CHAPTER 1

INTRODUCTION

Satellites provide humanity with data to infer properties of the earth that were impossible a century ago. Humanity can now easily monitor the amount of ice found on the polar caps, the size of forests and deserts, the earth's atmosphere, the seasonal variation on land and in the oceans and the surface temperature of the earth.

In this thesis satellite data are used to detect and estimate human settlement expansion. Anthropogenic changes to the environment are driven by the need to provide food, water and housing to more than 7 billion people. Unfortunately humanity's need to survive has a negative effect on the environment [8]. For example, human settlement expansion on the outskirts of Xalapa city, the capital city of the state of Veracruz in Mexico, is causing severe environmental damage in the region [9]. Monitoring the growth of settlements around the world is important as it could enable multiple governments to enforce sustainable development, which would decrease humanity's negative impact on the environment. Monitoring settlement expansion in South Africa is the primary focus of the thesis. Monitoring settlement expansion is especially important in South Africa as it is one of the most pervasive forms of land cover change found in southern Africa [10].

The chapter starts by explaining the importance of monitoring settlement expansion in South Africa (in greater detail than in the previous paragraph) and then continues by briefly introducing the techniques and the data that are used to detect settlement expansion in Section 1.2, Section 1.3 and Section 1.4. The main problem statement is given in Section 1.6, which also discusses the main contributions made by this thesis. The chapter ends with a list of publications written during the course of the study and a brief overview of the remaining chapters.

1.1 SETTLEMENT DETECTION

Human settlement expansion in South Africa is often unplanned and informal in nature, meaning that the settlements form in randomly selected places, without any provision for electricity, running water, refuse removal or water-borne sewage. These informal settlements usually develop as people move closer to employment opportunities [10].

According to a report from the nineteenth special session of the general assembly of the Unitited Nations (UN), the South African government needs to be empowered to plan, implement, develop and manage human settlements [11]. As mentioned before, predicting where human settlements will form is rather difficult. For this reason, the main focus of this thesis is to develop affordable techniques that will aid the South African government in monitoring the expansion of human settlements so that efficient decisions can be made regarding infrastructure development and resource allocation. Remote sensing provides an attractive solution to this problem, since the data of certain remote sensing sensors are free and provide large-scale monitoring capabilities. In this thesis remote sensing data will be the primary tool used to monitor settlement expansion. Two provinces of South Africa were selected as study regions, namely Gauteng and Limpopo. Gauteng was selected as it has a higher population growth rate than the remaining provinces of South Africa, which makes it an attractive study area. Limpopo was chosen due to the fact that it is a very poor province of South Africa [12]. Poverty usually leads to the formation of informal settlements.

1.2 HYPERTEMPORAL APPROACHES

Most of the remote sensing classification and change detection techniques available in literature are multi-temporal (usually bi-temporal or single date) techniques [13, 14]. In contrast, the approaches that are investigated in this thesis are all hypertemporal techniques. Hypertemporal techniques fully exploit the information located in hypertemporal time-series to classify or detect changes. A hypertemporal time-series is defined as a time-series that consists of frequent equal-spaced observations [10]. The benefit of using the temporal dimension effectively is that the date selection problem is circumvented [15]. When working with multi-temporal algorithms, selecting optimal dates is important, as class separability may be different during different seasons. Another possible benefit of hypertemporal techniques is that a time-series can provide phenological metrics, which are used for discerning between different vegetation types. There already exists a few hypertemporal classification

and change detection techniques in literature [5, 7, 10, 15–21]. The approaches cited are definitely not an exhaustive list.

1.3 DATA SELECTION

The MCD43A4 MODIS product consists of Bidirectional Reflectance Distribution Function (BRDF) corrected land surface reflectance (eight-day composite, 500 m resolution) time-series. The product was chosen to investigate the hypertemporal techniques discussed in this thesis, because the MCD43A4 product provides a long, reliable high temporal remote sensing time-series. Another benefit of the adjusted land spectral reflectance product is that it significantly reduces the anisotropic scattering effects of surfaces under different illumination and observation conditions [22]. Furthermore, MODIS data, when compared to Advanced Very High Resolution Radiometer (AVHRR) data, exhibit enhanced spectral and radiometric resolution, wide geographical coverage and improved atmospheric corrections, while preserving the same temporal resolution [15].

1.4 SEQUENTIAL APPROACHES

Sequential hypertemporal approaches are a relatively new subset of current hypertemporal remote sensing techniques that are available in literature. Up to now the focus in remote sensing has mainly been on sequential classification approaches [23]. A good literature review on the field of sequential analysis can be found in [24, 25]. Sequential approaches are threshold-based techniques. Sequential approaches keep sampling observations until an on-line statistic crosses a predefined threshold. The main advantage of sequential approaches is that on average, sequential approaches usually require fewer observations than fixed-sample-size approaches. The reason for the speed increase is that sequential approaches terminate uniquely for each observable sequence. Sequential approaches try to jointly optimise the accuracy and detection delay of the classifier or change detector. It should be clear that the accuracy of a classifier or change detector (which is the primary focus of remote sensing literature) is not the only design criterion to consider when designing classifiers and change detectors. The detection delay of a classifier or change detector is also an important design criterion [23]. Sequential classification and its application in remote sensing are studied in detail in [23]. One of the objectives of this thesis is to verify the preliminary sequential results of [23] and extend sequential analysis to the remote sensing change detection realm. Interest in sequential techniques is expressed in this thesis because, the South African government not only needs to detect settlement expansion,

but also needs to do so as quickly as possible.

1.5 INDUCTIVE SIMULATION

Sequential hypertemporal approaches usually rely on densities. As sequential approaches employ densities, they require large amounts of training data. For the current thesis, large amounts of training data are not available and therefore an efficient simulator that can augment datasets is required. Most remote sensing simulators in the literature are deductive simulators, which means that they employ the biophysical laws that govern the reflection of light [26, 27]. In contrast to deductive simulators, an inductive simulator uses a mathematical (inductive) model that is fitted directly on an existing dataset. The aim of an inductive model is to model the statistical characteristics of the original dataset and therefore it can be used for dataset augmentation.

1.6 PROBLEM STATEMENT

At this stage enough background has been discussed to formulate the fundamental problem statement of this thesis:

Problem Statement: Develop new sequential or non-sequential hypertemporal remote sensing techniques to detect settlement expansion in South Africa.

The existing techniques investigated and the contributions made by this thesis in trying to solve the above problem statement are discussed in the following three sections.

1.6.1 Existing hypertemporal techniques

The following existing hypertemporal techniques were investigated in this thesis:

1. Ackermann [23] applied sequential analysis to the remote sensing field and in doing so developed the time-varying maximum likelihood classifier. Ackermann also developed the minimum distance classifier [16].

2. Lhermitte et al. [5] showed that an efficient classifier could be created by using only the mean and seasonal harmonic components of a remotely sensed time-series. For the remainder of this thesis, the harmonic feature group proposed by Lhermitte et al. is denoted by $\boldsymbol{\iota}$.

3. Carrão et al. [15] showed that temporal features can provide good separability, which inspired the development of the temporal feature group $\boldsymbol{\zeta}$ (defined formally in Section 4.2.4.2).

4. Lunetta et al. [7] developed the band differencing algorithm (a hypertemporal change detection approach).

1.6.2 Contribution to hypertemporal classification

The following contribution was made in the hypertemporal remote sensing classification field:

1. A non-sequential hypertemporal SVM classifier was implemented. The SVM classifier uses a novel (an outcome of this thesis) noise-harmonic feature group $\boldsymbol{\theta}$ (where the symbol $\boldsymbol{\theta}$ is used to represent this noise-harmonic feature group), which is an extension (in size but also in classification capability) of the classic harmonic feature group $\boldsymbol{\iota}$ proposed in [5]. The feature group $\boldsymbol{\theta}$ is constructed from the CSHO [2]. The SVM using $\boldsymbol{\theta}$ is benchmarked against the minimum-distance classifier, the time-varying maximum likelihood classifier, an SVM classifier using the harmonic feature group $\boldsymbol{\iota}$ and an SVM using the temporal feature group $\boldsymbol{\zeta}$ (see Chapter 4 and Chapter 5 for more detail) [5, 15, 16, 23]. Generally the SVM classifier using $\boldsymbol{\theta}$ outperformed the minimum-distance classifier, the time-varying maximum likelihood classifier, the SVM classifier using the harmonic feature group $\boldsymbol{\iota}$ and the SVM using the temporal feature group $\boldsymbol{\zeta}$. The performance results of the new hypertemporal technique were published in [2].

1.6.3 Contribution to hypertemporal change detection

The following contributions were made to the hypertemporal remote sensing change detection field:

1. The sequential change detection algorithm called CUSUM (windowless version) [6] was introduced into the remote sensing field and benchmarked against the popular hypertemporal approach developed by Lunetta et al. (see Chapter 4 for more detail) [7, 10, 28]. This thesis therefore builds on and extends the work done by Ackermann [23], which mainly focused on sequential detection (had a smaller scope). Windowed versions of the CUSUM algorithm have been used in a remote sensing context [28, 29]. The problem of windowed approaches, is that it

becomes important to select an optimal window length. If the window is chosen either too small or too large then the change detection capability of the approach deteriorates. The windowless CUSUM approach presented here, is more flexible as it circumvents the optimal window length issue. The results of the windowless CUSUM approach were published in [30].

2. To implement the CUSUM algorithm effectively, an inductive simulator was developed (see Section 1.5 and Chapter 4 for more detail). In selective cases, inductive statistical models similar to the one used in this thesis have been used to simulate a single time-series [31]. The complex issue of replicating multispectral correlation and dependence was not undertaken in [31]. The inductive simulator developed in this thesis accurately enforces multispectral correlation and dependence. The details of this simulator were published in [32].

1.7 PUBLICATIONS AND RELATED WORK

The following conference papers (where C# implies a conference paper) were produced during the course of the PhD study:

[C1] E.R. Ackermann, T.L. Grobler, A.J. van Zyl, K.C. Steenkamp and J.C. Olivier, "Minimum error land cover separability analysis and classification of MODIS time series data", *IEEE International Geoscience and Remote Sensing Symposium*, Vancouver, Canada, July 2011, pp. 2999–3002.

[C2] T.L. Grobler, E.R. Ackermann, J.C. Olivier and A.J. van Zyl, "Systematic Luby Transform codes as incremental redundancy scheme", *IEEE AFRICON*, Livingston, Zambia, September 2011, pp. 1–5.

[C3] E.R. Ackermann, T.L. Grobler, A.J. van Zyl and J.C. Olivier, "Belief propagation for nonlinear block codes", *IEEE AFRICON*, Livingston, Zambia, September 2011, pp. 1–6.

[C4] T.L. Grobler, E.R. Ackermann, A.J. van Zyl, W. Kleynhans, B.P. Salmon and J.C. Olivier, "Sequential classification of MODIS time series", *IEEE International Geoscience and Remote Sensing Symposium*, Munich, Germany, July 2012, pp. 6236–6239.

[C5] B.P. Salmon, W. Kleynhans , F. van den Bergh, J.C. Olivier, W.J. Marais, T.L. Grobler, K.J. Wessels, "A search algorithm to meta-optimize the parameters for an extended Kalman filter to

improve classification on hyper-temporal images", *IEEE International Geoscience and Remote Sensing Symposium*, Munich, Germany, July 2012, pp. 4974–4977.

[C6] W. Kleynhans, B.P. Salmon, J.C. Olivier, F. van den Bergh, K.J. Wessels and T.L. Grobler, "Detecting land-cover change using a sliding window temporal autocorrelation approach", *IEEE International Geoscience and Remote Sensing Symposium*, Munich, Germany, July 2012, pp. 6765–6768.

The following journal papers (where J# implies a journal paper) were published during the course of the PhD study:

[J1] T.L. Grobler, A.J. van Zyl, J.C. Olivier, W. Kleynhans, B.P. Salmon and W.T. Penzhorn, "Wu's algorithm and its possible application in Cryptanalysis", *African Journal of Mathematics and Computer Science Research*, vol. 5, no. 1, pp. 1–8, January 2012.

[J2] T.L. Grobler, E.R. Ackermann, J.C. Olivier, A.J. van Zyl and W. Kleynhans, "Land-Cover separability analysis of MODIS time-series data using a combined Simple Harmonic Oscillator and a Mean Reverting Stochastic Process", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 3, pp. 857–866, June 2012.

[J3] W. Kleynhans, B.P. Salmon, J.C. Olivier, F. van den Bergh, K.J. Wessels, T.L. Grobler, K.C. Steenkamp, "Land cover change detection using autocorrelation analysis on MODIS time series data: detection of new human settlements in the Gauteng province of South Africa", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 3, pp. 777–783, June 2012.

[J4] T.L. Grobler, E.R. Ackermann, A.J. van Zyl, J.C. Olivier, W. Kleynhans and B.P. Salmon, "Using Page's Cumulative Sum Test on MODIS time-series to detect land-cover changes", *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 2, pp. 332–336, March 2013.

[J5] T.L. Grobler, E.R. Ackermann, A.J. van Zyl, J.C. Olivier, W. Kleynhans and B.P. Salmon, "An inductive approach to simulating multispectral MODIS surface reflectance time series", *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 3, pp. 446–450, May 2013.

[J6] E.R. Ackermann, T.L. Grobler, W. Kleynhans, J.C. Olivier, B.P. Salmon and A.J. van Zyl, "Cavalieri Integration", *Quaestiones Mathematicae*, vol. 35, no. 3, pp. 265–296, September

2012.

1.8 LAYOUT OF THESIS

The outline of the thesis is as follows:

Chapter 2: The chapter provides a broad overview of the remote sensing field, which includes a brief history of remote sensing, an introduction to the physical principles behind remote sensing, an overview of remote sensing platforms and an introduction to the MODIS sensor. The MODIS data used by the classification and change detection algorithms investigated in this thesis are also presented in this chapter.

Chapter 3: The chapter provides a broad overview of the sequential analysis field. It starts with the Neyman-Pearson optimal classification result, which is the predecessor of modern sequential analysis. The chapter then continues to the field of sequential classification, which is discussed by using two frameworks, namely Wald's framework and the Bayesian framework. From sequential classification the chapter progresses to a group of statistical change detection algorithms grouped under the collective name of quickest detection. The quickest detection techniques discussed in the chapter are divided into Bayesian and non-Bayesian approaches. Two main non-Bayesian approaches are discussed, namely the CUSUM stopping time and the Shiryaev-Roberts stopping time (as well as its variants). The main reason for including this chapter is to provide the theoretical background knowledge required to implement CUSUM as a sequential hypertemporal remote sensing change detection algorithm.

Chapter 4: The chapter provides the technical details of the newly proposed algorithms as well as the benchmarking sequential and non-sequential hypertemporal classification and change detection algorithms investigated in the thesis. The details of the inductive simulator developed for the CUSUM algorithm are also found in this chapter. Furthermore, the chapter contains literature reviews of remote sensing classification and change detection.

Chapter 5: The chapter starts with preliminary data analysis results obtained from the test datasets. These results can be used to predict the performance of the different classification and change detection approaches. The chapter then gives the classification and change detection accuracies and rankings of the different sequential and non-sequential hypertemporal classification and change de-

tection algorithms investigated in the thesis.

Chapter 6: In this chapter the main conclusions from Chapter 5 are summarised.

CHAPTER 2

REMOTE SENSING

The main aim of this chapter is to introduce two datasets. All the hypertemporal techniques presented in Chapter 4 are applied to these two datasets. The two datasets are discussed in Section 2.8. The first few sections of this chapter however introduce the basic principles of remote sensing, as well as the MODIS sensor. These introductory sections are needed to understand the content of the datasets in Section 2.8.

Remote sensing is the science of converting data about the earth's surface, recorded with remote (distant) sensor platforms, into usable information. The remote sensors archive how the earth's surface reflects or transmits electromagnetic energy at different wavelengths and thus record an electromagnetic spectral signature of the earth's surface [33].

2.1 HISTORY OF REMOTE SENSING

It can be argued that the moment in time which gave birth to photography was in fact also the starting point of spaceborne remote sensing. Photography was invented in 1839. Those early photographs were created by the photographic processes of Nicephore Niepce, William Henry Fox Talbot, and Louis Jacques Daguerre [33]. The first aerial photograph was taken (of Bievre, France) by a Parisian photographer named Gaspard Felix Tournachon (from a balloon). The earliest *existing* aerial photograph was taken from a balloon over Boston in 1860 by James Wallace Black and immortalised by Oliver Wendell Holmes [33]. The First and Second World War sparked the widespread use of aerial photography as a surveillance tool. The use of aerial photography for environmental purposes only became popular after the Second World War [10]. The term "remote sensing" was first coined by Evelyn Pruit after recognising that "aerial photography" no longer accurately described the different

images recorded, as some images (at that point) were recorded by using wavelengths outside the visible spectrum [34]. The next step in the evolution (starting in the 1960s) of remote sensing was when humans started using spaceborne platforms to house remote sensing sensors. The space era, which is still continuing, can be discussed under four headings, namely *military reconnaissance satellites*, *manned space flight*, *meteorological satellites* and *earth resource satellites* [35].

2.1.1 Military reconnaissance satellites

Before 1960, the United States of America (USA) and the former Union of Soviet Socialist Republics (USSR) used aerial photographs to keep track of each other's military capability. However, at the Surprise Attack Conference in Geneva (in 1958) it was proposed for the first time to use satellites to gather military information [35]. CORONA was one of the first military programmes under which satellites were launched into space to perform military reconnaissance (active during the 1960s). These missions were usually very short in duration, typically no longer than one or two weeks. These early systems were constrained, since they could only carry a limited amount of film. The film canister was ejected and picked up as it descended to earth [35]. Later systems could store images in digital format and transported the data to the earth via telemetry.

2.1.2 Manned space flight

On April 12, 1961 Yuri Gagarin became the first person to orbit the earth. Although no photos were taken during this mission, it became apparent that spaceborne earth observation had great potential. The USA also started manned space missions in the 1960s, culminating in the moon landing in 1969 during the Apollo programme [35]. The Mercury (1961-1963), Gemini (1965-1966) and Skylab (1973-1974) programmes were some of the manned American programmes that captured pictures of the earth. The Russians also conducted their own manned missions, which included the Vostok and Voskhod programmes, which were analogous to the Mercury and Gemini missions [35].

2.1.3 Meteorological satellites

Meteorological satellites (weather satellites) paved the way for the modern earth resource satellites. Meteorological satellite, TIROS-1, was the first satellite that was used for earth observation and was launched by the USA on April 1, 1960 [35]. Both *polar orbiting* and *geostationary satellites* are used for weather prediction. A geostationary satellite completes its orbit every 24 hours, so that it can

always monitor one specific place on earth and is usually found at a higher altitude than polar orbiting satellites. Polar orbiting satellites do not complete their orbit in 24 hours and can therefore survey the entire surface of the earth [36].

There are a few polar orbiting satellite programmes worth highlighting [36]:

1. ITOS/NOAA or POES: ITOS-1 was launched on January 23, 1970, while NOAA-1 was launched on December 11, 1970. It is noteworthy to mention that NOAA-6 (launched on June 27, 1979) contained the first in a series of AVHRRs, the predecessor of MODIS. The ITOS/NOAA programme is administrated by the National Oceanic and Atmospheric Administration (NOAA). The most recent Polar-orbiting Operational Environmental Satellite (POES) launched is NOAA-19, launched on June 2, 2009.

2. Nimbus: Nimbus-1 was launched on August 28, 1964. Nimbus-7 (launched on June 27, 1979) carries the Coastal Zone Color Scanner (CZCS), the Total Ozone Mapping Spectrometer (TOMS) and the Scanning Multichannel Microwave Radiometer (SMMR). The Nimbus satellites were put into space by the National American Aeronautics and Space Administration (NASA).

There are four important geostationary satellite programmes, that together provide complete coverage (weather) of the globe, namely [36]:

1. Meteosat: Meteosat-1 was launched on November 23, 1977. The Meteosat programme is administrated by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) and covers Europe and Africa.

2. GOES: GOES-1 was launched on October 16, 1975. The Geostationary Operational Environmental Satellites (GOESs) are operated by the National Environmental Satellite Data and Information Service (NESDIS) and have been developed by NOAA. There are two main GOES satellites in use, GOES-W, which services the western Americas and the Atlantic Ocean, and GOES-E, which covers the eastern Americas and the Pacific.

3. Indian INSAT: INSAT-1B was launched on August 30, 1983. The Indian National Satellite System (INSAT) satellites where launched by the Indian Space Research Organisation (ISRO) and provides coverage of India and the Indian Ocean.

4. Japanese GMS: GMS-1 was launched on July 14, 1977. The Geostationary Meteorological Satellite (GMS) programme is driven by the Japan Meteorological Agency and covers South-East Asia and Japan.

The coverage areas for each of the geostationary weather satellites are displayed in Figure 2.1.

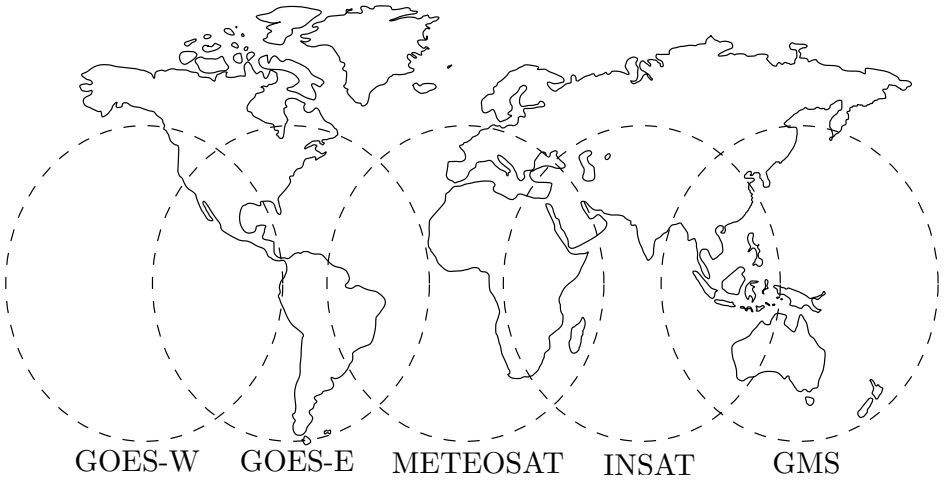


Figure 2.1: Worldwide coverage by international geostationary weather satellites.

2.1.4 Earth resource satellites

While weather satellites have been monitoring the earth's atmosphere since 1960 and have largely been considered useful, there was no real appreciation of land data from space before the development of earth resource satellites. The development of earth resource satellites can be divided into three generations. The first generation is mainly characterised by the fact that basic remote sensing sensors were used. The Landsat and Système Probatoire d'Observation de la Terre (SPOT) satellites are prime examples of first phase earth resource programmes. It can be argued that a second generation began with the inception of the Earth Observing System (EOS) programme, which hailed in the era of sophisticated remote sensing sensors that could survey the earth and allow humans to track climate change. The EOS was part of NASA's Mission to Planet Earth (MTPE), now called the Earth Science Enterprise (ESE). It is important to note that these generations are not mutually exclusive, and that some intersection does occur. At the moment earth resource satellite development is moving into a third generation that started with the launch of NPP. The third phase will be characterised by the use of remote sensing sensors that are far more advanced (giant leap) than the sensors housed in, for instance, Terra. These sensors will have the explicit goal of achieving the original vision of EOS,

which is to collect a 15-year global data set to address questions on climate change.

2.1.4.1 First generation

The idea of a civilian satellite that could be used for scientific earth surveillance was proposed in 1965 by William Pecora, director of the United States Geological Survey (USGS), and was inspired by the photographs taken on the Mercury, Gemini and Apollo missions in the 1960s. Unfortunately this idea was met with heavy criticism from the Bureau of Budget (BOB) and the Department of Defense (DOD), since the BOB thought high-altitude aircraft would be better suited to the task and the DOD was concerned that earth resource satellites would jeopardise its military reconnaissance missions.

In 1966 NASA felt pressure from the Department of the Interior (DOI), after USGS convinced the DOI to announce that the DOI would be starting its own earth resource satellite programme. This forced NASA to accelerate the building of an earth resource satellite. Unfortunately, a limited budget and sensor disagreements between application agencies again delayed the satellite construction process. Finally, by 1970 NASA received authorisation to build a satellite. The first earth resource satellite, Landsat-1, was launched on July 23, 1972 by NASA; at that time the satellite was known as the Earth Resources Technology Satellite (ERTS) [33]. Landsat-1 carried the Return Beam Vidicon (RBV) and Multispectral Scanner (MSS) systems. Seven satellites were launched in the Landsat series, some of which are still functioning today. Landsat-4, launched on July 16, 1982, carried the Thematic Mapper (TM), another predecessor of MODIS [33].

A few other earth resource satellite programmes (started as part of the first generation) worth mentioning are [35]:

1. SEASAT: SEASAT was managed by NASA's Jet Propulsion Laboratory and was launched on June 27, 1978. SEASAT had on board the first spaceborne Synthetic Aperture Radar (SAR), but unfortunately only functioned for three months.

2. SPOT: SPOT is a high-resolution, optical imaging earth observation satellite programme managed by Spot Image based in Toulouse, France. SPOT-1 was launched on February 22, 1986.

3. IRS: IRS are a series of earth observation satellites, built, launched and maintained by ISRO. IRS-1A was launched on March 17, 1988.

4. JERS: JERS-1 was launched on February 11, 1992 and was administrated by the Japan Aerospace Exploration Agency (JAXA).

2.1.4.2 Second generation

In the early 1980s, there was a merger between the human spaceflight missions and the earth science missions of NASA, which was termed System Z. System Z fostered the idea of having one giant spacecraft carrying a variety of sophisticated earth observation sensors, including radar. System Z changed into the EOS in 1983, after scientists realised that multiple small missions would lead to better results [37, 38]. In the beginning the System Z earth observation system was to consist of two large (15-ton) platforms called EOS-A and EOS-B. After a reduction in size, the original sun-synchronous system EOS-A was renamed to EOS-Terra to emphasise its main function, namely to make land observations. EOS-Terra was launched on December 18, 1999. EOS-B was renamed to EOS-Aqua, as EOS-B would focus on ocean observation, and was launched on May 4, 2002. EOS-Terra carries the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), the Multi-angle Imaging SpectroRadiometer (MISR), MODIS, the Measurements of Pollution in the Troposphere (MOPITT) and the the Clouds and the Earth's Radiant Energy System (CERES). EOS-Aqua carries the Advanced Microwave Scanning Radiometer-EOS (AMSR-E), the Atmospheric Infrared Sounder (AIRS), the Advanced Microwave Sounding Unit (AMSU), CERES, the Humidity Sounder for Brazil (HSB) and MODIS. Many EOS missions have been launched over the last decade [37, 38].

2.1.4.3 Third generation

During the mid 1990s a series of decisions were taken that NASA's EOS programme would only be regarded as a proof of concept and that NOAA would eventually be responsible for developing systems to study climate change. At about the same time the USA government decided to combine the low earth orbiting satellite programmes of NOAA and the DOD into the National Polar Orbiting Environmental Satellite Series (NPOESS). The responsibility of administrating NPOESS was assigned to the newly created Integrated Program Office (IPO) consisting of NASA/NOAA/DOD [37, 38]. The NPP satellite was originally developed by the IPO, until DOD participation in the project was dissolved. The NPOESS Preparatory Project (NPP) satellite is intended to bridge the gap between old (Terra) and new systems (still to be launched) and was launched on October 28, 2011. The NPP sa-

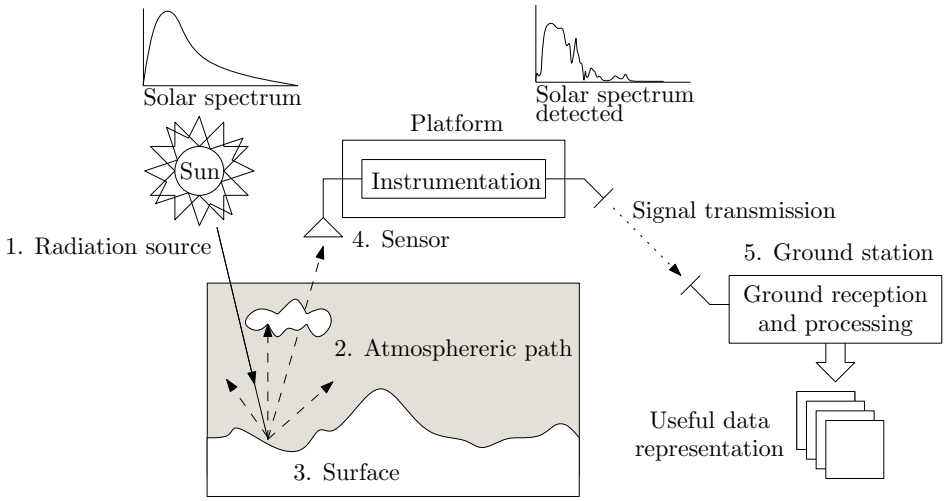


Figure 2.2: Signal and data flow in a typical remote sensing system (from [23]).

tellite carries five instruments, including the Advanced Technology Microwave Sounder (ATMS), the Cross-track Infrared Sounder (CrIS), CERES, the Visible Infrared Imager Radiometer Suite (VIIRS) and the Ozone Mapping and Profiler Suite (OMPS). MODIS is a predecessor of VIIRS. The Landsat Data Continuity Mission (LDCM) and Hyperspectral Infrared Imager (HyspIRI) are two new developments that will also form part of the third generation [37, 38].

2.2 A TYPICAL REMOTE SENSING SYSTEM

A general remote sensing platform is depicted in Figure 2.2 and consists of five main parts, namely the radiation source, the atmospheric path, the surface, the remote sensor and the ground reception station [39].

The sun is arguably the best known and most widely used source of electromagnetic energy and its energy is distributed throughout the electromagnetic spectrum. The electromagnetic energy from the sun travels through the atmosphere towards the surface of the earth. When the electromagnetic energy travels through the atmosphere, some of the energy is absorbed or scattered. The remaining energy arrives at the surface of the earth, where the energy is absorbed, reflected or transmitted. The absorbed energy can be re-emitted at a different wavelength. The reflected and emitted wavelengths travel back through the atmosphere towards the remote sensing sensor, where it is finally recorded. The atmosphere again absorbs and scatters some of the reflected and emitted energy. In the last step the recorded data are sent to a ground station where the data are processed to create useful

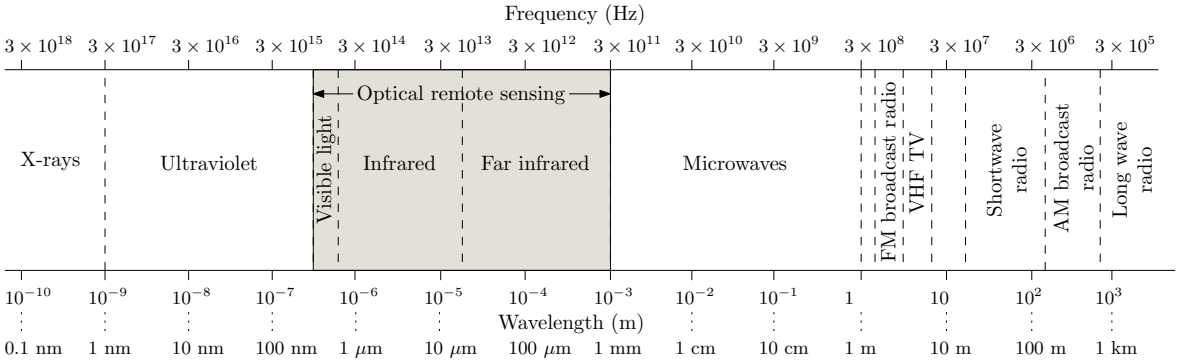


Figure 2.3: The electromagnetic spectrum, showing the region of interest for optical remote sensing (from [23]).

information [39].

2.3 ELECTROMAGNETIC RADIATION

The most important principles of electromagnetic radiation are discussed in this section. The section starts by introducing the electromagnetic spectrum and is followed by a section that explains how electromagnetic radiation propagates. Radiation units are discussed in Section 2.3.3, while Section 2.3.4 explains what blackbody radiation is.

2.3.1 Electromagnetic spectrum

In Figure 2.3 a segment of the electromagnetic spectrum is shown. Electromagnetic spectrum divisions were created for convenience and by tradition for each discipline and is therefore defined differently in other sources [34]. The ultraviolet, visible, infrared and microwave regions are usually used for remote sensing purposes.

Near ultraviolet radiation is known for its ability to induce fluorescence, emission of visible radiation, in some materials. Unfortunately ultraviolet radiation is severely scattered by the atmosphere and therefore not used very often in a remote sensing context [34].

The visible and infrared region together form the optical region [23]. The optical region is usually divided further into smaller regions. However more than one division are, used in literature, as shown in Table 2.1. The near infrared and mid-infrared regions are close to the visible region and have similar characteristics to visible light, and for this reason can be recorded via films, filters and cameras. The

far infrared region is reasonably far removed from the visible region. In everyday terminology this region is known as the thermal region, consisting of "heat" [34].

Table 2.1: Some common optical regions of the electromagnetic spectrum.

	Region	Wavelength (μm)
Visible	Blue	0.4 – 0.5
	Green	0.5–0.6
	Red	0.6–0.7
Infrared [23]	Near IR	0.7–1.4
	Short-wave IR	1.4–3.0
	Mid-wave IR	3.0–8.0
	Long-wave IR	8.0–15.0
	Far IR	15.0–1000
Infrared [35]	Photographic IR	0.7–0.9
	Very near IR	0.7–1.0
	Reflected IR	0.7–3.0
	Near IR	0.7–3.0
	Thermal IR	3.0–1000

The microwave region is usually used in an active remote sensing system.

2.3.2 Propagation of electromagnetic radiation

Electromagnetic radiation is produced through several means, including changes in the energy levels of electrons, acceleration of electrical charges, decay of radioactive substances and the thermal motion of atoms and molecules [34]. Electromagnetic radiation propagates by means of a transverse wave. Electromagnetic radiation consists of a perpendicular electric (E) and a magnetic field (H) that increase and decrease in phase with each other [35]. Transverse waves have a few important properties, namely [34]:

1. Wavelength (λ) is the distance between two successive peaks. Wavelength can be measured in everyday units of length, but the wavelengths of the electromagnetic radiation that is relevant to most remote sensing sensors are so short that less known units are usually employed, which

include the micrometre (μm: 10^{-6}) and the nanometre (nm: 10^{-9}).

2. Amplitude (A) is equal to the height of each peak. Amplitude is often measured as spectral irradiance (W.m^{-2}.μm^{-1}), expressed as watts per square metre per micrometre (as energy levels per wavelength interval).

3. Frequency (f) is measured in Hz and is defined as the number of crests passing a fixed point in a second.

All matter above 0 K produces electromagnetic radiation, and all electromagnetic radiation travels at the speed of light, $c = 2.99893 \times 10^8$ ms^{-1}. Because all electromagnetic radiation travels at the same speed, an inverse relation exists, between the wavelength and frequency of electromagnetic radiation, which is expressed mathematically as [35]

$$c = \lambda f. \tag{2.1}$$

In Equation 2.1, λ is measured in m and f is measured in Hz. Equation 2.1 explains why a transverse wave with a longer wavelength has a lower frequency when compared to a transverse wave with a shorter wavelength wave.

2.3.3 Radiation units

Although many electromagnetic radiation characteristics can be explained eloquently through wave theory, another theory offers useful insights when describing how electromagnetic energy reacts with matter. This theory, called the particle theory, states that electromagnetic radiation is absorbed and emitted in units called photons or quanta. The energy of a quantum is given by [33]

$$Q = hf,$$

where Q represents the energy of a quantum in joules (J), h is Planck's constant equal to 6.626×10^{-34} J.s and f represents frequency measured in Hz.

The rate $\frac{dQ}{dt}$ at which photons strike a surface is called radiant flux Φ measured in watts (W). Radiant exitance (M) and irradiance (E) are defined as $\frac{d\Phi}{dA}$, where A denotes area measured in m^2. The difference between radiant exitance and irradiance is that radiant exitance refers to the rate at which photons are emitted from a unit area, while irradiance refers to the rate at which photons strike a unit area. Spectral radiant exitance (M_λ) and spectral irradiance (E_λ) differ from M and E in that they des-

cribe how the energy is distributed with respect to wavelength across the electromagnetic spectrum and is therefore defined as $\frac{dM}{d\lambda}$ and $\frac{dE}{d\lambda}$ respectively and measured in W.m^{-2}.μm^{-1} [39].

The radiometric units introduced up to this stage, take into account energy, time, wavelength and area. A variable that is still unaccounted for is the viewing angle and radiance L thus takes into account the viewing angle and is defined as

$$L = \frac{d^2\Phi}{dAd\Omega\cos\theta},$$

where L is the observed or measured radiance (W.m^{-2}.sr^{-1}) in the direction θ and Ω is the solid angle (sr) subtended by the observation or measurement. As in the case of radiant exitance and irradiance, radiance also has a spectral counterpart called spectral radiance $L_\lambda = \frac{dL}{d\lambda}$ [39].

2.3.4 Blackbody radiation

A blackbody is an ideal body which, if it existed, would be a perfect absorber and a perfect radiator, absorbing all incident radiation, reflecting none, and emitting radiation at all wavelengths [39]. In remote sensing, the exitance curves of blackbodies at various temperatures can be used to model naturally occurring phenomena such as solar radiation and terrestrial emittance. The spectral radiant exitance M_λ (measured in W.m^{-2}.μm^{-1}) of a blackbody for different temperatures is described through Planck's law [39]

$$M_\lambda = \frac{\varepsilon c_1}{\lambda^5(e^{c_2/\lambda T} - 1)}, \tag{2.2}$$

where ε is emittance (dimensionless), c_1 is the first radiation constant and is equal to 3.7413×10^8 W.μm^4.m^{-2}, λ is radiation wavelength with units μm, c_2 is the second radiation constant, which is equal to 1.4388×10^4 μm.K, and T is the absolute radiant temperature in K.

Emittance (emissivity) is the ratio of the radiation given off by a surface to the radiation given off by a blackbody at the same temperature; a blackbody has an emissivity of 1, while a whitebody (perfect reflector) has an emissivity of 0. All other objects (greybodies) have an emissivity between 0 and 1 [39].

Alternatively, Planck's law can be described in terms of the radiation frequency f by using the follo-

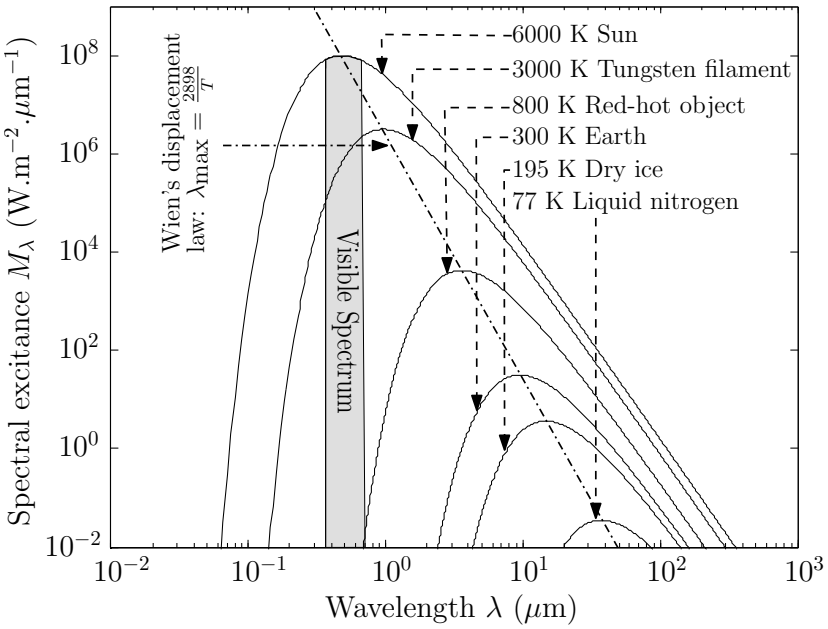


Figure 2.4: Blackbody radiation at various temperatures

wing substitutions

$$M_\lambda d\lambda = -M_f df$$
$$\lambda = \frac{c}{f}$$
$$\frac{d\lambda}{df} = -\frac{c}{f^2}.$$

After the above substitutions have been made, Equation 2.2 changes to

$$M_f = \frac{\varepsilon c_1 f^3}{c^4(e^{c_2 f/cT} - 1)}. \tag{2.3}$$

If Equation 2.3 is integrated over all frequencies, the radiant exitance M (measured in W.m^{-2}) will be obtained for a blackbody. That is

$$M = \int_0^\infty M_f df = \int_0^\infty \frac{\varepsilon c_1 f^3}{c^4(e^{c_2 f/cT} - 1)} df. \tag{2.4}$$

If $x = \frac{c_2 f}{cT}$ and $dx = \frac{c_2}{cT} df$ are substituted into Equation 2.4 the following is obtained

$$M = \frac{\varepsilon c_1 T^4}{c_2^4} \int_0^\infty \frac{x^3}{e^x - 1} dx \tag{2.5}$$

$$= \frac{\varepsilon c_1 T^4 \zeta(4)\Gamma(4)}{c_2^4} \tag{2.6}$$

$$= \frac{\varepsilon c_1 \pi^4}{15 c_2^4} T^4$$

$$= \varepsilon \sigma T^4, \tag{2.7}$$

where σ is the Stefan-Boltzmann radiation constant, which is equal to 5.6693×10^{-8} W.m^{-2}.K^{-4} and T is absolute temperature measured in K [39].

Equation 2.5 and Equation 2.6 are equal, since it is a well-known fact that $\int_0^\infty \frac{x^{n-1}}{e^x-1} dx$ is equal to the product $\zeta(n)\Gamma(n)$ for all $n \in \mathbb{N}$, where $\zeta(n) = \sum_i^\infty \frac{1}{i^n}$ is the well-known Riemann zeta function and $\Gamma(n) = (n-1)!$ is the gamma function [39, 40]. Equation 2.7 is known as the Stefan-Boltzmann radiation law, which states that the total radiation emitted from a blackbody is proportional to the fourth power of its absolute temperature [34].

If Equation 2.2 is differentiated with respect to wavelength, the derivative is set to 0, and the resulting equation solved $\lambda_{\max}$ (measured in μm) is obtained, which is the wavelength at which maximum emittance occurs for a given absolute temperature [39]. The result of this procedure is Wien's displacement law [34]

$$\lambda_{\max} = \frac{2898}{T}. \tag{2.8}$$

In Equation 2.8 the constant 2898 is measured in μm.K, while T represents absolute temperature, which is measured in K. Wien's displacement law states that the wavelength of maximum emittance for a blackbody is inversely proportional to absolute temperature [34].

2.4 ATMOSPHERIC INTERACTIONS

The atmospheric path is a critical component of any remote sensing system. When electromagnetic radiation travels through the atmosphere a lot of scattering and absorption takes place. Scattering alters the direction in which the electromagnetic radiation propagates, while absorption leads to attenuation in signal strength. When designing remote sensing systems it is of great importance to keep the effect of scattering and absorption in mind, as it would serve no purpose to record electromagnetic radiation in an atmospheric absorption window. Absorption and scattering are caused by particles and

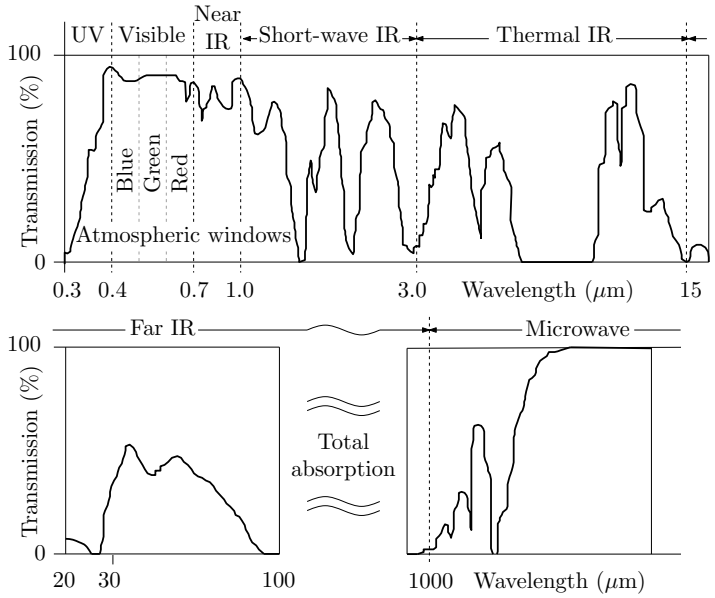


Figure 2.5: Atmospheric electromagnetic transmission windows (from [23]).

gases contained in the atmosphere.

2.4.1 Atmospheric absorption

As photons collide with atmospheric molecules, some of the radiation is absorbed through electron orbital transitions and induced vibrations, heating up the atmosphere. Nitrogen, oxygen, carbon dioxide, ozone and water vapour all absorb electromagnetic radiation at different wavelengths. The set of frequencies that a gaseous mixture can absorb consists of the union of all the the frequencies that the constituent gases of the gaseous mixture can absorb. The atmosphere contains nitrogen, oxygen, carbon dioxide, ozone and water vapour, therefore the net effect of this gaseous mixture in the atmosphere is atmospheric absorption windows. The parts of the spectrum that are not affected heavily by absorption (in which the transmission of electromagnetic radiation is high) are known as atmospheric transmission windows [23]. The atmospheric transmission windows are displayed in Figure 2.5.

2.4.2 Atmospheric scattering

Scattering is mainly caused when electromagnetic energy is redirected from its original propagation path via particulates or large gas molecules [23]. There are three basic types of scattering taking place in the atmosphere that affect electromagnetic radiation, namely *Rayleigh*, *Mie* and *nonselective*

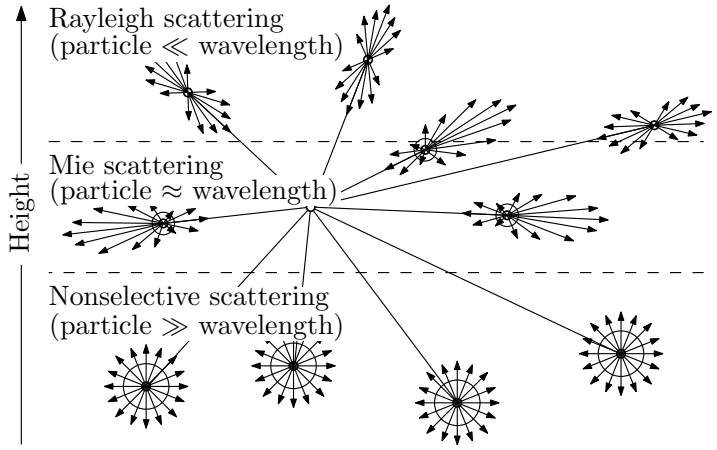


Figure 2.6: Atmospheric scattering (from [23]).

scattering, as shown in Figure 2.6.

2.4.2.1 Rayleigh scattering

Rayleigh scattering takes place at high altitudes, where the radiation wavelengths are much larger than the size of the particulates. In Rayleigh scattering, the volume-scattering coefficient σ_λ (with units cm^{-1}) is given by

$$\sigma_\lambda = \frac{4\pi^2 NV^2}{\lambda^4}\frac{(n^2-n_0^2)^2}{(n^2+n_0^2)^2}, \tag{2.9}$$

where N is the number of particles per cm^3, V is the volume of scattering particles (cm^3), λ is the radiation wavelength measured in cm, n is the refractive index of particles and n_0 is the refractive index of the medium. From Equation 2.9 it is clear that the scattering coefficient is proportional to the inverse fourth power of wavelength and this causes the shorter blue wavelengths to be scattered toward the ground much better than the longer red wavelengths, which makes the sky appear blue. As the sun approaches the horizon, the rays of the sun follow a longer path through the atmosphere, which in turn leads to an increase in blue wavelength scattering, leaving only the red wavelengths to reach the human eye (making a sunset appear orange/red) [39]. The primary components responsible for scattering at these altitudes include atmospheric gases such as oxygen and nitrogen or tiny specks of dust. Rayleigh scattering is symmetrical, with equal amounts of forwardscatter and backscatter [23].

2.4.2.2 Mie scattering

Mie scattering occurs closer to the ground than Rayleigh scattering, where the diameter of the particulates is about the same as the wavelength of radiation. For the most universal situation, in which there is a continuous particle-size distribution, the Mie scattering coefficient σ_λ (with units km^{-1}) is given by the following relationship

$$\sigma_\lambda = 10^5 \pi \int_{a_1}^{a_2} N(a) K(a,n) a^2 da,$$

where $N(a)$ is the number of particles in the interval a to $a+da$ (cm^{-3}), $K(a,n)$ is the scattering coefficient (cross-section measured in cm^{-1}), a represents the radius of the spherical particles (in cm) and n is the index of the refraction of particles [39]. Aerosols, dust particles, pollen, smoke and water vapour are the main causes of Mie scattering. Mie scattering is not as dependent on wavelength as Rayleigh scattering and mainly affects the visible spectrum. As can be seen from Figure 2.6, Mie scattering mainly causes forward scattering [23].

2.4.2.3 Nonselective scattering

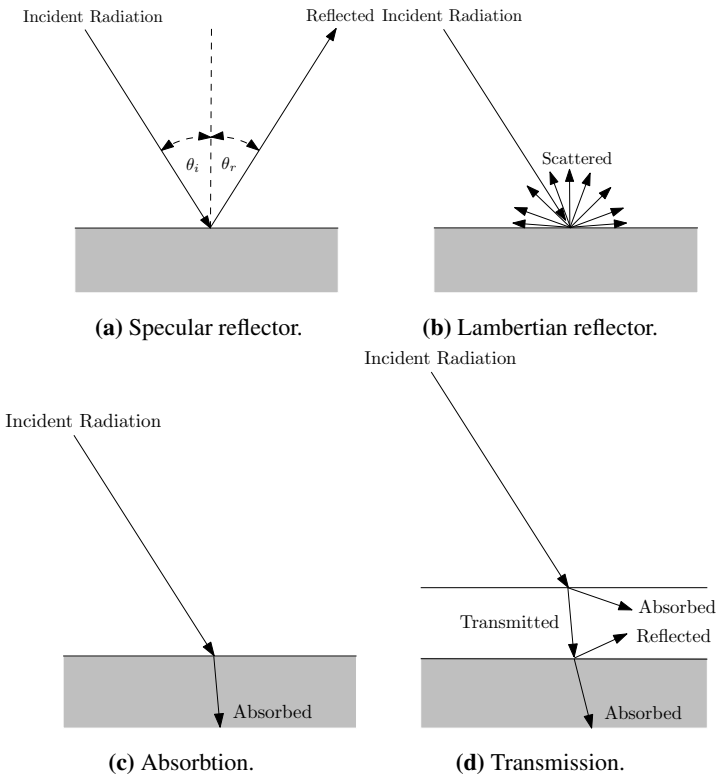
Nonselective scattering occurs at low altitudes, where the particles are usually much larger than the wavelength of radiation. Nonselective scattering scatters electromagnetic radiation uniformly and is not really dependent on wavelength. This kind of scattering is caused by large particulates such as dust, water droplets, ice crystals and hail. Since nonselective scattering scatters electromagnetic radiation uniformly, it is responsible for the fact that clouds appear white [23].

2.5 SURFACE INTERACTION

The previous section focused on the scattering and absorption effects of the atmosphere. In this section a closer look is taken at what happens to electromagnetic radiation when it reaches the earth's surface.

2.5.1 Reflection, absorption and transmission

When electromagnetic energy reaches the earth's surface, the incident radiation may be absorbed, reflected or transmitted [35]. The three phenomena introduced above are displayed in Figure. 2.7.



(a) Specular reflector. **(b)** Lambertian reflector.

(c) Absorbtion. **(d)** Transmission.

Figure 2.7: Interaction of electromagnetic radiation with a surface (from [10]).

Atoms and molecules contain electrons. To explain the concept of absorption, it is useful to imagine that the electrons are attached to the atoms via springs. The electrons vibrate continuously at a specific frequency, called the natural frequency. When an electromagnetic wave impedes on an atom that is vibrating at the same frequency as the frequency of the incident wave, the energy is absorbed by the atom via the resonance principle. Absorption can also occur due to electron orbital transitions and is not limited to vibration inducement. The absorbed radiation can then be re-emitted at a different wavelength. Reflection and transmission of electromagnetic waves occur because the frequencies of the incident waves do not match the natural frequencies of the objects. During reflection the atoms start vibrating for a short while, after which the energy is simply re-emitted at the same wavelength as the incident wave. In the case of transmission the radiation is passed on through the bulk of the material and emitted on the other side of the material at the same wavelength as the incident wave.

The roughness of the surface determines the type of reflection that will occur. A very smooth surface acts as a specular reflector, for which the reflection angle, θ_r, equals the incidence angle, θ_i. A very rough surface acts like a Lambertian reflector (diffuse reflector), which scatters the incident radiation

uniformly in all directions [35]. It should be clear that for remote sensing purposes a Lambertian reflector is preferred, since specular reflectors usually appear dark from most angles, as the incident radiation is not reflected uniformly [23].

2.5.2 Albedo

Absorption, reflectance and transmission are related via

$$E_I(\lambda) = E_A(\lambda) + E_R(\lambda) + E_T(\lambda), \tag{2.10}$$

due to the principle of conservation of energy. Where E_I denotes the incident radiation, E_A denotes the absorbed radiation, E_R denotes the reflected radiation and E_T denotes the transmitted radiation, with all energy components being a function of wavelength. Equation 2.10 states that all incident radiation is absorbed, reflected or transmitted. The albedo (spectral reflectance) of a surface is given by the ratio of the electromagnetic radiation reflected from a surface to the total electromagnetic radiation incident on the surface and is expressed mathematically as [33]

$$\rho(\lambda) = \frac{E_R(\lambda)}{E_I(\lambda)}.$$

2.5.3 Bidirectional Reflectance Distribution Function

There is a function that can describe the scattering characteristics of a surface much better than albedo can, namely the BRDF.

BRDF is usually denoted by the symbol f with units sr^{-1} and defined as

$$f(\theta, \phi, \theta', \phi') = \frac{dL'(\theta', \phi')}{dE(\theta, \phi)},$$

where dE is the irradiance (units W.m^{-2}), dL' is the reflected radiance (units $\text{W.m}^{-2}\text{.sr}^{-1}$), θ is the zenith angle of the radiation source, ϕ is the azimuthal angle of the radiation source and the primed angles refer to the location of the sensor. The relationship between irradiance dE and incident radiance dL is expressed mathematically as

$$\begin{aligned} dE(\theta, \phi) &= L(\theta, \phi)\cos(\theta)d\omega \\ &= L(\theta, \phi)\cos(\theta)\sin(\theta)d\theta d\phi \end{aligned}$$

where $d\omega$ is the solid angle defined as $\sin(\theta)d\theta d\phi$ [39].

2.5.4 Spectral signature of vegetation, soil and water

A graph of the spectral reflectance (albedo) of an object as a function of wavelength is termed a spectral reflectance curve (spectral signature) of the object [33]. Different types of objects have different spectral signatures and spectral signatures can therefore be used for classification. The spectral signature for three diverse types of objects are displayed in Figure 2.8.

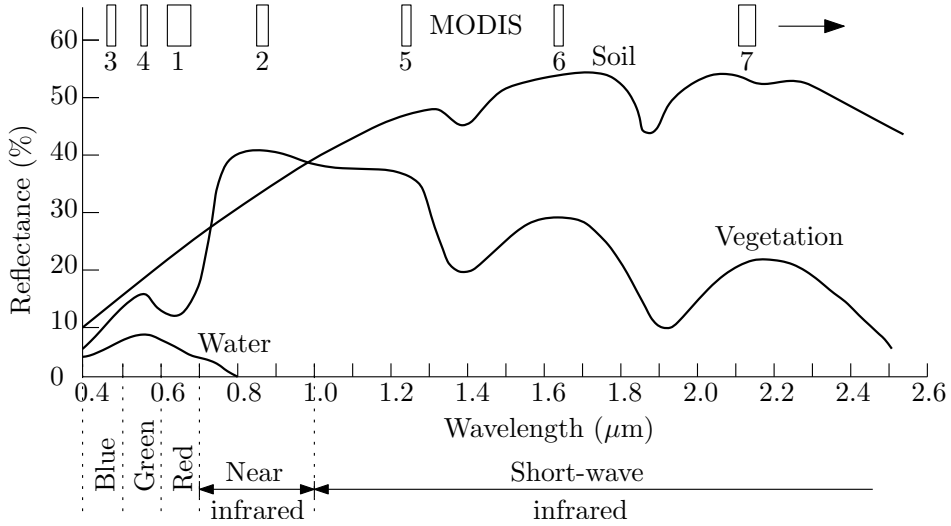


Figure 2.8: Spectral reflectance charachteristics of common earth surface materials in the visible and near-to-mid-infrared range. The positions of the MODIS spectral bands are also indicated (from [23]).

2.5.4.1 Vegetation

The spectral signature of lush green vegetation is characterised by a "peak-valley" configuration [33]. Vegetation appears green, since chlorophyll strongly absorbs radiation in the blue and red bands, while heavily reflecting radiation from the green band. As depicted in Figure 2.9, chlorophyll is a green pigment, which is contained in sacs called chloroplasts [35]. The peak in the spectral signature of the green band is clearly visible in Figure 2.8. Most of the Near-Infrared (NIR) radiation reaches the leaf's spongy mesophyll tissue, where 40 to 50 percent of the NIR radiation is reflected. The reflection caused by the spongy mesophyll tissue is responsible for the fairly flat spectral signature found in Figure 2.8 between 0.7 μm and 1.3 μm [33, 35]. The cell structures of different vegetation types vary a lot, which leads to discernible NIR reflection patterns [39]. The remaining NIR radiation is transmitted. Multiple layers of leaves in a plant canopy provide the opportunity for hierarchical layers of transmittance and reflectance. Hence the infrared reflectance increases with the number

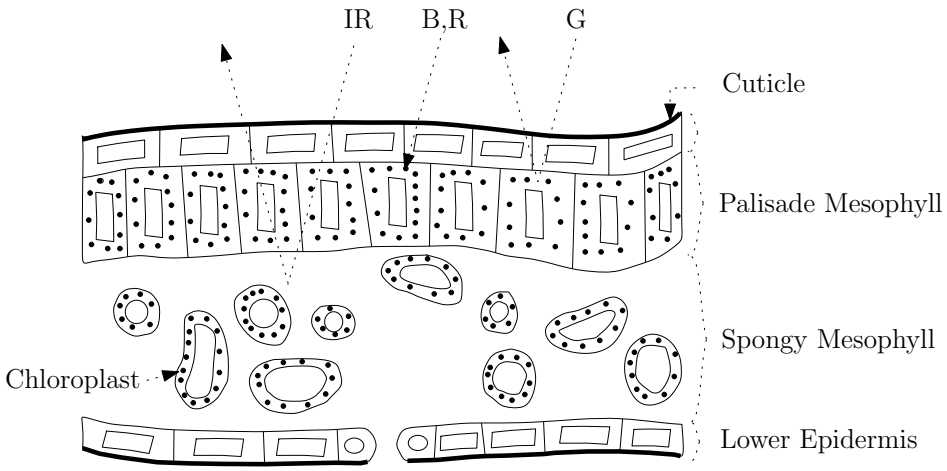


Figure 2.9: Generalised cross-section showing the cell structure of a green leaf.

of layers of leaves in a canopy [33]. Healthy vegetation consists mainly of water and the spectral signature of vegetation therefore also contains water absorption bands at 1.4μm, 1.9 μm and 2.7 μm which, are located in the Short-wave Infrared (SWIR) region.

Spectral ratioing (or image differencing) is a popular transform thst is used on remotely sensed data to enhance the interpretability of the data. Normalised Difference Vegetation Index (NDVI) is an example of such a spectral ratioing technique and for the MODIS sensor is calculated with

$$\text{NDVI} = \frac{(\text{Band 2}) - (\text{Band 1})}{(\text{Band 2}) + (\text{Band 1})}. \tag{2.11}$$

It can clearly be seen from Figure 2.8, that vegetation reflects much less radiation in the red band (MODIS 1) than in the NIR band (MODIS band 2) and vegetation will consequently have a large NDVI value. Soil will have a lower NDVI value, since there is not much difference between the amounts of radiation that are reflected in the red and NIR bands (in the case of soil). NDVI can be used to identify vegetative areas [23].

2.5.4.2 Soil

The soil curve in Figure 2.8 does not have a "peak-valley" appearance. Some of the factors affecting soil reflectance include moisture content, soil texture (proportion of sand, silt or clay), surface roughness, presence of iron oxide and organic matter content [33]. Moist soil has a lower reflectance if compared to dry soil. As with vegetation, the effect of moistness (water content) on reflectance is amplified in the water absorption bands. The soil texture influences the soil's capability of retaining water. Clay particles are smaller than those of silt, which in turn are smaller than those of sand. Sand

is thus more porous when compared to clay. Clay can retain water the best, while sand has the lowest retention capability due to its porous nature (which is caused by the size of the sand granules) [39]. The surface roughness of a soil and the presence of iron oxide and organic matter in a soil will also significantly decrease its reflectance capability [33].

2.5.4.3 Water

As can be seen from Figure 2.8, most of the radiation incident upon water is either absorbed or transmitted. The longer the wavelength of incident radiation, the better it is absorbed by water, and therefore water appears blue-green in the visible spectrum, and dark in the infrared range. Suspended sediments or shallow water bodies may cause increased reflection. The increased reflectance may even be observable in the NIR region [23].

2.6 REMOTE SENSING PLATFORMS

Different remote sensing platforms and systems are discussed in this section. The section concludes with the resolution of remote sensing sensors.

2.6.1 Ground-based, airborne and spaceborne platforms

There are three main types of remote sensing platforms, namely *ground-based*, *airborne* and *space-borne* platforms.

Ground-based sensors can usually only work on a small scale and are thus normally used for generating ground truth data. Balloons, aircraft and more recently Unmanned Aerial Vehicles (UAVs) are all airborne platforms. Satellites are the primary platforms used to host spaceborne sensors [23]. Active and passive remote sensing systems can be found on both airborne and spaceborne platforms.

2.6.2 Passive and active remote sensing systems

Remote sensing systems can be grouped into two mayor system categories, named *active* and *passive* systems. In a passive remote sensing system the sun is used as the source of electromagnetic radiation, while in an active system, such as radar, the system produces its own radiation. The difference between a passive and an active system is illustrated in Figure 2.10. Most active radiation systems

(radars) produce radiation in the spectrum bands where the sun does not radiate with high intensity, such as the microwave region. A passive sensor can also be designed to measure the thermal radiation produced by the earth, so that the earth becomes the radiation source of the passive system instead of the sun.

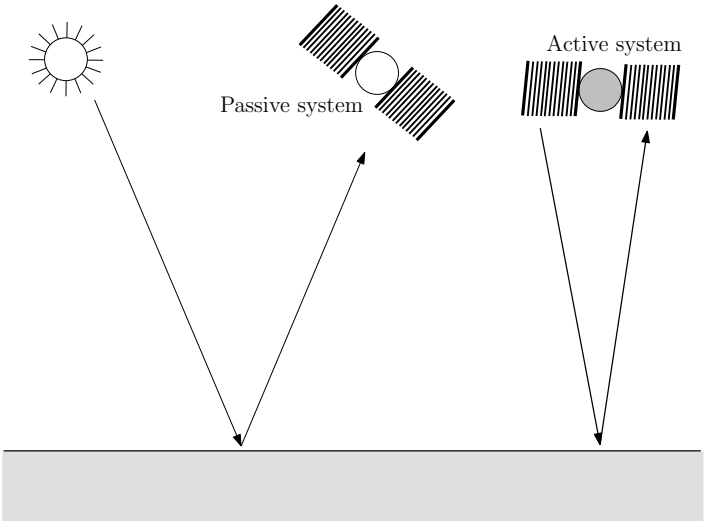


Figure 2.10: Active and passive remote sensing system.

Two main acquisition scanners are employed by passive remote sensing systems, namely the transverse scanner and the *pushbroom* scanner. A transverse (also termed across-track or *whiskbroom*) scanner is an electro-mechanical device that obtains data from narrow swaths of terrain (by using a scanning mirror), which are at right angles to the direction of movement. The scanning mirror sweeps across the satellite's ground track. The scanning mirror directs reflected (or emitted) radiation towards the on-board detectors. A pushbroom scanning system does not rely on a scanning mirror to direct radiation onto a detector, but instead employs a linear array of detectors. Each detector in the array measures the radiation reflected from a small area on the ground, which is known as a ground resolution cell [35].

2.6.3 Resolution of remote sensing sensors

One way of comparing different remote sensing sensors with one another is to compare their resolution. The resolution of a remote sensing sensor can be divided into four categories, namely its *spectral*, *spatial*, *temporal* and *radiometric* resolution [23].

2.6.3.1 Spectral resolution

Most remote sensing systems record data from different spectral bands. The width of these spectral bands is known as the spectral resolution of the sensor. If these spectral bands are not small (low resolution) and there are large gaps between them, the sensor is called a multispectral sensor; on the other hand, if the resolution is high and there are almost no gaps between bands, the sensor is known as a hyperspectral sensor [23]. The difference between multispectral and hyperspectral sensors is illustrated in Figure 2.11.

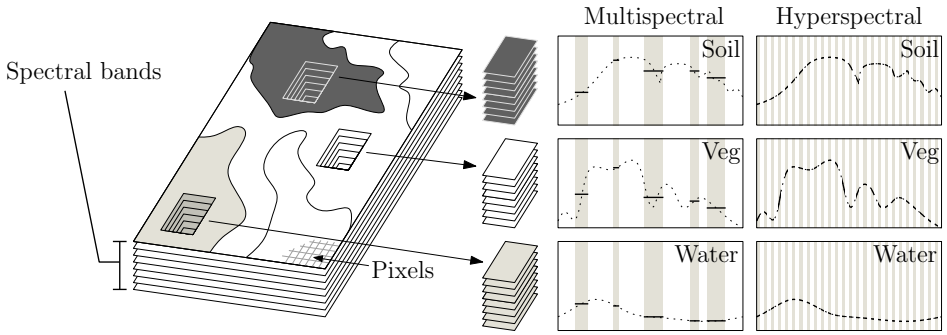


Figure 2.11: Interaction of reflected light with surface materials, showing multispectral and hyperspectral signatures (from [23]).

2.6.3.2 Spatial resolution

The spatial resolution depends on the Instantaneous Field of View (IFOV), which is the angular cone of visibility that describes the surface area from which radiation is recorded by the sensor at any instant in time.

Increasing the spatial resolution of a sensor can also decrease its spectral resolution. If the sensor's IFOV is made narrower (which increases its spatial resolution) the area which is monitored by the sensor is decreased. The smaller surveillance area means less energy reaches the sensor, which means that the Signal-to-Noise Ratio (SNR) of the sensor will also be lower. To compensate for this the scanning bandwidth must be increased, which in turn lowers the spectral resolution of the sensor [23].

2.6.3.3 Temporal resolution

The temporal resolution of a sensor refers to the time interval it needs to make successive measurements of the same physical location. The temporal resolution is actually dependent on the remote sensing platform carrying the sensor. There is normally a trade-off between the temporal and spatial resolution of a sensor. The higher the spatial resolution the lower the temporal resolution. This trade-off is due to the sensor's swath width. If the sensor has a large swath width, then the revisit time to a specific location will be shorter. The revisit time will be shorter, because the entire earth can be surveyed faster [23].

2.6.3.4 Radiometric resolution

The radiometric resolution of a sensor is measured in bits. The number of bits that a sensor has available to record values determines the number of unique levels of radiation it can measure. For example, MODIS, has a radiometric resolution of 12 bits, thus MODIS can detect $2^{12} = 4096$ unique levels of radiation.

Remote sensing data are commonly expressed with Digital Numbers (DNs) ranging from 0 to $2^b - 1$, where b is the radiometric resolution of the sensor, in bits [23].

2.6.4 Signal-to-noise ratio

Another important characteristic of a remote sensing sensor is its SNR. The SNR is defined as the the energy contained in the received signal divided by the energy in the noise that is generated by aberrations in the electronics. It is desirable to have a sensor with a high SNR [35].

2.7 MODERATE RESOLUTION IMAGING SPECTRORADIOMETER

In this section the MODIS sensor is discussed in detail. A literature review of land cover mapping applications with MODIS data can be found in [41]. The MODIS product selected for this thesis is discussed in Section 2.7.5.

2.7.1 History of MODIS

In 1983 NASA created the EOS Science and Mission Requirements Working Group (EOSMRWG) with the explicit purpose of developing a global concept for EOS. The EOSMRWG report delivered in 1984 called for several instruments to survey the earth. In 1984 NASA formed an instrument panel for each facility sensor proposed by the EOSMRWG report. The MODIS instrument panel suggested the development of two sensors, the MODIS-T and the MODIS-N. The original MODIS-N was a conventional imaging filtered radiometer capable of surveying 35 spectral bands, while MODIS-T was supposed to be a 64-band imagining spectroradiometer with the ability to tilt to avoid sun-glint from the oceans. The management and development of the MODIS sensor was assigned to Goddard Space Flight Center (GSFC), where it was decided to develop the MODIS-T sensor in-house and to outsource the MODIS-N sensor. In 1991 the Hughes/Santa Barbara Research Center (SBRC) was assigned the contract to build the MODIS-N sensor. Soon after the SBRC contract started, major restructuring of EOS occurred, which led to a decision to keep the MODIS-N design and to scrap the MODIS-T design. Over the next few years the SBRC developed and fabricated two MODIS flight models. The first of these models was installed in EOS-Terra, which was subsequently launched on December 18, 1999. The second model was installed in EOS-Aqua, which was launched on May 4, 2002 [37, 38].

2.7.2 MODIS sensor characteristics

Both EOS-Terra and EOS-Aqua are polar-orbiting sun-synchronous platforms. The orbital height of the EOS platforms is 705 km at the equator. Each MODIS instrument has a two-sided scan mirror with a maximum scan angle of 55° at either side of nadir, providing a nominal swath width of 2330 km. Because of to the large swath width, the MODIS sensor surveys the earth every one to two days [42, 43]. The predecessors of MODIS are NOAA's AVHRR and Landsat's TM. MODIS was compared to AVHRR in Section 1.3. Although the TM sensor provides a higher spatial resolution than MODIS, the TM sensor is charachterised by incomplete spatial coverage, low temporal resolution and cloud contamination [23]. A total of 36 spectral bands are surveyed by MODIS inside the spectral region 0.412-14.235 μm. The first two bands are located in the Red (R) (0.648 μm) and NIR (0.858 μm) regions and have a spatial resolution of 250 m. The next five MODIS bands (bands 3-7: 0.470 μm, 0.555 μm, 1.240 μm, 1.640 μm and 2.13 μm) have a spatial resolution of 500 m and are located in the visible to the SWIR regions. The remaining 29 bands (bands 8–36) have a spatial resolution

of 1000 m and are located in the the Mid-wave Infrared (MWIR) and Long-wave Infrared (LWIR) regions. The MODIS instrument has a 12-bit radiometric resolution [42]. The characteristics of the 36 spectral bands of MODIS are displayed in Table 2.2 and Table 2.3. Additional MODIS characteristics can be found in Table 2.4.

2.7.3 MODIS products

The data captured by the MODIS sensor can be subjected to a few levels of processing [42]:

- Level 0: The initial data set, which is automatically derived from the instrumental raw data.
- Level 1A: Contains geodetic information.
- Level 1B: Calibrated radiances for all bands and surface reflectance values for selective bands.
- Level 2: Derived geophysical variables at the same resolution and location as level 1 source data (swath products).
- Level 2G: Level 2 data mapped on a uniform space-time grid scale (sinusoidal).
- Level 3: Gridded variables in derived spatial and/or temporal resolutions.
- Level 4: Model output or results from analyses of lower-level data.

The raw MODIS data are transferred to ground stations in White Sands, New Mexico, via the Tracking and Data Relay Satellite System (TDRSS). The raw data are then forwarded to the EOS Data and Operations System (EDOS) at GSFC, where level 0 processing takes place. Level 1A and level 1B data are generated by GSFC Earth Sciences DAAC (GES DAAC). Higher-level MODIS land and atmosphere products are produced by the MODIS Adaptive Processing System (MODAPS), and distributed by three Distributed Active Archive Centers (DAACs), namely the L1 and Atmosphere Archive and Distribution System (LAADS), the Land Processes DAAC (LP DAAC) and the National Snow and Ice Data Center DAAC (NSIDC DAAC). Ocean colour products are produced and distributed by the Ocean Color Data Processing System (OCDPS) [37, 38].

There are close to 40 MODIS products available. Most MODIS product names start with three specific letters, which may be MOD, MYD or MCD. MOD indicates that the product was derived using only

Table 2.2: A summary of MODIS spectral bands 1-27.

Band	Wavelength (μm)	IFOV (m) [at nadir]	Primary use	Spectral region
Band 1	0.62–0.67	250 × 250	Land/Cloud/Aerosols Boundaries	R
Band 2	0.841–0.876	250 × 250	Land/Cloud/Aerosols Boundaries	NIR
Band 3	0.459–0.479	500 × 500	Land/Cloud/Aerosols Properties	B
Band 4	0.545–0.565	500 × 500	Land/Cloud/Aerosols Properties	G
Band 5	1.230–1.250	500 × 500	Land/Cloud/Aerosols Properties	SWIR
Band 6	1.628–1.652	500 × 500	Land/Cloud/Aerosols Properties	SWIR
Band 7	2.105–2.155	500 × 500	Land/Cloud/Aerosols Properties	SWIR
Band 8	0.405–0.420	1000 × 1000	Ocean Colour/Phytoplankton/ Biogeochemistry	B
Band 9	0.438–0.448	1000 × 1000		B
Band 10	0.483–0.493	1000 × 1000		B
Band 11	0.526–0.536	1000 × 1000		G
Band 12	0.546–0.556	1000 × 1000		G
Band 13	0.662–0.672	1000 × 1000		R
Band 14	0.673–0.683	1000 × 1000		R
Band 15	0.743–0.753	1000 × 1000		NIR
Band 16	0.862–0.877	1000 × 1000		NIR
Band 17	0.890–0.920	1000 × 1000	Atmospheric Water Vapour	NIR
Band 18	0.931–0.941	1000 × 1000	Atmospheric Water Vapour	NIR
Band 19	0.915–0.965	1000 × 1000	Atmospheric Water Vapour	NIR
Band 20	3.660–3.840	1000 × 1000	Surface/Cloud Temperature	MWIR
Band 21	3.929–3.989	1000 × 1000	Surface/Cloud Temperature	MWIR
Band 22	3.929–3.989	1000 × 1000	Surface/Cloud Temperature	MWIR
Band 23	4.020–4.080	1000 × 1000	Surface/Cloud Temperature	MWIR
Band 24	4.433–4.498	1000 × 1000	Atmospheric Temperature	MWIR
Band 25	4.482–4.549	1000 × 1000	Atmospheric Temperature	MWIR
Band 26	1.360–1.390	1000 × 1000	Cirrus Clouds Water Vapour	NIR
Band 27	6.535–6.895	1000 × 1000	Cirrus Clouds Water Vapour	MWIR

Table 2.3: A summary of MODIS spectral bands 28-36.

Band	Wavelength (μm)	IFOV (m) [at nadir]	Primary use	Spectral region
Band 28	7.175–7.475	1000×1000	Cirrus Clouds Water Vapour	LWIR
Band 29	8.400–8.700	1000×1000	Cloud Properties	LWIR
Band 30	9.580–9.880	1000×1000	Ozone	LWIR
Band 31	10.780–11.280	1000×1000	Surface/Cloud Temperature	LWIR
Band 32	11.770–12.270	1000×1000	Surface/Cloud Temperature	LWIR
Band 33	13.185–13.485	1000×1000	Cloud Top	LWIR
Band 34	13.485–13.785	1000×1000	Cloud Top	LWIR
Band 35	13.785–14.085	1000×1000	Cloud Top	LWIR
Band 36	14.085–14.385	1000×1000	Cloud Top	LWIR

Table 2.4: MODIS Design Specifications.

Orbit	705 km, 10:30 AM descending node or 1:30 PM ascending node, sun-synchronous, near polar, circular
Scan rate	20.3 rpm, cross track
Swath dimension	2330 km (cross track) by 10 km (along track at nadir)
Telescope	17.78 cm off-axis, a focal (collimated), with intermediately held stop
Size	$1.0 \times 1.6 \times 1.0$ m^3.
Weight	250 kg
Power	225 W (orbital average)
Data rate	11 Mbps (peak daytime)
Quantisation	12 bits
Spatial resolution (at nadir)	250m (bands 1–2) 500m (bands 3–7), 1000 m (bands 8–36)
Design life	5 years

data from EOS-Terra, MYD indicates that data from EOS-Aqua was used and MCD indicates that the product was generated using data from EOS-Terra and EOS-Aqua. Normally two numbers follow the three letters and indicate the intended application of the product. A list of these numbers with their appropriate descriptions can be found in Table 2.5 and Table 2.6.

Most standard MODIS land products use a sinusoidal grid tiling system. Tiles are 10 degrees by 10 degrees at the equator. The tile coordinate system starts at (0,0) (horizontal tile number, vertical tile number) in the upper left corner and proceeds right (horizontal) and downward (vertical). The tile in the bottom right corner is (35,17).

2.7.4 MODIS design

MODIS is a whiskbroom scanning radiometer with a double-sided paddle wheel scan mirror, which operates at 20.3 rpm. The incident radiation (earth view) is reflected from the scan mirror to a fold mirror. From the fold mirror the radiation is reflected onto the primary mirror of the Afocal Gregorian Telescope. The radiation is reflected by the primary mirror and then passes through a field stop. After passing through the field stop the radiation falls upon the secondary mirror of the telescope. The secondary mirror then reflects the incident radiation to three dichroic beamsplitters. The beamsplitters split the radiation into four regions: NIR, visible, SWIR and MWIR, and LWIR, after which the radiation ends up on four Focal Plane Assemblies (FPAs), one for each region. The MODIS sensor also houses four on-board calibrators. The name of each calibrator is given below [37, 38]:

1. Solar Diffuser (SD) and Solar Diffuser Stability Monitor (SDSM),
2. Spectral Radiometric Calibration Assembly (SRCA),
3. Blackbody (BB),
4. Space View (SV).

2.7.5 The MCD43A4 product

The MCD43A4 MODIS product consists of seven BRDF corrected land surface reflectance (eight-day composite, 500 m resolution) time-series [22]. BRDF is discussed in Section 2.5.3. The reason for selecting this product was discussed in Section 1.3. The product is built from a 16-day rolling window

Table 2.5: The MODIS Product Codes 01–28

Product Code#	Description
01	Level-1A Radiance Counts
02	Level-1B Calibrated, Geolocated Radiances
03	Geolocation Data Set
04	Aerosol Product
05	Total Precipitable Water
06	Cloud Product
07	Atmospheric Profiles
08	Gridded Atmosphere Products (Level 3)
09	Atmospherically Corrected Surface Reflectance
10	Snow Cover
11	Land Surface Temperature and Emissivity
12	Land Cover/ Land Change
13	Vegetation Indices
14	Thermal Anomalies, Fires and Biomass Burning
15	Leaf Area Index and FPAR
16	Surface Resistance and Evapotranspiration
17	Vegetation Production, Net Primary Productivity
18	Normalised Water Leaving Radiance
19	Pigment Concentration
20	Chlorophyll II, Fluorescene
21	Chlorophyll and Pigment Concentration
22	Photosynthetically Active Radiation
23	Suspended Solids Concentration in Ocean Water
24	Organic Matter Concentration
25	Coccolith Concentration
26	Ocean Water Attenuation Coefficient
27	Ocean Primary Productivity
28	Sea Surface Temperature

Table 2.6: The MODIS Product Codes 29–MODISALB.

Product Code#	Description
29	Sea Ice Cover
31	Phycoerythrin Concentration
35	Cloud Mask
36	Total Absorption Coefficient
36	Total Absorption Coefficient
37	Ocean Aerosol Properties
39	Clear Water Epsilon
43	Albedo-16 day (Level 3)
44	Vegetation Cover Conversion and Continuous Fields
MODISALB	Snow and Sea Ice Albedo

of acquisitions obtained from the Terra and Aqua satellites, which explains the use of "MCD" in the product name. An MCD43A4 pixel value consists of seven reflection ratios (at 500 m resolution). The seven reflection ratios are located in the seven MODIS land bands. The raw MCD43A4 data are DNs (16-bit unsigned integer values). The raw 16-bit data of MCD43A4 should not be confused with the raw radiation value (which is a 12-bit value) recorded by the MODIS sensor. The raw MCD43A4 data should be multiplied by 0.0001 to obtain reflection ratios. The temporal period of MODIS MCD43A4 (if an observation is produced every eight-days) roughly translates to 45 observations per year. NDVI is calculated from MCD43A4 by using Equation 2.11. In the remainder of this thesis the phrase "MODIS pixel" refers to the seven time-series at 500 m resolution that are associated with the MCD43A4 product.

2.8 DATASET DESCRIPTION

The datasets used in this thesis are constructed from the eight-day composite MODIS MCD43A4 BRDF corrected 500 m land surface reflectance product. The study areas associated with the MODIS MCD43A4 datasets span a total area of approximately 230 km^2 in Gauteng and 800 km^2 in Limpopo, South Africa. Gauteng and Limpopo are provinces in South Africa and their physical location is shown in Figure 2.12.

Gauteng is the smallest province in South Africa. The name "Gauteng" is derived from the Sesotho word meaning "Place of Gold". The name chosen for Gauteng is appropriate as it is the economic heart of South Africa. The capital of Gauteng is Johannesburg [12].

Limpopo is the northernmost province of South Africa. It is named after the Limpopo River. "Limpopo" is the Zulu word for "waterfalls". The Limpopo province houses the largest hunting industry in the country. The capital of Limpopo is Polokwane and was formerly known as Pietersburg [12].

The reasons for selecting Gauteng and Limpopo as study regions were discussed in Section 1.1.

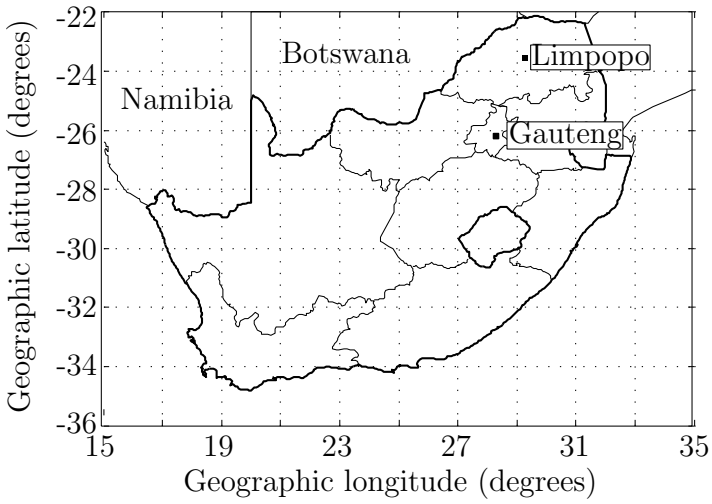


Figure 2.12: The physical location of the Gauteng province and the Limpopo province [2] © IEEE 2012.

Two land cover classes are considered: *settlements* and *natural vegetation*, denoted by s and v respectively. In this thesis the settlements class contains pixels that contain more than 50% buildings (construction), whereas the vegetation class contains pixels with more than 90% vegetation.

The above class classification rule is illustrated below with an example. Figure 2.13 is a Google Earth^{TM} image of a populated area in Gauteng. Four red parallelograms are visible in Figure 2.13. Each red parallelogram represents a pixel that is actually surveyed by the MODIS sensor and is 500 m$\times$500 m in size. If the vegetation settlement classification rule mentioned above is applied to Figure 2.13, only the bottom right pixel would be classified as a vegetation pixel, while the remaining three would be classified as settlement pixels.

Two MODIS MCD43A4 datasets are used to investigate the hypertemporal techniques discussed in

Figure 2.13: A Google Earth^{TM} image of a populated area in Gauteng (courtesy of Google Earth^{TM}).

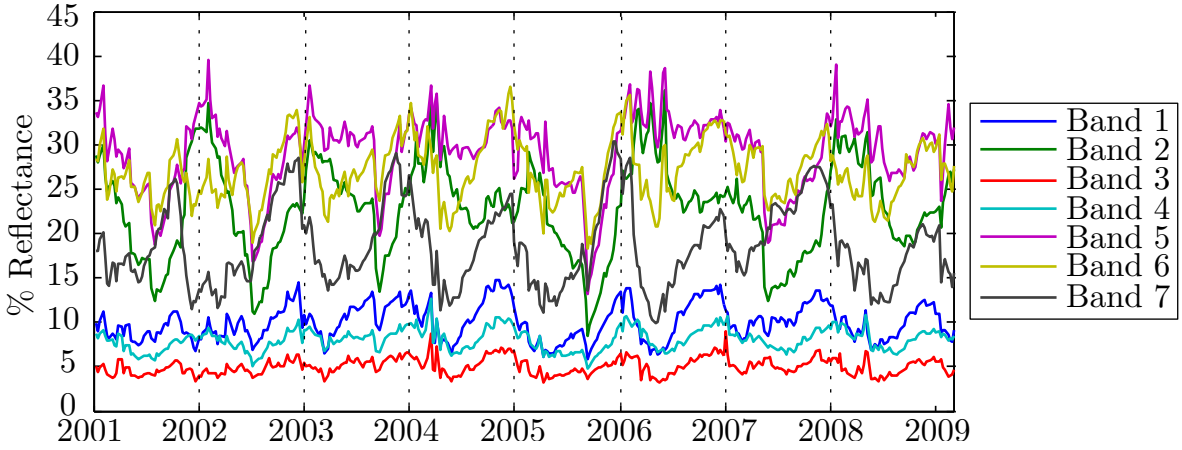


Figure 2.14: A random vegetation MODIS pixel in Gauteng.

this thesis. The Gauteng dataset consists of 1106 MODIS pixels, while the Limpopo dataset contains 3349 MODIS pixels and was selected with visual (human) interpretation of two high-resolution SPOT images from 2001 and 2009. MODIS pixels that, according to the SPOT images, either did not change or changed from vegetation to settlement were selected. Each MCD43A4 MODIS pixel contains seven time-series with $I = 368$ observations (extracted between January 2001 and March 2009). An NDVI time-series can be added to a MODIS pixel and is computed using the first two spectral land bands. The Gauteng and Limpopo datasets are respectively divided into the three classes: natural vegetation (592 Gauteng pixels and 1497 Limpopo pixels), settlements (333 Gauteng pixels and 1735 Limpopo pixels) and real land cover change from vegetation to settlement (181 Gauteng pixels and 117 Limpopo pixels). A random vegetation MODIS pixel in Gauteng is displayed in Figure 2.14.

2.9 CONCLUSION

The chapter provided a broad overview of the remote sensing field, which included a brief history of remote sensing, an introduction to the physical principles behind remote sensing, an overview of remote sensing platforms and an introduction to the MODIS sensor. The MODIS data used by the classification and change detection algorithms investigated in this thesis were also presented in this chapter.