CHAPTER ONE
INTRODUCTION

1.1 PROBLEM STATEMENT

Reliable monitoring of land cover and its transformation is an important component of environmental and natural resources management. Land cover is defined as the physical composition of material on the surface of the Earth, while land use is a description of how the land is used for socio-economic reasons [1]. Land cover is distinctly different from land use, but these two terms will be used interchangeably, as the focus of this thesis is the detection of land cover transformation of natural vegetation to newly formed human settlements. Several studies have investigated the global effects of anthropogenic activities on the planet, and it is estimated that more than a third of the Earth's land surface has been transformed by human activities [2]. The increase in human population is one of the major drivers of settlement expansion within geographical areas, which further increases the utilisation of the remaining natural resources [3]. Geographic information on land use and land cover change is highly sought after at local and global scales.

Land cover change often indicates land use change with major socio-economic impacts, while the transformation of vegetation cover (e.g. deforestation, agricultural expansion, urbanisation) has significant impacts on hydrology, ecosystems and climate [4, 5]. All these changes affect the environment and have a detrimental impact on the habitat of the human race. This raises the question whether the human's demand for natural resources is sustainable.

Sustainability is the long-term maintenance plan that will ensure the future of mankind's endeavours. The most widely quoted definition of sustainability and sustainable development was stated by the Brundtland Commission of the United Nations (UN) on March 20, 1987 as [6]:

> *Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs.*

The well-being of the environment is one of the major factors that contributes to sustainability. The UN General Assembly's discussion on sustainable human settlements concluded that countries' local governments need to plan, implement, develop, and manage human settlements [7]. It was further stated that the local government needs to manage existing settlements and prevent the establishment of any new unplanned settlements. The ability to determine where new settlements are formed, creates opportunities for the local government to provide running water supplies, sewage- and refuse removal services, which ties in directly with the UN's Millennium Development goals. The UN proposes a systematic development of sustainable cities for newly formed settlements. The South African government incorporated this vision into its local policies by focusing on service delivery to these newly formed settlements. Human settlement expansion is currently the most pervasive form of land cover change in South Africa [8]. Most of the new settlements are informal, unplanned and are usually built without the legal consent of the land owner [9, 10]. This thesis focuses on the detection of new human settlements formed in South Africa.

Satellite-based remote sensing is widely recognised by agencies, such as the United States Department of Agriculture (USDA)'s Farm Service Agency (FSA), the USDA's National Agricultural Statistics Services (NASS), and USDA's Foreign Agricultural Services (FAS), as a cost-effective method of acquiring information on the Earth's land surface [11]. Monitoring environmental dynamics, and classifying and detecting land cover change, require this type of cost-effective, systematic observations. The remote sensing science has thus progressed rapidly to meet the need to monitor global environmental change activities [12, 13]. Visually inspecting large volumes of high spatial resolution images for monitoring land cover is time-consuming and resource-intensive [14].

Earth observation satellites with wide swath widths provide the means of monitoring large areas on a frequent basis (high temporal resolution) [15]. These satellites are equipped with multiple coarse to medium spatial resolution sensors to record land surface information, in different spectral bands on a daily basis. Land cover surveillance of large geographical areas is augmented by the information inherent in the hyper-temporal satellite images, and therefore the analysis of these long-term data sets has attracted much attention [16, 17]. Owing to the complexity and non-parametric nature of land cover classification and change detection, machine learning methods are widely regarded as the most viable option for classification and change detection [14, 18]. The use of machine learning methods enables digital change detection, which encompasses the quantification of temporal phenomena from multi-date imagery that is most commonly acquired by satellite-based multi-spectral sensors [19].

Two types of land cover changes are usually investigated [20]: land cover modification and land cover transformation. Land cover modification is caused by internal changes within a particular land cover class. These changes affect the current state of the land cover class, but do not change the land

cover class, i.e. seasonal variation of natural vegetation. Land cover transformation of a particular geographical area involves change from one land cover class to another. This thesis focuses on land cover transformation of natural vegetation to newly formed human settlements, although the methods are applicable to other forms of land cover transformation. In the rest of this thesis the terms land cover transformation and land cover change are used interchangeably.

Change detection studies usually rely on image differencing, post-classification comparison methods, and change trajectory analysis [20–26], and the data are mostly treated as hyper-dimensional, but not necessarily as hyper-temporal. These methods therefore do not fully capitalise on the high temporal sampling rate which captures the dynamics of different land cover types. Satellites with high temporal acquisition rates provide information on the seasonal dynamics of a particular land cover type [15]. Incorporating the temporal information into a change detection algorithm allows the method to distinguish between land cover conversion and natural seasonal variations.

Main problem statement: *To detect land cover conversion of natural vegetation to newly formed human settlements reliably. The land cover change detection algorithm should incorporate temporal information to distinguish the change from seasonal variations. The land cover change detection algorithm should also be able to detect new human settlements that only span a small geographical area using coarse spatial resolution satellite imagery.*

1.2 OBJECTIVE OF THIS THESIS AND PROPOSED SOLUTION

Primary objective: *Develop a change detection algorithm that operates on multiple spectral bands, which exploits the richness of information inherent in hyper-temporal images*.

Secondary objective: *Develop a change detection algorithm that is sufficiently near automated, requiring minimal human interaction.*

As stated previously, machine learning methods are the more viable solution when analysing high dimensional data sets. A post-classification change detection approach detects change by classifying a geographical area into different classes over time. Land cover change is defined here as the transition in class label of a pixel's time series from one class to another class, after which it remains in the newly assigned class for the remainder of the time series [20]. A flow diagram for the proposed solution is shown in figure 1.1.

A set of images of a particular geographical area is obtained. The interval between two consecutive images must be short, which implies hyper-temporal acquisitions. The hyper-temporal images in this thesis were acquired by the MODerate resolution Imaging Spectroradiometer (MODIS) sensor

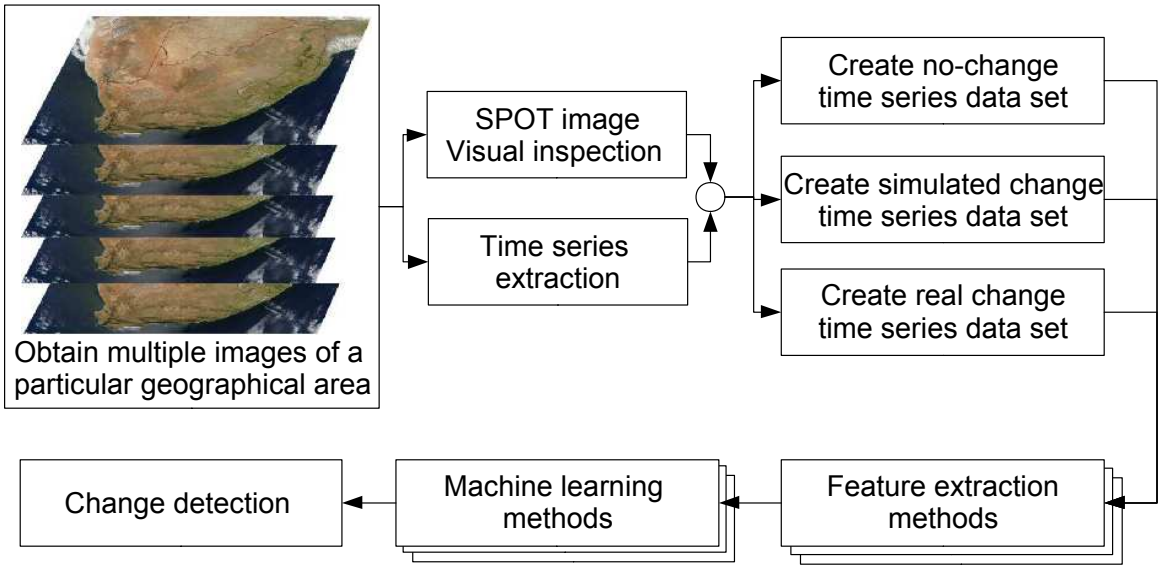


FIGURE 1.1: A flow diagram which depicts the steps followed to realise the proposed solution.

on board the Terra and Aqua satellites and are freely available. The MCD43A4 product provides hyper-temporal, multi-spectral (7 spectral bands) medium spatial resolution (500 metre) land surface reflectance data. The Bidirectional Reflectance Distribution Function (BRDF) correction models all the pixels in an image to a nadir view, which significantly reduces the anisotropic scattering effects of surfaces under different illumination and observation conditions [27, 28]. Time series of reflectance values were extracted for each spectral band over a particular geographical area (500 metre by 500 metre) from the multi-spectral hyper-temporal MODIS data set (February 2000 – January 2008).

Since the hyper-temporal images are coarse to medium spatial resolution, high spatial resolution satellite data are required for ground truth. Satellite Probatoire d'Obervation de la Terre (SPOT) images are high spatial resolution images, which are analysed by operators to identify areas that experienced land cover change or no land cover change.

Land cover change is a rare event on a regional scale and vital information, such as the date of change and rate of change, is usually not known. Therefore land cover change was simulated to enable a detailed assessment of change detection methods, which could not be performed on the real land cover change data set. A simulated land cover change time series data set is created by blending time series of two different land cover classes which did not change. The simulated land cover change data are used to test the functionality of the change detection methods, after which tests are performed on real examples of land cover change mapped using high spatial resolution images. Several contributions are made in this thesis that provide solutions to the primary and secondary objectives.

Contribution 1: *Develop of a novel land cover change detection method. The method is a post-classification approach and will operate on the Seasonal Fourier Features (SFF). SFF are*

hyper-temporal features extracted from time series.

The SFF are hyper-temporal features extracted without experiencing the usual pitfalls encountered with subsequence clustering [29]. The use of the SFF is then compared to another method proposed by Kleynhans *et al.* [30], referred to as the Extended Kalman Filter (EKF) feature extraction method. The drawback with this method is that it requires an offline optimisation phase, which must be performed by an operator. This does not satisfy the secondary objective (full automation) of this thesis, but has shown promising results.

Contribution 2: *Extend the EKF feature extraction method to a higher dimensions to improve change detection capabilities.*

The second objective concerned with full automation of the EKF extraction method is addressed in the following contribution.

Contribution 3: *Propose a novel criterion that is referred to as the Bias-Variance Equilibrium Point (BVEP). The BVEP is the point where the tracking of the reflectance values within time series are improved and the internal stability of the EKF is optimised. Define a Bias-Variance Score (BVS) that will measure the current system in relation to the BVEP.*

The BVEP criterion also provides statistical information on the seasonal vegetation activity cycle, which could provide vital insight into environmental dynamics [31, 32]. The optimisation of the BVS requires an unsupervised search method, which adjusts the variables to satisfy the BVEP criterion.

Contribution 4: *Design a new search algorithm, referred to as the Bias-Variance Search Algorithm (BVSA), that can effectively optimise the BVS to the BVEP criterion for optimal EKF performance.*

1.3 OUTLINE OF THESIS

The outline of the thesis is as follows:

- Chapter 2 gives a brief overview of the study area and an introduction to remote sensing principles. The chapter discusses several trade-offs that should be considered when selecting a sensor to solve the problem statement. The chapter concludes with an overview of some of the most common change detection methods found in the literature.

- Chapter 3 gives an introduction to supervised classification and in particular the Multilayer Perceptron (MLP). The chapter further discusses the pursuit of acceptable performance, and concludes with an overview on design considerations for a supervised classifier.
- Chapter 4 gives an introduction to unsupervised classification and provides several motivations for using an unsupervised classifier. The chapter also covers the disadvantages of unsupervised clustering and methods to mitigate them with proper cluster design.
- Chapter 5 defines four different feature extraction methods and their application to time series. These features are expected to provide good separation between natural vegetation and human settlement signals.
- Chapter 6 introduces the novel SFF and provides an in-depth investigation of the limitation of time series analysis mentioned by Keogh and Lin [29]. The chapter concludes with evidence of how the SFF provides a solution to this limitation.
- Chapter 7 introduces the BVEP, BVS, and Bias-Variance Search Algorithm (BVSA) used to optimise the EKF, in order to improve the quality of the extracted features.
- Chapter 8 presents the results of all experiments conducted in the thesis. These experiments report on classification accuracies, and change detection accuracies. These experiments are first conducted on a labelled data set within a particular province, and then expanded to run on a complete province, the Gauteng and Limpopo provinces of South Africa.
- Chapter 9 gives concluding remarks, as well as suggesting possible future research that could expand on the concepts introduced in this thesis.

CHAPTER TWO

REMOTE SENSING USED FOR LAND COVER CHANGE DETECTION

2.1 OVERVIEW

Remote sensing is the acquisition of information about an object without any direct contact with the object [33, Ch. 1]. Sensors are usually used to measure reflected wavelengths obtained from an object, which are then analysed for specific applications. A satellite-based sensor measures the reflected electromagnetic radiation of the Earth's surface and these measurements are then used to infer changes in surface reflectances caused by either environmental dynamics or anthropogenic activities.

Many international organisations and national governments have identified remote sensing as a beneficial field of study, and have made major joint investments in building better Earth observation systems. The objective of this chapter is to give the reader insight on how satellite-based sensors can be used to detect the formation of new human settlements on the Earth's surface.

2.2 SPONTANEOUS SETTLEMENTS

The standard of living in a country usually improves when sustainable economic growth is maintained. The government pursues a variety of projects to control the quality of economic growth [34]. Economical growth in developing countries is usually constrained by the lack of skilled labour, availability of resources, and necessary equipment. This lack of progress is aggravated by the pressure of a rapid growth in population and a backlog in housing development projects [9].

This backlog creates a shortage in the supply of affordable houses to the public, which results in the construction of temporary dwellings. These temporary dwellings are usually built without the legal consent of the land owner. The construction of temporary dwellings is not region-specific and has become a global phenomenon, although different characteristics are observed in the development of

these dwellings in each region [35]. A cluster of such temporary dwellings is formally known as a spontaneous settlement [9], which is a form of informal settlement [36, 37].

Social, economical, and political processes drive the migration of communities to certain regions, which often results in the development of informal settlement. This motivates the need for the local government to progressively track settlement expansion and migration [38, 39]. Settlement expansion is currently mapped on an irregular, ad hoc basis at great financial cost, using expensive visual interpretation of aerial photographs or satellite images. Regional information on settlement expansion gives the government the ability to plan the provision of services such as water, sanitation and electricity to these new or growing communities.

The behaviour of urban settlement migration and expansion has been empirically studied and predicted in various studies, but for several reasons cannot be applied to spontaneous settlements [9]. In this thesis no prior assumptions are made when attempting to find new or expanding settlements other than the decrease in seasonal behaviour associated with settlements.

Another motivation for tracking these spontaneous settlements is that their formation is currently one of the most pervasive forms of land cover change in South Africa [40]. The transformation of natural vegetation by practises such as deforestation, agricultural expansion and urbanisation has significant impacts on hydrology, ecosystems and the climate [4, 5, 41]. The area of interest in this thesis is the Limpopo province and Gauteng province located within South Africa.

2.2.1 Limpopo province

The Limpopo province is situated in the northern part of South Africa (Figure 2.1). The name of the province was derived from the river that separates South Africa from its neighbouring countries, Zimbabwe and Botswana. The province shares its southern borders with the Mpumalanga, Gauteng and North-West provinces.

The province is largely covered by natural vegetation, which is used for grazing by cattle and wildlife. It houses the largest hunting industry in South Africa. The province is also rich in numerous different tea and coffee plantations. The area is cultivated, with a range of agriculture focused on sunflowers, cotton, maize, peanuts, bananas, litchis, pineapples and mangoes.

The government departments within the province cannot currently capture and process all the necessary data on the different land cover types and anthropogenic activities throughout the province. This constraint is brought about by a limited budget, which motivates the pursuit of a less expensive alternative. Remote sensing (section 2.3) has been adopted by several governments as a less expensive option to augment the current processes of gathering information. If the government had access to more complete information, it could assist in the development of a management system to control and

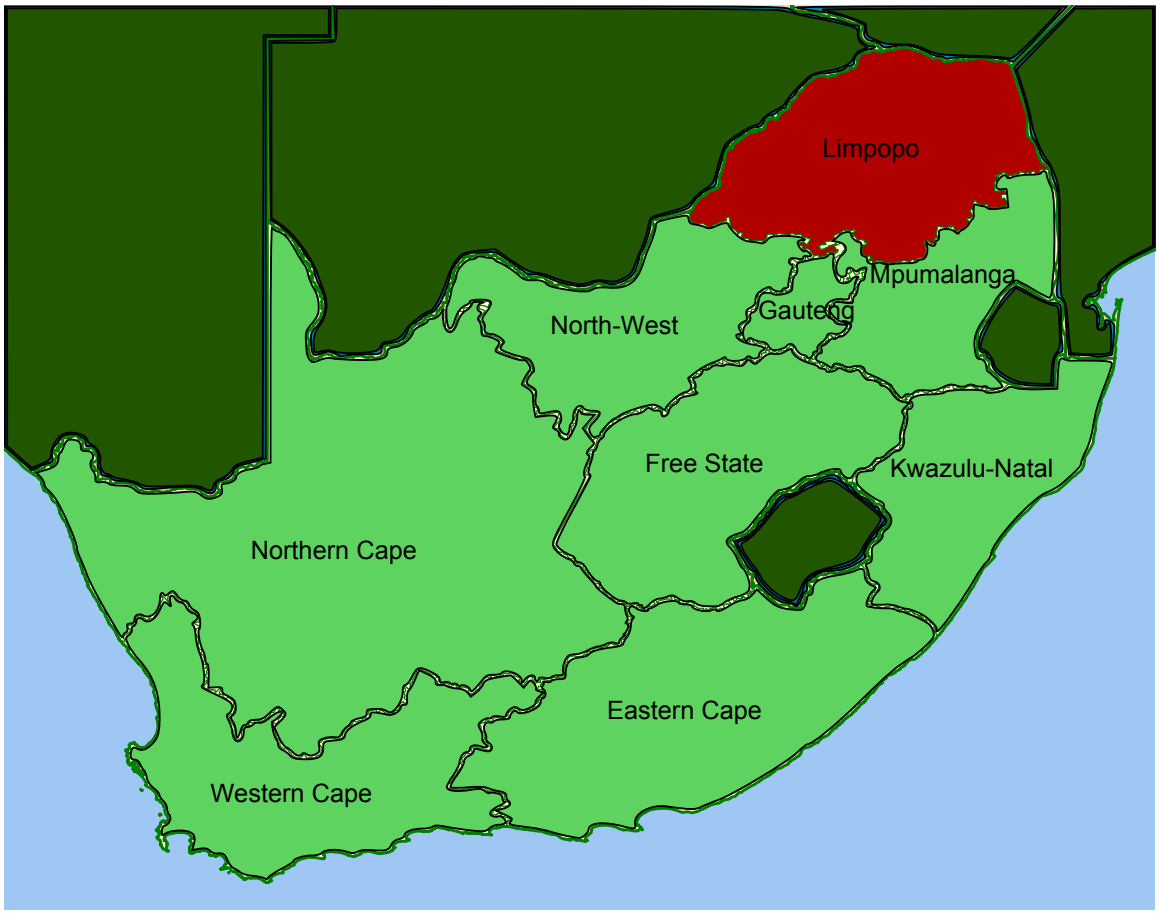


FIGURE 2.1: The Limpopo province is located in the northern part of South Africa.

monitor resources for the people throughout the province.

2.2.2 Gauteng province

The Gauteng province is situated in the highveld of South Africa (Figure 2.2). The name Gauteng comes from the Sesotho (indigenous language) word meaning *place of gold*. This is a common reference to the gold discovered in the city of Johannesburg in 1886. The province shares its borders with the Limpopo, Mpumalanga, North-West, and Free State provinces.

Gauteng is a landlocked province in the highveld, which is a high-altitude grassland. The province is the most urbanised one in the country. The province houses 20% of the country's population and only covers 1.4% of the country's total land area. A total population growth of over 30% was recorded between the years 2001 and 2010. Even though small in size, the province contributes 33.9% of South Africa's gross domestic product (GDP), which equates to 10% of the entire African continent.

In May 2008, the South African government identified problems caused by the massive influx of foreign nationals and provincial migration towards the Gauteng province. These problems range from

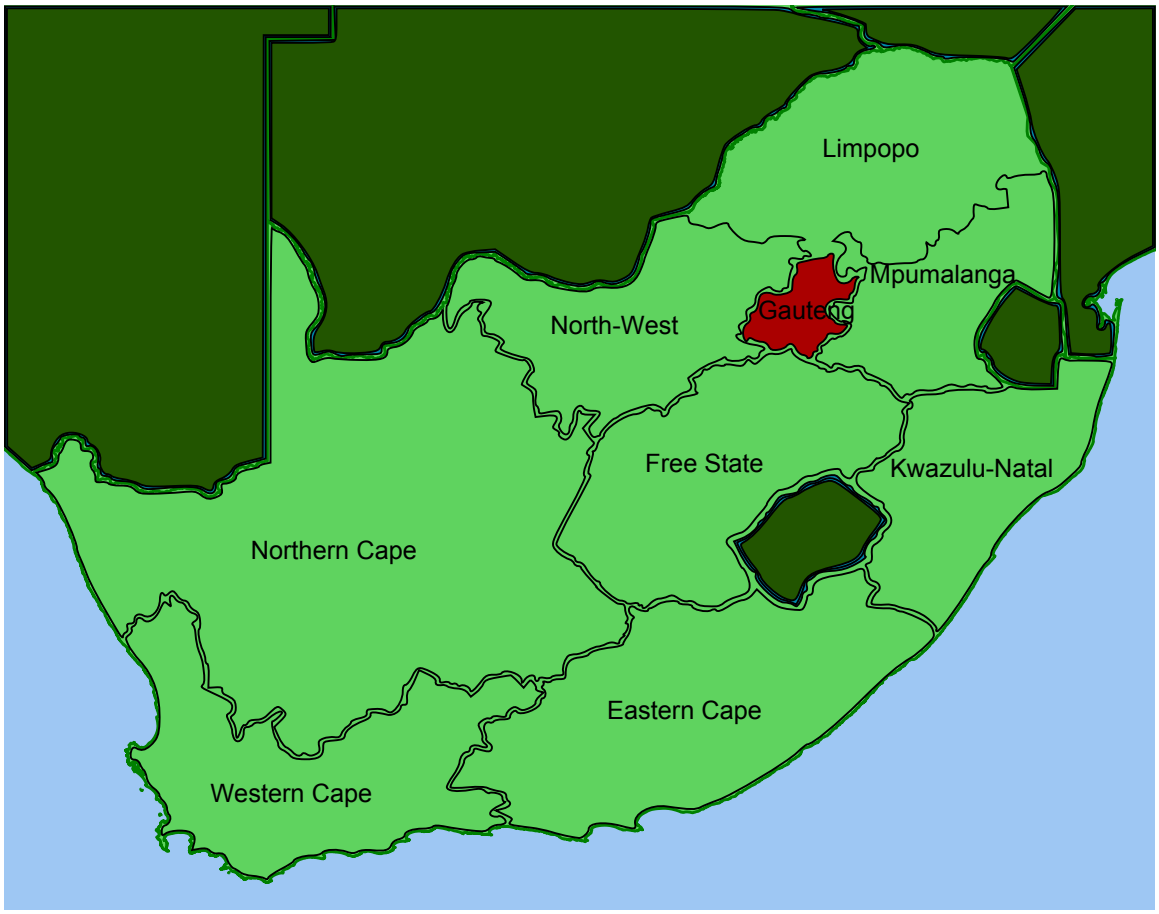


FIGURE 2.2: The Gauteng province is located in the highveld of South Africa.

social integration of multiple different cultures to proper service delivery. The active migration is motivated by a high median annual income for working adults and diverse employment opportunities. The province is rapidly growing to house cities that will be among the largest in the world. A projected population of 15 million people is expected by the year 2015.

2.3 OVERVIEW OF REMOTE SENSING

The Earth's surface is continually undergoing transformation caused by environmental change and anthropogenic activities. Many environmental problems stem from this continual transformation, of which some are; water shortage, soil degradation, greenhouse gas emissions, deforestation, biodiversity loss, etc. [33, Ch. 1].

The ability to evaluate the environmental dynamics will require periodic observation for analysis. Remote sensing is formally defined as the analysis of remotely acquired information on a particular object. This is usually accomplished using a sensor that is not in direct contact with the object [42, Ch. 1].

Earth observation satellites are non-military reconnaissance satellites that are used by the remote sensing community to acquire periodic observations of the Earth. These satellites use sensors to capture electromagnetic radiation which is reflected from or emitted by the Earth. The first Earth observation satellite that was developed was the Earth Resource Technology Satellite (ERTS-1), which was renamed to Landsat 1. It was designed to acquire multi-spectral medium resolution imagery on a systematic and recurring basis [43, Ch. 1].

Numerous additional remote sensing systems were commissioned and deployed through various agencies around the world after the success of the ERTS-1 mission. The Group on Earth Observations (GEO) was created in February 2005 to unite 60 national governments and 40 international organisations to implement the Global Earth Observation System of Systems (GEOSS). The main objective is to create high-quality, long-term, global observations in a timely fashion at minimal cost. The GEOSS system will ultimately monitor all aspects of the Earth's system to study global change.

A host of nations have launched hundreds of satellites into orbit since 1957, and this created a range of specifications that must be considered when choosing a sensor on a satellite for a specific application [43, Ch. 2]. The various permutations of the specifications are passive versus active sensors, the range of electromagnetic spectrum sensed, spectral bandwidth of each sensor, temporal acquisition rate, spatial resolution, radiometric resolution, etc. These specifications are discussed in successive sections along with the interaction of various components within a remote sensing system.

2.4 ELECTROMAGNETIC RADIATION

Electromagnetic radiation is a disturbance produced by an oscillation or acceleration of an electric charge. This disturbance consists of electromagnetic waves that comprise electric and magnetic fields which propagate perpendicular to one another with a set of time and spatial properties.

The electromagnetic wave oscillates through a medium with successive cycles and the distance between each completed cycle is called a wavelength. The energy density of the wave is defined by the amplitude. All electromagnetic waves radiate to the same wave theory and travel at the speed of light in a vacuum.

The electromagnetic wave acts according to its wavelength when it comes into contact with an object and can either reflect, refract, diffract or interfere. Electromagnetic radiation is classified into several categories according to wavelength: long waves, radio waves, microwaves, infrared, visible, ultraviolet, X-rays and Gamma rays. The categorised wavelengths are shown in figure 2.3.

One of the major sensor specifications on board a satellite is the deployment of either an active or passive sensor. An active sensor illuminates a scene with its own source of electromagnetic radiation.

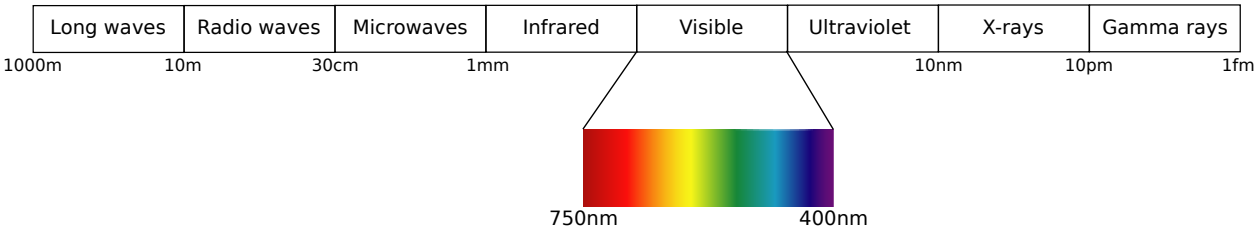


FIGURE 2.3: The electromagnetic spectrum [42, Ch. 1].

The source is set to a range of wavelengths of interest, which is typically in the 2.4 cm–107 cm range.

A passive sensor relies on the sun's radiation to illuminate a scene. A passive sensor is also called an optical sensor, as it operates in the visible and infrared spectrum. The visible spectrum is the most popular range in the electromagnetic spectrum, as it can be sensed by biological organisms.

The properties of the sun's radiance are of importance for a passive sensor, as it produces a wide range of wavelengths with a non-uniform energy distribution. Planck's law states that the spectral radiance is a function of the object's temperature and wavelength of the electromagnetic radiation [44]. The sun's peak emission is in the 400 nm–750 nm spectrum range, which is referred to as the visible spectrum. The spectral distribution across the spectrum remains relatively unchanged as it propagates through space [43, Ch. 2], but the reduction in intensity is subjected to the inverse-square law of the distance between the sun and the Earth [44].

2.5 EARTH'S ENERGY BUDGET

The Earth receives incoming energy from the sun and stars, while losing energy either through absorption, reflectance and transmittance [45, 46]. The conservation of energy states that an equilibrium between the incoming and outgoing energy must be preserved. This equilibrium is a function of the wavelength λ and is expressed as

$$\mathrm{E_I}(\lambda) = \mathrm{E_R}(\lambda) + \mathrm{E_A}(\lambda) + \mathrm{E_T}(\lambda), \tag{2.1}$$

where $\mathrm{E_I}(\lambda)$ denotes the incoming energy, $\mathrm{E_R}(\lambda)$ denotes the reflected energy, $\mathrm{E_A}(\lambda)$ denotes the absorbed energy and $\mathrm{E_T}(\lambda)$ denotes the transmitted energy. The total flux of the incoming energy $\mathrm{E_I}(\lambda)$ is a combination of solar radiation, geothermal energy, tidal energy (moon gravity) and heat energy (fossil fuel consumption). The outgoing energy is partitioned into either reflected, absorbed or transmitted radiation. The partitioning of the outgoing energy into either reflected, absorbed or transmitted radiation varies for different wavelengths, atmospheric conditions and geographical properties [42, Ch. 1].

A sensor on board a satellite measures only the reflected energy $\mathrm{E_R}$; to put the emphasis on the reflected energy, equation (2.1) is rewritten as

$$\mathrm{E_R}(\lambda) = \mathrm{E_I}(\lambda) - \mathrm{E_A}(\lambda) - \mathrm{E_T}(\lambda). \tag{2.2}$$

Approximately 30% of all incoming energy is reflected back into space. The contributions made to the reflected energy by geothermal energy, tidal energy and heat energy are negligibly small when compared to the reflected solar radiation [42, Ch. 1]. The average reflectance of 30% of the incoming energy $\mathrm{E_I}(\lambda)$ is further subdivided: atmospheric reflectance of 6%, cloud reflectance of 20% and the Earth's surface reflectance of 4% [47–49]. A brief overview is given of all the interacting media within the energy budget in the following sections.

2.5.1 Interaction with the atmosphere

Electromagnetic radiation penetrates the atmosphere, which consists of five layers of gases that are retained by the planet's gravitational field [50]. Power and spectral properties of electromagnetic radiation are altered as they propagate through the atmosphere. The atmosphere can either scatter or absorb electromagnetic radiation. The five layers of atmosphere are; the exosphere, thermosphere, mesosphere, stratosphere and troposphere.

The exosphere is the outer layer of the atmosphere. It is a very thin layer where the atoms and molecules leave the atmosphere and dissipate into outer space.

The thermosphere is the second layer that electromagnetic radiation penetrates and this is where most of the Earth Observation satellites orbit. The thermosphere extends between 90 km and 1000 km above sea level. The temperature in the layer is strongly affected by solar activities.

The mesosphere is the middle layer of the atmosphere and extends between 50 km to 90 km above sea level. The majority of the meteors originating from outer space burn up in this layer. It is difficult to measure the properties of the mesosphere, as only sounding rockets can be used at these altitudes.

The stratosphere is the second closest layer to the Earth's surface and is positioned at an altitude of between 8 km and 50 km. The ozone layer is situated within the stratosphere and absorbs most of the harmful solar radiation. An aircraft can fly through the stratosphere because of the temperature stratification within the layer.

The troposphere is the closest layer to the surface of the Earth and rises up to 20 km above sea level. Most weather activities occur within this layer, which holds nearly all water vapour and dust particles. Solar electromagnetic radiation heats up the surface of the Earth and in turn is transferred back to the troposphere.

The atmosphere alters the intensity and spectral composition of electromagnetic radiation before it

is sensed by a sensor on board a satellite. These effects are mainly categorised into either atmospheric scattering or absorption [42, 43].

2.5.1.1 Atmospheric scattering

The principal mechanisms affecting electromagnetic radiation as it propagates through the atmosphere are the scattering and absorption effects. Atmospheric scattering occurs when solar radiation is randomly diffused within the atmosphere. The behaviour of atmospheric scattering is determined by analysing the ratio of the particle's diameter to the wavelength of the electromagnetic wave. Atmospheric scattering is classified into three general categories [42, 43];

- *Rayleigh scattering* is the most common scattering effect in the atmosphere. This scattering occurs when a particle's diameter is much smaller than that of the interacting electromagnetic wave. Rayleigh scattering is inversely proportional to the fourth power of a radiating wavelength. This means that shorter wavelengths are more prone to scatter in the atmosphere than longer wavelengths.
- *Mie scattering* occurs when a particle's diameter is equal to an electromagnetic wave's wavelength. The major causes of Mie scattering are: pollen, dust, smoke, water vapour, and other particles situated in the lower portion of the atmosphere.
- *Non-selective scattering* occurs when an atmospheric particle's diameter is much larger than a radiating wavelength. Non-selective scattering mostly affects the visible, near infrared and mid-infrared spectrums. In this case, all the wavelengths are scattered equally regardless of their length. Non-selective scattering is found in water droplets, which give clouds and fog a white appearance.

2.5.1.2 Atmospheric absorption

Atmospheric absorption is caused by gaseous components that retain electromagnetic radiation within the atmosphere. Atmospheric absorption allows different wavelengths to be absorbed in different parts of the atmosphere. This absorption rate into different layers is illustrated in figure 2.4. The gases that absorb most solar radiation are: water vapour, carbon dioxide, and ozone [42, 43].

Earth observation satellites are limited, as they can only acquire images from wavelengths that are not absorbed into the atmosphere. The range of wavelengths that is not absorbed into the atmosphere is commonly referred to as the *atmospheric window* [42, Ch. 1]. A spectral sensor is usually set to measure a narrow band of spectrum within the atmospheric window.

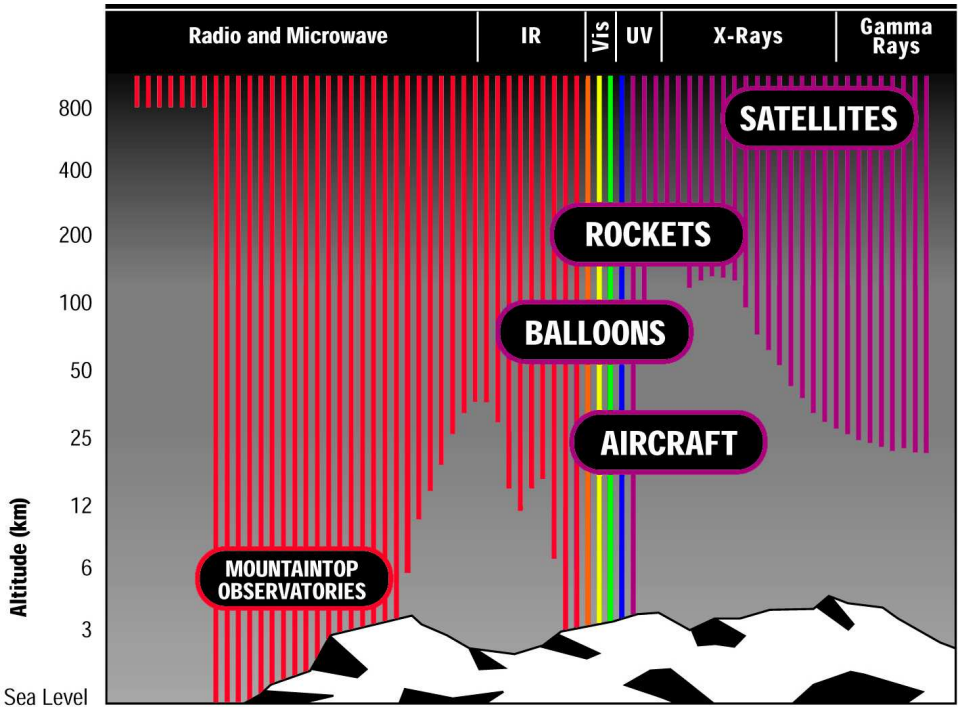


FIGURE 2.4: Atmospheric absorption allows different wavelengths to be absorbed in different parts of the atmosphere. This figure shows the different elevations at which electromagnetic radiation is absorbed into the atmosphere. Image supplied by NASA/CXC/SAO.

2.5.1.3 Atmospheric correction

The electromagnetic radiation recorded at a sensor is not a true reflection of the Earth's surface owing to the effects of atmospheric scattering and absorption. A critical preprocessing step for creating oceanic and land surface products is the correction of these atmospheric disturbances [51, 52].

Two general methods are used in correcting atmospheric disturbances: relative and absolute correction. Relative atmospheric correction is exactly as the term implies a relative histogram match of an image to a reference image. This method requires an accurate reference image for a specified geographical area and any adjoining areas.

Absolute atmospheric correction is further subdivided into empirical and physical methods. The absolute empirical method is not popular, as it has a tendency to over-simplify the corrections applied to an image.

The absolute physical method, on the other hand, uses a mathematical model to extract the effects of various gaseous components and then to compensate for these effects accordingly. A radiative transfer model is a form of the absolute physical method which extracts the gaseous concentrations directly from an image in order to estimate the corrected radiance for the image.

2.5.2 Interaction with the Earth's surface

The Earth's surface interacts with incoming electromagnetic radiation and can either absorb, reflect and/or transmit the radiation. The reflected electromagnetic radiation excites the components within the sensor. The amount of reflected electromagnetic radiation is a function of the wavelength and the properties of the surface. The surface has several properties that affect the amount of reflectance: mineral profile, surface contour, surface roughness, etc. Reflected electromagnetic waves are mostly affected by the surface's roughness and are divided into two general modes: specular (smooth) and diffuse (rough or Lambertian) [33, Ch. 4].

The Rayleigh criterion determines the level of roughness for a medium and is calculated as

$$h \leq \frac{\lambda}{8\cos(\theta)}. \tag{2.3}$$

The variable h denotes the surface irregularity height, λ denotes the wavelength and θ denotes the angle of incidence measured to the azimuth. If equation (2.3) is satisfied, then the surface is considered to be diffuse, otherwise it is specular [42, 43].

A specular surface reflects electromagnetic radiation according to Snell's law, which states that the outgoing energy is exactly reflected at a perpendicular angle to the azimuth of the incoming energy. A diffuse surface reflects the incoming electromagnetic radiation in all directions off the surface. A Lambertian (perfect diffuse) surface reflects the incoming energy uniformly in all directions off the surface.

Most natural surfaces are imperfect diffuse reflectors (specular component present) in the visible and near infrared spectrum. This makes remote sensing possible, as reflected electromagnetic radiation can be captured at most viewing angles. This would not be possible if the surface was completely specular, as it would have a high reflectance value at a single specific viewing angle and relatively low reflective values at all other viewing angles [53, 54].

2.5.3 Interaction with a satellite-based sensor

The principal concept of remote sensing is to observe an object remotely. In a satellite-based application it is the recording of electromagnetic radiation that has interacted with an object. A sensor, as defined in this thesis, is a device that measures a physical quantity and converts it into an electrical signal.

The advantage, when considering the interaction of radiation with the sensor, is that it can be designed to measure the environment optimally. A satellite sensor's specifications that will be discussed briefly are: the spatial, spectral, radiometric and temporal resolutions.

Spatial resolution is the geographical size that is recorded on a two-dimensional pixel in the image. The size of the area represented in a pixel is determined by the altitude, viewing angle and sensor characteristics. All these characteristics are influenced by the instantaneous field of view (IFOV) of the sensor [33, Ch. 4]. The IFOV of the sensor is time-dependent, as the satellite is not perfectly stable in its orbit. The distance between the satellite and the Earth varies continually, altering the physical size of the geographical area that is captured within a single pixel.

Another limiting factor is the point spread function (PSF) of the sensor. The PSF is the system impulse response between the geographical area and the sensor. This function describes the degree of illumination spreading from the adjacent area to the geographical area of interest. The PSF results in a blending or spreading effect on areas with relatively bright or dark objects within the IFOV of the sensor. This leads to high contrast features becoming indiscernible on satellite images even though their widths are less than the sensor's spatial resolution.

Spectral resolution is the bandwidth of the electromagnetic spectrum recorded by the sensor. A sensor that senses a shorter spectrum range of wavelengths (smaller bandwidth) has an improved ability to capture the spectral signature of an object within the spectral band when compared to a sensor that measures a larger spectrum range of wavelengths (larger bandwidth).

The disadvantage of increasing the spectral resolution is that the signal-to-noise ratio (SNR) decreases. Recorded radiance at the sensor is adversely affected by some form of noise. The physical propagation of electromagnetic radiation to the sensor can be seen as a time-variant multi-path propagation of the reflected electromagnetic wave of a geographical area with a certain level of additive noise. The additive noise in the sensor is made up mostly of thermal noise. The thermal noise does not decrease if a smaller bandwidth is sensed, although the instantaneous radiance in the sensor is reduced for a higher spectral resolution sensor as it is exposed to a shorter range of spectrum. The thermal noise remains the same regardless of the range of spectrum that is being sensed. To summarise: reducing the reflected power within the sensor (reducing the bandwidth) will inadvertently reduce the SNR.

Optimal spectral resolution is obtained when a sensor mitigates the effect of additive noise and has a spectral bandwidth that captures the best matched spectral signature for the intended remotely sensed object. Remote sensing systems usually use multi-spectral or hyper-spectral sensors. This is an array of sensors that capture different ranges of spectrum at the same time. A multi-spectral sensor has less than 100 unique spectral bands, while a hyper-spectral sensor has more than 100.

Radiometric resolution is the accuracy of converting electromagnetic radiation at the satellite sensor

to a digital binary format. A higher radiometric resolution enables the satellite sensor to distinguish between more levels of intensity.

It is possible to encode electromagnetic radiation as an information source at a rate that is close to its entropy [55, Ch. 6]. This is unfortunately limited by the storage space available on the satellite, which induces a certain level of distortion in the sampling of the electromagnetic radiation. The reason is that electromagnetic radiation is an analog source and requires an infinite number of binary bits to store.

A loss in precision is caused by the finite storage space, which induces a distortion that is directly related to the number of quantisation levels (number of binary bits per radiance sample). It should be noted that the number associated with each quantisation level is not a direct measure of the captured electromagnetic radiation, but rather the steps into which a range of physical values is divided.

In an effort to distribute the captured electromagnetic radiation more evenly over the range of quantisation levels, some sensors apply either non-linear quantisation mapping functions or an amplifier with an automatic gain control mechanism. This alters the intensity of the captured electromagnetic radiation and distributes it over a range of different quantisation levels without creating a saturated buffer in the remotely sensed image.

The total number of quantisation levels and the method of distributing radiation across the levels affect the level of distortion in the stored values. This rate of distortion is defined by the signal-to-quantisation-noise ratio (SQNR), which is expressed as

$$\mathrm{SQNR} = \frac{P_x}{P_{\tilde{x}}}. \tag{2.4}$$

The variable $P_{\tilde{x}}$ is the quantisation-noise power and P_x is the power of the radiation before quantisation.

Low-quality sensors have low SQNR, which equates to low radiometric resolution. The disadvantage in increasing the radiometric resolution is the costs and complexity of adding a higher resolution analogue-to-digital converter device and the increase in required storage space for storing the binary values of the digital image. For example, the Quickbird satellite owned by DigitalGlobe has a radiometric resolution of 11 bits. This enables the sensor to distinguish between 2048 (2^{11}) levels of radiance. The satellite has 128 Gb storage capacity, which equates to 57 images stored on board. The sensor can distinguish between 65536 (2^{16}) levels of radiance if the radiometric resolution is set to 16 bits. The problem is that only 39 images can be stored on board, which results in a 32% reduction in storage capacity.

Temporal resolution is the periodic rate of acquisition of a geographical area by the same satellite sensor. This is important for investigating any change in land surface and the monitoring of global environmental processes. The orbit, altitude, swath width, and priority tasking of the sensor on board

the satellite determines the temporal rate at which an area of interest can be imaged [42, Ch. 6]. Sensors are tasked from a mission control center to acquire images of geographical areas. Areas of interest are assigned a priority task, which improves the temporal acquisitions for this area. The temporal resolution varies from less than an hour to more than a few months [43, Ch. 2]. Fixed temporal resolution is a sensor that has a fixed viewing angle, repetitive orbital track and a fixed swath width.

The swath width is the trade-off between the temporal resolution and the spatial resolution. The wider the swath width, the shorter the revisit time period for a geographical area, while the narrower the swath width, the better the spatial resolution (for the same number of pixels).

TABLE 2.1: Specification of different remote sensing sensors.

Sensor	Temporal resolution (Revisit period)	Spatial resolution	Wavelength range	Number of spectral bands
Enhanced Thematic Mapper Plus (ETM+)	16 days	15 m – 60 m	0.45 μm–12.50 μm	8
MODerate-Resolution Imaging Spectroradiometer (MODIS)	1–2 days	250 m – 1000 m	0.405 μm–14.385 μm	36
Advanced Very High Resolution Radiometer (AVHRR)	Daily	1100 m – 4000 m	0.58 μm–12.50 μm	5

How to choose a sensor: This thesis focuses on expanding settlements. Finding newly developed housing requires several considerations when selecting the right remote sensing sensor.

High spatial resolution sensors have the ability to detect much smaller objects in an area. The drawback is that higher spatial resolution means lower temporal resolution. These images are thus not regularly acquired and are financially expensive.

Detecting new settlements is possible when comparing two high spatial resolution images taken at two different dates. The problem is that similar land cover types can appear significantly different at various times of the year. These seasonal changes in the land cover can be mitigated if the temporal resolution is high enough to capture these trends [15]. This makes the use of high temporal resolution sensors much more useful for change detection.

A list of specifications for three different satellites used to image the land surface is shown in table 2.1. The specifications for these three satellites are used to illustrate the range of trade-offs to consider when selecting a sensor.

The Enhanced Thematic Mapper Plus (ETM+) operates on a very high spatial resolution of 15 m – 60 m, with a low temporal revisit time of 16 days.

The Advanced Very High Resolution Radiometer (AVHRR) has a high temporal resolution of one day, but captures a geographical area at a spatial resolution of 1100 metres. The large swath width is necessary to obtain a high temporal resolution at the expense of the spatial resolution.

The MODerate-resolution Imaging Spectroradiometer (MODIS) is a newer instrument, which was

specifically designed for global land surface monitoring and is the chosen sensor for this study, as it has a high temporal resolution and medium spatial resolution capabilities [16]. MODIS has a temporal resolution of 1–2 days, which is close to the temporal resolution of the AVHRR sensor. MODIS also has a medium spatial resolution (250 m – 1000 m) and a wider variety of spectral bands.

2.6 MODERATE RESOLUTION IMAGING SPECTRORADIOMETER

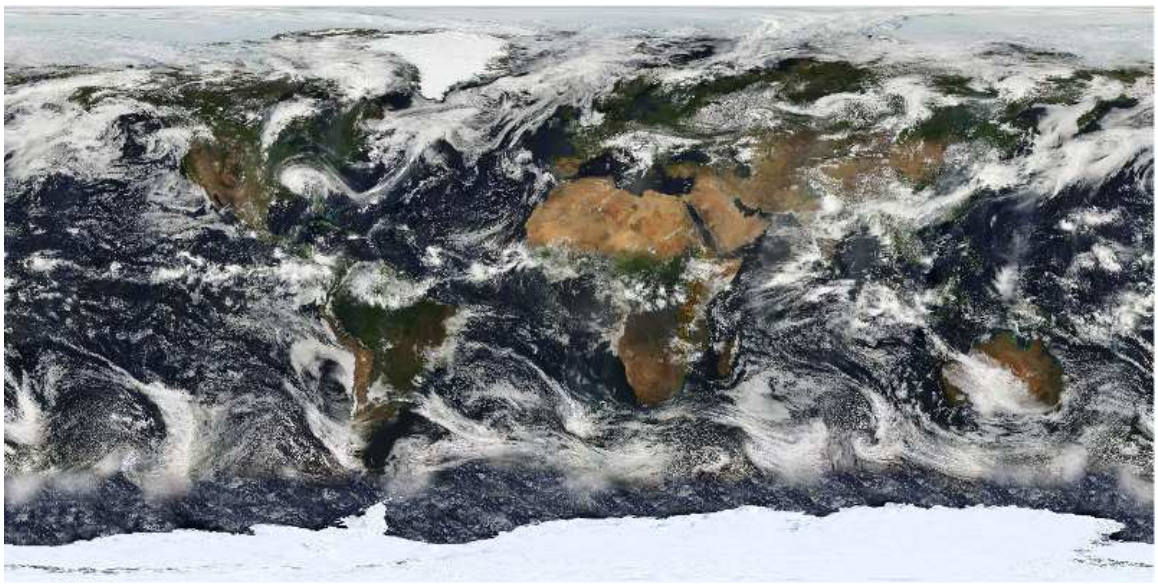


FIGURE 2.5: Multiple MODIS images concatenated to form a image of the Earth.

MODIS is an experimental scientific sensor launched into the Earth's thermosphere by NASA on board the Terra EOS-AM-1 satellite on December 18, 1999. A second MODIS sensor was launched on board the Aqua EOS-PM-1 satellite on May 4, 2002.

The Terra EOS satellite was the first NASA scientific research satellite to carry the MODIS instrument into orbit. The Terra satellite was launched from the Vandenberg Air Force base into a sun-synchronous orbit at an altitude of 705 km [56]. Terra is Latin for *Earth*. The Terra EOS satellite carries a total of five remote sensing sensors which record measurements of the Earth's climate system: Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), Clouds and the Earth's Radiant Energy System (CERES), Multi-angle Imaging SpectroRadiometer (MISR), MODIS and Measurements of Pollution in the Troposphere (MOPITT).

The Aqua EOS satellite was the second NASA scientific research satellite to carry a MODIS instrument into orbit. The Aqua satellite was launched from the Vandenberg Air Force base into an afternoon equatorial crossing orbit at an altitude of 705 km [56]. Aqua is Latin for *water*. The Aqua EOS satellite carries a total of six remote sensing sensors that collects information about the Earth's

TABLE 2.2: MODIS spectral bands properties and characteristics.

Spectral bands	Wavelengths (nanometres)	Resolution (metres)	Property or characteristic	Spectral range
Band 1	620–670	250	Absolute Land Cover Transformation, Vegetation Chlorophyll	Visible (Red)
Band 2	841–876	250	Cloud Amount, Vegetation Land Cover Transformation	Near Infrared
Band 3	459–479	500	Soil/Vegetation Differences	Visible (Blue)
Band 4	545–565	500	Green Vegetation	Visible (Green)
Band 5	1230–1250	500	Leaf/Canopy Differences	Short Infrared
Band 6	1628–1652	500	Snow/Cloud Differences	Short Infrared
Band 7	2105–2155	500	Cloud Properties, Land Properties	Short Infrared
Band 8	405–420	1000	Chlorophyll	Visible (Blue)
Band 9	438–448	1000	Chlorophyll	Visible (Blue)
Band 10	483–493	1000	Chlorophyll	Visible (Blue)
Band 11	526–536	1000	Chlorophyll	Visible (Green)
Band 12	546–556	1000	Sediments	Visible (Green)
Band 13	662–672	1000	Atmosphere, Sediments	Visible (Red)
Band 14	673–683	1000	Chlorophyll Fluorescence	Visible (Red)
Band 15	743–753	1000	Aerosol Properties	Near Infrared
Band 16	862–877	1000	Aerosol Properties, Atmospheric Properties	Near Infrared
Band 17	890–920	1000	Atmospheric Properties, Cloud Properties	Near Infrared
Band 18	931–941	1000	Atmospheric Properties, Cloud Properties	Near Infrared
Band 19	915–965	1000	Atmospheric Properties, Cloud Properties	Near Infrared
Band 20	3660–3840	1000	Sea Surface Temperature	Mid wave Infrared
Band 21	3929–3989	1000	Forest Fires & Volcanoes	Mid wave Infrared
Band 22	3929–3989	1000	Surface/Cloud Temperature	Mid wave Infrared
Band 23	4020–4080	1000	Surface/Cloud Temperature	Mid wave Infrared
Band 24	4433–4498	1000	Cloud Fraction, Troposphere Temperature	Mid wave Infrared
Band 25	4482–4549	1000	Cloud Fraction, Troposphere Temperature	Mid wave Infrared
Band 26	1360–1390	1000	Cloud Fraction (Thin Cirrus), Troposphere Temperature	Mid wave Infrared
Band 27	6535–6895	1000	Mid Troposphere Humidity	Mid wave Infrared
Band 28	7175–7475	1000	Upper Troposphere Humidity	Long wave Infrared
Band 29	8400–8700	1000	Surface Temperature	Long wave Infrared
Band 30	9580–9880	1000	Total Ozone	Long wave Infrared
Band 31	10780–11280	1000	Cloud Temperature, Forest Fires & Volcanoes, Surface Temperature	Long wave Infrared
Band 32	11770–12270	1000	Cloud Height, Forest Fires & Volcanoes, Surface Temperature	Long wave Infrared
Band 33	13185–13485	1000	Cloud Fraction, Cloud Height	Long wave Infrared
Band 34	13485–13785	1000	Cloud Fraction, Cloud Height	Long wave Infrared
Band 35	13785–14085	1000	Cloud Fraction, Cloud Height	Long wave Infrared
Band 36	14085–14385	1000	Cloud Fraction, Cloud Height	Long wave Infrared

water cycle. The six sensors are: the Atmospheric Infrared Sounder (AIRS), Advanced Microwave Sounding Unit (AMSU-A), Humidity Sounder for Brazil (HSB), Advanced Microwave Scanning Radiometer for EOS (AMSR-E), MODIS, and CERES.

NASA's strategy is to use the MODIS sensors to investigate and acquire hyper-temporal, multi-spectral and multi-angular observations of the Earth on a daily basis. MODIS was launched to continue the monitoring of the Earth from older sensors such as: Coastal Zone Colour Scanner (CZCS), the Advanced Very High Resolution Radiometer (AVHRR), the High Resolution Infrared Spectrometer (HIRS), and the Thematic Mapper (TM). The MODIS sensors were built by the Santa Barbara Remote Sensing Institute according to the specifications provided by NASA. NASA has gone to great lengths to ensure proper sensor calibration to generate an accurate long-term data set for global studies [57].

MODIS is a passive remote sensing instrument with 490 detectors, which are arranged to form 36 spectral bands that measure the 405 nm–14385 nm spectrum. Each detector in the sensor has a 12-bit

TABLE 2.3: Table description of the available MODIS land cover products.

Product	Short Description	Composition time	Spatial Resolution	Satellites	Product Code
Snow product	Snow cover land and snow albedo	Daily/8-day	500m/1km	Terra or Aqua	MOD10/MYD10 MOD29/MYD29
Land surface temperature	Land surface temperature and emissivity daily levels	Daily/8-day/ Monthly	1km/6km	Terra or Aqua	MOD11/MYD11
Land cover dynamic product	Decision tree classify 34 classes of land cover	Yearly	500m/1km	Terra or Aqua	MOD12/MYD12
Thermal Anomalies/ Fire products	Fire detection	Daily/8-day	1km	Terra or Aqua	MOD14/MYD14
LAI/FPAR products	Measure surface photosynthesis, evapotranspiration, and net primary production	8-day	1km	Terra, Aqua or combined	MOD15/MYD15/ MCD15
Gross Primary Production product	Measures growth of terrestrial vegetation	8-day	1km	Terra or Aqua	MOD17/MYD17
Surface Reflectance	Spectral reflectance and atmospheric scattering	Daily/8-day	250m/500m/ 1km	Terra or Aqua	MOD09/MYD09
Global Vegetation Indices	Calculates the NDVI and EVI indices	16-day/Monthly	250m/500m/ 1km	Terra or Aqua	MOD13/MYD13
Vegetation Cover Conversion	Estimate proportions of life form, leaf type, and leaf longevity	Yearly	500m	Terra	MOD44
BRDF/Albedo products	Mathematical models to describe BRDF and derive Albedo measurements	8-day/16-day	500m/1km	Terra, Aqua or combined	MOD43/MYD43/ MCD43
Burned Area product	Burning and quality information and survey for rapid changes on surfaces	Monthly	500m	Combined	MCD45

radiometric resolution and can acquire a swath of 2330 km (cross track) by 10 km (nadir track). The wide swath width of MODIS enables it to record the entire Earth's surface every two days. MODIS spectral bands are recorded at a different spatial resolutions: spectral bands 1–2 are measured at 250 m spatial resolution, spectral bands 3–7 are measured at 500 m spatial resolution and spectral bands 8–36 are measured at 1 km spatial resolution. The spatial resolution is reported at a nadir viewing angle. It should be noted that an increase in spatial resolution is experienced in the scan direction, which causes pixels to be partially overlapping at off-nadir angles. This phenomenon is known as the bowtie effect and is a source of variability over the revisit cycle.

The spectral bands are designed to provide observations of global environmental processes occurring in the troposphere: cloud activity, radiation budget, oceanographic occurrences and land cover monitoring (Full listing in Table 2.2). The images acquired by MODIS are converted with a set of preprocessing steps on a daily basis into terrestrial, atmospheric and oceanic products (Full product listing in Table 2.3).

The prefix MOD and MYD in the product code (table 2.3) refers to the product derived from the data acquired from the Terra and Aqua satellites respectively. The prefix MCD in the product code refers to the product derived using data from both satellites [27, 28, 58–60]. The composition time

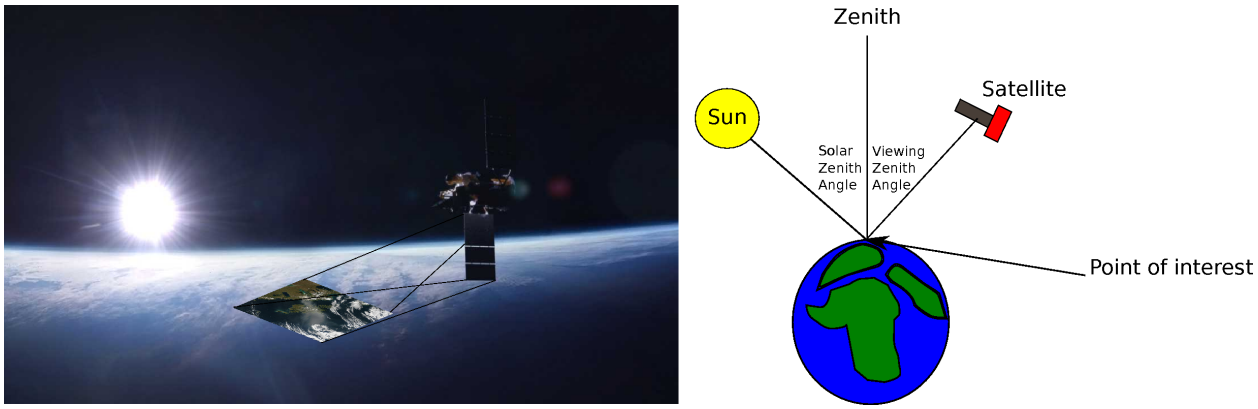


FIGURE 2.6: Example of a passive orbiting satellite acquiring an image from earth.

(table 2.3) reports the temporal resolution at which an acquisition for the product becomes available and the spatial resolution at which the products are produced.

The MODIS product chosen for this thesis is the MCD43A4 land surface reflectance product. The product is defined as a nadir viewed land surface reflectance, which is atmospherically corrected [61, 62]. The adjusted land spectral reflectance product significantly reduces the anisotropic scattering effects of surfaces under different illumination and observation conditions [27, 28]. This BRDF/Albedo product is also used as an input to derive land classifications for the *Land Cover Dynamic Product*. The MCD43A4 product uses the first 7 spectral bands, which are often referred to as the land surface bands. The 7 spectral bands are used because of the minimal atmospheric absorption of atmospheric gases.

The larger swath width on MODIS enables the surveying of every geographical area at least every two days. The MODIS instrument has an orbital repeat cycle of 16 days, which is a problem with the large swath width, as the viewing angles (at the same ground location) between successive observations might differ dramatically. This means that every 16 days an image is acquired of the same geographical area with similar viewing angles.

The disadvantage of acquiring images from a polar orbiting passive satellite is the variation in the reflected signal that is caused by the change in the surface reflectance during the composition period (Figure 2.6). This variation in signal is contributed by many different environmental and external sources such as: solar zenith angle, viewing zenith angle, seasonality, sensor angle, etc.

This disadvantage created the need to consider the distribution of the electromagnetic radiation as a function of the observation and illumination angles. The BRDF is a mathematical function which describes the variability in surface reflection based on the illumination and viewing angles [63].

Estimation of the BRDF enables the adjustment of the reflectance values as if they were taken from a nadir view. The MODIS MCD43A4 product uses a 16-day rolling window of acquisitions from both Terra and Aqua satellites, together with a semi-empirical kernel-driven bidirectional reflectance model

to determine the global set of parameters describing the BRDF. The hemispherical reflectance and the bi-hemispherical reflectance at the solar zenith angle are derived from the BRDF parameters to produce a coarse resolution composite image every 8 or 16 days [28].

A weighted linear sum of kernel functions is used for a BRDF model to correct for illumination and viewing angles. This BRDF model is a 4-variable function that sums together an isotropic parameter and two functions of viewing and illumination geometry to determine the reflectance [28]. The BRDF model is given by

$$\begin{aligned} R(\theta_{\text{sol}}, \theta_{\text{view}}, \theta_{\text{rel}}, \lambda) = f_{\text{iso}}(\lambda) + f_{\text{vol}}(\lambda) K_{\text{vol}}(\theta_{\text{sol}}, \theta_{\text{view}}, \theta_{\text{rel}}, \lambda) \\ + f_{\text{geo}}(\lambda) K_{\text{geo}}(\theta_{\text{sol}}, \theta_{\text{view}}, \theta_{\text{rel}}, \lambda), \end{aligned} \tag{2.5}$$

where θ_{sol} denotes the solar zenith angle and θ_{view} denotes the viewing angle. The variable θ_{rel} denotes the relative azimuth angle and λ denotes the wavelength.

The RossThick kernel function is currently best suited for the volume scattering radiative transfer model used in the kernel function $K_{\text{vol}}(\theta_{\text{sol}}, \theta_{\text{view}}, \theta_{\text{rel}}, \lambda)$ for the MODIS MCD43A4 product. The LiSparce kernel function is at present best suited for the geometric shadow casting theory used in the kernel function $K_{\text{geo}}(\theta_{\text{sol}}, \theta_{\text{view}}, \theta_{\text{rel}}, \lambda)$ [28].

The BRDF model's parameters are derived by the MODIS MOD43B1 product and are used to compute the albedos using the solar illumination geometry. The approximation of terrestrial albedo at a particular solar zenith angle, requires a weighted sum of the black-sky (directional-hemispherical) albedo and the white-sky (bi-hemispherical) albedo. The black-sky albedo is defined as albedo in the absence of a diffuse component and is a function of the solar zenith angle. The white-sky albedo is defined as albedo in the absence of a direct component when the diffuse component is isotropic [28]. The product uses the black-sky and white-sky model for albedo estimation.

The black-sky model is given as

$$\begin{aligned} \alpha_{\text{BS}} = f_{\text{iso}}(\lambda)(g_{0,\text{iso}} + g_{1,\text{iso}}\lambda^2 + g_{2,\text{iso}}\lambda^3) \\ + f_{\text{vol}}(\lambda)(g_{0,\text{vol}} + g_{1,\text{vol}}\lambda^2 + g_{2,\text{vol}}\lambda^3) \\ + f_{\text{geo}}(\lambda)(g_{0,\text{geo}} + g_{1,\text{geo}}\lambda^2 + g_{2,\text{geo}}\lambda^3). \end{aligned} \tag{2.6}$$

The coefficients for the black-sky model for the isotropic (iso), the RossThick (vol) and LiSparce (geo) can be substituted into equation (2.6) to simplify to

$$\alpha_{\mathrm{BS}} = f_{\mathrm{iso}}(\lambda) + f_{\mathrm{vol}}(\lambda)(-0.007574 - 0.070987\lambda^2 + 0.307588\lambda^3) + f_{\mathrm{geo}}(\lambda)(-1.284909 - 0.166314\lambda^2 + 0.04184\lambda^3). \quad (2.7)$$

The white-sky model is given as

$$\alpha_{\mathrm{WS}} = f_{\mathrm{iso}}(\lambda)g_{\mathrm{iso}} + f_{\mathrm{vol}}(\lambda)g_{\mathrm{vol}} + f_{\mathrm{geo}}(\lambda)g_{\mathrm{geo}}. \quad (2.8)$$

The coefficients for the white-sky model are also substituted into equation (2.8), which equates to

$$\alpha_{\mathrm{WS}} = f_{\mathrm{iso}}(\lambda) + 0.189184 f_{\mathrm{vol}}(\lambda) - 1.377622 f_{\mathrm{geo}}(\lambda). \quad (2.9)$$

The solar zenith angle is then transformed to a nadir angle at local sensor noon using the BRDF model.

Cloud obscuration reduces the number of observations that are available for processing even when both satellites are combined within a product. Fortunately, according to a global analysis conducted, South Africa has more than an 80% probability of acquiring enough non-cloudy images within 16 days to produce a reliable 8 day composite land reflectance MODIS product [64].

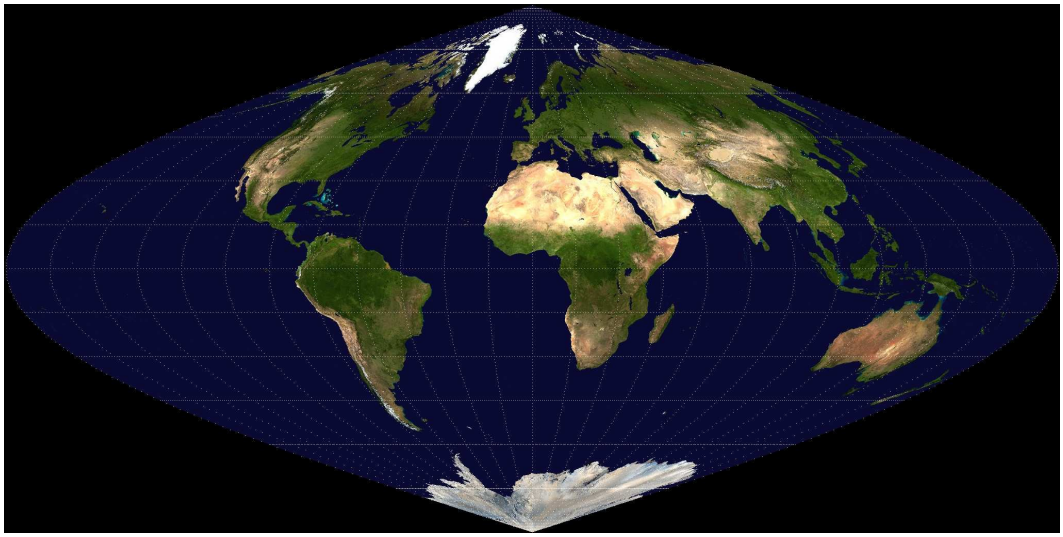


FIGURE 2.7: Sinusoidal projection of the the planet Earth.

The land surface reflectance products are sinusoidally projected and stored in a Hierarchical Data Format - Earth Observing System (HDF-EOS) format [65]. A sinusoidal projection of the planet Earth is shown in figure 2.7. The sinusoidal projection is a pseudocyclindrical projection, which translates images to retain relative geographical sizes between areas accurately. These images are then gridded to form an equal-sized gridded map. The disadvantage is that it distorts the shapes and orientation within the maps when viewing the images.

The PSF of the MODIS sensor was not measured prelaunch; instead a line spread function (LSF) was measured in the scan direction to derive the PSF [66]. The MODIS PSF induced radiation from adjacent areas which is mostly caused by clouds. A correction for this unwanted radiation entering the sensor is computed using both the PSF and the approximation of the radiance measured by the saturated spectral bands. This prior knowledge of the radiance received is usually discarded in most products, as it requires long computing times. The largest impact is the low radiance measured in MODIS oceanic products, which are in close proximity to highly reflective objects such as clouds, coastlines, or sun glint. The PSF introduces a small amount of straylight into the MODIS measurements, which does not have a large impact on land surface products.

2.7 VEGETATION INDICES

Vegetation indices were created to assist in the study of terrestrial vegetation in large-scale global environmental dynamics. Vegetation indices are spectral transformations of a set of spectral band combinations. The vegetation indices enhance the vegetation characteristics within an image, which facilitates the comparison of terrestrial photosynthetic activity variations [67].

2.7.1 Normalised Difference Vegetation Index

The Normalised Difference Vegetation Index (NDVI) is a scalar index that enhances vegetation characteristics in a multi-spectral image. The NDVI was inspired by phenology, which is the study of the periodical growth cycle of plants and how this cycle is influenced by seasonal and inter-annual variability in the ecosystem [68]. A global NDVI coverage map is shown in figure 2.8. NDVI is a normalised ratio that uses the λ_{RED} (Red spectrum band 0.63 μm – 0.69 μm) and λ_{NIR} (Near Infrared spectrum band 0.76 μm – 0.90 μm) spectral bands and is computed as

$$\mathrm{NDVI} = \frac{\lambda_{\mathrm{NIR}} - \lambda_{\mathrm{RED}}}{\lambda_{\mathrm{NIR}} + \lambda_{\mathrm{RED}}}. \tag{2.10}$$

The NDVI index capitalises on the differences in absorption rates between the two spectral bands when interacting with natural vegetation. The RED spectral band's electromagnetic radiation is absorbed by the natural vegetation for photosynthesis and the NIR spectral band's electromagnetic radiation is reflected by the natural vegetation because of the vegetation's cellular structure. The NDVI index exploits the low reflectance values in the RED spectral band and high reflectance values in the NIR spectral band for natural vegetation [69, 70]. The NDVI ratio shown in equation (2.10) produces positive values near 1 ($\mathrm{NDVI} \approx 1$) for areas containing a dense vegetation canopy and small positive values ($\mathrm{NDVI} \approx 0$) for bare soils.

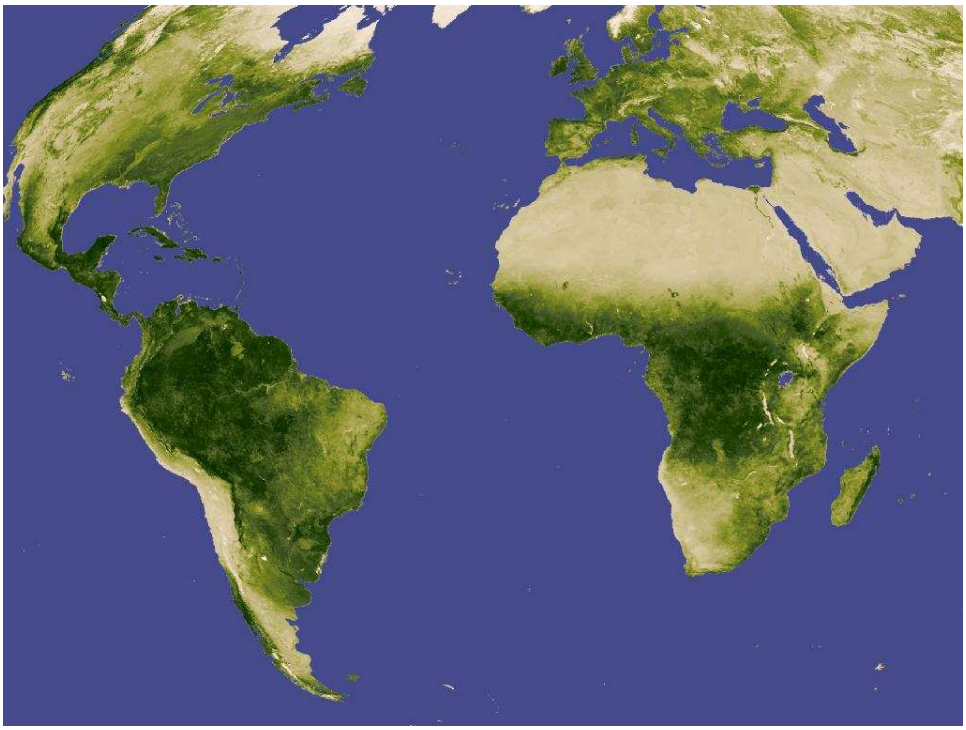


FIGURE 2.8: Global NDVI index coverage map created using MODIS. Image supplied by NASA.

The general use of the NDVI index is demonstrated in large regional environmental models, which include: leaf area index, biomass, chlorophyll, net plant productivity, fractional vegetation cover, accumulated rainfall, etc. Several studies tend to over-use the NDVI index in many applications for which it was not specifically designed [71]. The normalised difference between these two spectral bands only illustrates a relationship in the original information, while other important information is discarded. Whether the discarded information is relevant depends on the process of analysis and geographical area. The NDVI index is sensitive to numerous environmental factors, including atmospheric effects, thin cloud coverage (ubiquitous cirrus), moistness of the soil (precipitation or evaporation), difference in soil colour, anisotropic effects, and spectral effects (different sensors provide different NDVI).

Several alternatives to NDVI have been proposed to address a variety of limitations in analysing satellite acquired imagery. These include: the Perpendicular Vegetation Index [72], the Soil-adjusted Vegetation Index [73], the Atmospherically Resistant Vegetation Index [74], and the Global Environment Monitoring Index [71].

2.7.2 Enhanced Vegetation Index

The Enhanced Vegetation Index (EVI) is an improved version of the NDVI vegetation index. The EVI does not tend to saturate as quickly as the NDVI does in areas with high biomass. The EVI decouples

the canopy background reflectance, and is computed as

$$\mathrm{EVI} = G\,\frac{\lambda_{\mathrm{NIR}} - \lambda_{\mathrm{RED}}}{\lambda_{\mathrm{NIR}} + C_1\lambda_{\mathrm{RED}} - C_2\lambda_{\mathrm{BLUE}} + L}. \tag{2.11}$$

The variable λ_{NIR} denotes the surface reflectance of the near infrared band and λ_{RED} denotes the surface reflectance of the red spectral band. The variable λ_{BLUE} denotes the surface reflectance of the blue spectral band and L denotes the canopy background adjustment term. The coefficients C_1 and C_2 denote the aerosol resistance term and G is the gain coefficient.

The scaling coefficients are used to minimise the effects of aerosols. The blue spectral band is atmospherically sensitive and is used to adjust the red spectral band for aerosol influences. The coefficients used by MODIS to calculate EVI are substituted into equation (2.11) as

$$\mathrm{EVI} = 2.5\,\frac{\lambda_{\mathrm{NIR}} - \lambda_{\mathrm{RED}}}{\lambda_{\mathrm{NIR}} + 6\lambda_{\mathrm{RED}} - 7.5\lambda_{\mathrm{BLUE}} + 1}. \tag{2.12}$$

NDVI is the most widely used vegetation index, which could be attributed to its low computational costs. The use of EVI always raises two questions:

1. Does the sensor measure the blue spectral band independently?

2. Are the scaling coefficients used in computing EVI applicable to the current geographical area?

NDVI is a good vegetation index if properly used and was included in this thesis because of its popularity and to create a base performance for comparison [75, 76]. It should be noted that all methods proposed in this thesis could be adapted to operate with other sets of spectral bands and vegetation indices.

2.8 LAND COVER CHANGE DETECTION METHODS

Change detection can be viewed from a prototype theory mindset [77]. The prototype theory states that the performance of the results generated from a change detection method is based on the user's requirements. This creates a paradigm that there is no single solution for detecting change for all applications [18, 20]. Change detection methods are designed for a specific application and have their own merits and limitations.

An example to demonstrate the user's specific needs is shown in figure 2.9. A change in land cover type from natural vegetation to human settlement is experienced in the red polygon, while only seasonal change in the vegetation has occurred in the blue polygon. Applications and issues of change detection in the remote sensing community are summarised into several categories [24], namely:

1. land cover classification and change detection [78, 79],
2. forest monitoring [80, 81],
3. fire detection [82, 83],
4. urban expansion and change [84, 85],
5. natural environment change [86, 87], and
6. specialised applications [88, 89].

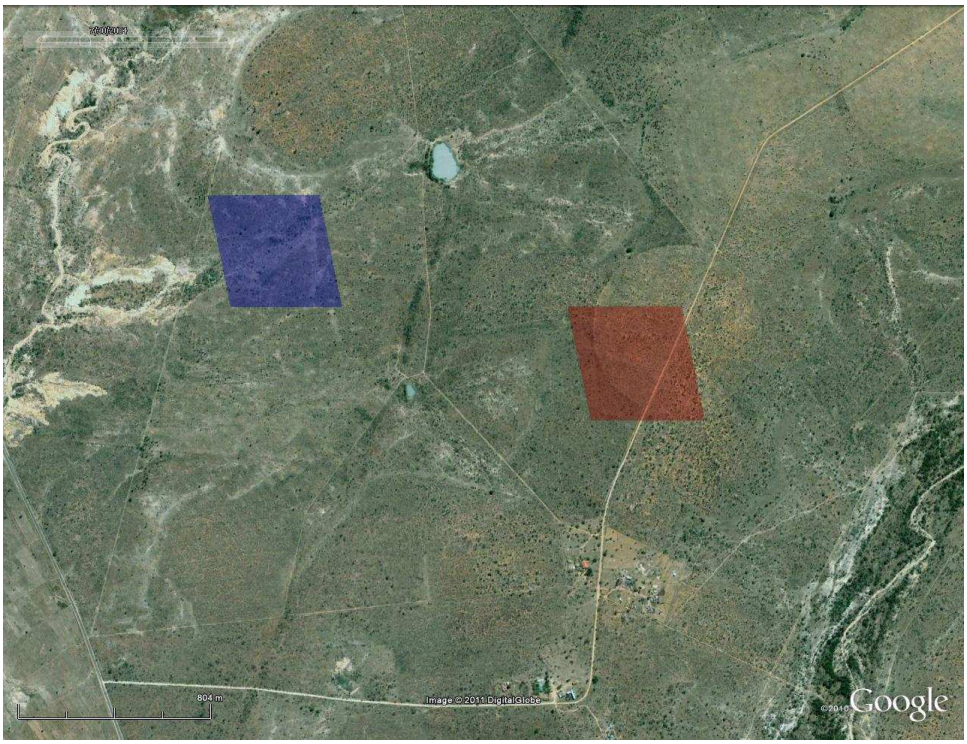
The remote sensing community's monitoring capabilities keep improving with the development and deployment of new technologies. Global data sets are becoming more accessible and computational resources are becoming more affordable [14]. These data sets come from several different sensors. The more popular are: Landsat Multi-Spectral Scanner (MSS), TM, MISR, SPOT, AVHRR and MODIS. The type of land cover change of interest also changes with technologies, which requires continuous pursuit of new change detection methods [18, 20].

There are four major steps involved when constructing a change detection framework [90]. The first step is image preprocessing to ensure the image is corrected by removing any unwanted artifacts [18, 20]. Preprocessing spatially registers and environmentally corrects each image to a minimum product's quality index. The product's quality index is reached by using topographical correction, spatial registration, radiometric calibration, atmospheric calibration and normalisation between multi-temporal imagery.

The purpose of the preprocessing is the assurance that the images acquired over a geographical area remain consistent through time and any changes in the reflectance values are not caused by processing artifacts. Incorrect preprocessing has adverse effects on the accuracy of the change detection methods [91, 92]. For example, if images are not correctly spatially registered, the geographical location of a pixel in one image will not correspond with the geographical location of the same pixel in another image.

The second step is proper feature extraction and selection. Suitable meaningful features must be obtained from the images to give the change detection method the ability to detect change. A renowned quotation is: *If you can measure it, you can improve it - William Thomson*. If no measurable feature exists to detect the change, then no change detection method will be able to detect it.

The third step is to develop a suitable change detection method that uses the features to detect changes according to the user's requirements. The method must be reliable and robust in most operating environments.



(a) Quickbird image taken on 30 July 2004 (courtesy of Google^{TM}Earth).

(b) Quickbird image taken on 31 December 2008 (courtesy of Google^{TM}Earth).

FIGURE 2.9: A change in land cover type is shown by the red polygon in (a) and (b), while only a seasonal change has occurred in the blue polygon.

The fourth step is the assessment of the previous three steps. How well did the change detection method satisfy the requirements set by the user? The overall accuracy assessed in the system is affected

by several factors, including [24]; (1) the quality of the preprocessing, (2) availability of reliable ground truth, (3) complexity of the environmental case study, (4) useful feature extraction, (5) feature analysis and processing, (6) change detection algorithms used, (7) the analyst's skills, (8) knowledge and information about the study area, (9) critical assessment of the system's outputs, and (10) time and cost constraints.

Standard statistical tests are used to measure the performance of the change detection algorithm quantitatively and are supported by visual assessment of the geographical areas. Change detection methods are divided into multi-temporal and hyper-temporal change detection methods. Change detection methods operating on multi-temporal images require only a few images; usually in the order of 2–5 images of the same geographical area. Change detection methods operating on hyper-temporal images usually requires hundreds of images taken at regular constant intervals; usually 8–30 days between acquisitions.

Most change detection methods found in the literature can either provide change information or a change alarm [93, 94]. A change alarm uses a threshold to provide binary *change*/*no change* information from the images. A change information algorithm uses post-classification to provide a *from-to* change.

Multi-temporal change detection methods evaluate local patterns in the reflectance values between images to detect change. The change detection method should compensate for the difference in environmental conditions, illumination conditions, and local trends in each of the images [95]. Multi-temporal change detection methods are grouped into several categories [24]: (1) algebra, (2) transformation, (3) classification, (4) advanced models, (5) Geographical Information System approach (GIS), (6) visual analysis, and (7) other methods.

The algebraic approach entails methods such as [24]: image differencing, image regression, image ratioing, index differencing, trajectory vector analysis, and background subtraction [93, 94]. These methods have low complexity and use manually adjusted thresholds to define change in the local vicinity.

The advantage of using an algebraic approach is the ease of interpreting the execution of the method. Another advantage is that it can operate on data sets which were captured in different environmental conditions. The disadvantage of these methods is that they have the potential to enhance the system noise, which effectively reduces the methods' performances. Another disadvantage is the setting of the threshold. The threshold has to be manually adjusted for each new data set. The methods are sensitive to features with little separability or features that are subjected to external events or time dependence.

The transformation approach uses methods to reduce the number of dimensions in the remote

sensing reflectance data set to create a new manifold [24]. The advantage of this approach is the removal of redundant dimensions and it puts emphasis on the information-carrying components [96, 97]. This approach includes transformation algorithms such as principal component analysis (PCA), Gramm-Schmidt, Chi-square, independent component analysis, etc. The disadvantage is the interpretation of the new manifold and the change trajectory of the geographical area.

The classification approach is characterised by classification methods such as: spectral combined analysis, expectation-maximisation (EM) algorithm, hybrid classification, hierarchical classification, and artificial neural networks (ANN). These methods require initial training on a set of labelled pixels. Afterwards the method is applied using the information gathered to classify a set of unknown labelled pixels. The advantage of using such a classification method is that it provides a change information matrix. These methods are robust to external environmental conditions [8, 98]. The disadvantage is the dependency on periodic updating of the training data sets.

The advanced model approach transforms the spectral reflectance values from multi-temporal spectral reflectance values to physical process parameters. The advantage is that the extracted process parameters are easier to interpret than the spectral reflectance values [99, 100]. Methods commonly used in this category are: Linear Spectral Mixture Analysis (LSMA), Li-Strahler reflectance model, spectral mixture models, and biophysical parameter estimation [24]. The disadvantage is finding a suitable model for the conversion and the intensive procedure of converting the reflectance values.

The GIS-based approach uses a GIS system to analyse satellite imagery. The advantage of a GIS system is the ability to incorporate several different layers of meta-data and satellite images for analysis [101]. The disadvantage is that different data sets have different product quality standards and when used together will degrade the results of the overall performance [24].

Visual interpretation of images can exploit the full capabilities of a remote sensing analyst's experience and knowledge. A skilled analyst can compensate for environmental conditions when looking for change [102]. The disadvantage of this approach is the processing time, and labour cost required for large geographical areas and the variability of skill level of the analyst.

There are many different change detection methods that cannot be grouped into the afore-mentioned categories. These methods produce new approaches to the field of change detection and have their associated advantages and disadvantages [103–105].

Land cover change is a function of time and can be abrupt or gradual. The ability to detect the difference between abrupt and gradual change is based on the temporal acquisition rate, the change detection method and the number of acquisitions.

Gradual change is defined as the slow change from one type of land cover to another. For example, settlement expansion is the process of clearing the indigenous vegetation and constructing a new human

settlement, which could take several months. Abrupt change is defined as a fast change in land cover type, for example, wild fire that can destroy all the natural vegetation in an area within a few hours [106].

Multi-temporal change detection methods flag all their land cover changes as abrupt. Previous studies have shown that multi-temporal change detection methods' performance is limited by the differences produced in the seasonal growth of vegetated areas [107]. Variations in surface reflectance values are observed in vegetated areas when the images are acquired at different times of the intra-annual growth cycle [19]. These phenological cycles induce variations that could raise the false change detection rate, as they are flagged as land cover change when it is only a natural seasonal variation. To overcome this limitation, a high temporal acquisition rate is required to capture the seasonal variations of a particular land cover [108]. This motivates the investigation into hyper-temporal change detection methods, as these methods can distinguish between phenological cycles, gradual and abrupt change [106].

Hyper-temporal change detection methods are used on multiple images acquired from a satellite with a short periodic revisit cycle and can be used to complement a multi-temporal change detection method [109]. The hyper-temporal acquisition rate provides continuous monitoring of the Earth, and is not limited by the availability of costly high-resolution images. This is used to augment information about which areas should rather be tasked for acquisition of high spatial resolution imagery. For example, a hyper-temporal change detection method maps the geographical areas with the highest probability of land cover change at low costs, after which a costly high-resolution image is acquired to confirm the change.

2.8.1 Hyper-temporal change detection methods

Majority of the change detection methods found in the literature are based on medium to high spatial resolution multi-temporal image analysis [18, 20]. Certain multi-temporal change detection methods can be extended to hyper-temporal images by applying the methods sequentially to subsets of multi-temporal images. The approaches that have been extended for the hyper-temporal case are: image differencing [110], image regression [111], image ratioing [112], index differencing [113], Principle Component Analysis (PCA) [75, 76], and Change Vector Analysis (CVA) [114].

These multi-temporal change detection studies rely on bi-temporal and trajectory analysis [20, 21, 24] and the data are mostly treated as hyper-dimensional, but not necessarily as hyper-temporal. These methods therefore do not fully capitalise on the temporal dimension, which captures the dynamics of different land cover types.

Hyper-temporal change detection methods attempt to understand the underlying force structuring

the data in the time dimension by identifying patterns and trends, detecting changes, clustering, modelling and forecasting [8, 40]. Hyper-temporal change detection methods are broadly divided into three categories: regression analysis, spectrum analysis, and temporal metrics.

2.8.1.1 Regression analysis

Regression analysis is a parametric method used to model the underlying structure of the data. The parameters of the model are estimated using the data set. For example, Kleynhans *et al.* assumed the MODIS NDVI time series could be modelled as a triply modulated cosine function [30]. The parameters for this model were estimated using an EKF. A labelled data set was used to estimate the models' covariance matrices manually to improve separability between different land cover classes. The estimated parameters were evaluated to detect changes in land cover.

Regression is also used to fit time series to a hypothetical temporal trajectory [109]. A temporal trajectory is a defined map of a finite sequence of points describing the expected observed values in a time series. The goodness of fit of a particular time series is computed for a set of hypothetical temporal trajectories and is measured using least squares. A set of hypothetical temporal trajectories is derived for forest disturbance dynamics in [109], which is used to describe the type of change.

The advantage of these methods is that there is no need to set a threshold. The disadvantage of both these methods is the assumption in the form of the model or temporal trajectories. Are all the changes that could realistically occur encapsulated in the model? Is the model able to adapt by inserting more parameters or creating a larger set of hypothetical temporal trajectories?

2.8.1.2 Spectrum analysis

Spectrum analysis is the analysis of harmonic frequencies within a time series. Fourier analysis is a type of spectral analysis which uses a Fourier transform to express a time series as a sum of a series of cosine and sine waves with varying frequencies, amplitudes and phases [115, Ch. 3]. The frequency of each wave component is related to the number of completed cycles defined in the time series. In many applications, the Fourier transform of time series is used for classification and segmentation [116]. Lhermitte *et al.* proposed a classification method that only evaluates the mean and seasonal Fourier transform components. The reason for this is due to the high sampling rate of a strong seasonal component in vegetation time series [116]. These components are then clustered using a post-classification change detection method [40].

Verbesselt *et al.* proposed the BFAST (Break For Additive Seasonal and Trend) approach, which uses trend, seasonal and remainder components to detect changes in the phenological cycles of plants [106]. The seasonal component is derived using the Fourier transform and has been shown to be more

stable than a piecewise linear seasonal model [117].

The advantage of these methods is that they are not dependent on a predefined model. They extract the harmonic frequencies from the time series, which means they allow the evaluation of all frequency components. The disadvantage of these methods is that the time series is assumed to be stationary and that enough harmonic frequencies are properly sampled within the time series.

2.8.1.3 Temporal metric

A temporal metric is derived from the time series by evaluating inter-annual differences in five temporal units: annual maximum, annual minimum, annual range, annual mean and temporal vector. Spatial information can also be included in some of these temporal metrics, such as: spatial mean and spatial standard deviation. The temporal metric is compared to a predefined threshold to determine whether change has occurred.

An example of a temporal metric is the evaluation of a moving average window's standard deviation on a time series. A time series is declared as a changed area when two different windows' standard deviation significantly differ from one another [118].

Another temporal metric is known as the disturbance index. The disturbance index is used to detect large-scale ecosystem distance [119]. The disturbance index measures the ratio between annual maximum land surface temperature and annual maximum EVI to the multiple year mean annual maximum land surface temperature and multiple year mean annual maximum EVI. If the current annual maximums are significantly higher than the long-term maximum, a disturbance is flagged. The difference between the two is evaluated with a predefined threshold to categorise the level of disturbance.

The annual NDVI differencing method is another temporal metric proposed by Lunetta *et al.* [19], which calculates the difference between consecutive summation of the annual NDVI time series. The pixel is flagged as change if a certain predefined threshold is exceeded in this difference. The threshold is usually determined using standard normal distribution statistics.

The EKF change detection method is a temporal metric proposed by Kleynhans *et al.* [120], which evaluates the Euclidean distance between parameters derived with an EKF within a spatio-temporal window. The EKF fits a triply modulated cosine function to a time series to model the seasonal variations. The pixel is flagged as change if the Euclidean distance exceeds a predefined threshold.

The autocorrelation function (ACF) change detection method is a temporal metric proposed by Kleynhans *et al.* [121], which evaluates the stationarity of a time series. The ACF of a time series in question is compared to the ACF of time series that did not change in the local geographical vicinity. The pixel is flagged as change if the deviation between the two ACFs exceeds a predefined threshold.

The advantage of using a temporal metric is that it operates on the raw time series data. This enables observation of abnormal behaviour that is usually filtered out by regression and spectrum analysis. The disadvantage of using a temporal metric is the selection of the threshold and the negative impact of the additive noise in the time series has on the performance.

The noise is reduced by creating methods that operate on annual statistics, which reduces the effective time series measurements significantly. For example, an original MODIS NDVI time series for 10 years (+450 time samples) can be reduced down to only 10 annual measurements represented by a temporal metric.

2.8.2 MODIS land cover change detection product

Since the launch of MODIS, several different products have been developed (see table 2.3 on page 22 for a listing). Only a few specific change detection products have been developed for a small range of applications. Thus there is currently no operational MODIS product to detect any changes in land cover. There have been two previous attempts to create an operational land cover change detection product [122–124].

The first attempt was the MODIS land use and land cover (LULC) algorithm, which detects land cover changes at a 1 km resolution using a CVA approach [114, 124]. The direction of the change vector is compared to a predefined threshold value and when exceeded, a change is flagged. It was suggested that neural network classifiers be used on a pixel-by-pixel basis to track the probability that a specific pixel changed over time [124]. The neural network is a supervised classifier and is used to derive a parameter for land cover classification. This parameter is used to determine if the new data of a geographical area are mapped to an existing category or to create a new category for the area. The monitoring of current and previous observations are used with the land cover parameter to declare change.

The second attempt at a MODIS LULC product was the MODIS Vegetative Cover Conversion (VCC) product. The VCC product uses the first two spectral bands of MODIS at a spatial resolution of 250 m to detect any changes caused by anthropogenic activities or extreme natural events [123]. Five different change detection methods were proposed in the VCC product:

1. RED-NIR space partitioning method: A two-dimensional map is created of the brightness and greenness at two separate time intervals and is used to detect change. The brightness is computed as the mean between the first two spectral bands. The greenness is computed as the difference between spectral bands 2 and 1.

2. RED-NIR space change vector: A change vector is mapped onto a spectral space (spectral band

1 and 2) between two different dates for the same pixel. The magnitude and trajectory of the change vector between the two dates are used to determine if changed occurred.

3. Modified Δ-space threshold: Uses a polar notation to define the differences in the RED and NIR values for a pixel at two different dates. The type of change is defined by the resulting vector in the polar plane.

4. Texture thresholding: Measures a coefficient of variation within a 3x3 spatial kernel at two different times. The coefficient of variation is calculated as the ratio between the standard deviation and mean within the kernel. Change is declared when the coefficient of variation exceeds a predefined threshold.

5. Linear feature thresholding: The method computes the mean and absolute difference of a pixel value for each neighbouring pixel in a 3x3 spatial kernel. A threshold determines whether a linear feature is present.

Neither the MODIS LULC [114] nor the MODIS VCC [123] product fully capitalises on the temporal dimension, as only two dates are compared. A multi-temporal change detection method was attempted, while disregarding the potential of a hyper-temporal change detection method, which has been used successfully in other fields [125, 126]: telecommunications, voice recognition, control systems, etc. Even though one of the primary objectives before the launch of the MODIS sensors was an operational land cover change detection product, to date no operational product has been developed.

2.9 SUMMARY

In this chapter, the use of remote sensing for monitoring geographical areas was discussed. The joint investment of many international organisations and national governments has led to the creation of numerous Earth observation satellites for various different applications. The chapter focused on the importance of using satellite remote sensing to detect new human settlement development in certain regions of South Africa.

The method of choosing a satellite-based sensor was discussed by considering the spatial, spectral, radiometric, and temporal resolutions. After considering multiple factors, the MODIS sensor was chosen, followed by a detailed description of its properties, with emphasis on the benefits of the BRDF corrected data products. The chapter concluded with a review of some of the popular multi-temporal change detection methods, and expanded to the use case of hyper-temporal change detection methods.