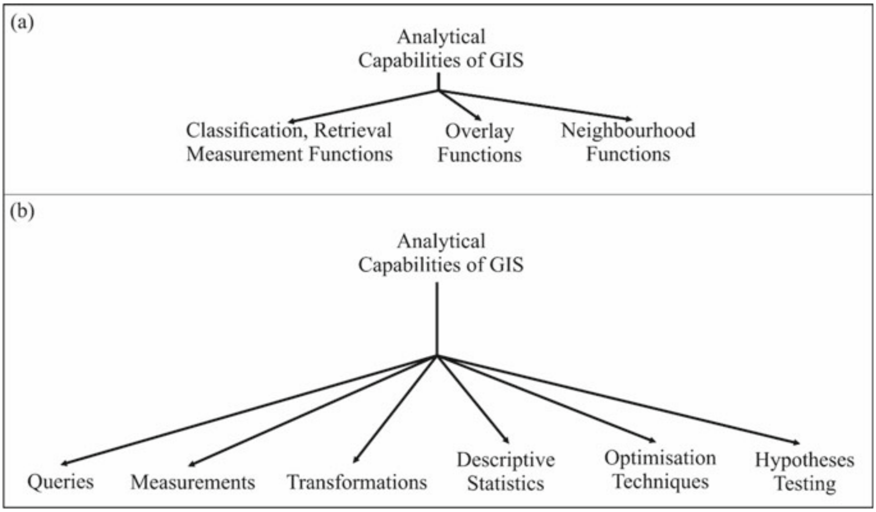


Fig. 6.2 Analytical capabilities of GIS: **a** Aronoff (1989) and **b** Longley et al. (2001)

are used to compute the points visible from a given location (viewshed modelling or viewshed mapping) with the help of a digital terrain model.

Longley et al. (2001) include six methods including queries, measurements, transformations, descriptive summaries, optimization techniques and hypothesis testing in spatial analysis (Fig. 6.2b). *Query* is a common and fundamental form of analysis, in which GIS responds to the simple questions posed by the user. No modifications are made to the data, and no new data is generated. The analyses range from simple query operations such as 'how many buildings are located within 10 km of location A' to complex questions such as 'which is the closest city to Srinagar heading north?', where the response depends on the capability of the system to comprehend what the user intends to mean by 'heading north'. *Measurements* are the numbers (numeric calculations) which characterize several dimensions of geographic data. They comprise measurements of simple attributes of an object/feature, such as lengths, areas and shapes, as well as measuring the relative variables with respect to two objects, such as distance and direction. *Transformation*, another basic spatial analyses technique, modifies the datasets by merging or comparing them to generate new datasets and, ultimately, new information. Transformation employs geometric, arithmetic or logical rules and comprise procedures that change rasters into vectors and vice-versa. Additionally, transformation can produce fields from a collection of objects or identify collection of objects in fields. In one or two numbers, *descriptive summaries* attempt to capture the core of a dataset. It is the spatial equivalent of the commonly employed descriptive statistics in statistical analysis, like mean and standard deviation. *Optimization* techniques are normative in nature and intended for selecting optimal locations of objects based on a set of predefined standards. These are utilized extensively in market-oriented studies, the delivery industry, and in a multitude of additional settings. *Hypothesis testing* refers to the method of drawing conclusions about the entire population based on the results of a small sample. On the basis of a sample, it enables researcher, for instance, to analyse that if a pattern of points may have formed by chance. The testing of hypotheses is the foundation of inferential statistics and heart of the statistical analysis, but its application to spatial data is considerably difficult.

6.3 Vector Data Analysis

The vector data model uses and integrates points and their *x, y* coordinates to create geometry of the features consisting of points, lines and polygons. The spatial geometries serve as inputs for vector data analysis. The precision of data analyses is contingent upon the precision of these features per se, in terms of location and shape, as well as their topological status. Spatial data analysis of the vector data involves Buffering, Overlay, Distance Measurement, Pattern Analysis, Feature Manipulation, etc.

Buffering refers to generating output polygon layer(s) with a zone (or zones) of a definite breadth surrounding a point, line, or polygon feature. It is ideally suitable for defining the area of influence surrounding a feature. It is predicated on the notion of proximity. It generates two zones, one that is within a predetermined distance of a particular feature and the other that is outside of that distance. The region inside the predetermined distance is the buffer zone. Typically, GIS modifies the value of an attribute to distinguish buffer zone from the region beyond buffer zone. The two zones are assigned the nominal designation values. In addition to the nominal designations, no additional attribute data modified. Point, line and polygon features can be subject to buffering operations (Fig. 6.3). Buffering creates circular buffer zones around points, extended buffer zones around line segments and buffer zones that extend beyond the perimeter of the polygons.

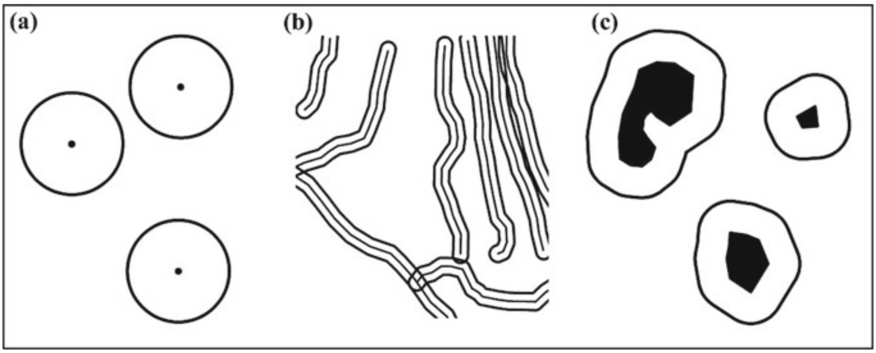


Fig. 6.3 Buffering vector features: **a** point feature buffering, **b** line feature buffering and **c** polygon feature buffering

Several buffering variations can be implemented in the GIS as per the needs of analysis (as depicted in Fig. 6.4). The buffer distance is not always required to be same but it can be variable. A feature may have only one buffer zone or multiple buffer zones (with constant or variable radial intervals). The buffering about a line feature does not need to be on both sides of the line; it could be placed on the left or the right side only. It is possible to extend buffer zones of polygons either inwards or outwards from the boundaries of the polygon. The boundaries of the overlapping buffer zones may be intact such that every buffer zone can be analysed separately or these boundaries can be dissolved to establish a single buffer zone with no overlapping areas in-between.

Overlay Operations combine the geometry and characteristics of two or more vector layers to produce the resultant layer. The shape of features in the resultant layer reflects the intersection of features from input layers. This is illustrated in Fig. 6.5. Each output feature has a unique combination of qualities from the input layers that vary from its neighbours.

The input vector layers used in an overlay operation must be georeferenced and based on a single coordinate system, which is a requirement for overlay in a Geographic Information System. Inputs for overlay operations can be the point, line or polygon layers, and the output is generally the lower-dimensional feature type; i.e.

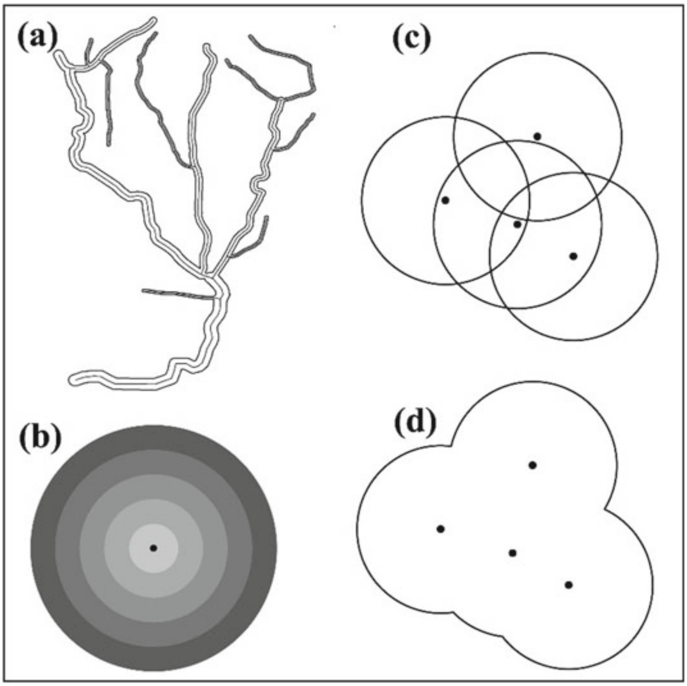


Fig. 6.4 Variants of buffering: **a** variable buffer distances, **b** multi-ring buffer, **c** undissolved buffer zones and **d** dissolved buffer zones

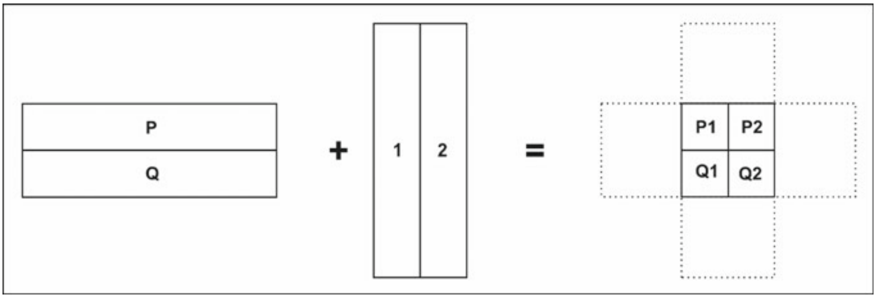


Fig. 6.5 Overlaying vector features

when polygon and line layers are used as inputs for the overlay operation, the output will be a line layer, and when line and point layers are used as inputs, the output will be a point layer. Let's examine three frequent GIS overlay operations: line-in-polygon, point-in-polygon and polygon-on-polygon. In a line-in-polygon overlay, the output comprises of line features identical to the input layer, with the exception that every line feature is cut up by the polygon borders used in the operation. The output layer consists of additional line segments as compared to the input layer (Fig. 6.6a). Every line segment of the output mixes properties of the input layer and the polygon underneath it. In a point-in-polygon overlay, identical point features from the input layer are replicated in the output, nevertheless each point is allocated the properties of the polygon in which it lies (Fig. 6.6b). Polygon-on-polygon, using two polygon layers, is the most frequent overlay operation in GIS. The output creates a fresh set of polygons by combining the polygon margins from the input layer and the overlay layer (Fig. 6.6c). New polygons incorporate properties from both levels, and these properties are distinct from those of neighbouring polygons.

In the GIS, several overlay operators are employed. They may pop up under distinct names in various GIS packages, but they are fundamentally based on the Boolean logical operators including union (OR), intersection (AND) and symmetrical difference (XOR). Only if the area extents of the inputs differ does the choice of overlay method become significant. If the input and output layers are of the same area extent, then the area extent of the larger input layer applies to both. Union (OR) maintains all characteristics of the inputs, as depicted in Fig. 6.7a. The output area extent of union is a combination of the area extents of two input layers. Intersect (AND) only maintains the properties that lie inside the common area extent of the input layers. This is illustrated in Fig. 6.7b. Symmetrical difference (XOR) maintains characteristics within the region extent shared by just one of the inputs (Fig. 6.7c). In terms of the output area extent, the symmetrical difference is the opposite of intersect.

Distance measurement is the process of determining the Euclidean (straight-line) distance between two features. Measurements can be taken from points in one layer to points in other layer, or from each point in one layer to its closest point or line in the next layer. Distance measurements are kept in a field in both instances. Data analysis can directly employ these distance metrics. Distance measurements can

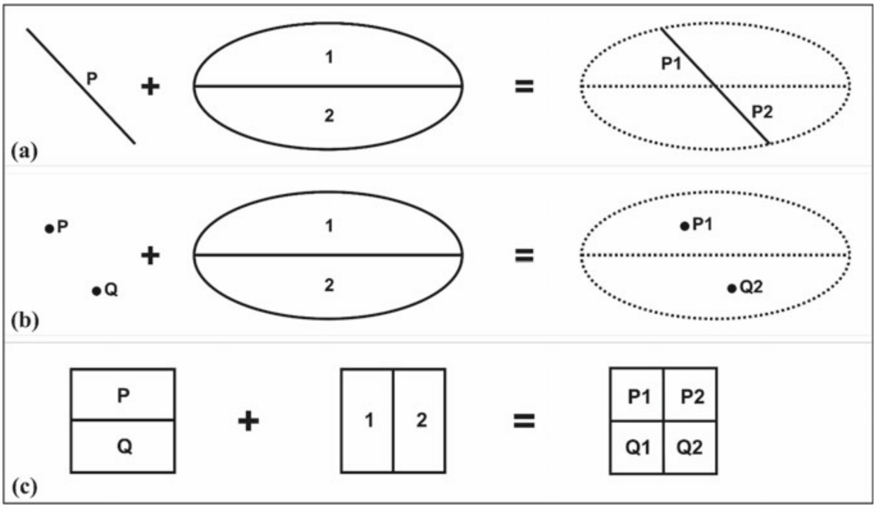


Fig. 6.6 Overlaying vector features: **a** line-in-polygon overlay, **b** point-in-polygon overlay and **c** polygon-in-polygon overlay

also be required as data analysis inputs. The input for the gravity model, a spatial interaction model often used in migration research, business applications, etc., is the distance between sites (Fig. 6.8).

GIS also offers capabilities for modifying and managing vector features within one or more data layers. These tools are frequently used for data preparation and data analysis, but they do not merge the geometry and attributes of input layers into one unified layer. These tools include *dissolve*, which aggregates features in a vector layer with a common attribute value(s); *clip*, which produces a fresh layer containing only the features of the input layer, plus their attributes, that lie within the extensions of the clip layer; *append* that generates a fresh layer by combining two or more layers representing the same feature with similar properties; *select* that generates new data layer containing the features specified by a user-defined query; *eliminate* that produces a fresh layer by deleting characteristics that satisfy a user-defined query; cut-and-paste-based *update* that substitutes the input layer with the updated layer and its feature; *erase*, which discards from the input layer any features that lie within the extensions of the erase layer; and *split*, which splits the input layer into two or more vector layers.

6.4 Raster Data Analysis

Raster data model uses an even grid of cells to transform the space. Value within each grid cell is a representation of the geospatial phenomenon at that cell location. The fundamental data structure of a raster with fixed cell placements is computationally

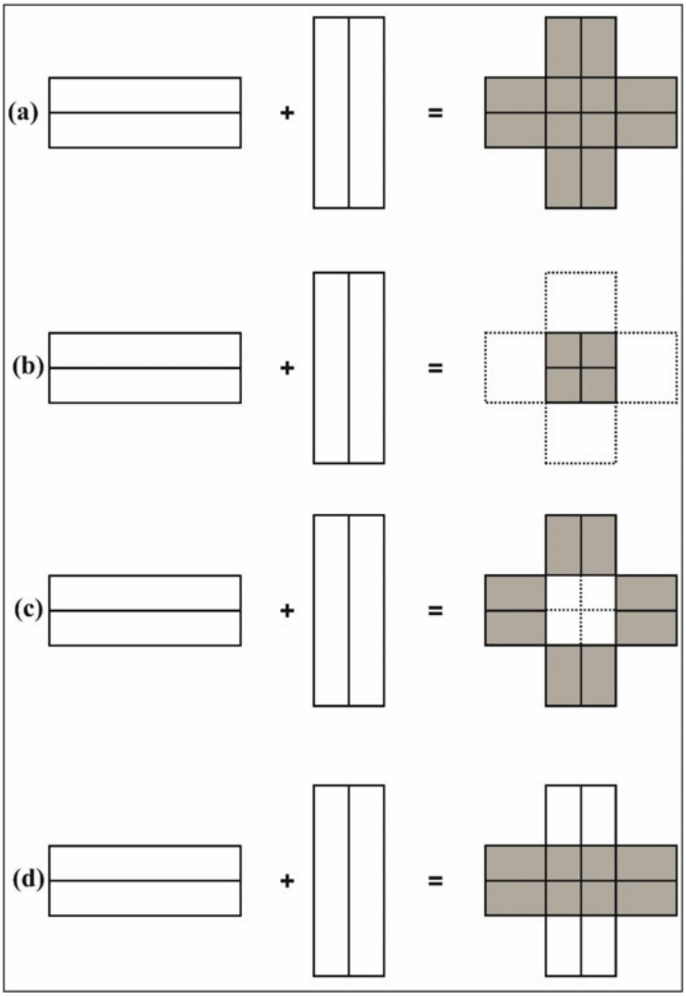


Fig. 6.7 Overlaying operators in GIS: **a** union, **b** intersection, **c** symmetrical difference and **d** identity

efficient and enables conducting a multitude of data analyses. As opposed to vector data analyses, which use point, line, and polygon geometries, raster data analyses use cells and raster layers. Individual cells, groups of cells or complete raster layers can be analysed while conducting raster data analysis. Raster data analysis may be carried out on a single raster or a group of several raster data layers. Given the fact that a raster operation may require two or more raster layers, it is required to describe the *raster data analysis environment* by defining its area extent and output cell size. The analysis space can resemble with a particular raster, an area specified by its minimum and maximum *x, y* coordinates. The area extent can also be derived using the union

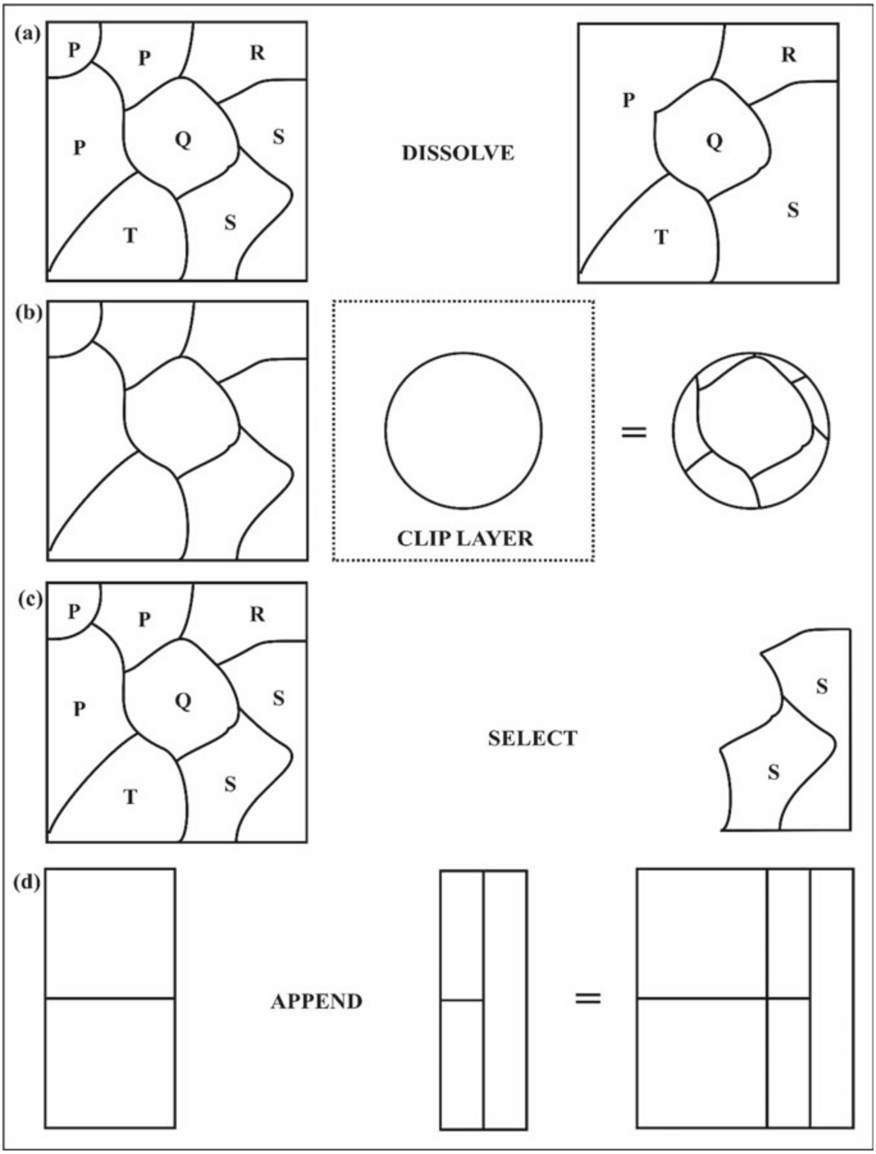


Fig. 6.8 Data management and modification in GIS: **a** dissolve, **b** clip, **c** select, **d** append, **e** eliminate, **f** update, **g** erase and **h** split

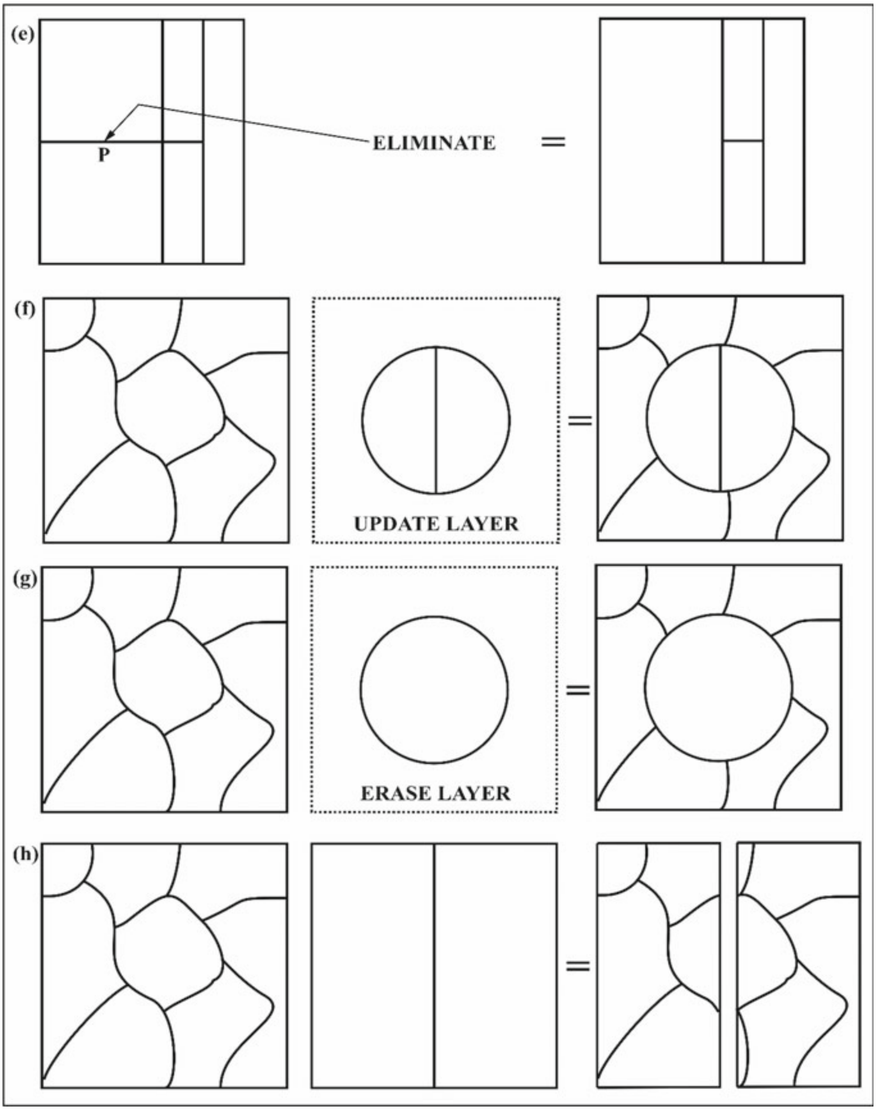


Fig. 6.8 (continued)

or intersection of the raster layers. Union employs an area extent that covers all input raster layers, but the intersection employs an area extent shared by all input raster layers. A vector layer or a raster analysis mask can also specify the region extent for analysis. It restricts analysis to its coverage region. Regarding the output cell size, GIS adheres to the principle that cell size of the output should match that of the input with the largest cell size. The output cell size can be defined at any scale considered

Table 6.1 Various functional operators used in GIS

Type	Operator
Arithmetic operators	+, −, ×, /, integer, floating-point, absolute
Logarithmic operators	exponentials, logarithms
Trigonometric operators	sin, cos, tan, arcsin, arccos, arctan
Power operators	square, cube, square-root, cube-root

appropriate; however, it is commonly configured to be equal to or greater than the largest cell size from among the input raster layers.

Local operations in raster data analyses are performed on cell-by-cell basis. Cell values in the fresh raster layer are generated either by a function connecting the input raster to the output raster or are allocated by a classification table. Local operations can generate a new raster from a single or multiple input raster layers. If a single raster is provided as input to the local operation, each cell value in the output raster is computed according to cell value of input raster at the matching place. This function that is used to derive cell values can be a GIS tool, a mathematical operator, or a constant. Several operators are available in the GIS, as shown in Table 6.1, and the user can also define the functions of raster analysis manually according to his requirement.

Multiple input raster layers enable a wider range of local operations than a single input raster. Local operations involving several raster layers are also known as compositing, overlaying and or superimposing (Tomlin 1990). In addition to the mathematical operator used for the single raster analysis, additional criteria based on the cell values or their frequencies in the input raster layers can be applied to the multiple raster analysis. Summary statistics, such as maximum, minimum (illustrated in Fig. 6.9), range, sum, mean, median and standard deviation, apply to raster layers containing numeric data.

For raster layers having numerical or categorical cell values, statistics such as majority (illustrated in Fig. 6.10), minority, and a number of unique values are applicable. If cells in one of the input raster layers have no data, the cells carry no data in the output raster by default.

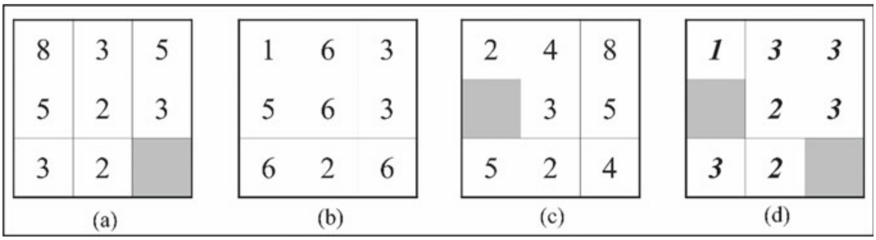


Fig. 6.9 Local operations on raster data: cell values of (**d**) are calculated as the MINIMUM of three input raster layers (**a**, **b**, **c**) and the grey-shaded cells have no data values

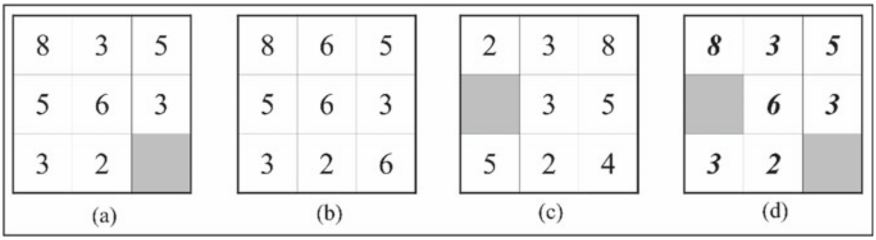


Fig. 6.10 Local operations on raster data: cell values of (**d**) are calculated as the MAJORITY of three input raster layers (**a**, **b**, **c**) and the grey-shaded cells have no data values

Reclassification also called recoding or transforming is another local operation that can generate a new raster by classification of the existing raster. While reclassifying a raster, new cell values are assigned to existing cell values. To illustrate this, a land use raster having forest, river, agriculture and built-up classes can be reclassified as a raster having classes named 1, 2, 3 and 4, respectively. There is a one-to-one transformation of the classes between input raster and output raster. Alternatively, a group of classes in the input raster can be clubbed together and assigned a single value in the output raster. As in, value 1 may be assigned to the forest and agriculture land use classes of the raster to create a single class instead of two in the output raster.

Focal operations, often referred to as neighbourhood operations, are based on the cell values of a neighbourhood area (cells) for the distance and/or directional connection to the focal cell. The neighbourhood deliberated about here may be of several forms primarily rectangular (defined by its height and width in cells, such as a 5-by-5 region centred at one focal cell) and circular (extending from focal cell with a specific radius). These are shown in Fig. 6.11. The choice of neighbourhood shape is not restricted to geometric variants of the rectangular and circular forms. Several studies have benefited from utilizing irregular and discontinuous neighbourhoods. However, the GIS does not include non-geometric neighbourhood options.

In general, the neighbourhood operations are the statistical operations that the user may apply to the preexisting raster layer(s) to do spatial analysis. A neighbourhood operation commonly analyses the cell values in the neighbourhood to compute the value, which is subsequently assigned to the focal cell. To perform neighbourhood analysis on a raster, the focal cell is transferred from one cell to the next until all cells have been covered. In peripheral regions, when a little or a substantial portion of the defined neighbourhood cannot be processed, alternative procedures for processing the focal cell values are implemented. A basic rule is to compute only using cell values accessible in the neighbourhood. Summary statistics, including maximum, minimum, range, total, mean, median, and standard deviation, and tabulation of measurements such as majority, minority and variety, can be created using focal statistics. These measurements and statistics are identical to those derived from local operations involving multiple raster layers.

Zones are groups of cells in a raster that have the same values (Chang 2018). They can be contiguous (cells that are physically connected as in a watershed raster), or

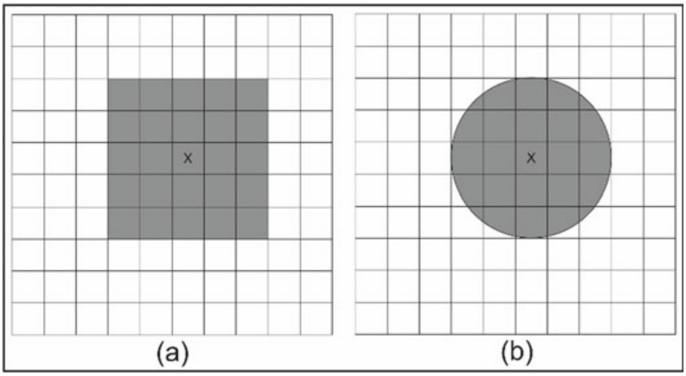


Fig. 6.11 Focal neighbourhoods for raster data analysis: **a** rectangular neighbourhood and **b** circular neighbourhood

they can be non-contiguous (cells dispersed over separate regions as in a land use raster). The *zonal operations* use these groups of cells (zones) and apply the zonal statistical procedures to the raster layer(s) to achieve the goals of spatial analysis. The zonal operation can be done with either one or two raster layers. When a single raster is used as input, zonal statistics measure the geometry of individual zones in the raster, like the area (the number of cells in the zone times the size of each cell), the perimeter (the length of the boundary of a contiguous zone and the sum of the lengths of the boundaries of several non-contiguous zones), the thickness (the diameter of the largest circle that can be drawn in each zone) and the centroid (the geometric centre of a zone, located at the intersection of the major axis and the minor axis of an ellipse that best approximates the zone). Alternatively, if two raster layers are used as input in a zonal operation, the result is a raster that summarizes the cell values of the input raster for each zone in the zonal raster. Statistics and measurements like area, minimum, maximum, sum, range, mean, standard deviation, median, majority, minority and variety may be used to make the summary. To illustrate this, Fig. 6.12 shows the zonal statistics computing the sum of the cell values of the zone.

Even though the local, focal and zonal operations cover the most important parts of raster data analysis, many other raster data operations are used for spatial analysis with raster data. Here, it's worth giving a brief description of these data operations. The area extents of the raster data obtained from the internet must typically be modified to accommodate the study area. Depending on the size of the research region, this may necessitate the clipping of a bigger raster file or the mosaicking of many smaller raster layers to create a larger raster. In order to clip a raster, we may define an analysis mask (study area vector) or the minimum and maximum *x, y* coordinates of a rectangular region for the analysis environment and then employ the larger raster as the input. Mosaic is a tool capable of combining several input raster layers into a single output raster. There are several possibilities for providing the values of falling cells in overlapping regions of various raster layers inside the raster overlay.

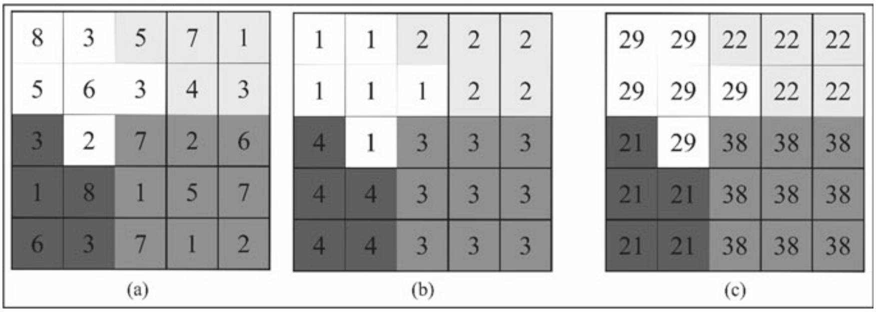


Fig. 6.12 Zonal statistics: cell values of (**c**) are calculated as the SUM of cell values of input raster layers (**a** and **b**)

Using a graphical object or a query expression, *raster data extraction* produces a new raster by extracting data from an existing raster. A dataset, a visual object (a collection of points or polygons such as rectangle and circle), or both, can be used to define the extraction space. If the data collection is a layer of point features, the extraction tool takes values from the point locations and adds them to a fresh field in the attribute table of point layer. If the data collection is a raster or polygon layer, the extraction tool derives the cell values inside the region indicated by the raster or polygon layer and assigns no data to cells outside of the area. In addition, extract-by-attribute builds a new raster with cell values that match the user-defined query expression.

6.5 Conclusion

Geographic Information System (GIS) has revolutionized the settings in which the geographical problems may be addressed and resolved. Spatial data analysis is the most significant function of GIS software. It offers an extensive variety of applications to refine an understanding of geographic phenomena. Spatial study of the spatial events in the digital world enables a panoramic visual and comprehensive understanding of the processes and associated factors. A document of *GI2000: Towards a European Geographic Information Infrastructure*, published by DG XIII of the European Commission (EGII, 1995) projected that spatial analysis will be used in an expanding number of application fields in the near future as GIS data manipulation tools grow increasingly powerful and user-friendly. Presently, GIS is supplying user-friendly tools and analytical possibilities for the full application of spatial analysis to the challenges at hand. Spatial analysis in GIS is obviously and function of the data model that the user works upon. Vector data analysis and raster data analysis are the two forms of GIS analyses. These are processed independently since GIS programme cannot execute these simultaneously in a single operation. Although some GIS applications enable the use of vector data in some raster data operations (e.g. extraction procedures), the data is transformed into raster data before the operation is started. Each GIS project is distinct in terms of data sources and objectives. Furthermore, vector data may be simply turned into raster data and vice-versa. We must thus choose the form of data analysis that is effective and suitable.

Questions

1. What is spatial analysis, and how is it generally carried out?
2. How does Aronoff enumerate the analytical capabilities of GIS?
3. What are different variants of buffering in vector data analysis?
4. Discuss overlaying in vector data analysis and raster data analysis?
5. How are local, focal and zonal operations different in raster data analysis?

References

Aronoff S (1989) Geographical information systems: a management perspective. WDL Publications, Ottawa, Canada

Chang KT (2018) Introduction to geographic information systems-ninth edition. McGraw-Hill Education, 2 Penn Plaza, New York, NY 10121. ISBN: 978-1-259-92964-9

Longley PA, Goodchild MF, Maguire DJ, Rhind DW (2001) Geographic information systems and science. Wiley, London

Tomlin CD (1990) Geographic information systems and cartographic modeling. Prentice Hall, Englewood Cliffs, NJ

Chapter 7
Non-spatial Data Management

Abstract In previous units, you learned about GIS concepts, spatial data, data models and data structures. In a typical GIS, both spatial and non-spatial data are used. Spatial data describes the location of features, whereas non-spatial data describes their characteristics. Attribute data is another name for non-spatial data. Geospatial data is a combination of both types of data. It means that both (spatial and non-spatial) data are required for a GIS to function properly. In this chapter, we will go over the concept of non-spatial data in depth. On the completion of this chapter students will be able to learn the following concepts about non-spatial data/attribute data:

- Introduction
- Basics of the Non-Spatial Data in geoinformatics
- Concept of the Relational Model
- Understanding the concept of joins, relates and relationship classes within GIS
- Spatial Join
- Concept of the entry of attribute (tabular data)
- Fields and attribute data modification.

Keywords Non-spatial data · Database · SSURGO · Normalization · Relationship · Data entry

7.1 Introduction

GIS deals with two types of data, first one is spatial data and the second one is non-spatial data. This is also known as geospatial data, refers to any data that is related to or contains information about a specific location on the earth's surface. Meanwhile the characteristic of the spatial features is described by the non-spatial data. This is not geographically bound.

This chapter is intended to provide detailed account of non-spatial data. By leveraging the power of spatial relationships, geospatial data can assist you in making better decisions. Making sense of the differences between spatial and non-spatial

M. Kumar et al., *Geographic Information Systems in Urban Planning and Management*, Advances in Geographical and Environmental Sciences,
https://doi.org/10.1007/978-981-19-7855-5_7

data will allow you to make even better decisions while also expanding your data knowledge (Chang 2019). We'll learn a lot about non-spatial data after we understand the differences between these two types of data.

7.1.1 Spatial Data

A geographical space is represented by the spatial data. They are distinguished by the use of points, lines and polygons. The point data represents the positional characteristics of some geographical features on the map, such as schools, hospitals, wells, towns, villages and so on; on a dimensionless scale, but with reference to location, we use points.

Lines are also used to represent linear features such as roads, railway lines, canals, rivers, power and communication lines and so on. Polygons are made up of a series of interconnected lines that encircle a specific area and are used to depict area features such as administrative units (countries, districts, states, blocks); land use types (cultivated area, forest land, degraded/wastelands, pastures, etc.); and features such as ponds and lakes. The spatial data is categorized in two primary formats:

(a) *Vector Data*

Vector data is graphical representations of reality. Vector data is classified into three types: points, lines and polygons. Lines are formed by connecting points, and polygons are formed by connecting lines that form an enclosed area. Vectors are representations of generalizations of objects or features on the earth's surface. Vector data and the shapefiles (.shp) file format are sometimes used interchangeably. Vector data is typically saved in ".shp" files. Figure 7.1 is representing the vector data.

(b) *Raster Data*

Raster data is information that is presented in the form of a grid of pixels. Each pixel in a raster has a value, which can be either a colour or a unit of measurement. This conveys information about the element under consideration. Rasters frequently refer to imagery. In the spatial world, however, this may specifically refer to orthoimage, which are photographs taken from satellites or other aerial devices. The quality of raster data varies depending on resolution and the task at hand. Figure 7.2 is representing the raster data.

7.1.2 Non-spatial Data

Non-spatial data, also known as attribute data, is information about spatial data. For example, if you have a map of your school's location, you can attach information such as the school's name, the subject stream it offers, the number of students in each class, the schedule of admissions, teaching, and examinations, available facilities such as

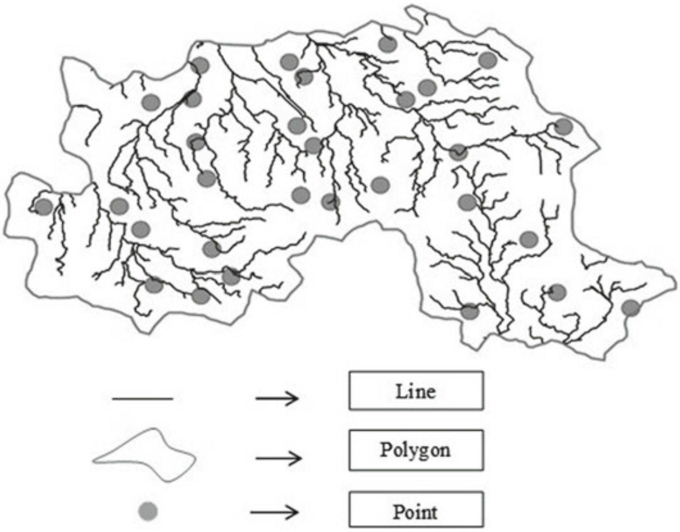


Fig. 7.1 Vector data (point, line, polygon)

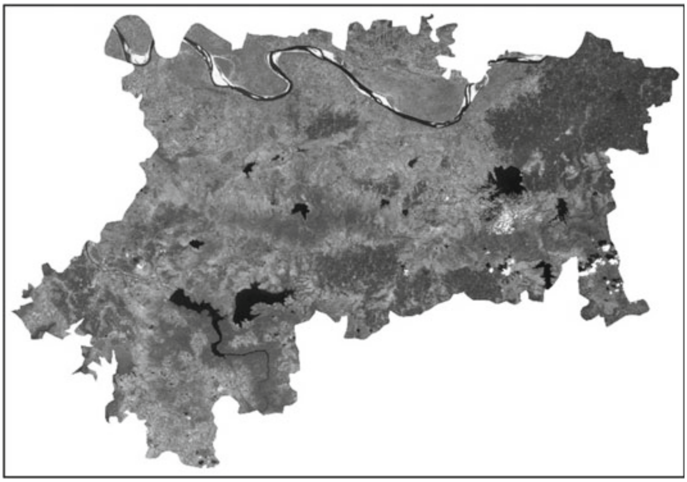


Fig. 7.2 Raster data

a library, laboratories, equipment and so on. In other words, you will be defining the attributes of the spatial data. As a result, non-spatial data is also known as attribute data. Sources of Geographic Data: Geographic data can be found in both analogue (maps and aerial photographs) and digital formats (scanned images).

An attribute may include both spatial and non-spatial data. Non-spatial data can help contextualize spatial data. It's identifying information. For example, this could

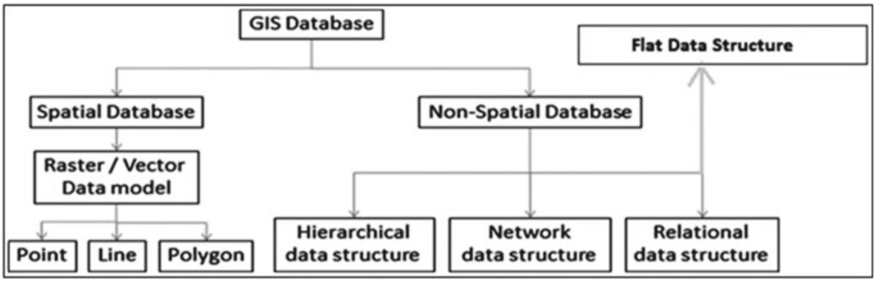


Fig. 7.3 Structural differences between various geospatial data

be someone's height. It has nothing to do with their geographical location. Figure 7.3 emphasizes the structural difference.

7.2 Non-spatial Data in GIS

In GIS, non-spatial data/attribute data is stored in tables. An attribute table is divided into rows and columns. Each row represents a spatial feature, each column a characteristic, and the intersection of a column and a row shows the value of a specific characteristic for a specific feature (Fig. 7.4). A row is also known as a record, and a column is known as a field.

7.2.1 Types of Attribute Tables

In GIS, there are two kinds of attribute tables for vector data. The feature attribute table, which has access to the feature geometry, is the first type. Each vector data contained the feature attribute with it. Feature ID of the attribute table is used to link the one feature with other in 'georelational data model'. Feature geometry is stored in the form of attribute table for object-based data. Meanwhile feature geometries of the vector data, for example length, perimeter, and area, is presented in the field of feature attribute table. Suppose a dataset has only a few attributes, a feature attribute

Label ID	pH	Depth	Fertility
1	6.3	12.7	High
2	5.2	11.3	Low

Row

Column

Fig. 7.4 Row and column of attribute table

table may be all that is required. But this doesn't happen always. Suppose a soil map consisted higher than hundred soil attribute properties. Repetition of the entries within an attribute table makes it too much complex and its quite time and computer memory consuming task. And this type of table is not so easy to handle. Hence, there is a need of another type of attribute table.

This new type of table is considered as non-spatial attribute table, which means it does not have direct access to the feature geometry but does have a field that connects it to the feature attribute table when necessary. Non-spatial data tables can be stored as delimited text files, dBASE files, Excel files, Access files or files managed by database software packages.

7.2.2 Database Management

Because a GIS contains feature attribute and non-spatial data tables, it necessitates the use of the "Database Management System" (DBMS) for the managing of the attribute tables. Database Management System (DBMS) could be understood as software package which allows us to create and manage databases (Oz 2004). Tools for the purpose of input, of data, searching of the data, retrieval, manipulation and output of data are provided by DBMS. For local databases, several GIS packages include DBMS tools. As example, ArcGIS manages personal geodatabases with Microsoft Access. Aside from its GIS applications, the use of a DBMS has other advantages. Big enterprises have to deal with big amount of data. To handle such big data, GIS is being used as integral part. Without taking the aid of GIS it will be difficult for such organizations and departments to handle such huge data. Hence keeping in mind, the applicability of the GIS, its totalistic performance could be enhanced if such geospatial technologies would be coupled with other contemporary and high yielding information technologies. Not only the database management tools hold the ability to manage the local databases but also several geospatial techniques are also able to access the remotely sensed database through its database connection. Those users who frequently use the data derived through the central database, for them it's critical.

For example, users of such types could receive the data from a national forest's headquarters office on a regular basis for the analysis. This explains the example of "client–server distributed database system" (Arvanitis et al. 2000). For example, this client could be state minister (say) conventionally send a request for retrieval of the data and to be processed on client's local computer system. As the adoptability of the cloud computing increases various option are arising to access the data from a centralized database. Nowadays web browser is being used as source to access the centralized database and to process it (Zhang et al. 2010).

7.2.3 *Attribute Data Types*

Attribute data could be categorized through data type. This described its way to be stored within GIS. The available data types vary depending on the GIS package. Most common attribute data types are number, text, date or 'Binary Large Object' (BLOB). Integers (without decimal) and floats (or floating points) are two types of numbers. Furthermore, an integer could be in the form of short/long; float could be in the form of single/double depending on computer memory capacity. 'BLOB' is not similar to traditional data types. Large block of data is stored in it in the form of binary system. Data access and retrieval from BLOP are more efficient and effective than other way. In an alternate way, it is possible to classify the data on the basis of measurement scale. With increasing degrees of sophistication, the measurement scale concept divides attribute data into nominal, ordinal, interval and ratio data (Stevens 1946).

Nominal data describes various types or categories of data, such as land use types or soil types. Ordinal data distinguishes data based on a ranking relationship. Soil erosion, for example, can be classified as severe, moderate or light. Interval data has well-defined intervals between values. A temperature of 70 °F, for example, is 10 °F warmer than 60 °F. The only difference between ratio data and interval data is that ratio data is based on a meaningful, or absolute, zero value. Since zero density is considered as absolute zero, hence population density could be considered as ratio data. Several tests, for example parametric and nonparametric tests, are intended to analyse the data at various scale. Hence it is crucial in statistical analysis to distinct the measurement scale. Nominal and ordinal data are examples of categorical data, while interval and ratio data are examples of numeric data.

7.3 The Relational Model

Database could be considered as group of the interrelated tables. There are four types of relational model, viz. flat file model, hierarchical model, network model and relational model (Jackson 1999). These all-relational models are shown in Fig. 7.5. Flat file is a considered as large table that contains all of the data. It is analogous to 'feature attribute table'. Spreadsheet could be considered as other example of it which contains only attribute type of data.

Second type of relational model is hierarchical type of databases which arrange and organizes the data at various levels and implement 'one-to-many' type of relationship among them. Figure 7.5 shows a simple example of the hierarchical levels of zoning, parcel and owner. Each level is divided into different branches based on the one-to-many relationship. Linkages between the nodes are represented in Fig. 7.5, i.e. a network database which creates connections across tables. Interlinkage among the tables should be acknowledged priorly and built within its respective database at designing time (Jackson 1999). This proceeds to complex and rigid database, as well

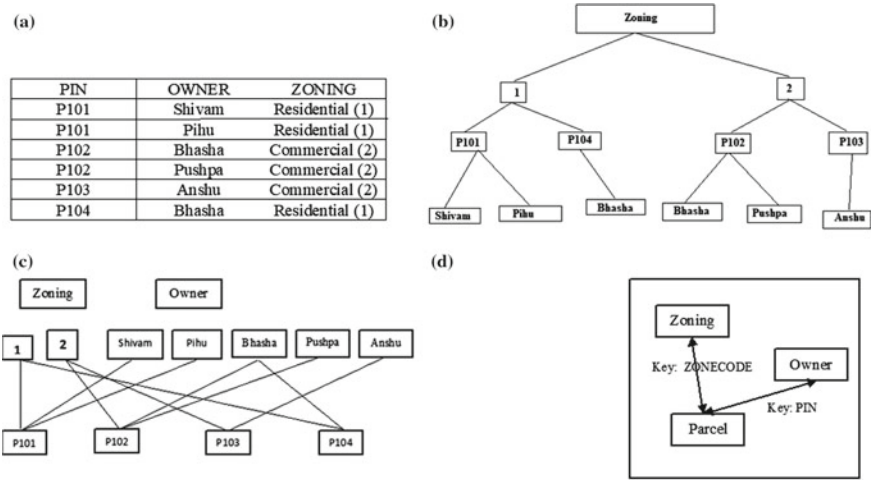


Fig. 7.5 **a** Flat model; **b** hierarchical model; **c**; network model **d** relational model

as limitations in database applications. For the purpose of database management most commercial and open-source GIS software are using the relational model (Codd 1970, 1990; Date 1995). Collection of the tables which is also acknowledged as relation is hence the relational database. It could be interlinked together through keys. Primary key is an important key which have either one or more than one attributes could be used uniquely for the identification of record within a table. It is not possible that the value of these primary key could ever be empty. These values are never ever altered. A foreign key is any attribute that refers to a primary key in another table. Name of the primary and foreign keys should not to be similar besides that case where both are serving the similar purpose. Figure 7.5 clearly depicts that, the common field which used for the interlinking of zone and parcel is zone code while PIN is the common field used for the interlinking of parcel and land owner. When used together, the fields could be used to link zoning and ownership. A relational database has two distinct advantages over other database designs (Carleton et al. 2005). Every table in the database should be formed, managed and manipulated independently of the others. With the passage of time geospatial technologies are becoming so popular which is in result to record and process the huge amount of data. Tables should be kept separate until and unless the analysis that should be performed by linking the multiple tables together. Relational database is found to be very efficient in the management and processing of the data because of the requirement of linking of the table are not permanent. Preplanning/advanced planning is still necessary, although relational database model is very simple and adaptable. GIS project might begin with a flat-database and with due processing it could migrate up to a relational database because as any project grew up it required to permit easier addition of the new data variables, due to this data volume of the project is increased (Blanton et al.

2006). GIS database (example of the spatial database) is referred as 'extended relational database' because it contains spatial data in addition to descriptive (attribute) data (Güting 1994). On the basis of this concept few scholars termed the GIS as "spatial database management system" (e.g. Shekhar and Chawla 2003). Georelational database management system is presented in Chap. 6. It deals with the spatial data within subsystem and attribute data within "Relational Database Management System" (RDBMS). Georelational database management system is able to correlate these two by using the pointer, e.g. Feature ID. Georelational database management system joins the spatial and attribute data into a common database and make easier to correlate and connect within these two data types. Besides the ability of GIS in feature attributes it should also be able to handle the geometrical features like points, lines and polygons and explore the spatial interrelationship (topographical relationship) between features. Hence process of linking becomes so important (Rigauz et al. 2002).

7.3.1 Example of Relational Database: SSURGO

Because of the benefits of relational databases, many government agencies have adopted them to store their data. For example, to manage the MAF/TIGER database, the 'United States Census Bureau' has also transferred from domestic database system to relational database system (Galdi 2005). The Natural Resources Conservation Service's (NRCS), Soil Survey Geographic (SSURGO) database is considered as example of the relational database system. NRCS collects the SSURGO data from mapping the field and save it as archives in the format by organizing the database. Such soil survey could be performed either for the whole country, part of country or multiple countries. This SSURGO database is the most extensive database of the USA. It includes both the spatial and tabular data. Spatial data collected from surveying the each and every area consisted the detailed information regarding soil maps. This soil map is prepared with the soil map unit made up with one or more than one non-contiguous polygon. For the purpose of developing the land use planning and management strategies, soil map should be considered as the smallest area unit of mapping. Description and of these tables and used key to connect them is very well provided by NRCS database. On first sight, one might observe the sheer size of the SSURGO database too much overwhelmed but on the basis of clear understanding of the relational database it will be easier to handle.

7.3.2 Normalization

Certain rules must be followed when creating a relational database, such as SSURGO. Normalization is an important rule. Normalization could be understood as process of decomposition of big complex tables into smaller tables units by preserving all the necessary properties and interlinkages (Vetter 1987). Performing the normalization of complex table is necessary to accomplish various goals:

- To remove the data redundancy in tables.
- To make sure that the data of the discrete tables should be maintained and keep it updated so that it could be easily linked together when required.
- Help to provide a distributed and organized database.

Let us understand the concept of normalization through an example. Figure 7.6 depicts four land parcels, and the attribute data for the parcels is displayed in Table 7.1. This table contains duplicate information: owner addresses of Bhasha are repeated, while residential and commercial zoning are entered two times. Records having changing values are also contained in the table. On the basis of parcel type, owner and his address could hold one or more than one values. Table 7.1, which is an unnormalized table, is difficult to manage or edit. Storing the attribute fields of owner and his address is not so easy in the initial steps. A change in ownership necessitates the updating of all attributes in the table. The same difficulty exists when it comes to operations like adding or deleting attribute values. The first step in normalization is represented by Fig. 7.7. This Fig. 7.7 is also known as the first normal form, has not multiple values in the attributes but the major problem of the redundancy of the data has worsened it. Except for the owner and his respective address P101 and P102 are identical. Address of the Bhasha appeared twice. Furthermore, the description of the zoning of residential and commercial fields each is listed thrice. For identifying the address of the owner, compound key is required which consists PIN and owner information. The second step in normalization is depicted in Fig. 7.8. Second step work upon three disintegrated tables parcel, owner and address information. First, to connect the land parcel with that of owner, PIN is working as a common field. Second, to connect the address with that of owner, owner's name is working as common field. The PIN and owner name can be used to establish a relationship between the parcel and address tables. Data redundancy in zone code and zoning of the fields is the only associated issue with second normal form. Figure 7.9 depicts the final step in normalization. A new table (say zone) is created to eliminate the data redundancy issue associated with zoning. To connect the parcel with that of zone tables, zone code will work as a common field. Unnormalized data of Table 7.1 has now been fully normalized. As the level of normality increases, capacity to achieve the aim of relational database model and maintenance cost increases while ability to access the data is reduced (Lee 1995). As a result, it could be said that not only the normalization is necessary to be managed within the conceptual design database but also other factors need to be considered in its physical design (Moore 1997). Last and final step in the normalization is represented in Fig. 7.9.

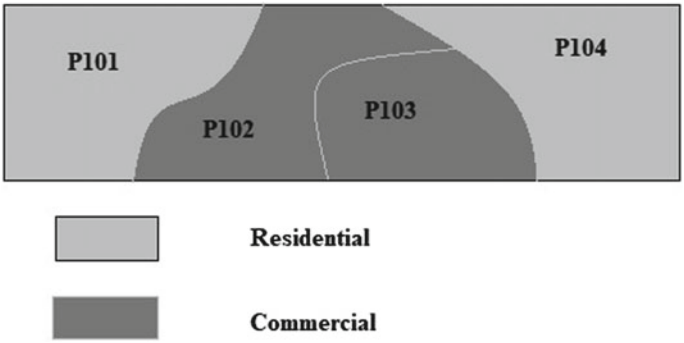


Fig. 7.6 Map depicts four parcels of land with the PINs P101, P102, P103 and P104. Two parcels are zoned residential, while the other two are zoned commercial

Table 7.1 An unnormalized table

Pin	Owner	Address	Date	Acres	Zone code	Zoning
P101	Shivam	Bihar, India	07-08-04	1.0	1	Residential
	Pihu	Asam, India				
P102	Bhasha	UP, India	21-11-99	3.0	2	Commercial
	Pushpa	Punjab, India				
P103	Anshu	Delhi, India	17-05-98	2.5	2	Commercial
P104	Bhasha	UP, India	18-07-97	1.0	1	Residential

Pin	Owner	Address	Date	Acres	Zone Code	Zoning
P101	Shivam	Bihar, India	07-08-04	1.0	1	Residential
P101	Pihu	Asam, India	07-08-04	1.0	1	Residential
P102	Bhasha	UP, India	21-11-99	3.0	2	Commercial
P102	Pushpa	Punjab, India	21-11-99	3.0	2	Commercial
P103	Anshu	Delhi, India	17-05-98	2.5	2	Commercial
P104	Bhasha	UP, India	18-07-97	1.0	1	Residential

Fig. 7.7 First step in normalization

7.3.3 Types of Relationships

Four different types of relationship are observed in the case of relational database model, viz. one-to-one, one-to-many, many-to-one and many-to-many (Fig. 7.10).

PIN	Date	Acres	Zone Code	Zoning
P101	07-08-04	1.0	1	Residential
P102	21-11-99	3.0	2	Commercial
P103	17-05-98	2.5	2	Commercial
P104	18-07-97	1.0	1	Residential

(a) Parcel Table

PIN	Owner
P101	Shivam
P101	Pihu
P102	Bhasha
P102	Pushpa
P103	Anshu
P104	Bhasha

(b) Owner

Owner	Address
Shivam	Bihar, India
Pihu	Asam, India
Bhasha	UP, India
Pushpa	Punjab, India
Anshu	Delhi, India
Bhasha	UP, India

(c) Address

Fig. 7.8 Separate tables are presented here from the second normalization step

When one record of the first table is related with only one record of the second table, then it will be considered as one-to-one type relationship. If one record of the first table is related with the more than one record of the second table, then it will be considered as one-to-many type relationship. An apartment complex's street address, for example, might linked with many households.

Reverse of the above type of relationship is considered as many-to-one type of relationship. In this relationship many records of one table are related with only record of the second table. It could be understood by reversing the example mentioned above. In other way we could say that if many households share the same street address, then it will be the example of many-to-one type of relationship. Last type of relationship is many-to-many type of relationship. In this type of relationship many records of the first table are related with many records of the second table. For example, many-to-many relationships imply more than one records from first table is be related with more than records in second table. For example, persons of house 'A' could support many persons of house 'B' at the same time person of house 'B' could support the many persons of house 'A' also. Such type of relationship is representing the many-to-many types of relationship.

PIN	Date	Acres	Zone Code
P101	07-08-04	1.0	1
P102	21-11-99	3.0	2
P103	17-05-98	2.5	2
P104	18-07-97	1.0	1

(a) Parcel Table

Owner	Address
Shivam	Bihar, India
Pihu	Asam, India
Bhasha	UP, India
Pushpa	Punjab, India
Anshu	Delhi, India

(b) Address Table

PIN	Owner
P101	Shivam
P101	Pihu
P102	Bhasha
P102	Pushpa
P103	Anshu
P104	Bhasha

(c) Owner Table

Zone Code	Zoning
1	Residential
2	Commercial

(d) Zone Table

Fig. 7.9 After normalization, separate the tables. The fields that are related to the tables are highlighted

7.4 Joins, Relates and Relationship Classes

To make use of a relational database, we can connect tables in the database for data query and management. In this section, we will look at three methods for connecting tables: (a). join, (b). relate, (c). relationship class.

Tables could be connected in the database for data query and management by using the relational database model. There are three methods for the connecting of the tables which is discussed here. Let us study them one by one.

7.4.1 Joins

This operation is performed through a common field and foreign key (Mishra and Eich 1992). Figure 7.11 represents an example of the joining of attribute data from two data tables. This process could also be performed if there are more than two data tables. In given example, soil-ID is working as a common parameter for the joining

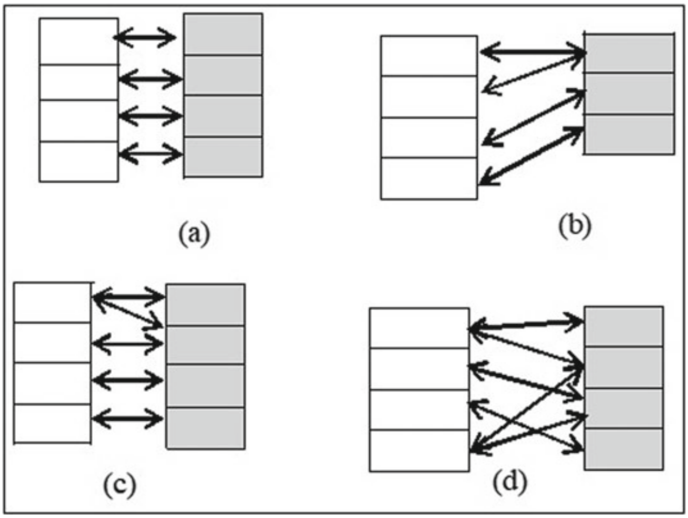


Fig. 7.10 **a** One-to-one relationship; **b** many-to-one relationship; **c** one-to-many relationships; **d** many-to-many relationship

of these two tables. Join operation is recommended to perform the one-to-one and many-to-one relationship. Two tables are joined by record in the case of a one-to-one relationship. In the case of many-to-one type of relationship, several records in the origin share the similar value as target in the destination table. With a one-to-many or many-to-many relationship, a join operation is ineffective because of only one matching value from the target destination could assigned a value in the origin table. Not all record could be assigned a value in destination table. That is the reason why, only one-to-one and many-to-one type of relationship hold join operation.

Record	Soil-ID	Area (km^2)	Perimeter (m)
1	1	1234.32	4212.47
2	2	1231.43	4193.84
3	3	423.63	793.23

Soil-ID	pH	Soil Code
1	5.6	Id3
2	6.1	Sg
3	6.7	Id3

Fig. 7.11 Common ground Soil-ID will connect the two table

7.4.2 Relates

If one requires to connect the two table temporarily by keeping them physically separate, then it is suggested to perform 'relate operation' between them. It is impossible to connect three or more than three table simultaneously by establishing the 'relates' in pair between the tables. Window based geospatial package would be the best option to work with relates because its ability to allow the display of multiple tables simultaneously. One major advantage of the relates is that they can be used all four relationships. Because a relational database is likely to contain various types of relationships, this is important for data query. If data is not at processing point and is stored at any remote database, then the process of 'relates' slow down the accessing capacity of data.

7.4.3 Relationship Classes

Relationships between objects can be supported by object-based data models such as geodatabases. In order to use for the management of attribute data, relationship is defined priorly and it is saved in geodatabase in the form of 'relationship class'. Relationship class is also able to hold all types of relationships as the 'relates' operation. To perform the many-to-many, type of relationship there is a need to setup an intermediate table in order to sort out any association and interconnection between the values of origin to the values of destinations. Rest the three types of relationship could be performed smoothly. If geodatabase consisted any relationship class, then it will automatically recognize it and could be used at the place of 'relate' operation.

7.5 Spatial Join

A spatial join operation joins two sets of spatial features and their attribute data using a spatial relationship (Güting 1994; Rigaux et al. 2002; Jacox and Samet 2007). For example, one could connect the school with the country where it is situated by using the topological relationship through the operation of spatial join. Result will represent that the school is located inside that country. Intersect is considered as other type of the topological relationship. Figure 7.12 also represents this type of topological relationship. Güting (1994), highlighted that the operation of the spatial join not only consider the concept of topological relationship but it is based on the direction and distance also.

Table relationships, as discussed in Sect. 7.3.3, could also applied for operations of spatial join.

For example, if each of the village have only one fault line, hence the relationship between that village and the fault line is of one-to-one type of relationship. If one state

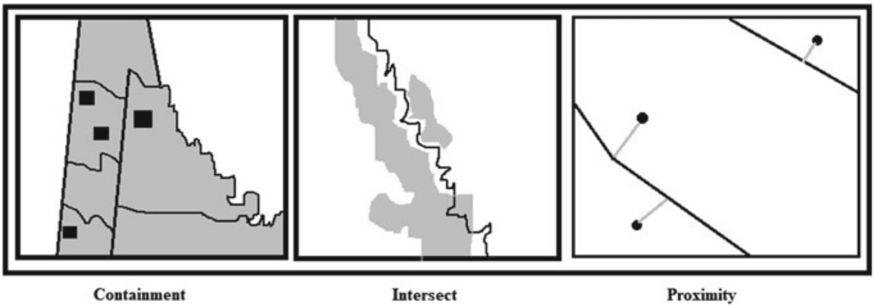


Fig. 7.12 Topological relationships of 'containment', 'intersect' and 'proximity' can be used to support spatial join operations

have more than one hospitals, then relationship between the state and the hospital is of one-to-many type.

7.6 Attribute Data Entry

Attribute data entry is analogous to perform a paper scanning for the purpose of data retrieval. This process necessitates the selection of appropriate method of digitizing for the process of attribute data entry and its verification.

7.6.1 Field Definition

First step in the field of attribute data entry starts with defining the fields of the attribute table. It is necessary to include the field length, type of data and the number of the decimals for the field definition. Field length specifies the numbers of the digits that need to be preserved for the field. This field should be sufficiently large to store the big amount of the data (e.g. sign, string). These data should have conformity with that of the supported GIS package. Decimal digits are described within the integral part of the definition of the float type of data. The field definition is now a field property. As a result, before defining a field, it is critical to consider how it will be used.

7.6.2 Methods of Data Entry

Assume a map has 4000 polygons, each with 50 attribute data fields. This may necessitate entering 200,000 values. Any GIS user is interested in how to save time and effort when entering attribute data. In the same way that we search for existing geospatial data, we should see if an agency has already entered attribute data in digital format. If this is the case, digital data could be easily loaded into GIS environment by us. To import a data in GIS environment it is very necessary to keep them in the appropriate format (e.g., .csv, or dBASE or excel format). Only option is to type in the case of absence of any attribute data files. This extent of typing might be vary depending on the method to be used. Single editing tool coming with GIS package, which edit the one record at one time is not much efficient for editing. This is very time-consuming task. In order to save the time associated in the processing one could use the relational database designing and lookup upon the common tables. It is preferable to enter map unit symbols or feature IDs directly into a GIS. This is due to the fact that a feature could be selected in the view window and look upon where the base map is consisted it and enter the symbol unit/ID. As far as non-spatial data entry is considered word processing software or spreadsheet could be the best option for it. These softwares include functions such as cut-and-paste, find-and-replace, and others many option that readily not found within GIS environment.

7.6.3 Attribute Data Verification

There are two steps to verifying attribute data. In first step of the verification, it is need to be make sure that the attribute data is interlinked with the spatial data properly. One should make sure that all the feature IDs are unique and it is not void. In second step of the data verification is to make sure that the used attribute data has not any type of falsification and provided data is correct. It should be noted that an attribute table should contain a variety of the errors due to observation, data entry process or by using the outdated (expired) data. Using attribute domains in the geodatabase to prevent data entry errors is an effective method (Zeiler 1999). By using an attribute domain, one could define the valid and acceptable range of values for the attribute table.

7.7 Manipulation of Fields and Attribute Data

Manipulation could be understood as the process of addition, deletion, and creation of the attributes by classifying and computing the current attribute data.

7.7.1 Adding and Deleting Fields

Data is commonly downloaded through internet resources for the purpose of GIS projects. These downloaded datasets have quite more attributes than we need. Field which is not necessary, need to be removed. Deletion of the unnecessary data, not only reduce the confusion but also the efficiency of computer to process such data is increased. It is simple to delete a field. It is necessary to specify an attribute table as well as the field to be deleted from the table. For the purpose of classification and computation addition of the field in the attribute data is necessary. The result of the classification and computation data must be filled in the new field. Added new field should be defined in the way of definition of attribute data entry.

7.7.2 Attribute Data Classification

During the process of classification of the attribute data from existing data, few new attributes could be generated. Suppose we have a dataset describing about the elevation of the area. New data could be generated by the reclassification of the elevations.

Classification creates the new attribute data in three steps.

1. New field is defined for saving the classification result.
2. Perform a query for dataset.
3. Value is assigned for selected data subset.

Second and third steps will be repeated until the completion of classification and new field to be assigned values, unless to make the automation of the procedure computer code is written. Data classification can help to simplify a dataset so that it can be used for display and analysis more easily.

7.7.3 Attribute Data Computation

Computation could be used to generate new attribute data from existing data. This process is done in two steps.

1. Process of new field definition
2. Calculation of the values of new fields on the basis of existing attributes.

A formula is used for this computation which could be coded manually or could be used with the mathematical algorithms. Conversion of the road map from miles to kilometre and then save it in new attribute is the example of the attribute data computation. Here in first step new attribute field is defined. Then field geometry will be calculated by using the appropriate mathematical conversion formula.

7.8 Conclusion

In this chapter we have discussed the detailed outlook about non-spatial/attribute data. Understanding of the attribute data would make user so efficient to deal with non-spatial data and generate valuable information using such data source. Students would be able to distinguish various data model and will able to incorporate them in their geospatial studies. Detailed understanding of the non-spatial data along with spatial data would provide a holistic approach to students to cope up with geospatial problems. In solving various problems, understanding of the concepts of joins, relates and relationship would make them able to deal with complex multidimensional problem. They would also be able to manipulate and make data entry in the attribute table to generate appropriate result and eliminate some inherent errors.

We have discussed the detailed concepts of GIS up to this chapter. In next chapter, we'll focus on the application of GIS in various dimensions. After understanding that we'll discuss the detailed working process in Geospatial fields through important case studies.

Questions

1. What is the difference between spatial data and non-spatial data?
2. Differentiate between raster and vector data.
3. Discuss the process of normalization of attribute data.
4. What are different types of relationship?
5. What are the different methods of attribute data entry? Elucidate.

References

Arvanitis LG, Ramachandran B, Brackett DP, Abd-El Rasol H, Du X (2000) Multiresource inventories incorporating GIS, GPS and database management systems: a conceptual model. Comput Electron Agri 28:89–100

Blanton JD, Manangan A, Manangan J, Hanlon CA, Slate D, Rupprecht CE (2006) Development of a GIS based, real-time internet mapping tool for rabies surveillance. Int J Health Geogr 5:47

Carleton CJ, Dahlgren RA, Tate KW (2005) A relational database for the monitoring and analysis of watershed hydrologic functions: I. Database design and pertinent queries. Comput Geosci 31:393–402

Chang KT (2019) Introduction to geographic information systems, vol 9. McGraw-Hill, Boston

Codd EF (1970) A relational model for large shared data banks. Commun Assoc Comput Mach 13:377–387

Codd EF 1990 The relational model for database management, Version 2. Addison-Wesley, Reading, MA

Date CJ (1995) An introduction to database systems. Addison-Wesley, Reading, MA

Galdi D (2005) Spatial data storage and topology in the redesigned MAF/TIGER system. U.S. Census Bureau, Geography Division

Güting RH (1994) An introduction to spatial database systems. VLDB J 3:357–399

Jackson M (1999) Thirty years (and more) of databases. Inf Softw Technol 41:969–978

Jacox E, Samet H (2007) Spatial join techniques. ACM Trans Database Syst 32, 1(7):44

Lee H (1995) Justifying database normalization: a cost/benefit model. Inf Process Manage 31:59–67
Mishra P, Eich MH (1992) Join processing in relational databases. ACM Comput Surv 24:64–113
Moore RV (1997) The logical and physical design of the land ocean interaction study database. Sci Total Environ 194(195):137–146
Oz E (2004) Management information systems, 4th edn. Course Technology, Boston, MA
Rigaux P, Scholl M, Voisard A (2002) Spatial databases with application to GIS. Morgan Kaufmann Publishers, San Francisco, CA
Shekhar S, Chawla S (2003) Spatial databases: a tour. Prentice Hall, Upper Saddle River, NJ
Stevens SS (1946) On the theory of scales of measurement. Science 103:677–680
Vetter M (1987) Strategy for data modelling. Wiley, New York
Zeiler M (1999) Modeling our world: the ESRI guide to geodatabase design. Esri Press, Redlands, CA
Zhang Q, Cheng L, Boutaba R (2010) Cloud computing: state-of the-art and research challenges. J Internet Serv Appl 1:7–18

Chapter 8
Application of GIS in Urban Policy/Planning/Management

Abstract The combined use of GIS and Remote Sensing has widespread tremendously over the last few decades. We have studied the concepts of GIS and Remote Sensing in our previous chapters. It's very necessary to understand the practical importance of these geospatial technologies rather than having only theoretical knowledge. In this chapter, it's attempted to present before you all the burning fields which are being controlled and monitored by these technologies. After finishing this chapter, we will be capable of understanding the following GIS and Remote Sensing applications.

- Geographic Information Systems in Microlevel Planning
- Geographic Information Systems in Water Resource Management
- Geographic Information Systems in Sustainable Development
- Geographic Information Systems in Agricultural and Natural Resource Management
- Geographic Information Systems in Sustainable Tourism Development
- Geographic Information Systems in Disaster Management.

Keywords Remote Sensing · Geographic information systems · Multi-sectoral sustainable development · Disaster management

8.1 Introduction

Nowadays, the arena of Remote Sensing, RS and Geographical Information System, GIS has become thrilling and attractive due to the massive growth of opportunities and essential tools that can be implemented at various levels aimed at sustainable decision-making in the field of socioeconomic development and preservation of natural resource. Remote Sensing and GIS technology, as well as their applications in various fields, have seen rapid growth in recent decades. This chapter describes the most commonly used methodologies for remotely sensed data, particularly image processing techniques, as well as the application capabilities of GIS technologies. GIS and Remote Sensing can help with the rapid planning of various control management programmes. Furthermore, RS and GIS have been used in conjunction in various

M. Kumar et al., *Geographic Information Systems in Urban Planning and Management*, Advances in Geographical and Environmental Sciences,
https://doi.org/10.1007/978-981-19-7855-5_8

applications to address planning and development-related issues. A strong database for the generation of baseline information in the field of natural resources is being also provided by RS, which is required for any development program's planning, implementation and monitoring. In many cases, Remote Sensing and Geographic Information Systems (GIS) have proven to be an effective tool for environmental monitoring. Satellite Remote Sensing has shown that it can monitor not only on the global horizon but also on the local.

GIS is an information system that works with data that is referenced by spatial or geographic coordinates. In other words, a Geographic Information System is both a database system and a set of operations for functioning (analysing) with spatially referenced data. GIS is the synthesis of technological and traditional disciplines. Because of the potential it holds for a broad range of disciplines that deal with spatial data, GIS has been dubbed a "enabling technology". GIS techniques are provided by many related fields of study. These related fields focus on data collection, whereas GIS brings them together by focusing on assimilation, modelling and evaluation. As a result, GIS is often referred to as the 'science of spatial information'.

In recent years, RS technology has proven to be highly useful in obtaining data for better resource management, and it could thus be used to monitor and manage the coastal environment (Ramachandran et al. 1993, 1996 and 1998). Furthermore, using a GIS to analyse trends and estimate changes in various themes aids in management decision-making. GIS and Remote Sensing technologies, in particular, enable rapid data collection, processing and integration, as well as the presentation of results in geographically referenced maps and reports. Huge scope is being associated with the geospatial technologies, and it is increasing rapidly with the passage of time. Natural hazard risk management can be managed using an integrated GIS-based approach. Several researchers have observed the theory underlying the use of these methods for 'aquatic weed monitoring' (Brown 1978; Gilmer et al. 1980; Berry 1986). Other authors (Campell 1987; Lillesand and Kieffer 1987) described in detail how Remote Sensing can be used as an instrument for environmental management. Urban terrain, physiognomy, lakes, plants, sights, traffic, land use, construction and population distribution can now all be investigated quickly using Remote Sensing technology.

Figure 8.1 depicts the scope of GIS in various fields.

8.2 Application of GIS in Microlevel Planning

8.2.1 Concept of the Microlevel Planning

Microlevel planning is one of the essential parts of the planning and development. It consists of the inclusion, participation and coordination of all people of the region. Microlevel planning region considers Gram Panchayat, local body or microlevel watershed as its planning unit. Planners across the world are concerned for the large area development plans which are created at the district level and neglecting the local

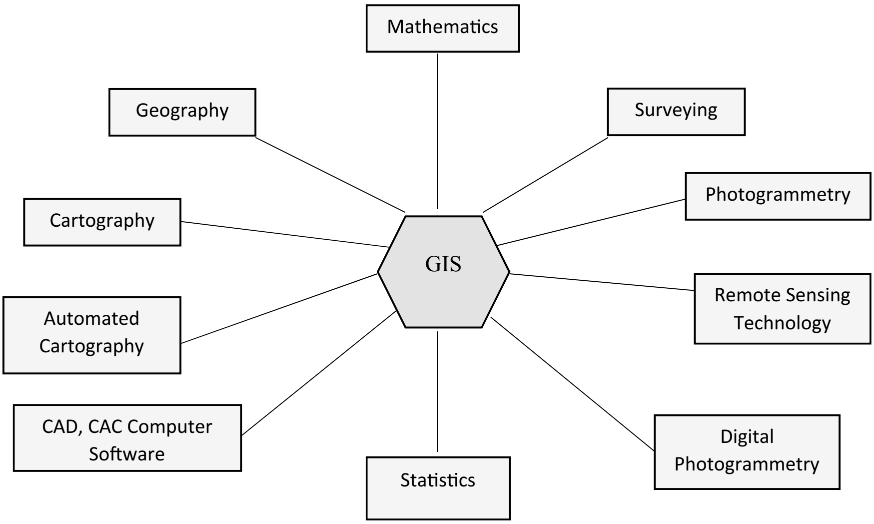


Fig. 8.1 Scope of GIS

issues. A district region of scale equal to or more than 1:50,000 is considered as large area. Such areas are not able to address the microlevel issues.

Microlevel planning is used to overcome these constraints. According to some, "macrolevel planning at the state or national level manifests broadly the measures to be taken, whereas microlevel planning at the local level manifests the specific measures to be taken". Specific action plans are generated as a result of level planning. On a larger scale, the integration of microlevel data long-term development would be ensured by planning, which is required to generate macrolevel results, and provide us with the tools we require to make sure balanced development (Kale 1992).

8.2.2 Use of Remote Sensing and GIS in Microlevel Planning

Remote Sensing is important for providing geospatial information as well as determining, improving and monitoring the earth's overall capacity. Geospatial technologies are highly applicable in the field of microlevel planning for the sustainable development and management of the region. Geospatial data is crucial to derive information in the field of hydrology, geology, agriculture and topography of the earth. Like many other countries in the world India is also moving fast in the field of Remote Sensing Technology. 'Indian Remote Sensing Programme' is intended to achieve functional ability in the field of earth resources exploration and conservation of the environment. Indian government is promoting the use of geospatial technologies for the planning and development of the region. These innate benefits of geospatial technologies are to precisely and accurately map the available natural resources of the region. This is needed for microlevel planning.

Traditional methods are very difficult, time-consuming, labour intensive and expensive for the construction of the data. Geospatial techniques are the most efficient tool for effectively and precisely mapping these features. The ability of satellite Remote Sensing sensors to acquire satellite data from the same location multiple times is exceedingly important for assessing the influence on the natural resources of the planning region, the amplitude and the overall direction of the changes. One major benefit of the implementation of geospatial techniques is its ability of real-time assessment of various types of data which is beneficial for the health management, planning development, resource management of the region (Congdon 1999; Gibson et.al. 2002; Hanchette 1998; Lovett et.al. 1998) . Undoubtedly, these all are very crucial parameters for the microlevel planning of the region. It works on the basis of spatial data which is obtained from Remote Sensing and non-spatial data which is received from other resources. GIS is capable to handle several layers in its domain and helps the decision-makers and planners to reach at a particular decision on the basis of processed data. One major benefit of the implementation of the geospatial technologies in planning activities is that it provides a newer perspective of visualization and presentation of the data. This will help to deduce the information in much efficient and productive way. Various new tools and techniques are available within GIS environment, for example by performing the network and proximity assessment one could easily deduce the information related to various social problems.

Within Sustainable Development Goal (SDG), 15, it is mandated to secure the "life on land". To secure the life on the land it is necessary to preserve and develop our living environment and conserve the available resource for our upcoming generations. These natural resources could be very well preserved by integrating the Remote Sensing and geospatial technologies. These techniques are time-saving and thus ideal for the spatial planning in the context of land use. These are also able to deal with the complex problems associated with the manipulation and retrieval of the large data.

At microlevel planning, block-level planning is acting as viable unit for land use development and planning performed at microlevel. In this way spatial technologies are evaluated as decision support system. Incorporating which we could easily manage and perform microlevel planning activities.

8.3 Use of Remote Sensing and GIS in Hydrological Management

The irrigation system is the world's largest consumer of freshwater. Approximately 70% of all water is used to produce 30–40% of the world's staple crops (Bastiaanssen 1998). Better irrigation management is required to achieve higher water use efficiency, because water scarcity is posing serious threat in front of hydrological planners (Welch et al. 1988). A comprehensive knowledge-based management system is required for the improvement of water productivity at the river basin scale (Bastiaanssen and Prathapar 2000). By estimating the performance of hydrological management in terms of quantitative and qualitative aspect we should invest in population increment and make the irrigation system much efficient so that it could be economically justified. Water scarcity is also exacerbated by poor water management and a lack of storage sites. 'Blue Water' also known as freshwater is the primary focus of traditional water resource planning and management. There is an urgent need to incorporate rainfall, particularly in arid and semiarid basins, which naturally infiltrates the soil and returns to the atmosphere via evapotranspiration.

Hydrological planning has always been an appealing exploration topic for the researchers. With the passage of time more than half of the requirement of drinking water is fulfilled by groundwater. Thus, groundwater is being extracted at a tremendous rate. Hence, it is an urgent requirement of groundwater management. Mismanagement of groundwater has had a negative impact on the ecosystem and, as a result, on food security. Hence sustainable management of the groundwater is very much needed. Lack of expertise in the field of surface and groundwater management is one of the major barriers. In contemporary world geospatial technologies have widen the scope for the hydrological management.

Implementation of geospatial technologies could play a crucial role in the water resources management of any nation. To achieve optimal water resource planning and operation, related projects and cutting-edge geospatial techniques must be combined with traditional groundwater management methods to secure the sustainability of water resources. Satellite data is one of the best options for hydrological modelling. It also provides decision-makers with spatial information. Data obtained using satellites sensors could not be manipulated by mediators. Thus, its originality is conserved in nature. These satellites' images are able to describe several visible landscape patterns over the earth surface. Direct measurements thus performed are often more reliable than secondary data because they are based on satellite observations.

In ungauged or data-scarce regions, satellites data-based hydrological modelling is being done by several scholars (Immerzeel et al. 2008). In order to calibrate and validate these models it is necessary to collect the long-term data series by performing rigorous measurements.

However, in basins such as the Indus, such data is scarce, resulting in a high level of uncertainty in model results. Remote Sensing-derived spatially variable information describing topography, crop types, land use, climatic data and leaf area index could be used for basins' modelling.

By integrating the Remote Sensing, Global Positioning System and Geographical Information System unbiased knowledge in the field of LULC, soil moisture, rainfall and precipitation, surface and groundwater availability, land surface temperature, etc., could be deduced easily which would extremely be helpful for the exploration of drought modelling, watershed management and modelling, urban heat island mapping and modelling, water resource mapping and modelling.

8.4 Application of Remote Sensing and GIS for Sustainable Development

As we all know the basic need of the human beings are food, cloth and shelter. Primary and major aim of any developmental activities is intended to fulfil these goals for human welfare and also ensure their quality education, cultural and recreational development. Economic development should be taken place to alleviate the poverty of the nation without affecting the sustainability of the nature. Maintaining the sustainability of the nature is the huge challenge in front of policymakers and government stakeholders. Rainforests, once thought to be the planet's lung, are being depleted faster than replantation. Tremendous population growth is resulting to extraordinary increment in the commercial activities. By increasing the commercial activities water demand of the nations is increased and particularly the small island countries are facing very serious threat due to it. Concerned stakeholders and planners have detailed information about the available resources, and they are authorized for the management of these resources so that sustainable development of the region could be assured. Furthermore, economic development and conservation could coexist. Attempt which is intended to separate the conservation of resources from the developmental activities should be strictly prohibited. Many natural resources, including land, have received insufficient or no sustainable management in the developing world. In past times, land managers are intended to maximize their output at the lowest expense which was adversely impacting the sustainability of the environment. Such trend of production was continuously expanding for many decades. This hugely impacts the available resources negatively, and production was increasing continuously without worrying about the environment. Several producers have not meant to repair the loss of environment which is done in the due process of their production. In order to formulate and successful implement of the reliable strategic framework for sustainable development it is necessary to collect all the necessary information on the condition of available resources which is need to be managed and conserved. This process includes the collection of information regarding physical, socioeconomic characteristics of the land and its available resources.

We understood the detailed about sustainability, but briefly it could be said that sustainable development is defined as development that meets the needs of the present without jeopardizing future generations' ability to meet their own needs. It contains two key concepts: the concept of needs, particularly the essential needs of the world's poor, to which top priority should be given; and the concept of limitations imposed on the environment's ability to meet present and future needs by the state of technology and social organization. Sustainable development entails striking a delicate balance between the human desire to improve lifestyles and feelings of well-being on the one hand, and the preservation of natural resources and ecosystems on which we and future generations rely on the other. GIS organizes geographic data so that a person reading a map can select the information needed for a specific project or task. GIS can be a valuable tool for assisting people in developing plans for implementing successful, long-term management strategies at both the local and global levels. A good GIS programme can process geographic data from various sources and integrate it into a map project. Except for the USA, many countries have an abundance of geographic data for analysis. GIS is an important tool in assessment, prioritization, mitigation, planning, science and training. Unsustainable development is characterized by economic inequality, social instability and environmental degradation.

Poor people bear the brunt of these issues because their livelihoods are precariously balanced on volatile economic opportunities and changing environments. What we need is a national strategy for sustainable development (a strategic and participatory process of analysis, debate, capacity building, planning and action towards sustainable development), a tool to help farmers overcome their problems and begin to strengthen their capacity for sustainable development. Using GIS in sustainable agriculture can result in significant benefits, particularly in terms of how they can be seen in all the cultures they represent, the soil which is substances in the soil, water availability, crop rotation and pests.

8.5 Use of Remote Sensing and GIS in Agricultural Resource Management

Agriculture is critical to the economy of every country. It is a significant trading industry for a country with a strong economy. The application of RS and GIS to examine and envision agrarian ecosystems has proven to be extremely important to both the farmers community and industry. In this section, we will attempt to provide an overview of RS and GIS applications in agricultural resource management. Real-time information on the variety of crops, their area and expected yield is critical for the agrarian economy. The use of various RS methodologies is critical in crop management and yield estimation to ensure agricultural resource conservation. Spectral information is a key component of RS data for modelling the cropping patterns, and it is linked to canopy parameters, which reflect crop health and growth stages. Remote Sensing and Geographic Information Systems can be used to assess and evaluate LULC as well as damage assessment due to the drought, river flooding and other weather extremes. Meteorological and vegetation data are the crucial inputs into agricultural meteorology. For detecting pest and disease infestations, Remote Sensing applications are an efficient and significant tool. It's one of the most efficient ways to assess and monitor water resources.

8.5.1 Remote Sensing and GIS in Inventory of Crops

Remote Sensing and Geographical Information System are important tools for identifying crops and areas where cropping patterns are changing, as well as for carrying out crop surveys and mapping. It is critical for the government of an agriculturally based country to have real-time information on the crops grown, their area and assumed crop yield. Spectral information is a key component of Remote Sensing data for crop modelling, and it's linked to canopy parameters, which reflect crop health and growth phases. With the help of satellite and survey data, crop-specific maps are created. They show the land layout as well as the owners, which is useful to agribusinesses like seed and fertilizer industries. Remote Sensing can be useful in creating the inventorying databases for various crops. The use of geospatial technologies reduces the amount of field data that must be gathered while increasing estimation precision (Kingra et al. 2016).

8.5.2 Use of Remote Sensing and GIS for the Crop Management

Flooding and moisture stress are regular events in water-logged areas, causing harm to crops, especially in regions where paddy is grown. Crop yield data from various crops is crucial intake for the accurately and efficiently prediction yield. 'Genotype' of each crop has a unique potential production per unit of area, that could be recognized in an ideal experimental investigation. Meanwhile, in the actual world, production of the crops is affected by several factors like soil, weather and cultivation practices. Biotic stresses, such as disease and pest, also have an impact on crop yield. RS on the basis of satellite data is an appropriate option for crop management because it estimates crop parameters in a prompt, precise, systematic and unbiased manner. One of the most important tools for yield modelling is Remote Sensing data (Dadhwal 2003). Crop vigour is a predictor of crop production. It could be evaluated by using the indices which are related to vegetation extracted from various parts of the spectrum. Crop growth, health and yield have all been tracked using plant growth simulation models. Since the majority of plant growth models were created at the field scale, hence their application in large areas has been limited. To estimate and forecast yield, synthetic aperture radar technology was combined with a crop modelling approach (Setiyono et al. 2014). Remotely sensed data could be used to identify crops and their health conditions by estimating leaf area. Attacks by insects typically act as reason for the breakdown of chlorophyll, and the decrease in concentration of chlorophyll in plants can be detected using Remote Sensing.

8.5.3 Nutrient and Water Stress Estimation Using Remote Sensing and GIS

Water, sunlight and adequate nutrients are required for plants to grow and thrive. Macronutrients are more required than that of micronutrients in terms of quantity to ensure the plant cell and tissue development (Mee et al. 2017). Geospatial technologies could be effectively implemented in the field of nutrient and water stress estimation. These technologies could be used to detect nutrient stresses which is important in site-specific nutrient management, which can reduce cultivation costs while increasing fertilizer use efficiency (Shanmugapriya et al. 2019). Water judicious use in arid regions is possible with the use of precision technologies. Drip irrigation is a very efficient way of irrigation in water-deficit regions, and water use efficiency could be increased by reducing runoff and percolation losses when one combined this information from remotely sensed data such as canopy temperature difference.

Nutrient deficiency is detected using multi-spectral and hyperspectral images. Spectral reflectance measurements can help in the selection of wavelengths that are sensitive to several kinds of nutrients and water stress (Mee et al. 2017). Detection of the water stress of the crops are critical in effective management of irrigation. Data obtained from satellites could produce spatiotemporal analysis of crop development and the impact of water stress condition on productivity.

8.5.4 Flood Monitoring Using Remote Sensing and GIS

Remotely sensed data allows measurements to be obtained, especially from space, over much larger spatial scales than those covered by ground tools and methodologies. The information on flooding was gathered using satellite data at various spatiotemporal scales. The use of 'automated spacecraft technology' has reduced flood detection and response time to a few hours. By incorporating Remote Sensing, one can better understand the flood process and its potential impact on the environment. (Schumann et al. 2018). River discharge estimations can be expanded to include a wider area. Large river surface water hydraulic characteristics, such as mean width of the river over a particular reach length and water surface slopes and elevations, and channel morphometry can be measured or evaluated globally using Remote Sensing techniques (Sun et al. 2010).

Forecasting floods and other hydrological management concerns rely heavily on hydrological data assimilation and modelling based on river flow measurements. Early warning systems can incorporate river precipitation and surface topography observations by using remotely sensed data to assess riverine sediments by estimating the alteration in river widths and rainfall (Acharya et al. 2018).

For the projection of riverine discharges, optimization method was implemented to minimize the extent of anomaly between the simulations and observations of the flood extent.

By incorporating RS data and surface energy balance algorithms, it is possible to estimate the spatial variability in evapotranspiration over a large area. Soil water and the evapotranspiration of the crops highly impact the temperature of cropped area. With the variation in the temperature of cropped area, energy emitted from it is also varied.

8.5.5 Remote Sensing and GIS-Based Assessment of Land Use/Land Cover (LULC)

Land use-land cover mapping involves finding surface features at different scales and putting them into hierarchical groups. This is important for studying how the world is changing. Deforestation and biodiversity loss are caused by human activity, and global warming has a significant impact on land use and land cover (Nix et al. 1987). As a result, available data on land use/land cover can be used to inform environmental management decisions and future planning. LULC is highly impacted by the population growth and the socioeconomic developmental activities. Population growth and socioeconomic development cause unplanned and uncontrolled changes in land use/land cover. Mismanagement of agricultural and forest lands typically causes LULC changes, resulting in severe problems such as floods and landslides. The satellite-derived LULC maps were used to perform a pixel-by-pixel change detection and comparison to deduce important piece of information (Reis 2008).

Understanding of the difference between land use and land cover is necessary. In simple language, land cover is referred as natural surface features, meanwhile land use is defined by those features all features which are being created and modified by the human beings. For example, rivers, mountains and plateau are land cover features while settlement, parks, artificial lakes, etc., are land cover features. Proper mapping of the LULC is necessary because by doing so we are able to perceive the general information about the surface features and having the detailed information about the surface features of any particular place one would be able to implement any plan and policy in an efficient way. Mapping of the LULC is important because it not only contains the fundamental information about surface features but also provide the detailed information on the spatial distribution pattern of the area under various classes.

Remote Sensing and Geographic Information Systems are broadly used in the preparation of information of LULC data for the region. Hence it is more preferable to perform; rather than doing surveys of large areas done by hand in terms of cost, accuracy and mistakes made by hand. Also, imagery or aerial photos give a complete picture of an area, so nothing is missed. On the other hand, surveys are more likely to miss some features. Satellite images can be taken at regular intervals, but surface features or events can't be tracked by surveys, which can't be done regularly or quickly (floods, deforestations, forest fires, etc.) By combining one piece of information with another, different geographical and socioeconomical aspects can be analysed in simple and efficient way. Digital detection refers to changes in land use-land cover properties based on georegistered multi-temporal Remote Sensing data. Collaboration of remotely sensed data and field observations can achieve land cover classification and change detection faster and cheaper than either alone (Diallo et al. 2009).

8.5.6 GIS and Remote Sensing in Agro-Metrological Application

Climate and meteorological phenomena have a significant impact on agriculture. Metrological data is collected using a variety of spatial networks of point station observations. Traditional agro-metrological techniques have significant limitations in terms of their data usage for the monitoring of crops in real time and predicting their yields.

Implementation of satellite data has made it possible to measure a number of basic agrometeorological parameters more accurately and more often. Rainfall data is collected at every two weeks, and minimum and maximum temperature data are few examples of inputs which are used in agro-metrological station for the analysis and forecasting.

Many people think that monitoring the vegetation, weather and climate through geostationary satellite Remote Sensing is the most important step forward in the last 25 years. These satellites gather information about the temperature of the ocean and the plants on land. In agricultural meteorology, vegetation and the meteorological data act as the most crucial parameters.

Two major meteorological satellites are used for meteorological purposes (Basso et al. 2008). The “Geosynchronous Meteorological Satellite” (GMS) is first, which orbits at 36,000 km height, and “Polar Orbiting Satellite” is second, which orbits at 750 km height in the Low Earth Orbit (LEO).

8.5.7 Remote Sensing and GIS in Pest Infestation

Remote Sensing applications are an efficient and significant method for detecting, mapping and monitoring invaders, as well as identifying pest-infested and diseased areas (Jensen et al. 1989 and 1992). There is vast spatial heterogeneity is found in nature. This complicates the study of the trends and patterns of biological invasion (Pest Attack). Geospatial technologies have ability to cope up with this problem (Joshi et al. 2004). Remote Sensing applications provide data that can be used to detect and map defoliation, characterize pattern disturbances and so on. By measuring the spectral responses through Remote Sensing, ‘Chlorosis’ leaves yellowing and reduction in foliage could be assessed on spatiotemporal basis.

At various flight altitudes, airborne Remote Sensing could attain varying spatial resolutions. Platforms which are situated at ground station are widely implemented for ‘pest management’, ‘crop disease detection’ and ‘weed infestation’, and they provide valuable data for management planning and decision-making (Huang et al. 2008). Aerial colour infrared photography with a standard camera has been successful in identifying damage caused by a variety of serious pests. (Rani et al. 2018).

8.6 Remote Sensing and GIS in Sustainable Tourism Development

For the development of GIS bases applications in the field of tourism one should closely monitor Crain and MacDonalds' three stages of GIS application development. Feature of interest and primary simple data queries like locational and conditional queries were assembled and organized firstly through inventory applications. Further, in second stage it will evolve into analysis type of application which would be able to incorporate much-complicated type of operations. At third and last final stage, such application will be created whose are more comprehensive in nature and able to better perform the decision-making process.

GIS has the greatest potential to support sustainable development approaches to tourism planning, development and management (also known as "sustainable tourism") during this final phase. Major geospatial technologies are intended to delineate the suitable tourism hotspots. GIS contributions were identified by Malczewski (1999) in all three phases of a planning and decision-making process: the intelligence phase for identifying opportunities or problems; the design phase for developing and analysing possible alternatives to the identified problem(s); and the choice phase for evaluating the alternatives. All three of these phases necessitate the identification and application of sustainable tourism indicators for assessing the current situation, identifying flaws, monitoring change and evaluating alternatives. GIS may be the most effective technology for identifying and monitoring these sustainability indicators. It could be argued that the potential of GIS in tourism development and management has yet to be fully explored, nor have the capabilities of GIS been fully utilized. Many of the aforementioned applications were created for recreational purposes rather than tourism. One reason for this was that they were associated with the environmental management of national parks and other tourist attractions. GIS has been widely used in environmental management and is regarded as an essential tool for environmental protection, monitoring, resource use optimization and allocation, and activity zoning (Aronoff 1991). Despite the fact that the environment is an important factor and resource for tourism development, GIS has not been widely used in environmental management in the tourism sector. Another inconsistency is that GIS is the primary technology specialized in handling geographical data and thus facilitating the study of geographic phenomena. As a result, tourism, which has a strong geographic component, could benefit from the use of this type of information system.

This capability is especially important in the context of sustainable tourism, which requires a balance of economic growth, environmental costs and societal benefits. What must be emphasized, however, is GIS's ability to connect disparate data and generate new information (Cowen 1988). McAdam (1999) stated the importance of geospatial technology by explaining its ability of computer-based mapping and analysis by displaying the geographical data and its readjustment to locate the spatial features and pattern of the actual earth surface without performing the huge manual survey. Ultimately, it could be said that supplementary analysis and information could be deduced by integrating the geospatial technologies with other one (Malczewski 1999). Remote Sensing and Global Positioning Systems are acting as a mean to collect the raw data to be processed within GIS environment. The visualization of the results of data entry and analysis is a widely recognized GIS capability. A variety of options are available within GIS environment to visualize any data, for example data could be displayed in attribute format (tabular) or in the form of map layout. Both visualization techniques help to communicate and interpret the result among the interested persons and also this process. Thus, it could be said that the visualization is important not only for the decision-making motto but also for the result obtained from the effective interpretation and analysis of the visualized data helps to manage the sustainable development of the region.

GIS and its visualization capabilities, as mentioned in the applications section above, were used in a number of cases to facilitate and enhance citizen and stakeholder participation. This aspect has the potential to make a significant contribution, as participatory processes are regarded as critical to achieving sustainable development (Harris et al. 1996; Nicholls 2001). This is related to the type of planning for sustainable development. Bottom-up hierarchies are preferable in terms of sustainability, and GIS technology can be used to aid in the decentralization of planning and policy formulation. One or more users performing relevant tasks may have access to stand-alone or distributed systems. Local government could use either its self-information system or it could collaborate with other agencies like state and central government depending on the requirement of process. By implementing the geospatial technologies, it is so easy to handle the thematic layers, data and various constraints which are the ultimate requirement for any result. Due to this characteristic, it is considered as a dynamic tool rather than static tool for the planners (Beedasy and Whyatt 1999). This fundamental characteristic of the geospatial tools helps a lot in the sustainable tourism development because the preferences and the target area could vary during the developmental and operationalization process.

8.7 Application of Remote Sensing and GIS in Disaster Management

Earth comprised lithosphere, hydrosphere, biosphere and atmosphere. Various events and phenomena are taking place upon it. But out of them few became so dangerous in nature resulting death or injury of humans, loss of properties like buildings and communication system. Such extreme events are known as disaster. These disasters leave the huge impact on the natural and socioeconomic environment of the earth system (Alexander 1995). Few disasters, like earthquake show its impact instantly at the same time some other disaster, like drought show its impact gradually.

This is very much necessary to differentiate the disaster with hazard. When a hazard (e.g. earthquake) comes in inhabited area, then it will be considered as disaster; otherwise it will be hazard. Disaster could be classified into natural and anthropogenic

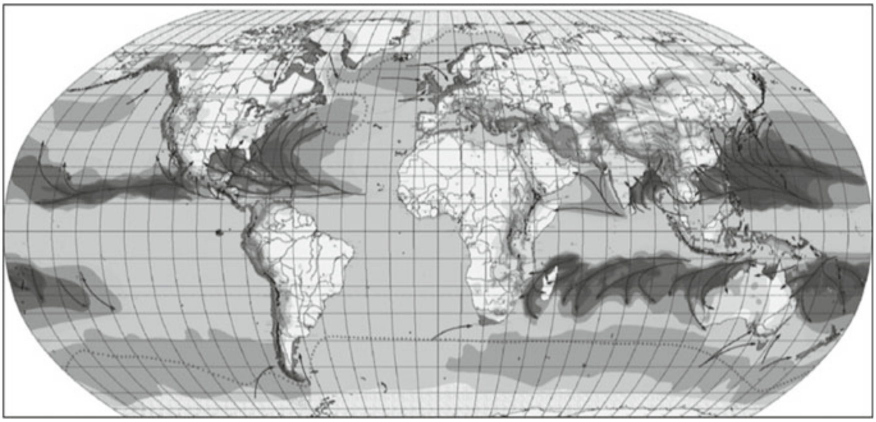


Fig. 8.2 Geographical distribution of various hazards. *Source* Berz et al. (2001)

disaster. Natural disasters occur in several parts of the world but specified types of disaster are occurred in specified area. Figure 8.2 represents the spatial distributions of several natural hazards like earthquakes and volcanoes got concentrated at the plate boundary.

There are several types of disasters:

- Natural disasters are events that occur as a result of purely natural phenomena and cause harm to human societies (such as earthquakes, volcanic eruptions and hurricanes).
- Anthropogenic disasters are those disasters which occur due to the influence of extreme human activities such as atmospheric pollution, nuclear testing, oil spill and biological testing.
- Natural disasters which are being induced and impacted by anthropogenic activities are known as human-induced disaster. Such types of disasters are got accelerated with the impact of human activities.

This classification is briefly described in Fig. 8.3.

Earthquakes cause the most damage. It accounts for 35% of total losses, ahead of floods (30%), windstorms (28%) and others (7%). Earthquakes are also the leading cause of fatalities (47%), followed by windstorms (45%), floods (7%) and others (1%) (Munich 1999).

Media play a major role in covering these disasters effectively which result to dramatic increment in loss because of extensive covering. This is representing the increased vulnerability of the disasters in recent times. With the population growth of the world, vulnerability towards the disaster is also increasing. Rapid growth of population is boosting the rate and intensity of the disasters. United Nation Development Programme (UNDP) has exposed the reality of population growth by its data where it has shown that population of the world has been doubled (3 billion to 6 billion) from 1960 to 2000s.

Natural	Some human influence	Mixed natural /human influence	Some natural influence	Human
▪ Earthquake ▪ Tsunami ▪ Volcanic eruption ▪ Snow storm / avalanche ▪ Glacial lake outburst ▪ Lightning ▪ Windstorm ▪ Thunderstorm ▪ Hailstorm Tornado ▪ Cyclone/ Hurricane ▪ Asteroid impact ▪ Aurora borealis	▪ Flood ▪ Dust storm ▪ Drought	▪ Landslides ▪ Subsidence ▪ Erosion ▪ Desertification ▪ Coal fires ▪ Coastal erosion ▪ Greenhouse effect ▪ Sealevel rise	▪ Crop disease ▪ Insect infestation ▪ Forest fire ▪ Mangrove decline ▪ Coral reef decline ▪ Acid rain ▪ Ozone depletion	▪ Armed conflict ▪ Land mines ▪ Major (air-, sea-, land-) ▪ traffic accidents ▪ Nuclear / chemical ▪ accidents ▪ Oil spill ▪ Water / soil / air ▪ pollution ▪ Groundwater pollution ▪ Electrical power ▪ breakdown ▪ Pesticides

Fig. 8.3 Classification of disaster

Another factor related to population pressure is that areas that were previously avoided due to their vulnerability to natural disasters become settled. Added to this is the significant trend of concentrating people and economic activities in large urban centres, the majority of which are located in vulnerable coastal areas. Poor newcomers frequently occupy marginal land, which is vulnerable to disasters, in megacities with rapid growth.

8.8 Conclusion

In this chapter, we have discussed the detailed application of the GIS and Remote Sensing in the various fields. For planning purposes, if the planning body is implementing the geospatial techniques, not only the functional cost but also manpower and operational cost could also be reduced with increased service potential. In the similar way, water is the prime essential for human survival. Use of geospatial techniques in the fields of water resource would ensure its very effective and progressive management. By this so water shortage problem could be reduced up to a certain extent. Unsystematic growth is posing severe threat on the human being. By promoting the implementation of the GIS and Remote Sensing in this sector could make it more sustainable without affecting the demands of future generations. Food is another chief requirement of human civilization. Rapidly growing population is creating a giant problem. Proper use of GIS could help the government officials and respected stakeholders to cope up with this issue. Frequency and intensity of disasters have also increased in recent decades due to several anthropogenic and natural reasons. By incorporating the geospatial technologies its effect over the human being could be certainly reduced. Health and tourism are the subsidiary problem associated with human beings. It's suggested to policymakers to implement the geospatial technologies as much as they could to overcome with these challenges.

Questions

1. Discuss meaning, concept and scope of the GIS.
2. What is difference between microlevel planning and macrolevel planning? How GIS is helpful in both levels of planning? Elucidate.
3. Discuss different types of disasters and why GIS should be used in disaster management?
4. Healthcare problem is major problem of the contemporary global world. Could GIS be treated as panacea in healthcare management? Discuss."

References

Acharya SM, Pawar SS, Wable NB (2018) Application of Remote Sensing GIS in agriculture. Int J Adv Eng Res Sci 5(4):237434

Alexander DE (1995) A survey of the field of natural hazards and disaster studies. In: Geographical information systems in assessing natural hazards. Springer, Dordrecht, pp 1–19

Aronoff S (1991) Geographic information systems: a management perspective. WDL Publications, Canada

Basso B, McVicar TR, Lee B (2008) Remote sensing and GIS applications in agrometeorology

Bastiaanssen WGM (1998) Remote sensing in water resources management: the state of the art. Int Water Manage Inst, Colombo, Sri Lanka

Bastiaanssen WGM, Prathapar SA (2000) Satellite observations of international river basins for all. In Florinsky I (ed) International archives of photogrammetry and remote sensing, ISPRO Congress, Amsterdam, the Netherlands, XXXIII Part B7, pp 439–451

Beedasy J, Whyatt D (1999) Diverting the tourists: a spatial decision-support system for tourism planning on a developing island. ITC-J 3–4:163–174

Berry JK (1986) Learning computer-assisted map analysis. J Forestry 84(10):39 43

Berz, G., Kron, W., Loster, T., Rauch, E., Schimetschek, J., Schmieder, J., Siebert, A., Smolka, A. & Wirtz, A. (2001). World map of natural hazards–a global view of the distribution and intensity of significant exposures. Natural hazards, 23(2), 443-465.

Brown WW (1978) Wetland mapping in New Jersey and New York. Photogram Eng Remote Sens 44(3):303–314

Campell JB (1987) Introduction to remote sensing. Guilford Press, New York

Congdon P (1999) Primary care needs assessment and resourcing: complementary practice and geographic perspectives. Health Place 5(1):59–82

Cowen DJ (1988) GIS versus CAD versus DBMS: What are the differences? Photogram Eng Remote Sens 54:1551

Dadhwal VK (2003) Crop growth and productivity monitoring and simulation using remote sensing and GIS. Satellite remote sensing and GIS applications in agricultural meteorology, pp 263–289

Diallo Y, Hu G, Wen X (2009) Applications of remote sensing in land use/land cover change detection in Puer and Simao Counties, Yunnan Province. J Am Sci 5(4):157–166

Gibson A, Asthana S, Brigham P, Moon G, Dicker J (2002) Geographies of need and the new NHS: methodological issues in the definition and measurement of the health needs of local populations. Health Place 8(1):47–60

Gilmer DS, Work EA, Colwell JE, Rebel DL (1980) Enumeration of prairie wetlands with Landsat and aircraft data. Photogram Eng Remote Sens 46:631–634

Hanchette C (1998) GIS implementation of 1997 CDC guidelines for childhood lead screening in North Carolina. In: GIS in public health, 3rd national conference abstract

Harris TM, Weiner D, Warner T, Levin R (1996) Pursuing social goals through participatory GIS: redressing South Africa's historical political ecology. In: Pickles J (ed) Ground truth: the social implications of geographic information systems. Guilford Press, New York, pp 196–222

Huang Y, Lan Y, Hoffmann WC (2008) Use of airborne multi-spectral imagery in pest management systems. Agri Eng Int: CIGR J

Immerzeel WW, Gaurand A, Zwart SJ (2008) Integrating remote sensing and a process-based hydrological model to evaluate water use and productivity in a south Indian catchment. Agric Water Manage 95:11–24

Jensen JR (1989) Remote sensing. In: Gaile GL, Willmott CL (eds) Geography in America, Merrill Publishing, Columbus

Jensen RJ, Narumalani S, Weatherbee O, Morris JKS, Mackey HE (1992) Predictive modeling of cattail and waterlily distribution in a South Carolina reservoir using GIS. Photogram Eng Remote Sens 58:1561–1568

Joshi C, De Leeuw J, van Duren IC (2004) Remote sensing and GIS applications for mapping and spatial modelling of invasive species. In: Proceedings of ISPRS (vol 35, p B7)

Kale P (1992) Sustainable development-critical issues. J Indian Soc Remote Sens 20(4):183–184

Kingra PK, Majumder D, Chandra B, Viswavidyalaya K, Singh SP (2016) Application of remote sensing and GIS in agriculture and natural resource management under changing climatic conditions. Agri Res J 53(3):295–302

Lillesand TM, Kieffer RW (1987) Remote sensing and image interpretation, 2nd edn. John Wiley and Sons, New York

Lovett A, Haynes R, Bentham G (1998) Improving health needs assessment using patient register information in. GIS and Health 6:191

Malczewski J (1999) GIS and multicriteria decision analysis. John Wiley, New York

McAdam D (1999) The value and scope of geographical information systems in tourism management. J Sustain Tour 7(1):77–92

Mee C, Siva KB, Ahmad HMH (2017) Detecting and monitoring plant nutrient stress using remote sensing approaches: a review. Asian J Plant Sci 16(1):1–8

Munich RE (1999) Topics 2000: natural catastrophes-the current position. Munich RE

Nicholls S (2001) Measuring the accessibility and equity of public parks: a case study using GIS. Manag Leis 6(4):201–219

Nix J, Hill P, Williams N (1987) Land and estate management. Packard Publishing, Chichester (UK)

Ramachandran S (1993) Coastal zone information system-pilot project for Rameswaram area. Report submitted to Department of Ocean Development. Government of India, 40

Ramachandran S, Krishnamoorthy R, Sundramoorthy S, Parviz ZF, Kalyanamuthiah A, Dharanirajan K (1996) Management of coastal environments in Tamil Nadu and Andaman and Nicobar Islands based on remote sensing and GIS approach. MAEER'S MIT Pune J 4(15–16):129–140

Ramachandran S, Sundramoorthy S, Krishnamoorthy R, Devasenapathy J, Thanikachalam M (1998) Application of remote sensing and GIS to coastal wetland ecology of Tamilnadu and Andaman and Nicobar group of islands with special reference to mangroves. Curr Sci 75(3):101–109

Rani DS, Venkatesh MN, Sri CNS, Kumar KA (2018) Remote sensing as pest forecasting model in agriculture. Int J Curr Microbiol Appl Sci 7(3):2680–2689

Reis S (2008) Analyzing land use/land cover changes using remote sensing and GIS in Rize, North-East Turkey. Sensors 8(10):6188–6202

Schumann GJ, Brakenridge GR, Kettner AJ, Kashif R, Niebuhr E (2018) Assisting flood disaster response with earth observation data and products: a critical assessment. Remote Sens 10(8):1230
Setiyono T, Nelson A, Holecz F (2014) Remote sensing based crop yield monitoring and forecasting. Crop monitoring for improved food security
Shanmugapriya P, Rathika S, Ramesh T, Janaki P (2019) Applications of remote sensing in agriculture—a review. Int J Curr Microbiol App Sci 8(01):2270–2283
Sun WC, Ishidaira H, Bastola S (2010) Towards improving river discharge estimation in ungauged basins: calibration of rainfall-runoff models based on satellite observations of river flow width at basin outlet. Hydrol Earth Syst Sci 14(10):2011–2022
Welch R, Remillard MM, Slack RB (1988) Remote sensing and geographic information system techniques for aquatic resource evaluation. Photogram Eng Remote Sens 54:177–185

Part II
Case Studies: Applications of Geographic Information Systems in Urban Planning and Management

Chapter 9
Case Study 1: Monitoring and Modelling of Urban Land Use Changes

Abstract Land use-land cover (LULC) changes are an important indicator for urban planning and management. Understanding the patterns of LULC change aids in the effective management of all available resources, particularly in regions where there is little or no reported data on the status of LULC. In this study, remotely sensed satellite imagery from Landsat 5 and Landsat 8 was obtained for two years, 2001 and 2020, respectively. Geographic Information Systems (GIS) were used to quantify past and present LULC changes in Mirzapur District of Uttar Pradesh. To achieve these goals, the maximum likelihood classifier (MCL) was used to generate LULC maps with six class categories (water body, built-up land, forest land, crop land, barren land and fallow land). The classified maps for 2001 and 2020 were used to perform a two-decade change analysis over the region. The change analysis revealed that over the last two decades, built-up areas increased by 23.55% between 2001 and 2020. Barren and fallow land decreased by 8.37 and 5.77%, respectively, during this time period. The area under waterbodies has also decreased by 1.82%. These findings provide invaluable baseline information with which the government and other concerned stakeholders, urban planners and decision-makers can better manage available resources and monitor environmental changes in order to ensure the sustainable living in the region.

Keywords Land use/Land cover · GIS · Maximum likelihood classification · Sustainable

9.1 Introduction

The land can be divided into two categories based on its utility: "Land use" and "Land cover". "Land use" refers to the sum of all arrangements, activities and inputs that people make in a specific land cover type (Khan and Jharia 2016). It includes both the sequence in which the land's biophysical attributes are manipulated and the purpose for which the land is used. It denotes how people have used the land and their living space, with a focus on the functional role of land for human needs and wellbeing.

In contrast, "land cover" refers to the observed physical and biological cover of the earth's land as vegetation, rocks, bodies of water or man-made features. The term

M. Kumar et al., *Geographic Information Systems in Urban Planning and Management*,
Advances in Geographical and Environmental Sciences,
https://doi.org/10.1007/978-981-19-7855-5_9

"land cover" referred to the various types of features found on the earth's surface. Land cover types include corn fields, lakes and concrete highways. The term "land use" refers to the human activity or economic functions associated with a particular plot of land (Lillesand and Kiefer 2009).

"Land cover" refers to the physical and biological cover over the earth's surface, including water, vegetation, bare soil or rock and ice, according to the Encyclopaedia of Geography. "Land use refers to how human activities such as farming, industry, transportation and urbanization have transformed the earth's surface".

With the increase in population human interventions in the nature is being increased gradually. Increased human interventions has largely impacted the spatiotemporal trends of land use-land cover change. Having knowledge of the land use/land cover is very much essential for the planning, development and natural resource management of the region. By understanding the information of spatiotemporal dynamics of LULC one could very well establish the relationship and inter-linkages between human and natural phenomenon. Thus, it is crucial to monitor and model the LULC change to manage the sustainability over the region. As a result, several studies are being done in the field of LULC monitoring and modelling (Reis 2008; Kaif et al. 2014; Kumari 2015). Remote sensing is playing a major role in this field. Remote sensing is defined as "The science and art of obtaining information about an object, area or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area or phenomenon under investigation" (Lillesand and Kiefer 2009). GIS is helping to process the data collected through the remote sensing. GIS is defined as a system for capturing, storing, verifying, manipulating, analysing and displaying spatially referenced data to the earth.

For the following reasons, remote sensing methods are becoming increasingly important for mapping land use and land cover:

- A large area image can be acquired quickly.
- The image can be acquired with spatial and spectral resolutions that correspond to the level of detail required.
- Provide specific information about both the accessible and inaccessible areas.
- Image interpretation is less expensive and faster than conducting a ground survey.
- Using different time series remote sensing images allows for temporal changes in land use and land cover.

GIS is the very crucial component for the land use-related studies in the recent world (Bhatta 2011). We all know that remote sensing data have wide applicability due to its synoptical view, repetitive-real-time data acquisition. In order to conserve the ecosystem and environmental management, various new tools are being offered by the RS & GIS technologies. Technological developments have shown significant impact on the planning and development-related activities. Planning in the country like India is very much essential because it has vast geographical extent and huge level of diversity is prevailing over the region. By using the GIS in the field of land use exploration a new dimension could be provided to the study. By incorporating the GIS and remote sensing technologies in the LULC study it could be make much holistic in nature. One major benefit of the GIS technologies is that it provides a detailed understanding of the natural phenomenon and their dynamics (Shahab and Amin 2012). By visualizing the spatial parameters, it is easy to analyse the ongoing trends in that region.

The proposed study is an attempt to identify a land use and land cover change pattern in the District of Mirzapur in the light of changing dynamics. Land is becoming a valuable resource as a result of massive agricultural and demographic pressures. Various natural and human activities altered the earth's surface, having both positive and negative consequences. We need land use-land cover information to access these impacts. Mirzapur District has seen significant positive and negative imprints such as city region expansion and development activities such as building, road construction, deforestation and many other anthropogenic activities since its inception. As a result, without any detailed and comprehensive effort, this has resulted in increased land consumption, modification and alterations in land use/land cover over time.

To evaluate these changes over time and to examine the existing land use pattern in order to detect the land consumption rate, so that planners have a basic planning tool. As a result, a study is required so that the Mirzapur District can avoid the problems associated with these changes.

9.2 Overview of the Study Area

Mirzapur District is located in the Indian state of Uttar Pradesh (Fig. 9.1). Mirzapur District is located in the south-eastern part of Uttar Pradesh, between 23,052 North and 25°32′ North latitude and 82°07′ East and 83°33′ East longitude, covering an area of 4521 km^2.

The district is bounded on the north by the districts of Bhadohi and Varanasi, on the east by the district of Chandauli, on the south by the district of Sonbhadra and on the northwest by Allahabad. The district headquarters are in Mirzapur. The city is surrounded by hills, and the district is well known for the holy shrines of Vindhyachal, Ashtbhuja and Kali khoh. It has a lot of waterfalls and natural areas. At first glance, the city appears to be a mash-up of town, villages and city life.

The district is divided into four Tehsils. Sadar, Lalganj, Marihan and Chunar are among them. In Mirzapur District, these four tehsils are further subdivided into 12 administrative blocks: Chhanbey, Kon, Majhawa, City, Pahari, Patehara, Hallia, Lalganj, Jamalpur, Narayanpur, Rajgarh, Shikhar and 1967 villages (Census of India 2011).

9.3 Database and Methodology

Since the 1950s, it has been common practise to map land use using satellite imagery. Small-scale satellite images have recently been used for large area mapping. The

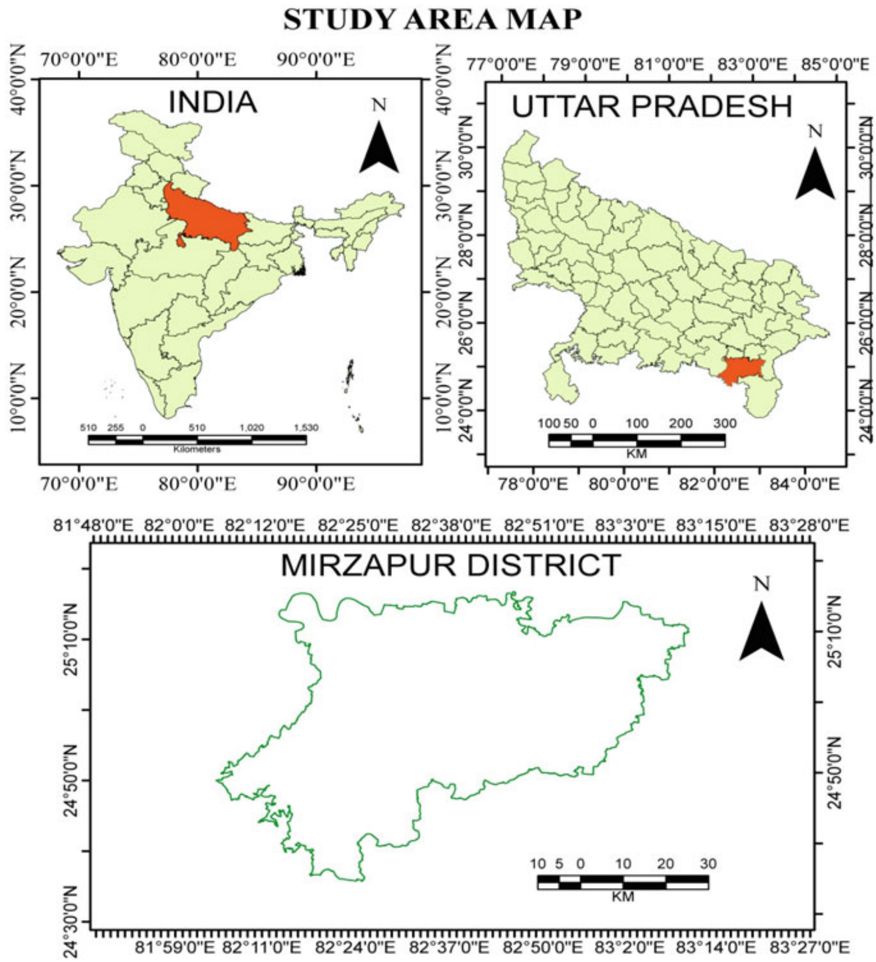


Fig. 9.1 Study area map

methodology used to create the LULC map was screen visual interpretation of satellite images. Visual interpretation keys such as tone, texture, size and pattern are validated using Google Earth and the Survey of India toposheet.

9.3.1 Data Used

Landsat satellite imagery of the Mirzapur area was obtained from the USGS site in 2001 and 2020. Other sources from which I obtained information include the United States Geological Survey's (USGS) land use-land cover classification

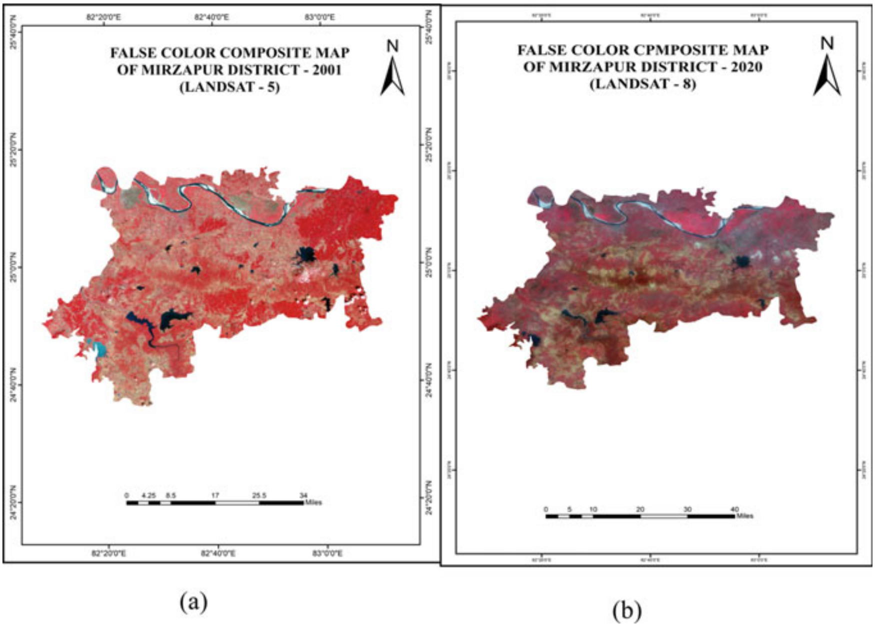


Fig. 9.2 **a** FCC map of Mirzapur district for 2001; **b** FCC map of Mirzapur district for 2020

system. The National Remote Sensing Agency (NRSA), Hyderabad, created a visual interpretation key and a land use-land cover classification scheme (Fig. 9.2).

9.3.2 Methodology

Figure 9.3 depicts the flowchart of the method used in this study. The study is based on satellite data and a limited field survey, the specifics of which are discussed in the following sections.

9.3.3 Image Acquisition and Preprocessing

Freely available Landsat 5(ETM+) for year 2001 and Landsat 8(OLI-TIRS) for year 2020 data was downloaded from the USGS earth explorer (earthexplorer.usgs.gov/).

It is very important to perform preprocessing of the data in order to rectify the errors during the process of scanning, transmission and data acquisition. After completion of the preprocessing classification of the data could be performed. Basically, preprocessing is intended to improve the radiometric and geometric qualities of the remotely sensed images.

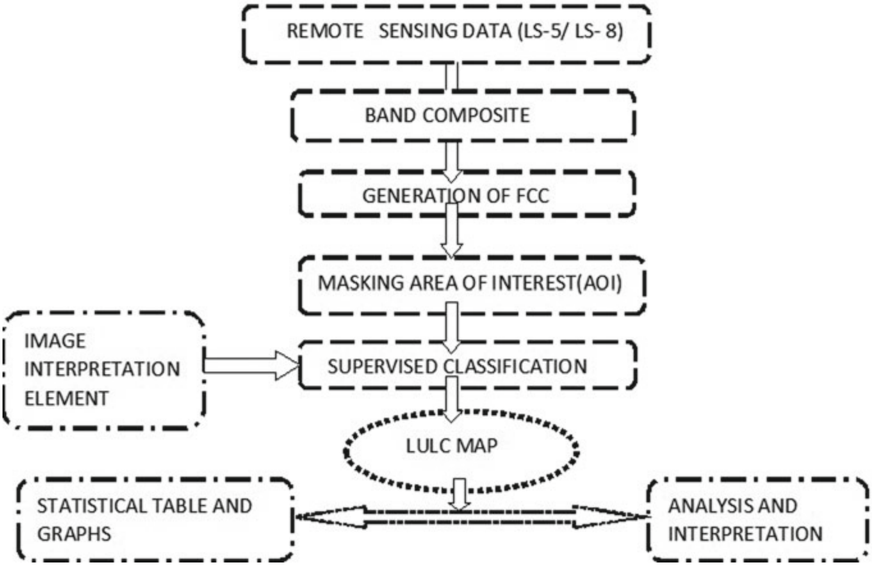


Fig. 9.3 Methodological flowchart

This could be done for the following purposes:

(i) Radiometric correction was performed to eliminate the effects of the atmosphere.
(ii) Geometric correction which is intended to register the image so that it could be used at any of the other applied reference system.
(iii) Noise removal is essential to eliminate any type of noise occurred due to limitation of the transmission and data acquisition process.

But, in this Landsat 5(ETM+)-2001 and Landsat 8(OLI/TIRS)-2020 image are already noise, and cloud-free, and also geometric corrected. Datasets converted into FCC by process of layer stacking and then mosaic prepared to cover the desired area of interest. The dataset is projected in UTM projection system (datum WGS 1984, 44N zone). Before the classification, the dataset is masked for the study area.

9.3.4 Image Classification

For the LULC classification, “Anderson Classification System” (Level 1) was adopted in this study. Land use/land cover of study area classified into built-up land, crop land, fallow land, barren land, forest and waterbodies. Same classes assign to both year datasets. Jensen (2006) explains the image classification system as a mapping process which is intended to develop a meaningful sense of each pixel of the raster. This generalizes the pixels of the raster into the certain meaningful categories. One typical method for the classification of RS image is known as pixel-based classification method. In this method, different pixels value is being considered by the classifiers and then makes their grouping on the basis of their spectral properties. In the present study supervised classification is used to classify the data. Image classification follows some criteria.

9.3.5 Criteria for Classification

LULC classification system which was designed and developed according to the USGS is mentioned below:

(1) Interpretation accuracy using remotely sensed data should be more than 85%.
(2) Accuracy level of the correctly interpreted classes should be more or less similar.
(3) For the one-time data, result from various users should be the similar.
(4) Such classification system should be applicable to the extensive area.
(5) It should be noticed the difference between the land use with that of the land cover types very clearly.
(6) Defined classification system should be able to implement the remotely sensed data over the different period of time in year.
(7) To classify a large-scale image, land use categories should be divided into large number of subcategories on the basis of the ground-based survey to reduce the overgeneralization of the data.
(8) Try to aggregate the categories as much as possible.
(9) Try to compare the data with that of the future land use-land cover images as much as possible.

9.3.6 Supervised Classification

For supervised classification, we "supervise" the pixel classification process by providing numerical descriptors of the various land cover categories present in a scene to the computer algorithm. To accomplish this, representative sample sites of known cover type are employed to create a numerical "interpretation key" that defines the spectral properties for each feature type of interest. Then, each pixel in the dataset is numerically compared to each category in the interpretation key and labelled with the name of the category to which it "most closely corresponds".

9.3.7 Post-classification Processing

Because of spectral confusion, supervised classification produces inadequate classification results. Confusion is caused by the spectral resemblance of various objects. For example, barren ground, agricultural fallow land and open woodland are frequently labelled incorrectly. The work used post classification processing of the classified maps to avoid such issues. This study's post classification processing includes recoding of classified erroneous classes for both year datasets. In recoding we identify the wrong classified area with the help of Google Earth and toposheet and digitize that area into polygons then digitized polygons are converted into AOI and within AOI, corresponding classes of LU/LC image are changed by recoding. Then final land use-land cover map is prepared.

- **Formula Used**: Changed Area = Area in 2020 (km^2) – Area in 2001 (km^2)
- **Software Used**: ArcGIS 10.8, MS Excel, MS Word

9.3.8 Result and Discussion

Land use-land cover classification deals primarily with man activity and non-activity on land. Here classification is based on what is observed on the land with remote sensing data, i.e. satellite imagery, on the background and minimum meaningful grouping is tried to make. Although to develop any land use-land cover classification system is essential to consider certain criteria limitations of satellite data and study area pertaining to the local conditions. In this mainly classified land use of study area are as follows:

(1) Waterbodies; (2) Built-Up Land; (3) Forest; (4) Crop Land; (5) Barren Land; (6) Fallow Land

In Table 9.1 and Fig. 9.4, there is a comparison of satellite image characteristics and LU/LC change from year 2001 to 2020 respectively, that how area under different categories transformed or naturally changed. This pattern of change can help for interpretation of changing pattern of LU/LC of Mirzapur District. After processing the data of year 2001 and 2020 following result has been come out;

Table 9.1 Satellite image characteristics

Characteristics	Year 2001	Year 2020
Name of the satellite	Landsat 5	Landsat 8
Sensor	ETM+	OLI/TIRS
Resolution	30 m	30 m
Path	142	142
Row	043	043
Date of acquired	23-10-2001	30-12-2020
Output format	GEOTIFF	GEOTIFF
Ellipsoid	WGS84	WGS84
UTM zone	44N	44N

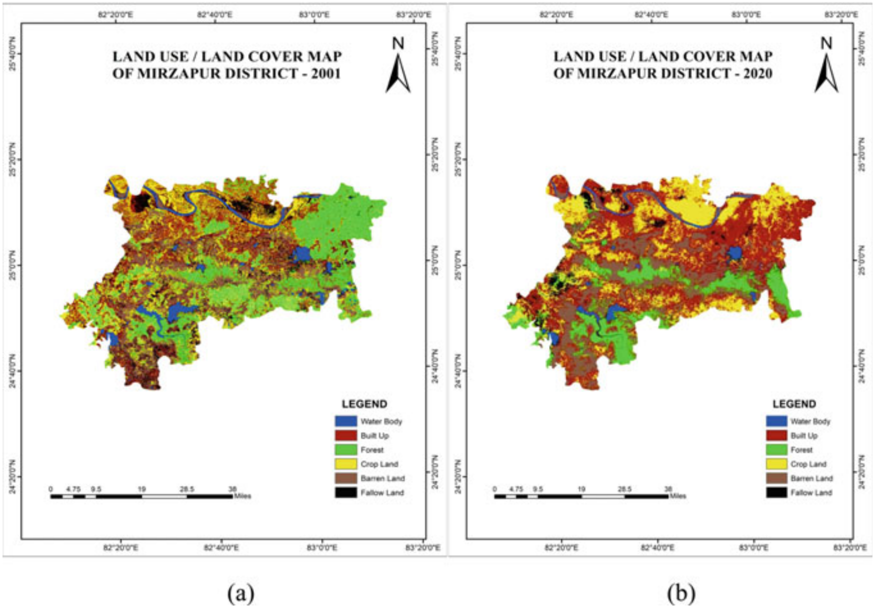


Fig. 9.4 **a** LULC map of 2001; **b** LULC map of 2020

9.3.8.1 Land Use-Land Cover Pattern of Mirzapur District in 2020

Figure 9.5 shows the land use-land cover pattern of Mirzapur District in year 2020. In this year area under built-up land becomes 36.74% of total area. This increase occurs mostly around city region. Area under crop was significantly high than the other categories. As population pressure increases, area being transformed into another use. Fellow land remains only 3.5%, and barren land remains only 20.05% of total. Area under forest is very low. Barren area decreases. Area under waterbodies was only 2.40%.

Figure 9.5 shows the relative area under different categories in year 2001 and year 2020, respectively.

Comparative study of these shows the relative change from 2001 to 2020 in Mirzapur District (Fig. 9.6).

9.4 Conclusion

Aim of study was to detect the land use-land cover change of Mirzapur District over a time period of almost 20 years using remote sensing and GIS techniques. Among the different categories of LULC crop land was dominant in both 2001 and 2020 whereas area under forests was abysmally low. Change analysis shows

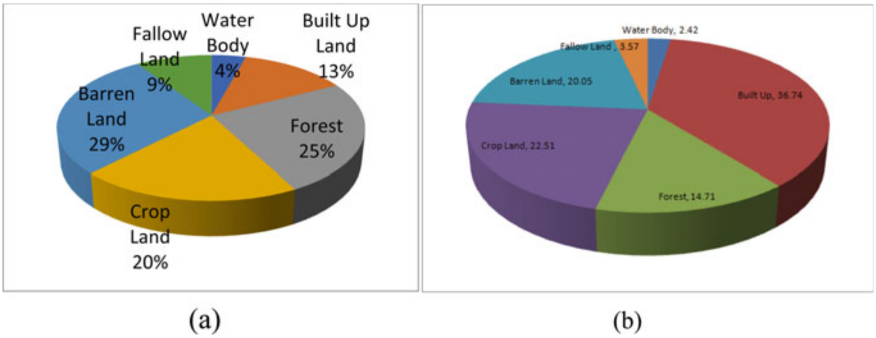


Fig. 9.5 **a** Area under different categories of LULC in Mirzapur district of year 2001; **b** Area under different categories of LULC in Mirzapur district of year 2020

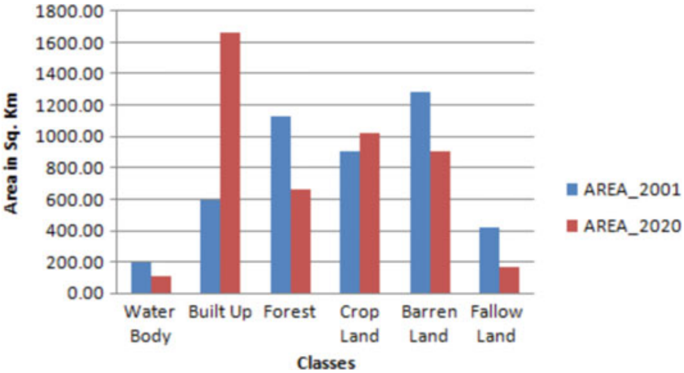


Fig. 9.6 Change in LULC of Mirzapur district from 2001 to 2020

increase in built-up area with fast rate which has grown by 23.55% in 20 years. This has negative impact on other categories of land use/land cover. Barren land and fallow land decreased by 8.37% and 5.77%, respectively. Reduction in area under fallow land and barren land encouraged agricultural activity, plantation and construction. There is small decrease only by 1.82% in area under waterbodies. A comprehensive LULC map was developed for two distinct years for a period of almost 20 years to study the changing pattern of land use/land cover. Therefore, LULC were distinctly developed for each study year. The main findings and results address the objectives specified earlier in chapter first. To prepare Multi-temporal GIS database for the future LULC change studies. The main objective of this study is to assess the spatial and temporal LULC change of Mirzapur District has been met. The LULC changes between 2001 and 2020 has been evaluated. The study explored the spatial and temporal characteristics of different land use expansion and contraction. The conversion of barren land and fallow land into agricultural land is signs of rational utilization of resources in Mirzapur District. But it is not desirable if material development takes place at the cost of replacement of natural resources. The extensive and excessive deforestation in the study area caused by the encroachment of agriculture and construction has rendered the development process unsustainable, uneco-friendly in character that dissertation.

References

Bhatta B (2011) Remote sensing and GIS, 2nd edn. Oxford University Press, New Delhi

District Census Handbook of Lucknow, Census of India (2011)

Jensen JR (2006) Remote sensing of the environment, 2nd edn. Published by Dorling Kindersley India. ISBN 10: 0131889508, ISBN 13: 9780131889507

Kaif KM et al (2014) An analysis of LULC change detection using remotely sensed data: a case study of Bauchi City. In: IOP conference on series: earth and environmental science, Selangor Malaysia

Khan R, Jharia DC (2016) Land use land cover change detection using remote sensing and geographic information system in Raipur Municipal Corporation Area, Chhattisgarh. SSARSC Int J Geo Sci Geo Inf 3(1)

Kumari K (2015) Urban sprawl: a case study of Lucknow city. Int J Human Soc Sci Invent 11–20

Lillesand TM, Kiefer RW (2009) Remote sensing and image interpretation, 5th edn. John Wiley and Sons, New York

Reis S (2008) Analysing land use/land cover changes using remote sensing and GIS in rise. North-East Turkey, Open access sensors, 8

Shahab F, Amin A (2012) Land transformation analysis using remote sensing and GIS techniques (a case study). J Geogr Inf Syst 229–236

Chapter 10
Case Study 2: Simulating Future Urban Growth Using Cellular Automata-Markov Chain Models

Abstract Developing nations are facing a serious threat of urbanization. India is one out of them. Keeping this concept on mind we tried to quantify the urban area expansion by using advanced geospatial technologies in this case study. Bangalore city was selected as our target study area. Here we assessed the urban area expansion from 2001 to 2025 in future. For that purpose, data for the years 2001, 2011 and 2016 was downloaded from Landsat Thematic Mapper (TM) and Operational Land Imager (OLI). ERDAS Imagine was used for land use-land cover (LULC) map classification for the years 2001, 2011 and 2016. LULC map was classified into five LULC classes, viz. urban, forest, vegetation, fellow land and water body. Cellular Automata-Markov (CA-Markov) Chain Modelling was used through IDRISI Selva software for the prediction of future land cover. Land Transition Probability matrix was also created using Markov model for the time period 2001–2025. 2016 LULC map was used for the validation of CA–Markov Model. Result revealed that urban area will be increased to 33% in 2001 to 55% in 2025. Nearly forest cover will become half from 30% in 2001 to 15% in 2025. Vegetation will also reduce from 33% in 2001 to 24% in 2025. This case study revealed the fact that rapid rate of the urban area expansion is being observed in the region. But this is happening at the cost of forest and vegetation loss. Result showed that these two land use classes will decrease substantially till 2025. Considering this result, it is the responsibility of the concerned authorities and government stakeholders to prepare a wholistic plan in such a way that proper urbanization could be managed at the least cost of green cover in the region.

Keywords Urbanization · Land use-land cover · Transition probability · CA-Markov

10.1 Introduction

Population is growing tremendously all over the world as a result urbanization is happening at very fast rate across the world (Angel et al. 2011). Various problems are being generated due to uncontrolled growth of population and urbanization. Some important of them are observed in the form of rush in traffic, air pollution,

M. Kumar et al., *Geographic Information Systems in Urban Planning and Management*,
Advances in Geographical and Environmental Sciences,
https://doi.org/10.1007/978-981-19-7855-5_10

deforestation, etc. (Jenerette et al. 2007). These problems could be overcome whether one accurately and efficiently models the potentiality of near-future urban growth.

In last few years, geospatial technologies such that Remote Sensing and Geographic Information System is being used extensively in the field of exploration of urban expansion and problems related to it. Geospatial tools and techniques are proved as very efficient techniques to deal with such type of problems.

As a common approach, the Markov Chain could perform an analysis of transformation of various land uses/land cover (LULC) and forecast their possible trends in near future (Mubea et al. 2011). This model consists of a series of random values, the probabilities of that which change over the period of time based on previous values. In other words, it measures the transition probability of the alteration of one land use class to other classes in a period of time. This model is efficient in the prediction of LULC alteration where it is not so easy to describe the alteration in the landscape (Kumar et al. 2014). Static transition anticipated by Markov chain models is one of the concerns from claim associated with adopting this strategy, which makes it more suitable for short-term prediction (Lambin et al. 2001). Nevertheless, one of the major constraints of Markov Chain is its 'lack of spatial dimension', and it is not able to recognize the spatial distribution of instances in each LULC type (Ye and Bai 2008).

Cellular Automate (CA) is widely used to estimate the expansion in urban areas. Use of Cellular Automata provides various benefits in the study:

- Help in the delineation of the potential urban expansion so that planners could implement their policies effectively.
- Help to reduce the adverse impact of uncontrolled and unsystematic urban expansion.
- Help to efficient management of available natural resources in order to ensure the sustainable growth of the region by estimation of the urban area expansion effectively.

CA is one of the most implemented methodologies for the assessment of LULC change analysis and its future prediction since the end of last century. CA modelling and forecasting are based on the assumption that "historical urban areas will affect future expansion patterns through interaction among different land use types" (Clarke et al. 1997). CA can imitate the emerging new macroscale occurrence (urban growth) from microlevel interaction due to its "bottom-up" framework. CA always begins with the cells with the maximum likelihood of change when assigning changes based on predetermined conditional rules. As a result, it is able to accurately predict the much more expected development sites, assess the probable amount of change and assign the estimated changes in the region. But CA is not absolute in nature. There are also certain limitations associated with it. One major drawback of the CA is that it only considers the nearby influence in spatial distribution without any quantifiable consideration of the role of drivers of urban expansion; however, this issue could be addressed by integrating CA with other techniques like Fuzzy Logic and Markov Chain (MC). In this case study to get through with this innate problem of CA we have integrated it with Markov Chain model. This approach to integrate one model with another model is very popular model for LULC change modelling process (Guan et al. 2011). Various other studies such that Guan et al. (2011) and Ye and Bai (2008) have used this integrated model for the LULC modelling. The use of combining CA with other models permits for quantitative evaluation of external factors in the GIS environment, where spatiotemporal data can function as citation analysis.

Integration of Cellular Automata model with Markov Chain model is very much efficient to assess and forecast future land use change, as it does not need vast amount of data. Even considering a little number of controlling factors one could also model the future urban growth accurately and precisely (Arsanjani et al. 2013).

In our study we have implemented the Markov Chain model integrated with Cellular Automata model to assess the future urban growth in Bangalore city. Bangalore is one of the major urban centres of India. With the passage of time, it is growing at very huge rate. Its tremendous growth could create a number of problems in front of policymakers and concerned stakeholders.

Main purpose of our case study is to evaluate the urban growth in the city in last two decades and forecast the future urban expansion after 5 years. This will help to understand the trend of urbanization in the region and help the policymakers to implement their plan effectively.

10.2 Overview of the Study Area

Capital of the Karnataka is 'Bangalore' which is also known as the "Garden City of India" and also the "Silicon City of India". Bangalore holds the 6th largest rank in the metropolitan cities of India. It is located in the southern part of the Indian Peninsular Deccan Plateau. It is extended between 77° 24′ E to 77° 47′ E longitude and 12° 45′ N to 13° 11′ N latitude.

Study region is expanded in the area of 1313 km^2 approximately. Elevation of the study area is varying from 637 to 902 m. Temperature is varying from 14 to 33 °C. Rainfall is observed around 1050 mm/year. Bangalore is famous for its elite-class education and technical hub. Various multi-national companies established their offices here. Its growth has increased up to 10 times since the independence. Bangalore is also observed as the administrative, cultural and commercial capital of Karnataka. It has witnessed significant urbanization and expansion of urban in current times as a result of intense development initiatives, with an emphasis on industrialization for the region's economic development. The intensive growth has led to a rise in population, putting pressure on infrastructure and natural resources, and ultimately giving rise to a slew of substantial environmental challenges, increased greenhouse gas emissions, a lack of proper infrastructure, traffic jams and a lack of basic amenities such as electricity and sanitation and hygiene in many areas, among others.

In this case study we will assess the trend of urbanization from 2001 to 2025. Study area is shown in Fig. 10.1.

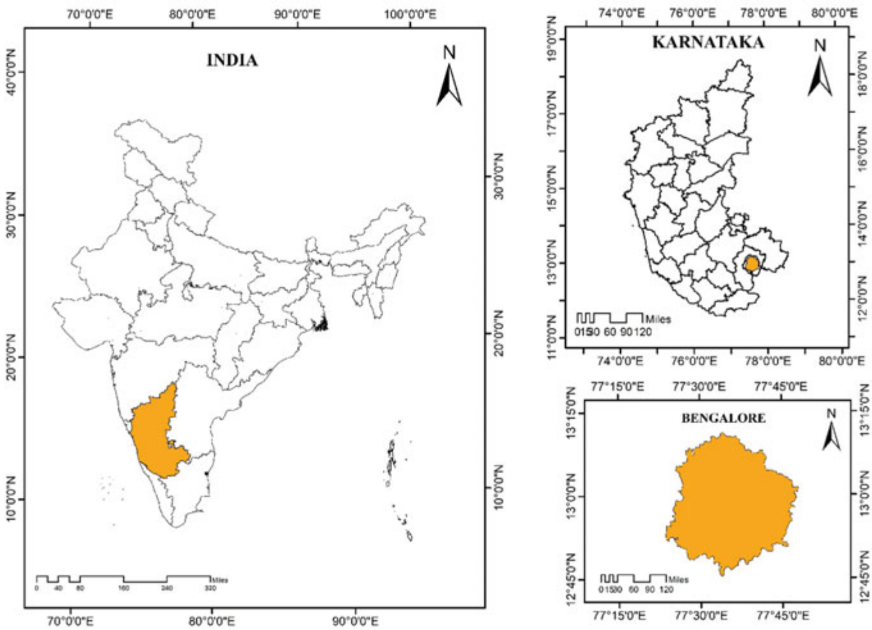


Fig. 10.1 Study area

10.3 Materials and Methods

10.3.1 Data Collections

In this study we are intended to perform future urban growth modelling. For that purpose, it is required to conduct an effective analysis of future LULC scenario. That is why a large amount of earth observation data is required to fulfil this objective (Araya and Cabral 2010). In this study we have collected the data from Landsat Thematic Mapper (TM) and Operational Land Imager (OLI) for the years 2001, 2011 and 2016. On the basis of these data LULC map was prepared for the years 2001, 2011 and 2016. LULC map was classified into five categories for all the three classifications, viz. water body, vegetation, urban, forest and fallow land.

10.3.2 Data Processing

10.3.2.1 Markov Chain Model

Being a stochastic model, Markov Chain models forecast the likelihood of the future land use change on the basis of past LULC change trend analysis on a spatiotemporal scale basis. In other words, it could be said that Markov Chain is nothing but simply a series of the random values whose probability for a defined time period is estimated by the previous value of the number (Surabuddin Mondal et al. 2013). Output of the Markov model is depending on the transition probability (Adhikari and Southworth 2012). To project the future LULC map, LULC transition probability matrix for the change from one period to another period is used. Probability matrix is defined as a set of the conditional probabilities for the model to change one cell from one class to another class during that interval (Akin et al. 2014). This model could be used to forecast how a specific LULC will change over the period of time (Fan et al. 2008; Hyandye and Martz 2017). The Markov model is able to predict the quantities of each LULC type or the temporal alteration of the LULC pattern, but it is not so appropriate at dealing with the spatial pattern of landscape change (Li et al. 2015), because the change in spatial distribution cannot be provided with the help of this model (Halmy et al. 2015).

The basic idea which is working behind the Markov chain model is that LULC at some point in the future ($n + 1$) can be predicted based on current LULC (n) (Iacono et al. 2015). Coppedge et al. (2007) supported this concept and stated that the LULC change for any location is not completely random, but it depends on the condition of the previous or current LULC situation. Following equations represent the formula used in the Markov Chain model.

$$L_{(n+1)} = r_{ij} \times L_n \tag{10.1}$$

$$r_{ij} = \begin{bmatrix} r_{11}\ r_{12}\ \ldots\ r_{1p} \\ r_{21}\ r_{22}\ \ldots\ r_{2p} \\ \vdots\ \ldots\ \vdots \\ r_{p1}\ r_{p2}\ r_{pp} \end{bmatrix} \tag{10.2}$$

In above Eq. (10.1), $L_{(n+1)}$ and L_n are representing the LULC at time span n and $(n + 1)$, respectively. Here, $0 \le r_{ij} \le 1$ and $\sum_{j+1}^{p} r_{ij} = 1, (i, j = 1, 2, \ldots p)$ is representing the transition probability matrix. In this paper LULC classification for the years 2001, 2011 and 2016 was done using ERDAS Imagine software. LULC map was classified into five categories as mentioned above. Markov Chain model was implemented in IDRISI Selva 17.0. For the processing of LULC map in the IDRISI, LULC map which was initially in.img file format was converted in.rst file format to process these data. First Markov was run on IDRISI between 2001 and 2016. Number of time period was defined between 2001 and 2011. 5 years were selected to project forward from the second (2011 LULC). Background cell optioned was assigned to 0.0, and proportional error was 0.15. This model provided the transition probability matrix in text format. This transition probability matrix was created by cross tabulating the two provided images of different time periods for the determination of the likelihood of a pixel in one land use class changing to another (Subedi et al. 2013). Markov transition area file and transition suitability image collection are also produced as output of the Markovian model. A transition area file is defined as a text file which holds the number of pixels which are supposed to change from one LULC class to another over a given time interval (Adhikari and Southworth 2012). Overall, the transition probability matrix, transition areas matrix, and a set of conditional probability images were generated for five different LULC classes such as forest, crop, barren, built up and water body in this study. This is done for the purpose of prediction of 2016 LULC map. This is done for the validation of Markov model.

Further, this model was iterated with 2001 and 2011 again for predicting the LULC map 14 years ahead from the second image (2011 LULC map) because we are intended to assess urban change in 2025.

10.3.2.2 Cellular Automata Model (CA Model)

A Cellular Automata Model is able to analyse the changing and controlling complex spatially distributed processes. This CA Model provides clear insights into local and global patterns of LULC change that link the new phase to its predecessors and neighbours (Al-sharif and Pradhan 2014). The CA model excels at replicating the spatiotemporal features of complex systems (Yang et al. 2014). This is the reason why this model has been widely used in the process of Urban Change modelling. This model can be thought of as a dynamic and relatively simple spatial system in which, according to a set of transition rules, the state of each cell of the matrix is determined by the former condition of the cells contained within a defined neighbourhood (Rocha et al. 2007). Ye and Bai (2008) described the CA model as an 'analytical engine' that enables the dynamic modelling within GIS. Besides its benefits, the CA has some issues with transition rule definition and model structure (Rocha et al. 2007). As a result, it is unable to forecast LULC dynamics. This technique's shortcoming can be addressed by combining it with other dynamic and empirical models, such as CA-Markov (Halmy et al. 2015). Besides its benefits, the CA has some issues with transition rule definition and model structure (Rocha et al. 2007). As a result, it is unable to forecast LULC dynamics. This technique's inadequacy can be addressed by combining it with other dynamic and empirical models, such as CA-Markov (Halmy et al. 2015). For the same reason, integrated CA-Markov Chain model has been applied in this case study. Sang et al. (2011) presented the CA-Markov Chain model in below-mentioned way.

$$S(n, n+1) = f(S(n), N) \tag{10.3}$$

Limited and Discrete Cellular States is represented by S, Cellular Field is represented by N, and n and $(n + 1)$ represent the different time interval. The rule of transformation of 'Cellular States' in local space is represented by f.

10.3.2.3 Cellular Automata-Markov Chain Integrated Model

Both the Cellular Automata and Markov Chain model are dynamic in nature in both, state as well as time. This model is capable of accurately predicting transitions or spatiotemporal dynamics among various LULC types. The CA-Markov model has been widely used in many scientific studies to forecast future LULC because it combines the benefits of Cellular Automata and the Markov Chain element of spatial adjacency, as well as knowledge of the likely spatial distribution of transitions (Arsanjani et al. 2011). This is reason why both the CA and Markov Chain are dependent on each other to predict the future LULC (Omar et al. 2014). This model is able to produce a more accurate spatial and temporal pattern of LULC change (Sayemuzzaman and Jha 2014).

Hence, we have applied the CA–Markov Model for the future LULC forecasting for the purpose of assessing urban growth in near future. Here we have simulated a predicted map for 2025 on the basis of 2001 and 2011 LULC. This modelling was done on "IDRISI-TerrSet Geospatial Monitoring and Modeling System software". The transition probabilities controlled by local rules supported the CA–Markov model's prediction of future LULC. For the purpose of CA-Markov modelling process three types of data are required. First one is basis land cover image (we used 2011 LULC in this case study), second one is Markov transition area file generated from Markov modelling, and third one is the transition suitability image collection generated from the Markov modelling. Cellular Automata iterations were fixed at 10. To define the neighbourhoods of each cell and to generate better results and create spatially explicit contiguous weighing factors, a predefined standard contiguity filter of 5 × 5 (Eq. 10.4) was used. The pixels that are further away from the existing LULC class are less suitable than those that are closer (Subedi et al. 2013).

$$5 \times 5\,\text{Contiguity Filter} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{10.4}$$

10.4 Result and Discussion

10.4.1 LULC Change Analysis and Urban Sprawl

On the basis of predicted LULC of 2025 and initial LULC of 2001, "Land Change Modeler" was used to perform the detailed change analysis and urban sprawl in the region. Result revealed that maximum urban change has occurred at the cost of vegetation and forest, respectively. These two classes are the major contributors for the urban change. Figure 10.2 represents the contribution of all factors for the net urban change from 2001 to 2025.

In Fig. 10.2 we can clearly observe the above-mentioned fact that vegetation contributed the most in urban change while water body contributed the least.

Fellow Land and Urban areas are increased substantially from 2001 to 2025. Result from CA-Markov model showed that the urban areas will be increased by approximately 289 km^2 and reach up to 730.42 km^2 till 2025. This will be 65% increment as compared to base year, 2001. Result revealed that highest increment will be observed in fellow land, urban areas and water body, respectively, while forest and vegetation have decreased substantially as compared to 2001. Figure 10.3 represents the net change in all land use categories from 2001 to 2025.

In 2025 urban areas will reach up to 730 km^2, forest cover will be 207.6 km^2, vegetation will be 314.7 km^2, fellow land will be 32.2 km^2 and water body will be 28.7 km^2. In the initial year of 2001, urban areas were 441.5 km^2, forest cover was 397.5 km^2, vegetation was 434.8 km^2, fellow land was 19 km^2 and water body was 21 km^2. Clear transition in from various classes to urban classes is shown in Fig. 10.4.

Spatial distribution of the transition from various classes to urban is obtained from Land Change Modeler and is represented in Fig. 10.5.

Temporal Distribution of different land use classes from 2001 to 2025 is presented in Table 10.1.

Urban sprawl could very well be observed from the produced map of 2025 by CA–Markov model. In Fig. 10.6 a comparative analysis could be done from the LULC map of 2001 and 2011 and projected LULC map of 2025.

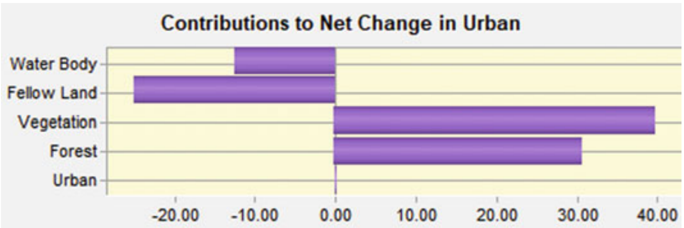


Fig. 10.2 Contributions to net change in urban areas from 2001 to 2025

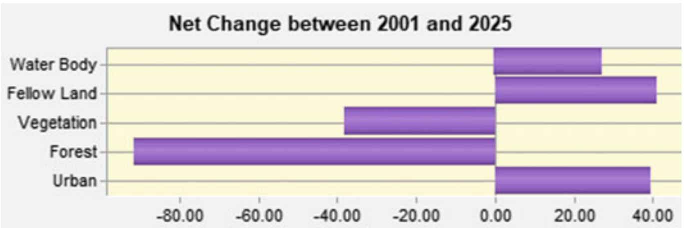


Fig. 10.3 Net change in all classes from 2001 to 2025

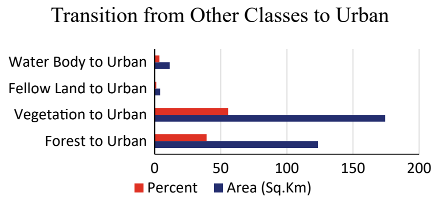


Fig. 10.4 Area-wise transition to urban class

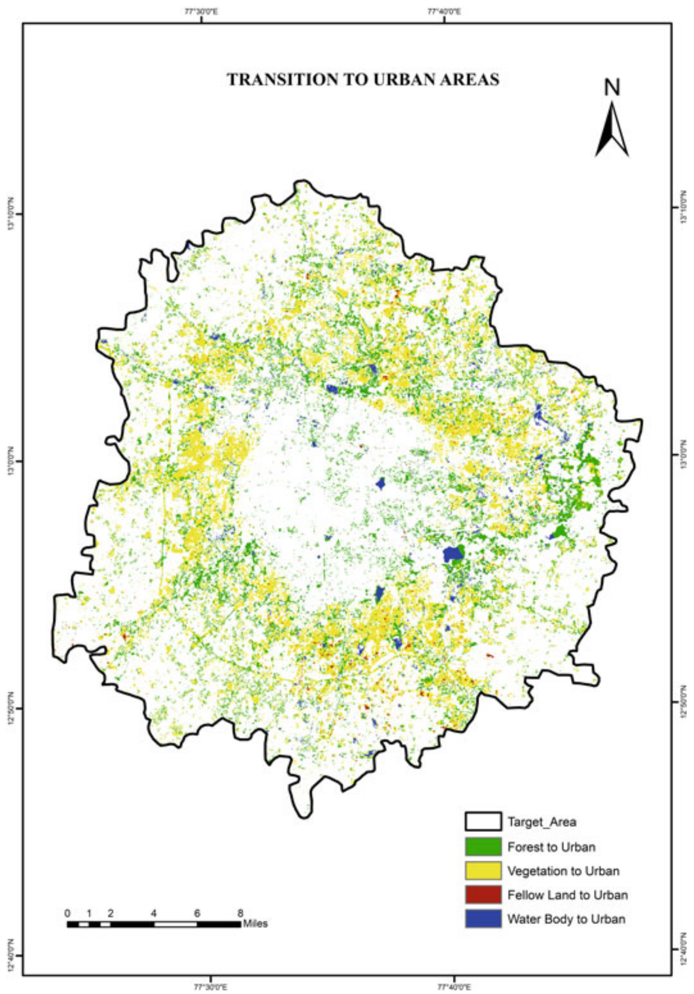


Fig. 10.5 Spatial distribution of transition from various classes to urban area

Table 10.1 Land use-land cover change from 2001 to 2025

Year	Urban		Forest		Vegetation		Fellow land		Water body	
	Area (km^2)	%	Area (km^2)	%	Area (km^2)	%	Area (km^2)	%	Area (km^2)	%
2001	441.49	33.61	397.50	30.26	434.81	33.10	19.02	1.45	20.84	1.59
2011	611.63	46.56	281.44	21.42	400.55	30.49	11.80	0.90	8.27	0.63
2025	730.42	55.60	207.62	15.80	314.68	23.95	32.22	2.45	28.73	2.19

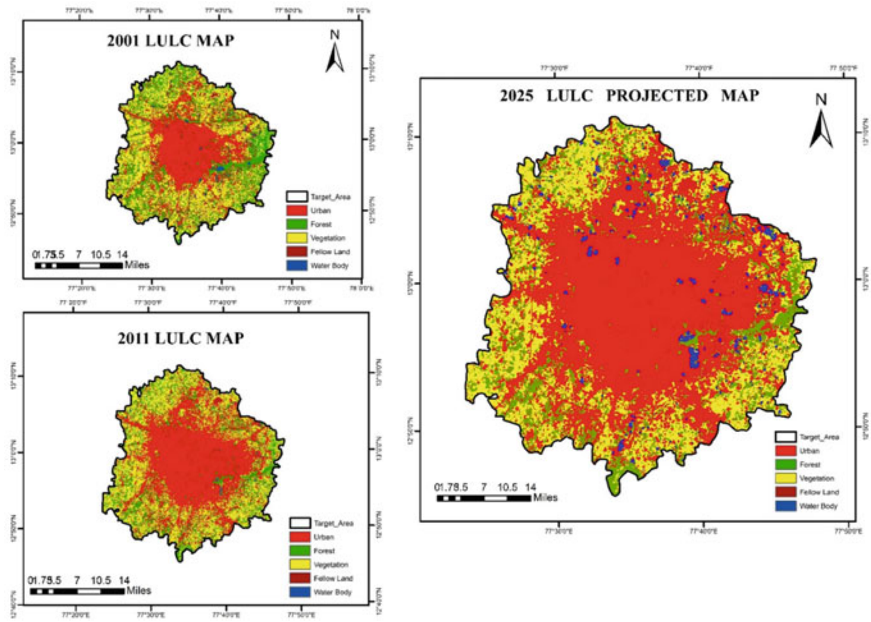


Fig. 10.6 LULC map of Bangalore city for the years 2001, 2011 and 2025, respectively

10.4.2 Analysis of the Markov Transition Probability Matrix

Using the IDRISI software, we can obtain the Markov transition probability matrix for Bangalore land use and analyse the land use change dynamics from 2001 to 2025. Table 10.2 shows the land use transition probability matrix, with values ranging from 0 to 1. The closer to 1, the more likely it is to transition from horizontal to vertical land use, and vice versa. Table 10.2 shows that the values on the axis of symmetry for urban, forest and fellow land are the largest in their respective rows (except for the other classes), implying that the probability of land use type transfer to its own type is the highest and that the land use type is most likely to remain unchanged. The probability of transferring urban land to urban land is 0.85. As can be seen, the stability of other land uses is poor and heavily influenced by other land use types.

Table 10.2 Transition probability matrix of various landuse classes from 2001 to 2025

2001	2025				
	Urban	Forest	Vegetation	Fellow land	Water body
Urban	0.85	0.04	0.03	0.04	0.04
Forest	0.24	0.33	0.41	0.01	0.01
Vegetation	0.34	0.22	0.42	0.01	0.01
Fellow land	0.27	0.11	0.57	0.04	0.01
Water body	0.47	0.3	0.21	0.01	0.01

Water bodies are observed as stable land use patterns in comparison with others. The urban area has seen the greatest increase in land use, primarily at the expense of vegetation and forest. Table 10.2 shows that there were mutual transformations between different land use patterns. Transition probabilities from fellow land and forest land to urban were nearly identical. Transition probability from water body to forest and forest to forest is comparable. According to the transition probability, the water body was the most unstable land use pattern.

10.4.3 Validation

To validate the performance of CA-Markov model, LULC map of 2016 was predicted by following the same procedure as discussed in Sect. 10.3. After prediction of 2016 LULC map, this was cross-validated by the actual LULC of 2016 which we have already classified using ERDAS Imagine. We have created the 1400 spatially balanced point throughout the region. And on the basis of these points Spearman's Rank Correlation method was calculated. Result revealed that both the maps are 78% correlated to each other. We can proceed with this level of accuracy.

Figure 10.7 is representing the projected 2016 LULC map and actual classified 2016 LULC map.

10.5 Conclusion

In the presented case study, we have discussed the urban growth modelling using CA-Markov integrated method. Detailed analysis and interpretation have been done in the result and discussion section. Here, it was identified that urban expansion is facing very tremendous rate by the Silicon Valley of India. This case study revealed that urban area expansion in Bangalore is observed at the rate of $\sim$ 1%/year. This situation is very alarming in nature because this vast expansion is taking place at the cost of forest and vegetation land. Existence of forest and vegetation is crucial for the survival of human being. Because without having enough amount of vegetation

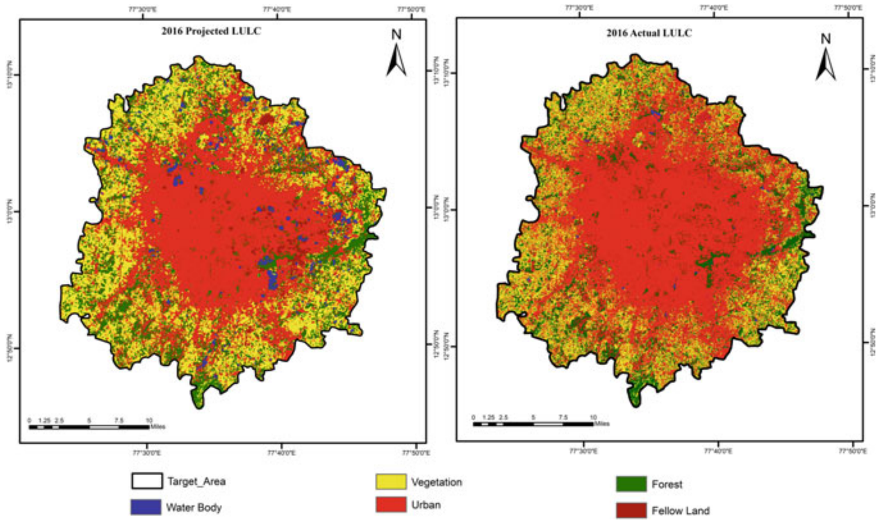


Fig. 10.7 Predicted and actual LULC map of 2016 for validation

and forest land natives will face not only problems related to their food but they will also face problem in the process of respiration. This situation will be accelerated by increased urban population.

Necessity of clean environment and requirement of food is essential for human life survival. Ensuring these all-natural resources is our moral duty and ask the government to do so as these are our 'Fundamental Duty' also. Keeping these all-stark realities in mind it is the humble duty of the concerned authorities and government stakeholders to prepare the plan in such a way that not only the proper growth of urbanization could be managed but also the sustainability of the forest and vegetation could be conserved for our upcoming generations.

References

Adhikari S, Southworth J (2012) Simulating forest cover changes of Bannerghatta National Park based on a CA-Markov model: a remote sensing approach. Remote Sens 4(10):3215–3243

Akin A, Aliffi S, Sunar F (2014) Spatio-temporal urban change analysis and the ecological threats concerning the third bridge in Istanbul City. Int Arch Photogrammetry Remote Sens Spat Inf Sci 40(7):9

Al-sharif AA, Pradhan B (2014) Monitoring and predicting land use change in Tripoli Metropolitan City using an integrated Markov chain and cellular automata models in GIS. Arab J Geosci 7(10):4291–4301

Angel S, Parent J, Civco DL, Blei A, Potere D (2011) The dimensions of global urban expansion: estimates and projections for all countries, 2000–2050. Prog Plan 75(2):53–107

Araya YH, Cabral P (2010) Analysis and modeling of urban land cover change in Setúbal and Sesimbra, Portugal. Remote Sens 2(6):1549–1563

Arsanjani JJ, Kainz W, Mousivand AJ (2011) Tracking dynamic land-use change using spatially explicit Markov Chain based on cellular automata: the case of Tehran. Int J Image Data Fusion 2(4):329–345

Arsanjani JJ, Helbich M, Kainz W, Boloorani AD (2013) Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. Int J Appl Earth Obs Geoinf 21:265–275

Clarke KC, Hoppen S, Gaydos L (1997) A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. Environ Plann B Plann Des 24(2):247–261

Coppedge BR, Engle DM, Fuhlendorf SD (2007) Markov models of land cover dynamics in a southern Great Plains grassland region. Landscape Ecol 22(9):1383–1393

Fan F, Wang Y, Wang Z (2008) Temporal and spatial change detecting (1998–2003) and predicting of land use and land cover in Core corridor of Pearl River Delta (China) by using TM and ETM+ images. Environ Monit Assess 137(1):127–147

Guan D, Li H, Inohae T, Su W, Nagaie T, Hokao K (2011) Modeling urban land use change by the integration of cellular automaton and Markov model. Ecol Model 222(20–22):3761–3772

Halmy MWA, Gessler PE, Hicke JA, Salem BB (2015) Land use/land cover change detection and prediction in the north-western coastal desert of Egypt using Markov-CA. Appl Geogr 63:101–112

Hyandye C, Martz LW (2017) A Markovian and cellular automata land-use change predictive model of the Usangu Catchment. Int J Remote Sens 38(1):64–81

Iacono M, Levinson D, El-Geneidy A, Wasfi R (2015) A Markov chain model of land use change. TeMA J Land Use Mobil Environ 8(3):263–276

Jenerette GD, Harlan SL, Brazel A, Jones N, Larsen L, Stefanov WL (2007) Regional relationships between surface temperature, vegetation, and human settlement in a rapidly urbanizing ecosystem. Landscape Ecol 22(3):353–365

Kumar S, Radhakrishnan N, Mathew S (2014) Land use change modelling using a Markov model and remote sensing. Geomat Nat Haz Risk 5(2):145–156

Lambin EF, Turner BL, Geist HJ, Agbola SB, Angelsen A, Bruce JW, Xu J (2001) The causes of land-use and land-cover change: moving beyond the myths. Glob Environ Chang 11(4):261–269

Li SH, Jin BX, Wei XY, Jiang YY, Wang JL (2015) Using Ca-Markov model to model the spatiotemporal change of land use/cover in Fuxian Lake for decision support. ISPRS Ann Photogrammetry Remote Sens Spat Inf Sci 2(4):163

Mubea KW, Ngigi TG, Mundia CN (2011) Assessing application of Markov chain analysis in predicting land cover change: a case study of Nakuru municipality. J Agric Sci Technol 12(2)

Omar NQ, Sanusi SAM, Hussin WMW, Samat N, Mohammed KS (2014). Markov-CA model using analytical hierarchy process and multiregression technique. In: IOP conference series: earth and environmental science, , June 2014, vol 20(1). IOP Publishing, p 012008

Rocha J, Ferreira JC, Simoes J, Tenedório JA (2007) Modelling coastal and land use evolution patterns through neural network and cellular automata integration. J Coast Res 827–831

Sang L, Zhang C, Yang J, Zhu D, Yun W (2011) Simulation of land use spatial pattern of towns and villages based on CA–Markov model. Math Comput Model 54(3–4):938–943

Sayemuzzaman M, Jha M (2014) Modeling of future land cover land use change in North Carolina using Markov chain and cellular automata model. Am J Eng Appl Sci 7(3):295

Subedi P, Subedi K, Thapa B (2013) Application of a hybrid cellular automaton–Markov (CA-Markov) model in land-use change prediction: a case study of Saddle Creek Drainage Basin, Florida. Appl Ecol Environ Sci 1(6):126–132

Surabuddin Mondal M, Sharma N, Kappas M, Garg PK (2013) Modeling of spatio-temporal dynamics of land use and land cover in a part of Brahmaputra River basin using Geoinformatic techniques. Geocarto Int 28(7):632–656

Yang X, Zheng XQ, Chen R (2014) A land use change model: integrating landscape pattern indexes and Markov-CA. Ecol Model 283:1–7

Ye B, Bai Z (2008) Simulating land use/cover changes of Nenjiang County based on CA-Markov model. In: International conference on computer and computing technologies in agriculture, August 2007. Springer, Boston, MA, pp 321–329

Chapter 11
Case Study 3: Identification of Potential Sites for Housing Development Using GIS-Based Multi-criteria Evaluation Technique

Abstract Due to the tremendous pace of population growth, the population is now expanding rapidly even in the hilly areas. It is now critical for policymakers and other stakeholders to look for potential built-up places to address the looming challenge. The process of choosing a built-up location is a multi-attribute decision problem. In hilly terrain, several heterogeneous criteria such as slope, road proximity, land use, distance from developed land, landslide, lithology, drainage proximity, lineament and aspect all play a role in site selection. The use of quantitative methodologies like Multi-Criteria Decision-Making (MCDM) techniques in land suitability procedures has expanded in the recent decade, allowing for the processing of diverse examined data. MCDM technique that was employed in this study to include uncertainties in decision-makers' opinions is Analytical Hierarchy Process, (AHP). Slope was given highest preference for the selection of built-up sites in hilly areas, while lowest preference was given to distance from landslide and distance to lineament and aspect by the model. Result of model validation by using ROC curve reveals that AHP is able to delineate suitability class with 92% of accuracy. Final suitability map was prepared classifying it into 5 classes. The results indicate that the AHP indicates 4.32% as very less suitable, 15.15% as less suitable, 21.64% as high suitable and 31.50% as very high suitable. In conclusion, considering the rapid pace of urbanization, the possible built-up sites selection will lead to domino effect to secure holistic hill area development and planning.

Keywords Suitability · GIS · MCDM · AHP · ROC

11.1 Introduction

To make the planning for the development of any region meaningful and holistic it's necessary to consider multi-level intra-state differences, potential and challenges of development in various regions. It is critical for the development of any region to plan through a series of interactive and multi-level exercises and to integrate decentralized strategies in order to make the optimum use of the available resources. Due to the tremendous pace of random population growth, the population is now expanding rapidly in unsystematic way and these resources are being exploited continuously.

M. Kumar et al., *Geographic Information Systems in Urban Planning and Management*, Advances in Geographical and Environmental Sciences,
https://doi.org/10.1007/978-981-19-7855-5_11

Unsystematic distribution of the population is creating a trap effect. It is now critical for policymakers and other stakeholders to look for potential built-up places to address the giant looming challenge. The process of choosing a built-up location is a multi-attribute decision problem. The use of quantitative methodologies like Multi-Criteria Evaluation Techniques/Multi-Criteria Decision-Making (MCDM) in land suitability procedures has expanded in the recent decade, able to deal with the multi-attribute decision problem. There are several MCDM techniques out of which some major techniques are AHP, Fuzzy AHP, ANP, BCM, COMET, etc.

A Multi-Criteria Decision-Making Problem is one in which we must make a decision regarding a problem based on multiple criteria. It was observed that first MCDM techniques emerged in 1960s aimed to relieve challenges in accepting multiple viewpoints and dealing with vast amounts of complex information during the decision-making process (Zopounidis and Doumpos 2002; Zopounidis and Pardalos 2010). Multi-Criteria Decision-Making entails a multi-stage procedure that includes (i) identifying objectives, (ii) selecting criteria to measure the objectives, (iii) outlining alternatives, (iv) giving weights to the criteria and (v) applying the proper mathematical method to rank alternatives. MCDM satisfies the requirement for unbiased integration of modern planning objectives for independent identification and evaluation of the best planning solutions (Ananda and Herath 2009; Herath and Prato 2006; Mosadeghi et al. 2009). These geographical MCDM techniques have the potential to improve the transparency and analytical rigour of land use decisions (Dunning et al. 2000; Hajkowicz and Collins 2007). It's difficult to assess the effectiveness of all of the criteria that apply to a given situation. It is crucial to assign an appropriate weight to them in order to assess them quantitatively. The most difficult aspect of the decision-making process is defining the weight of the criteria in relation to their importance. Saaty (1980) was the first to attempt to establish a mechanism for scaling the weight of criteria based on their importance in specific decision-making in the 1970s.

As stated earlier, we are intended to identify built-up suitability. Suitability requires effective management of land information on which such decisions should be based. Suitability approaches examine the interaction of location, development acts and environmental variables in order to classify observational units according to their suitability for a specific purpose (Collins et al. 2001; Kalogirou 2002; Keshavarzi and Heidari 2010; Malczewski 2004). Land suitability evaluation for different purposes is crucial as the effective tool for area development (Baja et al. 2001). FAO (1976) explains land suitability as the fitness of a given parcel of land for specific uses. Land suitability evaluation is one of the effective tools for such purposes (Baja et al. 2001). Generally, two approaches are used for land suitability assessment, i.e. qualitative and quantitative approaches. The qualitative approach is used to evaluate potential land on a large scale or as a preliminary tool for more intensive inquiry (Baja et al. 2002; Dent and Young 1981). The quantitative approach employs parametric procedures including more comprehensive land features, allowing for the performance of various statistical studies (Baja et al. 2002, 2001). In land suitability assessments, practical implementations of such spatial MCDM techniques have grown more common (Arciniegas et al. 2011; Chang et al. 2008; Chen et al. 2010; Greene et al. 2010; Kordi and Brandt 2012). One such quantitative method which can be used for suitability assessment under MCDM problem is the Analytic Hierarchy Process (AHP). The parameters for determining their relative weight are ranked on Saaty 9 points scale. It’s used to create ratio scales from paired comparisons, both discrete and continuous. Actual measurements or a fundamental scale that depicts the relative strength of preferences and sentiments can be used to make these comparisons. Multi-criteria decision-making, planning and resource allocation, and dispute resolution are among the areas where it has found the most use (Saaty 1977). The possible application of the AHP in public choice decisions involving complex, controversial and conflictual site selection processes was also highlighted by Banai-Kashani (1989). Popular AHP-based land use suitability analyses have been excoriated for requiring precise numerical values to express the strength of stakeholders’ preferences (Chang et al. 2008; Deng 1999; Kordi and Brandt 2012; Mikhailov 2003; Mosadeghi et al. 2013; Wang and Chen 2008).

Due to rapid pace of population development, its expansion is not being confined to plain areas nowadays. Even hilly terrains are also expected to experience huge population pressure in upcoming days. Several hill area development programmes are being run by governments since fifth five-year plan (Planning Commission). These plans are focused on the holistic socioeconomic development of the region. Rehabilitation of the people is one major objective of the hill area development programme.

Being well stabled and having good residential conditions social status of the people of hilly regions will be enriched. They will be able to join themselves in the mainstream of the society. The temperament of ‘being separatist’ will be vanished. Eco-rehabilitation practices such that Eco-restoration, Eco-development and Eco-preservation could be accelerated in the region. On the basis of requirements under Gadgil formula, currently central and state governments spent 90% and 10% of total fund for selected hilly areas. The application of these machine learning tools along with geospatial technology could reduce the economic pressure on government directly, and also, on the successful implementation of hill area development it’ll boost the nation’s economy indirectly. On the successful implementation of the work, we will be able to understand the performance of AHP in the selection of suitable sites in hilly terrain. This case study could be used as lodestar to find suitable locations under similar conditions.

Natural system limits, compatibility with existing land uses, social infrastructure and the availability of community services play crucial role in focusing on hill area development and planning. A set of criteria based on the expert’s opinion and local norms is used here to determine suitable sites for built-up purposes. These criteria include slope, distance to road, land use, distance from developed land, distance from landslide, lithology, distance from drainage, lineament and aspect. These different criteria have shown varying impact on built-up location. Thus, we are not able to implement simple arithmetic weightage approach to tackle this problem. The bulk of past MCDM applications has been observed at the national or regional level, with the primary goal of employing MCDM to prioritize established management options or such type of planning problem where we are unable to deal with simple weightage approach (Ananda and Herath 2003, 2008; Bryan et al. 2009; Hajkowicz 2002, 2008; Hajkowicz and McDonald 2006; Kodikara 2008; Qureshi and Harrison 2003; Xevi and Khan 2005).

Thus, we are intended to integrate GIS with MCDM for site suitability analysis. GIS-based site suitability analysis is critical for identifying suitable sites within least time and efforts.

Uttarakhand (Earlier known as Uttaranchal) is the 27th state of India which is formed by the Uttar Pradesh Reorganization Act, 2000, and is a state targeted to develop under the aegis of Hill Area Development Programme. Almora is its one of the 13th districts. Hawalbagh is one out of 11th block in Almora district of Uttarakhand. According to district census handbook data of Almora, it was observed that population density of Hawalbagh Block has increased from 310 person/km^2 in 2001 to 334.6 person/km^2 in 2011. Due to the tremendous pace of population growth, the population is now expanding rapidly. Here we performed a case study to discuss the above-mentioned rehabilitation problem and applicability of AHP to identify suitable built-up site with special reference to Hawalbagh Block in the Uttarakhand District.

11.2 Overview of the Study Area

The study area, namely the Hawalbagh block, is one of the eleven development blocks of Uttarakhand's District Almora, which is located in India's Lesser Himalayan topography. It is situated between 29° 32′ 30″ N–29° 44′ 23″ N latitudes and 79° 31′ 11″–79° 43′ 50″ E longitudes (Fig. 11.1). It's expanded over an area of 267.53 km^2, which is approximately 8% of the area of district Almora. Land Use-Land Cover map depicts that the vegetation covered around 54.75% (146.49 km^2), agriculture land covered 31.69% (84.73 km^2), barren land covered around 11.65% (31.17 km^2), built-up land covered 1.01% (2.72 km^2) and water body covered around 0.90% (2.42 km^2) of the Hawalbagh block. The land use and land cover of the region are the results of natural and social-economic factors, and the way people use them over time and space (Rawat and Kumar 2015). The Hawalbagh block is composed of rocks from two tectonic units. First one is 'Almora Group' in the centre and southern section, and the second one is 'Damtha Group' in the northern part (Valdiya 1980). Being situated in lesser Himalaya, cool temperate climatic condition is enjoyed over the region. Its average elevation from the mean sea level is 1405 m. Average slope of Hawalbagh block is 13.48°. Its average annual temperature from 1981 to 2020 is 20.06 °C. Average annual relative humidity at 2 m from 1981 to 2020 is 51.85%. Average annual precipitation from 1981 to 2020 is observed as 3.07 mm/day. Average annual direction of wind blow from 1981 to 2020 is observed as 275.39° with annual average speed of 2.55 m/s. The Kosi River which is the master stream of this block flows from north to south and divides the block in two.

The Hawalbagh block is made up of 10 nyaya panchayats, 124 grampanchyats and 234 villages. Population of the Hawalbagh in 2011 is 72,613 out of which 35,323 are

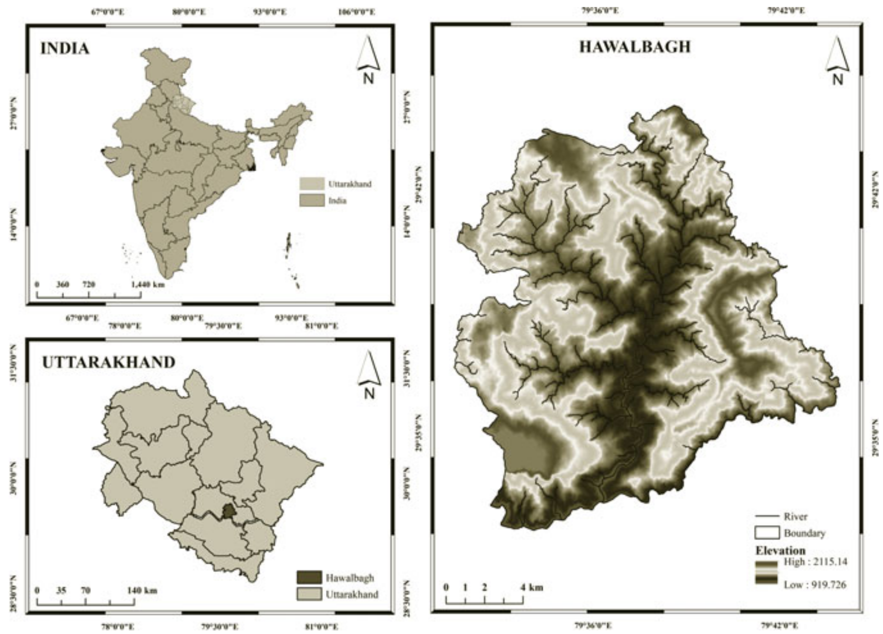


Fig. 11.1 Study area map

males and 37,290 are females. 67,447 people are living in rural areas while only 5166 people are living in urban areas. It takes into consideration intra-state differences in the level, potential and challenges of development in various regions. Keeping these all-stark realities in mind a comprehensive approach is applied to ensure harmonious and sustainable development of the region.

11.3 Database and Methodology

11.3.1 Database and Properties of Criterion

In hilly terrain several heterogeneous criteria such as slope, aspect, road proximity, land use, distance from developed land, landslide, lithology, drainage proximity, lineament and all play a role in site selection. We also have considered these criteria as essential criteria for site suitability analysis. Complete database is shown in Table 11.1.

Table 11.1 Data used and their sources

Data	Source	Used for
Slope & Aspect, 2020	CartoDEM	Slope map and aspect map
Lithology & Lineament, 2020	Lithological and Lineament Data from Bhukosh, GSI	Lithology map and lineament map
Land Use-Land Cover Map, 2020	Cartosat-1	Land use-land cover map
Road and Drainage Map, 2020	Digitized from Cartosat-1	Distance to road map and distance to drainage map
Landslide Map, 2020	Google Earth Pro	Distance to landslide map
Developed Land, 2020	LULC Map	Distance to developed map
Boundary, 2020	Hawalbagh block, Almora	Study area map

11.3.1.1 Slope and Aspect Map

Slope and Aspect map for the study region was derived from CartoDEM using ArcGIS 10.3. Figure 11.3a represents the slope map, and Fig. 11.3b represents the Aspect Map for Hawalbagh Block, Uttarakhand. Slope is the angular slope of terrain between the tops of hills and the bottoms of valleys, which is caused by a variety of factors such as geological structure. Mathematically slope is defined at the angle of inclination of the tangent drawn at any point of the curve. Because most development takes place on almost flat land, population density is observed to be higher than on steep slopes. As the slope increased, the likelihood of settlement reduced. Slope map was standardized by 'MSSmall Fuzzy membership function' (Fig. 11.2).

The horizontal direction in which a mountain slope faces is referred to as its aspect. In the northern hemisphere, in the middle of winter, north-facing slopes receive very little heat from the sun. South-facing hills, on the other hand, receive a lot more heat. As a result, south-facing slopes are generally hotter than north-facing ones. People in hilly locations choose to build their homes on sunny slopes. As a result, the value of southern-facing slopes is greater. East-facing slopes receive sunlight exclusively in the morning, when temperatures are cooler, but west-facing slopes receive sunlight in the afternoon, when temperatures are warmer. As a result, slopes facing east are colder than slopes facing west. Aspect Map was fuzzified by 'Gaussian Fuzzy membership function' with max at 180°.

11.3.1.2 Land Use-Land Cover Map

Land use-land cover map for the study region was prepared using Cartosat-1 data. Unsupervised classification was done on ArcGIS 10.3 to prepare land use-land cover (LULC) map. LULC map of Fig. 11.3c shows that forest and agriculture covered the maximum of area, i.e. 121.03 km^2. Figure 11.3d represents the LULC map for the study region. The land use-land cover map is a detailed representation of the

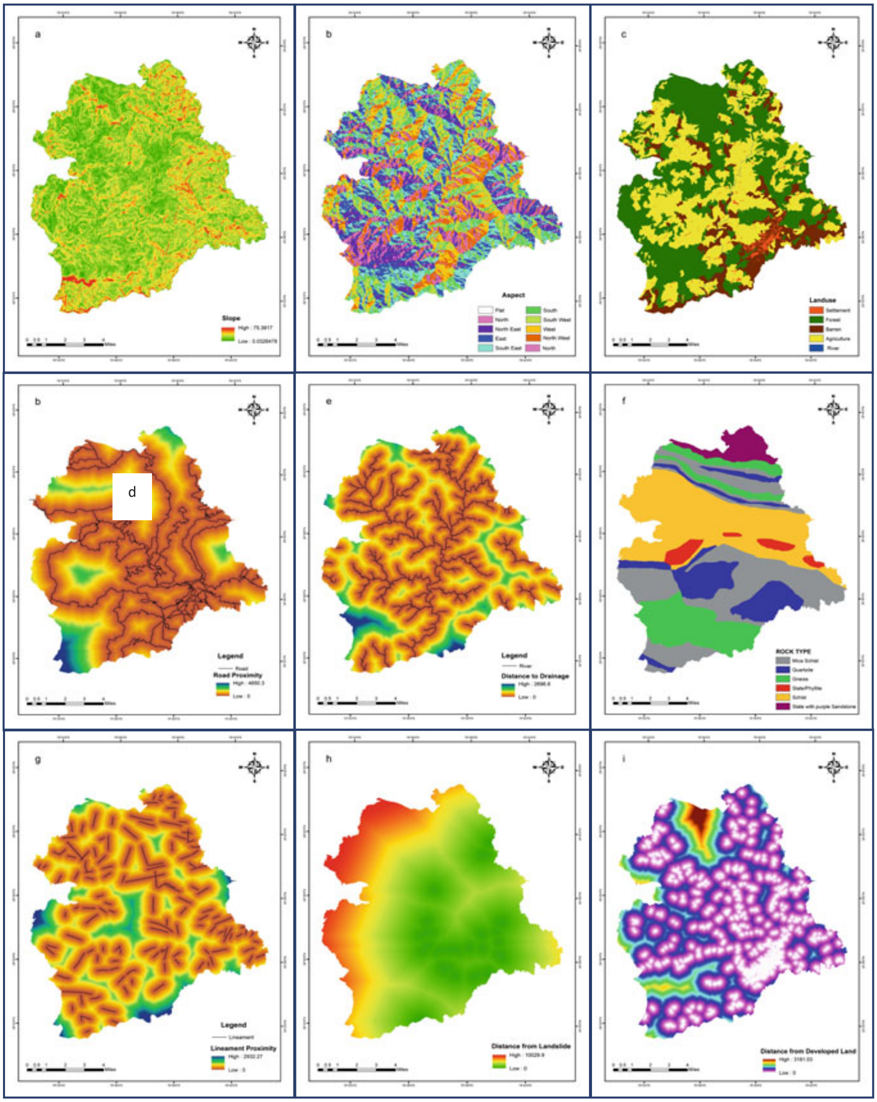


Fig. 11.2 **a** Slope map, **b** Aspect map, **c** Land use map, **d** Land proximity map, **e** Drainage proximity map, **f** Lithology map, **g** Lineament proximity map, **h** Landslide proximity map, **i** Developed land

classification of land use and land cover. A land use-land cover map is required for any development activity planning since it depicts the current state as well as the extent of future development. Built-up, barren, agricultural, river bed and forest have been classified on the LULC map for the study region. LULC map was first rescaled on the basis of scale value 0–4, where 0 means least suitable and 4 means very high

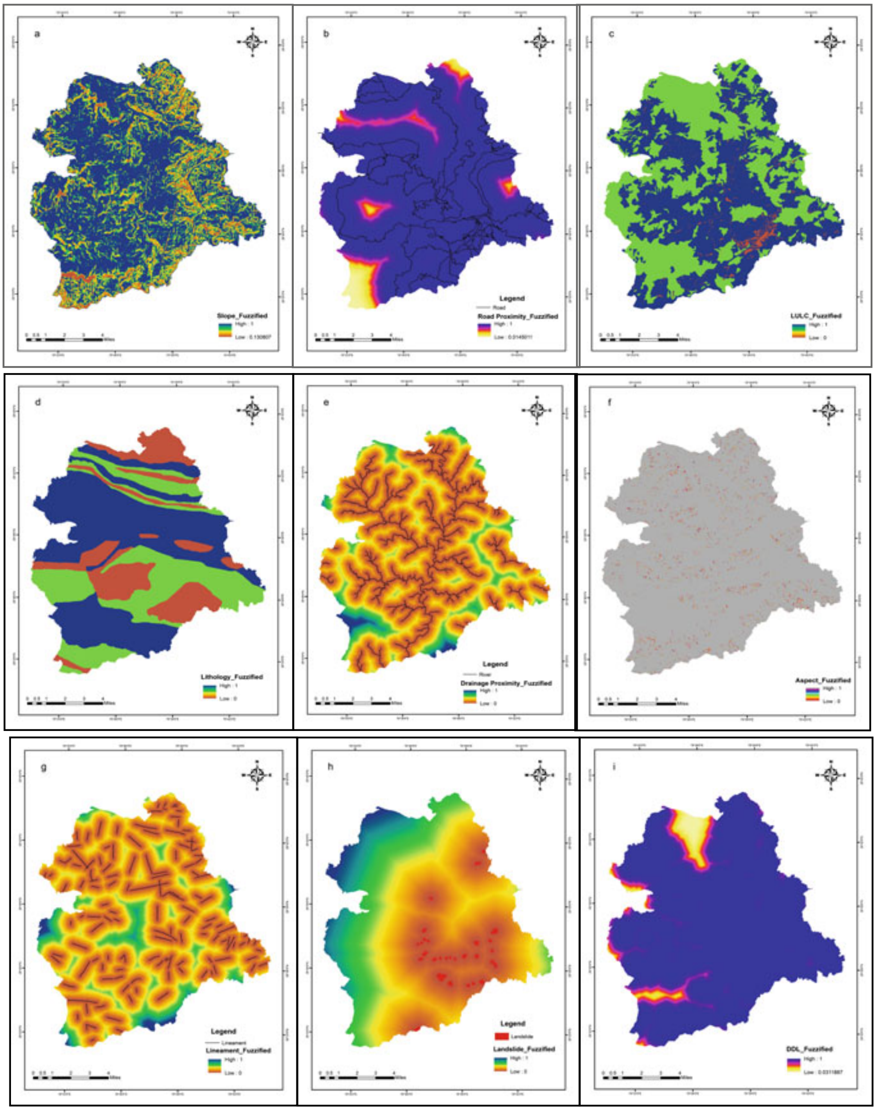


Fig. 11.3 **a** Standardized slope map, **b** Standardized distance to road (road proximity) map, **c** Rescaled and standardized LULC map, **d** Standardized lithology map, **e** Standardized distance to drainage (drainage proximity) map, **f** Standardized aspect map, **g** Standardized lineament map, **h** Standardized landslide map, **i** Standardized distance to developed land (settlement) map

suitable for built-up. After rescaling, 'linear Fuzzy membership function' was used to standardize the LULC criteria.

11.3.1.3 Road and Drainage Map

Road and drainage map was digitized using high-resolution CartoSat image using ArcGIS 10.3. After digitizing it distance to road and distance to drainage map was prepared. Distance to road and distance to drainage map is shown in Fig. 11.3d, e respectively.

Road accessibility is an important criterion for urban development because it connects towns. For the standardization of the Distance to road map 'small fuzzy membership function' is used having mid-value at 2000 m (2 km.)

In hilly places, proximity to drainage is significant because it causes landslides and flooding when runoff is high. According to many studies, the closeness of intensive gully erosion to drainage lines is an essential factor controlling the occurrence of landslides (Gokceoglu and Aksoy 1996; Pachauri et al. 1998). So, nobody prefers to reside near the rivers in hilly region. So, for standardizing the distance to drainage map 'linear fuzzy membership function' was used.

11.3.1.4 Lithology and Lineament Map

Lithology and lineament map was prepared using data obtained from Bhukosh portal of Geological Survey of India. Distance to lineament map was prepared using Arc GIS 10.3. Figure 11.3f represents the lithology map, and Fig. 11.3g represents the distance from lineament map.

One of the key causes of slope instability is lithology. Based on expert knowledge, the lithology map was initially rescaled into the range 0–4 where 0 means least suitable and 4 means very high suitable.

After rescaling, 'linear fuzzy membership function' was applied to standardize the lithology map.

Lineaments are linear or curvilinear features on the earth's surface that represent the weaker zone of bed rocks and are classified as secondary aquifers in hard rock locations. The intersection of lineaments is thought to be a favourable location for groundwater potential zones. Highly faulted zones have an unusually high occurrence of unstable slopes. Varnes (1984) determined that the degree of fracture and shearing plays a significant impact on slope stability. Thus, 'linear fuzzy membership function' was applied here to standardize the lineament map.

11.3.1.5 Distance Form Landslide and Developed Land (Settlement) Map

Distance maps were created in the same way as the previous distance maps. Landslide zone map was extracted using Google earth images and Arc GIS 10.3, and the map for developed land was created using the land use-land cover map. Figure 11.3h, i represent the Distance from Landslide and Developed Land, respectively.

Places affected by landslides are not appropriate for built-up development; however areas further away from the landslide spots can be regarded favourable for development. So, for the standardization of the landslide map 'linear fuzzy membership function' was used.

Everyone wishes to live in a society with surrounding people, infrastructure, facilities, accessibility and neighbourhoods, among other things. In other words, every developed land tends to be closer to non-developed land than it is to developed territory. To maintain the importance of this criteria 'small fuzzy membership function' with mid at 1600 m was used to standardize.

11.3.2 Criterion Standardization

Prior to associating the weights with each of the respective layers it's very necessary to standardize them so that all layers should bring up to a common level according to our requirement. Sets (or classes) with fuzzy boundaries are those in which the transition from membership to non-membership of a location in the set is gradual (Zadeh 1996). Fuzzy set membership is used to standardize criteria in multi-criteria evaluation. Selection of appropriate function to standardize the criterion is based on the understanding of the relationship between the criterion and the decision set, as well as how much information we have to infer fuzzy membership. The selection of appropriate membership function has been discussed earlier and tabulated here in Table 11.2. Each standardized criterion is shown in Fig. 11.3a–i.

11.3.3 Assigning Rank and Estimation of Criteria Weights

11.3.3.1 Analytical Hierarchy Process, AHP

AHP is one of the most well-known multi-criteria decision-making strategies (Saaty 1980). The classical AHP takes into account the decision-makers' specific judgements (Wang and Chen 2008). To construct a ratio matrix, this method uses pair-wise comparisons. It takes pair-wise comparisons as input and returns relative weights as output. The weights for each criterion are determined by normalizing the eigenvector associated with the maximum Eigen value of the reciprocal ratio matrix (Malczewski 1999). Pair-wise comparison matrix was developed on recommendation of experts

Table 11.2 Function used for standardization of criterion

Criteria	Membership function type
Slope	MSSmall
Distance to road	Small (decreasing) mid at 2000 m
Land use	Rescaled according to suitability and then linear
Lithology	Rescaled according to suitability and then linear
Distance to drainage	Linear
Aspect	Gaussian, max at 180°
Distance to lineament	Linear
Distance to landslide	Linear
Distance from developed/settlement land	Small (decreasing) mid at 1600 m

view on the basis of Saaty Scale (Saaty 1980). Pair-wise comparison matrix is shown in Table 11.3. Finally, arithmetic mean of their responses was used for the criteria weight calculation. Saaty Scale is presented in Table 11.4.

After creation of suitability matrix and criteria weight computation **Consistency Ratio (CR)** was calculated according to Eq. (11.1) to assess the consistency of the assessments in comparison with huge samples of completely random judgements.

$$\mathrm{CR} = \frac{\mathrm{CI}}{\mathrm{RI}\ (n)} \tag{11.1}$$

where CI is the Consistency Index, given by

$$\mathrm{CI} = \frac{(\lambda_{\max} - n)}{(n - 1)} \tag{11.2}$$

where $\lambda_{\max}$ is the principle eigenvalue of the suitability matrix and n is the number of criterion. *Random Consistency Index* is represented by RI(n). For 9 ordered matrix, its value will be 1.45 (Saaty 1980). The conventional CR threshold value of ≤ 0.10 was used in this study, which means that if CR $\leq$ 0.10, the pairwise comparison matrix is consistent and the weight values are valid and can be used (Ananda and Herath 2008; Kordi and Brandt 2012; Saaty and Vargas 2007). In this study, mean consistency ratio (CR) for all the nine categories is found less than 0.078. Final AHP weight for each criterion is presented in Table 11.5.

Table 11.3 Pairwise confusion matrix for AHP

	Slope	Road proximity	Land use	Lithology	Drainage proximity	Aspect	Lineament	Landslide location	Dis from developed land
Slope	1.000	2.333	4.000	5.667	6.000	8.333	6.333	4.333	3.000
Road proximity	0.429	1.000	3.667	4.000	5.333	7.667	6.000	3.333	1.667
Land use	0.250	0.273	1.000	3.667	4.333	6.667	6.000	3.333	1.667
Lithology	0.176	0.250	0.273	1.000	2.333	4.333	3.000	2.000	4.000
Drainage proximity	0.167	0.188	0.231	0.429	1.000	3.000	1.333	3.333	5.000
Aspect	0.120	0.130	0.150	0.231	0.333	1.000	2.667	5.333	6.667
Lineament	0.158	0.167	0.167	0.333	0.750	0.375	1.000	4.333	5.000
Landslide location	0.231	0.300	0.300	0.500	0.300	0.188	0.231	1.000	2.000
Dis from developed land	0.333	0.600	0.600	0.250	0.200	0.150	0.200	0.500	1.000

Table 11.4 Saaty 9-point scale for pairwise comparison matrix

Intensity of importance	Definition
1	Equal importance
3	Moderate importance
5	Strong importance
7	Very strong importance
9	Absolute/extreme importance
2,4,6,8	Intermediate values between above scale values
Reciprocals of the above	If any no *j* is assigned with any intensity of importance with respect to *I* then, intensity of importance of *I* with respect to *j* will be reciprocal of it

Table 11.5 AHP weight for each suitability criteria

Criteria	AHP weight
Slope	0.286724014
Distance to road	0.204901575
Land use	0.14282177
Lithology	0.085104919
Distance to drainage	0.069303756
Aspect	0.073234909
Distance to lineament	0.05928756
Distance to landslide	0.036661535
Distance from Developed/settlement land	0.041959963

11.3.4 Built-Up Suitability

Final suitability map was obtained by the overlays of each raster criterion, multiplying each with their respective AHP weight and summing them together. Resultant Map was classified into 5 categories. These 5 categories are Very High, High, Medium, Low and Very Low. Figure 11.4 represents the final suitability for housing development using Multi-Criteria Evaluation Technique, AHP.

11.4 Results

11.4.1 Criteria Influence Analysis for AHP

The criteria ranking order assigned according to the relative relevance weight of each criterion is a vital input to the spatial MCDM model. The study's findings

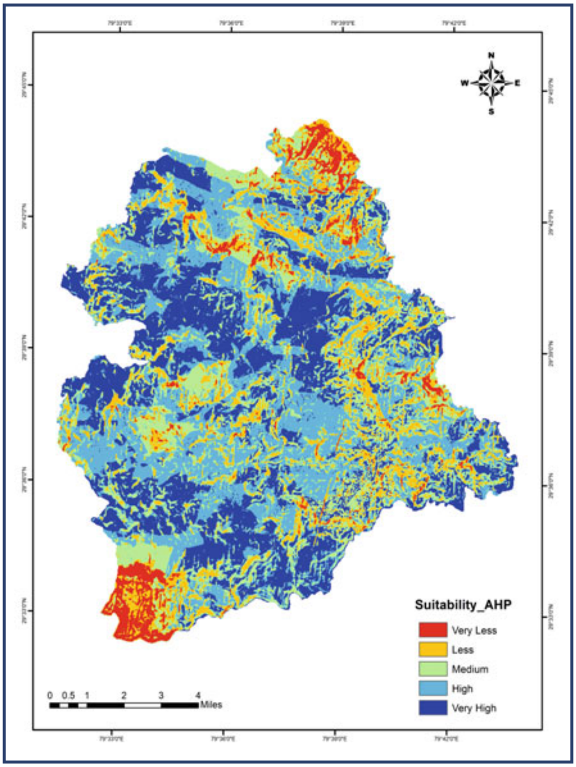


Fig. 11.4 Final suitability map by AHP

revealed discrepancies in the final weights obtained from each MCDM algorithm. The percentage influence of each criterion on the final outcomes of the model was determined according to Eq. (11.3) to see whether these divergences generated differences in the influence of each criterion on the final outcomes of the model.

$$\text{Percentage Influence} = \frac{W_i}{\sum_{i=1}^{n} W_i} \times 100 \tag{11.3}$$

where W_i is the weight of each criterion and n is the number of criteria.

Table 11.6 represents the criteria influence for AHP.

On analysis, we find that AHP has given highest weight to slope and lowest weight to distance from landslide for housing suitability. It has given 28.67% influence to slope meanwhile distance to landslide influenced only 3.66% for housing suitability.

Graph shown in Fig. 11.5 represents the graphical interpretation of the percentage influence of both MCDM techniques.

Table 11.6 Criteria ranking and percentage influence for AHP

Criteria	AHP	
	Percentage influence	Criteria ranking
Slope	28.67	1
Distance to road	20.49	2
Land use	14.28	3
Lithology	8.51	4
Distance to drainage	6.93	6
Aspect	7.32	5
Distance to lineament	5.92	7
Distance to landslide	3.66	9
Distance from developed/settlement land	4.19	8

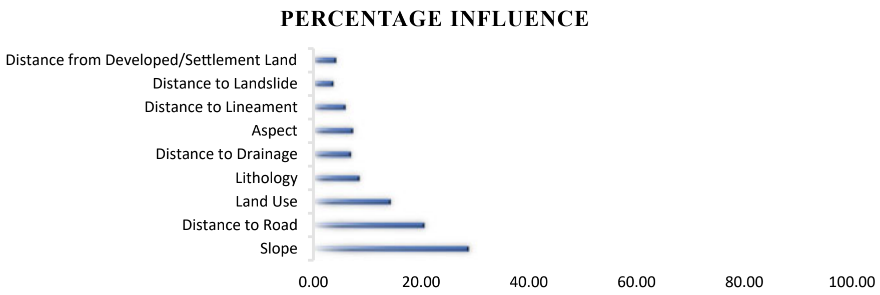


Fig. 11.5 Percentage influence of criteria

11.4.2 Suitability Area Analysis for AHP

Suitability maps were classified into 5 categories as mentioned earlier, viz. very high, high, medium, low and very low. Analysis reveals the fact that the MCDM techniques of AHP provided 73.09 km^2 (27.37% of the total area) for very high suitability, 84.12 km^2 (31.5% of the total area) for high suitability, 57.80 km^2 (21.65% of the total area) as medium suitable, 40.44 km^2 (15.15% of the total area) as less suitable and 11.54 km^2 (4.32% of the total area) as very less suitable. Table 11.7 and Fig. 11.6 represent the comparison of suitability area for both of the MCDM techniques.

Table 11.7 Area-wise suitability for AHP technique

Class	AHP	
	Area	Percentage
Very less	11.54	4.32
Less	40.44	15.15
Medium	57.80	21.65
High	84.12	31.51
Very high	73.09	27.37

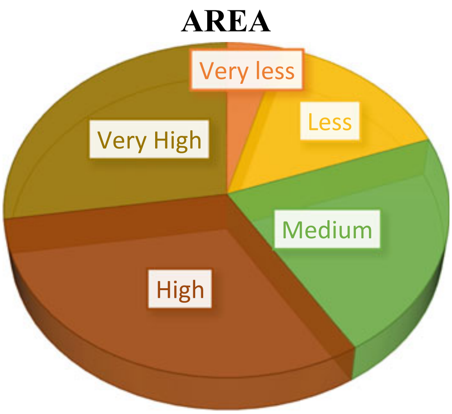


Fig. 11.6 Pie chart of the suitability classes of the region

11.4.3 Validation of the Result

Result of the AHP suitability was validated using ROC-AOC characteristics. ROC curve shows that AHP provided suitability area with 92.1% of accuracy level. Figure 11.7 shown is representing the ROC curve.

11.5 Discussion and Conclusion

The rapidly growing population of the hilly region, as well as a number of man-made and natural hazards, pose serious concerns. These issues could only be managed and reduced to a point. To address these issues, a region that can potentially be used for the development of human residential sites must be defined.

The majority of human settlements in hilly areas are influenced by nature, rather than socioeconomic considerations. Only natural aspects are investigated in this study to determine potential suitability. The use of built-up suitability mapping is crucial for locating unfavourable and favourable suitability locations. The current study mapped built-up suitability in Uttarakhand's Almora district's Hawalbagh Block using one

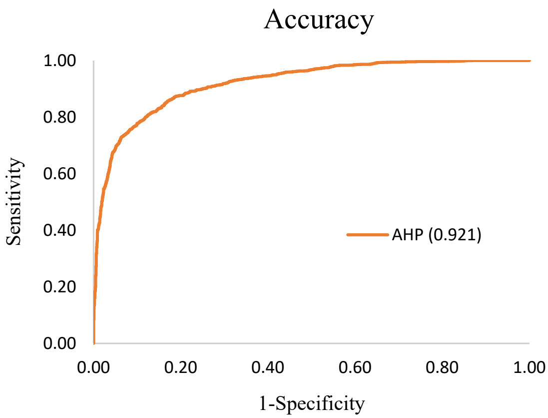


Fig. 11.7 ROC curve for the accuracy assessment

MCDM strategy, AHP. The suitability of the land for a specific use, as well as the values and interests of stakeholders in a region, determines land suitability (Steiner et al. 1991; Bojórquez-Tapia et al. 2019). The goal of land suitability analysis is to "assist decision-makers in identifying the most acceptable sites or patterns of locations for meeting the needs of the relevant stakeholders" (Bojórquez-Tapia et al. 2019). The identification of a criterion for built-up land suitability was a significant component of the research. Nine condition criteria were chosen based on literature reviews and expert comments (Chen et al. 2010; Ananda and Herath 2008, etc.).

The slope conditioning factor was identified as the most significant conditioning factor for built-up suitability mapping, while distance to drainage, lineament, aspect and lithology were the least effective conditioning factors. The results show that the middle to centre western sector of the Hawalbagh is very stable and suitable due to the lower slope, the availability of stable ground, the proximity to the road and the presence of cultivated fields. Study revealed that the extreme north, extreme south and eastern margins of the Hawalbagh Block are unsuitable for development in the foreseeable future.

ROC characteristics were used to validate the performance of AHP. The findings of the current study may be useful to a variety of policymakers, government stakeholders and researchers in choosing the best place for hillside development and planning.

This work employed the MCDM technique, AHP, to map the built-up suitability of the Hawalbagh Block in the Almora District of Uttarakhand, India. The models were built using a geographical database that included nine conditioning factors. Based on expert judgement and current literature, weightage was computed for all nine conditioning factors.

Finally, result was validated through true-positive and true-negative point using ROC curve. Validation results revealed that all models performed well and had accuracy levels above 90%. The suitability maps produced by this study may be useful for hillside housing development and planning of locations that are expected to be particularly suitable.

References

Ananda J, Herath G (2003) The use of analytic hierarchy process to incorporate stakeholder preferences into regional forest planning. Forest Policy Econ 5(1):13–26

Ananda J, Herath G (2008) Multi-attribute preference modelling and regional land-use planning. Ecol Econ 65(2):325–335

Ananda J, Herath G (2009) A critical review of multi-criteria decision-making methods with special reference to forest management and planning. Ecol Econ 68(10):2535–2548

Arciniegas G, Janssen R, Omtzigt N (2011) Map-based multicriteria analysis to support interactive land use allocation. Int J Geogr Inf Sci 25(12):1931–1947

Baja S, Chapman MD, Dragonvich D (2001) Fuzzy modeling of environmental suitability index for rural land use systems: an assessment using a GIS. http://www.geocomputation.org/2001. 16 May 2013

Baja S, Chapman MD, Dragonvich D (2002) A conceptual model for defining and assessing land management units using a fuzzy modeling approach in GIS environment. Environ Manage 29:647–661

Banai-Kashani R (1989) A new method for site suitability analysis: the analytic hierarchy process. Environ Manage 13(6):685–693

Bojórquez-Tapia LA, Díaz-Mondragon S, Gomez-Priego P (2019) GIS-approach for land suitability assessment in developing countries: a case study of forest development project in Mexico. In: Spatial multicriteria decision making and analysis. Routledge, pp 335–352

Bryan BA, Grandgirard A, Ward JR (2009) Managing natural capital and ecosystemservices: identifying strategic regional priorities in the South Australian Murray-Darling basin. CSIRO Water for a Healthy Country National Research Flagship. Available from: http://www.clw.csiro.au/publications/waterforahealthycountry/2009/wfhc-managing-naturalcapital-mdb.pdf.

Chang NB, Parvathinathan G, Breeden JB (2008) Combining GIS with fuzzy multicriteria decision-making for landfill siting in a fast-growing urban region. J Environ Manage 87(1):139–153

Chen Y, Yu J, Khan S (2010) Spatial sensitivity analysis of multi-criteria weights in GIS-based land suitability evaluation. Environ Model Softw 25(12):1582–1591

Collins MG, Steiner FR, Rushman MJ (2001) Land-use suitability analysis in the United States: historical development and promising technological achievements. Environ Manage 28(5):611–621

Deng H (1999) Multicriteria analysis with fuzzy pairwise comparison. Int J Approximate Reasoning 21(3):215–231

Dent D, Young A (1981) Soil survey and land evaluation. George Allen and Uniwin, Boston

Dunning DJ, Ross QE, Merkhofer MW (2000) Multiattribute utility analysis for addressing Section 316 (b) of the Clean Water Act. Environ Sci Policy 3:7–14. Ecosystem services: identifying strategic regional priorities in the South Australian Murray-Darling Basin. CSIRO Water for a Healthy Country National Research Flagship. Available from: www.clw.csiro.au/publications/waterforahealthycountry/2009/wfhc-managing-natural-capital-mdb.pdf

FAO (1976) A framework for land evaluation. Food and Agriculture Organization of the United Nations, Soils Bulletin No. 32. FAO, Rome

Gökceoglu C, Aksoy H (1996) Landslide susceptibility mapping of the slopes in the residual soils of the Mengen region (Turkey) by deterministic stability analyses and image processing techniques. Engg Geol 44(1–4):147–161

Greene R, Luther JE, Devillers R, Eddy B (2010) An approach to GIS-based multiple criteria decision analysis that integrates exploration and evaluation phases: case study in a forest-dominated landscape. For Ecol Manage 260(12):2102–2114

Hajkowicz S (2002) Regional priority setting in Queensland: a multi-criteria evaluation framework. CSIRO Land and Water. Available from: http://live.greeningaustralia.org.au/nativevegetation/pages/pdf/Authors%20H/3_Hajkowicz.pdf

Hajkowicz SA (2008) Supporting multi-stakeholder environmental decisions. J Environ Manage 88(4):607–614

Hajkowicz S, Collins K (2007) A review of multiple criteria analysis for water resource planning and management. Water Resour Manage 21(9):1553–1566
Hajkowicz S, Mcdonald G (2006) The assets, threats and solvability (ATS) model for setting environmental priorities. J Environ Planning Policy Manage 8(01):87–102
Herath G, Prato T (eds) (2006) Using multi-criteria decision analysis in natural resource management. Ashgate Publishing, Ltd
Kalogirou S (2002) Expert systems and GIS: an application of land suitability evaluation. Comput Environ Urban Syst 26(2–3):89–112
Keshavarzi A, Heidari A (2010) Land suitability evaluation using fuzzy continuous classification (a case study: Ziaran region). Mod Appl Sci 4(7):72–82
Kodikara PN (2008) Multi-objective optimal operation of urban water supply systems. Thesis (PhD), Victoria University
Kordi M, Brandt SA (2012) Effects of increasing fuzziness on analytic hierarchy process for spatial multicriteria decision analysis. Comput Environ Urban Syst 36(1):43–53
Malczewski J (1999) GIS and multicriteria decision analysis. Wiley
Malczewski J (2004) GIS-based land-use suitability analysis: a critical overview. Prog Plan 62(1):3–65
Mikhailov L (2003) Deriving priorities from fuzzy pairwise comparison judgements. Fuzzy Sets Syst 134(3):365–385
Mosadeghi R, Tomlinson R, Mirfenderesk H, Warnken J (2009) Coastal management issues in Queensland and application of the multi-criteria decision-making techniques. J Coastal Res 1252–1256
Mosadeghi R, Warnken J, Tomlinson R, Mirfenderesk H (2013) Uncertainty analysis in the application of multi-criteria decision-making methods in Australian strategic environmental decisions. J Environ Planning Manage 56(8):1097–1124
Qureshi ME, Harrison SR (2003) Application of the analytic hierarchy process to riparian revegetation policy options. Small-Scale Forest Econ Manage Policy 2(3):441–458
Rawat JS, Kumar M (2015) Monitoring land use/cover change using remote sensing and GIS techniques: a case study of Hawalbagh block, district Almora, Uttarakhand, India. Egypt J Remote Sens Space Sci 18(1):77–84
Saaty TL (1980) The analytic hierarchy process. McGraw-Hill, New York
Saaty TL, Vargas LG (2007) Dispersion of group judgments. Math Comput Model 46(7–8):918–925
Steiner F (1991) Landscape planning: a method applied to a growth management example. Environ Manage 15(4):519–529
Valdiya KS, KS V. (1980). Stratigraphic scheme of the sedimentary units of the Kumaun Lesser Himalaya
Varnes DJ (1984) Landslide hazard zonation: a review of principles and practice (No 3)
Wang TC, Chen YH (2008) Applying fuzzy linguistic preference relations to the improvement of consistency of fuzzy AHP. Inf Sci 178(19):3755–3765
Xevi E, Khan S (2005) A multi-objective optimisation approach to water management. J Environ Manage 77(4):269–277
Zadeh LA, Klir GJ, Yuan B (1996) Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers, vol 6. World Scientific
Zopounidis C, Doumpos M (2002) Multicriteria classification and sorting methods: a literature review. Eur J Oper Res 138(2):229–246
Zopounidis C, Pardalos PM (eds) (2010) Handbook of multicriteria analysis, vol 103. Springer Science & Business Media

Chapter 12
Case Study 4: Urban Green Space Analysis and Potential Site Selection for Green Space Expansion in NCT Delhi

Abstract Urban green space, a component of green infrastructure, refers to the areas covered with green vegetation including municipal parks, public playgrounds, etc., that are aimed to serve as the sites of recreation and environment enhancers in the city and planning regions. In this study, an urban green space analysis of NCT Delhi has been undertaken to understand its availability and distribution. Using the Sentinel-2A imagery, the existing green spaces of NCT Delhi were identified using the image classification. Vegetation, low vegetation, parks and golf course areas of NCT Delhi were extracted as urban green spaces. Per capita green space of NCT Delhi as a whole and for each district separately was quantified and mapped. According to World Health Organisation (WHO), a minimum of 9.5 m^2 of green space is required for each person for healthy living in the urban environment. NCT Delhi has per capita green space of 15.93 m^2/person, which is greater than the threshold. However, four districts of the study area, i.e. Central Delhi (5.23 m^2/person), Northeast Delhi (1.74 m^2/person), East Delhi (5.21 m^2/person) and West Delhi (5.84 m^2/person) showcase acute shortage of green spaces per capita, necessitating the expansion of urban green spaces. Potential site selection for the expansion of the urban green space was carried out by buffering existing green spaces, water bodies and roads as central elements. The buffered areas were assigned weights and combined to generate the suitability map for the development of urban green spaces. The suitability map yields that NCT Delhi has a 47.92 km^2 area that is highly suitable for the expansion of green spaces and another 320.48 km^2 area that is moderately suitable for green space expansion.

Keywords Urban green space · Image classification · Per capita urban green space analysis · Potential mapping

12.1 Introduction

Green spaces are the land use and land cover areas that are occupied by natural or anthropogenic vegetation features. The urban green spaces relate specifically to those sections of urban regions that have been designed to give recreation and aesthetic benefits to the urban population. These include municipal parks, public playgrounds,

M. Kumar et al., *Geographic Information Systems in Urban Planning and Management*, Advances in Geographical and Environmental Sciences,
https://doi.org/10.1007/978-981-19-7855-5_12

etc. Bayram and Gökyer (2012) define Urban Green Spaces as urban areas that are covered with natural or man-made vegetation in the city and planning regions. Urban green spaces are the components of 'green infrastructure'. It is an essential component of a city's public spaces and services and potentially the health-promoting space for all urban residents (World Health Organization 2017).

Urban green spaces are vital for mitigating the negative repercussions of urbanization (Ragab 2014). After the industrial revolution, the majority of population expansion occurred in urban areas. Over 50% of the world population currently resides in urban regions, and recent decades have witnessed tremendous urban expansion (Wu 2014). By 2050, the worldwide urban population is anticipated to reach 6.3 billion. This urban growth has increased the burden on natural resources and the environment, as the quantity of land utilized for infrastructure building has risen at the expense of natural green areas. It has increased landscape fragmentation, biodiversity loss, the formation of urban heat islands, greenhouse gas emissions, susceptibility to climate change events and the destruction of important or delicate ecosystems (Norton et al. 2015). Continuing urbanization necessitates the provision of suitable ecosystem services and expanding the extent and diversity of urban green spaces, to emerge as a significant means of enhancing environmental quality in cities and ascertaining sustainable development. Regarded as the 'green lung of the city', they typically perform important functions such as absorbing rainwater and pollutants, mitigating urban heat, preventing waterlogging and flooding, enhancing air quality and providing socioeconomic benefits (Ragab 2014). Researchers have demonstrated that improved access to UGS can also support mental health among the urban population.

As urbanization continues to accelerate and a higher part of the human population migrates to cities, the demand for urban green spaces is becoming increasingly vital. World Health Organisation (WHO) recommends a minimum of 9.5 m^2 of green space per capita for healthy living. This requires quantifying the green space area of an urban region and estimating the green space area of an urban region based on population. Consequently, the purpose of this study is to determine the overall and district-specific green space availability in NCT Delhi, India, as well as the green space deficits throughout the territory.

12.2 Advantages of the Urban Green Spaces (UGS)

Urban Green Spaces play a key role in urban areas due to their environmental, aesthetic, social and economic benefits to the health and well-being of inhabitants (Senanayake et al. 2013). Before proceeding with the case study, it is pertinent to discuss the advantages shortly. The benefits of UGS can broadly be discussed under two major themes.

Environmental Benefits: Urban green spaces have a substantial impact on the urban ecological setting of the relevant city; yet, the consequences are not restricted to the specific location exclusively. They affect the entire planet. Rakhshandehroo et al. (2017) highlight the environmental advantages of the urban green spaces under six major themes including nature conservation, biodiversity and wildlife, urban climate, air quality (reduced air pollution, carbon sequestration), noise reduction and cleaning up contaminants. The paper stresses that the interaction between cities and the environment is a core problem for environmental preservation as urban expansion continues to gain pace. The ecological footprint of cities influences ecosystems from a regional to a global scale because the ecological functions inside green spaces involve the overall system of plants, animals, soil and human activities. Also, there is a significant understanding of the necessity of installing urban green spaces as an innovative strategy to maintain and promote biodiversity as carefully designed urban open green spaces maintain ecosystems and preserve biodiversity and hence form particularly crucial hot spots for biodiversity. They can be 'wildlife corridors' or operate as 'urban forests' and show high connectivity (Byrne and Sipe 2010).

The green spaces cool the environment by evaporative cooling that uses energy and shading effects. It regulates the interception of solar radiation and its optical and thermal properties. Urban green spaces are of special importance since they give shade and cooling, to alleviate the urban heat island effects. They also significantly deal with the matter of air pollution. Many studies have demonstrated the beneficial influence of urban area vegetation in urban settings on ambient air quality as it helps remove air pollution and sequester carbon. It has also been established that soft walls offered by green spaces not only form places and provide aesthetic borders, but also have an essential influence on noise reduction. The green medians and vegetative buffers along the express lanes, ring roads and airport express routes, generate not only noise buffers but are also places of beautiful sight and buffers for air pollution.

Human Benefits: Interaction with green space has several mental, physical and socioeconomic advantages for people. Parks, garden areas, street trees and green traffic islands provide more than a nice view, successfully lowering our everyday stress by evoking a sense of calm. Studies have demonstrated that exposure to nature landscapes reduces stress in individuals. Consequently, green spaces also decrease aggressiveness and violence. Exposure to green environments is also associated with enhanced attention and focused concentration. This is due to the fact that focusing on green, tranquil surroundings allows involuntary attention (an effortless act) to take over and recharge the human brain, allowing voluntary attention (an effortful act) to rest. The physical advantages of green environments include improved health, quicker recovery, etc. Crime reduction, increased workplace productivity, safer driving, economic stimulation and positive impacts on children are additional socioeconomic benefits of green areas. In addition to generating economic advantages such as rising property values and stimulating tourism, green spaces also provide social benefits such as improved social ties and a healthier lifestyle (Senanayake et al. 2013). It also gives city dwellers opportunities for exercise, sociability, engagement with nature and access to areas with a rich cultural past (Crompton 2005).

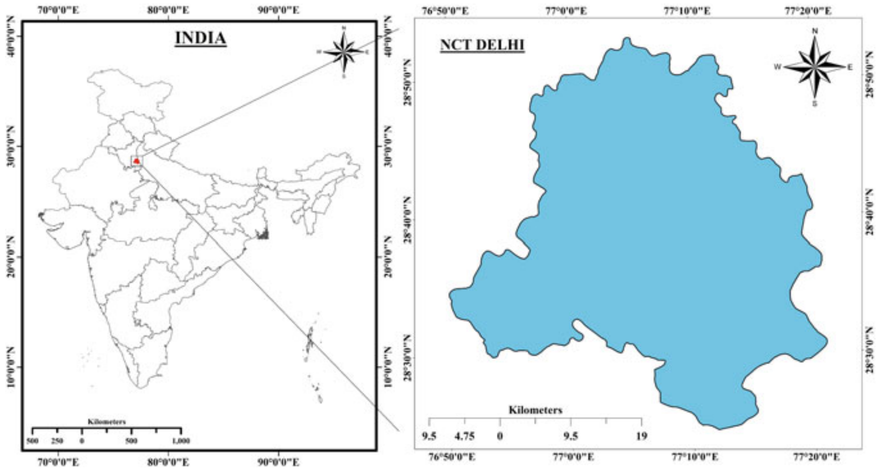


Fig. 12.1 Study area

12.3 Overview of the Study Area

NCT Delhi is bounded by latitudes 28° 24′ 24″ N and 28° 53′ 0″ N and longitudes 76° 49′ 59″ E and 77° 21′ 3″ E. It is the capital of India, located in the heart of the Indian subcontinent (Fig. 12.1). The region is topographically split into the Delhi Ridge, the Yamuna Flood Plain and the Plain. According to the 2011 Census. There are nine districts in the study region including Central Delhi, New Delhi, North Delhi, Northeast Delhi, East Delhi, South Delhi, Southwest Delhi, West Delhi and Northwest Delhi. With an area of 445.38 km^2, Northwest Delhi is the largest district and Central Delhi (15.10 km^2 area) is the smallest. Delhi has an overall built-up proportion of 33.54%. The districts with greater built-up generally have a lower occurrence of the urban green spaces. As per this axiom, Northeast Delhi has the highest urban built-up proportion 69.54% while Southwest Delhi has the lowest built-up proportion of 25.77%. Like other urban areas, NCT Delhi experiences the negative effects of concretization ranging from urban heat, waterlogging, etc., to human health issues including physical and mental problems.

12.4 Methodology

In this research work, as we aim to undertake the analysis and identification of the potential areas of the urban green space expansion in the region of NCT Delhi, the overall methodology can be segregated into the following sections.

12.4.1 Mapping of the Existing Urban Green Space

There are numerous methods of identifying the existing green space in a particular region. These include field survey-based mapping of the urban green spaces using conventional techniques and advanced remote sensing data-based methods like NDVI Thresholding and Extraction from Classified Images. In this study, we adopted the latter for it offers the advantage of comprehensive mapping of urban green space over an extensive region. To be specific, the green space mapping was approached using the method of extracting the urban green spaces from the classified satellite images. It is explained in Fig. 12.2.

To elaborate on this, the Sentinel-2A imagery specific to the NCT Delhi was downloaded from the Copernicus Open Access Hub of the European Space Agency (ESA). The image was processed at the characteristic resolution of 10 m. The bands 2 (B), 3 (G), 4 (R) and 8 (NIR) were *layer-stacked,* and the composite band image was subject to various enhancement operations, followed by *subsetting* the image according to the administrative boundary of NCT Delhi, India. The image was subject to the image classification procedure. Unsupervised classification of the image into 100 classes using *K*-means was used. These 100 classes were then narrowed down to 20 major classes of land use and land cover in the region of NCT Delhi. A significant amount of the pixels misclassified in the process of unsupervised classification were *recoded* to refine the quality of the classified image. The reclassified image was

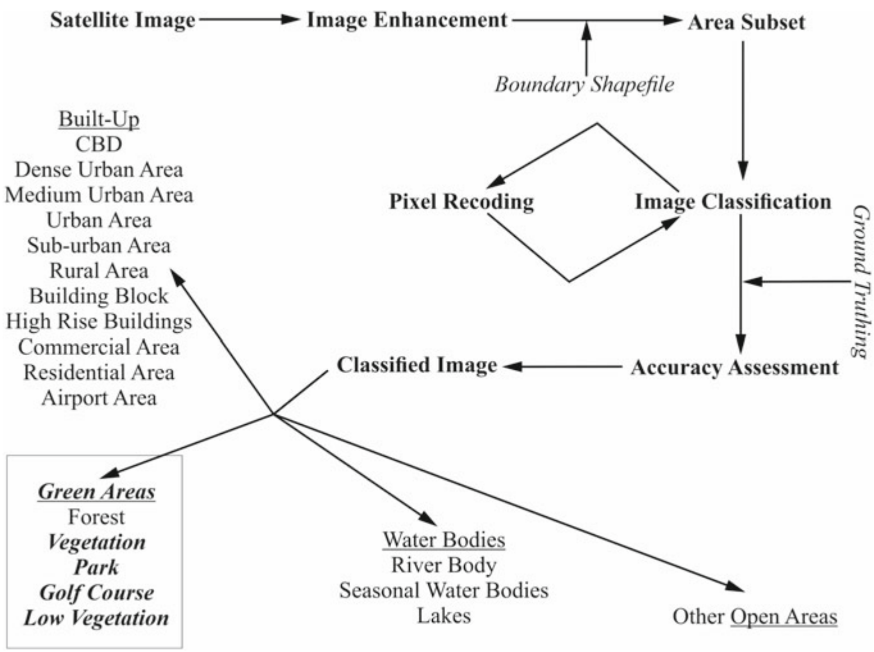


Fig. 12.2 Methodology: extraction of green spaces

validated for classification accuracy using the *ground-truthing* method. From the classified image, suitable classes of LU/LC were then identified as green spaces.

12.4.2 Urban Green Space Analysis

Urban green space analysis can be directed towards a broad variety of aims in research. However, in this study, the urban green spaces were assessed in terms of their availability to the population. The per capita urban green space is defined as the area of urban green space available to each inhabitant in the region. Firstly, the study of green space is conducted for the territory of NCT Delhi as a whole. In this regard, the following equation is applied:

$$\text{Per Capita Green Space} = \frac{\text{Total Urban Green Space Area}}{\text{Total Population}} \tag{12.1}$$

However, Eq. (12.1) produces a gross number for the amount of urban green space available relative to the population and does not show the disparate distribution. To refine this metric, the study was conducted at the level of districts of NCT Delhi, and Eq. (12.1) was adjusted as follows:

$$\text{Per Capita Green Space of each District} = \frac{\text{Urban Green Space Area of District}}{\text{Population of the District}} \tag{12.2}$$

In this study, the area of urban green spaces was mined from the previously mapped green spaces and the population data of the 2011 Census of India was used.

12.4.3 Potential Site Selection for Expansion of Urban Green Space

There are two aspects to the expansion of urban green space. The green areas may be increased in either the horizontal or vertical directions. However, the objectives of this study are confined to the horizontal expansion of urban green areas.

As part of a site suitability assessment, maps are made that show which areas are best for a certain activity (Hopkins 1977). To find the places where urban green spaces could be built, central elements that could be built around were found. Rivers, lakes and other bodies of water, as well as roads, were digitized. Aside from these, urban green spaces that were already there were vectorized. In ArcGIS 10.7@ESRI, buffers were set up around these features. Different features that were used as central elements were buffered in different ways, and different categories of buffers were given different weights on a scale from 1 to 5, with 1 being the most suitable and 5 being the least suitable. The vector layers were rasterized before combining. The cell statistics (overlay statistic = SUM) were then used to combine these raster layers to make a raster layer that shows different zones with different potential for developing urban green spaces.

12.5 Results and Discussion

The band composite of Sentinel-2A image, acquired from ESA, was classified into 20 LULC classes including CBD, Dense Urban Area, Medium Urban Area, Urban Area, Suburban Area, Rural Area, Building Block, High Rise Buildings, Residential Block, Commercial Area, Airport Area, River Body, Seasonal Water Bodies, Lakes, Forest, Vegetation, Park, Golf Course, Low Vegetation and Other Open Spaces. These can be grouped as built-up (including CBD, Dense Urban Area, Medium Urban Area, Urban Area, Suburban Area, Rural Area, Building Block, High Rise Buildings, Residential Block, Commercial Area, Airport Area), Water Bodies (including River Body, Seasonal Water Bodies, Lakes) Green Areas (Forest, Vegetation, Park, Golf Course, Low Vegetation) and Other Open Spaces. These areas are depicted in Fig. 12.3.

From the classified raster image of NCT Delhi, the urban green space was identified as the areas that comprised greenery in the urban settings, excluding the natural forests. As such Vegetation, Park, Golf Course and Low Vegetation classes of the LULC were selected and extracted as the urban green spaces and these are shown in Fig. 12.4**.** As per the map, NCT Delhi has 267.38 km^2 area as the urban green spaces. It is about 18.03% of the total area of the territory. The population of NCT

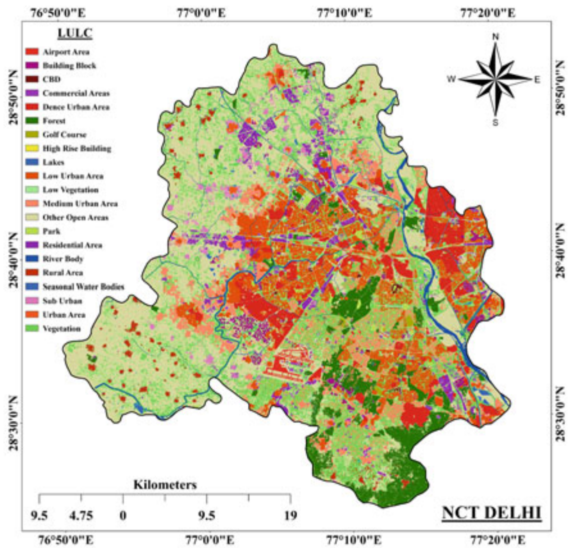


Fig. 12.3 Land use-land cover of NCT Delhi

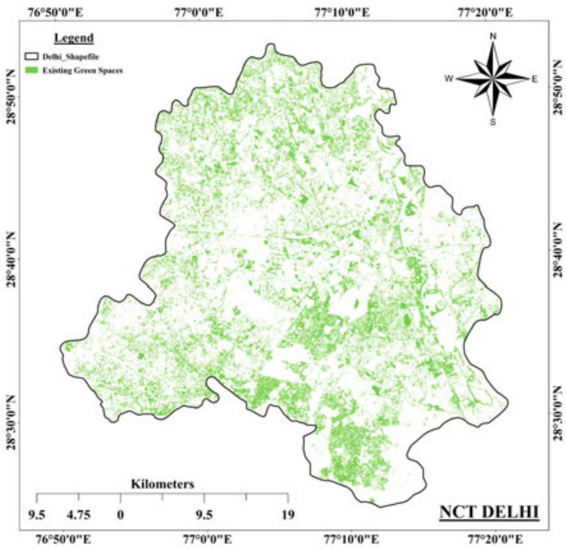


Fig. 12.4 Existing green spaces in NCT Delhi

Delhi, according to the 2011 Census, is 16,787,941 persons. Therefore, according to Eq. (12.1), the per capita urban green space of NCT Delhi is 15.93 m^2/person. As the WHO recommends that the per capita green space availability should be 9.5 m^2/person or more, NCT Delhi has sufficient urban green space.

Although the crude value of per capita green space of NCT Delhi (15.93 m^2/person) shows that the population is fairly served by the available urban green space of NCT Delhi, this crude figure undermines the area-specific disparities. Therefore, in this study, we adopted a district-level analysis of per capita urban green space in NCT Delhi. There are nine districts in NCT Delhi. Their per capita green space was calculated using Eq. (12.2). The population of these districts as per the 2011 Census is shown along with the urban green space area in Table 12.1. Four out of the nine districts of NCT Delhi, i.e. Central Delhi, Northeast Delhi, East Delhi and West Delhi, are facing a shortage of adequate urban green space as these districts have per capita green space of less than 9.5 m^2/person.

From Fig. 12.5, it is clear that there is a deficit in the urban green space in four districts of NCT Delhi, and sufficient urban green space in the remaining districts. The disparity of this distribution of urban green space vis-à-vis population necessitates that more urban green spaces shall be developed in the territory. As a result, the first pertinent step in the process is to identify the potential sites for the development of the urban green spaces. In this regard, buffers were created around the central elements as mentioned in the section on the methodology of this chapter, including the existing green spaces, water bodies and roads (Fig. 12.6).

Around the waterbodies, multiple ring buffers were demarcated at distances of 20, 40, 60, 80 and 100 m and weights of 1, 2, 3, 4 and 5 were assigned. Around the existing green spaces, a simple 20 m buffer zone was identified and given the

Table 12.1 Per capita urban green space of NCT Delhi

District	Green space area (km^2)	Population (2011)	Per capita UGS (m^2/person)
Central Delhi	3.04	582,320	5.23
New Delhi	12.00	142,004	84.52
Northwest Delhi	80.85	3,656,539	22.11
Northeast Delhi	3.89	2,241,624	1.74
East Delhi	8.91	1,709,346	5.21
South Delhi	44.57	2,731,929	16.32
Southwest Delhi	85.89	2,292,958	37.46
West Delhi	14.86	2,543,243	5.84
North Delhi	13.34	887,978	15.02
NCT Delhi (Total)	**267.38**	**16,787,941**	**15.93**

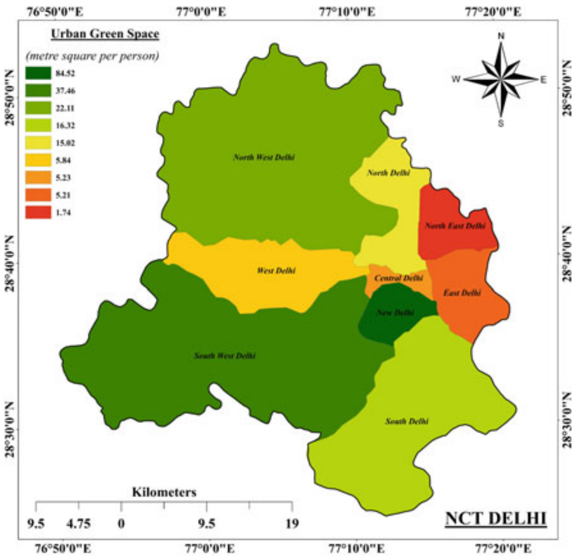


Fig. 12.5 Per capita urban green space (NCT Delhi-district wise)

weight 2. Similarly, around the roads, again a simple 20 m buffer area was identified and weighted as 2. The vector layers were rasterized, and the weights assigned to the classes of these layers were used in to combine these layers using cell statistics (local operation on raster layers) to deduce the urban green space expansion site suitability map (Fig. 12.7).

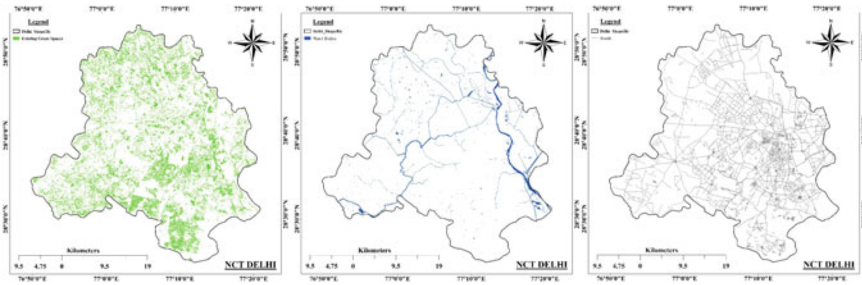
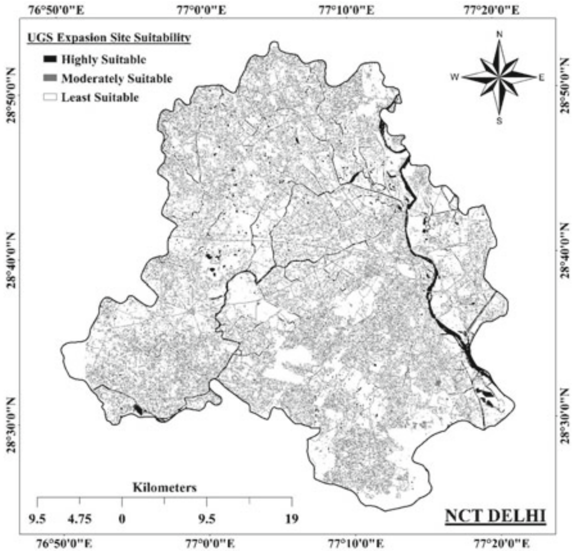


Fig. 12.6 Central elements for buffering to find potential sites for urban green space expansion

Fig. 12.7 Potential sites for urban green space expansion



The suitability map depicts the potential sites for the expansion of the urban green spaces. These sites have been classified into three categories based on the level of suitability, i.e. highly suitable, moderately suitable and least suitable. As such the highly suitable area available for green space expansion according to the suitability map given in Fig. 12.7 is 47.92 km^2 and the moderately suitable area is 320.48 km^2. Together these two categories of suitable areas developed would add 368.40 km^2 area of green space to the existing 267 km^2 area making it 635.40 km^2. This shall raise the overall per capita urban green space availability from 15.93 m^2/person to 37.85 m^2/person. The change in the scenario at the district level is given numerically in Table 12.2 and graphically in Fig. 12.8.

Table 12.2 Urban green space expansion potential of NCT Delhi

District	Green space area (km^2)	Suitable area (highly suitable)	Suitable area (moderately suitable)
Central Delhi	3.04	0.54	4.12
New Delhi	12.00	0.72	11.94
Northwest Delhi	80.85	11.97	100.6
Northeast Delhi	3.89	1.75	8.23
East Delhi	8.91	3.29	10.54
South Delhi	44.57	8.95	49.47
Southwest Delhi	85.89	8.38	99.98
West Delhi	14.86	4.54	21.29
North Delhi	13.34	7.79	14.28
NCT Delhi (Total)	**267.38**	**47.92**	**320.48**

12.6 Conclusion

In this study, an urban green space analysis of NCT Delhi has been undertaken to understand its availability and distribution using the Sentinel-2A imagery and GIS. Urban Green Space is small and has issues with its composition, distribution and character; yet, it has a significant growth potential. There is acute shortage of green spaces per capita in four districts of NCT Delhi, necessitating the expansion of urban green spaces. The suitability map reveals that NCT Delhi has a 47.92 km^2 area that is highly suitable for green space expansion, which when transformed into the urban green space, per capita urban green space available in the districts with deficit does not cross the threshold of 9.5 m^2/person. The created potential map may serve as a foundation for future decisions about the selection of regions for the establishment of UGS. However, there is a need to consider the expansion of green spaces along the vertical dimensions of NCT Delhi, in addition to the transformation of spaces available along the horizontal dimension.

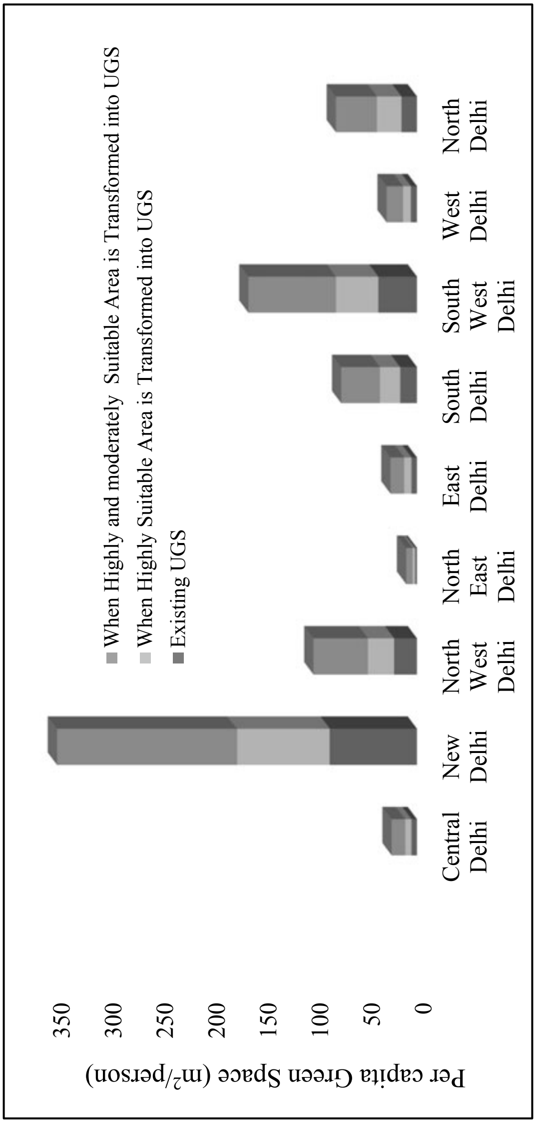


Fig. 12.8 Scenarios of per capita UGS at present and after UGS expansion

References

Bayram CB, Gökyer E (2012) Urban green space system planning. ÇankırıBartın University, Bartın and Turkey, pp 107–122

Byrne J, Sipe N (2010) Green and open space planning for urban consolidation—a review of the literature and best practice, vol 11. Griffith University

Crompton JL (2005) The impact of parks on property values: empirical evidence from the past two decades in the United States. Manag Leis 10(4):203–218

Hopkins LD (1977) Methods of generating land suitability maps: a comparative evaluation. J Am Plan Assoc 43(4):386–400

Norton BA, Coutts AM, Livesley SJ, Harris RJ, Hunter AM, Williams NS (2015) Planning for cooler cities: a framework to prioritise green infrastructure to mitigate high temperatures in urban landscapes. Landsc Urban Plan 134:127–138

Ragab K (2014) Quantitative evaluation of distribution and accessibility of urban green spaces—case study: City of Jeddah. Int J Geom Geos 4:526–535

Rakhshandehroo M, Yusof MJM, Arabi R, Parva M, Nochian A (2017) The environmental benefits of urban open green spaces. Alam Cipta 10(1):10–16

Senanayake IP, Welivitiya WDDP, Nadeeka PM (2013) Urban green spaces analysis for development planning in Colombo, Sri Lanka, utilizing THEOS satellite imagery—a remote sensing and GIS approach. Urban Forest Urban Green 12(3):307–314

World Health Organization (2017) Urban green spaces: a brief for action. World Health Organization, Regional Office for Europe: Copenhagen, Denmark, pp 1–24

Wu J (2014) Urban ecology and sustainability: the state-of-the-science and future directions. Landsc Urban Plan 125:209–221

Chapter 13
Case Study 5: A Multi-criteria Decision-Making for Alternative Landfill Site Selections Using Fuzzy TOPSIS Approach

Abstract Identification of suitable alternate landfill site is the Multi-Criteria Decision-Making (MCDM) problem. Several MCDM techniques are available to deal with such problem and used for decision-making. One such popular MCDM technique is Fuzzy TOPSIS which is applied in this case study for the purpose of alternate landfill site selection due to its ability to perform rational and appropriate decision-making in such types of problems. Varanasi is selected at the target region for assessing the best alternate landfill sites because Varanasi is facing serious threat of waste management in the region. To select the best alternate landfill sites in Varanasi, distance from waste production centre, distance from roads, depth to groundwater, distance from water bodies, distance from settlement land, types of soils and slope were considered. Out of these all seven criteria, four criteria, viz. distance from waste production centre, distance from roads, distance from settlement land and slope, were selected as beneficial criteria while rest three criteria, viz. depth to groundwater, distance from water bodies and types of soils, were considered as Non-beneficial (cost) Criteria. Five alternate sites, i.e. Khalispur, Bikapur, Rohania, Nuaon and Tikari, were selected to choose the best alternate among them. Result revealed that Rohania is the best alternate landfill site having highest Closeness Coefficient (0.899) and Tikari is the worst landfill site having lowest Closeness Coefficient (0.117).

Keywords Landfill sites · Fuzzy TOPSIS · FPIS · FNIS · Closeness coefficient

13.1 Introduction

Being a developing nation, India is witnessing a tremendous boom in the pace of urbanization, industrialization and population explosion. Due to this huge pressure is being exerted on the resources of the nation. Increased number of human activities in such developing nation are becoming the prime reason for the waste production. Production of proliferating amount of waste is creating severe threats over the region. Indian towns and city are not so much able to cope up with this huge problem, and the problem is increasing day by day. These towns and cities are not very much able to deal with the uncontrolled expansion of urbanization and industrialization. It is

M. Kumar et al., *Geographic Information Systems in Urban Planning and Management*, Advances in Geographical and Environmental Sciences,
https://doi.org/10.1007/978-981-19-7855-5_13

properly mentioned in the Indian Constitution under Article 21 that 'right to life' is the 'Fundamental Right' of the citizens. But this is the huge question in front of us that is, are we availing this right being surrounded by waste all around us?

It is the duty of the state to manage the proper amenities such that proper management of sewerage system, proper drainage management system and integrated approach to deal for the solid waste management. Day by day people are migrating from rural areas to urban areas in search of food, cloth and shelter. Thus, urbanization is increasing rapidly. Urbanization is working as influx for the population growth in the urban areas. All these factors are leading to increased amount of waste generation. And thus, burden over the governments and the concerned authorities is increasing day by day. Das et al. (1998) stated that larger than 90% of municipal solid waste are dispatched to landfill sites in most unhygienic manner.

The landfilling process of Municipal Solid Waste Management (MSWM) is too much unorganized in nature. The whole process is in omnishambles. One popular approach which is followed in India for landfilling of waste is done by simply dumping the waste material in the outside regions of the city without taking any concerns of sanitary measure. But main objective of landfilling is thus not so achieved because the main objective of landfilling is to minimize the exposure between humans and environments from the toxic wastes. Uncontrolled landfilling of solid waste could lead to increase soil, air and water pollution also.

Considering these all aspects in mind, Indian parliament has given approval to "Implementation of Municipal Solid Waste Management Rules, 2000". These rules mandated that the disposal of waste material should be done in scientific manner at suitable landfill sites, and the same is to be assured by concerned urban local bodies. But many cities in India have still not implemented this provision.

To consider the suitable landfilling sites, various technical, environmental, political and social factors need to be considered. Some major criteria are decided by Central Public Health and Environment Engineering Organization (CPHEEO) and Central Pollution Control Board (CPCB) to prefer the landfill sites in India.

To consider the suitable landfill sites there are various spatial and non-spatial factors should be considered for the final decision-making process. Hwang and Yoon developed the "Technique for Order Performance by Similarity to Ideal Solution (TOPSIS)" method in 1981. It is the most classical method to deal with Multi-Criteria Decision-Making (MCDM) Problems.

TOPSIS is based on the concept to select the best option among available alternates on the basis of certain important criteria. Best alternative solution in TOPSIS method is decided on the basis of largest distance from Negative Ideal Solution (NIS) and smallest distance from Positive Ideal Solutions (PIS). The benefit criteria are minimized, and the cost criteria are maximized by the NIS; at the same time PIS minimized the cost criteria and maximized the benefit criteria. Linguistic variables are used in this method to define the rank and weight of the criteria. But major problems associated with this technique was to handle such huge linguistic variables. To overcome this limitation Fuzzy TOPSIS method has been developed in recent years. Here Fuzzy number are chosen at the place of linguistic variables to reduce the problem associated. Chen and Hwang were the first who applied the Fuzzy TOPSIS for the decision-making. Liang proposed Fuzzy MCDM on the basis of ideal and non-ideal concept in the fray. This Fuzzy TOPSIS technique is easy to apply and gives satisfactory result in various studies.

Being one of the most used MCDM techniques in the literatures some major advantages of this technique are mentioned below.

(a) A sound logic that represents the reasoning behind human choice.
(b) A scaler value that takes into account both the best and worst alternatives at the same time.
(c) A simple computation process that is simple to programme,
(d) A performance measure for all alternatives that can be visualized on a polyhedron in any two dimensions.

There are several cities in India which are facing the huge problem of suitable landfilling sites. In this case study we are intended to discuss the landfill problem of the Varanasi District, Uttar Pradesh, India.

13.2 Overview of the Study Area

Varanasi popularly known as ‘Kashi’ or ‘Banaras’ is one of the oldest living cites in India. Name of the Varanasi is derived from the two tributaries of Ganga named “Varuna” and “Assi” which are flowing in northern and southern extents of the Varanasi. Its land has been the ultimate pilgrimage spot for the ‘Hindus’ from the ancient era. Varanasi is situated at an elevation of 80.72 m in central Ganga Valley of north India, in the state of Uttar Pradesh. Longitudinal extent of the Varanasi district is from 82° 39′ E to 83° 10′ E while its latitudinal extent is observed from 25° 10′ N to 25° 37′ N. This is situated in the eastern part of Uttar Pradesh. It is expanded over the area of 1454.11 km^2. According to 2011 census, population of 1,597,051 (43.44% of the total population) are residing in the urban regions of the Varanasi District. The topography of Varanasi is underlain by the alluvium deposits of quaternary age. Varanasi is situated on the left bank of Ganga River and lies in the heart of the middle Ganga plain. The land is very fertile because low-level floods in the Ganges replenish the soil on a regular basis.

It is one of the old luminous cities, which is also a successful industrial city, a place of knowledge, philosophy, education, culture, music and Indian traditional arts and crafts. Varanasi has provided the ideal environment for all cultural activities to thrive.

Varanasi experiences humid subtropical climate with large variation between summer and winter temperature. Fog is common in winter while hot dry winds blown in summer are called as loo. The soil of Varanasi is moderately alkaline in nature with a pH value of 8.2.

Varanasi attracts the tourists not only from India but also from throughout the world. According to Uttar Pradesh Tourism Department, Varanasi received over 233 million tourists (23 crores) in 2018! That is more than 56 times the city’s native population. This, combined with poor waste management, has resulted in a perpetual problem of filth and garbage with nowhere to dump it. Varanasi generated over 600 tonnes of waste per day that year, according to a 2013 report (no other calculations have been made since). That equates to more than 20,000 tonnes (20 million kilograms) of waste per month! Now, after nine years, this is very much necessary for proper scientific waste management at suitable landfill sites. Study area is shown in Fig. 13.1. According to Varanasi Municipal Corporation, there are several landfill sites in Varanasi, but major landfill sites are Khalispur, Bikapur, Rohania, Nuaon and Tikari. In this case study we are intended to obtain the best alternate landfill sites among these all using Fuzzy TOPSIS algorithm.

13.3 Database and Methodology

In this chapter we are intended to obtain the best landfill sites in Varanasi District of Uttar Pradesh among available 5 alternate landfill sites, i.e. Bikapur, Khalispur, Nuaon, Tikari and Rohania (Ohri et al. 2015). Information of these landfill sites was obtained from Varanasi District Municipal Corporation. We very well know that selection of alternate landfill sites is a multi-criteria decision-making problem. To cope up with such problems one such popular Multi-Criteria Decision-Making technique named Fuzzy TOPSIS was developed by Hwang and Yoon. Decision is taken on the basis of an index established on the base of similarity (or relative closeness) to the Fuzzy Positive Ideal Solution and distance from the Fuzzy Negative Ideal Solution. The alternative with the greatest closeness coefficient is then chosen as the best alternate (Yoon and Hwang 1995).

13.3.1 Criterion for the Selection of Landfill Sites

In order to select the best alternate landfill site in Varanasi District an extensive evaluation process needs to be done. Such identified best location must also meet the requirements of existing government regulations while also minimising economic, environmental, health and social costs (Siddiqui et al. 1996). A criterion is a decision-making of alternate landfill sites that can be measured and evaluated on the basis of two type of criterion: Beneficial Criterion and Non-beneficial Criterion (Cost Criterion). These two criteria are presented in Table 13.1.

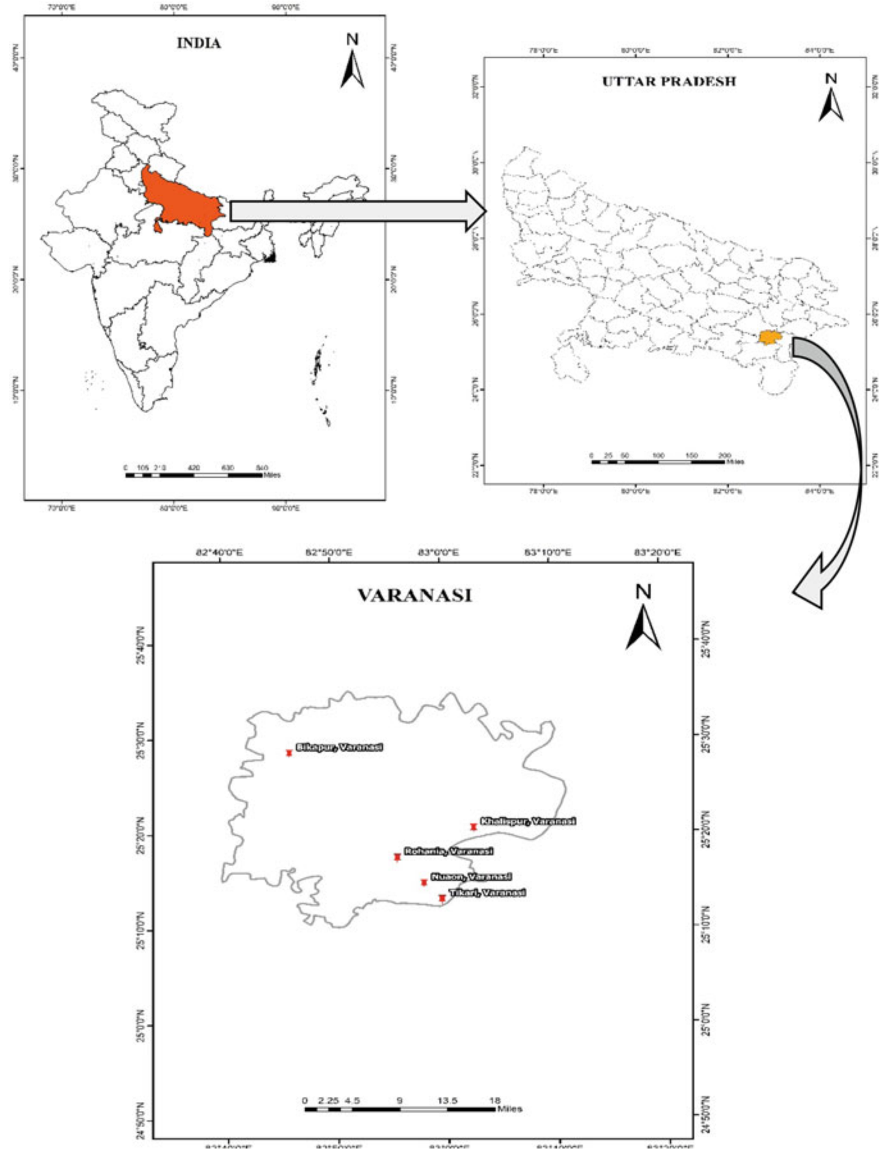


Fig. 13.1 Study area map

Table 13.1 Types of criteria for the alternate landfill site selection

Beneficial criterion	Non-beneficial criterion (cost criterion)
Distance from waste production centre	Depth to groundwater
Distance from roads	Distance from water bodies
Distance from settlement area	Type of soil
Slope	

13.3.2 Preparation of Fuzzy Rank Decision Matrix

After deciding the Beneficial and Non-beneficial Criterion in next step rank decision matrix is prepared on the basis of expert knowledge. In this case study we're considering only one alternate of Rank Decision Matrix. But in upcoming sections it's very well described to deal with more than one Rank Decision Matrix.

Rank Decision Matrix was prepared considering a 6-point scale which is listed in Table 13.2 (Beskese et al. 2015).

Rank decision matrix for each landfill site is listed in Table 13.3.

After defining the rank and weight to each criterion for each landfill sites as given in Table 13.3, corresponding numerical value is assigned to each rank according to Table 13.2 in our very next step.

Please consider that, if there are more than such Rank Decision Matrix (in the case of more than one expert) then corresponding numerical values will be assigned to each matrix.

Numerical values correspond to each rank which are mentioned in Table 13.4.

13.3.2.1 Preparation of Fuzzy Rank Decision Matrix in the Case of More Than One Decision-Makers

Suppose if we consider that more than one (say: n) decision-makers are participating in the process of decision-making. They give ranks to each criterion $\widetilde{x_{(i,j)n}} =$

Table 13.2 Linguistic scale to rank criteria

Rank	Numerical value
Just equal	(1, 1, 1)
Equally important	(1, 1, 3)
Weakly important	(1, 3, 5)
Essentially important	(3, 5, 7)
Very strong important	(5, 7, 9)
Absolutely important	(7, 9, 9)

Table 13.3 Rank decision matrix on the basis of expert opinion

Weight	Absolutely important	Essentially important	Essentially important	Essentially important	Very strong important	Weakly important	Equally important
Landfill Sites	Distance from waste production centre	Distance from roads	Depth to groundwater	Distance from water bodies	Distance from settlement area	Type of soil	Slope
Bikapur	Equally important	Essentially important	Weakly important	Weakly important	Essentially important	Weakly important	Equally important
Khalispur	Essentially important	Essentially important	Equally important	Very strong important	Very strong important	Weakly important	Equally important
Nuaon	Weakly important	Very strong important	Equally important	Essentially important	Weakly important	Essentially important	Essentially important
Tikari	Just equal	Just equal	Absolutely important	Absolutely important	Just equal	Weakly important	Very strong important
Rohania	Very strong important	Very strong important	Weakly important	Weakly important	Very strong important	Weakly important	Equally important

$\left(a_{(i,j)n}, b_{(i,j)n}, c_{(i,j)n}\right)$ and weight to each criterion $\tilde{w}_{(j)n} = \left(w_{1(j)n}, w_{2(j)n}, w_{3(j)n}\right)$ where n is the number of decision-makers; i and j are corresponding values to criteria.

Combined Fuzzy ranking for each criterion is calculated by

$$\tilde{x}_{(i,j)} = \left(a_{(i,j)}, b_{(i,j)}, c_{(i,j)}\right)$$

Here, $a_{(i,j)} = \min_n\left\{a_{(i,j)n}\right\}, b_{(i,j)} = \frac{1}{n}\sum_{n=1}^{n} b_{(i,j)n}$ & $c_{(i,j)} = \max_n\left\{c_{(i,j)n}\right\}$

Combined weight for each criterion is calculated by

$$\tilde{w}_{(j)} = \left(w_{1(j)}, w_{2(j)}, w_{3(j)}\right)$$

Here, $w_{1(j)} = \min_n\left\{w_{1(j)n}\right\}, w_{2(j)} = \frac{1}{n}\sum_{n=1}^{n} w_{2(j)n}$ & $w_{3(j)} = \max_n\left\{w_{3(j)n}\right\}$

13.3.3 Normalized Fuzzy Decision Matrix

After creation of the combined Fuzzy Decision Matrix next process is to create the Normalized Fuzzy Decision Matrix. Formula for the Normalization of the Fuzzy Decision Matrix is different for the Beneficial Criteria and Non-beneficial Criterion.

Table 13.4 Numerical values correspond to each rank on the basis of Beskese et al. (2015)

Weight	7	9	9	3	5	7	3	5	7	3	5	7	5	7	9	1	3	5	1	1	3
Landfill sites	Distance from waste production centre			Distance from roads			Depth to groundwater			Distance from water bodies			Distance from settlement area			Type of soil			Slope		
Bikapur	1	1	3	3	5	7	1	3	5	1	3	5	3	5	7	1	3	5	1	1	3
Khalispur	3	5	7	3	5	7	1	1	3	5	7	9	5	7	9	1	3	5	1	1	3
Nuaon	1	3	5	5	7	9	1	1	3	3	5	7	1	3	5	3	5	7	3	5	7
Tikari	1	1	1	1	1	1	7	9	9	7	9	9	1	1	1	1	3	5	5	7	9
Rohania	5	7	9	5	7	9	1	3	5	1	3	5	5	7	9	1	3	5	1	1	3

(i) *For Beneficial Criterion*:

$$\tilde{R}_{(i,j)} = \left[\tilde{r}_{(i,j)}\right]_{(m\times n)} = 1, 2 \ldots m \quad \text{and} \quad j = 1, 2 \ldots n$$

$$\text{Here, } \tilde{r}_{(i,j)} = \left(\frac{a_{(i,j)}}{c_j^*}, \frac{b_{(i,j)}}{c_j^*}, \frac{c_{(i,j)}}{c_j^*}\right); \quad \text{where, } c_j^* = \max_i\left\{c_{(i,j)}\right\}$$

(ii) *For Non-beneficial (Cost Criterion)*:

$$\tilde{r}_{(i,j)} = \left(\frac{a_j^-}{c_{(i,j)}}, \frac{a_j^-}{b_{(i,j)}}, \frac{a_j^i}{a_{(i,j)}}\right); \quad \text{where, } a_j^i = \min_i\left\{a_{(i,j)}\right\}$$

For this study Normalized Fuzzy Decision Matrix is shown in Table 13.5.

13.3.4 Weighted Normalized Fuzzy Decision Matrix

After preparation of the Normalized Fuzzy Decision Matrix, Weighted Normalized Fuzzy Decision Matrix needs to be calculated. Process of the computation of the Weighted Normalized Fuzzy Decision Matrix is given below.

$$\text{Weighted Normalized Fuzzy Decision Matrix } \tilde{V}_{(i,j)} = \tilde{r}_{(i,j)} \times \tilde{w}_{(j)}$$

Thus prepared Weighted Normalized Fuzzy Decision Matrix for alternate landfill site selection is presented in Table 13.6.

13.3.5 Fuzzy Positive Ideal Solution and Fizzy Negative Ideal Solution (FPIS & FNIS)

We have computed the Weighted Normalized Fuzzy Decision Matrix till here. In the next step it is necessary to compute the FPIS and FNIS for the calculation of the distance from FPIS and FNIS. Formula to compute the FPIS and FNIS is given below (Ashrafzadeh et al. 2012).

(i) *Fuzzy Positive Ideal Solution (FPIS), A^**

$$A^* = \left(\tilde{v}_1^*, \tilde{v}_2^*, \ldots, \tilde{v}_n^*\right), \text{ where } \tilde{v}_j^* = \max_i\left\{v_{(i,j3)}\right\};$$
$$i = 1, 2, \ldots m \;\&\; j = 1, 2, \ldots n$$

(ii) *Fuzzy Negative Ideal Solution (FNIS), A^-*

Table 13.5 Normalized fuzzy decision matrix

Weight	7	9	9	3	5	7	3	5	7	3	5	7
Landfill sites	Distance from waste production centre			Distance from roads			Depth to groundwater			Distance from water bodies		
Bikapur	0.111111	0.111111	0.333333	0.333333	0.555556	0.777778	0.2	0.333333	1	0.2	0.333333	1
Khalispur	0.333333	0.555556	0.777778	0.333333	0.555556	0.777778	0.333333	1	1	0.111111	0.142857	0.2
Nuaon	0.111111	0.333333	0.555556	0.555556	0.777778	1	0.33333	1	1	0.142857	0.2	0.333333
Tikari	0.111111	0.111111	0.111111	0.111111	0.111111	0.111111	0.111111	0.111111	0.142857	0.111111	0.111111	0.142857
Rohania	0.555556	0.777778	1	0.555556	0.777778	1	0.2	0.333333	1	0.2	0.333333	1

Weight	5	7	9	1	3	5	1	1	3
Landfill sites	Distance from settlement area			Type of soil			Slope		
Bikapur	0.333333	0.555556	0.777778	0.2	0.333333	1	0.111111	0.111111	0.333333
Khalispur	0.555556	0.777778	1	0.2	0.333333	1	0.111111	0.111111	0.333333
Nuaon	0.111111	0.333333	0.555556	0.142857	0.2	0.333333	0.333333	0.555556	0.777778
Tikari	0.111111	0.111111	0.111111	0.2	0.333333	1	0.555556	0.777778	1
Rohania	0.555556	0.777778	1	0.2	0.333333	1	0.111111	0.111111	0.333333

Table 13.6 Weighted normalized fuzzy decision matrix

	Distance from waste production centre			Distance from roads			Depth to ground water			Distance from water bodies		
Bikapur	0.777778	1	3	1	2.777778	5.444444	0.6	1.666667	7	0.6	1.666667	7
Khalispur	2.333333	5	7	1	2.777778	5.444444	1	5	7	0.333333	0.714286	1.4
Nuaon	0.777778	3	5	1.666667	3.888889	7	0.99999	5	7	0.428571	1	2.333333
Tikari	0.777778	1	1	0.333333	0.555556	0.777778	0.333333	0.555556	1	0.333333	0.555556	1
Rohania	3.888889	7	9	1.666667	3.888889	7	0.6	1.666667	7	0.6	1.666667	7

	Distance from settlement area			Type of soil			Slope		
Bikapur	1.66667	3.888889	7	0.2	1	5	0.111111	0.111111	1
Khalispur	2.777778	5.444444	9	0.2	1	5	0.111111	0.111111	1
Nuaon	0.555556	2.333333	5	0.142857	0.6	1.666667	0.333333	0.555556	2.333333
Tikari	0.555556	0.777778	1	0.2	1	5	0.555556	0.777778	3
Rohania	2.777778	5.444444	9	0.2	1	5	0.111111	0.111111	1

$$A^- = \left(\tilde{v}_1^-, \tilde{v}_2^-, \ldots, \tilde{v}_n^-\right), \text{ where } \tilde{v}_j^- = \min_i \left\{v_{(i,j1)}\right\};$$
$$i = 1, 2, \ldots m \; \& \; j = 1, 2, \ldots n$$

FPIS and FNIS for each criterion are presented in Table 13.7.

13.3.6 Distance from FPIS and FNIS

In our last step, we have calculated the FPIS and FNIS. Here we calculate the distance from FPIS and FNIS.

$$\text{Distance from FPIS } (D^*) = \tfrac{1}{3}\sqrt{\left[\left(v_1 - A_1^*\right)^2 + \left(v_2 - A_2^*\right)^2 + \left(v_3 - A_3^*\right)^2\right]}.$$

$$\text{Distance from FNIS } \left(D^-\right) = \frac{1}{3}\sqrt{\left[\left(v_1 - A_1^-\right)^2 + \left(v_2 - A_2^-\right)^2 + \left(v_3 - A_3^-\right)^2\right]}$$

Summation of FPIS is considered as d_i^*, and summation of FNIS is considered as d_i^- (Tables 13.8 and 13.9)

13.3.7 Closeness Coefficient and Suitability Rank

In the last step Closeness Coefficient (CC) is calculated on the basis of d_i^* and d_i^-. On the basis of obtained value of CC final suitability rank is decided. Higher value of CC should be given higher rank while lower values should be assigned lower values.

CC is obtained by following formula (Yildirim et al. 2018):

$$\text{Closeness Coefficient (CC)} = \frac{d_i^-}{d_i^* + d_i^-}$$

Final table which represents the CC and suitability rank is given below (Table 13.10).

13.4 Result and Discussion

Solid waste landfilling is the most widely used waste disposal method in the world. The first and most important step in any landfilling practice is site selection. The Fuzzy TOPSIS method for site selection was investigated in this study. The methods were used to investigate five landfilling alternatives for the city of Varanasi (Bikapur, Khalispur, Nuaon, Tikari and Rohania). The overall criterion was chosen in the first

Table 13.7 FPIS and FNIS for each criteria

	Distance from waste production centre			Distance from roads			Depth to groundwater			Distance from water bodies		
Bikapur	0.777778	1	3	1	2.777778	5.444444	0.6	1.666667	7	0.6	1.666667	7
Khalispur	2.333333	5	7	1	2.777778	5.444444	1	5	7	0.333333	0.714286	1.4
Nuaon	0.77778	3	5	1.666667	3.888889	7	0.99999	5	7	0.428571	1	2.333333
Tikari	0.777778	1	1	0.333333	0.555556	0.777778	0.333333	0.555556	1	0.333333	0.555556	1
Rohania	3.888889	7	9	1.666667	3.888889	7	0.6	1.666667	7	0.6	1.666667	7
A*	3.888889	7	9	1.666667	3.888889	7	1	5	7	0.6	1.666667	7
A-	0.777778	1	1	0.333333	0.555556	0.777778	0.333333	0.555556	1	0.333333	0.555556	1

	Distance from settlement area			Type of soil			Slope		
Bikapur	1.666667	3.888889	7	0.2	1	5	0.111111	0.111111	1
Khalispur	2.777778	5.444444	9	0.2	1	5	0.111111	0.111111	1
Nuaon	0.555556	2.333333	5	0.14257	0.6	1.666667	0.333333	0.555556	2.333333
Tikari	0.555556	0.777778	1	0.2	1	5	0.555556	0.777778	3
Rohania	2.777778	5.444444	9	0.2	1	5	0.111111	0.111111	1
A*	2.777778	5.444444	9	0.2	1	5	0.333333	0.555556	2.333333
A-	0.555556	0.777778	1	0.142857	0.6	1.666667	0.111111	0.111111	1

Table 13.8 Distance of each criteria from FPIS and d_i^*

	Distance from waste production centre	Distance from roads	Depth to groundwater	Distance from water bodies	Distance from settlement area	Type of soil	Slope	d_i^*
Bikapur	5.21788	1.16887	1.93830	1.94×10^{-6}	1.59730	1.93858	0.82151	***12.6824***
Khalispur	1.86366	1.168870	0	3.28319	2.21×10^{-5}	0	0.82151	***7.137***
Nuaon	3.727335	2.08×10^{-5}	5.77×10^{-6}	2.7234	3.19463	1.9385	3.74×10^{-6}	***1584***
Tikari	6.04646	4.14748	4.32810	3.52636	5.49895	0	02,552	***23.972***
Rohania	6.41×10^{-5}	2.02×10^{-6}	1.93830	1.94×10^{-6}	2.59×10^{-6}	0	0.82151	***2.7598***

Table 13.9 Distance of each criteria from FNIS and d_i^-

	Distance from waste production centre	Distance from roads	Depth to groundwater	Distance from water bodies	Distance from settlement area	Type of soil	Slope	d_i^-
Bikapur	1.15470	3.00890	3.5263	3.52636	3.95	1.9385	9.02×10^{-7}	17.109
Khalispur	4.25909	3.00890	4.3281	0.24857	5.50	1.9385	9.02×10^{-7}	19.2821
Nuaon	2.5819	4.14748	4.32805	0.81330	2.48	2.54×10^{-6}	0.8215	15.1702
Tikari	1.28×10^{-6}	3.58×10^{-6}	3.20×10^{-6}	3.27×10^{-6}	0.00	1.9385	1.2439	3.18256
Rohania	6.0464	4.14748	3.5263	3.52636	5.50	1.93858	9.72×10^{-7}	24.684

Table 13.10 Closeness coefficient and suitability rank

	d_i^*	d_i^-	Closeness coefficient	Rank
Bikapur	12.68248	17.10938	0.574297	3
Khalispur	7.137282	19.28212	0.729847	2
Nuaon	11.58402	15.17029	0.567022	4
Tikari	23.9729	3.182516	0.117196	5
Rohania	2.759864	24.68422	0.899437	1

section. Following that, these criteria were divided into Beneficial and Non-beneficial (cost) Criteria.

The closeness coefficients of the choices were calculated as 0.574 for Bikapur, 0.729 for Khalispur, 0.567 for Nuaon, 0.117 for Tikari and 0.899 for Rohania based on the Fuzzy TOPSIS Modeling results. Thus, it is very easy to conclude from this study that Rohania is the best alternative landfill site, while Tikari is the worst. The government could use these techniques to make their decisions more rational and appropriate.

One of the most significant benefits of the proposed methodology is its inherent flexibility.

It is recommended that such models be easily accessible, adapted to other cities around the world by changing the criteria, ranking and weighting based on importance in that region. As a result, the model structure used for Solving the Varanasi district's landfill site selection problem could also be applied to other locations having similar issue.

13.5 Conclusion

We know that the whole world is facing the problem of waste management tremendously. Varanasi is considered as the "*Devon Ki Nagari*" and the pilgrimage from all around the world reach here. Thus, waste management is not only important for the best survival of the local people but also to leave a good imprint on international horizon. Various techniques are available for the waste treatment. But one major question still remained unanswered and that is where to treat? In this study we are focused to answer the same question. We know that identification of the suitable landfill sites is multi-criteria decision problem. This decision-making process takes into account factors other than scientific evidence data mining and analysis, but also political factors and public approval. It is advised to readers that understanding of the criterion for the waste treatment landfill sites is very essential. Distance from the waste production centre, distance from the roads, distance from the water bodies, depth to groundwater, slope and nature of soil is such an important aspect which we have taken in our study for the estimation of the best landfill sites. Any important criterion other than this could be supplemented to make the result more effective and impressive.

References

Ashrafzadeh M, Rafiei FM, Isfahani NM, Zare Z (2012) Application of fuzzy TOPSIS method for the selection of warehouse location: a case study. Interdisc J Contemp Res Bus 3(9):655–671

Beskese A, Demir HH, Ozcan HK, Okten HE (2015) Landfill site selection using fuzzy AHP and fuzzy TOPSIS: a case study for Istanbul. Environ Earth Sci 73(7):3513–3521

Das D, Srinivasu MA, Bandyopadhyay M (1998) Solid state acidification of vegetable waste. Indian J Environ Health 40(4):333–342

Ohri A, Singh PK, Maurya SP, Mishra S (2015) Sanitary landfill site selection by using geographic information system. In: Proceedings of national conference on open-source GIS: opportunities and challenges, Department of Civil Engineering, IIT (BHU), Varanasi, , October 2015, pp 170–180

Siddiqui MZ, Everett JW, Vieux BE (1996) Landfill siting using geographic information systems: a demonstration. J Environ Eng 122(6):515–523

Yildirim V, Memisoglu T, Bediroglu S, Colak HE (2018) Municipal solid waste landfill site selection using multi-criteria decision making and GIS: case study of Bursa province. J Environ Eng Landsc Manag 26(2):107–119

Yoon KP, Hwang CL (1995) Multiple attribute decision making: an introduction. In: Sage University paper series on quantitative applications in the social sciences, Thousand Oaks, CA

Chapter 14
Case Study 6: Urban Flood Susceptibility Modelling of Srinagar Using Novel Fuzzy Multi-layer Perceptron Neural Network

Abstract Urban flooding (often referred to as water logging) is defined as the submergence of normally dry city areas by a considerable volume of water caused by heavy precipitation or overflowing of water bodies. Flood susceptibility modelling, by combining the effects of natural and human factors, determines the sensitivity of the space to flood hazard. Urban flood modelling has gained attention recently and since the incidence of urban floods has increased rapidly, due attention needs to be given to the urban flood studies. In this case study, urban flood susceptibility modelling of Srinagar City, Jammu and Kashmir, India, using Fuzzy MLPNN, has been carried out in Geographic Information System (GIS) environment. Fuzzy MLPNN is a simple and straightforward approach that unifies the complexity of the phenomenon of urban flooding by integrating fuzzy mathematics and machine learning to build a predictive model for the analysis of urban flood susceptibility using geospatial data. Eight flood conditioning factors (elevation, slope, profile curvature, plan curvature, geology, distance from natural streams, MFI and LULC) were used as independent variables along with urban flood locations as the dependent variable. A precursory FSZ map of Srinagar City was created using the frequency ratio technique, and non-flooded locations were accordingly determined. The developed model reveals the susceptibility of each and every pixel (12.5 $\times$ 12.5 m area) in the study area. The FSI, illustrated by the FSZ Map of Srinagar, demonstrates considerable susceptibility of the city to urban flood hazard. The dominant influence of spatiality of precipitation and water bodies is indicated by the conclusion that highly susceptible regions of the city are those where MFI is high and proximity to natural drainage is low. The FSZ map was validated using Area under the ROC Curve (AUC) Analysis, which substantiates the efficiency of the Fuzzy MLPNN model. With 0.931 and 0.922 AUC values, the success rate and predictive performance of the FSZ map come out to be excellent, respectively.

Keywords Flood susceptibility modelling · Urban flood · Fuzzy MLPNN · Srinagar city

M. Kumar et al., *Geographic Information Systems in Urban Planning and Management*, Advances in Geographical and Environmental Sciences,
https://doi.org/10.1007/978-981-19-7855-5_14

14.1 Introduction

Flooding is a natural hazard that affects goods, services, properties and the lives of humans and animals. It is the most common type of natural disaster in the world. Flooding in urban areas is linked to unplanned and haphazard urban construction that destroys the balance between drainage systems and the amount of precipitation. Urban floods are classified according to the source of occurrence as *rain-based, snow-based* and *mixed-based*, with rain-based urban floods being the most common in the Indian subcontinent. Rapid urbanization, combined with its suction effect on rural demography, has resulted in unsustainable city morphologies. The issue is further complicated by the issue of rapid climate change. As a result, the urban landscape is modified in an environmentally objectionable manner. Throughout the world, urban settlements are dealing with issues ranging from water entering the ground floor of residences to large-scale inundation of entire cities, which is frequently caused by a disbalance between precipitation and sewer systems.

Urban flooding leads to the disruption of urban functional systems. Among the costs of urban flooding are direct damage, which includes material loss; indirect damage, which includes work loss, traffic congestion, diseases and other issues; and social consequences, which include long-term psychological effects of floods that influence public perceptions (König et al. 2002). As such, it is appropriate to prevent social and economic losses through hazard management.

Flood susceptibility modelling is the process of determining the vulnerability of a space to flood hazard using historical flood records as the function of a combined effect of environmental and human factors. It divides the space into homogeneous units and ranks them according to their sensitivity to actual or potential hazards. It integrates geographical data, statistical tools, mathematical tools, artificial intelligence algorithms (machine learning and deep learning), etc., to create area-specific methodologies for the creation of susceptibility maps. Flood susceptibility modelling is not only a precondition for hazard management, but it also identifies the variables that cause and control disasters, which must be taken into consideration for sustainable development planning.

14.2 Overview of the Study Area

Srinagar city is a subdivision of the district of Srinagar, which serves as the summer capital of the Indian state of Jammu and Kashmir. Located between latitudes 33° 59′ 17″ N and 34° 12′ 43″ N and longitudes 74° 40′ 17″ E and 74° 57′ 41″ E, it is situated on the banks of Jhelum River, which is a remnant of the Great Karewa Lake that once flooded the Valley. The study area is depicted in Fig. 14.1. The city is prone to urban flooding for three reasons. Firstly, the landscape of the city is packed with urban infrastructure, preventing rainfall from penetrating the earth's surface. Secondly, the region is affected occasionally by the combined effect of barotropic southwest

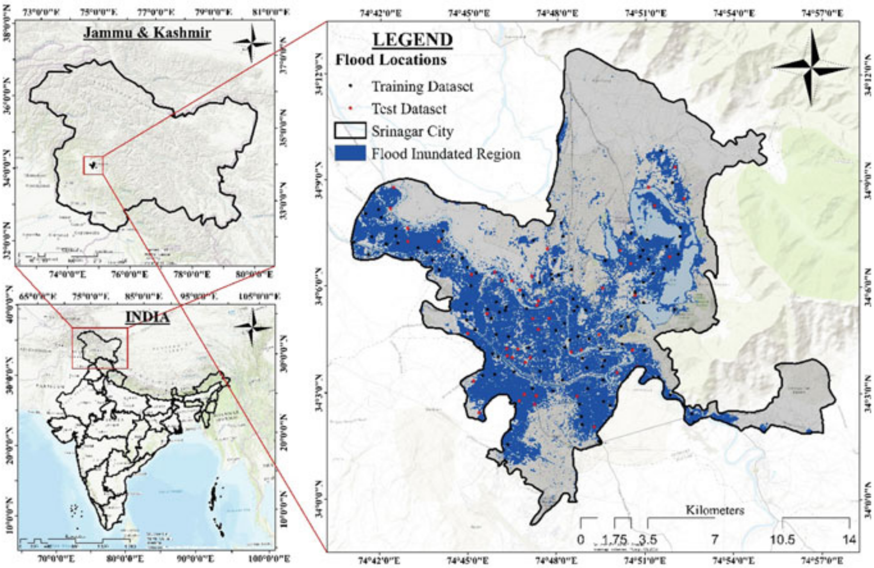


Fig. 14.1 Study area

monsoons and baroclinic western disturbances, resulting in higher precipitation (Bhat et al. 2019). Thirdly, the main river and its tributaries overflow during instances of heavy precipitation.

Geomorphologically, the city is dominated by flat terrain (slope less than 6.28°) except in the area of localized hills. Its elevation is 1528–1930 m (amsl). The Alluvial Screes and Karewa Formations dominate the region's geology. The climate is humid subtropical, with an average annual rainfall of 720 mm spread over 56 days. Srinagar city is generally at risk of flooding during protracted periods of heavy rain due to its location in the foothill zone of the Himalayan intermontane fault basin (Kashmir Valley), which is flanked on the northeast and southwest by the Greater Himalayas and Pir Panjal Mountain ranges.

14.3 Database and Methodology

In this section, we aim to elaborate on the prerequisites of the process of flood modelling to design the spatial database. The database includes flood inventory and flood conditioning factors. Suitable geostatistical or geomathematical approaches, which can be executed entirely or partially within the GIS environment, are also required in the process of flood modelling. To derive the comprehensive and highly precise flood susceptibility zonation (FSZ) map of Srinagar city, remote sensing (RS)

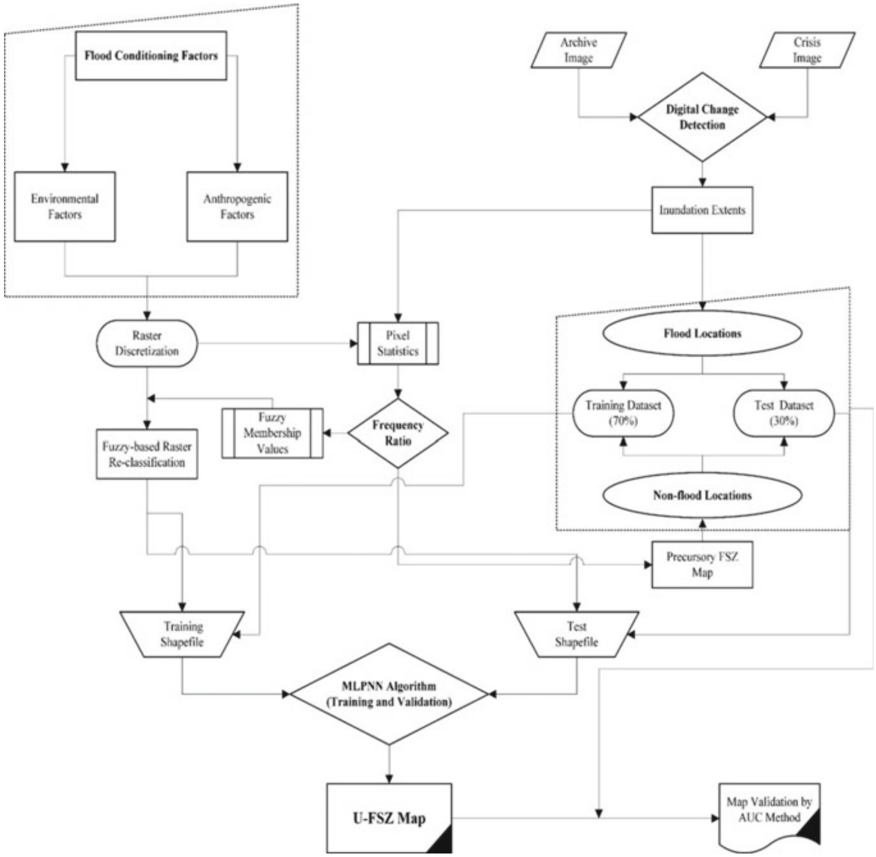


Fig. 14.2 Methodology

data, topographic maps and statistical data from secondary sources were used. The detailed methodology can be understood graphically from Fig. 14.2.

14.3.1 Flood Conditioning Factors

Flood conditioning factors (Ramesh and Iqbal 2020), also known as flood-influencing parameters (Tehrany et al. 2014a) and flood-related variables (Hong et al. 2017), are considered independent variables in flood susceptibility analysis (Tehrany et al. 2014a). It is important to choose carefully among the broad diversity of flood conditioning factors, keeping in mind the research region. A literature survey shows that environmental factors (geological, geomorphological hydrological and climatic) and anthropogenic factors are widely employed as flood conditioning factors. This study evaluated 8 flood conditioning factors based on related literature and the study area's conditions which are elevation, slope, profile curvature, plan curvature, geology, distance from natural streams, MFI and LULC. Sources of these flood conditioning factors are enumerated in Table 14.1.

The spatial database of the flood conditioning factors was prepared in ArcGIS **10.7@ESRI** and Erdas Imagine 2014 followed by classification and re-sampling into raster format with cell size 12.5 × 12.5 m covering the entire study area with 1,499,390 pixels. Figure 14.3 depicts the flood conditioning factors.

The environmental factors used in this study include topographic factors, geology, Modified Fournier Index (MFI) and distance from natural streams. The topographic factors were prepared using a 12.5 m resolution ALOS PALSAR DEM acquired from NASA's EarthData Portal. The research area was covered in a single data tile, which was *clipped* according to the study area boundary and *filled* to eliminate no-data

Table 14.1 Databases: parameters, their sources and source specifications

Parameter	Source	Source specifications
Flood inventory map	Landsat 8 OLI/TIRS images	Acquisition dates—August 25, 2014, and September 10, 2014
		Multi-spectral image
		Spatial resolution—30 m reflective, 15 m panchromatic and 100 m thermal
		Natural colour—4, 3, 2
		FCC combination—5, 4, 3
Topographic factors i. Elevation ii. Slope iii. Profile curvature iv. Plan curvature	ALOS PALSAR digital elevation model (DEM)	Spatial resolution—12.5 m
Distance from natural streams		
Geology	Directorate of Geology and Mining, J&K Government Srinagar—district survey report	Prepared as per Environment Impact Assessment (EIA) notification, 2016 of Ministry of Environment, Forest and Climate Change
		Published in 2017
Precipitation (MFI)	PRECTOT MERRA2 (NASA POWER)	Spatial resolution—0.5° × 0.5°
		Temporal resolution—1 Month
LULC	Landsat 5 image	Acquisition date—October 25, 2010
		Multi-spectral image
		Spatial resolution—30 m reflective and 120 m thermal
		Natural colour—3, 2, 1
		FCC combination—4, 3, 2

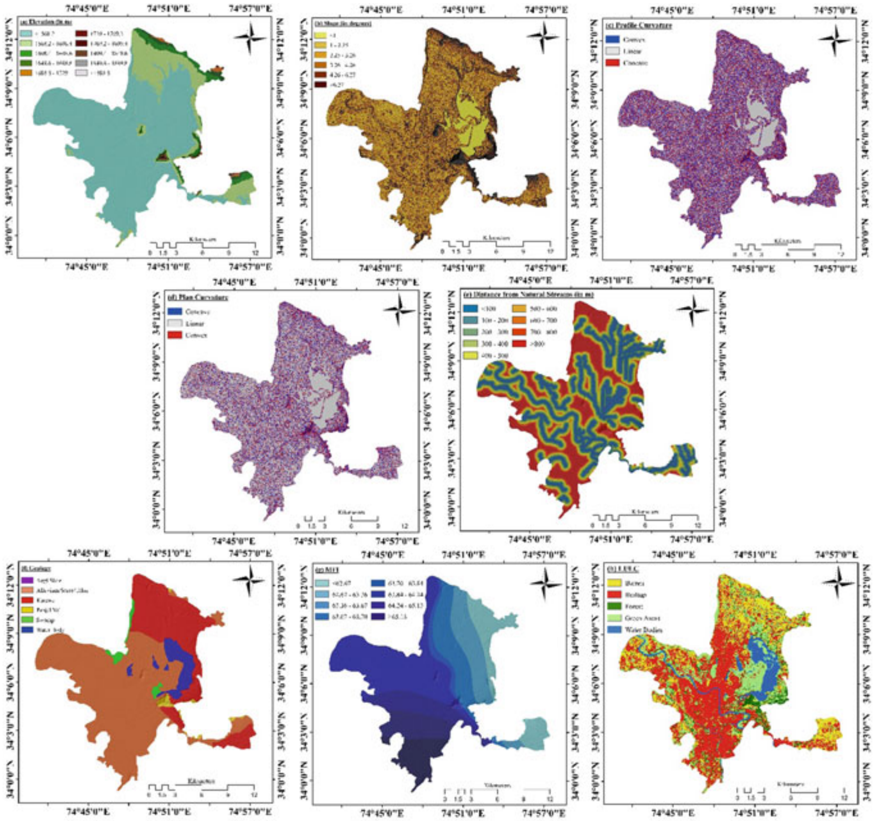


Fig. 14.3 Conditioning factors

values. Elevation, slope, profile curvature and plan curvature are the topographic factors used in the study. In general, higher elevation means less probability of floods and vice versa. The DEM was reclassified, and ten elevation categories were extracted by equal interval method in ArcGIS **10.7@ESRI**. The region is dominated by the elevation of 1528–1568.2 m, covering 76.32% area. Higher elevations can be observed in the northern and western periphery of the city and the interspersing local hills. The research region is characterized by flat terrain, which is a key factor in the occurrence of urban floods. Low-lying locations with gentle slopes are more susceptible to flooding than steep slopes that can rapidly drain rainwater (Ramesh and Iqbal 2020). The slope map was prepared and then reclassified into six classes based on quantile as it categorizes an equal number of entities (pixels) in each class. The study area has a flatter landscape with lower slope angles, where only a tiny proportion of pixels have slope angles greater than 6.27°. This slope shape reduces run-off speed and hence results in urban floods. Profile curvature is the curvature of the surface in the direction of the slope, and it controls the acceleration and deceleration of surface flow. Planform curvature, on the other hand, is the curvature of a surface perpendicular to the slope, and it relates to the convergence and divergence of flow across a surface. The profile and planform curvature were classified as concave, linear and convex.

Distance from Natural Streams impacts the magnitude of flooding, and area close to natural drainage is in general more susceptible to flooding. The streams were extracted from the DEM in ArcGIS **10.7@ESRI**, and using the *buffer* tool nine distance classes at a distance interval of 100 m were identified and rasterized with 12.5 × 12.5 m cell size.

Geology controls erodibility and permeability, which are crucial in urban flood mapping. Geological units occurring in Srinagar, as sourced from the District Survey Report (Srinagar District) by the Directorate of Geology and Mining, Jammu and Kashmir Government, are Cambrian Granite (Medium to coarse-grained biotite granite/gneiss, porphyritic granite, pegmatite, quartz vein, etc.); early Permian Panjal Trap (Andesitic and Basaltic formations), Nishat Formation (shale, slate, varvite and sandstone) and Agglomerate slate Sandstone, shales and slates); late Permian Zewan Formation (Cherts, shales, Limestone); Neogene-Quaternary Karewa Formation (clay, silt and sand, conglomerate); recent Alluvium (Sandy clay and gravel).

The amount of rain that falls in a specific region has a significant impact on the volume, level and kind of floods that occur in that area. Modified Fournier Index (MFI) is an alternative method of *R*-Factor estimation which was proposed by Arnoldus (1980) as an improved version of the Fournier Index as it considers rainfall data of all the months in the calculation as evident in Eq. (14.1).

$$\mathrm{MFI} = \sum p_i^2/P \tag{14.1}$$

where p_i stands for average rainfall of ith month of the year and P represents annual average rainfall.

To create the MFI map, rainfall data, i.e. PRECTOT MERRA2 (0.5° × 0.5°), Precipitation (mm day^{-1}) of twenty consecutive years (2000–2020) was downloaded, for 12-point locations inside and around the boundary of the study region, in.csv format from NASA Prediction of Worldwide Energy Resources (POWER) and statistically processed in Microsoft Excel. The calculated MFI values were integrated with the spatial data in ArcGIS **10.7@ESRI** and interpolated using the *inverse distance weighted* (IDW) method to derive the values for the entire area in the raster format of 12.5 × 12.5 m cell size, which ranged between 57.52 and 67.39 mm. The raster was then classified into ten classes by the quantile method. It is clear from the map that the western region of Srinagar receives a higher amount of rainfall than the eastern region.

Land use/land cover (LULC) is a dominant flood conditioning factor in urban areas, as the reduction in percolation enhances storm run-off (Tehrany et al. 2014b), and also there is a negative relationship between vegetation density and flooding. The LULC map was derived from a Landsat 5 satellite image of 30 m spatial resolution in ERDAS Imagine 2014. The actual image was clipped for the study area and classified into 100 classes by unsupervised method, followed by pixel recoding to yield a precise five-class raster (Built-up, Forest, Green Areas, Barren and Water Bodies). The image was then re-sampled in ArcGIS 10.7@ESRI for 12.5 × 12.5 m cell size.

14.3.2 Flood Inventory Databases

One of the most basic prerequisites for flood modelling is to locate and compile the urban flood inventory dataset (dataset of flood and non-flood locations). It is possible to create these datasets from primary data acquired in the field, secondary data sources or remote sensing data. The identification of flood-inundated region of Srinagar was made possible through the use of digital change detection analysis. Landsat 8 OLI/TIRS images acquired on August 25, 2014 (Archive Image), and September 10, 2014 (Crisis Image), were *imported* into ERDAS Imagine 2014 and *layer-stacked* to create the false colour composites (FCCs), which were *subset* using the shapefile of the study area. The FCCs were classified into two classes, one representing the water-covered parts of the study area and the other representing the dry parts. The class representing water in the archive image represented actual water bodies, but the class representing water in the crisis image represented areas that had been inundated by floodwaters as well as actual water bodies. Following classification, the images were subjected to *image differencing* using *image analysis* in ArcGIS **10.7@ESRI**. The resultant raster depicted the flood-inundated portion of the study area, which covered an area of 81.897 km^2. Prior to further processing, this flood inventory raster was re-sampled to a cell size of 12.5 × 12.5 m, thereby represented in 524,141 pixels. 163 flood locations were recorded in the flood-inundated region. Also, these points were categorized into the training dataset (70 per cent = 114) and test dataset (30 per cent = 49) using the *geostatistical analyst tools-utilities-subset features* tool. The inundated region and the datasets of flood locations are shown in Fig. 14.1.

Prior to being able to precisely identify the non-flood locations, a precursory flood susceptibility zonation map had to be created to identify the non-flood areas. This was achieved using the frequency ratio method. The Frequency Ratio embodies Hutton's concept of uniformitarianism, which argues that a past event (flood, earthquake, etc.) would reoccur in the same manner. FR is a data-driven, bivariate statistical analysis tool that illustrates the relationship between dependent and independent variables (Rahmati et al. 2016). FR is calculated as per Eq. (14.2). Apart from determining the relationship between the flood conditioning factors and the flood incidents, the frequency ratio model produces a flood susceptibility index (FSI) as per Eq. (14.3).

$$\mathrm{FR} = \frac{f(i)/\Sigma f}{p(i)/\Sigma p} \tag{14.2}$$

where $f(i)$ denotes flood pixels in each class of flood conditioning factor, Σf denotes the total number of flood pixels in the study area, $p(i)$ denotes the number of pixels in each class of flood conditioning factor and Σp denotes the total number of pixels in the study area.

$$\text{FSI} = \Sigma\text{FR} \tag{14.3}$$

14.3.3 Fuzzy Multi-layer Perceptron Neural Network (Fuzzy MLPNN)

Fuzzy MLPNN is an ensemble of fuzzy logic and Multi-layer Perceptron Neural Network that is developed as a novel tool in this case study for urban flood modelling purposes. According to the fuzzy sets theory of Zadeh (1965), fuzzy membership values varying between 0 and 1 determine the degree of membership (belongingness) of an element to a set. This is unlike the traditional Boolean Theory of Sets which indicates that an element is either completely a part of the set (degree of membership being 1) or not a part of the set (degree of membership being 0). Based on how they influence flooding, fuzzy membership values ranging between 0 and 1 are specified for each class of flood conditioning factor to create the flood susceptibility zonation map (Ramesh and Iqbal 2020). The method of determination of fuzzy membership values can be objective (data-driven) or subjective (expert judgement based). In this study, frequency ratio values of the flood conditioning factor classes are transformed into data-driven fuzzy membership values (μ) by *minimum–maximum normalization*, Eq. (14.4). Using these fuzzy values, the discretized flood conditioning factor rasters are reclassified to generate geospatial fuzzy layers, values of which could be used for input to the MLPNN algorithm.

$$\mu = \frac{x - \min}{\max - \min} \tag{14.4}$$

where x denotes FR value and min and max denote minimum and maximum values of FR of a particular flood conditioning factor.

An artificial neural network (ANN) is a machine learning algorithm designed to replicate the behaviour of human neural networks. The most widely used artificial neural network, the MLPNN, contains neurons on three levels: input, hidden and output. At the input level, flood conditioning components are involved that contribute to the occurrence of urban floods, which is the output. The hidden layer represents the input–output relationship in the MLPNN architecture. Classification problems can be effectively modelled using the MLPNN supervised learning algorithm (Meng et al. 2013). However, the number of neurons in the hidden layer must be specified to construct the most precise classification model possible.

To train the Fuzzy MLPNN model, the training dataset vector (consisting of 114 training flood locations and 114 training non-flood locations) was prepared in ArcGIS 10.7@ESRI. The attribute table of this vector layer was populated with the values of each flood conditioning factor fuzzy values. This geospatial data layer (shapefile) was

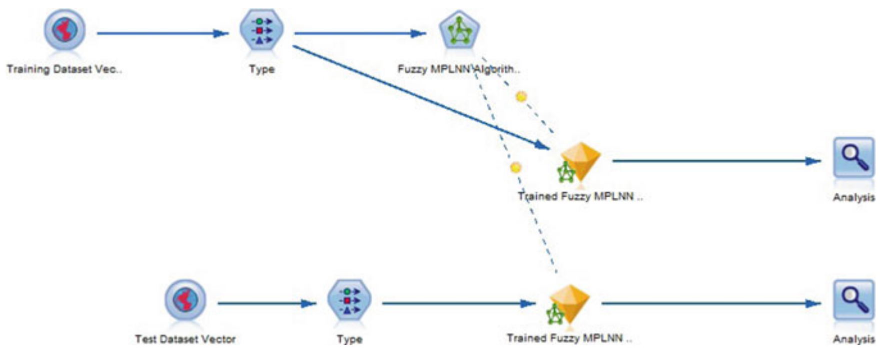


Fig. 14.4 Training and testing MLPNN in SPSS

directly used as input for MLPNN model designing in SPSS Modeller. The trained model was calibrated by the determination of the number of neurons in the hidden layer by hit and trial. The test dataset vector (consisting of 49 training flood locations and 49 training non-flood locations) was prepared in the same manner as the training dataset vector to access the predictive potential of the trained Fuzzy MLPNN model before the model could be applied to the prediction of flood susceptibility of each and every location within the study area. Figure 14.4 illustrates the Fuzzy MLPNN model preparation and validation framework. After the validation, this model was applied to predict the urban flood probability at each and every location (1,499,390 pixels of 12.5 × 12.5 m dimensions) of Srinagar city, which can be depicted as the Urban Flood Risk Zonation Map.

14.3.4 Accuracy Assessment of the Flood Risk Map

Accuracy Assessment involves the determination of the success rate and predictive performance of the resultant FSZ map. There are various methods of accuracy assessment, but the AUC method is commonly used. The area under the curve (AUC) refers to the area falling under the Receiver operating characteristic (ROC) curve, produced by plotting sensitivity values on *y*-axis and 1-specificity values on *x*-axis. Sensitivity is calculated as per Eq. (14.5) and represents the incidences that are correctly classified as floods in the final map. Specificity is measured as per Eq. (14.6) and signifies incidences that are correctly classified as non-flood.

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{14.5}$$

$$\text{Specificity} = \frac{\text{True Negative}}{\text{False Positive} + \text{True Negative}} \tag{14.6}$$

where the true positive and the true negative are correctly classified instances while false positive and false negative are incorrectly classified instances (Bui et al. 2019).

The area under the ROC curve determines the performance of the susceptibility maps. The value of AUC close to 1 validates an FSZ map as excellent, while its value close to 0.5 indicates failure. Accordingly, Pradhan and Kim (2017), classified the accuracy results as excellent (0.90–1), good (0.80–0.90), fair (0.70–0.80), poor (0.60–0.70) and fail (0.50–0.60). The success rate and prediction rate of maps are determined by overlaying the training dataset and test dataset, respectively, on it.

14.4 Results

Urban flood modelling of Srinagar City is carried out in this case study, using 8 flood conditioning factors and the Fuzzy MLPNN methodology. The flood locations were identified using the digital change detection of Landsat 8 OLI/TIRS imagery, as previously cited. For each discrete class of all flood conditioning factors, FR values were calculated as per Eq. (14.2) and spatially represented in the GIS environment. The summation of these FR layers according to Eq. (14.3) yielded FSI values and hence the precursory urban flood susceptibility map. The non-flood locations were extracted from this precursory map. The preparatory map and the non-flood locations are shown in Fig. 14.5. The calculation of the frequency ratio is given in Table 14.2.

The fuzzy values required to build the Fuzzy MLPNN model were calculated using the minimum–maximum normalization of FR values, as defined in Eq. (14.4). The calculated values are given in Table 14.2. All the flood conditioning factor rasters

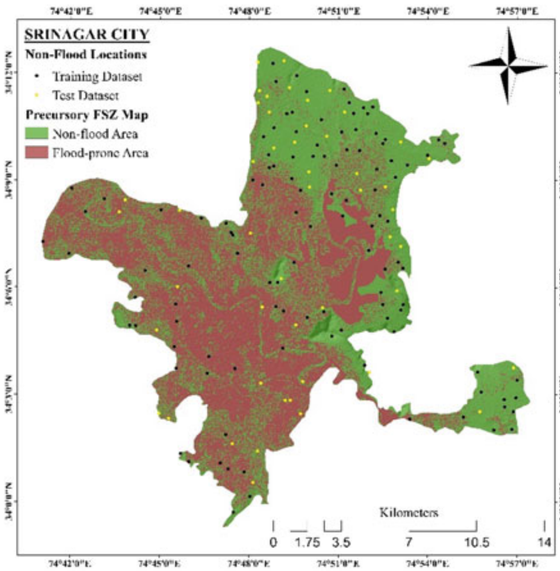


Fig. 14.5 Precursory FSZ map depicting the non-flood locations

Table 14.2 Calculation of frequency ratio, fuzzy values and tabulated conditioning factor importance

Conditioning factor	Class	Total class pixels %age	Class flood pixels %age	Frequency ratio	Fuzzy values	Predictor importance (fuzzy MLPNN)
Elevation	1528–1568.2	76.32	99.12	1.30	1.00	0.01
	1568.2–1608.4	16.83	0.39	0.02	0.02	
	1608.4–1648.6	4.12	0.14	0.03	0.02	
	1648.6–1688.8	1.67	0.14	0.08	0.06	
	1688.8–1729	0.74	0.1	0.13	0.10	
	1729–1769.2	0.21	0.06	0.29	0.22	
	1769.2–1809.4	0.07	0.04	0.57	0.44	
	1809.4–1849.6	0.03	0.02	0.50	0.38	
	1849.6–1889.8	0.01	0	0.00	0.00	
	1889.8–1930	0	0	0.00	0.00	
Slope	0–1.00	16.5	17.87	1.08	0.90	0.00
	1.00–2.26	18.11	20.46	1.13	1.00	
	2.26–3.26	22.47	24.78	1.10	0.94	
	3.26–4.26	13.99	14.31	1.02	0.78	
	4.26–6.26	15.54	14.31	0.92	0.59	
	6.26–63.95	12.72	7.92	0.62	0.00	
Profile curvature	Convex	32.55	32.08	0.99	0.20	0.41
	Flat	34.3	35.28	1.03	1.00	
	Concave	33.15	32.64	0.98	0.00	
Plan curvature	Concave	22.52	21.21	0.94	0.08	0.17
	Flat	54.58	57.52	1.05	1.00	
	Convex	22.91	21.27	0.93	0.00	
Distance from natural streams	< 100	15.66	17.81	1.14	0.96	0.99
	100–200	14.29	16.57	1.16	1.00	
	200–300	12.63	14.18	1.12	0.93	
	300–400	10.82	12.07	1.12	0.93	
	400–500	9.23	10.08	1.09	0.87	
	500–600	7.73	7.98	1.03	0.76	
	600–700	6.51	5.95	0.91	0.54	
	700–800	5.44	4.53	0.83	0.39	
	> 800	17.61	10.87	0.62	0.00	
GEOLOGY	Alluvium/Scree/Talus	63.42	86.4	1.36	1.00	0.08
	Panjal Vol	1.29	0.55	0.42	0.31	

(continued)

Table 14.2 (continued)

Conditioning factor	Class	Total class pixels %age	Class flood pixels %age	Frequency ratio	Fuzzy values	Predictor importance (fuzzy MLPNN)
	Karewa	27.8	6.36	0.23	0.17	
	Water body	5.7	5.44	0.95	0.70	
	Aggl State	0.01	0	0.00	0.00	
	Swamp	1.7	1.3	0.77	0.57	
MFI	60.10–62.68	12.21	1.24	0.10	0.00	1.00
	62.68–63.36	12.73	8.48	0.67	0.27	
	63.36–63.68	12.81	7.76	0.61	0.24	
	63.68–63.70	3.43	1.22	0.35	0.12	
	63.70–63.84	31.56	33.07	1.05	0.44	
	63.84–64.25	9.68	20.18	2.08	0.93	
	64.25–65.14	8.92	19.93	2.24	1.00	
	65.14–67.40	8.67	8.11	0.94	0.39	
LULC	Unclassified	0	0	1.25	0.95	0.15
	Forest	9.23	7.48	0.81	0.23	
	Green areas	19.14	18.55	0.97	0.49	
	Barren	22.04	14.91	0.68	0.02	
	Built-up	42.43	54.41	1.28	1.00	
	Water bodies	7.04	4.69	0.67	0.00	

were reclassified according to these values in ArcGIS 10.7 (ESRI). The training and test point datasets for training the MLPNN were prepared using these using the reclassified raster layers in ArcGIS 10.7@ESRI in vector format.

14.4.1 Flood Susceptibility Modelling

MPLNN was trained in SPSS Modeller with the prepared training dataset, consisting of the flood and non-flood location-specific data from Fuzzy layers. The trained model was calibrated by the determination of the number of neurons in the hidden layer using the hit-and-trial method. When the number of hidden layer neurons was set at 4, the algorithm yielded a maximum classification accuracy of 84.6%. An increase or decrease in the number of neurons leads to a reduction in the efficiency of the model. This Fuzzy MLPNN model was applied to the deduced susceptibility of Srinagar City to urban flood hazard. It predicted the susceptibility of an entire region consisting of 1,499,390 pixels on the basis of the susceptibility scale that is developed using the training data. The Urban Flood Susceptibility Zonation Map of Srinagar City

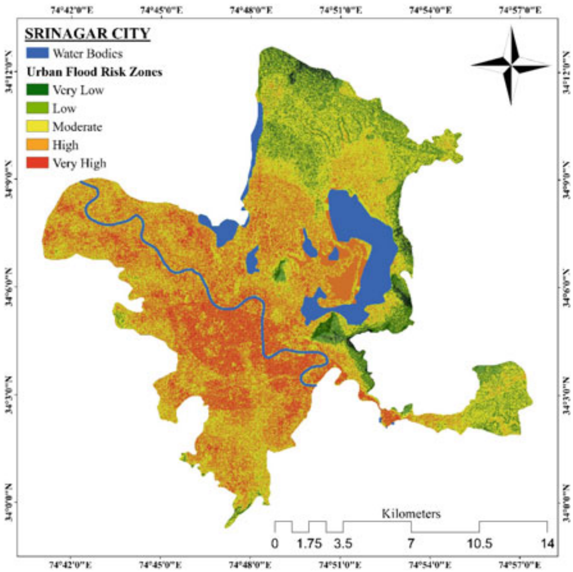


Fig. 14.6 Urban flood risk zonation of srinagar using fuzzy MLPNN

produced by using the Fuzzy MLPNN methodology is given in Fig. 14.6. It depicts the susceptibility of the region into five classes, i.e. very low*, low, moderate, high* and *very high,* by the quantile method.

14.4.2 Role of the Flood Conditioning Factors

Eight flood conditioning factors, including elevation, slope, profile curvature, plan curvature, geology, distance from natural streams, MFI and LULC, were used in this case study. According to the frequency ratio statistics, regions of Srinagar characterized by elevation < 1568.2 m, slope < 4.26°, flat and concave curvature, distance from natural streams 600 m, alluvium, slate and talus geology, MFI > 63.70 mm and built-up land use classes have a positive association (high susceptibility) with the occurrence of urban flooding.

The role of flood conditioning factors in the genesis and modification of urban floods in Srinagar, according to the Fuzzy MLPNN model, can be understood from predictor importance ranking as shown in Fig. 14.7. MFI indicates that rainfall is the most dominant flood conditioning factor that induces urban flooding in the region. Drainage overflow is a secondary cause of urban flooding, while the remaining influencing factors determine the locational characteristics of water logging. Although elevation and slope are significant determinants of the spatial aspect of water logging on a larger geographic scale, curvature takes precedence as a more significant flood conditioning factor in this case study's smaller study area. Accordingly, the Fuzzy

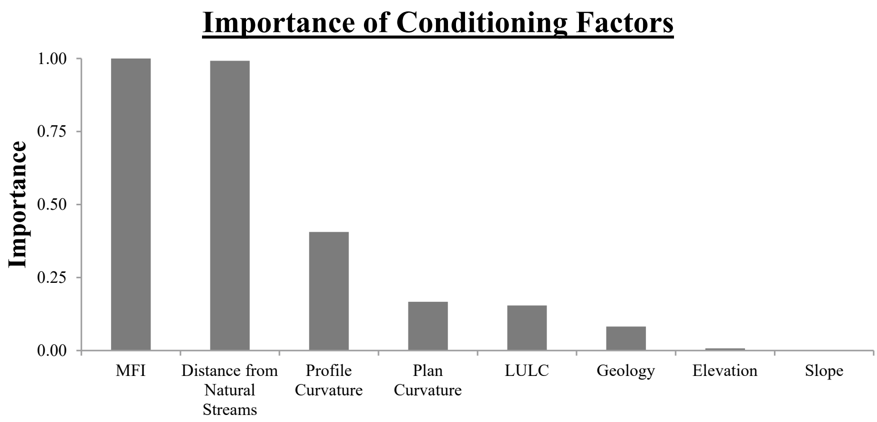


Fig. 14.7 Predictor importance ranking

MLPNN classifies profile curvature and plan curvature as significant factors, indicating whether the location is concave (hence water accumulating tendency) or convex (hence water dispersing tendency) in the direction or perpendicular to the direction of the general slope, respectively.

14.4.3 Map Validation by AUC Analysis

Generally, predictive FSZ maps generated using geospatial modelling are validated by producing their ROC curves in ArcGIS 10.7@ESRI using the ROC tool included in the Spatial Data Modeller (ArcSDM) toolbox and comparing the resulting AUC values to the AUC scale, as previously cited. The success rate of the FSZ map was determined in this study using a training dataset of 114 flood locations and 114 non-flood locations, while the predictive performance of the FSZ map was determined using a test dataset of 49 flood locations and 49 non-flood locations. The findings of the accuracy assessment using AUC values (as shown in Fig. 14.8) indicate that all maps have an *excellent* success rate and prediction rate. The Fuzzy MLPNN map achieves a high rate of accuracy (AUC = 0.931) and a high rate of prediction (AUC = 0.922).

14.5 Discussion

The objective of this case study was to estimate the urban flood susceptibility of Srinagar City, Jammu and Kashmir, India, using a novel Fuzzy MLPNN methodology. The rationale for developing and using such modelling approaches stems

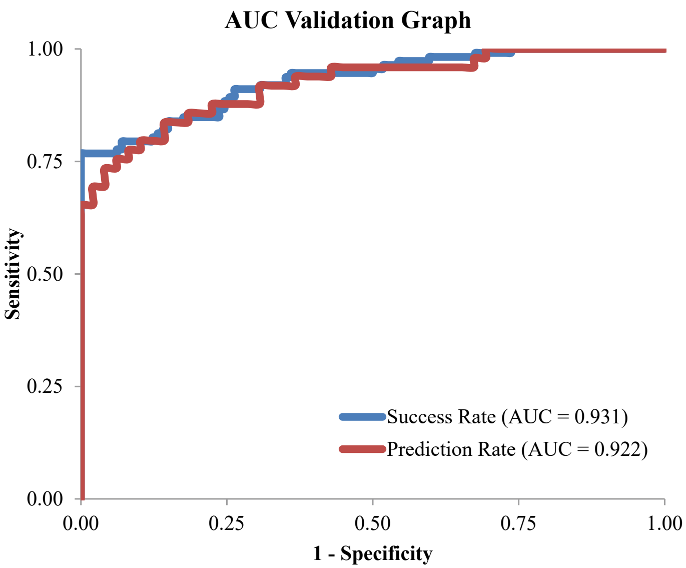


Fig. 14.8 Map Validation using AUC analysis

from the reality that climate change is increasing the occurrence of sudden extreme precipitation events across the globe, thereby endangering natural and human space through increased flood hazard occurrences. Fuzzy MLPNN integrates Fuzzy Logic and Multi-layer Perceptron Neural Network as a predictive methodology for hazard mapping. The model is trained using geospatial data and then applied to the entire study area to create the urban flood susceptibility zones. The advantage of this model is that it is simple to prepare, train and apply in the case of hazard studies.

The flood hazard map of Srinagar city produced by the application of Fuzzy MLPNN, validated using the AUC analysis as an *excellent* urban flood model, reveals that the region is sufficiently susceptible to urban floods under heavy precipitation and due to the backwater effects of the Jhelum River. As the map indicates, the most susceptible regions of the city are those where MFI is high and proximity to natural drainage is low.

Urban flooding is a complex phenomenon that needs multi-dimensional analysis. Numerous environmental and anthropologic parameters play specific roles in the causation, intensification and geographical distribution of urban floods. Elevation, slope, profile curvature, plan curvature, geology, distance from natural streams, MFI and LULC were used in this case study. The Fuzzy MLPNN model also explains the role of the flood conditioning factors on an ordinal scale. Generally, it is understood that the factors controlling the permeability, i.e. LULC, Geology, etc., are the basic causes of urban floods, but in this study, the designed model has shown that, in the first place, it is water sources (precipitation and drainage) that bring about the problem, followed by the effective role of topographic curvature in guiding the accumulation.

14.6 Conclusion

This case study explains the general procedure for urban flood modelling and applies the novel Fuzzy MLPNN methodology for urban flood modelling in Srinagar City, Jammu and Kashmir, India. The approach is simple and straightforward. It unifies the complexity of the phenomenon of urban flooding and can be replicated in other urban flood studies. Eight flood conditioning factors (elevation, slope, profile curvature, plan curvature, geology, distance from natural streams, MFI and LULC) are used as independent variables, with water-logging locations as the dependent variable. Fuzzy MLPNN is a machine learning algorithm that also requires non-flooded location data for training purposes. A precursory FSZ map of Srinagar City was created using the frequency ratio technique, and non-flooded locations were accordingly determined. The Fuzzy MLPNN model developed for and applied to Srinagar City yielded the urban flood susceptibility index for each and every pixel (12.5×12.5 m) in the study area. The FSI, represented by the FSZ Map of Srinagar city, reveals the significant susceptibility of the region to urban flood hazard, and the spatial distribution of the susceptibility is demonstrated. As the map indicates, the most susceptible regions of the city are those where MFI is high and proximity to natural drainage is low. The outcome proves to be a substantial basis from which the state can be able to formulate suitable preventive strategies.

References

Arnoldus HMJ (1980) An approximation of the rainfall factor in the Universal Soil Loss Equation, pp 127–132

Bhat MS, Ahmad B, Alam A, Farooq H, Ahmad S (2019) Flood hazard assessment of the Kashmir Valley using historical hydrology. J Flood Risk Manage 12:e12521

Bui DT, Tsnagaratos P, Ngo P, Pham TD, Pham TB (2019) Flash Flood Susceptibility modelling using an optimized fuzzy rule-based feature selection technique and tree based ensemble methods. Sci Total Environ 668:1038–1054

Hong H, Ilia I, Tsangaratos P, Chen W, Xu C (2017) A hybrid fuzzy weight of evidence method in landslide susceptibility analysis on the Wuyuan area, China. Geomorphology 290:1–16

König A, Sægrov S, Schilling W (2002) Damage assessment for urban flooding. In: Global solutions for urban drainage, pp 1–11

Meng XH, Huang YX, Rao DP, Zhang Q, Liu Q (2013) Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. Kaohsiung J Med Sci. 29(2):93–99.

Pradhan AMS, Kim YT (2017) Spatial data analysis and application of evidential belief functions to shallow landslide susceptibility mapping at Mt. Umyeon, Seoul, Korea. Bull Eng Geol Environ 76(4):1263–1279s

Rahmati O, Pourghasemi HR, Zeinivand H (2016) Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. Geocarto Int 31(1):42–70

Ramesh V, Iqbal SS (2020) Urban flood susceptibility zonation mapping using evidential belief function, frequency ratio and fuzzy gamma operator models in GIS: a case study of Greater Mumbai, Maharashtra, India. Geocarto Int

Tehrany MS, Lee MJ, Pradhan B, Jebur MN, Lee S (2014a) Flood susceptibility mapping using integrated bivariate and multivariate statistical models. Environ Earth Sci 72(10):4001–4015

Tehrany MS, Pradhan B, Jebur MN (2014b) Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. J Hydrol 512:332–343

Zadeh LA (1965) Zadeh, fuzzy sets. Inform Control 8:338–353

Chapter 15
Case Study 7: Assessment, Mapping and Prediction of Urban Heat Island in Srinagar City Region

Abstract Urban Heat Island (UHI) refers to the manifestation of relatively higher surface and air temperatures in the urban centres with respect to the immediate surroundings where land use/land cover is generally green in nature (forest or agriculture). It is a local climate system in which the temperatures are observed to follow a slope in the direction radially outwards from the city centre. In this study, UHI of the Srinagar City Region has been mapped along the time series from 1991 to 2020 and the scenario has been predicted also for 2030. Land Surface Temperature (LST), retrieved from the thermal bands of Landsat 5 images for 1991, 1999 and 2010, and from Landsat 8 images of 2020 using Mono-window (MW) Algorithm was used to understand the temperature trend and the evolution of UHI zones in the Srinagar City region. The results show that the mean surface temperature of the study area has shifted from 16.04 °C in 1991 to 26.21 °C in 2020. In the same time period, the area of UHI zone has grown at a rate of 2.85 km^2 per year from 1991 to 2020. It expanded from 13.57 km^2 (1.82% of total area) in 1991 to 96.18 km^2 (12.90% of total area) in 2020. According to Multi-layer Perceptron Neural Network (MLPNN), which was used to predict the mean surface temperature, Srinagar City Region will experience 27.21 °C in 2030. This study also predicted the potential scenario of UHI zones for 2030 using Cellular Automata-Markov Chain Integrated Model (CA-Markov). It forecasts that in 2030, with an expansion rate of 12.57 km^2 per year, UHI zone of Srinagar City Region will expand to 221.94 km^2 covering 29.77% of the total area, if the present scenarios continue.

Keywords Urban heat island (UHI) mapping · Land surface temperature (LST) · Mono-window algorithm · Cellular automata-Markov chain integrated model (CA-Markov) · Multi-layer perceptron neural network (MLPNN)

15.1 Introduction

A metropolitan area that experiences much warmer air and surface temperatures than its surroundings is referred to as an urban heat island (UHI). The urban growth modifies the heat balance at the local level, and the City Center tends to encounter high temperatures, and in the thermal cross-section of a city, it would appear that

M. Kumar et al., *Geographic Information Systems in Urban Planning and Management*, Advances in Geographical and Environmental Sciences,
https://doi.org/10.1007/978-981-19-7855-5_15

the temperature decreases radially from the city centre to periphery. UHI is also characterized as closed isotherms showing a relatively warm region at local level (or at regional level) caused by human influence on natural climatology. The development of urban heat islands (UHI) is one of the most prevalent manifestations of local anthropogenic climate interference (de Faria Peres et al. 2018). Land use and land cover (LULC) changes play a significant part in the formation of urban heat islands, as a result of increased population and infrastructural demands in rapidly expanding megacities, reduce the green spaces and water bodies and increase concrete impervious surfaces (Mohan et al. 2012). According to Imhoff et al. (2010), the UHI phenomenon is related with changes in surface, where the drainage system rapidly takes the majority of rainwater from impermeable regions, and as a result, only a small percentage of net radiation is used for evaporation (latent heat flux) while the majority of available radiation is used to warm the land surface and air directly above it (sensible heat flux). Also, construction of high-rise buildings inhibits the out-radiation of heat from the hot surfaces during the nights as trap the radiation. Urban construction materials have different thermal capacities and conductivities; building structure and configurations trap radiation and effluents and create an inhibition cover surface that affects flow of air and diffusion of heat; engineering structures strip away surface water and change flow regimes and humidity (de Faria Peres et al. 2018).

Mapping of the Urban Heat Islands is a complex process. There is paucity of adequate data, and on-ground methods of data collection are not efficient in terms of coverage, cost and time. Remote sensing (RS) offers the advantageous data for analysing land surface temperature (LST) and UHI phenomenon on multi-temporal scale and comprehensively for each and every point of the area under investigation. Thermal satellite sensors deliver detailed temperature data for large cities and beyond (Nichol and To 2012). In this case study, attempt has been made to map the present-time UHI zones of Srinagar City Region using thermal remote sensing data products and by adopting the methodologies based on LST thresholding. Also, this study aims to predict the UHI zone extents and the mean surface temperature in the study area by 2030 using Cellular Automata-Markov Chain Integrated (CA-Markov) Model and Multi-layer Perceptron Neural Network (MLPNN), respectively.

15.2 Overview of the Study Area

Srinagar City Region can be defined as the area that includes the Central City and the peripheral region (including outgrowth and natural landscape). It is located between the latitudinal extents of 33° 57′ 46″ N and 34° 14′ 26″ N and the longitudinal extents of 74° 39′ 26″ E and 74° 59′ 27″ E. It is the administrative capital city of the Union Territory of Jammu and Kashmir, India. According to 2011 Census figures, there are 108 major cities and towns in the Union Territory of Jammu and Kashmir of which Srinagar is the primate city and the population of Srinagar City Region is about 12.06 lakh. The total area of the Srinagar City Region (study area as shown in

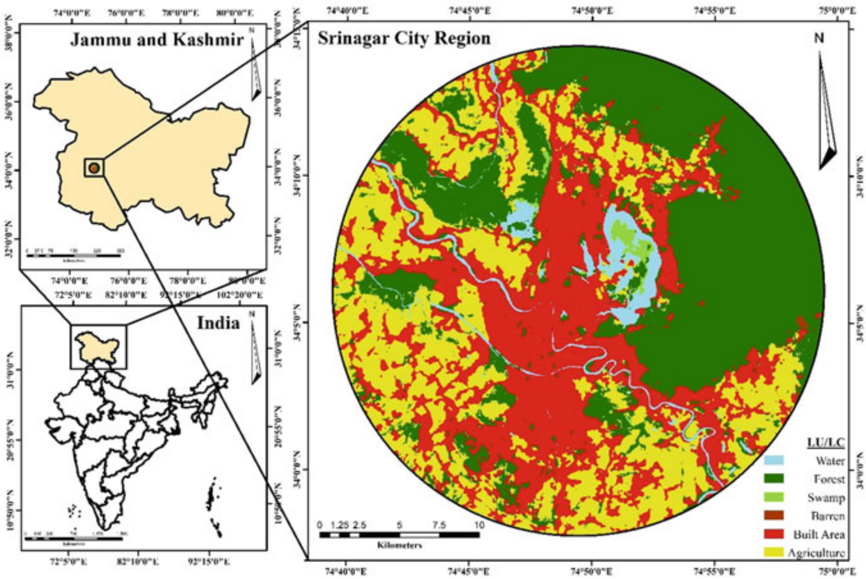


Fig. 15.1 Study area (Srinagar city region)

Fig. 15.1) is 745.02 km^2 of which 36.04% (268.50 km^2) is forest, 31.50% (234.67 km^2) is built-up, 28.11% (209.45 km^2) is agriculture, 3.61% (26.89 km^2) is water bodies, 0.61% (4.55 km^2) and 0.13% (0.96 km^2) is barren. Urban built-up is the second most dominant land use category after the forest cover, and it is expanding at a rapid rate. The high concentration of population and cumulative urban development of region can be attributed to the administrative status of the city and to magnetic effect of the city on working population and the student population. Owing to this, central city of the study area has lately experienced fast increase in the temperature as compared to the rest of the Jhelum Valley. Therefore, it is pertinent to undertake the UHI assessment, mapping and prediction analysis of urban heat island effect in the Srinagar City Region.

15.3 Data and Methods

The methodology of this case study is divided into two broad sections: (1) urban heat island mapping and assessment and (2) urban heat island prediction. The methodology is executed in ArcGIS 10.7@ESRI, IDRISI-TerrSet Geospatial Monitoring and Modelling System software and WEKA 3.8.6.

15.3.1 Urban Heat Island Mapping and Assessment

For urban heat island mapping and assessment, freely available Landsat images of four years (1991, 1999, 2010 and 2020) spanned across 30 years at a rough interval of 10 years are used for Land Surface Temperature (LST) retrieval and demarcation of the Urban Heat Island (UHI) zone in the Srinagar City Region. These images were downloaded from United States Geological Survey (USGS) Earth Explorer, and their characteristics are detailed in Table 15.1.

Retrieving LST can generally be accomplished through the use of established mathematic algorithms like mono-window (MW), split-window (SW), single channel (SC) and multi-angle (MA) algorithms (Rongali et al. 2018**).** In this study, LST of Srinagar City Region is retrieved from the Landsat Images using the Mono-Window (MW) Algorithm. It relies chiefly on the thermal bands (bandwidth = 10.40 − 12.50 µm) of satellite imagery. The procedure for the LST retrieval starts with the conversion of the Digital Number (DN) Values of satellite image into the spectral radiance values using Eq. (15.1).

$$L_\lambda = M_L Q_{\text{cal}} + A_L \tag{15.1}$$

where L_λ stands for spectral radiance (TOA), M_L is the band-specific multiplicative rescaling factor (RADIANCE_MULT_BAND_Y), Q_{cal} is the quantized and calibrated standard product pixel value (Thermal Band) and A_L is the band-specific additive rescaling factor (RADIANCE_ADD_BAND_Y). Spectral radiance at the sensor aperture is calculated in watts/m^2.ster.µm. This spectral radiance (L_λ) is then converted into the Brightness Temperature using Eq. (2)

$$\text{BT} = \left(\frac{K2}{\ln(K1/L_\lambda + 1)}\right) - 273.15 \tag{15.2}$$

where BT refers to brightness temperature at satellite (in Kelvin) and $K1$ and $K2$ are the band-specific thermal conversion constants (K1_CONSTANT_BAND_Y and K1_CONSTANT_BAND_Y, respectively). Subtraction of 273.15 term, as shown in Eq. (2), transforms the brightness temperature from Kelvin (K) to degree Celsius (°C). Following this, normalized differential vegetation index (NDVI) is calculated using Eq. (15.3), in which NIR and R refer to near-infrared band and red band of the

Table 15.1 Satellite images used in the study

Data	Path/row	Year	Spatial resolution	Source
Landsat 5 TM	149, 036	April, 1991	30	Earth explorer
Landsat 5 TM	149, 036	August, 1999	30	Earth explorer
Landsat 5 TM	149, 036	May, 2010	30	Earth explorer
Landsat 8 OLI/TIRS	149, 036	June, 2021	30	Earth explorer

satellite imagery, respectively.

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - R}{\mathrm{NIR} + R} \tag{15.3}$$

In the next step, Proportional Vegetation Index (P_v) is calculated as minimum–maximum normalization of NDVI, as shown in Eq. (15.4). The P_v is used for the estimation of the land surface emissivity (e) according to Eq. (15.5).

$$P_v = \left(\frac{\mathrm{NDVI} - \mathrm{NDVI}_{\min}}{\mathrm{NDVI}_{\max} - \mathrm{NDVI}_{\min}} \right)^2 \tag{15.4}$$

$$e = 0.004.P_v + 0.986 \tag{15.5}$$

Finally, the Land Surface Temperature is calculated using the brightness temperature (BT) and the land surface emissivity value (e) according to Eq. (15.6).

$$\mathrm{LST} = \frac{\mathrm{BT}}{1 + \left(\lambda_r . \frac{\mathrm{BT}}{p} \right)} . \ln(e) \tag{15.6}$$

where certain additional parameters involved are $\lambda_r = 10.8\mu m$ and $p = h.\frac{c}{s} = 14388$. The calculation LST using Mono-window (MW) algorithm requires the use of various parameters in the equations, and the values of these parameters are specific to the satellite data products (images) and their bands. Table 15.2 lists the major parameters and their specific values used in the equations henceforth.

In this study, LST retrieved from four different Landsat images are used to identify, demarcate and map the Urban Heat Island (UHI) zones across the timeline of 1991 to 2020. The LST raster layers created using Eq. (15.6) used as the base and pixels with the values greater than the mean LST are identified as the potential UHI zones. However, in addition to the urban built-up, topography also controls the distribution of heat in LST raster layers. As such, high surface temperature zones may fall outside the urban zones and it is pertinent to eliminate these zones for accurate demarcation of the actual UHI zones in the study area. Srinagar city is located in the lower elevation (local) plain terrain while the higher elevations are mostly covered with forest or barren lands. The barren lands of the adret slopes also show

Table 15.2 Band-specific parameters and their values for different landsat images

Parameter	Landsat 5 (TM) value	Landsat 5 (OLI/TIRS) value
M_L	0.055375	0.0003342
A_L	1.18243	0.10000
$K1$	607.76	774.8853
$K2$	1260.56	1321.0789

greater land surface temperatures. Accordingly, in this study, the elevation raster of ALOS PALSAR DEM, classified into two classes based on the manually determined threshold elevation, intersects with the potential UHI zone raster layers to map the actual UHI zones of Srinagar.

15.3.2 Urban Heat Island (UHI) Prediction

This study also seeks to estimate the temperature mean and spatial extent of the urban heat island in Srinagar City Region by 2030. Firstly, the mean surface temperature of the study area for 2030 is calculated on the basis of mean surface temperature of 29 years preceding 2020. The mean surface temperature values of 1991, 1999, 2010 and 2020 are interpolated by additive average to yield the mean surface temperature values of each year from 1991 to 2020. The mean surface temperatures for 2030 are then forecasted using Multi-layer Perceptron Neural Network (MLPNN) in WEKA 3.8.6.

Secondly, prediction of the spatial extent of UHI by 2030 requires an effective analysis of historical LST scenarios. Cellular Automata-Markov Chain Integrated Model (CA-Markov) is applied for the purpose in this study. The CA–Markov model has been widely used in many scientific studies to forecast future land use/land cover types (LULC) because it combines the benefits of Cellular Automata and the Markov Chain element of spatial adjacency, as well as knowledge of the likely spatial distribution of transitions (Arsanjani et al. 2013). On one hand, Markov Chain models forecast the spatiotemporal possibility of the future on the basis of historical instances and trend analysis. In other words, it can be said that Markov Chain is simply a series of the random values whose probability for a defined time period is estimated by the previous value of the number (Surabuddin Mondal et al. 2013). The basic principle behind the Markov chain model is that a phenomenon at some point in future can be predicted using past and present scenarios of the same phenomenon (Iacono et al. 2015). Coppedge et al. (2007) support this concept and state that the change in the phenomenon at any location is not completely random, but a function of the conditions of the past and present time. The formula used in the Markov Chain model is defined mathematically in Eq. (15.7).

$$L_{(n+o)} = r_{ij} \times L_n \tag{15.7}$$

where $L_{(n+1)}$ and L_n represent the phenomenon (LULC generally and LST in this study) at time span n and $(n + o)$, respectively, and r_{ij} is defined mathematically in Eq. (15.8).

$$r_{ij} = \begin{bmatrix} r_{11} \; r_{12} \; \ldots \; r_{1p} \\ r_{21} \; r_{22} \; \ldots \; r_{2p} \\ \vdots \; \ldots \; \vdots \\ r_{p1} \; r_{p2} \; r_{pp} \end{bmatrix} \tag{15.8}$$

where $0 \leq r_{ij} \leq 1$ and $\sum_{j+1}^{p} r_{ij} = 1, (i, j = 1, 2, \ldots p)$. is representing the transition probability matrix.

On the other hand, Cellular Automata Model is able to analyse the changing and controlling complex spatially distributed processes. It predicts the values for each cell of the matrix (raster) on the basis of the former conditions of the cells contained within a defined neighbourhood (Rocha et al. 2007). Sang et al. (2011) presented the CA-Markov Chain model in as follows:

$$S(n, n+1) = f(S(n), N) \tag{15.9}$$

Limited and Discrete Cellular States is represented by S, Cellular Field is represented by N, and n and $(n + 1)$ represent the different time interval. The rule of transformation of 'Cellular States' in local space is represented by f.

Cellular Automata-Markov Chain Integrated Model (CA-Markov) is capable of accurately predicting spatiotemporal dynamics of LULC, but in this study we replace the LULC raster layers with the LST raster layers for we aim to predict the spatiotemporal dynamics of surface temperatures. The past LST pattern determined using the Landsat data is used as the base to predict future UHI of Srinagar City Region. LST raster layers of the past and present are classified into suitable number of surface temperature classes to depict the corresponding on-ground zones. The LST values for future are predicted by executing CA-Markov in IDRISI-TerrSet Geospatial Monitoring and Modelling System software. The transition probabilities controlled by local rules supported the CA-Markov's prediction of future LST for Srinagar City Region. Three types of data are required for CA-Markov modelling process: firstly, the basis land cover image; secondly, Markov transition area file generated from Markov modelling and, thirdly, the transition suitability image collection generated from the Markov modelling.

15.4 Results and Discussion

This study aims to identify and demarcate the Urban Heat Islands in the Srinagar City Region using the Land Surface Temperature (LST). LST was obtained for four images of 1991, 1999, 2010 and 2020, using Eq. (15.6). The LST of these four times is given in Fig. 15.2, which shows the change in the surface temperatures.

The LST is classified into six classes, and it is vivid that the study region shows the dominant prevalence of 15–20 °C in 1991, 20–25 °C in 1999 and 2010 and 25–30 °C

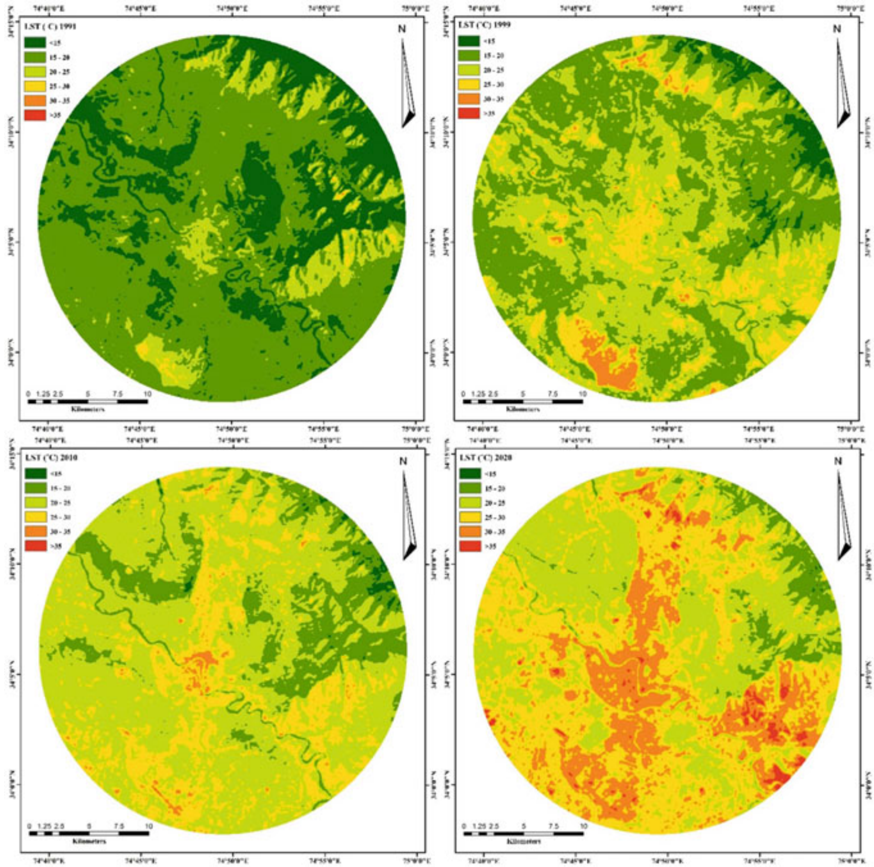


Fig. 15.2 Land surface temperature (LST) of Srinagar city region in 1991, 1999, 2010 and 2020

in 2020. This is also depicted in Table 15.3 and validated by mean LST values of the respective years, given in Table 15.4.

Table 15.3 Area in different LST ranges (year-wise)

LST range	1991	1999	2010	2020
< 15 °C	185.46	23.76	5.40	0.22
15–20 °C	485.86	305.80	123.60	39.33
20–25 °C	69.64	317.53	453.51	262.52
25–30 °C	4.59	85.35	156.13	297.74
30–35 °C	0.05	13.15	6.94	136.45
> 35 °C	0.00	0.00	0.02	9.33

Table 15.4 LST statistics (year-wise)

Year	Min surface temperature (°C)	Max surface temperature (°C)	Mean surface temperature (°C)	Standard deviation
1991	−8.16	33.18	16.04	3.90
1999	5.38	34.78	20.99	3.59
2010	4.36	35.97	22.88	3.10
2020	13.20	40.01	26.21	4.02

Figure 15.2 also shows that in 1991, the surface temperatures greater than mean (16.04 °C) are found along the Srinagar City's built-up (central region of the study area) and along the south-facing adret slopes of local mountains in 1991. These regions majorly have temperatures in the range of 20–25 °C. In 1999, the same regions again show surface temperatures greater than the mean (20.99 °C) but the surface temperature range of these regions shifts to 25–30 °C. The trend is followed in 2010 and 2020 also as the surface temperature range shifts to 30–35 °C. The areas of LST raster layers with surface temperatures greater than the mean category were extracted as the potential UHI zone raster layers.

The control of urban built-up and topography is simultaneously shown in the time series LST maps of Fig. 15.2. But in this study, as we primarily focus on the Urban Heat Island (UHI), spatial distinction needs to be made between the influence of topography and built-up on LST. To eliminate the high surface temperatures resulting from topography and to demarcate the actual UHI in the study area, intersection of potential UHI raster and the binary elevation raster was executed using the *Raster Calculator*. The elevation raster was classified into two classes with 1575 m (amsl) as threshold for the built-up is majorly located in the elevation range of 1500–1575 m. The UHI zones extracted from four time periods of 1991, 1999, 2010 and 2020 are shown in Fig. 15.3.

The urban heat island (UHI) area in Srinagar City region has grown from 13.57 km^2 in 1991 to 96.18 km^2 in 2020. UHI expansion has taken place at the rate of 2.85 km^2 per year from 1991 to 2020 in Srinagar City Region. In terms of percentage of total area covered, it has changed from 1.82% in 1991 to 12.90% in 2020 (Fig. 15.4; Table 15.5).

The prediction of the mean surface temperature of Srinagar City Region for 2030 by Multi-layer Perceptron Neural Network (MLPNN) was based on the mean surface temperatures of 1991, 1999, 2010 and 2020 given in Table 15.4. The mean surface temperature values of the years between 1991 and 1999, 1999 and 2010, and 2010 and 2020 were interpolated by additive average method and are given in Tables 15.6, 15.7 and 15.8, respectively. The mean surface temperatures of 29 years were input into WEKA 3.8.6, which predicted the mean surface temperature for each year that matched with the calculated mean surface temperatures of respective years. The Multi-layer Perceptron Neural Network (MLPNN) predicted that the mean surface temperature of Srinagar City Region will rise from 26.21˚C in 2020 to 27.21˚C in 2030. This is shown graphically in Fig. 15.5 (Table 15.9).

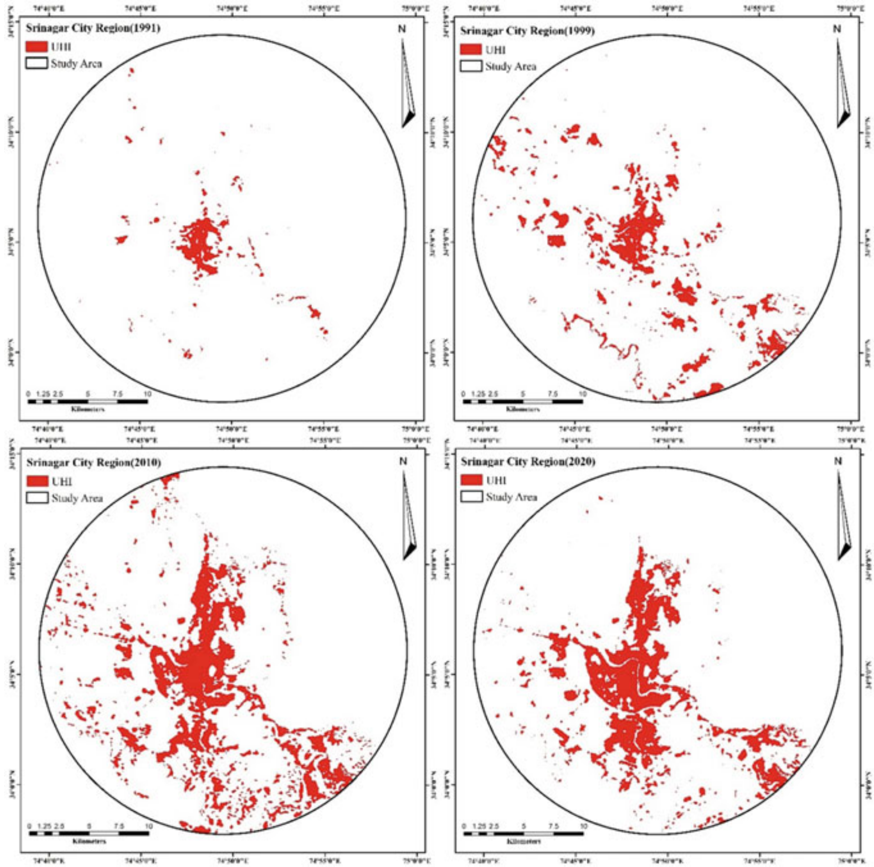


Fig. 15.3 Urban heat island (UHI) zones of Srinagar city region in 1991, 1999, 2010 and 2020

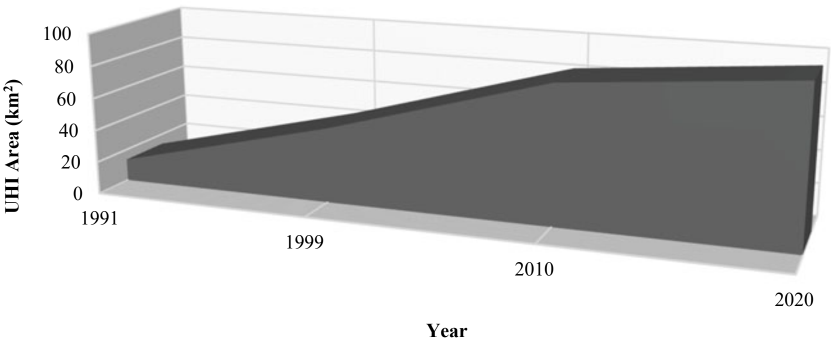


Fig. 15.4 Urban heat island (UHI) zone expansion in Srinagar city region from 1991 to 2020

Table 15.5 UHI area statistics (year-wise)

Year	UHI area (in km^2)	Percentage of total
1991	13.57	1.82
1999	46.04	6.17
2010	84.57	11.34
2020	96.18	12.90

Table 15.6 Calculated mean LST (1991–1998)

Year	1991	1992	1993	1994	1995	1996	1997	1998
Mean LST (°C)	16.04	16.66	17.28	17.90	18.51	619.13	19.75	20.38

Table 15.7 Calculated mean LST (1999–2009)

Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Mean LST (°C)	20.99	21.16	21.33	21.50	21.67	21.84	22.02	22.19	22.36	22.53	22.70

Table 15.8 Calculated mean LST (2010–2019)

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Mean LST (°C)	22.8	23.21	23.54	23.87	24.21	24.54	24.87	25.21	25.54	25.87

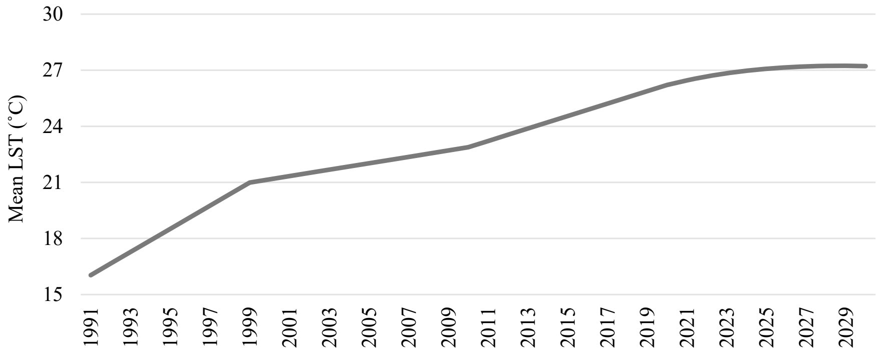


Fig. 15.5 Mean LST trend in Srinagar city region from 1991 to 2020 and predicted mean LST from 2021 to 2030 using multi-layer perceptron neural network

Table 15.9 Predicted mean LST (2020–2030)

Year	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
Mean LST (°C)	26.21	26.45	26.66	26.83	26.96	27.07	27.14	27.20	27.22	27.23	27.21

Cellular Automata-Markov Chain Integrated Model (CA-Markov) was sued to predict the spatial extent of the UHI in the Srinagar City Region for 2030. The LST raster layers of 2010 and 2020 were classified into six classes showing surface temperatures in the ranges of < 15 °C, 15–20 °C, 20–25 °C, 25–30 °C, 30–35 °C and > 35 °C. 2010 LST raster was used as earlier LST image, and 2020 was used as later LST image. The LST predicted by the CA-Markov for 2030 is shown in Fig. 15.6. It depicts that the surface temperatures of a major portion of the study area will rise and contribute to urban heat. It predicts that area of the Srinagar City Region experiencing different surface temperatures will be 0.01 km^2 in < 15 °C category, 9.71 km^2 in 15–20 °C category, 122.76 km^2 in 20–25 °C category, 244.81 km^2 in 25–30 °C category, 317.97 km^2 in 30–35 °C category and 49.82 km^2 > 35 °C category. Evidently greater surface temperatures (30 °C or more) will be observed in extensive portions of the study area.

The potential UHI zones of 2030 were again identified as those areas of Srinagar City Region, where predicted surface temperatures are greater than the predicated mean surface temperature (27.21). This potential UHI raster was extracted and intersected with the binary elevation raster (threshold = 1575 m) to map out the actual UHI zones of 2030, which is shown in Fig. 15.7. It reveals that urban heat island will cover 221.94 km^2 (29.77%) area of Srinagar City Region by 2030. The rate of expansion of UHI which was 2.85 km^2 per year will accelerate to 12.57 km^2 per year, and it can be predicted further that if this predicted rate continues till and after 2030, then the surface temperatures of Srinagar City Region will rapidly degrade.

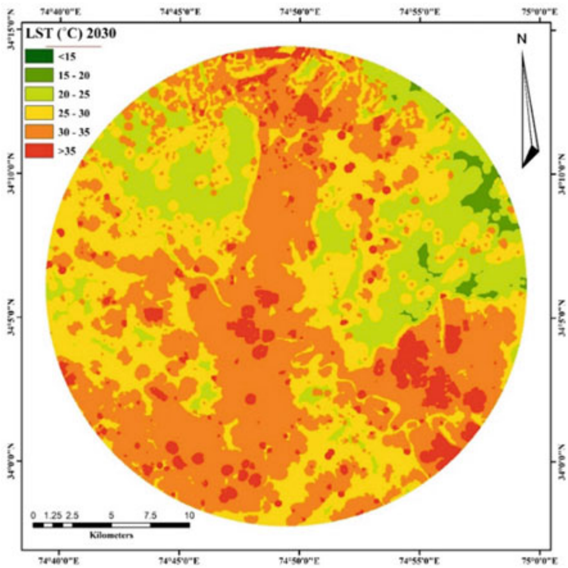


Fig. 15.6 Predicted LST of Srinagar city region for 2030

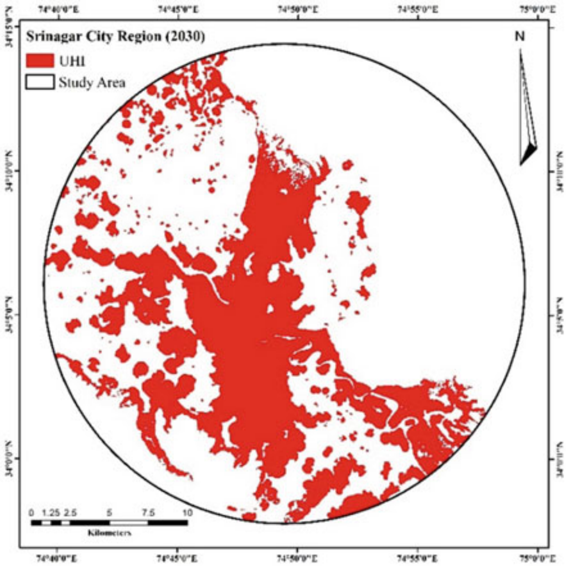


Fig. 15.7 Predicted UHI Zones of Srinagar city region for 2030

15.5 Conclusion

Urban Heat Island (UHI) effect is aggravating the environmental quality in the cities across the globe. Mapping of the UHI zones is a complex process as it is easy to collect temperature data across an urban region and its periphery. The data specific to well-distributed point locations of the study region can be collected manually using suitable instrumentation and interpolated for the entire region to generate the continuous temperature map that can further be used for demarcation of the UHI zones. But the time and cost of such a mapping project will render it unfeasible. However remote sensing data can be taken advantage of in the matter. To identify and demarcate these zones, the base maps of surface temperature can be created easily using the Land Surface Temperature (LST), over which one can determine the thresholds of delineating the UHI zones. In this study, mapping of the UHI zones in Srinagar City Region and assessment of their spatiotemporal dynamics has been carried out for the time period of 1991, 1999, 2010 and 2020. Mono-window (MW) Algorithm was used to retrieve the LST from the thermal bands of Landsat 5 and Landsat 8 imagery. Remote sensing data-based assessment of the LST, and consequently the UHI zones, also offers the advantage of predicting the future scenarios at pixel level. In this study, it is predicted that along with the rapid rise in the mean surface temperatures of Srinagar City Region, UHI zones are going to expand across the major portion of the region. It is pertinent to suggest here that efforts need to be taken in the right direction to control the prevailing trends of rising mean surface temperatures and to control the expanding nature of the UHI zones of Srinagar City Region.

References

Arsanjani JJ, Helbich M, Kainz W, Boloorani AD (2013) Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. Int J Appl Earth Obs Geoinf 21:265–275

Coppedge BR, Engle DM, Fuhlendorf SD (2007) Markov models of land cover dynamics in a southern Great Plains grassland region. Landscape Ecol 22(9):1383–1393

de Faria Peres L, de Lucena AJ, Rotunno Filho OC, de Almeida Franca JR (2018) The urban heat island in Rio de Janeiro, Brazil, in the last 30 years using remote sensing data. Int J Appl Earth Obs Geoinf 64:104–116

Iacono M, Levinson D, El-Geneidy A, Wasfi R (2015) A Markov chain model of land use change. TEMA J Land Use Mobil Environ 8(3):263–276

Imhoff ML, Zhang P, Wolfe RE, Bounoua L (2010) Remote sensing of the urban heat island effect across biomes in the continental USA. Remote Sens Environ 114(3):504–513

Mohan M, Kikegawa Y, Gurjar BR, Bhati S, Kandya A, Ogawa K (2012) Urban heat island assessment for a tropical urban airshed in India

Nichol JE, To PH (2012) Temporal characteristics of thermal satellite images for urban heat stress and heat island mapping. ISPRS J Photogramm Remote Sens 74:153–162

Rocha J, Ferreira JC, Simoes J, Tenedório JA (2007) Modelling coastal and land use evolution patterns through neural network and cellular automata integration. J Coastal Res 827–831

Rongali G, Keshari AK, Gosain AK, Khosa R (2018) A mono-window algorithm for land surface temperature estimation from Landsat 8 thermal infrared sensor data: a case study of the Beas River Basin, India. Pertanika J Sci Technol 26(2):829–840

Sang L, Zhang C, Yang J, Zhu D, Yun W (2011) Simulation of land use spatial pattern of towns and villages based on CA–Markov model. Math Comput Model 54(3–4):938–943

Surabuddin Mondal M, Sharma N, Kappas M, Garg PK (2013) Modeling of spatio-temporal dynamics of land use and land cover in a part of Brahmaputra River basin using Geoinformatic techniques. Geocarto Int 28(7):632–656