The Potential of Remote Sensing Data for Water Budget Evaluation in Arid Areas of Sinai, Egypt

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Mona Ahmad Mahmoud Morsy

aus Kairo, Ägypten

Tübingen

2020

Gedruckt mit Genehmigung der Mathematisch- Naturwissenschaftlichen
Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation	22.4.2020
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Peter Dietrich
2. Berichterstatter:	Prof. Dr. Thomas Scholten

To my late Dad,

Who inspired me to follow his direction

ACKNOWLEDGMENTS

I wish to express my gratitude to Prof. Dr. Peter Dietrich, the head of the Department of Monitoring-und Erkundungstechnologien at the UFZ, Leipzig and the main supervisor of this work, who supported me with his invaluable encouragement, continuous advice, follow-up, and generous financial support.

I wish also to express my sincere appreciation to Prof. Dr. Thomas Scholten, professor of Soil Science and Geomorphology at Eberhard Karls Universität Tübingen, for his supervision and continuous advice and assistance during the progress of this research, as well as for the ability to complete my work at the institute.

Special thanks to Dr. Karsten Schmidt and Dr. Ruhollah Taghizadeh-Mehrjardi from the Geography Department at Eberhard Karls Universität Tübingen, who offered continuous help, support, and advice during the final stages of my work.

I wish to express gratefulness to my colleagues and professors at the Geology Department of Suez Canal University, especially Prof. Dr. Mohamed Helmi Geriesh, whose help facilitated my acceptance at Tübingen.

Deepest thanks to Aya El-Wazir, who supported me with all the needed knowledge about RStudio software.

Special thanks to Dr. Mai Abd Allah and Dr. Hadya Adel, who supported me with valuable information during my work progress.

I also wish to express my deepest appreciation for Mrs. Margaretha Baur, who supported me in all aspects of life during my time at the University of Tübingen.

Deepest thanks to all the friends who supported me during this work: Lamia Abdalla, Fatma Refat, Shadia Trrak, Nurgül Düzenli and Katharina Kilian.

Unlimited gratitude for my uncle (Ahmed Hosny), who supported me in gathering all the needed climatic data from the Egyptian Meteorological Authority and provided his continuous support and love.

Deepest thanks to Milena Mihovilović who, added her magical touch by proofreading this work.

Finally, I would like to thank my beloved husband (Ahmed El-Masry), my son (Omar El-Masry), my lovely and supportive mother (Sohir Hosny), and all other family members for their endless support and encouragement through the thick and thin of the last few years.

ABSTRACT

Water scarcity is a growing concern in arid and semi-arid regions of the world, locations where groundwater is the main source of freshwater. In order to preserve local water budgets, it is critical that accurate climatic data be acquired. Unfortunately, the majority of these arid regions feature a very limited number of rain gauges, reducing the reliability of the data produced. The present study offers a series of steps for overcoming the issue of data scarcity. Once resolved, this could then promote greatly needed hydrological studies on topics such as the spatiotemporal distribution of rainfall, the mitigation of flash floods hazards, or the minimization of soil erosion. These studies would be directly applicable to the study site, which was the eastern side of the Gulf of Suez in the Sinai Peninsula in Egypt.

Data collected by Landsat5 and Landsat8 during the years 2000, 2010, and 2019 were downloaded from Google Earth Engine in order to determine changes in land cover and land use. This allowed the author to answer three main questions: Is the area under study active or dormant? Was there a change in the rate of water consumption over time? If so, was this change in consumption positive or negative?

Following the acquisition of land cover and land use data, the spatiotemporal distribution of rainfall at the study site was analyzed using two types of remote sensors, the TRMM(3B42V7) and the GPM(IMERG). The performances of each were compared to determine the optimal type for utilization in this study.

A final step included the use of DEM file and GPM(IMERG) data to identify the most suitable locations for a new network of rain gauges. Results suggested that, during the period of 2000 to 2019, there was an increase in both vegetation and land development in the areas tested. They further indicated an increase in the rate of water consumption. Data acquired via both remote sensors during light intensity precipitation events produced significant conclusions. However, data recorded during heavy intensity events was not able to achieve this, irrespective of the type of sensor used. Overall, GPM(IMERG) data produced results closer to those of the limited rain gauge records. It was therefore selected for use with the SRTM90×90m DEM file, which aided the author in identifying 31 new locations for rain gauges, together able to produce more accurate and reliable information on the climate at the test site in the future.

The conclusions reached by this study recommend the installation of the abovementioned rain gauge network, allowing the community to properly assess proposed water management procedures (including withdrawal portions and expansion rates in existing settlements) as well as to develop improved warning systems for flash flood hazards.

ZUSAMMENFASSUNG

Wasserknappheit ist ein wachsendes Problem in ariden und semiariden Regionen der Welt, an Orten, an denen Grundwasser die Hauptquelle für Süßwasser ist. Um die lokalen Wasservorkommen zu erhalten, ist es wichtig, dass genaue Klimadaten erfasst werden. Leider verfügen die meisten dieser Trockengebiete über eine sehr begrenzte Anzahl von Niederschlagsmessern, was die Zuverlässigkeit der erzeugten Daten verringert. Die vorliegende Studie bietet eine Reihe von Schritten zur Überwindung des Problems der Datenknappheit. Einmal gelöst, könnte dies dringend benötigte hydrologische Studien zu Themen wie der räumlichen und zeitlichen Verteilung von Niederschlägen, der Minderung der Gefahr von Sturzfluten oder der Minimierung von Bodenerosion fördern. Diese Studien wären direkt auf das Untersuchungsgebiet, die östliche Seite des Golfs von Suez auf der ägyptischen Sinai-Halbinsel, anwendbar.

Die von Landsat5 und Landsat8 in den Jahren 2000, 2010 und 2019 gesammelten Daten wurden von Google Earth heruntergeladen, um Änderungen der Landbedeckung und der Landnutzung zu ermitteln. Dies ermöglichte der Autorin die Beantwortung von drei Hauptfragen: Ist das untersuchte Gebiet aktiv oder ruhend? Hat sich der Wasserverbrauch im Laufe der Zeit verändert? Wenn ja, war diese Veränderung des Verbrauchs positiv oder negativ?

Nach der Erfassung von Landbedeckungs- und Landnutzungsdaten wurde die räumlich-zeitliche Verteilung des Niederschlags am Untersuchungsort mit zwei Arten von Fernsensoren analysiert, dem TRMM(3B42V7) und dem GPM(IMERG). Die Leistungen von jedem wurden verglichen, um den optimale Sensortyp für die Verwendung in dieser Studie zu bestimmen.

Ein letzter Schritt umfasste die Verwendung von DEM-Dateidaten des optimalen Fernsensors, um die am besten geeigneten Standorte für ein neues Netzwerk von Niederschlagsmessern zu ermitteln. Die Ergebnisse lassen darauf schließen, dass im Zeitraum von 2000 bis 2019 sowohl die Vegetation als auch die Landentwicklung in den untersuchten Gebieten zunahm. Sie wiesen ferner auf einen Anstieg des Wasserverbrauchs hin. Daten, die über beide Fernsensoren während
Lichtintensitätniederschlagsereignissen erfasst wurden, führten zu signifikanten Ergebnissen. Daten, die während schwerer Intensitätsereignisse aufgezeichnet wurden, konnten dies jedoch ungeachtet des verwendeten Sensortyps nicht erreichen. Insgesamt lieferten die GPM(IMERG)-Daten Ergebnisse, die denen der Aufzeichnungen der begrenzten Niederschlagsmessgeräte näher kamen. Er wurde daher für die Verwendung mit der SRTM90×90m-DEM-Datei ausgewählt, die der Autorin dabei half, 31 neue Standorte für Niederschlagsmesser zu identifizieren, die in der Lage sind, genauere und zuverlässigere Informationen über das Klima auf dem Testgelände zu erhalten.

In den Schlussfolgerungen dieser Studie wird die Installation des oben genannten Netzwerks von Niederschlagsmessern empfohlen, damit alle die vorgeschlagenen Wassermanagementverfahren (einschließlich der Entnahmeanteile und Expansionsraten in bestehenden Siedlungen) richtig einschätzen und verbesserte Warnsysteme für Sturzflutgefahr entwickeln kann.

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

Arid regions are defined as areas receiving only light and irregular precipitation, with rates falling below those of evaporation. In contrast, semi-arid regions are those receiving a relatively greater amount of precipitation, which can occur for several months out of the year and allow soil moisture to reach levels that can support grass and shrubland (Malagnoux, 2007; Ezzahar *et al*., 2007). As described by UNESCO (1979) and Pilgrim *et al*. (2009), degree of aridity can be determined by calculating the ratio of mean annual precipitation to mean annual evaporation. This degree helps define distinct aridity zones, such as semi-arid, arid, and hyper-arid. Their means can vary considerably, with semi-arid zones presenting a ratio of 0.20-0.50, arid zones a ratio of 0.03-0.20, and hyper-arid falling to 0.03 or below. Regions characterized as arid or semi-arid display greater climate instability and variation than hyper-arid zones. Depending on the season, they can experience both drought and flooding, which leads to environmental disasters as well as severe water shortages that heavily stress local aquifers.

Arid and semi-arid regions represent 30% of the world´s terrestrial area (Scanlon *et al*., 2006; Dregne, 1991) (Figure 1.1). These areas have recently experienced a rapid increase in population density, with over one billion inhabitants globally (Yin *et al*., 2013). This increase has led to higher land cover and usage, pressures that both local governments and international scientific communities are carefully monitoring (Malagnoux, 2007; Scanlon *et al*., 2006).

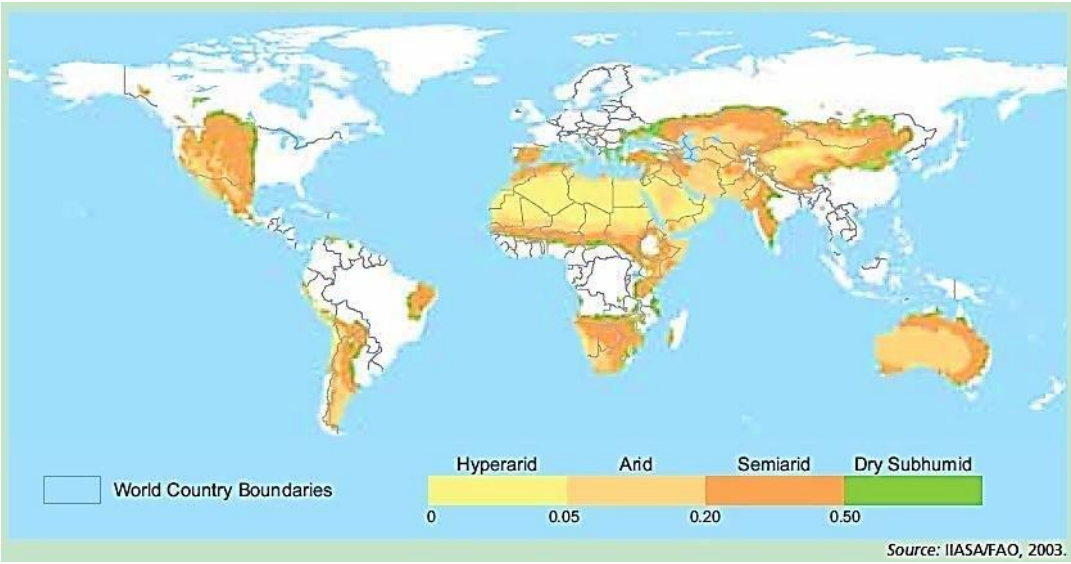


Figure 1.1: Global distribution of arid regions (after Scanlon *et al*., 2006).

1.1 Water Scarcity

Increased freshwater demand has become a growing problem in arid and semi-arid zones. Population growth in these areas has surpassed that of more humid regions (Scanlon *et al*.,

2006), despite local water supply being much lower. In fact, the majority of arid and semi-arid areas worldwide rely mainly on ground water, which is primarily recharged by rainfall (Figure 1.2). This recharge, however, is infrequent and unpredictable, with precipitation occurring only once or twice per year. This insufficient recharge in turn lowers the quality of available ground water and leads to increased salinization.

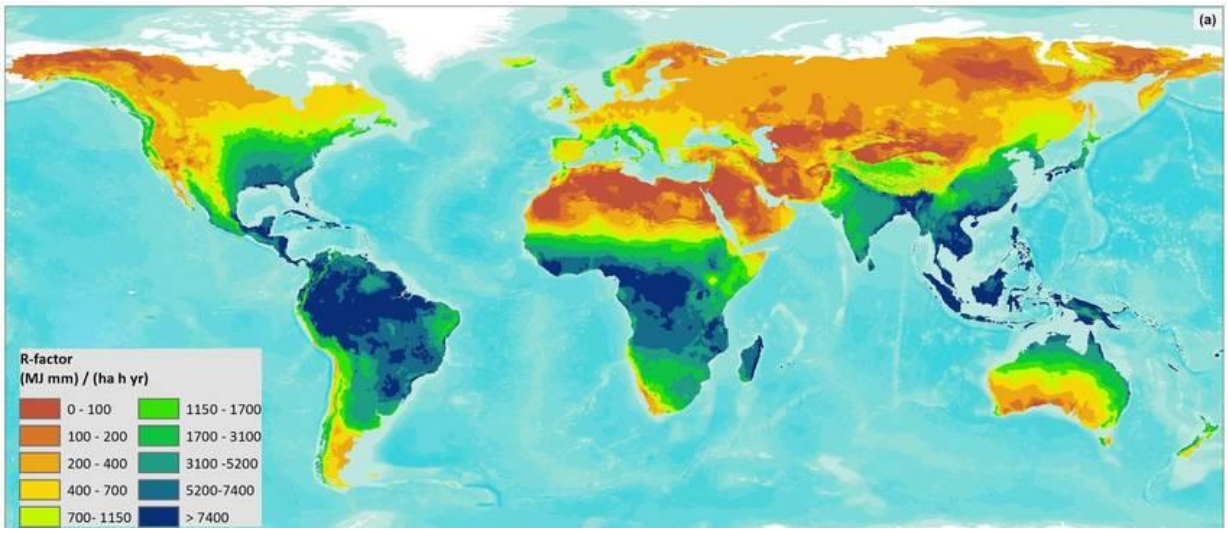


Figure 1.2: Global distribution of annual rainfall (after Panagos *et al*., 2017). Based on a Gaussian interpolation of 3,625 stations in 63 countries.

1.2 Data Scarcity

Data availability is a limiting factor when deriving scientific conclusions of studies on arid and semi-arid regions, with insufficient data reducing the quality of results and leading to misguided decisions and policies. Data influencing these regions can be divided into two groups, both of which can be difficult to estimate. These consist of natural and man-made factors. Natural factors are those influencing the water cycle and directly include: precipitation rate, evapotranspiration, runoff, and infiltration. Indirect factors include temperature, relative humidity, and wind speed (Sherief, 2008). In contrast, man-made factors can describe, for example, water consumption rate, population expansion, land cover increase, and intensification of land use.

Collecting sufficient usable data on the abovementioned factors is critical for sustainable groundwater management in arid and semi-arid zones. However, rain gauges in most mountainous arid regions are few and sparse, if present at all (Pomeon *et al*., 2018, Figure 1.3). Additionally, these existing gauges have a limited capability for capturing continuous records (e.g. hourly changes might not be recorded). Furthermore, the gauges are largely isolated and represent areas of low population density (Pilgrim *et al*., 2009), which results in a low frequency

of maintenance and rapid deterioration. Taken together, these factors significantly reduce the efficacy of water management strategies in arid areas, affecting the water table and general development in the region.

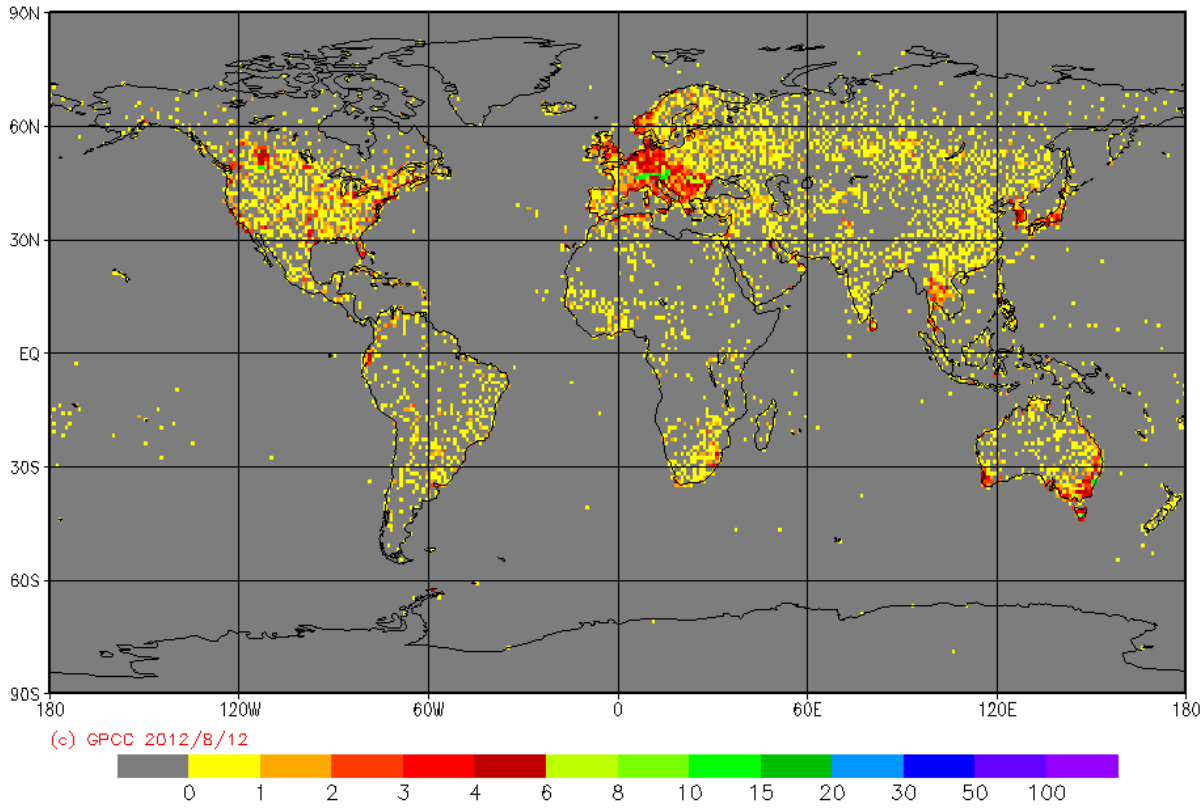


Figure 1.3: Number of rain gauges per 1° grid. (Source: Global Precipitation Climatology Centre, 2018).

1.3 Research Plan

The present study selected the eastern side of the Gulf of Suez as a sample site for the evaluation of the suitability of remote sensing-derived data as alternatives to ground truth data in arid and semi-arid regions. As previously mentioned, data scarcity has typically negatively impacted climatic studies in these regions. Results were intended to serve as a basis for further hydrological assessments, which could then be used for optimal settlement planning and development. Such future studies could utilize these conclusions to better investigate flash flood zones, ground water recharge, spatiotemporal distribution of rainfall, soil erosion, changes in climate, or ground water localization.

The present study featured four main objectives. The first was primarily concerned with the identification of factors affecting water cycle equilibrium at the test site. This was done in order to identify which additional types of data should be collected, as well as to compile a rough overview of the mechanisms at play in the local water cycle. To accomplish this, we

adapted data collected by Sherief (2008) over six-decade period (1934-1989) to examine trends in precipitation, evaporation, and infiltration.

The second objective sought to describe changes in land cover and land use at the test site. This provided a direct impression of critical test site features and allowed us to determine whether, and which, future studies were required. This objective examined characteristics such as active vs. dormant states as well as changes in ground water consumption rate. Moreover, it demonstrated the strengths and weaknesses of using remote sensing data for the quantitative tracking of water consumption rate over time in arid and semi-arid regions. In order to accomplish this, we utilized Landsat 5 and 8 across the years 2000, 2010, and 2019.

The third objective involved the identification of alternatives to the use of data derived from rain gauges. This was achieved by testing and comparing two types of remote sensing data in the accurate detection of the spatiotemporal distribution of rainfall at the test site. Through the use of statistical metrics and the incorporation of limited rain gauge records from 2015-2018, we were able to determine the best-performing type of remote sensing data.

The fourth objective focused on the identification of new potent+ial locations for rain gauges. In order to accomplish this objective, GPM(IMERG) and SRTM 90×90 data with definite statistical algorithms were used.

2. THE WATER CYCLE IN ARID REGIONS

Adequate management of water resources has recently become an issue of intense focus in arid and semi-arid regions. Local fresh water supply in these climate zones is generally highly limited and is mainly derived from groundwater, which is susceptible to depletion (Sheffield *et al*., 2018). Consequently, the initial step toward sustainable groundwater control is an assessment of local water cycle equilibrium, in combination with identification of groundwater consumption rates. The results of such studies can be used to inspire rules and regulations for the maintenance and preservation of groundwater sources in (semi-)arid areas. Potential regulations could, for example, mandate that withdrawal from aquifers do not exceed natural recharge rates, which would in turn reduce land use and limit population growth. The current chapter approaches this issue through a discussion of the water cycle, water storage, and water consumption patterns in arid and semi-arid regions.

2.1 Background

In general, the hydrological cycle describes the movement of water between the biosphere, atmosphere, lithosphere, and hydrosphere (Pagano and Sorooshian, 2002; Kuchment, 2004). Fresh water can accumulate and be stored in various natural reservoirs, such as oceans, lakes, rivers, soil, glaciers, groundwater, and the atmosphere. Water is also able to transfer between reservoirs by precipitation, evaporation, condensation, deposition, runoff, and infiltration (Kuchment, 2004). The reservoir contributing most to evaporation are the oceans, where water vapor transfers to the atmosphere in the form of clouds that are then propelled great distances by wind, before finally condensing and precipitating, furthering the cycle (Pagano and Sorooshian, 2002). Although wind can promote the transportation of cloud water, the vast majority (91%) of precipitation occurs over the oceans themselves. The other nine percent of precipitation falls over land masses, where it then either infiltrates the ground or becomes surface runoff (Pagano and Sorooshian, 2002; Kuchment, 2004). This precipitation can result in three general outcomes: replenishment of atmospheric water reservoirs via evaporation, recharging of groundwater, or returning to the ocean (Figure 2.1) (Pagano and Sorooshian, 2002; Kuchment, 2004).

The balance of water entering and exiting a particular environment can be described as its water cycle equilibrium. Taking into account multiple factors affecting water availability, this can be quantified by the following formula (Pitman, 2003; Hansen *et al*., 1983; Niu *et al*., 2006; Han *et al*., 2010):

$$P = E + R + \Delta S \quad Eq(2.1)$$

where (P) represents the rate of precipitation, (E) the rate of evapotranspiration, (R) the amount of runoff, and (ΔS) the change in storage capacity of soil moisture. In recent decades, exploitation of groundwater has intensified as a result of climate change and global warming. This has led to alterations in local hydrological cycles that are increasingly destabilizing regional water balances (Shen and Chen, 2010).

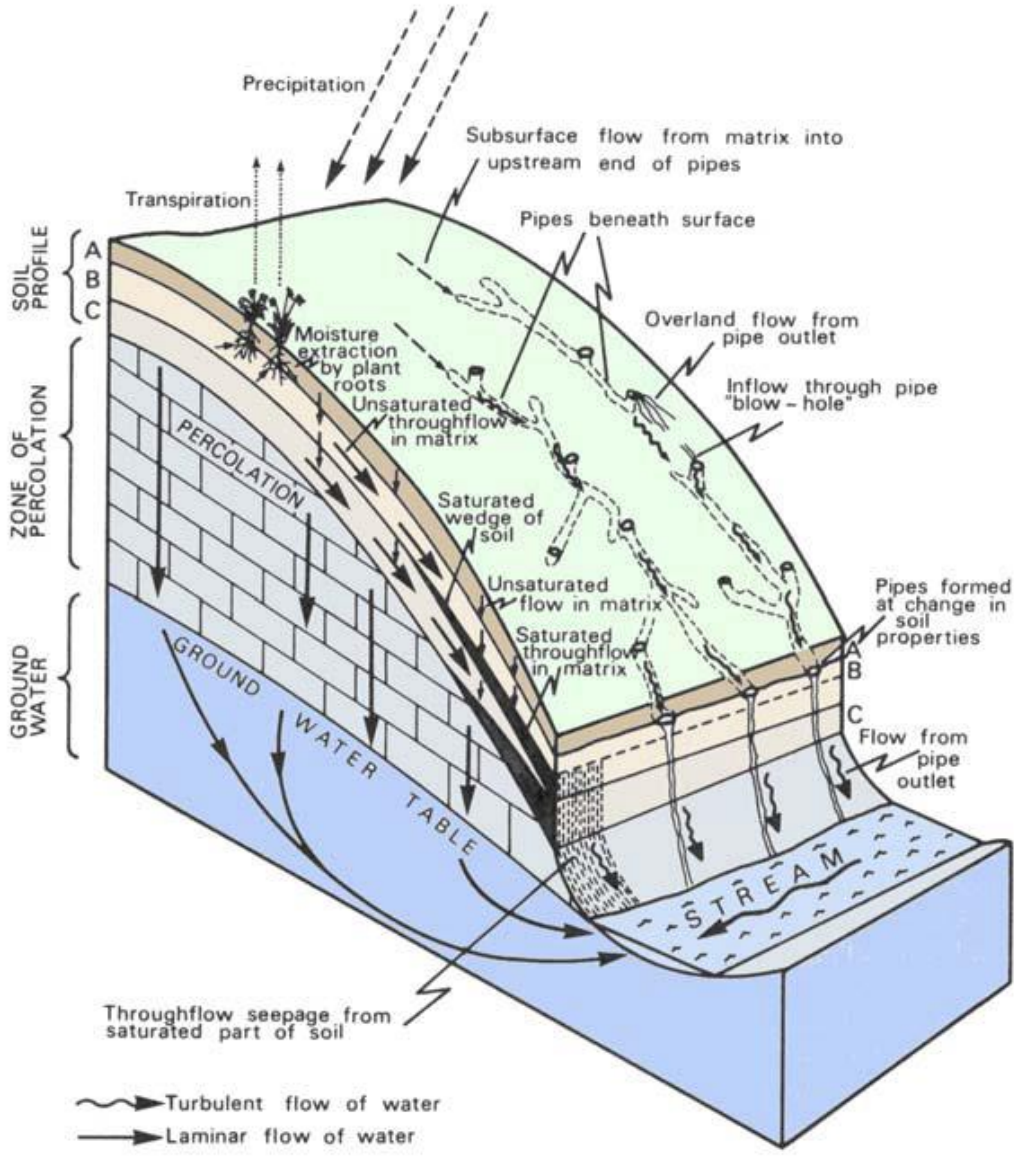


Figure 2.1: The various mechanisms impacting the water cycle (after Kirkby, 1978; Sherief, 2008).

2.1.1 Precipitation

Precipitation functions as the primary factor maintaining water cycle equilibrium (Equation 2.1). Accordingly, it has served as a dominant subject in the majority of hydrological studies

on flash flood risk assessment, groundwater localization, climate change, and forecasting (Tapiador *et al*., 2012). The intensity of a precipitation event can be influenced by the extent, strength, and trajectory of its storm. In arid regions, such storms display distinct variations over even small-scale areas (Niu *et al*., 2006). The possibility of very low local precipitation rates can then negatively affect the continuity of reclamation of degraded land (Kiros *et al*., 2016). Conventional rainfall data acquisition relies on the spatial distribution and number of rain gauges and radar stations. However, in many arid regions, these devices are insufficient for adequate data delivery (Alexakis and Tsanis, 2016) due to their limited number, sporadic distribution, or even nonexistence. In such cases, both active and passive remote sensing technologies are relied upon for data delivery. These data can then be used for assessment of local precipitation rates and spatiotemporal distribution, as well as other hydrological parameters. Remote sensing technologies can provide data at both high temporal and spatial resolution, all while representing vast areas (including considerably elevated land masses).

2.1.2 Infiltration

Infiltration describes the first hydrological consequence of precipitation, occurring when rainfall hits the ground and percolates the soil surface (Beven, 2004; Thornes, 2009). Several factors controlling rainfall percolation rate and its spatial variability include soil type, texture, moisture, and hydraulic properties; vegetation; animal activities; and climate (Beven, 2004; Khan *et al.*, 2014). Infiltration and runoff in arid and semi-arid regions display more complex characteristics when compared with less dry climates, as several additional factors exist that influence the two. For example, they can be affected by the relationship between bedrock slope, curvature, porosity, permeability, and extent versus the degree of soil cover (Beven, 2004; Khan *et al.*, 2014). In order to semiquantify infiltration, internationally-documented models incorporate several of these aforementioned factors (e.g. bedrock qualities and soil cover) as input parameters.

2.1.3 Runoff

Runoff is defined as the outflow of precipitated water from landmasses to the open ocean. As described by Horton (1941), runoff is caused by excess rainfall not percolating the soil surface. This can be due to several reasons, such as soil saturation or the closing off of openings in the soil. As a hydrological phenomenon, it produces both constructive and destructive consequences. Its presence can negatively affect settlements, vegetation cover, road infrastructure, and, in some cases, lead to soil erosion and devastating landslides. Alternatively,

it can also be exploited as a source of freshwater in arid regions. In fact, it is a recent target of interest for addressing the increasing demand for potable water and electricity in these areas (Masoud, 2011). To quantify the relationship between rainfall and runoff, researchers utilize several techniques, among them: simple correlation, area-based methods, regional regressional methods, and GIS-based models (e.g. SCS and SCS-CN) (Masoud, 2011; Abuzied *et al*., 2016; Bo *et al*., 2011). These models are based on water cycle equilibrium and incorporate land use, soil type, terrain slope, soil moisture, and antecedent moisture as primary input parameters.

2.1.4 Evapotranspiration

Evapotranspiration concerns the movement of water and energy from the lithosphere and hydrosphere to the atmosphere (Li *et al*., 2014). Evapotranspiration consists of two processes: the evaporation of liquid water from land masses and large water bodies and the transpiration of water from plant leaves (Vinukollu *el al*., 2011). Evapotranspiration strongly influences water cycle equilibrium, especially in arid and semi-arid regions, where the evaporation rate can regularly exceed precipitation rate. Consequently, estimation and semiquantification of evapotranspiration is another target of focus when determining strategies for efficient water resource management in arid areas (Shen and Chen, 2010). Unfortunately, data on evapotranspiration cannot be retrieved directly by remote sensing products (Kalma *et al*., 2008). It can, however, be estimated by its dependence on various factors, namely local temperature, relative humidity, wind speed, vegetation characteristics, and plant phenology (Kalma *et al*., 2008). As a result, the estimation of evapotranspiration requires input from a variety of sensors, ground observations, and models (Kustas *et al*., 1996; Su, 2002; Kalma *et al*., 2008).

2.2 Storage

The global availability of stored water can be separated into distinct reservoirs, including both on-land and in-land freshwater resources. Among the on-land sources are glaciers, snow, lakes, marshes, and rivers. In-land freshwater resources, on the other hand, exist as soil moisture and ground water (Hartmann *et al*., 2002). The amount of global water supply that is stored on- and in-land is relatively small, though the water flux through these systems is relatively great (Pagano and Sorooshian, 2002; Hartmann *et al*., 2002). In the majority of arid regions, aquifers represent the predominant source of stored fresh water, and storage rate depends mainly on aquifer type, water table level, and degree of water flux (Pagano and Sorooshian, 2002; Hartmann *et al*., 2002).

2.2.1 Aquifers

Aquifers serve as the primary in-land reservoirs of stored freshwater in arid and semi-arid regions. They can be categorized into three distinct types: confined, unconfined, and leaky aquifers (with the type depending mainly on the local lithology). Confined aquifers are both over- and underlaid by a confining bed and yield usable quantities of freshwater to wells or springs (Heath, 1983). Unconfined aquifers, in contrast, are overlaid by permeable beds and underlaid by confining beds with very low hydraulic conductivity (Heath, 1983; Prasad, 2002). Leaky aquifers over- or underlaid by a semipermeable layer through which vertical leakage can take place (Prasad, 2002) (Figure 2.2).

Water stored by aquifers in arid areas can originate as either 'modern' or 'fossil' ground water (Sultan *et al*., 2011), two types distinguishable by distinct isotopic signatures. Modern water describes water that recharges aquifers during current and ongoing precipitation events. Fossil ground water, however, is that which formerly recharged the aquifer during previous decades under different climatic conditions (Sultan *et al*., 2011). Naturally, the majority of recent precipitation in currently arid regions tends to limited and low in intensity. It insufficiently recharges local aquifers and cannot provide for the increasing water demands of growing populations, shifting reliance toward fossil ground water.

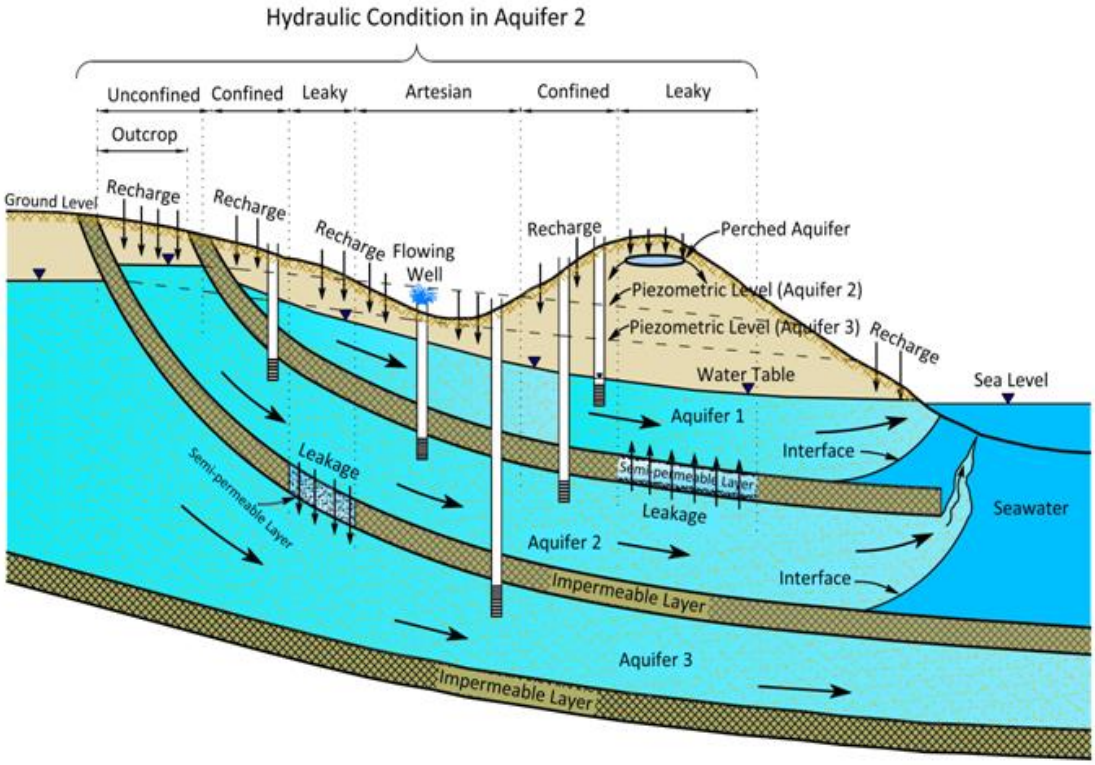


Figure 2.2: Illustration of the different types of aquifers (after Bear, 1979).

2.2.2 Soil moisture

The secondary reservoir for in-land water storage derives from soil moisture, which is responsible for the interaction between the lithosphere and atmosphere. According to the Global Climate Observing System (2010), it is considered to be one of the 50 most critical variables for determining climate (Parinussa *et al*., 2017). This variable is often used to highlight the differences between drought and flood seasons (Cao *et al.*, 2019) and is required for the modeling of important hydrological factors, such as infiltration and runoff (Parinussa *et al*., 2017). Soil moisture displays high temporal variation, as well variation between topographies, soil properties, vegetation, and climate (Crow *et al*., 2012). In order to obtain continuous data on soil moisture, scientists use in-situ measurements along with microwave sensors to produce data sets with considerable accuracy and spatial resolution, as well as a high capture frequency (Liu *et al*., 2012).

2.2.3 Rivers and lakes

On-land reservoirs of stored water consist of rivers and lakes. Furthermore, there exists a hydraulic interaction between surface and ground water in many watersheds, with streams, rivers, and lakes both feeding and withdrawing from local ground water aquifers (Pagano and Sorooshian, 2002; Kuchment, 2000). The entire process depends on the aquifer's ground water level, which itself is reliant on both precipitation and irrigation rate (Massoud *et al*., 2010). However, if water inflow and outflow are under equilibrium, the absolute change in water storage will be zero. While rivers and lakes are uncommon in arid regions, aquifers can occasionally lie adjacent to seaside coastlines. Hydraulic connection between the two can lead to issues with water contamination and saltwater intrusion, especially when ground water levels drop below those of the sea surface (Eissa *et al*., 2016; El Sayed, 2006). In such situations, limits on water withdrawal should be implemented to avoid these consequences, taking precipitation and water recharge rates into consideration.

2.3 Water consumption

Water consumption is the driving force unbalancing the water budget in arid regions. Consumption rates gradually but directly increase with population mass and subsequent land cover and land development. (Scanlon *et al*., 2006). As a result, it is critical for the continuity of arid communities that population (and consequently water withdrawal) be limited.

Six percent of the world's forests are located in arid zones (Malagnoux, 2007) and, despite natural climate constraints, are increasingly being used for agriculture. In fact, 85% of available

water in these regions is diverted for crop irrigation (Ezzahar *et al*., 2007). In order to combat this, several projects have been established for the promotion of sustainable management of irrigation water in arid climates (Malagnoux, 2007).

One billion people reside in arid regions worldwide and as a group represent the world's poorest (Malagnoux, 2007). As this population grows and water needs increase, the overexploitation of trees and forests required to sustain the population will lead to further desertification. Additionally, reduced rainfall due to climate change and global warming will fail to adequately recharge aquifers, also leading to the insufficient natural irrigation of the abovementioned forests (Malagnoux, 2007).

3. REFERENCE SITE

The Sinai Peninsula is considered one of the most unique regions in Egypt. It is known as a prime sightseeing destination, partly due to its location between the Mediterranean and Red Seas. It also contains vast natural wealth in the form of gemstones, gold, coal, and other resources. Like the majority of Egypt, it is classified as arid and semi-arid and relies on groundwater as its source of freshwater. The eastern side of the Gulf of Suez presents one of the most promising locations for future urban expansion and might in the future be able to house a considerable portion of the growing population, in addition to tourist accommodation. In fact, it currently is already demonstrating a gradual increase in number of residences, along with the associated land exploitation. In terms of land geology, this eastern region features variable geological settings, with lithological units even appearing in fascinating outcrops. Moreover, aspect, slope, and elevation vary greatly, which in turn directly affect precipitation rates, evaporation, infiltration, and runoff. Additionally, the entire area of study relies primarily on a single aquifer. The following chapter discusses the characteristics of this region.

3.1 Location and topography

The Sinai Peninsula features a triangular shape, covering an area of 61,000km^2 (El-Tahlawi, 2014). It is bound by the Gulfs of Aqaba and Suez on its eastern and western sides, respectively, and the Mediterranean Sea delimits its northern coast. This peninsula is separated from the Eastern Desert by the Gulf of Suez and the Suez Canal (Said, 1962). The central portion of the peninsula comprises the highest igneous and metamorphic mountains of the region, with elevations ranging from 2070m at Mount Serbal to 2641m at Mount Catherine (Said, 1962). This central region is dissected by numerous wadies, or dry valleys, all of which include intense drainage networks (Said, 1962). The Eastern side of the Gulf of Suez is nearly 350km long and 80km wide (McClay, 1998). Regional formations feature complete stratigraphy from Precambrian to Quaternary periods (McClay, 1998). It is located between latitudes 29°54' N and 27°42' N and longitudes 32°42' and 34°06' E. In terms of urban areas, the region is populated by Sharm El-Sheikh at its southern vertex, Ras-Sudr and Abu-Rudeis in the north, and the cities of El-Tor and Saint Catherine in the center (Figure 3.1). Middle and Southern vertex of this region comprises the El-Qaa Plain, located between latitudes 28°30' and 28°40' N and longitudes 33°17' and 33° 37' E (Sayed *et al*., 2004). The overall area of the El-Qaa Plain is roughly estimated to be 6070km^2, with a maximum length of 150km, and a maximum width of 20km in the north (Ghodeif and Gorski, 2001) (Figure 3.1). It is also narrowest in the south (Azab and El-Khadragy, 2013). According to Sayed *et al.* (2004), the

eastern portion of the El-Qaa Plain includes a Precambrian mountain region with elevation varying from 300m to 2624m. This region contains various types of igneous rocks, such as diorite, granite, metagabbro, and volcanic varieties (Sherief, 2008; Han *et al*., 2010). Its dominantly sedimentary sector can be found in Gabal Qabaliat in northwestern sector, where elevation reaches approximately 250m and where the terrain moderately slopes toward the El-Qaa Plain. It is also this northwestern site that separates the Gulf of Suez from the El-Qaa Plain. Local sedimentary outcrops include limestone, sandstone, siltstone, gypsum, and anhydrite formations. The central Plain is composed mainly of Quaternary deposits, which are generally not perfectly flat and are often dissected by various wadies, alluvial fans, palaya, and terraces (Said, 1962). A 2008 study by Sherief divides the area between types of deposits, whether alluvial or Wadi-derived. (Figure 3.2, and 3.3).

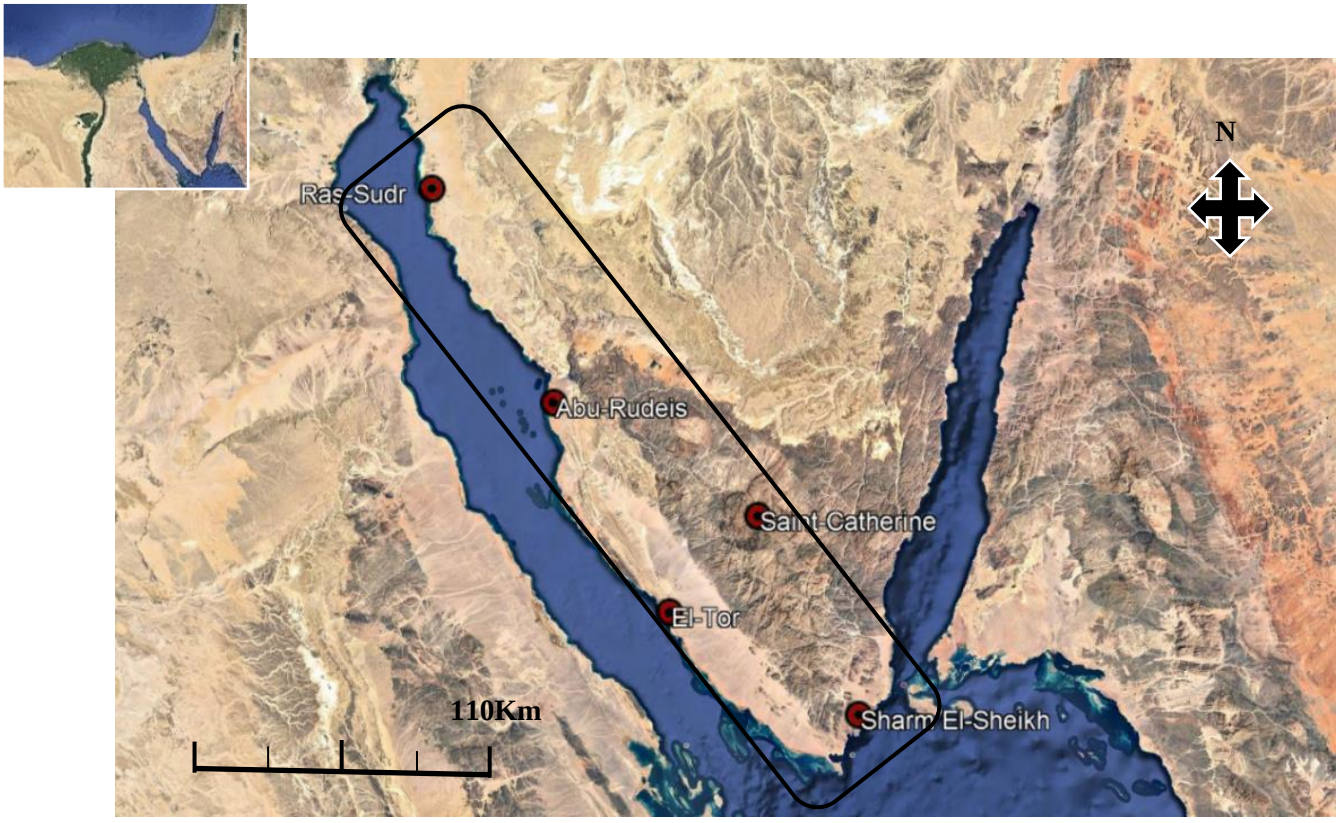


Figure 3.1: Satellite map of the eastern portion of the Gulf of Suez, with its five ground-based stations identified. (Source: satellite image from Google Earth, 2017).

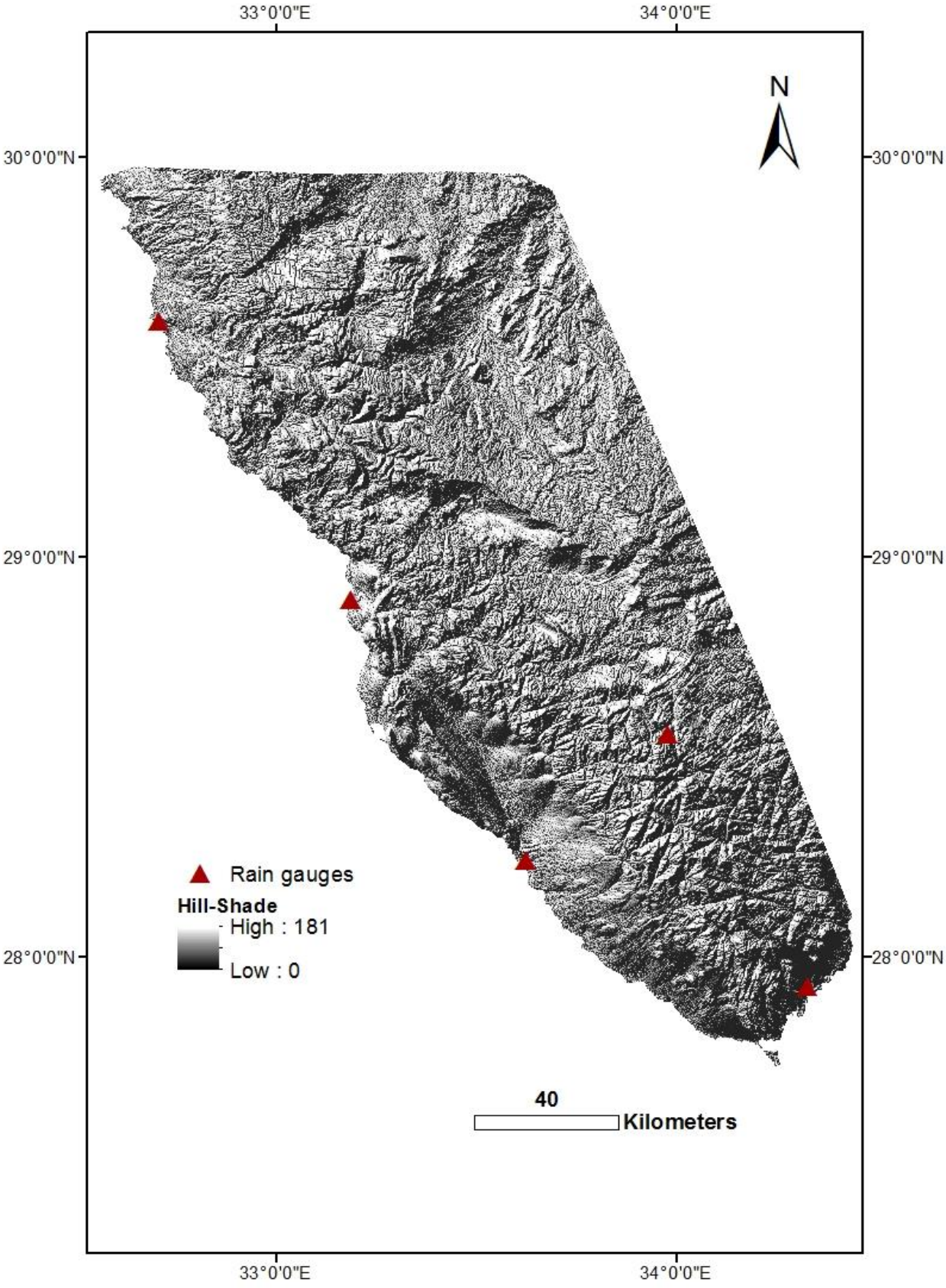


Figure 3.2: 3D morphological Hill-Shade map of the El-Qaa Plain and its surroundings. (Created by ArcGIS 10.5).

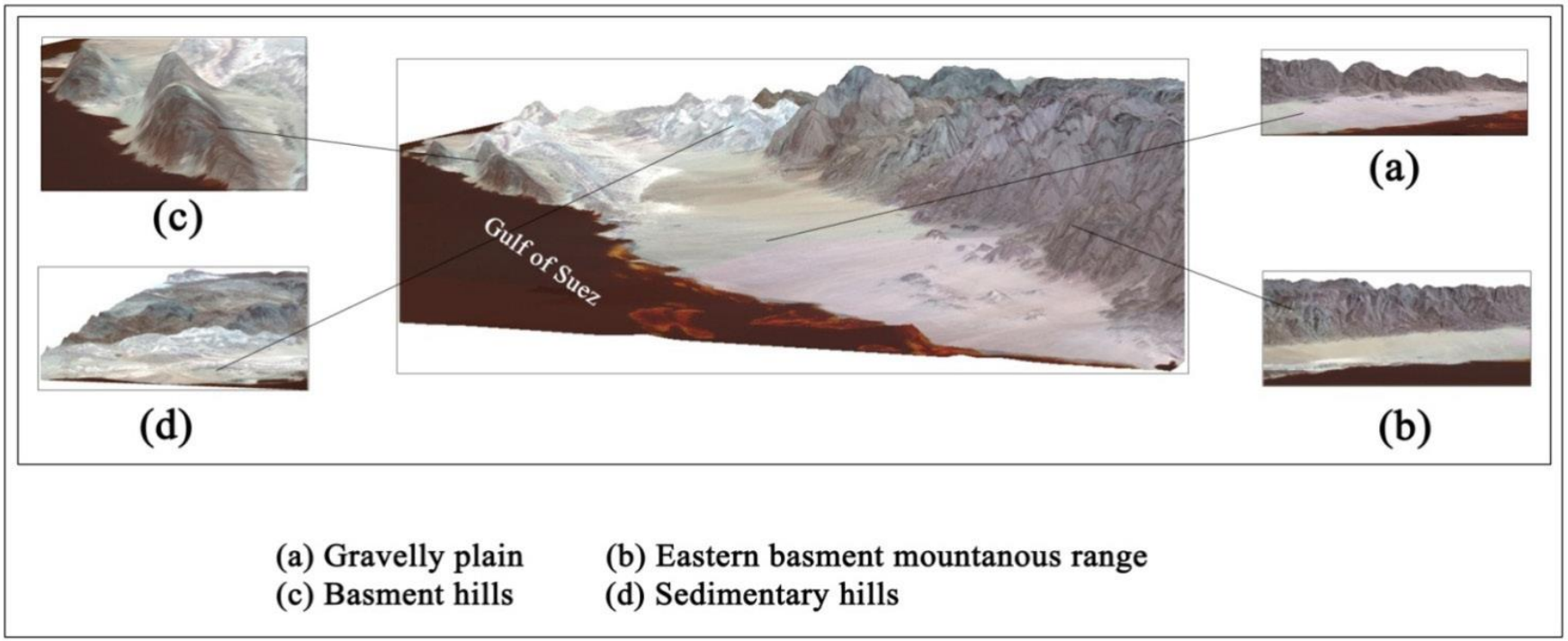


Figure 3.3: 3D views of major geomorphologic units of the El-Qaa Plain (after Wahid *et al*., 2016).

3.1.1 Aspect

Aspect can be defined as the cardinal direction a slope faces. It represents a critical factor in many hydrological studies, particularly those involving infiltration, runoff, and flashflood hazard mitigation. When analyzing the factors influencing the water cycle at a test site, it is paramount to include the site's slope aspect. This factor largely depends on the lithological composition of the site as well as its geomorphology. As seen in Figure (3.4), the alluvial plain feature aspects that fluctuate between South, Southwest, North, and West. In contrast, the basement complex of this area features additional aspects that fluctuate between Southeast, East, Northwest, and North. Ultimately, each of the abovementioned aspects face the seaside. Gabal Qabaliat, for example, has an aspect toward the East, which influences the recharge of the El-Qaa Plain. Taken together, the aspects mentioned confirm that the infiltration taking place in the El-Qaa Plain occurs via microstructures within its basement complex and that excess water travels toward the direction of the Gulf and Red Sea. The map presented was prepared using SRTM at a 90×90m resolution in ArcGIS 10.5.

3.1.2 Slope degree

Slope describes the rate of increase or decrease in altitude against horizontal distance. It can be expressed as a ratio, decimal, fraction, percentage, or tangent of the angle of inclination (Sherief, 2008) and produces a noticeable effect on the motion of the water cycle. For example, runoff intensity and infiltration rate are not only determined by rainfall intensity, but also depend the slope of the terrain on which they fall (Mu *et al*., 2015). Many practical applications exist for the measurement of slope degree, including quantification of runoff rate and peak, and the prediction of destructive flashfloods. Figure (3.5) displays the slope of the area under study in degrees, with the alluvial plain ranging from 0 to 10 degrees, and the basement complex more steeply ranging from 10 to 70 degrees. This map indicates that infiltration rate over the alluvial plain is generally higher than that over the basement complex. Moreover, runoff rate may decelerate when reaching the plain. This slope map was prepared using the same DEM file and software as previously (Figure 3.5).

3.1.3 Elevation

Elevation can be described as the difference in height between regions. It directly and proportionately influences the precipitation and evaporation rates of a particular area. The elevation of the area under study are shown in Figure (3.6). It can be seen that the elevation of the alluvial plain is lower than that of the basement complex, ranging from 0 to 300m vs. 300

to 2496m. This results in the alluvial plain receiving the lowest amount of rainfall and the basement complex the highest. The upcoming elevation map was created using the same DEM file and software as used to create the previous two.

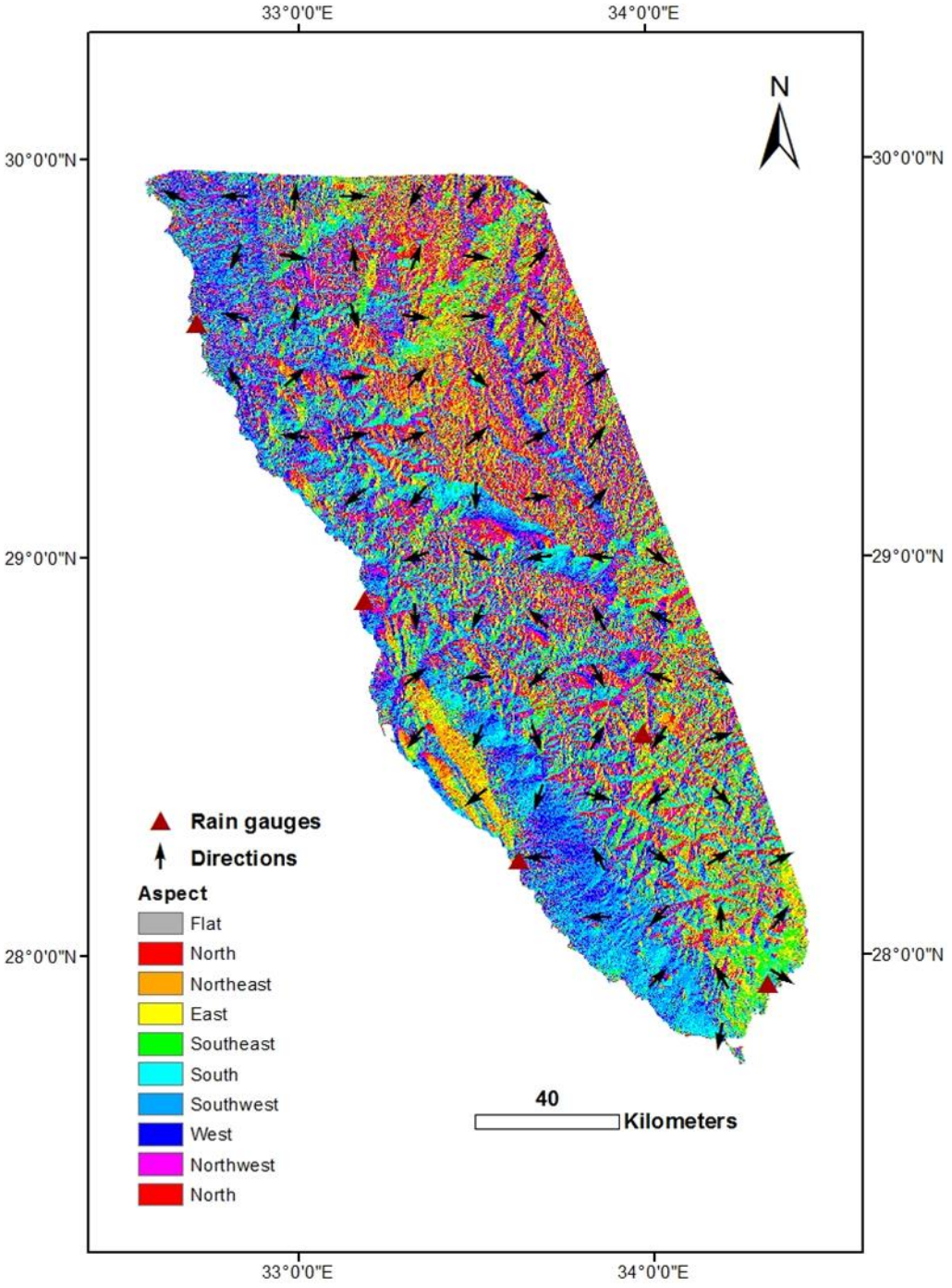


Figure 3.4: Display of aspects detected at the study site.

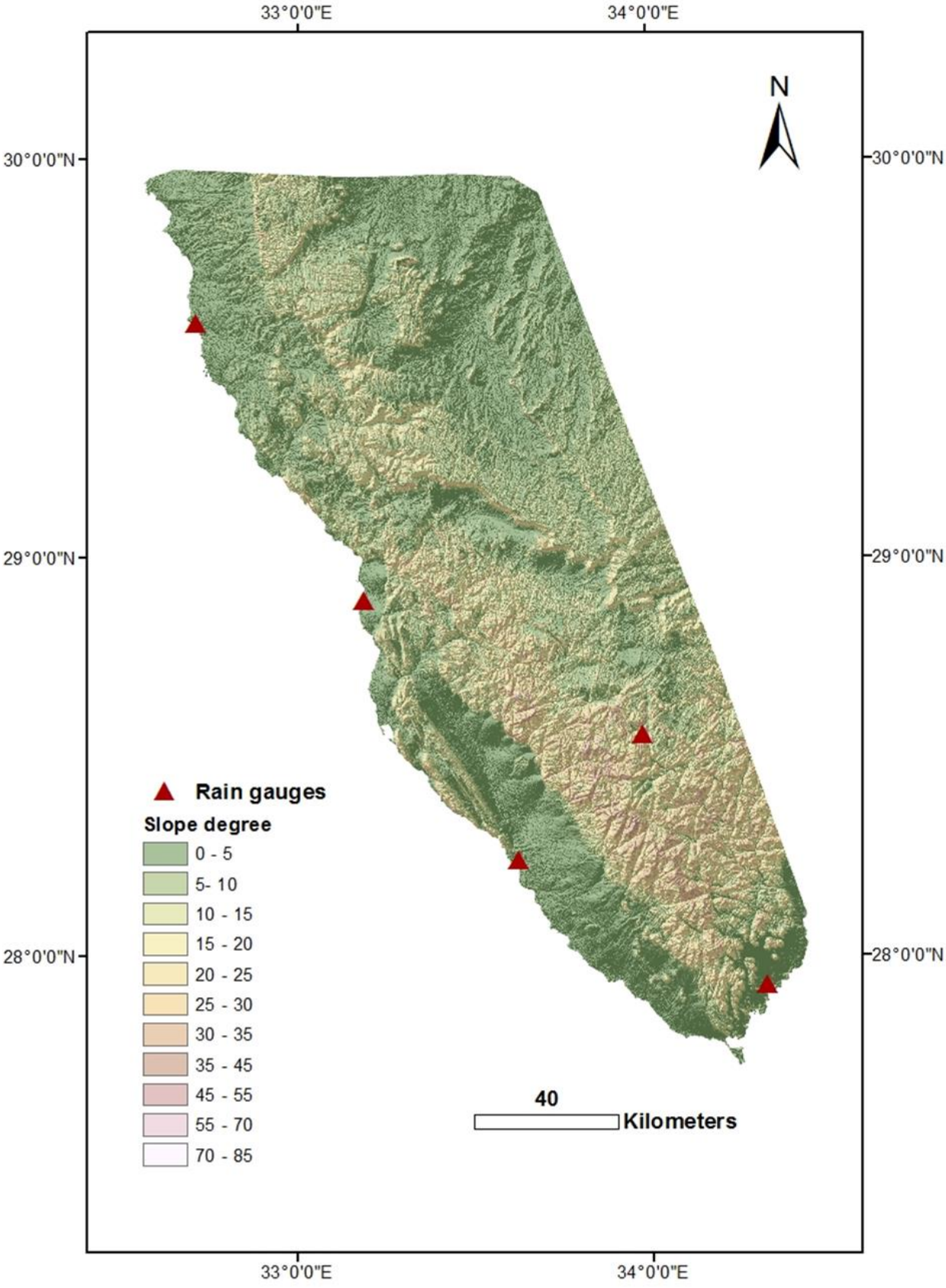


Figure 3.5: Display of slopes at the study site.

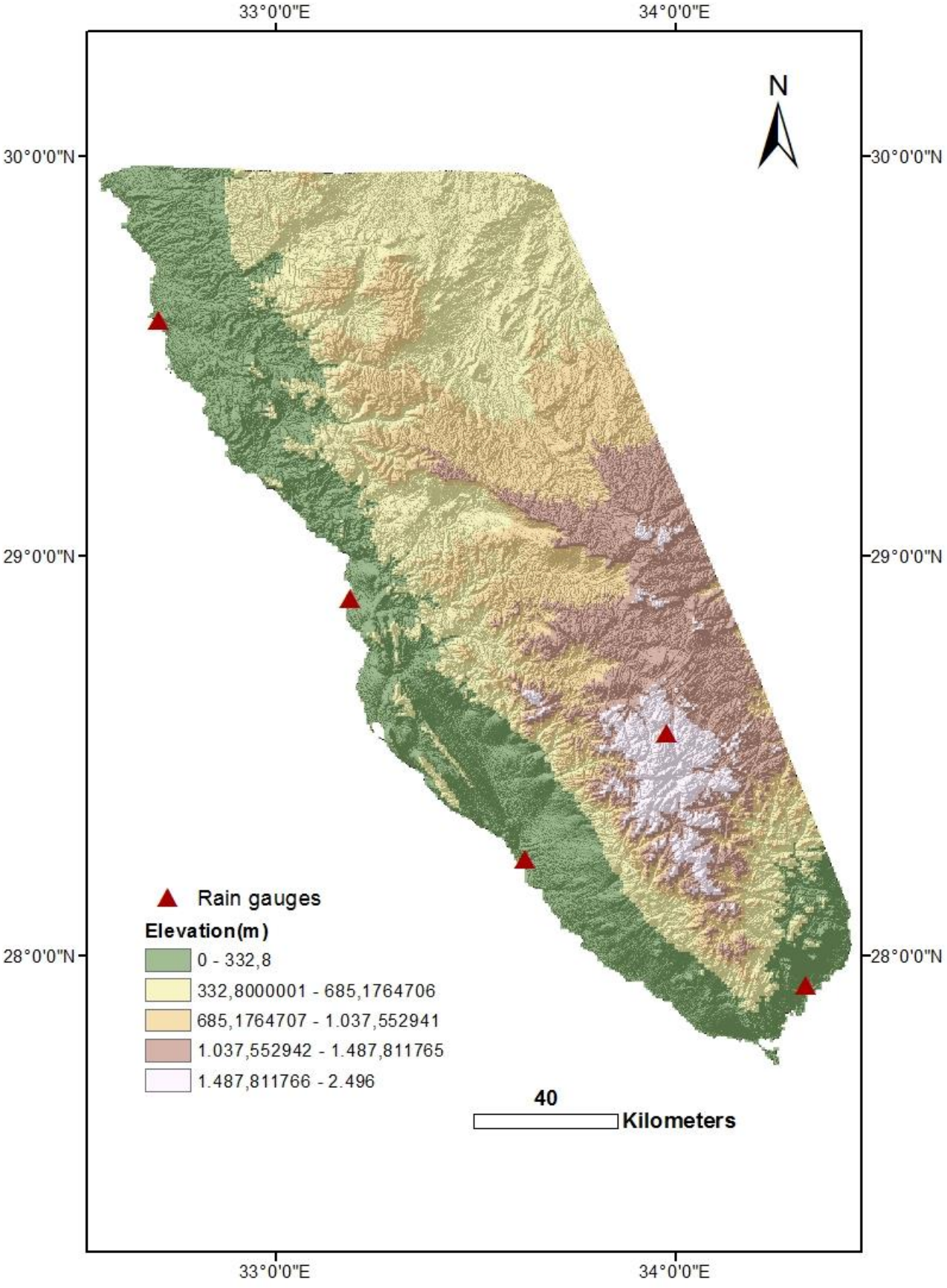


Figure 3.6: Elevation in the El-Qaa Plain.

3.2 Geologic setting

3.2.1 Geology

The geology of the eastern side of the Gulf of Suez ranges in age from Precambrian basement rocks to Quaternary deposits (Selim *et al*., 2016; McClay, 1998), with plate motion between crust and mantle forming the Precambrian basement complex in the southern Sinai. These complexes become more recent towards the Mediterranean Sea, as they are covered with sedimentary succession (Selim *et al*., 2016; Sayed *et al*., 2004). Public offshore data are quite limited (McClay, 1998). However, sufficient studies regarding the geology of El-Qaa Plain do exist, with a number focused the southern portion of the study site, from Wadi Feiran to Ras-Mohamed. The Precambrian basement complex in the plain comprises several rock types. Igneous plutonic rocks cover about 43.3% of the test site and consist of felsic, intermediate, and mafic groups. The felsic group can be found in granitic rocks, which cover 26.8% of the area under study and consist of course- to medium-grained alkaline granite, riebeckite, and monzogranite, which are locally-foliated and megacrystic (Sherief, 2008). Intermediate group rocks were present as isolated diorite masses, which varied in composition (El-Masry *et al*., 1992) from coarse- to medium-grained quartz diorite as well as hornblende biotite granodiorite. (Sherief, 2008). Mafic group consisted of metagabbro rocks that covered 1.5% of the study site. Among these were deformed course- to medium-grained hornblende and olivine-bearing varieties (Sherief, 2008). Volcanic rocks consisted of metamorphosed, pyroclastic, and volcaniclastic types. These included tuffs, agglomerates, tuffaceous laminates, turbidites, and conglomerates (Sherief, 2008). Metamorphic rocks covered 3% of the study site and consisted of coarse- to medium-grained banded hornblende-biotite gneiss that was partly porphyroblastic and mylonitic, and which contained interbanded migmatite amphibolite and metabasite dykes. Sedimentary-derived areas consisted of limestone exposed in various forms, including Wata, Sudr, Thebes's, Darat, and Samalot formations. These appear as thin layers ranging in thickness from 0.5 to 2m that contain fossil fragments of gastropods, echinoderms, and foraminifers (NOWEIR and EL-SHISHTAWY, 1996; Sherief, 2008). Of note, one of the largest sandstone bodies on Earth, the Araba formation, exists in the area under study. This formation is composed of varicolored Nubian sandstone in shades of white, yellow, red, brown and gray and is up to 120m thick (Sherief, 2008). The siltstone-based Matulla formation consists of brown to reddish siltstone, shale, sandy shale, sandstone, and, less abundantly, limestone and dolomite. In contrast, the Sudr formation is composed of gypsum, anhydrite with marl interbeds, a 40m-thick section of white chalk, chalky porcelanite, and calcareous and siliceous shale (Sherief,

2008). Alluvial deposits cover nearly 39.59% of the study site and represent important lithological outcrops in this area (Figure 3.7).

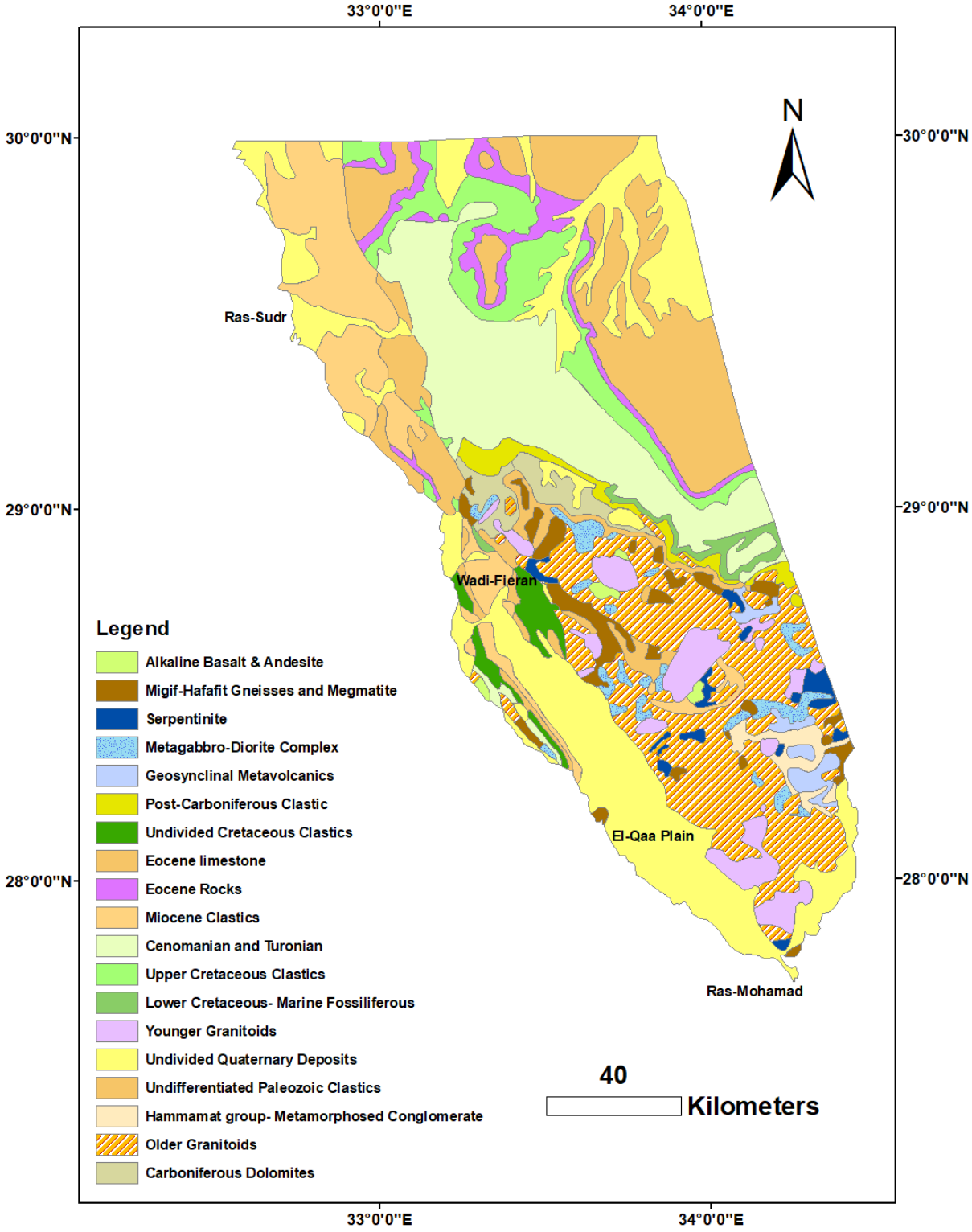


Figure 3.7: Illustration of outcrop lithology at the study site. (Source: Geological Map of Egypt, 1981).

3.2.2 Hydrology

The Quaternary aquifer is the main source of ground water at the study site (Mohamed *et al*., 2004; El Refai 1984; El-Fakharany *et al*., 2003). It extends from Wadi Feiran to the head of Ras-Mohamed (Wahid *et al*., 2016) (Figure 3.8). This aquifer is recharged by rainwater infiltration of the basement complex on its eastern side, where water is able to percolate through its fractured systems to seep through buried wadies (Massoud *et al*., 2010; Ahmed *et al*., 2014). Moreover, according to Sultan *et al*. (2007) there is a second source of recharge existing beneath the Quaternary aquifer known as the Nubian Sandstone Aquifer. Here, water ascends through deep-seated fault systems into relatively thick alluvial aquifers bounded by the fault complex within the El-Qaa Plain.

The Quaternary aquifer composed of (El-Tor group), that composed mainly of gravel, sand, silt, and clay. Its eastern portion, however, contains cobble and boulder-sized debris of weathered Precambrian basement rocks. Its western portion features carbonates and evaporites, with presence increasing toward the more sedimentary Gabal Qabiliat in the West (Massoud *et al*., 2010; Mohamed *et al*., 2004). The freshwater contained within the Quaternary aquifer covers an area of roughly 3300km^2 (Sultan *et al*., 2009) and water depth varies between 4m near the city of El-Tor and 73-80m towards the southern end of the aquifer.

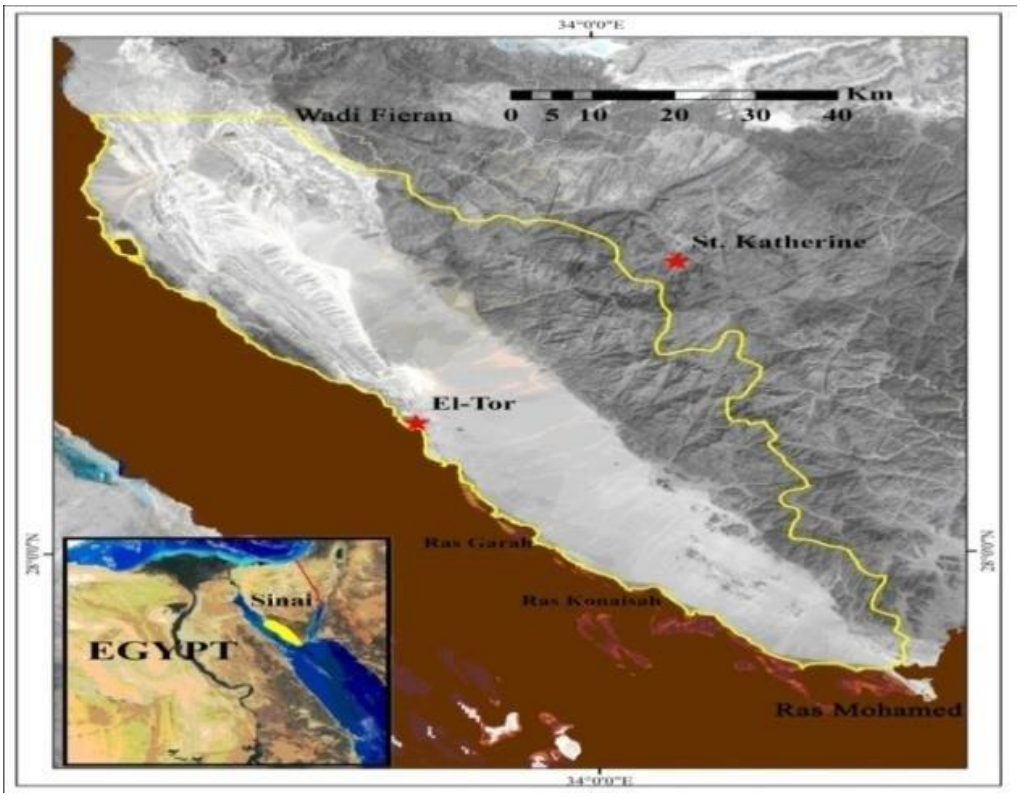


Figure 3.8: Location of the quaternary aquifer extending from Wadi Feiran to Ras Mohamed (after Wahid *et al*., 2016).

4. CLIMATIC CONDITIONS

As discussed in Chapter 2, factors influencing water cycle equilibrium can also directly impact water budgets in arid regions, communities largely dependent on groundwater as their source of freshwater. The climate in these areas, however, can vary greatly (Section 2.1.1). The purpose of this chapter is to briefly discuss the factors at play in the various regions of the study site (namely precipitation, evapotranspiration, infiltration, and runoff) and what branching out of and influences these factors (e.g. precipitation intensity, differing lithologies, and runoff volume and peak). Using these analyses, this chapter seeks to identify the climatic data further required for adequate studies. The data used to form these conclusions is derived from a 2008 study by Sherief.

4.1 Background

The abovementioned factors were analyzed by Sherief in a 2008 study using data collected by the Egyptian Meteorological Authority over a period of 55 years (1934-1989). As arid regions are mostly reliant on groundwater as a freshwater source, and precipitation is the key parameter influencing groundwater levels, establishing rainfall intensity at the study site was highly important. In his study, Sherief categorized precipitation into light (0.1 to 1mm), moderate, (1 to 10mm) and heavy (>10mm) intensity events. Intensity distribution showed that 61% of yearly events were light, 34% were moderate, and 5% were heavy. The area under study received the vast majority of its yearly precipitation during Autumn and Winter (October to March). Its southern end received 98.4% of its yearly total during December, while the middle region experienced 71.5% of its precipitation between the months of December and January. The northernmost side received 61.5% of its total between December and January (Sherief, 2008) (Table 4.1).

Table 4.1: Months of highest precipitation for each of the El-Qaa Plain regions and their associated yearly precipitable portions. (Modified after Sherief, 2008).

El-Qaa Plain Regions	Date	Yearly precipitable portion
Southern-Region	December	98.4%
Middle-Region	December - January	71.5%
Northern-Region	December- January	61.5%

As previously discussed, evapotranspiration serves as one of the most crucial outgoing factors influencing water cycle equilibrium in arid regions (Davarzani *et al*., 2014). In the El-Qaa Plain, the effect of vegetation is insignificant due to its limited land cover. However, during

the rainy season, total daily evaporation loss (as calculated by Sherief, 2008) can amount to (16.4mm/d). The southern end of the study site experienced the highest evaporation rate, while the middle region recorded the lowest (Figure 4.1). Among the factors controlling evaporation rate at the test site were temperature, relative humidity, and wind speed (Davarzani *et al*., 2014).

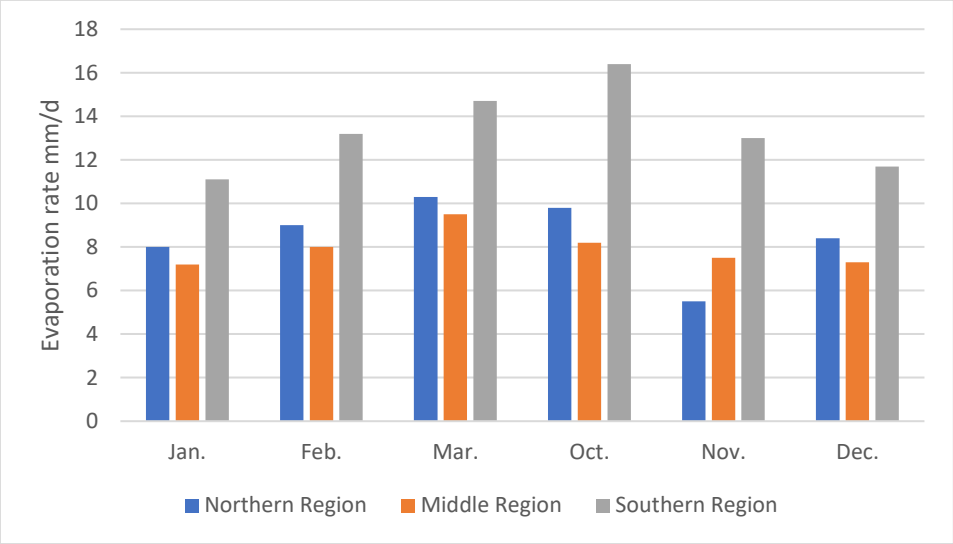


Figure 4.1: Daily evaporation rate during the rainy season in the each of the three studied regions. (Modified after Sherief, 2008).

Temperature was the primary factor affecting evaporation rate at the test site. In Summer and Spring, the rate of evaporation reaches its maximum level owing to temperature increase. In contrast, during the Fall and Winter, the minimum rates were recorded (due to temperature decrease) (Sherief, 2008). According to data collected by Sherief (2008) between 1934 and 2002, May, June, July, and August featured the highest evaporation rates, while March, April, September and October experienced moderate evaporation rates and January, February, November, and December recorded the lowest (Sherief, 2008).

Relative humidity is defined as the ratio of the partial pressure of water vapor to the equilibrium vapor pressure of water at a given temperature (Zhao *et al*., 2012). Relative humidity naturally shifts according to changes in temperature and pressure, meaning that, given a fixed amount of water vapor, an increase in temperature will result in lower relative humidity. (Zhao *et al*., 2012). The southern region of the study site invariably demonstrated lower relative humidity with comparison to the northern region, which featured the highest relative humidity. Relative humidity at the site under study increase during Autumn and Winter and decreased in the Spring and Summer. Values ranged from <50% in Spring and Summer to 60-70% in Autumn and Winter, classifying the site as arid and semi-arid (Sherief, 2008).

The final factor affecting rate of evaporation is wind speed. Moreover, it has noticable effect on precipitation and relative humidity (Back and Bretherton, 2005). Wind direction at the test site generally fluctuated between North, Northeast, East, Southeast, South, Southwest, West, and Northwest, differing between regions and seasons. In the Northern, Middle, and Southern regions, the predominant wind direction was Northwest. Additional information can be found in Sherief's 2008 study.

In 1973, Chorley calculated the infiltration ratios of the various lithological units at the study site, values which were later modified by Sherief (2008) (Table 4.2). According to Sherief (2008), greatest infiltration took place in the plain that covers 28.8% of the total El-Qaa Plain. This region ranged in elevation from 0 to 300m, presented slopes from 0 to 15°, and featured permeabilities ranging from 15140 to 151400 $m^3/day/km^2$. Lowest infiltration rates were seen in the basement units over the eastern mountains, ranging from 3.79 to 37.9 $m^3/day/km^2$. This area ranged in elevation from 400 to 2000m, with slopes between 20 to 70°.

Table 4.2: The infiltration losses, porosities, and permeabilities of the different lithological materials found in the El-Qaa Plain (after Sherief, 2008).

Lithologic material	Area km^2	porosity	permeability	Infiltration million m^3
Granite, Slate, Schist, Gneiss, Tuff	2598.5	0.0001 - 1	3.79	9.9
Volcanic	221.6	0.001 - 50	37.9	8.4
Clay, Silt, Shale, Sabkha	277.4	15 - 50	75.7	21.0
Sandstone	403.8	5 - 25	151.4	61.1
Limestone	282.7	0.1 - 10	1514	482
Alluvial sands, Wadi deposits	543.4	30 - 40	15140	822.7
Alluvial fans (gravels and boulders)	1750.4	25 - 35	151400	265.011

Sherief (2008) calculated both total runoff volume and peak runoff in the main basins of the study site. This was achieved by taking into consideration the area, drainage density, slope, rainfall depth, and geological surface of the basins. The northern region of the El-Qaa Plain experienced runoff volume ranging from 500 to 2000 m^3, while the middle region recorded 200 to 499 m^3 and the southern region <70 to 499 m^3. The northern region recorded >150m^3/s, whereas the middle and southern regions experienced a runoff peak of <5m^3/s (Table 4.3).

Table 4.3: Runoff volume and runoff peak in the three regions of the El-Qaa Plain. (Modified after Sherief 2008).

Region	Runoff Volume m^3	Runoff Peak m^3/s
Northern Region	500 to 2000	>150
Middle Region	200 to 499	<5
Southern Region	<70 to 499	<5

4.2 Data and methods

Rain gauge-based data published in 2008 by Sherief (covering average monthly precipitation rates from 1934-1989) were used to determine rainfall intensities and calculate the 5th, 50th, and 95th percentiles (the value below which a given percentage of observations falls) (Figure 4.2). They represent a benchmark for both normal and non-normal data distributions. Five percent of rainfall data fall before line (a) in Figure 4.2, 95% of data fall between lines (a) and (c), and 100% of data fall after line (c). Line (b) represents the median of the complete data set.

The selection of the three abovementioned percentiles was based on the rainfall intensity limits published by Sherief (2008). These intensities, moreover, corresponded to actual values observed at the study site, as detected by remote-sensing data (Chapter 6). Frequency of intensities were calculated via the creation of histograms derived from data captured by rain gauges during the period of 1934-1989. Histograms were calculated in RStudio using the code below (Code 4.1).

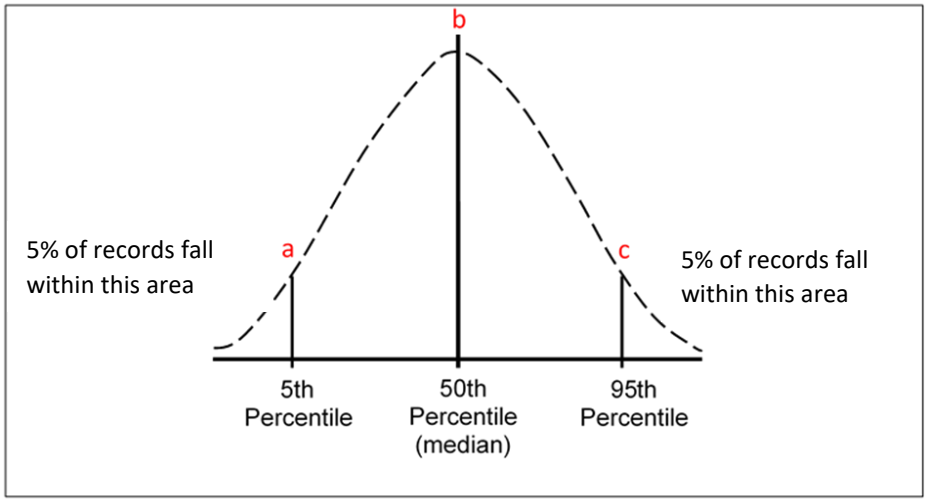


Figure 4.2: 5th, 50th, and 95th percentiles and their positions on normally distributed data set.

To identify the factors affecting evaporation rates at the test site, the correlation coefficient (R) and calculated probability (p) values were determined. This allowed for the evaluation of the relationship between temperature and evaporation, wind speed and evaporation, and relative humidity and evaporation (drawing from the same 2008 data published by Sherief). Correlation describes a statistical technique that can established whether and how strongly pairs of variables are related. The strength of the relationship varies with R and p values. At the study site, a correlational test was applied for the northern, middle, and southern regions. RStudio was used

to calculate correlation coefficients, as well as to create correlation graphs (via ggplot2 package). The code used for these purposes follows below (Codes 4.2, 4.3, and 4.4).

Sherief's 2008 study provided valuable information on the porosity, permeability, and infiltration properties of lithological units at the study site. Using this, we constructed correlational relationships between infiltration ratios and permeabilities of these lithological units. This was achieved using Codes 4.2, 4.3, and 4.4. These relationships provided a better understanding of the influence of permeability on infiltration rate at the area under study.

```
hist((rainfall),col = "lightblue", border = "pink",main = "",xlab = "")

title(main = "Ras-Sudr", cex.main=1.2, font.main=1, xlab = "Rainfall(mm)")
```

Code (4.1): Used to produce histograms of the frequency of rainy events of different intensities.

```
Cor.test(file$evaporation, file$temperature)
```

Code (4.2): Used to calculate correlation coefficient values.

```
ggplot(file,aes(x=evaporation,y=temperature))+geom_point(size=3,colour="blue")+geom_smooth(method=lm,linetype=1,size=1,alpha=0.5,colour="black")+annotate("text",1.5,18,label="R=0.962,p<5.539e07")+labs(y="Temperature(C°)",x="Evaporation(mm)")+ggtitle("Northern Region")+theme_gray()+theme(text = element_text(size=15),axis.text = element_text(size=17,colour="black"),axis.title.x = element_text(size = 15),plot.title = element_text(hjust = 0.5))+theme(legend.position = "none",panel.grid.major = element_blank(),panel.grid.minor = element_blank())
```

Code (4.3): Used to create correlation graphs, including R and p-values.

```
ggsave(file="northernregion.png")
```

Code (4.4): Used to save graphs with an enhanced correlation line.

4.3 Results

Based on calculations of 5th, 50th, and 95th percentiles, precipitation intensity was classified into light, moderate, and heavy. Light intensity events were defined as 0 to 0.25mm of precipitation, while moderate events ranged from 0.25 to 8.9mm and heavy events more than 8.9mm (Figure 4.3). Naturally, the impact of each intensity category on the environment varies. For example, heavy intensity events are considered the most dangerous, as they can deposit 5%

of total yearly precipitation at once and lead to intense flash flooding that results in property damage and loss of life. In comparison, light intensity events supported water cycle continuity at the study site through changes in soil moisture and infiltration rates. The frequencies of the abovementioned intensities were confirmed by the calculation of the histogram constructed from the data derived from Sherief's 2008 study (Figure 4.4). All gauges recorded light intensity events to be the most frequent and heavy events to be the least frequent. An overall histogram representing the entire test site was subsequently calculated, producing nearly identical results.

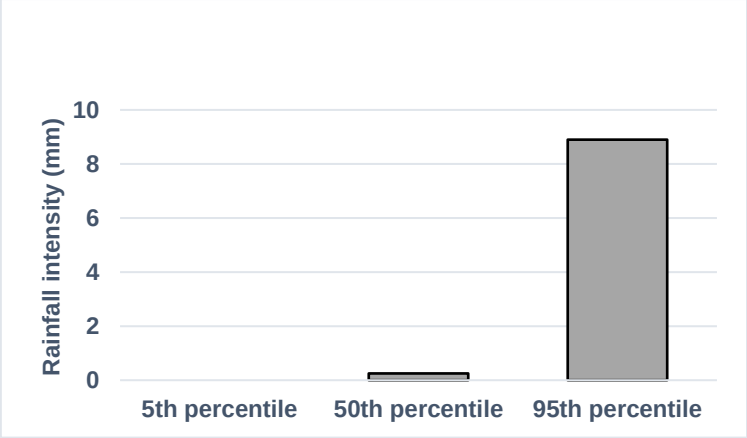
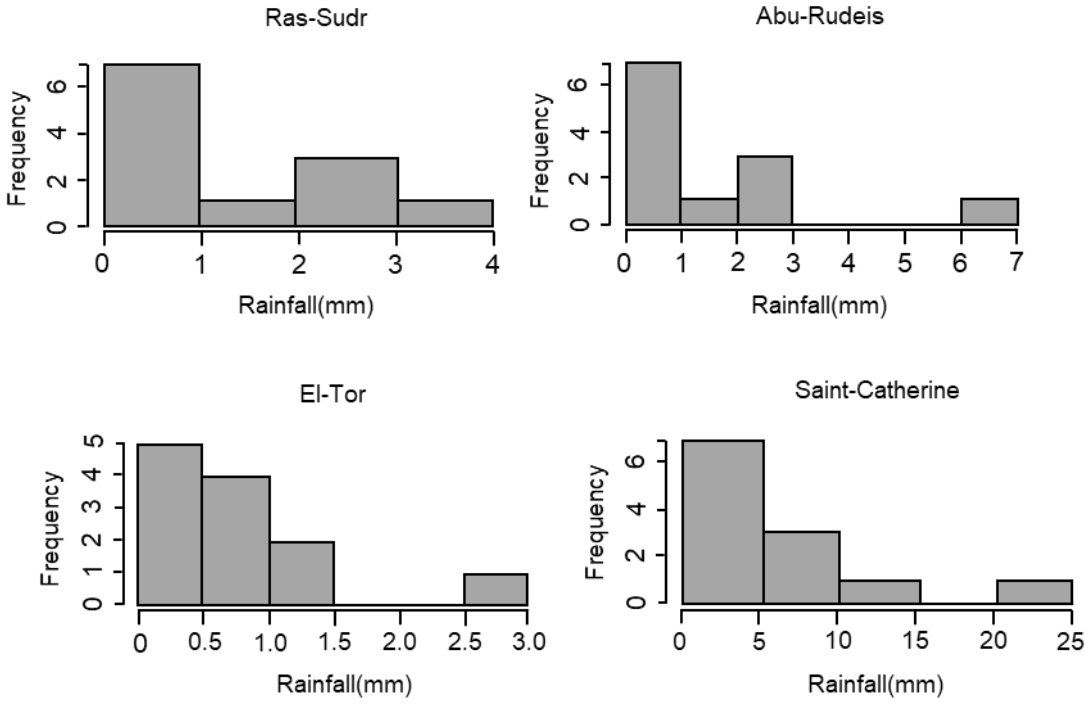


Figure 4.3: Classification of monthly rainfall intensity in the El-Qaa Plain. Based on mean monthly data from 1934-1989.



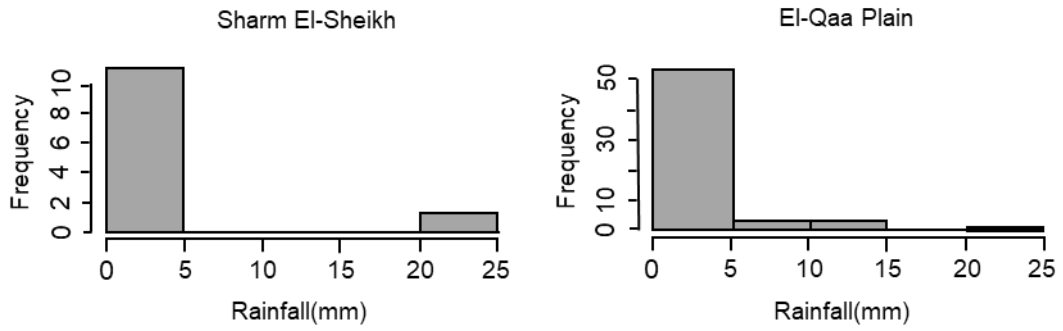


Figure 4.4: Histograms illustrating the frequency of monthly rainfall events in each intensity category. Based on data collected from five rain gauges between 1934-1989.

The relationship between temperature and evaporation rate was determined by calculation of the corresponding correlation coefficient (Figure 4.5). This coefficient, R, was found to be 0.962, 0.779, and 0.966 in the Northern, Middle, and Southern regions. P was calculated as 5.539e-07, 0.0028, and, 3.009e-07 for the same respective regions. These values indicate a strong positive correlation between the variables tested. In other words, evaporation rate strongly depended on temperature. In the Northern and Southern regions, evaporation rate showed strong negative correlation with relative humidity (R being -0.605 and -0.8325, respectively), while in the Middle region, this relationship was not observed (P = 0.1325) (Figure 4.6). Wind speed also demonstrated a strong correlation with evaporation rates at the study site. The strength of the relationship was nearly equal over the Northern (R = 0.778) and Southern (R = 0.786) regions. However, an even stronger correlation existed in the Middle region (R = 0.919) (Figure 4.7 and Table 4.4). In summary, temperature, relative humidity, and wind speed all had a significant effect on evaporation rate, save for relative humidity in the Middle region, which did not.

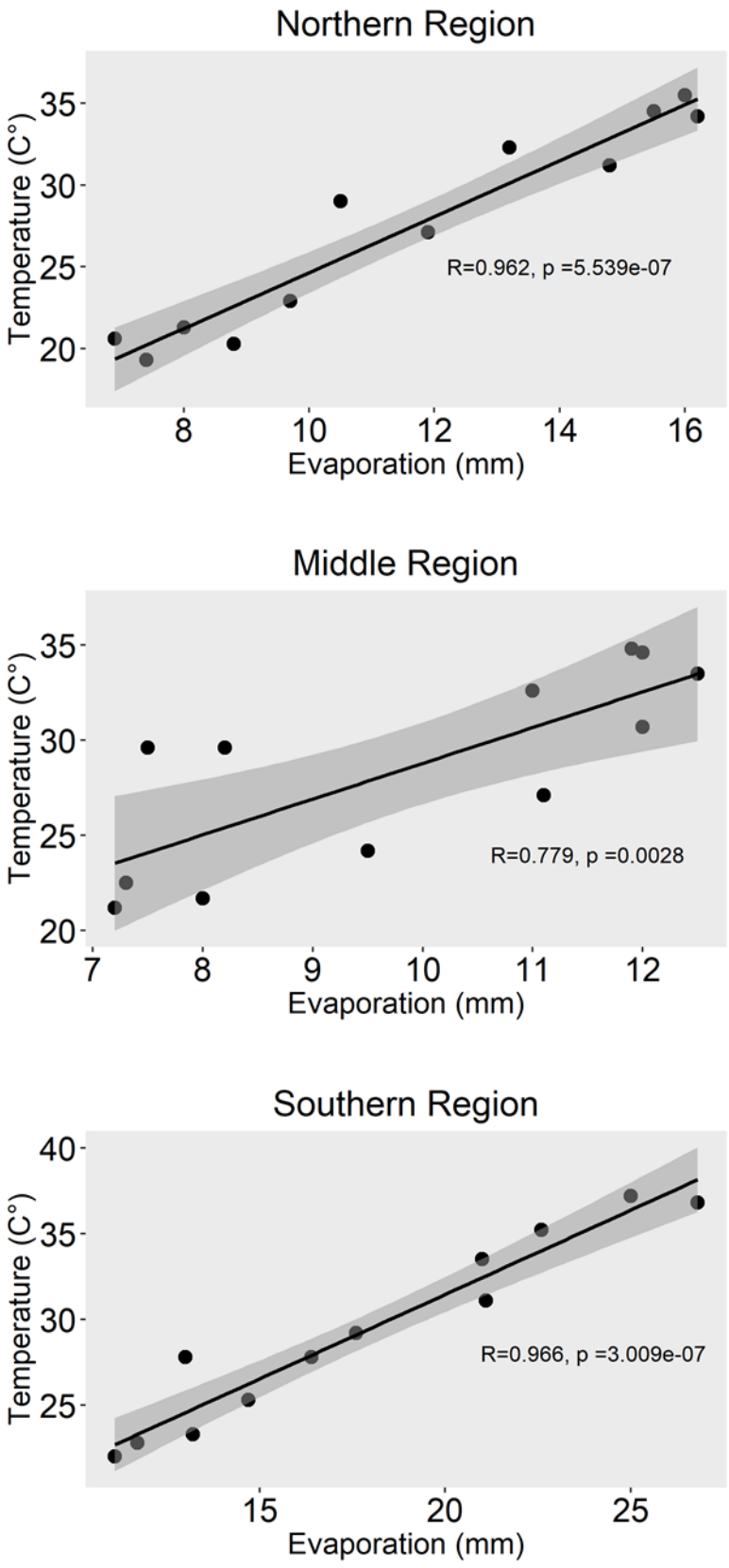


Figure 4.5: Correlation coefficient graph illustrating the effect of temperature on changes in evaporation rate. The data represent the monthly mean of daily data collected by three ground stations between 1943 and 2002 (Ras-Sudr in the northern region, El-Tor in the middle region, and Sharm El-Sheikh in the southern region).

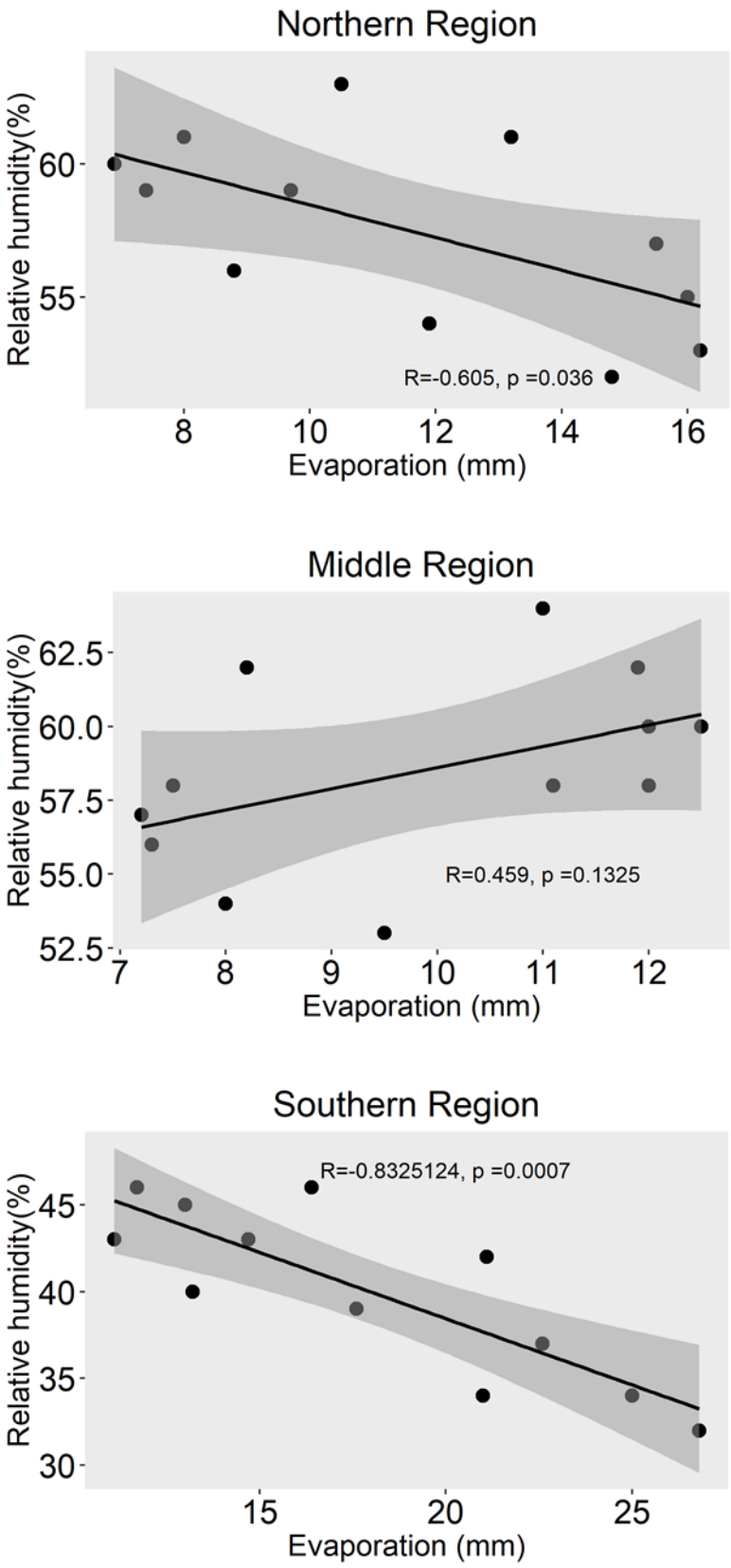


Figure 4.6: Correlation coefficient graph describing the effect of relative humidity on evaporation rate. The data represent the monthly mean of daily data collected by three ground stations between 1943 and 2002 (Ras-Sudr in the northern region, El-Tor in the middle region, and Sharm El-Sheikh in the southern region).

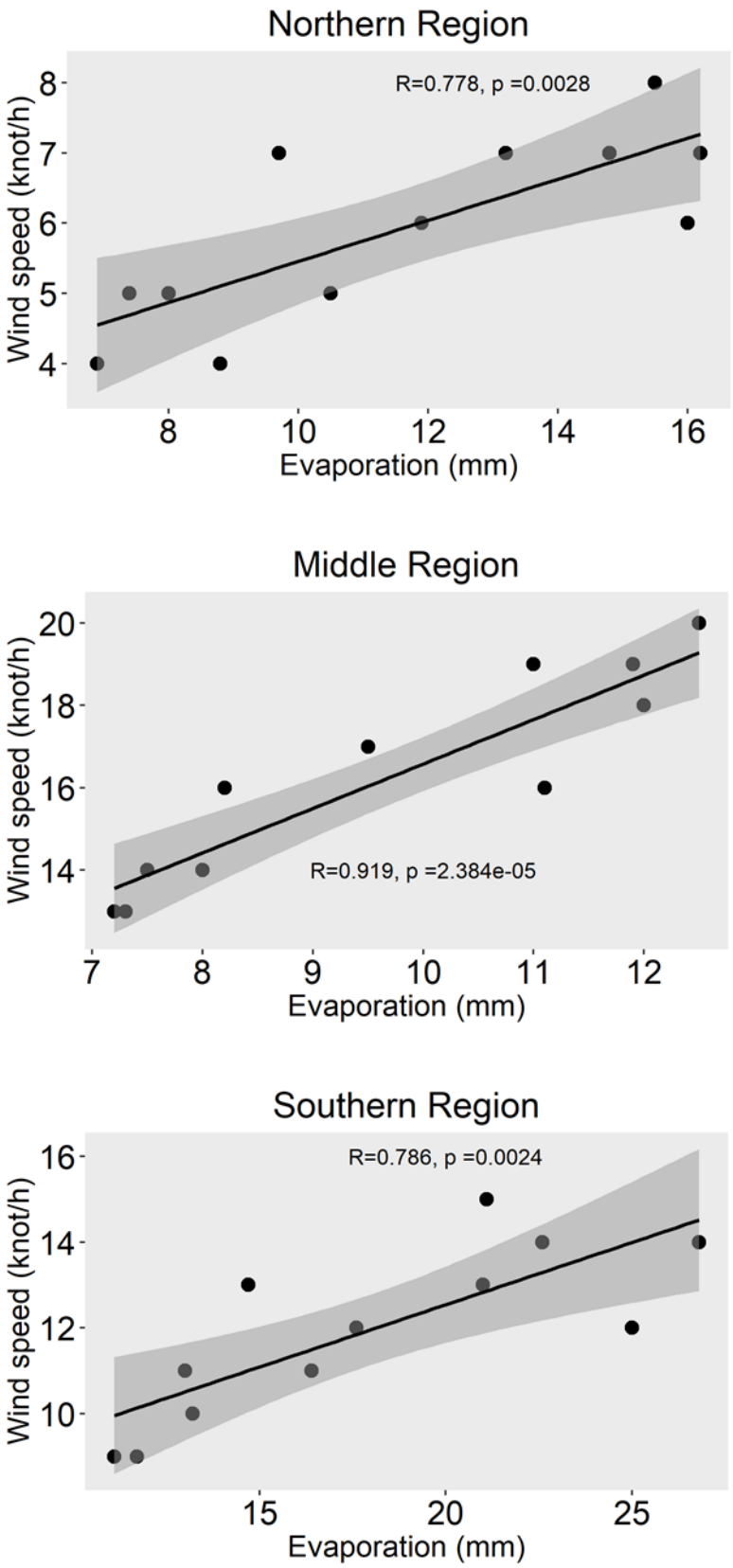


Figure 4.7: Correlation coefficient graph describing the effect of relative wind speed on evaporation rate. The data represent the monthly mean of daily data collected by three ground stations between 1943 and 2002 (Ras-Sudr in the northern region, El-Tor in the middle region, and Sharm El-Sheikh in the southern region).

Table 4.4: Correlational strengths between temperature and evaporation (T/Evap), relative humidity and evaporation (RH/Evap), and wind speed and evaporation (WS/Evap).

Region	T/Evap	RH/Evap	WS/Evap
Northern-Region	R=0.962 P=5.539e-07	R=-0.605 P=0.036	R=0.778 P=0.0028
Middle-Region	R=0.779 P=0.0028	R=0.459 P=0.1325	R=0.919 P=2.384e-05
Southern-Region	R=0.966 P=3.009e-07	R=-0.8325 P=0.0007	R=0.786 P=0.0024

The correlation coefficient was calculated between the permeability and infiltration for all lithological materials found at the study site. A strongly significant relationship was shown between the two factors, where R = 0.916 and p = 0.010 (when taking into consideration the existence of one outlier) (Figure 4.8).

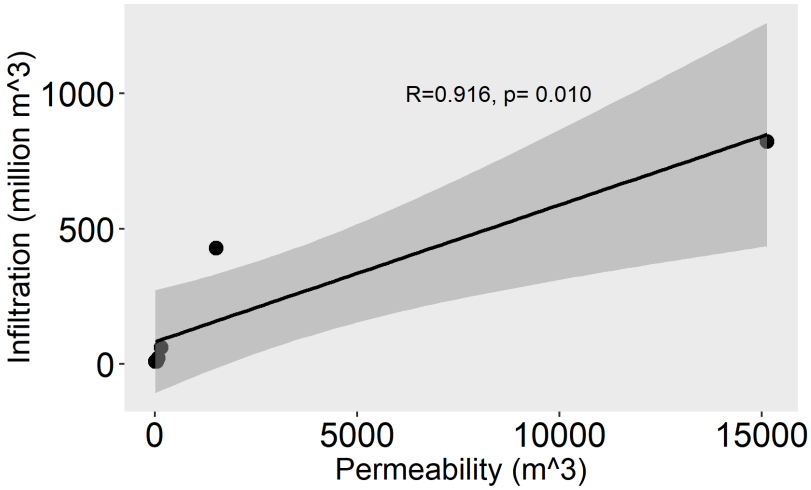


Figure 4.8: Correlation coefficients between the permeabilities and infiltration rates of the different lithological materials.

Chapters 3 and 4 discussed the geomorphology and climatic conditions of the study site for the purpose of visualizing the progression of the water cycle in that area (Figure 4.9) Primarily driving this cycle is the Gulf of Suez (c), where the highest evaporation rates can be found. Other sources of evaporation include rainfall and soil moisture. This evaporation continuously generates clouds until they are overloaded and precipitate over the study site. A portion of the rainfall occurs directly over the plain (b), where infiltration can take place. A larger portion precipitates over the complex mountainous area (a) referred to as the basement complex. Despite indeed containing micro- and megastructures that could allow for infiltration, this complex features a solid formation and steep slope that counteracts these opportunities.

Instead, this slope accelerates the movement of water during high intensity precipitation events, often resulting in flash floods.

When rain falls directly over the plain (b), it quickly becomes partially saturated, which leads to additional water instead being returned to the Gulf of Suez. This process is in accordance with Kuchmann (2000), who states that 80% of evaporation returns to the ocean/seas.

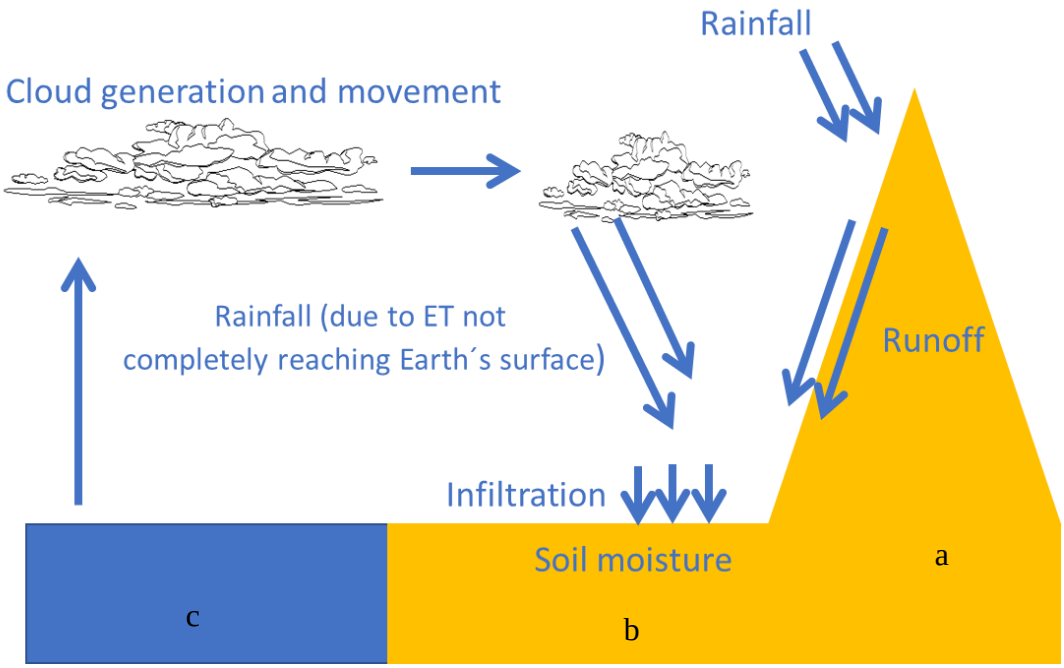


Figure 4.9: Visualization of the water cycle at the test site, featuring the basement complex (a), the plain (b), and the Gulf of Suez (c).

4.4 Discussion

The identification of the factors influencing water cycle equilibrium at the test site was of paramount importance, given that the entire area primarily depends on groundwater for all aspects of life. However, this identification is not easily achieved and is often disregarded in literature due to the limited number of ground stations, the high price of available data, and the lack of previous work available.

In addition to the climatic variables generally at play in arid areas, the test site featured geographic variability. For example, the Middle and Southern regions received more rainfall than the Northern region. The effect of temperature on evaporation rate was more pronounced in the Northern and Southern regions when compared to the Middle region. Moreover, relative humidity in the Northern and Southern regions displayed a negative correlation with evaporation, while no significant correlation at all was found for the Middle region. In all three

regions, there was a strong positive correlation between wind speed and evaporation, with the correlation being strongest for the Middle region. The highest runoff volume and peak was seen in the Northern region, while the lowest was found in the Southern region (Sherief, 2008).

Taken together, these results point toward a need for a greater understanding of the factors influencing the water budget at the study site, with emphasis on the differences between regions. This could be achieved via alternatives to the limited rain gauges currently installed, which would ideally provide wider spatiotemporal coverage of the area.

5. CHANGES IN LAND COVER AND LAND USE

Determining whether the test site is active or dormant is highly important, as it establishes whether or not additional studies are required. A dormant site would not require as frequent studies as an active site, for example. This determination can be done by studying the increase or decrease in the groundwater consumption rate. Calculations by Attia (1930) determined the extraction rate from the Quaternary aquifer in 1930 to be 0.18×10^6 m^3 /year. Then, in 1972 Gilboa declared that consumption was to reach 1.1×10^6 m^3 /year in 1972. Gorski and Ghodeif (2000) recorded the increase in the consumption rate to be 9.5×10^6 m^3 /year in 2000 and 11.0×10^6 m^3 /year in 2011, using ground water level measurements. Finally, Ahmed *et al*. in 2014 incorporated these values into calculations for future extractions via application of a second-order polynomial function. He estimated the yearly extraction to be 15×10^6, 30×10^6, and 65×10^6 m^3 /year in 2020, 2050, and 2100, respectively. Ahmed also predicted that water quality would begin to deteriorate due to saltwater intrusion in 45 years. Controlling withdrawal portions is therefore critical for avoiding the risk of depletion and for maintaining a balanced water level in the Quaternary aquifer. While consumption rates at the test site have indeed increased over time, indicating an associated increase in human activity, no published data is available regarding changes in the local ground water level over time. By tracking changes in land cover and land use over time at the test site, one could possibly confirm or dismiss the previous claims regarding local water consumption, namely, that it is gradually increasing. These measurements could also provide additional information on the area, all verifying the strength of remote sensing in reliable data acquisition.

5.1 General Background

Change detection can be described as the method used to track the changes between images at various time points (Singh, 1989; Zhu, 2017). It serves as one of the most important applications of remote sensing, as it can provide a foundation for the understanding of the relationships and interactions between humans and natural phenomena, which can then be used to better manage natural resources (Afify, 2011) Change detection has been prominently used in literature to measure changes in forest sizes, crop varieties, buildings, soil erosion, shorelines, and wadi fillings (Coppin and Bauer, 2009; Green *et al*., 1994; Beurs and Henebry, 2004). The vast amount of remote sensing imagery available allows for better monitoring of spatiotemporal changes at local, regional, and global levels (Jianya, 2008). This type of remote sensing-derived data is also especially useful for acquiring information about areas that are difficult or dangerous to access (Kennedy *et al*., 2009).

Data selection serves as one of the most impactful steps when establishing a successful technique for change detection (Weber, 2001). For efficient selection of data, it is necessary to first investigate the area under study. This can be done through analyzation of the spatial distribution and temporal scales of land changes, the spectral characteristics of the images, and the identification of the areas that can be sensed remotely (Li *et al*., 2008). For comparison of multitemporal images, it is recommended to select images acquired from the same type of sensors, with the same spectral and spatial resolution, and from the same seasonal time frame. This is done in order to minimize unwanted variance due to changing factors such as sun angle, as well as seasonal and phenological differences (Li *et al*., 2008; Coppin and Bauer, 1996).

5.1.1 Pre-processing

Image selection must be followed by radiometric correction. Radiometric conditions are affected easily and by various factors, among them differences in imaging sensors and dates, solar altitude, acquisition angles, meteorological conditions, and cloud cover (Paolini *et al*., 2006; Jianyaa, 2008). These factors can influence the accuracy of change detection algorithms. Therefore, to reduce these effects, radiometric correction should be applied (Paolini *et al*., 2006; Jianyaa, 2008). Radiometric correlation describes the process of reducing or removing the inconsistency between the values recorded by sensors and the spectral reflectance and radiance of the object. According to Paolini *et al*. (2006) and Jianyaa (2008), two types of radiometric correction can include relative radiometric correction and absolute radiometric correction. The typical method, absolute radiometric correction, involves adjusting the radiation value to the standard value using, for example: atmospheric radiation transmission codes, spectral curves (in the lab), dark object transmission codes, and rectification of the scene by removal of dark objects. The authors also, however, referred to these methods as impractical and expensive, also mentioning that the majority of authors rely on relative radiometric correction when an image is regarded as a reference image. The radiation characteristics of a later image is then adjusted to match the former one. This can be done by correction with either histogram regularization or with a fixed object. This type of correction can then remove or reduce the effects of the atmosphere and sensor using a simple algorithm, making it widely popular.

Image registration follows and involves adjusting the different datasets to one coordinate system, regardless of whether the data are derived from multiple sensors, times, or viewpoints. Jianyaa, (2008) and Li *et al*. (2008) emphasize that image registration accuracy should be high in order to avoid false positives of change in the area. However, applying a high-accuracy resolution to multitemporal and multisensor-derived remote sensing images is

challenging due to several factors. These include the curvature and rotation of the earth as well as differences in imaging models, angles, and conditions. This is especially the case in mountainous regions and urban areas. In these cases, general image registration methods are ineffective, and ortho-rectification is required (Paolini *et al*., 2006; Jianyaa, 2008).

5.1.2 Information Extraction

The final step of radiometric correction involves the clustering of existing units by the using carefully selected remote sensing imagery. This step can be performed through three procedures: unsupervised classification, supervised classification, and application of NDVI and NDBI equations. All of the procedures mentioned are machine learning algorithms used to identify the useful unknown classes (Kotsiantis, 2007). Unsupervised classification can help build a rough idea about the classes at the test site (Afify, 2011), but should then be followed by supervised classification in order to automatically categorize all pixels in the image into land cover classes or themes (Lillesand& Kiefer, 1994). Supervised classification is a very sensitive method that requires previous knowledge about the test site (Rogan and Chen, 2004). As stated by Franklin *et al*. (2003), most studies monitoring changes in land cover and land use depend on traditional image classification algorithms that assume that image data are normally distributed, that the images are H-resolution2, and that pixels are composed entirely of single land cover or land use types. In most studies, this was achieved by maximum likelihood supervised classification.

The Normalized Difference Vegetation Index (NDVI) is a remotely sensed measure of vegetation. It depends on both multispectral remotely sensed images and various band combinations (Forkel *et al*., 2013; Gandhi *et al*., 2015). The Normalized Difference Building Index (NDBI), on the other hand, allows for the mapping of changes in built-up areas (Varshney, 2013). The Normalized Difference Water Index (NDWI) allows for the mapping of changes in water bodies (Varshney, 2012). The vegetation band combination for the Landsat5 data was (4,3,2) and (5,4,3) for the Landsat8. The buildings band combination was (5,4,3) and (6,5,4), respectively. The water band combination for the Landsat5 was (4,5,3) and for the Landsat8, (5,6,4). Gandhi *et al*. (2015), and Estoque and Murayama, (2015) define the equations of the NDVI, NDBI, and NDWI as the following:

$$NDVI = \frac{NIR - RED}{NIR + RED} \qquad \text{eq (5.1)}$$

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \qquad \text{eq (5.2)}$$

$$NDWI = \frac{NIR - SWIR}{NIR + SWIR} \qquad \text{eq (5.3)}$$

In eq (5.1), (5.2), and (5.3), (NIR) represents bands number 4 and 5, (RED) represents bands number 3 and 4, and (SWIR) represents bands number 5 and 6 in the Landsat5 and Landsat8, respectively.

5.1.3. Change Detection Algorithms

To extract the changed/unchanged pixels from the remote sensing imagery over time, three types of algorithms are commonly used. These types include image differencing, image rationing, and image regression. The image differencing algorithm is based on the pixelwise subtraction of the remote sensing imagery of time t^2 from time t^1, for the same area. The mathematical formulation of the algorithm was stated by Ilsever and Ünsalan (2013) and Théau (2006) as:

$$I\,d(x,y) = I^1(x,y) - I^2(x,y) \qquad \text{eq (5.4)}$$

where I^1 and I^2 are the images obtained from t^1 and t^2 and (x, y) represent the coordinates of the pixels. The resulting image, *Id*, reflects the intensity difference between I^1 and I^2. However, this technique will only work if images are properly registered. If no change in time is present, this results in pixel values of 0. Should a change occur, these values will be either positive or negative (Ilsever and Ünsalan, 2013) (Figure 5.1).

Image rationing is the second algorithm mentioned by Ilsever and Ünsalan (2013) and was described to be as simple as image differencing method. This method depends on the ratio between remote sensing imageries (both pixel-wise and band-wise ratios) at two time points, t^1 and t^2. However, it is not as widely used as the first algorithm (Ilsever and Ünsalan, 2013). According to Théau (2006), the formula of the algorithm is the following:

$$Ir(x,y) = I^1(x,y) \,/\, I^2(x,y) \qquad \text{eq (5.5)}$$

The Ir, or ratio image, ranges in values from 0 to ∞. If the intensities of the pixel values are equal, the resultant pixel takes the value of 1 (Figure 5.1).

The third algorithm, the image regression algorithm, assumes a linear relationship between pixel values from the same area at two different time points (Ilsever and Ünsalan, 2013). This algorithm depends on the assumption that unchanged pixels between two dates will create a uniform line. The outliers of the regression are then considered to comprise the changed area (Figure 5.2).

$$I_2(x, y) = aI_1(x, y) + b \qquad \text{eq (5.6)}$$

In equation 5.6, I_2 represents the image obtained at time t_2 and I_1 the image obtained at time t_1. The value of I_2 can be easily calculated by application of the regression equation. This method has the advantage of reducing the impact of radiometric heterogeneity (e.g. the atmosphere, sun angle, sensor calibration) between multidate images (Théau, 2006).

5.1.4 Post-processing

The accuracy of the results formulated by the change detection algorithms depends on several factors. These include the precision of the image registration, the availability of ground reference data, the complexity of the area, the algorithm used, and the skills of the analyst (Jianyaa *et al*., 2008). As it can be difficult to acquire multitemporal ground reference datasets, three alternatives exist. These were stated by Jianyaa *et al*. (2008) to be: field surveying with the assistance of historical GIS data (simultaneously or within temporal proximity), use of high-resolution images, and visual interpretation. The selection of a suitable method depends on the intended application.

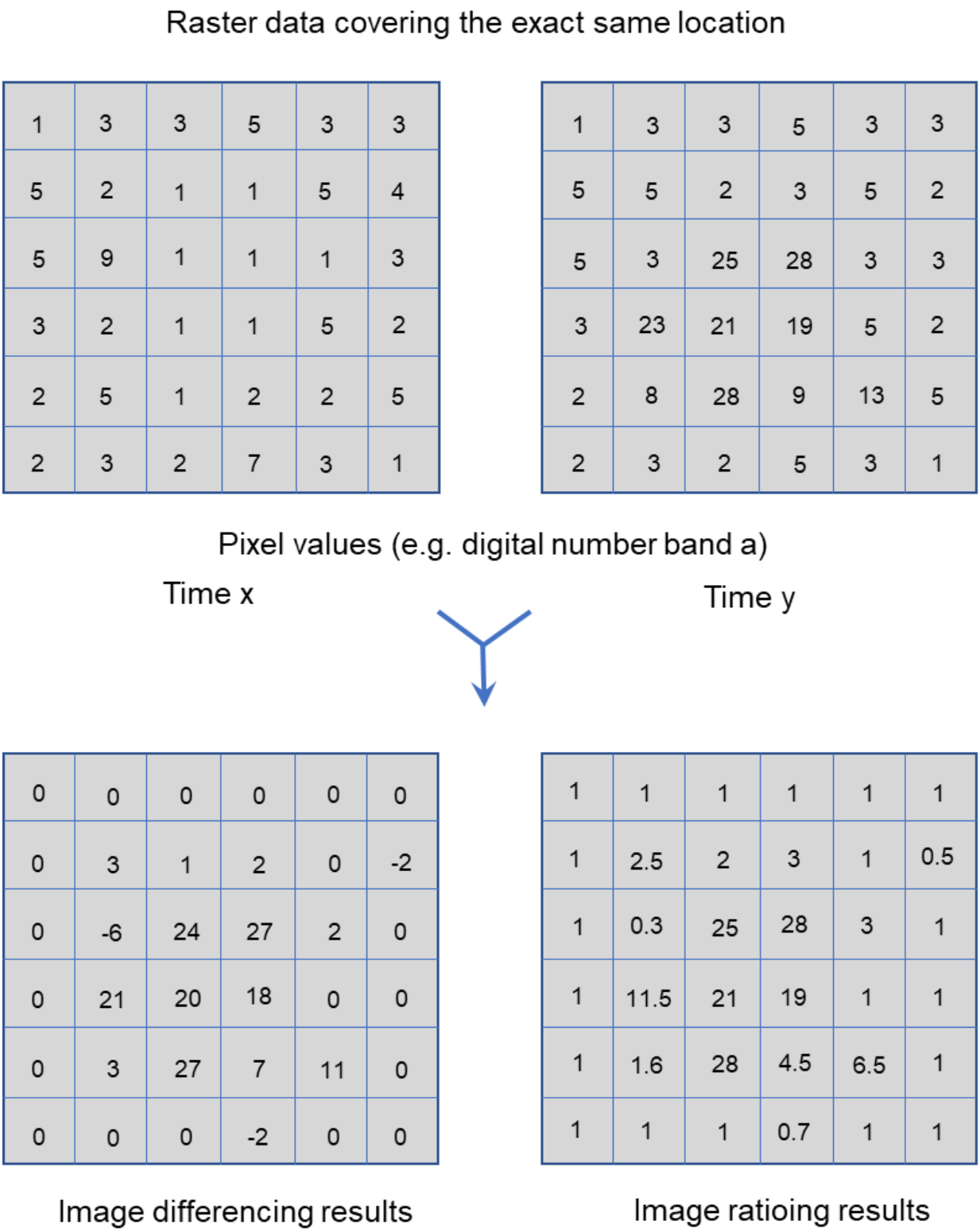


Figure 5.1: Example of image differencing and image ratioing procedures. (after Théau, 2006).

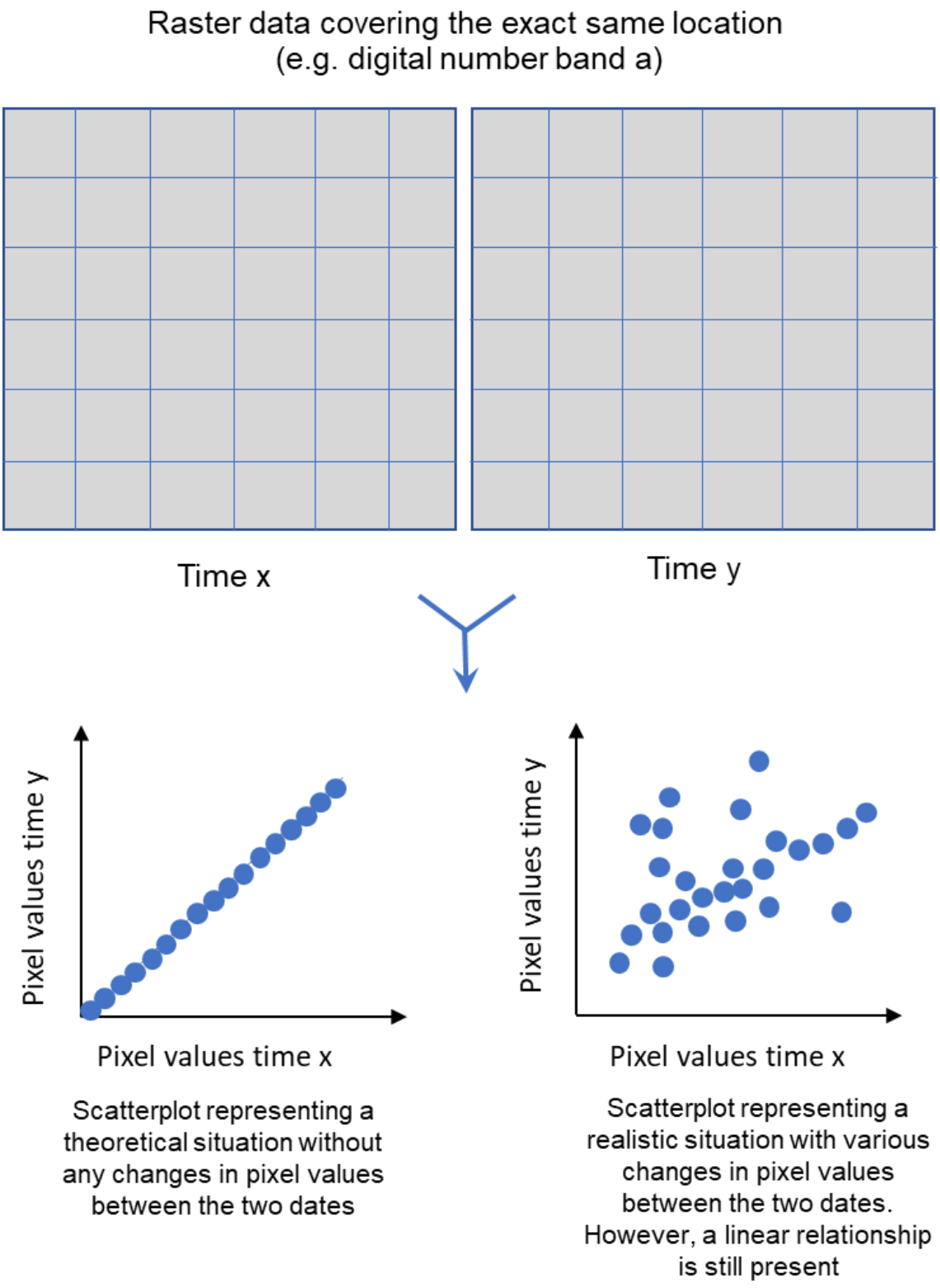


Figure 5.2: Examples and principles of image regression (after Théau, 2006).

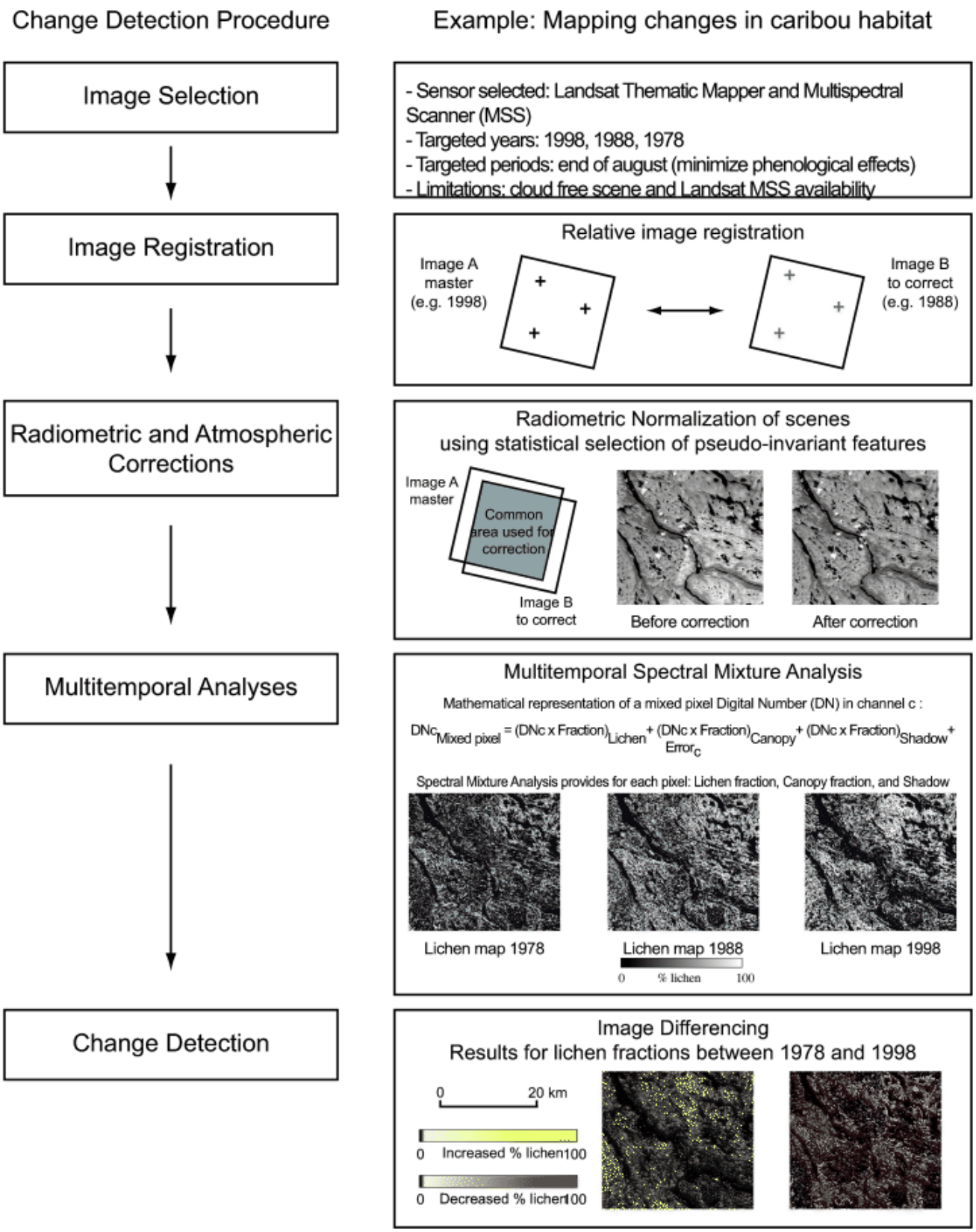


Figure 5.3: Elaboration on change detection procedures (after Théau, 2006).

5.2 Data and Methods

Data was provided by two Landsat5 scenes downloaded from the Google Earth Engine covering the years 2000 and 2010. Further data was sourced from a Landsat8 scene from 2019. Image registration involved georeferencing the data to (WGS 1984 UTM Zone 36N). The data was also treated by an image analysis tool in the ArcGIS10.5 software, where gamma stretch, brightness, contrast, and transparency were brought to the same values for the three scenes. Additionally, the statistical relationships between the pixels of each scene were translated into standard deviations. The three scenes were then used to determine the changes that took place along the eastern side of the Gulf of Suez. Each scene was checked in bands (5,4,3) and (6,5,4) for the Landsat5 and Landsat8, respectively. Unfortunately, over large areas, changes were difficult to notice due to their sparce coverage in the desert. Therefore, three separate locations were selected by the author to represent the changes occurring between the years 2000 and 2019 (Figure 5.4). The first location was a section of the city of Ras-Sudr, which was located from 32°93′0″E to 32°49′30″E longitude, and 29°43′0″N to 29°30′0″N latitude and featured an area of 20.6km^2. The second location was a portion of the city of El-Tor located between 33°34′0″E and 33°44′30″E longitude and 28°23′30″N and 28°10′30″N latitude, with an area of 36.9km^2. The third location consisted of a part of the city of Sharm El-Sheikh located between 34°21′30″E and 34°24′30″E longitude and 27°59′30″ and 27°54′30″ latitude, with an area of 1.7km^2.

As previously mentioned, the primary goal behind detecting changes at the test site was to obtain a rough idea about the decreases or increases in water consumption in that area, which would then allow for the formulation of future studies. The relevant units for this purpose were vegetation and buildings, which were quantified via application of the NDVI and NDBI formulas (Section 5.1.2) The number of pixels belonging to each class were counted and entered into an excel sheet, converted into areas (km^2), and finally, represented by bar plots. Following determination of the increases/decreases of each class of cover (vegetation, buildings), the resultant classes were verified to assure the efficacy of the procedure. To accomplish this, accuracy assesment tools in Arctoolbox were utilized using 50 random points for Sharm El-Shiekh and Ras-Sudr, and 100 points for El-Tor. The selected tool automatically calculated the values for each class at each point and saved it to an attribute table. The author, in contrast, determined the class of each point using the main Landsat scenes. The accuracy of the resultant classes were subsequently calculated using the confusion matrix tool in Arctoolbox.

The image differencing algorithm was then used to calculate the added, removed and unchanged pixels between the 2000 and 2019 scenes. A raster calculator tool in Arctoolbox (ArcGIS 10.5) was used to calculate the differences between each individual pixel in the year 2000 scene with the compatible pixel of the 2019 scene. The number of pixels in each class was counted and entered into an excel sheet and then converted into areas. This conversion was achieved by multiplying each value by 30×30, converting to square kilometers, and then converted square kilometers in percentages. These percentages were then represented as bar plots created using RStudio software and its related packages (ggplot2 and reshape2).

A rough numerical estimate of the water consumption rate in El-Tor was calculated using the results of the change detection algorithms together with the information collected regarding crop types. Three types of crops were found in El-Tor: mango, creep, and olive. No information, however, was available regarding the locations of each type of crop or the irrigation rate required for each. Therefore, the mean consumption rate for all crops was calculated for one acre, followed by a calculated for the entire El-Tor area, using the irrigation rates published on public Egyptian websites.

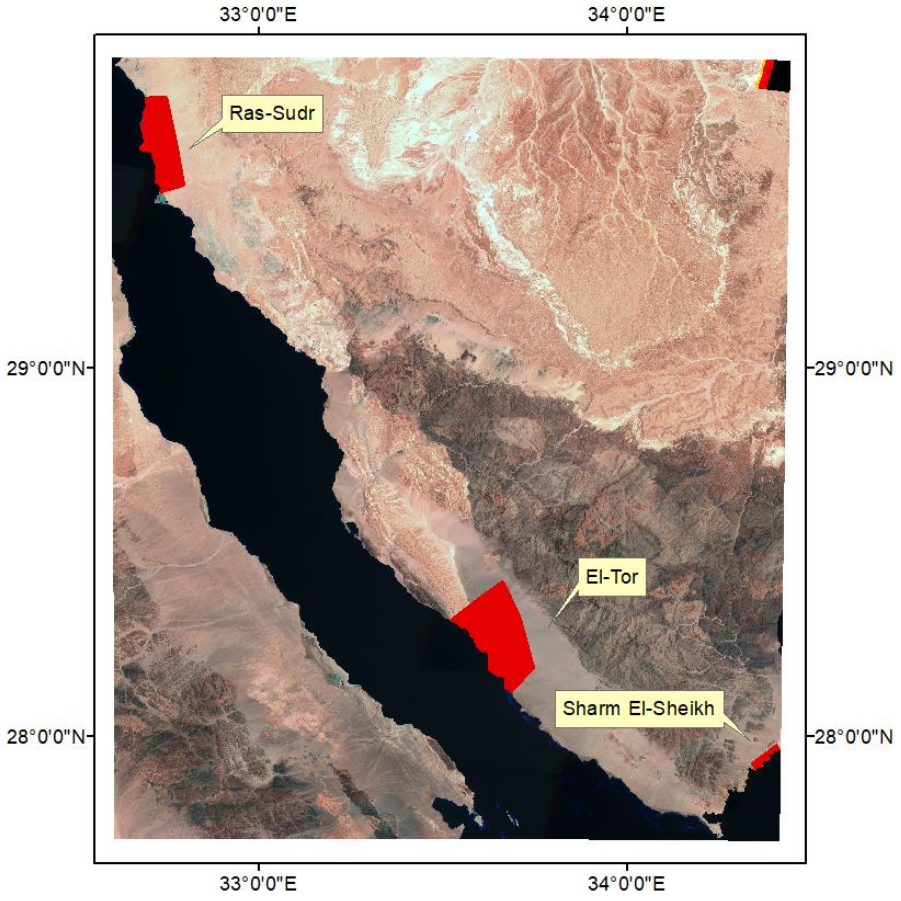


Figure 5.4: Locations of the areas selected for testing of the change detection technique.

5.3 Results:

The cities of Ras-Sudr, El-Tor, and Sharm El-Sheikh all experienced an increase in land cover and land use between the period of 2000 to 2019, as determined by data from bands (5,4,3) and (6,5,4) of the Landsat5 and Landsat8, respectively (Figures 5.5, 5.6, and 5.7). Due to this increase, the exact changes were calculated for each city using the NDVI and NDBI algorithms, resulting in five classes per scene. The classes were manually adjusted by the author in order to better compare against the data from the Landsat scenes. Ras-Sudr and El-Tor featured two classes of vegetation, one of bare soil, and one class of buildings. Sharm El-Sheikh featured two classes of vegetation and one class each of streets, buildings, and bare soil (Figures 5.8, 5.9, and 5.10; Table 5.1).

In the present study, vegetation type was insignificant. However, the total area of vegetation and buildings combined was the most significant target for change. The total area of vegetation in Ras-Sudr increased from 0.072km^2 in 2000 to 1.395km^2 in 2019. The area covered by buildings also increased, from 1.304km^2 to 3.292km^2. Total vegetation in El-Tor increased from 0.128km^2 to 0.959km^2 and the area covered by buildings increased from 0.390km^2 to 1.809km^2. The vegetated area in Sharm El-Sheikh increased from 0.117km^2 to 0.234km^2, while the area covered by streets increased from 0km to 0.342km^2, and that by buildings from 0.102km^2 to 0.2268km^2. These increases were calculated through comparison of scenes from 2000 and 2019, taking into consideration the different total areas of each location at that time point (Table 5.1, and Figure 5.11). In El-Tor, one type of vegetation appeared in the 2000 and 2010 scenes and second type of vegetation appeared in 2019. Three types of bare soil appeared in the 2000 and 2010 scenes, with one of them disappearing by 2019. This could, however, be related to differences in sensors (Table 5.1).

Classification was followed by a verification process intended to test the degree of correspondence between the original Landsat scenes and the images that were classified. The random verification points chosen are shown below (Figures 5.12, 5.13, and 5.14). Varying accuracy was observed at current stage. The classified image of Ras-Sudr featured an 80%, 74%, and 84% accuracy in 2000, 2010, and 2019, respectively. Each class of each scene also resulted in differing accuracies. The buildings class of Ras-Sudr in 2000, 2010, and 2019 resulted in the lowest accuracies, when compared to the remainder of the classes in the same scenes. In contrast, the vegetation and bare soil classes resulted in the highest accuracies (Table 5.2). The classes seen at El-Tor exhibited generally higher accuracies than those in Ras-Sudr, with 90%, 83%, and 82% accuracies in the 2000, 2010, and 2019 scenes, respectively. Of all

classes, the El-Tor vegetation class presented the lowest accuracy. The bare soil and buildings classes produced higher accuracies (Table 5.3).

As the lithology of the region generally produces pixel colors that lead to an underestimation of other classes, the author blinded bare soil T3 with bare soil T2 prior to beginning the entire process. Classes found in Sharm El-Sheikh featured accuracies of 84%, 80%, and 76% during the same time span. Of lowest accuracy were the buildings and streets classes, while vegetation and bare soil showed higher accuracy (Table 5.4 and Figure 5.15).

Classification accuracy was estimated by the calculation of the number of changed, unchanged, and added pixels, followed by their percentages of change. When the image differencing algorithm was applied to the three locations tested, the smallest percentage of removed area was found in Ras-Sudr (0.0831%), followed by El-Tor (2.8569%), and Sharm El-Sheikh (3.3689%). Percentages of unchanged areas were 59.3699%, 83.0683%, and 68.5026% for Ras-Sudr, El-Tor, and Sharm El-Sheikh, respectively. Area added was 40.5469km^2 for Ras-Sudr, 14.0746km^2 for El-Tor, and 28.1283km^2 for Sharm El-Sheikh (Figures 5.16, 5.17, 5.18 and 5.19; Table 5.5). The results of the changed, unchanged, and added pixels include the changes that took place over all existing classes, including bare soil.

Water consumption values were roughly calculated using the results from the change detection analysis, along with information collected in 2018 from local El-Tor farmers about crop types. The three types of crops being grown in El-Tor City included mangos, grapes, and olives. Mangos consumed 4331m^3 to 5724m^3 of water per acre, while grapes consumed 500m^3 to 1200m^3, and olives, 3000m^3 to 3500m^3. The mean of all three crop types was therefore 2610m^3 to 3474m^3 per acre per year, taking into consideration that 1 acre = 0.00404km^2. Therefore, the water consumption rate of the vegetated areas in El-Tor ranged from 83520m^3 to 111168m^3 in 2000, fluctuated between 117450m^3 and 156330m^3 in 2010, and ranged from 618570m^3 to 823338m^3 in 2019.

The entire Southern Sinai region covers 31272km^2 and was inhabited by 169822 individuals in 2016. Therefore, each km^2, on average, contained roughly 5 to 6 people. Each person consumed between 30.28L and 37.85L per day, or between 11052L and 13815L per year. This indicated that, for each km^2, water consumption ranged from 55.26m^3 to 69.075m^3 per year. The consumption per person in El-Tor ranged from 21.55m^3 to 26.939m^3 in 2000, from 29.044m^3 to 36.3058m^3 in 2010, and from 100m^3 to 125m^3 in 2019, assuming 5 people per km^2 (Table 5.6).

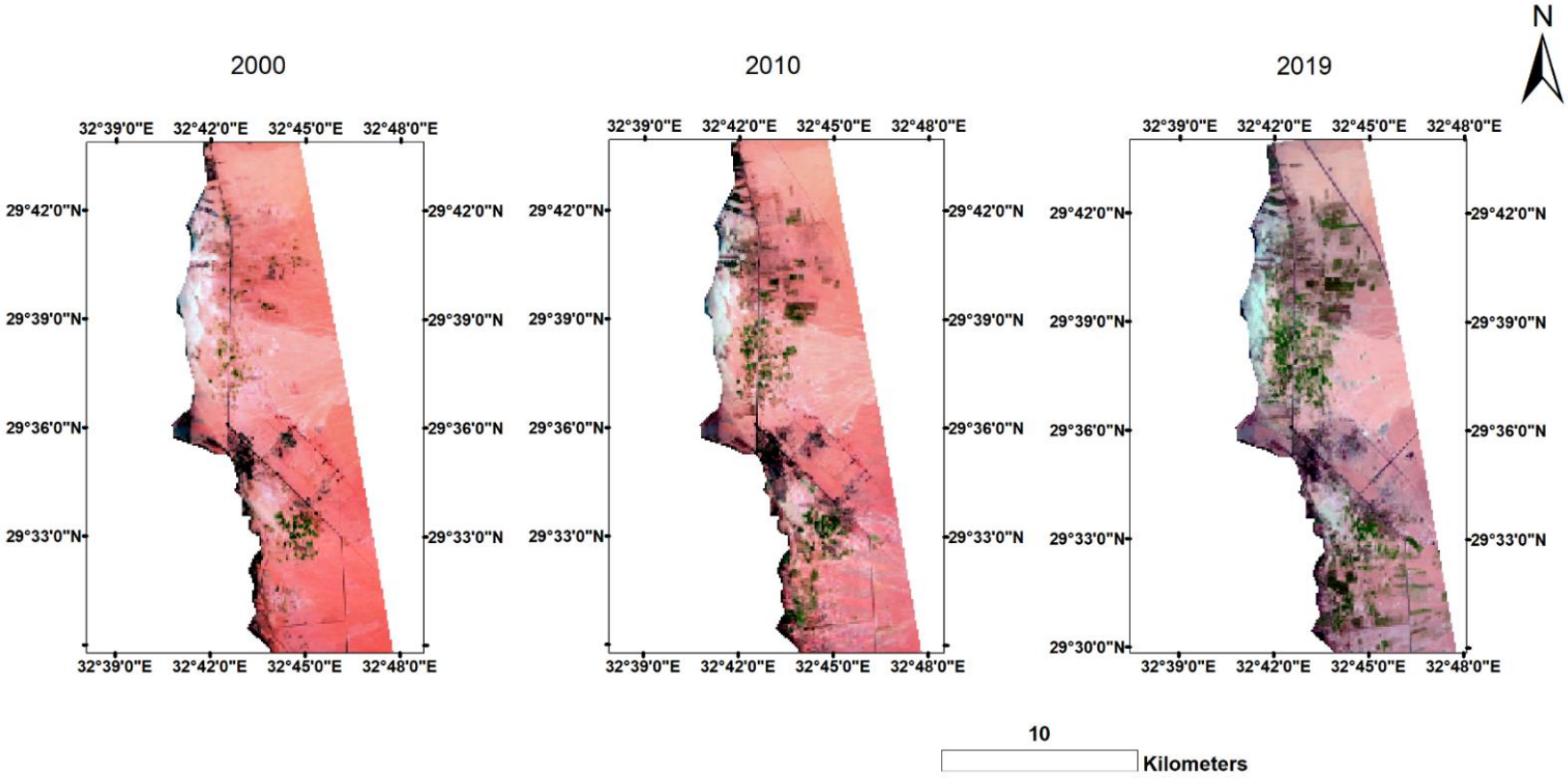


Figure 5.5: Two Landsat5 scenes showing Ras-Sudr in 2000 and 2010, and one Landsat8 scene showing Ras-Sudr in 2019. A (5,4,3) band combination was used for the first two scenes, while (6,5,4) band combination was used for the last scene.

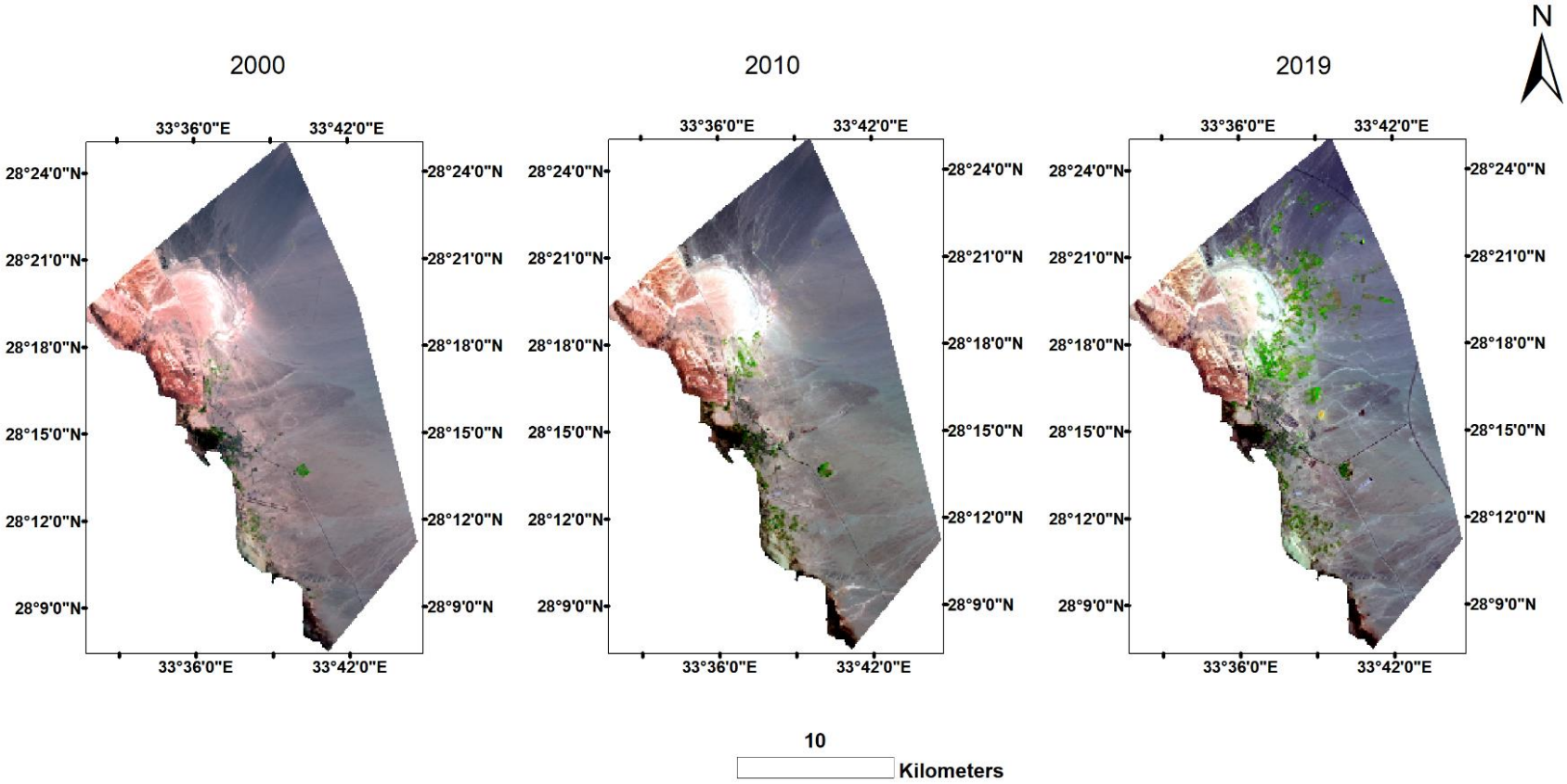


Figure 5.6: Two Landsat5 scenes showing the city of El-Tor in 2000 and 2010, along with one Landsat8 scene showing El-Tor in 2019. Acquired using (5,4,3) and (6,5,4) band combinations, respectively.

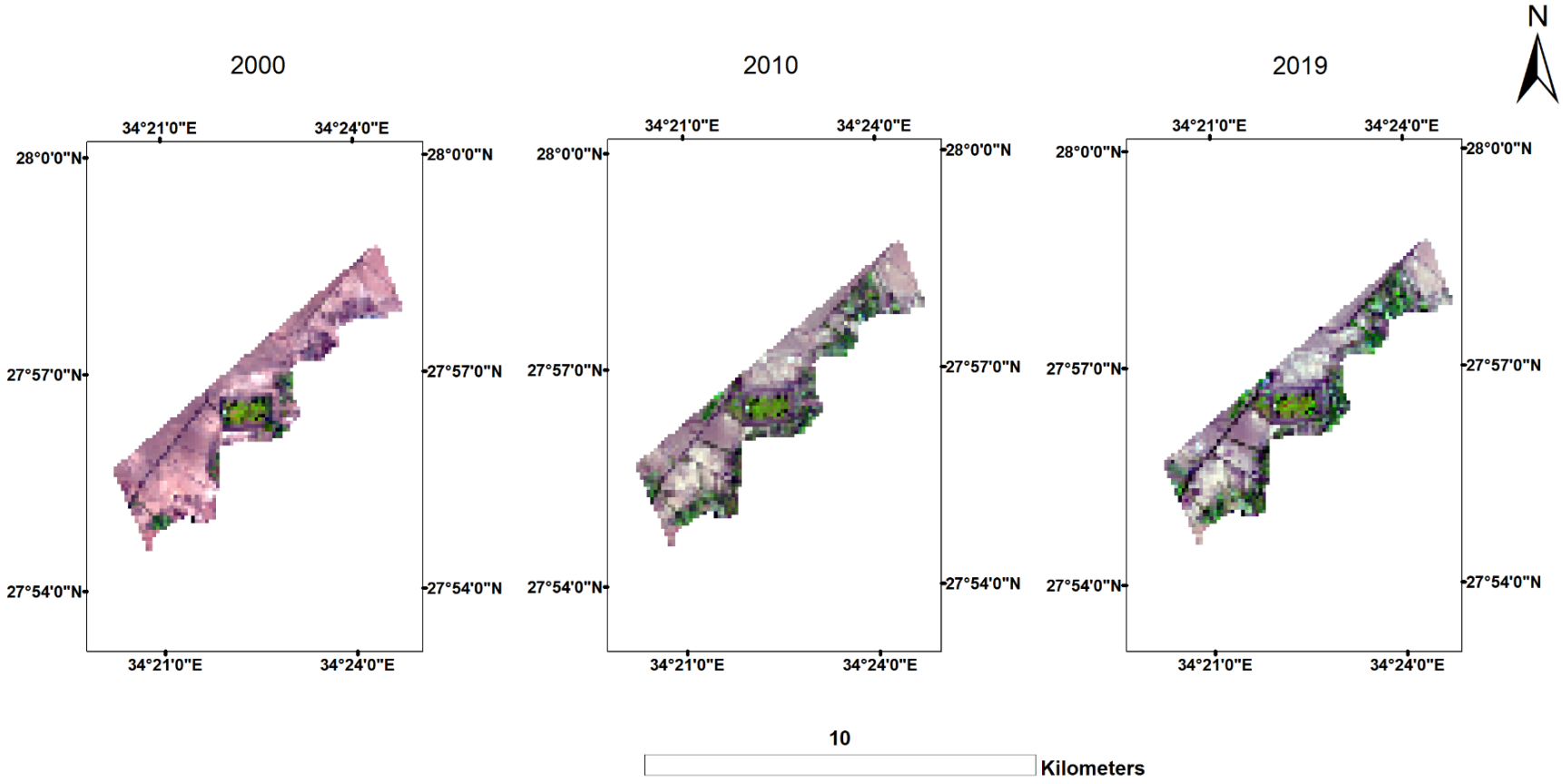


Figure 5.7: Two Landsat5 scenes showing Sharm El-Sheikh in 2000 and 2010, and one Landsat8 scene from 2019. Acquired using (5,4,3) and (6,5,4) band combinations.

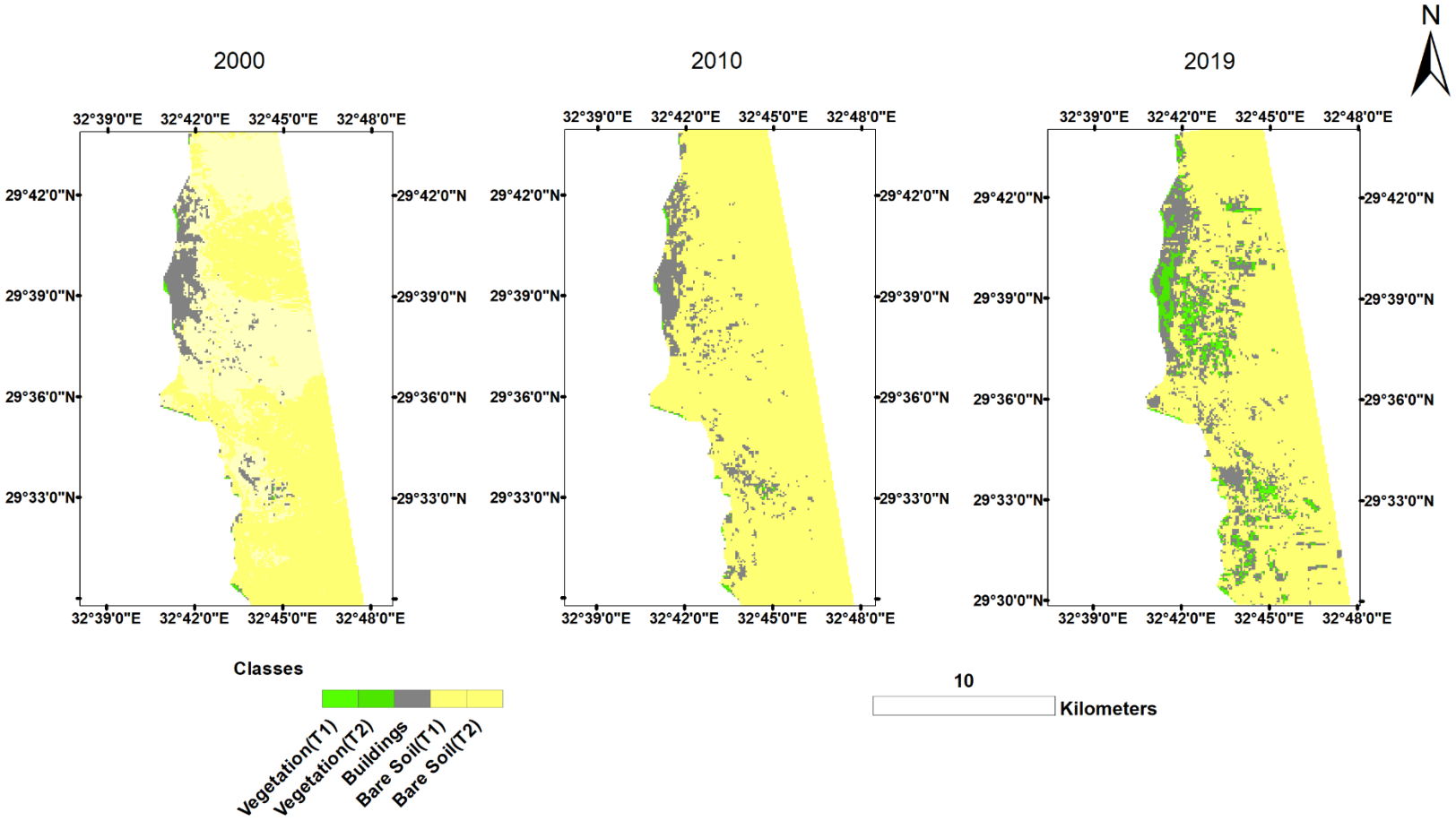


Figure 5.8: Classes found in each scene along with changes in area. Based on application of NDVI and NDBI formulas on data from Ras-Sudr.

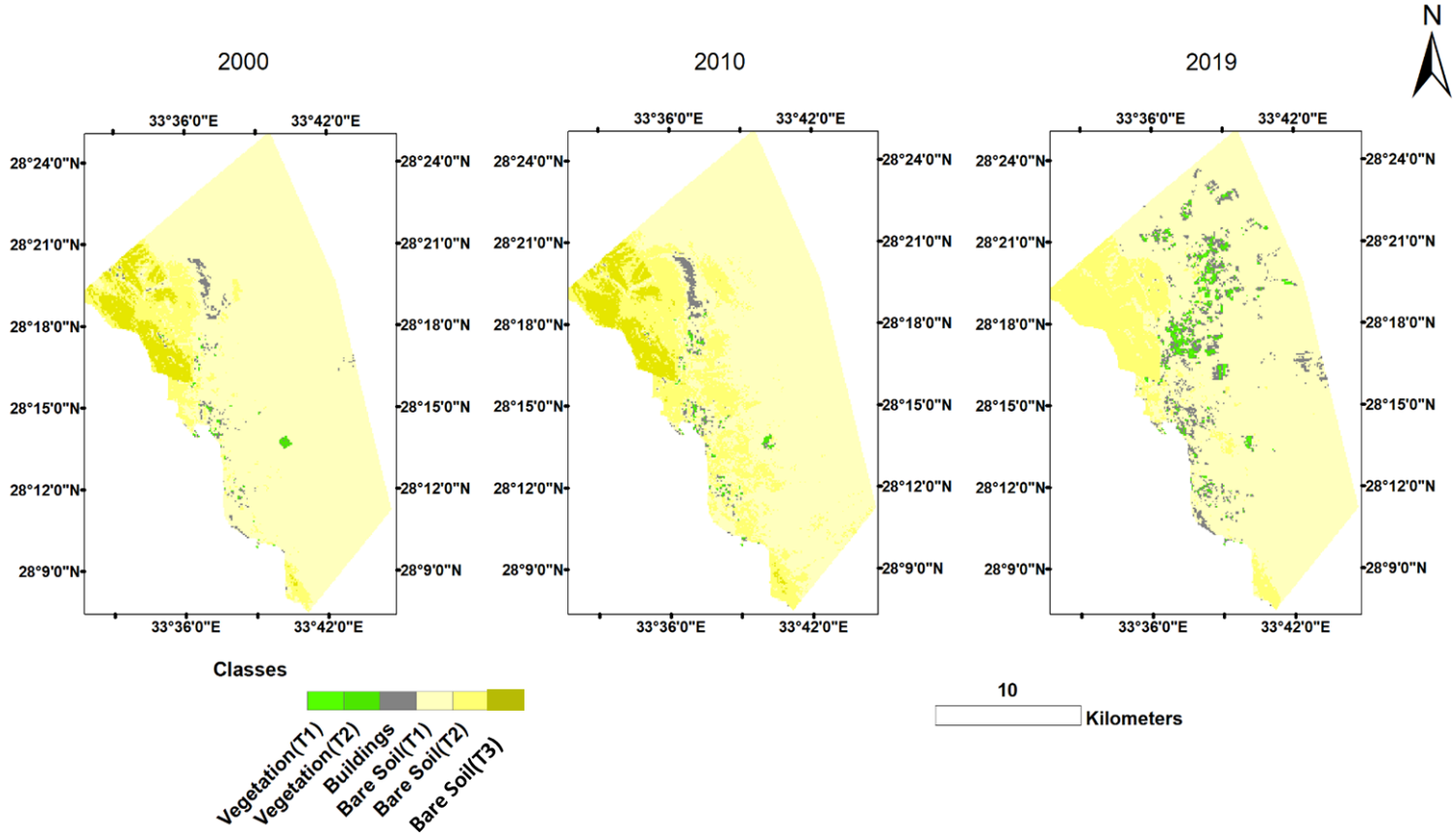


Figure 5.9: Classes found in each scene along with changes in area. Based on application of the NDVI and NDBI formulas on data from El-Tor.

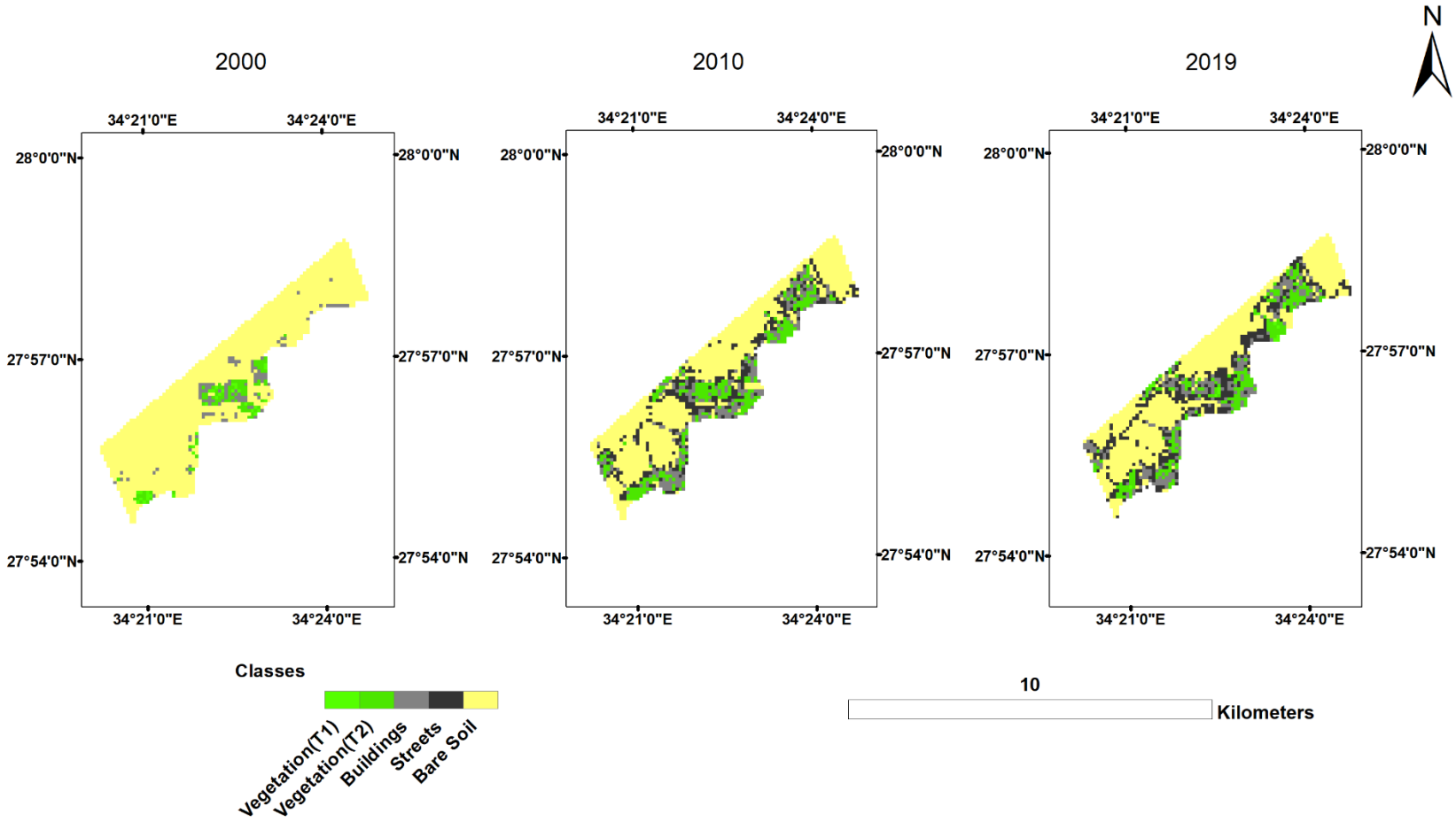


Figure 5.10: Classes found in each scene along with changes in area. Based on the application of the NDVI and NDBI formulas on data from Sharm El-Sheikh.

Table 5.1: Changes in area for each class from 2000 to 2019. Total areas were 20.6km^2, 36.9km^2, and 1.7km^2 for Ras-Sudr, El-Tor, and Sharm El-Sheikh, respectively.

Ras-Sudr			
	2000	2010	2019
VegetationT1(km^2)	0.0099	0.0153	0.2403
VegetationT2(km^2)	0.0621	0.0945	1.1556
Buildings(km^2)	1.3041	1.6371	3.2922
Bare Soil T1(km^2)	8.2764	8.7417	7.7436
Bare Soil T2 (km^2)	10.917	10.0809	8.1378
El-Tor			
	2000	2010	2019
VegetationT1 (km^2)	0.1287	0.1836	0.1908
VegetationT2 (km^2)	0	0	0.7806
Buildings (km^2)	0.3906	0.5256	1.8099
Bare Soil T1(km^2)	31.032	26.3547	28.7829
Bare Soil T2 (km^2)	3.429	7.6167	5.3055
Bare Soil T3 (km^2)	1.8774	2.1771	0
Sharm El-Sheikh			
	2000	2010	2019
Vegetation T1(km^2)	0.0297	0.0612	0.0144
Vegetation T2 (km^2)	0.0873	0.1863	0.2196
Buildings	0.1026	0.234	0.2268
Street	0	0.2718	0.3429
Bare soil	1.4859	0.9522	0.8793

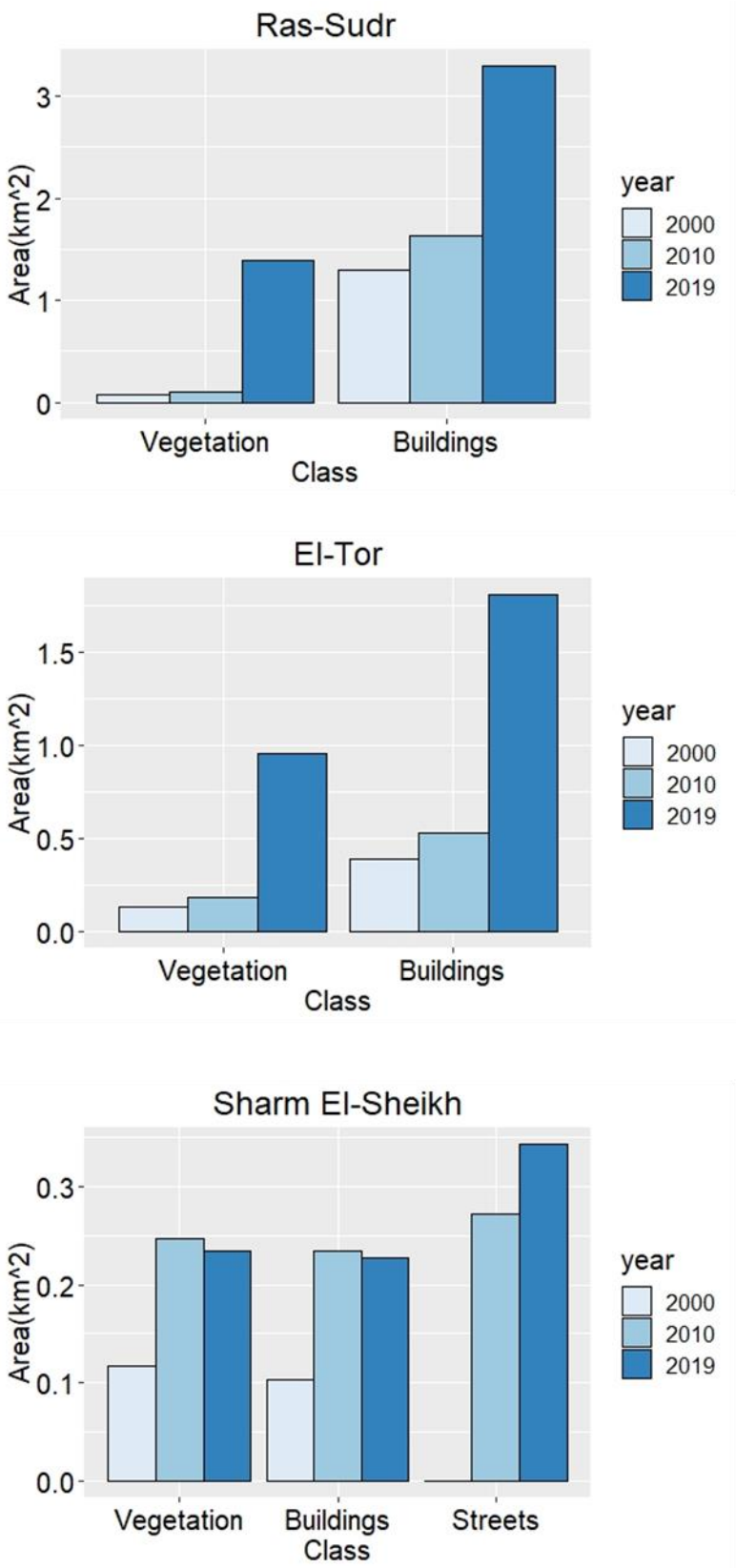


Figure 5.11: Bar plots showing changes in vegetation, building cover, and street cover. Based on three Landsat scenes from 2000, 2010, and 2019 focused on the three selected locations. Total tested areas were 20.6km^2, 36.9km^2, and 1.7km^2 for Ras-Sudr, El-Tor, and Sharm El-Sheikh, respectively.

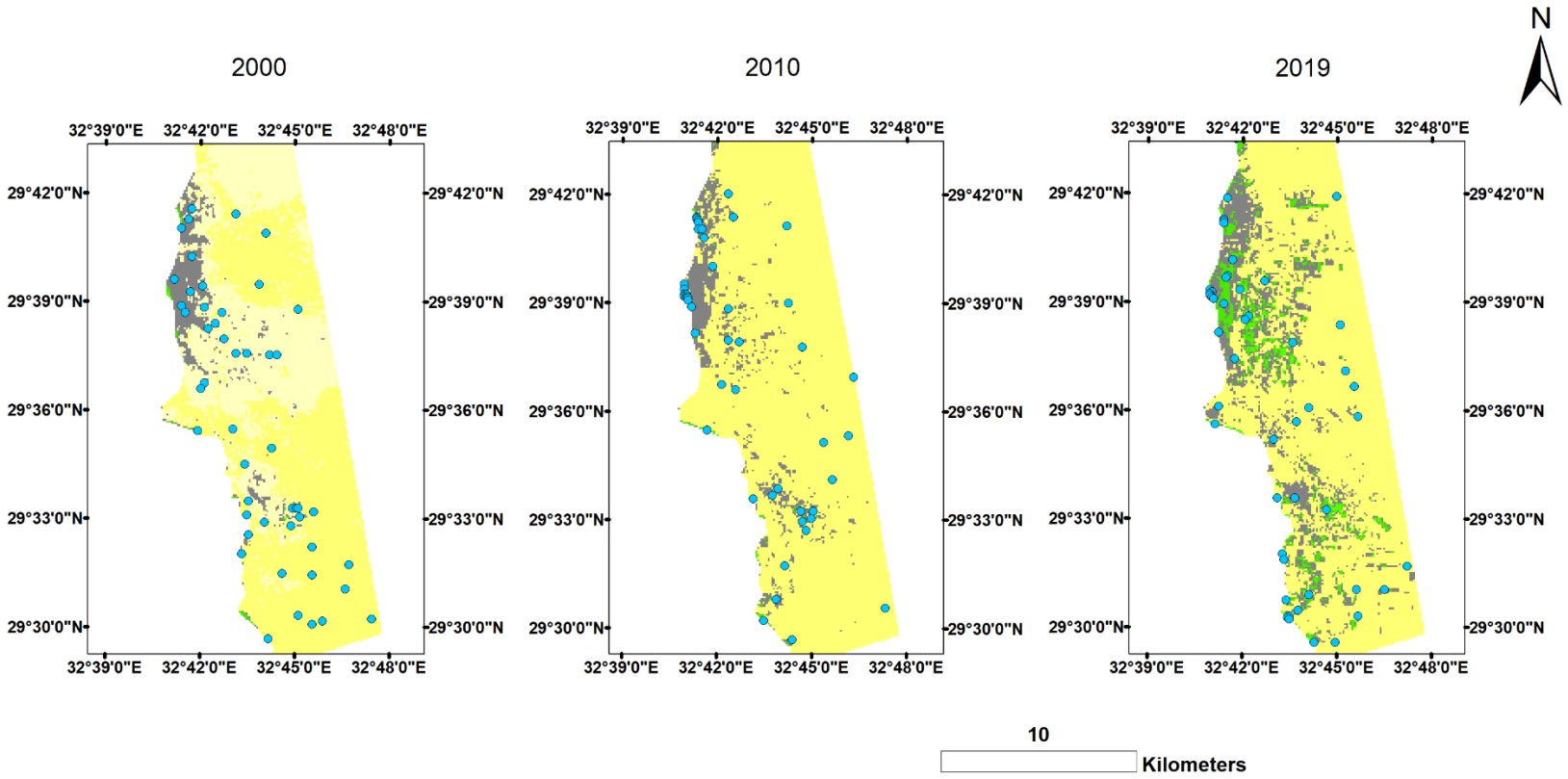


Figure 5.12: Locations of the 50 verification points used for Ras-Sudr.

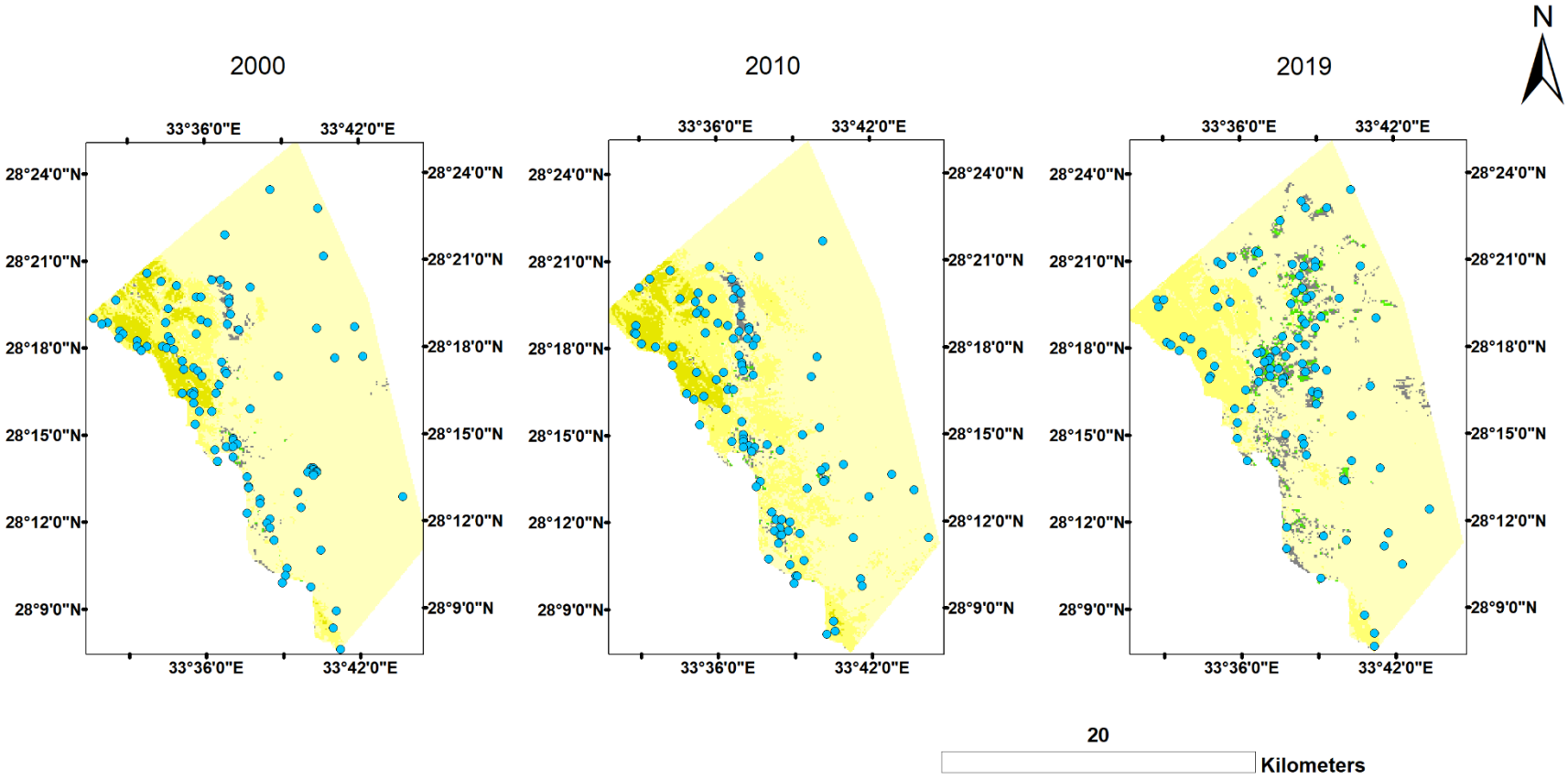


Figure 5.13: Locations of the 100 verification points used for El-Tor.

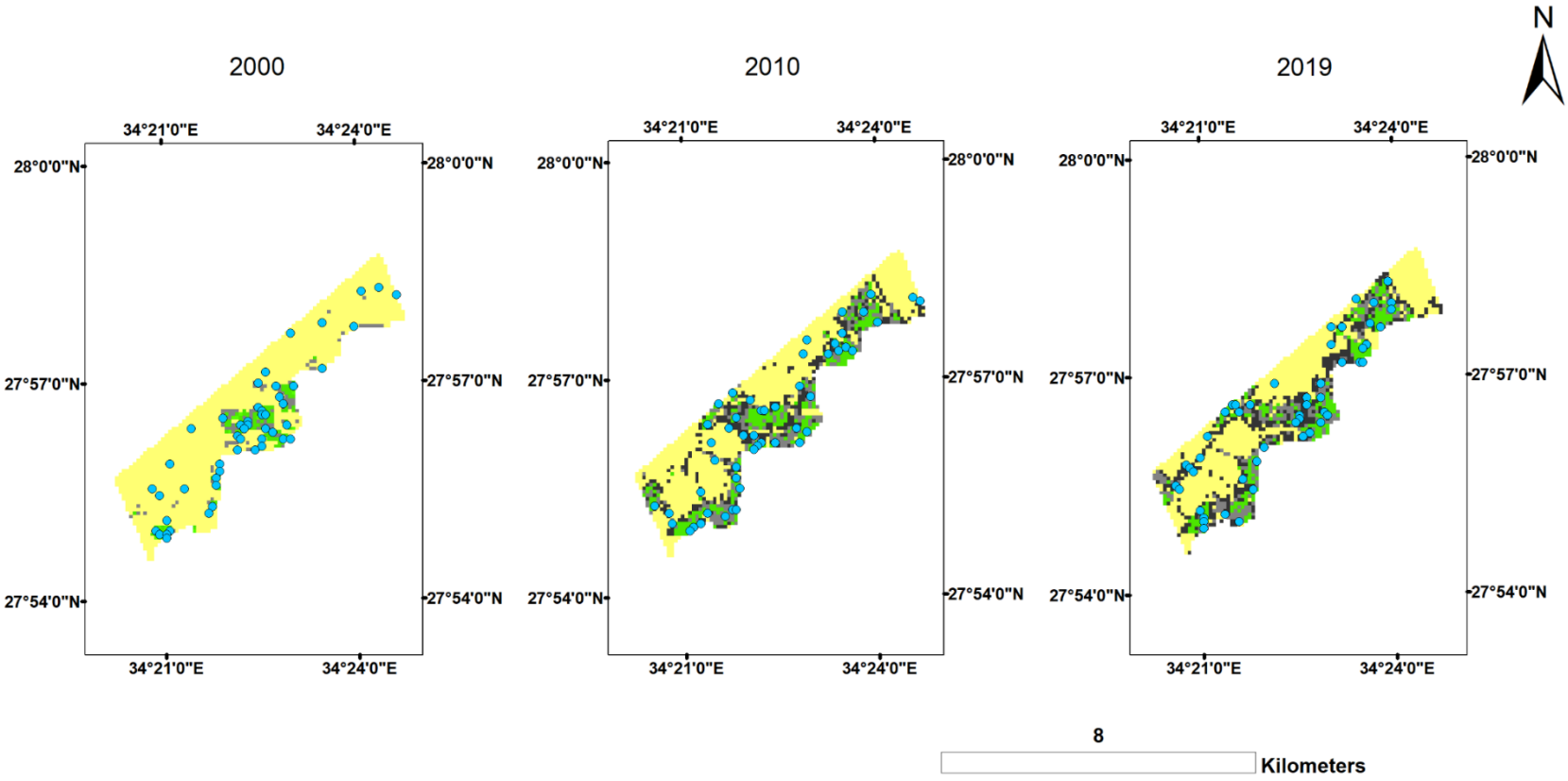


Figure 5.14: The locations of the 50 verification points used for Sharm El-Sheikh.

Table 5.2: Accuracy estimation of each class and the overall classes of Ras-Sudr.

ClassValue (Ras-Sudr2000)	Vegetation(T1)	Vegetation(T2)	Buildings	Bare Soil(T1)	Bare Soil(T2)	Total	U_Accuracy	Kappa
Vegetation(T1)	10	0	0	0	0	10	100%	0
Vegetation(T2)	0	4	6	0	0	10	40%	0
Buildings	1	1	8	0	0	10	80%	0
Bare Soil(T1)	0	1	1	8	0	10	80%	0
Bare Soil(T2)	0	0	0	0	10	10	100%	0
Total	11	6	15	8	10	50	0	0
P_Accuracy	90%	66%	53%	100%	100%	0	80%	0
Kappa	0	0	0	0	0	0	0	75%
ClassValue (Ras-Sudr2010)	Vegetation(T1)	Vegetation(T2)	Buildings	Bare Soil(T1)	Bare Soil(T2)	Total	U_Accuracy	Kappa
Vegetation(T1)	7	0	3	0	0	10	0.7	0
Vegetation(T2)	0	3	7	0	0	10	0.3	0
Buildings	0	0	10	0	0	10	1	0
Bare Soil(T1)	0	0	2	8	0	10	0.8	0
Bare Soil(T2)	0	0	1	0	9	10	0.9	0
Total	7	3	23	8	9	50	0	0
P_Accuracy	100%	100%	43%	100%	100%	0	74%	0
Kappa	0	0	0	0	0	0	0	68%
ClassValue (Ras-Sudr2019)	Vegetation(T1)	Vegetation(T2)	Buildings	Bare Soil(T1)	Bare Soil(T2)	Total	U_Accuracy	Kappa
Vegetation(T1)	14	0	6	0	0	20	0.7	0
Vegetation(T2)	0	14	5	1	0	20	0.7	0
Buildings	0	1	19	0	0	20	0.95	0
Bare Soil(T1)	0	0	3	17	0	20	0.85	0
Bare Soil(T2)	0	0	0	0	20	20	1	0
Total	14	15	33	18	20	100	0	0
P_Accuracy	100%	93%	57%	94%	100%	0	84%	0
Kappa	0	0	0	0	0	0	0	80%

Table 5.3: Accuracy estimation of each class and the overall classes for El-Tor.

ClassValue(El-Tor2000)	Vegetation(T1)	Vegetation(T2)	Buildings	Bare Soil(T1)	Bare Soil(T2)	Total	U_Accuracy	Kappa
Vegetation(T1)	16	2	0	2	0	20	80%	0
Vegetation(T2)	3	16	1	0	0	20	80%	0
Buildings	0	1	19	0	0	20	95%	0
Bare Soil(T1)	0	0	0	20	0	20	100%	0
Bare Soil(T2)	0	0	0	1	19	20	95%	0
Total	19	19	20	23	19	100	0	0
P_Accuracy	84%	84%	95%	86%	100%	0	90%	0
Kappa	0	0	0	0	0	0	0	88%
ClassValue(El-Tor2010)	Vegetation(T1)	Vegetation(T2)	Buildings	Bare Soil(T1)	Bare Soil(T2)	Total	U_Accuracy	Kappa
Vegetation(T1)	17	2	0	0	1	20	85%	0
Vegetation(T2)	5	11	2	2	0	20	55%	0
Buildings	0	1	19	0	0	20	95%	0
Bare Soil(T1)	0	2	0	16	1	19	84%	0
Bare Soil(T2)	0	0	0	0	19	19	100%	0
Total	22	16	21	18	21	98	0	0
P_Accuracy	77%	68%	91%	88%	90%	0	83%	0
Kappa	0	0	0	0	0	0	0	79%
ClassValue(El-Tor2019)	Vegetation(T1)	Vegetation(T2)	Buildings	Bare Soil(T1)	Bare Soil(T2)	Total	U_Accuracy	Kappa
Vegetation(T1)	12	7	1	0	0	20	60%	0
Vegetation(T2)	2	17	1	0	0	20	85%	0
Buildings	3	2	15	0	0	20	75%	0
Bare Soil(T1)	0	0	0	20	0	20	100%	0
Bare Soil(T2)	0	0	1	1	18	20	90%	0
Total	17	26	18	21	18	100	0	0
P_Accuracy	70%	65%	83%	95%	100%	0	82%	0
Kappa	0	0	0	0	0	0	0	77%

Table 5.4: Accuracy estimation of each class and the overall classes for Sharm El-Sheikh.

ClassValue(Sharm2000)	Vegetation(T1)	Vegetation(T2)	Buildings	Streets	Bare Soil	Total	U_Accuracy	Kappa
Vegetation(T1)	9	0	1	0	0	10	90%	0
Vegetation(T2)	0	5	5	0	0	10	50%	0
Buildings	0	0	9	1	0	10	90%	0
Streets	0	0	0	10	0	10	100%	0
Bare Soil	0	0	0	1	9	10	90%	0
Total	9	5	15	12	9	50	0	0
P_Accuracy	100%	100%	60%	83%	100%	0	84%	0
Kappa	0	0	0	0	0	0	0	80%
ClassValue(Sharm2010)	Vegetation(T1)	Vegetation(T2)	Buildings	Streets	Bare Soil	Total	U_Accuracy	Kappa
Vegetation(T1)	10	0	0	0	0	10	100%	0
Vegetation(T2)	0	6	2	0	2	10	60%	0
Buildings	0	0	7	2	1	10	70%	0
Streets	0	0	2	7	1	10	70%	0
Bare Soil	0	0	0	0	10	10	100%	0
Total	10	6	11	9	14	50	0	0
P_Accuracy	100%	100%	63%	77%	71%	0	80%	0
Kappa	0	0	0	0	0	0	0	75%
ClassValue(Sharm2019)	Vegetation(T1)	Vegetation(T2)	Buildings	Streets	Bare Soil	Total	U_Accuracy	Kappa
Vegetation(T1)	6	0	1	3	0	10	60%	0
Vegetation(T2)	0	7	1	2	0	10	70%	0
Buildings	1	1	7	1	0	10	70%	0
Streets	1	1	0	8	0	10	80%	0
Bare Soil	0	0	0	0	10	10	100%	0
Total	8	9	9	14	10	50	0	0
P_Accuracy	75%	77%	77%	57%	100%	0	76%	0
Kappa	0	0	0	0	0	0	0	70%

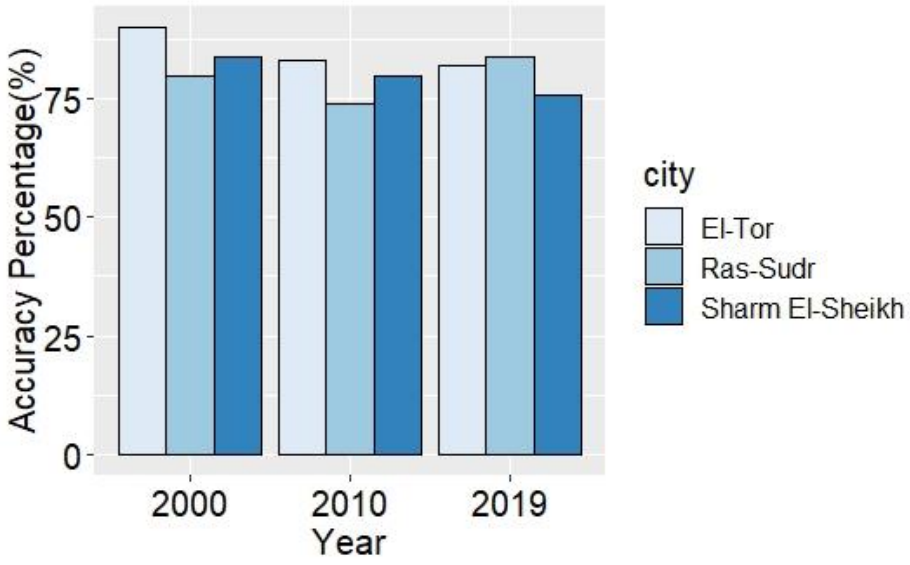


Figure 5.15: The accuracy percentage of the tested locations, based on randomly selected points by ArcGIS 10.5 software.

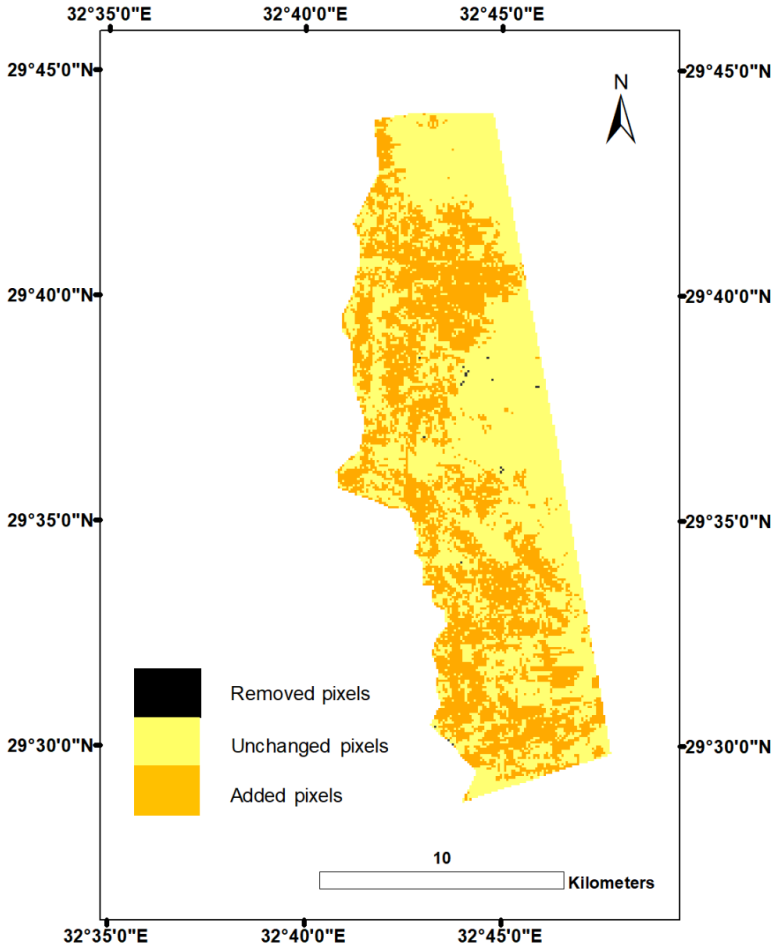


Figure 5.16: Results of the image differencing algorithm applied on Ras-Sudr. In black are pixels that existed in 2000 but were removed by 2019. Orange indicates pixels that did not exist in 2000, yet were added by 2019. Yellow indicates pixels that were present in both 2000 and 2019.

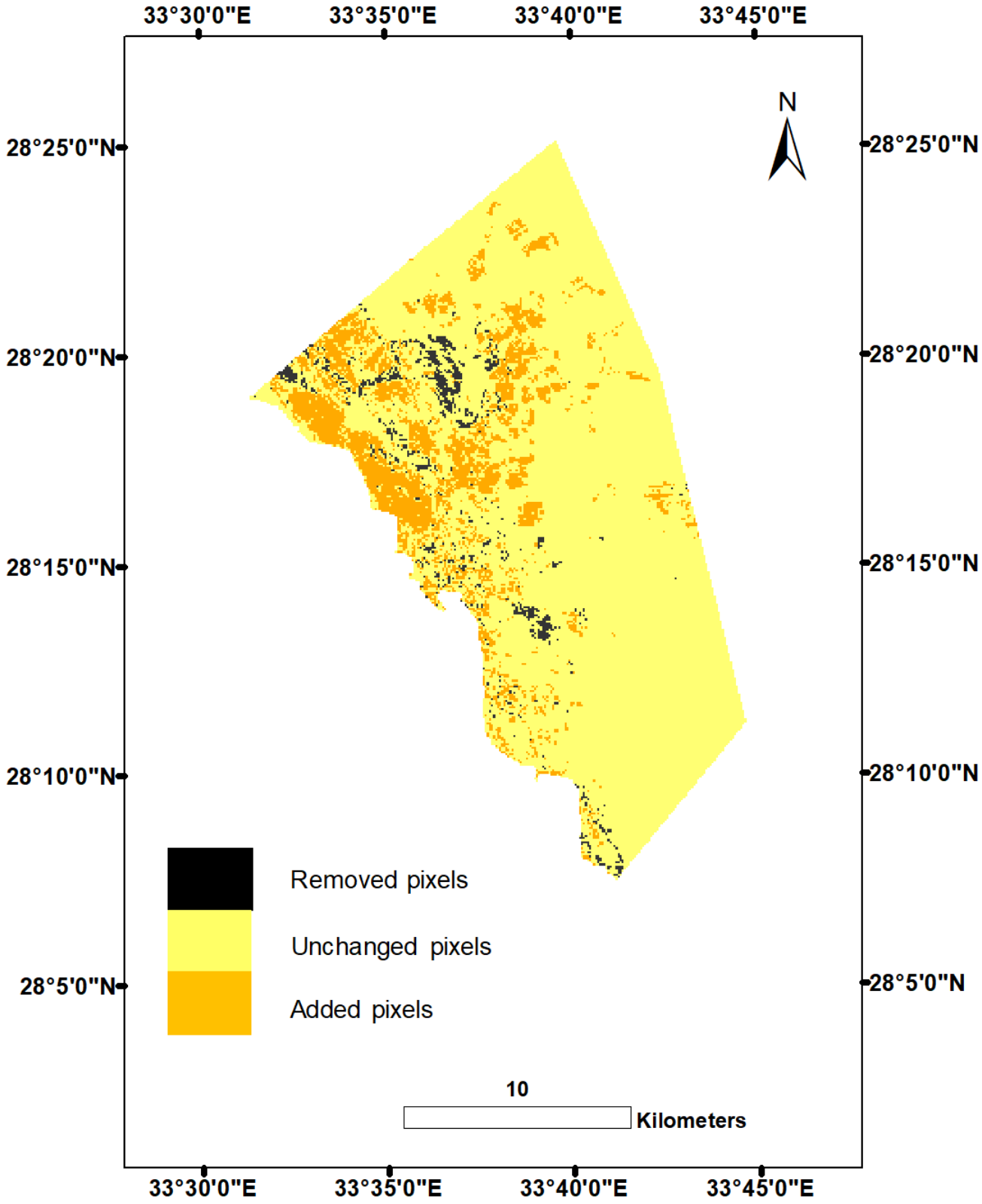


Figure 5.17: Results of the image differencing algorithm applied on El-Tor. In black are pixels that existed in 2000 but were removed by 2019. Orange indicates pixels that did not exist in 2000, yet were added by 2019. Yellow indicates pixels that were present in both 2000 and 2019.

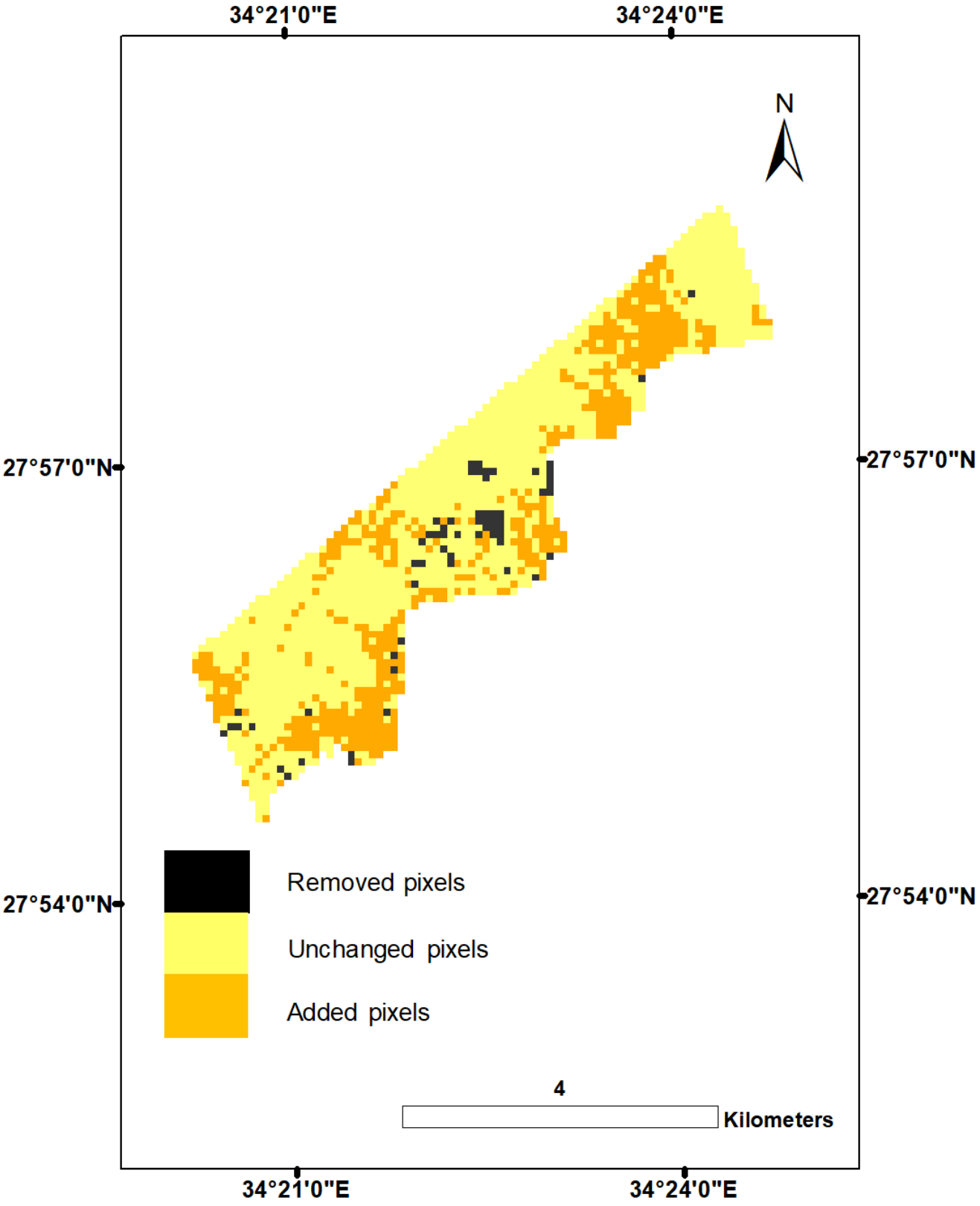


Figure 5.18: Results of the image differencing algorithm applied on Sharm El-Sheikh. In black are pixels that existed in 2000 but were removed by 2019. Orange indicates pixels that did not exist in 2000, yet were added by 2019. Yellow indicates pixels that were present in both 2000 and 2019.

Table 5.5: Added, removed, and unchanged areas between 2000 and 2019 at each of the selected test locations.

	Added Areas (%)	Removed Areas (%)	Unchanged Areas (%)
Ras-Sudr	40.54692627	0.083132794	59.36994093
El-Tor	14.07467096	2.856933558	83.06839548
Sharm El-Sheikh	28.12834225	3.368983957	68.5026738

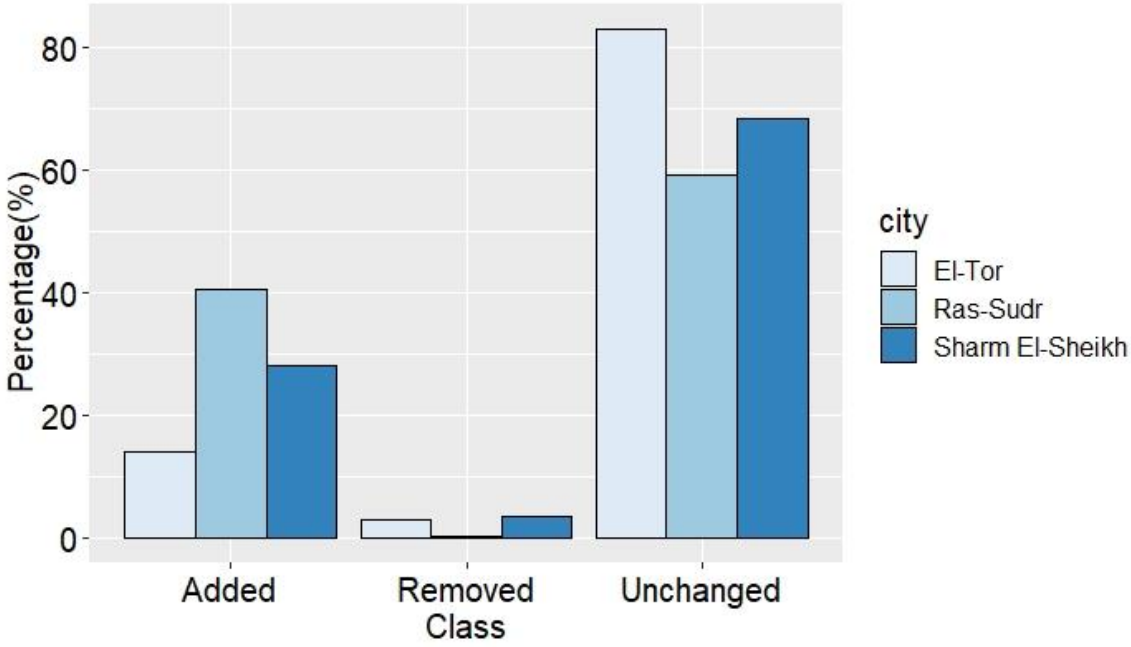


Figure 5.19: Percentages of the added, removed, and unchanged areas between 2000 to 2019 at each of the selected test sites.

Table 5.6: Rough estimation of the increase in groundwater consumption rate at the test site. Determined through use of the change detection algorithm along with rough information collected from local farmers and public websites.

Year	2000	2010	2019
Vegetated areas	83520: 111168m^3	117450:156330m^3	618570:823338m^3
Buildings	21.55: 26.939m^3	29.044:36.3058m^3	100:125m3
Total	83541.55:111194.94m^3	146494:156366.3m^3	618670:823463m^3

5.4 Discussion:

The regional classification of arid areas is often insignificant, as these locations typically feature very sparse urban development or farmlands, which can then appear as bare soil following classification. It is therefore recommended to crop the remotely-sensed scenes of interest to acquire a better view of the local details.

The classification of arid regions featuring widely ranging lithological units can be difficult, as the colors they present often overlap with the colors of building in the area, leading to an underestimation of the number of pixels belonging to the building. Swimming pools and artificial lakes in tourist areas can also overlap with vegetation to produce further misinterpretation.

A very slight increase in vegetation was seen in Ras-Sudr from 2000 to 2010. From 2010 to 2019, however, this type of land cover doubled. The same pattern was seen for building cover, which increased slightly from 2000 to 2010 and doubled during the second time frame. In El-Tor, there was also a slight increase in vegetation from 2000 to 2010, but a very large increase from 2010 to 2019. A similar pattern was seen for building cover. This can potentially be explained by the rise in the tourism industry in these cities during that time span. A considerable increase in vegetation, buildings, and street cover was visible in Sharm El-Sheikh between 2000 and 2010. However, unlike the previous two cities, vegetation and building cover slightly decreased between 2010 and 2019, with street cover slightly increasing.

Water consumption in Ras-Sudr and El-Tor increased slightly from 2000 to 2010. However, this nearly doubled between 2010 and 2019 as the vegetation and building numbers increased. In Sharm El-Sheikh, the water consumption increased significantly from 2000 to 2010, but barely changed between 2010 and 2019. This could be due to the decrease in vegetation and buildings that were a result of a population shift to other cities following a decline in tourism around 2011. However, when interpreting these results, it is important to make clear that the changes over the entire cities were not calculated, due their large areas and the complicating overlap between buildings and lithologies.

Water consumption did not necessarily increase in all cases, but a definite consumption level cannot be provided due to the lack of both public data on consumption rates and continuous water table levels

. In El-Tor, however, a rough estimate of water consumption could be established from information on crops provided by local farmers. Unfortunately, this did not include information

on the exact areas of each crop type, their individual water consumption rates, the total vegetated areas, the yearly increase in population mass, or how many gallons of water were being used. This information would have been necessary for an accurate calculation of the differences in water consumption over time as determined via change detection.

The remote sensing data demonstrated relatively high potential for the detection of changes at the test site over time. These results provided insight into the type of climate studies that should be performed in the future in this area, especially those regarding water cycle equilibrium. Moreover, it succeeded in providing a rough numerical estimate of water consumption over time in one of the cities reliant on the Quaternary aquifer, El-Tor. This was even achieved in the absence of water table level measurements. More accurate values could be established, however, if sufficient information on the test site were to made available, whether via higher-resolution remote sensing or simply water table level measurements, both of which are highly recommended for such future studies.

6. TESTING REMOTE SENSING DATA PERFORMANCE

Precipitation serves as the key parameter of water cycle equilibrium in arid regions, primarily due to its role in recharging ground aquifers and compensating for human consumption (Chapter 2). As stated in Chapter 4, there was a notable increase the area covered by vegetation and buildings at the test site, which in turn affects water consumption rates. However, the data and information captured by both Sherief (2008) and the author (Chapter 4) are insufficient for a full understanding of the spatiotemporal distribution of rainfall at the test site. The number of rain gauges at this site are highly limited, leading to deficiencies in the records captured. Therefore, the present chapter evaluates the strengths and weaknesses of using data derived from remote sensing to fill in the gaps in the rain gauge records. Additionally, this chapter also presents a comparison of data from two types of remote sensors, with the purpose of identifying the source that provides the best-performing data.

6.1 General background

6.1.1 Rain gauges

Rain gauges are the most accurate tools for measuring both precipitation rate at a physical point scale and for measuring rainfall depth as it accumulates over time (Sun *et al*., 2018; Tapiador *et al.*, 2012). Several types of rain gauges exist, including accumulation gauges, tipping bucket gauges, weighing gauges, and optical gauges, each with their own strengths and weaknesses (*Sun et al*., 2018; Tapiador *et al.*, 2012). The most commonly used type is the tipping bucket gauge, which is used to estimate rainfall rate and volume. It has the capability to measure trace amounts of rain, as little as 0.2, 0.5, or 1 mm (Das and Prakash, 2011). The instrument consists of a funnel that receives the rain and sections it into smaller containers. These containers then dump the rainwater after collecting a certain quantity. The dumping procedure is accompanied by an electrical signal that is recorded. In older versions, this signal would be recorded by a pen mounted on an arm attached to a geared wheel (Das and Prakash, 2011). However, tipping bucket gauges do contribute a source of error when measuring heavy rainfall, as the water can accumulate in the containers faster than the dumping process can take place, leading to an underestimation of the heavy rainfall rate. This can occur when the precipitation rate is higher than 300mm/h. This type of gauge can also underestimate a light rainfall rate when water evaporates out of the containers prior to the dumping step (Tapiador *et al*., 2012). A less commonly used type of rain gauge depends on the weighing of the rainfall accumulate at different sampling rates. The saturation effect is therefore not relevant (Tapiador

et al., 2012). One of the challenges faced when attempting the accurate estimation of rainfall rate by rain gauges in arid regions is the wind effect, especially during light rainfall. Wind can transfer these sparse raindrops between locations, disturbing the point scale measuring function of the rain gauges. This can lead to two gauges in close proximity recording different quantities of rainfall (Tapiador *et al*., 2012)

6.1.2 Remote Sensing Data

The Tropical Rainfall Measurements Mission (TRMM) was the first and most widely-used source of remotely-sensed data on rainfall in tropical and subtropical areas. A joint space mission between NASA and the Japan Aerospace Exploration Agency (JAXA) (Immerzeel *et al.*, 2009; Jeniffer *et al*., 2010), it was launched on November 27th, 1997. The TRMM contained several instruments on board the satellite, most prominently a precipitation radar (PR) in the form of a simple one-parameter radar that operated at one transmitting/receiving frequency and a single polarization in order to provide information about rain type, strength, and distribution (Immerzeel *et al.*, 2009). The TRMM Microwave Imager (TMI**)** provided quantitative information about rainfall, water vapor, cloud water content, and sea surface temperature (SST) (Immerzeel *et al.*, 2009). The (PR) complemented the results of the (TMI). The Microwave Imager (TMI) contained passive microwave sensors able to provide measurements of the radiance (the product of the absorption, emission, and scattering) of the precipitating clouds along the sensor view path. The radiance frequency reflected the properties of the clouds and precipitation particles (Kummerow and Barnes, 1998). The satellite also contained active microwave sensors, which provided information on cloud altitudes by measuring backscatter delay (Kummerow and Barnes, 1998). A Visible and Infrared Scanner (VIRS) provided indirect measurements of rainfall intensity, distribution, and type (Immerzeel *et al.*, 2009; Jeniffer *et al*., 2010). While the VIRS might produce less reliable data (Kummerow and Barnes, 1998), it provides more frequent data compared to the infrequent data captured by the (TMI and PR). Lightning Imaging Sensors (LIS**)** detected lighting specifically, which naturally plays an important role in precipitation events. The Clouds and the Earth's Radiant Energy System (CERES) allowed the determination of the radiant energy balance of the Earth. This, together with an understanding of the latent heating derived from precipitation, helped provide a significantly improved picture of the atmosphere's energy system (Kummerow and Barnes, 1998).

A Special Sensor Microwave/Imager (SSM/I) exists onboard the Defense Meteorological Satellite Program (DMSP). It collects data regarding the Earth´s atmosphere through its

microwave instrument (Alemohammad *et al*., 2013). The microwave radiometer is passive and has the capability of measuring radiation emitted at four frequencies, in both ascending and descending overpasses. SSM/I provides valuable information on precipitation rate, water vapor, cloud liquid water, wind speed, and soil moisture (Berg *et al*., 2012). However, no studies have yet assessed its performance. Therefore, it is suggested that it be calibrated to a reference satellite or a stable reference system (Yang *et al*., 2011).

The CloudSat satellite, launched in 2006, carries the first Cloud Profiling Radar (CPR) system, which is able to identify cloud properties and quantify light rain at a 94 GHz frequency. It follows a 16-day return period with equatorial crossing times (local time). This radiometer, previously onboard the 'A-Train' constellation of afternoon satellites, served a complementary mission to the TRMM (Tapiador *et al*., 2012). Level 2 retrieval products used included the Level 2B Cloud Geometries Profile (2B-GEOPROF) and the Level 2B Cloud Water Content Radar-Visible Optical Depth (2B-CWC-RVOD), which features a 250m vertical and 1.1km horizontal resolution (Tompkins and Adebiyi, 2012).

The Global Precipitation Measurement (GPM) mission is the most recent collaboration between NASA and JAXA and is focused on observation of global precipitation (Kim *et al*., 2016). It was launched on February 27, 2014 with its two sensors: the GPM Microwave Imager (GMI), which measures the intensity, type, and size of the precipitation, and the Dual-frequency Precipitation Radar (DPR), which observes the structure of storms within and under clouds (Kim *et al*., 2016). It is widely used in remote sensing of data on precipitation rate (Libertino *et al*., 2016). The GPM uses a higher spatial resolution, wider swath, and higher capturing frequency than the TRMM. It has recently been used to explore the water cycle effect around the globe (Libertino *et al*., 2016). The instruments onboard the GMI measure various frequencies ranging from 10.65 to 183 GHz and with resolutions ranging from 11.2×18.3 Km to 4.4×7.3. In addition, the DPR measures in frequencies of 13.6 and 35.55 GHz, with a spatial resolution equal to 5×5 Km, and with swath area ranging from 245km to 120km (Libertino *et al*., 2016). The TRMM and GPM missions are described in greater detail below.

There also exist merged techniques that are based on a combination of several sources of microwave observation. One example of a merged technique is the Climate Prediction Center's morphing technique (GMORPH), a high-resolution precipitation analysis product with an 8km spatial resolution and a half-hourly temporal resolution, produced since November 22nd, 2002. The GMORPH is a combined effort from the US Air Force Defense Meteorological Satellite Program (DMSP) and the US National Oceanic and Atmospheric Administration (NOAA). The

DMSP provides the Special Sensor Microwave Imager (SSM/I) and the Special Sensor Microwave Imager/Sounder (SSMI/S). NOAA provides the Advanced Microwave Sounding Unit (AMSU), Aqua Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), and the TRMM Microwave Imager (TMI) (Zeweldi and Gebremichael, 2008; Sheffield *et al*., 2018; Tapiador *et al*., 2012). Another example is JAXA's Global Satellite Mapping of Precipitation (GSMaP) satellite, which combines various passive microwave sources (including TMI, SSM/I, and AMSU) to produce hourly products with a 0.1° spatial resolution (Zeweldi and Gebremichael, 2008; Sheffield *et al*., 2018; Tapiador *et al*., 2012). Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) fills in the gaps left by passive microwave observations (TMI, SSM/I, and AMSU) with infrared observations (TIR) to create 0.25° data with a one-hour temporal resolution. PERSIANN can also measure historical precipitation over the past three decades. Other groupings of products, such as CHIRPS and MSWEP, combine available satellite products (CMORPH, GridSat, GSMaP, and TMPA 3B42RT) with rain gauges and model outputs (Sheffield *et al*., 2018), to produce data with a 5km spatial resolution and a once-daily temporal resolution (Sheffield *et al*., 2018) (Table 6.1)

The TRMM (3B42V7) provided some of the most recommended and used radar satellite data (Abera *et al*., 2015). It allowed for high spatiotemporal coverage, despite some uncertainties due to cloud effects as well as limitations in remote sensor performance and retrieval algorithms (Long *et al*., 2016). However, the process of data merging can help eliminate these uncertainties. Data from the (3B42V7) was acquired with a 0.25°×0.25° spatial resolution, a three-hour temporal resolution (starting from the day before at 22:30:00 to the data day at 22:29:59), spatial coverage from 50S to 50N, and a relative bias of 2.37% (Fensterseifer *et al*., 2016). The 3B42 algorithm estimated precipitation by integrating the estimates from the multispectral microwave (MW) sensors with infrared data to fill in any gaps between MW overpasses (Fensterseifer *et al*., 2016). For the study site, eight scenes of TRMM (3B42V7) data (with nc extensions) were downloaded from the official NASA website (mirador.gsfc.nasa.gov). The data was then treated by ArcGIS 10.3 software (Table 6.2).

The Integrated Multi-Satellite Retrievals for GPM (IMERG) algorithm is the Level 3 multi-satellite precipitation algorithm of the GPM, which combines intermittent precipitation estimates from all constellation microwave sensors, IR-based observations from geosynchronous satellites, and monthly rain gauge precipitation data (Sayed *et al*., 2004; Ghodeif and Gorski, 2001). Three different daily IMERG products are offered: IMERG Day 1

Early Run (near real-time, with a latency of 6h), IMERG Day 1 Late Run (reprocessed near real-time with a latency of 18h), and IMERG Day 1 Final Run (gauged-adjusted with a latency of four months) (Guo *et al*., 2017). The IMERG Final Run product provides more accurate precipitation information than the near real-time products across GPCC-gauged regions (Ghodeif and Gorski, 2001) (Table 6.2).

Table 6.1: Summary of currently-available precipitation satellite products (after Sun *et al*, 2018).

Data Products	Res.	Freq.	coverage	period	Data source	References
GPCP	2.5°	Monthly	Global	1979–present	GPI, OPI, SSM/ I scattering, SSM/ I emission, TOVS	(Adler *et al*.,2003)
GPCP1dd	1.0°	Daily	Global	1996–present	SSM/I-TMPI, TOVS	(Huffman and Bolvin,2013)
GPCP_PEN_v2.2	2.5°	5-daily	Global	1979–2014	OPI, SSM/I, GPI, MSU	(Xie *et al*.,2003)
CMAP	2.5°	Monthly	Global	1979–present	GPI, OPI, SSM/I scattering, SSM/I emission, MSU, NCEP–NCAR	(Xie *et al*.,2003)
CPC-Global	0.5°	Daily	Global land	2006–present	GTS, COOP, NMAs	(Xie *et al*.,2010)
TRMM3B43	0.25°	Monthly	50°S–50°N	1998–present	TMI, TRMM Combined Instrument, SSM/I, SSMIS, AMSR-E, AMSU-B, MHS, and GEO IR	(Huffman *et al*.,2007)
TRMM3B42	0.25°	3h/Daily	50°S–50°N	1998–present	TMI, TRMM Combined Instrument, SSM/I, SSMIS, AMSR-E, AMSU-B, MHS, and GEO IR	(Huffman *et al*.,2007)
GSMaP	0.1°	1h/daily	60°S–60°N	2002–2012	TMI, AMSR-E, SSM/I, multi-functional transport satellites (MTSAT), Meteosat-7/8, GOES11/12	(Ushio *et al*.,2009)
PERSIANN-CCS	0.04°	30min/3,6h	60°S–60°N	2003–present	Meteosat, GOES, GMS, SSM/I, polar/near polar precipitation radar, TMI, AMSR	(Sorooshian *et al*.,2000)
PERSIANN-CDR	0.25°	3,6h/Daily	60°S–60°N	1983–present	GOES8, GOES10, GMS-5, Metsat-6, andMetsat-7, TRMM, NOAA15,16, 17, DMSPF13, F14, F15.	(Ashouri *et al*.,2015)
CMORPH	0.25°/8km	30min/3h/Daily	60°S–60°N	2002–present	TMI, SSM/I, AMSR-E, AMSU-B, Meteosat, GOES, MTSAT	(Joyce *et al*.,2004)
GPM	0.1°	30min/3h/daily	60°S–60°N	2015–present	GMI, AMSR-2, SSMIS, Madaras, MHS, Advanced Technology Microwave Sounder	(Hou *et al*.,2014)
MSWEP & CHIRPS	0.1°/0.5°	3h/daily	Global	1979–present	CPC, GPCC, CMORPH, GSMaP-MVK, TMPA, ERA-Interim, JRA-55	(Beck *et al*.,2017)

Table 6.2: Summary of the differences between TRMM(3B42) and GPM(IMERG) data. (after Kim *et al*., 2015).

Product	Temporal resolution	Spatial resolution	Spatial coverage	Time of image	Official start	Main product data source
TRMM (3B42V7)	3h	0.25°	50°N-50°S	Time±1.5h	1998/1/1	Geostationary IR, TMI, TCI, SSMI, AMSR-E, AMSU, SSMI/S, MHS
GPM (IMERG)	0.5h	0.10°	60°N-60°S	Start time	2014/3/12	Geostationary IR, GMI, GCI, TMI, SSMI/S, AMSR2, MHS, GPCC

6.2 Data and methods

6.2.1 Rain gauge data

The Egyptian Meteorological Authority provided the study with a second group of data. This data revealed the rainiest days and the number of rainy days per month for the period of 2014 to 2018, along with the duration (in days) of each rain event. This information was then used to evaluate the performance of the data derived from the remote sensors. The most significant dates data-wise were the 9th of March, 2014, the 25th of October, 2015, the 27th of October, 2016, the 12th of April, 2017, and the 28th of June, 2018. Data from these dates were those used to complete the statistical metrics presented in Section (6.2.4). Although the distribution and number of current rain gauges are insufficient for constructing an adequate impression about the spatiotemporal distribution of rainfall at the study site, they were used in the present study as a benchmark, to gain a general idea about the accuracy of the data derived from the TRMM (3B42V7) and GPM(IMERG). This was done using coherent statistical tests, to determine whether the data tied to the test site could be used without further validation.

6.2.2 TRMM(3B42V7)

For each event, eight scenes from the TRMM (3B42V7) were downloaded from the official NASA website (mirador.gsfc.nasa.gov) and ArcGIS 10.5 software was used to process the data. This was achieved in four steps complementing the first stage of the statistical metrics. The data were opened as a raster layer and clipped to match the study site. The data's pixel size was resampled to match the GPM(IMERG) data. Finally, the value of each pixel was calculated and recorded in an Excel sheet, with values derived from the following time points: the starting point of the events (0h), three hours later (3h), six hours later (6h), nine hours later (9h), twelve hours later (12h), and one day later (24h). Next, data were divided into Plain and Hill groups according to the elevation of the area represented by the pixel. The values of the pixels whose locations coincided with those of the rain gauges were entered into Excel sheets daily at both the 0.25° and 0.1° resolutions.

6.2.3 GPM(IMERG)

Fifty daily scenes of GPM(IMERG) data were downloaded to encompass rainy events from 2015 to 2018. This did not include data from 2014, as the GPM mission had yet to officially start. Therefore, the 2014 event was excluded from all statistical metrics. The official NASA Mirador website was used to download scenes with 'nc' extensions. The data were opened and clipped using ArcGIS10.5 software. The value of each pixel from the (0h), (3h), (6h), (9h), (12h), and (24h) scenes was calculated and stored in an Excel sheet. Next, the values of pixels whose locations coincided with those of the rain gauges were collected in a separate Excel sheet for further statistical metrics.

6.2.4 Statistical Metrics

The first group of statistical tests were performed with the purpose of evaluating the difference, coherence, and correlation between the data derived from the TRMM (3B42V7) and that from the GPM (IMERG), with a 0.1° spatial resolution. These tests included the Shapiro-Wilk test (a test of the normal distribution of samples with sizes between 3 and 5000), the Wilcoxon signed-ranked test (to determine the variance and differences between the two data sets), and the Spearman correlation coefficients test (r), which is ideal for determining the r-value between two non-parametric data sets. All tests, as well as their related results and graphs, were created using the ggplot2 and reshape2 packages in RStudio.

The second group of statistical analyses was selected with the purpose of identifying the remote sensing product with higher compatibility to the physical rain gauges. A Spearman correlation coefficient test was applied between the rain gauge data and the TRMM (0.25°), TRMM (0.1°), and GPM (0.1°), using data collected between 2015 and 2018. This was done to determine the correlational strength between the remote sensing data and the benchmark. Following this, a root mean square error (RMSE) test was performed to evaluate the distribution of the error. A (Bias%) test was used to evaluate the size of the differences between the two data sets, and a mean absolute error (MAE) test revealed the mean magnitude of the errors without considering their direction (Equation 6.1, 6.2, and 6.3) (Kim *et al*., 2017; Chen *et al*., 2018). The Metrics package was used to calculate the results of all of the aforementioned verification tests.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(P_{sat_i} - P_{gau_i}\right)^2} \qquad \text{Eq (6.1)}$$

$$Bias = \frac{1}{n}\sum_{i=1}^{n}\left(P_{sat_i} - P_{gau_i}\right) \qquad \text{Eq (6.2)}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|P_{sat_i} - P_{gau_i}\right| \qquad \text{Eq (6.3)}$$

In the equations above, P_{sat} refers to satellite precipitation records, P_{gau} represents the records derived from the physical rain gauges, and n is defined as the number of samples.

A third group of categorical statistics was used to verify the accuracy and efficiency of the selected data on the remote sensing of rainfall detection. This group consisted of: the probability of detection (POD), the false alarm ratio (FAR), and the critical success index (CSI). These were calculated for each single event to examine the potential of the satellite products at various rainfall thresholds (Chen *et al*., 2018; Kim *et al*., 2015). The POD implied the fraction of the correctly detected precipitation events (Ebert *et al*., 2006), the FAR provided the fraction of false alarms (Kim *et al*., 2015), and the CSI calculated the correct number of detected events divided by total number of false alarms, hits, and misses (Chen *et al*., 2018). Chen *et al*. (2018) defined hits as rain detected by both gauges and satellites and misses as rain observed by gauges but not detected by satellite. False alarms were described as rain detected by satellites but not observed by physical gauges (Table 6.3). In the present study the author analyzed the ability of the data from both the TRMM(3B42V7) and GPM(IMERG) to correctly detect rainfall rates at three different threshold values (0.1, 1, and 10). Those values represented the rainfall intensity limits previously established in Chapter 4. The following equations (6.4, 6.5, and 6.6) were used to fulfill the purpose of the third statistical group:

$$POD = \frac{Hits}{Hits + Misses} \qquad \text{Eq (6.4)}$$

$$FAR = \frac{false\ alarms}{Hits + false\ alarms} \qquad \text{Eq (6.5)}$$

$$CSI = \frac{Hits}{Hits + false\ alarm + Misses} \qquad \text{Eq (6.6)}$$

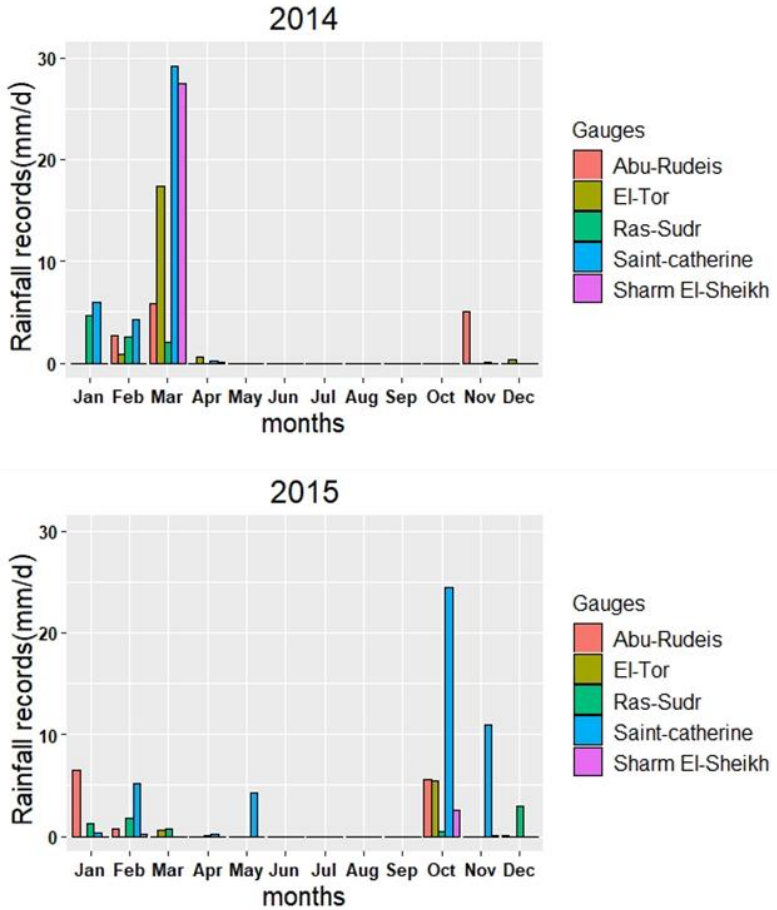
Table 6.3: showing the compatibility between rain gauges and satellite precipitation products for each precipitation threshold (after Chen *et al*., 2018; Kim *et al*., 2015).

	Gauge ≥ threshold	Gauge < threshold
Satellite ≥ threshold	Hits	False alarm
Satellite < threshold	Misses	Correct negatives

6.3 Results and discussion

6.3.1 Rain gauges

The first precipitation event (March 9th, 2014) ranked as a heavy intensity event, as three rain gauges recorded more than 10mm/d, and two of them recorded 1-10mm/d. The second event (October 25th, 2015) ranked as a moderate intensity event, as three rain gauges recorded 1-10mm/d, one rain gauge recorded >10mm/d, and one gauge recorded 0.1-1mm/d. The third event (October 27th, 2016) event ranked as a heavy to moderate intensity event, as two rain gauges recorded >10mm, while three gauges recorded 1-10mm/d. The fourth and fifth events (April 12th, 2017 and April 28th, 2018) ranked as light intensity events, as the majority of gauges recorded 0.1-1mm/d (Table 6.4, Figure 6.1).



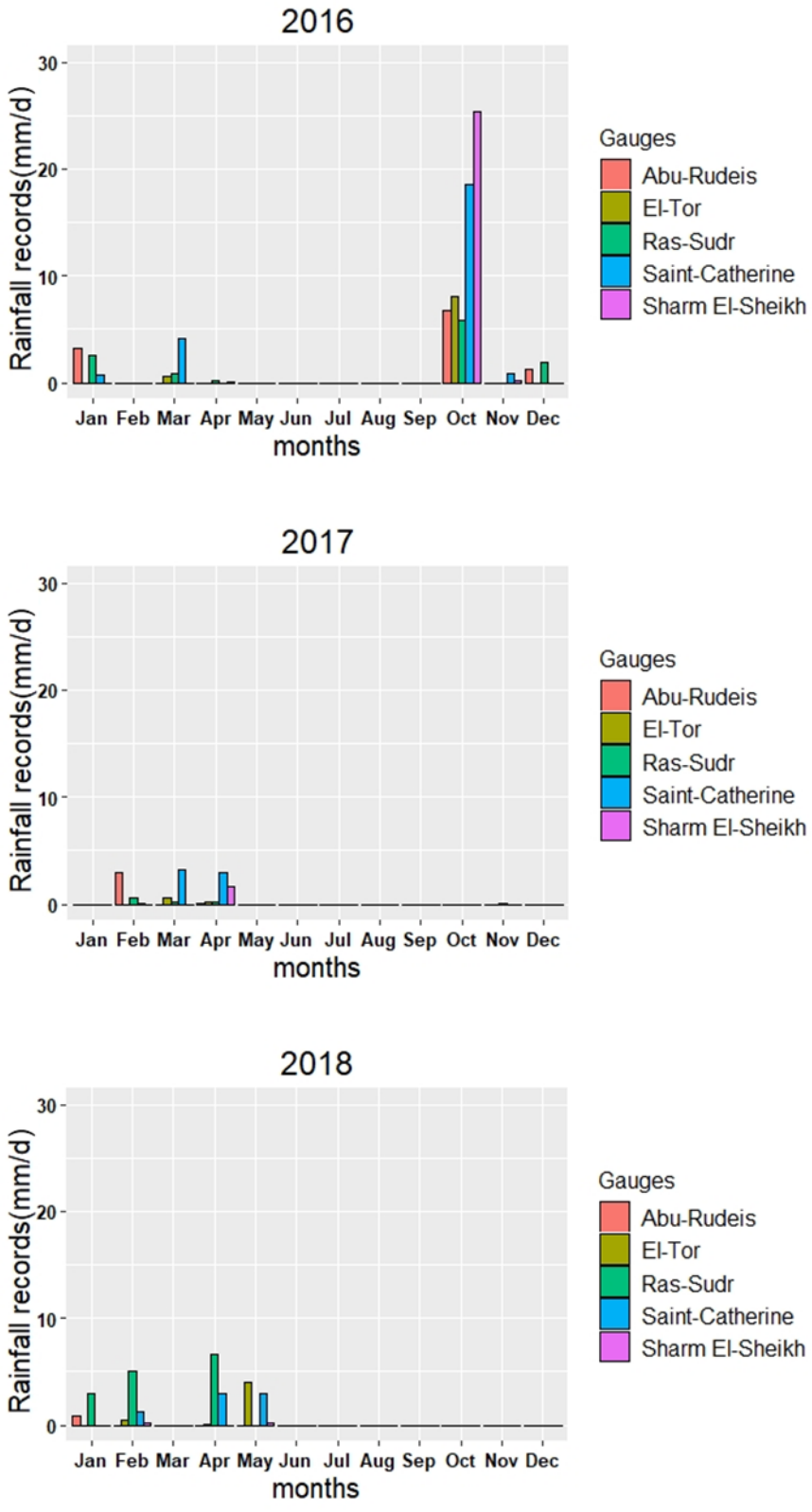


Figure 6.1: Monthly rain gauge records from the period of 2014 to 2018.

Table 6.4: Intensities of rainfall events recorded in the El-Qaa Plain during the period of 2014 to 2018.

El-Qaa Plain Regions	2014 Event Intensity	2015 Event Intensity	2016 Event Intensity	2017 Events Intensity	2018 Events Intensity
Southern-Region	Heavy	Moderate	Heavy	Light	Light
Middle-Region	Heavy	Heavy	Heavy	Moderate	Moderate
Northern-Region	Moderate	Moderate	Moderate	Light	Light

6.3.2 TRMM(3B42V7) and GPM(IMERG)

Precipitation maps were created for the three data sources: the TRMM(3B42V7) data with 3-hour temporal resolution and 0.25° spatial resolution, the TRMM(3B42V7) data with 3-hour temporal resolution and 0.1° spatial resolution, and the GPM(IMERG) data with half-hour temporal resolution and 0.1° spatial resolution. Individual maps were created for all precipitation events mentioned between 2015 to 2018 (Figure 6.2). The distribution maps illustrated the differences between the three resolution-based data sets. TRMM (0.25° and 0.1°) revealed very similar results. However, noticeable changes were seen between the TRMM data sets and that of the GPM(IMERG), especially in the 2016 event, which was the event exhibiting the highest intensity (Figure 6.2).

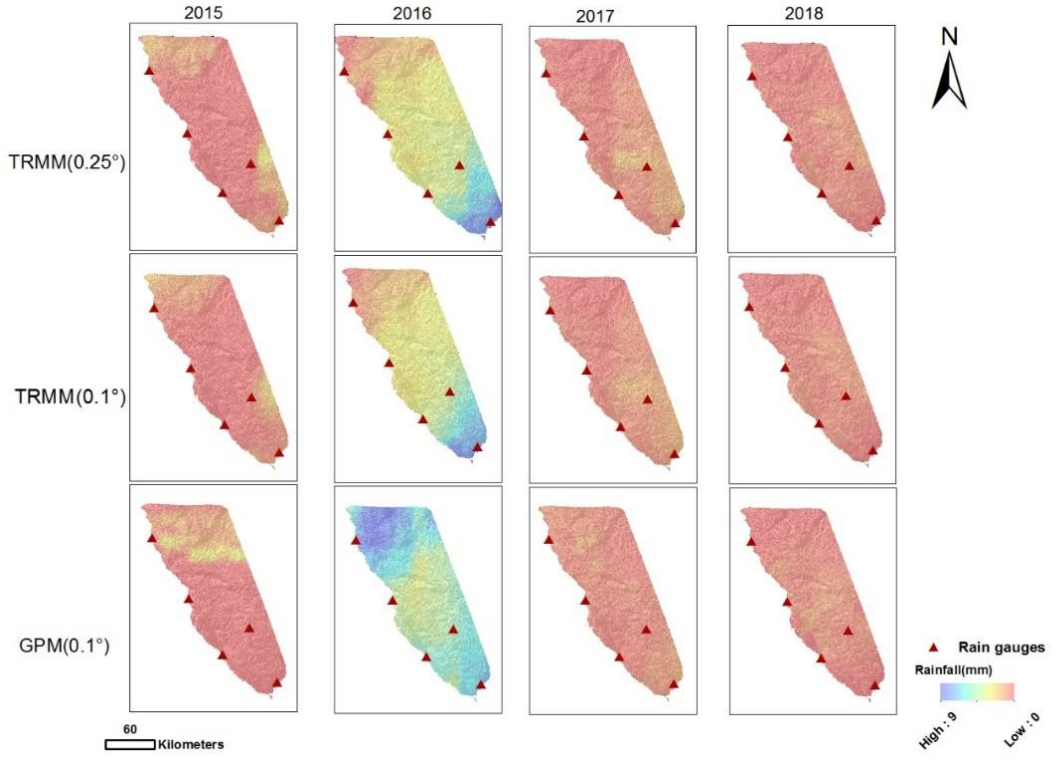


Figure 6.2: Spatial distribution overlay of rainfall over an area hill shade map, recorded during the period of 2015 to 2018, based on TRMM(3B42V7) and GPM(IMERG) accumulation scenes (mm/d).

6.3.3 Statistical Metrics

The results of the first application of statistical metrics are visible in Table 6.5. The Shapiro-Wilk normality test between the pixel values of the TRMM(3B42V7) and GPM(IMERG) (both at 0.1° resolution) showed these non-parametric data sets to have p-values mostly below 0.05. Boxplots provided below illustrate the differences between pixel values from the TRMM(3B42V7) and GPM(IMERG) datasets (again both at 0.1° spatial resolution) at various successive times (0h, 3h, 6h, 9h, 12h, 24h) (Figure 6.3). The highest difference could be found in the October 27th, 2016 event. Moreover, all of the mentioned boxplots displayed a gradual increase in disparity as the precipitation duration increased (9h, 12h, and 24h). The lowest difference appeared between the Plain and Hill precipitable portions of the June 28th, 2018 event, and between the onset (0h) and third hour (3h) of all mentioned events. Spearman's rank correlation coefficient revealed a weak correlation between the two remote sensing data sets for the years 2015, 2016, and 2017. However, a strong correlation was found between the two data sets in 2018 (Figure 6.4 and Table 6.5). The most highly-correlated values derived from the first few hours of precipitation (0h, 3h, and 6h). Unfortunately, in most cases, correlation between the two datasets at later hours (9h, 12h, and 24h) was weak, with nonsignificant r-values and negative correlations. These unexpected results from the first group of statistical analyses are, however, not sufficient for providing a full understanding of the performance of the two datasets.

The second group of statistical metrics produced Spearman correlation coefficients as follows: for the relationship between the physical rain gauge records and the TRMM with a 0.25° resolution, R= 0.328 and p = 0.157; for the relationship between the physical rain gauge records and the TRMM with a 0.1° resolution, R= 0.546 and p = 0.012; for the relationship between the physical rain gauge records and the GPM(IMERG) with a 0.1° resolution, R= 0.745 and p = 0.00016 (Figure 6.5). The GPM(IMERG) showed the strongest correlation with the rain gauges and the TRMM (0.25°) no correlation.

These tests were consequently followed by the previously-mentioned metrics (RMSE, BIAS, and MAE). These metrics were calculated for each event and plotted as a boxplot graph featuring the maximum and minimum limits, the 25th percentile, the 75th percentile, and the median of each test (Figure 6.5). The GPM(IMERG) dataset displayed the lowest RMSE values for the 2015, 2016, and 2018 precipitation events (10.677, 10.562, and 1.883, respectively). Furthermore, it recorded the values closest to 0 in the BIAS (%) test for the 2015 and 2016 events, and the lowest MAE values for the 2015, 2016, and 2018 events (6.726, 8.076, and

1.367, respectively). The values from the TRMM 0.1° dataset were close those of the TRMM 0.25° dataset, but with better performance. All values are listed in Table 6.6. The metrics boxplots reveal the best performing remote sensor to be the GPM(IMERG), as it recorded the lowest RMSE and MAE median and maximum values, when compared to the other data types (Figure 6.6). Moreover, its median was very close to 0 in the BIAS (%) test, in comparison to the TRMM 0.25° and TRMM 0.1° data.

The third group of categorical statistics was applied with three different thresholds (0.1, 1.0, and 10). Results confirmed the high capabilities of the TRMM(3B42V7) and GPM(IMERG) remote sensors for the detection of low-intensity events, as the 0.1 threshold was the best-performing threshold for both remote sensing datasets, both achieving 1 in the POD and CSI tests (Figure 6.7). The FAR test revealed results of 0.4 and 0.2, respectively. The second threshold, 1.0, produced a 1 in the POD test for both data sets, with the CSI values were calculated to be 0.8 and 1, and the FAR results to be in 0.4 and 0.5, respectively. The third threshold, 10mm, offered the worst results. The TRMM(3B42V7) achieved a 0 on all tests mentioned. The GPM(IMERG) received vales of 1, 1, and 0.3 for the POD, FAR, and CSI tests, respectively. In general, the GPM(IMERG) provided more robust results than the TRMM(3B42V7), and higher certainty was seen for light-intensity events in both datasets. TRMM 0.1° data are close to values of the TRMM 0.25° data, but with better performance.

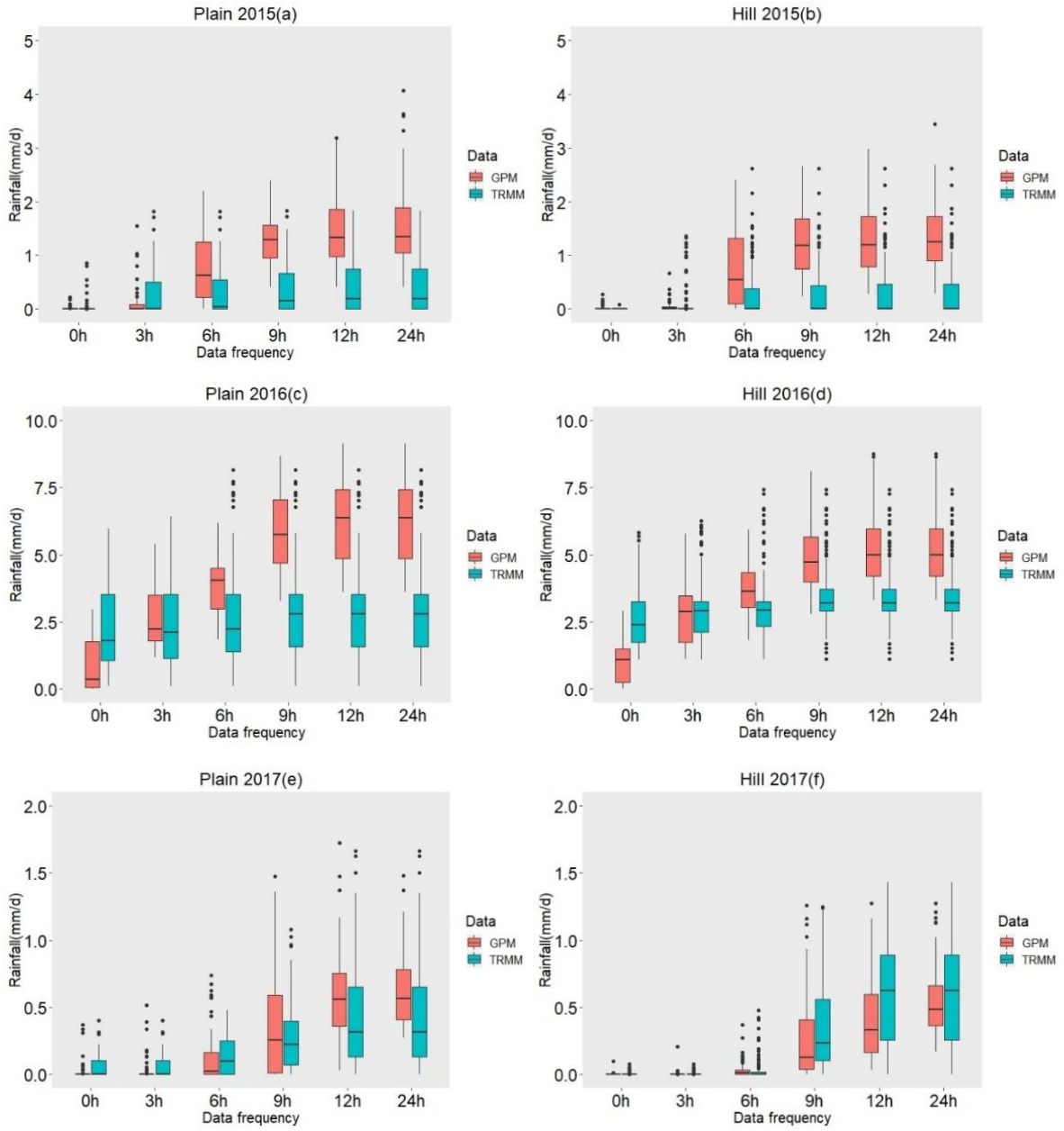
Table 6.5: Results of the Shapiro-Wilk test, the Wilcoxon signed-rank test, and the correlation coefficients (R and p values) between the TRMM(3B42V7) and the GPM(IMERG) over the Hill and Plain regions during successive timepoints (0h, 3h, 6h, 9h, 12h, 24h).

	Shapiro.test	*Wilcox.test*	*R*	*cc*
2015plain(0h)	*p < 2.2e-16*	*p = 0.1873*	*-0.1638265*	*p = 0.1922*
2015plain(3h)	*p < 2.2e-16*	*p = 0.5814*	*0.6058509*	*p = 8.919e-08*
2015plain(6h)	*p = 6.16e-10*	*p = 3.325e-06*	*0.3865752*	*p = 0.001469*
2015plain(9h)	*p = 5.935e-06*	*p = 3.189e-15*	*0.2821433*	*p = 0.02278*
2015plain(12h)	*p = 3.341e-10*	*p = 1.62e-15*	*0.4333235*	*p = 0.0003114*
2015plain(24h)	*p = 4.574e-12*	*p = 2.894e-16*	*0.456812*	*0.0001308*
2015hill(0h)	*p < 2.2e-16*	*p = 0.0001557*	*-0.03583405*	*p = 0.6976*
2015hill(3h)	*p < 2.2e-16*	*p = 0.0002039*	*-0.02549188*	*p = 0.7823*
2015hill(6h)	*p < 2.2e-16*	*p = 9.49e-14*	*-0.3282374*	*p = 0.0002522*
2015hill(9h)	*p = 1.296e-11*	*p < 2.2e-16*	*-0.5227105*	*p = 9.125e-10*
2015hill(12h)	*p = 2.545e-11*	*p < 2.2e-16*	*-0.4434222*	*p = 3.934e-07*
2015hill(24h)	*p = 3.924e-11*	*p < 2.2e-16*	*-0.2816123*	*p = 0.001834*
2016plain(0h)	*p = 1.149e-10*	*p = 1.722e-07*	*0.6792872*	*p = 3.609e-10*

2016plain(3h)	*p = 5.315e-07*	*p = 0.063*	*0.4389378*	*p = 0.0002271*
2016plain(6h)	*p = 0.0002912*	*p = 7.602e-06*	*0.03273418*	*p = 0.7942*
2016plain(9h)	*p = 0.00157*	*p = 1.763e-12*	*-0.5122959*	*p = 1.097e-05*
2016plain(12h)	*p = 0.0006267*	*p = 1.641e-13*	*-0.5211682*	*p = 7.236e-06*
2016plain(24h)	*p = 0.0006267*	*p = 1.641e-13*	*-0.5211682*	*p = 7.236e-06*
2016hill(0h)	*p = 2.042e-07*	*p < 2.2e-16*	*0.8662037*	*p < 2.2e-16*
2016hill(3h)	*p = 1.051e-08*	*p = 0.4478*	*0.9131037*	*p < 2.2e-16*
2016hill(6h)	*p = 1.101e-06*	*p = 1.541e-07*	*0.4897346*	*p = 3.234e-08*
2016hill(9h)	*p = 3.023e-06*	*p < 2.2e-16*	*-0.1419997*	*p = 0.1266*
2016hill(12h)	*p = 1.414e-05*	*p < 2.2e-16*	*-0.2082603*	*p = 0.02441*
2016hill(24h)	*p = 1.414e-05*	*p < 2.2e-16*	*-0.2082603*	*p = 0.02441*
2017plain(0h)	*p < 2.2e-16*	*p = 0.2178*	*0.5628374*	*p = 1.06e-06*
2017plain(3h)	*p < 2.2e-16*	*p = 0.02497*	*0.3758659*	*p = 0.002032*
2017plain(6h)	*p = 9.843e-13*	*p = 0.7156*	*0.5213674*	*p = 8.462e-06*
2017plain(9h)	*p = 1.133e-08*	*p = 0.9647*	*-0.2703409*	*p = 0.02941*

2017plain(12h)	*p = 4.464e-07*	*p = 0.0004236*	*0.143249*	*p = 0.255*
2017plain(24h)	*p = 4.64e-08*	*p = 2.039e-06*	*0.2284996*	*p = 0.06713*
2017hill(0h)	*p < 2.2e-16*	*p = 0.0012*	*0.1485082*	*p = 0.107*
2017hill(3h)	*p < 2.2e-16*	*p = 0.1134*	*0.005078644*	*p = 0.9563*
2017hill(6h)	*p < 2.2e-16*	*p = 8.091e-06*	*-0.02085452*	*p = 0.8219*
2017hill(9h)	*p = 3.283e-15*	*p = 0.0001393*	*-0.5469143*	*p = 1.234e-10*
2017hill(12h)	*p = 1.722e-08*	*p = 0.0002391*	*-0.4631567*	*p = 1.133e-07*
2017hill(24h)	*p = 1.638e-05*	*p = 0.261*	*-0.09604863*	*p = 0.2988*
2018plain(0h)	*p = 1.585e-14*	*p = 0.06119*	*0.4234261*	*p = 0.008459*
2018plain(3h)	*p = 2.306e-05*	*p = 0.05561*	*0.7068321*	*p = 0.000246*
2018plain(6h)	*p = 0.0002393*	*p = 0.004568*	*0.703494*	*p = 5.82e-05*
2018plain(9h)	*p = 0.006646*	*p = 0.1368*	*0.6359128*	*p = 0.0007103*
2018plain(12h)	*p = 0.006646*	*p = 0.1368*	*0.6359128*	*p = 0.0007103*
2018plain(24h)	*p = 0.006646*	*p = 0.1368*	*0.6359128*	*p = 0.0007103*
2018hill(0h)	*p = 1.585e-14*	*p < 2.2e-16*	*0.4234261*	*p = 1.776e-06*

2018hill(3h)	*p = 2.306e-05*	*p = 0.7851*	*0.7068321*	*p < 2.2e-16*
2018hill(6h)	*p = 0.0002393*	*p-value = 0.3289*	*0.703494*	*p < 2.2e-16*
2018hill(9h)	*p = 0.006646*	*p = 0.03293*	*0.6359128*	*p < 2.2e-16*
2018hill(12h)	*p = 0.006646*	*p = 0.03293*	*0.6359128*	*p < 2.2e-16*
2018hill(24h)	*p = 0.006646*	*p = 0.03293*	*0.6359128*	*p < 2.2e-16*



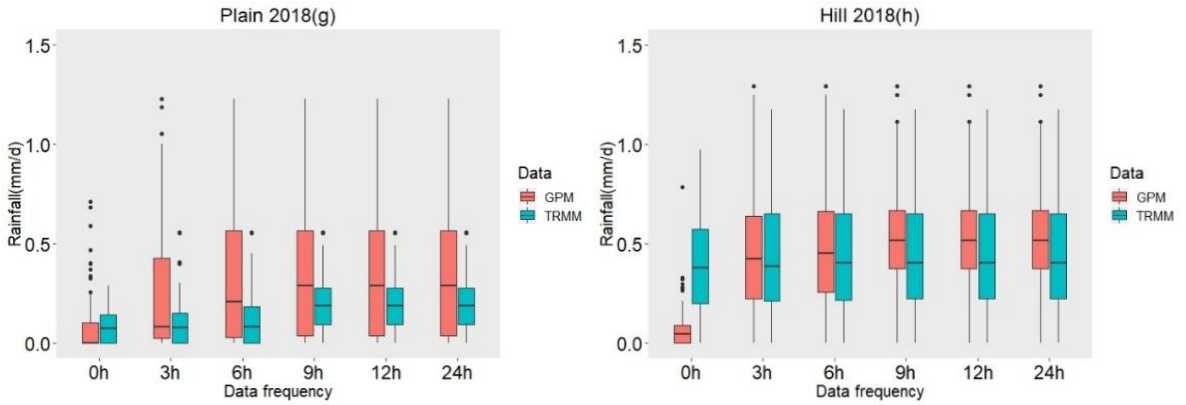


Figure 6.3: Boxplots illustrating the differences between the TRMM(3B42V7) and the GPM(IMERG) at a 0.1° spatial resolution, over Hill and Plain regions at successive timepoints (0h, 3h, 6h, 9h, 12h, 24h).

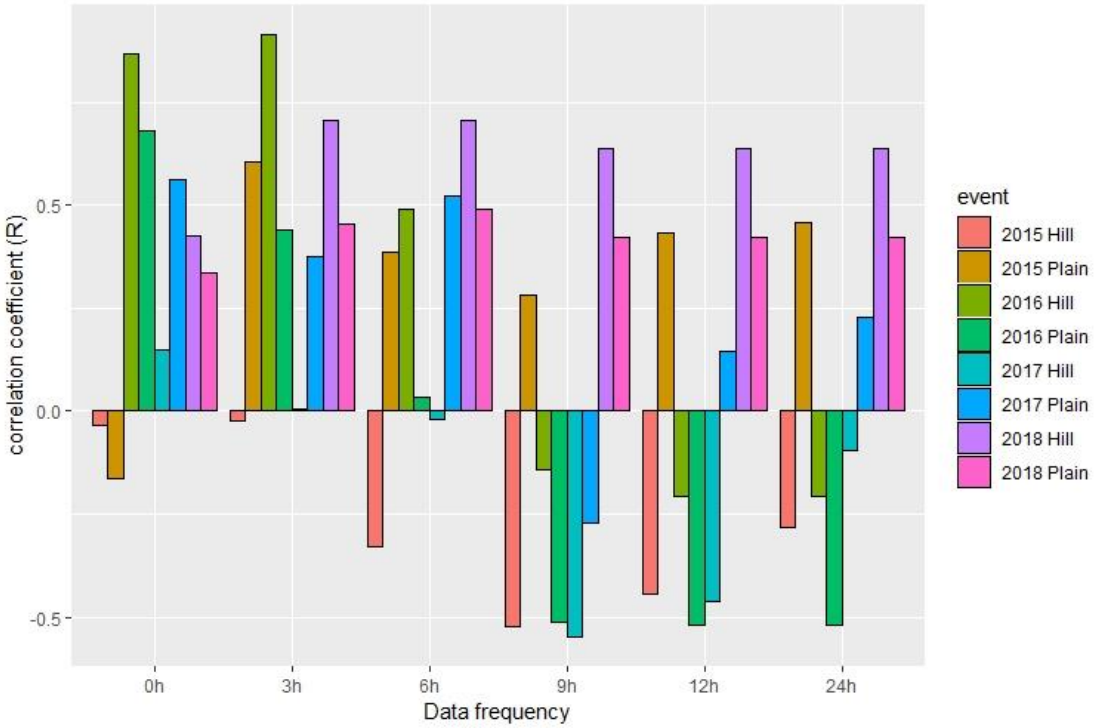
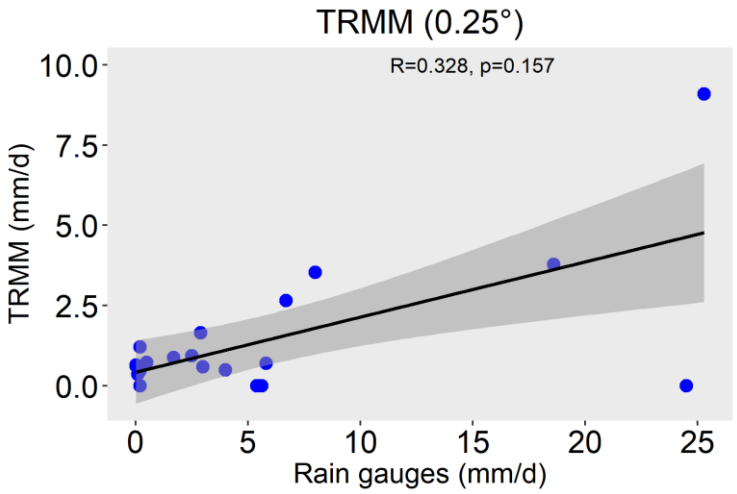


Figure 6.4: Bar plot of the correlation coefficients of the two sets of remote sensing data, TRMM(3B42V7) and GPM(IMERG), over the Hill and Plain regions between onset to 24h.



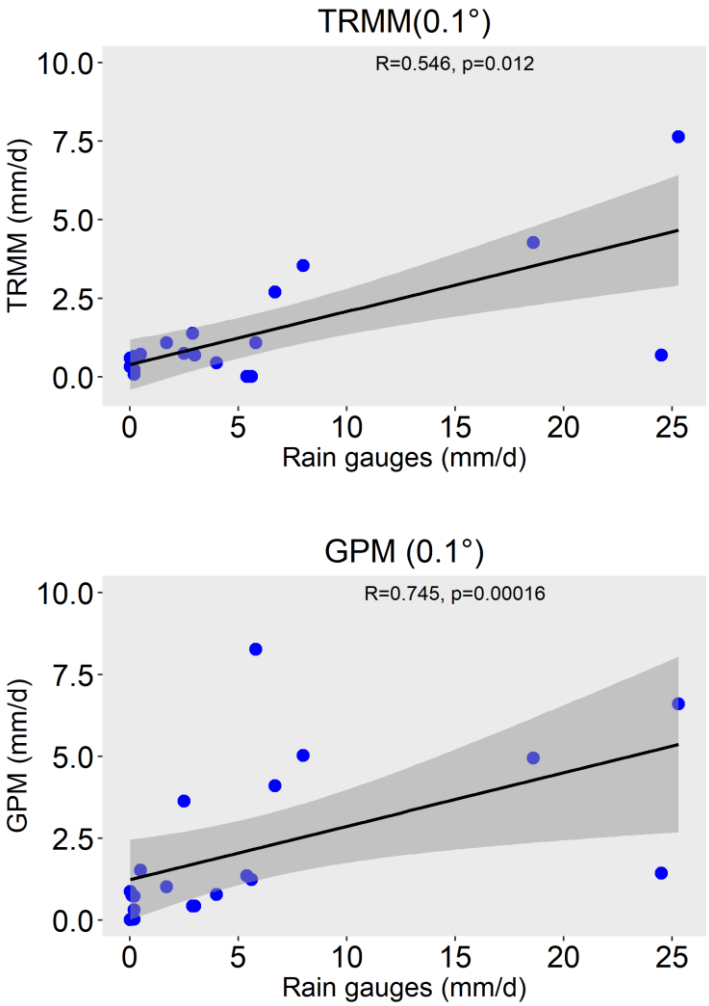


Figure 6.5: Correlation between remote sensing data (at spatial resolutions of both 0.25° and 0.1°) with rain gauge records.

Table 6.6: RMSE, BIAS%, and MAE values for each recorded event, with spatial resolutions specified.

Metrics	RMSE	BIAS	MAE
2015(TRMM 0.25°)	11.51743	0.6328	7.458
2015(TRMM 0.1°)	11.23075	0.6465437	7.3568
2015(GPM 0.1°)	10.67772	-0.003797354	6.7262
2016(TRMM 0.25°)	10.43829	0.6960047	8.93
2016(TRMM 0.1°)	10.72723	0.6873562	9.034
2016(GPM 0.1°)	10.56291	0.3615109	8.0762
2017(TRMM 0.25°)	0.8247361	-1.623736	0.7228
2017(TRMM 0.1°)	0.7661856	-0.8150872	0.5724
2017(GPM 0.1°)	1.207487	-1.718658	0.89222
2018(TRMM 0.25°)	1.94539	0.9652	1.4788
2018(TRMM 0.1°)	1.919342	1.012	1.378
2018(GPM 0.1°)	1.883479	1.01694	1.36774

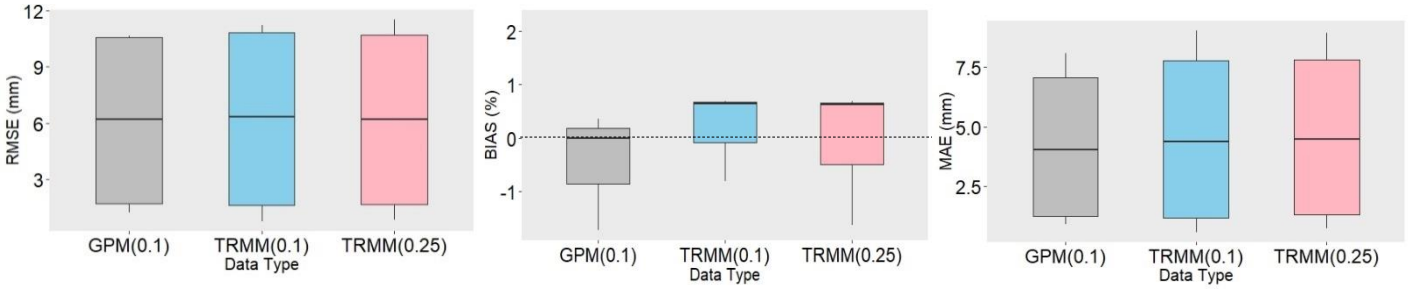


Figure 6.6: Boxplots of RMSE, BIAS%, and MAE values recorded by each single event.

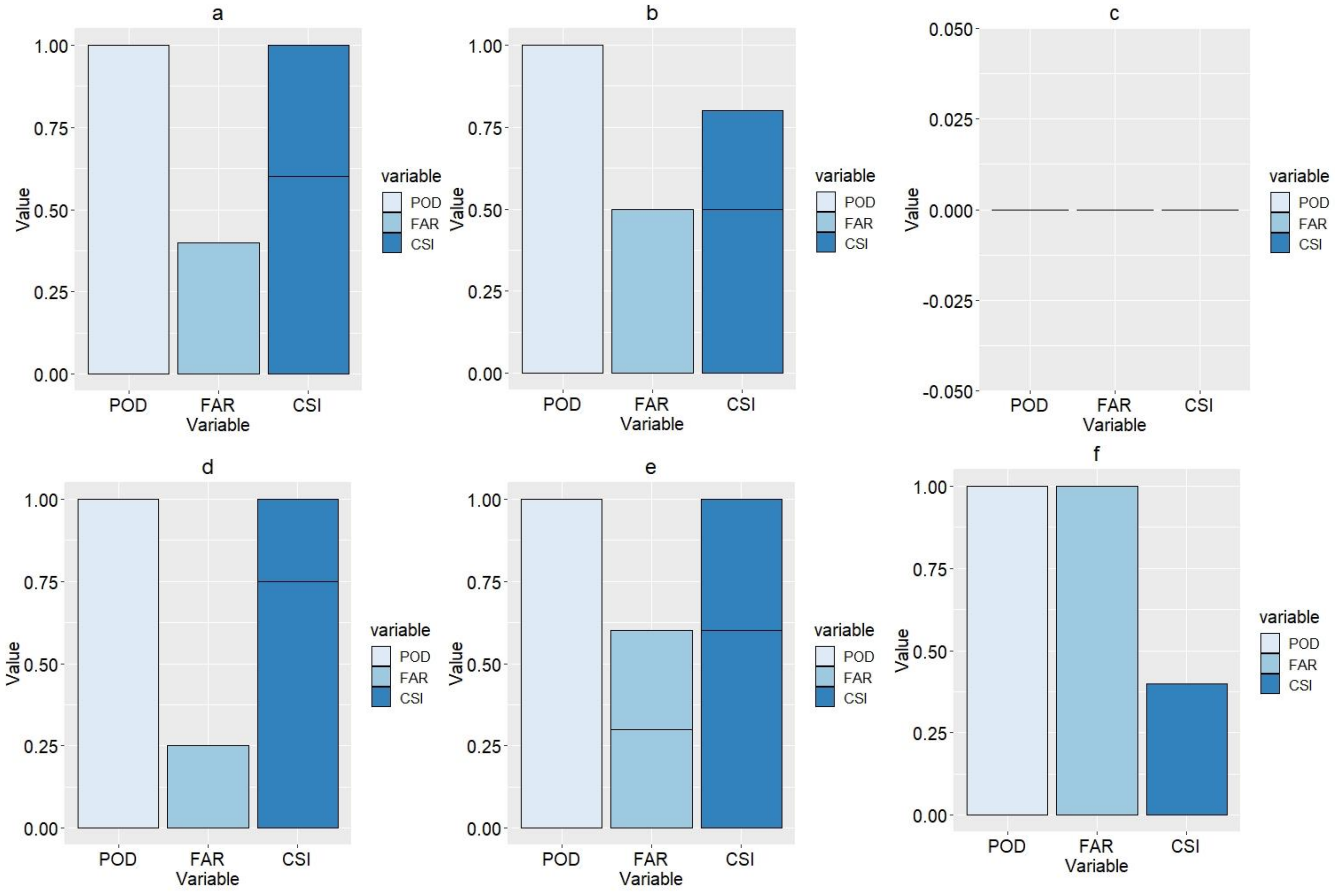


Figure 6.7: Bar plots of POD, FAR, and CSI results of the TRMM(3B42V7) and GPM(IMERG) at three different thresholds (0.1, 1, and 10) using data from all mentioned events. A, B, and C represent the thresholds of 0.1, 1, and 10 for the TRMM(3B42V7) data, while D, E, and F represent the thresholds of 0.1, 1, and for the GPM(IMERG) data.

6.4 Discussion

The statistical metrics used revealed high correlation and low differences between the values of the pixels derived from the TRMM(3B42V7) and GPM(IMERG) at the precipitation event onsets, as well as 3h and 6h after, in some cases. However, as the event durations increased and precipitation accumulated, the correlation between both datasets decreased, and differences increased. Consequently, values from each data type were compared with records

from the limited rain gauges, using several statistical metrics to evaluate the performance of each type. Both exhibited better results when compared to the rain gauges records, but with different performances. The GPM(IMERG) data demonstrated the best performance among all data types, while the TRMM (0.25°) exhibited the lowest performance.

Categorical statistics confirmed the improved performance of both the TRMM(3B42V7) and GPM(IMERG) when recording low as opposed to high-intensity events. Moreover, the GPM(IMERG) demonstrated better performance than the TRMM(3B42V7), regardless of threshold or event intensity. These results provide an overview of the strong performance of the GPM(IMERG) in arid and semi-arid areas, performance that is also seen in humid regions, confirming suitability in spite of differences in topographic effects or other factors (Chen *et al*., 2018; Kim *et al*., 2017).

7. OPTIMIZATION OF RAIN GAUGES

Data acquired by GPM(IMERG) performed well at the study site. However, it tended toward underestimation of values when recording over mountainous areas and during heavy intensity precipitation events. It can therefore be said that, despite the spatiotemporal advantages of this type of data type, limitations do exist (see Chapter 6). The results of the previous chapter confirm the need for a new network of rain gauges for the purpose of providing continuous information about the test site. As terrestrial data could not be collected directly at this site, remote sensing data was used to determine optimal locations for new rain gauges. The gauges in this new network could provide a promising basis for future hydrological research in the area, such as studies on rainfall, runoff, and settlement planning.

7.1 General background

7.1.1 Data clustering

Data clustering has been described as an unsupervised machine learning technique that gathers and divides datasets into small partitions, each with nearly identical values and characteristics. This procedure also facilitates the mathematical, and especially statistical, analysis of the dataset (Mann and Kaur, 2013; Inaba *et al*., 1994). When applied, this clustering technique produces clusters with high intra-class similarity and low intra-class dissimilarity (Mann and Kaur, 2013). Data clustering techniques can be divided into various unique methods, including partition algorithms, hierarchical algorithms, spectral algorithms, grid-based algorithms, and density-based algorithms (Mann and Kaur, 2013; Inaba *et al*., 1994; Elavarasi *et al*., 2011.). The current chapter details the use of partitioning algorithms for the division of the dataset into K partitions, with each partition representing a cluster. This algorithm relies on a mean square error-minimizing objective function (Mann and Kaur, 2013; Elavarasi *et al*., 2011) defined as:

$$E = \sum\sum \|P - m_i\|^2 \qquad \text{Eq (7.1)}$$

The point in the cluster is referred to as (P) the cluster mean is denoted by (m_i). The cluster should exhibit two properties: each cluster must contain at least one point, and each point must belong to one cluster (Mann and Kaur, 2013; Elavarasi *et al*., 2011). However, the main weakness of this procedure can be seen when one point is found close to the center of another cluster, where the overlap can lead to poor results (Mann and Kaur, 2013). The partitioning algorithm itself consists of several procedures, including k-means clustering, bisecting k-means

clustering, k-medoids clustering, PAM (Partitioning Around Medoids), CLARA (Clustering Large Applications) and probabilistic clustering (Mann and Kaur, 2013; Elavarasi *et al*., 2011). K-means clustering represents the simplest and most commonly used technique in literature. The classification of this algorithm mainly depends on the selection of several dots for the construction of the initial centroids of the clusters, with the remaining dots being placed at their focal points in accordance with a minimum distance criterion (Li and Wu, 2012; Kanungo *et al*., 2002). Once all dots are assigned, the positions of the k centroids must be recalculated and dots rearranged to their closest centroids. These two steps must be repeated several times to assure the accurate positioning of the centroids (Kodinariya and Makwana, 2013) (Figure 7.1). The quality of the clustering procedure depends on how packed the clusters are, although results invariably feature dots located outside of the resultant clusters (Oyelade *et al*., 2010). This step could therefore be followed by the merging or splitting of resultant clusters based on similarity (Jain *et al*., 1999).

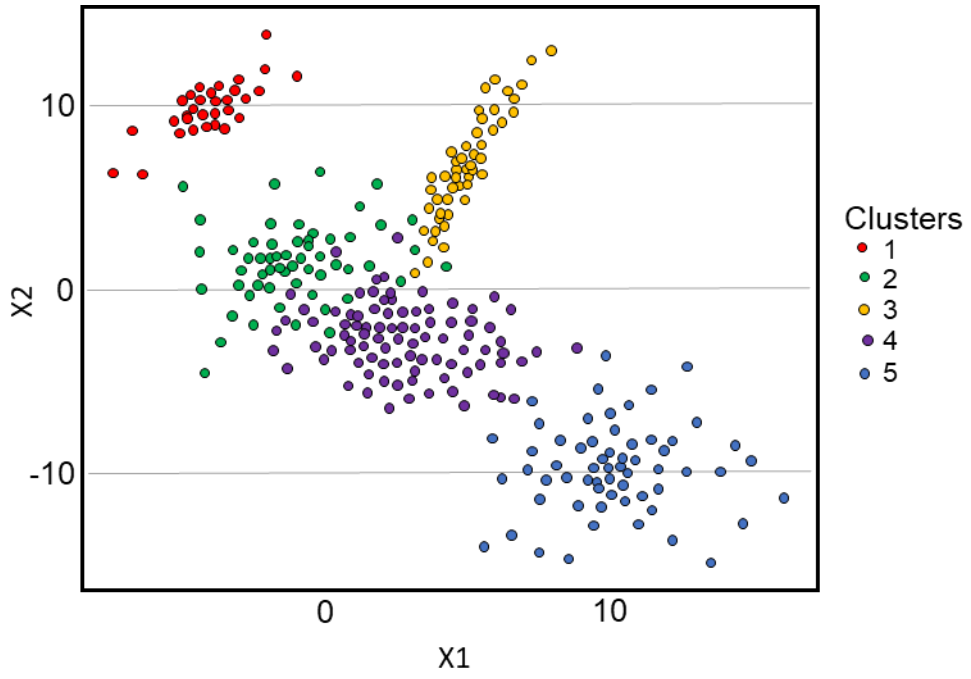


Figure 7.1: K-means clustering outlook. (Modified after Likas *et al*., 2003).

The k-means algorithm is more suited to numerical data where the number of clusters (k) has been specified in advance. This can be achieved through use of the elbow method of clustering, which is one of the oldest methods for calculating the correct number of clusters for a certain dataset (Kodinariya and Makwana, 2013). This method is a visual technique that begins with a *k* equal to 2 and increases in increments of 1. At a certain value of *k*, the numerical

scale drops significantly to form an elbow shape followed by a plateau. The point at which the plateau begins represents the *k* value (Kodinariya and Makwana, 2013) (Figure 7.2).

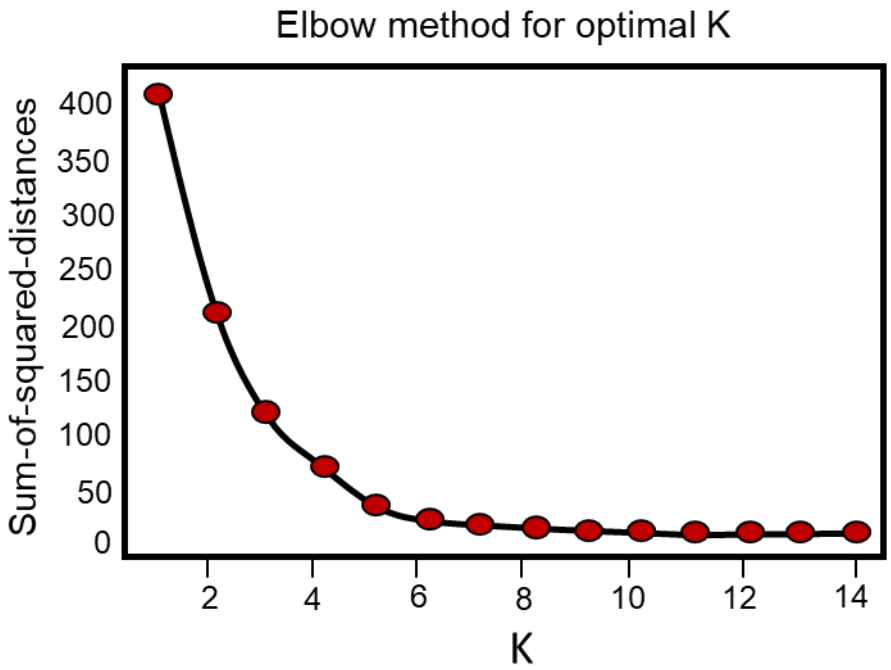


Figure 7.2: Elbow graph outlook. (Modified after Likas *et al*., 2003).

7.1.2 Kriging of Standard Error

Kriging refers to an interpolation technique that uses a given set of data points located at random positions in space to estimate its leased-biased intermediate values (Virdee and Kottegoda, 2009). This technique offers the advantage of providing significant results even when the data features high natural variability (Virdee and Kottegoda, 2009). Kriging relies on estimation by local weighted averaging, as illustrated by the following:

$$\hat{z}(B) = \sum_{i=1}^{n} \lambda_i z\,(x_i) \qquad \text{Eq (7.2)}$$

Here, $\hat{z}(B)$ represents the estimation over a fixed block of land and λ_i refers to the weights, which sum to one to ensure that no bias exists and to minimize variance (Virdee and Kottegoda, 2009). As an example (provided by GIS Geography), the location of the purple point in Figure 7.3 can be predicted by taking the inverse weighted distance of the closest three input points (the values of 12, 10 and 10). Based on these distances, once can calculate how far each input point is from the unknown location, resulting in a value of 11.1. The calculation follows below:

((12/350) + (10/750) + (10/850)) / ((1/350) + (1/750) + (1/850)) = 11.1 Eq (7.3)

Kriging-based geostatistics was a procedure developed for the determination of optimal networks (Adhikary *et al*., 2015). An important feature of this approach is the provision of a defined measure of error, which in the present study helped form the basis for the rain gauge network design, as a controlled minimization of kriging error allowed for the optimal configuration to be determined (Adhikary *et al*., 2015). Kriging of standard error can be calculated on a separate layer in the ArcGIS 10.5 software. This provides a measure of the confidence of how likely a prediction is to be true and was used to both estimate locations for new samples and to minimize the number of samples required for optimum results (Siska *et.al*., 2005; Aziz1 *et.al*., 2019; Webster and Oliver, 2007). The mathematical background of this method also takes into account the optimal weighting of each sample, in order to achieve an optimum interpolation of sample values (Kassim and Kottegoda, 1991). Furthermore, the area where the least data is input becomes the region of highest standard error.

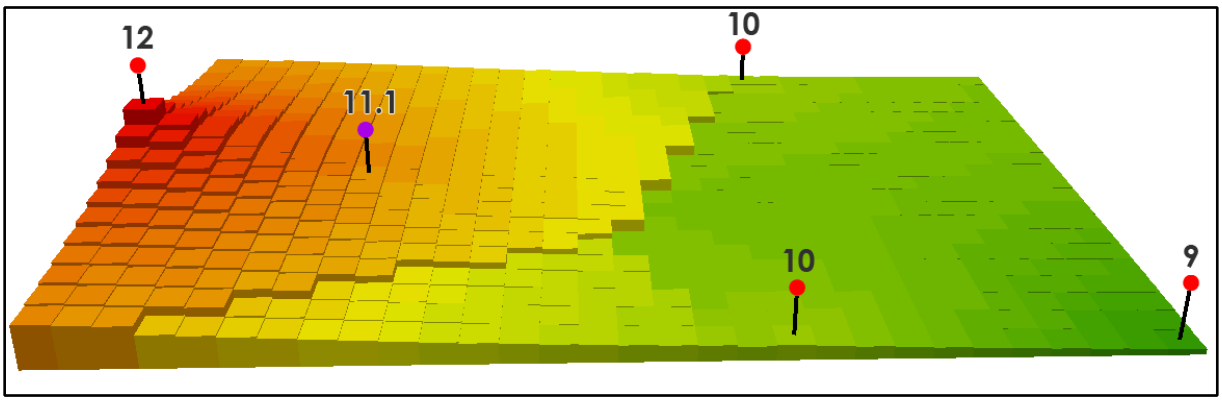


Figure 7.3: Illustration of the mechanism of the interpolation technique used. (Source: GIS Geography, 2020).

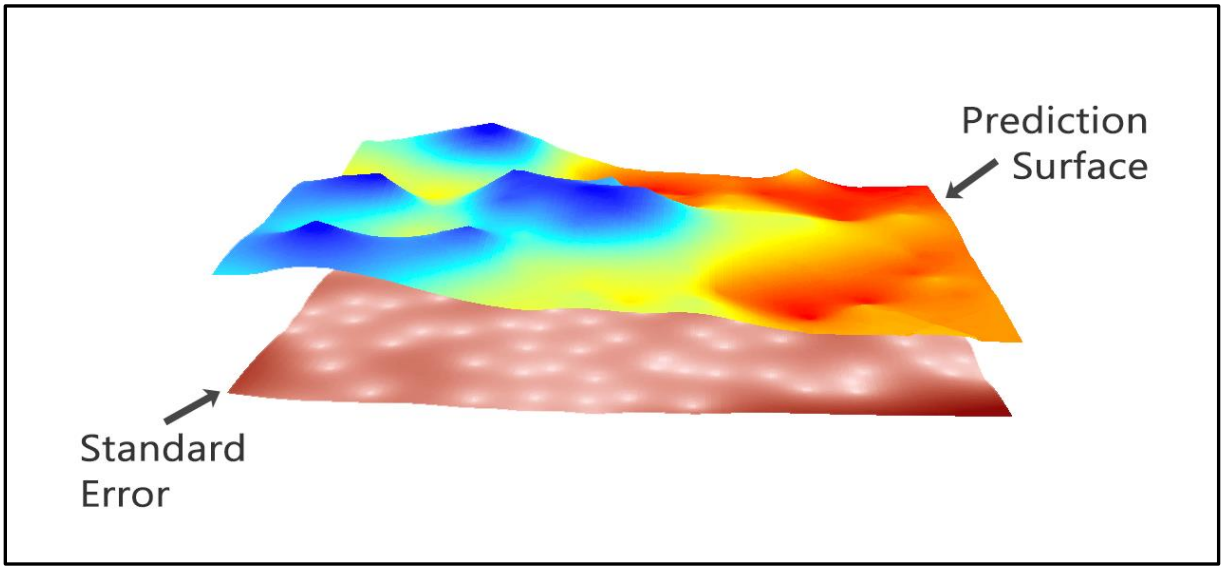


Figure 7.4: Display of the surface predicting rainfall via interpolation, with the standard error surface below. (Source: GIS Geography, 2020).

To efficiently apply kriging, it is important to incorporate a variogram model that defines the spatial structure of the samples (Adhikary *et al*., 2015). A variogram is a statistical function that tests the spatial correlation between samples measured at a certain location (Aziz *et al*., 2019). It provides graphical illustration of the differences in distance between all pairs of sampled locations (Aziz *et al*.,2019). Bachmaier and Backes (2008) and Robinson and Metternicht (2006) stated the equation of the theoretical variogram as the following, with $Z(\vec{x})$ and $z(\vec{x}+\vec{h})$ representing the spatial positions separated by a vector $\vec{h}$:

$$y[\vec{h}] = \frac{1}{2}E\left[\left[z(\vec{x}+\vec{h}) - Z(\vec{x})\right]^2\right] = \frac{1}{2}Var\left[z(\vec{x}+\vec{h}) - Z(\vec{x})\right] \qquad \text{Eq (7.4)}$$

$Z(\vec{x})$ and $z(\vec{x}+\vec{h})$ symbolize random variables. $y[\vec{h}]$ depends mainly on the separation vector $\vec{h}$, but not on the location ($\vec{x}$). The additions $z(\vec{x}+\vec{h}) - Z(\vec{x})$ are assumed to have no drift $E\left[\left[z(\vec{x}+\vec{h}) - Z(\vec{x})\right]\right]$=0 for all $\vec{h}$ and $\vec{x}$. The above theoretical variogram permits the identification of all sample values at any distance. Additionally, it offers a unique solution for the calculation of the weight of each sample, which can then be used to prepare the associated kriging interpolation maps (Adhikary *et al*., 2015). Several types of variograms exist and are primarily defined by their shape. However, exponential, gaussian, and spherical models are those most commonly used in hydrology (Cecinati, 2017; Adhikary *et al*., 2015). The equations of the aforementioned models (per the authors) follow:

$$y(h) = c_0 + c_1\left[1 - e\left(-\frac{3h}{a}\right)\right] \qquad \text{Eq(7.5)}$$

$$y(h) = c_0 + c_1\left[1 - e\left(-\frac{3h^2}{a^2}\right)\right] \qquad \text{Eq(7.6)}$$

$$y(h) = c_0 + c_1\left[1.5\left(\frac{h}{a}\right) - 0\cdot 5\left(\frac{h^3}{a^3}\right)\right] \qquad \text{Eq(7.7)}$$

Equation (7.5) serves as the formula of the exponential model (Adhikary *et al*., 2015; Aziz *et al*.,2019), while equations (7.6) and (7.7) represent the gaussian and spherical models, respectively (Adhikary *et al*., 2015). Nugget, range, and sill as denoted by c_0, a, and $c_0 + c_1$, respectively. Figure (7.5) identifies the different components of a variogram, illustrating how each assist in the accuracy of the kriging interpolation. The nugget effect refers to the nonzero intercept of the variogram. It serves as an overall estimate of the error caused by both measurement inaccuracy and environmental variability at scales too fine to be resolvable by the

sampling interval (Clark, 2010). Range describes the distance beyond which no autocorrelation between variables exists (Adhikary *et al*., 2015). Sill is defined as the constant semi-variance of relative values beyond the range (Adhikary *et al*., 2015).

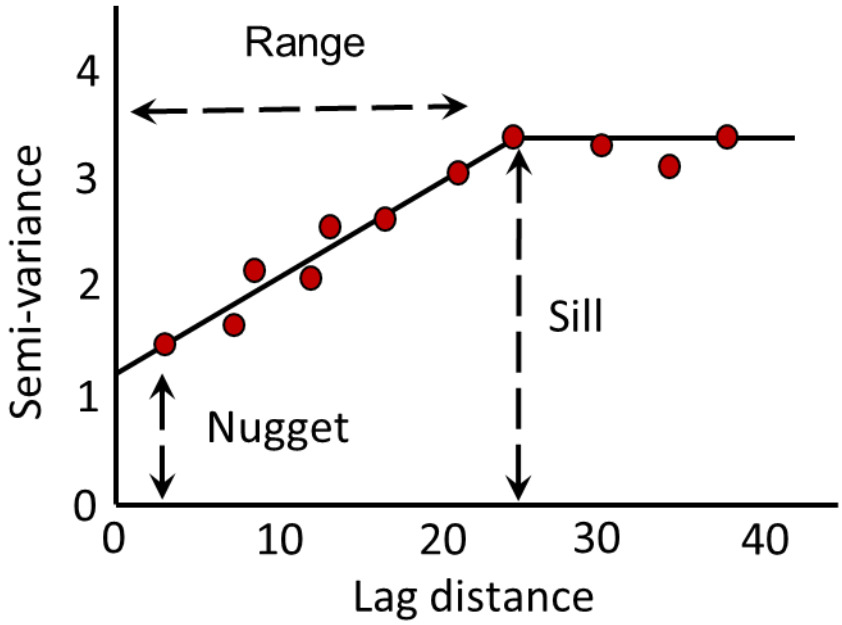


Figure 7.5: The components of a variogram. (Modified after GIS Geography, 2020).

7.1.3 The empirical cumulative distribution function (ECDF)

The ECDF models empirical (*i.e.* observed) data. It represents the probability distribution of the sample, as opposed to the population (Hammerla *et al*., 2013). It was constructed by Lahiri *et al*. in 1999 as a random distribution function for providing a statistical summary of a random field over a given region. It allows one to plot an ordered sample of a population from its minimum to maximum values and subsequently create a visualization of how the sample is distributed across the population (Hammerla *et al*., 2013) (Figure 7.6).

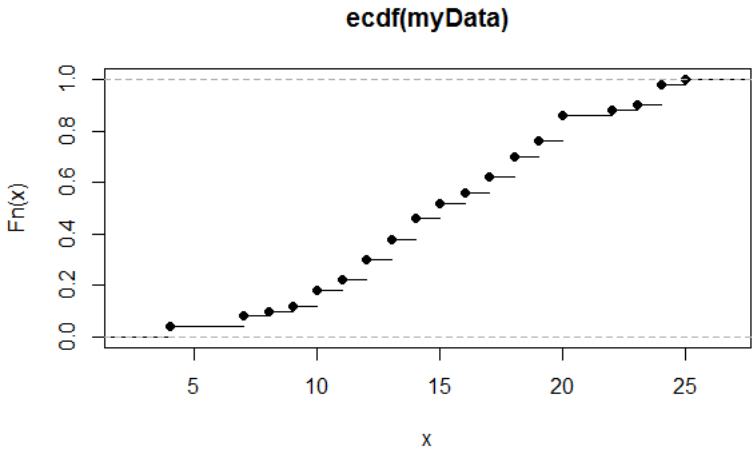


Figure 7.6: example for Empirical cumulative distribution function curve produced in RStudio.

7.2 Data and methods

To achieve the goal of this chapter, two types of remote sensing data were used. The first was derived from GPM(IMERG) and featured a 0.1° (or 10×10km) spatial resolution, which was the resolution that had demonstrated the greatest accuracy in detecting the spatiotemporal resolution of rainfall at the test site (Chapter 6). Four scenes of this data type were used, covering four significant events during the period of 2015 to 2018 Each of the abovementioned scenes represented the daily composite of half-hourly scenes downloaded from the official site of the NASA Mirador. These four scenes were used due to each one reflecting unique intensities, which provided varied effects over the test site (Chapter 6). The second data source was a 90x90m-resolution SRTM3 DEM file that was used to obtain nearly global data on elevation (Chapter 3), which was directly related to rainfall intensity (Figure 7.7). Two types of software, RStudio, and ArcGIS10.5, were used to process the relevant data.

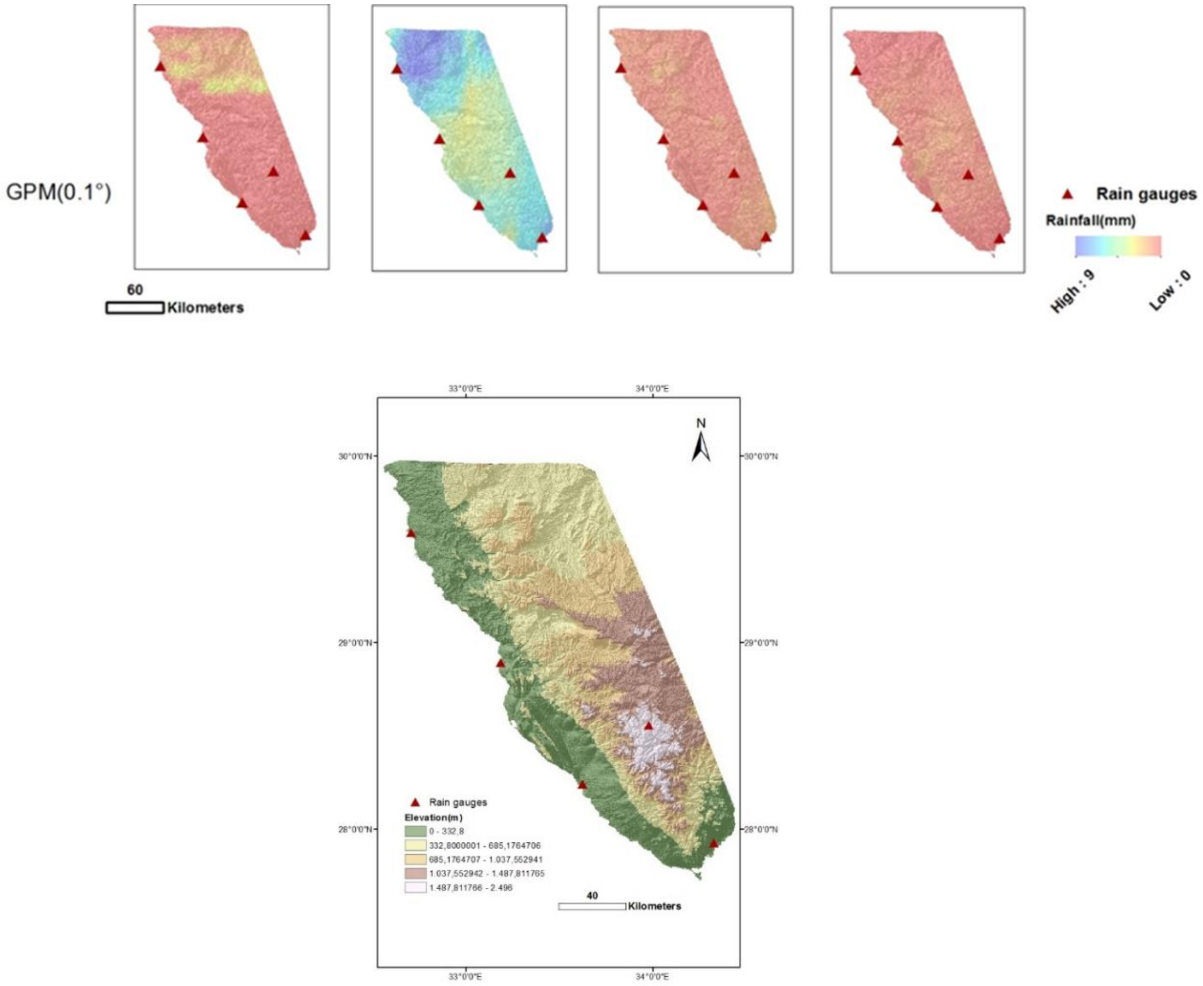


Figure 7.7: Visualization of the datasets used to optimize the number of rain gauges, from GPM(IMERG) and SRTM DEM (90×90m).

7.2.1 Data clustering

The DEM file was resampled to match the pixel size of the GPM(IMERG) data, (10×10km). All previously mentioned remote sensing-derived data scenes were saved as TIFF files, stacked, and converted to point data. The clustering procedure began by calculating an elbow graph for the determination of the optimum number of clusters (section 7.1.2). The elbow graph is a visual procedure, with the 'elbow' portion of the graph displaying a wide area prior to plateauing. Three different numbers of clusters were compared. Each number produced a single point file that was converted into a raster image file and then a shape file. The centroids were calculated for the three resultant shape files. Three packages of RStudio were used to complete the clustering procedure (raster, rgdal, and ggplot2). The raster package was used to add the four raster layers to the workspace variables. It also contains a function, 'stack', that creates a raster stack for multiple layers. The regdal package enables the user to read different types of files, including shape files, point files, and vector files. In the case of this study, it was used to read the point files that were produce. The third package, ggplot2, can be used to create advanced graphs and, in the present study, was utilized to create the k-mean point graph (Hijmans *et al*., 2019; Bivand *et al*., 2017). The code responsible for clustering procedures is provided below.

```
library(raster)
library(rgdal)
# load data raster
dem <- raster("DEM.tif")
rainfall1 <- raster("Rainfall2015.tif")
rainfall2 <- raster("Rainfall2016.tif")
rainfall3 <- raster("Rainfall2017.tif")
rainfall4 <- raster("Rainfall2018.tif")
# stack raster files
RasterFinal1 = stack(rainfall1, rainfall2)
RasterFinal2 = stack(rainfall3, rainfall4)
RasterFinal3 = stack(RasterFinal1, RasterFinal2)
RasterFinal4 = stack(RasterFinal3, dem)
# convert to csv
PointData <- as.data.frame( rasterToPoints(RasterFina4) )
PointData = PointData[complete.cases(PointData),]
str(PointData)
#Elbow Method for finding the optimal number of clusters
```

```
set.seed(123)
# Compute and plot wss for k = 2 to k = 20.
k.max <- 20
data <- PointData [,5:6]
wss <- sapply(1:k.max,function(k){kmeans(data, k, nstart=50, iter.m15)$tot.withinss})
plot(1:k.max, wss, type="b", pch = 19, frame = FALSE, xlab="Number of clusters K",
ylab="Total within-clusters sum of squares")
OptimumCluster = 6
# k means
clusters <- kmeans(PointData[,3:4], OptimumCluster)
PointData$OptimumNuCluster <- as.factor(clusters$cluster)
str(PointData)
# plot it here
library(ggplot2)
ggplot() + geom_point(aes(x = x, y = y, colour = as.factor(OptimumNuCluster)),data = PointData)
+ggtitle("Cluster using KMean")
# export data
write.csv(PointData,"PointData.csv")
```

Code (7.1): Loading, stacking, and clustering raster data in RStudio.

7.2.2 ECDF and kriging of standard error

The locations of the calculated centroids (for the three selected k values) were tested by an empirical cumulative distribution function curve (ECDF) (section 7.1.2). This was done to select the optimum k value, or number of clusters, for the test site. This evaluation depended on the aforementioned raster layers that were input (Figure 7.8). The sum of events from 2015, 2016, 2017, and 2018, the upper and lower limits of the cumulative rainfall scene, and the DEM file were all recorded. The spatial coverage of each k value was tested by ECDF graph for both rainfall and elevation. Although the optimum number of clusters had been chosen, there remained spaces of insufficient coverage, as visible on the ECDF curve. Consequently, an additional procedure was used in order to compensate for the shortcomings of the original clustering procedure (kriging of standard error). This new procedure incorporated both the existing gauges and those resulting from the predictive clustering technique to complete the interpolation of the standard error. As the area containing the least data produces the highest error, it is important to increase the number of points there. Consequently, looping was used to add a sufficient number of points (gauges) in ArcGIS 10.5. The points added were tested by

ECDF following each trial in order to detect the least costly point at which the procedure could be terminated.

```
plot(ecdf(Finalgauges$F2018gpm_1), xlab = "Rainfall (m)", ylab =
"ECDF",cex.lab=1.5,cex.axis=1.5,lty=1,lwd=2, main="2018",pch=19,cex=1)
```

Code (7.2): Calculating the ECDF curve for rainfall and elevation data.

7.3 Results

7.3.1 Data clustering

The elbow graph that was produced featured a wide elbow-shaped area, which lead to doubts about the optimum cluster number (Figure 7.8). As a result, three different k-values (3, 6, and 9) were visually selected and tested. Clustering of the three selected values in RStudio resulted in three point-shape files containing 3, 6, and 9 classes, which were converted to raster files (Figure 7.9) and then polygon-shape files by ArcGIS10.5 software (Figure 7.10) This resulted in 3, 6, and 9 centroids for each shape file (Figure 7.10).

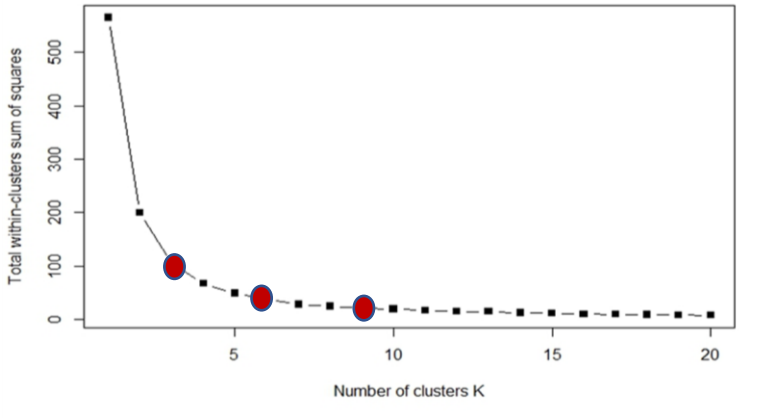


Figure 7.8: Cluster number as selected by elbow graph analysis.

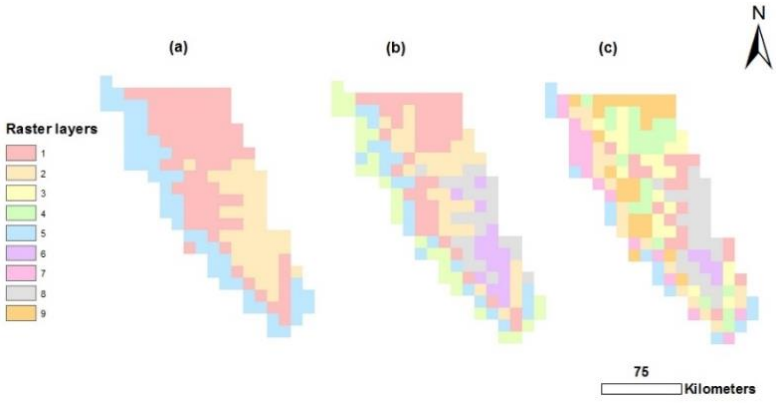


Figure 7.9: Three raster layers based on the point file created by RStudio software. (a) represents the 3-cluster raster layer, (b) the 6-cluster layer, and (c) the 9-clusters layer.

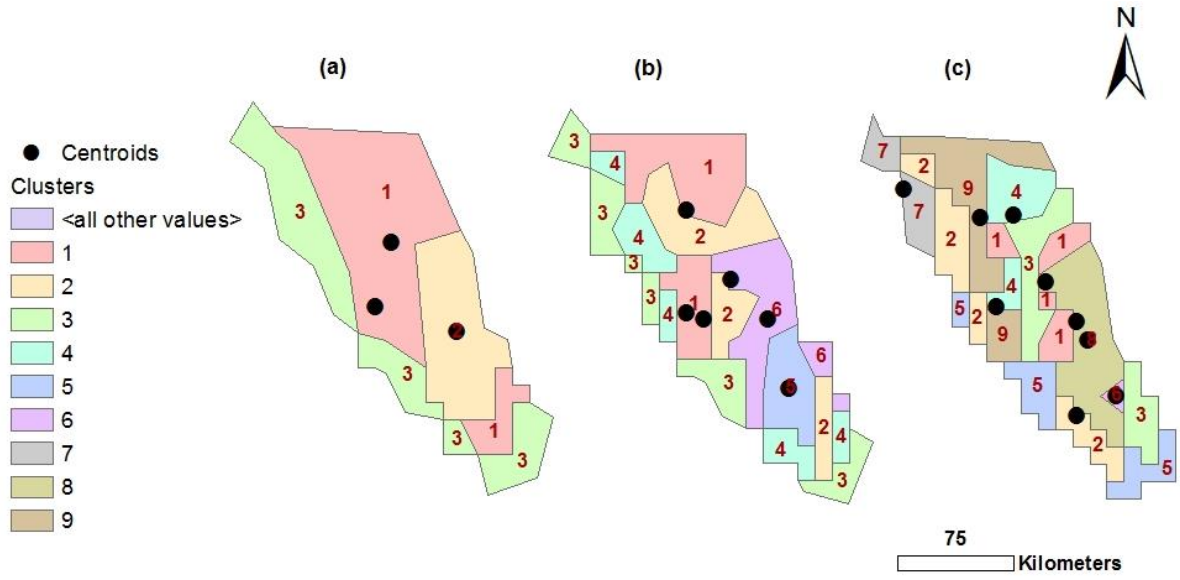


Figure 7.10: Converted raster-to-polygon file with centroid locations shown (a, b, and c representing the 3-, 6-, and 9-centroid clusters, respectively).

7.3.2 Checking by ECDF

To determine the optimum k value as well as evaluate the spatial coverage of the proposed centroids over the rainfall and elevation spectrums, ECDF was applied to the previously mentioned 3-, 6-, and 9-cluster centroids. This required the input of the upper and lower limits of rainfall (4-16mm) and elevation (0-2000m), which were both used as x-axis scales (Figure 7.11).

The ECDF of the 3 centroids revealed extensive vacant space between the 7mm and 16mm values in the rainfall map as well as between the 0m to 500m, 500m to 1000m, and 1200m to 2000m elevations in the associated map. The ECDF of the 6-centroid distribution demonstrated vacant space between the 7mm and 9mm as well as the 9mm and 16mm values. The elevation curve also revealed space between the values of 0m to 500m and 600m to 1400m. The ECDF of the 9-centroid distribution displayed less vacant space, this time being present in the rainfall range of 9mm to 16mm and in the elevation curve from 100m to 600m and 1300m to 1700m (Figure 7.11). While all the abovementioned results revealed gaps in the distribution of the centroids, the nine-centroid distribution featured the best coverage when compared to the 3- and 6-centroid distributions. Consequently, the nine centroids (plus the five existing gauges, for a total of 14), were used as a foundation for a subsequent procedure intended to enhance their spectral coverage and efficiency.

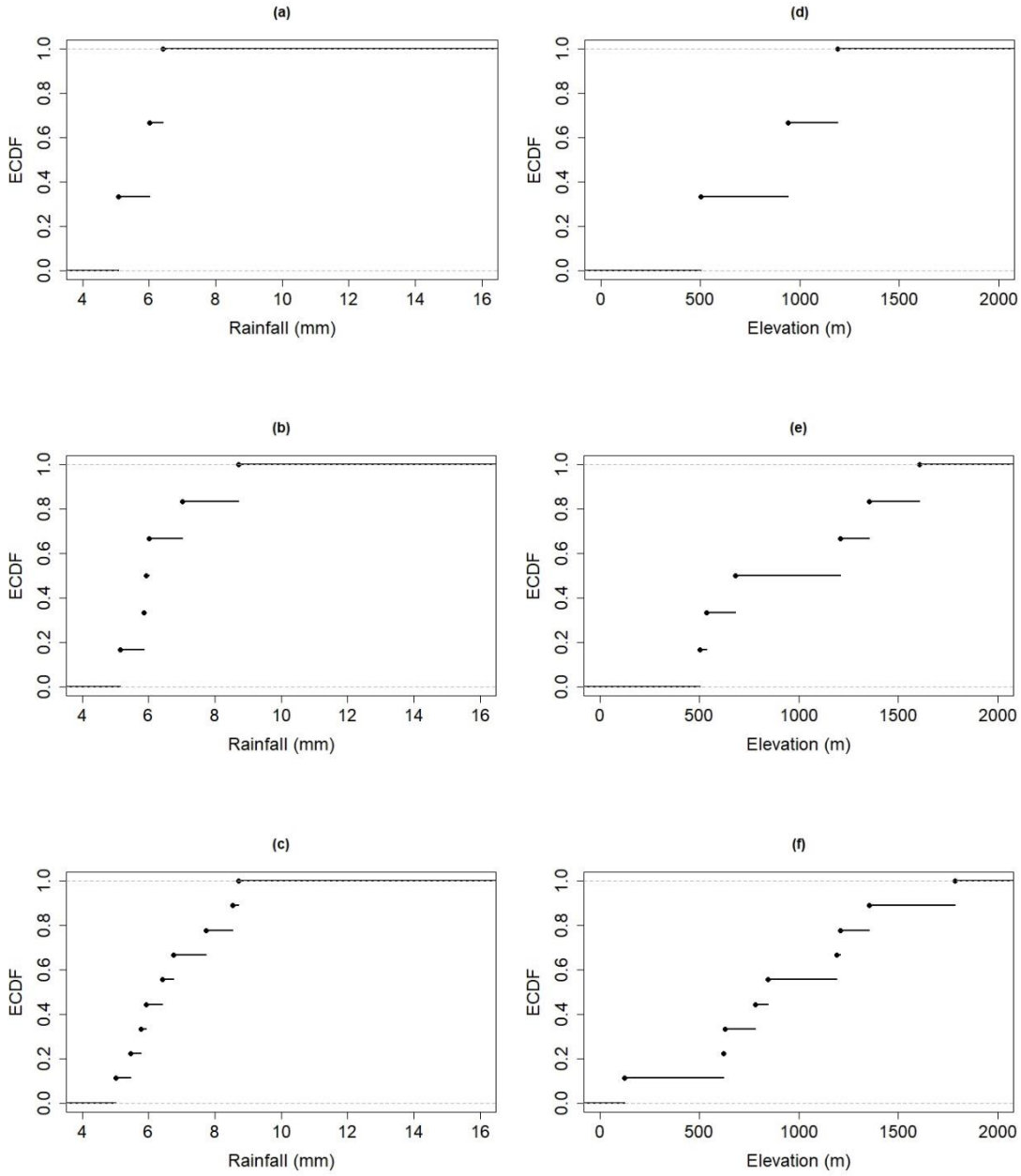


Figure 7.11: ECDF of rainfall and elevation spectrums as covered by the previously mentioned 3-, 6-, and 9-centroid clusters. The x-axis of (a), (b), and (c) feature the same limits as the sum of the rainfall events recorded between 2015 and 2018 (as captured by GPM(IMERG)). The x-axis of (d), (e), and (f) is compatible with the test site elevation.

To identify the reason behind insufficient coverage despite optimum cluster number, the coverage limits of each cluster were recorded (Tables 7.1 and 7.2). Cluster number one covered the precipitation rates ranging from 5 to 11.5mm, although the complete range was from 4 to 16mm. This focal point was therefore not sufficient for the reflection of the complete cluster range. As elevation presented a greater range in values than rainfall, each associated cluster covered a smaller range of elevations.

Table 7.1: Lower and upper limits of each resultant cluster as determined by using the complete range of rainfall (4-16mm).

Cluster No.	Rainfall Min.	Rainfall Max.
1	5	11.5
2	3	13.3
3	5	9.7
4	3.7	10.3
5	5	13.6
6	5.2	7.7
7	5	14
8	4.8	10.3
9	3.7	12.7

Table 7.2: Lower and upper limits of each resultant cluster as determined by using the complete range of elevation (0-2000m).

Cluster No.	Elevation Min.	Elevation Max.
1	721	1115
2	262	404
3	448	860
4	522	691

5	6	103
6	1609	2000
7	124	369
8	1172	1524
9	436	556

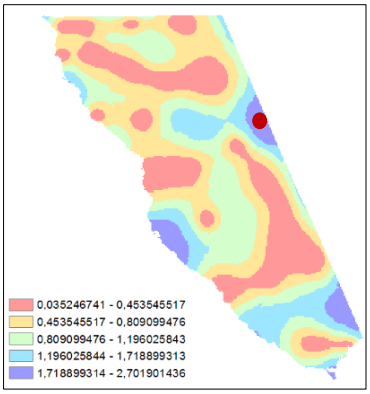
7.3.3 Kriging of standard error

To eliminate the standard error found in the previously-mentioned 14-gauge design, a second technique was utilized. This involved adding a single gauge a time, up through 22 iterations (Figure 7.12). Following the addition of the first point, mean standard error decreased from 7.2 to 1.2 (Figure 7.13). By the third trial, the resultant graph formed a peak, as the standard error increased from 1.2 to 2.0. Following this, it decreased down to 0.7 by the fifth trial, after which a small depression formed under the sixth and seventh trials (values of 0.6 and 1.0, respectively. During the eighth and ninth trials, a plateau was reached, with very small negative deviations in the 10^{th}, 11^{th}, and 12^{th} trials (0.79, 0.77, and 0.96, respectively). This plateau continued for the 13^{th}, 14^{th}, and 15^{th} trials, with a slightly depression visible for the 16^{th} to 19^{th} trials (0.92, 0.84, 0.80, 0.80, and 0.4, respectively). By the 20^{th} trial, the curve again began to rise slightly, with values of 0.84, 0.86, and 1.08 for the 20^{th}, 21^{st} and 22^{nd} iterations. The final trial recorded a value of 1.02.

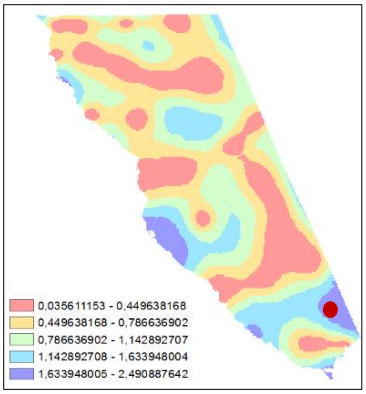
To evaluate the resultant Gaussian variograms, Tobler's First Law of Geography was followed. This law states that everything is related to everything else, but that nearer things are more related than distant things (GIS Geography, 2020). The variograms that were produced displayed distance in degrees on the x-axis and variance between variables on the y-axis. From here one can see that by the 22^{nd} trial, as the distance between two points (h) increased, the variance (y) also increased. Moreover, 22 of the graphs exhibited binned points dispersed around the model, indicating high variance. Nineteen graphs displayed positive autocorrelation, while four showed none (non-raising model) (Figure 7.14). Variograms also revealed fluctuating nugget, sill, and range values (Table 7.3). However, in the final trial, binned points were fitted around the model, indicating that the least variance (between neighboring and distant

points) could be found here. Trial number 22 displayed no nugget effect at the origin along with the highest sill value, resulting in a random field with a variance of 0.97 and a range of 0.56. Given these values, the 22nd trial was considered optimum and further iterations were not performed. The best performance at lowest cost had been identified.

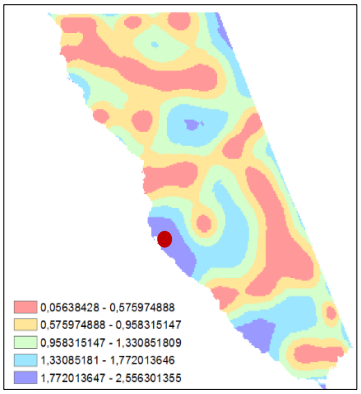
Trial 9



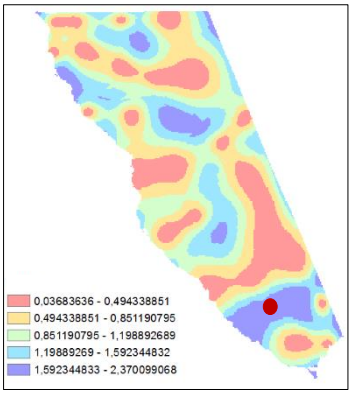
Trial 10



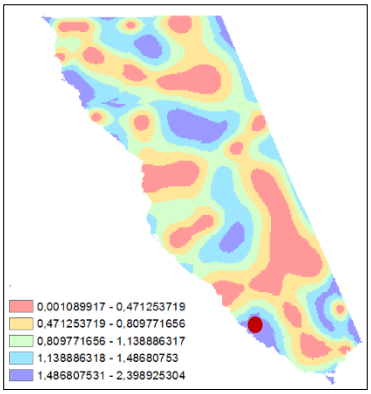
Trial 11



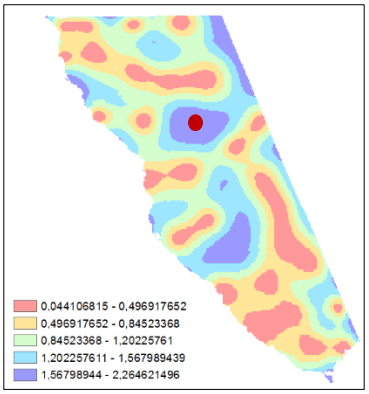
Trial 12



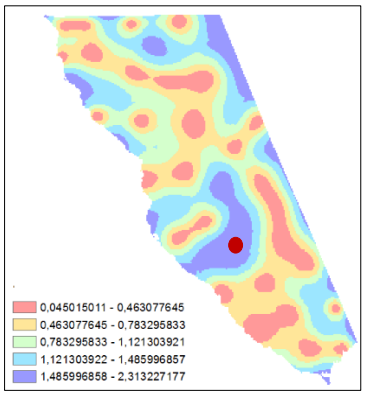
Trial 13



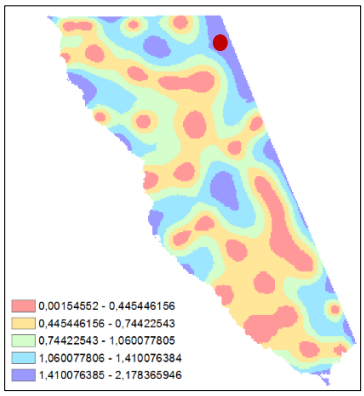
Trial 14



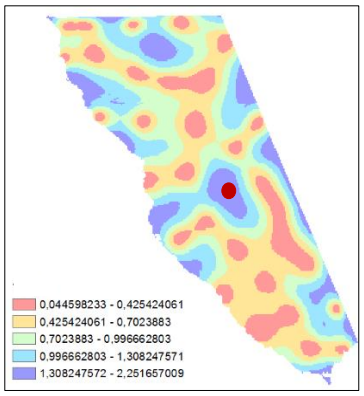
Trial 15



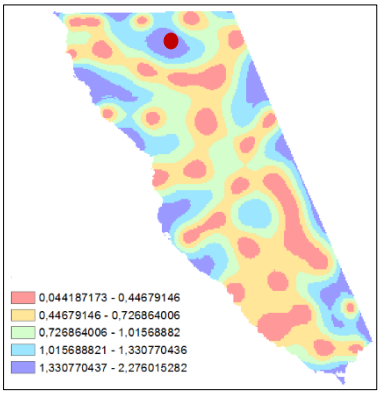
Trial 16



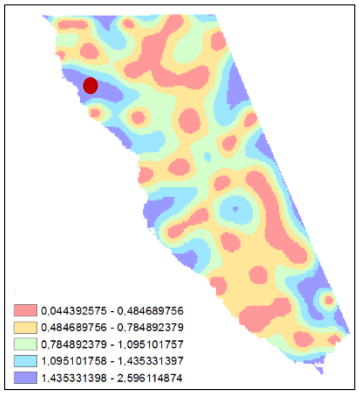
Trial 17



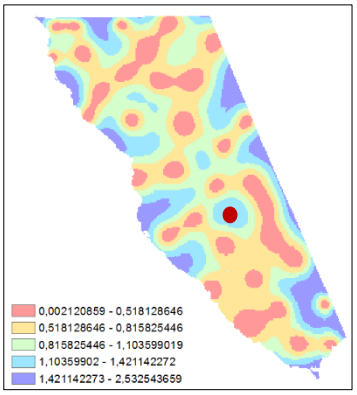
Trial 18



Trial 19



Trial 20



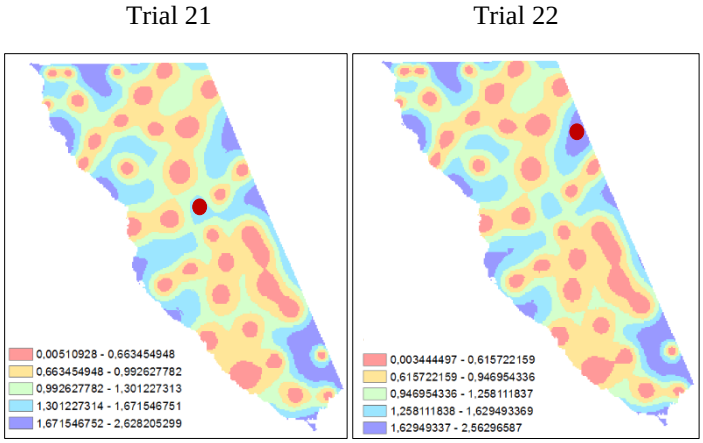


Figure 7.12: Results of a loop run to reduce standard error via the addition of one gauge at a time to the rain gauge configuration. The complete procedure involved 23 trials, with the first being the configuration established by clustering, in addition to the existing gauges.

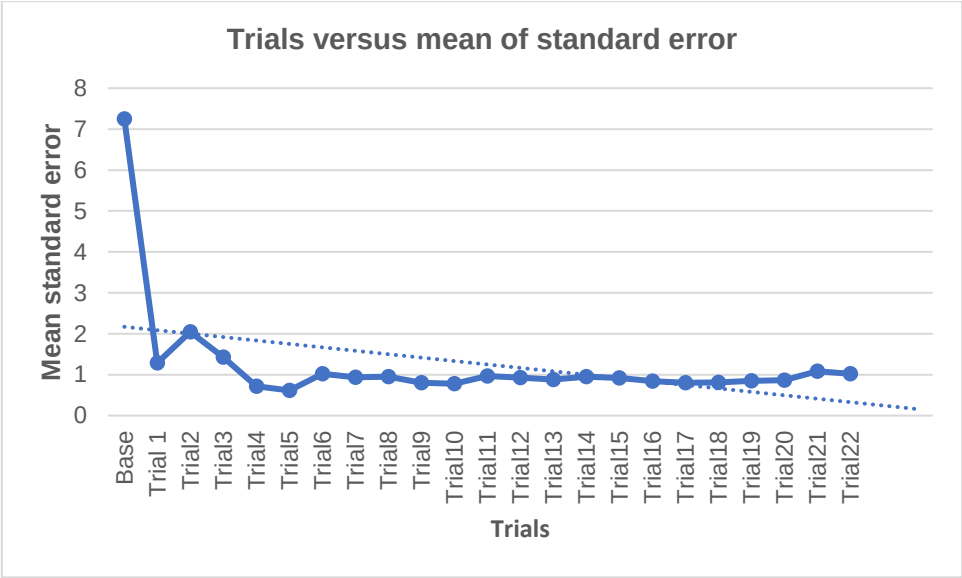
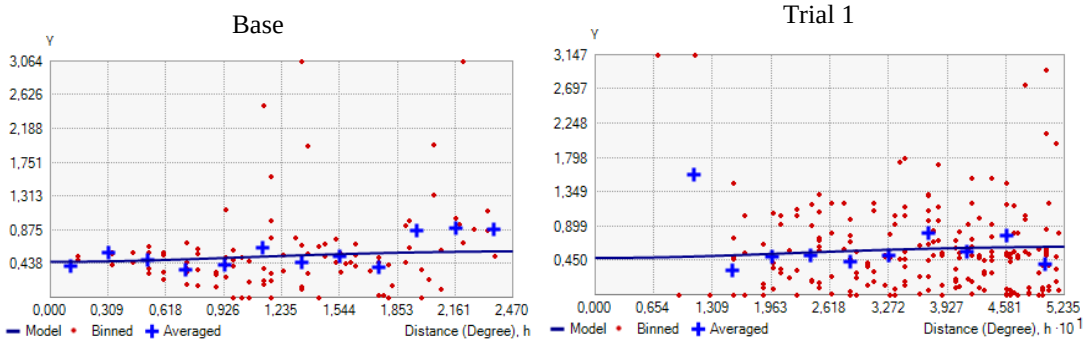
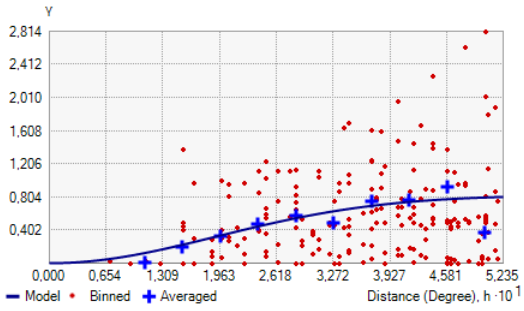


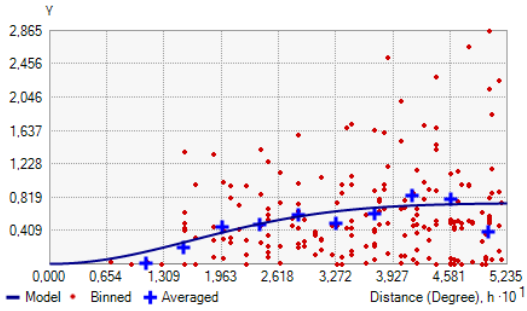
Figure 7.13: Changes in captured mean error values upon each subsequent iteration.



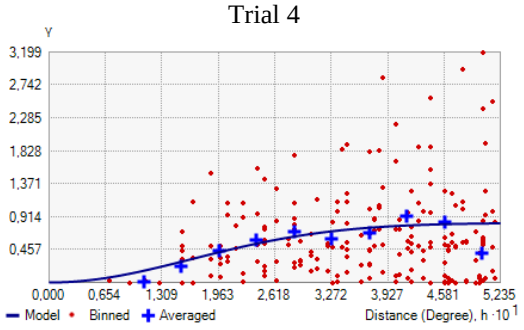
Trial 2



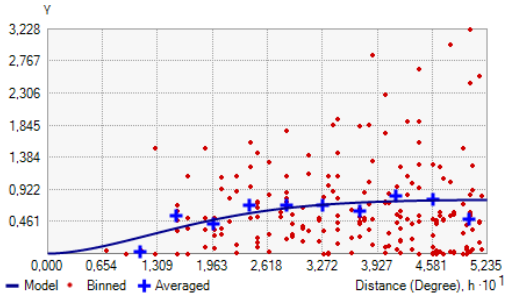
Trial 3



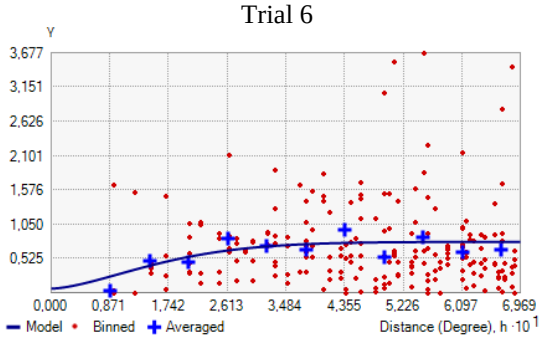
Trial 4



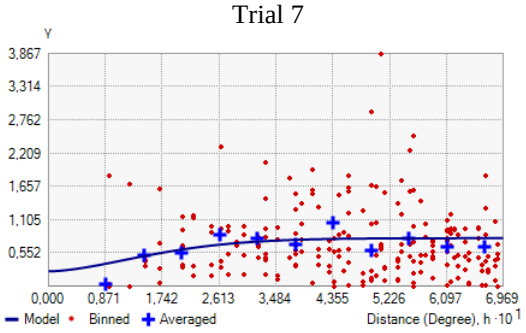
Trial 5



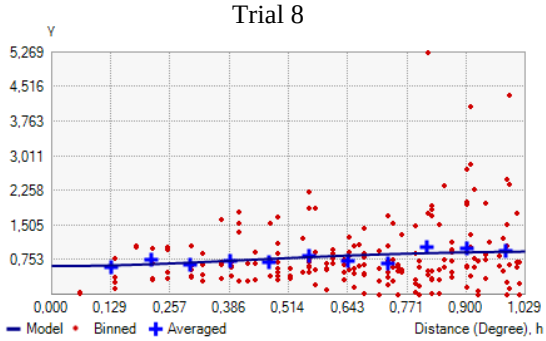
Trial 6



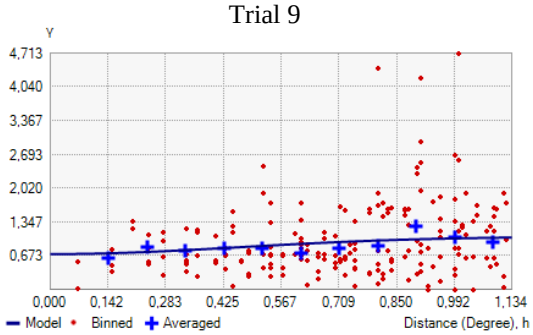
Trial 7



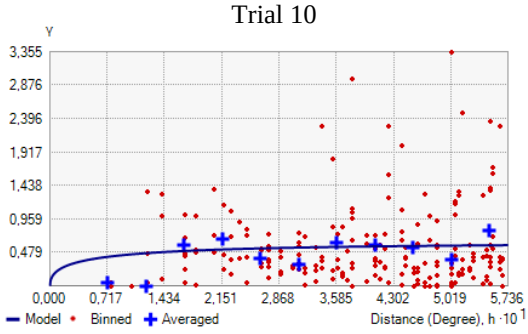
Trial 8



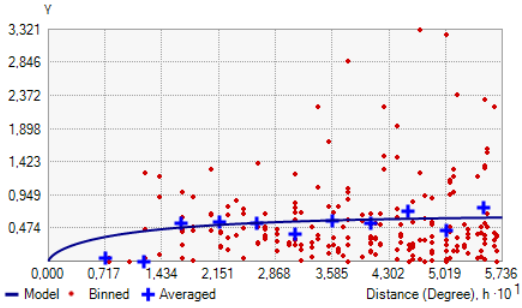
Trial 9



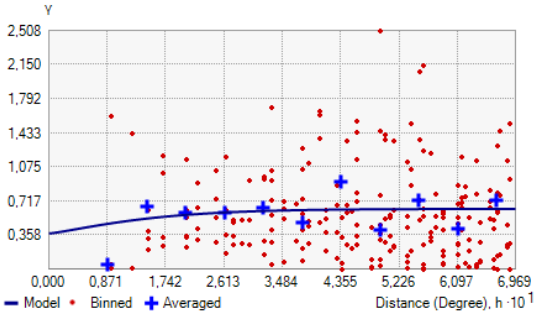
Trial 10



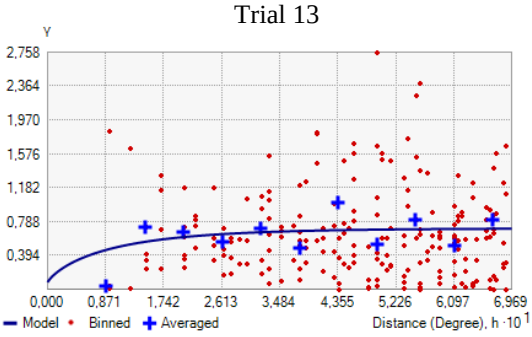
Trial 11



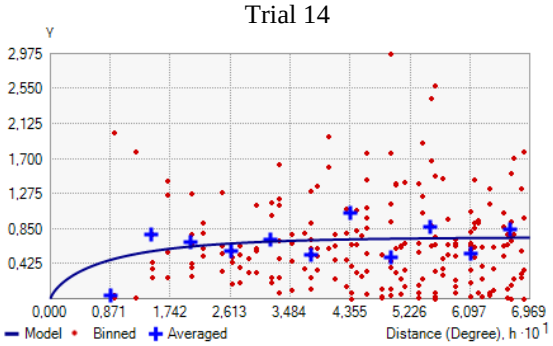
Trial 12



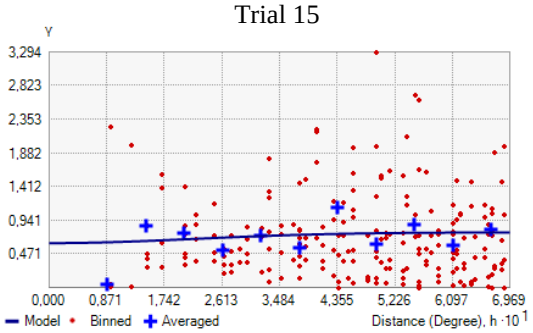
Trial 13



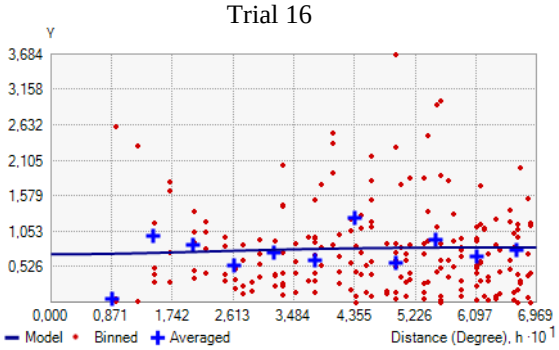
Trial 14



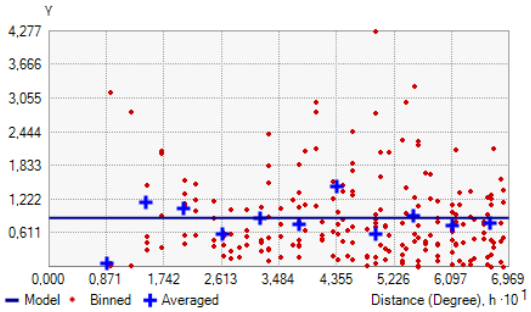
Trial 15



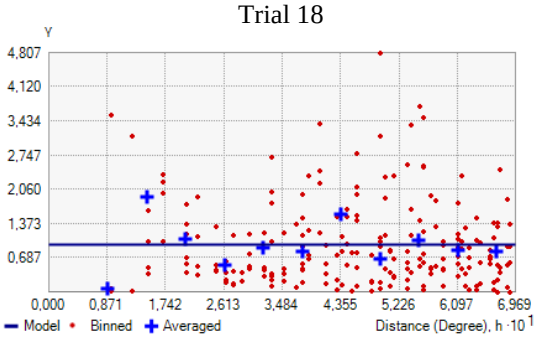
Trial 16



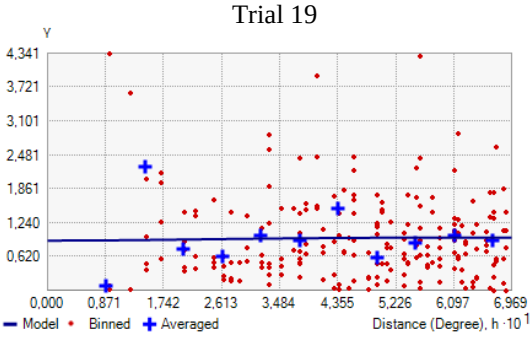
Trial 17



Trial 18



Trial 19



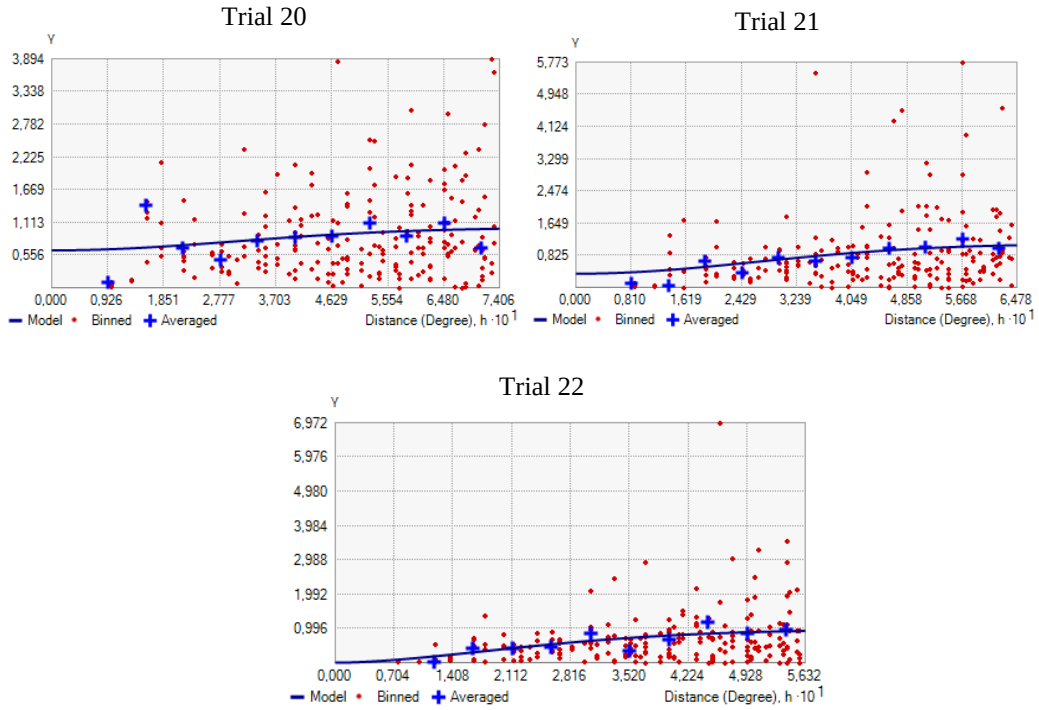


Figure 7.14: Variograms produced upon each trial.

Table 7.3: Nugget, sill, and range values as recorded by each variogram over the 22 trials.

	Nugget	Sill	Major Range
Base	0,46	0,32	2,47
Trial1	0,48	0,33	0,52
Trial2	0,008	0,81	0,47
Trial3	0,008	0,73	0,40
Trial4	0,007	0,81	0,40
Trial5	0,00	0,77	0,37
Trial6	0,06	0,65	0,37
Trial7	0,24	0,31	0,37
Trial8	0,61	0,28	1,02
Trial9	0,70	0,36	1,13
Trial10	0,00	0,61	0,57
Trial11	0,00	0,65	0,57
Trial12	0,37	0,12	0,37
Trial13	0,08	0,54	0,37
Trial14	0,00	0,74	0,37
Trial15	0,62	0,48	0,53
Trial16	0,71	0,61	0,54
Trial17	0,88	0,88	0,69
Trial18	0,95	0,95	0,69
Trial19	0,91	0,86	0,55
Trial20	0,63	0,25	0,74
Trial21	0,35	0,41	0,64
Trial22	0,00	0,97	0,56

7.3.4 Double checking by ECDF

The ECDF was again utilized to test the combined locations of the existing and proposed rain gauges, which totaled 36 in number. Figure (7.15) (a) illustrates the results of an ECDF

test on the extent of rainfall spectrum coverage provided by the entire proposed gauge network, with very promising results supporting the location selection technique. Figure (7.16) (b) displays the results of a test on the extent of coverage by elevation, with the graph again presenting promising results, aside from a small vacant area between 1350 and 1650m. To further verify the techniques, gauge coverage was tested for each precipitation event. Complete coverage was seen for the 2016, 2017, and 2018 events (Figure 7.16). However, a gap between 3.8 and 6mm did exist for the 2015 event. The combined proposed and existing gauges are presented below (Figure 7.17).

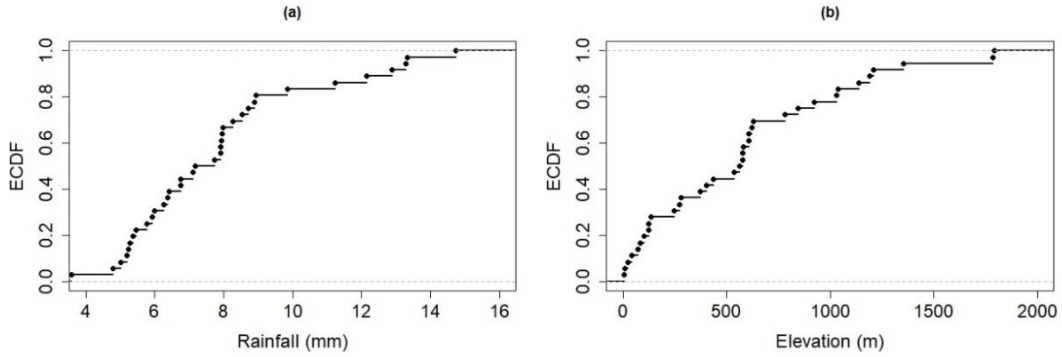


Figure 7.15: ECDF tests on the efficiency of the proposed gauges. (a) presents the rainfall spectrum, defined as the sum of the four previously mentioned events (2015, 2016, 2017, and 2018). (b) presents the elevation spectrum.

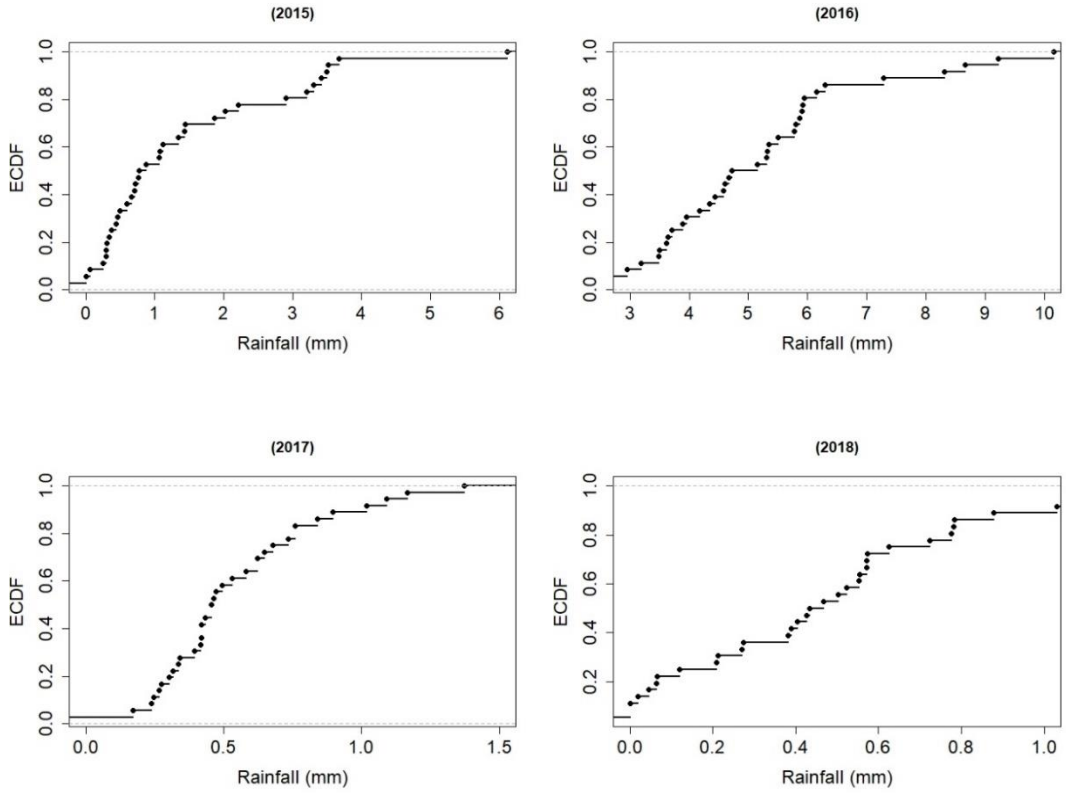


Figure 7.16: ECDF tests on the efficiency of the proposed gauges in covering each single event. The limits on the x-axis are compatible with the limits of each single event, as recorded by GPM(IMERG).

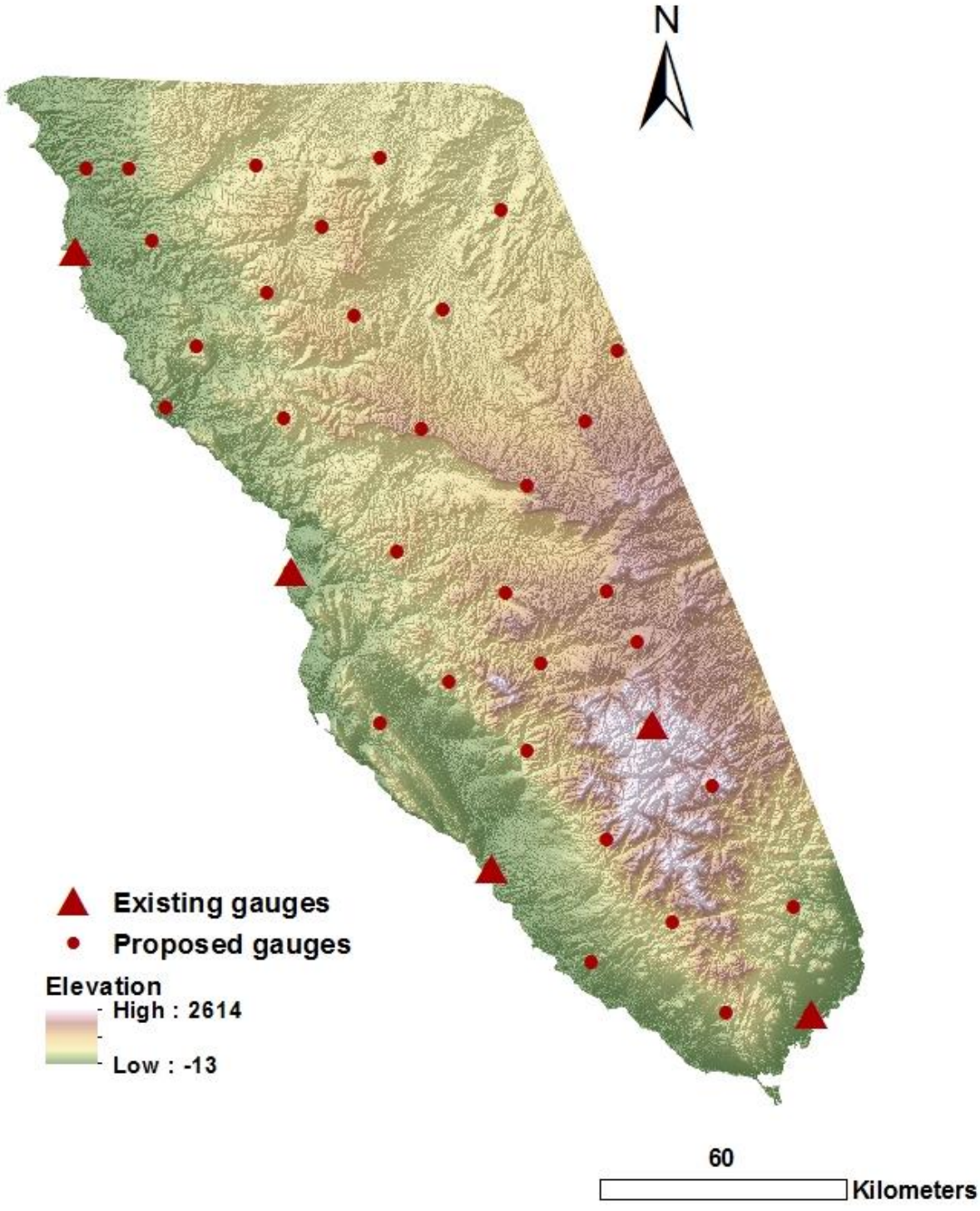


Figure 7.17: Visualization of the thirty-one proposed gauges overlaid upon the DEM file of the study area.

Table 7.4: The locations of the proposed stations in decimal degrees.

No	x	y
1	34.109079	28.422905
2	33.87783	28.306249
3	33.942809	28.740471
4	33.87684	28.8479
5	33.701398	29.079282
6	33.417455	28.936052
7	33.517566	29.464686
8	33.325042	29.451862
9	32.881869	29.617134
10	32.736867	29.773296
11	32.832045	29.77627
12	33.108658	29.782219
13	34.137775	27.929214
14	32.912352	29.252789
15	33.168144	29.228994
16	33.382296	29.797091
17	33.132452	29.502632
18	33.531012	28.651975
19	33.831419	29.220071
20	34.286491	28.158237
21	33.381676	28.560019
22	34.0219	28.126362
23	33.84344	28.037132
24	33.470817	29.206161
25	33.704004	28.501841
26	33.646254	29.684079
27	33.655772	28.846509
28	33.253643	29.646008
29	32.980005	29.384267
30	33.734294	28.691844
31	33.900857	29.374749

7.4 Discussion

In the present study, a coarse resolution of 10x10km was chosen in order to minimize the number of clusters and resultant proposed gauges, in an effort to minimize the required budget. Moreover, the intensity of the rainfall at the study site did not vary greatly, with the highest variation existing between the plain and hill regions. Two main techniques were used to establish new locations for rain gauges. The first involved k-means clustering, which was initially intended to be sufficient for the present study. However, the number of locations that resulted were insufficient for covering the complete spectrum of rainfall and elevation. Consequently, a second technique, the kriging of standard error, was performed, using the

gauges determined by k-means clustering as a foundation. The kriging of standard error was looped over 22 iterations and ultimately detected new locations for 22 gauges. The 31 gauges that resulted, in addition to the 5 existing gauges, were then tested by ECDF, with results revealing excellent coverage across the entire range of cumulative rainfall, elevation and for each separate precipitation event (from the years 2016, 2017, and 2018). However, limited vacant spaces existed for the 2015 event. These results were sufficiently reliable to be applied to the test site. Exact proposed locations in decimal degrees are provided above (Table 7.4). Taking into consideration that the future location validation is going to be achieved by k-means clustering of the new gauges records. Regarding the issue of maintenance, the proposed gauges furthest from existing settlements and roads could be excluded, keeping in mind that this could affect the coverage quality of the proposed rain gauge network.

8. CONCLUSION AND RECOMMENDATIONS

8.1 Conclusion

Arid and semi-arid zones worldwide are currently facing growing water shortages. This is partly due to their reliance on groundwater which, when combined with insufficient water distribution data due to poorly distributed rain gauges, leads to excessive water withdrawal by expanding populations. The agricultural and industrial activity then required by these populations leads to water table depletion, destructively influencing the continuity of these communities. Consequently, the presented study focused on the determination of suitable steps for overcoming the issue of data scarcity in this context. These steps were consistent with those proposed by most hydrological studies focused on topics such as groundwater localization, spatiotemporal water distribution, soil erosion minimization, and mitigation of flash flood hazards.

The presented study chose the eastern side of the Gulf of Suez, one of the (semi-)arid regions of Egypt, as a test site. This site was chosen due to its growing tourism and agriculture industries, which are increasingly impacting the environment. This area also features varied climate conditions and complex topology, which makes it a suitable region for studying the factors that influence water cycle equilibrium. The identification of these factors relied on data collected from 1934-1989 and published in 2008 by Sherief. Precipitation was chosen by the author to be the main focus of the presented study, as it represents the key parameter influencing water cycle equilibrium in arid regions and is the main resource responsible for recharging aquifers chronically depleted by excessive consumption.

Sherief (2008) divided rainfall intensities into, light, moderate and heavy precipitation events, with light indicating 0.1 to 1mm of precipitation, moderate reflecting 1 to 10mm, and greater than 10mm denoting heavy events. Light intensity events were most frequent, accounting for 61% of the total portion of precipitation events. Heavy events represented the lowest frequency, at 5% of the total. Moderate events represent 34% of the total number of events. The Southern and the Middle regions of the test site showed the highest precipitation rate, according to data recorded at Sharm El-Sheikh and Saint-Catherine stations. The lowest rates were seen in the Northern region, according to data collected by Ras-Sudr and Abu-Rudies stations. The majority of precipitation events at the test site took place between December and January.

Following the rough identification of rainfall intensities, the author analyzed how changes in land cover and land use were affecting ground water consumption rates. This was done to determine whether groundwater reserves would still be capable of meeting increasing water demands at the test site, independent of rainfall. This was achieved by use of Landsat5 and 8 data collected from the years 2000, 2010, and 2019. As settlements and vegetation covered only a small proportion of the test site, it was not possible to analyze the entire area. Instead, three locations displaying clear changes were selected. These were spots of the cities of Ras-Sudr, El-Tor, and Sharm El-Sheikh. In Ras-Sudr, there was a very slight increase in vegetation and settled areas during the period of 2000 to 2010. From 2010 to 2019, however, extensive increases could be seen. El-Tor also showed a slight increase in the abovementioned areas during the period of 2000 to 2010, and similarly increased cover significantly from 2010 to 2019. In contrast, the city of Sharm El-Sheikh saw a noticeable increase in vegetated areas, buildings, and streets from 2000 to 2010 and a decrease from 2010 to 2019. The results of this chapter indicated that water consumption was not uniformly increasing. In some cases, it decreased or remained stable, such as in Sharm El-Sheikh from 2010 to 2019. Moreover, it was clear that the groundwater derived from the Quaternary aquifer was and is still capable of covering the increasing water demands of the cities of Ras-Sudr and El-Tor.

The author then turned to data derived by remote sensors in order to further analyze the spatiotemporal distribution of rainfall. Remote sensing was chosen as it provided data that was higher in both spatial and temporal resolution. Two types of remote sensing data collection were used, TRMM(3B42V7) and GPM(IMERG). These data from four rainy events between 2015 and 2018. However, the collected two data types produced highly different and weakly correlated values. The values of 0.1°-resolution pixels were measured at successive times (0h, 3h, 6h, 9h, 12h, 24h) for both the TRMM(3B42V7) and GPM(IMERG) and subsequently compared in Excel via box plot graphs. The greatest difference in values appeared for the October 27, 2016 precipitation event. Moreover, the boxplots for all events grew increasingly disparate with the progression of time (from 9h to 12h to 24h). The lowest difference was seen in the Plain and Hill for the June 28, 2018 event between onset and the third hour. The data sets derived from the two remote sensors displayed a weak correlation for the years 2015, 2016, and 2017. However, a strong correlation was indeed found for the 2018 dataset. Higher correlation was generally seen when the duration of the precipitation events was lower, with data collected at the onset, at the third hour, and at the sixth hour. Most cases, however, involved events with data collection at 9h, 12h, and 24h and featured insignificant r values and negative correlations.

The author sought to compare the performance of the two mentioned data sets. GPM(IMERG) provided the strongest correlation with the physical rain gauges, while TRMM (0.25°) provided the weakest. For this reason, these measures were followed by the RMSE, BIAS, and MAE metrics. The GPM(IMERG) data performed best, with the lowest RMSE and MAE median and maximum values. The BIAS median was also near 0% for this data type, compared to the TRMM 0.25° and TRMM 0.1° data. A third category of statistical tests applied three different thresholds to FAR, POD, and CSI analysis. Taken together, the results of these thresholded tests confirmed the superior performance of the GPM (IMERG). Moreover, both data sets displayed their best performance when calculated using the lowest threshold and worst when using the highest threshold. This indicates that both data sets captured low intensity events with greater accuracy than high intensity events, with uncertainty increasing proportionately with intensity. Overall, the GPM(IMERG) better captured both light and heavy intensity events.

These results reiterate the need for additional data on precipitation. Under current circumstances, with a low number of existing rain gauges and highly-priced and limited data available, acquiring additional data is difficult. The author therefore constructed a plan for a new rain gauge network design. To accomplish this, data from GPM(IMERG) (which performed best at the test site) along with 90x90 SRTM elevation data were used. These two datasets were clustered using k-means clustering to produce an elbow graph whose elbow-shaped region offered several possible options for the number of optimum clusters at the test site. The author chose three different cluster sizes (3, 6, and 9) and calculated the possible centroids for each size. Calculations resulted in three centroids, 6 centroids, and 9 centroids. These centroids were tested using the empirical cumulative distribution function (ECDF), once the sum of the GPM(IMERG) scenes, the scene limits, and the elevation map limits was determined. This test revealed gaps in all centroids mentioned, as each cluster covered such a big range of rainfall compared to it in elevation. Consequently, the author established nine clusters as the optimal size. Nine centroids were therefore taken, along with the existing five gauges, as a basis for standard error kriging. This allowed the author to locate the pixel with the highest error and add a gauge to that location, gradually minimizing error via looping. This procedure was repeated 22 times and the added points were tested with an ECDF. The complete spectrum of rainfall and elevation was efficiently covered by the proposed rain gauge locations, which included both the points from clustering and the five existing gauges.

Unfortunately, number of the presented gauges in the mountainous test site, located far from existing cities and settlements, where the proposed gauges would likely rapidly deteriorate

due to lack of maintenance. It might therefore be possible to exclude these gauges from the network, while keeping in mind the potential decrease in coverage quality.

8.2 Recommendations

Higher-resolution remote sensing data are recommended for capturing changes in land cover and land use in (semi-)arid regions. This will reduce sources of error, increase certainty, and provide additional details. As a result, more accurate estimations of ground water consumption rates will be possible.

Continuous measurements of water table levels are strongly suggested. These will allow for a greater understanding of groundwater consumption, which can then be explained to local farmers in an effort to recruit them as effective parties in the preservation of water resources.

As primarily TRMM data was used by the authors, the utilization of GPM(IMERG) is recommended for rainfall analysis at the test site in Sinai Peninsula.

The present study recommends the prompt installation of the rain gauge network mentioned in Chapter 7. This will allow for the rapid capture of the data needed for the evaluation of the success of the proposed procedures.

While remote sensing-derived data is highly valuable, it is important to verify results using ground truth data in order to refine rough data into accurate and reliable values.

REFERENCES

Abera, W., Brocca, L. and Rigon, R., 2016. Comparative evaluation of different satellite rainfall estimation products and bias correction in the Upper Blue Nile (UBN) basin. Atmospheric research, 178, pp.471-483.

Adhikary, S.K., Muttil, N. and Yilmaz, A.G., 2015. Genetic programming-based ordinary kriging for spatial interpolation of rainfall. Journal of Hydrologic Engineering, 21(2), p.04015062.

Adler, R.F., Huffman, G.J., Chang, A., Ferraro, R., Xie, P.P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D. and Gruber, A., 2003. The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). Journal of hydrometeorology, 4(6), pp.1147-1167.

Afify, H.A., 2011. Evaluation of change detection techniques for monitoring land-cover changes: a case study in new Burg El-Arab area. Alexandria Engineering Journal, 50(2), pp.187-195.

Ahmed, M., Sauck, W., Sultan, M., Yan, E., Soliman, F. and Rashed, M., 2014. Geophysical constraints on the hydrogeologic and structural settings of the Gulf of Suez rift-related basins: case study from the El Qaa Plain, Sinai, Egypt. Surveys in Geophysics, 35(2), pp.415-430.

Alemohammad, S.H., Entekhabi, D. and McLaughlin, D.B., 2014. Evaluation of long-term SSM/I-based precipitation records over land. Journal of Hydrometeorology, 15(5), pp.2012-2029.

Alexakis, D.D. and Tsanis, I.K., 2016. Comparison of multiple linear regression and artificial neural network models for downscaling TRMM precipitation products using MODIS data. Environmental Earth Sciences, 75(14), p.1077.

Ashouri, H., Hsu, K.L., Sorooshian, S., Braithwaite, D.K., Knapp, K.R., Cecil, L.D., Nelson, B.R. and Prat, O.P., 2015. PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. Bulletin of the American Meteorological Society, 96(1), pp.69-83.

Attia MA (1930) Report on fresh water of El-Tor Quarry. Geological Survey of Egypt, Cairo.

Azab, A.A. and El-Khadragy, A.A., 2013. 2.5-D gravity/magnetic model studies in Sahl El Qaa area, southwestern Sinai, Egypt. Pure and Applied Geophysics, 170(12), pp.2207-2229.

Azizi, M.J., Seifi, F. and Moghadam, S., 2019. A robust simulation optimization algorithm using kriging and particle swarm optimization: Application to surgery room optimization. Communications in Statistics-Simulation and Computation, pp.1-17.

Bachmaier, M. and Backes, M., 2008. Variogram or semivariogram? Understanding the variances in a variogram. Precision Agriculture, 9(3), pp.173-175.

Back, L.E. and Bretherton, C.S., 2005. The relationship between wind speed and precipitation in the Pacific ITCZ. Journal of climate, 18(20), pp.4317-4328.

Bear, J., 1979. Hydraulics of groundwater New York. Mc GrawHill Inc.

Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A.I., Weedon, G.P., Brocca, L., Pappenberger, F., Huffman, G.J. and Wood, E.F., 2017. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. Hydrology and Earth System Sciences, 21(12), pp.6201-6217.

Be'eri-Shlevin, Y., Katzir, Y. and Whitehouse, M., 2009. Post-collisional tectonomagmatic evolution in the northern Arabian–Nubian Shield: time constraints from ion-probe U–Pb dating of zircon. Journal of the Geological Society, 166(1), pp.71-85.

Berg, W., Sapiano, M.R., Horsman, J. and Kummerow, C., 2012. Improved geolocation and Earth incidence angle information for a fundamental climate data record of the SSM/I sensors. IEEE Transactions on Geoscience and Remote Sensing, 51(3), pp.1504-1513.

Beven, K., 2004. Robert E. Horton's perceptual model of infiltration processes. Hydrological processes, 18(17), pp.3447-3460.

Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M. and Hijmans, R., 2017. Rouault E. rgdal: Bindings for the 'Geospatial'data Abstraction Library. R package version 0.9-1< http://cran. r-project. org/web/packages/rgdal.

Blasband, B., White, S., Brooijmans, P., De Boorder, H. and Visser, W., 2000. Late Proterozoic extensional collapse in the Arabian–Nubian shield. Journal of the Geological Society, 157(3), pp.615-628.

Cao, Q., Hao, Z., Zhou, J., Wang, W., Yuan, F., Zhu, W. and Yu, C., 2019. Impacts of various types of El Niño–Southern Oscillation (ENSO) and ENSO Modoki on the rainy season over the Huaihe River basin. International Journal of Climatology, 39(5), pp.2811-2824.

Cecinati, F., Wani, O. and Rico-Ramirez, M.A., 2017. Comparing approaches to deal with non-gaussianity of rainfall data in kriging-based radar-gauge rainfall merging. Water Resources Research, 53(11), pp.8999-9018.

Chen, Y., Huang, J., Sheng, S., Mansaray, L.R., Liu, Z., Wu, H. and Wang, X., 2018. A new downscaling-integration framework for high-resolution monthly precipitation estimates: Combining rain gauge observations, satellite-derived precipitation data and geographical ancillary data. Remote Sensing of Environment, 214, pp.154-172.

Chorley, R.J., Beckinsale, R.P. and Dunn, A.J., 1973. The history of the study of landforms or the development of geomorphology: the life and work of william morris davis (Vol. 2). Psychology Press.

Clark, I., 2010. Statistics or geostatistics? Sampling error or nugget effect?. Journal of the Southern African Institute of Mining and Metallurgy, 110(6), pp.307-312.

Cochran, J.R. and Karner, G.D., 2007. Constraints on the deformation and rupturing of continental lithosphere of the Red Sea: the transition from rifting to drifting. Geological Society, London, Special Publications, 282(1), pp.265-289.

Coppin, P.R. and Bauer, M.E., 1996. Digital change detection in forest ecosystems with remote sensing imagery. Remote sensing reviews, 13(3-4), pp.207-234.

Crow, W.T., Berg, A.A., Cosh, M.H., Loew, A., Mohanty, B.P., Panciera, R., de Rosnay, P., Ryu, D. and Walker, J.P., 2012. Upscaling sparse ground-based soil moisture observations for

the validation of coarse-resolution satellite soil moisture products. Reviews of Geophysics, 50(2).

Das, R.K. and Prakash, N.R., 2011. Design of an improvised tipping bucket rain gauge for measurement of rain and snow precipitation. International Journal of Instrumentation Technology, 1(1), pp.44-59.

Davarzani, H., Smits, K., Tolene, R.M. and Illangasekare, T., 2014. Study of the effect of wind speed on evaporation from soil through integrated modeling of the atmospheric boundary layer and shallow subsurface. Water resources research, 50(1), pp.661-680.

De Beurs, K.M. and Henebry, G.M., 2004. Land surface phenology, climatic variation, and institutional change: Analyzing agricultural land cover change in Kazakhstan. Remote Sensing of Environment, 89(4), pp.497-509.

Dregne, H., 1991. A new assessment of the world status of desertification. Desertification Control Bull, 20, pp.6-29.

Ebert, E.E., Janowiak, J.E. and Kidd, C., 2007. Comparison of near-real-time precipitation estimates from satellite observations and numerical models. Bulletin of the American Meteorological Society, 88(1), pp.47-64.

Eissa, M.A., Mahmoud, H.H., Shouakar-Stash, O., El-Shiekh, A. and Parker, B., 2016. Geophysical and geochemical studies to delineate seawater intrusion in Bagoush area, Northwestern coast, Egypt. Journal of African Earth Sciences, 121, pp.365-381.

El Tahlawi, M.R., 2014. Sinai Peninsula: an overview of geology and thermal groundwater potentialities. In Thermal and Mineral Waters (pp. 25-38). Springer, Berlin, Heidelberg.

Elavarasi, S.A., Akilandeswari, J. and Sathiyabhama, B., 2011. A survey on partition clustering algorithms. International Journal of Enterprise Computing and Business Systems, 1(1).

El-Masry, N.N., El-Kaliouby, B.A., Khawasik, S.M., and El-Ghawaby, M.A., 1992. Reconsideration of the geologic evolution of Saint Catherine ring dyke, South Sinai. Con. Geol. Sinai Develop., Ismailia,, pp 229-238.

EL-Refai A. A; 1992. Water resources of Southern Sinai, Egypt. (Geomorphological and Hydrogeological studies). Ph.D. Thesis, Fac. Sci. Cairo. Univ., p. 357

Estoque, R.C. and Murayama, Y., 2015. Classification and change detection of built-up lands from Landsat-7 ETM+ and Landsat-8 OLI/TIRS imageries: A comparative assessment of various spectral indices. Ecological Indicators, 56, pp.205-217.

Ezzahar, J., Chehbouni, A., Hoedjes, J.C., Er-Raki, S., Chehbouni, A., Boulet, G., Bonnefond, J.M. and De Bruin, H.A.R., 2007. The use of the scintillation technique for monitoring seasonal water consumption of olive orchards in a semi-arid region. Agricultural Water Management, 89(3), pp.173-184.

Fensterseifer, C., Allasia, D.G. and Paz, A.R., 2016. Assessment of the TRMM 3B42 precipitation product in southern Brazil. JAWRA Journal of the American Water Resources Association, 52(2), pp.367-375.

Forkel, M., Carvalhais, N., Verbesselt, J., Mahecha, M., Neigh, C. and Reichstein, M., 2013. Trend change detection in NDVI time series: Effects of inter-annual variability and methodology. Remote Sensing, 5(5), pp.2113-2144.

Franklin, J., Phinn, S.R., Woodcock, C.E., Rogan, J., 2003. Rationale and conceptual framework for classification approaches to assess forest resources and properties. In: Wulder, M., Franklin, S.E. (Eds.), methods and applications for remote sensing of forests: concepts and case studies, Kluwer Academic Publishers, Dordrecht, pp. 279–300.

Gandhi, G.M., Parthiban, S., Thummalu, N. and Christy, A., 2015. NDVI: vegetation change detection using remote sensing and GIS–a case study of Vellore District. Procedia Computer Science, 57, pp.1199-1210.

Ghodeif, K. and Gorski, J., 2001. Protection of fresh ground water in El-Qaa Quaternary aquifer, Sinai, Egypt. In New approaches characterizing groundwater flow (pp. 71-76).

Gilboa, Y., 1972. Survey of water resources at Al-Tor. Water Planning for Israel Ltd., Tel Aviv.

GIS Geography Kriging Interpolation–The Prediction Is Strong in this One. Published on February 4, 2017, Last Updated 7 Jan 2020. Retrieved from https://gisgeography.com/kriging-interpolation-prediction/

Gorski, j. and Ghodeif, K., 2000. Salinization of shallow water aquifer in El-Qaa coastal plain, SI.

Green, K., Kempka, D., and Lackey, L., 1994. Using remote sensing to detect and monitor land-cover and land-use change. Photogrammetric Engineering & Remote Sensing, 60(3), pp. 331-337.

Hammerla, N.Y., Kirkham, R., Andras, P. and Ploetz, T., 2013, September. On preserving statistical characteristics of accelerometry data using their empirical cumulative distribution. In Proceedings of the 2013 international symposium on wearable computers (pp. 65-68).

Han, S.C., Yeo, I.Y., Alsdorf, D., Bates, P., Boy, J.P., Kim, H., Oki, T. and Rodell, M., 2010. Movement of Amazon surface water from time-variable satellite gravity measurements and implications for water cycle parameters in land surface models. Geochemistry, Geophysics, Geosystems, 11(9).

Hansen, J., Russell, G., Rind, D., Stone, P., Lacis, A., Lebedeff, S., Ruedy, R. and Travis, L., 1983. Efficient three-dimensional global models for climate studies: Models I and II. Monthly Weather Review, 111(4), pp.609-662.

Hartmann, H.C., Pagano, T.C., Sorooshian, S. and Bales, R., 2002. Confidence builders: Evaluating seasonal climate forecasts from user perspectives. Bulletin of the American Meteorological Society, 83(5), pp.683-698.

Heath, R.C., 1983. Basic ground-water hydrology: US geological survey water-supply paper 2220. US Geological Survey, Washington, DC.

Hijmans, R.J., van Etten, J., Mattiuzzi, M., Sumner, M., Greenberg, J.A., Lamigueiro, O.P., Bevan, A., Racine, E.B. and Shortridge, A., 2013. Raster package in R.

Horton, R.E., 1941. An approach toward a physical interpretation of infiltration-capacity 1. Soil science society of America journal, 5(C), pp.399-417.

Hou, A.Y., Kakar, R.K., Neeck, S., Azarbarzin, A.A., Kummerow, C.D., Kojima, M., Oki, R., Nakamura, K. and Iguchi, T., 2014. The global precipitation measurement mission. Bulletin of the American Meteorological Society, 95(5), pp.701-722.

Huffman, G. and Bolvin, D., 2012. GPCP version 1.2 One-Degree Daily (1DD) precipitation dataset documentation. WDC-A, NCDC, Asheville.

Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Wolff, D.B., Adler, R.F., Gu, G., Hong, Y., Bowman, K.P. and Stocker, E.F., 2007. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. Journal of hydrometeorology, 8(1), pp.38-55.

İlsever, M. and Ünsalan, C., 2013, June. Locating the urban area in satellite images to detect changes in them. In 2013 6th International Conference on Recent Advances in Space Technologies (RAST) (pp. 109-114). IEEE.

Immerzeel, W.W., Rutten, M.M. and Droogers, P., 2009. Spatial downscaling of TRMM precipitation using vegetative response on the Iberian Peninsula. Remote Sensing of Environment, 113(2), pp.362-370.

Inaba, M.; Katoh, N. & Imai H. (1994). Applications of weighted voronoi diagrams and randomization to variance-based k-clustering. In: Proceedings of the 10th ACM Symposium on Computational Geometry, 332-339.

Jain, A.K., Murty, M.N. and Flynn, P.J., 1999. Data clustering: a review. ACM computing surveys (CSUR), 31(3), pp.264-323.

Jeniffer, K., Su, Z., Woldai, T. and Maathuis, B., 2010. Estimation of spatial–temporal rainfall distribution using remote sensing techniques: A case study of Makanya catchment, Tanzania. International journal of applied earth observation and geoinformation, 12, pp.S90-S99.

Jianya, G., Haigang, S., Guorui, M. and Qiming, Z., 2008. A review of multi-temporal remote sensing data change detection algorithms. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 37(B7), pp.757-762.

Joyce, R.J., Janowiak, J.E., Arkin, P.A. and Xie, P., 2004. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. Journal of hydrometeorology, 5(3), pp.487-503.

Kalma, J.D., McVicar, T.R. and McCabe, M.F., 2008. Estimating land surface evaporation: A review of methods using remotely sensed surface temperature data. Surveys in Geophysics, 29(4-5), pp.421-469.

Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R. and Wu, A.Y., 2002. An efficient k-means clustering algorithm: Analysis and implementation. IEEE Transactions on Pattern Analysis & Machine Intelligence, (7), pp.881-892.

Kassim, A.H.M. and Kottegoda, N.T., 1991. Rainfall network design through comparative kriging methods. Hydrological Sciences Journal, 36(3), pp.223-240.

Kennedy, R.E., Townsend, P.A., Gross, J.E., Cohen, W.B., Bolstad, P., Wang, Y.Q. and Adams, P., 2009. Remote sensing change detection tools for natural resource managers: Understanding

concepts and tradeoffs in the design of landscape monitoring projects. Remote sensing of environment, 113(7), pp.1382-1396.

Khan, A., Chatterjee, S., Bisai, D. and Barman, N.K., 2014. Analysis of change point in surface temperature time series using cumulative sum chart and bootstrapping for Asansol Weather Observation Station, West Bengal, India. American Journal of Climate Change, 3(01), p.83.

Kim, K., Park, J., Baik, J. and Choi, M., 2017. Evaluation of topographical and seasonal feature using GPM IMERG and TRMM 3B42 over Far-East Asia. Atmospheric research, 187, pp.95-105.

Kirkby, M.J., 1978. The stream head as a significant geomorphic threshold. School of Geography, University of Leeds.

Kiros, G., Shetty, A. and Nandagiri, L., 2016. Analysis of variability and trends in rainfall over northern Ethiopia. Arabian Journal of Geosciences, 9(6), p.451.

Kodinariya, T.M. and Makwana, P.R., 2013. Review on determining number of Cluster in K-Means Clustering. International Journal, 1(6), pp.90-95.

Kotsiantis, S.B., Zaharakis, I. and Pintelas, P., 2007. Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, 160, pp.3-24.

Kuchment, L.S., 2004. The hydrological cycle and human impact on it. Water Resources Management.

Kummerow, C., Barnes, W., Kozu, T., Shiue, J. and Simpson, J., 1998. The tropical rainfall measuring mission (TRMM) sensor package. Journal of atmospheric and oceanic technology, 15(3), pp.809-817.

Kustas, W.P. and Norman, J.M., 1996. Use of remote sensing for evapotranspiration monitoring over land surfaces. Hydrological Sciences Journal, 41(4), pp.495-516.

Lahiri, S.N., Kaiser, M.S., Cressie, N. and Hsu, N.J., 1999. Prediction of spatial cumulative distribution functions using subsampling. Journal of the American Statistical Association, 94(445), pp.86-97.

Li, C., Sinha, E., Horton, D.E., Diffenbaugh, N.S. and Michalak, A.M., 2014. Joint bias correction of temperature and precipitation in climate model simulations. Journal of Geophysical Research: Atmospheres, 119(23), pp.13-153.

Li, Y. and Wu, H., 2012. A clustering method based on K-means algorithm. Physics Procedia, 25, pp.1104-1109.

Li, Z., Chen, J. and Baltsavias, E. eds., 2008. Advances in photogrammetry, remote sensing and spatial information sciences: 2008 ISPRS congress book (Vol. 7). CRC Press.

Libertino, A., Sharma, A., Lakshmi, V. and Claps, P., 2016. A global assessment of the timing of extreme rainfall from TRMM and GPM for improving hydrologic design. Environmental Research Letters, 11(5), p.054003.

Likas, A., Vlassis, N. and Verbeek, J.J., 2003. The global k-means clustering algorithm. Pattern recognition, 36(2), pp.451-461.

Lillesand, R.W., and Kiefer, R.W., 1994. Remote sensing and image interpretation. Wiley, New York, (4th ed.).

Liu, Y.Y., Dorigo, W.A., Parinussa, R.M., de Jeu, R.A., Wagner, W., McCabe, M.F., Evans, J.P. and Van Dijk, A.I.J.M., 2012. Trend-preserving blending of passive and active microwave soil moisture retrievals. Remote Sensing of Environment, 123, pp.280-297.

Mahmoud, S., Reilinger, R., McClusky, S., Vernant, P. and Tealeb, A., 2005. GPS evidence for northward motion of the Sinai Block: implications for E. Mediterranean tectonics. Earth and Planetary Science Letters, 238(1-2), pp.217-224.

Malagnoux, M., 2007, April. Arid land forests of the world: global environmental perspectives. In International Conference on Afforestation and Sustainable Forests as a Means to Combat Desertification, Jerusalem, Israel (pp. 16-19).

Mann, A.K. and Kaur, N., 2013. Review paper on clustering techniques. Global Journal of Computer Science and Technology.

Massoud, U., Santos, F., Qady, G. El, Atya, M., & Soliman, M., 2010. Identification of the shallow subsurface succession and investigation of the seawater invasion to the Quaternary aquifer at the northern part of El Qaa plain , Southern Sinai , Egypt by transient electromagnetic data, pp.267–277.

McClay, K.R., Nichols, G.J., Khalil, S.M., Darwish, M. and Bosworth, W., 1998. Extensional tectonics and sedimentation, eastern Gulf of Suez, Egypt. In Sedimentation and Tectonics in Rift Basins Red Sea: Gulf of Aden (pp. 223-238). Springer, Dordrecht.

Meert, J.G., 2003. A synopsis of events related to the assembly of eastern Gondwana. Tectonophysics, 362(1-4), pp.1-40.

Mu, W., Yu, F., Li, C., Xie, Y., Tian, J., Liu, J. and Zhao, N., 2015. Effects of rainfall intensity and slope gradient on runoff and soil moisture content on different growing stages of spring maize. Water, 7(6), pp.2990-3008.

National Center for Atmospheric Research Staff (Eds). Last modified 20 Sep 2018. "The Climate Data Guide: GPCC: Global Precipitation Climatology Centre." Retrieved from https://climatedataguide.ucar.edu/climate-data/gpcc-global-precipitation-climatology-centre.

Niu, G.Y., Yang, Z.L., Dickinson, R.E., Gulden, L.E. and Su, H., 2007. Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data. Journal of Geophysical Research: Atmospheres, 112(D7).

Noweir, A., EL Shishtawy A (1996) Structure setting and stratigraphy of the area east of El-Qaà Plain, southwestern Sinai, Egypt. J Geol, 40(1), pp.1-22.

Oliver, M.A. and Webster, R., 2014. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. Catena, 113, pp.56-69.

Oyelade, O.J., Oladipupo, O.O. and Obagbuwa, I.C., 2010. Application of k Means Clustering algorithm for prediction of Students Academic Performance. arXiv preprint arXiv:1002.2425.

Pagano, T. and Sorooshian, S., 2002. Hydrologic cycle. Encyclopedia of global environmental change. The earth system: Physical and Chemical Dimensions of Global Environmental Change, 1, pp.450-464.

Panagos, P., Borrelli, P., Meusburger, K., Yu, B., Klik, A., Lim, K.J., Yang, J.E., Ni, J., Miao, C., Chattopadhyay, N. and Sadeghi, S.H., 2017. Global rainfall erosivity assessment based on high-temporal resolution rainfall records. Scientific reports, 7(1), p.4175.

Paolini, L., Grings, F., Sobrino, J.A., Jiménez Muñoz, J.C. and Karszenbaum, H., 2006. Radiometric correction effects in Landsat multi-date/multi-sensor change detection studies. International Journal of Remote Sensing, 27(4), pp.685-704.

Parinussa, R., Wang, G., Liu, Y., Hagan, D., Lin, F., van der Schalie, R. and de Jeu, R., 2017. The Evaluation of Single-Sensor Surface Soil Moisture Anomalies over the Mainland of the People's Republic of China. Remote Sensing, 9(2), p.149.

Pilgrim, D. H., T. G. Chapman, and D. G. Doran. "Problems of rainfall-runoff modelling in arid and semiarid regions." Hydrological Sciences Journal 33, no. 4 (1988): 379-400.

Pitman, A.J., 2003. The evolution of, and revolution in, land surface schemes designed for climate models. International Journal of Climatology: A Journal of the Royal Meteorological Society, 23(5), pp.479-510.

Poméon, T., Diekkrüger, B. and Kumar, R., 2018. Computationally Efficient Multivariate Calibration and Validation of a Grid-Based Hydrologic Model in Sparsely Gauged West African River Basins. Water, 10(10), p.1418.

Prasad, M., 2002. Acoustic measurements in unconsolidated sands at low effective pressure and overpressure detection. Geophysics, 67(2), pp.405-412.

Robinson, T.P. and Metternicht, G., 2006. Testing the performance of spatial interpolation techniques for mapping soil properties. Computers and electronics in agriculture, 50(2), pp.97-108.

Rogan, J. and Chen, D., 2004. Remote sensing technology for mapping and monitoring land-cover and land-use change. Progress in planning, 61(4), pp.301-325.

Said, R., 1962. The Geology of Egypt, El-Sevier Publ. Co., Amsterdam, New York.

Sayed, M.A.A., El-Fakharany, M.A., and Hamed, M.F., 2004. Integrated geophysical and hydrogeological studies on the quaternary aquifer at the middle part of El-Qaa Plain, SW Sinai, Egypt. Journal of Egyptian Geophysical Society, 2(1), pp. 135-145.

Scanlon, B.R., Keese, K.E., Flint, A.L., Flint, L.E., Gaye, C.B., Edmunds, W.M. and Simmers, I., 2006. Global synthesis of groundwater recharge in semiarid and arid regions. Hydrological Processes: An International Journal, 20(15), pp.3335-3370.

Selim, S.S., Darwish, M. and Khadrah, A.A., 2016. Architecture and evolution of a tectonically-induced Middle Eocene clastic wedge on the southern Tethyan carbonate shelf, North Eastern Desert, Egypt. Proceedings of the Geologists' Association, 127(3), pp.377-390.

Sheffield, J., Wood, E.F., Chaney, N., Guan, K., Sadri, S., Yuan, X., Olang, L., Amani, A., Ali, A., Demuth, S. and Ogallo, L., 2014. A drought monitoring and forecasting system for sub-Sahara African water resources and food security. Bulletin of the American Meteorological Society, 95(6), pp.861-882.

Sheffield, J., Wood, E.F., Pan, M., Beck, H., Coccia, G., Serrat-Capdevila, A. and Verbist, K., 2018. Satellite Remote Sensing for Water Resources Management: Potential for Supporting

Sustainable Development in Data-Poor Regions. Water Resources Research, 54(12), pp.9724-9758.

Shen, Y. and Chen, Y., 2010. Global perspective on hydrology, water balance, and water resources management in arid basins. Hydrological Processes: An International Journal, 24(2), pp.129-135.

Sherief, Y., 2008. Flash floods and their effects on the development in El-Qaá plain area in south Sinai, Egypt, a study in applied geomorphology using GIS and remote sensing (Doctoral dissertation, PhD Dissertation, Mainz University, Germany).

Singh, A., 1989. Review article digital change detection techniques using remotely-sensed data. International journal of remote sensing, 10(6), pp.989-1003.

Siska, P.P., Goovaerts, P., Hung, I.K. and Bryant, V.M., 2005. Predicting ordinary kriging errors caused by surface roughness and dissectivity. Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group, 30(5), pp.601-612.

Sorooshian, S., Hsu, K.L., Gao, X., Gupta, H.V., Imam, B. and Braithwaite, D., 2000. Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. Bulletin of the American Meteorological Society, 81(9), pp.2035-2046.

Su, Z., 2002. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. Hydrology and earth system sciences, 6(1), pp.85-100.

Sultan, M., Metwally, S., Milewski, A., Becker, D., Ahmed, M., Sauck, W., Soliman, F., Sturchio, N., Yan, E., Rashed, M. and Wagdy, A., 2011. Modern recharge to fossil aquifers: Geochemical, geophysical, and modeling constraints. Journal of hydrology, 403(1-2), pp.14-24.

Sultan, M., Yan, E., Sturchio, N., Wagdy, A., Gelil, K.A., Becker, R., Manocha, N. and Milewski, A., 2007. Natural discharge: a key to sustainable utilization of fossil groundwater. Journal of Hydrology, 335(1-2), pp.25-36.

Sultan, S.A., Mekhemer, H.M., Santos, F.A.M., and Abd Alla, M., 2007. Geophysical measurements for subsurface mapping and groundwater exploration at the central part of Sinai Peninsula, Egypt. The Arabian Journal for Science and Engineering, 34, pp. 104-119.

Sultan, S.A., Mohameden, M.I. and Santos, F.M., 2009. Hydrogeophysical study of the El Qaa Plain, Sinai, Egypt. Bulletin of engineering geology and the environment, 68(4), p.525.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S. and Hsu, K.L., 2018. A review of global precipitation data sets: Data sources, estimation, and intercomparisons. Reviews of Geophysics, 56(1), pp.79-107.

Tang, G., Zeng, Z., Long, D., Guo, X., Yong, B., Zhang, W. and Hong, Y., 2016. Statistical and hydrological comparisons between TRMM and GPM level-3 products over a midlatitude basin: Is day-1 IMERG a good successor for TMPA 3B42V7?. Journal of Hydrometeorology, 17(1), pp.121-137.

Tapiador, F.J., Turk, F.J., Petersen, W., Hou, A.Y., García-Ortega, E., Machado, L.A., Angelis, C.F., Salio, P., Kidd, C., Huffman, G.J. and De Castro, M., 2012. Global precipitation measurement: Methods, datasets and applications. Atmospheric Research, 104, pp.70-97.

Théau, J., 2006, August. Detection of changes using remote sensing: an overview of principles and applications. In Geo-Spatial and Range Sciences Conference, online1.

Thornes, J.B., 2009. Catchment and channel hydrology. In Geomorphology of desert environments (pp. 303-332). Springer, Dordrecht.

Tompkins, A.M. and Adebiyi, A.A., 2012, April. Using CloudSat cloud retrievals to assess differences between satellite rainfall products over Africa. In EGU General Assembly Conference Abstracts (Vol. 14, p. 12508).

UNESCO (1979) Map of the world distribution of arid regions. MAB Tech. Notes 7, UNESCO, Paris.

Ushio, T., Sasashige, K., Kubota, T., Shige, S., Okamoto, K.I., Aonashi, K., Inoue, T., Takahashi, N., Iguchi, T., Kachi, M. and Oki, R., 2009. A Kalman filter approach to the Global Satellite Mapping of Precipitation (GSMaP) from combined passive microwave and infrared radiometric data. Journal of the Meteorological Society of Japan. Ser. II, 87, pp.137-151.

Varshney, A., 2013. Improved NDBI differencing algorithm for built-up regions change detection from remote-sensing data: an automated approach. Remote sensing letters, 4(5), pp.504-512.

Vinukollu, R.K., Meynadier, R., Sheffield, J. and Wood, E.F., 2011. Multi-model, multi-sensor estimates of global evapotranspiration: Climatology, uncertainties and trends. Hydrological Processes, 25(26), pp.3993-4010.

Virdee, T.S. and Kottegoda, N. T. (2009). A brief review of kriging and its application to optimal interpolation and observation well selection. Hydrological Sciences Journal. Taylor & Francis, London, England.

Wahid, A., Madden, M., Khalaf, F. and Fathy, I., 2016. Geospatial analysis for the determination of hydro-morphological characteristics and assessment of flash flood potentiality in arid coastal plains: a case in southwestern Sinai, Egypt. Earth Sciences Research Journal, 20(1), pp.1-9.

Weber, K.T., 2001. A method to incorporate phenology into land cover change analysis. J. Range Manage, 54, pp. A1-A7.

Xie, P., Chen, M. and Shi, W., 2010, January. CPC unified gauge-based analysis of global daily precipitation. In Preprints, 24th Conf. on Hydrology, Atlanta, GA, Amer. Meteor. Soc (Vol. 2).

Xie, P., Janowiak, J.E., Arkin, P.A., Adler, R., Gruber, A., Ferraro, R., Huffman, G.J. and Curtis, S., 2003. GPCP pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates. Journal of Climate, 16(13), pp.2197-2214.

Yang, S., Weng, F., Yan, B., Sun, N. and Goldberg, M., 2011. Special Sensor Microwave Imager (SSM/I) intersensor calibration using a simultaneous conical overpass technique. Journal of Applied Meteorology and Climatology, 50(1), pp.77-95.

Yin, L., Zhou, Y., Ge, S., Wen, D., Zhang, E. and Dong, J., 2013. Comparison and modification of methods for estimating evapotranspiration using diurnal groundwater level fluctuations in arid and semiarid regions. Journal of hydrology, 496, pp.9-16.

Zeweldi, D.A. and Gebremichael, M., 2009. Evaluation of CMORPH precipitation products at fine space–time scales. Journal of Hydrometeorology, 10(1), pp.300-307.

Zhao, X., Zhou, Y., Min, J., Wang, S., Shi, W. and Xing, G., 2012. Nitrogen runoff dominates water nitrogen pollution from rice-wheat rotation in the Taihu Lake region of China. Agriculture, ecosystems & environment, 156, pp.1-11.

Zhu, Z., 2017. Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. ISPRS Journal of Photogrammetry and Remote Sensing, 130, pp.370-384.