Remote sensing for detecting rapid post-fire recovery as Groundwater-Dependent Ecosystems in the Cape Floristic Region

Simcelile Chenge

A dissertation submitted in fulfilment
of the requirements for the degree of
Master of Science (Geomatics Engineering)

Supervised by Julian Smit, Adam G. West, Kaveer Singh

Geomatics Division
The School of Architecture, Planning & Geomatics
Faculty of Engineering & the Built Environment
University of Cape Town

2021-06-14

Declaration

I know the meaning of plagiarism and declare that all the work in the document, save for that which is properly acknowledged, is my own. This thesis/dissertation has been submitted to the Turnitin module (or equivalent similarity and originality checking software) and I confirm that my supervisor has seen my report and any concerns revealed by such have been resolved with my supervisor.

Signed by candidate

Signature _____

Date 23 February 2021

APPLICATION FORM

Please Note:
Any person planning to undertake research in the Faculty of Engineering and the Built Environment (EBE) at the University of Cape Town is required to complete this form **before** collecting or analysing data. The objective of submitting this application *prior* to embarking on research is to ensure that the highest ethical standards in research, conducted under the auspices of the EBE Faculty, are met. Please ensure that you have read, and understood the **EBE Ethics in Research Handbook** (available from the UCT EBE, Research Ethics website) prior to completing this application form: http://www.ebe.uct.ac.za/ebe/research/ethics1

APPLICANT'S DETAILS		
Name of principal researcher, student or external applicant		Simcelile Chenge
Department		Department of Geomatics (School of APG)
Preferred email address of applicant:		CHNSIM007@myuct.ac.za
If Student	Your Degree: e.g., MSc, PhD, etc.	MSc
	Credit Value of Research: e.g., 60/120/180/360 etc.	180
	Name of Supervisor (if supervised):	A/Professor Julian Smit
If this is a researchcontract, indicate the source of funding/sponsorship		SANSA
Project Title		"Application of geospatial methods to identify and map vulnerable Groundwater Dependent Ecosystems (GDEs) in the Fynbos Mountains"

I hereby undertake to carry out my research in such a way that:

- there is no apparent legal objection to the nature or the method of research; and
- the research will not compromise staff or students or the other responsibilities of the University;
- the stated objective will be achieved, and the findings will have a high degree of validity;
- limitations and alternative interpretations will be considered;
- the findings could be subject to peer review and publicly available; and
- I will comply with the conventions of copyright and avoid any practice that would constitute plagiarism.

SIGNED BY	Full name	Signature	Date
Principal Researcher/ Student/External applicant	Simcelile Chenge	Signed by candidate	**02 Oct 2018**

APPLICATION APPROVED BY	Full name	Signature	Date
Supervisor (where applicable)	A/Professor Julian Smit	✓	02 Oct 2018
HOD (or delegated nominee) Final authority for all applicants who have answered NO to all questions in Section1; and for all Undergraduateresearch (Including Honours).	Patruba Odera Click here to enter text		3/10/2018 Click here to enter a date.
Chair : Faculty EIR Committee For applicants other than undergraduate students who have answered YES to any of the above questions.			

Acknowledgements

It took me a great deal of effort and frustration to put together this dissertation. Conversely, the period was also enjoyable. I made professional friends and very personal ones. I would like to acknowledge their support and encouragement during this period.

I am particularly thankful to my supervisor, Julian Smit. We have a very chilled and nurturing working relationship. Julian brought me to UCT and introduced me to my now many professional friends. One of them is my co-supervisor, Adam West. Adam taught and uplifted me academically and personally. His passion for teaching and contributing meaningfully to Society through Science saw me right to the end, especially in the pandemic's mist. I am also grateful to my third co-supervisor Kaveer Singh for his advice and help in writing and data analysis for always having fruits in his desk for me during our meetings.

Our Principal Technical Officer in the Geomatics Division, Mignon Wells, for her technical support and the way she does it - with care and love. From hardware to software licenses, she made sure I had every thing I needed to do my work. While doing the work, she ensured I was warm and safe from any cold in the lab by giving me her blanket. I promise to bring it back;).

I am also grateful to the librarians in the Built Environment Library and Science & Engineering librarians in the main library for showing me how the library works and for fast tracking my requests for books not currently available in the library. Finally, I appreciate my colleagues in the Geomatics Division and Westlab in the Biological Sciences department for their companion in this journey and keeping each other sane during this period.

Table of Contents

List of Tables

List of Figures

Abstract

Groundwater Dependent Ecosystems (GDEs) concentrate high levels of biodiversity and several species not found anywhere else. They prevail in the landscape through the ecological contribution of groundwater. They, GDEs, are vulnerable to drastic changes in groundwater depth. If, for example, bulk groundwater pumping drastically increases the groundwater depth and GDEs can no longer access it, they would die out. In the Cape Floristic Region (CFR), South Africa, there is limited information about the spatial distribution of groundwater dependent ecosystems. With the CFR having multiple locations with current and subsequent bulk groundwater pumping, identifying the spatial distribution of GDEs is a prerequisite for establishing their groundwater requirements. This dissertation presents a proposed novel method to identify rapid recovering wetlands predicted to be GDEs and uses Random Forest (RF) to predict their spatial distribution. The proposed novel approach leveraged the periodic fire disturbances in the CFR and applied the remote sensing index; Normalised Difference Vegetation Index (NDVI) extracted from high spatial resolution (1 m) aerial orthoimages. The proposed novel approach involves three levels of analysis. The first two levels used a one-way Analysis of Variance (ANOVA) to analyse the sensitivity of mean NDVI to discriminate wetland and non-wetland classes in burned and unburned study sites, and a post-hoc test: Tukey's Honest Significant Differences (HSD) pair-wise comparison to detect differences between the wetland and non-wetland mean NDVI and infer an NDVI threshold of wetland classes. In unburned sites, ANOVA showed no statistical significance between wetland and non-wetland classes, $F\ (2,15) = 3.53$, $p = 0.055$. In burned sites, however, ANOVA showed there was a significant difference between wetland and non-wetland classes, $F\ (2,15) = 9.66$, $p = 0.002$. ANOVA and Tukey showed there were significant differences between

wetland and non-wetland classes, with wetlands having between 0.22 and 0.37 greater NDVI than non-wetlands. The last level of analysis employed a kernel density estimator function to assess the recovery rate post-burn and use it to detect faster recovery as potential of wetlands to be GDEs; results showed that potential wetland GDEs experience rapid NDVI recovery $>$ 236 days post-fire. In the fire prone CFR, leveraging fire data to detect GDEs provides a potentially simple and efficient way of building a local database for GDEs. The proposed novel approach showed leveraging fire data is a simple alternative to laborious field data to identify and map GDEs in the CFR. But because of the finite spectral bands in aerial orthoimages, Sentinel-2A multi-epochs dataset was utilised to carry out random forest for predicting the spatial distribution of potential wetland GDEs in the Kogelberg Nature Reserve. Sentinel-2A bands: Short-Wave Infrared (SWIR), Near-Infrared (NIR), Red-edge, Red, Green, NDVI and Normalised Difference Wetness Index (NDWI) predictors and the potential wetland GDEs/non-wetland classes as the response. I tuned RF using five-fold repeated spatial cross-validation instead of the typical cross-validation tuning to account for the spatial structure of the data. The overall predictive accuracy of RF was between 59%-71%. This predictive accuracy may have been reduced by the application of spatial cross-validation that accounted for the spatial autocorrelation in the multi-date data. The dissertation showed that Sentinel-2A multi-date data applies in predicting the distribution of potential wetland GDEs but might not be effective for smaller ($< 100\ m^2$) wetlands. These small wetlands showed rapid post-fire recovery (less than a year post-fire) and were effectively detected with high resolution aerial orthoimages (1 m) spatial resolution.

Keywords: GDEs, groundwater, spatial, NDVI, fire, Cape Floristic Region (CFR)

Abbreviations

ANOVA	Analysis of Variance
ADEs	Aquifer Dependent Ecosystems
AOA	Area of Applicability
BGIS	Biodiversity GIS
BOA	Bottom of Atmosphere reflectance
CARET	Classification and Regression Training
CAST	Caret Applications for Spatio-Temporal
CFR	Cape Floristic Region
df	Degrees of freedom
DI	Dissimilarity Index
Dim	Dimension
EM	Electromagnetic (EM) spectrum
ESA	European Space Agency
EU	European Union
GDEs	Groundwater Dependent Ecosystems
GEM	Groundwater Dependent Ecosystem Mapping
GIS	Geographic Information Systems
JPEG	Joint Photographic Experts Group
L1C	Level-1C
L2A	Level-2A
m	Metres
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalised Difference Vegetation Index
NDWI	Normalised Difference Wetness Index
NIR	Near- Infrared

NIR	Near Infrared Reflectance
nm	Nanometre
OOB	Out-of-Bag
PCA	Principal Component Analysis
PWC	Plant Water Content
RF	Random Forest
RGB	Red Green Blue
S2A	Sentinel-2A
S2B	Sentinel-2B
SAFE	Standard Archive Format for Europe
SANBI	South African National Biodiversity Institute
SAR	Synthetic Aperture Radar
SAST	South African Standard Time
SNAP	Sentinel Application Platform
SVM	Support vector machine
SWIR	Short-Wave Infrared
SWIR	Short-wave infrared
TC	Tasseled Cap
TMG	Table Mountain Group Aquifer System
TOA	Top-of-Atmosphere Reflectance
Tukey's HSD	Tukey's Honest Significant Differences
UTC	Coordinated Universal Time
WCNCB, CapeNature	Western Cape Nature Conservation Board
WRC	Water Research Commission

Chapter 1

Introduction

1.1 Introduction to GDEs research

Like other semi-arid regions, South Africa is shifting increasingly from surface water resources, as the primary water supply, to groundwater resources. But if human abstraction from these groundwater reserves exceeds recharge, that may lower groundwater levels beyond the reach of Groundwater Dependent Ecosystems (GDEs). GDEs may vanish without groundwater. Eamus & Froend (2006) said that to protect GDEs, catchment water managers must decide, "How much water can be taken from a groundwater system/aquifer while still maintaining a low level of risk to GDEs?" The answer to this question guides catchment water managers when they allocate water use. They undertake to balance environmental water requirements and the growing urgency for bulk groundwater abstraction. If groundwater abstraction grows, the precondition to allocate water equitably between the environment and meet basic human needs and other human requirements embodied in environmental and water laws/policies becomes essential.

1.2 Research problem

The South African National Water Act (No. 36) of 1998 defines and makes provisions for the ecological reserve: water required to maintain the health of an ecosystem and

meet the need for groundwater abstraction. Experts must determine the ecological reserve relying on data about the water requirements of ecosystems. Ecosystems require water to maintain their health, but when a groundwater abstraction scheme subverts the ecological reserve, consequences for ecosystems might be catastrophic. These ecosystems might lose plant communities in vast numbers, and usually, the harm is irreversible. For example, Western Australia, in 1985, suffered a loss close to 80% of its native *Banksia* tree species (Groom, Froend & Mattiske, 2000). The Water Authority of Western Australia later blamed their loss to drastic drops in groundwater levels (or drawdown) induced by bulk groundwater abstraction. Thereafter, in another well-field before undertaking a groundwater abstraction project, the Water Authority of Western Australia investigated the impact on *Banksia* tree species. Groom, Froend & Mattiske (2000) reported that groundwater monitoring started a year before beginning with the groundwater abstraction project in 1990. Between 1990 and 1991, the groundwater abstraction resulted in a 2.2 m decline in groundwater levels. Within a 200 m radius of the well-field, the groundwater drawdown coincided with high mortality, 80% and 64%, of individual *Banksia* species overstorey and understorey, respectively. Such loss of *Banksia* species linked to the impact of groundwater drawdown within a space of one year typifies the need to establish an ecological reserve before beginning a groundwater abstraction project.

The impact of groundwater abstraction on groundwater dependent plant species, as reported in Western Australia, maybe clear after the fact. In most instances, however, impact on GDEs entails a gradual loss in plant communities. Groundwater dependent plant species, once their access to groundwater is cut-off, just never bounce back. So, they find their safe-haven in areas of reliable groundwater supply (Keppel et al., 2012). These species rely on groundwater more than their surroundings and assemble into GDEs that consistently appear as geographically isolated ecological patches of green islands with high moisture content (e.g., springs, seeps and wetlands). Safe-havens like these, create conducive environments for these sensitive species to thrive. But when the safe-havens undergo radical changes in groundwater depth, they may vanish. The threat of groundwater drawdown is worsening under the changing climate. McLaughlin et al. (2017) claimed

that freshwater (e.g., groundwater) more than temperature regulates the spatial distribution of plant species and noted that areas able to sustain their greenness in the face of the changing climate are hydrologic refugia. GDEs are a type of hydrologic refugium because of their access to the local/regional groundwater. Groundwater access and/or use may lead to dependency on this resource. But access alone may not equate to dependence on groundwater; the dependence to groundwater is when ecosystems cannot survive without the resource (i.e., groundwater; Colvin et al., 2007). The latter narrows what GDEs mean but still poses risks of misclassifying actual GDEs as non-GDEs. Han et al. (2015) provided a more biologically accurate and comprehensive alternative meaning and suggested that access and/or use of groundwater lead to dependency in groundwater. So, GDEs may have partial and/or complete dependence in groundwater. When you define the nature of ecosystem dependence on groundwater as partial and/or complete, the urgency to establish the ecological reserve as a tool to protect GDEs is clearer.

The ecological reserve as a tool helps groundwater abstraction projects minimise ecological and environmental effects, even more so on vulnerable GDEs. In the Western Cape, South Africa, the Water Research Commission (WRC) funded the project "Ecological and Environmental Impacts of Large-scale Groundwater Development in the Table Mountain Group (TMG) Aquifer System" to identify and characterise the impact of bulk groundwater abstraction on GDEs (Colvin et al., 2009). The project tested its potential impact on GDEs at specific well-fields. Colvin et al. (2009) showed that GDEs experienced distress coinciding with groundwater drawdown but required long-term monitoring to establish the degree and nature of dependence in groundwater. The South African National Water Act (No. 36) of 1998 requires projects abstracting groundwater to identify GDEs and determine their ecological reserve and then continue with groundwater abstraction. Groundwater abstraction is the primary impetus behind such projects, so they might be inadequate to identify GDEs and to develop a comprehensive database of the spatial distribution of GDEs. As shown by Colvin et al. (2009) to establish dependence in groundwater requires longer monitoring periods. The case-by-case approach, identifying GDEs to carry out groundwater abstraction,

might risk discovering groundwater reliance once already vulnerable endemic plant species have died out. This case-by-case approach dominates GDEs research in the Cape Floristic Region (CFR): a biodiversity hotspot underlaid by the TMG Aquifer system with ongoing and planned bulk groundwater abstractions. The Aquifer system sustains the hyper-diverse fynbos vegetation of the CFR, often in small patches (e.g., 100-4000 m^2). These small patches are potential GDEs (February et al., 2004). There is a current lack of detailed mapping of GDEs in the CFR at the relevant management scale (i.e., catchment). Without such data, we cannot monitor potential GDEs to measure the increase in groundwater depth below which they cannot survive.

1.3 Research question

How can detailed mapping of GDEs at catchment scale be achieved to better monitor the impact of groundwater abstraction?

1.4 Aim and objectives

In this dissertation, the aim is to identify and map potential GDEs by proposing a novel approach that leverage the recurrent fires in the CFR. The objectives are: (i) to employ fire data to detect wetland classes in post-fire conditions. (ii) estimate their (wetland classes) potential to be GDEs and predict their spatial distribution by employing the supervised machine learning classification, Random Forest (RF).

1.5 Research contribution

This dissertation adds to GDEs research by advancing the local database of GDEs in the CFR through the proposed novel method. In two respects, the proposed method is unlike any other previous attempts at identifying the areas of GDEs. First, it leverages the frequent fires (accessible data) in the CFR and integrates the data with the remote sensing index Normalised Difference Vegetation Index

(NDVI) to simplify how we locate GDEs in the landscape. Second, the approach is straightforward, and since orthoimages could entirely delineate fire scars; it requires no laborious field visits to collect data.

1.6 Layout

The current chapter provides the backdrop against which to propose the novel approach, the context of the study, research problem and response to the research problem. The chapter gives evidence on the worsening state of vulnerable GDEs plant species due to increased demand for bulk groundwater abstraction. Beyond characterising and assessing the potential harm of groundwater abstraction projects, the chapter emphasises the urgency to establish the ecological contribution of groundwater to GDEs. Research dedicated to establishing a GDEs database, beyond just completing an ecological assessment report to go ahead with a groundwater abstraction, serves to develop monitoring networks and thus, a sustainable approach to protect vulnerable GDEs.

Chapter 2 provides the current landscape of GDEs research focusing on identifying/mapping of GDEs. And further presents the GDEs that this study focuses on, wetland GDEs. The first part of the chapter, introduces the relationship between GDEs and groundwater. The meaning of groundwater, its ecological contribution, and how species use and/or depend on groundwater. The second part, covers two key themes. Further explains geospatial technologies (e.g., remote sensing) employed to find GDEs. Then gives a review of geospatial technologies in GDEs research and present the full picture of what we know about finding GDEs in the landscape. The chapter conclude by presenting a review the current approaches to finding and mapping GDEs in the CFR.

Chapter 3 outlines the methodological approach and describe the geospatial data employed in the dissertation. Overall, the chapter has three sections. The first section develops and explain the logic behind the proposed novel approach. The second section, deals with Sentinel-2A data application for modelling potential GDEs with random forest. Then the last section of the chapter, describes the key shortfalls of the methodological approach, focusing on proposed novel approach

and the spatial modelling, and their meaning to the validity of the dissertation claim.

Chapter 4 presents the results and analyses. The chapter has two primary sections that follow the methods. The first section presents the results of the post-fire NDVI analysis method. This follows three levels of analysis. In level one of analysis, the method determines if post-fire NDVI can detect the difference between wetlands and non-wetlands. The second level of analysis defines a quantitative NDVI baseline for delineating wetlands, and the last level analysis is an NDVI seasonal distribution estimate analysis as a way of estimating groundwater dependence of the delineated wetland classes. Last section of the chapter presents results and analysis of the machine learning (i.e., predictive modelling with random forest). Firstly, oulines the Principal Component Analysis (PCA) for the multi-date Sentinel-2A dataset. Secondly, shows the results of the predictive modelling and spatial prediction with random forest.

Chapter 5 concludes the dissertation and look ahead at the future of GDEs research with the proposed novel approach as a method of finding and mapping GDEs.

Chapter 2

Literature review of GDEs research

2.1 Introduction

This chapter presents a review of published GDEs studies to establish the current landscape of GDEs research. Its focus is on applications of geospatial technologies in GDEs research. Past studies have explored several approaches to identify GDEs in the landscape including geospatial technologies.

Geospatial technologies (e.g., Remote Sensing and Geographic Information Systems; GIS) gather data about environmental properties (e.g., vegetation greenness) without being in contact with the environment. In environmental/ecosystem studies remote sensing reduces the need for laborious field visits to collect data and by collecting data, even in the most remote areas. The advantages of remote sensing are well documented in the literature and its growing use in groundwater-related research. But the groundwater phenomena still present a unique challenge for remote sensing, since groundwater is an unseen resource below the ground and soil surface (Blomquist & Ingram, 2003). Above the surface, groundwater has ecological indicators. Remote sensing relies on ecological indicators to gather data on groundwater, a way around its unseen nature. The demand for groundwater is growing and equally the need to establish efficient conservation strategies for

vulnerable GDEs particularly in world biodiversity hotspots (i.e., areas with high biological diversity, containing $\geq$ 1500 endemic plant species and having lost $\geq$ 70 % of its native vegetation; Myers et al., 2000). Plant species in biodiversity hotspots may depend on groundwater, with the growing urgency for groundwater pumping, the imperative is to define their relationship with the groundwater. How does groundwater contribute to maintaining GDEs?

2.2 What is groundwater?

Knowing what groundwater means would help to understand its ecological contribution. The technical definition of groundwater differs based on context; it means different things to different people (Holmes, 2000). Past studies show no unanimity on how to define groundwater. Howbeit, the past studies, focused on recurring themes in specific disciplines (e.g., biology or geohydrology). These disciplines place significance in key attributes of groundwater. For example, geohydrology emphasises on the volume of water for potential abstraction while ecology focuses on its ecological role. Regardless of its role in the environment, groundwater makes up a stage in the hydrological cycle. In the hydrological cycle, water originates as precipitation and drains under the influence of gravity into underground storage through the small surface and subsurface voids, pores and cracks (Doody et al., 2017). These make up the weathered bedrock (or aquifer) and above it lies the vadose zone, an area of unsaturated rated rock and soil moisture (Dawson, Hahm & Crutchfield-Peters, 2020). The cross section in Figure 2.1 illustrate this nicely. The two zones are separated by the water table (i.e., an upper layer of groundwater).

The groundwater contest in ecology and geohydrology exemplifies the context-based definition of groundwater. But even the South African National Water Act (No. 36) of 1998,, one of the most progressive water legislations in the world does not define groundwater. The Act instead defines an aquifer, then grounds the groundwater definition to the geohydrology perspective. Either perspective cannot capture groundwater meaning from the GDEs perspective. The studies on GDEs require some guidelines in defining groundwater to identify ecosystems that

rely on groundwater for their existence in the landscape. Serov & Kuginis (2017) proposed settling to a comprehensive definition from both ecology and geohydrology perspectives and, defines groundwater accessible to GDEs as:

> "Water naturally available in the subsurface, inclusive of both the saturated and unsaturated zones"

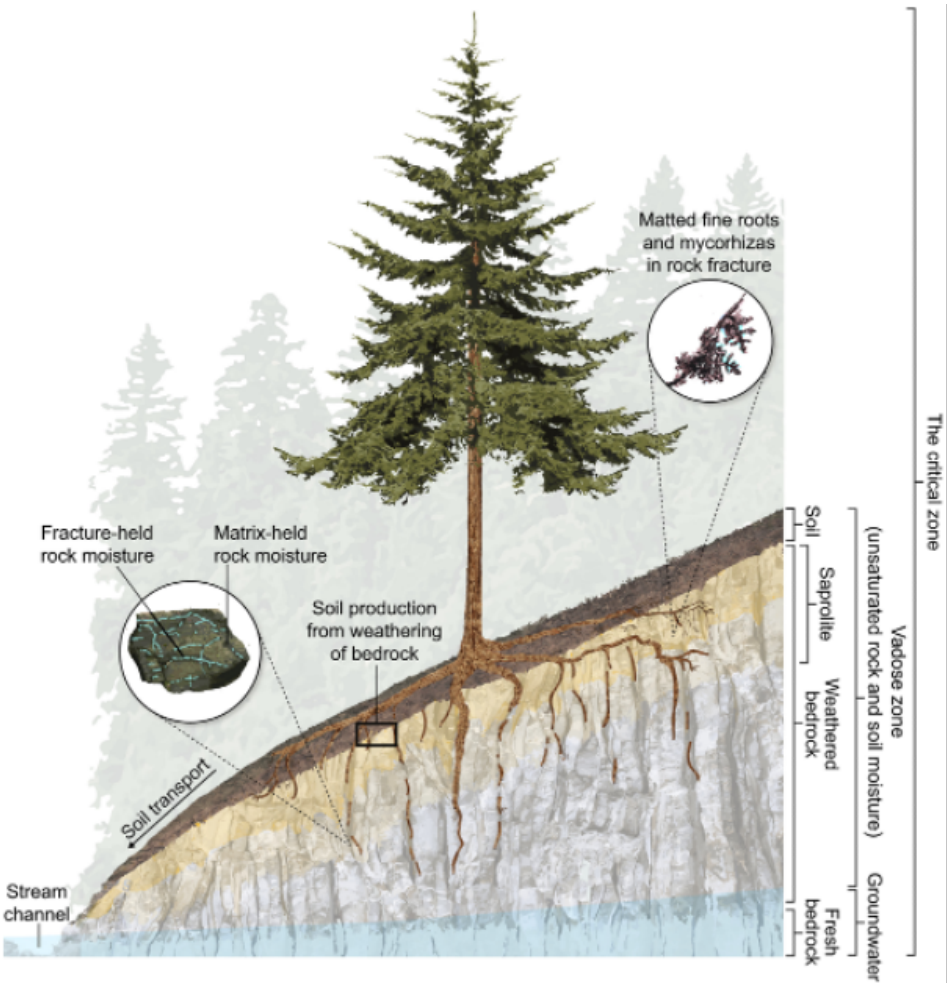

Figure 2.1: A cross sectional look of the above and below-ground sections, the Critical Zone, that influence the availability of groundwater (Dawson et al., 2020).

2.3 Groundwater dependent ecosystems

GDEs include both native plant and animal communities. This dissertation focuses on the plant communities of GDEs. The term describes ecosystems that might not withstand conditions if groundwater were no longer available. Past studies have defined it to mean:

2.3.1 Definitions

(i) ecosystems that partially and or completely depend on groundwater availability for their persistence in the landscape (Knight Merz & Sinclair Knight Merz, 2001 and Murray et al., 2003).

(ii) ecological communities of species that depend on water emerging from aquifers or groundwater (Colvin et al., 2007).

(iii) ecosystems that depend on groundwater and could significantly degrade even irreversibly be degraded if groundwater was no longer accessible (Howard & Merrifield, 2010).

(iv) ecosystems that require access to groundwater to meet all or some of their water requirements to maintain the communities of biota, ecological processes they support, and ecosystems services they provide (Adams, Smith & Yang, 2015).

2.3.2 Classification of GDEs

Past studies classify GDEs based on access to groundwater, seasonal use of groundwater, and the nature of the porous medium (i.e., aquifers): primary and secondary aquifers. Hatton & Evans (1998), in Australia, established the first classification system of GDEs. Their classification defined a criterion of dependence in groundwater as: "the degree of dependence in groundwater proportional to the fraction of the annual water budget that GDEs derive from groundwater sources". This criterion was aimed at identifying ecosystems that would degrade if groundwater were unavailable because of abstraction. The classification system broadly grouped GDEs into four groups:

(i) terrestrial GDEs: land areas that rely on groundwater more than surface water bodies.

(ii) river base-flow Systems: this group comprise surface water- groundwater interactions in river base flows. The base-flow in river systems may have a groundwater component that sustains its health.

(iii) aquifer and cave Ecosystems: population or species that resides in aquifers and caves.

(iv) wetlands: the wetlands with known and likely components of groundwater.

But this classification system lacked experimental evidence validating the accuracy of the grouping. So, they further sub-divided the classification system by ecosystems based on groundwater access and how often in a year, an ecosystem uses and/or relies on groundwater: (i) obligate GDEs: highly dependent on groundwater, (ii) facultative GDEs: proportionally rely less on groundwater and (iii) opportunistic GDEs: no obvious reliance on groundwater but use it when available. The last two classes make up GDEs that use groundwater (or ephemeral users) and those that appear to be reliant on groundwater but are almost entirely fed by surface water. Unlike the first class, the last two classes do not capture that without groundwater GDEs may not survive in the landscape. To detect them in the landscape, past studies have relied on this first class of GDEs. But the class broad nature is a major weakness and fails to account for why groundwater use does not automatically equate to dependency (Colvin et al., 2007).

In South Africa Colvin et al. (2007) developed a narrower classification system of GDEs, known as Aquifer Dependent Ecosystems (ADEs). They defined ADEs as having a direct connection with an aquifer system and whose structure and functions would be altered without groundwater. And grouped them based on known South African aquifer systems (i.e., considering their lithology; physical properties): (i) in-aquifer ADEs, (ii) springs ADES, (iii) riverine aquatic ADEs, (iv) riparian ADEs, (v) terrestrial ADEs and (vi) estuarine ADEs. The Colvin et al. (2007) classification system is unlikely to apply outside South Africa. Even in South Africa, ADEs intentionally exclude other types of GDEs not directly connected to

an aquifer system.

An earlier classification system developed by Eamus et al. (2006) is more inclusive and applies in settings outside Australia. Their aim was to develop a functional methodology for practitioners to identify GDEs (or potential GDEs) in the field. So, they grouped GDEs into classes based on their use/dependence to a common groundwater source. They proposed three primary classes of GDEs:

(i) aquifers and cave ecosystems: GDEs that live below the surface, in aquifers or in caves (stygofauna). This class also includes the hyporheic zone of the river systems.

(ii) ecosystems relying on surface expression of groundwater: these GDEs include base-flow, wetlands, floodplains and estuarine.

(iii) ecosystems relying on subsurface expression of groundwater: GDEs that do not require surface expression of groundwater since they can access groundwater in aquifers through deep roots or via capillary fringe. These GDEs often access groundwater by extending their roots into the aquifer. They include terrestrial communities (e.g., phreatophytes).

Eamus et al. (2006) went beyond just classifying GDEs. They also developed a toolbox for identifying each class of GDEs and key sets of questions that might help practitioners to select appropriate methods to identify the GDEs in the field. The key questions included but not limited to: (i) "which population or species of an ecosystem are groundwater dependent", (ii) "if some populations or species of an ecosystem are GW-dependent, what degree of dependency is expressed?".

The functional methodology is widely adopted in the research literature (Kløve et al., 2011; Barron et al., 2014). Howbeit, in the research literature current classifications systems (Table 2.1), including the functional methodology, lack simplistic guidelines on how to define and detect areas of GDEs in the landscape . To bridge the knowledge gap, Serov & Kuginis (2017) suggested a comprehensive sub-ecosystem classification guideline of GDEs using Australia as an example. Their recommended approach concentrated on the ecological role of groundwater, like

the system recommended by (Bertrand et al., 2012). Although an effective step towards comprehensive GDEs classification, their suggested classification applies in one country, Australia. So, whether the detailed site-specific guideline classification system proposed by Serov & Kuginis (2017) or the universal approaches established by Hatton & Evans (1998) and Eamus et al. (2006) a need exists to improve our understanding of the interaction between groundwater and related ecosystems using groundwater. "What?" and "where?" of GDEs are fundamental attributes or elements for a holistic and simplistic classification system (Eamus & Froend, 2006).

To detect the "where?" of GDEs you often rely on existing classification systems that define GDEs (i.e., the what? part). GDEs have a unifying element; they appear greener and contain more moisture than their surroundings because of their access to a groundwater source often decoupled from the prevailing regional climate. During prolonged drought seasons, the unseen (i.e., underground) groundwater sources experience delayed response. GDEs access and/or use of the groundwater means they appear ecologically greener and moisture abundant than their surroundings relying on surface water bodies. Eamus et al. (2016) define the phenomena as the "Green Island" hypothesis.

Such Green Island (or safe-havens) characterised by stable climate conditions and potentially house highly diverse plant species in small patches (e.g., < 0.1 hectares; Keppel et al., 2012). Such conditions also characterise hydrologic refugia, areas of stable moisture content compared to their surroundings due to a decoupled water source like groundwater (McLaughlin et al., 2017)

Table 2.1: The existing GDEs classification systems.

Year	Approach of classification & studies
1998	The first proposed classification system based on groundwater access and type of dependence (Hatton & Evans, 1998)
2006	Update of the initial GDE classification (Eamus et al., 2006)
2007	Classification of GDEs based on geological settings and development of the term "Aquifer Dependent Ecoystems (ADEs)" (Colvin et al., 2007)
2017	Review of existing classification systems and defining GDEs down to sub-ecosystem units (Serov & Kuginis, 2017)

2.3.3 Inventories of GDEs

GDEs functional methodology underlies one of the most comprehensive GDEs inventory in the world, the Australian online GDEs atlas. The Australian Bureau of Meteorology started it as a tool to inform groundwater planning and management (Eamus et al., 2006). The online GDEs atlas describes three major classes of GDEs: (i) aquatic GDEs, (ii) terrestrial GDEs and (iii) subterranean GDEs. It is a web-based mapping application that you can interact with to view key information about GDEs, for example, type of GDEs, their status (known or potential GDEs) and degrees that potential GDEs are actual GDEs.

Although the functional methodology underlies the classification system of GDEs in the atlas. Doody et al. (2017) built the actual online GDEs atlas as a continental (or national, i.e., Australian) GDEs inventory of GDEs. They built the inventory following five iterative steps: (i) collated past studies and defined rules of dependency, (ii) collected field data, (iii) integrated field data with geospatial data (remote sensing and GIS), (iv) derived potential GDEs, and (v) drafted the atlas. In the first step, they conducted a national literature review of past GDEs studies (i.e., field-based and desk-top studies) and reviewed expert opinion about the interaction of ecosystems with groundwater in Australian landscape. Experts employed

the literature and their expertise to come up with 'rules of GDEs dependency' that provided a framework for identifying GDEs. The rules of GDEs dependency included provisions like: "vegetation that uses water besides rainfall is more likely to be using groundwater", "vegetation in areas of shallow groundwater is likely to be GDEs", and "where groundwater is shallow vegetation, water use cannot be satisfied by soil water, and therefore groundwater use is more likely". These rules of GDEs dependency helped when collecting field (e.g., evapotranspiration and rainfall) and geospatial (MODIS) data across Australia. Then, they integrated the data to define known GDEs and for modelling potential GDEs. Thereafter, produced a provisional GDEs draft for expert feedback. This process was iterative until experts reviewed the derived potential GDEs map/model and approved the final online GDEs atlas (Figure 2.2).[1] The Bureau often revises and updates the atlas as more studies are conducted at the national, regional and then local level.

[1]The Australian online GDEs atlas from the Australian Bureau of Meteorology

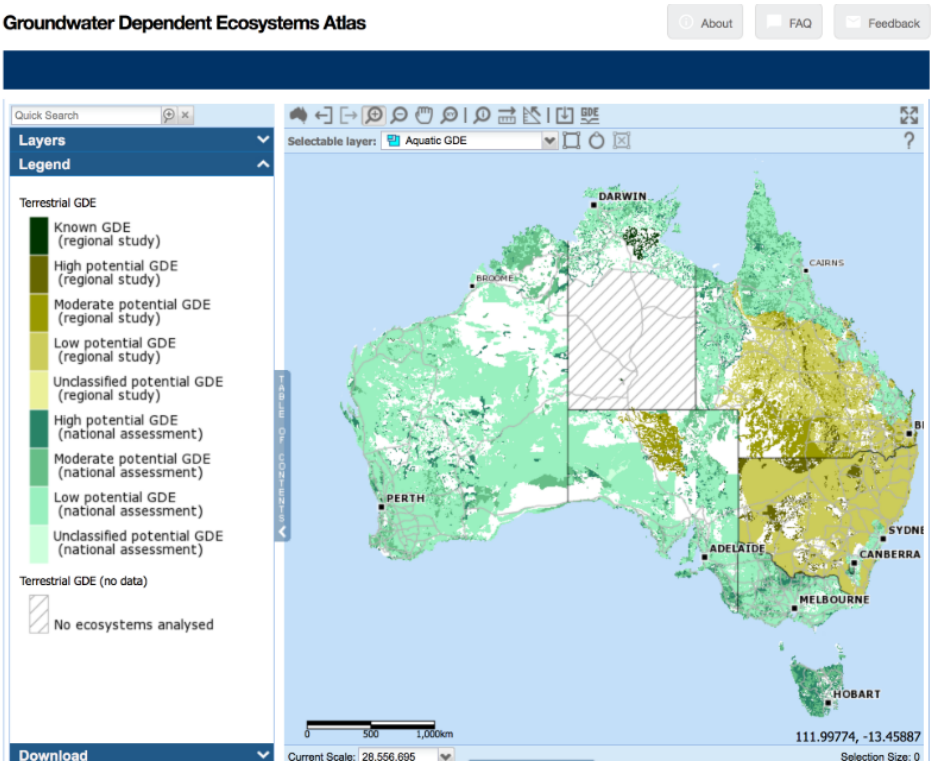

Figure 2.2: Australian national GDEs atlas. An online screenshoot from the Australian Bureau of Meteorology.

Countries often lack such thorough national inventories of GDEs, but conservation provisions are in place to preserve known and potential GDEs. In Europe, for example, member states protect GDEs under the European Union (EU) Water Framework Directive of 2000. It is a directive that obligates member states to safeguard the surface ecosystems with potential dependency on groundwater (Umweltbundesamt, 2011). Under the directive, the EU requires state members by law to allocate water to possible GDEs in the landscape. The Water Framework Directive of 2000 provisions share parallels with the ecological reserve provisions under the South African National Water Act (No. 36 of 1998). Ecological reserve provisions relate to the water required by ecosystems to survive and meet basic human needs before other purposes (e.g., industrial water use). It, the reserve, defines the potential dependency to groundwater instead of working out the water levels necessary to maintain GDEs. The latter requires data of known and/or

potential GDEs, but such data is limited. Colvin et al. (2007) developed a South African national probable GDEs map in key hydrological type settings. They proposed a national map with preliminary spatial data of probable GDEs and could be further revised to show probable GDEs at appropriate management scale (i.e., catchments). The preliminary GDEs map/model identified likely areas of both terrestrial and coastal GDEs as Aquifer Dependent Ecosystems (ADEs) based on environmental habitat, water source, and lithology as seen in Figure 2.3.[2]

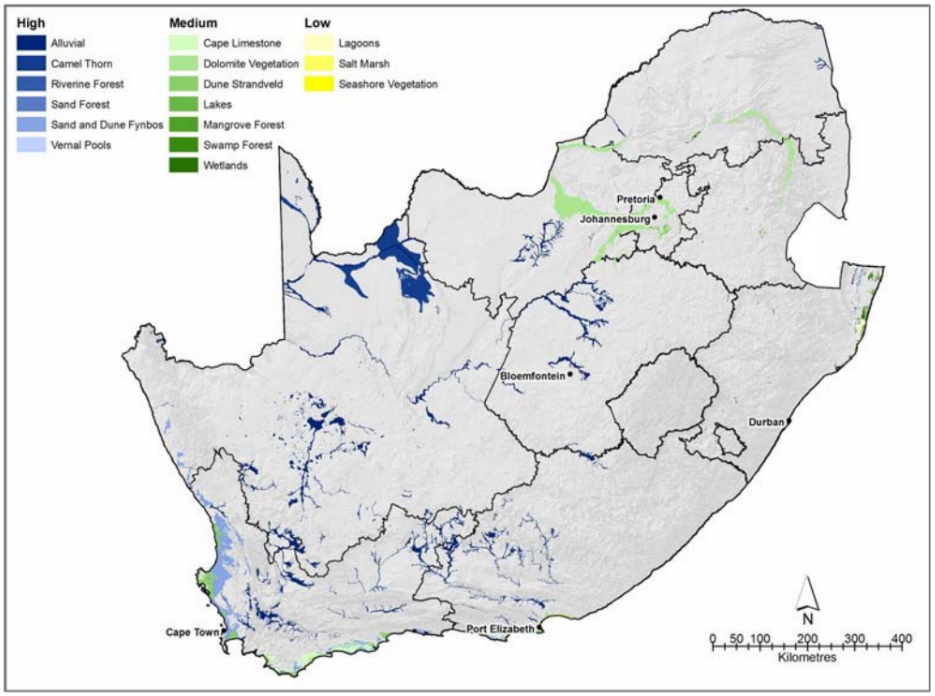

Figure 2.3: South Africa national probable locations of GDEs (Colvin et al., 2007).

2.3.4 Wetland GDEs

Existing GDEs classification systems put wetlands in the class of GDEs that require the surface expression of groundwater and label them as terrestrial GDEs (Eamus et al., 2016). Wetlands (or wetland GDEs) in the landscape typify the Green Island hypothesis. The Green Island hypothesis describes GDEs as areas that access

[2]preliminary map of where GDEs are likely to be in the country but far from being accurate at catchment and landscape scale Colvin et al. (2007)

groundwater and experience delayed drought effects compared to areas with no access to groundwater. Based on the Green Island hypothesis such ecosystems form pockets of moisture abundance and greenness and, are easier to find (Davies & Day, 1998). The Green Island hypothesis helps to identify locations of wetlands. Wetlands may form ecotones - transitional ecosystems from two extreme ends of the hydrological cycle, wet land, and dry land (Mitsch & Gosselink, 2000). Ecotones separate the dry and wet ecosystems. When wetlands form ecotones, they often appear disconnected from surface water bodies (Tiner, 2003). Such geographically isolated wetlands are potential GDEs, where existence of springs and/or seeps shows their usage of groundwater (Münch & Conrad, 2007). The International Convention on Wetlands, Secretariat (2010), defines them as:

"Areas of marsh, fens, peat land or water, whether natural or artificial, permanent or temporary, with water that is static or flowing fresh, brackish or salt including areas of marine water the depth of which at low tide does not exceed 6 metres".

Secretariat (2010) gives an agreed-upon and standardised method of defining wetlands. But their approach is unduly broad and not always relevant for delineating wetlands at a local scale. It also appears not to conform to the wetland ecological principle. The ecological principle poses that no two wetlands are the same, even if they appear in the same landscape (Klemas, 2011). Unlike the broad approach by the Ramsar Convention, Cowardin et al. (1979) accounts for both the ecological principle and Green Island hypothesis to define wetlands as:

"Areas where the water table is at near or above-ground long enough to promote formation of hydric soils that supports the growth of hydrophytes".

The characteristics of wetlands include:

(i) frequent water inundation-varying from wet to dry state

(ii) the presence or formation of hydric soils and

(iii) vegetation adapted to such conditions (e.g., phreatophytes or hydrophytes)

2.4 Geospatial technology in GDEs research

The Green Island hypothesis typified by wetlands is key to adopting geospatial technologies (e.g., remote sensing and GIS) to find areas of GDEs (e.g., wetland GDEs) in the landscape. Wetland GDEs as expressed by the Green Island hypothesis keep their moisture or greenness longer compared to their surroundings, even during prolonged droughts. With remote sensing data, such areas are detectable in the landscape. Based on the Eamus et al. (2006) classification system in Sub-section 2.3.3, remote sensing has potential to find classes (ii) and (iii) of GDEs, these are GDEs that: (ii) need the surface expression of groundwater (e.g., wetland GDEs) and (iii) rely on deep roots or capillary action. To find these classes of GDEs, remote sensing relies on vegetation proxy models to detect vegetation growth/abundance and/or water stress/water content to estimate the areas of GDEs. The primary vegetation proxy used to find groundwater dependent wetlands based on the green island hypothesis is the NDVI. NDVI use dates back to the 1970s and is the most investigated, and most suitable index in modelling photosynthetically active plant species (Frohn et al., 2012; Pettorelli, 2013).

NDVI characterises plant biomass based on solar energy that interact with plant species either with charts (e.g., tables, map) or quantitatively through GIS models. The model Equation (2.1) illustrates the quantitative approach to compute the NDVI, and the equation takes in electromagnetic (EM) spectrum wavelengths (or bands) sensitive to plant biomass. Plant species' reflectance properties depend on their physiology and pigmentation characteristics (Pettorelli, 2013). For example, most healthy plant species appear green because their chlorophyll pigment has a strong absorption in the blue-red wavelengths of the solar radiation energy spectrum. But the region of the EM spectrum bands applied to compute the NDVI is the Red (600-nm) and Near-infrared (NIR; 700-nm). As the measure of healthy vegetation, NDVI enhances the detection precision by applying the two bands (Red and NIR) as a ratio to extract the vegetation difference (Pettorelli, 2013).

The NDVI ratio is dimensionless with values between (-1 and +1), high NDVI values represent healthy vegetation while small values are typical of dry vegetation and bare soil. But such index proxy values still need field verification to make sure its

spatial accuracy in the landscape. It, NDVI, is the most dominant remote sensing index for measurement for vegetation greenness, but its spectral band combination limits it for sensing vegetation moisture. In its spectral band combination, only NIR wavelengths is sensitive to vegetation moisture. Gao (1996) developed the Normalised Difference Wetness Index (NDWI) to extract the vegetation wetness difference. The ratio, NDWI, applies the NIR and the Short-Wave Infrared (SWIR) region (Equation (2.2)). The SWIR is more sensitive to plant water content, but Liu et al. (2004) proposed that the Red-edge (650-nm), although outside the normal water content bands, is far more accurate in measuring plant water content. But the NDWI is well established in the research literature as the plant water stress/content proxy. The value range of NDWI is like that of NDVI (-1 and +1), where positive index values represent an accurate identification of high moisture healthy vegetation. The negative values of the index represent other surfaces (e.g., dry vegetation and bare soil).

$$NDVI = \frac{(NIR - R)}{(NIR + R)}, -1 \geq NDVI \leq 1 \tag{2.1}$$

NIR = Near-infrared radiance,

R = Red spectral radiance

$$NDWI = \frac{(NIR - SWIR)}{(NIR + SWIR)}, -1 \geq NDWI \leq 1 \tag{2.2}$$

NIR = Near-infrared radiance,

SWIR = Short-Wave Infrared radiance

2.5 Review of geospatial technology reasearch on GDEs

The rest of the chapter presents a review of previous research using remote sensing indices to find GDEs. In GDEs research, the use of indices is not well settled in the literature, but it is expanding as more studies are being published on the subject. So, this review included studies within the growing published research pool. The research reviewed focuses on how to find the locations of GDEs through remotely sensed data and GIS. The aim was not to examine everything but unpack the current landscape in the advances of remote sensing indices in GDEs research and identify knowledge gaps to contribute towards the proposed novel approach.

2.5.1 Review of GDEs research

In EllenBrook, Western Australia, Barron et al. (2014) employed the multi-spectral approach with both NDVI and NDWI indices derived from Landsat to identify potential GDEs. They developed a new method to identify potential GDEs: Groundwater Dependent Ecosystem Mapping (GEM). The GEM method assumed areas that sustained their greenness and moisture content during a prolonged drought are potential GDEs. It, GEM, was a two-step verification method. The first step comprised supervised classification with the NDVI and NDWI, and the second was ground-truth verification using maps and aerial images. In the first step, the GEM method classified potential GDEs in the study region into five land cover classes (LCs) based on their response to limitations in groundwater:

(i) LC1 “not drying”: areas with vegetation that have permanent access to groundwater, so they maintain their greenness and moisture content during and after a prolonged drought.

(ii) LC2 “slow drying”: without precipitation groundwater levels eventually subside, unlike non-GDEs, these areas have a delayed response to drought because groundwater takes longer to subside compared to the dramatic response of surface water bodies to drought.

(iii) LC3 “fast drying”: areas that rely on soil moisture storage.

(iv) LC4 (water). Area that are permanent water bodies and may have groundwater connection through base-flow.

(v) LC5 (non-GDEs): areas that show low or no greenness.

Finally, the ground-truth verification step defined three classes of potential GDEs in the study region: (i) GDEs with permanent access to groundwater, (ii) GDEs with diminishing access to groundwater, (iii) GDEs that are permanent water bodies. Overall, the GEM method is a simple implementation of the Green Island hypothesis to identify potential GDEs. Key potential GDEs found by the GEM method also coincided with known GDEs. GEM method provides a simple framework to detect GDEs by employing the main remote sensing indices for vegetation health and vegetation water content, NDVI and NDWI, respectively. Although, GEM provides a simple framework, but the applied indices produce inferential results. These results often require further field verification and long-term monitoring to move beyond inferences.

Lv et al. (2013) integrated NDVI and field groundwater level data to find the areas of potential GDEs in the semi-arid Hailiutu River catchment of Northern China. They calculated groundwater levels to create groundwater potential maps/models and correlated them with seasonal NDVI derived from Landsat-5 Thematic Mapper (TM) spatial imagery to locate potential GDEs in the catchment. The maximum NDVI value obtained in the dry summer imagery > 0.6 with a mean of 0.29, and these NDVI values correlated with groundwater depth of less than 10 m below ground. The results showed that areas with high NDVI values when correlated with shallow groundwater depths were likely to be potential GDEs. This research assessed NDVI changes as a function of groundwater levels and verified the results in the field that NDVI proved a reliable indicator of vegetation cover. The areas with greater vegetation cover coincided with shallow groundwater depths and were then predicted as potential GDEs. They concluded that NDVI as a function of groundwater depth was effective and showed that potential GDEs are likely to be in areas of shallow groundwater depth. But, if there is a lack of groundwater depth data, this approach may not work, notably in hard-to-reach remote areas.

Where groundwater depth data exist, this method has merit for use and generate reasonable accuracy for identifying potential GDEs.

With groundwater depth data from 13 boreholes, Jin et al. (2011) applied the method in the Ejina catchment, China. They also correlated MODIS imagery NDVI to assess the spatial distribution of possible groundwater dependent vegetation (i.e., grassland, scrubland and woodland). In the three-vegetation class, higher NDVI values did not correspond with shallow groundwater depths. In groundwater depths between 2.5–3.5 m, vegetation cover was observed, but groundwater depth > 5.5 m NDVI showed a low vegetation cover. This study showed NDVI can be integrated with groundwater depth to detect the presence of vegetation cover. The access to groundwater depth data is key to this method but is often nonexistent. Accurate groundwater depth data requires field acquisition, but the data might be nonexistent because of several reasons (e.g., hard-to-reach sites, expensive or laborious to instal recording equipment). Fieldwork is extensively reduced with applications of remote sensing indices like the NDVI, but this method still requires field groundwater depth data.

The reviewed studies apply spectral analysis, which measures GDEs as pixel objects. But the spectral analysis approach ignores the natural spatial distribution of GDEs. The approach presumes a pixel represents a phenomenon, but a more practical method considers other elements of an image such as texture and patterns to extract actual objects (Blaschke, 2010).

Guirado et al. (2018) integrated the pixel approach with the object-based analysis to detect areas of potential groundwater dependent vegetation in the Cabo de Gata-Nijar catchment, Spain. They used land fractures data and aerial photograph imagery to identify *Ziziphus* shrubs coinciding with fractures as potential GDEs. They collected the imagery using an RGB (Red Green Blue) H4D Hasselblad camera aboard a Helicopter flying at 550 m above-ground. And the software, eCognition (v.8.9) to segment the image scenes into recognisable *Ziziphus* shrubs based on the object-based image analysis approach. They superimposed the recognisable *Ziziphus* shrubs with a map of known geographical faults (or fractures) in the landscape. The fractures map showed points of weakness in the landscape which

might encourage groundwater discharge or recharge. The research supported the claim that areas of GDEs coincide with areas of seeps or faults showed by fractures in the landscape (Münch & Conrad, 2007). They further supported the claim with seasonal NDVI derived from Sentinel-2A and used to analyse the relationship between *Ziziphus* patches and seeps and/or faults. *Ziziphus* patches with > 0.5 NDVI values and within < 50 m of seeps and/or faults maintained their greenness during wet seasons and prolonged dry summers and were then predicted to be potential groundwater dependent vegetation.

The two major approaches behind the studies reviwed so far have data requirements that might not be available in other areas. The first, NDVI multi-temporal analysis is the principal approach adopted in the literature to identify GDEs. Remote sensing multi-temporal analysis often requires available data (e.g., freely available satellite images). White et al. (2016) in the Great Artesian Basin, Australia used a simple NDVI analysis to show the effectiveness of multi-temporal analysis to detect groundwater dependent wetland vegetation. They extracted the NDVI for the analysis from Quick-bird and World View-2 Very High-resolution imagery for both the dry and wet seasons. Multi-temporal NDVI was correlated with field data (i.e., soil and springs or seeps flow measurements). Field measurements showed that areas of high NDVI values also maintained their moisture levels throughout wet and dry seasons. The research also established NDVI thresholds (0.37 - 0.51) for delineating potential groundwater-fed vegetation.

2.5.2 GDEs research in the Cape Floristic Region (CFR)

Münch & Conrad (2007) defined potential GDEs using the multi-temporal analysis method in three catchments of the West Coast (northern Sandveld) in the CFR. Their multi- temporal analysis approach relied on the Green Island hypothesis to show the probable presence of GDEs in the landscape. They applied the GIS model; landscape wetness potential model based on terrain features and augmented it with the NDVI, and Tasseled Cap (TC) transformation obtained from Enhance Landsat Thematic Mapper (30 m). The TC and NDVI determined wetland vegetation seasonal variations during a dry and wet year, and wetness index defined their (i.e.,

wetland vegetation) potential to be GDEs. They classified land cover as either probable areas of groundwater dependence or 'none' GDEs. This study showed integrated remote sensing and GIS applicability to find GDEs and concluded by producing a regional map of GDEs.

The CFR has vegetation that adapt well to fire disturbances. Its (i.e., CFR) recurrent fires act as more than disturbances, but is integral to regulating the hyper-diverse and endemic fynbos vegetation in the region (Bond & Keeley, 2005). Life cycle of fynbos vegetation is inter-linked to the fire regime. Fires serve as soil-mineralizing agents and maintain vegetation diversity (Rutherford et al., 2011). The fynbos diversity relies on the harmonious relationship with fire. But this harmony rest on a balance due to increased sensitivity to fire disturbances caused by the changing climate (Slingsby et al., 2017). A chunk of research done in the CFR focuses on post-fire regeneration and resilience as it affects vegetation diversity.

Wilson, Latimer & Silander (2015) leveraged the many fires to model and map post-fire recovery at a regional scale employing the remote sensing index, NDVI. They extracted the NDVI from MODIS imagery with a 10-year spatial coverage to simulate post-fire vegetation regeneration. They modelled post-fire vegetation recovery with climate, topography, and soil as predictors across fires of different ages and, found that depending on the environmental condition (e.g., model predictors; climate, topography, and soil) post-fire recovery varied from < 5 to > 25 years. More, they found that moisture availability was significant to the vegetation recovery rate, with dry parts of the regions experiencing slower recovery rates compared to those with more moisture. The moisture presence in quicker post-fire recovering areas confirmed the relevance of the Green Island hypothesis. Elsewhere Chen et al. (2011); Díaz-Delgado, Lloret & Pons (2004); Escuin, Navarro & Fernández (2008) and Hope, Albers & Bart (2012) also proved its applicability in post-fire conditions. In post-fire conditions, the Green Island hypothesis might curtail the demand for laborious field data to find GDEs, but past studies have not applied it. This dissertation contributes towards this knowledge gap by presenting a proposed method that applies this hypothesis to find GDEs. Münch & Conrad (2007) improved spatial data of GDEs in the CFR but across the region little is

known about GDEs at management scale (i.e., catchment scale; Colvin et al., 2007). The dissertation further serves as a contribution towards improving spatial data of GDEs in the CFR.

Chapter 3

Methodology

3.1 Methodological approach

The main impetus behind this dissertation was to leverage the frequent fires in the CFR to identify wetlands with rapid post-fire recovery. The dissertation claim was that rapid post-fire recovering wetlands are potential GDEs. Such wetlands were identified by proposing a novel approach: Post-fire NDVI analysis. Considering the recurrent fires in the CFR, post-fire recovery could be key in finding groundwater dependent wetlands. The post-fire NDVI analysis method leveraged the recurrent fires in the CFR. After developing the method, the machine learning algorithm random forest, was run for predicting the spatial distribution of potential wetland GDEs. The study implemented the novel approach and machine learning with quantitative remote sensing and ground-based geospatial data. The geospatial dataset employed was taken from secondary sources.

3.2 Study region - Kogelberg and Steenbras Nature Reserves

After developing the method, the supervised machine learning algorithm random forest, was applied to predict the spatial distribution of potential wetland GDEs.

The random forest and the novel approach were both implemented by employing secondary geospatial data. The secondary geospatial dataset comprised remote sensing data and ground-based spatial data (Table 3.1). Ground-based spatial dataset comprised the fire disturbance data and wetland boundaries data, and remote sensing data composed of aerial images and sentinel- 2A images. The study integrated the orthoimages, both aerial and sentinel-2A, with the ground-based dataset to implement the proposed novel post-fire NDVI analysis and machine learning methods. Aside from sentinel-2A, the secondary data was gathered from an ongoing wetland monitoring project TMG Aquifer System. The current monitoring in the TMG entails acquiring aerial imagery (1 m) from limited field surveyed wetland sites in the CFR including Kogelberg, Steenbras, Palmiet, Villiersdorp and Wemmershoek regions.

Table 3.1: The geosptial data employed in the study.

Data	Source
Fire disturbance data	Western Cape Nature Conservation Board (WCNCB, CapeNature)
Aerial imagery	Table Mountain Group (TMG) Aquifer System wetland monitoring project
Wetland boundaries	Table Mountain Group (TMG) Aquifer System wetland monitoring project
Sentinel-2A images	ESA, Copernicus Open Access Hub

In the regions of the CFR with the current monitoring, Kogelberg and Steenbras Nature Reserves were chosen for this study. The Reserves have three wetland sites under monitoring, two of which experienced fire disturbances during the monitoring period. Leveraging fire data is key to this study since the aim is to identify and map potential GDEs by proposing a novel approach that leverage the recurrent fires in the CFR. For ease of reference, henceforth the wetland study sites are referred to as I, IIa and IIb (Table 3.2). Two of the wetland study sites recorded fires on 16-March 2011 (site IIa) and 03-June 2010 (site IIb). Site I had no fires and hence selected as a reference site. But the fire database shows that the larger areas of the reference site was burned at least 3 years (06-December 2008) before starting with the aerial imagery acquisition. The first seasonal imagery, 27-March 2011, showed no sign of a fire disturbance.

For all three wetland sites, two adjacent non-wetland classes (or controls) were selected just outside the perimeter of the known wetland (Figure 3.1). The extent indicators zoomed-in to the wetland study sites, from bottom-right, bottom-left and middle-left: wetland study site I, IIa and IIb. Table 3.3 illustrates the training classes (i.e., wetland and non-wetlands) by their location, study site, and their area size differences.

Table 3.2: Aerial imagery and fire disturbances in the three wetland sites.

Sites	UTM.X	UTM.Y	Fires
site I	-8614.555	-3783777	NA
Site IIa	-6435.521	-3795951	16-Mar 2011
Site IIa	-3629.143	-3800055	03-Jun 2010

Figure 3.1: Kogelberg and Steenbras Nature Reserves indicating location of known wetlands and adjacent controls. Wetland classes are illustrated by solid white lines and adjacent non-wetland classes by solid black lines. From Bottom right, bottom left and middle left respectively: I, IIa and IIb. Study sites I and IIa are zoomed-in at scales 1:4 000 and site IIb at 1:8 00.

Table 3.3: List and areal size of known wetland classes and selected non-wetland classes.

Class	Area (m^2)
Site I	
Wetland	1076.597
Control A	652.5783
Control B	750.8505
Site IIa	
Wetland	1429.543
Control A	1012.48
Control B	1253.499
Site IIb	
Wetland	3313.409
Control A	2056.7
Control B	610.3947

3.3 Geospatial data

The geospatial dataset covered the training sites, wetland study sites, to implement the proposed novel post-fire NDVI and thereafter, spatial predictive mapping with machine learning. In its entirety, the dissertation was written in R using the bookdown package with the R Markdown-syntax (Xie, 2020). The geospatial dataset in this dissertation was processed and analysed using the free software environment and programming language R (R Core Team, 2021). Other packages used to pre-process the data include ArcGIS Pro and Sentinel toolbox (viz., Sen2Cor, Sentinel Application Platform; SNAP).

3.3.1 Obtaining fire disturbance data

The fire disturbance data was downloaded from the South African National Biodiversity Institute (SANBI) online spatial data portal Biodiversity GIS (BGIS). The BGIS portal house the data, but the Western Cape Nature Conservation

Board (WCNCB; CapeNature) are the custodians. Access to the fire database from (CapeNature, 2018) is hyperlinked to the phrase BGIS whereever it appears in this dissertation.

CapeNature uses it as a fire management tool in all their Nature reserves in the CFR. They maintain it as a multi-polygon geometry file. It has recorded yearly fires in the CFR, with fires dating back as distant as the year 1927. For this study, the latest update of 2018/19 fire data was downloaded. The downloading process entailed just searching for "CapeNature fires" in the BGIS portal. Selected fires burned in the wetland study site (Table 3.2) and took place during or near to the start of monitoring period by the TMG project. The fire data (or database) provides the outermost fire boundaries. As such, the TMG project aerial images described bellow along Google Earth time series imagery on Esri's ArcGIS v10.2 were used to verify that burned and unburned areas were actually burned or unburned.

3.3.2 Obtaining aerial images data

The wetland monitoring project in the TMG is an interdisciplinary effort surveying and monitoring surveyed wetlands. For this study, selected wetland study sites are in the Kogelberg and Steenbras regions of the CFR encompassed by the greater Kogelberg Nature Reserve (Figure 3.1). The TMG project started acquiring aerial images (i.e., end of summer and end of winter) in 2011-13. After this period, the project continued but shifted to an annual acquisition. From personal contacts: at the outset of the project, they aimed to analyse wetland vegetation, end of winter and end of summer, end-of-wet and end-of-dry seasons, respectively. They derived NDVI for the two seasons as end-of-wet and end-of-dry NDVI images. Their NDVI analysis revealed no significant variations in wetland vegetation in the two seasons (i.e., end-of-wet and end-of-dry seasons). Thereafter, the project turned to annual acquisition to curtail the cost of acquiring seasonal aerial images. From the TMG project, this dissertation employed the seasonal dataset (Table 3.4) to implement the proposed post-fire NDVI analysis method since some areas in the CFR have been shown to experience rapid post-fire recovery (i.e., less than a year post-fire). From the TMG project the images were already calibrated, orthorectified and radiometrically corrected.

Table 3.4: Seasonal aerial orthoimages data employed to compute Normalised Difference Vegetation Index (NDVI) for post-fire NDVI analysis.

Date	Time	Season	Region	Sites
27-Mar-11	12:04	end of summer	Steenbras	I
27-Mar-11	12:18	end of summer	Kogelberg	IIa & IIb
07-Nov-11	10:32	end of winter	Kogelberg	IIa & IIb
07-Nov-11	10:53	end of winter	Steenbras	I
20-Mar-12	11:01	end of summer	Kogelberg	IIa & IIb
20-Mar-12	11:14	end of summer	Steenbras	I
18-Nov-12	11:14	end of winter	Steenbras	I
18-Nov-12	11:27	end of winter	Kogelberg	IIa & IIb
02-Mar-13	10:51	end of summer	Steenbras	I
02-Mar-13	11:02	end of summer	Kogelberg	IIa & IIb
23-Nov-13	11:22	end of winter	Steenbras	I
23-Nov-13	11:35	end of winter	Kogelberg	IIa & IIb

Note:
South African Standard Time (SAST).
[1] The day, time, season, and region where the images were acquired.

3.3.3 Obtaining of sentinel-2A data

The sentinel-2A (S2A) satellite imagery was downloaded from the European Space Agency (ESA). You may click Copernicus Open Access Hub to access the Geohub. The procedure to obtain involved five basic steps: step one defined the region of interest- the Kogelberg Nature Reserve. This step focused the search on the global archive to the region of interest - Kogelberg Nature Reserve. Then, step two defined the sensing period. The sensing period was 2017-18, expecting to capture the effects of the recent drought. The Sentinel program monitor such natural phenomena with both Synthetic Aperture Radar (SAR) and optical imaging (Copernicus, 2017). SAR imagery has several advantages over optical imagery (e.g., collects data day and night while latter requires solar energy). But step three selected optical imaging over SAR for its broad spectral coverage. Like all active sentinel missions, Sentinel-2 is a constellation of two satellite sensors: S2A and Sentinel-2B (S2B). However, Sentinel-2B orthoimages were unavailable in December 2017 because of a payload

irregularity (Copernicus, 2017). Without S2B sensor, the fourth step selected S2A satellite. The fifth and ultimate step determined the acceptable cloud cover-cloud-free (0%) orthoimages.

Table 3.5: The file naming system of the Sentinel-2A (S2A) product. The naming system details the nature of the retrieved orthoimages.

Product	Description
S2A	Sensor ID
MSIL1C	Reference to the level of processing
20180327T081601	Sensing start time, Coordinated Universal Time (UTC)
N9999	Production baseline number
R121	Relative orbit number
T34HCH	Absolute orbit number (e.g 34 is a UTM zone for the study region)
20191215T193704	Sensing stop time (UTC)
SAFE	File-name extension (Standard Archive Format for Europe)

The downloading procedure retrieved cloud-free S2A orthoimages, Level-1C images. S2A was launched in 23-June 2015 and its products (Level-1C orthoimages) were available in the latter parts of 2016 for the study region (Kogelberg and Steenbras Nature Reserves). For this study, S2A orthoimages were downloaded for the period 2017-18 coinciding with a drought period in the CFR. The timeline was selected based on the hypothesis that GDEs maintain their greenness/moisture relatively longer than their surroundings, even in during prolonged drought periods. For the selected timeline (2017-18), available cloud-free multi-date S2A orthoimages were downloaded. The dates approximately coincided with the end-of-dry season: 02-March 2017, 27-March 2018 and end-of-wet season: 17-November 2017 and 23-October 2018. Level-1C orthoimages are packaged as a compressed Standard Archive Format for Europe (SAFE) file. SAFE files contain 13 Joint Photographic Experts Group (JPEG)-2000 orthoimages (or spectral bands). The file naming system shows details about the orthoimages and how they were collected by the satellite sensor. The details including the imagery level of processing (e.g., S2A_MSIL2A_20180327T081601_N9999_R121_T34HCH_20191215T193704.SAFE and Table 3.5). For example, Multi-spectral Instrument (MSI) in "MSIL1C" refers to the sensor payload and Level-1C (L1C) to the level of processing (Drusch et al., 2012). ESA calibrates S2A spectral bands to radiance and transforms them into reflectance orthoimages shown by "Level-1C". But their reflectance is top-of-atmosphere reflectance (TOA) containing values of interference like haze,

clouds, and atmospheric aerosols. The S2A Level-1C orthoimages, TOA, were pre-processed using Sen2Cor v2.8 algorithm, a Sentinel-2 Toolbox. The input in Sen2Cor was the compressed "SAFE" file and performed atmospheric correction. After the atmospheric processing, the level of processing changed from Level-1C, TOA, to Level-2A showing Bottom of Atmosphere reflectance (BOA). BOA dataset contains minimal atmospheric interference and was the data used for predicting the spatial distribution of GDEs.

3.4 NDVI from aerial orthoimages data

The aerial orthoimages dataset from the TMG project (Table 3.4) was employed to generate the post-fire multi-date NDVI data samples for the wetland study sites (or training sites). The mechanics of generating NDVI has been presented and explained in Section 2.4. Using the NDVI Equation (2.1), the R package "raster" was adopted to create a nested function for generating NDVI for the study period (i.e., 2011-13; Hijmans et al., 2015). For each training site (n=3) the nested function enabled cropping and masking of the NDVI to the training wetland and non-wetland classes (n=9). Thereafter, the results for each training class were merged into a single Table 4.1 showing the NDVI data sample by training site.

3.5 The novel post-fire NDVI analysis

The novel post-fire NDVI analysis method employed the NDVI data sample (Table 4.1). Its goal, novel post-fire NDVI analysis method, was to answer the question: Can we detect more rapid recovery of NDVI in wetlands than neighbouring vegetation? It addressed the question in three levels of analysis: (i) determine if there is a difference between wetland and non-wetland classes in their post-fire NDVI, (ii) defined a quantitative NDVI baseline as a proxy for delineating wetlands, and (iii) analysed post-fire NDVI seasonal distribution to assess the potential of delineated wetlands to be GDEs.

In the first level of analysis, one-way Analysis of Variance (ANOVA) tool was adopted to deal with the question "Is post-fire NDVI sensitive enough to detect

rapid recovery in wetlands predicted to be GDEs?" Input data for the ANOVA was the NDVI data sample. The mean NDVI was the dependent variable, and the independent variable was the training class (or groups/levels): known wetland (n=1) and non-wetland (A and B; n=2) classes in each of the two burned wetland study sites (IIa and IIb), and the unburned reference site (site I). ANOVA tested the null hypothesis that wetland and non-wetland classes have similar post-fire NDVI spectral signatures. The alternative hypothesis was that wetland and non-wetland classes have different post-fire NDVI spectral signatures. The null and alternative hypothesis restated in symbolic forms are:

$H_0 : \mu_w = \mu_a = \mu_b$

$H_1 :\neq ...\mu_w = \mu_a = \mu_b$, $(\neq)$ indicating that at least one of the three land "classes" is different (sensitivity of post-fire NDVI).

μ_w : wetland mean NDVI

$\mu_a = control_A$: non-wetland (A) mean NDVI

$\mu_b = control_B$: non-wetland (B) mean NDVI

The ANOVA tool performs a hypothesis test by conducting a statistical significance test: F-test (or F-ratio) named after the developer Ronald Fisher (Hector, 2015). The F-test (or F-ratio) evaluates the statistical difference between mean groups of independent variables (s). With the F-ratio defined as the statistical signal, $MS_{Between}$ Group Variance to statistical noise, MS_{Within} Group Variance ratio (Hector, 2015).

$$F = \frac{MS_b}{MS_w} \tag{3.1}$$

MS (b&w) = mean square between and within groups.

A larger F-ratio increases the confidence that a distinction between the group means is real (Hector, 2015). The F-value, and related degrees of freedom (df), are incorporated to work out a p-value that defines statistical significance. A p-value < 0.05 is sufficient evidence to reject the null hypothesis. A false (or rejected) null hypothesis pointed towards post-fire mean NDVI being sensitive enough to discriminate between wetlands and non-wetland classes. but it does not specify

which means differ from each other. As the second level of analysis of the novel approach, a post-hoc test: Tukey's Honest Significant Differences (HSD), was applied to compare and infer which mean NDVI groups (or training class) were significantly different. The Tukey HSD pair-wise comparison inferred a mean NDVI baseline (or thresholds) for delineating wetland classes.

3.5.1 One-way ANOVA assumptions

The subsequent levels of analysis in the proposed novel approach relied on the assumptions of ANOVA to be true. But if the data did not fulfil these assumptions, that nullified the test. ANOVA makes three key assumptions: data carries a normal distribution, has a homogeneity of variance within each group, and comprises independent observations (Sawyer, 2009). The simplest technique to check the first assumption involves drawing a histogram. A histogram of a data with normal distribution will display a bell like shape (Sawyer, 2009). While homogeneity of variance means the histogram distribution curves are similar and have the same standard deviation. The independent observations assumption might simply require random sampling of the data.

Assumption of independence was controlled for by randomly selecting adjacent non-wetland classes and ensuring that their pixels do not overlap with each other and with the known wetland pixels. However, the observations (NDVI values) in each pixel class are spatially autocorrelated. The spatial autocorrelation is an issue here because the observations do not conform to the assumptions of ANOVA. Since the data show clear boundaries between wetland and non-wetland pixels, the test was still considered valid. Then, temporal aggregation mean NDVI data for the three wetland study sites was checked for the homogeneity of variance with the Levene's test of variance (Fox, 2015). A p-value < 0.05 in Levene's test violates the homogeneity of variance assumption. In all three wetland sites, site I ($p = 0.84$), site IIa ($p = 0.79$) and site IIb ($p = 0.89$) homogeneity of variance was not violated. Last, the normal distribution assumption was checked by examining the NDVI density curve distributions and they largely resembled a bell like shape as expected for normally distributed data. But such visual representation might be unreliable, so Shapiro-Wilk test of normality was applied. The Shapiro-Wilk

test is a significance test for normal distribution where (p-value > 0.05) signifies a normal distribution and confirms normal distribution (Mohd Razali & Yap, 2011). A Shapiro-Wilk test (W = 0.93, p-value = 0.27) confirmed the normal distribution of the mean NDVI data in wetland study site I and wetland study site IIb (W = 0.93, $p = 0.27$). But Shapiro-Wilk test was violated (W = 0.88, $p = 0.035$) in wetland study site IIa, the site that burned 11-days before obtaining the initial NDVI. Given the drastic drop in NDVI data post-burn, violating the Shapiro-Wilk test was expected.

3.5.2 Post-fire/multi-date NDVI distribution

The third and last level of analysis of the proposed post-fire NDVI analysis method was to analyse the underlying NDVI distributions of delineated wetland classes to assess their potential to be groundwater dependent. A kernel density estimator function, through the R package ggplot2 (Wickham, 2016), was utilised to illustrate (density plots) multi-date distribution of post-fire NDVI in the two burned sites. Instead of a histogram that just bins the multi-date NDVI distribution, the kernel estimate function computed the underlying NDVI density distribution. The goal was to visualise the NDVI multi-date distribution post-fire to interpret the underlying NDVI density curves of wetland and non-wetland classes and assess their post-fire recovery rate.

3.6 Random forest with sentinel-2A data

The spatial modelling of potential wetland GDEs still relied on the Green Island hypothesis but with Sentinel-2A instead of the aerial images and the fire data. These data were limited in their scope. First, the aerial orthoimages were not radiometrically normalised to get the accurate surface reflectance of wetland biomass in the study region. Second, their spectral bands did not extend to the Red-edge and SWIR range, which measures the spectral signature of plant water molecules (or plant water content). The next best thing was S2A data. S2A has a broad spectral band extending to the Red-edge, SWIR range and significantly, in its class of sensors, mid-resolution, it has the highest spatial resolution and global spatial

coverage. In the time S2A data was accessible, the wetland study sites had no fires, so instead of the post-fire NDVI approach, the assumptions based on the Green Island hypothesis applied to guide the spatial prediction of GDEs with random forest. The random forest algorithm was trained with S2A spectral data. The criteria of inclusion for spectral bands was their sensitivity to wetland plant species (or wetland biomass) and water stress (or plant water content). Based on the defined inclusion criteria, the selected spectral bands comprised SWIR (B11), Red-edge (B05), Near-infrared (NIR; B08), Red (B04), and (B03) Green (Table 3.6). In the Green band (and to a slighter extent the Red band) the chlorophyll in normal plant species reflect a lot of Electromagnetic radiation energy, and appropriately, appears green. But particularly the Red and NIR spectral bands that are established in remote sensing research for discriminating healthy plant species. Healthy (or photosynthetically) active plant species are negatively correlated with water stress (Liu et al., 2004). They are measured with the Plant Water Content (PWC) 900-2500-nm EM energy spectral bands. But Liu et al. (2004) found Red-edge (680-nm) outside the PWC to be a more reliable indicator of plant species water content. Both the SWIR (1500-nm) and Red-edge (680-nm) spectral bands were selected to increase the predictors that measure biomass in wetland GDEs. Gao (1996) in the PWC bands applied SWIR and combined it with NIR to develop the NDWI to measure vegetation water difference. Compared to NDVI, the NDWI less sensitive to atmospheric scattering effects (Adam, Mutanga & Rugege, 2010). So here both indices were generated with sentinel toolbox SANP to integrate each other and enhance the measuring wetland GDEs biomass spectral signatures.

Table 3.6: Selected S2A spectral bands and indices.

Dataset	Description	Indicator	Resolution
NDVI	Normalised Difference Vegetation Index	Plant health	10 m
NDWI	Normalised Difference Wetness Index	Water stress	20 m
B03	Green	Natural green	10 m
B04	Red	Plant health	10 m
B05	Red Edge	Plant water content	20 m
B08	Near Infrared	Plant health/water content	10 m
B011	Short Wave Infrared	Plant water content	20 m

Note:
Low spatial resolution (i.e., 20 m) data was scaled to match S2A higher resolution (10 m).

The indices (i.e., NDWI and NDVI) and selected single spectral bands were stacked together using the raster package by Hijmans et al. (2015), to create a raster stacked object data (Table 3.6). The multi-date data comprised seven predictor variables (or bands) in each training classes (or binary response), wetland or non-wetland classes. So, the predictors in the two-potential wetland GDEs, identified by the post-fire NDVI method, plus the known wetland class in the reference site (included to increase the training data) comprised the multi-date data employed to train the random forest. The sample size (i.e., number of pixels) of multi-date data after removing NA (empty pixels) was 135 pixels. The 135 pixels comprised the binary response (i.e., wetland or non-wetland classes), and Appendix A has the complete multi-date data (summer and winter seasons) for 2017-18 employed to train the random forest algorithm to predict the spatial distribution of GDEs in the study region.

Machine learning random forest algorithm (or classifier), unlike traditional maximum likelihood, is intuitive, makes no assumptions about the data and requires less time to train and perform spatial predictions (Pal & Mather, 2003). Tree-based classifiers unlike other machine learning algorithms can handle qualitative predictors and are easier to explain as they adopt a method analogous to the growth of trees (James et al., 2013). Random forest is an ensemble (i.e., aggregate of decision trees) tree-based classifier and outperforms the decision tree classifier including other well-established ensemble methods (e.g., SVM; Schratz et al., 2019).

3.6.1 Mechanics of Random Forest

The ensemble random forest has two types of algorithms, classification and regression. This study applied the classification algorithm to classify the categorical binary response (wetland or non-wetland classes) of the multi-date data. The main mechanic behind the random forest classifier is to aggregate decision trees into a forest. It aggregates the decision trees through an approach called bagging (or bootstrap), partitioning the training data into two samples: In-bag and out-of-bag samples (Breiman, 2017). The random forest classifier uses the former to train the trees on, and the latter applies an internal cross-validation and estimates the random forest model performance (Belgiu & Drăguţ, 2016). Literature refers to the estimated error as Out-of-Bag (OOB) error. Random forest classifier has two key hyper-parameters: randomly selected user-defined number of features (m_{try}), defined as a square root of the training data variables $m_{try} = \sqrt{p}$, for splitting the decision trees at each node and the number of trees ($Ntree$) in the forest (Breiman, 2017). The algorithm then averages out the class assignment probabilities. Thereafter, the forest applies an internal cross-validation and decision trees in the forest vote for a class membership, and it will choose one with maximum votes (Maxwell, Warner & Fang, 2018). The research is not clear about the optimum (m_{try}) value and the number of trees enough to produce an optimal random forest classifier. However, several studies have identified that the algorithm requires several trees, after which random forest classifier accuracy plateaus (Maxwell, Warner & Fang, 2018). In remote sensing, Belgiu & Drăguţ (2016) suggest 500 is the appropriate number of decision trees. But likely that ($Ntree$) is case specific (Maxwell, Warner & Fang, 2018). So, random forest hyper-parameters are optimised on a case-by-case basis. In such cases, the statistical software R Core Team (2021) provides several packages to optimise/tune the random forest hyper-parameters.

The several tuning methods in R include K-fold Cross-Validation (CV). CV is an interesting approach in small data samples like the one employed in this study since it uses the complete dataset in model training (Maxwell, Warner & Fang, 2018). The approach randomly splits the data into k disjunct subsets, before running the classifier k times. In each run one subset is withheld for validation (Lovelace, Nowosad & Muenchow, 2019). CV assumes statistical independence,

randomly partitioning the data into training and test data to ascertain reduced bias-assessment of the generalisability of the classifier to unseen (independent test data). Considering the multi-date data in this study is a geospatial dataset governed by the first law of geography that states: close objects are more similar than those farther apart (Miller, 2004). So, Schratz et al. (2019) proposed a spatial component to the conventional CV approach. Spatial Cross-validation differs from conventional CV by spatial partitioning, splitting the data into spatially disjointed subsets (Lovelace, Nowosad & Muenchow, 2019). The effect of applying spatial cross-validation considers the spatial nature of geographic data, unlike the conventional cross-validation approach of random selection (Figure 3.2). So, the random forest classifier in this study adopted a five-fold repeated spatial cross- validation tuning (SpRepCV) to optimise the hyper-parameters. The spatial component to the multi-date data was added by incorporating the geometry metadata of the binary response.

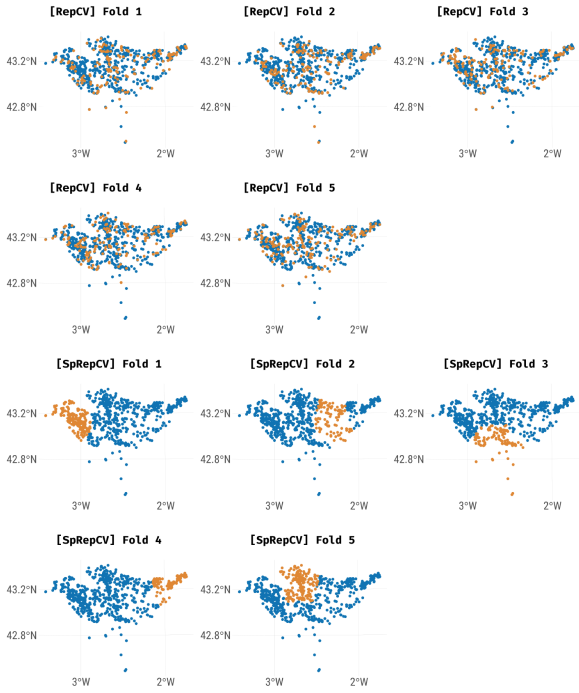

Figure 3.2: Comparing the effects of spatial cross-validation to the random selection of the conventional cross-validation. The diagram shows 5-fold repeated non-spatial (RepCV) and 5-fold repeated spatial cross-validation (SpRepCV) effects, the dots yellow (training dataset), and purple (test dataset; Schratz et al., 2019).

3.6.2 Feature space: Principal Component Analysis

The multi-date data was assessed by applying PCA a pre-analysis technique. Its purpose was to explain the variation in the feature space (predictor variables) of the multi-date data. So, FactoMineR: An R package for multivariate analysis conducted the PCA with the multi-date data (Le & Husson, 2008). PCA eliminates the noise and redundancy in the feature space by reducing the n-dimensions of the data into lower Dimensions (Dim) or Principal Components (PCs) to explain the variation in the dataset (Jolliffe, 2002). It considers the original variables in the dataset as Dimensions (or PCs). The number of PCs at least equal the variables in the data and combines the data (Kassambara, 2017). PCA has three key outputs: eigenvalues and variance table, correlation plot and loading vector

coefficients (Jolliffe, 2002). Eigenvalues/variances measure variation maintained by each principal component in the data and an eigenvalue > 1 shows that a PC accounts for most of the variation in the data. Correlation plot and loading vector show original variables that more correlated and quality of the contribution (or loading coefficients) cos^2 to the PCs that explained most of the variation in the data. The variables that explained most of the variation were the most important contributors (i.e., large loading coefficients) to the PCs that explained most of the variance in the data. Total Variance Explained Table illustrates the total variance explained by each principal component and was adopted to determine the most important PCs (eigenvalue > 1 and 70 - 90% variance) in explaining the variation in the dataset.

3.6.3 Training random forest algorithm

The supervised random forest classification employed sentinel-2A data to test its spatial accuracy to model potential wetland GDEs at a local scale. The random forest algorithm was trained on the multi-date data (or multi-epochs S2A data: 02-March 2017, 17-November 2017, 27-March 2018, and 23-October 2018), full dataset provided in Appendix A. Training of the random forest algorithm was carried out with the R package Classification and Regression Training (CARET) by Max (2020). Its random forest algorithm, hyper-parameters were optimised (or tuned) adopting a five-fold SpRepCV (or five-fold spatial repeated cross- validation) with 100 iterations. The SpRepCV tuning was carried out with the R package CAST: caret Applications for Spatio-Temporal models (Meyer, 2020). Caret trained 500 random forest models per date in the multi-epoch data, and it selected the model with the best predictive performance. The best model optimised the hyper-parameters and was good at learning from unseen observations (i.e., test data). Then the predictive performance of the best (or optimal model) was measured by applying the overall Accuracy metric.

The optimal model contained coefficients that learned the relationship between wetland and non-wetland classes (i.e., binary response), and the S2A bands (i.e., predictors). The optimal model then derived spatial prediction maps of potential wetland GDEs for each date in the dataset. Spatial prediction maps were derived

by applying the model coefficients to the original stacked S2A data (for the entire study region) using the R base predict function and the raster package. The spatial prediction maps were derived from a binary classification random forest model. They, spatial prediction maps, only recognised wetland and non-wetland pixels. But the original stacked data for the entire study region had other pixels besides the binary (wetland/non-wetland) classes. So, applying spatial predictions to the entire study region might be over-optimistic and invalidate the spatial model/maps. Meyer & Pebesma (2020) found an optimal way to spatially model binary classes without being over-optimistic. They developed the Dissimilarity Index (DI) to make predictions only in the areas where the model had learned the predictor-response relationship. When the DI was applied to the multi-date spatial models it worked out areas where the model learned the relationship of pixels as either wetland or non-wetland classes and defined them as an Area of Applicability (AOA). In the end, the final spatial predictive maps were presented as one showing the study region without a defined AOA and one with a defined AOA side-by-side.

3.7 Methodological limitations

The multi-date data employed to train the random forest algorithm comprised more non- wetland classes (n = 6) than wetland classes (n = 3).For a meaningful analysis, a minimum 2^7X 12 = 1536 is recommended. In such small samples the data is imbalanced, the algorithm often ignores the minority classes and focus on the overall performance by labelling all the pixels as the majority class (Maxwell, Warner & Fang, 2018). Such models will produce meaningless spatial prediction maps. To overcome this issue, this study tuned the random forest algorithm with the SpRepCV strategy to make sure model training considered the spatial nature of the data. But a major source of uncertainty was the data spatial autocorrelation. The feature space of the multi-date data (seven bands) is related, and in correlated data the random forest algorithm can be bias when assessing relative variable importance (Schratz et al., 2019). So, the study also applied the PCA besides the relative importance generated by random forest algorithm to peek through the feature space of the data. Overall, the S2A data was a test data since its spatial resolution (10 m) might not be ideal for predicting wetland and non-wetland pixels

$<$ 100 m.

Although the proposed post-fire NDVI analysis method was implemented using a higher spatial resolution data, temporal coverage limited it to post-burn index analysis. The uni-temporal (post-fire) design of the approach was accidental because aerial orthoimages lacked data before the fire. A pre-/post-fire index analysis approach was likely to have an increased detection accuracy, since it would measure rapid recovery against pre-fire conditions. Escuin, Navarro & Fernández (2008) recommended using pre-/post-fire indices to detect burned and unburned pixels since post-fire indices do not discriminate well between burned pixels and water bodies or bare soils or pixels containing very little vegetation. So, knowing burned and unburned pixels may improve the reliability of detecting rapid post-fire recovering wetland GDEs. But the cost and use of pre-/post-fire data determine if a study will use uni-temporal or bi-temporal approach. For this study post-fire NDVI analysis was developed from the data collected from an going project and so a bi-temporal approach was not considered because of data limitations. Within the accessible data, sources of error in uni-temporal data (e.g., shadow effects) were preemptively dealt with by capturing aerial images at the same time (e.g., late morning) and on or close to the same day. Overall, the sources of uncertainty because of the uni-temporal design do not negate the proposed novel approach. The dissertation proposes it as the first step to identify wetland GDEs. It is a contribution in GDEs research for expanding the Toolkit for identifying GDEs, along with existing approaches like the probability rating map for finding GDEs at regional scale proposed by (Münch & Conrad, 2007).

Chapter 4

Results and discussion

4.1 Post-fire NDVI analysis results

4.1.1 Aerial orthoimages NDVI

Post-fire NDVI recovery in wetland classes differed based on the date of the initial images. At site IIa there was a littel difference between wetland and non-wetland (A and B) class pixels recorded NDVI values after the fire. These pixels were almost completely burned by the recent fire disturbance (Figure 4.1). By 236 days after fire, the wetland at this site had much higher NDVI, suggesting that the wetland vegetation had recovered faster than the non-wetland vegetation. The non-wetland biomass displayed a gradual recovery throughout the three-year multi-date duration (2011-13). Within the multi-date, 236-days post-burn, the wetland biomass appeared to plateau for the remaining 2 years. The biomass vigour in both the wetland and non-wetland classes showed no sign of a fire disturbance 236-days ago in study site IIb (Figure 4.2). In contrast to site IIa, wetland and non-wetland NDVI values in site IIb appeared to have already plateaued. Biomass plateauing > 297-days post-fire (site IIb) and > 236-days post-fire (site IIa) suggest a recovery that occurred earlier but could not be captured because of the spatial coverage of the seasonal multi-date data. In the multi-date data, the reference study site illustrates the typical seasonal behaviour of wetland and non-wetland biomass

absent any fire disturbances (Figure 4.3). These results showed that, absent a fire distance, wetland and non-wetland classes were harder to distinguish.

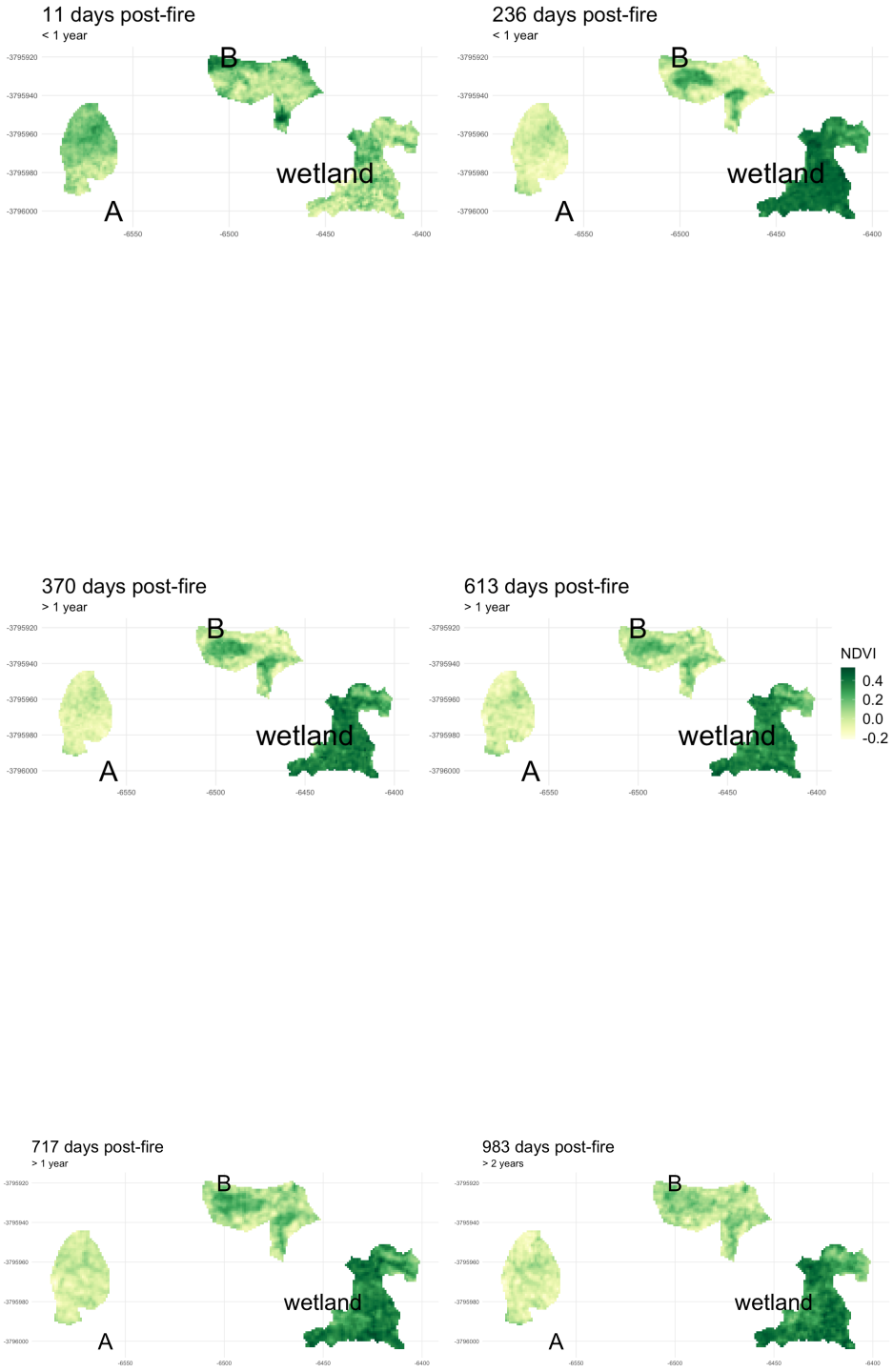

Figure 4.1: Aerial orthoimages NDVI- wetland and non-wetland (A and B) in study site IIa. Yellow tints illustrate bare soil and darker shades of green normal biomass vigour.

Figure 4.2: Aerial orthoimages NDVI- wetland and non-wetland (A and B) in study site IIb. First NDVI date: 27-March 2011 and fire date: 03-June 2010.

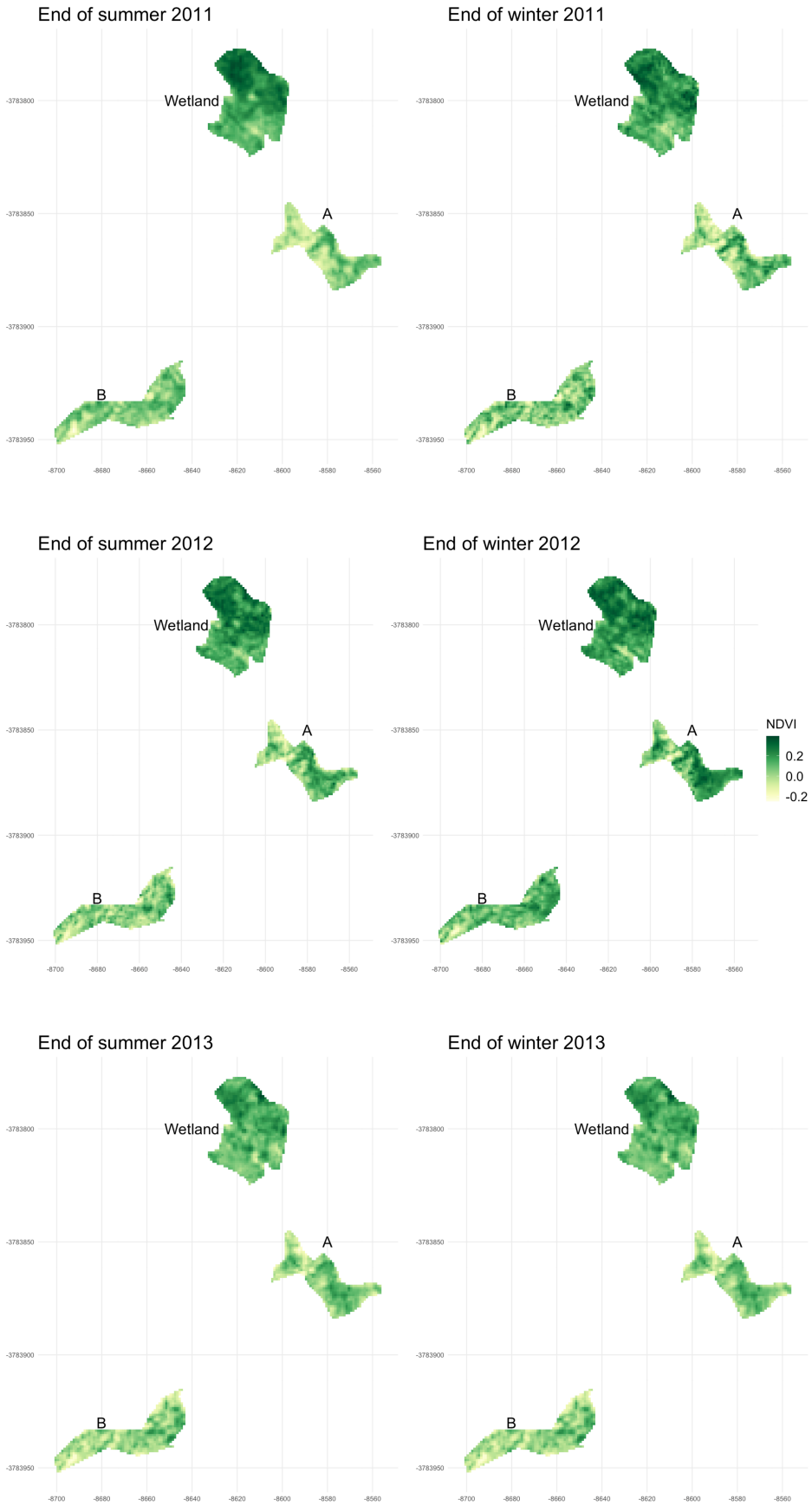

Figure 4.3: Aerial orthoimages NDVI- wetland and non-wetland (A and B) in reference study site (no fire disturbance).

The NDVI figures illustrate the overall greenness of the wetland and non-wetland pixels. But their mean NDVI gives an accurate and a quantitative description of the wetland and non-wetland class biomass post-fire (Table 4.1). The mean NDVI sample shows their values for the multi-date duration and including for the reference study site. The reference study site shows vegetation biomass without a fire disturbance. In the burned study sites (site IIa and site IIb), the mean NDVI values post-fire for both the wetland and non-wetland classes illustrate some increase in vegetation health/abundance recovery. Initially, the wetland class in study site IIa showed a mean NDVI of -0.01 indicative of the recent fire (11-days) before acquiring the aerial image. But 236-days post-fire its mean NDVI showed a rapid recovery to 0.35. Thereafter, the wetland maintained its mean NDVI approximately > 0.30 except for the decrease in 02-March 2013 (end of summer). This mean NDVI decrease could not be entirely explained. It could be a seasonal variation and therefore a sign the wetland is partially or completely not groundwater dependent. Also, the wetland class in site IIb recorded a drop in mean NDVI (0.14) from a consistent > 0.20 value. At site IIb, initial wetland mean NDVI 297-days post-fire started at 0.35 and maintained a > 0.20 value. The wetland class recorded a fully recovered mean NDVI but showed no clear sign of dropping, increasing, or plateauing for the remainder of the multi-date. While the wetland with the complete burn data showed rapid recovery its next NDVI image, 236-days after the fire. At 236-days post-fire and beyond, its NDVI some stage appeared have plateaued. In the CFR other studies in post-fire NDVI recovery have also reported NDVI plateauing, although approximately two–seven years post-fire (Hope, Albers & Bart, 2012; Wilson, Latimer & Silander, 2015).

Table 4.1: Mean NDVI data sample derived from aerial orthoimages.

Ref-site		Site IIa		Site IIb	
meanNDVI	variable	meanNDVI	variable	meanNDVI	variable
0.3320072	wetland	-0.0104144	wetland	0.3514740	wetland
0.1361984	wetland	0.3202167	wetland	0.2062898	wetland
0.2010823	wetland	0.3418561	wetland	0.2234546	wetland
0.0945657	wetland	0.2306208	wetland	0.1419637	wetland
0.2405314	wetland	0.3271533	wetland	0.2806956	wetland
0.1824560	wetland	0.2888033	wetland	0.2247174	wetland
0.1541612	A	0.0133412	A	0.1599868	A
0.0091122	A	-0.0403232	A	0.0533940	A
0.0621797	A	-0.0450105	A	0.0601898	A
0.0413285	A	-0.0294797	A	0.0377317	A
0.1717447	A	0.1021098	A	0.1815241	A
0.1514379	A	0.1010526	A	0.1458637	A
0.1911391	B	0.0157489	B	0.1600693	B
0.0176322	B	0.0150315	B	0.0609574	B
0.0353709	B	0.0594145	B	0.0784488	B
0.0126602	B	0.0364973	B	0.0550531	B
0.1497616	B	0.1726432	B	0.1946105	B
0.1464945	B	0.1503879	B	0.1350704	B

Note:

non-wetland class (A), and non-wetland class (B).

4.1.2 One-way Analysis of Variance

The temporal mean NDVI data sample in Table 4.1 was employed to carry out one-way ANOVA test (or the F-ratio) for the three study sites. The ANOVA test showed there was a significant difference between wetland and non-wetland class pixels in the burned study sites (IIa and IIb). In study site IIa (Figure 4.4), F-ratio revealed there was a significant difference between the known wetland and non-wetland classes (F (2,15) = 9.66, $p = 0.002$). In study site IIb F-ratio also showed there was a significant difference between the known wetland and non-wetland (A and B) classes (F (2,15) = 7.97, $p = 0.004$). The difference in the class means is illustrated by the post ad hoc test: Tukey HSD results (Figure 4.5). So, rejecting the null hypothesis that post-fire NDVI spectral signatures of known wetlands and non-wetland classes are similar in both burned study sites. And, accepting the alternative hypothesis that wetlands and non-wetlands have different post-fire NDVI values. Their mean NDVI difference was 0.37 (Tukey HSD) in study site IIa and was adopted to infer a mean NDVI baseline value post-fire above which pixels show wetland classes. This value was 0.22 in study site IIb suggesting NDVI plateaus after a specific period post-burn and becomes harder to differentiate between wetland and non-wetland classes. These results affirm applicability of the first level of analysis of the proposed novel post-fire NDVI analysis method; post-fire NDVI can detect the difference between wetland and non-wetland classes. For the non-wetland and wetland classes in the reference study site, the F-ratio tested the null hypothesis but for "normal" NDVI (unburned) instead of post-fire NDVI. The F-ratio revealed no significant difference between the known wetland and non-wetland classes (F (2,15) = 3.53, $p = 0.055$). This $p = 0.055$ might some signicance but strong enough compared to the sites got burned. This statistical correlation between the wetland and non-wetland class mean NDVI (Tukey HSD) is showen in Figure 4.6. Overall, rejected the null hypothesis that normal wetland NDVI and non-wetland class NDVI even absent a fire disturbance is similar.

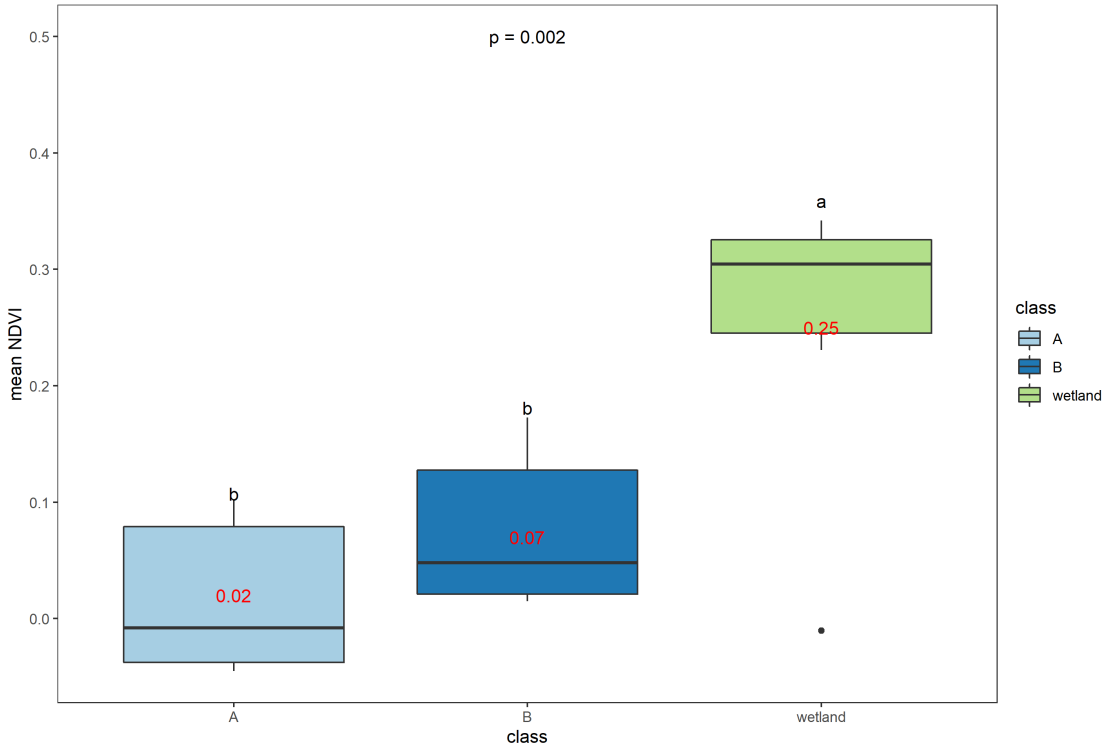

Figure 4.4: Study site IIa (burned). ANOVA showed there was a significant difference between wetland and non-wetland classes (F (2,15) = 9.66, p = 0.002). The different letters indicate significantly different means (Tukey HSD) and their mean indicated by the red lines.

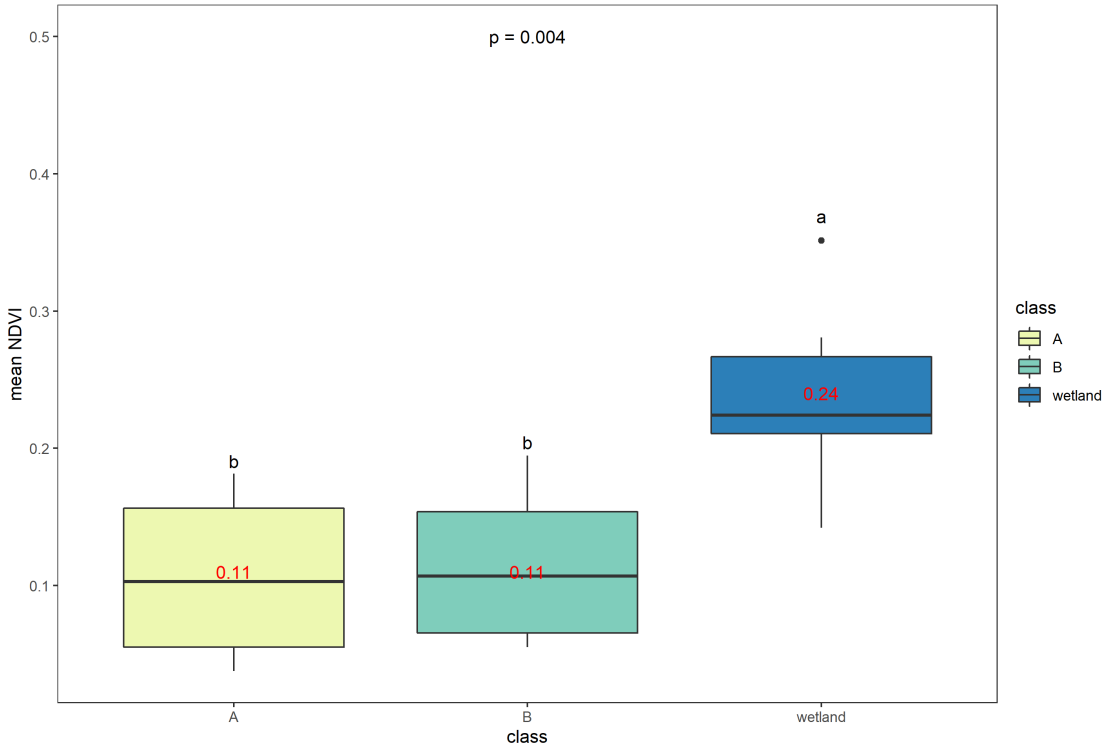

Figure 4.5: Study site IIb (burned). ANOVA showed there was a significant difference between wetland and non-wetland classes (F (2,15) = 7.97, p = 0.004). The different letters indicate significantly different means (Tukey HSD) and their mean indicated by the red lines.

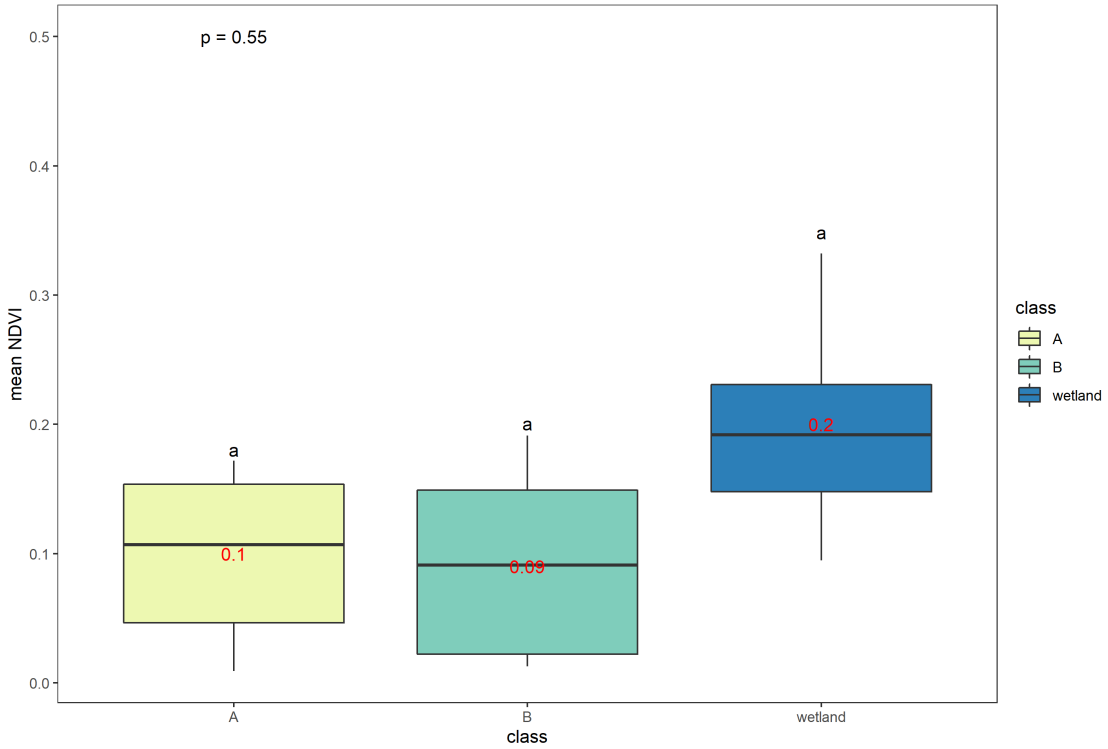

Figure 4.6: Reference study site (unburned). ANOVA showed no significant difference between wetland and non-wetland classes (F (2,15) = 3.53, p = 0.055). The different letters indicate significantly different means (Tukey HSD) and their mean indicated by the red lines.

4.1.3 Post-fire/multi-date NDVI distribution

The third and last level of analysis in the post-fire NDVI analysis approach derived multi-date NDVI distribution of the ANOVA test inferred wetland classes to assess their potential to be GDEs based on the recovery rate post-burn. For study site IIa, the NDVI distribution fell below > 0.1 in the kernel density curves for all the classes (i.e., wetland and non-wetland) 11-days post-fire (Figure 4.7). But, 236-days post-fire, the kernel density curve for the wetland class recovered rapidly while the curves for the non-wetland (A and B) classes more or less remained stable. Their recovery rate was lower compared to the rapid recovery in the known wetland. The wetland NDVI rapid recovery was observed on the next NDVI image, 236-days post-fire, but might have taken place earlier. After 236-days post-burn, the kernel density curve maintained an NDVI range of 0.25 for the remaining 2-years in the multi-date. The wetland kernel density curve plateaued from the first NDVI, 297-days post-fire, in study site IIb (Figure 4.8). Its NDVI maintained a range > 0.25 except for an anomaly on the 02-March 2013, likewise, noticed in the other two study sites. In both burned study sites, the spectral coverage of the data showed an already recovered wetland class. The NDVI kernel density curve of the wetland and non-wetland classes in the reference study site remain well established and appeared constant throughout the study period (end of summer and end of winter) in the multi-date duration, 2011-13 (Figure 4.9). Kernel density curve of non-wetland classes was distributed towards the low positive to the negative NDVI value range, and positive NDVI range for the wetland class.

The class area or number of pixels contributed to the physical properties of the NDVI density curves. So, the kernel density estimator function normalised the area under the curves for all the classes (i.e., wetland and non-wetland) to be equal to one. But the approach still did not account for the area difference (different number of pixels) of the wetland and non-wetland classes. With even more robust normalisation approaches, the trend in the results would remain. A key caveat is that the relative size of the NDVI density curves for non-wetland classes could likewise be attributed to their small area; site IIa (A = 1012 m^2) and (B = 1253 m^2), and for study site IIb (A = 2056 m^2) and (B = 610 m^2), compared to the larger known wetland areas of 1429 and 3313 m^2 for the first and second study

site. The area difference was noted as a potential source of error in the presented results. But, even accounting for area difference, the observed pattern will remain apparent. So, these preliminary kernel density distribution curves are fascinating because they suggest that post-fire NDVI is sensitive to detect rapid recovery in wetlands predicted to be GDEs.

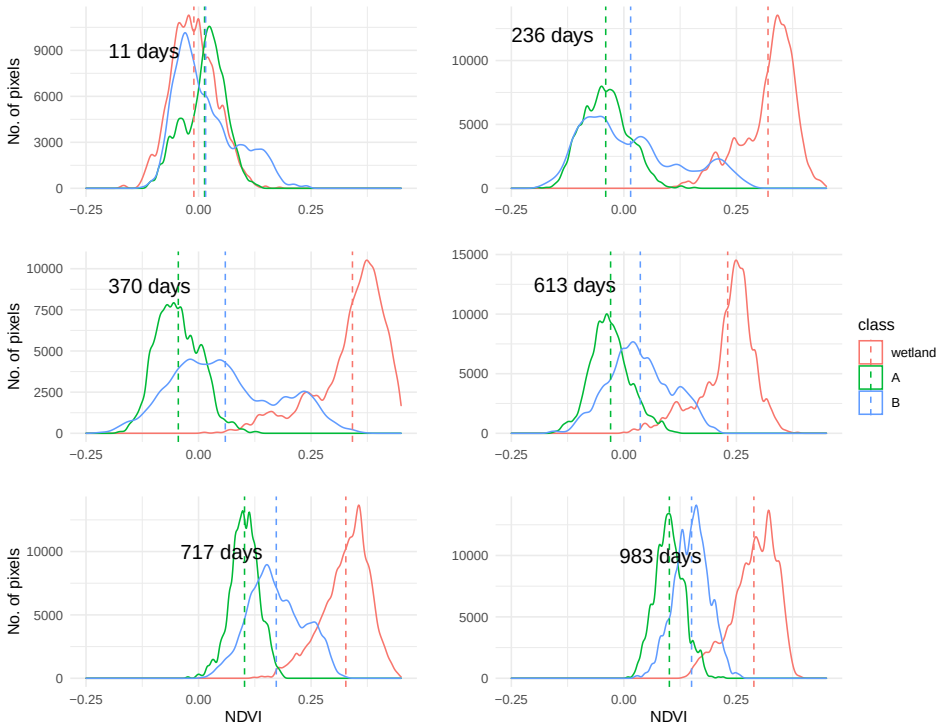

Figure 4.7: NDVI seasonal distribution (site IIa) 11-days post-fire. I - 27-March 2011, II - 07-November 2011, II - 20-March 2012, IV - 18-November 2012, V - 02-March 2013, VI - 23-November 2013.

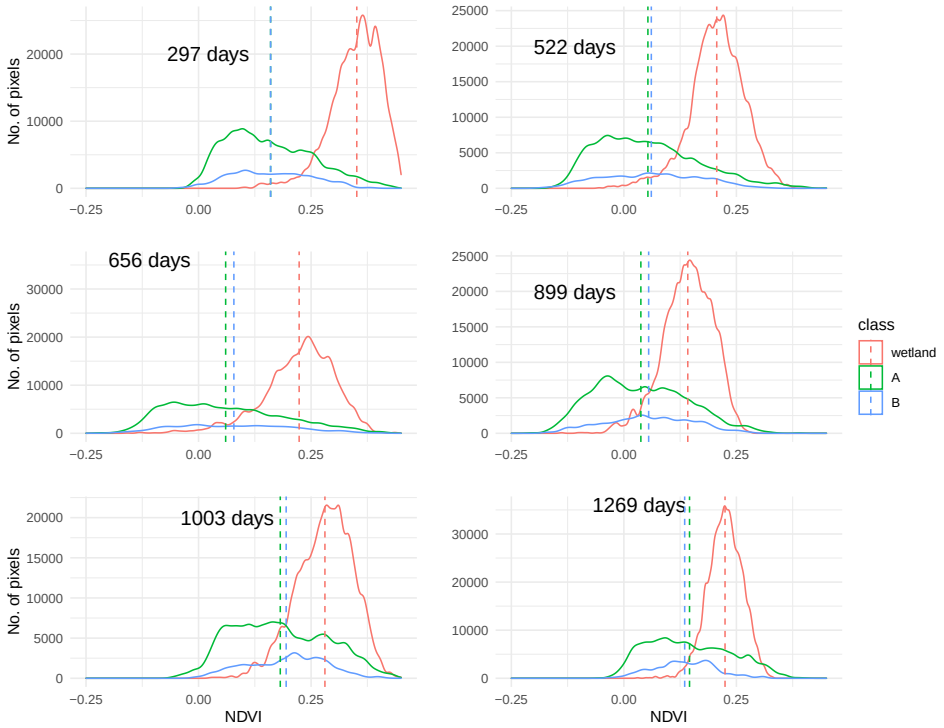

Figure 4.8: NDVI seasonal distribution (site IIb) 269-days post-fire. I - 27-March 2011, II - 07-November 2011, II - 20-March 2012, IV - 18-November 2012, V - 02-March 2013, VI - 23-November 2013.

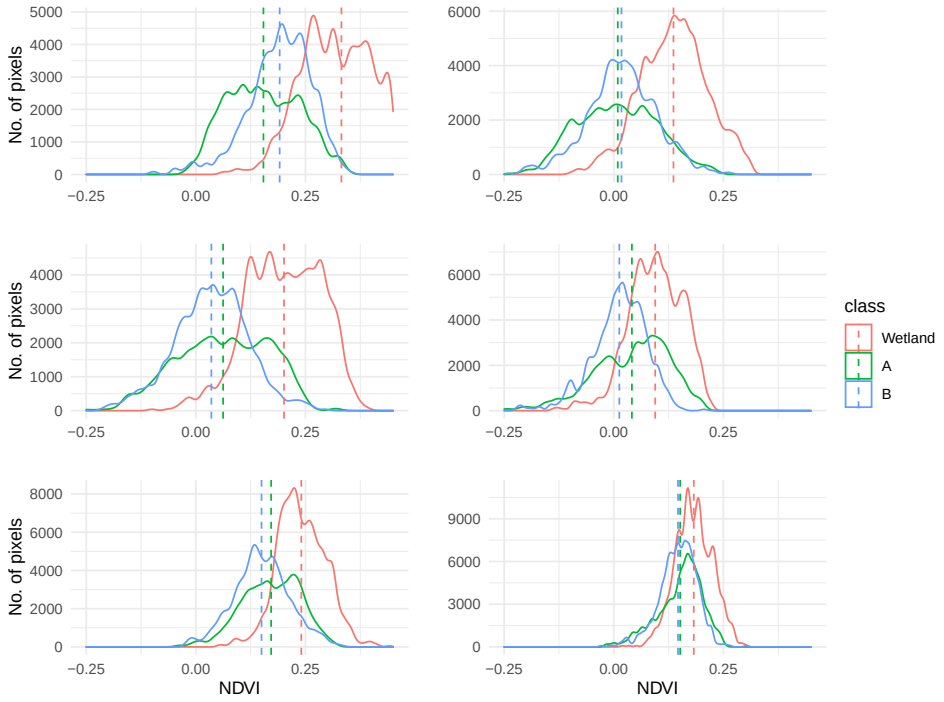

Figure 4.9: NDVI seasonal distribution in the reference study site (unburned; 3 years before the start of the images). The doted/dashed lines indicate the mean NDVI, wetlands and non-wetlands (A and B).

4.2 Post-fire NDVI analysis- discussion

The proposed novel approach found that temporal coverage post-fire is key to detect wetland classes. To detect the wetland classes, the approach applied the ANOVA test with larger confidence (or statistical significance) in study site IIa, with initial NDVI 11-days post-fire, contrasted to study site IIb with initial NDVI 297-days post-fire. So, the ANOVA test detected the mean spectral signature between wetland and non-wetland classes in the former and the latter with a p-value = 0.002 and p-value = 0.004, respectively. In the study site IIb, initial wetland NDVI appeared to have already plateaued. Between the two study sites, with plateaued and plateauing NDVI, the post-fire analysis results showed that significant difference of the mean NDVI between wetland and non-wetland classes drops with time post-fire. Their statistical difference (Tukey HSD) applied to infer an NDVI threshold range above which an NDVI spectral signature is likely to show a wetland class. Based on multi-date NDVI kernel density curves, if they show rapid post-fire recovery, the dissertation hypothesised that they are likely to be potential GDEs.

To detect significant difference between the wetland and non-wetland class in the reference study site, relatively stable NDVI (potential vegetation health and age post-burn) presented a challenge for the ANOVA test. In NDVI analysis, such variation without a fire disturbance in the case of this study often obscures other spatio- temporal variations (Lim & Kafatos, 2002). Leveraging the fire disturbances disrupted the strong seasonal biomass variations to assess the recovery rate of the delineated wetland classes hypothesised to be GDEs. I inferred their potential to be GDEs based on their rapid post-fire recovery rate. In the CFR, Wilson, Latimer & Silander (2015) showed that rapid post-fire recovery takes at least 1-year to 5-years and $>$ 25 years post-fire to reach maturity. Their study covered a regional scale, and they based the findings on the low spatial resolution MODIS NDVI. The proposed novel post-fire NDVI analysis approach was implemented with a finer spatial resolution image NDVI (i.e., analogous to zooming inside the MODIS pixels) and found that wetland classes had already recovered 1-year post-fire. These detected wetlands were hypothesised to be groundwater dependent. The proposed approach then derived multi-date NDVI kernel density curves to

assess their potential wetland GDEs, recovery rate to infer their dependency to groundwater. But their initial recovery rate could not be detected since their NDVI values plateaued 236 and 297-days post-fire. Before plateauing, NDVI in these wetland classes displayed rapid post-fire recovery. The results pointed to a rapid post-fire recovery that took place between 11-days and 236-days post-fire. The employed data failed to capture a more accurate date range of the rapid recovery in potential wetland GDEs. But overall, the general implications of these results are that the temporal coverage of the employed aerial orthoimages (seasonal acquisition) was inadequate to detect the rapid post-fire recovery of potential wetland GDEs in the CFR.

4.3 Random forest predictive modeling

4.3.1 Principal component analysis

Random forest mapped the potential wetland GDEs defined by the proposed novel approach using S2A multi-date data. But before training random forest with the multi-date (2017-18) data, its feature space was explained with the PCA. The first two principal components made up more than 92% of the variance for the multi-date data and its feature space. The results for each multi-date data are shown in Table 4.2; 02-March 2017, Table 4.3; 17-November 2017, Table 4.4; 27-March 2018, and Table 4.5; 23-October 2018.

Table 4.2: PCA results for S2A multi-date data- 02-March 2017.

	eigenvalue	variance (%)	cumulative variance (%)
Dim.1	4.3549242	62.2132028	62.21320
Dim.2	2.2872588	32.6751262	94.88833
Dim.3	0.2007886	2.8684086	97.75674
Dim.4	0.0749301	1.0704307	98.82717
Dim.5	0.0477227	0.6817523	99.50892
Dim.6	0.0252399	0.3605702	99.86949
Dim.7	0.0091357	0.1305093	100.00000

Table 4.3: PCA results for S2A multi-date data- 17-November 2017.

	eigenvalue	variance (%)	cumulative variance (%)
Dim.1	4.0538175	57.9116779	57.91168
Dim.2	2.4981480	35.6878290	93.59951
Dim.3	0.2228110	3.1830144	96.78252
Dim.4	0.1359503	1.9421467	98.72467
Dim.5	0.0564356	0.8062225	99.53089
Dim.6	0.0291316	0.4161654	99.94706
Dim.7	0.0037061	0.0529442	100.00000

The loadings coefficients defined input of the original features (or bands) to the most important principal components. The loadings (or coefficients quantity) table showed the coefficient values of each predictors (bands) that contribute to the selected PCs. The loadings result and positive correlation coefficient are illustrated in Table 4.6; 02-March 2017, Table 4.7; On 17-November 2017, Table 4.8; 27-March 2018, and Table 4.9; 27-March 2018. In the last step of PCA, correlation quality (cos^2) results illustrated the correlation of the variables contributing to principal components (PC1 and PC2). The top plot (end of summer) in Figure 4.10, the first loading vector (i.e., PC1) placed equal weight on B03, B04, B05, B05 and B11, with much less weight on NDVI and NDWI. The second loading vector placed more weight on B08 and the spectral indices (NDVI and NDWI) and less weight on the other individual spectral bands. Then, in the bottom plot (end of winter) spectral indices showed close correlation with each other than the

Table 4.4: PCA results for S2A multi-date data- 27-March 2018.

	eigenvalue	variance (%)	cumulative variance (%)
Dim.1	4.3936362	62.7662308	62.76623
Dim.2	2.2817629	32.5966132	95.36284
Dim.3	0.2008688	2.8695550	98.23240
Dim.4	0.0657528	0.9393263	99.17173
Dim.5	0.0438292	0.6261311	99.79786
Dim.6	0.0124244	0.1774912	99.97535
Dim.7	0.0017257	0.0246523	100.00000

Table 4.5: PCA results for S2A multi-date data- 23-October 2018.

	eigenvalue	variance (%)	cumulative variance (%)
Dim.1	4.3906670	62.7238144	62.72381
Dim.2	2.2306033	31.8657612	94.58958
Dim.3	0.2595325	3.7076075	98.29718
Dim.4	0.0804683	1.1495466	99.44673
Dim.5	0.0345188	0.4931257	99.93986
Dim.6	0.0033114	0.0473057	99.98716
Dim.7	0.0008987	0.0128388	100.00000

Table 4.6: Coefficient values of the original variables for the first two PCs (02-March 2017).

	PC1	PC2
B04	0.9810447	0.0794118
B11	0.9566314	-0.1006670
B03	0.8664150	0.4256412
B05	0.8545445	0.4636803
B08	0.1661561	0.9670116
ndwi	-0.5770599	0.7832421
ndvi	-0.7973735	0.5710247

Table 4.7: Coefficient values of the original variables for the first two PCs (17-November 2017).

	PC1	PC2
B11	0.9403715	0.0250723
B04	0.9380354	0.3037778
B03	0.6003112	0.7409787
B05	0.4894326	0.8161913
B08	-0.4810944	0.8453219
ndwi	-0.7542371	0.6249407
ndvi	-0.9430616	0.2913775

Table 4.8: Coefficient values of the original variables for the first two PCs (27-March 2018).

	PC1	PC2
B04	0.9810974	0.0556029
B11	0.9492514	-0.1118025
B03	0.8889412	0.3860108
B05	0.8492117	0.4764079
B08	0.2101692	0.9626852
ndvi	-0.5548253	0.8155057
ndwi	-0.8164720	0.5462511

Table 4.9: Coefficient values of the original variables for the first two PCs (23-October 2018)

	PC1	PC2
B04	0.9640078	0.1793213
B11	0.9533168	0.0049142
B03	0.8156114	0.4983729
B05	0.7546256	0.5795379
B08	-0.1666151	0.9746604
ndwi	-0.6964967	0.7037280
ndvi	-0.8972141	0.4110805

individual bands. For the summer (top plot) and end winter seasons (bottom plot) in 2018 Figure 4.11, the multi-date data showed a similar correlation between spectral indices and individual spectral bands. As expected, individual bands showed close correlation with each other and indices with each other, but the key difference was the values of their coefficient quality (cos^2) input to either PC1 or PC2. The apparent behaviour of the bands/indices was used to explain the nature (or statistical independence) of the multi-date data feature space and how such correlation affected the reliability of the random forest. Beyond explaining the feature space of the multi-date data principal components were not incorporated in training the random forest. Random forest trained on the original variables multi-date data to avoid lowering the dimension of the training data.

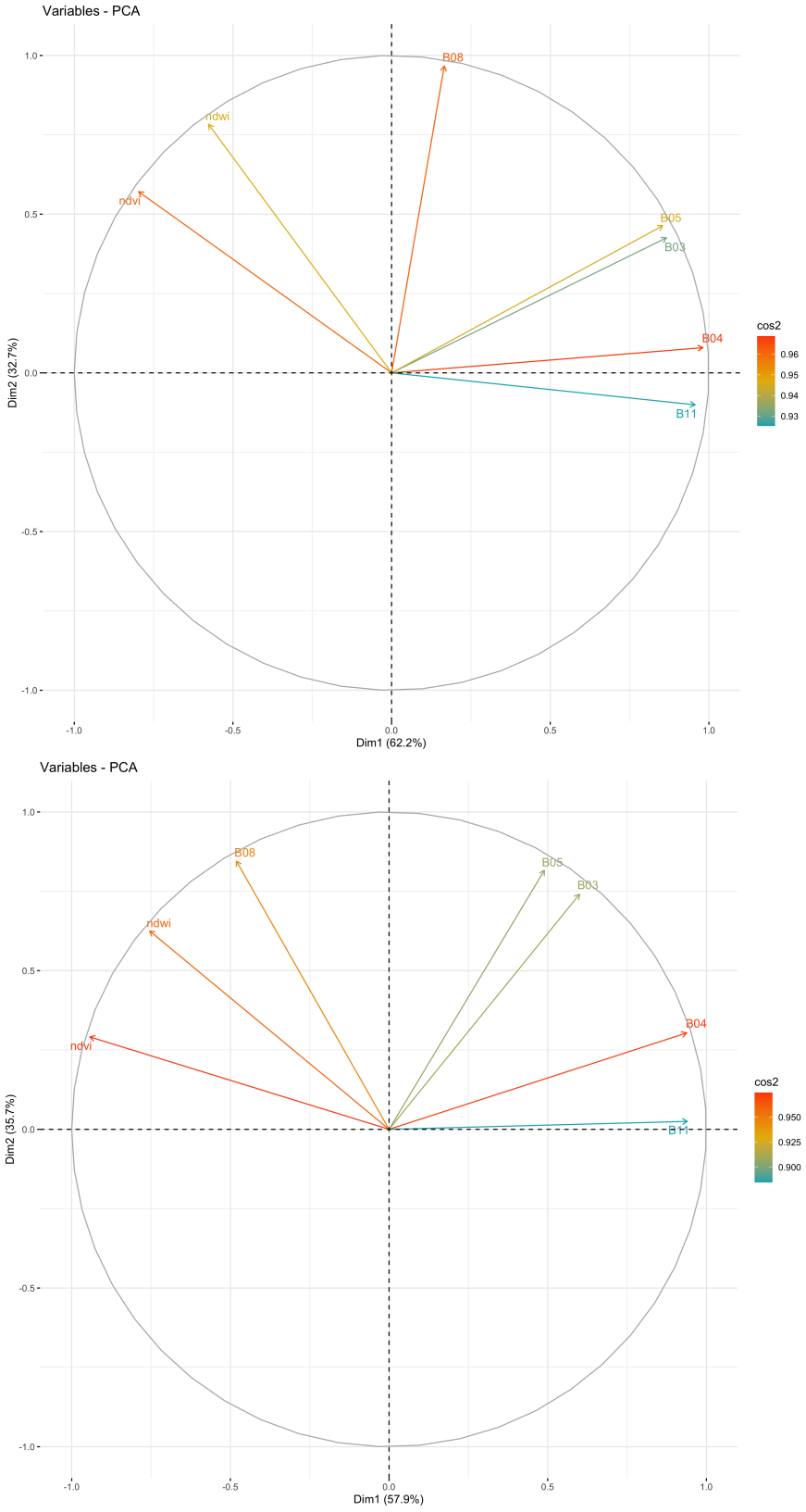

Figure 4.10: Variable correlation quality for the dates; 02-March 2017 (top), 17-November 2017 (bottom).

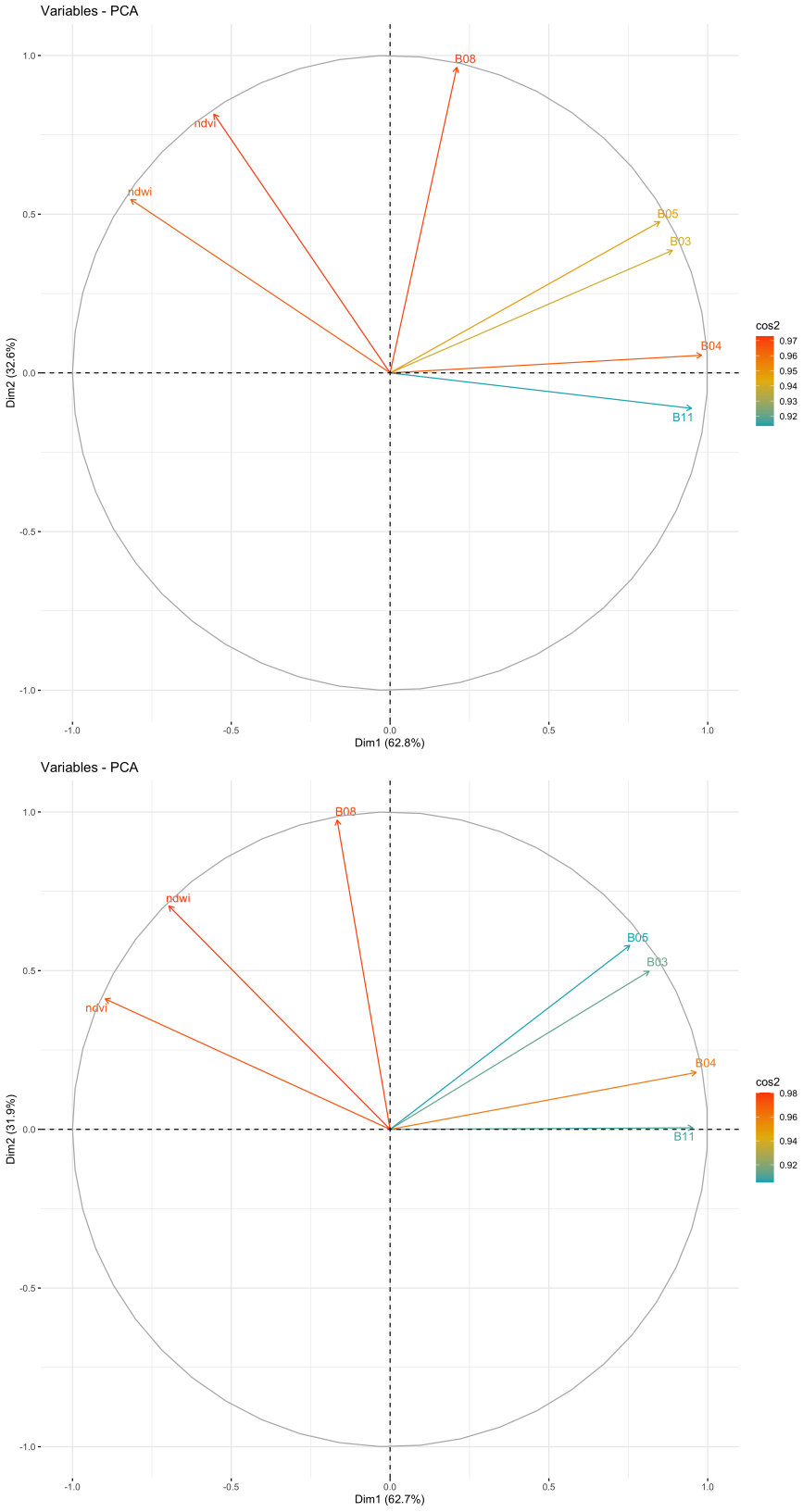

Figure 4.11: Variable correlation quality for the dates; 7-March 2018 (top), 23-October 2018 (bottom).

PCA explained the structure and feature space of the multi-date S2A dataset. The key results of the PCA is that the first two PC axes explained > 90% of the variation the data. In spectrometry data processing applying techniques like PCA are critical to understanding the feature space and for reducing its dimensions (Adam & Mutanga, 2009). The eigenvalues (> 1.00) from the PCA results showed that only the first two principal components explained most of the variation in the multi-date data. The decision for how many principal components to keep is subjective and differs from study to study. In this study they, principal components, were not intended for training random forest but to show the correlation in the data. Overall, the first two PCs explained 94% of the variance in the multi-date data on 02-March 2017, 95% on 17-November 2017, 98% on 27-March 2018 and 96% on 23-October 2018. For the duration of the multi-date, loading coefficients of the original variables showed that the individual spectral bands contributed to the first principal component. The new index (i.e., PC1) could estimate wetland biomass having spectral bands for both species' health/abundance and water content. But PC1 is more likely to be sensitive to photosynthetic activities more than biomass water content since both water content bands (SWIR; B11) had the lowest loading coefficient for it. PC2 had its largest loading coefficients from NDWI and NDVI, respectively. So, PC1 and PC2 failed to distinguish between the traditional plant water content bands and the bands for estimating plant health because of the high correlation nature of the multi-date data. But based on the correlation quality, the most important (correlated) variables for discriminating wetlands predicted to be GDEs could be classified into two groups: NIR, Red-edge, Green, Red and SWIR, and the second grouped into: NDVI, NDWI and NIR. In the second PCA defined group, the indices (NDVI and NDWI) were more correlated with each other than the NIR band. Previous studies have shown that indices are more effective for estimating vegetation spectral signatures since they measure differences from individual bands (Pettorelli, 2013). The classification in the first group is consistent with previous findings that variations in the NIR, Red-edge, Green, Red and SWIR are effective for estimating wetland biomass (Liu et al., 2004; Adam & Mutanga, 2009). But the Green band appeared to be an anomaly since it has low reflectance and transmittance given its high absorption by plant species because of pigments like chlorophyll and carotene (Adam, Mutanga & Rugege, 2010). The Green band

might not be effective in estimating the spectral signature of plant species, but its inclusion for modelling wetland GDEs is justifiable. Gangat et al. (2020) found that the Green band had the highest coefficient for predicting soil moisture content of wetlands.

4.3.2 Random forest modelling

In the training data, the random forest algorithm generated results of relative variable importance. But the algorithm found no increase in performance when using over three features (bands). The algorithm generated variable importance showed different variable were important to define the predictor-response relationship depending on the season: winter or summer. Red-edge, NDWI, and NIR were the most important variables in summer, SWIR, and Red replaced Red-edge and NIR in winter (Figures 4.12). SWIR, Red was also important in the winter of 2018, but NDVI replaced NDWI (Figure 4.13). But in summer random forest showed that only SWIR and NDWI were the important variables. The variable importance measure in random forest is not robust, its results give inference of the key variable to the predictor-response relationship. Their measure was appropriate to test the application of random forest with S2A multi-date data to predict wetland GDEs.

The predictive accuracy of the optimal random forest model in 02-March 2017 was 71% and with a final $m_{try} = 2$. In 17-November 2017 the predictive accuracy and optimal hyper-parameters of the best model was 63% and $m_{try} = 3$. For 07-March 2018, the final $m_{try} = 2$ (held at constant) and the predictive accuracy of the best model was 62%. Random forest algorithm found no increase in predictive performance when using over two variables, and since m_{try} is dependent to the number of important variables to the classification outcome (Probst, Wright & Boulesteix, 2019). And, for 23-October 2018 the best predictive model had an accuracy of 59% and a final $m_{try} = 2$. The overall accuracy of the optimal models for the multi-date data was not impressive and this was expected since spatial cross-validation tuning has been shown to reduce predictive performance at least by 24-39% to account for spatial autocorrelation (Schratz et al., 2019). Last, for each optimal model spatial prediction results were illustrated using a spatial

prediction map: 02-March 2017 (Figure 4.14), 07-November 2017 (Figure 4.15), 07-March 2018 (Figure 4.16), and 23-October 2018 (Figure 4.17). Each figure illustrates the original spatial prediction first and then the spatial prediction map corrected for the AOA. The environmental conditions outside the AOA differ from the predictor-response relationship the model learned during model training. So, to define the AOA ensured the predictions of the model were optimistic and did not extend to areas unknown to the model.

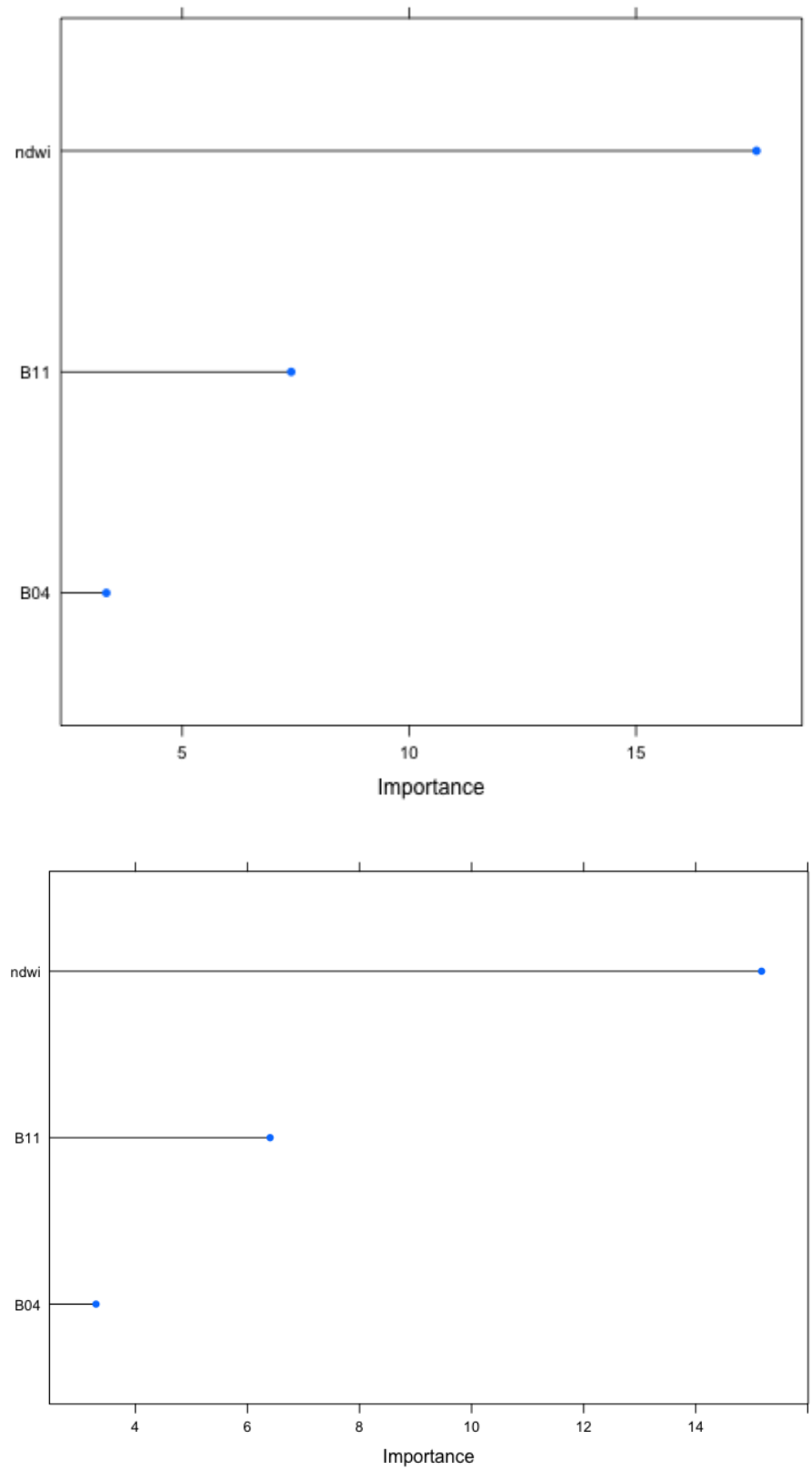

Figure 4.12: Variable importance- 02-March 2017 (top), 17-November 2017 (bottom).

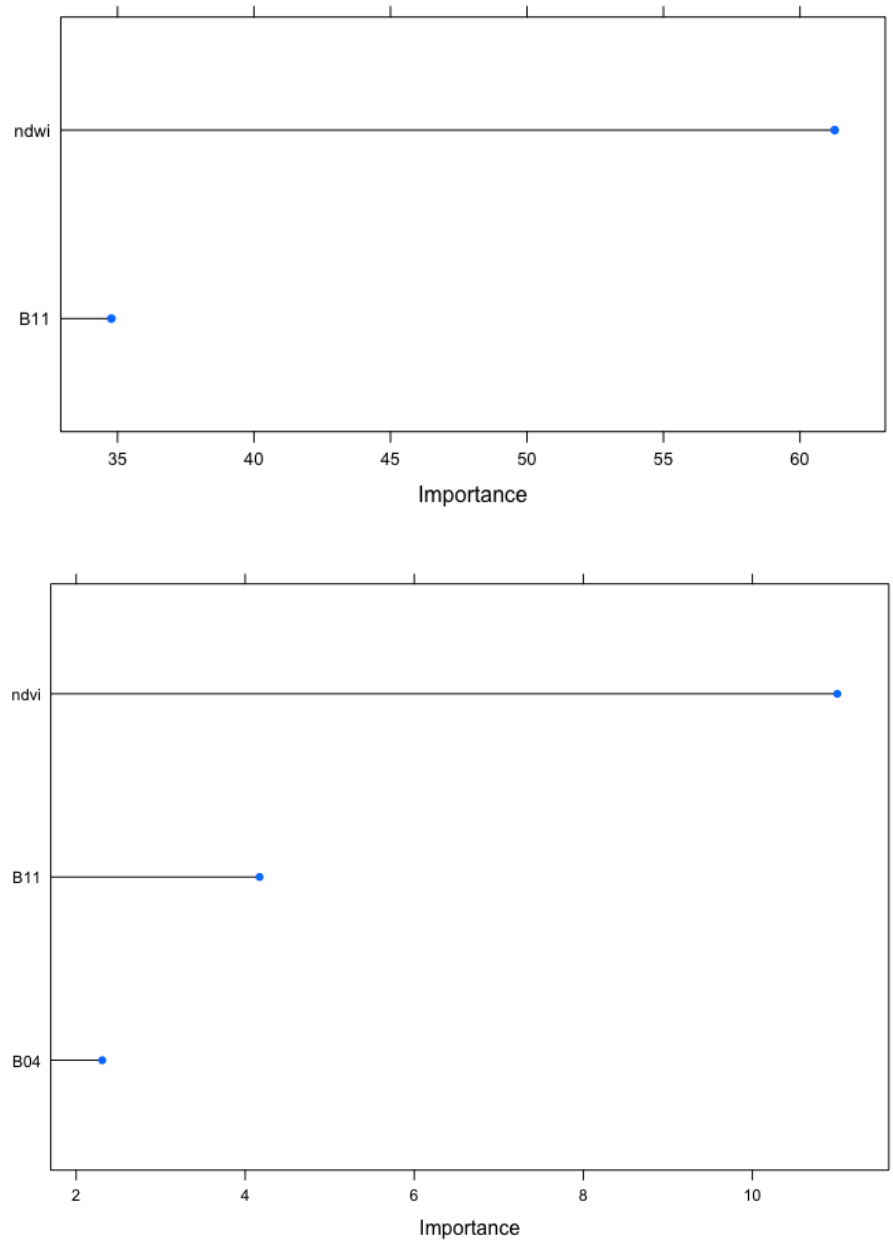

Figure 4.13: Variable importance- 07-March 2018 (top), 23-October 2018 (bottom).

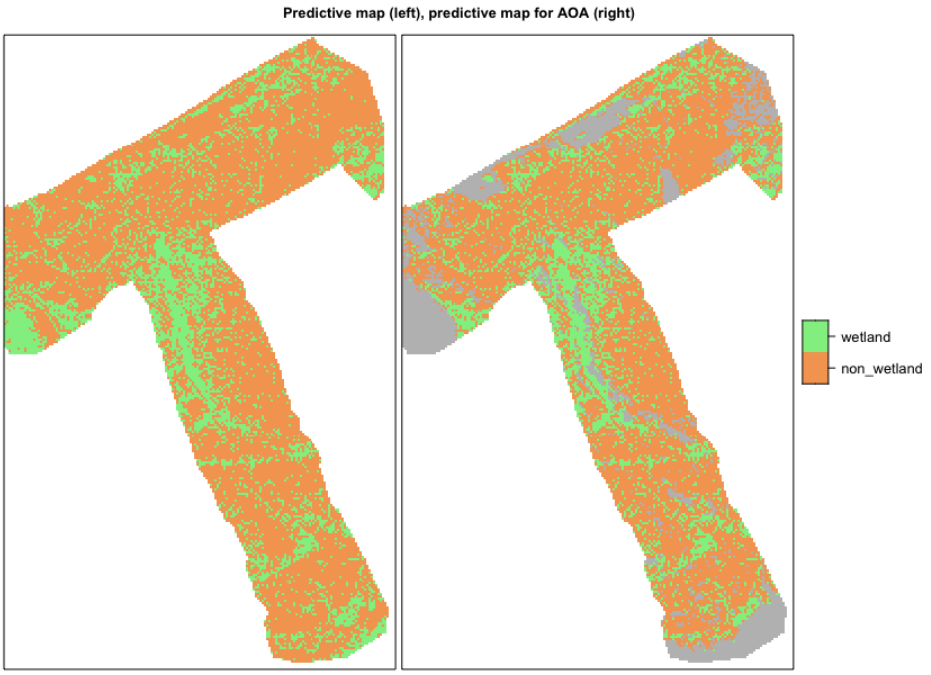

Figure 4.14: Predicted map of wetland GDEs- 02-March 2017. The model identified pixels as either wetland or non wetland was therefore not applicable across the study region. Outside the Area of Applicability (AOA) are pixels the model could not classify into wetland and non wetland.

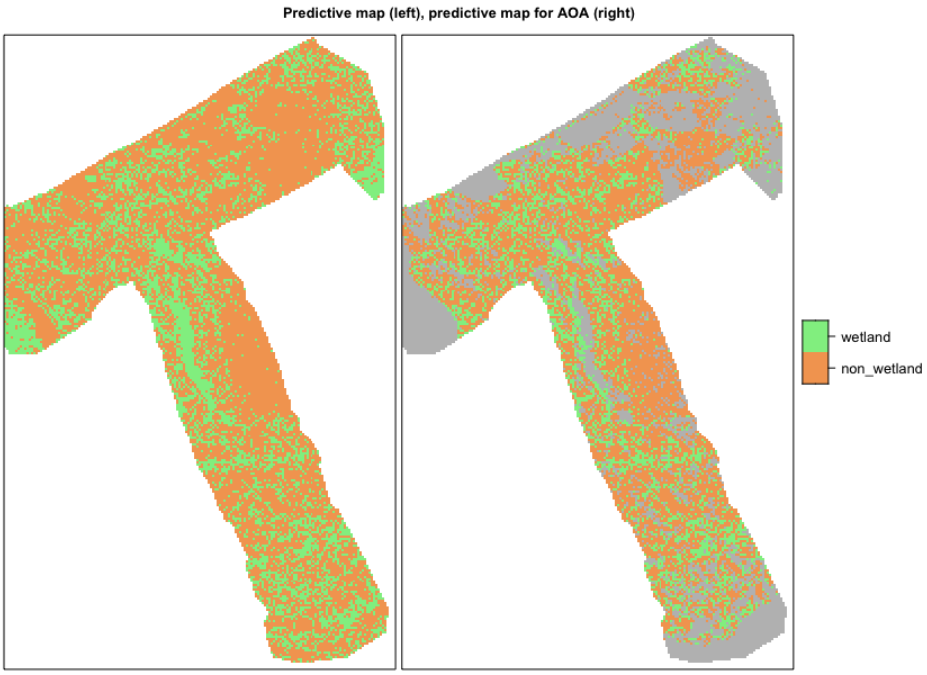

Figure 4.15: Predictive map of wetland GDEs- 07-November 2017.

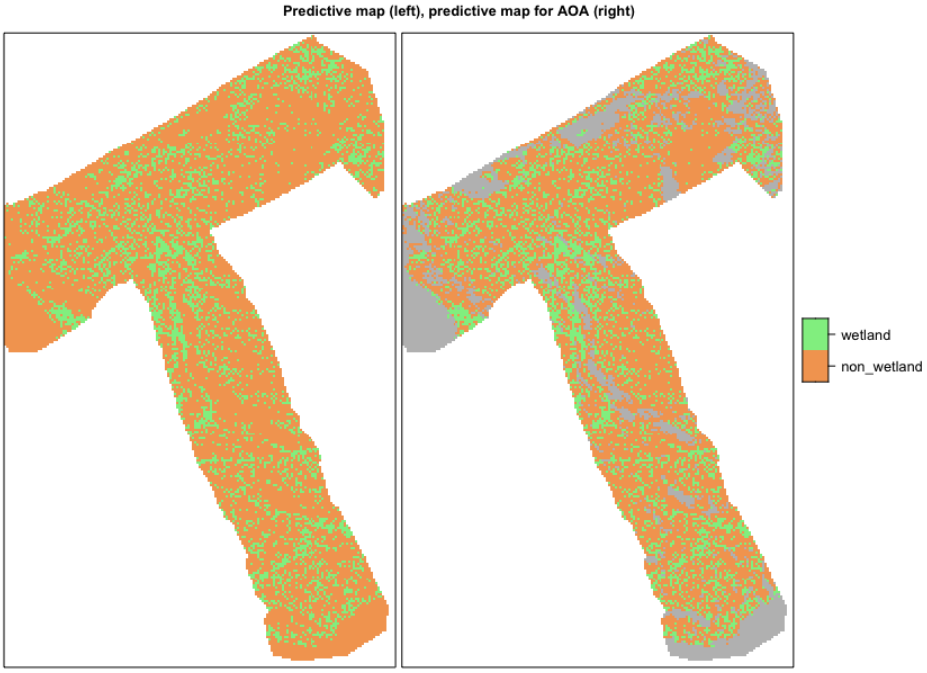

Figure 4.16: Predictive map of wetland GDEs- 07 March 2018.

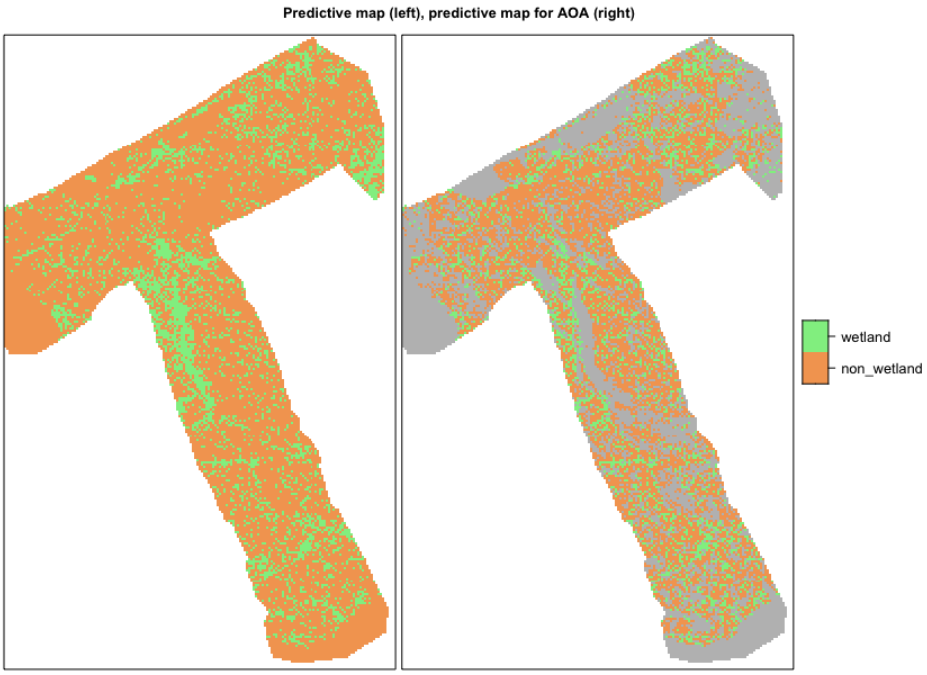

Figure 4.17: Predictive map of wetland GDEs- 23 October 2018. The pixels maintaining their (greenness) throughout the multi-date (dry/wet periods) are potential GDEs.

The potential wetland GDEs were predicted from a random forest algorithm trained on the entire multi-date data. For model training, a 5-fold SpRepCV tuned the model hyper-parameters instead of the conventional repeated cross-validation. A non-spatial cross-validation technique cannot consider the spatial nature of the data and may produce over-optimistic predictions (Lovelace, Nowosad & Muenchow, 2019; Schratz et al., 2019). The best model predictive optimised random forest hyper-parameters, a final m_{try} was between 2-3 and $Ntree$ 500 and 20 trees. User-defined $Ntree$ hyper-parameters influenced the predictive accuracy of the model but beyond, 500 trees and below 20 trees, its impact plateaued. For 07-March 2018, regardless of the $Ntree$ value, only two variables were defined as important and it held m_{try} at a constant value of 2. No model hypothesis could not explain

these results. Beyond the anomaly, the overall accuracy across the multi-date was relatively low, highest 71% and lowest 59%. When using SpRepCV autocorrelation is accounted for, low model predictive accuracy has also been reported in previous studies (Brenning, 2012; Lovelace, Nowosad & Muenchow, 2019).

The most optimal random forest model produces spatial prediction maps/models that model the learned predictor-response relationship in the entire study region/site. The study region, Kogelberg and Steenbras Nature Reserves, were experiencing a drought during the study multi-date. The tuned random forest model was used to conduct spatial modelling during this multi-date and predicted the pixels that maintained their greenness throughout the multi-date to be GDEs. Accuracy of the spatial prediction maps of GDEs was further improved by defining the AOA as proposed by (Meyer & Pebesma, 2020). The AOA defined pixels that could not be GDEs or non-GDEs because the environmental (pixel classes) conditions were unknown to the model (e.g., water pixels), and those were defined as being outside of the AOA for potential wetland GDEs.

Chapter 5

Conclusions

The bulk groundwater pumping in the TMG aquifer system to augment the primary source of water supply has unknown effects on GDEs. The worst-case scenario would be lowering the water table to depths that might put groundwater out-of-reach for GDEs. Regardless, GDEs require monitoring to determine groundwater depths below which they cannot survive. This is especially the case in the CFR with a hyper-diverse, endemic and potentially vulnerable fynbos vegetation. But the first step is knowing where the GDEs are in the landscape (Eamus, 2016). The proposed novel approach leveraged the recurrent fires in the CFR to detect and show that rapid post-fire recovering wetlands are potential GDEs. In the context of the research question; how can detailed mapping of GDEs at catchment scale be achieved to better monitor the impact of groundwater abstraction? The proposed novel approach potentially provides a path to answering this question in the CFR. The aim of the dissertation was also achieved by employing fire data and the remote sensing index NDVI, the proposed novel approach provides a simple, effective and improved way to detect GDEs from their surroundings at the local scale. More significantly, though, in a fire prone region like the CFR, leveraging fire data to detect GDEs has potential to reduce the reliance on groundwater depth data and other field observation techniques that require laborious ground calibration to integrate with remote sensing data. Such a simple and efficient approach is necessary and timely given the ongoing bulk groundwater abstraction could have

an irreversible impact on GDEs more so on those not yet categorised (i.e., unknown GDEs).

However, this approach is limited to post-burn environments. It further assumes that burned regions have been monitored by remote sensing at least after the fire, but it should be under monitoring before the fire to compare pre/-post-fire conditions. Also, it should be employed as the first step of identification with further ground-truth. In this dissertation, its implementation was tested in 3 sites, so it requires a larger site selection to test its applicability to discriminate wetland classes. In this study, ANOVA and Tukey detected significant differences between wetland and non-wetlands post-fire, with wetland classes having NDVI between 0.22 and 0.37 greater than non-wetland classes. The known wetland classes showed rapid post-fire recovery in $<$ 236-days post-fire in one burned site, and in $<$ 297-days post-fire in another burned study site. Within the CFR, Wilson, Latimer & Silander (2015) reported that post-fire recovery takes at least 1-year to 5-years. Their work employed NDVI images with coarse spatial resolution (i.e., MODIS) and naturally it failed to pick up the small size wetland classes detected in this study. Unlike Wilson, Latimer & Silander (2015), this study showed the CFR has small wetlands that might have recovery as rapidly as within a year post-fire.

Fire data had limited dates that coincided with aerial orthoimages and they, aerial orthoimages, had a limited feature space (spectral bands/variables). Instead, the supervised machine learning random forest algorithm was trained with the sentinel-2A orthoimages. The random forest achieved the last objective of the dissertation; was effective in mapping the spatial distribution of potential GDEs. Previous studies have found that adding more spectral bands to the training data improves the predictive performance of the model, but because of high correlation in the multi-date data, more bands did not improve the predictive performance in this study. Only three spectral bands were important to explain the predictor-response (spectral signature of wetland and non-wetland GDE classes) relationship in the dataset. These results showed that training the model with correlated spectral bands lowers the dimension of the training data and affects its statistical independence. So, the key recommendation when using S2A (or similar dataset) to train random forest would be to improve the dimensions of the data but preserve its statistical

independence. Last, to better detect GDEs in the CFR, it is recommended acquiring monthly images, at least in the first 1-2 years post-fire is likely to capture their rapid recovery far better than seasonal acquisition. Studies employing pre- /post-fire images at high spatial resolution will reduce the need for laborious ground-truthing, since they will compare detected rapid recovery with pre-fire conditions. And areas that are non-GDEs might also be resilient to fire, so to account for spatial heterogeneity of the physical landscape use of SAR data is recommended to produce improve accuracy.

Appendix A

Sentinel-2A multi-date data

In Chapter 3:

Table A.1: Multi-date data, 02-March 2017 (summer)- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve.

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.3645603	-0.1401152	495	719	951	1544	2079	control
0.3990729	-0.1292945	673	713	931	1660	2153	control
0.4085779	-0.1292945	473	655	931	1560	2153	control
0.5572858	-0.1594116	352	398	751	1400	1931	control
0.5374149	-0.1594116	379	408	751	1356	1931	control
0.5816794	-0.1594116	341	411	751	1554	1931	control
0.5755060	-0.0937697	365	409	815	1518	1738	control
0.4886318	-0.0071297	615	641	1053	1866	1907	wetland
0.5208060	-0.0252667	506	547	966	1736	1826	wetland
0.5903166	-0.0252667	436	440	966	1708	1826	wetland
0.5759186	-0.0095883	463	479	917	1780	1790	wetland
0.5884297	0.0231568	474	498	1009	1922	1835	wetland
0.5885246	0.0231568	491	502	1009	1938	1835	wetland
0.6915960	0.1532194	493	400	964	2194	1611	wetland

Table A.1: Multi-date data, 02-March 2017 (summer)- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.7236936	0.0909091	488	386	915	2408	1530	wetland
0.7465913	0.2340536	484	381	950	2626	1471	wetland
0.7455540	0.2340536	507	372	950	2552	1471	wetland
0.7497456	0.1532194	510	369	964	2580	1611	wetland
0.7235436	0.1532194	511	401	964	2500	1611	wetland
0.7422050	0.2954111	522	401	932	2710	1474	wetland
0.7506631	0.2954111	489	376	932	2640	1474	wetland
0.7294995	0.2777340	458	381	930	2436	1377	wetland
0.7423566	0.2777340	484	375	930	2536	1377	wetland
0.7619804	0.3253760	484	375	1016	2776	1413	wetland
0.7341350	0.3253760	472	398	1016	2596	1413	wetland
0.7561917	0.2954111	469	379	932	2730	1474	wetland
0.7237027	0.2954111	484	394	932	2458	1474	wetland
0.7011494	0.2777340	437	390	930	2220	1377	wetland
0.7485549	0.2777340	451	348	930	2420	1377	wetland
0.7435897	0.3253760	477	385	1016	2618	1413	wetland
0.7098151	0.3253760	473	408	1016	2404	1413	wetland
0.7046929	0.2333001	453	387	886	2234	1538	wetland
0.7056711	0.1809145	474	410	922	2376	1648	wetland
0.6796992	0.1809145	485	426	922	2234	1648	wetland
0.6794240	0.1657022	458	423	918	2216	1586	wetland
0.7110239	0.1657022	495	405	918	2398	1586	wetland
0.7111111	0.1882716	485	390	977	2310	1578	wetland
0.6809816	0.1882716	467	416	977	2192	1578	wetland
0.6425918	0.1809145	464	433	922	1990	1648	wetland
0.6716647	0.1657022	484	427	918	2174	1586	wetland

Table A.1: Multi-date data, 02-March 2017 (summer)- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.6528925	0.1657022	505	462	918	2200	1586	wetland
0.6033333	0.1882716	489	476	977	1924	1578	wetland
0.5705833	-0.0969760	501	508	1015	1858	2104	wetland
0.6564083	0.0550868	484	441	961	2126	1904	wetland
0.4801125	-0.0969760	599	647	1015	1842	2104	wetland
0.6500396	0.0550868	494	442	961	2084	1904	wetland
0.5548645	-0.1752148	504	501	1011	1750	2257	wetland
0.3936034	-0.1698710	619	711	970	1634	2221	wetland
0.5184714	-0.0582038	511	567	988	1788	2009	wetland
0.6232108	-0.0582038	506	487	988	2098	2009	wetland
0.4878049	-0.1698710	566	588	970	1708	2221	wetland
0.5926319	0.0568266	587	575	1145	2248	1917	control
0.5464780	0.0568266	658	705	1145	2404	1917	control
0.5447977	0.0409688	607	630	1128	2138	1861	control
0.5537059	0.0409688	552	563	1128	1960	1861	control
0.5826264	0.1081017	625	615	1137	2332	1877	control
0.6390495	0.1081017	500	562	1137	2552	1877	control
0.6396461	0.1308594	500	509	1088	2316	1780	control
0.5824345	0.1308594	516	542	1088	2054	1780	control
0.5352160	0.1081017	706	683	1137	2256	1877	control
0.6586307	0.1081017	523	531	1137	2580	1877	control
0.6692806	0.1308594	500	485	1088	2448	1780	control
0.5294523	0.0292072	700	683	1246	2220	2094	control
0.6535184	0.0292072	514	549	1246	2620	2094	control
0.6285622	0.1088314	542	567	1193	2486	1998	control
0.5541561	0.0292072	636	708	1246	2468	2094	control

Table A.1: Multi-date data, 02-March 2017 (summer)- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.5588723	0.1088314	609	665	1193	2350	1998	control
0.4642969	-0.0232080	821	859	1449	2348	2491	control
0.5312290	0.0064004	598	698	1254	2280	2251	control
0.4614687	-0.0232080	839	891	1449	2418	2491	control
0.4666666	-0.0349510	825	872	1402	2398	2591	control
0.4805389	-0.0689491	658	694	1297	1978	2248	control
0.5805330	0.0237679	551	543	1091	2046	1951	control
0.5575757	0.0237679	615	584	1091	2056	1951	control
0.5712648	0.0237679	555	561	1091	2056	1951	control
0.5833333	-0.0005179	494	540	1056	2052	1932	control
0.5325609	-0.0005179	515	585	1056	1918	1932	control
0.5402125	-0.0252101	599	606	1146	2030	2135	control
0.5848733	-0.0231023	434	483	947	1844	1705	control
0.5738831	-0.0231023	479	496	947	1832	1705	control
0.6130536	0.0486772	502	498	1070	2076	1798	control
0.5742972	-0.0153483	457	477	948	1764	1720	control
0.6461105	0.0934404	426	439	999	2042	1693	control
0.5850567	-0.0214168	307	311	653	1188	1240	control
0.5979044	-0.0214168	301	307	653	1220	1240	control
0.6000000	0.0204415	313	312	655	1248	1198	control
0.6272041	0.0204415	328	296	655	1292	1198	control
0.6275000	0.0272189	335	298	688	1302	1233	control
0.6070752	0.0204415	328	311	655	1272	1198	control
0.6362476	0.0204415	331	285	655	1282	1198	control
0.6038882	0.0272189	337	326	688	1320	1233	control
0.5872595	0.0174394	340	311	645	1196	1155	control

Table A.1: Multi-date data, 02-March 2017 (summer)- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.6538675	0.0980109	368	311	724	1486	1247	control
0.6428572	0.0716555	368	330	743	1518	1315	control
0.6493776	0.0716555	380	338	743	1590	1315	control
0.6385788	0.0934960	397	356	765	1614	1338	control
0.6109647	0.0934960	406	369	765	1528	1338	control
0.6547945	0.0716555	352	315	743	1510	1315	control
0.6907216	0.0716555	352	300	743	1640	1315	control
0.6713035	0.0934960	367	319	765	1622	1338	control
0.6475496	0.0934960	394	338	765	1580	1338	control
0.6334340	0.0825750	394	364	777	1622	1361	control
0.6675869	0.0820532	353	301	700	1510	1225	control
0.7691999	0.1854951	393	287	703	2200	1168	wetland
0.7662786	0.1854951	374	271	703	2048	1168	wetland
0.7717056	0.1581769	348	259	638	2010	1099	wetland
0.7839155	0.3493088	353	266	639	2196	1059	wetland
0.7614314	0.1581769	319	240	638	1772	1099	wetland
0.7720392	0.3493088	323	256	639	1990	1059	wetland
0.7351694	0.2278861	332	250	601	1638	1030	wetland
0.7557626	0.2278861	332	249	601	1790	1030	wetland
0.7605364	0.2980226	310	250	621	1838	994	wetland
0.7636276	0.1848617	327	245	602	1828	1120	wetland
0.7629063	0.2278861	323	248	601	1844	1030	wetland
0.7840459	0.2278861	336	245	601	2024	1030	wetland
0.7894737	0.2980226	339	260	621	2210	994	wetland
0.4957627	-0.1645190	435	476	813	1412	1943	control
0.4911894	-0.1645190	392	462	813	1354	1943	control

Table A.1: Multi-date data, 02-March 2017 (summer)- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.4692231	-0.1899300	469	526	954	1456	2127	control
0.4402391	-0.1722953	498	562	852	1446	2048	control
0.4806981	-0.1722953	422	491	852	1400	2048	control
0.4761120	-0.2073135	427	477	827	1344	2047	control
0.4401709	-0.2073135	430	524	827	1348	2047	control
0.7102297	0.1206566	414	347	814	2048	1607	wetland
0.6443629	0.1206566	401	388	814	1794	1607	wetland
0.6639419	0.1206566	370	370	814	1832	1607	wetland
0.6817982	0.1049483	374	361	1003	1908	1646	wetland
0.6108767	-0.0143431	387	415	795	1718	1768	wetland
0.6161417	-0.0143431	366	390	795	1642	1768	wetland
0.6094441	-0.0432748	362	397	946	1636	1784	wetland
0.5568983	-0.1325842	394	440	872	1546	2016	wetland
0.6027132	-0.0143431	388	410	795	1654	1768	wetland
0.5959900	-0.0143431	383	403	795	1592	1768	wetland
0.5688564	-0.0432748	365	443	946	1612	1784	wetland
0.5696203	-0.0895692	415	442	872	1612	1922	wetland

Note:

Data for the first date

Table A.2: Multi-date data, 17-November 2017- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve.

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.4048257	-0.1396735	509	666	927	1572	2199	control
0.3599349	-0.1429305	697	786	881	1670	2227	control

Table A.2: Multi-date data, 17-November 2017- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.3914674	-0.1429305	462	649	881	1484	2227	control
0.5056180	-0.2078037	335	440	718	1340	2043	control
0.5136911	-0.2078037	325	444	718	1382	2043	control
0.5568241	-0.2078037	308	427	718	1500	2043	control
0.5446382	-0.1066231	332	431	761	1462	1863	control
0.5024304	-0.0277306	519	563	911	1700	1816	wetland
0.5466418	-0.0295581	444	486	857	1658	1759	wetland
0.5617977	-0.0295581	437	468	857	1668	1759	wetland
0.5613187	-0.0005893	459	499	823	1776	1698	wetland
0.5715539	0.0047713	456	488	911	1790	1773	wetland
0.5896521	0.0047713	484	460	911	1782	1773	wetland
0.6920249	0.1780962	470	419	976	2302	1606	wetland
0.7134711	0.1076072	478	402	936	2404	1613	wetland
0.7212819	0.2330529	455	387	943	2390	1550	wetland
0.7165873	0.2330529	462	387	943	2344	1550	wetland
0.7303852	0.1780962	442	378	976	2426	1606	wetland
0.7107012	0.1780962	450	392	976	2318	1606	wetland
0.7337183	0.2500578	503	415	909	2702	1621	wetland
0.7153311	0.2500578	452	389	909	2344	1621	wetland
0.7186496	0.1815418	408	350	863	2138	1481	wetland
0.7230065	0.1815418	432	363	863	2258	1481	wetland
0.7375631	0.2261511	429	364	951	2410	1521	wetland
0.6984006	0.2261511	422	396	951	2230	1521	wetland
0.7077690	0.2500578	466	410	909	2396	1621	wetland
0.6944335	0.2500578	432	387	909	2146	1621	wetland
0.7230769	0.1815418	392	333	863	2072	1481	wetland

Table A.2: Multi-date data, 17-November 2017- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.7288528	0.1815418	404	351	863	2238	1481	wetland
0.7215190	0.2261511	418	363	951	2244	1521	wetland
0.6849476	0.2261511	450	405	951	2166	1521	wetland
0.6691909	0.1422935	457	415	941	2094	1739	wetland
0.6533546	0.0686629	461	434	937	2070	1804	wetland
0.6705787	0.0686629	472	407	937	2064	1804	wetland
0.6888533	0.1119324	452	388	923	2106	1682	wetland
0.6760563	0.1119324	461	414	923	2142	1682	wetland
0.6636037	0.1042164	461	421	955	2082	1689	wetland
0.6488906	0.1042164	491	451	955	2118	1689	wetland
0.6078431	0.0686629	499	480	937	1968	1804	wetland
0.6382145	0.1119324	471	462	923	2092	1682	wetland
0.6158730	0.1119324	483	484	923	2036	1682	wetland
0.6157240	0.1042164	490	479	955	2014	1689	wetland
0.5319802	-0.0811474	573	611	1082	2000	2205	wetland
0.6498873	0.0539957	471	466	1011	2196	1971	wetland
0.4389009	-0.0811474	724	776	1082	1990	2205	wetland
0.6166038	0.0539957	538	508	1011	2142	1971	wetland
0.5966244	-0.1302876	477	478	909	1892	2004	wetland
0.4704154	-0.1004435	594	631	989	1752	2109	wetland
0.5802469	0.0188379	502	510	954	1920	1849	wetland
0.6054097	0.0188379	504	496	954	2018	1849	wetland
0.5329140	-0.1004435	564	557	989	1828	2109	wetland
0.5903421	0.0370765	514	509	975	1976	1805	control
0.5808198	0.0370765	490	542	975	2044	1805	control
0.5817875	0.0411410	501	496	981	1876	1748	control

Table A.2: Multi-date data, 17-November 2017- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.5671830	0.0411410	489	517	981	1872	1748	control
0.5737958	0.0781415	527	553	922	2042	1746	control
0.6466902	0.0781415	447	451	922	2102	1746	control
0.6240481	0.0739464	458	469	992	2026	1747	control
0.5378549	0.0739464	570	586	992	1950	1747	control
0.5656214	0.0781415	584	561	922	2022	1746	control
0.6455973	0.0781415	472	485	922	2252	1746	control
0.6070922	0.0739464	521	554	992	2266	1747	control
0.5085886	0.0054754	634	658	1093	2020	1998	control
0.6219931	0.0054754	464	495	1093	2124	1998	control
0.5579869	0.0437332	562	606	1061	2136	1957	control
0.5158997	0.0054754	612	647	1093	2026	1998	control
0.5100821	0.0437332	581	656	1061	2022	1957	control
0.4425062	-0.0338379	689	783	1271	2026	2322	control
0.5322033	-0.0284213	559	621	1086	2034	2153	control
0.4550602	-0.0338379	656	770	1271	2056	2322	control
0.4442134	-0.0440363	685	802	1272	2084	2359	control
0.4888889	-0.0519119	602	644	1131	1876	2077	control
0.5701245	0.0169309	476	518	981	1892	1829	control
0.5042603	0.0169309	602	640	981	1942	1829	control
0.5414794	0.0169309	526	561	981	1886	1829	control
0.5495496	-0.0436038	446	500	914	1720	1807	control
0.4840794	-0.0436038	458	559	914	1608	1807	control
0.4872638	-0.0560626	600	624	1061	1810	2025	control
0.5342206	-0.0845387	421	490	835	1614	1687	control
0.5573384	-0.0845387	406	469	835	1650	1687	control

Table A.2: Multi-date data, 17-November 2017- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.6136662	0.0136209	440	441	927	1842	1738	control
0.5670732	-0.0335246	391	426	834	1542	1711	control
0.5690822	-0.0067278	402	446	844	1624	1646	control
0.5614576	-0.1138153	336	355	650	1264	1556	control
0.5913721	-0.1138153	355	341	650	1328	1556	control
0.6018465	-0.0427586	352	345	700	1388	1512	control
0.6276957	-0.0427586	352	328	700	1434	1512	control
0.6062118	-0.0307898	357	355	726	1448	1540	control
0.5865922	-0.0427586	400	370	700	1420	1512	control
0.5960805	-0.0427586	381	371	700	1466	1512	control
0.5915789	-0.0307898	396	388	726	1512	1540	control
0.6058547	0.0240000	420	377	768	1536	1464	control
0.6272252	0.0108875	369	356	731	1554	1499	control
0.6250000	0.0231106	365	378	789	1638	1564	control
0.6369516	0.0231106	361	393	789	1772	1564	control
0.6244384	0.0576192	411	418	831	1808	1611	control
0.6040516	0.0576192	451	430	831	1742	1611	control
0.6163022	0.0231106	377	386	789	1626	1564	control
0.6367816	0.0231106	399	395	789	1780	1564	control
0.6198347	0.0576192	411	414	831	1764	1611	control
0.6218721	0.0576192	421	408	831	1750	1611	control
0.6238362	0.0252707	412	404	822	1744	1620	control
0.6137592	0.0028240	416	393	808	1642	1589	control
0.7531045	0.2066738	448	338	835	2400	1474	wetland
0.7546244	0.2066738	430	325	835	2324	1474	wetland
0.7405159	0.1064843	417	342	822	2294	1502	wetland

Table A.2: Multi-date data, 17-November 2017- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.7665260	0.2718987	434	332	832	2512	1438	wetland
0.7436281	0.1064843	427	342	822	2326	1502	wetland
0.7734402	0.2718987	425	325	832	2544	1438	wetland
0.7377567	0.2114538	425	332	826	2200	1432	wetland
0.7603363	0.2114538	456	342	826	2512	1432	wetland
0.7750258	0.3020439	464	327	819	2580	1383	wetland
0.7721335	0.1662490	434	314	812	2442	1497	wetland
0.7852029	0.2114538	429	315	826	2618	1432	wetland
0.7886179	0.2114538	434	312	826	2640	1432	wetland
0.7979188	0.3020439	444	301	819	2678	1383	wetland
0.4933605	-0.1482112	410	496	807	1462	2022	control
0.4839039	-0.1482112	410	505	807	1452	2022	control
0.4611540	-0.1587715	516	593	945	1608	2226	control
0.4581006	-0.1616702	523	582	837	1566	2170	control
0.4665987	-0.1616702	470	523	837	1438	2170	control
0.4570519	-0.2114777	435	512	822	1374	2111	control
0.4658265	-0.2114777	452	551	822	1512	2111	control
0.6210721	0.0077564	378	410	757	1754	1727	wetland
0.5693327	0.0077564	377	455	757	1658	1727	wetland
0.5844930	0.0077564	339	418	757	1594	1727	wetland
0.5773295	-0.0071784	389	440	938	1642	1824	wetland
0.5765853	-0.0685879	384	434	758	1616	1854	wetland
0.5802469	-0.0685879	332	391	758	1472	1854	wetland
0.5330578	-0.1255156	336	452	862	1484	1910	wetland
0.5490873	-0.1230769	389	457	827	1570	2044	wetland
0.5703125	-0.0685879	366	440	758	1608	1854	wetland

Table A.2: Multi-date data, 17-November 2017- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.5698267	-0.0685879	351	422	758	1540	1854	wetland
0.5150340	-0.1255156	373	500	862	1562	1910	wetland
0.5224451	-0.1184391	436	500	875	1594	2035	wetland

Note:

Data for the second date

Table A.3: Multi-date data, 27 March 2018- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve.

ndvi	ndwi	B03	B04	B05	B08	B11	class
-0.0768392	0.3965375	521	732	898	1694	1976	control
-0.1120903	0.4380018	546	630	891	1612	2019	control
-0.1059983	0.4026644	510	695	891	1632	2019	control
-0.1644222	0.5740413	299	361	636	1334	1859	control
-0.1747235	0.5464772	283	383	636	1306	1859	control
-0.1195423	0.6119074	288	352	636	1462	1859	control
-0.0510172	0.6415809	280	331	692	1516	1679	control
0.0064300	0.5483872	495	525	894	1800	1777	wetland
0.0026541	0.5799256	392	452	826	1700	1691	wetland
0.0211288	0.6310679	371	399	826	1764	1691	wetland
0.0308802	0.6199547	378	419	783	1786	1679	wetland
0.0443745	0.6213334	396	426	843	1824	1669	wetland
0.0344229	0.6464088	393	384	843	1788	1669	wetland
0.2134463	0.7296455	438	347	909	2220	1439	wetland
0.2507772	0.7423313	416	357	851	2414	1446	wetland
0.3074961	0.7694962	428	334	914	2564	1358	wetland

Table A.3: Multi-date data, 27 March 2018- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.3057260	0.7705372	435	331	914	2554	1358	wetland
0.2767027	0.7535382	436	357	909	2540	1439	wetland
0.2867410	0.7493261	413	372	909	2596	1439	wetland
0.3041096	0.7743137	420	333	865	2618	1397	wetland
0.3092707	0.7777777	420	331	865	2648	1397	wetland
0.3276231	0.7421848	414	367	907	2480	1256	wetland
0.3207139	0.7467812	425	354	907	2442	1256	wetland
0.3498608	0.7792597	430	331	915	2668	1285	wetland
0.3319469	0.7482088	423	369	915	2562	1285	wetland
0.3160343	0.7854534	441	323	865	2688	1397	wetland
0.2885154	0.7563345	437	351	865	2530	1397	wetland
0.2863637	0.7190585	399	370	907	2264	1256	wetland
0.3068433	0.7443831	394	347	907	2368	1256	wetland
0.3357456	0.7704694	406	335	915	2584	1285	wetland
0.3040888	0.7417722	414	357	915	2408	1285	wetland
0.2027435	0.7328064	390	338	822	2192	1453	wetland
0.1767196	0.7429467	403	328	851	2224	1556	wetland
0.1878914	0.7106351	423	385	851	2276	1556	wetland
0.1939021	0.7057700	416	385	862	2232	1507	wetland
0.2281690	0.7497264	409	343	862	2398	1507	wetland
0.1964621	0.7215580	418	361	884	2232	1499	wetland
0.1729655	0.6886417	413	392	884	2126	1499	wetland
0.1000579	0.6618611	391	387	851	1902	1556	wetland
0.1449645	0.6746889	430	392	862	2018	1507	wetland
0.1891310	0.7218543	434	357	862	2210	1507	wetland
0.1241601	0.6586207	406	396	884	1924	1499	wetland

Table A.3: Multi-date data, 27 March 2018- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
-0.0341947	0.5744681	449	500	944	1850	1981	wetland
0.0728953	0.6978067	418	372	878	2090	1806	wetland
-0.0512072	0.4407736	618	694	944	1788	1981	wetland
0.0373134	0.6162791	459	462	878	1946	1806	wetland
-0.0815120	0.6046721	439	440	876	1786	2103	wetland
-0.0975412	0.4396552	597	650	941	1670	2031	wetland
-0.0226645	0.5639098	460	493	907	1768	1850	wetland
0.0609137	0.6547902	442	436	907	2090	1850	wetland
-0.0743190	0.4944492	529	592	941	1750	2031	wetland
0.0733689	0.5918675	501	542	1010	2114	1825	control
0.1060494	0.6209620	485	528	1010	2258	1825	control
0.0987212	0.5782513	525	574	961	2148	1762	control
0.0455038	0.6157388	441	459	961	1930	1762	control
0.0823648	0.5738428	547	580	1010	2142	1816	control
0.1290168	0.6571630	398	487	1010	2354	1816	control
0.1418616	0.6769006	399	442	980	2294	1724	control
0.0878307	0.6298057	475	467	980	2056	1724	control
0.0865191	0.5567568	620	615	1010	2160	1816	control
0.1506080	0.6621622	457	500	1010	2460	1816	control
0.1813865	0.7011966	433	437	980	2488	1724	control
0.0321961	0.5616140	595	603	1143	2148	2014	control
0.0878623	0.6514266	456	507	1143	2402	2014	control
0.1355476	0.6611949	460	516	1035	2530	1926	control
0.0710332	0.5346993	632	704	1143	2322	2014	control
0.0893617	0.5906109	472	593	1035	2304	1926	control
-0.0279735	0.4459974	781	872	1363	2276	2407	control

Table A.3: Multi-date data, 27 March 2018- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.0101289	0.5632348	477	613	1062	2194	2150	control
-0.0372765	0.4567982	763	833	1363	2234	2407	control
-0.0574122	0.5078821	690	718	1314	2200	2468	control
-0.0851170	0.4822581	571	642	1090	1838	2180	control
0.0194200	0.6148335	461	457	897	1916	1843	control
0.0067367	0.5631799	533	522	897	1868	1843	control
0.0183755	0.6148649	425	456	897	1912	1843	control
0.0092542	0.6234676	381	430	890	1854	1820	control
-0.0156250	0.5631369	375	493	890	1764	1820	control
-0.0432544	0.5431894	509	550	1004	1858	2026	control
0.0168878	0.5920680	372	432	796	1686	1630	control
-0.0111663	0.5955956	357	404	796	1594	1630	control
0.0696903	0.6643717	397	390	868	1934	1682	control
-0.0061125	0.5871156	360	423	811	1626	1646	control
0.0689056	0.6519015	376	389	855	1846	1608	control
-0.0151650	0.6477611	212	236	551	1104	1138	control
-0.0215440	0.6317365	209	246	551	1090	1138	control
0.0314465	0.6470588	238	246	533	1148	1078	control
0.0366398	0.6477273	239	248	533	1160	1078	control
0.0337463	0.6518772	241	255	549	1210	1131	control
0.0191083	0.6220130	253	261	533	1120	1078	control
0.0349149	0.6443812	255	250	533	1156	1078	control
0.0411191	0.6472166	260	263	549	1228	1131	control
0.0085547	0.6458486	247	241	530	1120	1101	control
0.1176006	0.6867183	281	263	611	1416	1118	control
0.0819165	0.6577857	255	289	624	1400	1188	control

Table A.3: Multi-date data, 27 March 2018- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.1219512	0.6764218	296	293	624	1518	1188	control
0.1285715	0.6719577	336	310	669	1580	1220	control
0.1081871	0.6586434	341	312	669	1516	1220	control
0.0733230	0.6842105	270	258	624	1376	1188	control
0.1206514	0.7088036	282	258	624	1514	1188	control
0.1291934	0.7102702	287	268	669	1582	1220	control
0.1153009	0.6975717	318	274	669	1538	1220	control
0.1164773	0.6750134	325	305	705	1572	1244	control
0.0682005	0.6666667	283	260	622	1300	1134	control
0.3181818	0.7904143	350	258	608	2204	1140	wetland
0.2490119	0.7836312	303	230	608	1896	1140	wetland
0.2293040	0.7851064	288	202	586	1678	1052	wetland
0.3359632	0.7990261	291	227	527	2032	1010	wetland
0.2218935	0.7811320	262	203	586	1652	1052	wetland
0.2902319	0.7938446	273	211	527	1836	1010	wetland
0.2148953	0.7785349	252	195	530	1566	1012	wetland
0.2274809	0.7836938	239	195	530	1608	1012	wetland
0.2821997	0.7953394	258	202	494	1772	992	wetland
0.2603878	0.8037662	234	198	530	1820	1068	wetland
0.3096862	0.8019709	245	211	530	1920	1012	wetland
0.3257828	0.8173516	273	200	530	1990	1012	wetland
0.3880320	0.8329939	291	205	494	2250	992	wetland
-0.1203732	0.5547004	354	405	708	1414	1801	control
-0.1380727	0.5473624	348	399	708	1364	1801	control
-0.1332956	0.5257114	409	475	878	1528	1998	control
-0.1066399	0.5198044	452	491	794	1554	1925	control

Table A.3: Multi-date data, 27 March 2018- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
-0.1427723	0.5353536	411	437	794	1444	1925	control
-0.1749154	0.5416427	338	399	692	1342	1911	control
-0.1627624	0.5095996	372	447	692	1376	1911	control
0.1070195	0.7049180	346	333	714	1924	1552	wetland
0.0324189	0.6299213	346	376	714	1656	1552	wetland
0.0761905	0.6952649	340	325	714	1808	1552	wetland
0.0725327	0.7230182	331	290	805	1804	1560	wetland
-0.0167529	0.6311269	329	365	688	1614	1669	wetland
0.0086130	0.6704378	334	335	688	1698	1669	wetland
-0.0476191	0.6391698	301	339	825	1540	1694	wetland
-0.1282808	0.5706695	365	404	821	1478	1913	wetland
-0.0112087	0.6360902	331	363	688	1632	1669	wetland
-0.0081546	0.6444666	332	355	688	1642	1669	wetland
-0.0399018	0.6224066	311	364	825	1564	1694	wetland
-0.0803782	0.5885656	376	403	750	1556	1828	wetland

Note:

Data for the third date

Table A.4: Multi-date data, 23 October 2018- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve.

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.3491215	-0.1174413	837	852	1010	1766	2236	control
0.2968868	-0.1423550	685	926	996	1708	2275	control
0.3652789	-0.1338151	784	808	996	1738	2275	control
0.4561403	-0.2301449	383	496	835	1328	2122	control

Table A.4: Multi-date data, 23 October 2018- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.4441386	-0.2358765	403	505	835	1312	2122	control
0.5082988	-0.1868009	392	474	835	1454	2122	control
0.5188583	-0.1196455	365	472	806	1490	1895	control
0.4733096	-0.0840708	563	592	972	1656	1960	wetland
0.5289719	-0.0709824	487	504	901	1636	1886	wetland
0.5337808	-0.0477778	487	521	901	1714	1886	wetland
0.5515695	-0.0351367	495	500	869	1730	1856	wetland
0.5537634	-0.0219966	473	498	919	1734	1812	wetland
0.5910113	-0.0117253	481	455	919	1770	1812	wetland
0.6535190	0.1306187	518	448	964	2138	1644	wetland
0.6793314	0.1368313	509	422	938	2210	1678	wetland
0.6809816	0.1561182	482	416	929	2192	1600	wetland
0.6736402	0.1578947	467	429	929	2200	1600	wetland
0.6775170	0.1481865	498	426	964	2216	1644	wetland
0.6828429	0.1450858	487	415	964	2202	1644	wetland
0.7062750	0.2148289	505	440	894	2556	1652	wetland
0.6904003	0.1817732	509	437	894	2386	1652	wetland
0.6884584	0.1545254	454	386	899	2092	1532	wetland
0.6831844	0.1498335	454	390	899	2072	1532	wetland
0.6960970	0.1788254	484	401	940	2238	1559	wetland
0.6868217	0.1651941	457	404	940	2176	1559	wetland
0.6770414	0.1618468	488	441	894	2290	1652	wetland
0.6767831	0.1444847	487	426	894	2210	1652	wetland
0.6873449	0.1422172	441	378	899	2040	1532	wetland
0.6918768	0.1596270	460	385	899	2114	1532	wetland
0.6933281	0.1638509	458	393	940	2170	1559	wetland

Table A.4: Multi-date data, 23 October 2018- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.6632695	0.1336482	437	413	940	2040	1559	wetland
0.6562248	0.0787340	477	428	905	2062	1761	wetland
0.6283042	0.0389206	462	457	924	2002	1852	wetland
0.6225031	0.0359188	480	463	924	1990	1852	wetland
0.6352306	0.0385224	485	439	943	1968	1822	wetland
0.6478873	0.0723014	487	450	943	2106	1822	wetland
0.6249494	0.0530184	495	463	969	2006	1804	wetland
0.6010166	0.0232810	489	471	969	1890	1804	wetland
0.5600170	-0.0070691	527	515	924	1826	1852	wetland
0.5964052	0.0349576	525	494	943	1954	1822	wetland
0.6159681	0.0525221	498	481	943	2024	1822	wetland
0.5570033	0.0290635	522	544	969	1912	1804	wetland
0.4838966	-0.1080306	628	649	1085	1866	2318	wetland
0.5602946	-0.0322270	535	567	1018	2012	2146	wetland
0.3737373	-0.1160328	752	837	1085	1836	2318	wetland
0.5072243	-0.0397287	634	648	1018	1982	2146	wetland
0.5266666	-0.1230253	541	568	1010	1832	2346	wetland
0.3919813	-0.1224540	714	781	1065	1788	2287	wetland
0.5411184	-0.0499366	548	558	1004	1874	2071	wetland
0.5803642	-0.0184411	534	530	1004	1996	2071	wetland
0.4621578	-0.1147940	653	668	1065	1816	2287	wetland
0.5310016	0.0020812	557	590	1008	1926	1918	control
0.5766082	0.0263959	493	543	1008	2022	1918	control
0.5648670	0.0178197	524	540	1000	1942	1874	control
0.5632378	-0.0053648	505	518	1000	1854	1874	control
0.5154052	0.0254826	623	637	1001	1992	1893	control

Table A.4: Multi-date data, 23 October 2018- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.5993716	0.0363960	505	510	1001	2036	1893	control
0.6099920	0.0466213	474	484	957	1998	1820	control
0.5725638	0.0162162	508	511	957	1880	1820	control
0.5028377	0.0239752	614	657	1001	1986	1893	control
0.6075206	0.0403042	524	501	1001	2052	1893	control
0.6354827	0.0931739	513	489	957	2194	1820	control
0.4928945	-0.0280010	637	678	1128	1996	2111	control
0.5647468	-0.0210399	526	563	1128	2024	2111	control
0.5880779	0.0058083	499	539	1027	2078	2054	control
0.4163424	-0.0265013	729	825	1128	2002	2111	control
0.5596478	-0.0264868	528	550	1027	1948	2054	control
0.3757109	-0.0912708	811	933	1332	2056	2469	control
0.4959096	-0.0889680	593	647	1128	1920	2295	control
0.4094707	-0.0990430	774	848	1332	2024	2469	control
0.4403409	-0.1034482	700	788	1287	2028	2496	control
0.4372294	-0.1055596	691	715	1201	1826	2257	control
0.5419723	-0.0150963	561	562	1016	1892	1950	control
0.4871592	-0.0177453	653	649	1016	1882	1950	control
0.5148314	-0.0225485	559	597	1016	1864	1950	control
0.5318401	-0.0503675	494	533	917	1744	1929	control
0.5277517	-0.0858429	482	502	917	1624	1929	control
0.4600574	-0.0830975	633	659	1067	1782	2105	control
0.4685599	-0.1269219	471	524	841	1448	1869	control
0.4686741	-0.1055901	478	547	841	1512	1869	control
0.5560599	-0.0413870	485	489	918	1714	1862	control
0.4956521	-0.0859167	498	522	901	1548	1839	control

Table A.4: Multi-date data, 23 October 2018- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.5443410	-0.0511740	430	465	845	1576	1746	control
0.5293376	-0.1188658	365	373	669	1212	1539	control
0.5536159	-0.1052065	350	358	669	1246	1539	control
0.5564854	-0.0616216	372	371	695	1302	1473	control
0.5776323	-0.0525188	360	355	695	1326	1473	control
0.5981254	-0.0448180	372	343	702	1364	1492	control
0.5534665	-0.0540251	371	380	695	1322	1473	control
0.5773017	-0.0391535	394	365	695	1362	1473	control
0.5919956	-0.0205199	400	367	702	1432	1492	control
0.5810956	-0.0003396	417	390	750	1472	1473	control
0.5961436	-0.0013405	391	377	759	1490	1494	control
0.5738375	-0.0025907	397	417	804	1540	1548	control
0.5972291	0.0208729	378	407	804	1614	1548	control
0.5840667	0.0344828	440	449	812	1710	1596	control
0.5827068	0.0268293	460	444	812	1684	1596	control
0.5864979	-0.0144168	408	392	804	1504	1548	control
0.6023739	0.0227273	395	402	804	1620	1548	control
0.5988566	0.0250458	407	421	812	1678	1596	control
0.5963946	-0.0006270	420	403	812	1594	1596	control
0.6054158	0.0112701	427	408	812	1660	1623	control
0.5836392	-0.0153244	411	397	809	1510	1557	control
0.6974339	0.1837607	463	395	858	2216	1528	wetland
0.6970671	0.1604396	456	377	858	2112	1528	wetland
0.6693192	0.1174904	451	374	802	1888	1491	wetland
0.7101670	0.2144397	451	382	796	2254	1458	wetland
0.6843034	0.1231990	428	358	802	1910	1491	wetland

Table A.4: Multi-date data, 23 October 2018- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.7172607	0.1904498	445	353	796	2144	1458	wetland
0.6987036	0.1312891	424	337	799	1900	1459	wetland
0.7023060	0.1636572	437	355	799	2030	1459	wetland
0.7174515	0.2065610	450	357	825	2170	1427	wetland
0.7153906	0.1832489	443	368	792	2218	1531	wetland
0.7343629	0.2124157	446	344	799	2246	1459	wetland
0.7289041	0.2162235	461	355	799	2264	1459	wetland
0.7223251	0.2395417	467	375	825	2326	1427	wetland
0.4760222	-0.1500000	454	519	829	1462	1978	control
0.4782173	-0.1453387	458	521	829	1476	1978	control
0.4663024	-0.1570681	545	586	999	1610	2210	control
0.4584451	-0.1412786	568	606	925	1632	2169	control
0.4524159	-0.1862182	515	561	925	1488	2169	control
0.4648478	-0.2070435	482	510	902	1396	2125	control
0.4534474	-0.1908098	491	543	902	1444	2125	control
0.5543278	-0.0225825	453	484	804	1688	1766	wetland
0.5191841	-0.0606607	453	495	804	1564	1766	wetland
0.5627452	-0.0511905	402	446	804	1594	1766	wetland
0.5482587	-0.0784720	395	454	953	1556	1821	wetland
0.5165496	-0.1079646	423	482	818	1512	1878	wetland
0.5551020	-0.1040564	403	436	818	1524	1878	wetland
0.5245365	-0.1552870	373	436	866	1398	1912	wetland
0.5000000	-0.1586957	434	516	931	1548	2132	wetland
0.5290168	-0.0817972	445	491	818	1594	1878	wetland
0.5620947	-0.0905924	435	439	818	1566	1878	wetland
0.5305598	-0.1240447	402	457	866	1490	1912	wetland

Table A.4: Multi-date data, 23 October 2018- S2A spectral bands/indices for wetland-non-wetland (control) classes at the Kogelberg Nature Reserve. *(continued)*

ndvi	ndwi	B03	B04	B05	B08	B11	class
0.4919431	-0.1313466	502	536	911	1574	2050	wetland

Note:

Data for the fourth date

References

Adam, E. & Mutanga, O. 2009. Spectral discrimination of papyrus vegetation (Cyperus papyrus L.) In swamp wetlands using field spectrometry. *ISPRS Journal of Photogrammetry and Remote Sensing.* 64:612–620. DOI: 10.1016/j.isprsjprs.2009.04.004.

Adam, E., Mutanga, O. & Rugege, D. 2010. Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review. *Wetlands Ecol Manage.* 18(3):281–296. DOI: 10.1007/s11273-009-9169-z.

Barron, O.V., Emelyanova, I., Niel, T.G.V., Pollock, D. & Hodgson, G. 2014. Mapping groundwater-dependent ecosystems using remote sensing measures of vegetation and moisture dynamics. *Hydrological Processes.* 28(2):372–385. DOI: 10.1002/hyp.9609.

Belgiu, M. & Drăguţ, L. 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing.* 114:24–31. DOI: 10.1016/j.isprsjprs.2016.01.011.

Bertrand, G., Goldscheider, N., Gobat, J.M. & Hunkeler, D. 2012. Review: From multi-scale conceptualization to a classification system for inland groundwater-dependent ecosystems. *Hydrogeol. J.* 20(1):5–25. DOI: 10.1007/s10040-011-0791-5.

Blaschke, T. 2010. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* 65(1):2–16. DOI: 10.1016/j.isprsjprs.2009.06.004.

Blomquist, W. & Ingram, H.M. 2003. Boundaries Seen and Unseen. *Water*

International. 28(2):162–169. DOI: 10.1080/02508060308691681.

Bond, W.J. & Keeley, J.E. 2005. Fire as a global "herbivore": The ecology and evolution of flammable ecosystems. *Trends in Ecology & Evolution*. 20(7):387–394. DOI: 10.1016/j.tree.2005.04.025.

Breiman, L. 2017. *Classification and Regression Trees*. ed. Routledge. DOI: 10.1201/9781315139470.

Brenning, A. 2012. Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sperrorest. In *2012 IEEE International Geoscience and Remote Sensing Symposium*. ed. 5372–5375. DOI: 10.1109/IGARSS.2012.6352393.

CapeNature. 2018. *BGIS (Biodiversity GIS)*. ed. SANBI. Available: `https://www.sanbi.org/link/bgis-biodiversity-gis/` [2018, June 10].

Chen, X., Vogelmann, J.E., Rollins, M., Ohlen, D., Key, C.H., Yang, L., Huang, C. & Shi, H. 2011. Detecting post-fire burn severity and vegetation recovery using multitemporal remote sensing spectral indices and field-collected composite burn index data in a ponderosa pine forest. *International Journal of Remote Sensing*. 32(23):7905–7927. DOI: 10.1080/01431161.2010.524678.

Colvin, C., Le Maitre, D.C., Saayman, I.C., Hughes, S., South Africa, Water Research Commission & CSIR Natural Resources and the Environment (Organization). 2007. *Aquifer dependent ecosystems in key hydrogeological typesettings in South Africa: Report to the Water Research Commission*. ed.

Colvin, C., Riemann, K., Brown, C., Le Maitre, D., Mlisa, A., Blake, D., Aston, T., Maherry, A., et al. 2009. Ecological And Environmental Impacts of large-scale groundwater development in the Table Mountain Group (TMG) aquifer system. *Water Research Commission. Gezina*.

Copernicus, E.S.A. 2017. *Copernicus Open Access Hub*. ed.

Cowardin, L.M., Carter, V., Golet, F.C. & LaRoe, E.T. 1979. *Classification of wetlands and deepwater habitats of the United States*. (Technical Report). U.S. Department of the Interior, U.S. Fish and Wildlife Service. Available:

`https://tamug-ir.tdl.org/handle/1969.3/20139` [2019, February 17].

Davies, B. & Day, J. 1998. *Vanishing Waters*. 1st ed. V. Extract fr. Cape Town: Cape Town: University of Cape Town.

Dawson, T.E., Hahm, W.J. & Crutchfield-Peters, K. 2020. Digging deeper: What the critical zone perspective adds to the study of plant ecophysiology. *New Phytologist*. 226(3):666–671. DOI: 10.1111/nph.16410.

Díaz-Delgado, R., Lloret, F. & Pons, X. 2004. Spatial patterns of fire occurrence in Catalonia, NE, Spain. *Landscape Ecol.* 19(7):731–745. DOI: 10.1007/s10980-005-0183-1.

Doody, T.M., Barron, O.V., Dowsley, K., Emelyanova, I., Fawcett, J., Overton, I.C., Pritchard, J.L., Van Dijk, A.I.J.M., et al. 2017. Continental mapping of groundwater dependent ecosystems: A methodological framework to integrate diverse data and expert opinion. *J. Hydrol. Reg. Stud.* 10(1):61–81. DOI: 10.1016/j.ejrh.2017.01.003.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., et al. 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*. 120:25–36. DOI: 10.1016/j.rse.2011.11.026.

Eamus, D. 2016. *Vegetation dynamics: A synthesis of plant ecophysiology, remote sensing and modelling*. ed. New York, NY: Cambridge University Press.

Eamus, D., Froend, R., Loomes, R., Hose, G. & Murray, B. 2006. A functional methodology for determining the groundwater regime needed to maintain the health of groundwater-dependent vegetation. *Aust. J. Bot.* 54(2):97–114. DOI: 10.1071/BT05031.

Eamus, D., Fu, B., Springer, A.E. & Stevens, L.E. 2016. Groundwater dependent ecosystems: Classification, identification techniques and threats. In *Integr. Groundw. Manag. Concepts, Approaches Challenges*. ed. Springer International Publishing. 313–346. DOI: 10.1007/978-3-319-23576-9_13.

Eamus, D. & Froend, R. 2006. Groundwater-dependent ecosystems: The

where, what and why of GDEs. *Aust. J. Bot.* 54(2):91–96. DOI: 10.1071/BT06029.

Escuin, S., Navarro, R. & Fernández, P. 2008. Fire severity assessment by using NBR (Normalized Burn Ratio) and NDVI (Normalized Difference Vegetation Index) derived from LANDSAT TM/ETM images. *Int. J. Remote Sens.* 29(4):1053–1073. DOI: 10.1080/01431160701281072.

February, E.C., Bond, W., Taylor, R. & Newton, R. 2004. Will water abstraction from the Table Mountain aquifer threaten endemic species? - A case study at Cape Point, Cape Town SO - South African Journal of Science. 100(5-6):253-255, 2004 May-Jun. *S. Afr. J. Sci.* 100(5-6):253–255.

Fox, J. 2015. *Applied regression analysis and generalized linear models.* ed. Sage Publications.

Frohn, R.C., D'amico, E., Lane, C., Autrey, B., Rhodus, J. & Liu, H. 2012. Multi-temporal sub-pixel landsat ETM+ classification of isolated wetlands in cuyahoga county, OHIO, USA. *Wetlands.* 32(2):289–299. DOI: 10.1007/s13157-011-0254-8.

Gangat, R., Deventer, H. van, Naidoo, L. & Adam, E. 2020. Estimating soil moisture using Sentinel-1 and Sentinel-2 sensors for dryland and palustrine wetland areas. *South African Journal of Science.* 116(7/8, 7/8). DOI: 10.17159/sajs.2020/6535.

Gao, B.C. 1996. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* 58(3):257–266. DOI: 10.1016/S0034-4257(96)00067-3.

Groom, P.K., Froend, R.H. & Mattiske, E.M. 2000. Impact of groundwater abstraction on a Banksia woodland, Swan Coastal Plain, Western Australia. *Ecol. Manag. Restor.* 1(2):117–124. DOI: 10.1046/j.1442-8903.2000.00033.x.

Guirado, E., Alcaraz-Segura, D., Rigol-Sánchez, J.P., Gisbert, J., Martínez-Moreno, F.J., Galindo-Zaldívar, J., González-Castillo, L. & Cabello, J. 2018. Remote-sensing-derived fractures and shrub patterns to identify groundwater dependence.

In *Ecohydrology*. ed. V. 11. DOI: 10.1002/eco.1933.

Han, M., Zhao, C., Feng, G., Disse, M., Shi, F. & Li, J. 2015. *An eco-hydrological approach to predicting regional vegetation and groundwater response to ecological water conveyance in dryland riparian ecosystems*. ed. V. 380-381. DOI: 10.1016/j.quaint.2015.02.032.

Hatton, T.J. & Evans, R. 1998. *Dependence of Ecosystems on Groundwater and its Significance to Australia. Land and Water Resources Research and Development Corporation Occasional Paper No 12/98*. (12). Land and Water Resources Research and Development Corporation.

Hector, A. 2015. *The new statistics with R: An introduction for biologists*. ed. Oxford: University Press.

Hijmans, R.J., Van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J.A., Lamigueiro, O.P., Bevan, A., et al. 2015. Package "raster". *R package*.

Holmes, R.M. 2000. The importance of ground water to stream ecosystem function. *Streams and Ground Waters*. 137–148.

Hope, A., Albers, N. & Bart, R. 2012. Characterizing post-fire recovery of fynbos vegetation in the Western Cape Region of South Africa using MODIS data. *International Journal of Remote Sensing*. 33(4):979–999. DOI: 10.1080/01431161.2010.543184.

James, G., Witten, D., Hastie, T. & Tibshirani, R. 2013. *An introduction to statistical learning: With applications in R*. ed. (Springer texts in statistics, 103). New York: Springer.

Jin, X.M., Schaepman, M.E., Clevers, J.G.P.W., Su, Z.B. & Hu, G.C. 2011. Groundwater depth and vegetation in the Ejina area, China. *Arid L. Res. Manag.* 25(2):194–199. DOI: 10.1080/15324982.2011.554953.

Jolliffe, I.T. Ed. 2002. Principal Components in Regression Analysis. In *Principal Component Analysis*. ed. (Springer Series in Statistics). New York, NY: Springer. 167–198. DOI: 10.1007/0-387-22440-8_8.

Kassambara, A. 2017. *Practical guide to cluster analysis in R: Unsupervised machine learning.* ed. V. 1. Sthda.

Keppel, G., Van Niel, K.P., Wardell-Johnson, G.W., Yates, C.J., Byrne, M., Mucina, L., Schut, A.G.T., Hopper, S.D., et al. 2012. Refugia: Identifying and understanding safe havens for biodiversity under climate change. *Glob. Ecol. Biogeogr.* 21(4):393–404. DOI: 10.1111/j.1466-8238.2011.00686.x.

Klemas, V. 2011. Remote Sensing of Wetlands: Case Studies Comparing Practical Techniques. *J. Coast. Res.* 27(3):418–427. DOI: 10.2112/JCOASTRES-D-10-00174.1.

Kløve, B., Ala-aho, P., Bertrand, G., Boukalova, Z., Ertürk, A., Goldscheider, N., Ilmonen, J., Karakaya, N., et al. 2011. Groundwater dependent ecosystems. Part I: Hydroecological status and trends. *Environ. Sci. Policy.* 14(7):770–781. DOI: 10.1016/j.envsci.2011.04.002.

Le, S. & Husson, F. 2008. FactoMineR: An R Package for Multivariate Analysis. 25:1–18. DOI: 10.18637/jss.v025.i01.

Lim, C. & Kafatos, M. 2002. Frequency analysis of natural vegetation distribution using NDVI/AVHRR data from 1981 to 2000 for North America: Correlations with SOI. *International Journal of Remote Sensing.* 23(17):3347–3383. DOI: 10.1080/01431160110110956.

Liu, L., Wang, J., Huang, W., Zhao, C., Zhang, B. & Tong, Q. 2004. Estimating winter wheat plant water content using red edge parameters. *International Journal of Remote Sensing.* 25(17):3331–3342. DOI: 10.1080/01431160310001654365.

Lovelace, R., Nowosad, J. & Muenchow, J. 2019. *Geocomputation with R.* ed. CRC Press. Available: `http://books.google.com?id=8W2PDwAAQBAJ`.

Lv, J., Wang, X.S., Zhou, Y., Qian, K., Wan, L., Eamus, D. & Tao, Z. 2013. Groundwater-dependent distribution of vegetation in Hailiutu River catchment, a semi-arid region in China. *Ecohydrology.* 6(1):142–149. DOI: 10.1002/eco.1254.

Max, K. 2020. *Caret: Classification and Regression Training*. ed. Available: `https://CRAN.R-project.org/package=caret`.

Maxwell, A.E., Warner, T.A. & Fang, F. 2018. Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*. 39(9):2784–2817. DOI: 10.1080/01431161.2018.1433343.

McLaughlin, B.C., Ackerly, D.D., Klos, P.Z., Natali, J., Dawson, T.E. & Thompson, S.E. 2017. Hydrologic refugia, plants, and climate change. *Glob. Chang. Biol.* 23(8):2941–2961. DOI: 10.1111/gcb.13629.

Meyer, H. 2020. *CAST: 'Caret' Applications for Spatial-Temporal Models*. ed. Available: `https://github.com/HannaMeyer/CAST`.

Meyer, H. & Pebesma, E. 2020. *Predicting into unknown space? Estimating the area of applicability of spatial prediction models*. ed. Available: `http://arxiv.org/abs/2005.07939` [2020, November 15].

Miller, H.J. 2004. Tobler's First Law and Spatial Analysis. *Annals of the Association of American Geographers*. 94(2):284–289. DOI: 10.1111/j.1467-8306.2004.09402005.x.

Mitsch, W.J. & Gosselink, J.G. 2000. The value of wetlands: Importance of scale and landscape setting. *Ecological Economics*. 35(1):25–33. DOI: 10.1016/S0921-8009(00)00165-8.

Mohd Razali, N. & Yap, B. 2011. Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *J. Stat. Model. Analytics*. 2.

Münch, Z. & Conrad, J. 2007. Remote sensing and GIS based determination of groundwater dependent ecosystems in the Western Cape, South Africa. *Hydrogeol. J.* 15(1):19–28. DOI: 10.1007/s10040-006-0125-1.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., Fonseca, G.A.B. da & Kent, J. 2000. Biodiversity hotspots for conservation priorities. *Nature*. 403(6772):853–858. DOI: 10.1038/35002501.

Pal, M. & Mather, P.M. 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*. 86(4):554–565. DOI: 10.1016/S0034-4257(03)00132-9.

Pettorelli, N. 2013. *The Normalized Difference Vegetation Index*. ed. Oxford: Oxford University Press. DOI: 10.1093/acprof:osobl/9780199693160.001.0001.

Probst, P., Wright, M. & Boulesteix, A.-L. 2019. Hyperparameters and Tuning Strategies for Random Forest. *WIREs Data Mining Knowl Discov*. 9(3). DOI: 10.1002/widm.1301.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. ed. Vienna, Austria: R Foundation for Statistical Computing. Available: `https://www.R-project.org/`.

Rutherford, M.C., Powrie, L.W., Husted, L.B. & Turner, R.C. 2011. Early post-fire plant succession in Peninsula Sandstone Fynbos: The first three years after disturbance. *South African J. Bot.* 77(3):665–674. DOI: 10.1016/j.sajb.2011.02.002.

Sawyer, S.F. 2009. Analysis of Variance: The Fundamental Concepts. *Journal of Manual & Manipulative Therapy*. 17(2):27E–38E. DOI: 10.1179/jmt.2009.17.2.27E.

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J. & Brenning, A. 2019. Performance evaluation and hyperparameter tuning of statistical and machine-learning models using spatial data. *Ecological Modelling*. 406:109–120. DOI: 10.1016/j.ecolmodel.2019.06.002.

Secretariat, R. 2010. *The Ramsar Handbooks for the Wise Use of Wetlands*. ed. Gland, Switzerland: Ramsar Secretariat.

Serov, P.A. & Kuginis, L. 2017. A groundwater ecosystem classification - the next steps. *Int. J. Water*. 11(4):328. DOI: 10.1504/IJW.2017.10008847.

Slingsby, J.A., Merow, C., Aiello-Lammens, M., Allsopp, N., Hall, S., Kilroy Mollmann, H., Turner, R., Wilson, A.M., et al. 2017. Intensifying postfire weather and biological invasion drive species loss in a Mediterranean-type

biodiversity hotspot. *Proc. Natl. Acad. Sci.* 114(18):4697–4702. DOI: 10.1073/pnas.1619014114.

Tiner, R.W. 2003. Geographically isolated wetlands of the United States. *Wetlands.* 23(3):494–516. DOI: 10.1672/0277-5212(2003)023[0494:GIWOTU]2.0.CO;2.

Umweltbundesamt. 2011. *Water Framework Directive - The way towards healthy waters.* ed. Umweltbundesamt. Available: `https://www.umweltbundesamt.de/en/publikationen/water-framework-directive-way-towards-healthy-0` [2019, October 23].

White, D.C., Lewis, M.M., Green, G. & Gotch, T.B. 2016. *A generalizable NDVI-based wetland delineation indicator for remote monitoring of groundwater flows in the Australian Great Artesian Basin.* ed. V. 60. DOI: 10.1016/j.ecolind.2015.01.032.

Wickham, H. 2016. *Ggplot2: Elegant graphics for data analysis.* ed. springer.

Wilson, A.M., Latimer, A.M. & Silander, J.A. 2015. Climatic controls on ecosystem resilience: Postfire regeneration in the Cape Floristic Region of South Africa. *Proc Natl Acad Sci USA.* 112(29):9058–9063. DOI: 10.1073/pnas.1416710112.

Xie, Y. 2020. *Bookdown: Authoring Books and Technical Documents with R Markdown.* ed.