Remote sensing as a monitoring solution for water hyacinth (*Pontederia crassipes*) in the context of the biological control programme at Hartbeespoort Dam

A thesis submitted in fulfilment of the requirements for the degree of

Master of Science

At

Rhodes University, Department of Geography

By

David Kinsler

Supervised by

Ms. Gillian McGregor and Prof. Julie Coetzee

January 2023

Abstract

Water hyacinth (*Pontederia crassipes* (C.Mart.) Solms (Pontederiaceae)) is a significant aquatic weed both globally and in South Africa. Despite notable success with biological control of other invasive macrophytes, the plant remains as a problematic weed in many aquatic systems in South Africa, particularly due to the eutrophic status of many of its water systems, as well as the plant's tolerance to cooler climatic conditions than most of its existing biological control agents. Hartbeespoort Dam, located about 30 kilometres west of Pretoria, South Africa, has been infamously infested with water hyacinth for decades, which impacts the important socioeconomic utility of the dam and functioning of natural ecological processes in the system.

The dam has a long history of efforts to control water hyacinth, which include widespread herbicidal spray, mechanical removal and classical biological control programmes since the early 1990s - mostly with limited or short-lived success. However, after the introduction of a new, cold-tolerant biological control agent, *Megamelus scutellaris* Berg (Hemiptera: Delphacidae) in 2018 with an inundative release strategy, the water hyacinth dropped significantly from a maximum cover of about 45 percent (819 hectares) down to less than two percent (40 hectares) over a three-month period (November 2019 – January 2020). This was significant, as it marked the first successful biological control of water hyacinth in a eutrophic, temperate system in South Africa.

However, due to the scale of Hartbeespoort Dam (1820 hectares) and the high spatiotemporal variation of the floating mats across time and space, quantifying and monitoring these rapid changes has proved difficult. In response to this problem, this thesis proposed a remote sensing solution to address the need for accurate, timely and readily accessible monitoring data of the water hyacinth population on the dam. Leveraging the temporally frequent (< 5 days revisit time) Sentinel-2 multispectral satellite data, as well as the powerful cloud-computing resources of Google Earth Engine, this thesis developed and deployed a relatively simple and robust index-based decision tree classification method to demonstrate the value of these technologies as an effective monitoring and analysis tool for monitoring large macrophyte infestations.

To this end, several challenges had to be overcome in order to produce easily accessible data that was accurate and reliable. For example, due to the size of the Sentinel-2 Level-1C image dataset from August 2015 to March 2021 (n = 654), an automated process of filtering out clouded images was required. Additionally, the co-presence of algal and cyanobacterial blooms necessitated the development of a novel index, coined the Algae Resistant Macrophyte Index (ARMI), to deal with the challenges of accurate macrophyte detection. The high spatiotemporal variability of the floating mats meant that a typical, location-based confusion matrix as a means of assessing the accuracy of the decision tree classifier required a different approach which compared the total classified areas with higher resolution images.

This thesis aims to demonstrate the utility of remote sensing tools to provide effective monitoring information to managers, researchers and other stakeholders. There is scope to expand to more areas in South Africa and beyond and may prove an invaluable tool to augment and support on-going and future macrophyte monitoring programmes.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisors, Gillian McGregor and Julie Coetzee for their personal input, continued enthusiasm and support for the research conducted in this thesis, as well as Martin Hill who provided much clarity and direction through this process.

Thank you to the Centre for Biological Control, Rhodes University, for creating the opportunity and providing the funding for me to do this research.

I would like to thank Mark Thompson, of GeoTerra Image (Pty) Ltd, who kindly provided me with invaluable input in the beginning stages of this project, as well as introducing me to Google Earth Engine.

I'd like to express gratitude to Geethen Singh, who endured hours upon hours of long, technical conversations about everything remote sensing and Google Earth Engine, and proved an essential sounding board for much of the work conducted herein.

Finally, and most importantly, I'd like to acknowledge my dear partner, Anna Talbot, whose heroic patience and support made this all possible. Without you, I would not have done this.

Chapter 1: Introduction to the research

1.1 Introduction

We are currently witnessing a global ecological crisis at a scale never before seen in human history, as biodiversity loss occurs at unprecedented rates (Lim, 2019). Widely termed the "sixth mass extinction" (e.g. Barnosky *et al.*, 2011), ample evidence from multiple disciplines point to anthropogenic actions as the primary cause (IPBES, 2019). The spread of non-native invasive species (henceforth "invasive species") is considered one of the direct causes of global biodiversity loss and detailed records from 21 countries indicate that the number of invasive species has increased by an average of 70 % over the last 50 years (IPBES, 2019). Indicators of this biodiversity loss include changes in the number of wild species, the distinctness of ecological communities, and ecological alteration of terrestrial and, notably, aquatic ecosystems.

South Africa is an ecologically diverse country that boasts nine terrestrial biomes covering a wide range of ecological niches (Rutherford *et al.*, 2006) and is identified as one of 17 megadiverse nations globally (Skowno *et al.*, 2019). Within the nine biomes are 458 ecosystem types, of which 80 % are endemic (Dayaram *et al.*, 2019). This biodiversity is due, in part, to the country's wide-ranging climate and topographic diversity: from its semi-arid western region to the high rainfall coastal belt in the east. The rainfall gradient and the contrasting seasonality results in a varied climate that is similar to many other regions throughout the globe (Richardson & Thuiller, 2007). This endowment has created high levels of biodiversity with many endemic species, but also has left South Africa vulnerable to a wide range of potential biological invaders.

South Africa has a long history of biological invasions. There are more than 1400 identified non-native species in South Africa, across all biomes, that can survive and reproduce outside human cultivation or captivity (van Wilgen *et al.*, 2020). With the absence of natural predators and widespread disturbance, many of these species have become invasive (Coetzee *et al.*, 2011). Freshwater ecosystems are among the most threatened (Figure 1.1), with 65 of the 191 known non-native aquatic species listed as invasive (Skowno *et al.*, 2019). If left unchecked, invasive species have the potential to transform entire ecosystems into long-term degraded stable states (Scheffer *et al.*, 2003). The effects have wide-ranging impacts and cost the country at least USD $1 billion per year (van Wilgen *et al.*, 2020).

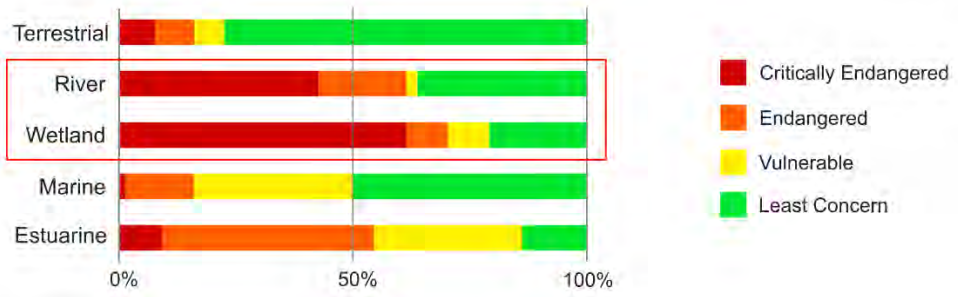


Figure 1.1: Percentage of threat status of each ecosystem type in South Africa, with wetlands and rivers ranking the as the two most threatened systems. Graph data adapted from Skowno *et al.*, (2019).

As a drought-prone country, the degradation of aquatic systems has significant adverse ecological, economic and social impacts. Since the early 20th century, South African aquatic ecosystems have been prone to macrophyte (a term used to refer to macroscopic freshwater plants) invasions (Hill & Coetzee, 2017; Hill *et al*., 2020). South Africa's landscape is typically erosive, resulting in the formation of very few natural lakes and other large water bodies (Coetzee & Hill, 2012). As a result, no native floating macrophyte species have evolved in the country (Coetzee & Hill, 2012). With the construction of large impoundments, the new freshwater habitats thus become inherently vulnerable to invasion by non-native macrophyte species (Coetzee & Hill, 2012). Compounding this issue, low water quality standards for wastewater treatment works, agricultural runoff and untreated waste from densely populated urban areas produce high pollutant levels that flow into freshwater ecosystems, producing eutrophic conditions that cause excessive growth of floating macrophytes once they enter a system (Coetzee & Hill, 2012).

Whilst most of the worst invasive macrophytes in South Africa have been effectively controlled with biological control, the biological control of *Pontederia crassipes* (C.Mart.) Solms (Pontederiaceae) (water hyacinth) has remained variable (Hill & Coetzee, 2017). Due to the widespread eutrophic conditions and the plant's tolerance to a range of climatic conditions, water hyacinth remains particularly prevalent in many aquatic ecosystems in South Africa. It can adversely impact both ecological and socio-economic systems significantly and is expensive to control, particularly by traditional mechanical and herbicidal means (Hill *et al*., 2020). Consequently, it remains South Africa's most important invasive aquatic weed and thus remains as an item of key research interest.

1.1.1 Water Hyacinth on Hartbeespoort Dam

Hartbeespoort Dam, the site on which this research is based, is located 30 kilometres west of Pretoria, South Africa. It is one of the most eutrophic dams in the country (Matthews, 2014) and has been infamously infested with water hyacinth for decades, with the first record from 1959 (Ashton *et al*., 1979). The hypertrophic conditions are also responsible for many cyanobacterial blooms over the decades (Matthews, 2014). The catchment of the largest tributary, the Crocodile River, includes the highly industrial and urbanised areas of northern Johannesburg and is responsible for most of the nutrient load, primarily nitrogen (N) and phosphorus (P), into the dam (Twinch, 1986). The inundation of water hyacinth is a product of these hypertrophic conditions and the floating plant thrives in the nutrient-abundant system.

Many attempts to control the water hyacinth on the dam, from both public and private sectors, have occurred over the years with limited or short-lived success. The most notable recent effort was by the government funded 'Harties Metsi a Me' programme, beginning in 2006, which focused on reducing water hyacinth by means of mechanical removal (Mitchell & Crafford, 2016). The programme ended in 2016, and the dam remains periodically infested by water hyacinth to the present day. Costly aerial herbicidal spray efforts had achieved some success in reducing the plant cover, but the effects were temporary as remaining plants rapidly re-propagate alongside the large seedling fecundity. In 2017, a moratorium on herbicide spray on the dam was enacted and the water hyacinth continued to flourish, largely unabated.

Biological control efforts had begun on the dam from the early 1990s, with several agent species being introduced, including the two *Neochetina* spp. Weevils, onto the dam during that period (Coetzee *et al.,* 2022). Although they successfully established and are still present on the dam, successful biological

control was limited due to the hypertrophic water and temperate climate; conditions which favoured the plants over their agents.

1.1.2 The introduction of *Megamelus scutellaris* on Hartbeespoort Dam

First introduced in South Africa in 2013, *Megamelus scutellaris* Berg (Hemiptera: Delphacidae) is the most recent biological control introduction for water hyacinth in the country (Hill & Coetzee, 2017). Although of tropical origin, this species was observed to still thrive in the cooler regions of its native and introduced ranges (Hill & Olckers, 2001).

Given its favourable climatic tolerances, *M. scutellaris* was introduced to Hartbeespoort in 2018 by the Centre for Biological Control (CBC), Rhodes University, in collaboration with local stakeholders. The release strategy was to inundate the water hyacinth with insects, which took place over several months in late 2018 and early 2020, with more than 197 000 individuals released in total (Coetzee *et al.*, 2022). The insects established on the dam and the population rapidly increased in the following months, successfully over-wintering in that year. Subsequently, in the summer period of 2019/2020, the water hyacinth populations on the reservoir declined precipitously for the first time without significant herbicidal application. This event marked the first case of successful biological control of water hyacinth in a temperate, eutrophic system in South Africa (Coetzee *et al.,* 2022). The success of this programme is noteworthy and may have far-reaching consequences as water hyacinth remains regarded as the world's worst aquatic weed (Figure 1.2).

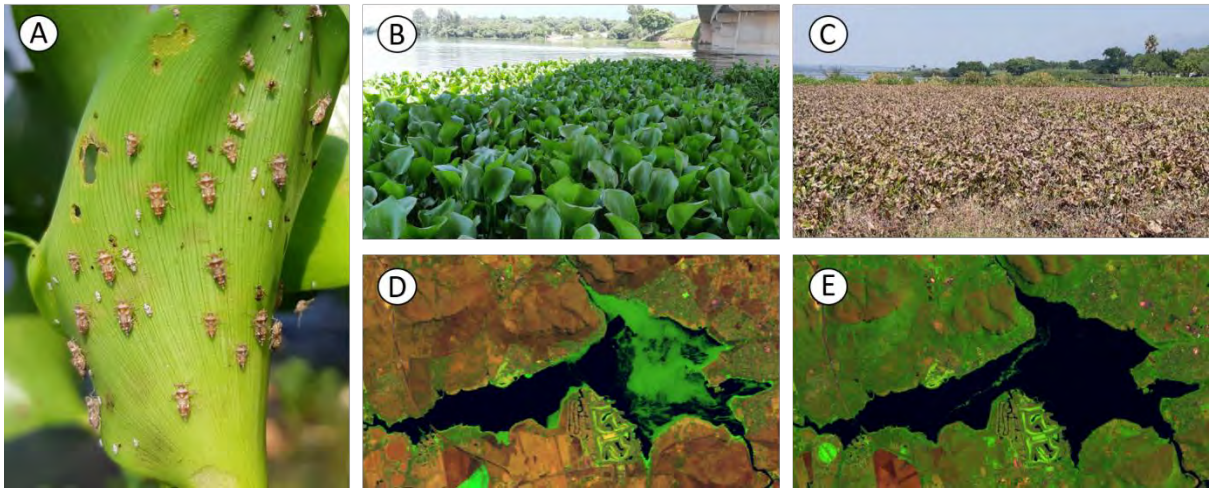


Figure 1.2: (A) *Megamelus scutellaris* at various life stage cycles, Hartbeespoort Dam. Examples of (B) healthy and (C) severely impacted water hyacinth mats on Hartbeespoort Dam. (D) Abundant water hyacinth cover on Hartbeespoort Dam in October 2019 followed by (E) a significant reduction in water hyacinth cover in February 2020.

1.2 Motivation for the research

Effective monitoring is a crucial part of a successful biological control programme, providing an important feedback loop between management decisions and system responses. While significant funding is put towards invasive plant control in South Africa, there is a lack of monitoring and reporting on the outcome of these projects (van Wilgen *et al.*, 2022). Continuous monitoring of invasive aquatic plants is essential for proper management and effective control, however this remains variable in South Africa (Thamaga & Dube, 2018a). Therefore, there is an opportunity and a need for a national effort towards the development and application of monitoring solutions, on-going feedback, and quantification of management strategies.

On Hartbeespoort Dam, the rapid and large-scale changes in water hyacinth population size necessitate timely and accessible monitoring information of the on-going status of macrophyte invasion levels on the dam. Due to the physical scale and high spatiotemporal variability of the floating mats on the dam, the feasibility of traditional infield monitoring methods is limited, particularly at a dam-wide scale. Remote sensing (RS) technologies provide an opportunity to fill this monitoring gap. Recent advancements in the RS field, particularly a) the increased availability of open source, frequent, and high-resolution satellite imagery, and b) the availability of free, cloud-based remote sensing software for rapid analysis of larger data volumes have significantly improved the monitoring potential of invasive plant populations at larger scales. While the use of remote sensing to observe and measure changes in invasive populations has been in practise for decades, the emergence of these new technologies has considerably improved the scope of RS application as a near-real-time and accurate monitoring tool to support the limited monitoring opportunities of South African invasive macrophytes. Acquiring and presenting detailed, accurate monitoring data of CBC's biological control programme on Hartbeespoort Dam would improve the ability of stakeholders to make informed management decisions, compare the results of different management strategies and increase research potential to guide future biological control efforts of water hyacinth both locally and beyond.

1.3 Aims and Objectives

The aim of this thesis was to develop an accurate and reliable remote sensing method to map and monitor water hyacinth on Hartbeespoort Dam, in the context of the successful biological control programme implemented by CBC, to provide both historical and near-real-time monitoring information to stakeholders.

Objective 1**:** In the context of existing literature and exploratory investigation, identify challenges and opportunities associated with remote sensing of water hyacinth on Hartbeespoort Dam.

Objective 2: Develop a remote sensing solution that addresses the challenges of mapping a dynamic natural feature, water hyacinth, to produce time-series data that can be readily disseminated to stakeholders.

Objective 3: Develop and conduct a suitable accuracy assessment for validating the results of objective two.

1.4 Thesis Structure

This dissertation consists of five chapters. The first chapter contextualises and provides the motivation for this research by introducing the threat of non-native invasive species to biodiversity and socioeconomic well-being, provides background context to the water hyacinth problem on Hartbeespoort Dam and highlights the recent success of the *Megamelus scutellaris* biological control programme on the dam. The case for monitoring with remote sensing methods is then made, highlighting the relatively under-utilised potential of remote sensing to fill the monitoring deficit in South Africa. The aims and objectives of the thesis were then outlined from this context.

The second chapter provides background information on the status of water hyacinth in South Africa, outline drivers behind its continued invasiveness and contextualise the invasion on Hartbeespoort Dam. Then it addresses the existing literature on water hyacinth remote sensing, the methods used in previous research, and identifies the opportunity to leverage this with improved remote sensing software and data sources. It highlights the paradigm shift in remote sensing, where increased temporal and spatial resolution satellites, combined with cloud storage and computing platforms, have shifted RS from being simply a change detection tool into a near-real-time monitoring solution.

The third chapter is the main methods chapter of this dissertation. It investigates and identifies challenges with monitoring water hyacinth on Hartbeespoort Dam and provides solutions to overcome these where possible. It then develops an empirically determined, index-based threshold classification tree that was designed to automatically and efficiently produce accurate, dense time-series data on the area of water hyacinth on Hartbeespoort Dam.

The fourth chapter addresses the challenges associated with a regular RS accuracy assessment, given the high spatio-temporal variability of the floating mats, and develops an accuracy assessment alternative for the results of Chapter 3. Instead of using a confusion matrix, it compares the total classified area values with data acquired from higher resolution satellite imagery.

The fifth chapter revisits the original aim of this thesis and discusses how remote sensing monitoring approach can fill a need for better monitoring of invasive plant species. It addresses the limitations of the work of this dissertation and suggests areas for future research and development.

Chapter 2: Literature and background

2.1 Introduction

The geography of South Africa makes it particularly vulnerable to macrophyte invasions. The country has an unusually high median altitude of over 1000 metres above sea level which has produced a predominantly erosive geomorphology with a near-complete absence of natural lakes. Given this lack of suitable habitat, no native floating plants have evolved in the region (Coetzee & Hill, 2012). However, South Africa is a water scarce country and is required to build numerous and often large artificial impoundments throughout the country to secure reliable water supply. This forms a previously absent ecological niche for non-native invasive species free from competition of native species.

Additionally, South Africa adopted an unusually high Phosphorus limit of < 1 mg/litre for waste water treatment works (WWTWs) in the 1970s (Department of Water Affairs and Forestry (DWAF), 1996). Despite the lenient concentration standard, WWTWs frequently fail to meet this requirement (Quayle *et al.,* 2010; Dennis & Dennis, 2019). This has produced severe levels of eutrophication in many South African freshwater systems, resulting in fertile, nutrient abundant habitat for aquatic invaders. Many South African reservoirs (e.g. Hartbeespoort Dam) occur downstream of populated, urban settlements where the industrial waste and sewerage effluent can flow into the system.

In response to these conditions, South Africa has a long history of macrophyte invasions. The first records of naturalised water hyacinth populations date back as early as 1910, and over the subsequent decades, it has spread widely across the country, invading streams, rivers, wetlands and artificial impoundments (Hill *et al.* 2020) (Figure 2.1). In addition to water hyacinth, there are four other historically important invasive macrophytes in South Africa: *Pistia stratiotes* (water lettuce), *Salvinia molesta* (Kariba weed), *Myriophyllum aquaticum* (parrot's feather) and *Azolla filiculoides* (red water fern). Together, they are referred to as the "Big Bad Five" macrophyte invaders of South Africa. Introduced as ornamental plants, all of these species originate from South America and evolved in large, warm and slow-moving waters, features largely absent in South Africa outside of artificial impoundments. They have flourished in large part due to the abundant nutrient supply and lack of natural enemies found in their native range (Coetzee *et al.,* 2011).

With the exception of water hyacinth, the "Big Bad Five" macrophyte species are considered to be under successful biological control (Coetzee & Hill, 2012). The biological control of water hyacinth has remained variable, particularly in the eutrophic and temperate systems of South Africa (Coetzee & Hill, 2012). Therefore, along with newer targets for biological control, such as the submerged species *Egeria densa* and *Hydrilla verticillata*, and emergent species *Sagittaria platyphylla*, water hyacinth remains a key focus of biological control research in South Africa.

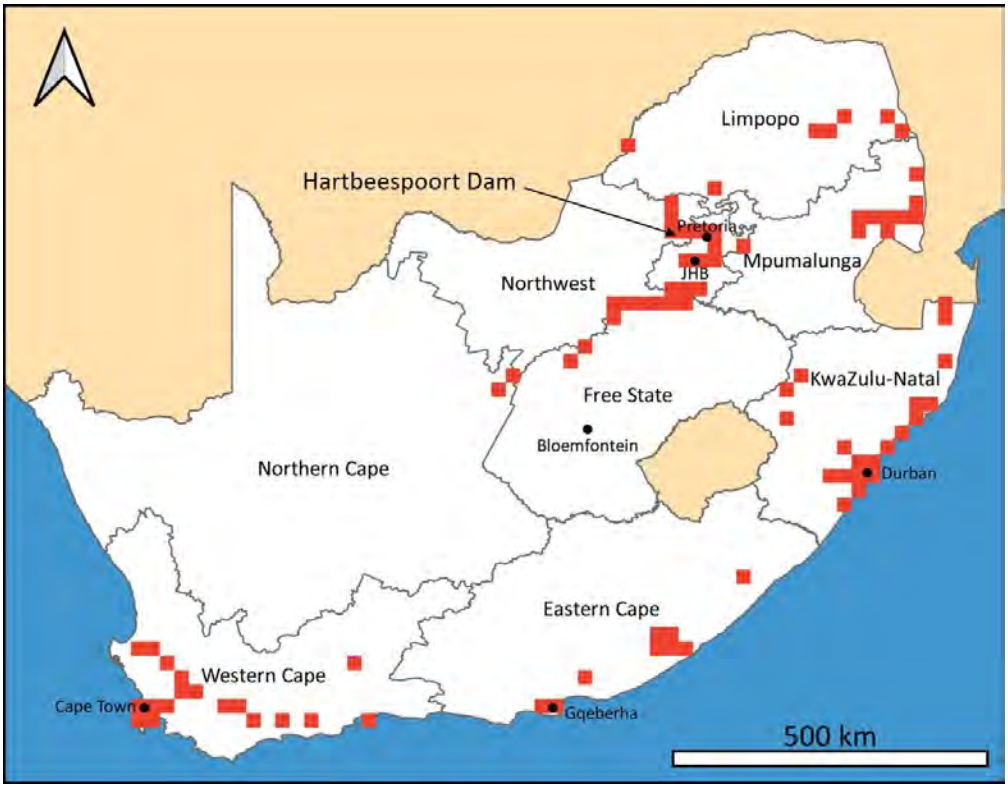


Figure 2.1: Water hyacinth distribution in South Africa (quarter degree squares). Data from the Centre for Biological Control, Rhodes University.

2.2 Water Hyacinth

Water hyacinth is a flowering plant that originates from the Amazon Basin, South America, but has since been transported, originally as a prized ornamental plant, throughout the globe. In its natural environment, it plays an important ecological role, providing habitat, shelter and feeding grounds for many species (Thamaga & Dube, 2018b).

However, outside its natural range, it has become invasive in many tropical and subtropical regions and is regarded as one of the top 10 worst weeds in the world (Shanab *et al.*, 2010). It is a rapid growing, free-floating plant that forms dense mats that alter the physical and chemical composition of aquatic ecosystems (Shanab *et al.*, 2010). These mats block waterways, obstructing the function of transport systems, tourism activities, hydroelectric plants and irrigation schemes (Shanab *et al.*, 2010). Ecologically, the dense mats reduce dissolved oxygen levels in the environment, altering phytoplankton levels and reducing fish numbers (Shanab *et al.*, 2010). The dense mats are also well suited for potential disease vectors such as mosquito larvae and the bilharzia snail (van Wyk & van Wilgen, 2002).

The plant propagates profusely through asexual, vegetative reproduction (Byrne *et al*., 2010) but also produces long-lived seeds (van Wyk & van Wilgen, 2002; Hill *et al.*, 2020). However, the seeds require relatively specific conditions (i.e. warm, shallow water) to germinate and thus most populations primarily grow and spread by vegetative means (van Wyk & van Wilgen, 2002). Under optimum conditions, water hyacinth has a very high biomass growth rate of up to 6 % per day (Ashton *et al.* 1979), and once established, can quickly colonise aquatic ecosystems and become difficult to control.

2.3 The effect of environmental variables on the growth of water hyacinth

It is well established that nutrients and temperature are the two most important factors influencing water hyacinth growth dynamics (Wilson *et al.*, 2005; Byrne *et al.,* 2010) and South Africa suffers from severe eutrophication levels in many of its water systems. For example, Matthews (2014) found that 31 of South Africa's 50 largest lakes are hypertrophic. The growth of water hyacinth has a direct correlation with concentrations of phosphorus (P) and nitrogen (N) specifically, and research shows that water hyacinth biomass at nutrient-rich sites can increase eightfold compared to biomass nutrient poor sites (Coetzee & Hill, 2012). Thus, the abundant nutrient supply has created conditions for water hyacinth and other macrophytes to proliferate. As a result, macrophyte invasions of eutrophic systems are in many cases regarded as a symptom of the eutrophication, rather than a unique condition on their own (Hill & Olckers, 2001; Byrne *et al.* 2010).

Temperature, like nutrient supply, has a significant and direct effect on water hyacinth growth rates. The optimum temperature is about 30°C, with no growth occurring once temperature drops below 10°C (Wilson *et al.,* 2000; Wilson, 2002). This accounts for the lower water hyacinth growth rates in more temperate, cooler regions of South Africa, particularly in winter. Given its natural range in the tropical regions of South America, the plant has not evolved the necessary mechanisms for surviving low temperatures (Owens & Madsen, 1995). Despite this, it can withstand near freezing temperatures (<5°C) for a limited time (maximum of 3 weeks) owing to its carbohydrate stores near the stem base, although these kinds of conditions reduce its regrowth potential over time (Owens & Madsen, 1995). For this reason, Wilson *et al.,* (2000) note that frost is likely to have the most significant effect on population growth, as frost increases the loss of biomass and is a major cause of leaf mortality in temperate regions (Wilson *et al.,* 2005).

Aside from nutrients and temperature, other environmental factors have been observed to impact water hyacinth population. Flooding can disrupt large mats and leave plants stranded on land or flush them downstream, although small pockets will likely remain on the edges enabling the population to re-establish (Wilson *et al.,* 2000). Wind and wave action, on larger water bodies, can also break up mats and cause them to sink, but this appears to be only significantly effective after other (e.g. agent) damage to the plant has taken place (Hill & Olckers, 2001). Salinity is a major limiting factor in coastal regions, where a concentration of more than 0.2 % is enough to kill the plants (Wilson *et al.,* 2000).

2.4 Control Methods of Water Hyacinth

2.4.1 Mechanical

Mechanical control, such as using rakes and pitchforks to manually remove the plant from the water, is one of the key pillars of Working for Water (WfW), a government funded programme which aims to provide jobs through invasive plant removal (Hill & Coetzee, 2017). However, this method of water hyacinth control is generally considered ineffective and is not promoted (Hill & Coetzee, 2017; Hill *et al.*, 2020). For example, up to EUR 26,000,000 (~ ZAR 514,000,000) was spent to manually remove water hyacinth from the Guadiana River, Spain, between 2005 and 2015 (Hill *et al.*, 2020). Despite

this, water hyacinth has reinvaded, most likely from seed or overlooked plants, and has spread further along the river. Thus, that management option for water hyacinth had failed (Hill *et al.*, 2020).

2.4.2 Herbicide

Herbicidal control, chiefly in the form of glyphosate, has historically been used to control larger water hyacinth populations in South Africa (Hill & Coetzee, 2017). Whilst its effect is almost immediately observable, the key to successful herbicide programmes is continued, long-term follow-up applications to control re-emergence from seed and scattered plants (van Wyk & van Wilgen, 2002). The lack of a long-term follow-up programme is a significant cause of failure of herbicide control strategies (Hill & Olckers, 2001). The costs of this method are also significant, comprising the majority of the ZAR 42,000,000 spent by the Department of Environmental Affairs between 2010 and 2018 in controlling freshwater macrophytes (Hill *et al.*, 2020). Herbicidal damage to non-target species also present a challenge to this approach, but as Byrne *et al*. (2010) note, the presence of water hyacinth in some cases may be potentially more ecologically damaging than the glyphosate herbicide itself.

2.4.3 Biological Control

One of the key factors that contribute to a species' invasiveness is the absence of its natural enemies. Biological control takes advantage of this by introducing the organism's host-specific natural enemies from its native range (van Wyk & van Wilgen, 2002). The first biological control agent for water hyacinth in South Africa, *Neochetina eichhorniae* (Coleoptera: Curculionidae), was introduced in 1974 (Hill & Olckers, 2001). Since then, eight additional agents have been introduced: *Orthogalumna terebrantis* (Acari: Sarcoptiformes: Glamunidae), in 1989; *Niphograpta albiguttalis* (Lepidoptera: Crambidae), in 1990; *Neochetina bruchi* (Coleoptera: Curculionidae), also in 1990; *Cercospora piaropi* (Mycosphaerellales: Mycosphaerellaceae), in 1992*; Eccritotarsus catarinensis* (Hemiptera: Miridae), in 1996; *Eccritotarsus eichhorniae* (Hemiptera: Miridae), in 2008; *Cornops aquaticum* (Orthoptera: Acrididae), in 2011; and most recently, *Megamelus scutellaris* (Hemiptera: Delphacidae), in 2013 (Hill & Coetzee, 2017).

Despite the high number of biocontrol agents introduced, the highest of any country worldwide, the results until recently have been variable (Hill & Olckers, 2001; van Wyk & van Wilgen, 2002) and water hyacinth remains South Africa's worst water weed (Byrne *et al.*, 2010; Coetzee & Hill, 2012). This contrasts with notable success in other regions such as Lake Victoria, Uganda, where *Neochetina* spp., aided by stormy weather from an El Nino event, caused a significant decline in the water hyacinth population (Wilson *et al.*, 2007). It is worth noting that in areas in South Africa where populations have not been significantly reduced, biological control has still lessened the overall impact and plants generally become smaller in size (Hill & Olckers, 2001) and biological control of water hyacinth is still considered the most sustainable method for reducing population size (Coetzee & Hill, 2012).

2.4.4 Integrated Control Strategies

Integrated control combines two or more of the control methods outlined above. In real-world conditions, biological control does not happen in silos. Particularly under certain eutrophic conditions, an approach that combines biological control and careful herbicide application is potentially the most effective (Coetzee *et al*., 2011; Coetzee & Hill, 2012). Sub-lethal doses of herbicide, on the periphery of the main spray strips for example, stunt the weeds sufficiently to halt growth and reproduction, but still allows for agent populations to persist on the plants (Byrne *et al*., 2010; Coetzee *et al*., 2011). van

Wyk and van Wilgen (2002) compared the costs of three different control methods for water hyacinth: herbicide only, biocontrol only and an integration of the two. They found that integrated control was slightly more cost-effective than only biological control, and five times more cost-effective than only herbicide. However, this was a single case study and the costs of different methods will most likely vary by site.

2.5 Hartbeespoort Dam

Hartbeespoort Dam is situated near the eastern boundary of the North West province and about 30 kilometres west of Pretoria. Located in the highveld, summer rainfall region of South Africa, the area experiences warm summers (mean 26°C) and cool winters (mean 11°C) with occasional frost in the winter season (South African Weather Service data). The dam's surface area is about 1820 hectares and has a capacity of about 193 million m^3 (Scott *et al.*, 1977). Hartbeespoort Dam has two main tributaries: the Crocodile River, the primary inflow, from the south and the Magalies River from the west. The tributary catchment area spans about 410 000 hectares, extending southward of the dam, and its southern boundary covers the northern half of Johannesburg (Figure 2.2: A). Its primary function is to provide irrigation and domestic water (van Wyk & van Wilgen, 2002) but the dam is also a popular and accessible tourist destination, particularly from the nearby cities. There are also several large housing estates around the dam.

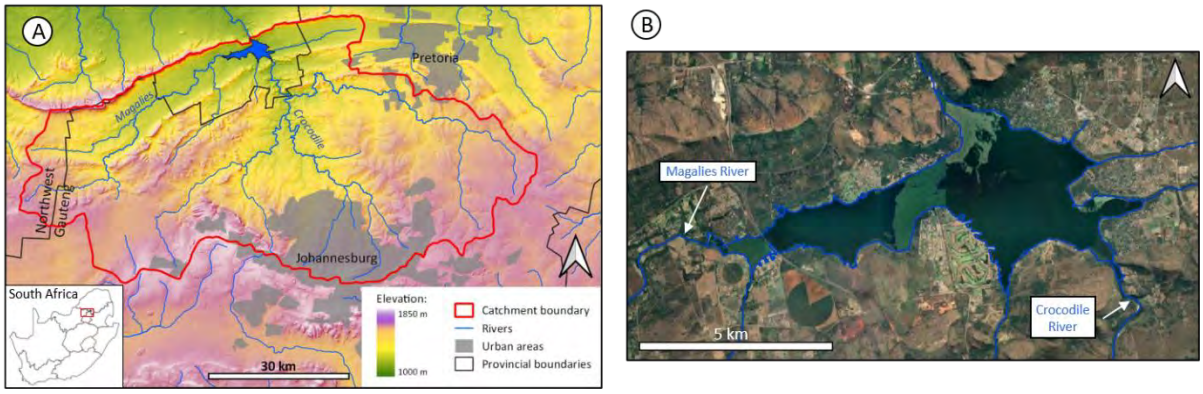


Figure 2.2: (A) Elevation map of the catchment area of Hartbeespoort dam and (B) Hartbeespoort Dam with the two main tributaries listed.

Hartbeespoort Dam is severely eutrophic and is regarded as one of the most polluted water systems in the country (Harding, 2008; Matthews, 2014). The high nutrient influx, notably phosphorus (P) and nitrogen (N), is derived from a combination of wastewater treatment works, sewerage leakages from informal settlements, industrial waste and agricultural runoff into the river system, with the outflows from WWTWs being recognised as the main source of nutrients into the dam (Dennis & Dennis, 2019). The Crocodile River, which originates from northern Johannesburg, is the catchment source for most of these treatment plants. Of the 10 WWTWs in the Hartbeespoort catchment, eight of them feed into the Crocodile River (Dennis & Dennis, 2019).

Between 2010 and 2017, an average influx of 582 and 1687 tonnes per annum of total phosphorus (TP) and total nitrogen (TN), respectively, were calculated on the dam by Carroll & Curtis (2021). The mean annual influx also increased significantly over the same period and the corresponding outflux was significantly lower for both TP (39 %) and TN (53 %). The dam is therefore a significant nutrient

sink, with much of the excess TP and TN being stored in the sediment of the dam and the water column (to a lesser extent).

The eutrophic conditions have substantial down-stream effects on the ecological functioning of Hartbeespoort Dam. Toxic cyanobacterial blooms, catalysed by high nutrient levels, have frequented the dam for decades, particularly during the warm summer months (Zohary, 1985). The presence of these cyanobacteria blooms severely degrades the water quality, to the extent that swimming in the dam poses a serious threat to human health. Tourism is an important economic activity at Hartbeespoort Dam, but toxic water limits the kinds of recreational activities that can take place on the dam.

Reducing the nutrient flux from its sources, before it is deposited into the dam, is the only viable long-term solution for reducing eutrophication on Hartbeespoort Dam (Curtis & Carroll, 2021). Nevertheless, there have been significant attempts to reduce existing nutrient loads on the dam. For example, the 'Harties Metsi a Me' programme, which ran from 2006 to 2016, used bioremediation – by physically removing water hyacinth and cyanobacterial mats from the water – to reduce the nutrient load and plant abundance on the system (Mitchell & Crafford, 2016). However, while the programme may have achieved some localised improvements within the dam, the changes to the total nutrient load would have been proportionately small compared with the total nutrient influx (Carroll & Curtis, 2021).

2.5.1 Control efforts of water hyacinth on Hartbeespoort Dam

Alongside cyanobacteria, water hyacinth has flourished in response to the eutrophication on Hartbeespoort Dam for decades. The first local records date back to 1959, with high levels of water hyacinth cover, as much as 60 %, reported by the late 1970s (Ashton *et al*., 1979). The dam, until very recently, periodically achieved similar cover levels to those reported almost 50 years ago.

Numerous mechanical, chemical and biological control efforts to address the water hyacinth problem on Hartbeespoort Dam have occurred over the decades. The 'Harties Metsi a Me' programme, funded by the South African government through the Water Research Commission (WRC), used manual and mechanised methods to physically remove the plant from the water as part of their bioremediation programme. From 2006 to 2014, the project expenditure totalled over ZAR 158 million (Mitchell & Crafford, 2016). Despite this considerable cost and time spent on the project, which was terminated in 2016, the effectiveness of its approach to controlling water hyacinth was limited, despite small-scale successes. Furthermore, job creation from the project was proportionately low for a programme of this budget size (less than 90 full time staff in any given year) and the expenditure fell short of the ZAR 504 million budget allocated towards the programme (Mitchell & Crafford, 2016). Therefore, full utilisation from the project was arguably not achieved, which may have contributed to its limited success.

The use of herbicidal control on water hyacinth on Hartbeespoort Dam was first trialled in 1977 and showed promising results (Ashton *et al.,* 1979). Although costly, the practice produced rapid results and was used intermittently until a moratorium on herbicide spraying on the dam was put in place in 2017. Anecdotally, the practice has continued since then, although at a limited scale. Herbicidal spray, chiefly in the form of glysophate, suffers from similar limitations as mechanical removal, whereby remaining water hyacinth plants from sheltered areas or upstream sources, in conjunction with a high seedling fecundity, can rapidly re-establish in the months following the spray event. However,

retardant doses of glysophate, which do not kill the plants but induce severe stress, has been showed to work synergistically with biological control methods (Jadhav *et al*., 2008).

A biocontrol programme was implemented at Hartbeespoort Dam for water hyacinth in the 1990s, where five agent species were introduced: *Neochetina eichorniae* (mottled water hyacinth weevil) and *Neochetina bruchi* (chevroned water hyacinth weevil), *Niphograpta albiguttalis* (water hyacinth moth), *Orthogalumna terebrantis* (water hyacinth mite) and *Eccritotarsus catarinensis* (water hyacinth mirid) (Coetzee *et al*., 2022). The method employed a classical biological control strategy, where relatively low numbers of agents are introduced onto the plants, which incrementally grow in number and eventually establish self-sustaining population numbers on the target host. Although these agents did establish, all of which occur on the dam to present, their population growth was limited by cold winters and herbicidal spray (Coetzee *et al*., 2022). Given these conditions and the eutrophic state of the dam, the agent population growth lagged behind the water hyacinth growth, and therefore effective biological control was limited.

Introduced into South Africa in 2013, the plant hopper *Megamelus scutellaris* is the most recent addition to the suite of water hyacinth biological control agents in the country (Hill & Coetzee, 2017). It is a phloem-feeding bug that induces significant tissue damage to water hyacinth, reducing its growth rate as well as increasing plant mortality (Sutton *et al*., 2016). Since there has been limited success of water hyacinth control in the more temperate regions of South Africa, *M. scutellaris* was of particular interest due to its observed ability to still thrive in the cooler regions of its native habitat in South America (Hill & Olckers, 2001). The cold tolerance of *M. scutellaris* outside its native range was observed by Tipping *et al*. (2014), who reported that the species had successfully over-wintered in frost-prone areas for three consecutive years in Florida, United States of America.

Besides its tolerance to cold, there is a synergism between *M. scutellaris* and naturally occurring pathogens, as the insects can facilitate pathogen infection which results in a greater control impact on the water hyacinth (Sutton *et al*., 2016). This effect is most evident under eutrophic conditions, which makes this approach well suited to many South African waterbodies.

2.6 Remote Sensing of Water Hyacinth

Remote sensing is the science of obtaining information about a spatial object, without coming into direct contact with it (Lillesand *et al*., 2015). Some satellite sensors receive sun electromagnetic (EM) radiation reflected from the earth surface and record it in an image for analysis. Different objects reflect different EM frequencies, which allow for the objects to be distinguished and mapped either manually or, more frequently, by computer algorithms. These ‘passive’ satellite sensing techniques, which require solar illumination, have been used for decades. The most well-known example is perhaps National Aeronautics and Space Administration’s (NASA) Landsat satellite series, which has conducted continuous earth observation from the 1970s and is still currently in operation. Active remote sensing methods, such as synthetic aperture radar (SAR), are increasingly utilised too. Due to their own illumination source, and relatively long wave frequencies, SAR sensors have the ability to conduct observations at night and through cloud cover, a distinct advantage over passive sensors (Datta *et al*., 2021).

The advent of accessible low-cost and freely available satellite data in recent years, along with increased frequency and coverage, has significantly increased the monitoring potential for water hyacinth with remote sensing (Datta *et al*., 2021). Traditionally, water hyacinth has been monitored using conventional methods, typically in the form of field surveys (Dube *et al*., 2017). However, this approach has several significant drawbacks, such as being spatially restricted, time consuming, labour-intensive and expensive (Ritchie *et al*., 2003). Conversely, archived and near-real-time satellite imagery provides opportunity for rapid and comprehensive spatiotemporal analyses of water hyacinth and other aquatic weeds (Thamaga & Dube, 2018b). Due to the existing image archives, this method of monitoring, besides being more time-efficient and cost-effective, also allows for retrospective analysis of past events (Byrne *et al.*, 2010).

2.6.1 Data sources and methods of existing water hyacinth remote sensing studies

Remote sensing has been used to measure water hyacinth from at least the 1990s (Venugopal, 1998; Everitt *et al*., 1999) and the number of published literature on the topic has increased significantly in recent years (Thamga & Dube, 2018b). Wilson *et al.* (2000) suggest the most useful parameters for monitoring water hyacinth population dynamics are biomass density and cover area, variables that can be derived with remote sensing. Whilst some research has explored using remote sensing to estimate water hyacinth biomass density (e.g.: Robles *et al.,* 2015), there is a significant body of literature on detecting and measuring area cover of water hyacinth and other macrophytes with remote sensing techniques (Thamaga & Dube, 2018b). As part of an integrated management control plan for water hyacinth, Byrne *et al*. (2010) noted that analysis of free satellite imagery had the potential for use in monitoring both the physiological status and physical extent of water hyacinth, provided that it had a temporal frequency of at least twice yearly. In the 12 years since that publication, satellite image availability has progressed significantly, most notably with the Sentinel-2 satellite constellation, which has a revisit time of at least 5 days (Matongera *et al*., 2017).

The three most popular satellite platforms for detecting water hyacinth are Sentinel-2 (Thamaga & Dube, 2019; Singh *et al*., 2020; Mucheye *et al*., 2022), the Landsat series (Oyama *et al*., 2015; Dube *et al*., 2017; Mukarugwiro *et al*., 2019; Singh *et al*., 2020) and the Moderate Resolution Imaging Spectroradiometer (MODIS) (Kiage & Obuoyo, 2011; Fusilli *et al*., 2013). Similar to MODIS, the Medium Resolution Imaging Sensor (MERIS) sensor was also a popular choice (Cheruiyot *et al*., 2014; Liang *et al*., 2017) but was discontinued in 2012. While all three of the existing platforms are freely available to the public, each present different trade-offs, primarily in the number of spectral bands, spatial resolution, temporal resolution and historical catalogue (Table 2.1).

Table 2.1: A comparison of the three main satellite platforms used in RS for water hyacinth

Satellite(s)	Sensor	Spectral bands	Spatial resolution (m)	Temporal resolution (days)	Year of launch
Sentinel-2A/B	MSI	13	10, 20, 60	5	2015/2017
Landsat 8/9	OLI/OLI-2	9	30	16 (8 when combined)	2013/2021
Aqua/Terra	MODIS	36	250, 500, 1000	1-2	1999

MSI: Multispectral Instrument; OLI: Operational Land Imager; MODIS: Moderate Resolution Imaging Spectroradiometer.

While MODIS contains the longest temporal dataset and highest temporal resolution, its spatial resolution was disqualifying for the purposes of this study. Sentinel-2 was most suitable for this study, as its relatively high temporal resolution (5 days) and 10/20 metre spatial resolution gave it a distinct advantage over the Landsat series. Crucially, the high temporal resolution presented the best opportunity, of all the freely available data sources, for the information to be used in a monitoring capacity.

There is a wide range of methods used to detect water hyacinth and other floating macrophytes in the remote sensing literature (Table 2.2). The simplest, and most popular, is index thresholding (also known as index slicing) (Kiage & Obuoyo, 2011; Fusilli *et al*., 2013; Cheruiyot *et al.*, 2014; Oyama *et al.*, 2015; Liang *et al.,* 2017). More complex machine learning approaches such as discriminant analysis (Dube et al., 2017; Thamaga & Dube, 2019), random forest (Mukarugwiro *et al*., 2019; Singh *et al*., 2020), support vector machine (Mukarugwiro *et al*., 2019) and maximum likelihood (Cheruiyot *et al*., 2014) are also used.

Table 2.2: A comparison of ten water hyacinth RS studies, highlighting the different satellite platforms, bands, indices, classifications and accuracies obtained in each.

Study	Platform, sensor	Significant bands	Indices	Classification method	Accuracy
Dube *et al*. (2017)	Landsat 8, OLI	Green (band 3), NIR (band 5), SWIR (bands 7 and 8)	-	Discriminant analysis	Overall accuracy = 72%
Cheruiyot *et al*. (2014)	Envisat, MERIS	-	NDVI	Index slicing, maximum likelihood	Overall accuracy of 75 % (52 % on sparse vegetation) and 80 % (72 % on sparse vegetation), respectively
Liang *et al*. (2017)	Envisat, MERIS	Blue, green, SWIR	CMI, FAI, TWI	Index slicing	Overall accuracy of 86 %
Kiage & Obuoyo (2011)	Aqua/Terra, MODIS	-	NDVI	Index slicing	Not reported
Fusilli *et al*. (2013)	Aqua/Terra, MODIS	-	NDVI	Index slicing	Visually inspected (qualitative)
Mukarugwiro *et al*. (2019)	Landsat 8, OLI	-	-	Random Forest, Support Vector Machine	Overall accuracy of 85 % and 65 %, respectively
Thamaga & Dube (2019)	Sentinel-2, MSI	Red edge (bands 5, 6, 7), NIR (band 8), NIR-narrow (band 8A), SWIR (bands 11, 12)	16 different indices, including NDVI, NDWI and SAVI	Linear discriminant analysis	Overall accuracy of 77 %
Singh *et al*. (2020)	Sentinel-2, MSI and Landsat 8, OLI	-	MNDWI, NDVI, GARI, LSWI	Index slicing and random forest	Overall accuracy of 80 %
Oyama *et al*. (2015)	Landsat 5, TM and Landsat 7, ETM+	SWIR (bands 7 and 8)	NDVI, NDWI, FAI,	Index slicing	Not reported
Mucheye *et al*. (2022)	Sentinel-2, MSI	-	NDVI	Index slicing	Not reported

NIR: near-infrared; SWIR: shortwave infrared; NDVI: normalised difference vegetation index; CMI: cyanobacteria and macrophyte index; FAI: floating algae index; TWI: topographic wetness index; NDWI: normalised difference water index; SAVI: soil-adjusted vegetation index; MNDWI: modified normalised difference vegetation index; GARI: green atmospherically resistant index; LSWI: land surface water index.

Index thresholding is popular for its computational simplicity and proven results (Cheruiyot *et al*., 2014). For example, Liang *et al.* (2017) used three moderate resolution imaging spectroradiometer (MODIS)-based indices, Cyanobacteria and Macrophytes Index (CMI), Floating Algae Index (FAI) and Turbid Water Index (TWI), to successfully discriminate water, macrophytes and algae distributions on Lake Taihu, a large waterbody in China. On Lake Victoria, several studies have used Normalised Difference Vegetation Index (NDVI) thresholds to map water hyacinth on the lake (Kiage & Obuoyo, 2011; Fusilli *et al.,* 2013; Cheruiyot *et al.*, 2014). While NDVI is the most popular index for water hyacinth, it has poor performance with algae/cyanobacteria and water hyacinth separation (Liang *et al.*, 2017). In this scenario, indices that use the shortwave infrared (SWIR) bands have been preferred due to the spectral absorbency of water in this frequency (Oyama *et al*., 2015; Liang *et al*., 2017). The

variety of methods reflects the varying spatial and temporal scales requirements of different studies, available data sources and software options, and is indicative of a larger phenomenon in remote sensing where there is a continued development of a wide variety of classification algorithms (Woodcock *et al*., 2020).

Remote sensing has been used to produce time-series of water hyacinth on large water bodies in the past. For example, Kiage & Obuoyo (2011) used MODIS-derived NDVI to produce a time-series of water hyacinth area on Lake Victoria from 1997 to 2010, with one image per year (n = 14). This low sample rate, by contemporary standards, is indicative of the technological limitations that RS has faced as large image analysis has been limited by the available computing and storage capacities of local desktop computers (Woodcock *et al*., 2020). With the exception of Singh *et al*. (2020), the studies outlined above have been used on relatively small sets of image collections, highlighting the opportunity for the development of a dense time-series algorithm for water hyacinth.

2.6.2 A 'paradigm shift' in remote sensing

In the last few years, due the development of two key technologies, the remote sensing field has undergone a 'paradigm shift' in its scope and capability (Azzari & Lobell, 2017; Woodcock *et al*., 2020). The first is the exponential increase of available data from satellite platforms, both freely available and proprietary. This provides the benefit of increased coverage area as well as higher revisit times, allowing for finer scale temporal analysis. In addition, the spatial resolution of these technologies has improved. Historically, satellite images tended to have resolutions measured in hundreds of metres, but with the advent of satellites like Landsat 8 & 9, Sentinel-2 and PlanetScope, the multispectral ground resolutions have increased to 30, 10 and 3 metres respectively.

The second key technology is the introduction of cloud-based remote sensing platforms, such as Google Earth Engine (GEE) (Gorelick *et al.*, 2017). These systems allow for the storage and analysis of large (petabyte scale) datasets that previously would have been prohibitively expensive or time-consuming for most analysts (Gorelick *et al*., 2017). Although other cloud computing remote sensing platforms are available, such as Microsoft's Planetary Computer, GEE is currently the most established platform and thus the more popular choice for most scientists. As an indication of its growing popularity, the number of remote sensing publications that use GEE has increased from five in 2015 to 299 by 2020 (Zhao *et al*., 2021).

In combination, these two technologies have resulted in an exponential increase of new use cases and applications of remote sensing, expanding its scope and utility in a wide range of fields. As more data and computational power become available, analysts can train better classification algorithms with more data, and at faster speeds and larger scales than before. Importantly, many of these resources are openly available, which has allowed the democratisation of these technologies (Gorelick *et al*., 2017).

This enables the production of dense time-series analysis in a way that has not been possible previously (Azzari & Lobell, 2017). Along with the increased quality and quantity of data being produced, dense time-series allows for more subtle changes and how those changes respond to other variables to be measured (Woodcock *et al*., 2020). Further, reduced delays between image acquisition and availability, often within 24 hours from observation, provides near-real-time information. As a result of this, a new application of remote sensing has emerged, shifting from primarily a change detection tool into a monitoring tool (Woodcock *et al*., 2020).

2.6.3 Challenges and opportunities

Remote sensing, as an observation tool for the extent and density of water hyacinth, does present some challenges, however. Data from Sentinel-2, which represents the best resolution for openly available data, has a pixel resolution of 10x10 meters for its near-infrared, red, green and blue band channels. This introduces an inherent level of granularity into the data and thus potential difficulty in discriminating water hyacinth from other vegetation and other spectrally similar phenomena (e.g. cyanobacteria) (Oyama *et al*., 2015; Liang *et al*., 2017). The approach also has some barriers to entry, as hardware and software can be expensive, and the use of these systems requires appropriate training and expertise (Gorelick *et al*., 2017; Thamaga & Dube, 2018b). However, free and open-source software (such as QGIS and Google Earth Engine) and sensor platforms (such as the Sentinel and Landsat constellations) are becoming increasingly available, and accessible.

Data quality also presents a challenge. Floating macrophytes often co-occur with other floating macrophytes species, algae and cyanobacteria blooms (Oyama *et al*., 2015; Liang *et al.*, 2017). Sun glint, cloud cover and thick aerosols also pose challenges for accurate classification which requires scientists to balance data quality versus data quantity (Liang *et al*., 2017). Robust filtering algorithms and integration with other data sources, such as synthetic aperture radar, present viable solutions to ensuring the production of high quality and continuous RS data (Nagendra *et al*., 2013)

More broadly, there are opportunities for RS to provide impactful information to scientists in other fields of research. Nagendra *et al*. (2013) note that the lack of integration of knowledge between specialists and remote sensing analysts often leads to this approach being underused and undervalued. Further, as Thamaga & Dube (2018b) point out, hydrologists, aquatic scientists and environmentalists would benefit greatly from integrating remote sensing into their work, and that effective environmental management is likely best achieved when these two fields collaborate.

Timely monitoring information may prove crucial to waterbody managers, scientists and other stakeholders as it provides real time information about sensitive ecosystem responses to management interventions. The research contained within this thesis aims to bridge this gap, particularly as remote sensing offers significant potential for water hyacinth monitoring on Hartbeespoort Dam and beyond.

Chapter 3: The development of a mapping and monitoring method of water hyacinth

3.1 Introduction

The use of remote sensing to measure and monitor water hyacinth has rapidly increased in recent years due to the increased data availability and improved computational resources available to remote sensing software (Thamaga & Dube, 2018b; Singh et al., 2020). Compared to traditional, infield monitoring methods, RS is advantageous in several important areas. Firstly, it provides a view of the entire study site, which can be measured accurately and objectively. Secondly, it drastically reduces the need for time spent monitoring in the field at sites - which are often difficult to access. This reduces the required man-hours and other expenses associated with infield work. Thirdly, the stored image datasets allow for retrospective analysis.

However, many past analyses have often employed a 'snapshot' approach, whereby a few images, periodically spaced over time, are measured and the change is observed and quantified. This is partially due to the computational limitations of traditional desktop-based RS software, but also because smaller systems invaded by water hyacinth are more spatially stable. However, given the high spatio-temporal characteristics of the water hyacinth mats on Hartbeespoort Dam combined with the rapid population fluctuations, these traditional change detection techniques for mapping water hyacinth were found to be inadequate. This provided an opportunity to investigate and develop a tailored remote sensing technique that leveraged large datasets and cloud computing to improve data resolution to account for the high distribution variability. Additionally, the information produced needed to be accessible in a timely fashion, to inform the on-going water hyacinth control programmes on the dam.

3.1.1 Aim and objectives

This chapter aimed to develop a remote sensing method to create a detailed time-series of water hyacinth on Hartbeespoort Dam and make data from newly acquired images available as a near-real-time monitoring resource through three objectives:

1) The first objective identified the site-specific classification challenges and proposed solutions where appropriate.
2) The second objective developed a repeatable, computationally efficient classification algorithm that accounts for the challenges highlighted in objective 1. This includes features such as cloud filtering and suitable spectral indices.
3) The third objective was to provide an accessible version of the data to stakeholders with near-real-time updates via a WebApp interface to make the data easily accessible.

3.2 The challenges of remote sensing water hyacinth on Hartbeespoort Dam

In the context of Hartbeespoort Dam, this section investigated the potential challenges of using remote sensing to map water hyacinth. This was undertaken in order to gain an understanding of study site and identify potential obstacles that may need to be overcome to create accurate time-series and monitoring data. These challenges were identified from understanding other similar studies (e.g. Thamaga & Dube, 2018a; Liang *et al.,* 2017; Singh *et al.*, 2020) as well as through the exploratory analysis of satellite images of Hartbeespoort Dam. Solutions to these problems, when applicable, were suggested.

3.2.1 High spatio-temporal variability

A key concern was the high spatial variability of the floating plants which produce significantly different measurements between consecutive satellite observation events. Environmental variables, primarily wind speed and direction, have a notable effect on its distribution over short periods of time, as illustrated by Figure 3.1. This variability is somewhat random but is most notable during the summer months. Most importantly, besides the changing position of the floating mats, their absolute area was highly variable from wind-induced compression.

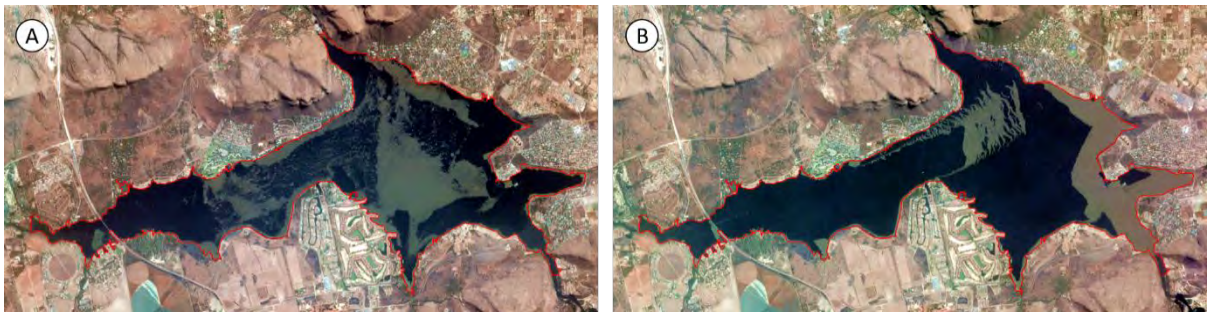


Figure 3.1: PlanetScope satellite images of water hyacinth on Hartbeespoort Dam, captured 24 hours apart on 30 August 2019 (A) and 31 August 2019 (B). Images © 2023 Planet Labs PBC.

In the literature, this issue is left mostly unaddressed, in part due to an absence of suitable data sources with a revisit frequency high enough to highlight this problem, particularly in older studies, or simply from a lack of need. For example, water hyacinth mats constrained by tighter river boundaries (e.g. Thamaga & Dube, 2018a) do not suffer as much from this issue. The solution to this was to increase the number of observations, substituting the poor reliability of individual observations with a higher quantity of total observations, producing an average trend that better reflects the true coverage area.

3.2.2 Cloud cover

Clouds are an inherent limitation to optical remote sensing. In a monitoring context where up-to-date information is required, this becomes a significant limitation. Hartbeespoort Dam lies within a summer rainfall region, resulting in an increase in cloud cover over the summer months which coincides with the peak water hyacinth growth period. Periods of weeks, sometimes months, of cloudy observations introduced large gaps in the time-series data. Some studies, such as Dong *et al.* (2020), have combined optical data with Synthetic Aperture Radar (SAR) satellite data to account for cloud cover and to provide better granularity and fill missing gaps in time-series. SAR is resistant to most cloud cover and therefore was proposed as an alternative data source to optical data during cloudy periods.

3.2.3 Algae and Cyanobacterial blooms

Hartbeespoort Dam has a history of algae and cyanobacterial blooms due to the high levels of eutrophication. Although its occurrence was limited on the dam during the study period (August 2015 – March 2021), it shares many spectral similarities with water hyacinth (Oyama *et al.*, 2015; Liang *et al.*, 2017) and careful attention was required to account for these blooms to avoid misclassification errors. Investigation into the spectral differences between the water hyacinth and the blooms was required to determine a suitable index that what reliably distinguish them apart.

3.3 Methods

3.3.1 Overview

The entire remote sensing workflow was conducted on the GEE platform. GEE is a freely available, cloud-based computing platform developed by Google for RS scientists (Gorelick *et al.*, 2017). The script-based interface with a cloud-computing backend allows for powerful RS workflows and boasts an extensive data catalogue with a well-documented application programming interface (API) reference library.

During the development of the classification method, ensuring that the workflow could be executed without the need for further manual input was crucial, as it allowed for the successful deployment of the method as an ongoing monitoring tool. Therefore, data preparation, analysis, and output processes were designed to be fully automated, with the goal of ensuring scalability and repeatable execution functionality without the need of further specialised input or knowledge once established.

3.3.2 Data sources

The primary data source was the Sentinel-2 Level-1C top of atmosphere (TOA) from the GEE data catalogue (ImageCollection ID: COPERNICUS/S2_HARMONIZED) for the time period between August 2015 and March 2021. It is a proven remote sensing data source that has been successfully used to map water hyacinth in South Africa in recent years (Thamaga & Dube, 2018b; Singh *et al.*, 2020).

The Sentinel-2 constellation comprises two satellites, arranged in a sun-synchronous orbit with a phase difference of 180 degrees (European Space Agency, 2022). Each satellite has an overpass frequency of 10 days, producing a combined 5-day revisit time at the equator. In higher and lower latitudes, including the area over Hartbeespoort Dam, the revisit time was reduced even further (2-3 days) due to overlapping swath paths. The MultiSpectral Instrument (MSI) onboard the satellite constellation is sensitive to a range of multispectral frequencies, of which the blue, green, red, near-infrared and short-wave infrared frequency bands were used in this study (Table 3.1).

Table 3.1: Detail of the Sentinel-2 image bands used in this thesis

Description	Band name	Wavelength (nm)	Resolution (m)
Blue	B2	494	10
Green	B3	560	10
Red	B4	665	10
Near-infrared (NIR)	B8	835	10
Short-wave infrared 1 (SWIR)	B11	1612	20

Sentinel-1 C-band Ground Range Detected (GRD) Synthetic Aperture Radar (SAR) (ImageCollection ID: COPERNICUS/S1) was used as an additional source of data, particularly over periods of high cloud-cover. The Synthetic Aperture Radar (SAR) radio waves are resistant to all weather conditions except for very dense rain cells, presenting a feasible data substitute over cloudy periods. GEE currently only supports Sentinel-1 Ground Range Detected (GRD) data and not the Single Look Complex (SLC) data which are required for decomposing and creating a scatter matrix. Therefore, values such as scatter entropy and alpha (reflectance angle), which are valuable indicators of surface structure (e.g. Prudente *et al.,* 2019) were unavailable.

However, polarimetric parameters obtainable from GRD data are still useful for basic analysis. The standard GRD data contain four primary bands, comprising two single co-polarisation bands: horizontal transmit and receive (HH) and vertical transmit and receive (VV); and two dual-band cross-polarisation bands: horizontal transmit/vertical receive (HV) and vertical transmit/horizontal receive (VH) (European Space Agency, 2016). The availability of the horizontal transmit bands (HH, HV) is limited in GEE and was not available for South Africa. Each scene was pre-processed by GEE using the Sentinel-1 Toolbox, where they undergo thermal noise removal, terrain correction and backscatter coefficient (σ°) radiometric calibration. The final GRD values are converted to decibels (dB) with log scaling and fall within a maximum and minimum range of +1 and -50. All bands are co-registered, allowing for direct comparison.

Regarded as an additional data source, the Sentinel-1 data were not incorporated into the primary classification method outlined in this chapter. This ensured algorithmic simplicity by reducing the potential for process errors once published and computational complexity. A smaller, adjacent Sentinel-1 remote sensing method was conducted, primarily to highlight the potential for multi-source data harmonization in future work.

3.3.3 Creating a boundary mask

The first step in the time-series classification was to mask out Hartbeespoort Dam from the rest of the image scene. The masked images were significantly smaller – a Sentinel-2 tile covers > 1.2 million hectares while Hartbeespoort Dam occupies only ~ 0.15 % (1820 ha) of that area. By masking out the dam, it effectively reduced the required classification groups down to three; water hyacinth, algae/cyanobacteria and water, which simplified the classification process and reduced computation times.

Existing boundary masks were available, but their resolution did not match the Sentinel-2 10 m pixel scale and thus introduced significant edge-pixel contamination. Further contamination was introduced by imperfect georeferencing between existing masks and Sentinel-2 images. Therefore, a

Sentinel-2-derived boundary layer that masked the dam at an equivalent resolution to the analysis images was created.

With reference to Figure 3.2 below, the mask was produced with all 654 Sentinel-2 images (a) from the study period (August 2015 – March 2021), filtered for clouds (b) and converted to Normalised Difference Water Index (NDWI, Gao (1996)) to identify water pixels (c). A per-pixel ordering function available in GEE ('qualityMosaic') composited the NDWI images to create a single mosaicked image (d) that contained the maximum NDWI value for each cell stack over the time period. NDWI pixels with a value greater than 0.1 were extracted from the composite (e) to create the final mask product that represented the full water extent of Hartbeespoort and was free from macrophyte obstruction (f). The largest continuous area by size was extracted to remove adjacent, smaller water bodies within the scene (such as near-by ponds and heavy hillside shadows).

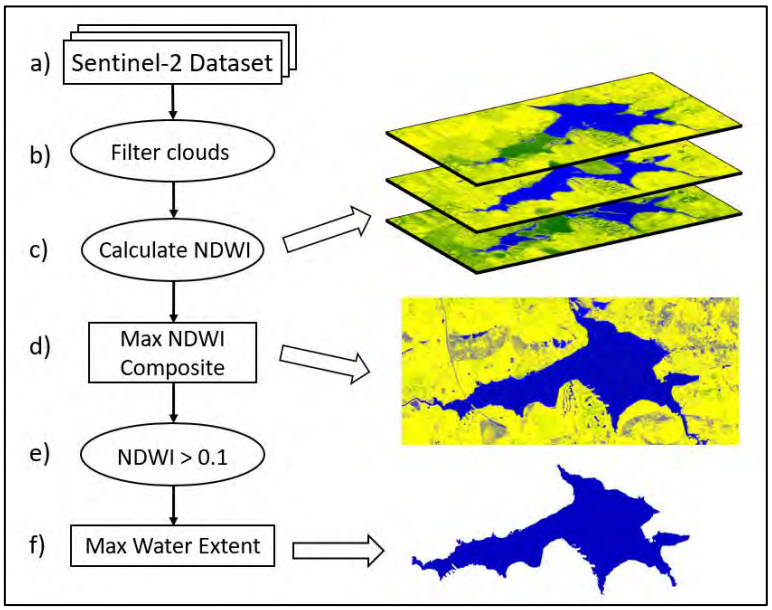


Figure 3.2: Workflow to calculate boundary mask. NDWI represents the normalised difference water index (Gao, 1996).

3.3.4 Filtering clouded images

As the classification measured the total macrophyte area, it was important that the input images were completely cloud-free. Each Sentinel-2 image contains information in the metadata (property: CLOUD_COVERAGE_ASSESSMENT) about the total cloud percentage of the whole scene. This is useful for exploratory analysis of image collections but excluded instances where the study site may be cloud-free despite most of the scene containing clouds, and vice versa. Therefore, a method that assessed the cloud cover only over the dam was used instead.

With the Sentinel-2 analysis, the quality band (QA60) is frequently used to mask or filter clouds over the study area (e.g. Sun *et al.,* 2019, Liu *et al.*, 2020, Wang *et al.*, 2020, Jia *et al.*, 2021). The band, which has a 60 m pixel resolution, was too coarse at the scale of Hartbeespoort Dam and did not detect smaller clouds.

Instead, a pre-calculated cloud probability dataset called 's2cloudless' (https://github.com/sentinel-hub/sentinel2-cloud-detector) was used. The algorithm, developed by researchers at Sinergise, is

based on a gradient boosting machine learning algorithm (Zupanc, 2017). S2cloudless was advantageous in this study as it had a native pixel resolution of 10 m, a pseudo-probability (0 – 100) that could be adjusted to the specific use-case, and was available in GEE's data catalogue (ImageCollection ID: COPERNICUS/S2_CLOUD_PROBABILITY). In an extensive comparison of 10 cloud masking algorithms for Landsat 8 and Sentinel-2, s2cloudless performed exceptionally well in most contexts (Skakun *et al.*, 2022).

To determine the presence of clouds over the dam in each image, the cloud probability layer was joined with its corresponding Sentinel-2 image by date (Figure 3.3: a) and masked with the maximum water extent layer of Hartbeespoort Dam (Figure 3.3: b). The mean value of the remaining cloud probability pixels was extracted (Figure 3.3: c) and images with an average cloud probability greater than 0.1 for the entire the dam area were marked as containing clouds and removed (Figure 3.3: d). A threshold of 0.4 is typically used for s2cloudless (Zupanc, 2017), but the more conservative 0.1 threshold was used to ensure strict cloud exclusion.

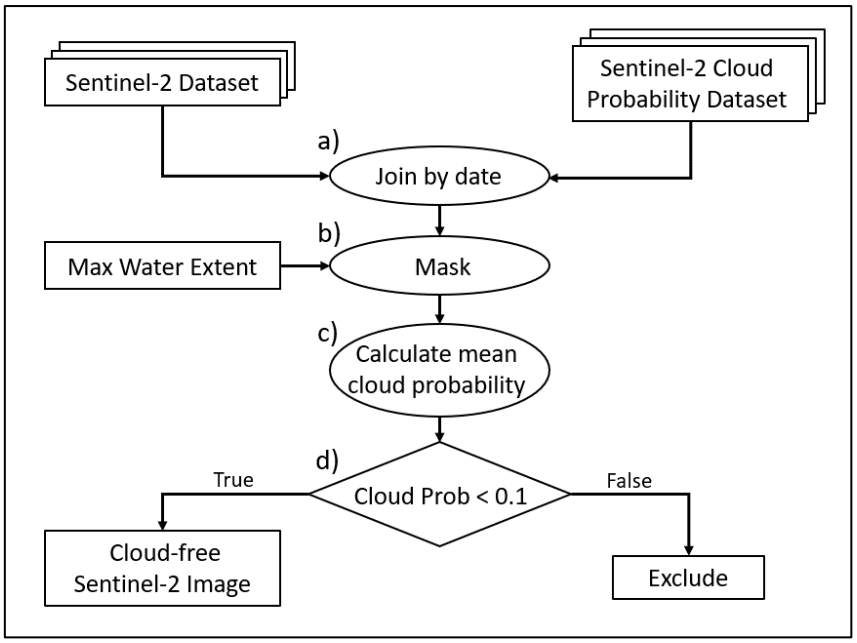


Figure 3.3: Cloud-removal workflow

3.3.5 Spectral index selection

This section details the selection of the two indices that were used in the decision tree classifier below. In order to determine suitable indices for detecting water hyacinth, a list of common spectral indices, compatible with the available Sentinel-2 bands (Table 3.2) were tested and examples are given in Table 3.3 below. Each index was visually inspected over three different scenarios on the dam: water hyacinth only, mixed water hyacinth and algae/cyanobacteria, and algae/cyanobacteria only (Table 3.3).

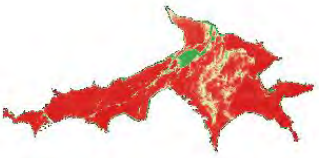
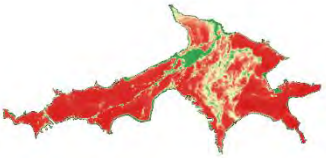
The Normalised Difference Water Index (NDWI) (Gao, 1996) was chosen to identify water pixels in each image scene as it was found to be reliable index even during observation instances with high glare and water turbidity. However, due to the similar spectral properties in most bands between water hyacinth and algae blooms on Hartbeespoort Dam, a separate index was required that further classified the remaining non-water pixels after the NDWI threshold classification.

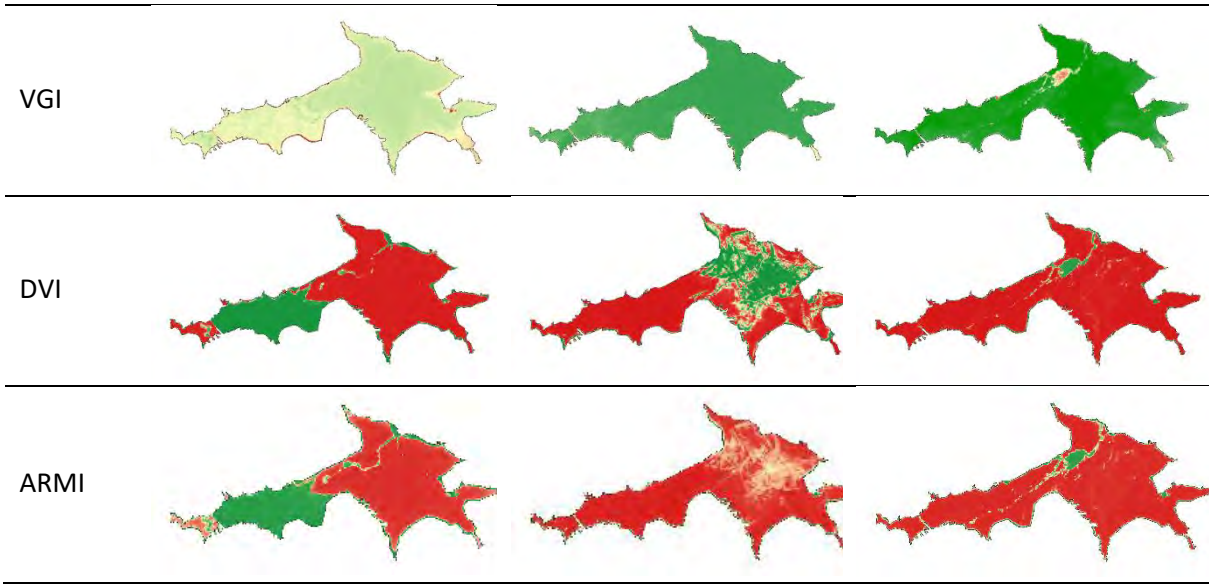
While NDWI performed well in identifying water presence, none of the indices satisfactorily distinguished between water hyacinth and algae/cyanobacteria blooms. It was noted that, consistent with previous research by Oyama *et al.* (2015) and Liang *et al.* (2017), certain existing algorithms and indices are inadequate in accurately distinguishing between macrophytes and algae as a result of their comparable reflectance spectra, particularly in the red and near-infrared frequencies. Therefore, this presented an opportunity to develop an alternative index, based on the available Sentinel-2 bands, to discriminate between the two classes for use in the second threshold step of the decision tree classifier used in this chapter.

Table 3.2: A list of common Sentinel-2 indices tested for macrophyte and algae detection and separation

Index	Full name	Formula	Reference
CLgreen	Chlorophyll Index Green	$\left(\frac{NIR}{Green}\right) - 1$	Gitelson *et al.*, (2003)
DVI	Difference Vegetation Index	$NIR - Green$	Tucker *et al.*, (1977)
GNDVI	Green Normalised Vegetation Index	$\frac{NIR - Green}{NIR + Green}$	Metternicht (2003)
NBR	Normalised Burn Ratio	$\frac{NIR - SWIR\,1}{NIR + SWIR\,1}$	Key & Benson (2006)
NDVI	Normalised Difference Vegetation Index	$\frac{NIR - Red}{NIR + Red}$	Tucker (1979)
NDWI	Normalised Difference Water Index	$\frac{Green - NIR}{Green + NIR}$	Gao (1996)
RDVI	Renormalised Difference Vegetation Index	$\sqrt{NDVI * DVI}$	Chen (1996)
SAVI	Soil-Adjusted Vegetation Index	$\frac{(NIR - Red)(1 + L)}{NIR2 + Red + L}$	Huete (1988)
SRI	Simple Ratio Index	$\left(\frac{NIR}{Red}\right)$	Chen (1996)
TVI	Transformed Vegetation Index	$\sqrt{\frac{NIR - Red}{NIR + Red} + 0.5}$	Bannari *et al.*, (2002)
VGI	Vegetation Greenness Index	$\frac{Green - Red}{Green + Red}$	Bannari *et al.*, (1995)

Table 3.3: An example of each index applied to three different image scenarios on Hartbeespoort Dam. Note: green represents high values, whilst red represents low values.

Index	Water Hyacinth	Cyanobacteria	Mixed
NDVI			
NDWI			
NBR			
SRI			
SAVI			
GNDVI			
CLgreen			
TVI			
RDVI			



3.3.5.1 Development of a novel index for distinguishing macrophytes from algal blooms

To develop an effective index, it was necessary to understand the spectral properties between water hyacinth and algae to potentially identify and leverage those differences between the two classes. The two classes were sampled across 10 Sentinel-2 bands to identify potentially useful spectral patterns. In a desktop analysis, spectral reflectance values were sampled across multiple images that contained high water hyacinth NDVI, low water hyacinth NDVI, algae, and water. The results were plotted on a line graph (Figure 3.4). Band reflectance values of the lowest and highest NDVI water hyacinth periods, observed in winter and summer respectively, were plotted to show the seasonal variation of the spectral reflectance. The shaded green area represents the spectral range of water hyacinth between that is found between the two seasonal extremes.

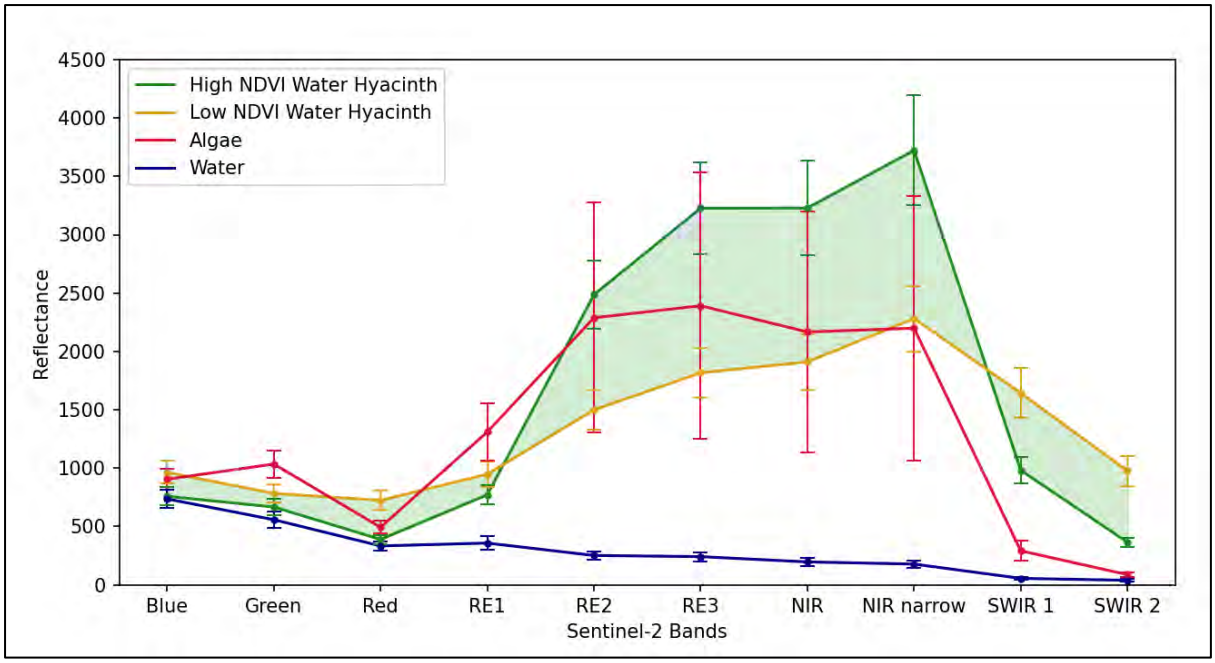


Figure 3.4: Sentinel-2 spectral reflectance properties of high/low NDVI water hyacinth, algae and water. The green shaded area represents the variation of water hyacinth reflectance values between the maximum (high) and minimum (low) NDVI observations. Error bars represent standard deviation.

The largest spectral distinction between macrophytes and algae was observed in the Short-Wave Infrared (SWIR) bands, and to a lesser extent in the Green and Red Edge 1 (RE1) bands. The shortwave infrared (SWIR) spectral separation properties can be attributed to the structural dissimilarities between the aquatic plant water hyacinth and algal mats. Surface water, which is highly absorbent in SWIR wavelengths, is partially or completely obscured by water hyacinth (Figure 3.5: a). Conversely, algae/cyanobacteria occur at and below the water surface (Figure 3.5: b) resulting in a lower SWIR reflection compared to water hyacinth.

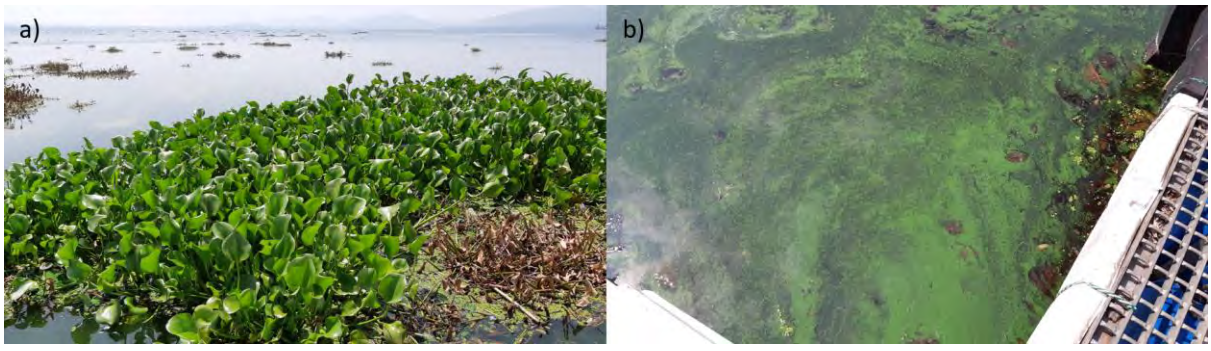


Figure 3.5: Photographs of water hyacinth (a) and algae (b) on Hartbeespoort Dam. Note the obscuration of water by the water hyacinth mat, compared to the algae/cyanobacteria, which sits just below the water surface.

Other studies have observed this spectral distinction in the SWIR wavelength too. For example, Oyama *et al.* (2015) used the SWIR bands from the Landsat TM/ETM+ sensors and Liang *et al.* (2017) used the MODIS SWIR bands to distinguish macrophytes from algae, due to the almost full spectral absorption of water in that frequency range. However, no published literature using the Sentinel-2 SWIR bands for macrophyte and algae discrimination was found at the time of writing. More broadly, several

indices have been developed for monitoring algal blooms, such as the cyanobacteria index (CI), the maximum chlorophyll index (MCI) and the floating algal index (FAI) (Oyama *et al.*, 2015). However, the shortfall of all these indices is their use of the NIR band, resulting in poor distinction between algae and macrophytes.

In Figure 3.6, the inverse relationship between water hyacinth and algae in the SWIR 1 and green bands is illustrated more clearly by plotting the spectral values of the different classes in the green and SWIR bands on the y and x axes respectively. Whilst the Red Edge 1 band presented a similar inverse relationship with SWIR 1, the Green band was preferred for its higher spatial resolution.

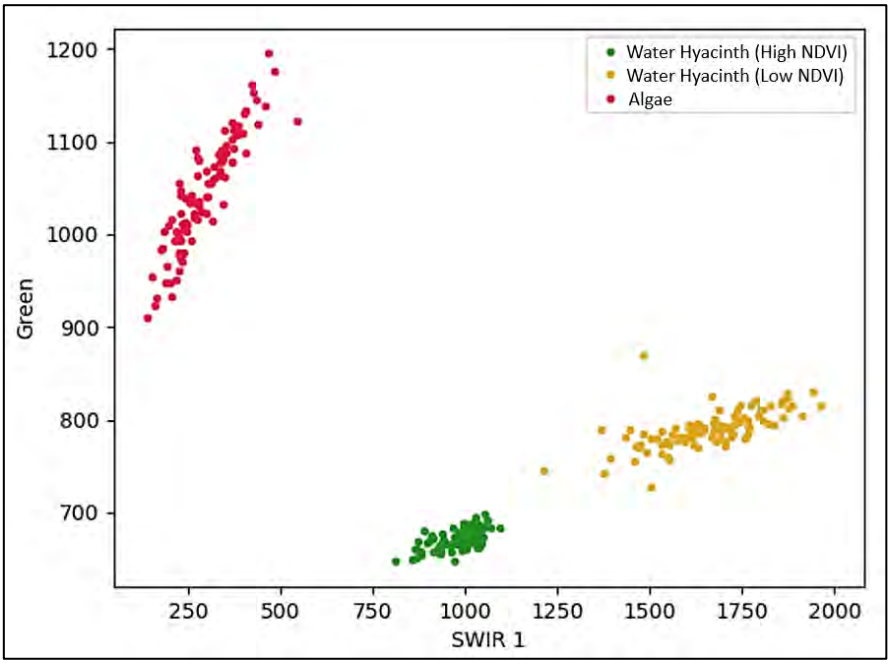


Figure 3.6: The inverse spectral relationship of water hyacinth and algae in the shortwave infrared (SWIR) 1 and Green Sentinel-2 bands.

Based off this observation, a new index was created with these two bands. For the purpose of this thesis, it was termed the Algae Resistant Macrophyte Index (ARMI) however it should be noted that it applies to cyanobacterial blooms too, as these are often co-present and are referred to interchangeably here. The formula for ARMI is:

$$ARMI = \frac{SWIR\ 1\ -\ Green}{SWIR\ 1\ +\ Green}$$

Following the same methodological reasoning behind other normalised indices such as NDVI (Tucker, 1979) and NDWI, the ARMI index takes advantage of the inverse relationship between the SWIR 1 and Green bands by calculating their difference through subtracting the Green band from the SWIR 1 band, and then dividing them by their sum to normalise the values between the ranges of +1 and -1. Positive values indicate high water hyacinth probability (and other macrophytes) and are resistant against detecting algal mats due to the differences in surface water obscuration in the SWIR frequency. Negative values are likely to indicate either algal or water presence. Figure 3.7 illustrates the difference between the NDWI and ARMI indices on Hartbeespoort Dam. The term ‘macrophyte’ was used instead of ‘water hyacinth’, as this index was not required to distinguish between macrophyte

species and can be applied to a wider range of floating macrophytes that occur above the water surface.

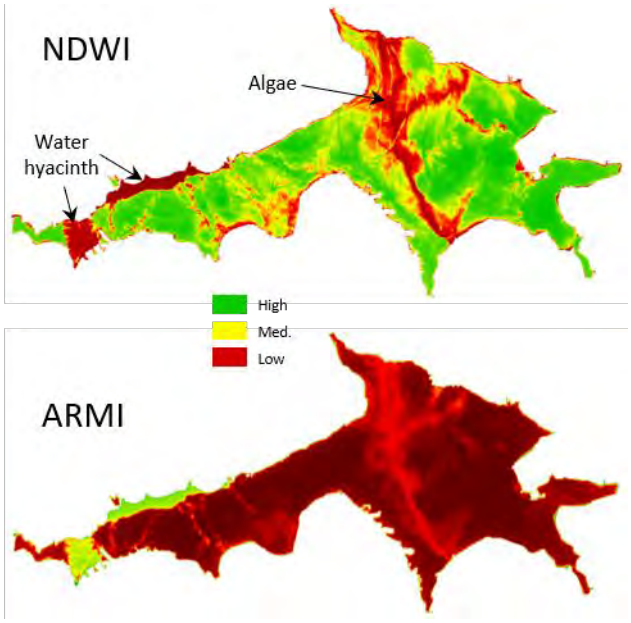


Figure 3.7: A direct comparison of NDWI and ARMI on Hartbeespoort Dam, 16 December 2016. High NDWI values represent water, whereas high ARMI indices represent macrophyte mats. Note the clear distinction between water hyacinth and algae pixels within the ARMI image.

ARMI enabled the separation of non-water pixels (from the product of the NWDI threshold in the decision tree classifier) into separate water hyacinth and algae classes, which accounted for much of the water hyacinth classification inaccuracy when algal blooms were present on the dam (Figure 3.8).

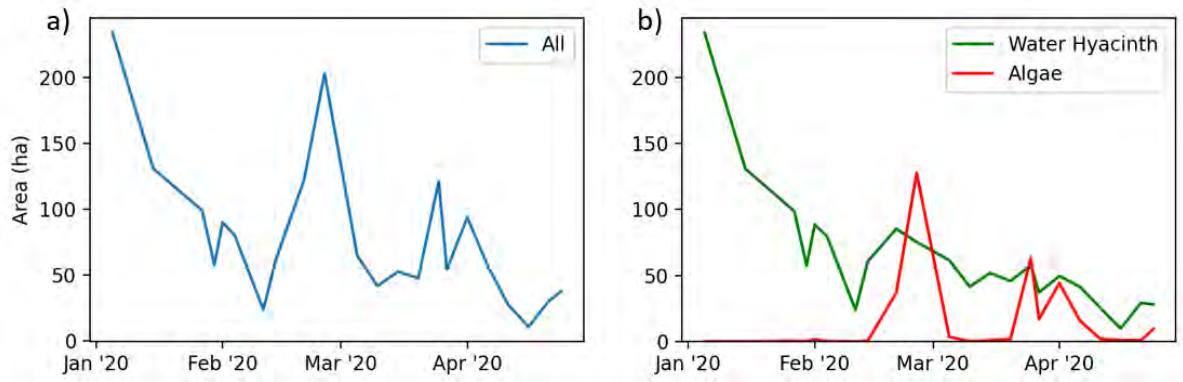


Figure 3.8: Time-series data after NDWI thresholding (a) and then ARMI thresholding (b) to separate pixels into water hyacinth and algae. Note the classification of algae in b) accounts for most of the variation seen in a) over this period as water hyacinth steadily decreases.

3.3.6 Decision tree classifier to distinguish water hyacinth, algal and water pixels

This thesis used an empirical, index-based threshold classification method which is a popular classification approach for detecting macrophytes (Kiage & Obuoyo, 2011; Fusilli *et al.*, 2013; Cheruiyot *et al.*, 2014; Oyama *et al.*, 2015; Liang *et al.*, 2017). Due to the required multi-class output (water, macrophytes/water hyacinth and cyanobacteria/algae), a decision tree was used with two different index-based threshold classifications. This was similar to the study by Liang *et al.*, (2017), where the authors created a decision tree that thresholded MODIS-derived spectral indices to distinguish cyanobacterial scums and macrophytes on Lake Taihu, China. The benefit of an index-based approach, in part, is their relative computational simplicity (Cheruiyot *et al.*, 2014).

Due to the eutrophic conditions, algal/cyanobacterial blooms periodically occur on Hartbeespoort Dam. Although reductions in large blooming events were observed once water hyacinth became dominant on the dam in 2017, it was still occasionally present over the study period and thus the classifier needed to detect and discriminate these blooms. Therefore, NDWI was used to identify water pixels in the image scene, and ARMI was used on the remaining pixels to identify macrophytes only.

The NDWI (Gao, 1996) and ARMI were calculated as:

$$NDWI = \frac{Green - NIR}{Green + NIR} \quad and \quad ARMI = \frac{SWIR\ 1\ -\ Green}{SWIR\ 1\ +\ Green}$$

In the first step of the decision tree (Figure 3.9: a), NDWI was calculated from the masked and filtered Sentinel-2 images. The NDWI pixels ($NDWI_i$) with values greater than the NDWI threshold ($NDWI_{th}$) were classed as water. In the second step (Figure 3.9: b), the remaining NDWI pixels formed a mask for the ARMI layer, which was calculated from the same original image. ARMI pixels ($ARMI_i$) greater than the ARMI threshold ($ARMI_{th}$) were classed as water hyacinth, while the remainder were classed as algae. The calculation of the threshold values for NDWI and ARMI are described in the next section.

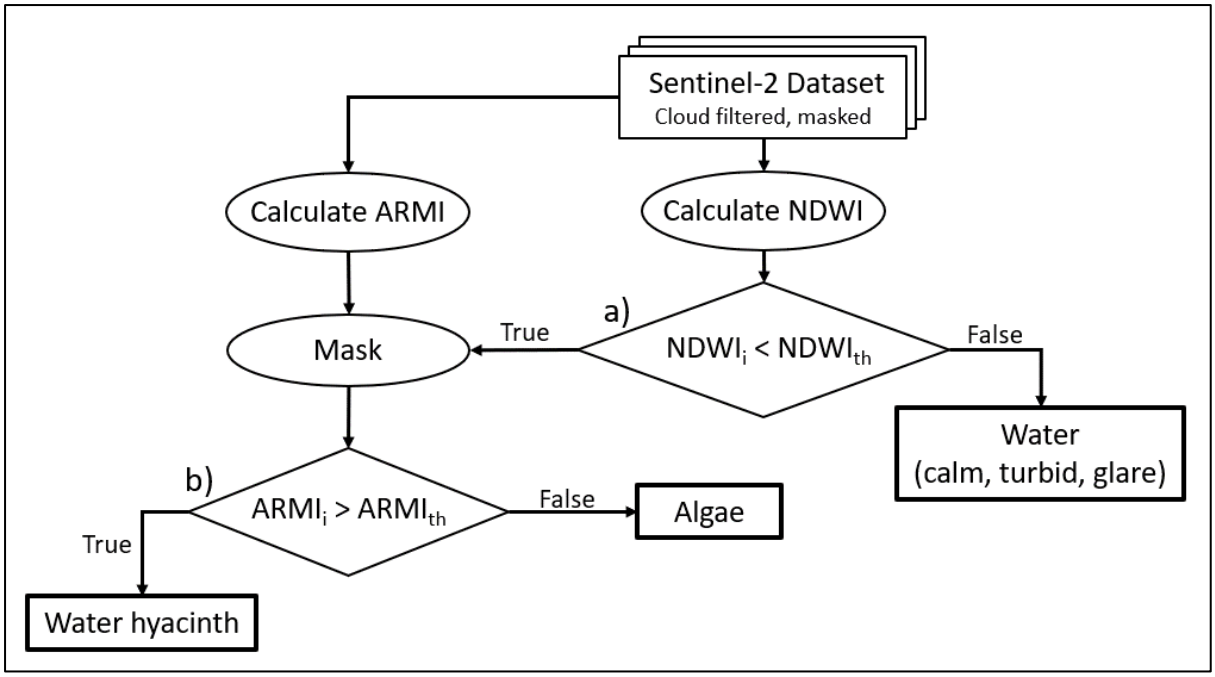


Figure 3.9: Classification decision tree to distinguish water hyacinth, algae and water. ARMI and NDWI are acronyms for Algal Resistant Macrophyte Index and Normalised Difference Water Index, respectively.

This classifier model iterated over each qualifying Sentinel-2 image over the study period, from August 2015 until March 2021. The output values for each class, presented as pixel counts, were converted to hectares (n pixels / 100) and percentages (pixel area / dam area x 100).

3.3.7 NDWI and ARMI Thresholds for the decision tree

The decision tree classifier used empirically determined NDWI and ARMI thresholds, also known as 'index slicing', to classify water, water hyacinth and cyanobacteria pixels in each image scene. To calculate the NDWI threshold, 14 cloud-free images of Hartbeespoort Dam were manually selected to represent a variety of dam conditions, such as instances of high glare and turbidity, as well as representation throughout the seasons. NDWI was calculated for each representative image and sampled with a 20 m point grid covering the entire dam (n = 45510). The same process was conducted with ARMI but with 6 images due to limited instances of definitive algae presence on the dam. The point values from each image set were collated into a single dataset for statistical classification.

There is no single method for determining a threshold and methods vary between different studies (Fusilli *et al.*, 2013; Cheruiyot *et al.*, 2014; Oyama *et al.*, 2015; Liang *et al.*, 2017). For example, Oyama *et al.*, (2015) used a threshold classification method for separating water, macrophytes and cyanobacterial blooms with Landsat-derived NDVI and NDWI indices. However, their threshold values were determined by the averaged minimum and maximum values between each class based on manually sampled pixels (n = 150 per class) from a single Landsat image. Given the significantly higher number of samples over multiple reference images that were used in this study, there was no clear separation between classes, but rather a continuum, and thus a statistical clustering method was used instead. Spectral index clusters were calculated with the Elkan version of the K-means clustering algorithm with the Scikit-Learn Python library (Pedregosa *et al.*, 2011). K-means clustering is an unsupervised machine learning technique and is one of the most common algorithms for finding

clusters in data (Elkan, 2002). The value that separated the K-means clusters served as static thresholds in the decision tree classifier (Figures 3.13 & 3.14).

3.3.8 Detecting Vegetation Health

Normalised Difference Vegetation Index (NDVI) (Tucker, 1979) is a common metric used to indicate plant health and provides managers with valuable information on the status of the plants on the dam. It utilises the red and near infrared wavelengths to measure photosynthetic activity to produce a relative metric that is normalised between -1 and +1.

To accurately determine the mean NDVI value of the water hyacinth, a dynamic mask that covered only the densest water hyacinth distribution in each image was used. This reduced contamination of surface water spectra that were present in less dense mat areas and edges. To achieve this, an NDWI version of each cloud-free, masked image was calculated (Figure 3.10: a) whereby all pixels with a value greater than -0.5 were extracted to form a mask (Figure 3.10: b). An NDVI version of the same image was then calculated (Figure 3.10: c) and masked with the dense water hyacinth mask (Figure 3.10: d). The NDVI values of the remaining pixels were averaged (Figure 3.10: e) to produce a single mean NDVI value per image scene.

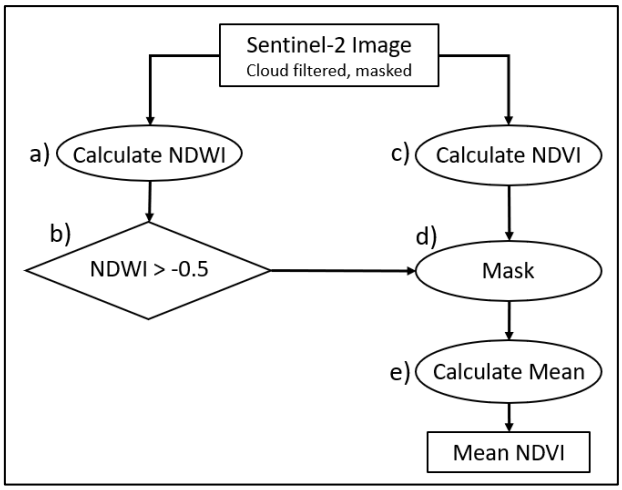


Figure 3.10: Workflow for calculating mean NDVI for each Sentinel-2 image

3.3.9 Sentinel-1 Classification

The dam boundaries were masked with the same boundary used for the Sentinel-2 images, leaving the two remaining classes of water and water hyacinth. As algae occurs below the water surface, which mostly reflects SAR waves, it was not detectable and thus did not need to be included in the classification. Vertical transmit/vertical receive (VV) images from the descending pass were used as they contained the lowest noise level and distortion over Hartbeespoort Dam. As with the Sentinel-2 classifier, the threshold between the two classes was identified with K-means clustering from pixel samples across multiple image scenes.

For the accuracy assessment of the Sentinel-1 results, agreement between Sentinel-1 and Sentinel-2 datasets were compared and the days of coincident observation were evaluated with a linear regression. All Sentinel-1 observations occurred around 03:34 am, resulting in no closely timed optical reference images. Therefore, there were no coincident high-resolution observations to further

compare the accuracy of the results. However, this was deemed sufficient given its ancillary role as a supporting source of data to the main Sentinel-2 classification.

3.4 Results

3.4.1 Cloud-free images

From the 1st August 2015 to the 31st March 2021, there were a total of 655 Sentinel-2 images of Hartbeespoort Dam, of which 386 (58.93 %) were classed as being cloud-free. Cloud-free images followed a strong seasonal pattern and were most abundant during the winter period, as the summer rainfall season is frequented by cloud cover (Figure 3.11). The number of available images doubled after March 2017, following the launch of a second satellite (Sentinel-2B) which reduced the revisit time from 10 to 5 days or less.

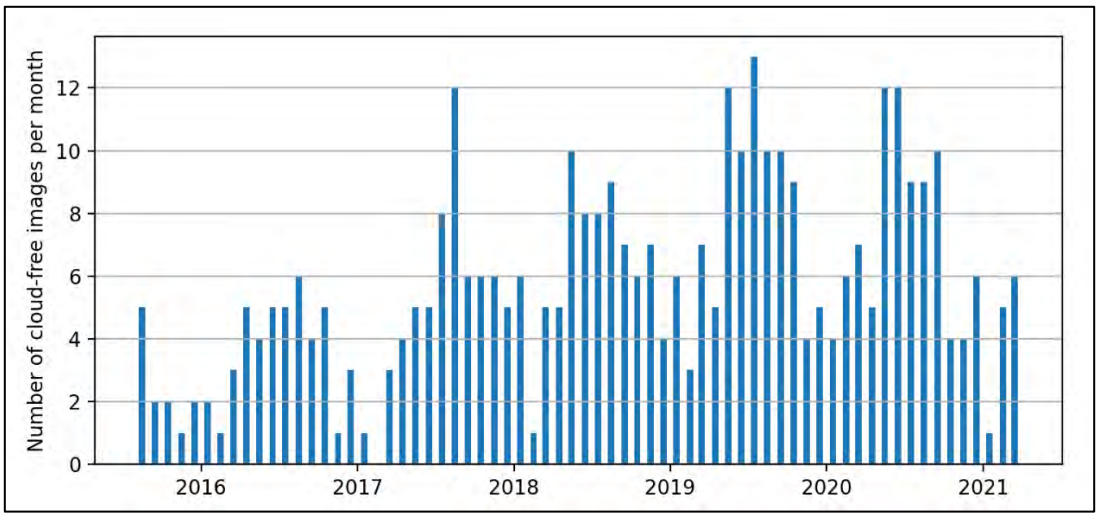


Figure 3.11: Number of cloud free images per month from August 2015 to March 2021.

When grouped by month, January and February had the lowest proportion of cloud-free images at 33 % and 30 % over the entire period, respectively (Figure 3.12). The winter months from May to August typically had an image availability over 80 %.

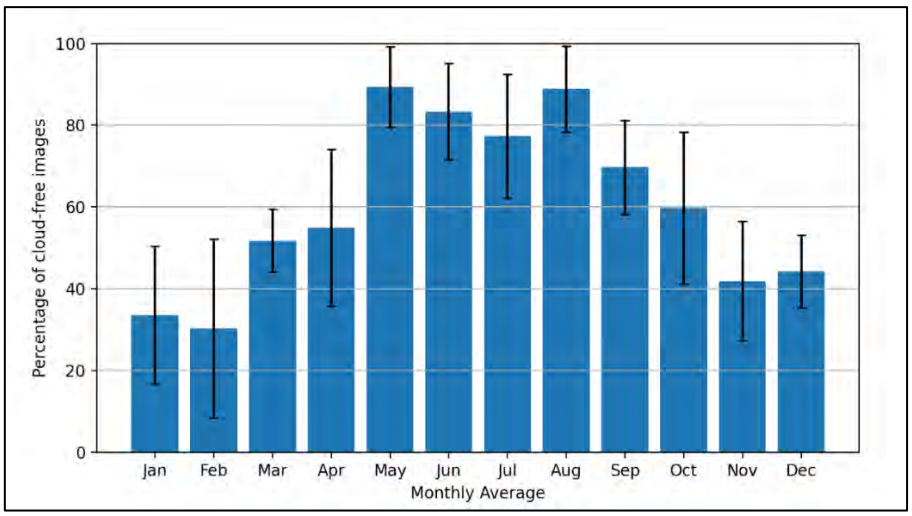


Figure 3.12: Monthly percentage of cloud-free images over Hartbeespoort Dam for the period of August 2015 to March 2021. Error bars were calculated using standard deviation.

3.4.2 Threshold values for the decision tree classifier

3.4.2.1 NDWI

A total of 14 images were sampled with the 20 m grid, totalling 638 140 sample points across all the images. The K-means cluster indentified three distinct clusters in the NDWI spectral data of Hartbeespoort Dam (Figure 3.13). The low NDVI value cluster was identified as water hyacinth and algae, whilst the two higher values represented high turbid/glare water and calm water. Although turbid and calm water were distinct, they were well seperated from water hyacinth and algae with an NDWI threshold of -0.26.

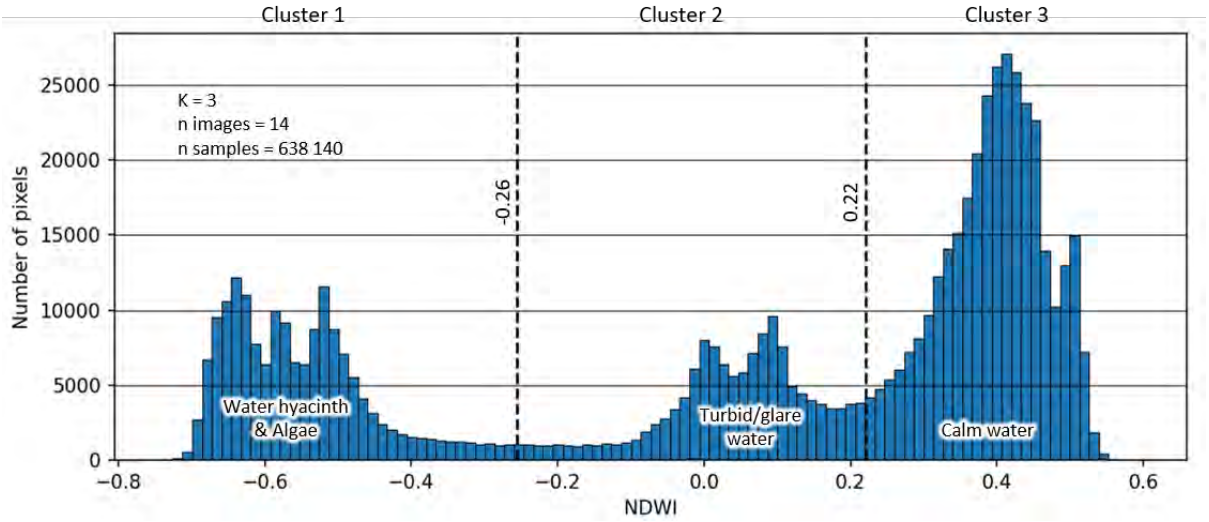


Figure 3.13: K-means clusters of the NDWI values on Hartbeespoort Dam. The boundary between the left and centre clusters (-0.26) represents the threshold for separating water from water hyacinth and algae.

3.4.2.2 ARMI

The number of images in this test was limited to six due to the low frequency of coincidental algal blooms with water hyacinth where both classes were sufficiently abundant. The K-means classifier indentified an ARMI threshold of -0.23 between the two classes. The class separation was less distinct than the NDWI classification, likely due to mixing between water hyacinth and algae.

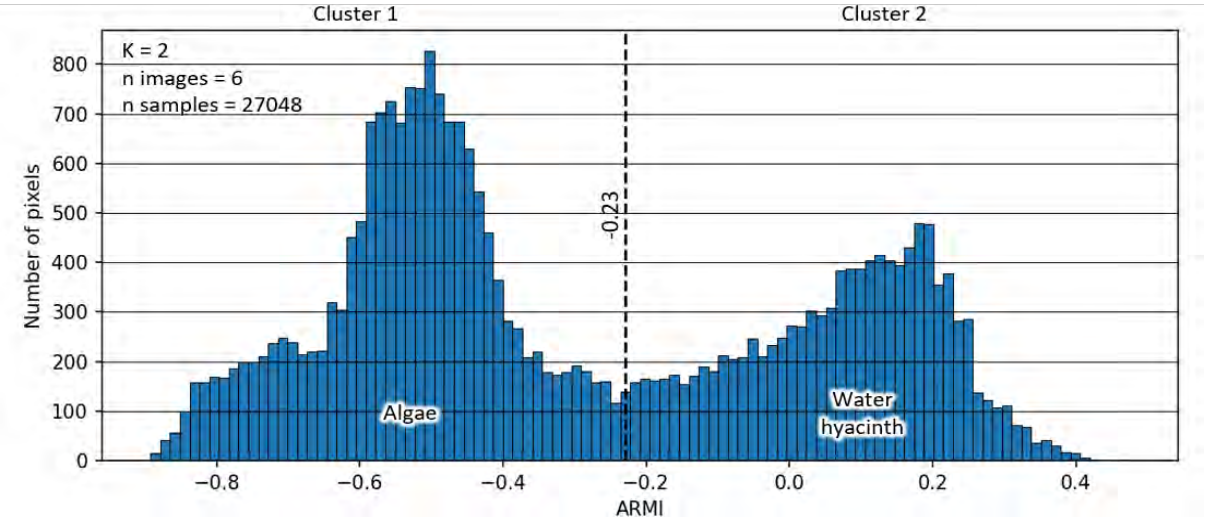


Figure 3.14: K-means clusters of the ARMI index on Hartbeespoort Dam, providing a threshold value of -0.23 between the water hyacinth and algae classes.

3.4.3 Decision tree outputs

Figure 3.15 visualises the indices and classification results of a single Sentinel-2 scene of Hartbeespoort Dam from 16 December 2016. This specific image was chosen to illustrate a concurrent instance of water hyacinth and algae cover classes.

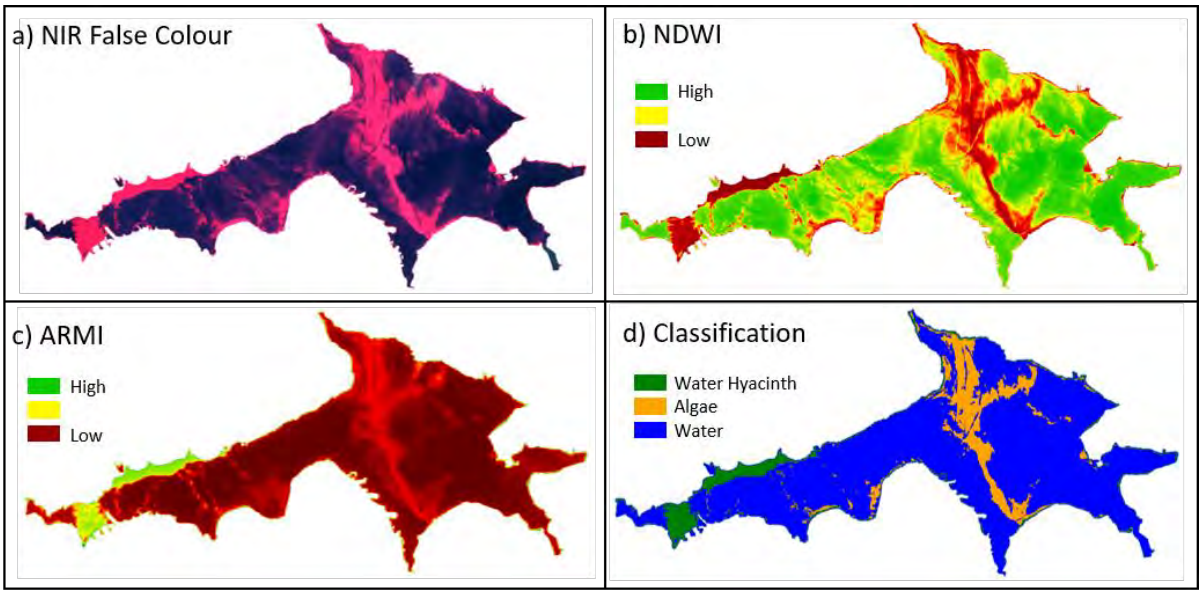


Figure 3.15: a) Near-infrared visualisation of a Sentinel-2 image, 16 December 2016. b) and c) represent the NDWI and ARMI indices of the same image. Note the difference between algae and water hyacinth with the ARMI index. d) Classification output from the decision tree classifier.

NDWI does not distinguish between water hyacinth and algae blooms as it is only intended to detect water pixels, while the ARMI was robust against algae/cyanobacteria by only enhancing the water hyacinth reflectance value.

Water hyacinth and algae were spectrally similar in NIR false colour images but differed in their structure. Water hyacinth mats typically featured well-defined borders, relative to algal mats with

more graduated borders. Algae is more easily navigated and disturbed by boats, which can create visible wake trails in the satellite imagery.

3.4.4 Sentinel-1

3.4.4.1 Image threshold and classification

Figure 3.16 below represents an example Sentinel-1 (VV band, descending) image from the 6th December 2020 (a), its resultant classification (b) and the K-means classification used to determine the threshold between the classes (c). Only descending Sentinel-1 images were classified, as they presented the most consistent classification results over time.

The northern hill slopes that are perpendicular to the satellite have high reflectance values, whereas the low values of the surface water are near black as the flat surface reflects most of the radar waves away from the receiver. As a result, the textured water hyacinth mats are significantly more reflective than the water surface. For consistency, only data from the descending orbit direction were used, which had a revisit frequency of 12 days.

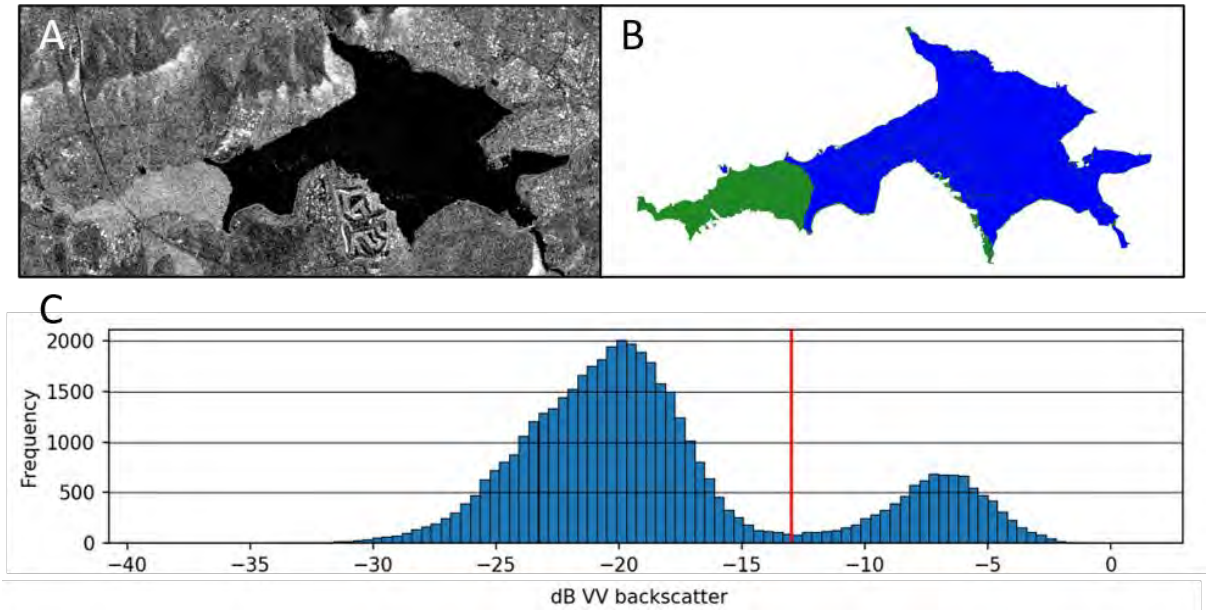


Figure 3.16: (a) Unprocessed Sentinel-1 VV backscatter of Hartbeespoort Dam, 6 December 2020. (b) The same image after masking and classification (green = water hyacinth). (c) Frequency of VV backscatter of the dam, featuring two distinct classes. Pixels values higher than -13 dB VV (red line) are classed as water hyacinth.

3.4.4.2 Agreement between Sentinel-1 and Sentinel-2 results

The results of the linear regression analysis between Sentinel-1 (S1) and Sentinel-2 (S2) indicate that S2 was found to be a statistically significant predictor of S1 ($p < 0.001$), with an R-squared value of 0.98 and a regression equation of $y = -2.6415 + 1.0905x$. Furthermore, the time-series illustrates the general agreement between the datasets over time, even though the exact measurements of similarly timed observations differed due to the spatiotemporal variability of the mats.

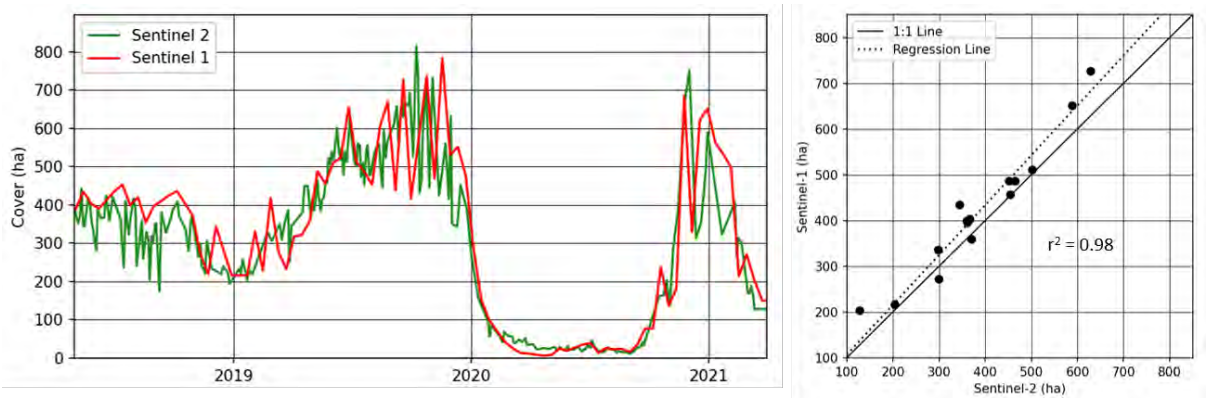


Figure 3.17: A comparison of Sentinel-1 and Sentinel-2 time-series on Hartbeespoort Dam (left). Linear regression comparing same-day Sentinel-2 and Sentinel-1 classifications of water hyacinth (right).

3.4.5 Water hyacinth time-series on Hartbeespoort Dam

The time-series below (Figure 3.18) represents the classification and NDVI results of every qualifying Sentinel-2 image (386) of Hartbeespoort Dam from 1 August 2015 to 31 March 2021. The cover peaked at 817 ha (44.85 %) in August 2017 and again 817 ha (44.71 %) in October 2019. Following the second peak, the water hyacinth declined significantly over the 2019/2020 summer period, where it remained below 40 ha (2 %) for most of that year. It increased again in late 2020, almost returning to original cover levels, but in early 2021 it quickly decreased once again. The mean NDVI of the water hyacinth follows a strong seasonal fluctuation between 0.8 in the summer and 0.55 in the winter. The decline in water hyacinth cover is matched with an unseasonal decrease in mean NDVI over the same summer period.

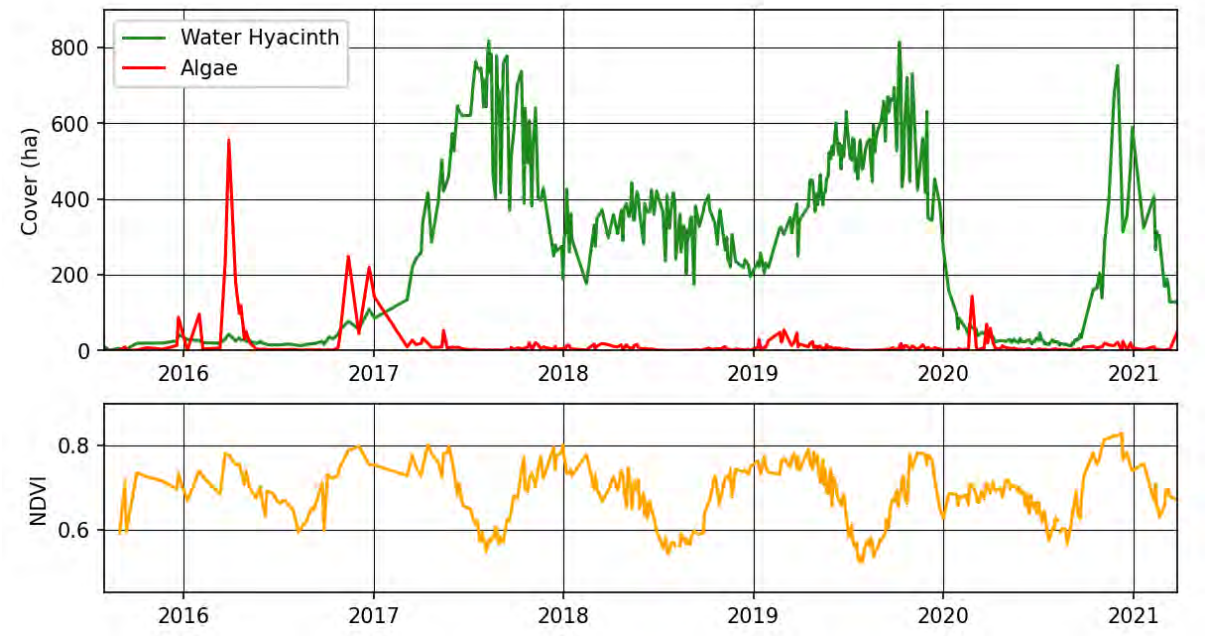


Figure 3.18: A time-series produced by the decision tree classifier of water hyacinth and algae cover on Hartbeespoort Dam from 1 August 2015 to 31 March 2021. The bottom graph represents mean water hyacinth NDVI over the same period.

3.5 Geovisualisation and data accessibility through a WebApp interface

GIS and RS software requires expertise, training and time to use and interpret effectively. It was therefore important to ensure that the monitoring data produced could be presented in a convenient, accessible format to end-users unfamiliar with RS software. There is a growing suite of available geovisualisation tools and libraries that has allowed geospatial scientists to effectively present their data to the public (Balla *et al.*, 2020).

GEE is designed to provide researchers with tools to easily share their results with other researchers, policy makers and the general public (Gorelick *et al.*, 2017). GEE has built-in functionality that allows developers to code a simple user interface to interact with their GEE scripts, backed by GEE's resources, without the need for specialist knowledge or coding ability from the end-user. The WebApp does not require end-users to sign-in and is accessible on any desktop browser. The app building features gives developers options, although sometimes limited, to curate user interaction with the underlying script and present the spatial data in an interactive map window as well as through charts that can be downloaded in CSV format, giving access to raw data that is produced by the GEE script on the back end. This drastically improves data dissemination, by providing end-users with a curated GEE webapp interface that can be easily accessed. This removes time delays that otherwise would occur from data requests that would need to be manually acquired and then shared by an RS user.

With reference to Figure 3.19: A, the "Single Image Analysis" section allows users to select a dam and a Sentinel-2 image for analysis, presented as drop-down menus. The interface automatically filters for images of the selected dam site, with the most recent image of the dam selected by default. Users can then select the "Get Cover Estimate" button, which produces a classified image of water hyacinth on the dam. Numerical data of the classified results are presented in the top centre of the map window (Figure 3.19: B).

In Figure 3.19: C, time-series options are made available for historical analysis. Users can specify their start and end dates for the selected dam, and the output is presented in a responsive graph within the left windowpane. The raw data from the graph can be downloaded in a CSV format, allowing for further data analysis.

Figure 3.19: A screenshot of the WebApp built in GEE. A) Shows the options for a single image analysis. B) The map window that displays the selected image, with a cover estimate in both hectares and a percent. C) The time-series section, where users can request cover and NDVI time-series between specific data.

3.6 Discussion

The aim of this chapter was to develop a robust classification method that could be repeated across a large dataset of Sentinel images to produce a detailed, up-to-date time-series of water hyacinth on Hartbeespoort Dam that could then be readily disseminated to stakeholders. It builds on previous studies that have used index-based thresholding techniques to classify floating macrophytes from a variety of different satellite sources, such as MERIS (Cheruiyot *et al.*, 2014), MODIS (Liang *et al.*, 2017) and Landsat 7 (Oyama *et al.*, 2015) and adapts it to the more recent Sentinel-2 satellite constellation in a cloud-computing environment. Due to the only relatively recent development of cloud-based remote sensing platforms such as GEE, limited research has taken advantage of the new capabilities to process large datasets in this space until recently (Singh *et al.*, 2020). This provided an opportunity to develop a classification algorithm that leveraged the increased temporal and spatial availability of Sentinel datasets with GEE software to produce dense time-series and near-real-time monitoring information of floating macrophytes.

The first objective addressed the case-specific challenges of conducting a time-series classification of water hyacinth on Hartbeespoort Dam. The primary challenge was acquiring accurate cover measurements, due to the high spatio-temporal variation of the floating mats. Wind-induced compression and decompression frequently produced different area and positional measurements between consecutive observations. The solution, made feasible by GEE, was to classify as many images as possible in order to produce a trendline that averaged out the high variations between observations. To ensure the highest number of quality images were available, a method was developed to filter out images within the dataset that contained any cloud cover within boundary of the dam. This was important as classification results depended on reporting the total extent of the water hyacinth presence on the dam, which meant even minor cloud presence had the potential to skew the classification result. Additionally, emphasis was placed on reducing false negatives – with generalised, coarse cloud presence layers, images that were cloud-free over the dam might still be miscategorised due to nearby cloud presence. Incorporation a precise cloud filtering method ensured the maximum number of suitable images for analysis were included in the time-series.

The occasional presence of cyanobacterial blooms on the dam also needed to be addressed, as they are spectrally similar to macrophytes in several band wavelengths. This meant that the NDWI threshold method for classifying water hyacinth also falsely detected cyanobacteria blooms. Given the index-based structure of the decision tree classifier, a suitable index that was sensitive to one class, but not the other, was required. Several established, off-the-shelve indices were tested, but none performed notably well for this specific task. Oyama *et al.* (2015), used a specific index to separate macrophytes and cyanobacteria, the floating algal index (FAI), but used spectrally different bands to those available with Sentinel-2. This led to the development of a new index, ARMI, which leveraged the inverse spectral relationship of macrophytes and cyanobacteria observed in the shortwave infrared (SWIR1) and green bands.

The second objective used the challenges and opportunities identified above to produce a classification algorithm that could efficiently and accurately process as many suitable images as possible over multi-year time periods. Previous studies on the remote sensing of water hyacinth have successfully used a range of classification methods, as highlighted in Chapter 2. Preliminary exercises during the investigational stage of this thesis tested different classification models, such as the supervised Classification And Regression Tree (CART) machine learning model. CART is readily

available in GEE and therefore is a popular choice for image classification on the platform. It produced comparable results to the index threshold decision tree method presented here, but at a higher computational cost. The empirically determined classification thresholds of the NDWI and ARMI indices with k-means clustering had an additional benefit of providing explicit insight in the classification process, eliminating the 'black box' nature of many statistical classifier methods.

Given the relatively simple and fixed structure of the decision tree classifier, several key assumptions were made. Firstly, that water hyacinth was the only floating macrophyte species present on the dam over the study period from August 2015 to March 2021. This was a valid assumption during this study, as on-site monitoring confirmed that water hyacinth was the dominant species over this period. However, since the middle of 2021, *Salvinia minima* began to rapidly increase and is now the dominant macrophyte species on the dam. While the classifier still detects this new species, it does not distinguish between them. Secondly, it assumed that the water boundary was relatively stable as a fixed boundary mask was used over the entire period. Again, this was observed to a valid assumption in this case, although other sites and time periods may need a more dynamic masking approach. Thirdly, that the water hyacinth mats were free floating as open water was required to be observed at least once during the period to be included in the boundary mask. This was intentional, as it excluded emergent and other spatially permanent macrophyte species on the dam, thus reducing false positives in the classification outputs.

All of these assumptions are justified in the context of this study on Hartbeespoort Dam, but preliminary application of the classifier to other water hyacinth sites in South Africa highlighted potential challenges for scalability and application beyond its intended context of Hartbeespoort Dam. For example, it was found that the algorithm underperformed on sites where water levels were highly variable, and therefore accurate and stable boundary mask could not be determined.

More generally, as with remote sensing more broadly, there were limitations with the scale of satellite data used in this study. Due to the pixel resolution (10 m) of the Sentinel-2 images, the presence of small or sparse macrophyte mats could not be reliably detected due to minimal detection unit and edge-pixel effects. However, given the intended use-case, which was to map and monitor the overall population trends, the impact of this limitation was reduced by maximising the number of observations over time.

Conducting an accuracy assessment on these results proved to be a challenge. Due to the high spatio-temporal variability of the water hyacinth mats, combined with the limited accessibility, a traditional ground-truthing exercise of the accuracy of the classification method was not feasible. To ensure accuracy and robustness of the algorithm developed in this chapter, a significant part of the research focused on developing and conducting a suitable accuracy assessment method, which has its own dedicated chapter (Chapter 4).

The third objective was to make these data accessible to a wider audience by creating a WebApp interface that both visualises the data and provides access to download the time-series data for further analysis. Google Earth Engine has created utilities to facilitate this, by providing tools that enable researchers to build interactive UIs that interface with their scripted algorithms (Gorelick *et al.*, 2017). This is particularly useful for RS researchers, as it reduces the requisite expertise typically required for bottom-up application development and web programming (Gorelick *et al.*, 2017).

The WebApp was published in 2021 (titled as 'Macrophyte Monitoring Tool') and has become an important part of the monitoring strategy for CBC at Hartbeespoort Dam. Further, it currently includes nine other waterbodies of interest in South Africa that meet the algorithmic assumptions of the classifier outlined above. Additional sites are readily added by request, should they meet these requirements.

This WebApp forms part of a growing suite of apps produced by researchers in GEE, covering a wide variety of applications such as urban mapping, crop yield estimation, flood mapping and malaria risk mapping (Gorelick *et al.*, 2017). Google Earth Engine has also been integrated with third-party web applications, such as the popular Global Forest Watch (www.globalforestwatch.org) which provides an interactive map of the annualised global forest change dataset produced by Hansen *et al.* (2013).

There is scope for improvement of the current WebApp solution, but it will likely require a new software solution due to the limited functionality available within the GEE UI. For example, there is a wait time of minutes for long time-series requests as GEE classifies the entire dataset for each request. Periodical (e.g. daily) pre-processing and storing of the data would significantly reduce query times, along with other features such as a presence alert/notification function would be useful additions. Nevertheless, in the context of limited alternatives, the current WebApp solution is more than sufficient in the interim. The WebApp is the subject to on-going development and improvement and access is available on request through the author.

Overall, the results from this study demonstrated the viable application of remote sensing, by leveraging frequent, freely available satellite data with the cloud-computing capability of GEE, to measure and monitor water hyacinth on Hartbeespoort Dam. The classification algorithm presented here proved to be a relatively simple and robust means to classify water hyacinth mats that has the potential to be scaled beyond Hartbeespoort Dam. However, due to the assumptions that are made in this classification algorithm, further work would be required to successfully scale this method to a more diverse range of waterbodies going forward.

Chapter 4: An accuracy assessment of the remote sensing methods

4.1 Introduction

Accuracy assessments are a cornerstone of the remote sensing process. Remotely acquired information contains an inherent level of generalisation as a function of variables such as image resolution, atmospheric effects and ambiguous spectral signatures (Steven, 1987). As such, sources of error can be introduced into the workflow during image preprocessing, image interpretation and classification (Story & Congalton, 1986).

The most common method to measure accuracy in RS is to compare the classified image with reference data, expressed as a confusion matrix (Ma & Redmond, 1995; Lewis & Brown, 2001). Reference, or 'ground truth', data points are typically collected either directly in the field or from high resolution aerial photographs that provide surface context to the synoptic overview provided by satellites (Steven, 1987). A confusion matrix is a 2-dimensional tally that compares categorical reference and predicted data, with the major diagonal indicating the level of agreement between the two datasets (Story & Congalton, 1986).

The confusion matrix method assumes that the geographic positions of the predicted and reference data points are consistent. However, acquiring location-based reference data of floating macrophytes on Hartbeespoort Dam was particularly challenging for two reasons: a) direct field observations have accessibility constraints in an aquatic environment, and b) high spatio-temporal variation of the floating mats in large waterbodies. High spatio-temporal variation causes changes in both mat position and area as it expands and contracts in response to environmental conditions, primarily from wind in the case of Hartbeespoort Dam. Therefore, matching high-resolution images with the Sentinel-2 images with the required level of spatial consistency for a confusion matrix was unfeasible.

Additionally, due to the use of categorical data, confusion matrices are well-suited to comparing the accuracy of landcover (mis)classifications between classes. In the case of detecting macrophytes mats from water, categorical inaccuracy was not the primary source of error given the high spectral distinction between these two classes. Instead, accurate boundary delineation was the key concern due to a relatively high edge-to-area ratio of the floating mats which produced mixed pixel boundaries of the 10 metre Sentinel-2 pixels.

To overcome these challenges, an approach which compared the total classified area values of a Sentinel-2 scene to a higher resolution data source was used. High-resolution PlanetScope satellite images were used as the reference data source as the satellites have near-daily revisit times with a high pixel resolution of 3 metres, more than an order of magnitude higher than Sentinel-2's 10 metre resolution. Despite the precedence for higher resolution satellite imagery to assess water hyacinth remote sensing accuracy via a typical confusion matrix (e.g. Cheruiyot *et al*., 2014; Dong *et al.*, 2020), there is a lack of research that conducts an accuracy assessment that uses the total classified area values for its accuracy metric.

It was uncertain whether coarser resolution images would under-predict macrophyte presence due to minimum detectable units and edge-pixel effects, potentially producing a non-linear result from different resolution images from the same observation date. Whilst research has been conducted on resolution compatibility in remote sensing (e.g. Claverie *et al.,* 2018 on combining Landsat 8 and

Sentinel-2 images), the niche characteristics of floating macrophytes warranted its own brief investigation.

4.2 Aim and objectives

The aim of this chapter was to develop and conduct a suitable accuracy assessment of the remote sensing classification results produced in Chapter 3.

The first objective was to determine whether water hyacinth classification areas could be suitably compared across different image resolutions.

The second objective was to identify a suitable set of PlanetScope images that were coincident with Sentinel-2 image acquisition dates (referred to as 'image pairs') and to measure the extent of water hyacinth within those images.

The third objective was to statistically compare the reference data produced in objective 2, referred to as 'observed' data, to the classified Sentinel-2 data, referred to as 'predicted' data, in order to produce an error metric for the remote sensing results of Chapter 3.

4.3 Materials and Methods

4.3.1 Testing for resolution compatibility

To address the first objective, three atmospherically corrected (preprocessed by Planet.com) PlanetScope images of Hartbeespoort Dam were selected for this exercise. Each image met the selection criteria of comprising a well-represented mixture of dense and sparse macrophyte mats and of varying sizes.

4.3.1.1 Classification preparation

To prepare them for classification, the images were masked to a fixed Hartbeespoort Dam boundary. The images were converted to NDVI to better distinguish macrophytes from water (Figure 4.1: 3) and resampled to four different resolutions (Figure 4.1: 4) with a cell area weighted mean method (as per Claverie *et al*., 2018). The resampled images represented the resolutions of Sentinel-2 MSI (10 and 20 m), Landsat 8 OLI multispectral (30 m) and pansharpened (15 m) bands (Figure 4.1: 5). Although Landsat 8 data were not used in this thesis, its band resolutions were included here to evaluate its feasibility for potential future inclusion.

All the images were classified with an Iso Cluster Unsupervised (n clusters = 2) method (Figure 4.1: 6) in ArcMap 10.8.1 (Environmental Systems Research Institute (ESRI), 2011). Close visual inspection was used to validate the classification accuracy.

4.3.1.2 Extraction of macrophyte cover

The macrophyte cover was extracted and converted to a cover percentage by dividing the number of water hyacinth raster cells with the total number of raster cells in the image. This produced a relative metric that was independent of the raster cell size (Table 4.1). The average difference between the resampled and reference classifications was calculated in order to indicate compatibility.

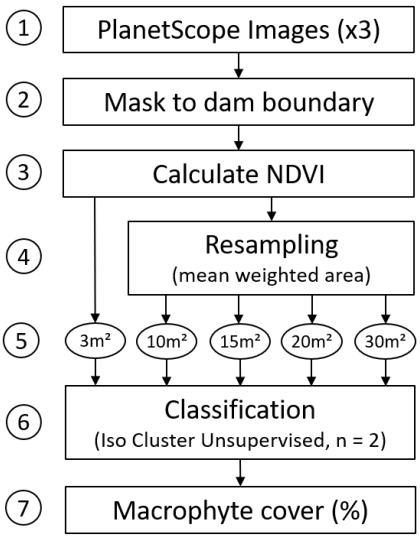


Figure 4.1: Workflow for testing scale transferability of water hyacinth images

4.3.2 PlanetScope image selection and classification

The second objective was to identify coincident observations of Sentinel-2 and PlanetScope events in order to compare their classified cover metrics directly. Due to the high spatiotemporal variability of the floating macrophyte mats, the acquisition date of the PlanetScope images needed to match the time and date of the Sentinel-2 images as closely as possible. All PlanetScope images were required to have full dam coverage, either as a single tile or as a composite image and be fully cloud-free. The variability of PlanetScope image quality necessitated careful inspection to ensure consistency and completeness of all the images used.

PlanetScope images were used due to their relatively high resolution of 3 metres and near-daily revisit times which provided multiple instances of coincident image acquisition days with Sentinel-2. Orthorectified, atmospherically corrected images were downloaded from Planet.com under an academic license.

Each PlanetScope image was masked to the dam boundary, effectively leaving two remaining cover classes (macrophyte mats and water) as no other significant landcover classes were present on the dams during this period. In limited instances where algal blooms were present, the images were disqualified.

NDVI was calculated for each image as:

$$NDVI = \frac{NIR - Red}{NIR + Red} = \frac{Band\ 4 - Band\ 3}{Band\ 4 + Band\ 3}$$

This provided a single band that distinguished the macrophytes (high NDVI) from water (low NDVI). NDVI was chosen, as opposed to shortwave infrared-based indices used with Sentinel-2 in Chapter 3, due to the availability of only four multispectral bands (blue, green, red and near infrared) in the PlanetScope images.

The NDVI images were classified with an Iso Cluster unsupervised method (n clusters = 2) in ArcMap 10.8.1. The Iso Cluster classification was an accessible, repeatable classification method that avoided

boundary generalisation errors that were encountered during efforts to manually classify the detailed reference images. Each classification result underwent detailed visual inspection for quality control.

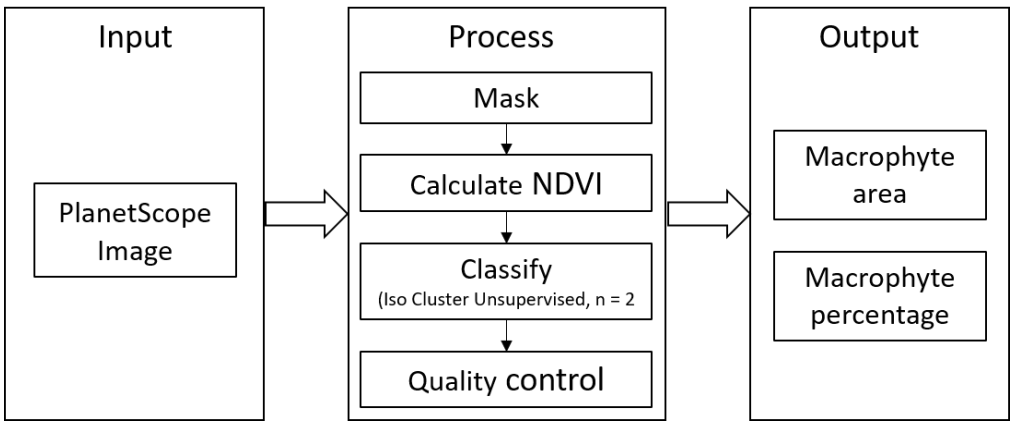


Figure 4.2: Outline of the image classification workflow for the PlanetScope images.

4.3.3 Statistical analysis

The final objective was to conduct an out-of-sample statistical analysis to compare the classification results between each Sentinel-2/PlanetScope image pair. Out-of-sample tests, which compare the results of the classifier to new data (i.e. classified images from a different satellite source) are generally considered more robust compared to in-sample tests. All statistical functions were performed in the Python programming language (Python Software Foundation. Python Language Reference, version 3.9. Available at http://www.python.org) with the 'StatsModels' package (Seabold & Perktold, 2010). These methods follow a 2-dimensional predicted vs. observed model performance validation technique that, while unusual in the remote sensing field, is commonly used throughout the natural science literature (Willmott & Matsuura, 2005; Pineiro *et al.*, 2008).

The results of the 'observed' PlanetScope classifications were compared against their corresponding 'predicted' Sentinel-2 classifications with a linear regression. Linear regressions are a popular method used in statistics to evaluate models, whereby the values and slope of the 'predicted' vs. 'observed' are compared against a 1:1 line (Pineiro *et al.*, 2008). Two forms of the data were analysed concurrently: the absolute values (in hectares) and relative values (as a percentage of the overall dam area). The latter provided an area-independent error metric that was more intuitive to understand and interpret.

The agreement between the predicted and observed datasets was evaluated with a determination coefficient (R-squared) and the errors quantified with root mean square error (RMSE), mean absolute error (MAE) and mean bias error (MBE). R-squared is a common method to measure the extent to which the independent variables can explain the dependent variables (Filgueiras *et al.,* 2020). RMSE represents the standard deviation between predicted and observed data by comparing the quadratic difference between variables. As such, RMSE can be sensitive to outliers, making it a misunderstood error metric despite wide application in the natural sciences (Willmott & Matsuura, 2005). Instead, an MAE metric was better suited in this case, as it unambiguously provides a simple mean error between the variables (Willmott & Matsuura, 2005). Nevertheless, both measures were used for comparison. Additionally, MBE was included as it indicates the overall direction of the prediction error.

4.4 Results

4.4.1 Testing for resolution compatibility

The three PlanetScope images had similar macrophyte cover classifications at all five resolutions with an overall mean difference of 0.16 % between the reference and resampled images. Table 4.1 presents the full results for each image in both pixel count for each resolution and percentage. Figure 4.3 provides an illustrated example of the different resolutions. As indicated by this near-linear relationship, it was determined that macrophyte classifications derived from images of different resolutions (i.e. Sentinel-2 and PlanetScope) could be directly compared. These results supported the use of high-resolution imagery as a means for accuracy assessment in this thesis.

Table 4.1: Results from the comparison of the Iso Cluster classified PlanetScope images at different resolutions from Chapter 4

Image Resolution (metres)	Macrophyte Pixels	Water Pixels	Macrophyte Cover
Image: 20190904_074408_0e26_3B_AnalyticMS_SR			
3 (native)	825134	1238967	39.98%
10	74164	111613	39.92%
15	32846	49759	39.76%
20	18650	27814	40.14%
30	8428	12225	40.81%
Image: 20190813_081120_87_1064_3B_AnalyticMS_SR			
3 (native)	752477	1311624	36.46%
10	68662	117115	36.96%
15	30375	52230	36.77%
20	17159	29305	36.93%
30	7520	13133	36.41%
Image: 20191103_075713_0f34_3B_AnalyticMS_SR			
3 (native)	857434	1206667	41.54%
10	78137	107640	42.06%
15	34325	48280	41.55%
20	18875	27589	40.62%
30	8654	11999	41.90%

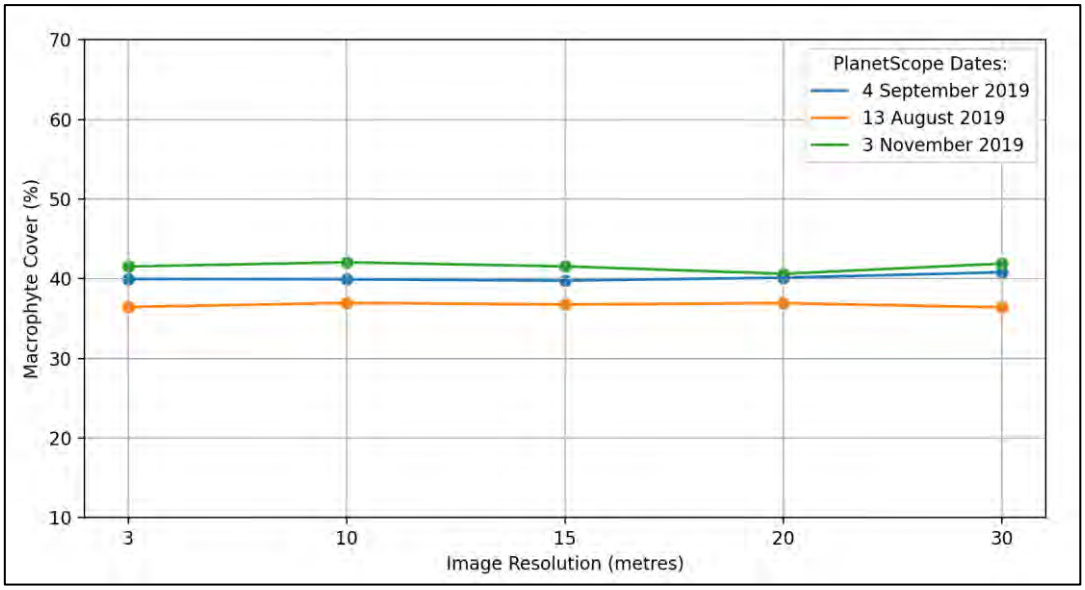


Figure 4.3: Classified macrophyte cover for three images at 3, 10, 15, 20 and 30 m resolutions

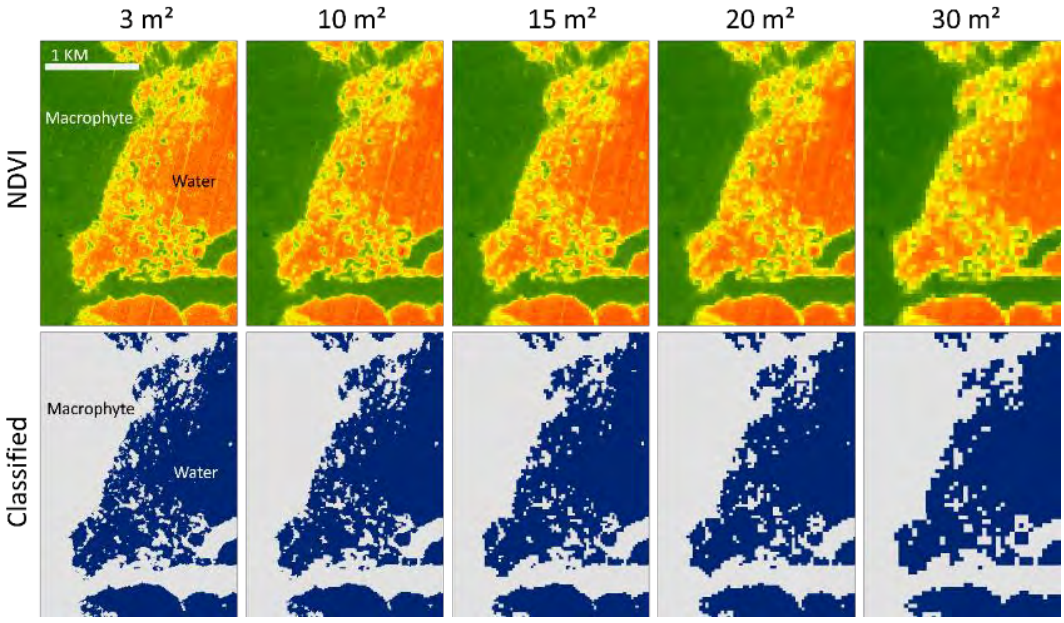


Figure 4.4: An example of different resolutions of a PlanetScope image (cropped) to test classification consistency between macrophytes (water hyacinth) and water on Hartbeespoort Dam

4.4.2 Image pairs and classification

Fifteen images from May 2018 to December 2020 were selected based on the criteria outlined in the methods section, with an effort to ensure year-round representation. High mat drift, incomplete coverage and cloud cover limited the final number of images pairs that were suitable for comparison. Table 4.2 lists the 15 qualifying coincidental Sentinel-2 and PlanetScope observation days. Figure 4.5 illustrates one of these days, 2 December 2020, as an example. In this example, the GEE classification of the Sentinel-2 'predicted' 799 hectares, whilst the PlanetScope classification 'observed' 770 hectares.

Table 4.2: Classification results of all the image pairs with the corresponding observation dates.

Image Date	Observed Area (ha)	Predicted Area (ha)	Observed (%)	Predicted (%)	Difference (Observed - Predicted)
20 May 2018	261.422	265.844	14.0625	14.3	-0.24
24 June 2018	423.342	362.505	22.7726	19.5	3.27
11 July 2018	416.777	358.799	22.4219	19.3	3.12
24 July 2018	375.293	325.333	20.1988	17.5	2.69
30 August 2018	348.644	301.166	18.7544	16.2	2.55
01 July 2019	609.744	591.162	32.8799	31.8	1
24 July 2019	476.177	533.533	25.6142	28.7	-3.09
31 July 2019	537.355	483.34	28.9105	26	2.91
13 August 2019	6477	540.977	34.8034	29.1	5.7
04 September 2019	673.199	606.034	36.2123	32.6	3.61
09 September 2019	683.122	648.791	36.7547	34.9	1.85
14 October 2019	449.399	436.877	24.1735	23.5	0.67
18 October 2020	220.314	167.31	11.8512	9	2.85
25 November 2020	651.971	669.24	35.0711	36	-0.93
02 December 2020	769.591	799.37	41.4398	43	-1.6
				Mean	0.84
				Std Dev	2.84

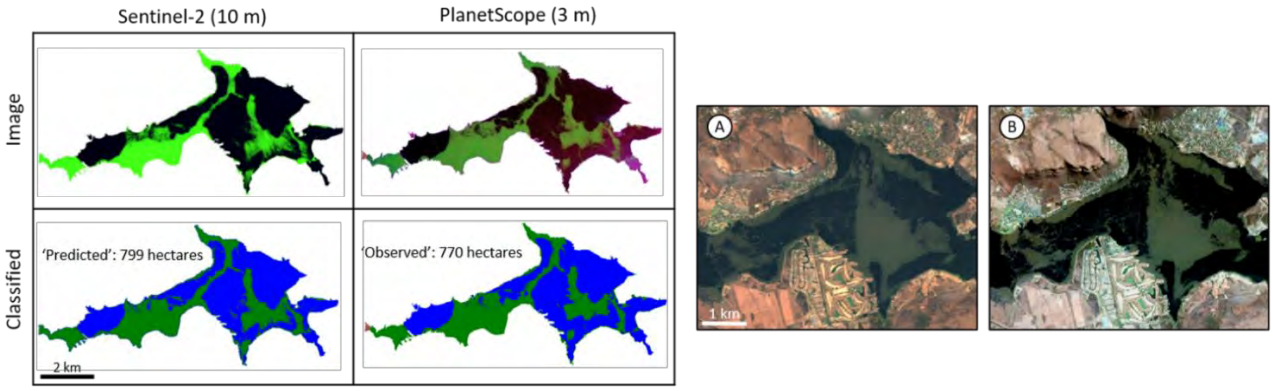


Figure 4.5: Top: an example of an image pair and the corresponding classification results from 2 December 2020. Bottom: an image pair from 30 August 2019, captured 11 minutes apart, illustrating noticeable mat drift between the Sentinel-2 image at 09:46 (a) and the PlanetScope image at 09:57 (b).

4.4.3 Statistical comparison between observed and predicted classifications

The results of the linear regression analysis between the predicted Sentinel-2 dataset and the observed PlanetScope dataset revealed a strong, linear correlation ($p < 0.001$) with an R-squared value of 0.941 and a regression equation of $y = -43.7653 + 1.027x$.

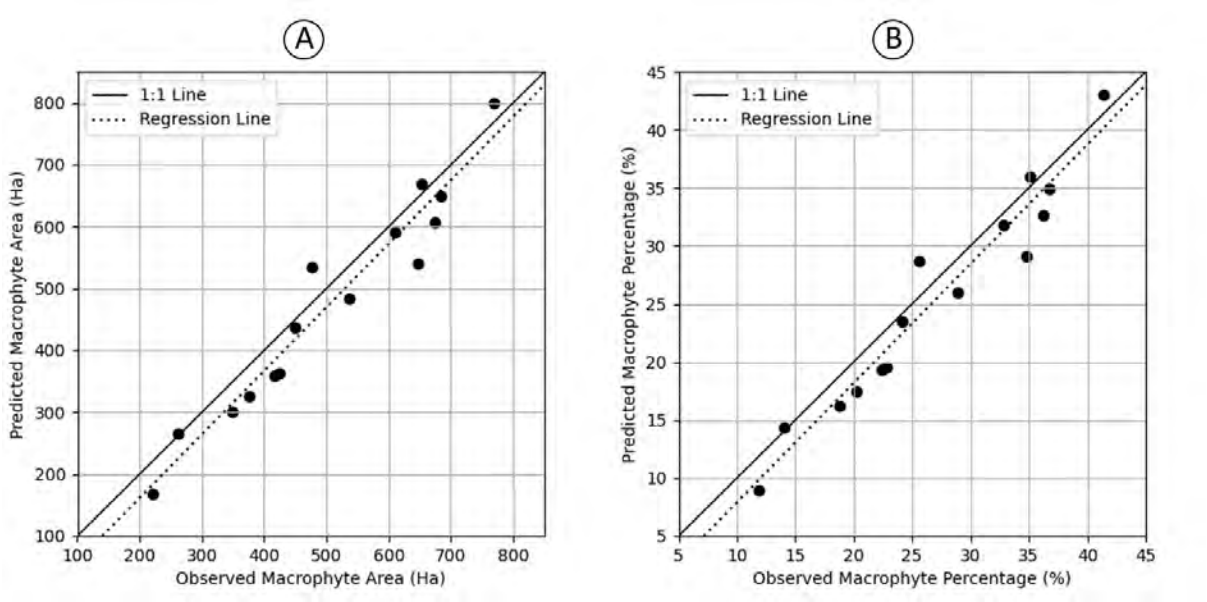


Figure 4.6: Linear regression of predicted vs. observed values expressed as both area (a) and percentage (b) forms of the data. Both regressions have an R-squared of 0.941. As illustrated by the regression line, the predictions tended to slightly under-estimate the observations, which is further supported by an MBE result of -30.2 ha (-1.63 %).

The RMSE showed a difference of 51.295 ha (2.76 %) between predicted and observed data, however with the MAE, this number reduced to 44.714 ha (2.41 %). An MBE of -30.204 ha (-1.625 %) suggests that the predicted data, on average, tend to slightly underpredict the observed values.

Table 4.3: Evaluation metrics of the regression model

	R-squared	RMSE	MAE	MBE
Area (ha)	0.941	51.295	44.714	-30.204
Percentage (%)	0.941	2.759	2.405	-1.625

4.5 Discussion

This chapter aimed to conduct an accuracy assessment on the classification results from Chapter 3 by comparing the total classified macrophyte areas with higher resolution satellite imagery (PlanetScope) in order to address the issue of constant mat drift that rendered a typical position-based error matrix method unfeasible. This included testing whether different image resolution scales could be directly comparable, as there was concern that the complex and detailed mat boundaries would render different area results as resolution was decreased. Once it was determined that the area classification remained relatively stable across different resolutions, PlanetScope images were used as the reference data for the accuracy assessment.

A high R-squared of 0.941 and a mean absolute error of 44.7 ha (2.4 %) between predicted and observed datasets suggests a high level of accuracy of the Sentinel-2 classification method used in Chapter 3. There are two likely factors contributing to this high level of accuracy. Firstly, the 'observed' data from the PlanetScope classifications are themselves predicted values, and therefore are not a complete representation of the true ground values. This may have introduced a bias in the reference data, falsely increasing the agreement between the data. However, their use as observed values are justified by a) their significantly higher ground resolution to that of the Sentinel-2 data (9 m^2 and 100 m^2, respectively), b) the use of many image pairs to overcome potential classification anomalies introduced by varying mat densities, and c) careful selection of the PlanetScope images and close visual examination of the classified image outputs. The initial plan was to manually trace the boundaries of the water hyacinth mats, but due to the complexity and precision needed to consistently identify the boundaries in multiple images, this approach was found to be infeasible.

Secondly, the dam boundary masking step in the Sentinel-2 classification algorithm resulted in a straightforward discrimination process between the two remaining landcover classes. With the exception of occasional algal blooms, no other spectrally similar landcover class was present during the study period, eliminating the risk of misclassification errors. This is contrasted to other remote sensing studies of water hyacinth that contend with multiple weed species (e.g. Cavalli *et al.*, 2009; Albright *et al.*, 2004) and additional terrestrial cover classes (e.g. Venugopal, 1998; Thamaga & Dube, 2018a).

Continuous mat drift may have produced an underestimation between the true correlation between the datasets. The constant wind-induced movement causes the mats to expand and contract, changing their overall coverage on the water surface over a short period of time. Although all the PlanetScope/Sentinel-2 image pairs were captured within the same day, and typically within the same hour, there were varying discrepancies between each image pair that was apparent upon visual inspection (Figure 4.5). This will have resulted in a slight difference in classification outputs of the image pairs.

Other researchers, such as Keebine (2019), have also noted the challenge of conducting accuracy assessments of water hyacinth classification on Hartbeespoort Dam due the high spatio-temporal variability. However, solutions have been left largely unaddressed in existing literature, as either the problem has not been recognised or, in many cases, the mats in smaller water bodies and river systems tend to remain more stable, permitting the use of more traditional accuracy assessment methods. As such, these accuracy metrics cannot be directly compared to other similar remote sensing studies of

water hyacinth in South Africa, such as that by Thamaga & Dube (2018a), due to the use of fundamentally different accuracy assessment metrics used herein.

A limitation of this study was the resolution of the ground-truth images. Although the PlanetScope images have a considerably higher resolution than Sentinel-2, an abundant presence of water hyacinth mats smaller than Planet Scope's 9 m^2 ground resolution, which go undetected, may cumulatively skew the overall cover area value. Higher resolution images would have remedied this problem. However, there are some practical and economical obstacles to obtaining significantly higher resolution imagery. Openly available Google Earth imagery was too infrequent, and other high resolution (< 0.5 m) commercial satellite images are costly and have long revisit times resulting in infrequent coincident observation days with Sentinel-2 images.

An alternative to high resolution satellite data are images collected by Unmanned Aerial Vehicles (UAVs). However, survey quotes obtained during this study ranged between ZAR 6000 and ZAR 45000 (USD 397 and USD 2977, respectively) depending on the service provider and sensors used. UAV surveys are typically difficult and expensive to conduct in South Africa due to strict regulations by the Civil Aviation Authority. The available quadcopter UAVs were reportedly limited to approximately 30 hectares per battery, making full coverage of the 1821-hectare dam unfeasible. Fixed-wing, long range drones are better suited for this scale of survey, however mat drift between adjacent passes will still pose challenges on windy days. Further, the survey would need to be timed with a Sentinel-2 overpass on a cloud-free day, and a single survey would only represent one Sentinel-2/PlanetScope image pair. Alternatively, a smaller water hyacinth-infested study site with better UAV mapping potential could stand in for an accuracy assessment exercise, but this would exclude the high spatio-temporal variability component that is a significant variable on larger water bodies.

There were a relatively low number of suitable image-pairs that were available for analysis on Hartbeespoort Dam, which meant each individual image carried a large weight and the potential to skew the overall results of the statistics. More images from different satellite sources and study sites would have helped improve this. Although not included in this thesis, ancillary research used additional water hyacinth cover data on Roodeplaat Dam which increased the accuracy result further.

In conclusion, this chapter aimed to conduct an alternative method to quantifying the error and uncertainty of the remote sensing classification data. Despite the limitations, it demonstrates the opportunities of high-resolution image data as a valuable source of reference data for classification validation and presents a potential solution for other studies that are presented with similar challenges.

Chapter 5: Synthesis

5.1 Introduction

This research set out to address a problem raised by the researchers at CBC: the need to monitor water hyacinth, a dynamic natural feature, at a population-wide scale on Hartbeespoort Dam. More broadly, there is a monitoring deficit of aquatic weeds in South Africa (Thamaga & Dube, 2018a; van Wilgen *et al.*, 2022) and this thesis aimed to use the Hartbeespoort Dam case study to illustrate the potential for remote sensing technology to fill this gap, particularly as the field has rapidly advanced in recent years due to increased satellite availability and powerful software.

Water hyacinth is a significant invasive aquatic weed both globally and in South Africa. Although considerable biological control success has been achieved with many other important invasive macrophytes, the plant remains a problematic weed in many aquatic systems in the country. This is due in large part to the eutrophic conditions which promote high growth rates, and its greater ability to withstand cooler climatic conditions than most of the plant's existing biological control agents (Coetzee & Hill, 2012). Excessive herbicidal applications further hamper the effectiveness of the biological control programme, attenuating the agents' ability to develop the requisite population densities for effective control.

In response to the hypertrophic conditions, Hartbeespoort Dam has been infamously infested with water hyacinth for decades (Ashton *et al*., 1979). This reduces the important socioeconomic utility of the dam and disrupts natural ecological processes in the system. Along with repeated physical and herbicidal control efforts, past attempts of classical biological control from a suite of agents had failed to produce complete control. However, the release of the new biological control agent, *Megamelus scutellaris*, has shown promising results on Hartbeespoort Dam in recent years. Through an inundative introduction strategy starting in late 2018 by the CBC, in collaboration with local stakeholders, considerable reduction of the water hyacinth cover on the dam was observed. This marked the first case of successful biological control in such a system in South Africa (Coetzee *et al.*, 2022).

Monitoring these population-scale events is in an integral part of a successful control programme, providing important feedback on system responses to management actions. However, given the scale of the dam, and the rapid spatiotemporal fluctuations of the water hyacinth mats, these changes were difficult to measure and quantify in the field. This necessitated the need for a quantitative monitoring solution that was accurate, timely and readily accessible. This thesis proposed a remote sensing solution to address this need. Technological advances in the RS field, particularly the increased availability of moderate-to-high resolution satellite data and cloud-computing resources, have produced a paradigm shift that expands the application of RS from a relatively coarse change detection tool, into a solution that can provide the detailed monitoring data that was required at Hartbeespoort Dam.

Research around the application of using new RS technologies such as GEE have grown in recent years, indicating a recognition of the application potential and value it provides in the RS sector (Zhao *et al.*, 2021). In the literature, there has been a steadily growing body of literature focused on mapping water hyacinth and other macrophytes in recent years (Thamaga & Dube, 2018b). However, technologies such as GEE have remained relatively under-utilised as a monitoring solution in for invasive

macrophytes (Singh et al, 2020), which currently suffer from a lack of on-going monitoring solutions (Thamaga & Dube, 2018a; van Wilgen *et al.*, 2022). Therefore, this research gap presented an opportunity to build on the existing work and adapt it to new satellites and software to produce a robust, dense time-series dataset of water hyacinth on Hartbeespoort Dam.

High spatio-temporal variability of the water hyacinth necessitated the need for unique solutions to address particular challenges. For example, the limited physical accessibility of the floating mats, combined with the near constant wind-drift, meant a typical ground-truthing method for assessing the accuracy of the data was unfeasible. Therefore, this thesis tested and conducted an alternative accuracy assessment method that gauged the classification accuracy of water hyacinth on the dam by comparing the total classified area to a higher resolution reference image, rather than on geographic position.

5.2 Opportunities, limitations and recommendations

While this thesis focused solely on water hyacinth at Hartbeespoort Dam, there is scope to scale these methods to other nearby sites and beyond. The scalability of Google Earth Engine presents an opportunity for existing algorithms to be transferred to other areas with minimal additional adjustment. Therefore, the insights gained in this thesis have the potential to be scaled significantly. Although not a topic of this thesis, the WebApp currently has been scaled to include several other nearby water bodies invaded by water hyacinth, such a Roodeplaat Dam, northeast of Pretoria, South Africa and provides ongoing information about the status of macrophytes on these sites.

Further abroad, the methods conducted in this thesis have been adapted by the author to monitor the status of the *Salvinia molesta* on Lake Ossa, Cameroon, after the introduction of the salvinia weevil (*Cyrtobagous salviniae*) which preliminary data shows promising biological control at the time of writing. Lake Ossa is a large lake complex with many areas only accessible by boat, which provided an ideal opportunity augment RS data into the monitoring strategy.

Besides providing valuable feedback on the status of on-going biological control interventions, the dense time-series information could be used to measure the recovery rate of different removal strategies (e.g. mechanical, herbicide) to calculate a monetary value for managers to assess the cost and benefits of different management approaches.

Due to the remote and near-real-time nature of RS monitoring, it has the potential to provide timely updates for rapid response to new macrophyte invasions. For example, García-de-Lomas *et al.*, (2022) conducted rapid mechanical removal of water hyacinth from Guadalquivir river branch in Seville, Spain, which significantly reduced the cost that would have been incurred had the same intervention been delayed to later in the growing season. However, a significant portion (83 %) of the €22 500 budget were spent on post-removal boat monitoring and follow-up treatments. In-field monitoring was further complicated by mass colonization of wasp nests along the shores and the presence of West Nile disease which limited the feasibility for monitoring exercises to take place. In projects such as this, while still effective, the addition of up-to-date RS data could be a useful tool to target in-field monitoring efforts and reduce the required monitoring time spent in the field.

As covered in the discussion of Chapter 3, there are several limitations of the current algorithm that require further work to improve its application over a more diverse set of aquatic systems. Current

assumptions, such as a relatively constant water level, a low boundary-to-area ratio and single-species occurrence, while suited to the Hartbeespoort Dam context over the study period, require further research and development for robust deployment to larger areas. Since the middle of 2021, a new invasive aquatic plant, *Salvinia minima*, has established on Hartbeespoort Dam. The algorithm is currently not developed to distinguish between these two species. Further research is required to determine if the current algorithm could be suitably adapted to respond to these limitations. Deploying more complex machine-learning approaches, such as deep learning image segmentation techniques, may also prove beneficial, particularly for discrimination between spectrally similar macrophyte species.

Future improvements to data accessibility and the WebApp are recommended. While it is currently sufficient for its intended scope and use, the existing GEE user interface tool suite prohibits more advanced functionality, such as alert notifications and automated reports. Periodical computational preprocessing of the data would also reduce delays between user data requests and data delivery on the WebApp. Investigation into alternative data storage, cloud computation and data interface solutions may be beneficial in providing more functionality than what is currently available on the GEE platform.

Investigating a potential correlation between NDVI and other plant health spectral indices and agent-induced stress could provide valuable early indicators of agent impacts on the plants as Byrne *et al.* (2010) noted a correlation between NDVI and agent herbivory on water hyacinth. Biomass estimation is another important variable in water hyacinth monitoring (Wilson *et al.*, 2000), and some research has suggested the possibility for RS to derive this information (Robles *et al.,* 2015) but it warrants further investigation and will require the collection of significant supporting field data. There is promising scope for the time-series data produced in this thesis to be analysed in the context of other collected field data, especially key variables such as agent density, nutrient supply and weather data, in order to develop a holistic model of the key drivers of the water hyacinth population dynamics on Hartbeespoort Dam.

5.3 Conclusion

The aim of this thesis was to develop an accurate, reliable method of mapping and monitoring water hyacinth on Hartbeespoort Dam, within the context of the recent successful biological control of water hyacinth by *Megamelus scutellaris*. It identified the need for monitoring as an important management and scientific resource and proposed a remote sensing solution to fill this gap by highlighting the opportunities presented by the recent technological advances in the field. Further, it illustrates how RS can be used to improve monitoring programmes in the aquatic sciences.

Challenges and opportunities were identified by examining the existing literature and through exploratory investigation of the study site. Novel solutions were developed where needed, such as the development of a new spectral index. This aided the development and production of an accurate classification algorithm that could be automated across large datasets over time. Importantly, to fill the monitoring deficit that has been noted by van Wilgen *et al.*, (2022) and others, the data were disseminated to researchers via an accessible WebApp.

This thesis demonstrated the potential of RS to effectively produce a detailed, accurate overview of water hyacinth on Hartbeespoort Dam over time, as well as serve as a useful near-real-time monitoring tool that could be presented in an accessible way to stakeholders. This provides information for improved management strategies and insight for future biological control monitoring strategies and research.

References

Ashton, P.J., Scott, W.E., Steyn, D.J. and Wells, R.J. 1979. The chemical control programme against the water hyacinth *Eichhornia crassipes* (Mart.) Solms on Hartbeespoort Dam: Historical and practical aspects. *South African Journal of Science,* 75(3): 303-306.

Azzari, G. and Lobell, D.B. 2017. Landsat-based classification in the cloud: An opportunity for a paradigm shift in land cover monitoring. *Remote Sensing of Environment*, 202: 64-74.

Balla, D., Zichar, M., Tóth, R., Kiss, E., Karancsi, G. and Mester, T. 2020. Geovisualization techniques of spatial environmental data using different visualization tools. *Applied Sciences*, 10(19). Article number: 6701.

Bannari, A., Asalhi, H. and Teillet, P.M. 2002. Transformed difference vegetation index (TDVI) for vegetation cover mapping. In *IEEE International geoscience and remote sensing symposium (5).* Toronto, Candada: 3053-3055.

Bannari, A., Morin, D., Bonn, F. and Huete, A. 1995. A review of vegetation indices. *Remote Sensing Reviews*, 13(1-2): 95-120.

Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L., Maguire, K.C. and Mersey, B. 2011. Has the Earth's sixth mass extinction already arrived? *Nature*, 471(7336): 51-57.

Byrne, M.J., Hill, M.P., Robertson, M., King, A., Jadhav, A., Katembo, N., Wilson, J., Brudvig, R. & Fisher, J. 2010. *Integrated Management of Water Hyacinth in South Africa: Development of an integrated management plan for water hyacinth control, combining biological control, herbicidal control and nutrient control, tailored to the climatic regions of South Africa.* Report to the Water Research Commission. WRC Report No. No TT 454/10. Pretoria: Water Research Commission.

Carroll, A.S.D. and Curtis, C.J. 2021. Increasing nutrient influx trends and remediation options at Hartbeespoort Dam, South Africa: a mass-balance approach. *Water SA*, 47(2): 210-220.

Cavalli, R.M., Laneve, G., Fusilli, L., Pignatti, S. and Santini, F. 2009. Remote sensing water observation for supporting Lake Victoria weed management. *Journal of environmental management*, 90(7): 2199-2211.

Chen, J.M. 1996. Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Canadian Journal of Remote Sensing*, 22(3): 229-242.

Cheruiyot, E.K., Mito, C., Menenti, M., Gorte, B., Koenders, R. and Akdim, N. 2014. Evaluating MERIS-based aquatic vegetation mapping in Lake Victoria. *Remote Sensing*, 6(8): 7762-7782.

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.C., Skakun, S.V. and Justice, C. 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote sensing of environment*, 219: 145-161.

Coetzee, J.A. and Hill, M.P. 2012. The role of eutrophication in the biological control of water hyacinth, *Eichhornia crassipes*, in South Africa. *BioControl*, 57(2): 247-261.

Coetzee, J.A., Hill, M.P., Byrne, M.J. & Bownes, A. 2011. A review of the biological control programmes on *Eichhornia crassipes* (C. Mart.) Solms (Pontederiaceae), *Salvinia molesta* DS Mitch.(Salviniaceae), *Pistia stratiotes* L.(Araceae), *Myriophyllum aquaticum* (Vell.) Verdc.(Haloragaceae) and *Azolla filiculoides* Lam.(Azollaceae) in South Africa. *African Entomology*, 19(2): 451-468.

Coetzee, J.A., Miller, B.E., Kinsler, D., Sebola, K. & Hill, M.P. 2022. It's a numbers game: inundative biological control of water hyacinth (*Pontederia crassipes*), using *Megamelus scutellaris* (Hemiptera: Delphacidae) yields success at a high elevation, hypertrophic reservoir in South Africa. *Biocontrol Science and Technology*: 1302-1311.

Datta, S., Das, P., Dutta, D. & Giri, R.K. 2021. Estimation of surface moisture content using Sentinel-1 C-band SAR data through machine learning models. *Journal of the Indian Society of Remote Sensing*, 49(4): 887-896.

Dayaram, A., Harris, L.R., Grobler, B.A., Van der Merwe, S., Rebelo, A.G., Powrie, L.W., Vlok, J.H.J., Desmet, P.G., Qabaqaba, M., Hlahane, K.M. & Skowno, A.L. 2019. Vegetation Map of South Africa, Lesotho and Swaziland 2018: A description of changes since 2006. *Bothalia*, 49(1).

Dennis, I. & Dennis, S.R. 2019. Social paradigm shift required to counter the eutrophication of the Hartbeespoort Dam in South Africa. *Water and Society V*, 239: 159-172.

Department of Water Affairs and Forestry (DWAF). 1996. *South African Water Quality Guidelines (second edition). Volume 7: Aquatic Ecosystems*. Pretoria: Department of Water Affairs and Forestry.

Dong, D., Wang, C., Yan, J., He, Q., Zeng, J. & Wei, Z. 2020. Combing Sentinel-1 and Sentinel-2 image time series for invasive *Spartina alterniflora* mapping on Google Earth Engine: a case study in Zhangjiang Estuary. *Journal of Applied Remote Sensing*, 14(4). Article number: 044504.

Dube, T., Mutanga, O., Sibanda, M., Bangamwabo, V. & Shoko, C. 2017. Testing the detection and discrimination potential of the new Landsat 8 satellite data on the challenging water hyacinth (*Eichhornia crassipes*) in freshwater ecosystems. *Applied geography*, 84: 11-22.

Elkan, C. 2003. Using the triangle inequality to accelerate k-means, in *Proceedings of the 20th international conference on Machine Learning (ICML-03).* California: University of California: 147-153.

Environmental Systems Research Institute (ESRI). 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

European Space Agency. 2016. *User Guides*. [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/user-guides/ [2022, October 20].

European Space Agency. 2022. *Sentinel-2*. [Online]. Available: https://sentinel.esa.int/web/sentinel/missions/sentinel-2 [2022, October 20].

Filgueiras, R., Mantovani, E.C., Fernandes-Filho, E.I., Cunha, F.F.D., Althoff, D. & Dias, S.H.B. 2020. Fusion of MODIS and Landsat-Like images for daily high spatial resolution NDVI. *Remote Sensing*, 12(8). Article number: 1297.

Fusilli, L., Collins, M.O., Laneve, G., Palombo, A., Pignatti, S. & Santini, F. 2013. Assessment of the abnormal growth of floating macrophytes in Winam Gulf (Kenya) by using MODIS imagery time series. *International journal of applied earth observation and geoinformation*, 20: 33-41.

Gao, B.C. 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment*, 58(3): 257-266.

García-de-Lomas, J., Dana, E.D., Borrero, J., Yuste, J., Corpas, A., Boniquito, J.M., Castilleja, F.J., Martínez, J.M., Rodríguez, C. & Verloove, F. 2022. Rapid response to water hyacinth (*Eichhornia crassipes*) invasion in the Guadalquivir river branch in Seville (southern Spain). *Management of Biological Invasions*, 13(4): 724-736.

Gitelson, A.A., Viña, A., Arkebauer, T.J., Rundquist, D.C., Keydan, G. & Leavitt, B. 2003. Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophysical research letters*, 30(5): 521-524.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. & Moore, R. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202: 18-27.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R. & Kommareddy, A. 2013. High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160): 850-853.

Harding, W.R. 2008. *The determination of annual phosphorus loading limits for South African dams*. *WRC Report No. 1687/1/08*. Pretoria: Water Research Commission.

Hill, M.P. & Coetzee, J. 2017. The biological control of aquatic weeds in South Africa: Current status and future challenges. *Bothalia-African Biodiversity & Conservation*, 47(2): 1-12.

Hill, M.P., Coetzee, J.A., Martin, G.D., Smith, R. & Strange, E.F. 2020. Invasive alien aquatic plants in South African freshwater ecosystems. In B.W. van Wilgen, J. Measey, D.M. Richardson, J.R. Wilson & T.A. Zengeya (eds.). *Biological Invasions in South Africa*. South Africa: Springer Nature. 97-114.

Hill, M.P. & Olckers, T. 2001. October. Biological control initiatives against water hyacinth in South Africa: constraining factors, success and new courses of action. In *ACIAR proceedings*: 33-38. ACIAR; 1998.

Huete, A.R. 1988. A soil-adjusted vegetation index (SAVI). *Remote sensing of environment*, 25(3): 295-309.

IPBES. 2019. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. S. Díaz, J. Settele, E. S. Brondízio E.S., H. T. Ngo, M. Guèze, J. Agard, A. Arneth, P. Balvanera, K. A. Brauman, S. H. M. Butchart, K. M. A. Chan, L. A. Garibaldi, K. Ichii,

J. Liu, S. M. Subramanian, G. F. Midgley, P. Miloslavich, Z. Molnár, D. Obura, A. Pfaff, S. Polasky, A. Purvis, J. Razzaque, B. Reyers, R. Roy Chowdhury, Y. J. Shin, I. J. Visseren-Hamakers, K. J. Willis, & C. N. Zayas (eds.). Bonn, Germany: IPBES secretariat.

Jadhav, A., Hill, M. & Byrne, M. 2008. Identification of a retardant dose of glyphosate with potential for integrated control of water hyacinth, *Eichhornia crassipes* (Mart.) Solms-Laubach. *Biological Control*, 47(2): 154-158.

Jia, M., Wang, Z., Mao, D., Ren, C., Wang, C. & Wang, Y. 2021. Rapid, robust, and automated mapping of tidal flats in China using time series Sentinel-2 images and Google Earth Engine. *Remote Sensing of Environment*, 255. Article number: 112285.

Keebine, G.L. 2019. Mapping and monitoring the spatial distribution of Eichhornia crassipes (water hyacinth) in the Hartbeespoort Dam, South Africa, using remote sensing data. Unpublished master's thesis. Johannesburg: University of the Witwatersrand.

Key, C.H. & Benson, N.C. 2006. Landscape assessment (LA). D.C. Lutes, R.E. Keane, J.F. Caratti, C.H. Key, N.C. Benson, S. Sutherland & L.J. Gangi (eds.). *FIREMON: Fire effects monitoring and inventory system*. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. 155-164.

Kiage, L.M. & Obuoyo, J. 2011. The potential link between El Nino and water hyacinth blooms in Winam Gulf of Lake Victoria, East Africa: evidence from satellite imagery. *Water resources management*, 25(14): 3931-3945.

Lewis, H.G. & Brown, M. 2001. A generalized confusion matrix for assessing area estimates from remotely sensed data. *International journal of remote sensing*, 22(16): 3223-3235.

Liang, Q., Zhang, Y., Ma, R., Loiselle, S., Li, J. & Hu, M. 2017. A MODIS-based novel method to distinguish surface cyanobacterial scums and aquatic macrophytes in Lake Taihu. *Remote Sensing*, 9(2). Article number: 133.

Lillesand, T., Kiefer, R.W. & Chipman, J. 2015. *Remote sensing and image interpretation*. New York: John Wiley & Sons.

Lim, M. 2019. 'Revolutionary change' needed to stop unprecedented global extinction crisis. *Science Education News*, 68(2): 29-31.

Liu, X., Zhai, H., Shen, Y., Lou, B., Jiang, C., Li, T., Hussain, S.B. & Shen, G. 2020. Large-scale crop mapping from multisource remote sensing images in google earth engine. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13: 414-427.

Ma, Z. & Redmond, R.L. 1995. Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing*, 61(4): 435-439.

Matongera, T.N., Mutanga, O., Dube, T. & Sibanda, M. 2017. Detection and mapping the spatial distribution of bracken fern weeds using the Landsat 8 OLI new generation sensor. *International journal of applied earth observation and geoinformation*, 57: 93-103.

Matthews, M.W. 2014. Eutrophication and cyanobacterial blooms in South African inland waters: 10 years of MERIS observations. *Remote Sensing of Environment*, 155: 161-177.

Metternicht, G. 2003. Vegetation indices derived from high-resolution airborne videography for precision crop management. *International Journal of Remote Sensing*, 24(14): 2855-2877.

Mitchell, S.A. & Crafford, J.G. 2016. *Review of the Hartbeespoort Dam integrated biological remediation programme (Harties Metsi a Me).* Pretoria: Water Research Commission.

Mucheye, T., Haro, S., Papaspyrou, S. & Caballero, I. 2022. Water quality and water hyacinth monitoring with the Sentinel-2A/B satellites in Lake Tana (Ethiopia). *Remote Sensing*, 14(19): 4921.

Mukarugwiro, J.A., Newete, S.W., Adam, E., Nsanganwimana, F., Abutaleb, K.A. & Byrne, M.J. 2019. Mapping distribution of water hyacinth (*Eichhornia crassipes*) in Rwanda using multispectral remote sensing imagery. *African Journal of Aquatic Science*, 44(4): 339-348.

Nagendra, H., Lucas, R., Honrado, J.P., Jongman, R.H., Tarantino, C., Adamo, M. & Mairota, P. 2013. Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecological Indicators*, 33: 45-59.

Owens, C.S. & Madsen, J.D. 1995. Low temperature limits of water hyacinth. *Journal of Aquatic Plant Management*, 33: 63-68.

Oyama, Y., Matsushita, B. & Fukushima, T. 2015. Distinguishing surface cyanobacterial blooms and aquatic macrophytes using Landsat/TM and ETM+ shortwave infrared bands. *Remote Sensing of Environment*, 157: 35-47.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12: 2815-2830.

Piñeiro, G., Perelman, S., Guerschman, J.P. & Paruelo, J.M. 2008. How to evaluate models: observed vs. predicted or predicted vs. observed? *Ecological modelling*, 216(3-4): 316-322.

Prudente, V.H.R., Oldoni, L.V., Vieira, D.C., Cattani, C.E.V. & Sanches, I.D. 2019. Relationship between SAR/Sentinel-1 polarimetric and interferometric data with biophysical parameters of agricultural crops. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42: 599-607.

Quayle, L.M., Dickens, C.W.S., Graham, M., Simpson, D., Goliger, A., Dickens, J.K., Freese, S. & Blignaut, J. 2010. *Investigation of the positive and negative consequences associated with the introduction of zero-phosphate detergents into South Africa. WRC report no. TT*, *446(10).* Pretoria: Water Research Commission.

Richardson, D.M. & Thuiller, W. 2007. Home away from home—objective mapping of high-risk source areas for plant introductions. *Diversity and Distributions*, 13(3): 299-312.

Ritchie, J.C., Zimba, P.V. & Everitt, J.H. 2003. Remote sensing techniques to assess water quality. *Photogrammetric Engineering & Remote Sensing*, 69(6): 695-704.

Robles, W., Madsen, J.D. & Wersal, R.M. 2015. Estimating the biomass of waterhyacinth (*Eichhornia crassipes*) using the normalized difference vegetation index derived from simulated landsat 5 TM. *Invasive Plant Science and Management*, 8(2): 203-211.

Rutherford, M.C., Mucina, L. & Powrie L.W. 2006. Biomes and bioregions of southern Africa, in L. Mucina & M.C. Rutherford (eds.). *The vegetation of South Africa, Lesotho and Swaziland.* Pretoria: South African National Biodiversity Institute. 30-51.

Scheffer, M., Szabo, S., Gragnani, A., Van Nes, E.H., Rinaldi, S., Kautsky, N., Norberg, J., Roijackers, R.M. & Franken, R.J. 2003. Floating plant dominance as a stable state. *Proceedings of the national academy of sciences*, 100(7): 4040-4045.

Scott, W.E., Seaman, M.T., Connell, A.D., Kohlmeyer, S.I. & Toerien, D.F. 1977. The limnology of some South African impoundments I. The physico-chemical limnology of Hartbeespoort Dam. *Journal of the Limnological Society of Southern Africa*, 3(2): 43-58.

Seabold, S. & Perktold, J. 2010. Statsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference,* 57(61). Article number: 10-25080.

Shanab, S.M., Shalaby, E.A., Lightfoot, D.A. & El-Shemy, H.A. 2010. Allelopathic effects of water hyacinth [*Eichhornia crassipes*]. *PloS one*, 5(10). Article number: e13200.

Singh, G., Reynolds, C., Byrne, M. & Rosman, B. 2020. A remote sensing method to monitor water, aquatic vegetation, and invasive water hyacinth at national extents. *Remote Sensing*, 12(24). Article number: 4021.

Skakun, S., Wevers, J., Brockmann, C., Doxani, G., Aleksandrov, M., Batič, M., Frantz, D., Gascon, F., Gómez-Chova, L., Hagolle, O. & López-Puigdollers, D. 2022. Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. *Remote Sensing of Environment*, 274. Article number: 112990.

Skakun, S., Wevers, J., Brockmann, C., Doxani, G., Aleksandrov, M., Batič, M., Frantz, D., Gascon, F., Gómez-Chova, L., Hagolle, O. & López-Puigdollers, D. 2022. Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. *Remote Sensing of Environment*, 274. Article number: 112990.

Skowno, A.L., Poole, C.J., Raimondo, D.C., Sink, K.J., Van Deventer, H., Van Niekerk, L., Harris, L.R., SmithAdao, L.B., Tolley, K.A., Zengeya, T.A., Foden, W.B., Midgley, G.F. & Driver, A. 2019. *National Biodiversity Assessment 2018: The status of South Africa's ecosystems and biodiversity.* Synthesis Report. South African National Biodiversity Institute, an entity of the Department of Environment, Forestry and Fisheries, Pretoria: 1-214.

Steven, M.D. 1987. Ground truth an underview. *International Journal of Remote Sensing*, 8(7): 1033-1038.

Story, M. & Congalton, R.G. 1986. Accuracy assessment: a user's perspective. *Photogrammetric Engineering and remote sensing*, 52(3): 397-399.

Sun, Z., Xu, R., Du, W., Wang, L. & Lu, D. 2019. High-resolution urban land mapping in China from sentinel 1A/2 imagery based on Google Earth Engine. *Remote Sensing*, 11(7). Article number: 752.

Sutton, G.F., Compton, S.G. & Coetzee, J.A. 2016. Naturally occurring phytopathogens enhance biological control of water hyacinth (*Eichhornia crassipes*) by *Megamelus scutellaris* (Hemiptera: Delphacidae), even in eutrophic water. *Biological Control*, 103: 261-268.

Thamaga, K.H. & Dube, T. 2018a. Testing two methods for mapping water hyacinth (*Eichhornia crassipes*) in the Greater Letaba river system, South Africa: discrimination and mapping potential of the polar-orbiting Sentinel-2 MSI and Landsat 8 OLI sensors. *International journal of remote sensing*, 39(22): 8041-8059.

Thamaga, K.H. & Dube, T. 2018b. Remote sensing of invasive water hyacinth (*Eichhornia crassipes*): A review on applications and challenges. *Remote Sensing Applications: Society and Environment*, 10: 36-46.

Thamaga, K.H. & Dube, T. 2019. Understanding seasonal dynamics of invasive water hyacinth (*Eichhornia crassipes*) in the Greater Letaba river system using Sentinel-2 satellite data. *GIScience & Remote Sensing*, 56(8): 1355-1377.

Tipping, P.W., Sosa, A., Pokorny, E.N., Foley, J., Schmitz, D.C., Lane, J.S., Rodgers, L., Mccloud, L., Livingston-Way, P., Cole, M.S. & Nichols, G. 2014. Release and establishment of *Megamelus scutellaris* (Hemiptera: Delphacidae) on water hyacinth in Florida. *Florida Entomologist*, 97(2): 804-806.

Tucker, C.J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2): 127-150.

Tucker, C.J., Elgin Jr, J.H., McMurtrey Iii, J.E. & Fan, C.J. 1979. Monitoring corn and soybean crop development with hand-held radiometer spectral data. *Remote Sensing of Environment*, 8(3): 237-248.

Twinch, A.J. 1986. The phosphorus status of sediments in a hypertrophic impoundment (Hartbeespoort Dam): implications for eutrophication management. *Hydrobiologia*, 135(1-2): 23-34.

van Wilgen, B.W., Measey, J., Richardson, D.M., Wilson, J.R. & Zengeya, T.A. 2020. Biological invasions in South Africa: an overview. In B.W. van Wilgen, J. Measey, D.M. Richardson, J.R. Wilson & T.A. Zengeya (eds.). *Biological Invasions in South Africa*. South Africa: Springer Nature. 3-31.

van Wilgen, B.W., Wannenburgh, A. & Wilson, J.R. 2022. A review of two decades of government support for managing alien plant invasions in South Africa. *Biological Conservation*, 274. Article number: 109471.

van Wyk, E. & van Wilgen, B.W. 2002. The cost of water hyacinth control in South Africa: a case study of three options. *African Journal of Aquatic Science*, 27(2): 141-149.

Venugopal, G. 1998. Monitoring the effects of biological control of water hyacinths using remotely sensed data: a case study of Bangalore, India. *Singapore Journal of Tropical Geography*, 19(1): 91-105.

Wang, X., Xiao, X., Zou, Z., Hou, L., Qin, Y., Dong, J., Doughty, R.B., Chen, B., Zhang, X., Chen, Y. & Ma, J. 2020. Mapping coastal wetlands of China using time series Landsat images in 2018 and Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, 163: 312-326.

Willmott, C.J. & Matsuura, K. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1): 79-82.

Wilson, J.R. 2002. *Modelling the dynamics and control of water hyacinth, Eichhornia crassipes (Martius) Solms-Laubach.* Published doctoral dissertation. London: University of London.

Wilson, J.R., Ajuonu, O., Center, T.D., Hill, M.P., Julien, M.H., Katagira, F.F., Neuenschwander, P., Njoka, S.W., Ogwang, J., Reeder, R.H. & Van, T. 2007. The decline of water hyacinth on Lake Victoria was due to biological control by *Neochetina* spp. *Aquatic Botany*, 87(1): 90-93.

Wilson, J.R., Holst, N. & Rees, M. 2005. Determinants and patterns of population growth in water hyacinth. *Aquatic Botany*, 81(1): 51-67.

Wilson, J.R., Rees, M., Holst, N., Thomas, M.B. & Hill, G. 2000, October. Water hyacinth population dynamics. In *Aciar Proceedings: 96-104. ACIAR (1998).* United Kingdom: Imperial College at Silwood Park.

Woodcock, C.E., Loveland, T.R., Herold, M. & Bauer, M.E. 2020. Transitioning from change detection to monitoring with remote sensing: A paradigm shift. *Remote Sensing of Environment*, 238. Article number: 111558.

Zhao, Q., Yu, L., Li, X., Peng, D., Zhang, Y. & Gong, P. 2021. Progress and trends in the application of Google Earth and Google Earth Engine. *Remote Sensing*, 13(18). Article number: 3778.

Zohary, T. 1985. Hyperscums of the cyanobacterium *Microcystis aeruginosa* in a hypertrophic lake (Hartbeespoort Dam, South Africa). *Journal of Plankton Research*, 7(3): 399-409.

Zupanc, A. 2020. *Improving cloud detection with machine learning*, *Medium*. Sentinel Hub Blog. Available at: https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13 [2022, October 20].