LAND-COVER MAPPING IN AN AGRICULTURE ZONE USING SIMULATED SENTINEL-2 DATA

LOGAN S. PRYOR

B.Sc. University of Lethbridge, 2009

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfillment of the
Requirements of the Degree

MASTER OF SCIENCE

Department of Geography
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

ABSTRACT

Remote sensing technologies are used to assist in the mapping and monitoring of land cover in space and time. The European Space Agency’s (ESA) upcoming Sentinel-2 MultiSpectral Instrument (MSI) to be launched in 2013 has improved spatial and spectral properties compared to the current large-swath medium-resolution satellite sensors. Prior to the deployment of future sensors it is important to simulate and test the sensor data to evaluate the sensor's potential performance in producing the existing data products and develop new algorithms. This study simulated Sentinel-2 MSI data from airborne hyperspectral data over an agriculture area in northern Alberta, Canada. The standard Sentinel-2 MSI land-cover product was evaluated by comparing it to one created from the standard Landsat 5 TM and SPOT 5 HRV data products. Furthermore the standard Sentinel-2 MSI water column content band configuration and algorithm was evaluated for atmospheric correction purposes.

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Dr. Karl Staenz and Dr. Craig Coburn, for their ongoing support and insight. I would also like to thank Dr. Jinkai Zhang and Mr. Trevor Armstrong for many conversations throughout this process. I would like to thank Mr. David Rolfson and Mr. Subir Chowdhury for their assistance. The research for this project would not have been able to occur if not for the University of Lethbridge School of Graduate Studies as well as the funding provided from the NSERC CREATE AMETHYST program, to which I am grateful. Also, I acknowledge the members of my thesis committee, Dr. Derek Peddle and Dr. Adriana Predoi-Cross, and the defense chair, Dr. Ian MacLachlan, as well as Dr. Phil Teillet.

In addition I would like to thank the European Space Agency, specifically Dr. Michael Berger for his guidance during my time at ESRIN in Rome. Furthermore, I am grateful for the friends and special individuals that I met during my time there and the professional and personal relationships that have formed.

Lastly, I would like to thank my family for their support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAFC- Agriculture and Agri-Foods Canada
AISA - Airborne Imaging Spectrometer for Application
ASTER - Advanced Space borne Thermal Emission and Reflection Radiometer
AVIRIS - Airborne Visible/Infrared Imaging Spectrometer
CCRS - Canada Centre for Remote Sensing
CFS - Canadian Forest Service
CORINE - Coordination of Information on the Environment
DEM - Digital Elevation Model
DN - Digital Number
EMR - Electromagnetic Radiation
EnMAP - Environmental Mapping and Analysis Program
EO - Earth Observation
EOSD - Earth Observation for Sustainable Development of Forests
ESA - European Space Agency
FIR - Finite Impulse Response
FLAASH - Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes
FTS - Fast Track Services
FWHM - Full Width Half Maximum
GIFOV - Ground Instantaneous Field of View
GLC2000 - Global Land Cover 2000
GMES - Global Monitoring for Environment and Security program
GSE - GMES Service Elements
GSD - Ground Sampling Distance
HRV - High Resolution Visible
ISDAS - Imaging Spectrometer Data Analysis System
LCC2000 - Land Cover Circa 2000
LCCS - Land Cover Classification System
LSD - Local Standard Deviation
LUT - Look-up table
MERIS - MEdium Resolution Imaging Spectrometer
MNF - Minimum Noise Fraction
MODIS - Moderate Resolution Imaging Spectrometer
MODTRAN - MODerate resolution atmospheric TRANsmission
MSI - MultiSpectral Imager
MSS - Multispectral Scanner
NDVI - Normalized Difference Vegetation Index
NIR - Near infrared
NLCD - National Land Cover Dataset
NLWIS - National Land and Water Information Services

PC - Principle Component
PFRA - Prairie Farm Rehabilitation Administration
PSF - Point Spread Function
ROI - Region of Interest
RS - Remote Sensing
RSRF - Relative Spectral Response Function
SNR - Signal-to-Noise Ratio
SPOT - Satellite Pour l'Observation de la Terre
SSA - Singular Spectral Analysis
SVM - Support Vector Machine
TM - Thematic Mapper
TOA - Top of Atmosphere
WF - Wiener Filter

1. INTRODUCTION

Global human population growth and development are consuming an increasing quantity of natural resources. Modern agriculture land-use practices are now being performed on a scale greater than at any point in history. For example in the last 40 years, irrigation and fertilizer application rates have both increased substantially (Tilman *et al.*, 2001, Gleick, 2003). Despite the magnitude of this growth, few studies have been carried out on the effects that these land-use practices are having on global environmental processes (Asner *et al.*, 2004). In general, land-cover and land-use change have caused loss of biodiversity through fragmentation of habitats and the degradation of soil and water (Pimm and Raven, 2000).

Natural ecosystems, such as wetland areas, are considered integral parts of the global ecosystem as they play important roles in soil conservation, maintaining water quality, pollution prevention, and economic activities (Mitsch and Gosselink, 1993). Specifically, wetland areas within the agricultural zone are important as they reduce nutrient loading caused by agricultural runoff (Blankenberg *et al.*, 2008). Therefore, the quantity and size of wetland areas in an agricultural catchment area can improve the region's ability to process agricultural run-off nutrients (Moreno *et al.*, 2007) and wetlands appear to function best as a spatially distributed connected system in a region (Mitsch and Gosselink, 2000). However, agriculture chemicals introduced to a wetland environment impact the its productivity (Johnson, 1986). Furthermore, wetland areas around the world are under threat from conversion to agriculture production. In a recent census of wetland loss to agriculture, was estimated that over half of the world’s wetlands have been lost (Bartzen *et al.*, 2010).

Remote sensing technologies have been used and evaluated extensively for environmental monitoring and mapping since their introduction in the 1970s (Lent and Thorley, 1969, Dey and Richards, 1981). More specifically, they have been applied to map and monitor wetland areas (Jensen *et al.*, 1984) and agricultural areas (Kauth and Thomas, 1976). Large-coverage multi-spectral sensors, such as Landsat Thematic Mapper (TM) and Satellite Pour l'Observation de la Terre (SPOT) High Resolution Visible (HRV), have been used extensively to map and monitor land cover (Homer *et al.*, 2007, Cihlar *et al.*, 1996, Vogelmann *et al.*, 1998, Wulder *et al.*, 2003). For example, the Landsat program has been running for over 40 years and has collected the world's largest assemblage of temporal satellite data (Wulder *et al.*, 2008b).

European nations, among others, have relied upon Landsat data products as an integral component of large area land-cover monitoring and mapping (Wulder *et al.*, 2008b). However, a joint initiative between the European Space Agency (ESA) and the European Commission was established to create a European capacity for the provision and use of operational monitoring information and security applications, called the Global Monitoring for Environment and Security (GMES) program (ESA, 2007). ESA is currently developing the Sentinel satellite missions, which will comprise 5 different satellites when complete. Sentinel-1 will be a radar satellite, Sentinel-2 will provide medium-resolution spatial data for land services, Sentinel-3 will provide data for ocean and land applications and Sentinel-4/5 will provide data for atmospheric composition modelling. This constellation will reduce the European community's reliance on other satellite systems and insure continuity of Earth observation data.

Of specific interest to this research was the Sentinel-2 sensor, the MultiSpectral Imager (MSI). It will have enhanced radiometric and spectral characteristics and spatial coverage compared to existing sensors such as SPOT 5 HRV and Landsat 5 TM. High-level data products, including land cover/use, will be developed through GMES projects such as the GMES Fast Track Services (FTS) and GMES Service Elements (GSE) to address global environmental issues (ESA, 2007).

1.1 Satellite Sensor Simulation

It has become common practice to simulate satellite data prior to sensor deployment (Börner *et al.*, 2001, Guanter *et al.*, 2009). Data simulation offers the development and validation of data processing algorithms, both for pre-processing and scientific exploitation (Guanter *et al.*, 2009). For example, prior to the launch of SPOT 1, a simulation campaign was performed to assess the performance of the new sensor. Applications such as land-cover mapping, land-use planning and ecological investigations were evaluated using simulated data (Betts *et al.*, 1986, Buchan and Hubbard, 1986, Essery and Wilcock, 1986, Connors *et al.*, 1987). Moderate Resolution Imaging Spectrometer (MODIS) data were also simulated prior to launch to optimize the algorithm for atmospheric water vapour estimation (Kaufman and Gao, 1992).

Advancements in sensor simulation have introduced simulation techniques that include modules to simulate solar/viewing geometry and atmospheric effects (Börner *et al.*, 2001). Recently, hyperspectral data from the called Environmental Mapping and Analysis Program (EnMAP) instrument were simulated using a five-module system called “EnMAP Simulator” (Guanter *et al.*, 2009). The multiple-module system, which takes into account the characteristics of the sensors and the scene, successfully simulated

EnMAP scenes and it will serve as a basis for data pre-processing and information extraction algorithm development (Guanter *et al.*, 2009)

This thesis research developed a procedure to simulate a Sentinel-2 MSI data set and implemented an approach to assess a standard Sentinel-2 MSI land-cover information product in an agriculture zone. An airborne hyperspectral data set with a high-spectral and spatial resolution was used to simulate the MSI data.

The performance of the MSI data was evaluated by comparing the land-cover classification map to those generated from data acquired from current Earth observation satellite sensors, SPOT 5 HRV and Landsat 5 TM. The simulation and evaluation of the Sentinel-2 MSI data prior to launch is essential for the development of information products that can be used by organizations interested in land-cover mapping.

1.2 Objectives and Hypothesis

The main research objectives of this study were to:

(a) Develop a method to simulate the Sentinel-2 MSI data from airborne hyperspectral data and evaluate the Sentinel-2 standard land-cover product generated from simulated Sentinel-2 data against ground reference information and existing information products derived from Landsat 5 TM and SPOT 5 HRV data, and

(b) Evaluate the performance of simulated Sentinel-2 data for estimating atmospheric water column content.

The hypothesis of this study is as follows:

With enhanced spectral and spatial characteristics, Sentinel-2 MSI will be able to perform land cover mapping as well as or better than current medium-resolution multispectral sensors, Landsat 5 TM and SPOT 5 HRV. Furthermore, Sentinel-2 MSI will be able to estimate water column content to an acceptable level of error.

The following chapter reviews the current research with respect to sensor scene simulation, hyperspectral image processing, and image classification to generate land cover products in the context of the above research objective.

2. LITERATURE REVIEW

2.1 Introduction

Remote Sensing (RS) is the field of study associated with extracting information about an object without coming into physical contact with it (Schott, 2007). Traditionally terrestrial RS relies on measuring and understanding the electromagnetic radiation (EMR) that is reflected or emitted from the surface of the Earth and received by an RS system. For passive RS systems, by measuring the solar irradiance onto a surface in addition to the radiant exitance from a surface, the hemispherical reflectance (ρ), of the area can be calculated as shown in Equation 2.1,

$$\rho = \frac{M}{E} \tag{2.1}$$

where M is the exitance from a surface and E is the irradiance onto the surface.

The surface reflectance derived from RS data can then be used to identify surface properties due to absorption characteristics certain molecules have. This calculation and these measurements concerned with vegetation (Hill and Sturm, 1991, Friedl *et al.*, 1995), minerals (Clark *et al.*, 1990, Hunt, 1977), water (Kutser *et al.*, 2001) and snow and ice (Dozier, 1989).

Data simulation studies require knowledge related to specific sensor characteristics and the characteristics of the scene as well as pre-processing of data that ensures data quality. The generation of land-cover information products from RS systems makes use of classification algorithms. The following sections will review the literature related to satellite data simulation, noise in satellite data, and RS data classification.

2.2 Remote Sensing Systems

Remote sensing systems can defined by their fundamental resolutions: Spectral resolution, spatial resolution, temporal resolution and radiometric resolution (Jensen, 2005). The following sections will discuss these fundamental resolutions and present the characteristics of the MSI as they apply.

2.2.1 Spectral Resolution

A RS system has a sensor that samples radiation at one or more locations within the electromagnetic (EM) spectrum, which is referred to as bands. A band is defined by its Relative Spectral Response Function (RSRF) which is characterized by the centre wavelength and bandwidth at Full-Width Half-Maximum (FWHM) (Figure 2.1). The FWHM is the width of the band at the point where the band has a response equal to half its maximum.

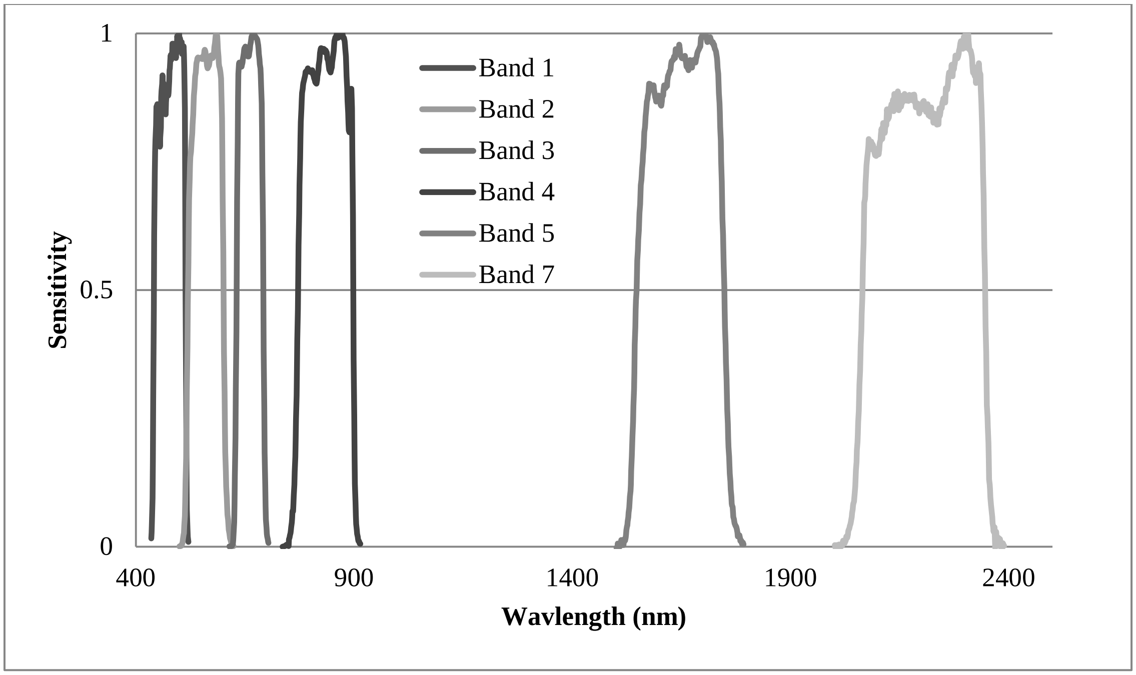


Figure 2.1: Landsat 5 TM multi-spectral sensor RSRF curves. The bandwidth for each band at 0.5 sensitivity is the FWHM.

RS systems has developed to position or add sensors bands so that they measure EM radiation within certain wavelength regions (Jensen, 2005). For example, the multi-spectral TM sensor that was added to the Landsat series of Earth observation satellites in 1982 was superior for vegetation applications due to enhanced spectral resolution compared to other systems with fewer spectral bands (Tucker, 1978). The continual evolution of optical remote sensing sensor technology has led to the development of spaceborne hyperspectral instruments. These instruments have hundreds of contiguous, narrow spectral bands, compared with multi-spectral sensors that usually only have several discontinuous and wide spectral bands (Hu *et al.*, 2009). These sensors usually have reduced spatial resolution due to data volume and are not yet sufficiently reliable to be used in an operational context.

The Sentinel-2 MSI is a multispectral sensor and will have 13 bands between 443 nm and 2190 nm. The band centers and spectral widths are presented as follows:

Table 2.1: Spectral characteristics of Sentinel-2 MSI.

Band #	Center λ (nm)	Spectral width $\Delta\lambda$ at FWHM (nm)
1	443	20
2	490	65
3	560	35
4	665	30
5	705	15
6	740	15
7	783	20
8	842	115
8a	865	20
9	945	20
10	1375	30
11	1610	90
12	2190	180

2.2.2 Spatial Resolution

The spatial resolution of a terrestrial RS instrument is defined by the Ground Instantaneous Field-Of-View (GIFOV). The GIFOV is the area on the ground being measured by a sensor at a given time of scanning (Schott, 2007). It is determined by the distance from the sensor to the surface and the instantaneous field of view of the sensor. The GIFOV is also referred to as ground sampling distance (GSD). Airborne sensors generally have a higher spatial resolution than satellite sensors due to their closer proximity to the target. The Point Spread Function (PSF) of an imaging system characterizes how a sensor responds to point objects (Schott, 2007) (Figure 2.2). In an ideal system there would be no spread of a measured signal as such effects constitute an inherent source of uncertainty in satellite images. This is because radiance from neighbouring pixels will contribute to the signal received by a sensor from a single pixel (Huang *et al.*, 2002, Justice *et al.*, 1989).

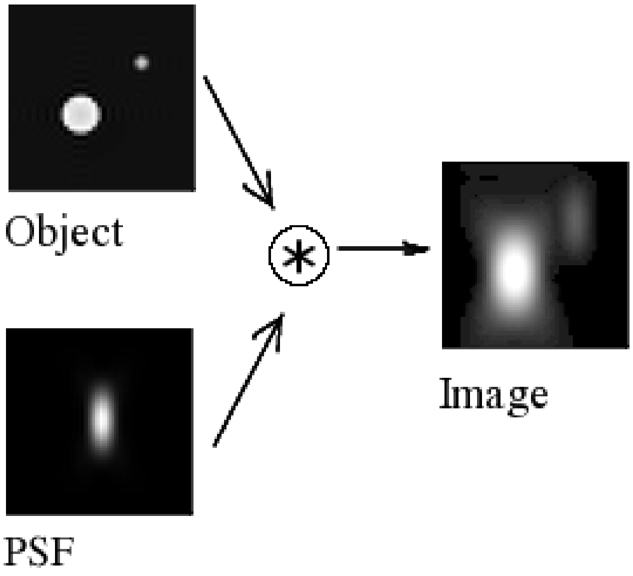


Figure 2.2: An object when being sampled by an optical system will be blurred on the focal plane depending on the PSF of the optical system.

In order to accurately simulate the spatial characteristics of an image, the PSF of the sensor system has to be characterized. The FWHM of the PSF, when projected on the ground, is referred to as the instrument's ground sampling size (Schott, 2007). The GSD is the sample distance in each direction projected on the ground is a good estimate of the sensors ground sampling size (Schott, 2007). If the PSF of a sensor being simulated has not been measured it can be modelled using a 3-dimensional Gaussian function (Kavzoglu, 2004).

Sentinel-2 MSI will use three different spatial ground sampling sizes, 10 m, 20 m, and 60 m. The 10 m bands are to ensure compatibility with SPOT 4 and 5 in addition to meet user requirements for land cover classification (Drusch *et al.*, 2012). The 60 m bands will be used for atmospheric correction (443 nm for aerosols retrieval, 940 nm for water vapour correction and 1375 nm for cirrus cloud detection.) (Drusch *et al.*, 2012).

Table 2.2: Spatial characteristics of Sentinel-2 MSI.

Band #	**Ground Sampling Size (m)**
1	60
2	10
3	10
4	10
5	20
6	20
7	20
8	10
8a	20
9	60
10	60
11	20
12	20

2.2.3 Temporal Resolution

The temporal resolution of a sensor defines how the data are collected in time. This is referred to as the revisit-time of the sensor, or the length of time it takes a sensor to revisit the same spatial location to collect data. The revisit time can be reduced if the sensor has a large swath width, is part of a constellation of the same sensor or if the sensor has off-nadir viewing capabilities.

With two satellites, Sentinel-2 MSI will have a revisit time of 5 days at the equator and 2-3 days at mid-latitudes. At an orbit altitude of 786 km and with its swath width of 290 km, full and systematic coverage of all land surfaces from -56 degrees to +83 degrees is expected (Aschbacher and Milagro-Pérez, 2012).

2.2.4 Radiometric Resolution

As an RS system collects data when it is sampling radiance, in a simple sense, the incoming radiance signal is converted to an electronic signal. This continuous signal is then partitioned or quantized to a numerical scale that is dependent on the bit-depth of the sensor. For example the raw data recorded by an 8-bit sensor would have a possible data range of 256 values (2^8). The data would then require a multiplicative factor and an additive factor to convert the data to radiance units. These factors are called the gain (multiplicative) and offset (additive).

2.3 Sensor Noise

Unwanted variation of a signal that should be at a constant level is defined as noise (Schott, 2007). These effects interfere with the true radiance signal of the surface that is being measured. Noise affects the quality of a signal and can be expressed, when viewed relative to the corresponding signal, as the signal-to-noise ratio (SNR) (Schott, 2007).

Noise effects in remotely sensed data can be divided into two main groups, random and systematic noise. Random noise can be defined as the unpredictable variation of the recorded signal occurring indiscriminately in a dataset. Systematic noise, on the other hand, has a pattern to its distribution and variation.

2.3.1 Random Noise

As EMR is received by a sensor, incoming photons interact with a sensor's detector and the output is an electrical signal. It is then transferred by an analogue/digital converter to a digital number (DN). The input photons (the output electrical signal) are referred to as the signal.

Random noise in RS systems occurs from elements that add uncertainty to the raw-signal level of a target, such as random fluctuation of photon events, sensor temperature change, electronic noise from each component in the imaging system, acoustic vibration of electrical components, static charge and the quantization of the signal. It is quantified by the standard deviation of its statistical distributions (Fiete and Tantalo, 2001). Furthermore, the sum of the variances of all noise contributors is the variance of the total noise (Fiete and Tantalo, 2001). Random noise in optical RS systems is typically independent of the data and is generally additive to DN in nature (Corner *et al.*, 2003). It can be represented by a random zero-mean Gaussian process with a probability density function $f_x(x)$ given as:

$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma_n}} e^{-x^2/2\sigma_n^2} \tag{2.2}$$

where σ_n is the standard deviation of the total system noise and x is the amount of noise variation.

2.3.1.1 Random Noise Estimation

In many cases, the technical details of the detector or sensor are not available to measure the amount of the different types of noise that may contribute to a signal. Therefore, an estimation of the standard deviation of the noise of the image data is desired so that the data can be suitably filtered to remove the noise (Corner *et al.*, 2003). If the desired signal is known, the estimation of the noise can be carried out simply by calculating the root mean square error of the recorded signal relative to the real signal (Schott, 2007). Unfortunately in remotely sensed images, the true values of the scene are usually unknown. Therefore, spatial autocorrelation can be used to estimate the noise in an image in the spatial domain.

Methods for calculating the SNR generally involve an area in a remotely sensed image that is homogenous (Gao, 1993). By calculating the signal mean, μ, and standard deviation, σ, of the homogenous area, the ratio of the mean to the standard deviation gives an estimate of the SNR of the sensor (Fujimotor *et al.*, 1989) as follows:

$$SNR = \frac{\mu}{\sigma} \tag{2.3}$$

This method may yield SNR values that are not representative of the entire scene because only a few homogenous areas are being used in the calculation due to the fact that generally not enough homogenous areas are present in a image.

2.3.2 Systematic Noise

Striping in images occurs when the individual detectors are not calibrated correctly, or due to temperature change or functional failure (Sun *et al.*, 2008). Miscalibration of the gains and offsets of the detectors in a push broom sensor's array can create distinct stripes. Striping artefacts, if not corrected, can lead to improper image classification and incorrect retrieval of quantitative information, such as biophysical and biochemical variables (Simpson *et al.*, 1995).

2.3.3 Noise Simulation

Additive noise can be effectively simulated using the probability function presented in equation (2.2). The generated noise can then be added to the image expressed as:

$$x(i, j) = S(i.j) + n(i, j) \tag{2.4}$$

where x is the recorded value in DN, S is the digital signal and n is the additive noise for a pixel at location (i,j). This simulation method has been used to evaluate noise estimation techniques (Corner *et al.*, 2003).

Noise simulation is also used when creating simulated satellite-type data (Börner *et al.*, 2001, Guanter *et al.*, 2009). Simulated satellite data require the use of re-sampling, both spectrally and spatially, of high-spatial resolution hyperspectral data. This process creates an unrealistically high SNR and, therefore, noise is required to be added (Börner *et al.*, 2001).

2.4 Pre-processing Remotely Sensed Data

Data that are initially recorded by a RS system are in a raw, un-processed format. The raw data needs to be processed before the data can be used. This process includes the reduction of sensor artefacts, the conversion to physical units and the correction for other unwanted effects on the data such as atmospheric attenuation.

2.4.1 Reduction of Random Noise

Noise can be removed in either the spatial domain using smoothing kernels, or in the spectral domain using convolution filters, or by first transforming the data to another state, such as frequency space.

2.4.1.1 Filter-based Methods

One of the earliest methods for smoothing continuous data was done using a least-square method (Savitzky and Golay, 1964). Savitzky and Golay's method involved moving a user-defined window in the spectral domain. A best-fit polynomial equation was created through the points in the window. By using the coefficients for the equation, a new single point is calculated. The main disadvantage to this method is the difficulty in determining the window size so that only noise is removed and not small spectral features.

The Savitzky-Golay filter assumes that random noise has similar characteristics throughout the spectrum and can be handled with an invariant procedure over the spectrum (Tsai and Philpot, 1998). However, this assumption may not always be appropriate as noise in hyperspectral detectors may be wavelength dependent (Kawata and Minami, 1984). This technique improved the least mean-square method by developing an algorithm that attempts to filter out wavelength dependent random noise. In this algorithm, the variance of the noise of the detector at each wavelength must be known or estimated. The main issue with the two above mentioned filters, as well as a mean filter, is that as the filter size increases, the spectral details of the dataset are suppressed (Tsai and Philpot, 1998).

Smoothing of data spectrally can be improved by using characteristics that are sensor dependent (Khurshid *et al.*, 2006). For example, using the gain values of the sensor to determine a noise model. A noise cube is then generated after calculating the noise for all pixels. Subsequently, a user-defined number of similar spectra are selected using a moving window. The average of the spectra is calculated and then used to replace the original DN of a pixel. This method may affect the fraction mixture of a pixel in the scene and the authors note that the effects may need to be evaluated in the future (Khurshid *et al.*, 2006).

2.4.1.2 Transformation Methods

Noise can also be removed by first transforming the data-set. One method utilizes the principal component (PC) transformation such as used in hyperspectral data dimensionality reduction. A PC analysis of multi-spectral data transforms the data (de-correlate and maximize the variability) into components that are arranged in descending order with decreasing image variability (Schott, 2007). For image data, the maximum noise fraction (MNF) transform was developed to arrange the components in order with decreasing image noise (Green *et al.*, 1988).

Once the data has been transformed, the noise can be reduced by smoothing. Although it is not completely removed from the data, the noise variance is equalized across all the bands, which is the best alternative (Schott, 2007). The process is completed when the data are then transformed back into the image space. The noise-adjusted PC function (Lee *et al.*, 1990) is functionally similar to the MNF process differing in the direction that the components are ordered. Therefore, it has become convention to reverse the order of the components and use the term minimum-noise-fraction (MNF) (Schott, 2007).

Noise can also be removed using singular spectral analysis (SSA). With the SSA-based approach, the reflectance spectrum of a given pixel in a hyperspectral data cube is transformed into its state space. This space is created using a roving window that consists of overlapping ”snap shots” of the reflectance spectrum of a given pixel. The state space is characterized by irregular bases, which are the resultant products of the roving window being applied to the reflectance spectrum. This approach allows the reduction of noise, while keeping absorption features (Hu *et al.*, 2009). Each base can be separated into a group that contains mostly signal and a group that contains mostly noise. Therefore, the original spectrum can be recreated with only the signal portion of the bases. This method has shown improvement compared to the MNF method (Hu *et al.*, 2009).

By transforming image data from the spatial domain to the frequency domain by using a Fourier transform, the application of a filter can smooth noise (Kosarev and Pantos, 1983). In that paper, the Fourier transform was performed and the standard deviation of the high frequency portion of the Fourier spectrum was calculated. An optimal Wiener filter (WF) (Kosarev and Pantos, 1983) was computed from the mixed signal, and the noise Fourier spectrum that was used to filter noise. After multiplying the WF coefficients to the Fourier amplitudes, an inverse Fourier transform was performed to recreate the original data. This method of filtering, in addition to noise reduction, is also well suited to data compression as only a small portion of the Fourier spectrum is used in reconstruction (Kosarev and Pantos, 1983).

Similar to the Fourier method, noise can also be reduced in spectra by using wavelets. The main difference between the wavelet transform and the Fourier transform is that the wavelet transform is in the spatial domain. The discrete wavelet transformations have shown to out-perform other smoothing methods, such as the Savitsky-Golay filter, with little tuning of the transform (Schmidt and Skidmore, 2004).

2.4.2 De-Striping

By using multiple detector elements along track, push-broom sensors (two dimensional detector arrays), can create linear along track errors in the image. This could be caused by radiometric calibration of the detectors changing over time. In sensors with linear detector arrays, the striping will occur in the across-track dimension. This well-known phenomenon has occurred with many prominent sensors, both linear-detector type such as: Landsat Multispectral Scanner (MSS), and Landsat TM, and push-broom type: SPOT HRV and Hyperion (Sun *et al.*, 2008). The algorithms for de-striping images can be grouped into two categories: digital filtering and statistical approaches (histogram matching, and moment matching ; (Sun *et al.*, 2008).

2.4.2.1 Digital Filtering

Digital filtering of striping in the frequency or spatial domain, by use of a Finite Impulse Response (FIR) filter, is excellent for artefacts that are consistent and periodic in nature such as striping (Simpson *et al.*, 1995). However, the spatial frequency components that are filtered out are not purely determined by the striping effect. Therefore, this method will lose some information. Furthermore, it is difficult to determine the filter size. In the Fourier domain, periodic striping noise will be shown as peaks at certain frequencies. By placing a Gaussian function, the FIR filter, the striping can be filtered out over the noise peak. This method is best used in situations where the sensor's striping characteristics are not dynamic due to the manual nature of frequency selection for striping removal.

2.4.2.2 Statistical Approaches

Histogram matching, as one of the earliest methods used for de-striping images, assumes that the probability function of the radiance of the scene that is measured by the detector is the same (Horn and Woodham, 1979). Therefore, the statistics of the sensor outputs can be used to adjust the gain and offsets for the detectors to eliminate striping. The disadvantage of this method is that the corrections made are only relative to the other detectors and may not be representative of the actual radiance values of the scene.

Lastly, moment matching has been used to remove striping in both airborne and satellite images. Moment matching, which is similar to histogram matching, uses local statistics to correct imagery for striping. The main difference between histogram matching and moment matching is that moment matching only adjusts the moment statistics, mean and standard deviation, in a given band to match a reference band or bands (Sun *et al.*, 2008). Moment matching can be carried out in the spatial domain taking advantage of spatial autocorrelation. With hyperspectral imagery, it can also be processed in the spectral domain, taking advantage of the spectral autocorrelation properties of hyperspectral bands. Spectral moment matching is superior to spatial moment matching when de-striping data (Sun *et al.*, 2008).

2.5 Atmospheric Effects

Radiant energy that reaches an imaging sensor system is converted into data that capture that energy's intensity and spectral characteristics (Peddle *et al.*, 2003). Raw data that are originally acquired by a remote sensor contains systematic and non-systematic errors inherent to both the process of signal acquisition and the intervening atmosphere

(Peddle *et al.*, 2003). Raw data are converted to at-sensor radiance through the use of calibration coefficients typically supplied with the data as follows:

$$Q = (G * L_\lambda) + B \tag{2.7}$$

where Q is the DN recorded, G is the gain, B is the offset and L_λ is the at-sensor radiance at a given wavelength λ.

At-sensor radiance of radiant energy reflected from the Earth and the atmosphere is composed of three components (Schaepman-Strub *et al.*, 2006) :

$$L_\lambda = L_\lambda^{atm} + L_\lambda^{dir} + L_\lambda^{diff} \tag{2.8}$$

where L_λ^{atm} is the scattered light from the atmosphere, but not reflected from the surface, (also called the path radiance) and L_λ^{dir} is the direct reflected radiance passing straight through the atmosphere to the sensor shown in Figure 2.3. The diffuse portion of reflected radiant energy, L_λ^{diff}, reaching a sensor is composed of three components as follows:

$$L_\lambda^{diff} = L_\lambda^{diff1} + L_\lambda^{diff2} + L_\lambda^{diff3} \tag{2.9}$$

where L_λ^{diff1} is the contribution of diffuse radiant energy transmitted directly to the surface and then scattered by the atmosphere towards the sensor after being reflected by the surface , L_λ^{diff2}, is the contribution of diffuse radiant energy scattered by the atmosphere before reaching the surface and then directly transmitted to the sensor and L_λ^{diff3} is the contribution of radiant energy scattered by the atmosphere both before and after being reflected from the surface as illustrated in Figure 2.4 (Schaepman-Strub *et al.*, 2006).

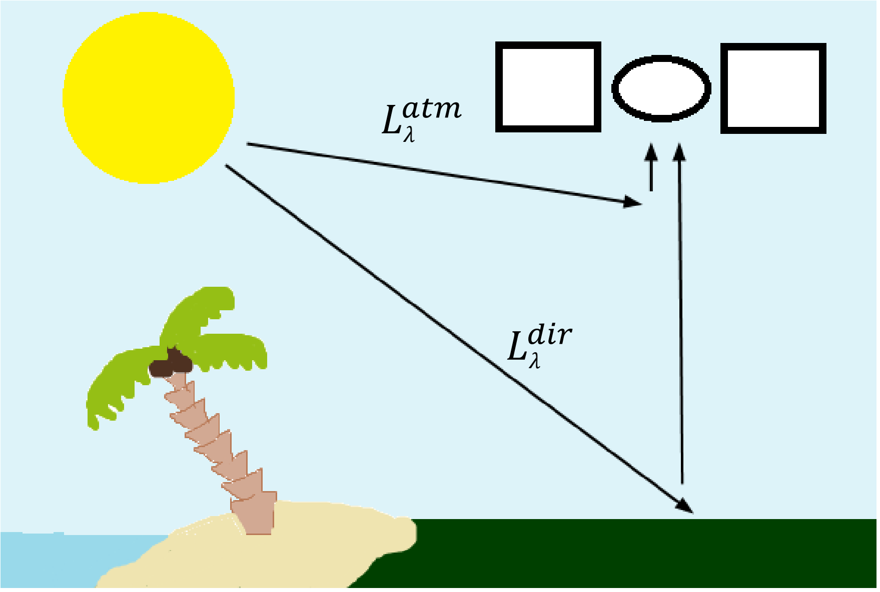


Figure 2.3: Atmospheric component also referred to as path radiance, L_λ^{atm}, and ground direct reflectance radiance, L_λ^{dir}, as measured by a remote sensing sensor.

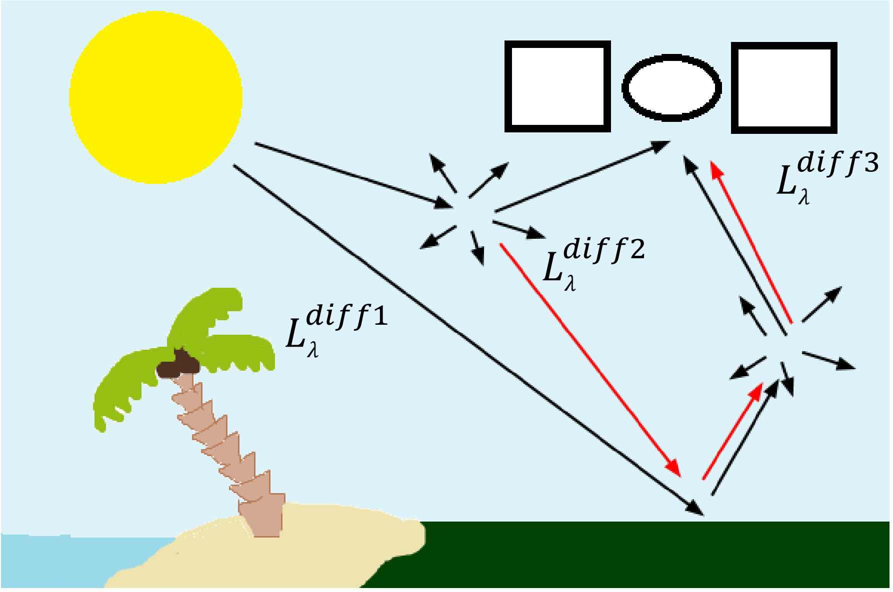


Figure 2.4: The three diffuse portions of radiation measured by a remote sensing system: (L_λ^{diff1}) reflected from the surface prior to being scattered, (L_λ^{diff2}) radiant energy scattered and directly transmitted into the sensor, and (L_λ^{diff3}) is radiant energy that is scattered before and after being reflected from the surface.

The measured amount of the at-sensor radiance when taking into account the target's reflective properties, atmospheric conditions and geographic location can be expressed as:

$$L_{\lambda} = E * \rho * T + L_P \ , \tag{2.10}$$

where E is the amount of down-welling radiance, which is a function of geographic location and image acquisition time, ρ is the reflective amount of the surface, and T is the atmospheric transmittance which is governed by many factors including, temperature, molecular abundances, particular matter, and atmospheric gases. L_P is the path radiance which is also governed by the composition of the atmosphere and the quantity of down-welling radiance, all of which is wavelength dependent. L_{λ}^{diff1} and L_{λ}^{diff3} are included in the $E * \rho * T$ component, and L_{λ}^{diff2} is included in the L_P component in this equation.

2.5.1 Absolute Atmospheric Correction

Measured at-sensor radiance of the same land surface can vary due to time of acquisition and/or atmospheric properties. Therefore, it has become common practice to convert the measured at-sensor radiance to surface reflectance to allow for the accurate estimation of biophysical or biochemical parameters of interest. In general, this correction is carried out using two different approaches. The first uses information collected on the ground to remove the effects of illumination empirically and atmospheric scattering and gaseous absorption. The second makes use of atmospheric radiative transfer models, which simulate the radiation transfer through the atmosphere. They remove the effects of atmospheric scattering and gaseous absorption under varying illumination and viewing conditions (Staenz *et al.*, 2002).

2.5.1.1 Empirical Line Method

The empirical line calibration (ELC) method is a procedure that will absolutely calibrate raw RS data to surface reflectance using a linear function. The ELC method is based on the following equation:

$$DN_b = \rho_b A_b + B_b \tag{2.11}$$

where DN_b is the digital number for a given pixel in band *b*, ρ_b is the reflectance of the surface materials of a pixel in band *b*, A_b is the multiplicative term which affects the DN (transmittance and instrumental factors used for calibration), and B_b is the additive term (primarily atmospheric path radiance and instrumental calibration offset) (Farrand *et al.*, 1994). ELC is used to solve for the gain value of A and the offset value of B. By plotting the recorded data values for a dark homogeneous target and a light homogeneous target collected on the ground against data from the same targets collected by a sensor, which is affected by an atmosphere, the coefficients for the linear equation can be calculated as shown in Figure 2.5.

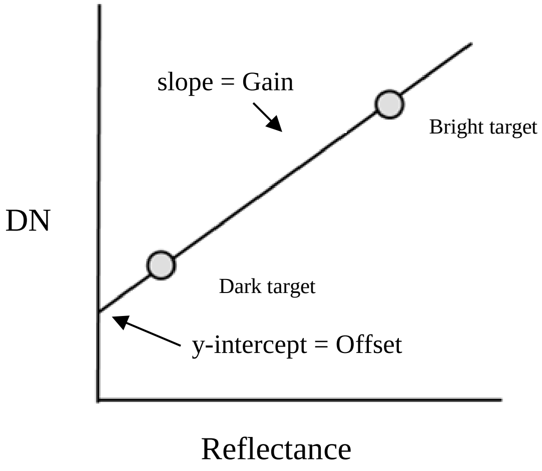


Figure 2.5: ELC method used to determine gain and offset values for atmospheric correction of remotely sensed data using ground-collected data.

2.5.1.2 Radiative Transfer Model

The use of a radiative transfer (RT) model to remove the atmospheric scattering and absorption effects has become an effective tool when processing remotely sensed data. A RT model estimates the atmospheric transfer, the amount of irradiance and other atmospheric parameters for a given location and time. When used with RS data to compensate for the unwanted atmospheric effects, a look-up table can be generated that contains atmospheric transfer for varying viewing geometry, solar irradiances, and atmospheric parameters, such as aerosol optical depth and columnar water vapour (Staenz and Williams, 1997).

Although RT codes contain atmospheric models with standard atmospheric parameters that can be chosen for data correction (Berk *et al.*, 1999), the effectiveness of the RT model is increased if the user has access to measured atmospheric parameters to be used in the computation (Peddle *et al.*, 2003). The values for atmospheric parameters can be collected in-situ, retrieved from ground stations, or estimated from the image itself.

2.5.2 Water Vapour Column Content

Water vapour in the atmosphere absorbs EM radiation at specific wavelengths. Atmospheric water vapour content is an important parameter for atmospheric correction. Furthermore, the effectiveness of using an RT model for the purpose of atmospheric correction is limited by the accuracy of the atmospheric input parameters used. The accuracy of RT model-based atmospheric corrections increases when accurate, representational values for this parameter are available (Farrand *et al.*, 1994). Water column content measurements can be collected in the field using a sun photometer or obtained from permanent base-stations, radiosonde data, or estimated from optical remote sensing data.

In cases where sensors are designed to collect data for world-scale temporally continuous information products, it may not be practical to collect field data. Therefore, the best option is to develop sensors that can extract atmospheric parameters independently for atmospheric correction purposes. For example, Sentinel-2 MSI has been designed with the potential to extract atmospheric parameters, like atmospheric water column content and aerosol optical depth, for correction purposes.

Atmospheric water column content can be estimated from optical remote sensing systems using differential spectral absorption techniques. In general, differential spectral absorption techniques involve a ratio between radiance in bands located in a water vapour absorption feature and an interpolated radiance of a near reference band unaffected by water vapour absorption as depicted in Figure 2.6 (Schläpfer *et al.*, 1998).

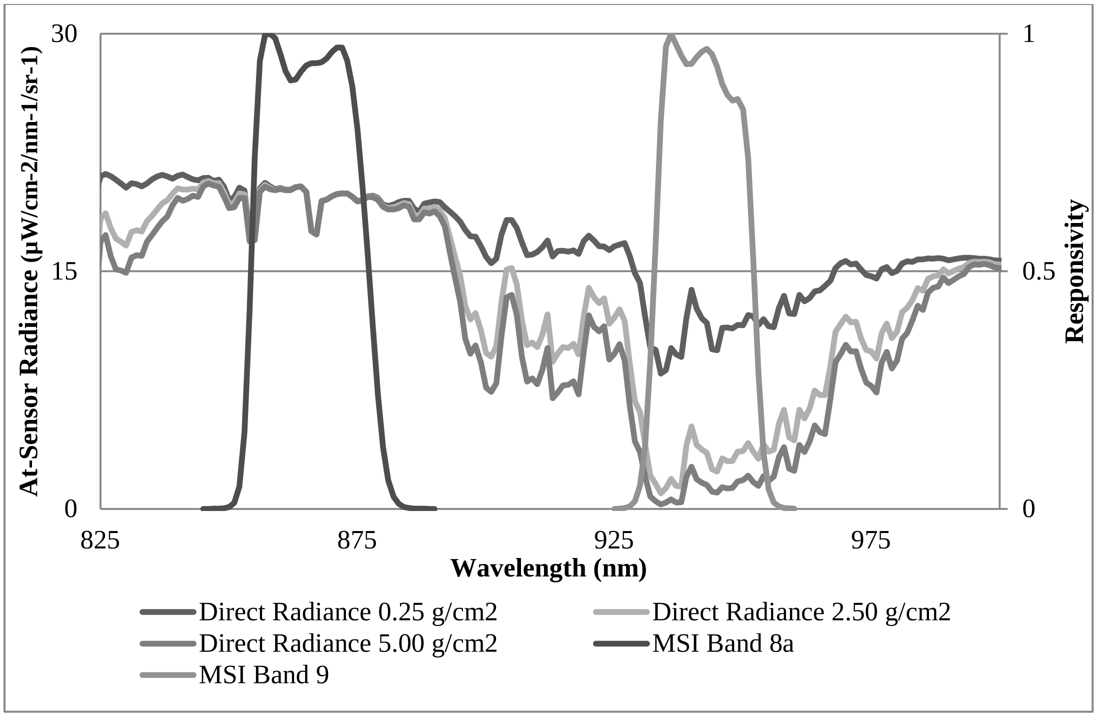


Figure 2.6: At-sensor radiance for three quantities of atmospheric water column content. The RSRF for MSI band 9 (water absorption feature) and band 8a (non-absorption feature) are shown.

Which differential absorption method is used is controlled by the number of spectral bands and their widths. This method has been applied to both hyperspectral and multispectral sensor data and is commonly used in model-based atmospheric corrections (Staenz *et al.*, 2002, Gao and Goetz, 1990, Sobrino *et al.*, 2003, Schläpfer *et al.*, 1998, Richter and Schläpfer, 2008). Using hyperspectral data can increase the accuracy of estimations when using the curve fitting method, which uses multiple bands, over other methods (Staenz and Williams, 1997).

Scattering of radiation due to particles in the atmosphere can affect the atmospheric transmission. Aerosol optical depth of the atmosphere defines the amount of scattering caused by the suspended particles, such as small dust particles, aerosols, fossil fuel combustion products or suspended sea salts (Schott, 2007). Because the aerosol optical depth is related to the transmission of the atmosphere, being able to estimate its magnitude has been an important parameter for atmospheric correction of remotely sensed data. Methods have been developed and implemented to estimate aerosol optical depth over land (Kaufman *et al.*, 1997) as well as water (Tanré *et al.*, 1997).

2.6 Land-Cover Mapping

Land-cover information at global, continental, and regional scales is an essential product for use in scientific research, economic and government planning. This information is essential for governments to overcome the problems of uncontrolled development, deteriorating environmental quality, loss of prime agricultural lands, destruction of important wetlands, and loss of fish and wildlife habitat (Anderson *et al.*, 1976).

Terrestrial ecosystems exert considerable control on the planet's biogeochemical cycles, which in turn significantly influence the climate system through the radiative properties of greenhouse gases and reactive species (Friedl *et al.*, 2002). Vegetation cover affects current climate state and plays an important role in climate forcing. At the same time, the main factor controlling the distribution of natural vegetation is climate (Bartholomé and Belward, 2005).

If living conditions and standards are to be improved or maintained at current levels, land-use data are needed in the analysis of environmental processes (Anderson *et al.*, 1976). Knowledge of the present distribution and area of agricultural, recreational, and urban lands, as well as information on their changing relative proportions, is needed to determine better land-use policy, to project transportation and utility demand, to identify future development pressure areas and to implement effective plans for regional development (Anderson *et al.*, 1976).

Although recognized as an important source of information required for policy development research and planning, the utility of suitable land-cover and land-use information is limited by temporal variability, mapping scale, inaccuracies, or regional availability (Foody, 2002). Still, many land-cover products continue to be produced from RS data (Bartholomé and Belward, 2005, Giri *et al.*, 2005, Friedl *et al.*, 2002, Latifovic *et al.*, 2004, Hansen *et al.*, 2000, Cihlar *et al.*, 1996).

2.6.1 Land-Cover Classification Systems

Anderson (1976) published a framework outlining the major reasons to perform land-cover mapping using remotely sensed data. This initial research has been the baseline for modern land-cover class schemes. For example, the United States' national land-cover dataset (NLCD), which uses a 21 class system, is based on the proposed Anderson classes (Homer *et al.*, 2007).

The Coordination of Information on the Environment (CORINE) land-cover mapping scheme was developed between 1985 and 1990 by the European Commission to collect and store environmental information. The CORINE scheme has been used with multiple types of remote sensing systems including the Landsat, and SPOT series (See and Fritz, 2006). It has three tiers of classes with 5, 13 and 44 classes, respectively. This class system is primarily used in Europe. It can also use ancillary data that is not necessarily derived from remotely sensed data to aid in the classification process (Bossard *et al.*, 2000).

Similar to the CORINE systems, the United Nations introduced the Land-Cover Classification System (LCCS), which was developed not only to rely on the spectral information that is collected by remote sensing systems, but also data that can be collected *in-situ*, such as biophysical parameters. The LCCS is also unique as there is no standard set of land-cover categories. Because this system is dynamic, it is highly transferable to other land-cover classification systems (Di Gregorio and Jansen, 2000). It was the foundation for ESA's Globcover project (GLC2000), a global land-cover classification project (Arino *et al.*, 2007), which utilized MERIS data collected over a 18-month period to create a 300-m land-cover classification of the entire surface of the Earth

A global land-cover product was also produced from MODIS data (Friedl *et al.*, 2002). This product has been compared for similarities and differences to the GLC2000 product. The issues that have arisen are transferable to other land-cover classification systems and sensors. The main differences between the MODIS land-cover product and other products is that by using a single sensor, the methodology of the MODIS land-cover product is repeatable due to the fact that the it does not involve experts outside of the development team, and it does not have international ownership. However, it can impede the utilization of the data by other academics or interested parties. On the other hand, the GLC2000 product is produced at a regional to national scales by a multitude of different organizations using different methodologies which may not be reproducible (Giri *et al.*, 2005).

Major land-cover products have been created in Canada in recent years to meet the needs of specific sectors (DB-Geoservices, 2008). The Canadian Forest Service (CFS) created the Earth Observation for Sustainable Development of Forests (EOSD) land-cover product (Wulder *et al.*, 2008a), while the Agriculture and Agri-Food Canada (AAFC) generated a land-cover product for the National Land and Water Information Services (NLWIS) (Fisette *et al.*, 2006). Both land-cover products are 30-m resolution and were created with Landsat TM data with the assistance from provinces. Although both land-cover classes were independently developed, the CFS legend was made compatible with the LCCS class scheme (Wulder *et al.*, 2008a). Because the products were sector-specific, national coverage was not necessary. However, the products were joined with an Arctic ecosystem land-cover product produced by the Canada Centre for Remote Sensing (CCRS), using a harmonized legend so that a single continuous coverage of Canada was created (DB-Geoservices, 2008), entitled the Land Cover Circa 2000 (LCC2000) land-cover product (Olthof *et al.*, 2009).

2.6.2 Importance of Remotely Sensed Data

The ability of remote sensing technologies to determine biophysical and biochemical parameters using the spectral reflectance properties of different vegetation materials has been the motivation for studies to map large regions. For example, chlorophyll and carotenoid content in leaves both influence the reflectance in the 400-nm to 750-nm wavelength interval (Gitelson *et al.*, 2003, Thomas and Gausman, 1976).

This knowledge of specific important wavelength regions has been used to generate vegetation indices, such as the Normalized Difference Vegetation Index (NDVI) to assist in the mapping of vegetated surfaces using remote sensing technologies (Rouse *et al.*, 1973, Herold *et al.*, 2008). In addition, minerals exhibit specific absorption features at certain wavelengths depending on the mineral properties, such as chemical structure (Hunt, 1977). Therefore, these unique spectral features have the potential to be used as diagnostic features for identification purposes (Clark *et al.*, 1990). Overstory and understory vegetation, soils, water bodies, and other surface features, which together make up land cover, often have differences in their reflectance characteristics that help to distinguish them (Peddle *et al.*, 2003).

2.6.3 Data Classification

Land-cover classes are, by definition, generally homogenous in nature within the context of a given classification. A land-cover product that is to be generated from an RS image data requires the pixels in the image to be grouped into categories that are analogous with the land-cover classification system. When classifying an image on a per-pixel basis, the individual pixels in RS image can be categorized into groups based on the spectral characteristics of each pixel. Image data can be classified using either supervised or unsupervised classification algorithms. Unsupervised classification algorithms do not require the input of training data. The user inputs the number of classes the image is to be grouped into, and the algorithm processes the data, grouping the pixels into statistically similar classes. This method does not label classes. Supervised classification on the other hand requires the input of training data that will be used in the classification process. Different algorithms may be better suited for different types of data. Therefore, new types of data require appropriate classification algorithms (Smits *et al.*, 1999). Furthermore, it is important to evaluate the use of classification algorithms in addition to the accuracy of classification.

Sampling of image pixels is used to train supervised classifiers as well as validate the classification. Important requisites of a sample scheme are that it sufficiently represents the land-cover or land-use category to be mapped, and that no biases will be introduced by the sample scheme itself (Smits *et al.*, 1999). Classically, the practical minimum is to have sampling sizes that are 10 times the number of bands in a dataset (Schott, 2007).

Smits *et al*. (1999) reviewed various schemes addressing the issues of sampling procedure and sample size. The paper states that simple random sampling is not suitable for categories that are small as they will become under-sampled. However, this problem can be remedied by using prior knowledge of the area in combination with stratified random sampling. Lastly, sampling of categories this way may create bias as the locations sampled are generally in close proximity to roads and may not accurately represent the area being sampled.

2.6.3.1 Support Vector Machine

Sampling schemes for training data for use with parametric classifiers such as maximum likelihood require large sample sizes so that the classes are accurately represented. This process can be very time-consuming and costly. One alternative to this has been to use the support vector machine (SVM) classifier, which due to the nature of the algorithm requires less training samples.

SVM classifiers make use of a hyper plane to separate classes. The points (pixels) that are close in proximity to the hyperplane are the support vectors that are used to train the classifier. Studies have shown that SVM classifiers can outperform maximum-likelihood, neural network, and decision tree classifiers (Bartzen *et al.*, 2010, Foody and Mathur, 2004). Furthermore, similar accurate results can be obtained using small sample sizes that contain small amounts of pixels with mixed spectral characteristics (Foody and Mathur, 2006, Mathur and Foody, 2008).

2.7 Summary

Simulation studies involve a thorough understanding of the sensor characteristics, how these are simulated from other data types, and the effects of the atmosphere on the signal being simulated. Pre-processing of data is an essential step in the simulation of future satellite data. To understand the components and types of sensor noise and how they affect data quality is essential for simulation studies where the goal is to create realistic and accurate data. Understanding the affect the atmosphere has on remotely sensed data, specifically the atmospheric gases and aerosols are important for the process of correcting remotely sensed data. Furthermore, knowledge of these phenomena is required in data simulation studies.

Land-cover products generated from remote sensing systems are an essential source of information for environmental and government organizations. However, these information products need to be generated in such a way that they are representative of the area under consideration. Furthermore, this type of product can be used to evaluate future remote sensing sensor systems through simulation. By performing the evaluation prior to launch, information products can be assessed immediately and sensor performance can be determined. The SVM classification approach has shown that it is a robust method for classification of multispectral data with minimum training data inputs.

3. METHODS

3.1 Introduction

This chapter will outline the procedures and equipment used through the process of simulating Sentinel-2 MSI data, as well as the generation of land-cover data-products and the evaluation procedures used. The conversion of the airborne data from its raw form into an at sensor radiance product is the essential first step in the pre-processing procedures. These data were then used in concert with radiative transfer code to simulate Sentinel-2 MSI at-sensor radiance. As no commercially available system exists to simulate the spectral and spatial properties of the Sentinel-2 MSI sensor, the procedures used for simulating the spectral and spatial aspects of the sensor were derived from mission documents and implemented in ENVI. This simulation included a full simulation of known noise parameters for this instrument. The full simulation processes are outlined in Figure 3.1.

For comparison purposes, Landsat 5 TM and SPOT 5 HRV data were used to create standard land-cover products to evaluate the simulated Sentinel-2 MSI land-cover product.

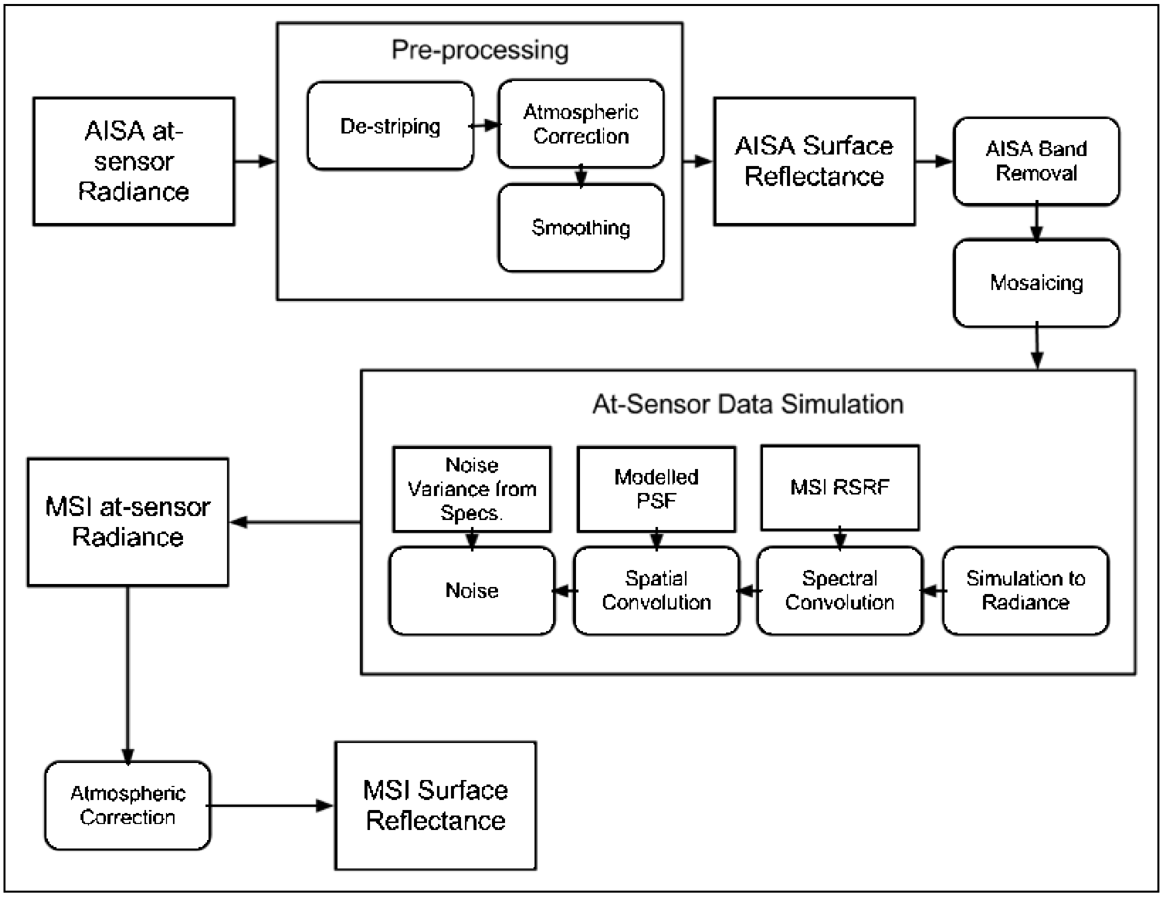


Figure 3.1: Flow-chart outlining AISA pre-processing steps and simulation of MSI data.

3.2 Study Site

The test site is located in the Vermillion River watershed near Edmonton, Alberta: 53° 30’ 16” N: 111° 29’ 31” W (Figure 3.2). This watershed is within the North Saskatchewan watershed. It covers approximately 7,826 km^2, which is about 11% of the North Saskatchewan watershed. The area receives on average 373.6 mm in precipitation a year, making it a semi-arid environment, where the majority of the precipitation falls in the summer months. Currently, the land use in the area is predominantly agricultural with 51% of the watershed is being used for crops, 46% is grassland, 1% for forage crops the remaining land use is forested. Data from the Prairie Farm Rehabilitation Administration

(PFRA) Land Classification estimates 0.2% of the land is wetlands. However, another inventory report, completed by Ducks Unlimited Canada, estimated that the wetland percentage was closer to 5.6% (NSWA, 2005).

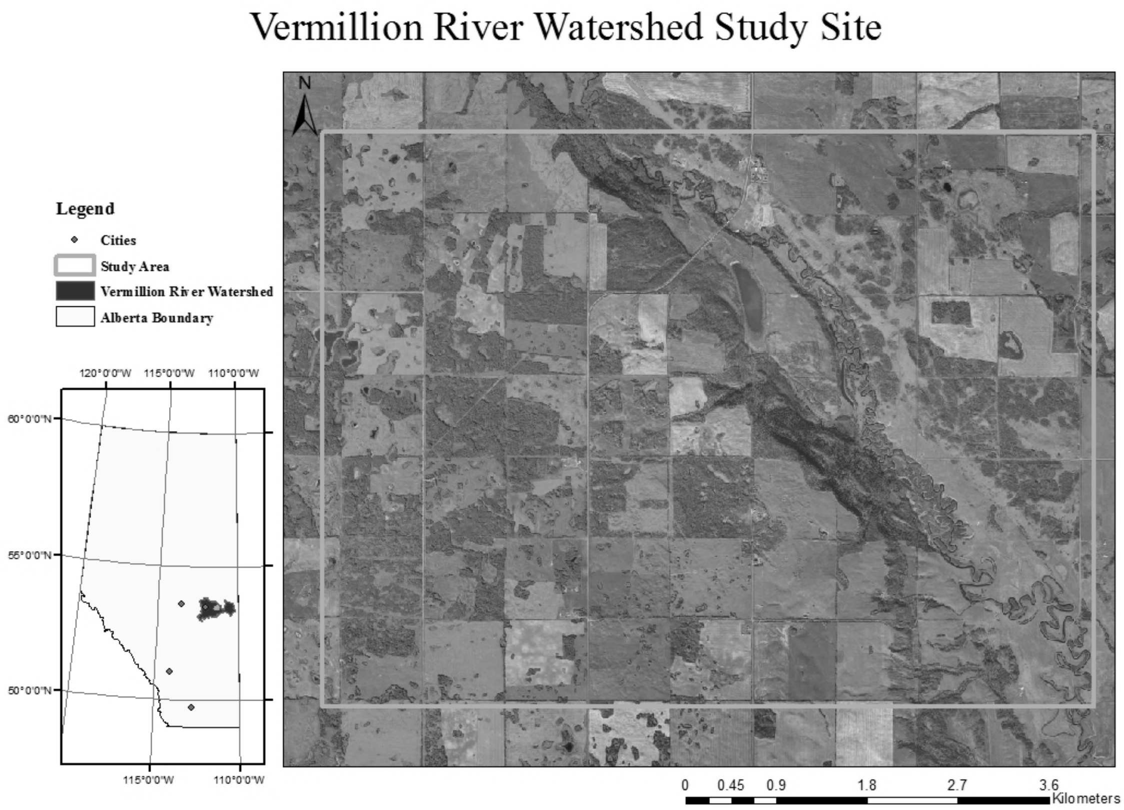


Figure 3.2: Orthophoto of study site. The box indicates the area in which the AISA data was collected that was used in the simulation.

3.3 Hyperspectral Airborne Data

The Airborne Imaging Spectrometer for Application (AISA) sensor was used to collect the hyperspectral image data over the study site. AISA is a hyperspectral push-broom remote sensing instrument with 256 detector elements on each of the two sensors, one in the visible near infrared (VNIR) and one in the short wave infrared (SWIR). Sensor arrangement for this mission consisted of 472 bands between 400 nm and 2400 nm. The bands in the VNIR have a bandwidth at FWHM of approximately 2.3 nm at a spectral sampling equal to the FWHM. In the SWIR region, the spectral resolution increases to approximately 6 nm with a sampling frequency again equal to the FWHM so as to maintain spectral band overlap of at least 50%.

The flying height above ground of the aircraft was approximately 1500 m . At this flying height, the 256-sensor array has a GIFOV of 2 m. The entire area was imaged on July 27th, 2010 and required 13 flight lines flown in an East-West orientation. The metadata for these flight lines are presented in Table 3.1. Metadata for the flight line 12 was interpolated from the other data due to not being recorded when the data was collected.

Table 3.1: Metadata for the 13 AISA hyperspectral flight lines that cover the study site including the heading direction, acquisition time, and the geographic centre of the flight line. The metadata for flight line 12 has been interpolated.

Line	**Heading (°)**	**UTC Time**	**Latitude North (° N)**	**Longitude (° W)**	**Flying height above ground (km)**
1	86	15:52:11	53.509	-111.359	1.561
2	278	15:56:01	53.506	-111.359	1.560
3	86	16:02:53	53.501	-111.359	1.558
4	277	16:08:11	53.497	-111.358	1.553
5	87	16:13:20	53.493	-111.360	1.496
6	278	16:18:46	53.489	-111.360	1.506
7	87	16:24:18	53.485	-111.361	1.498
8	277	16:29:42	53.481	-111.360	1.488
9	88	16:35:10	53.477	-111.360	1.474
10	88	16:45:44	53.469	-111.362	1.476
11	277	16:51:12	53.465	-111.360	1.486
12	88	16:56:37	53.462	-111.363	1.481
13	277	17:01:18	53.473	-111.361	1.478

3.3.1 Data Pre-processing

The pre-processing of the data was performed using the Imaging Spectrometer Data Analysis System (ISDAS) developed at CCRS (Staenz *et al.*, 1998). ISDAS is a suite of tools for processing and analysing hyperspectral data .

3.3.1.1 Noise Removal

Systematic and random noise was present in the hyperspectral data. It was important to remove or limit the amount of noise prior to simulation so that the noise was not compounded further during simulation (Khurshid *et al.*, 2006). The noise present was both random Gaussian and striping which was caused by the detectors.

Spectral moment matching was used to remove the striping in the hyperspectral airborne data. It uses local statistics to correct imagery for striping. Random noise caused by the sensor electronics was reduced by smoothing the radiometrically and atmospherically corrected reflectance data using a 9-band wide window.

3.3.1.2 Correction to Surface Reflectance

As electromagnetic radiation is emitted from the Sun, it interacts with the Earth's atmosphere and scattering, transmittance, and absorption occur. Atmospheric models, which simulate RT through the atmosphere, were used to remove the effects of atmospheric scattering and gaseous absorption under varying illumination and viewing conditions. The majority of the atmosphere is close to the Earth's surface. Therefore, due to the aircraft collecting data in this region the effects of the atmosphere need to be removed. For this study MODTRAN 4.2 RT code was used.

An atmospheric correction was applied to the airborne radiance data using MODTRAN 4.2 and a reverse look-up table approach was used to determine water vapour amounts and viewing angle differences (Staenz and Williams, 1997). Solar irradiance quantities at the sensor altitude and acquisition time were calculated in MODTRAN using the metadata presented in Table 3.1. Additional MODTRAN input parameters are shown in Table 3.2. Water vapour estimates were calculated using an absorption curve fitting approach centered over the water absorption feature at 940 nm following the techniques used by (Gao and Goetz, 1990).

Table 3.2: MODTRAN input parameters used in the atmospheric correction.

Wavenumber step	1cm^{-1}
Atmospheric layers	33
Atmospheric model	Sub-arctic summer
Aerosol model	Continental (rural)
Aerosol optical depth	60 km
Water vapour	10 values (0.5 g/cm^2 – 3.0 g/cm^2)
C02 mixing ratio	As per model
Ozone column	As per model

3.3.1.3 Geometric Correction

Geometric correction was conducted to reduce or eliminate the effects of the aircraft movement during collection. The 13 flight lines were geometrically corrected using data recorded by the plane's inertial navigation system relating to pitch, yaw, and, roll of the aircraft. A lidar-derived digital elevation model (DEM) sampled at the same time as the images was used for orthorectification of the image data.

3.3.1.4 Mosaicking

Mosaicking is the process of assembling a single image from the individual flight-line images. Because solar and viewing angles changed during data acquisition, or errors may be introduced in the previous processing steps, there may be differences between flight-lines. By performing histogram-matching to the mosaic, these differences were adjusted to match the flight-line’s radiometery. Histogram-matching calculates the statistical histogram within the overlap region of two flight-lines and then will calculate and apply a transfer function to one of the images. This processing step was carried out using the Geomatica image analysis software from PCI Geomatics 2012.

3.4 Atmospheric Simulation

In order to simulate realistic at-sensor radiance, the MODTRAN 4.2 RT code was applied to the reflectance data of the scene to move the latter to the MSI altitude. Atmospheric parameters used in the MODTRAN calculation were the same as in the previous atmospheric correction. The exceptions are the water column content and the sensor viewing angle. The former was held constant at 2.11 g/cm2 and the effect of the viewing angle on the pixel was ignored due to the small size of the test site. Accordingly, it was impossible to simulate the full swath width of MSI (290 km).

3.5 Sensor Simulation

3.5.1 Spectral Simulation

Simulation of multi-spectral sensor data requires the convolution of higher resolution hyperspectral data using measured or modelled RSRF of the multi-spectral instrument. RSRFs can be modelled with a 2-dimensional Gaussian distribution function where the FWHM of the curve is equal to the proposed sampling width of the band at 50% signal strength (Verhoef and Bach, 2003).

For this study the RSRF of the MSI, which have been measured at 1 nm intervals, were provided by ESA (ESA, 2007) (Table 2.1).

3.5.2 Spatial Simulation

As the AISA hyperspectral airborne data were captured at a smaller spatial resolution than the MSI data, a spatial simulation was required to provide an estimate of the resultant data products. Spatial simulation of satellite data from airborne data is performed by a spatial convolution of the high-spatial resolution airborne data to the lower spatial resolution of the satellite data. The spatial convolution for this simulation was performed in the spatial domain using a modelled PSF. The modelled PSF is a three dimensional Gaussian function where the spatial FWHM of the function is equal to the ground sampling size of the various bands of the sensor. The ground sampling size and the converted Gaussian function to kernel size for the 2 m airborne data are presented in Table 3.4., while the convolution process is shown in Figure 3.3.

Table 3.4: Spatial characteristics of Sentinel-2 MSI including convolution filter size

Band #	**Ground Sampling Size (m)**	**PSF Kernel Dimensions**
1	60	77 x 77
2	10	15 x 15
3	10	15 x 15
4	10	15 x 15
5	20	25 x 25
6	20	25 x 25
7	20	25 x 25
8	10	15 x 15
8a	20	25 x 25
9	60	77 x 77
10	60	77 x 77
11	20	25 x 25
12	20	25 x 25

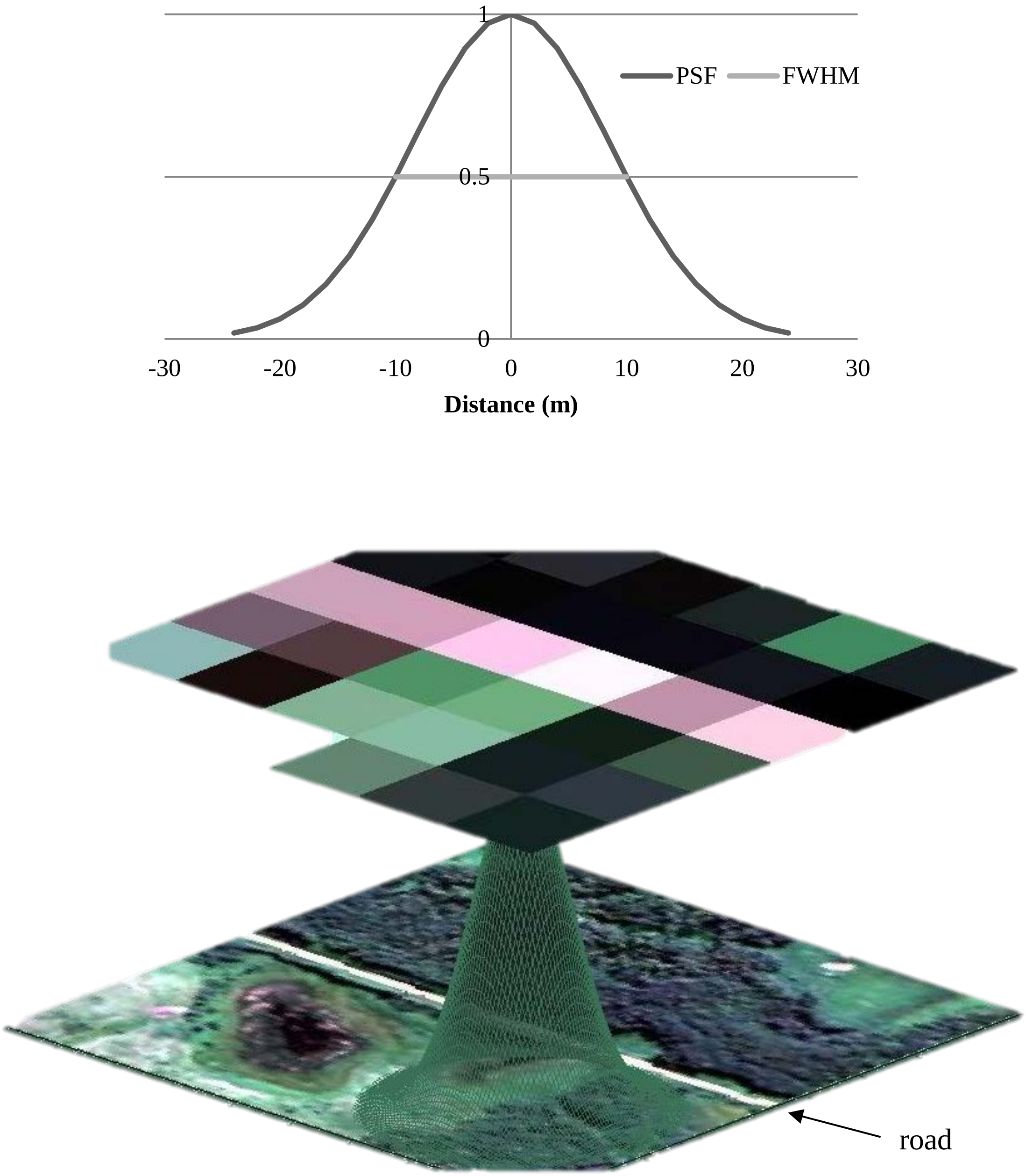


Figure 3.3: Top: example PSF of 2-m to 20-m data filter displayed in 2 dimension. Bottom: 2-m spatial resolution data (below) are convolved to 60 m (above) using a modelled PSF. Notice that the road is still visible in the 60 m pixel despite being much smaller than the pixel dimension.

3.5.3 Sensor Gain and Offset

As radiance is measured by a sensor, the output is a series of digital values quantized over the bit-depth of the sensor. The 12-bit MSI enables the radiance values to be scaled over 12 bits. The gain *G* and offset *B* were calculated as follows:

$$G = \frac{4095}{L_{min} - L_{max}} \tag{3.1}$$

$$B = 1 \text{ -} L_{min} \tag{3.2}$$

Where L_{min} and L_{max} correspond to the range of radiances the proposed detector will measure. The calculated values for G and B are presented in Table 3.5. The at-sensor radiance was then converted to digital values using gains and offsets calculated from the minimum and maximum radiance values presented in the mission document (ESA, 2007) (Table 3.5) as follows:

$$Q = (G * L_{\lambda}) + B \tag{3.3}$$

Where Q is the digital value, and L_{λ} is the measured at-sensor radiance.

Table 3.5: Maximum and Minimum radiance values to be measured by MSI and corresponding simulated gain and offset values.

Band #	$L_{\min}$ **$Wm^{-2}sr^{-1}\mu m^{-1}$**	$L_{\max}$ **$Wm^{-2}sr^{-1}\mu m^{-1}$**	**Gain**	**Offset**
1	15.97	587.87	7.16	-114.35
2	11.70	615.48	6.78	-79.35
3	6.49	559.01	7.41	-48.10
4	3.31	484.13	8.52	-28.19
5	2.61	449.55	9.16	-23.911
6	2.06	412.92	9.97	-20.53
7	1.67	387.08	10.63	-17.74
8	0.95	307.80	13.34	-12.68
8a	0.95	307.80	13.35	-12.68
9	0.51	232.91	17.62	-8.99
10	0.06	83.00	49.37	-2.96
11	0.40	69.78	59.02	-23.61
12	0.10	24.60	167.14	-16.71

3.5.4 Noise

When performing satellite data simulation where the spectral and spatial resolution of the sensor is lower than the input data, the simulated data will have an unrealistically high SNR (Börner *et al.*, 2001), due to the fact that the input instrument's sensor noise will be filtered out. Therefore, it is required to add noise back to the simulated data to simulate the resultant data properly. Although noise in optical remote sensing systems is additive in nature, due to the calibration transfer function (equ. 3.3), additive noise when applied to raw DN will become multiplicative when the raw DNs are converted to scaled radiance. It can be represented by a random zero-mean Gaussian process. Noise was added to the simulated DN data by calculating the standard deviation of the noise levels, σ, for reference radiance after being computed to DN values using equation 3.4.

$$\sigma = \frac{DN_{ref}}{SNR} \tag{3.4}$$

Where DN_{ref} is the reference radiance values presented in the mission requirement document (ESA, 2007) (Table 3.6) after being converted to DN. For each band, a zero-mean Gaussian probability function was then generated and a random number was added to each pixel to simulate noise. This process was done in MATLAB; example of the script is presented in Appendix A.

Table 3.6: Metadata used for simulating random Gaussian noise.

Band #	**L_{ref} $Wm^{-2}sr^{-1}\mu m^{-1}$**	**SNR @ L_{ref}**	**L_{ref} in DN**	**Value of σ**
1	129.11	129	810	6.28
2	128.00	154	789	5.12
3	128.00	168	901	5.36
4	108.00	142	892	6.28
5	74.60	117	660	5.64
6	68.23	89	660	7.42
7	66.70	105	691	6.58
8	103.00	174	1362	7.83
8a	52.39	72	686	9.53
9	8.77	114	146	1.28
10	6.00	50	293	5.86
11	4.00	100	212	2.12
12	1.00	100	267	2.67

3.6 Ancillary Satellite Data

SPOT 5 HRV and Landsat 5 TM data were acquired over the study area to create land-cover products to be used to compare the accuracy of the simulated Sentinel-2 MSI land-cover product (Figure 3.4). The SPOT 5 HRV image was acquired on July 15, 2010, and the Landsat 5 TM image was acquired on August 25, 2010. The characteristics of these sensors compared to the specifications of the MSI are displayed in Table 3.7.

Table 3.7: Sensor characteristics of the Sentinel-2 MSI compared with SPOT 5 HRV and Landsat TM.

	SPOT 5 HRV	**Landsat 5 TM**	**Sentinel-2 MSI**
Bands	5	8	13
Wavelength Range (nm)	480 - 1750	450 - 2350	443 – 2190
Spectral Resolution (nm)	70 - 170	70 -210	15 - 180
Spatial Resolution (m)	10	30, 120	10, 20, 60
Swath Width (km)	60	185	290

Both images were converted to at-sensor radiance using the calibration gains and offsets provided in the metadata. The images were then clipped to the same geographic extent as the simulated MSI scene.

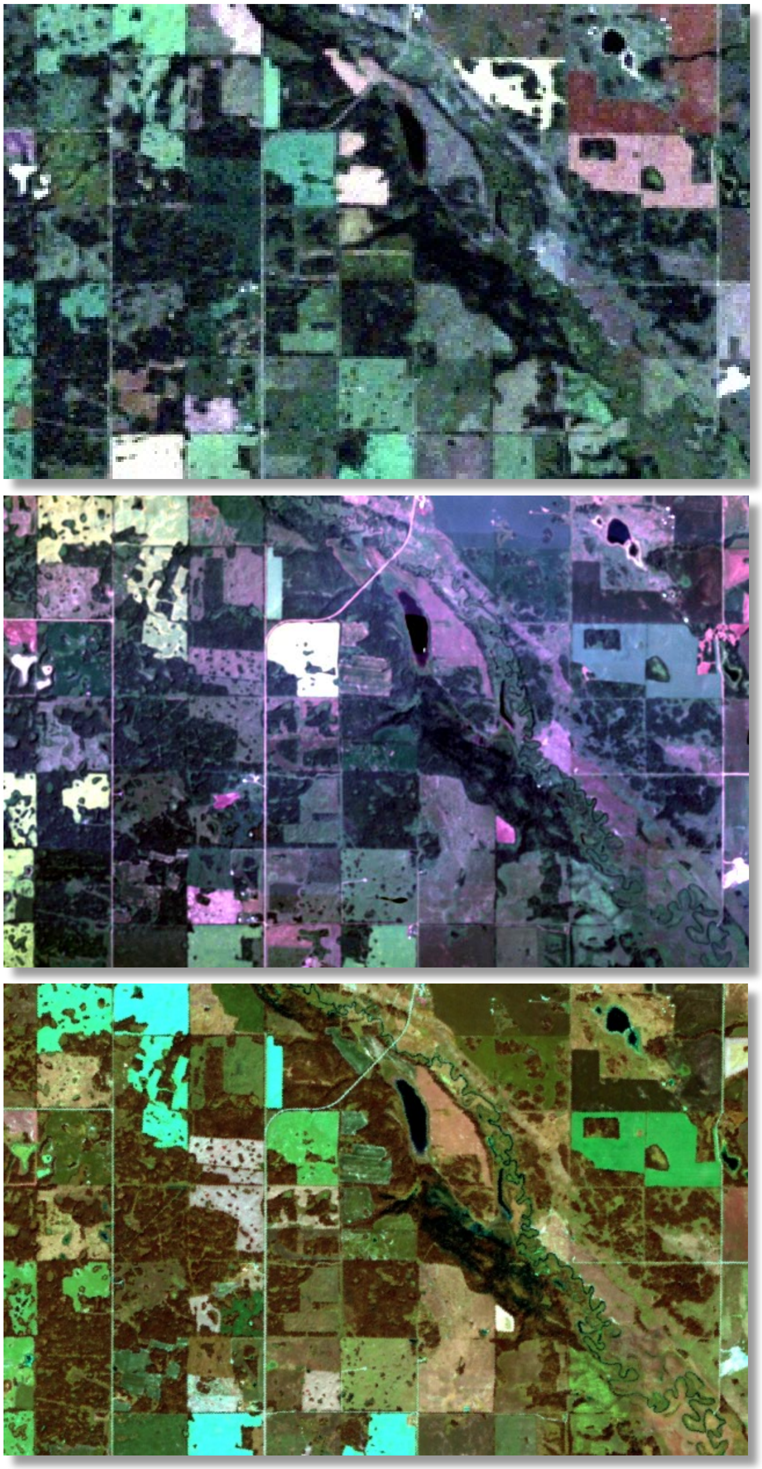


Figure 3.4: True-colour Landsat TM (top) R: Band 3 G: Band 2 B: Band 1, Simulated Sentinel-2 MSI (middle) R: Band 4 G: Band 3 B: Band 2, and Natural-colour SPOT HRV image R: Band 2 G: 0.75B1+0.25B3 and B: 0.75B1+0.25B3 (bottom).

3.7 Atmospheric Correction

The data from the three different sensors were corrected to surface reflectance using the atmospheric module FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) in ENVI/IDL software package (http://www.exelisvis.com). This atmospheric correction procedure, which uses the MODTRAN RT code, allows the user to input atmospheric, viewing and solar illumination parameters at the time of data acquisition.

3.8 Land-cover Classification

Supervised SVM classification was performed to classify the scenes using the spectral information into classes required for land-cover mapping. SVM was chosen due to its superior classification accuracy over other supervised algorithms (Foody and Mathur, 2004) and small training data requirements (Foody and Mathur, 2006). Parameters for the SVM classifier were determined to return optimal results. Seven land-cover classification categories were independently selected based on harmonized EOSD-NLWSI LCC2000 class legend to best fit the available ground reference data for the scene (Table 3.8).

Table 3.8: LCC2000 land-cover classes used in classification of simulated MSI data combined with description of class.

LCC2000 Class	Description
20 - Water	Rivers, lakes, streams
30 - Exposed land	Tilled land, gravel, road, exposed rock
80 - Wetlands	Woody, non-woody
110 – Grassland	Rangeland
121 – Annual Croplands	Hay, alfalfa
122 – Perennial Croplands	Wheat, Oats, Barely, Canola
200 – Forest	Open/closed, conifer, deciduous, mixed.

3.8.1 Training Sites

A ground survey of agricultural fields, including wetland sites, forested areas and water bodies, were collected slightly prior to the airborne data at the time of acquisition (Figure 3.5). The ground survey consisted of *in-situ* crop verification, combined with ground survey of planted crops by local landowners, and confirmed areas of wetlands and forests by use of aerial photographs and in-situ verification. From this data, classes of homogenous regions were created using Regions of Interest (ROIs) (Figure 3.6). They covered approximately 30% of the study site as shown in Table 3.9. From these ROIs 360 pixels (10 m) were collected based on an equal random sampling rate from all classes to be used for training the SVM classifier. Because the training sites used were at the 10 m resolution, the Landsat TM bands were partitioned into 9 pixels each with a spatial resolution of 10 m.

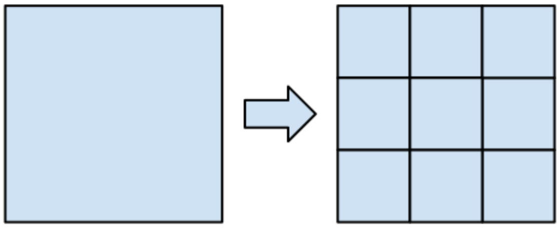


Figure 3.5: Landsat TM data was partitioned into 9 pixels per pixel so that it lined up with the ROI data. The Landsat TM data remained unchanged in space.

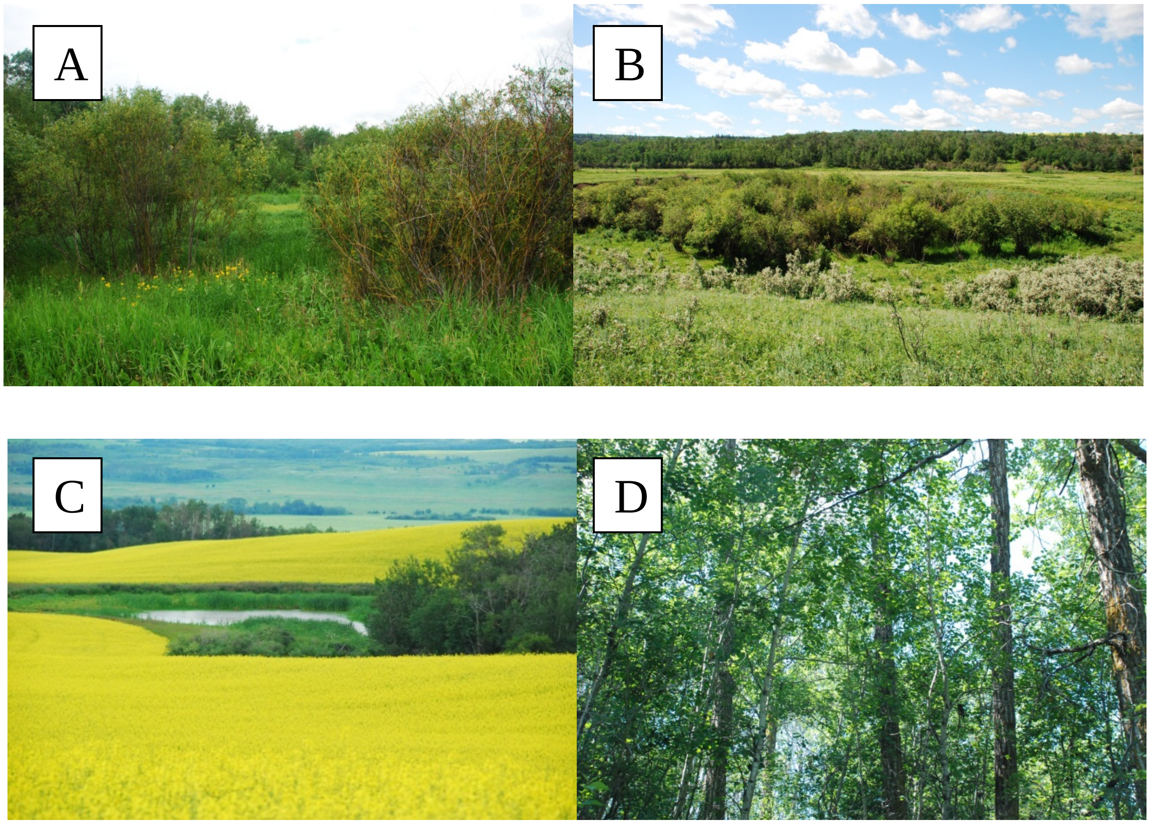


Figure 3.6: A: Wetland area with both woody and herbaceous vegetation. B: Willows surrounding a wetland area. C: Canola field D: Deciduous forest.

3.8.2 Test Pixels

Test pixels were randomly selected following the same process as the training pixels. Accordingly, 360 test pixels were randomly selected from each class to be used in the accuracy assessment of the land-cover maps. The 50/50 distribution of the training and testing pixels was adapted from Foody et. al (2006). These pixels and the training set of pixels are exclusive.

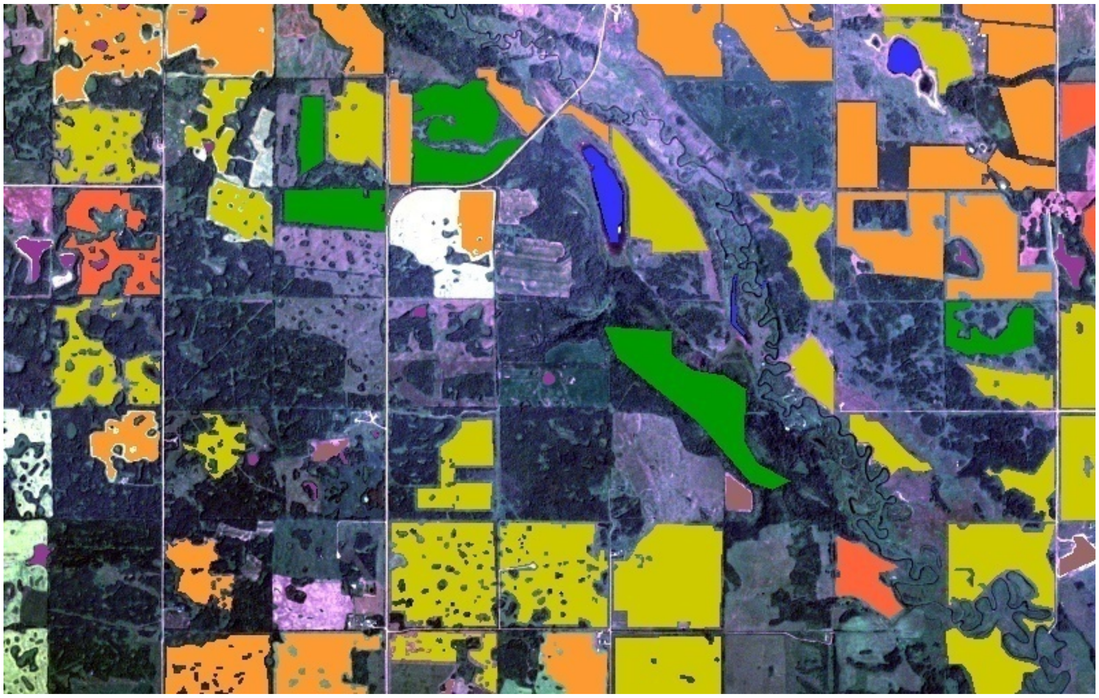


Figure 3.7: Distribution of ROIs used in the random selection of training pixels and test pixel.

Table 3.9: Quantity of image at 10 m spatial resolution verified for each land-cover class.

Land-Cover Class		Pixel #/Polygon #	Percent of Image (%)
20- Water	🟦	1392/4	0.35
30 – Exposed Land	🟫	1056/5	0.26
80 - Wetlands	🟪	1249/16	0.31
110 - Grasslands	🟨	52726/24	13.14
121 – Annual Croplands	🟧	41904/22	10.45
122 – Perennial Croplands	🟥	5328/5	1.33
200 - Forest	🟩	12030/6	3.00
Total		115682	28.83

3.8.3 Classification Accuracy

Classification accuracy was evaluated using the calculated confusion matrices for the land-cover maps. Producer and user classification accuracies were determined from the matrices. Additionally, the Kappa coefficient, which represents the overall accuracy taking into consideration random probability of correct classification also was used. Kappa analysis is a technique to determine if one confusion matrix is statistically different from another (Congalton and Green, 2008).

3.9 Water Column Content Simulation

To evaluate MSI's ability to estimate atmospheric column water vapour, 21 datasets were simulated with water vapour amounts ranging from 0 g/cm2 to 5.0 g/cm2 at a resolution of 0.25 g/cm2. These Sentinel-2 MSI at-sensor datasets were simulated consistently with the spectral, spatial and radiometric procedures outlined previously from the pre-processed AISA hyperspectral data.

3.9.1 MODTRAN Atmospheres

To create the simulated atmospheres, 21 runs of MODTRAN 4.2 were executed at a 1-nm spectral resolution to generate the atmospheric transfer spectra that will be used to calculate the reflected and path radiance values to be used in this simulation. MODTRAN settings for atmospheric properties were left at defaults with only the water column content amount being changed as to make it a controlled experiment. For each water vapour amount, two MODTRAN runs were calculated for a ground reflectance of 0.0 and 1.0. From the output TAPE7 of MODTRAN, the ground reflectance of 1.0 and the total radiance for a reflectance of 0.0 were used for the calculations. The latter represents the atmospheric path radiance effects as well as diffuse radiance contributions. These data were then spectrally convolved to the original AISA characteristics prior to being applied to the AISA reflectance data.

3.9.2 Differential Water Absorption Algorithms

To evaluate the MSI ability to estimate water column content, a band quotient was created using the MSI bands 9 (absorption region) and 8a (non-absorption region) (Table 3.3). After the at-sensor radiance for the 21 atmospheres for bands 8a and 9 were convolved (Figure 2.6) using the previous explained approach in Section 3.9. Following this, the band quotient, R_{BQ}, was calculated by dividing band 9 by band 8a on a pixel-by-pixel basis.

$$R_{BQ} = \frac{L_m}{L_r} \tag{3.5}$$

Where L_m is the at-sensor radiance measured in the water absorption band and L_r is the at-sensor radiance measured in the non-absorption band. Two regression functions were chosen to evaluate the bands: logarithmic polynomial function (Fischer and Bennartz, 1997) and an exponential function (Gao and Goetz, 1990). These functions were chosen because they have been used in the calculation using other sensors. For each atmosphere, the mean value of the band quotient was used in the regression process. The goodness-of-fit of the chosen functions was described by the coefficient of determination.

3.10 Summary

This chapter has presented the methods used in this simulation study. The airborne data that the MSI data were simulated from and its characteristics were presented. The process involving spectral and spatial simulation as well as the inclusion of noise are also presented. The process of creating a land-cover product was presented along with the proposed methods for classification accuracy assessment. Lastly, The method used for simulating atmospheres using MODTRAN was shown. The next chapter will present the results.

4. RESULTS

4.1 Introduction

Airborne hyperspectral data were initially pre-processed to remove atmospheric effects and noise so that the data could be used in the simulation of Sentinel-2 MSI data. The processed airborne data were simulated to MSI data using spectral resampling, spatial convolution methods and random noise addition. Two products were investigated: the accuracy of a land-cover product generated from the simulated MSI data and the water column content estimation utility.

The accuracy of the simulated MSI land-cover product was evaluated by comparing it with land-cover products generated from Landsat 5 TM data and SPOT 5 HRV data. The method used for classification purposes was an SVM classifier using 360 stratified random pixels for training and 360 independent pixels for testing. The same training and testing data sets were used for the Landsat and SPOT imagery. Evaluation of the accuracy was carried out by assessing confusion matrices.

MSI's atmospheric water column content estimation performance was tested by creating band quotient datasets using 21 simulated at-sensor radiance datasets with incremental increases of water column content. The average band quotient value for each scene was recorded and plotted against the simulated water column content amount. Two different regression functions were used to determine the estimation accuracy of MSI: exponential and 2nd degree logarithmic polynomial.

The results are presented in four sections. The first one presents qualitatively the results of pre-processing the hyperspectral AISA data. Next, the results and data from the

MSI simulation are presented, followed by the classification results for the simulated MSI, HRV and TM data. Lastly, the results of the atmospheric water column content estimation process are presented.

4.2 Pre-processing

The AISA hyperspectral data were initially pre-processed to remove systematic and non-systematic errors prior to simulation of the MSI data. De-striping was performed using the spectral moment matching approach, which by taking advantage of the spectral autocorrelation between hyperspectral bands uses moment statistics to correct the data (Figure 4.1).

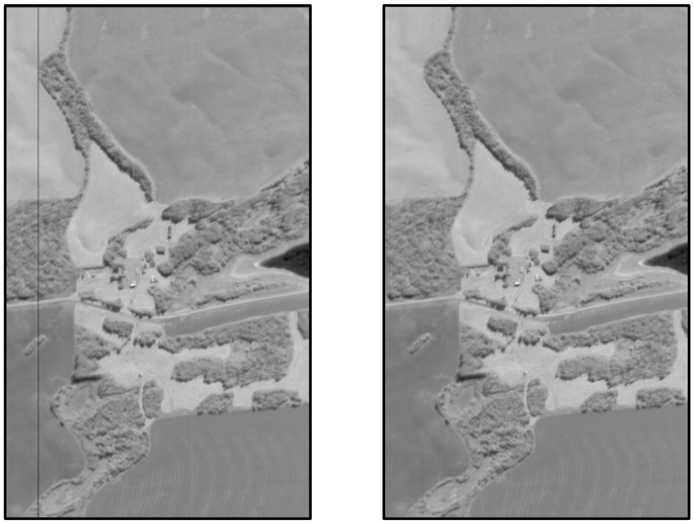


Figure 4.1: De-striping was performed using spectral moment matching. The image (band at 450 nm), on the left is prior to de-striping and the image to right is the same hyperspectral band after de-striping.

The AISA hyperspectral data were then corrected to surface reflectance using MODTRAN 4.2. AISA hyper-spectral bands that contained excessive noise were masked. The corrected surface reflectance data showed spikes, that were due to the uncertainties of the atmospheric correction, the calibration of the raw data, and the instrument noise. These spikes were removed by filtering it with a 9-nm wide window size. (Figure 4.2).

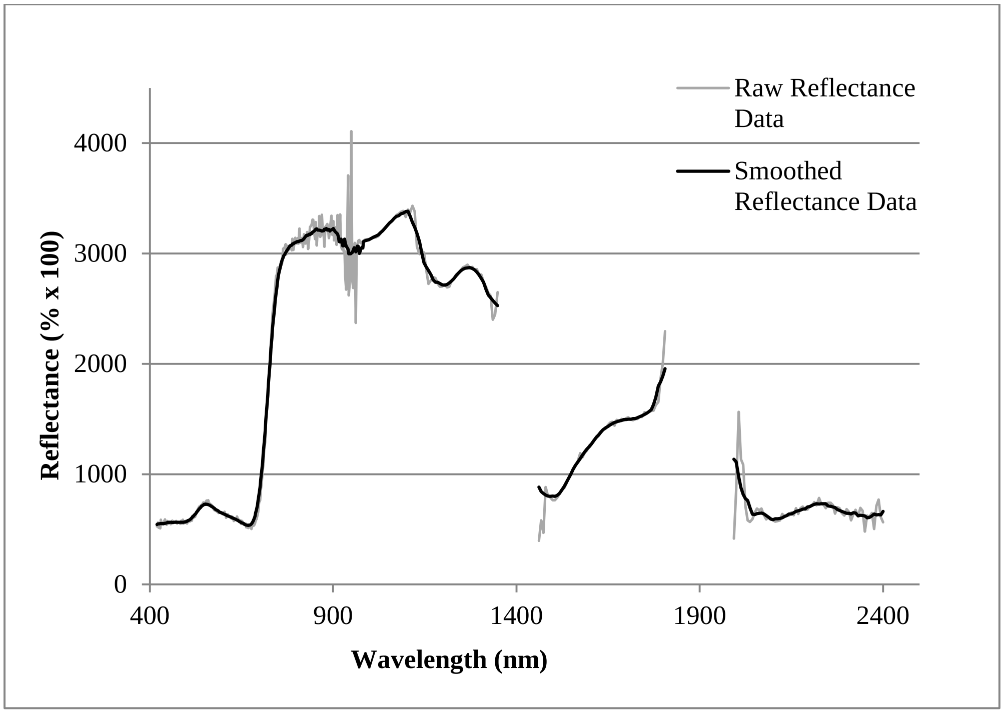


Figure 4.2: AISA 'raw' and corrected surface reflectance for a typical vegetation spectrum. Excessive noise is present at around 940 nm due to water absorption and close to 2400 nm due to the sensor.

A total of 13 AISA hyperspectral flight-lines were then mosaicked and the area was clipped to a uniform boundary. A gap in the flight lines occurred during mosaicking in which there were no data present (Figure 4.3). This area was masked in the dataset after the MSI data had been simulated and does not contain pixels used for training or validation purposes. The resultant image at 2-m spatial resolution was 3982 lines and 2522 pixels in size (Figure 4.4). Bands that were not necessary for the spectral simulation of MSI data were removed. This resulted in the AISA reflectance data contained 281 spectral bands.

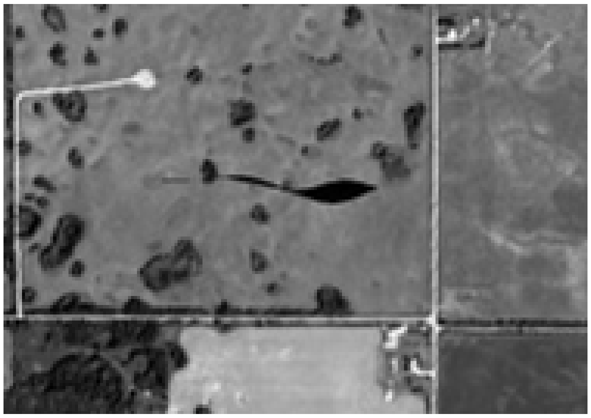


Figure 4.3: After the flight lines were mosaicked an area of no data due to insufficient flight-line overlap was discovered, shown as the black area in the middle.

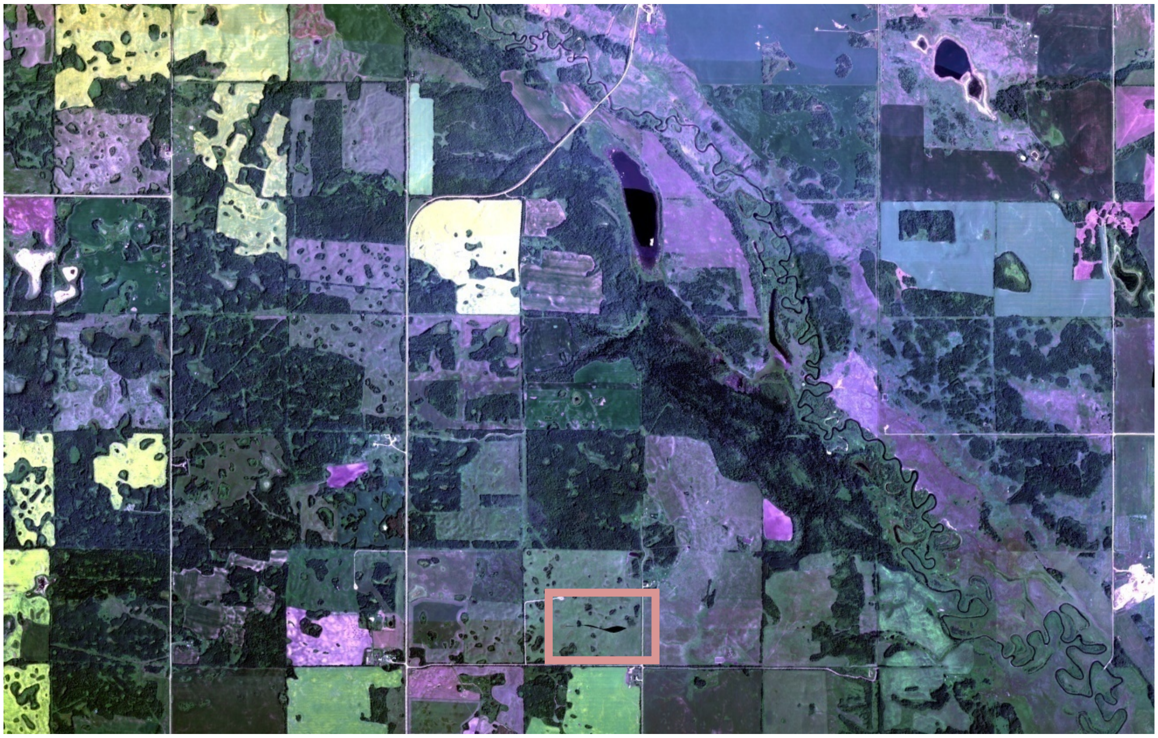


Figure 4.4: RGB colour mosaic of AISA hyperspectral airborne. Area of no data, is hilighted by the box.

4.3 MSI Simulation

Simulation of MSI data first required the AISA data to be transformed from surface reflectance to at-sensor radiance using the MODTRAN RT code. Of the 13 bands, only 12 could be successfully simulated. Band 10 (1375 nm) could not be simulated because insufficient hyperspectral data points were available after the AISA bands with excessive noise were removed.

Spatial simulation was performed on the 12 spectrally simulated bands for each spatial resolution (10 m, 20 m, and 60 m) separately. Due to the Gaussian shape of the PSF filter used, the data on the edges of the image were not properly simulated. This was compensated by removing the pixels from the edges after the data had been convolved and resampled. An example of the results for each spatial simulation is shown in Figure 4.5.

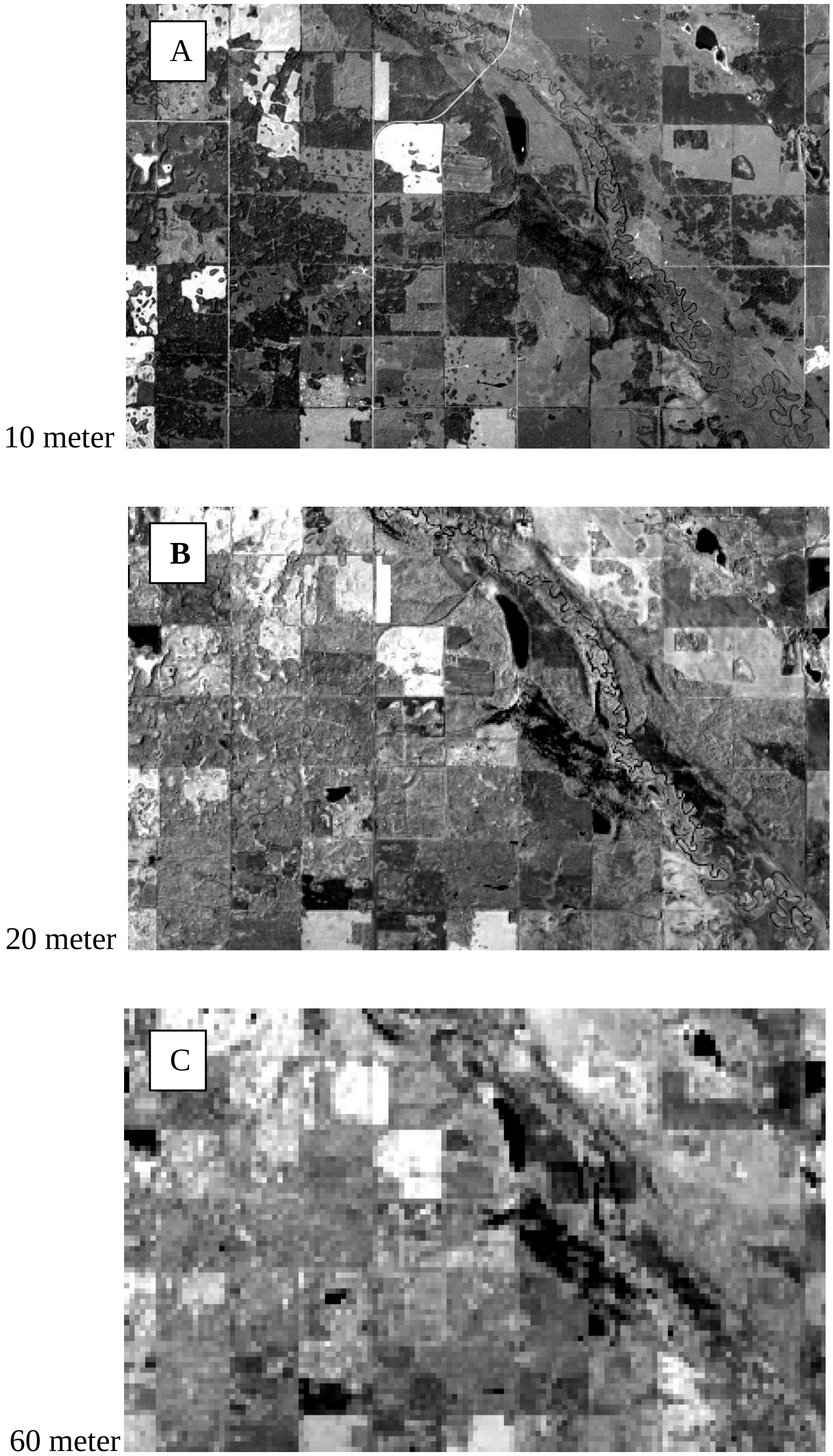


Figure 4.5: A: Band 3 with 10-m spatial resolution (796 x 504 pixels). B: Band 8 with 20-m spatial resolution (398 x 252 pixels). C: Band 9 with 60-m spatial resolution (133 x 84 pixels).

Radiance values were converted to DN prior to adding random gaussian noise for each band as shown in Figure 4.6.

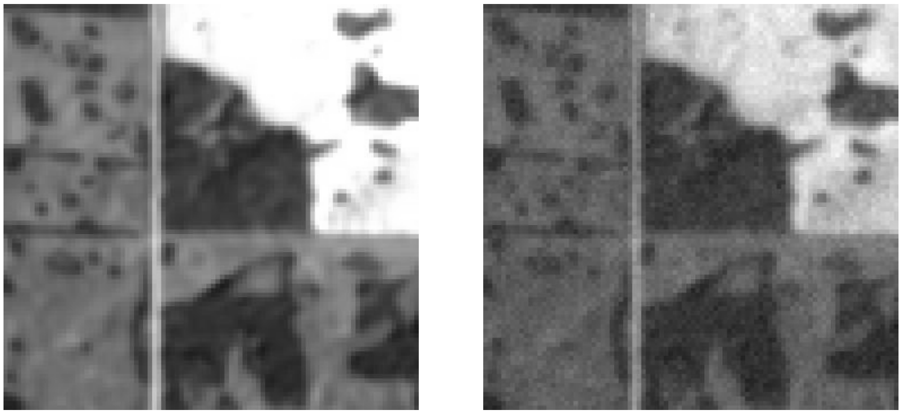


Figure 4.6: Random Gaussian noise is added to simulated raw MSI DN Band 3 (left) to create realistic data (right) with the appropriate signal-to-noise ratio.

Reflectance spectra for atmospherically corrected MSI data are presented in Figures 4.7 and 4.8 for a non-vegetation and vegetation target, repetitively, in comparison with AISA reflectance data.

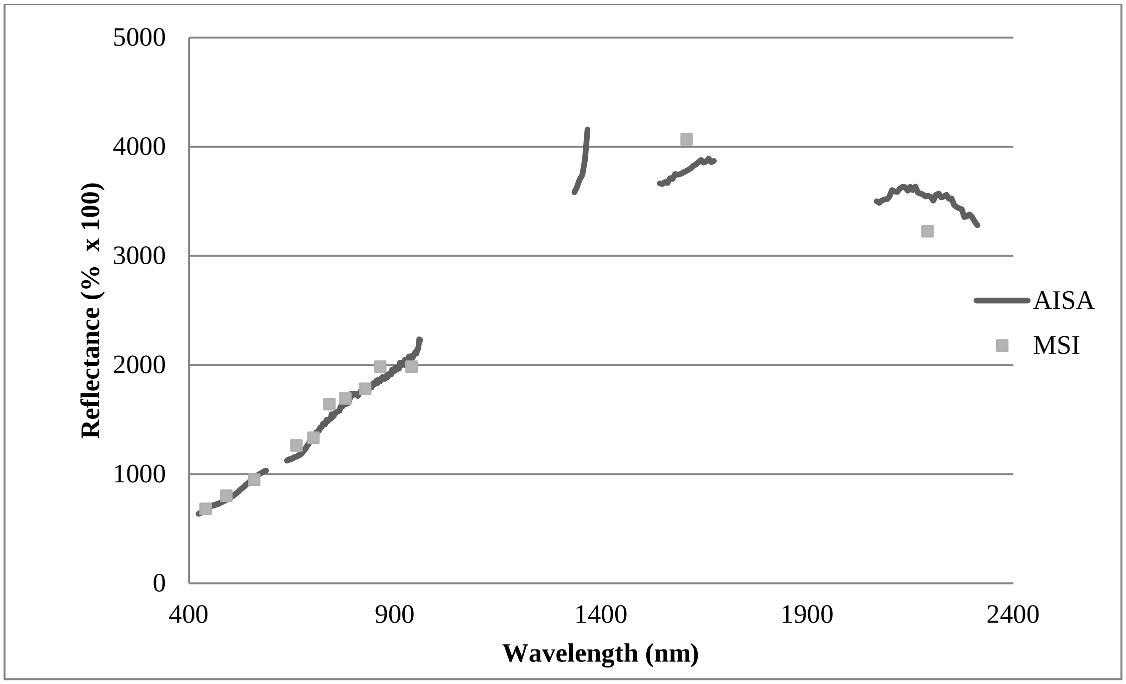


Figure 4.7: Comparison of non-vegetation reflectance spectra of simulated MSI and AISA data.

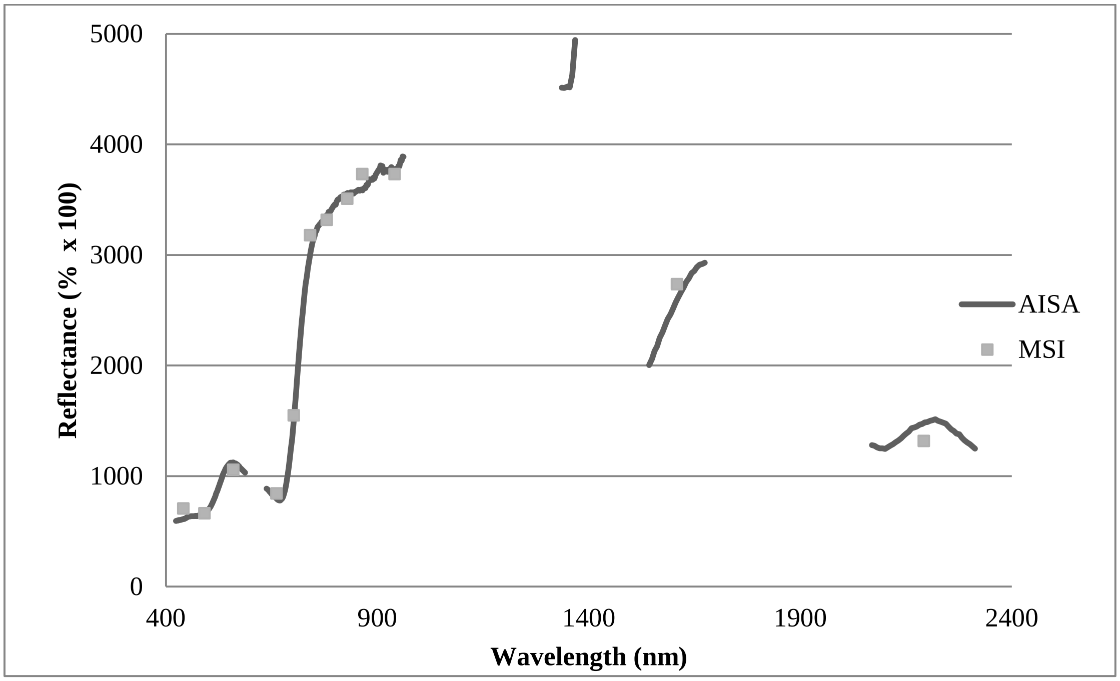


Figure 4.8: Comparison of vegetation reflectance spectra of atmospherically corrected MSI and AISA data. The AISA data at 1375 nm with no MSI data represents MSI band 10 which was unable to be simulated.

4.4 Water Column Content Estimation

A dataset was created with simulated atmospheric water column content for the purpose to evaluate Sentinel-2 MSI's ability to estimate water column content using a differential absorption algorithm (ratio of band 9 and band 8a). The average band quotient pixel value for each water column content dataset was plotted against the water column content value used in the simulation.

The regression of the band quotient data to an exponential function resulted in the following equation (5.1)

$$y = 6.82e^{-5.08x} \tag{5.1}$$

where *y* is the estimated water column content and *x* is band quotient value. This regression had a R^2 of 0.995 (Figure 4.9).

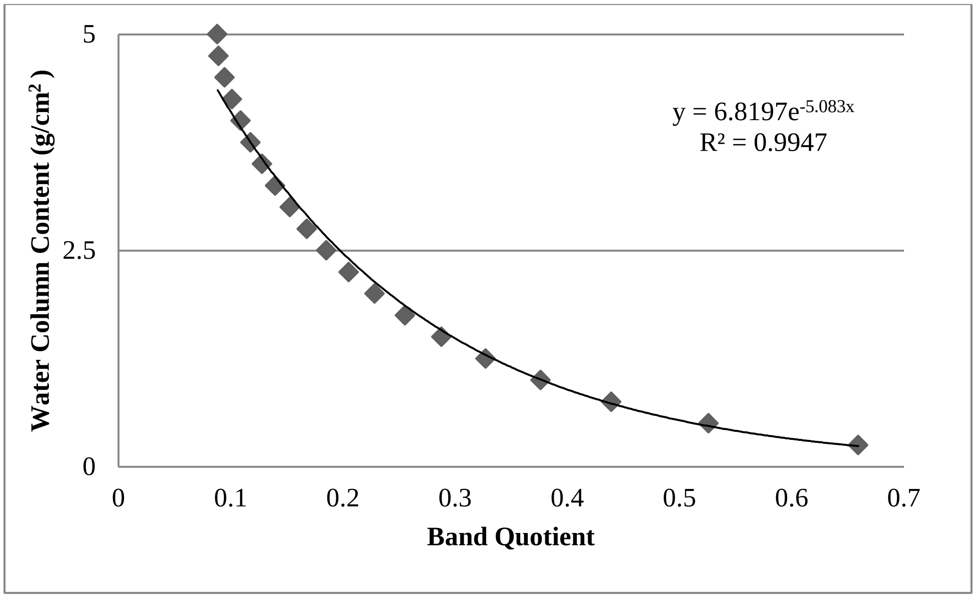


Figure 4.9: Results of the exponential function plotted against the average band quotient of each simulated atmosphere.

An alternative is the regression of the band quotient data after being converted to logarithmic base 10 in the second-degree polynomial function. This resulted in the following equation (5.2):

$$y = 3.56x^2 - 0.78x + 0.015 \tag{5.2}$$

where x is the logarithmic base 10 band quotient value. This regression had a R^2 of 0.998 (Figure 4.10).

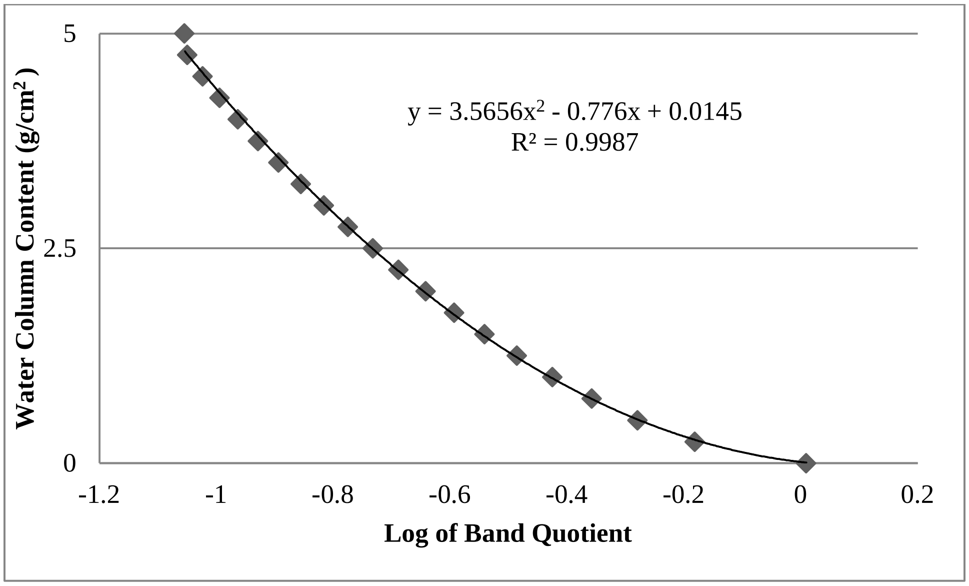


Figure 4.10: Results of 2nd degree polynomial function plotted created from the regression of the log of band quotient of each simulated atmosphere.

These functions were then applied to the simulated data. The average per-pixel percent error values were calculated between the observed and predicted values as shown in Figure 4.11. The average error per-pixel is lower than the exponential function. However the error amount of both functions increased at very high or very low levels of water column content.

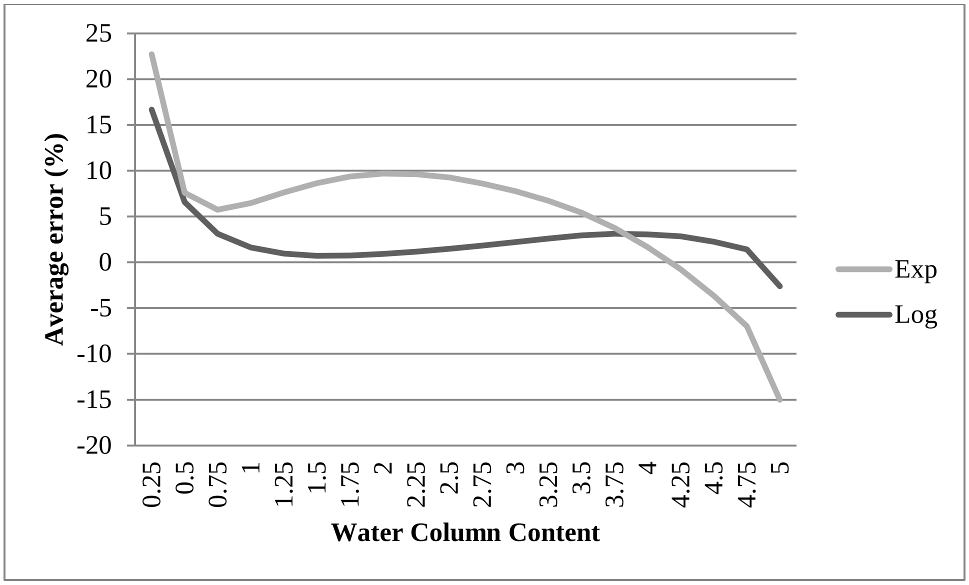


Figure 4.11: Average per-pixel percent errors amount for each simulated data-set, comparing the two regression functions.

4.5 Classification

Supervised SVM classification was performed on simulated Sentinel-2 MSI data, SPOT 5 HRV data and Landsat 5 TM data. A total of 360 randomly selected pixels from the ground reference ROIs for each land-cover class were used to train the classifier. Using random selection, 360 pixels were selected from the same ROIs that were used for testing the classification. Classified land-cover maps are presented in Figures 4.12 - 4.14. Accuracies for all land-cover classes for the three classifications are presented in the confusion matrices as shown in Tables 4.1 – 4.6 in both number of pixels and percentage of pixels.

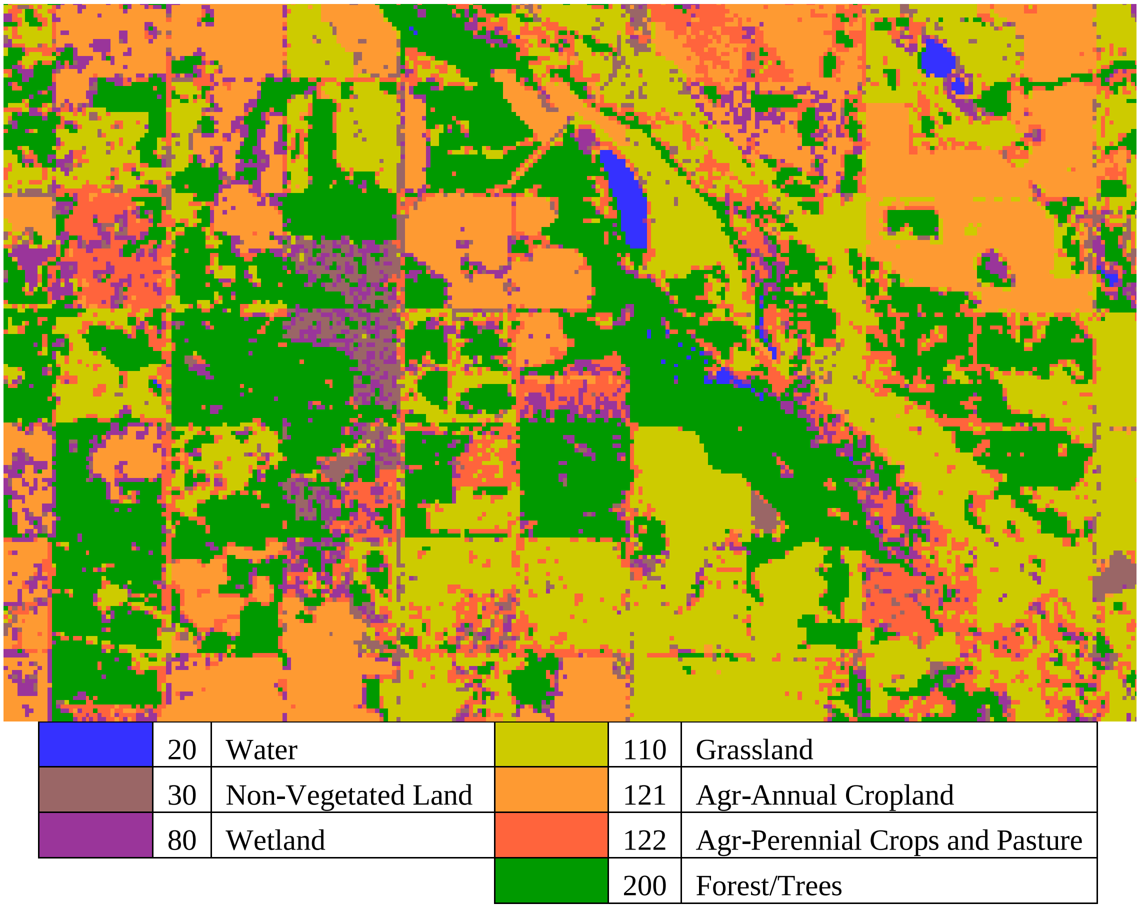


Figure 4.12: Land-cover map created from Landsat TM data using the SVM classifier.

Table 4.1: Landsat 5 TM classification accuracies.

	Reference								User Accuracy
Map Prediction	20	30	80	110	121	122	200	Total	(%)
20 - Water	349		18				6	373	93.6
30 - Non - Vegetated		325	7	12		14		358	90.8
80 - Wetlands	5	13	244	5	7	19	1	294	83.0
110 - Grasslands		4	10	303	3	44		364	83.2
121 - Annual		3	6	12	335	10		366	91.5
122 - Perennial	1	14	67	24	15	269	7	397	67.8
200 - Forest	5	1	8	4		4	346	368	94.0
Total	360	360	360	360	360	360	360	2520	
Producer Accuracy (%)	96.9	90.3	67.8	84.2	93.1	74.7	96.1		86.2

Table 4.2: Landsat 5 TM classification accuracies.

	Reference							User Accuracy
Map Prediction	20	30	80	110	121	122	200	(%)
20 - Water	96.9						1.7	93.6
30 - Non - Vegetated		90.3	1.9	3.3		3.9		90.8
80 - Wetlands	1.4	3.6	67.8	1.4	1.9	5.3	0.3	83.0
110 - Grasslands		1.1	2.8	84.2	0.8	12.2		83.2
121 - Annual		0.8	1.7	3.3	93.1	2.8		91.5
122 - Perennial	0.3	3.9	18.6	6.7	4.2	74.7	1.9	67.8
200 - Forest	1.4	0.3	2.2	1.1		1.1	96.1	94.0
Producer Accuracy (%)	96.9	90.3	67.8	84.2	93.1	74.7	96.1	86.2

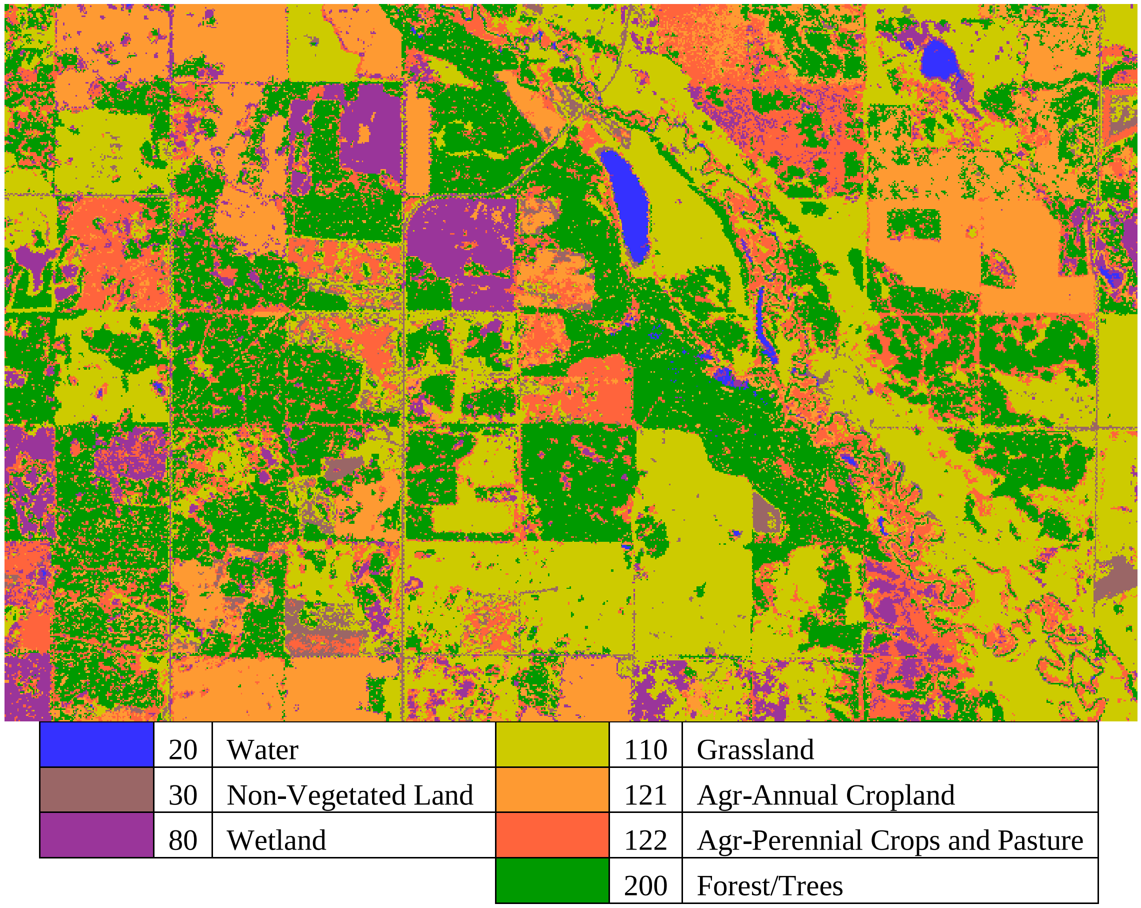


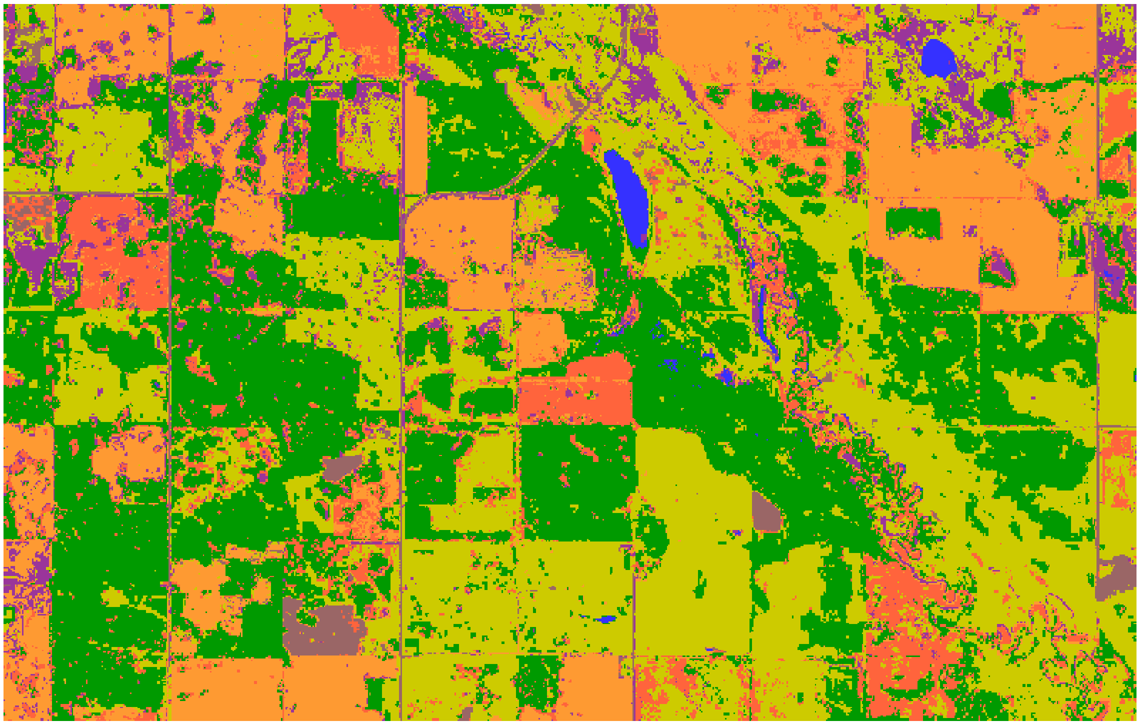
Figure 4.13: Land-cover map created from SPOT 5 HRV data using the SVM classifier.

Table 4.3: SPOT 5 HRV classification accuracies.

	Reference								User Accuracy
Map Prediction	20	30	80	110	121	122	200	Total	(%)
20 - Water	356		19					375	94.9
30 - Non - Vegetated		316		2	2	21		341	92.7
80 - Wetlands	3	3	225	33	18	58	3	343	65.6
110 - Grasslands		34	16	292	2	35		379	77.0
121 - Annual			49	16	276	13	10	364	75.8
122 - Perennial		7	49	15	24	232	14	341	68.0
200 - Forest	1		2	2	38	1	333	377	88.3
Total	360	360	360	360	360	360	360	2520	
Producer Accuracy (%)	98.9	87.8	62.5	81.1	76.7	64.4	92.5		80.6

Table 4.4: SPOT 5 HRV classification accuracies.

	Reference							User Accuracy
Map Prediction	20	30	80	110	121	122	200	(%)
20 - Water	98.9		5.3					94.9
30 - Non - Vegetated		87.8		0.6	0.6	5.8		92.7
80 - Wetlands	0.8	0.8	62.5	9.2	5.0	16.1	0.8	65.6
110 - Grasslands		9.4	4.4	81.1	0.6	9.7		77.0
121 - Annual		0.0	13.6	4.4	76.7	3.6	2.8	75.8
122 - Perennial		1.9	13.6	4.2	6.7	64.4	3.9	68.0
200 - Forest	0.3		0.6	0.6	10.6	0.3	92.5	88.3
Producer Accuracy (%)	98.9	87.8	62.5	81.1	76.7	64.4	92.5	80.6



🟦	20	Water	🟨	110	Grassland
🟫	30	Non-Vegetated Land	🟧	121	Agr-Annual Cropland
🟪	80	Wetland	🟥	122	Agr-Perennial Crops and Pasture
			🟩	200	Forest/Trees

Figure 4.14: Land-cover map created from simulated Sentinel-2 MSI 10-m data using the SVM classifier.

Table 4.5: MSI classification accuracies.

	Reference								User Accuracy
Map Prediction	20	30	80	110	121	122	200	Total	(%)
20 - Water	353		5					358	98.6
30 - Non - Vegetated		349		1		1		351	99.4
80 - Wetlands	2	3	307	9	2	20	1	344	89.2
110 - Grasslands			7	307	11	42	7	379	81.0
121 - Annual			6	12	328	12		358	91.6
122 - Perennial			34	19	19	284	1	360	78.9
200 - Forest	5		1	12		1	351	370	94.9
Total	360	360	360	360	360	360	360	2520	
Producer Accuracy (%)	98.1	96.9	85.3	85.3	91.1	78.9	97.5		90.4

Table 4.6: MSI classification accuracies.

Map Prediction	Reference 20	30	80	110	121	122	200	Total	User Accuracy (%)
20 - Water	98.1		1.4						98.6
30 - Non - Vegetated		96.9		0.3		0.3			99.4
80 - Wetlands	0.6	0.8	85.3	2.5	0.6	5.6	0.3		89.2
110 - Grasslands			1.9	85.3	3.1	11.7	1.9		81.0
121 - Annual			1.7	3.3	91.1	3.3			91.6
122 - Perennial			9.4	5.3	5.3	78.9	0.3		78.9
200 - Forest	1.4		0.3	3.3		0.3	97.5		94.9
Producer Accuracy (%)	98.1	96.9	85.3	85.3	91.1	78.9	97.5		90.4

Classification accuracies are presented using confusion matrices for the SVM classifications of the three data-sets. Overall classification accuracy was the highest with the simulated Sentinel-2 MSI data at 90.4% followed by Landsat 5 TM (86.2%) and SPOT 5 HRV (80.6%).

The results of all classifications showed that grasslands and forest classes were the largest land-cover classes. The accuracy for the MSI dataset for these classes were 97.5% and 85.3%, respectively. These values were higher than the corresponding values from Landsat TM 96.1% and 84.2% and the SPOT HRV dataset 92.5% and 81.1% (Figure 4.15).

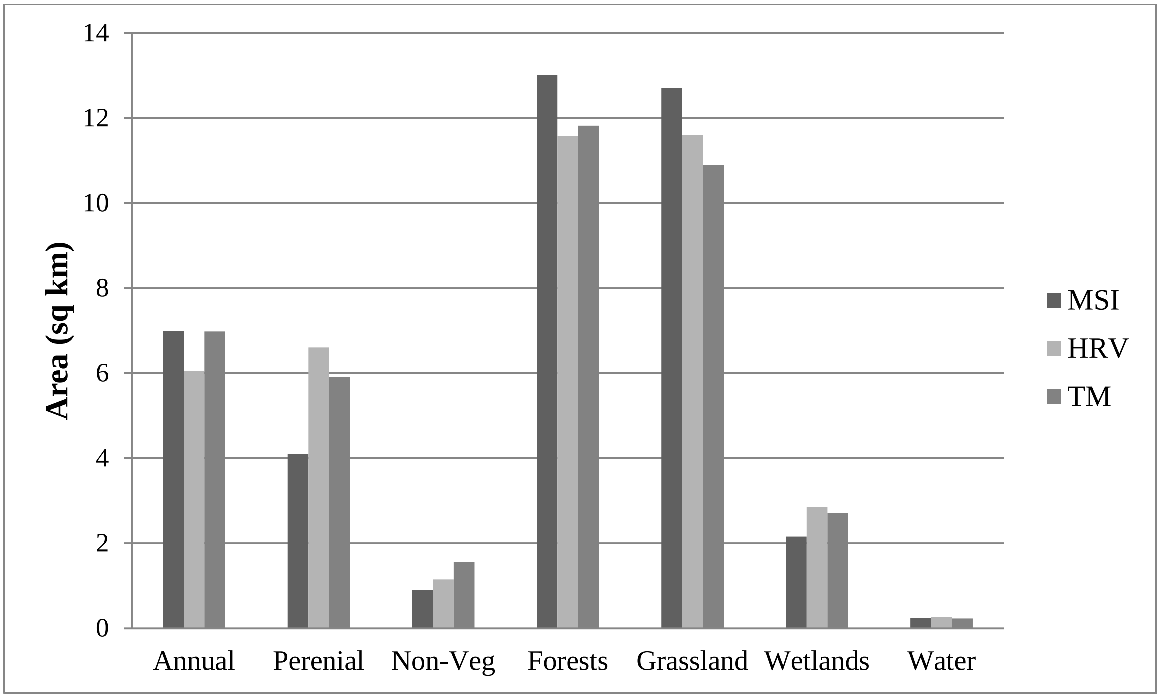


Figure 4.15: Histogram showing the differences in the estimated areas of the land-cover classes using the classification results from Landsat 5 TM, SPOT 5 HRV and Sentinel-2 MSI.

Of all the classes, the MSI accuracy was the highest with the exception of the water classes where SPOT 5 HRV accuracy was 98.9% compared to 98.1% for MSI and the perennial class, where the Landsat 5 TM accuracy was 93.1% compared to 91.1% for the MSI classification (Figure 4.16).

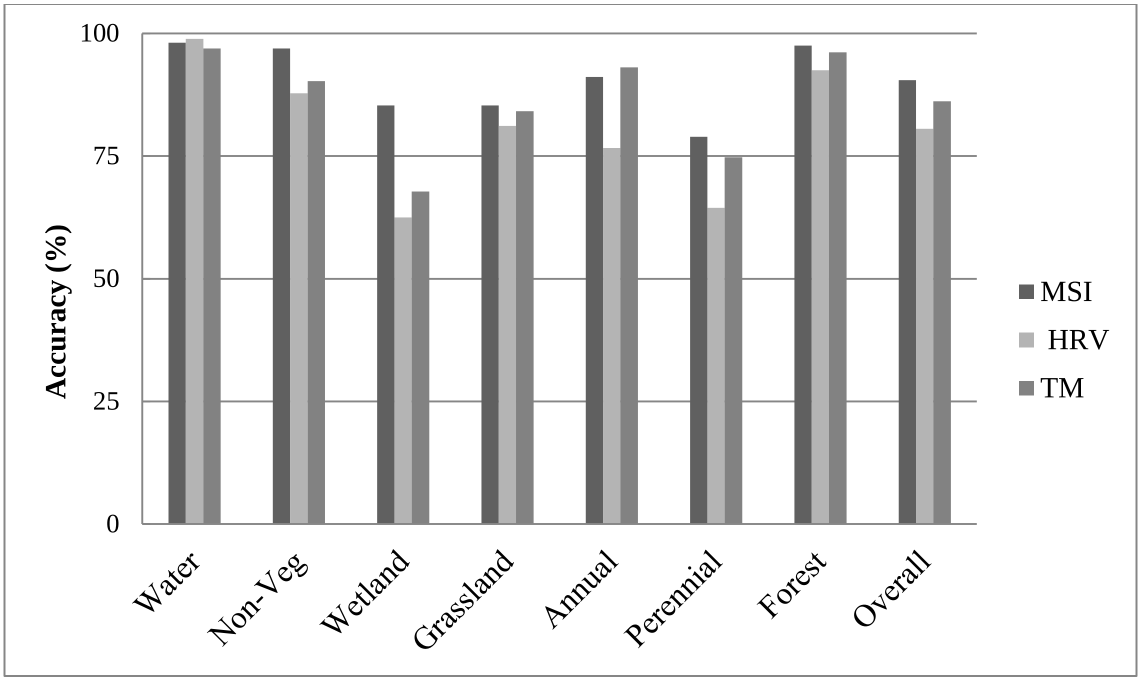


Figure 4.16: Histogram showing the producer accuracy of the classes for each dataset. Sentinel-2 MSI land cover product has an higher accuracy for all classes except water and Annual Cropland (Annual).

Wetland areas, which are regions of interest in agriculture zones, were classified with the lowest overall accuracy for all sensors. MSI accuracy for the wetland class was 85.3%, which is superior to Landsat 5 TM with a classification accuracy of 67.8% and SPOT 5 HRV with 62.5%. The second worst class in terms of accuracy is perennial crops. The overall classification accuracy for this land-cover type was 78.9%, 74.7% and 64.4% for MSI, TM and HRV, respectively. The Kappa coefficients of the three classifications are presented in Table 4.4. The results of the comparison of the classifications by use of Z-statistic is presented in Table 4.5.

Table 4.4: Kappa analysis results for the individual error matrix.

Sensor	Kappa Coefficient	Variance	Z statistic
MSI	0.89	0.0000511	124.3
HRV	0.77	0.0000921	80.5
TM	0.84	0.0000703	100.0

Table 4.5: Pairwise comparison results for the Kappa analysis.

Pairwise Comparison	Z statistics
MSI vs TM	4.54
MSI vs HRV	9.63
HRV vs TM	5.12

By evaluating the z statistics results from the Kappa analysis (Table 4.4), it was shown that the results of all classifications are significant. Furthermore, it is also presented that there are significant differences between the confusion matrix results when comparing the classifications against each other

5. DISCUSSION

5.1 Introduction

This chapter discusses the results from pre-processing, and simulation of the MSI data, classification of the simulated MSI data, and atmospheric water column content simulation. The findings as they relate to remote sensing data simulation and the MSI's capability to create a land-cover product in an agricultural area are presented. Lastly, a summary of the research, the major findings, and potential future research is presented.

5.2 Pre-processing of the Hyperspectral Data

Pre-processing the hyperspectral data in preparation prior to sensor simulation is an essential necessity. The reduction of random and systematic noise ensures that errors are not magnified throughout the simulation process. Furthermore, it is important that the hyperspectral data first be corrected to surface reflectance so that an accurate at-sensor radiance simulation can be performed. De-striping was performed using spectral moment matching. This method seemed to be satisfactory in removing systematic striping as it was evaluated qualitatively. The hyperspectral AISA data were corrected to surface reflectance using MODTRAN 4.2 in combination with a LUT approach.

5.3 Simulated Sentinel-2 Data

All simulations of sensors that are not yet operational represent rough approximations of sensor parameters that cannot be evaluated until the sensor is on orbit. For example, the values used for the gain and offset in the simulation were determined independently and the actual values of the sensor may differ when launched. These values affect the multiplicative effects of noise on the data when converting the raw data to radiance. The noise that was added to the simulated data was the maximum ideal amount as proposed in the mission requirements document (ESA, 2007). After launch, the operational sensor could realistically have much different SNRs for the specific bands (Green *et al.*, 1998). This could either improve or decrease the quality of the data. Lastly, the noise was also simulated on an entire image; the study did not consider that the independent detectors in the MSI array could have varying SNRs when the data is being acquired.

5.4 Image Classification

Image classification was performed on the simulated MSI data as well as a SPOT HRV and a Landsat TM datasets to be used for comparison. Superior overall accuracy was achieved for the simulated MSI land-cover product (90.4%) compared to over SPOT HRV (80.6%) and Landsat TM (86.2%) classification.

The least accurate land-cover type for all sensors was wetlands. Producer accuracies for the wetland land-cover were 85.3%, 67.8% and 62.5% for MSI, TM and HRV data. With respect to the MSI data, the wetland class was mostly misclassified as perennial crops, with 9.4% of ground reference wetland pixels being misclassified as perennial crops. The simulated MSI data only misclassified 5.5% of the pixels of perennial ground reference data as wetlands. With regard to the SPOT HRV data, the wetland class was also significantly misclassified as perennial crop. This land-cover type was also misclassified as wetland areas. In this case, 16.1% of the perennial crops ground reference pixels were classified as wetland areas and 13.6% of ground reference pixels of wetland land-cover were misclassified as perennial crops. In addition, 13.6% of wetland ground reference pixels were misclassified as annual crops.

Misclassification of wetland areas as perennial crops was also consistent with the TM data, where 18.6% of wetland ground reference pixels were misclassified as perennial crops and 5.2% perennial crops pixels were misclassified as wetlands. Although the omission errors for MSI and TM are the same for wetlands and perennials, MSI data had higher classification accuracies for wetland and perennial land-cover types compared to the Landsat TM and SPOT HRV classifications.

The perennial land-cover type was also severely confused with grasslands for MSI and TM data. For the MSI land-cover accuracy, 11.6% of the perennial crops were misclassified as grassland. However, this is a slight improvement over the Landsat TM classification with an 12.2%, which is greater than SPOT HRV at 9.7%. The commission error for grasslands for the MSI data was the highest for perennial crops at 5.2%. This is also the case with the TM data where the commission error is slightly higher at 6.6%.

The main reason for the improved accuracy from the MSI data was due to the increased number of spectral bands which were narrower than those of the Landsat TM and SPOT HRV sensors. The red-edge, the transition zone from the visual portion of the spectrum to the NIR portion of the spectrum, is well represented by the increased number of MSI bands, compared to those of the SPOT HRV and Landsat TM sensors. Increased sampling of the red-edge region with narrow bands has been shown to improve spectral classification for vegetation (De Backer *et al.*, 2005). Furthermore, the increased spatial resolution compared to the Landsat TM data improved the MSI's ability to map smaller areas of land-cover such as especially the wetland areas.

In addition to the aforementioned sensor characteristics, the accuracy differences could be attributed to the difference in the temporal acquisition of the datasets. The HRV data was acquired on July 15, 2010, which is two weeks before the airborne data was flown on July 27, 2010. The TM data were acquired later on August 25, 2010. Due to the different data acquisition dates, there could be variation in the plant phenology between scenes.

5.5 Water Column Content Estimation

The simulation results suggested that Sentinel-2 MSI will be able to perform water column content estimation. Due to the nature of the spectral band configurations (20-m and 60-m ground sampling distance), the effects of random noise on the estimation process were limited due to the greater ground sampling distance of band 9 (60-m) created a smoothing effect when the band was calculated. The accuracy of the simulation results must be tempered with the assumed uncertainties that have been included in the processes, mainly the spatial simulation which assumed a 3-dimensional Gaussian function and the atmospheric modelling using MODTRAN. The water column content simulation was also performed over an agriculture zone. This means that the results are only accurate over this type of general land cover. A eco-zone with a different composition of land-cover types could have different results.

6. CONCLUSIONS

Land-cover products that are created from large-swath Earth observation satellites enable land-cover change monitoring. The accuracy of land-cover products are a function of the areal extent of the land-cover types being classified and the spatial resolution of the sensor in addition to the spectral resolution of the sensor.

The main objective of this thesis was to simulate data from ESA's upcoming Sentinel-2 MSI satellite from airborne hyperspectral data. Spectral spatial and noise simulation was applied to the airborne data to create realistic MSI data. While it is difficult to evaluate the absolute accuracy of a simulated image product, the procedures outlined in this thesis demonstrate an approach to sensor simulation for the user community when detailed information about the sensor radiometric and spatial properties are not fully available.

The simulated data were evaluated by comparing the land-cover product generated from the data to similar land-cover products generated from SPOT HRV and Landsat TM data. The data of all three sensors were classified using the SVM approach. In addition, the performance of the MSI sensor to estimate water column content was evaluated by calculating the band quotient (band 9 and band 8a) of multiple simulated atmospheres with varying water column content.

The main results show that the land-cover product generated from the simulated MSI data had superior accuracy (90.4%) when compared to the other two sensors, Landsat TM (86.2%) and SPOT HRV (80.6%). Therefore the simulated Sentinel-2 MSI data will perform as well as current systems or perhaps even provide enhanced data products. All classifications exhibited confusion between perennial crops and grassland areas. Wetland areas were the least accurate class for all sensors. However, the simulated MSI data showed the highest classification accuracy for this land-cover type. This is mainly due to the addition of spectral bands as well as the improved spatial resolution compared to Landsat TM.

The MSI simulation approach performed in this study can be used in the development of future information extraction algorithms, such as for the retrieval of biophysical and biochemical parameters in vegetation canopies. These are essential steps required for the preparation of new satellite missions. Furthermore, this study was performed on a single dataset from a single year. It would be ideal to perform this simulation study on a temporal dataset to discover how MSI data can detect temporal variation, which is essential for monitoring and management of natural resources and the environment.

Addressing the hypothesis, this study's results suggest Sentinel-2 MSI will be able to perform land cover mapping as well as or better than current medium-resolution multispectral sensors, Landsat 5 TM and SPOT 5 HRV. Furthermore, Sentinel-2 MSI will be able to estimate water column content to an acceptable level of error.

This study has shown that the Sentinel-2 MSI will maintain and potentially improve accuracy levels when generating land-cover product in an agriculture zone. The development of these kind of simulation experiments is important for information product continuity, the further development of remote sensing algorithms, and exploitation of future satellite data.

REFERENCES CITED

Anderson, J. R., Hordy, E. E., Roach, J. T. & Witmer, R. E. (1976). *A Land Use and Land Cover Classification System for use with Remote Sensor Data*. pages: 28.

Arino, O., Gross, D., Ranera, F., Bourg, L., Leroy, M., Bicheron, P., Latham, J., Di Gregorio, A., Brockman, C., Witt, R., Defourny, P., Vancutsem, C., Herold, M., Sambale, J., Achard, F., Durieux, L., Plummer, S. & Weber, J. L. (2007). GlobCover: ESA service for global land cover from MERIS. In *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, 2412-2415.

Aschbacher, J. & Milagro-Pérez, M. P. (2012) The European Earth monitoring (GMES) programme: Status and perspectives. *Remote Sensing of Environment,* 120: 3-8.

Asner, P., G., Elmore, J., A., Olander, P., L., Martin, E., R., Harris & Thomas, A. (2004). *Grazing systems, ecosystem responses, and global change*. Palo Alto, CA, ETATS-UNIS: Annual Reviews.pages: 39.

Bartholomé, E. & Belward, A. S. (2005) GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing,* 26: 1959-1977.

Bartzen, B. A., Dufour, K. W., Clark, R. G. & Caswell, F. D. (2010) Trends in agricultural impact and recovery of wetlands in prairie Canada. *Ecological Applications,* 20: 525-538.

Berk, A., Anderson, G. P., Bernstein, L. S., Acharya, P. K., Dothe, H., Matthew, M. W., Adler-Golden, S. M., James H. Chetwynd, J., Richtsmeier, S. C., Pukall, B., Allred, C. L., Jeong, L. S. & Hoke, M. L. (1999). MODTRAN4 radiative transfer modeling for atmospheric correction. ed. M. L. Allen, 348-353. SPIE.

Betts, N. L., Cruickshank, M. M. & Tomlinson, R. W. (1986) An evaluation of SPOT-simulation imagery for land-use mapping and ecological investigations in upland areas of Northern Ireland. *International Journal of Remote Sensing,* 7: 779-790.

Blankenberg, A.-G. B., Haarstad, K. & Søvik, A.-K. (2008) Nitrogen retention in constructed wetland filters treating diffuse agriculture pollution. *Desalination,* 226: 114-120.

Börner, A., Wiest, L., Keller, P., Reulke, R., Richter, R., Schaepman, M. & Schläpfer, D. (2001) SENSOR: a tool for the simulation of hyperspectral remote sensing systems. *ISPRS Journal of Photogrammetry and Remote Sensing,* 55: 299-312.

Bossard, M., Feranec, J. & Otahel, J. (2000). *CORINE land cover technical guide – Addendum 2000*. European Environment Agency.

Buchan, G. M. & Hubbard, N. K. (1986) Remote sensing in land-use planning: an application in west central Scotland using SPOT-simulation data. *International Journal of Remote Sensing,* 7: 767-777.

Cihlar, J., Ly, H. & Xiao, Q. (1996) Land cover classification with AVHRR multichannel composites in northern environments. *Remote Sensing of Environment,* 58: 36-51.

Clark, R. N., King, T. V. V., Klejwa, M., Swayze, G. A. & Vergo, N. (1990) High Spectral Resolution Reflectance Spectroscopy of Minerals. *Journal of Geophysical Research,* 95: 12653-12680.

Congalton, R. G. & Green, K. (2008). *Assessing the accuracy of remotely sensed data: principles and practices*. CRC.

Connors, K. F., Gardner, T. W. & Petersen, G. W. (1987) Classification of geomorphic features and landscape stability in northwestern New Mexico using simulated spot imagery. *Remote Sensing of Environment,* 22: 187-207.

Corner, B. R., Narayanan, R. M. & Reichenbach, S. E. (2003) Noise estimation in remote sensing imagery using data masking. *International Journal of Remote Sensing,* 24: 689-702.

DB-Geoservices. (2008). GeoBase Land Cover Product User Assessment.

De Backer, S., Kempeneers, P., Debruyn, W. & Scheunders, P. (2005) A band selection technique for spectral classification. *Geoscience and Remote Sensing Letters, IEEE,* 2: 319-323.

Dey, B. & Richards, J. H. (1981) The canadian north: utility of remote sensing for environmental monitoring. *Remote Sensing of Environment,* 11: 57-72.

Di Gregorio, A. & Jansen, L. J. M. (2000). *Land Cover Classification System (LCCS) : classification concepts and user manual*. [Rome]: FAO.

Dozier, J. (1989) Spectral signature of alpine snow cover from the landsat thematic mapper. *Remote Sensing of Environment,* 28: 9-22.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P. & Martimort, P. (2012) Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment,* 120: 25-36.

ESA (2007) GMES Sentinel-2 Mission Requirements Document. 14 May 2012.

Essery, C. I. & Wilcock, D. N. (1986) SPOT-simulation campaign: a preliminary land-use classification for a 200 km2 river catchment. *International Journal of Remote Sensing,* 7: 801-814.

Farrand, W. H., Singer, R. B. & Merényi, E. (1994) Retrieval of apparent surface reflectance from AVIRIS data: A comparison of empirical line, radiative transfer, and spectral mixture methods. *Remote Sensing of Environment,* 47: 311-321.

Fiete, R. D. & Tantalo, T. (2001) Comparison of SNR image quality metrics for remote sensing systems. *Optical Engineering,* 40: 574-585.

Fischer, J. & Bennartz, R. (1997). *Retrieval of Total Water Vapour Content from MERIS Measurements: Algorithm Theoretical Basis Document ; ATBD 2.4*. Freie Univ.

Fisette, T., Chenier, R., Maloley, M., Gasser, P., Huffman, T., White, L., Ogston, R. & Elgarawany, A. (2006) Methodology for a Canadian Agricultural Land Cover Classification. *In: Proceedings of the 1st International Conference on Object-based Image Analysis, Salzburg University, Austria*.

Foody, G. M. (2002) Status of land cover classification accuracy assessment. *Remote Sensing of Environment,* 80: 185-201.

Foody, G. M. & Mathur, A. (2004) A relative evaluation of multiclass image classification by support vector machines. *Geoscience and Remote Sensing, IEEE Transactions on,* 42: 1335-1343.

Foody, G. M. & Mathur, A. (2006) The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment,* 103: 179-189.

Friedl, M. A., Davis, F. W., Michaelsen, J. & Moritz, M. A. (1995) Scaling and uncertainty in the relationship between the NDVI and land surface biophysical variables: An analysis using a scene simulation model and data from FIFE. *Remote Sensing of Environment,* 54: 233-246.

Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F. & Schaaf, C. (2002) Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment,* 83: 287-302.

Fujimotor, N., Takahashi, Y., Moriyama, T., Shimada, M., Wakabayashi, H., Nakatani, Y. & Obayashi, S. (1989). Evaluation of Spot HRV image data received in Japan. In *International Geoscience and Remote Sensing Symposium (IGARSS'89). and the 12th Canadian Symposium on Remote Sensing.*, 463-466. Vancouver, Canada.

Gao, B.-C. (1993) An operational method for estimating signal to noise ratios from data acquired with imaging spectrometers. *Remote Sensing of Environment,* 43: 23-33.

Gao, B.-C. & Goetz, A. F. H. (1990) Column Atmospheric Water Vapor and Vegetation Liquid Water Retrievals From Airborne Imaging Spectrometer Data. *Journal of Geophysical Research,* 95: 3549-3564.

Giri, C., Zhu, Z. & Reed, B. (2005) A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets. *Remote Sensing of Environment,* 94: 123-132.

Gitelson, A. A., Gritz, Y. & Merzlyak, M. N. (2003) Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology,* 160: 271-282.

Gleick, P. H. (2003). *Water use*. Palo Alto, CA, ETATS-UNIS: Annual Reviews.pages: 40.

Green, A. A., Berman, M., Switzer, P. & Craig, M. D. (1988) A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing,* 26: 65-74.

Green, R. O., Eastwood, M. L., Sarture, C. M., Chrien, T. G., Aronsson, M., Chippendale, B. J., Faust, J. A., Pavri, B. E., Chovit, C. J. & Solis, M. (1998) Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sensing of Environment,* 65: 227-248.

Guanter, L., Segl, K. & Kaufmann, H. (2009) Simulation of Optical Remote-Sensing Scenes With Application to the EnMAP Hyperspectral Mission. *Geoscience and Remote Sensing, IEEE Transactions on,* 47: 2340-2351.

Hansen, M. C., Defries, R. S., Townshend, J. R. G. & Sohlberg, R. (2000) Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing,* 21: 1331-1364.

Herold, M., Mayaux, P., Woodcock, C. E., Baccini, A. & Schmullius, C. (2008) Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1Â km datasets. *Remote Sensing of Environment,* 112: 2538-2556.

Hill, J. & Sturm, B. (1991) Radiometric correction of multitemporal Thematic Mapper data for use in agricultural land-cover classification and vegetation monitoring. *International Journal of Remote Sensing,* 12: 1471-1491.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., Vandriel, J. N. & Wickham, J. (2007) Completion of the 2001 National Land Cover Database for the Conterminous United States. *Photogrammetric Engineering and Remote Sensing,* 73: 337-341.

Horn, B. K. P. & Woodham, R. J. (1979) Destriping LANDSAT MSS images by histogram modification. *Computer Graphics and Image Processing,* 10: 69-83.

Hu, B., Li, Q. & Smith, A. (2009) Noise reduction of hyperspectral data using singular spectral analysis. *International Journal of Remote Sensing,* 30: 2277-2296.

Huang, C., Townshend, J. R. G., Liang, S., Kalluri, S. N. V. & DeFries, R. S. (2002) Impact of sensor's point spread function on land cover characterization: assessment and deconvolution. *Remote Sensing of Environment,* 80: 203-212.

Hunt, G. R. (1977) Spectral signatures of particulate minerals in the visible and near infrared. *Geophysics,* 42: 501-513.

Jensen, J. R. (2005). *Introductory digital image processing : a remote sensing perspective*. Upper Saddle River, N.J.: Prentice Hall.pages:

Jensen, J. R., Christensen, E. J. & Sharitz, R. (1984) Nontidal wetland mapping in South Carolina using airborne multispectral scanner data. *Remote Sensing of Environment,* 16: 1-12.

Johnson, B. T. (1986) Potential impact of selected agricultural chemical contaminants on a northern prairie wetland: A microcosm evaluation. *Environmental Toxicology and Chemistry,* 5: 473-485.

Justice, C. O., Markham, B. L., Townshend, J. R. G. & Kennard, R. L. (1989) Spatial degradation of satellite data. *International Journal of Remote Sensing,* 10: 1539 - 1561.

Kaufman, Y., Tanré, D., Remer, L., Vermote, E., Chu, A. & Holben, B. (1997) Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. *J. Geophys. Res,* 102: 051-17.

Kaufman, Y. J. & Gao, B. C. (1992) Remote sensing of water vapor in the near IR from EOS/MODIS. *Geoscience and Remote Sensing, IEEE Transactions on,* 30: 871-884.

Kauth, R. J. & Thomas, G. S. (1976). The Tasselled Cap -- A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by LANDSAT. Purdue University.

Kavzoglu, T. (2004) Simulating Landsat ETM+ imagery using DAIS 7915 hyperspectral scanner data. *International Journal of Remote Sensing*, 25: 5049 - 5067.

Kawata, S. & Minami, S. (1984) Adaptive smoothing of spectroscopic data by a linear mean-square estimation. *Applied Spectroscopy,* 38: 49-58.

Khurshid, K. S., Staenz, K., Sun, L., Neville, R., White, H. P., Bannari, A., Champagne, C. M. & Hitchcock, R. (2006) Preprocessing of EO-1 Hyperion data. *Canadian Journal of Remote Sensing*, 32: 84-97.

Kosarev, E. L. & Pantos, E. (1983) Optimal smoothing of 'noisy' data by fast Fourier transform. *Journal of Physics E: Scientific Instruments,* 16: 537.

Kutser, T., Herlevi, A., Kallio, K. & Arst, H. (2001) A hyperspectral model for interpretation of passive optical remote sensing data from turbid lakes. *Science of The Total Environment*, 268: 47-58.

Latifovic, R., Zhu, Z.-L., Cihlar, J., Giri, C. & Olthof, I. (2004) Land cover mapping of North and Central America—Global Land Cover 2000. *Remote Sensing of Environment,* 89: 116-127.

Lee, J. B., Woodyatt, A. S. & Berman, M. (1990) Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform. *IEEE Transactions on Geoscience and Remote Sensing,* 28: 295-304.

Lent, J. D. & Thorley, G. A. (1969) Some observations on the use of multiband spectral reconnaissance for the inventory of wildland resources. *Remote Sensing of Environment,* 1: 31-45.

Mathur, A. & Foody, G. M. (2008) Crop classification by support vector machine with intelligently selected training data for an operational application. *International Journal of Remote Sensing*, 29: 2227-2240.

Mitsch, W. J. & Gosselink, J. G. (1993). *Wetlands*. Van Nostrand Reinhold, New York.pages: 722.

Mitsch, W. J. & Gosselink, J. G. (2000) The value of wetlands: importance of scale and landscape setting. *Ecological Economics,* 35: 25-33.

Moreno, D., Pedrocchi, C., Comín, F. A., García, M. & Cabezas, A. (2007) Creating wetlands for the improvement of water quality and landscape restoration in semi-arid zones degraded by intensive agricultural use. *Ecological Engineering*, 30: 103-111.

NSWA. (2005). *State of the North Saskatchewan Watershed Report 2005: A foundation for collaborative watershed management*. North Saskatchewan Watershed Alliance.

Olthof, I., Latifovic, R. & Pouliot, D. (2009) Development of a circa 2000 land cover map of northern Canada at 30 m resolution from Landsat. *Canadian Journal of Remote Sensing,* 35: 152-165.

Peddle, D. R., Teillet, P. M. & Wulder, M. A. 2003. Radiometric image processing. In *Remote Sensing of Forest Environments: Concepts and Case Studies*, eds. M. A. Wulder & S. E. Franklin, 553. Boston, Massachusetts, USA: Kluwar Academic Publishers.

Pimm, S. L. & Raven, P. (2000) Biodiversity: Extinction by numbers. *Nature,* 403: 843-845.

Richter, R. & Schläpfer, D. (2008) Considerations on Water Vapor and Surface Reflectance Retrievals for a Spaceborne Imaging Spectrometer. *Geoscience and Remote Sensing, IEEE Transactions on,* 46: 1958-1966.

Rouse, J. W., Haas, R. H., Schell, J. A. & Deering, D. W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. In *Third ERTS Symposium*, 309-317. Washington DC.

Savitzky, A. & Golay, M. J. E. (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry,* 36: 1627-1639.

Schaepman-Strub, G., Schaepman, M. E., Painter, T. H., Dangel, S. & Martonchik, J. V. (2006) Reflectance quantities in optical remote sensing—definitions and case studies. *Remote Sensing of Environment,* 103: 27-42.

Schläpfer, D., Borel, C. C., Keller, J. & Itten, K. I. (1998) Atmospheric Precorrected Differential Absorption Technique to Retrieve Columnar Water Vapor. *Remote Sensing of Environment,* 65: 353-366.

Schmidt, K. S. & Skidmore, A. K. (2004) Smoothing vegetation spectra with wavelets. *International Journal of Remote Sensing,* 25: 1167-1184.

Schott, J. (2007). *Remote sensing : the image chain approach*. New York: Oxford University Press.pages: 666.

See, L. M. & Fritz, S. (2006) A method to compare and improve land cover datasets: application to the GLC-2000 and MODIS land cover products. *Geoscience and Remote Sensing, IEEE Transactions on,* 44: 1740-1746.

Simpson, J. J., Gobat, J. I. & Frouin, R. (1995) Improved destriping of GOES images using finite impulse response filters. *Remote Sensing of Environment,* 52: 15-35.

Smits, P. C., Dellepiane, S. G. & Schowengerdt, R. A. (1999) Quality assessment of image classification algorithms for land-cover mapping: A review and a proposal for a cost-based approach. *International Journal of Remote Sensing,* 20: 1461-1486.

Sobrino, J. A., El Kharraz, J. & Li, Z. L. (2003) Surface temperature and water vapour retrieval from MODIS data. *International Journal of Remote Sensing,* 24: 5161-5182.

Staenz, K., Secker, J., Gao, B. C., Davis, C. & Nadeau, C. (2002) Radiative transfer codes applied to hyperspectral data for the retrieval of surface reflectance. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57: 194-203.

Staenz, K., Szeredi, T. & Schwarz, J. (1998) ISDAS - A system for processing/analyzing hyperspectral data. *Canadian Journal of Remote Sensing,* 24: 99-113.

Staenz, K. & Williams, D. J. (1997) Retrieval of surface reflectance from hyperspectral data using a look-up table approach. *Canadian Journal of Remote Sensing*, 23: 354-368.

Sun, L., Neville, R., Staenz, K. & White, H. P. (2008) Automatic destriping of Hyperion imagery based on spectral moment matching. *Canadian Journal of Remote Sensing,* 34: S68-S81.

Tanré, D., Kaufman, Y., Herman, M. & Mattoo, S. (1997) Remote sensing of aerosol properties over oceans using the MODIS/EOS spectral radiances. *J. Geophys. Res*, 102: 16.

Thomas, J. R. & Gausman, H. W. (1976) Leaf reflectance vs. leaf chlorophyll and carotenoid concentrations for eight crops. *Agronomy Journal,* 69: 799-802.

Tilman, D., Fargione, J., Wolff, B., D'Antonio, C., Dobson, A., Howarth, R., Schindler, D., Schlesinger, W. H., Simberloff, D. & Swackhamer, D. (2001) Forecasting Agriculturally Driven Global Environmental Change. *Science*, 292: 281-284.

Tsai, F. & Philpot, W. (1998) Derivative Analysis of Hyperspectral Data. *Remote Sensing of Environment,* 66: 41-51.

Tucker, C. J. (1978). An Evaluation of the First Four Landsat-D Thematic Mapper Reflective Sensors for Monitoring Vegetation: A Comparison with Other Satellite Sensor Systems, NASA technical memorandum 79617, Goddard Space Flight Center, Greenbelt, Maryland.

Verhoef, W. & Bach, H. (2003) Simulation of hyperspectral and directional radiance images using coupled biophysical and atmospheric radiative transfer models. *Remote Sensing of Environment,* 87: 23-41.

Vogelmann, J. E., Sohl, T. L., Campbell, P. V. & Shaw, D. M. (1998) Regional Land Cover Characterization Using Landsat Thematic Mapper Data and Ancillary Data Sources. *Environmental Monitoring and Assessment,* 51: 415-428.

Wulder, M. A., Dechka, J. A., Gillis, M. A., Luther, J. E., Hall, R. J., Beaudoin, A. & Franklin, S. E. (2003) Operational mapping of the land cover of the forested area of Canada with Landsat data: EOSD land cover program. *The Forestry Chronicle*, 79: 1075-1083.

Wulder, M. A., White, J. C., Cranny, M., Hall, R. J., Luther, J. E., Beaudoin, A., Goodenough, D. G. & Dechka, J. A. (2008a) Monitoring Canada’s forests. Part 1:

Completion of the EOSD land cover project. *Canadian Journal of Remote Sensing*, 34: 549-562.

Wulder, M. A., White, J. C., Goward, S. N., Masek, J. G., Irons, J. R., Herold, M., Cohen, W. B., Loveland, T. R. & Woodcock, C. E. (2008b) Landsat continuity: Issues and opportunities for land cover monitoring. *Remote Sensing of Environment*, 112: 955-969.

APPENDIX A

Example MATLAB code for generating random zero-mean Gaussian noise applied to Sentinel-2 MSI band 12.

```
noise = 2.8424.*randn([252,398]);
band12noise = data + noise;
name = ['band-12' '.bsq'];
multibandwrite(band12noise,name,'bsq')
name = ['band-12-noise' '.bsq'];
multibandwrite(nose,name,'bsq')
```