Predicting mire distribution using species distribution models (SDMs) and GISc: a case study of the Prince Edward Islands (PEIs)

By

Maleho Mpho Sadiki

Submitted in fulfilment of the requirements for the degree

Master of Geoinformatics

in the

Faculty of Natural and Agricultural Sciences

University of Pretoria

February 2022

Declaration

I, Maleho Mpho Sadiki, that the thesis/dissertation, which I hereby submit for the degree Master of Geoinformatics at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE: 

DATE: 07 February 2022

Predicting mire distribution using species distribution models (SDMs) and GISc: a case study of the Prince Edward Islands (PEIs)

Student:	Maleho Mpho Sadiki
Supervisor:	Dr. C.D Hansen (*Department of Geography, Geoinformatics and Meteorology, University of Pretoria)*
Co-supervisor:	Prof. M Greve (*Department of Plant and Soil Sciences, University of Pretoria)*
Degree:	MSc Geoinformatics
	Faculty of Natural and Agricultural Sciences, Department of Geography, Geoinformatics and Meteorology, University of Pretoria

Abstract

Peatlands are wetlands with peat-producing plants that account for one-third of wetlands worldwide and provide a variety of ecological functions and ecosystem services such as carbon storage, biomass production, biodiversity conservation, and climate regulation. As important ecological systems that are vulnerable to climate change, it is critical to assess what drives mire distribution in order to predict the potential impact of climate change on their distribution. As a Special Nature Reserve, the Prince Edward Islands (PEIs) are important conservation areas for South Africa. They support extensive peatlands, which are actively accumulating peat, and are, therefore, referred to as mires. The Islands have experienced severe reductions in precipitation and significant warming in the last decades; anecdotal evidence suggests that these have affected the occurrence and extent of mires on the PEIs. Factors that drive mire occurrence are unclear and must be identified in order to improve the ability to monitor them over time and this can be achieved using Species distribution models (SDMs). SDMs are a significant tool in studies on species distribution, the ecological consequences of climate change, and efforts to protect specific species or biodiversity as a whole. Predictive models have been used effectively to map and detect wetlands at both the local and regional levels. The aim of this study was to use species distribution modelling to understand the drivers and predict the island-wide distribution of mires on the PEIs.

A total of 1415 mire presence-absence points from a vegetation field survey conducted on Marion Island from 2018 to 2020 were used. As there is no single best SDM algorithm, and it is difficult to accurately identify which environmental variables drive the distribution of mires on the PEIs, multiple regression-based and machine learning SDMs based on six different combinations of environmental factors were investigated. The environmental variables combinations included climate variables, topographic, geology and soils and satellite imagery variables, a combination thereof and wetland classification proxy variables from three wetland classification systems (Ramsar, Hydrogeomorphic (HGM), International Union for Conservation of Nature (IUCN) Global Ecosystem Typology 2.0 wetland classification systems).

Random Forest model performed the best, only performing fairly in terms of the AUC (0.74) and TSS (0.42) metrics but managing a 99% correct classification rate (CCR) of all the mire presence-absence

observations when trained and tested on Marion Island. The distribution of mires was largely influenced by surface wetness and slope. Low annual mean temperature, low temperature and precipitation seasonality, and increasing distance from coast (up to 7.2 km inland) also influenced the distribution of mires on the PEIs, though less strongly than surface wetness and slope. According to the model predictions, mires occupy 8.7 km² (of ~290 km²; ~ 3%) of Marion Island and 2.63 km² (of ~ 45 km²; ~6%) of Prince Edward Island respectively.

The predictive performance and reliability of the models can be improved by making enhancements to the datasets of environmental variables in terms of resolution. This is especially true for the spatial, temporal, and spectral resolutions of satellite imagery used to model environmental variables, the spatial resolution of the WorldClim climate data (which is currently based on data from only one meteorological station on Marion Island), and the spatial resolution and accuracy of the geology dataset. The inclusion of other environmental variables may also improve the predictive ability of the models in this study.

Acknowledgements

This thesis would not have been possible without the help and support of many people, and I would like to convey my profound appreciation to each and every one of them. To my friends and family, thank you for your unrelenting support. Thank you to my supervisors, Dr. Christel Hansen and Professor Michelle Greve, for your knowledge and for continuously making important suggestions to improve this work. Also, thank you to the South African National Space Agency (SANSA) for your ongoing financial and personal support.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AMERL	Automated Method of Extracting Rivers and Lakes
AUC	Area Under the Curve
AWEI	Automated Water Extraction Index
BRT	Boosted Regression Trees
CART	Classification and Regression Trees
DEM	Digital Elevation Model
GIS	Geographic Information Systems
GLM	Generalised Linear Model
HGM	Hydrogeomorphic
IDW	Inverse Distance Weighting
IPCC	Intergovernmental Panel for Climate Change
IUCN	International Union for Conservation of Nature
MARS	Multivariate Adaptive Regression Splines
m a.s.l.	Metres above sea level
MIR	Middle Infrared
MNDWI	Modified Normalised Difference Water Index
NDVI	Normalised Difference Vegetation Index
NDWI	Normalised Difference Water Index
NIR	Near-Infrared
PEIs	Prince Edward Islands
RF	Random Forest
ROC	Receiver Operating Characteristics
SDM	Species Distribution Model
SVM	Support Vector Machine
TPI	Topographic Position Index
TWI	Topographic Water Index
WI	Water Index

Chapter 1: Introduction

1.1. Background

The natural greenhouse effect is affected and enhanced by anthropogenic activities, which release greenhouse gases into the atmosphere, increasing global temperatures (IPCC, 2021). As the global radiative forcing becomes increasingly positive, the global climate system changes in an unprecedented manner, causing greater variability in the climate and frequent and more intense weather events (IPCC, 2021). Climate influences ecosystems, species, and processes such as primary production rates and chemical element input–output balance (Grimm et al., 2013), and are threatened by the changing global climate. Water, an integral component of the climate system, is critical for human and ecological systems (Huang et al., 2018a), and simultaneously affected by climate change through various mechanisms (Bates et al., 2008; Stagl et al., 2014; Huang et al., 2018a). Extreme changes and variation to the water cycle processes such as precipitation, water vapour contents, evaporation, soil moisture, and runoff are closely associated with a warming climate, affecting the hydrological cycle (Bates et al., 2008).

The effect of climate change of water can be seen through the loss of wetland areas globally. Wetlands are essential ecosystems, offering invaluable ecosystem services such as food, fresh water, energy resources, controlling erosion, providing habitats for wetland-dependent species, and contribute to the well-being of people and environments globally (MEA, 2005; Amler et al., 2015; Hu et al., 2017; Ramsar Convention on Wetlands, 2018). Although wetland ecosystems are naturally dynamic over seasons, years, and longer climatic cycles, anthropogenic pressures such as wetland drainage, pollution, vegetation clearance, livestock grazing, land use changes, tourism, and overexploitation of their natural resources, as well as climate change, which will have an impact on the hydrology of specific wetland ecosystems primarily through changes in precipitation and temperature regimes, have increased rate of change in wetlands, resulting in faster degradation and greater biodiversity losses than other ecosystems (MEA, 2005; Erwin, 2009; Bassi et al., 2014; Finlayson et al., 2017).

Mires are peatlands. They are described as wetlands with vegetation that forms peat, that is, vegetation that has not completely decomposed and has accumulated due to a water table at or near the surface (Rydin et al., 1999; Joosten and Clarke, 2002; Dartnall and Smith, 2012; Rydin et al., 2013). Their formation is governed by the climate; depending on cool and humid climatic conditions (Essl et al., 2012). As mires depend on cool and humid climate conditions, they are vulnerable to any changes thereof (Essl et al., 2012). Carbon storage, biomass production, and climate regulation are just a few of the ecosystem services provided by peatlands and, as they deteriorate due to various factors, including those of climate change or land use change, they release stored carbon into the atmosphere, contributing to greenhouse gas emissions (Joosten and Clarke, 2002; Minasny et al., 2019). Consequently, peatlands are important in the global carbon cycle and climate regulation (FAO, 2020).

This study looks at the distribution of mires on the Prince Edward Islands (PEIs). The PEIs are South African-owned, remote islands in the sub-Antarctic Ocean made up of Marion Island and Prince Edward

Island (Figure 1-1, pg. 2). They are located approximately 2180 km southeast of Cape Town and are politically part of the South African province of the Western Cape (Smith and Mucina, 2006).

Figure 1-1: The sub-Antarctic Islands and surrounding Southern Ocean (Data source: Matsuoka et al. (2021)).

The PEIs have a fairly stable annual climate, characterised by low temperature, regular rainfall, humidity and strong winds (Smith, 2002; Pakhomov and Chown, 2003; Smith and Mucina, 2006; le Roux and McGeoch, 2008). As a result of the oceanic climate, water bodies are common (Dartnall and Smith, 2012), and peat formation is favoured. Subsequently, terrestrial vegetation of the PEIs is characterised by the presence of mires (Gremmen, 1981; Essl et al., 2012). Figure 1-2 depicts a mire on Marion Island.

The Islands are mostly unoccupied except for a few research scientists visiting for short periods, or a handful of personnel residents there to ensure research installations and the research base on Marion Island remain operational. The Islands are, thus, isolated and minimally affected by direct human impacts. However, studies on Marion Island and other sub-Antarctic islands indicate changes to their climate (Frenot et al., 1997; Kirkpatrick and Scott, 2002; Smith, 2002; Pendlebury and Barnes-Keoghan, 2007; le Roux and McGeoch, 2008). While Marion Island has seen substantial warming, decreased precipitation and increased wind speeds since the 1960s (le Roux, 2008; le Roux and McGeoch, 2008), the changes to climate on the sub-Antarctic islands have been heterogenous across the region. Kerguelen Island, Macquarie Island and Heard Island have also seen a decrease in temperature, although at different rates compared to Marion Island (Frenot et al., 1997; Budd, 2000; Kirkpatrick and Scott, 2002; Pendlebury and Barnes-Keoghan, 2007). Furthermore, unlike on Marion Island, precipitation on Macquarie Island has increased (Pendlebury and Barnes-Keoghan, 2007). Such

changes in the climate will have a direct impact on the indigenous biota (Smith and Steenkamp, 1990; Smith et al., 2001; Smith, 2002). Furthermore, anecdotal evidence suggests that water bodies on Marion Islands are shrinking and that the Island is becoming drier, including mires where peat moisture content is decreasing (Chown and Smith, 1993; Sadiki, 2019). Selkirk (2007) reported drier conditions in mires on the sub-Antarctic islands due to decreased precipitation in some parts of the region and increasing windspeed.

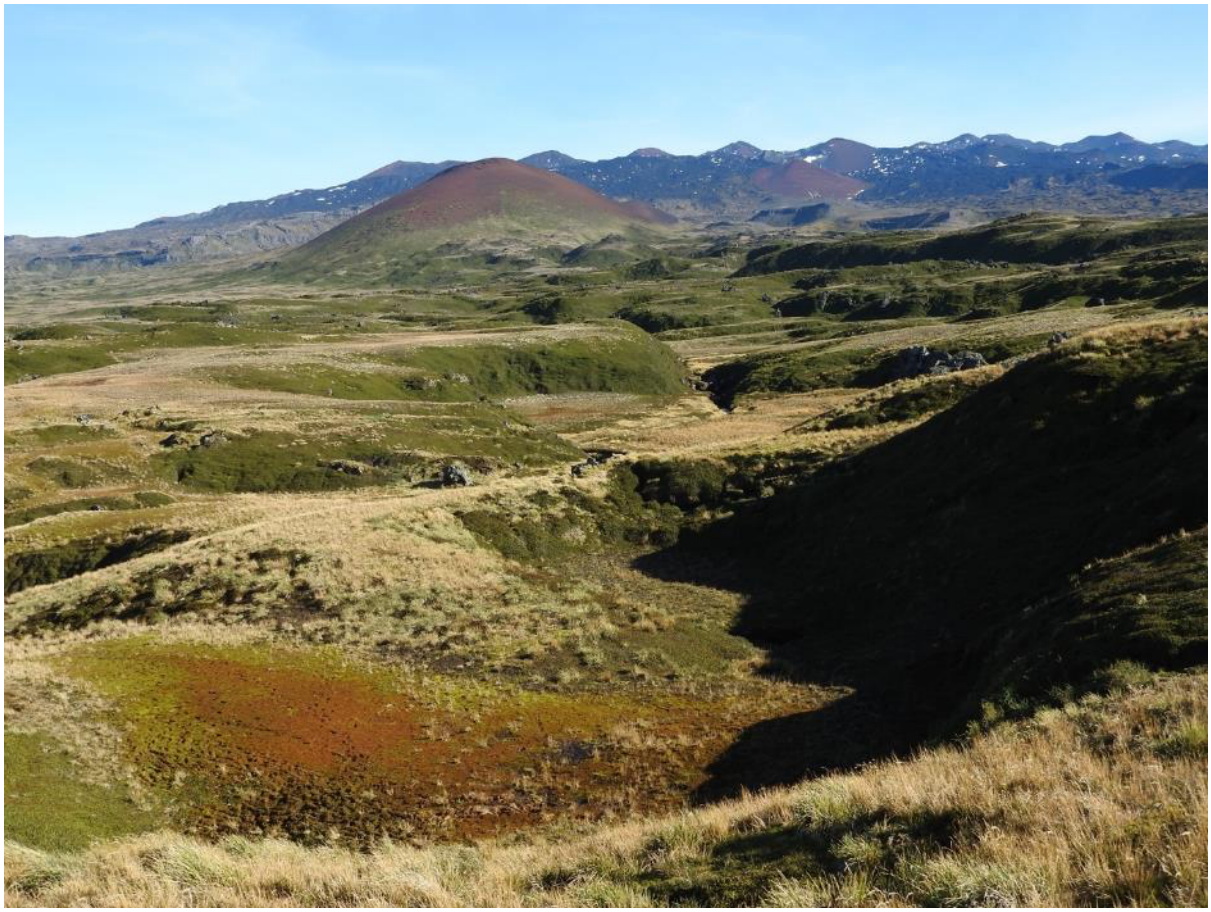


Figure 1-2: A mire on Marion Island (Image source: M Greve).

Mire ecosystems' geographical distribution is crucial for their conservation and management. In the face of climate change and related ecological changes, it is critical to understand what causes mire distribution in order to comprehend their current and future dynamics (Stagl et al., 2014; Harenda et al., 2018). Compared to resource intensive, on-the-ground methods, remote sensing offers efficient, effective, and reliable ways to identify and observe surface water (Torbick et al., 2007; Zhai et al., 2015). Geographic Information Systems (GIS) in turn provide powerful platforms to model environmental parameters that may be used as proxies for identifying mire locations. However, the identification and delineation of mires from satellite imagery is more complex than delineating normal open water bodies, as mires constitute systems with different spectral properties, including open water, vegetation dominated by different species, and a mixture thereof (Kaplan and Avdan, 2017). The vegetation of the Prince Edward Islands varies greatly in an area, with small areas, especially at low elevations below 200 m, often containing several vegetation types (Smith and Mucina, 2006), making it difficult to discern between mire vegetation and other vegetation types using only satellite imagery. As a result, other factors besides the existence of plant and water are required to monitor their change through time. As

such, factors that influence their distribution must be established in order to determine how they might be tracked through time.

Species distribution models (SDMs), which characterise and predict the occurrence of particular species by evaluating the connection between known species occurrences and environmental factors thought to affect their occurrence, are often used in research into the distribution of species, ecological repercussions of climate change, as well as attempts to conserve particular species or biodiversity as a whole (Guisan and Zimmermann, 2000; McPherson et al., 2004; Franklin, 2009). Predictive models have been used successfully at both the local and regional levels to map and detect wetlands (Hunter et al., 2012; Hiestermann and Rivers-Moore, 2015; Rebelo et al., 2017). Therefore, this study employs SDMs to determine the drivers of mire distribution of the PEIs. Almost always, the data necessary to train SDMs requires substantial preparation in GIS (Brown, 2014). As such, while SDMs were built outside of GIS, the data in this study were pre-processed, visualised and results prepared (into binary prediction maps) in a GIS (QGIS or ArcGIS Pro). In addition, remote sensing methods were used to create and pre-process data required for input into the SDMs. As such, this study uses SDMs, GIS and remote sensing to simulate distribution of mires on the PEIs landscapes.

1.2. Problem Statement

The PEIs were declared a Special Nature Reserve by South Africa in 1995 and thus hold a special protection status regarding South African territory, with human activities limited to conservation and scientific research (Chown et al., 2006; Smith and Mucina, 2006). Although they are among the most isolated locations on the planet with limited direct human impact, they have been experiencing a warmer and drier climate. Change to the climate has a direct influence on the hydrological system. Anecdotal evidence suggests wetlands on Marion Island are changing, with lakes shrinking, while mires become drier and shallower (Hedding and Greve, 2018). Furthermore, preliminary evidence suggests that lakes are becoming smaller in extent (Sadiki, 2019). Mires represent an important habitat and are integral to the hydrological and biogeochemical water cycles on the PEIs, consequently affecting the ecology of the Islands.

In the context of climate change and associated ecological changes, it is important to estimate the future dynamics and spatial distribution of mire ecosystems. Knowing where mires are located and how they might change under different climate scenarios, in turn, is fundamental to conserve and better manage these features. Although research has been conducted on the mires on Marion Island, the focus has been on features such as their overall location or distribution, the vegetation dynamics within them, and the influence of climate change (Gremmen, 1981; Smith and Steenkamp, 1990; Chown and Smith, 1993; Smith and Mucina, 2006). No research has been conducted to attempt to understand what factors dictate their occurrence. This research applies remote sensing, GIS, and species distribution modelling to determine the key ecological drivers for the existence of mires on the PEIs, to provide a reference state of mire distribution that can be used in future studies to monitor changes to their distribution over time.

1.3. Aim and Objectives

The aim of this study was to use species distribution modelling, utilising variables prepared using GIS and remote sensing methods, to predict and understand the distribution of mires on the PEIs. To achieve this aim, two key objectives were identified.

Objective 1: Identify the best algorithm to predict the distribution of mires on the PEIs.
This involves testing the performance of different species distribution models based on different sets of predictor variables to determine the best model and predictor variables to be used to predict the distribution of mires on the PEIs.

Objective 2: Determine and understand the main drivers of the distribution of mires on the PEIs.
This involves in-depth analysis of the contributions of the predictor variables.

Chapter 2: Literature Review

2.1. Wetlands

Water is a vital part of life on earth; thus, there is a widespread need for understanding the spatial and temporal variations of surface water as changes thereof impact the human population, the environment, and other natural resources (Alsdorf et al., 2007; Huang et al., 2018a). Surface water is dynamic, changing between states of matter, and continuously being redistributed as a result of both natural and anthropogenic factors (Coppin et al., 2004). With advances in earth observations and Geographic Information Systems (GIS), the observation of water from space becomes an elementary remote sensing application and a fundamental part of exploring hydrological processes (Huang et al., 2018a). As such, research on the impacts of different factors such as climate change, urbanisation, and pollution on surface water have been conducted in many parts of the world.

Wetland ecosystems are defined as “permanently or seasonally inundated freshwater habitats ranging from lakes and rivers to marshes, as well as coastal and marine areas such as estuaries, lagoons, mangroves, and reefs” (Ramsar Convention on Wetlands, 2018, p. 11).They are essential to life, offering invaluable ecosystem services such as food, fresh water, energy resources, controlling erosion, providing habitats for wetland-dependent species, and contribute to the well-being of people and environments globally (MEA, 2005; Amler et al., 2015; Ramsar Convention on Wetlands, 2018). In a natural cycle, wetland ecosystems are particularly dynamic across seasons, years, and larger climatic cycles; however, anthropogenic pressures and the impacts of climate change have exacerbated their rate of change, leading to a rapid degradation and higher losses in biodiversity compared to other ecosystems (MEA, 2005). To monitor wetlands and prevent further loss, their distribution across the landscape should be known.

Peatlands are wetlands with vegetation that creates peat, that is, plants depositing at the surface without entirely decomposing due to deposition occurring at or near the water table (Rydin et al., 1999; Joosten and Clarke, 2002; Dartnall and Smith, 2012; Rydin et al., 2013). They depend on cool and humid climatic conditions, along with low evaporation rates, and high effective moisture preferred (Yu et al., 2009; Essl et al., 2012; Harenda et al., 2018). Peatlands account for one-third of all wetlands globally (approximately 3% of the earth’s surface), and they provide a variety of ecological services such as carbon storage, biomass production, biodiversity conservation, and climate regulation (Joosten, 2012; Grundling et al., 2017; Minasny et al., 2019). However, as they are deteriorating as a result of climate change and land use change, stored carbon is released into the atmosphere, contributing to greenhouse gas emissions (Joosten and Clarke, 2002; Harenda et al., 2018; Minasny et al., 2019). In areas that are experiencing drying as a result of climate change, the high water table level required for peatlands is lowered, enabling oxygen to permeate the peatlands, increasing peat degradation and consequently a fast release of stored carbon into the atmosphere (FAO, 2020).

Mires are peatlands where the peat is actively accumulating (Joosten and Clarke, 2002). Gremmen (1981) studied the vegetation on the PEIs, distinguishing 41 plant communities, which were grouped into six community complexes, identifying the peatlands on the Islands as mires in the process. Smith

and Mucina (2006) further classified the vegetation of the Islands under three main vegetation groups, namely, Marine Macroalgal vegetation, Subantarctic Tundra and Subantarctic Polar Desert and created five mapping units, which include sub-Antarctic cinder cones, coastal vegetation, fellfield, mire-slope vegetation, and polar desert, to compile maps of the vegetation on the Islands. Their work provides a description of the distribution of mires on Marion Island, stating that "mire vegetation is found in most lowland areas, being most extensive below 200 m, but found up to 400 m altitude. On Marion Island approximately 30% of the area below 100 m and approximately 3% of that between 100 and 300 m is occupied by mire vegetation; the largest mires on Marion Island are found on the coastal plain between Repetto's Hill and Long Ridge, inland of East Cape, Macaroni Bay and on the western coastal plain between Kleinkoppie and Kampkoppie" (Smith and Mucina, 2006, p. 716). Mire complexes encompass almost half of the Island below 300 metres above sea level (m a.s.l) (Van Zinderen Bakker Sr, 1970; Smith et al., 2001; Yeloff et al., 2007). These mires have a smooth surface with no pattern of depressions (Gremmen, 1981; Yeloff et al., 2007) and many are covered by a thick layer of gelatinous algae (Van Zinderen Bakker Sr, 1970). Figure 2-1 shows the location of the largest mires on Marion Island as per Smith and Mucina (2006) description.

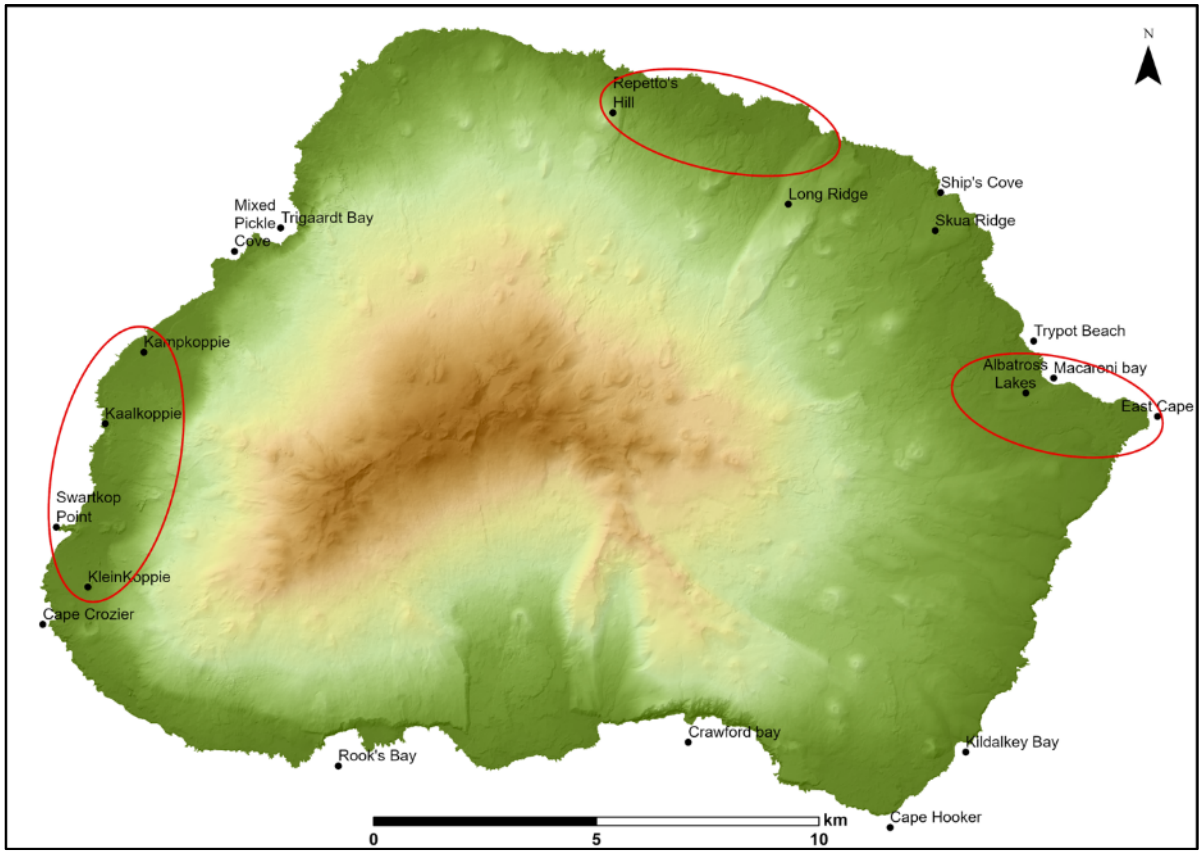


Figure 2-1: Location of the largest wetlands as per the Smith and Mucina (2006) description. The locations of the largest mires are circled by red ovals.

Pekel et al. (2016) derived worldwide surface water locations and seasonality using inventories and national descriptions, statistical extrapolation of regional data, and satellite images; however, there are no data for the Prince Edward Islands in the global datasets. Therefore, the only documentation highlighting the distribution of the mires on the PEIS is provided by Gremmen (1981), Smith et al. (2001) and Smith and Mucina (2006).

2.2. Wetland Mapping

Remote sensing of water has long been a subject of interest. In 1973, McGinnis and Rango (1975) monitored floods using earth resources satellite systems. In the 1970s, open water was mapped as an indicator of waterfowl habitat quality. This study was conducted to assist in devising appropriate management decisions of migratory waterfowl, resulting in maps of locations and frequency of surface water bodies (Work et al., 1976). In most cases, remote sensing has been used to model the spatial and temporal changes of lakes and rivers. One such example, a study carried out by Rokni et al. (2014), evaluated surface area changes of Lake Urmia, Iran, over 13 years (2000-2013) after observing increasing salinity and decreasing surface area in the lake. A similar study was conducted by Sarp and Ozcelik (2018) where they investigated the spatiotemporal changes in Lake Burdur, Turkey, over 24 years (1987-2011) using multiple spectral water indices.

By using remote sensing to observe and monitor spatiotemporal changes in wetlands, these important land features can be preserved and better managed (Ozesmi and Bauer, 2002). Wetlands can be monitored using conventional field mapping; however, remote sensing offers a better alternative, especially over large geographic areas (Torbick et al., 2007). Furthermore, wetlands can be monitored through the interpretation of aerial photographs (Minasny et al., 2019). Due to the high spatial resolution aerial photographs offer, change detection is likely to be more precise than most classification results from satellite imagery as they allow for the identification of small, long or narrow wetlands (Ozesmi and Bauer, 2002). However, it works better once the wetland of interest has been identified; hence, it cannot be used to identify wetlands but rather to determine their extent (Rebelo et al., 2017). Additionally, change detection using aerial photographs may become a time-consuming process, especially when it is being conducted over multiple years (Baker et al., 2007). Aerial photograph interpretation is also subjective and prone to human error, as well as variation in interpretation. As such, results are difficult to replicate and can be inconsistent (Coppin et al., 2004; Baker et al., 2007).

As the spatial, spectral and temporal resolution of satellite data continues to improve, modern day remote sensing relies on satellite remote sensing for land cover and land use change detection; wetland detection and delineation is no exception (Baker et al., 2007; Rebelo et al., 2017). Satellite remote sensing offers advantages over aerial photographs in that it offers repeated coverage allowing for seasonal or yearly change detection, provides the data in a format that is easier to integrate into a GIS, and does so at a lesser cost (Ozesmi and Bauer, 2002). Most of the research regarding the remote sensing of water is aimed at attaining more efficient and effective methods of detecting and monitoring water. Methods of extracting water areas from remotely sensed imagery are developed based on the principle that water has a lower reflectance when compared to other land cover types (Huang et al., 2018a). The boundary between water and non-water features, specifically land, is defined by their respective responses to the near-infrared radiation (NIR) band, with water appearing darker than land as it absorbs the NIR (Klemas and Pieterse, 2015).

Methods of analysing remotely sensed data to assess water resources date as far back as 1976. Work et al. (1976) used satellite data to map open surface water (ponds and lakes) in the glaciated prairie where they focussed on identifying water features using the principle that in a single infrared waveband, water possesses a low radiance. With this principle, they were able to examine each scene pixel in the imagery as either water or non-water using a 'high-speed digital computer', resulting in quick and accurate recognition of the open surface water (ponds and lakes). The study also adopted a multiple waveband approach, which they referred to as proportion estimation, which estimated the fraction of a cell that may be composed of open water, allowing for the identification of more open water features.

The most popular current method of detecting surface water from satellite imagery, an adaptation of the Normalised Difference Vegetation Index (NDVI), is the Normalised Difference Water Index (NDWI), as suggested by McFeeters (1996). The NDVI is an index intended to assess surface biomass and productivity using the near-infrared (NIR) and red bands (refer to equation (*4*), pg. 32), to enhance the presence of vegetation, while it does not affect water features (McFeeters, 1996). The near-infrared band (750–900 nm) is used primarily for imaging vegetation, with the red band (600-690 nm) mainly used to imaging man-made objects, soil, and vegetation. McFeeters (1996) suggested that if the green band (515-600 nm), used for imaging vegetation and deep-water structures, were to be substituted in for the red band, only water features would be enhanced. McFeeters (1996) was able to offer a method for identifying open water areas from satellite data while suppressing soil and vegetation elements at the same time. The NDWI (equation (*5*), pg. 33) is a first-generation water index based on the idea that water has the highest absorption capabilities and vegetation has the highest reflectance in the near infrared (NIR) (Huang et al., 2018a).

Although the NDWI can differentiate between water and vegetation and successfully extract water from remotely sensed images, the index has some shortcomings (Xu, 2006; Gautam et al., 2015; Du et al., 2016; Huang et al., 2018a). Specifically, it is unable to distinguish between water and built-up areas because built-up areas return positive values like those of water due to NIR reflectance being lower than the green reflectance (Xu, 2006; Du et al., 2016). Previous work has demonstrated that the modification to the NDWI by Xu (2006), as illustrated by equation (*1*), makes the index more stable and reliable and thus more suitable to enhance and extract water bodies by replacing the NIR band with the short-wave infrared (SWIR) band, which is used for imaging vegetation and soil moisture content, and is better at distinguishing between water, soil and built-up areas.

$$\text{MNDWI} = \frac{Green - SWIR}{Green + SWIR} \qquad (1)$$

Although the modified Normalised Difference Water Index (MNDWI) also has shortcomings in that it is unable to discriminate between water and snow because of the high reflectivity of snow compared to water, the index is still suitable for surface water extraction from images, if this limitation is acknowledged. In the first ten years of the twenty-first century, the NDWI was the most popular and extensively utilised approach, and the modified NDWI has since become the acknowledged standard for water detection and monitoring (Huang et al., 2018a).

There are other water extraction indices such as the Automated Water Extraction Index (AWEI) proposed by Feyisa et al. (2014), which uses a different strategy of extracting surface water compared to the NDWI and MNDWI. The index was proposed to use band differencing, addition, and the application of different coefficients to maximise the distinguishability of water features from the non-water feature pixels. It was designed with shadows and urban backgrounds in mind as these are major problems that result in low accuracy of surface water feature classification. The authors used Landsat 5 imagery to develop two versions of the AWEI, one for areas that do not have shadows ($AWEI_{nsh}$), and one for areas that have shadows ($AWEI_{sh}$). After comparing these with the performance of the MNDWI on the same data, they concluded that the AWEI was an improved water index, especially when applied to areas that have shadows and are built-up. The improvement is credited to the index's ability to accurately classify edge pixels, using a stable threshold (Feyisa et al., 2014).

The performance of the NDVI, NDWI, MNDWI, and AWEI were compared from their application on Open Land Imager (OLI) imagery (Zhai et al., 2015). The results showed that AWEI and MNDWI are better indices for surface water extraction. A different result is achieved when a Support Vector Machine (SVM) classification method coupled with spectral water indexing (NDWI, MNDWI, AWEI) is used as a method of extracting water bodies and detecting change (Sarp and Ozcelik, 2018). The results show the method produced the best results for all the spectral water indices that were applied after the SVM classification. Thus, the results suggest that three of the indices (NDWI, MNDWI, and AWEI) are capable of extracting water if the conditions are favourable. Therefore, the appropriateness of an index is determined by the need and conditions of the study area.

Water indices (WIs), namely, NDWI, MNDWI, and AWEI, are good detectors of surface water features from remotely sensed imagery. However, Jiang et al. (2014) suggest that there are problems associated with these WIs that are based on the method of thresholding, that is, deciding which pixels are water, non-water, and mixture features. Inefficient detection and identification of surface water features; water features mixed up with non-water features (background noise); and variations in the extraction threshold that are based on the characteristics of each scene are all issues (Jiang et al., 2014). Thus, the Automated Method for Extracting Rivers and Lakes (AMERL) index was created with the idea that mixed water pixels are typically found in narrow rivers or along the edge of lakes or wide rivers. The index simplifies threshold optimisation while also removing noise that may be caused by shadows. Compared to the other WIs over three study areas in China, the AMERL performed the best, thus improving the extraction of narrow rivers (Jiang et al., 2014).

As mentioned previously, the identification and delineation of vegetated wetlands (such as peatlands) from satellite imagery is more complex than that of open water bodies, due to them occurring in open water, vegetation, or a mixture thereof (Kaplan and Avdan, 2017). These can be mapped using vegetation or water spectral signatures (Kameyama et al., 2001; Thomas et al., 2015). This becomes a concern with pixel-based indices such as NDVI and NDWI being used to identify vegetated wetlands from satellite imagery as the threshold is not a fixed number; rather, it must be computed dynamically in order to differentiate between water, non-water, and mixed features (Ji et al., 2009). This is illustrated by Kaplan and Avdan (2017), who investigated Sentinel-2 imagery's ability to map and monitor

wetlands. The authors were successful in mapping and monitoring wetlands by extracting their boundaries using an object-based classification method and using two indices, the Normalised Difference Vegetation Index (NDVI) and the Normalised Difference Water Index (NDWI) to classify the contents (i.e., fully vegetated, mixed, and water areas) within the wetland boundaries. Regardless, other challenges to using remote sensing to map wetlands remain. Although wetlands are defined by the presence of water, this does not indicate that the water is constantly there or that it is on the earth's surface. Water may be present seasonally or at specific times of the year, and it may also occur below the surface, where the plans are rooted (Gallant, 2015). Furthermore, the distribution of wetlands is influenced by a variety of abiotic and biotic elements such as elevation, slope, geology, soils, and climate, all of which interact and interconnect, making it difficult to map the distribution (Hiestermann and Rivers-Moore, 2015; Rebelo et al., 2017). All these factors present difficulties when using remote sensing to map open water features. Huang et al. (2018a) further highlight that one of the biggest limitations is the unavoidable effect of cloud cover on the imagery. This is of concern since the PEIs are often covered by clouds, making cloud-free images exceedingly rare. The authors further suggest that another limitation is that sensors are unable to penetrate vegetation cover, thus preventing the detection of water bodies that may be underneath vegetation. This can be managed by the inclusion of topographic information to supplement remote sensing-derived data. As such, while remote sensing has introduced new methods of detecting, extracting, monitoring, and mapping surface water features, it is important to remember that there are limitations to using remotely sensed data, which may result in under representation of wetlands (Hiestermann and Rivers-Moore, 2015). Therefore, auxiliary data such as topographic information can help to improve the accuracy of wetlands mapping (Hiestermann and Rivers-Moore, 2015).

There is a close relation between topography and surface water because of its fluid characteristics. Thus, surface characteristics can play a crucial role in hydrological modelling. The shape and permeability of a landscape controls the movement of surface water (Wolock et al., 2004; MacMillan and Shary, 2009). Water flows from higher to lower places, thus the flow and accumulation of water are closely related to the terrain (Huang et al., 2018a). For instance, when the profile curvature is concave, surface flows converge to form areas of accumulation or deposition while they diverge where the profile curvature is convex, moving at greater speeds that facilitate surface runoff and erosion thus reducing infiltration and groundwater recharge (Moore et al., 1991; MacMillan and Shary, 2009). A convex plan curvature suggests the presence of ridges, from which water flows outward (diverge) while it converges on concave slopes forming drainage lines (valleys) since water flows inward and accumulates at those points (Milevski, 2007). As such, the classification of the landscape aids in extraction of surface water bodies by providing the determination of areas of possible water accumulation. Considerations are the permeability of the underlying geology and soil, as this affects the flow and accumulation of water. Surface runoff is extensive in areas of low permeability, limiting groundwater recharge while in landscapes with high permeability, groundwater recharge is significant as a result of limited surface runoff (Wolock et al., 2004). The geology on the Prince Edward Islands (PEIs) consists primarily of grey and black lava flows. The predominant younger black lava flows are more porous while grey lava is more impermeable and less susceptible to weathering (Gremmen, 1981). Andosols, which are the predominant soil type on Marion Island, are also highly porous and have good internal drainage

(Driessen et al., 2001; Lubbe, 2010). Therefore, the permeability of the surface also affects the amount of surface water, surface water runoff, surface water infiltration, and groundwater storage and recharge. The analysis of the terrain factors such as slope, aspect, curvature, relief (hills, mountains, and valleys), the drainage network, as well as the underlying soil and geology, thus provides valuable information in the study of spatial and temporal patterns of surface water.

These characteristics of the surface can be extracted from digital elevation models (DEMs) (Giles and Franklin, 1998). Lower areas where water is likely to accumulate can be identified from DEMs (Huang et al., 2017). DEMs have been used in several studies to assist in the detection of surface water from remotely sensed imagery. Uuemaa et al. (2018) effectively identify potential places for wetland construction or restoration in New Zealand using topographic parameters (slope, topographic wetness, and stream network). Birsrat and Berhanu (2018) used the Topographic Wetness Index (TWI) to determine sites that store water in Ethiopia. Of note is that peatland development areas have distinct hydrologic regimes, climates, chemistry, landforms, substrates, and flora (Bourgeau-Chavez et al., 2018; Minasny et al., 2019). Topographic information that describes hydrologic regimes can be derived using methods of GIS. Primary and secondary terrain analysis methods, such as slope, the TWI, or landforms (Topographic Position Index (TPI), based landform classification), can be used as supporting information when identifying peatlands. Therefore, from a DEM, it is possible to create a reliable and usable hydrology model, or a landform model, that can assist in the delineation of surface water from satellite imagery. DEM data can also be used to supplement the detection and extraction of water by providing supplementary information to derive water depth, eliminate terrain shadows, and map water underneath vegetation (Huang et al., 2018a). Wang et al. (2002) presented a method for mapping flood extent in a coastal floodplain using Landsat 7 TM data. Due to Landsat TM's inability to penetrate vegetation, the flooded areas were underestimated. Therefore, DEM data were incorporated to overcome this limitation. Although DEM data do not provide information of the actual surface water, it can accurately represent the landscape, and surface water accumulation areas can be more precisely inferred.

Hydrology is an important part of the formation of both lakes and mires. There are several indices that model surface water content well. These model the influence of the terrain by mathematically categorising the landscape from a DEM (Hjerdt et al., 2004), to identify wet areas. Such indices include the TWI, which was introduced by Beven and Kirkby (1979). The index quantifies the effect of the local topography on the runoff flow direction and accumulation by considering flow direction, flow accumulation, and slope (Beven and Kirkby, 1979; Qin et al., 2011; Ballerine, 2017). High values of topographic wetness indicate areas with high potential of water accumulation while lower (negative) values indicate drier areas (Mattivi et al., 2019). Beven and Kirkby (1979) define TWI using equation (*2*).

$$\mathrm{TWI} = \ln\left(\frac{\boldsymbol{Catchment\ area}}{\tan\beta}\right), \textit{ where } \beta \textit{ = slope in degrees} \qquad (2)$$

The TPI is an algorithm that classifies landscapes into slope positions (Figure 2-2). It compares each cell's elevation value to the mean elevation of neighbouring cells determined by a radius (Weiss, 2001; Jenness and Enterprises, 2006). The resultant TPI values can range from negative to positive values.

Positive TPI values indicate that the cell is higher in elevation than the neighbouring cells (ridgetops or hilltops), in contrast, the negative values indicate the cell is lower in elevation than neighbouring cells (valleys or canyon bottoms). TPI values near zero indicate locations of flat areas or mid-slope, depending on the slope of the cell (Weiss, 2001; Jenness and Enterprises, 2006).

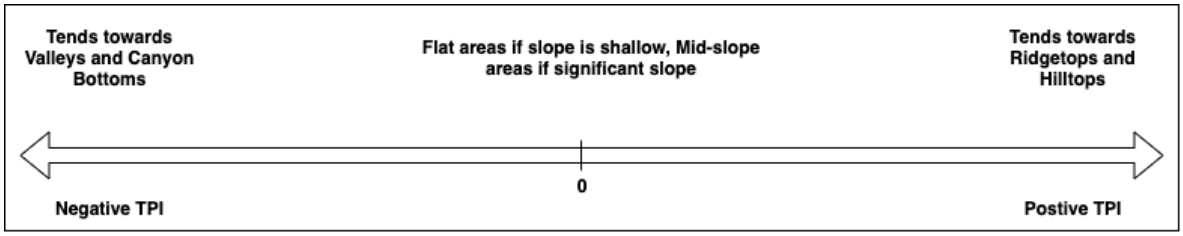


Figure 2-2: Summary of landform classifications according to TPI values (Jenness and Enterprises, 2006).

This algorithm is also scale dependent. The level of detail in the classification depends on the shape and size of the neighbourhood of analysis (Tağıl and Jenness, 2008). Different shapes (circle, rectangle, annulus, wedge or irregular) can be used for the neighbourhood. Weiss (2001) used an annulus (ring-shaped object) neighbourhood created by specifying two different radii. Different sized neighbourhoods reveal different landforms (Skentos and Ourania, 2017). Small neighbourhoods are used to classify small features, such as small streams and lakes (Figure 2-3), while larger neighbourhoods are used to classify larger features, such as canyons or mountains (Figure 2-4).

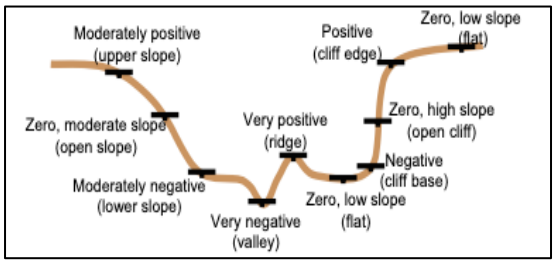


Figure 2-3: Small neighbourhood TPI and slope position classification (Weiss, 2001).

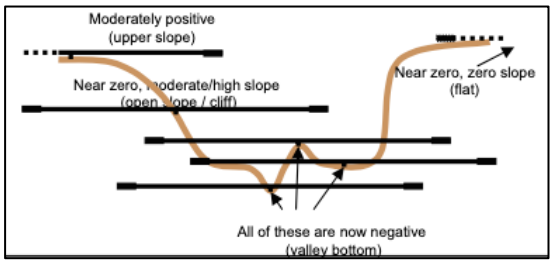


Figure 2-4: Large neighbourhood TPI and slope position classification (Weiss, 2001).

TPI values can be classified into discrete slope position classes based on how extreme they are, as well as the slope for values close to zero. The classes may be defined in terms of the actual TPI values or in terms of standardised TPI values based on the mean and standard deviation of the TPI grid (Jenness and Enterprises, 2006). The latter method was suggested by Weiss (2001), and is the commonly used method used to classify slope positions. By standardising TPI grids to a mean of zero (mean = 0), and standard deviation of one (stdev = 1), any slope combination of TPI grids may be classified using the same fundamental algorithm (Weiss, 2001; Jenness and Enterprises, 2006). Weiss (2001) suggested a 6-category slope position classification (Table 2-1) while Tağıl and Jenness (2008) used a 4-category slope position classification (Table 2-2).

Table 2-1: TPI and slope criteria for 6-slope position classes (Weiss, 2001).

Class	Class description	TPI and slope criteria
1	Ridge	TPI ≥ 1 SD
2	Upper Slope	0.5 SD < TPI ≤ 1SD

3	Middle Slope	-0.5 SD < TPI < 0.5 SD, Slope > 5°
4	Flat Slope	-0.5 SD < TPI < 0.5 SD, Slope ≤ 5°
5	Lower Slope	-1 SD < TPI ≤ -0.5 SD
6	Valley	TPI ≤ -1 SD

Table 2-2: TPI and slope criteria for 4-slope position classes (Tağıl and Jenness, 2008).

Class	Class description	TPI and slope criteria
1	Ridge	TPI ≥ 1 SD
2	Mid Slope	-1 SD < TPI < 1 SD, Slope ≥ 6°
3	Flat Surface	-1 SD < TPI < 1 SD, Slope < 6°
4	Valley	TPI ≤ -1 SD

A single neighbourhood can classify the landscape into slope position classes, however, TPI values that are calculated from two neighbourhood sizes are better as they provide more information about the general shape of the landscape, thus making it possible to identify more complex features (Weiss, 2001; Tağıl and Jenness, 2008; Skentos, 2017). Combining TPI at two scales classifies the landscape into 10 landform classes (Table 2-3). As elevation is spatially autocorrelated, with the ranges of the TPI values increase with scale, standardising TPI grids to a mean of zero (mean = 0), and standard deviation of one (stdev = 1) makes it possible to classify the landscape into landform classes (Weiss, 2001; Jenness and Enterprises, 2006).

Table 2-3: TPI and slope criteria for 10 landform classification (Weiss, 2001).

Class	Class description	TPI and slope criteria	
		Small neighbourhood	**Large neighbourhood**
1	Canyons, deeply incised streams	TPI ≤ -1 SD	TPI ≤ -1 SD
2	Midslope drainages, shallow valleys	TPI ≤ -1 SD	-1 SD < TPI < 1 SD
3	Upland drainages, headwaters	TPI ≤ -1 SD	TPI ≥ 1 SD
4	U-shaped valleys	-1 SD < TPI < 1 SD	TPI ≤ -1 SD
5	Plains	-1 SD < TPI < 1 SD	-1 SD < TPI < 1 SD
		Slope ≤ 5°	
6	Open slopes	-1 SD < TPI < 1 SD	-1 SD < TPI < 1 SD
		Slope ≥ 6	
7	Upper slopes, mesas	-1 SD < TPI < 1 SD	TPI ≥ 1 SD
8	Local Ridges, hills in valleys	TPI ≥ 1 SD	TPI ≤ -1 SD
9	Midslope ridges, small hills in plains	TPI ≥ 1 SD	-1 SD < TPI < 1 SD
10	Mountain tops, high ridges	TPI ≥ 1 SD	TPI ≥ 1 SD

2.3. Species Distribution Modelling

A strong link exists between a species and its surroundings; therefore, supplementary environmental and climatic data can be used to explain the distribution of a species across the landscape. Species distribution models (SDMs) employ algorithms to estimate a species' geographic range by evaluating the link between known species occurrences and environmental variables presumed to influence its occurrence (Guisan and Zimmermann, 2000; Elith and Leathwick, 2009; Franklin, 2009). Correlations between the distributions of species and the physical environment have long been a subject of interest

(Guisan and Zimmermann, 2000; Elith and Leathwick, 2009; Hallstan, 2011). In the late 1970s, when computational capacity was limited, species distribution modelling was first developed (Zimmermann et al., 2010), and as a result of technology improvements, such as improved statistical approaches, GIS, more digital data, and a greater range of tools available to analyse them, new and improved modelling methods have been developed (Elith and Leathwick, 2009; Miller, 2010). Species distribution modelling typically involves four steps, namely 1) collecting the location of species occurrence; 2) amassing environmental predictor variables; 3) fitting occurrence and environmental data into a model or algorithm to identify the conditions associated with the species' occurrence; and 4) predicting the species' distribution to a landscape (Guisan and Zimmermann, 2000; Pearson, 2007; Elith and Leathwick, 2009). Wetland occurrence is governed by complicated interactions between geographic characteristics such as altitude, gradient, and geology, all of which impact groundwater, soils, and climatic variables, and SDMs have proved to be effective in mapping and identifying wetlands of varied sizes (Hiestermann and Rivers-Moore, 2015; Rebelo et al., 2017; Zhong et al., 2021). Essl et al. (2012) also looked at the impact of a changing climate on the distribution of Austria's nine main types of peatlands. As such, SDMs, supported by pre-analyses using GIS and remote sensing, are a useful tool for mapping and predicting the location of wetlands.

The initial step in developing SDMs is the collection of sufficient and accurate locations of the occurrence of the species of interest (Guisan and Zimmermann, 2000; Pearson, 2007; Elith and Leathwick, 2009). This information can either be presence-only data, which provide locations where the species were observed, or both presence and absence data providing locations where species were observed and were not observed (Pearson, 2007). In most cases, presence data are plentiful, however, absence data are often unavailable or insufficient due to a lack of surveying effort (Gomes et al., 2018). However, some SDMs allow for absence data to be randomly drawn from the area of interest (Hijmans and Elith, 2013; Gomes et al., 2018). Field surveys conducted by an individual or a small group of individuals (personal collection), surveys conducted by a large number of people (large surveys), museum collections, and Internet resources can all be used to collect occurrence data (Pearson, 2007). Thereafter, the most relevant environmental data (predictors variables) (Jiménez-Valverde et al., 2008; Hijmans and Elith, 2013) are collected. In most cases, SDMs are developed using existing predictor data in the hopes that the model is able to discern which of those factors are critical to the species' survival; however, others have argued that only ecologically relevant predictor data, which can be continuous or categorical, should be used instead (Elith and Leathwick, 2009; Naimi and Araújo, 2016). SDM data may also be prone to multicollinearity among predictors, spatial autocorrelation in both response and predictor variables, and positional uncertainty, which must be addressed before fitting a model with the data (Naimi and Araújo, 2016).

In the modelling of species distribution, a variety of algorithms are available that are classified into three classes, namely, profile, regression, and machine learning methods (Hijmans and Elith, 2013). Profile methods solely analyse presence data, whereas regression and machine learning methods employ both presence and absence or background data (Hijmans and Elith, 2013). In the early phases of SDM use and development, regression models were often utilised; however, more advanced statistical techniques for SDMs are currently used, based on improved methodology and ecological

understanding, enhancing the accuracy of model predictions (Guisan and Zimmermann, 2000; Yu et al., 2020). Machine learning approaches have become more popular in SDMs during the last two decades (Früh et al., 2018; Yu et al., 2020). Machine learning model methods include Artificial Neural Networks (ANN), Random Forests (RF), Boosted Regression Trees (BRT), and Support Vector Machines (SVMs) (Hijmans and Elith, 2013). Ensemble models, which combine individual models to create one predictive output, are often used when it is difficult to discern which model would be best to use in a new situation and should produce more accurate predictions than any single model (Araujo and New, 2007; Marmion et al., 2009; Kaky et al., 2020). The accuracy of the prediction of the species distribution depends on the model or algorithm used, so the choice of model must be made carefully (Elith and Leathwick, 2009). Therefore, some have resorted to comparing multiple models' ability to predict to select the best one for application in a particular study (Essl et al., 2012; Hiestermann and Rivers-Moore, 2015; Früh et al., 2018). In terms of predicting the occurrence of wetlands, the machine learning MaxEnt model is the most popular SDM and has been successfully used to predict the occurrence of wetlands (Hunter et al., 2012; Rebelo et al., 2017). This model is based on the principle maximum entropy and is suitable for prediction of species occurrence from presence-only data (Phillips et al., 2006; Kaky et al., 2020).

Following the fitting of the models, further processes such as model evaluation, prediction, and variable importance assessment can be used (Naimi and Araújo, 2016). By evaluating the models, their ability to predict the occurrence of a species can be assessed and this can assist in choosing the optimal model for the situation. For model evaluation, testing data is required to test the models, and in rare cases, there is enough data to have a completely independent test dataset, that is species occurrence data that were not used in the training of the model (Guisan and Zimmermann, 2000; Pearson, 2007; Naimi and Araújo, 2016). However, in most cases, a data resampling method is required to create a test data set. These methods include, 1) random subsampling, 2) k-fold cross-validation, and 3) bootstrapping. Such methods are ideal when the species occurrence dataset is too small to be split into two independent training and testing datasets (Guisan and Zimmermann, 2000). The first method, subsampling, is a hold-out strategy that relies on data splitting into training and testing data (Kohavi, 1995). The method is repeated k times, and the estimated accuracy is the mean accuracy of each run (Kohavi, 1995). The second method, k-fold cross-validation, is used by the majority of modelers as it reduces variability and is less sensitive to erroneous findings from a single random pick because it cycles through all the data (Hijmans and Elith, 2013). Finally, bootstrapping is a sampling approach that repeats a sampling with a replacement method, drawing a sample with the same size as the original data and using it for training data each time (Naimi and Araújo, 2016). Guisan and Zimmermann (2000) suggest that a resampling method is used in conjunction with an independent test dataset, when possible, to assess the stability of the model and the quality of the model predictions.

SDM accuracy may be determined by their ability to correctly categorize presence and absence occurrences (discrimination metrics) or prediction accuracy (reliability metrics) (Leroy et al., 2018). The area under the receiver operating characteristic (ROC), Cohen's Kappa statistic, and true skill statistic (TSS) are the most popular classification metrics that are calculated on presence-absence data (Leroy et al., 2018). The ROC curve is the most popular metric to compare SDM performance (Hijmans and

Elith, 2013; Yu et al., 2020). When continuous probability produced scores are translated to a binary presence-absence variable, the ROC reduces subjectivity in the threshold selection process by summing entire model performance across all feasible thresholds (Lobo et al., 2008; Peterson et al., 2008). Area under the curve (AUC) values vary from 0 to 1, with an AUC score between 0.9 and 1 indicating an excellent model, 0.8 to 0.9 indicating a good model, 0.7 to 0.8 indicating a fair model, 0.6 to 0.7 indicating a poor model, and 0.5 to 0.6 indicating a failed model (Swets, 1988; González-Ferreras et al., 2016). As such, an AUC of 0.5 indicates that the model is as good as a random guess (i.e., imply that the model is no better than randomly predicting presences against absences), and values larger than 0.7 indicate that the model is sufficient for modelling species distributions (Swets, 1988). Although it is widely accepted as the standard technique for assessing SDM correctness, others do not advocate using this metric as a comparison measure of model accuracy. They argue that models with similar AUC values can predict vastly diverse distribution patterns, and that AUC should be used with caution when comparing model performance for different species and for models using different data sets (Termansen et al., 2006; Austin, 2007; Lobo et al., 2008). They also caution against comparing the accuracy of multiple models for the same species if the total extent investigated differs, because the extent to which models are carried out has a significant impact on the AUC values (Lobo et al., 2008). Other studies have also criticized the use of AUC as a comparative measure of SDM performance (Peterson et al., 2008; Jiménez-Valverde, 2012). As a result, the AUC is frequently employed in combination with another metric when utilised as a measure of accuracy (Mainali et al., 2015; Leroy et al., 2018).

Another metric that is frequently used to evaluate model performance is Cohen's Kappa statistics. This metric adjusts the total accuracy of model predictions to account for the accuracy predicted to occur through chance (Allouche et al., 2006; Raes and ter Steege, 2007). The Kappa statistic ranges from -1 to +1, with a value of +1 showing perfect agreement between observation and the prediction of the testing dataset and values of zero and below suggesting no better than random performance (Cohen, 1960; Landis and Koch, 1977). As with the AUC, the Kappa statistic has also been criticised. This metric is affected by the prevalence of the modelled species, which is the proportion of the study area occupied by the species (McPherson et al., 2004; Allouche et al., 2006; Pontius Jr and Millones, 2011). As an alternative, Allouche et al. (2006) suggest using the more straightforward and intuitive TSS as a measure of SDM success. The metric compares the proportion of correct predictions to the proportion of hypothetical predictions, disregarding any predictions that may be due to random guess (i.e., when the model is no better than predicting presences versus absences at random) (Allouche et al., 2006). TSS has the advantages offered by the Kappa statistic, while not affected by species prevalence (Allouche et al., 2006). Furthermore, the size of the validation set has no effect on TSS (Allouche et al., 2006). TSS values range between -1 and +1, with larger values indicating a better model. TSS values less than 0.2 are considered fail or null models, between 0.2 and 0.4 are poor, between 0.4 and 0.6 are fair and those greater than 0.6 are good to excellent (González-Ferreras et al., 2016).

Unlike the AUC, Kappa and TSS require that the continuous model predictions be transformed into binary predictions based on a threshold (Fielding and Bell, 1997). To account for species prevalence, sensitivity and specificity can be used. Sensitivity (true positive proportion) and specificity (false positive

proportion) are two reliability metrics, which are independent of prevalence. Sensitivity is the proportion of correctly predicted species presences, whereas specificity measures the proportion of correctly predicted absences (Swets, 1988; Allouche et al., 2006). The sensitivity and specificity values range from -1 to +1, with values closer to +1 indicating better model performance (Swets, 1988). Swets (1988) classified sensitivity and specificity values between 0.5 and 0.7 as poor accuracy, those between 0.7 and 0.9 as appropriate for some uses, and those greater than 0.9 as high accuracy. By applying a statistically determined threshold to the continuous surface and selecting the value that maximises the sum of sensitivity and specificity (i.e., sensitivity + specificity/ 2), binary presence or absence maps can be constructed (Liu et al., 2005).

SDMs make the most of ancillary information such as topographic attributes and climatic data to explain the relationship that exists between a species and its surrounding environment, and offer a practical, reliable, and cost-effective method of estimating species distribution. However, there is no single best modelling algorithm, therefore, an appropriate modelling technique for the species under investigation must be selected in order achieve high accuracy predictions of species occurrence (Li and Wang, 2013).

In this study, the ability of multiple regression-based and machine learning SDM algorithms to predict the distribution of mires on the PEIs were tested. The accuracy of the models was assessed using three metrics, the AUC, TSS and the correct classification rate (CCR), which crosses all the model presence-absences with the model's binary predictions to measure overall model accuracy (Fielding and Bell, 1997; Martínez-Freiría et al., 2016).

Chapter 3: Study Area

3.1. Geographical and Oceanographic Position

The Prince Edward Islands (PEIs), which comprise Marion Island (46° 54´ S, 37° 45´ E) and Prince Edward Island (46° 38´ S, 37° 57´ E) (Figure 3-1), are South African-owned, remote islands in the sub-Antarctic Ocean. The PEIs have fairly stable annual climates, characterised by low temperature, regular rainfall and strong winds (le Roux, 2008). They are mostly unoccupied except for a few research scientists visiting some of these islands for short periods, or a small handful of personnel residents on these islands to ensure research bases remain operational.

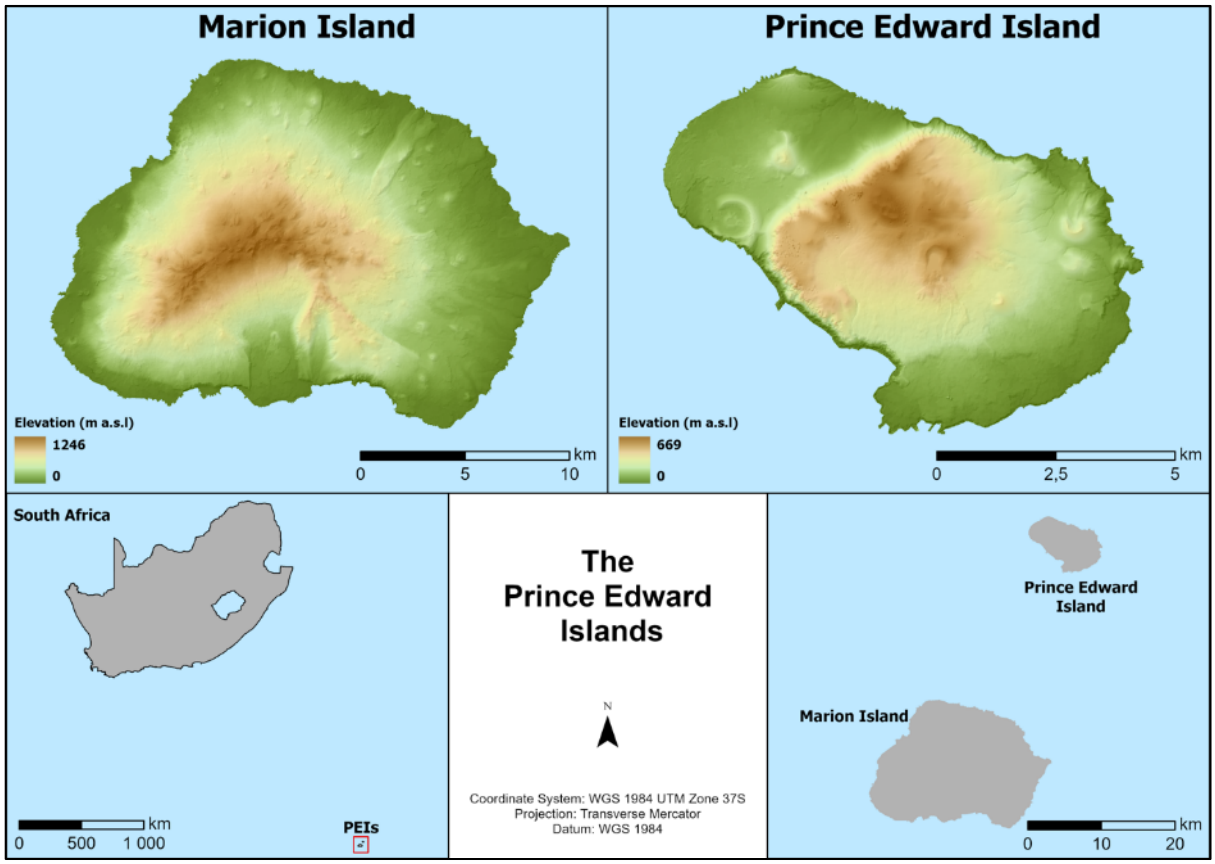


Figure 3-1: The location of the study area, the Prince Edward Islands (PEIs), Marion Island, and Prince Edward Island, in relation to South Africa.

The Islands are the summits of geologically young (Quaternary Period) coalescing shield volcanoes, rising from the West Indian Ocean Ridge, with the closest land being the sub-Antarctic Crozet Islands, situated approximately 900 km to the east (Gremmen, 1981; Smith and Mucina, 2006; Chown and Froneman, 2008; Chown et al., 2008). Marion Island has a low roughly oval (dome-like) shape, covering an area of 290 km², and rising to approximately 1230 metres above sea level (m a.s.l.) (Mascarin peak) (Smith and Mucina, 2006), and is the larger of the two Islands. Prince Edward Island lies 19 km northeast of Marion Island, covering 46 km², with van Zinderen Bakker Peak rising to approximately 672 m a.s.l (Smith and Mucina, 2006; Chown and Froneman, 2008). Prince Edward Island has a distinctive asymmetric form and extensive vertical relief, with cliffs up to 400 m high on the western side and up to 500 m to the north and south of the central block (Gremmen, 1981; Rudolph et al., 2020).

3.2. Climate

The Islands are located in the Roaring Forties, and thus have an oceanic climate that is characterised by low temperatures with small seasonal variations, heavy rain, snow, strong prevailing westerly winds (50 km per hour or greater), humidity and frequent cloud cover (Smith, 2002; Pakhomov and Chown, 2003; Smith and Mucina, 2006; le Roux, 2008; le Roux and McGeoch, 2008). High precipitation, humidity, strong winds, and cloud cover are similar characteristics of other sub-Antarctic islands (Bergstrom and Chown, 1999). The climate on Marion Island varies across the landscape due to variations in aspect, altitude and recording height (le Roux, 2008). It is speculated that the prevailing westerly winds travelling up the windward side (western) of the mountain, cause the warm, moist air to lose moisture and result in warmer, drier air on the leeward side (eastern) of the mountain, resulting in the windward side experiencing more humidity, cloudiness and precipitation than the leeward side of the island (Rouault, 2005; le Roux, 2008).

The permanent meteorological station on Marion Island has recorded mostly uninterrupted weather observations since 1948 (Chown and Smith, 1993; le Roux, 2008). Due to the surrounding ocean, the Islands experience small temperature and precipitation variations (Smith and Mucina, 2006; le Roux, 2008). Although Prince Edward Island has no meteorological records, it lies 19 km from Marion Island and is assumed to have a similar climate to Marion Island, with slightly lower diurnal temperature variation (le Roux, 2008). They are considered 'thermally stable', with Marion Island experiencing a difference of 4.1 °C between the coldest and warmest month, while the diurnal temperature varies by only 1.9 °C (Smith, 2002; Smith and Mucina, 2006). There is a high prevalence of cloudiness and sporadic direct sunshine in the Southern Ocean, which is associated with the high precipitation on the Islands (Smith and Mucina, 2006). Precipitation falls almost daily , mostly in the form of rainfall rather than snow or hail, with the rainfall distributed almost equally throughout the year (Smith and Mucina, 2006).

However, since the 1960s a steady increase in the mean diurnal and annual temperatures and a decrease in precipitation has been observed, resulting in a drier and warmer climate (le Roux, 2008; le Roux and McGeoch, 2008). Mean annual temperature increased from 5.4 °C in the 1950s to 6.4 °C in the 1990s, with an average increase of 0.28 °C and 0.24 °C to daily maximum and minimum daily temperature per decade respectively, resulting in an increase from a maximum daily temperature of 7.6 °C in the 1950s to 8.6 °C in the 1990s and an increase from a minimum daily temperature of 2.8 °C (1950s) to 3.7 °C (1990s) (le Roux and McGeoch, 2008). Over the same period (1950s-1990s), Marion Island saw a decline in annual rainfall, from approximately 3000 mm per annum to 2000 mm per annum (le Roux and McGeoch, 2008). Furthermore, there has been a decrease in the annual rainfall days on the Island, with the number of days without rainfall increasing from an average of 49.1 days to 89.1 days (le Roux and McGeoch, 2008). The increasing temperatures and decreasing precipitation is linked to the warming of the surrounding ocean and changing atmospheric circulation patterns (Smith and Mucina, 2006; le Roux, 2008). Wind speed and the number of days with potential evapotranspiration also increased, while cloud cover varied with a peak in the 1970s (le Roux and McGeoch, 2008). The changes to climate on the sub-Antarctic islands have been heterogenous across the region. While some of the islands, such as Kerguelen Island, Macquarie Island and Heard Island, have seen a decrease in

temperature, the rate of change differs to that on Marion Island (Frenot et al., 1997; Budd, 2000; Kirkpatrick and Scott, 2002; Pendlebury and Barnes-Keoghan, 2007). Furthermore, unlike on Marion Island, precipitation on Macquarie Island has increased (Pendlebury and Barnes-Keoghan, 2007).

3.3. Topography, Geology and Soils

Marion Island's central highland area slopes gradually on the eastern and northern side, forming coastal plain areas approximately 4 to 5 km towards the sea, while it is much more abrupt on the western and southern side of the Island, dropping to a low plain (Figure 3-2) (Gremmen, 1981; Smith and Mucina, 2006). The plains and valleys on these slopes are separated by grey lava plateaus and ridges, while the lower-lying areas are covered in black lava (Gremmen, 1981). Some of the scoria cones on Marion Island rise to 200 m above the surroundings, forming major landscape features (Hedding, 2006).

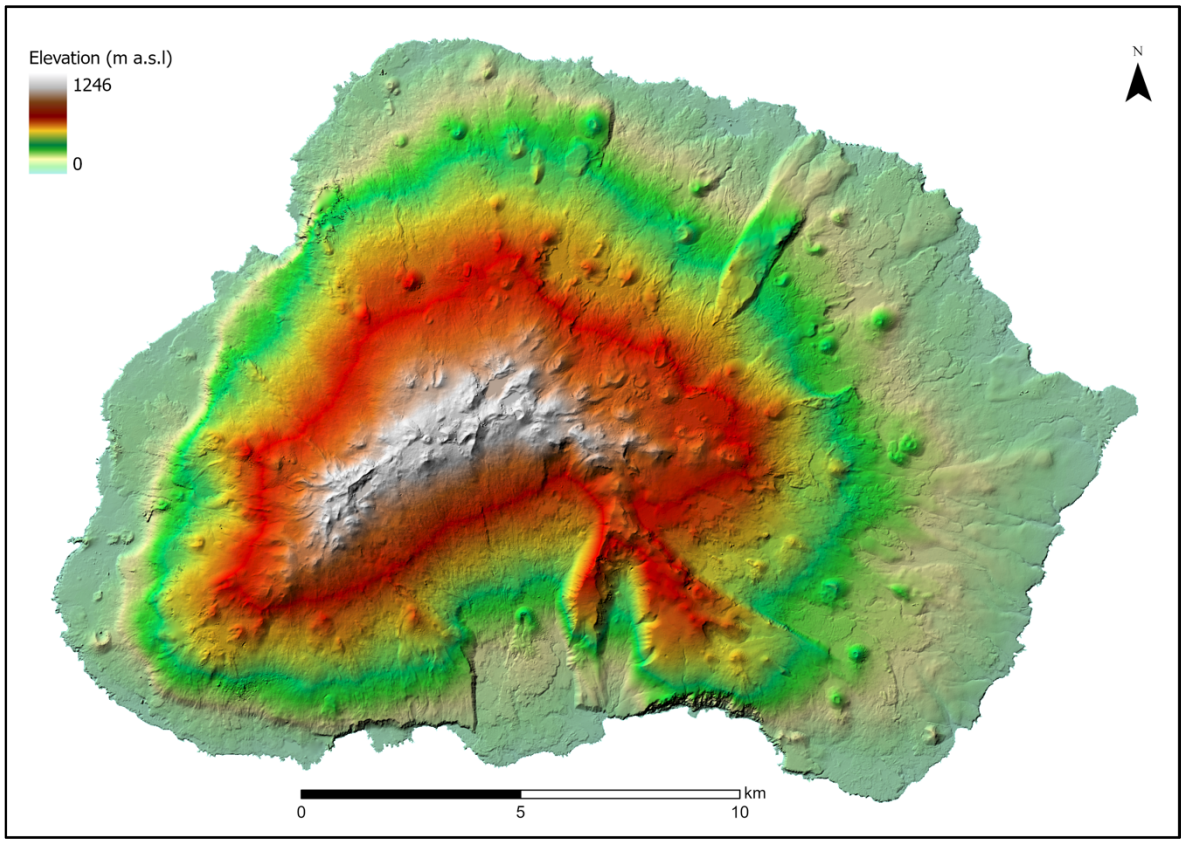


Figure 3-2: Shaded relief depicting Marion Island's topography.

Much of Prince Edward Island is a central block that gently slopes towards the east coast, covering almost 7 km (Figure 3-3). However, the central block creates a steep escarpment to the west of the Island, while the Island's central plateau descends into the sea as cliffs to the north and the south (Gremmen, 1981; Smith and Mucina, 2006).

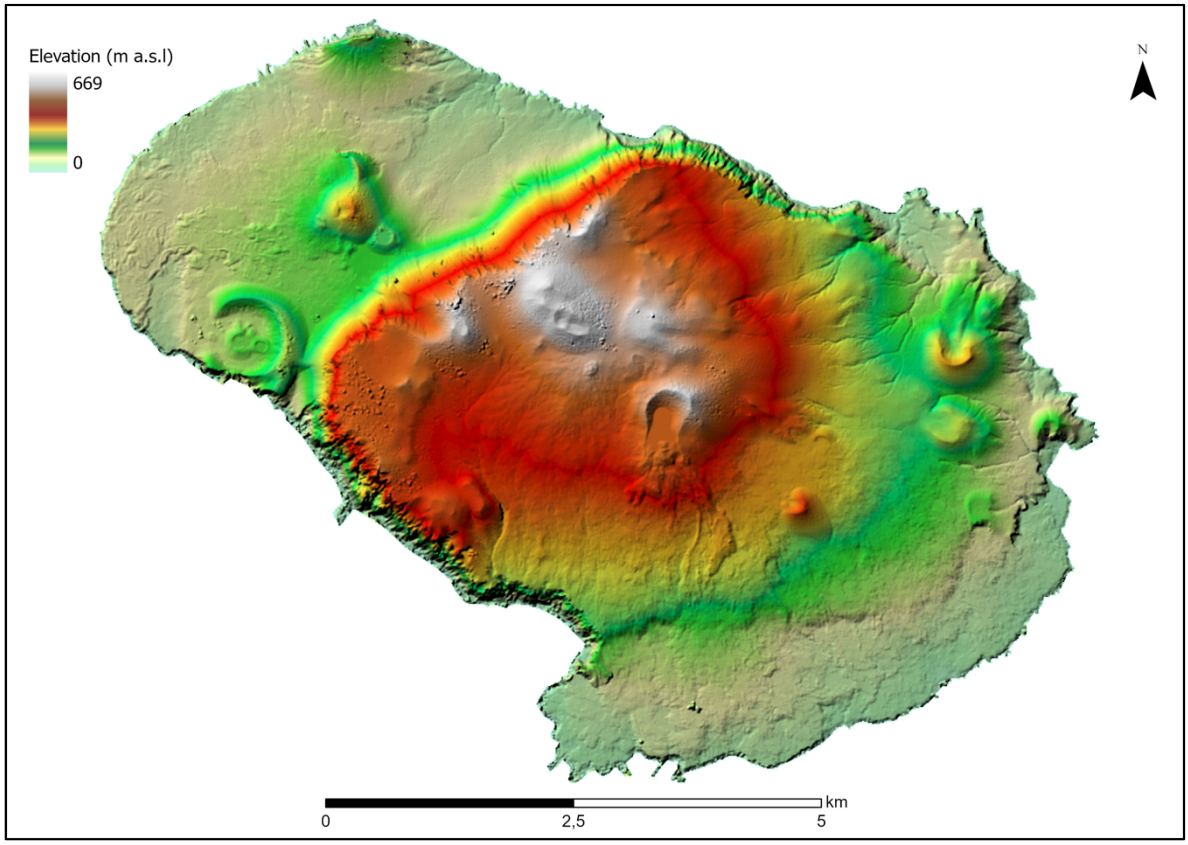


Figure 3-3: Shaded relief depicting Prince Edward Island's topography.

The Islands were formed as a result of different periods of volcanic activity, resulting in two main types of basaltic lava, namely black and grey lava (Øvstedal and Gremmen, 2001). The original volcanology and geology maps, with low spatial precision and detail, for the Prince Edward Islands (PEIs) were developed by Verwoerd and Langenegger (1968), and were recently modified by Rudolph et al. (2020). As seen in Figure 3-4 and Figure 3-5, the bedrock of the Islands is mainly grey lava, which is covered by the younger black lava (Rudolph et al., 2020). It can also be seen that the younger black lava is the dominant geology on the Islands (Table 3-1, pg. 23) (Gremmen, 1981; Øvstedal and Gremmen, 2001; Rudolph et al., 2020). There are 130 red and orange conical hills of volcanic cinder, known as scoria cones, which are the product of several lava flows from explosive volcanic activity on Marion Island, whereas on Prince Edward Island, there are approximately 15 scoria cones scattered over the landscape (Gremmen, 1981; Heymann et al., 1987; Boelhouwers et al., 2008; Hedding, 2020). Black lava flows and scoria cones are porous while grey lava is more impermeable, making them less susceptible to the processes of freezing and thawing and erosion (Gremmen, 1981).

Table 3-1: The total surface area cover of the different geological units on Marion and Prince Edward Island (Rudolph et al., 2020).

Geological unit	Marion Island	Prince Edward Island
	Surface area (km²)	Surface area (km²)
Black lava	183.29	23.16
Grey lava	58.24	11. 46
Scoria	37.15	4.21
Wind-blown ash	3.41	4.24
Grey-bedded ash	0.35	2.02
Black lava: Post 1980	11.58	0

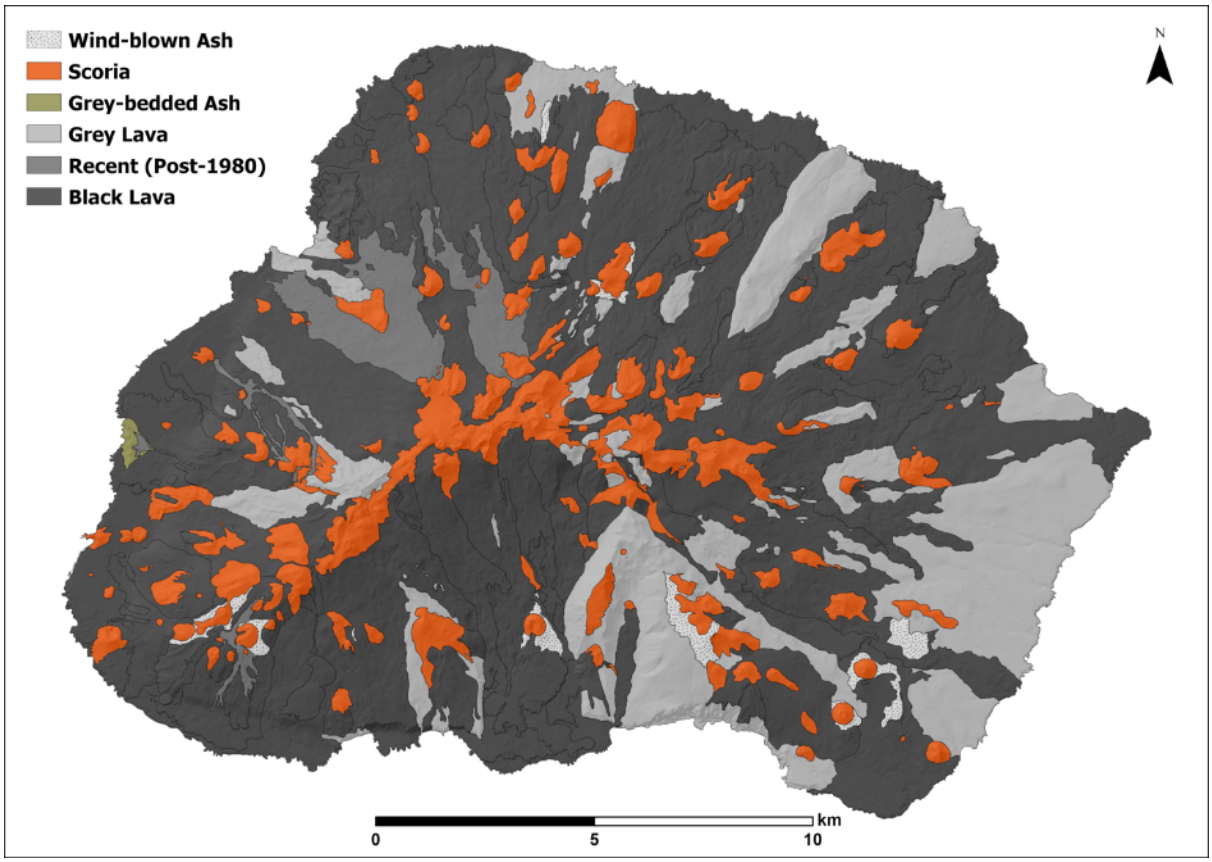


Figure 3-4: A simplified depiction of the geology of Marion Island using a simplified categorization scheme (Rudolph et al., 2020), adapted here to improve discernibility of features.

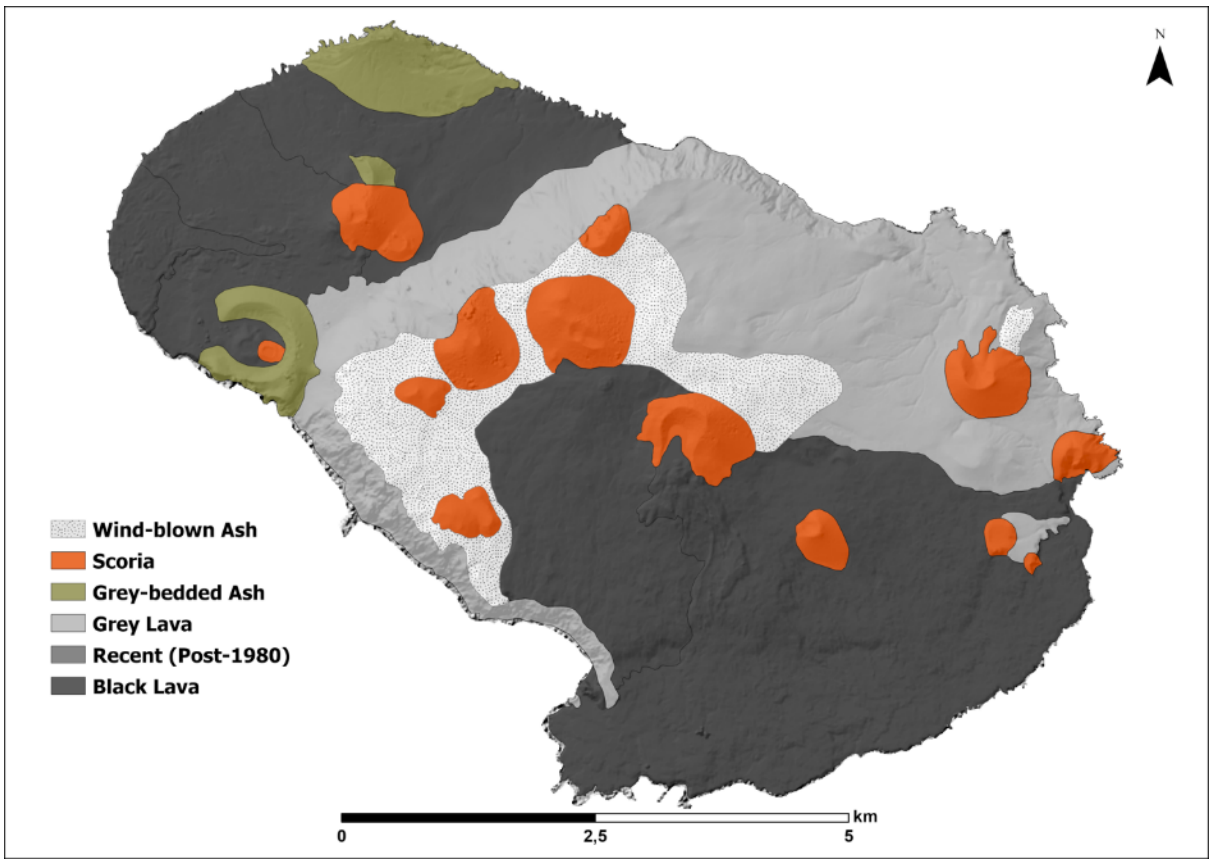


Figure 3-5: A simplified depiction of the geology of Prince Edward Island using a simplified categorization scheme (Rudolph et al., 2020), adapted here to improve discernibility of features.

Organic matter is likely to accumulate where decomposition is slow as a result of a thermally stable climate and moist substrate, therefore, the oceanic climate of the Prince Edward Islands is conducive to the accumulation of organic matter (Gremmen, 1981; Smith and Mucina, 2006). Therefore, peats are the dominant soil formations on these Islands. On Marion Island, the peat deposits range between a thickness of a few centimetres and 4 m in the poorly drained lowland areas, and thicknesses between 5 and 30 cm in well-drained areas (Gremmen, 1981). Although peats can lie directly on rock, there is usually loamy clay between the peat and the underlying rock (Gremmen, 1981; Smith and Mucina, 2006). Peats are much thicker on older, impermeable grey lava flows (up to 3 m), while they are shallower on younger, more porous black lava flows (generally less than 1 m) (Smith and Mucina, 2006). No substantial soil formation has occurred in some areas of young black lava flows and scoria deposits (Gremmen, 1981).

Lubbe (2010) used the World Reference Soil Classification System to classify the soils found on Marion Island, as the soils are of organic and volcanic nature, thus different from those found in South Africa, and the soil groups in the South African Soil Classification System. Lubbe (2012) classified soils into three reference soil groups, namely, Andosols, Histosols, and Regosols (Lubbe, 2010). Histosols, also known as ‘peat soils’, ‘mulch soils’, ‘bog soils’, and ‘organic soils’, are soils characterised by high organic matter (20% and 30%), and formed under waterlogged conditions from decomposed plant remains (Driessen et al., 2001; The Editors of Encyclopaedia Britannica, 2016b). Andosols are typically dark coloured soils that develop on volcanic landscapes from mainly volcanic ash (Driessen et al., 2001; The Editors of Encyclopaedia Britannica, 2016a). These soils are highly porous and typically occur in high

terrains, and as a result, have good internal drainage (Driessen et al., 2001). On average, the organic matter in Andosols is approximately 8%, however, some of the Andosols on Marion Island have higher organic matter (16% or greater). Although the soils are characteristics of Andosols, they have organic matter similar to Histosols. As such they are classified into an intermediary category, Histic Andosols (Driessen et al., 2001; Lubbe, 2010). Soils that cannot be classified into any soil groups in the World Reference Soil Classification System are classified as Regosols (Driessen et al., 2001; The Editors of Encyclopaedia Britannica, 2016c). A non-soil solid rock group was further added to describe recent lava flows, resulting in 5 final soil categories, namely, Andosols, Histosols, Histic Andosols, Regosols and Solid rock, depicted in Figure 3-6 (Lubbe, 2010).

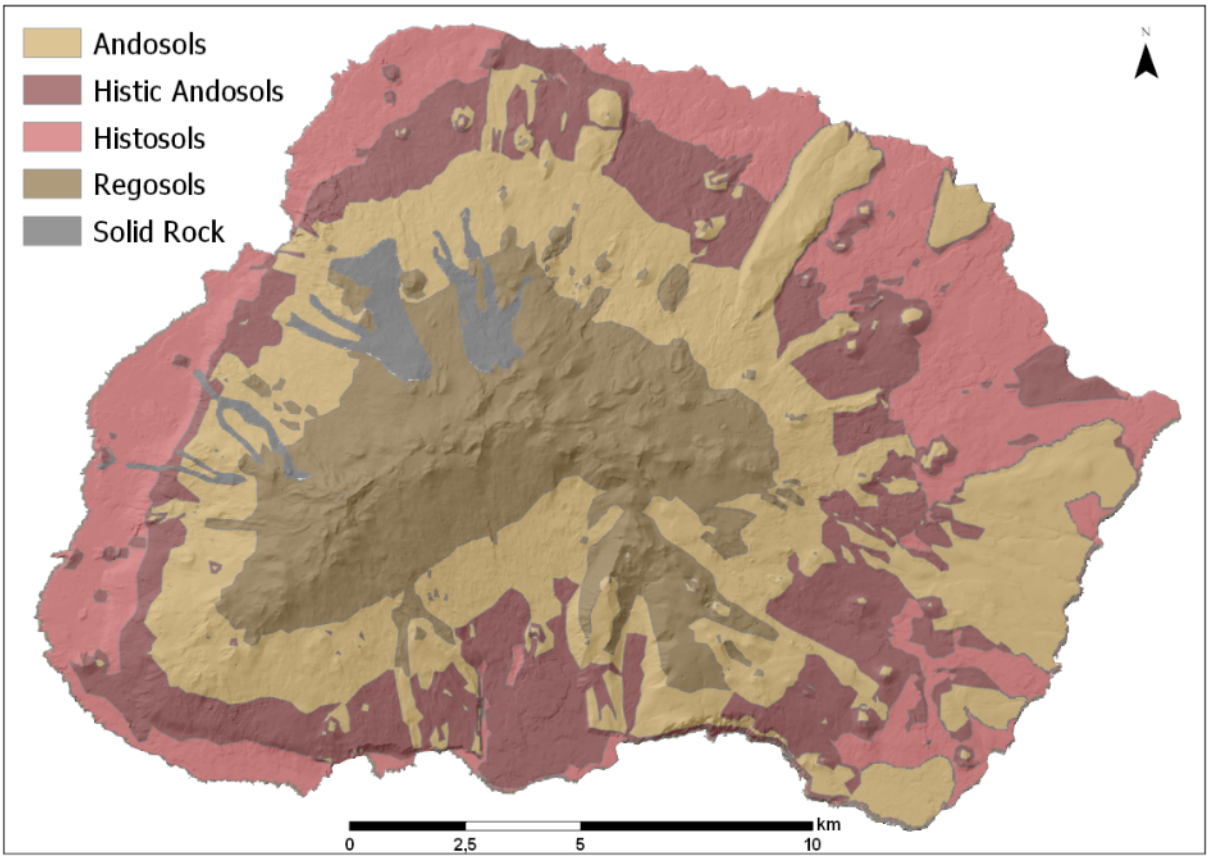


Figure 3-6: Soils of Marion Island (Lubbe, 2010).

3.2. Hydrology

On the PEIs, the predominant, porous black lava facilitates substantial underground drainage, while the impermeable grey lava flows allow for a more developed above-ground drainage system (Gremmen, 1981). Therefore, the Islands experience high rates of sub-surface drainage. Tarns, lakes, and ponds can be found on both Islands (Smith and Mucina, 2006). Marion Island has 35 streams, a few of which are perennial due to the porosity of the dominant geology, while none of the few small streams on Prince Edward Island are perennial (Gremmen, 1981). Lakes between 1000 and 30 000 m² in area, with depths ranging from 0.5 to 2.5 m, are considered large lakes and are usually restricted to the impermeable grey lava flows (Gremmen, 1981; Smith, 2008). Water also moves through the soil; however, it is concentrated in drainage lines. Small water springs occur at the bottom of some grey lava slopes, with water tracts providing water to the springs also feeding into streams (Gremmen, 1981). Lakes do occur on the more blocky and porous black lava flows, albeit they are less common(Smith, 2008). Bog ponds,

which are generally small in size, are frequent in mire areas (areas of peat deposit accumulation) (Gremmen, 1981).

3.3. Vegetation

Due to the low temperatures, low incident solar radiation, and strong winds, the Islands are barren of trees (Van Zinderen Bakker Sr, 1970; Smith and Lewis Smith, 1987; Smith and Mucina, 2006). The Island has a tundra-type biome (Crafford et al., 1986), characterised by tundra-like (low-growing) vegetation consisting mainly of mosses, ferns, and some lichens in the central areas, as well as vascular plants dominating in some areas (Heymann et al., 1987; Smith and Mucina, 2006). While not all parts of the Island are vegetated, in particular the higher elevations, certain areas are characterised by dense vegetation cover, with 100% of the ground covered by plants.

Due to the cool and oceanic climate on the PEIs, peat formation is common (Gremmen, 1981). Peat formation is possible where the water table is close to the surface for most of the time (Rydin et al., 1999; Raeymaekers et al., 2000). This results in the formation of waterlogged mires, mostly in lowland areas, which can range from a depth of a few centimetres to more than 4 m where drainage is impeded (Gremmen, 1981). The *Juncus scheuchzerioides–Blepharidophyllum densifolium* complex is made up of mire plant communities and graminoid plants, particularly *Agrostis magellanica, Uncinia compacta*, and *Juncus scheuchzerioide*s, are significant peat-forming plants (Gremmen, 1981; Yeloff et al., 2007).

Smith and Mucina (2006) used field research to group vegetation on the PEIs and create five mapping units, which include sub-Antarctic cinder cones, coastal vegetation, fellfield, mire-slope vegetation, and polar desert, to compile maps of the vegetation on the Islands. The five units can be seen in the simplified representation of Marion Island's vegetation presented in Figure 3-7.

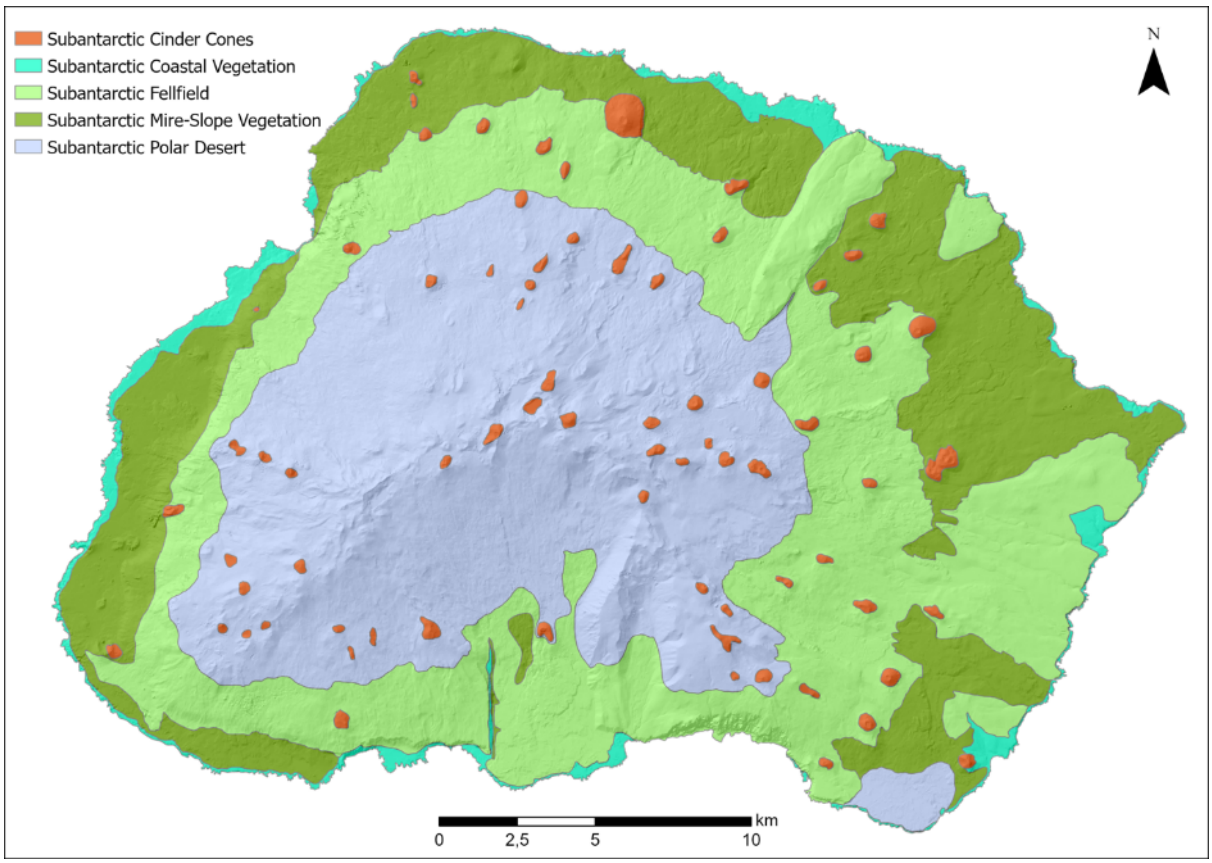


Figure 3-7: A simplified representation of Marion Island's vegetation (after Smith and Mucina (2006)).

3.4. Human Influence on the Island's Ecosystems

The Islands are important conservation areas and were declared a Special Nature Reserve in 1995, the ultimate protection under South African legislation (Smith and Mucina, 2006). As a result, access to the Islands is restricted to ensure a maximum level of conservation. For 13 months at a time, Marion Island allows 20-25 permanent researchers and/or workers to stay on the island. This number may increase up to 80 people to accommodate visiting personnel during the annual relief voyage to the Island. During the recent Covid-19 pandemic, however, these numbers have been less per 13-month period. Prince Edward Island's protection is much stricter, with a maximum of 10 people allowed a short visit of 8 days every 4 years, allowing the Island to remain relatively pristine compared to Marion Island (Chown et al., 2006). Although the Islands are now protected by law and have limited human interactions, they were subject to the culling of seals in the 19th century and well into the 20th century, with sealers taking temporary settlement on the Islands (Gremmen, 1971; Cooper, 2008; Greve et al., 2020). This has directly impacted the Islands' native wildlife and ecosystems, with up to 29 known alien species of terrestrial taxa (vertebrates, invertebrates, plants, and microorganisms) being introduced as a result (Greve et al., 2020). The Islands have remained reasonably undisturbed since the introduction of the Sea Birds and Seals Protection Act (Act 46 of 1973), which not only provides protection for most of the seabirds and seal species through its control of their capture and hunting but also ensures that the Islands are isolated from ship, aircraft, motor vehicle, and human traffic (Smith and Lewis Smith, 1987; Smith and Mucina, 2006).

Chapter 4: Methodology

4.1. Data Acquisition and Pre-Processing using GIS

The data required for species distribution modelling was prepared in various Geographic Information Systems (GIS). Figure 4-1 depicts and explains a workflow that summarizes the activities used to obtain and prepare data for inclusion in Species Distribution Models (SDMs).

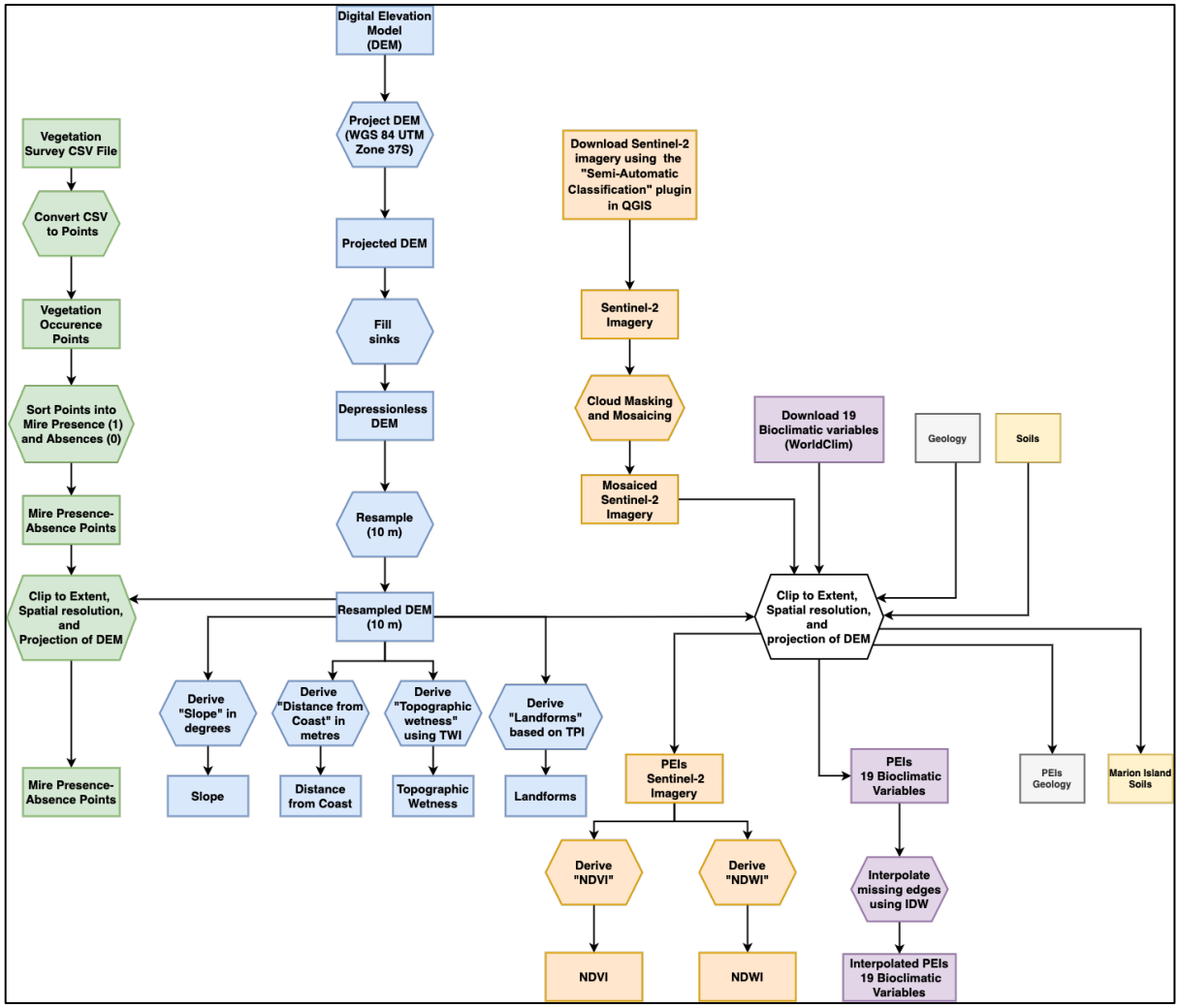


Figure 4-1: Workflow of the operations undertaken in a Geographic Information System (GIS) to acquire and prepare data for use in Species Distribution Models (SDMs).

A vegetation field survey was conducted from 2018 to 2020, using rapid vegetation plots that cover all the main vegetation types on Marion Island. The vegetation complexes proposed by Gremmen and Smith (2008) were used to classify the different vegetation types. The complexes included a Mire, Slope, Fellfield. Polar Desert, Saltspray and Biotic (Gremmen and Smith, 2008). The observations were collected using two sampling methods:

1. Stratified random sampling of plots, using geology to stratify plots, was used to locate plots around Marion Island. Vegetation complex was estimated from species abundance estimates inside the plot area.

2. Transects were walked and the vegetation scored after vegetation type had changed. The vegetation complex was estimated from the dominant species. To classify mires, waterlogged ground was an additional requirement.

The observations are available as a list of coordinates, with a total of 1415 points, 255 mire presence points (based on the descriptions and classification given by Gremmen (1981)), and 1 160 absence points. For this study the species occurrence observations were transformed into binary presence-absence mire observations. The mire complex was employed to indicate the existence of mires, whereas the other vegetation complexes were classified as mire absences. Figure 4-2 shows the distribution of the mire occurrence observations on Marion Island. No training data (mire presence-absence) exists for Prince Edward Island, thus, the model trained on Marion Island was projected onto Prince Edward Island.

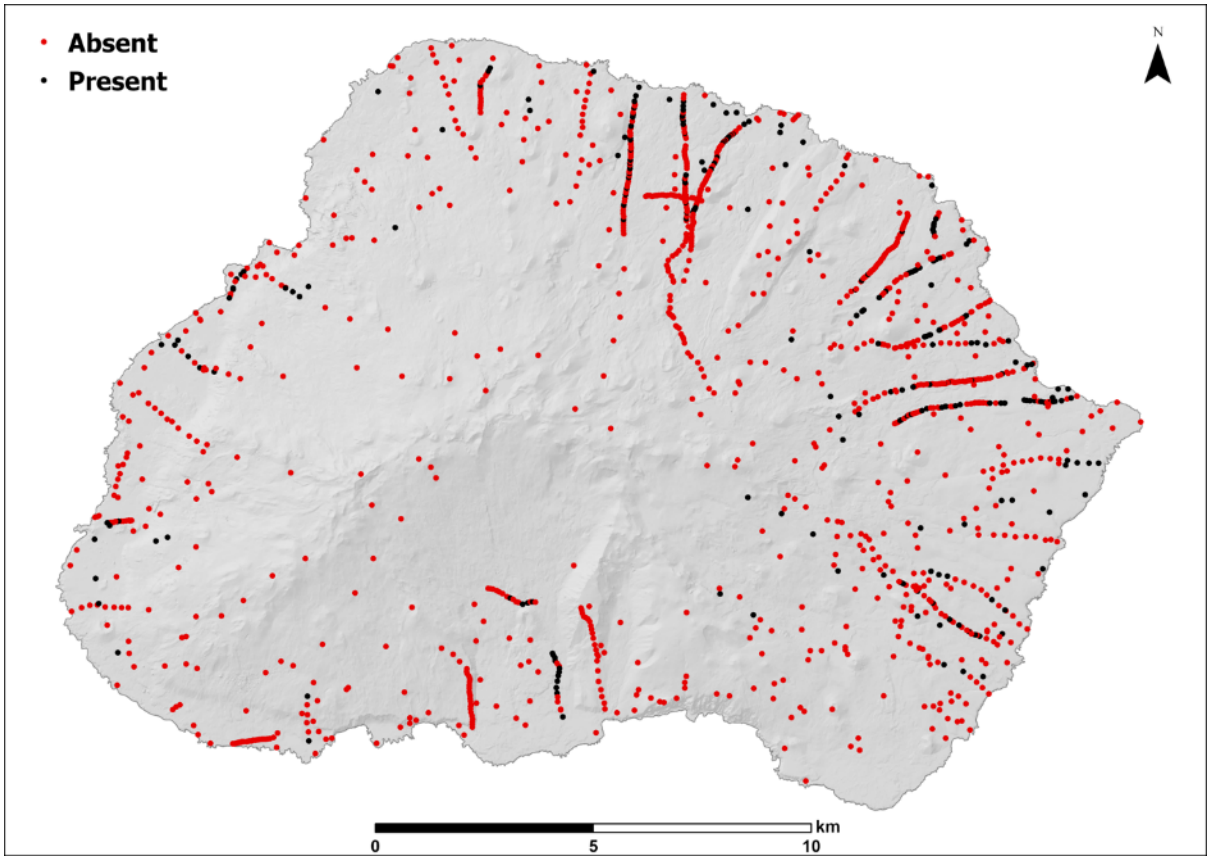


Figure 4-2: All mire presences and absences observations collected on Marion Island.

The explanatory variables to the SDMs were chosen based on their availability and prediction potential for mires on the PEIs (Naimi et al., 2011). Digital Elevation Models (DEMs) of the Prince Edward Islands were acquired from the National Geo-spatial Information (NGI) with 1 m spatial resolution. Random errors in DEM elevation values may result in some cells being lower than the surrounding cells. These cells are referred to as sinks or pits, and hamper flow routing. Such cells are often problematic when determining the flow direction (upon which many hydrological functions depend) across the terrain as there is an erroneous inward flow into the cell without any outward flow (Wang and Liu, 2006; Li et al., 2011; Lindsay, 2016; Hariri et al., 2021). DEM depressions are not always artificial depressions that indicate defects in the DEM, but can be true, real-world terrain features such as lakes, ponds, caves,

or hollows of a karst region that can be easily spotted with high resolution LIDAR DEMs (Wang and Liu, 2006). As a result, sinks were removed from the DEM at a spatial resolution of 1 m using the default parameters of the Fill Sinks (Wang & Liu) tool in SAGA GIS (Wang and Liu, 2006). Standard procedures for creating a depressionless DEM identify the depressions, delineating their catchments and identifying their pour points before determining which cell must be filled and to what height (Wang and Liu, 2006). Using the least-cost search algorithm, the Wang and Liu technique computes a spill elevation for each grid cell without previous delineation of depression catchments, resulting in a depressionless DEM and identifying the locations and depth of surface depressions (Wang and Liu, 2006). The DEMs were resampled to 10 m to match the resolution of the Sentinel-2 Imagery that was also used in this study. Environmental variables, namely, elevation, slope, aspect, distance from coast, topographic positions, and topographic wetness were derived from the DEM. The DEM represents elevation, and the slope was extracted from the DEM using built-in raster functions in ArcGIS Pro 2.8.0. A raster depicting distance from the coast (inward of the island) was created using the Euclidean Distance tool in ArcGIS Pro 2.8.0. Equation (*2*) was used to extract the topographic wetness (TWI) from the DEM using the ArcPy script written by Wolf and Fricker (2013) in ArcGIS Pro.

The classification of landforms on the PEIs was separated into two stages. The initial step was to classify the terrain into four categories of slope positions (1) ridges, 2) mid slopes, 3) flat surfaces, 4) valleys) (Table 2-2, pg. 14) based on the five neighbourhoods for both islands (10 m, 50 m, 100 m, 150 m, and 200 m), as suggested by Becker et al. (2014). This was done to identify the small and large neighbourhood radii that would be utilized in the second stage, where combined TPI at two scales were used to classify the landscape into 10 landform classes (Table 2-3). The workflow of the 4-category slope positions classification is given by Figure 4-3. Code as used in the Esri ModelBuilder is provided below.

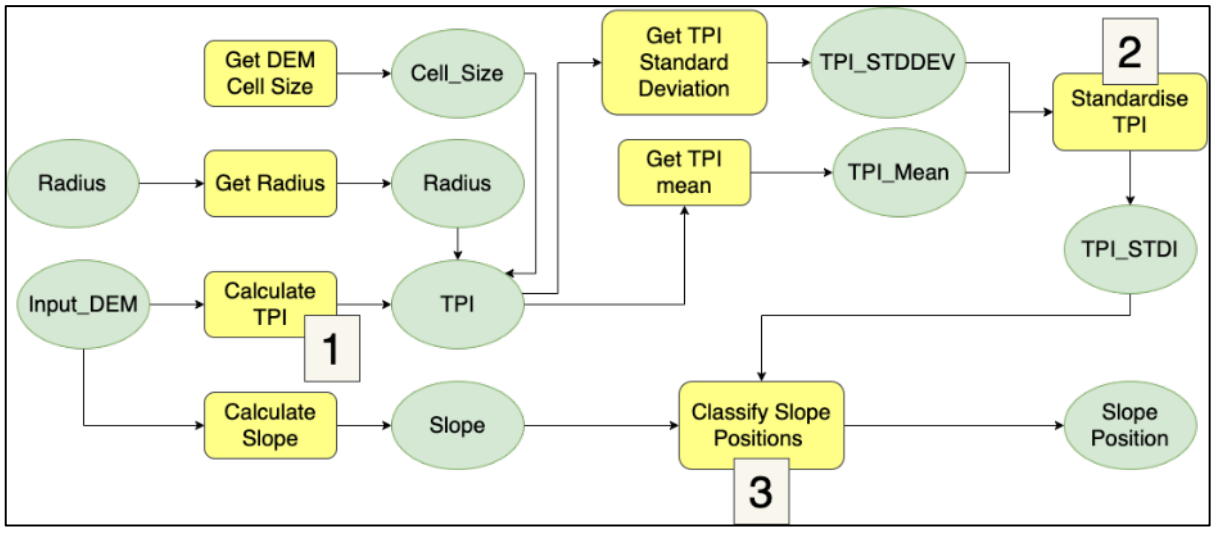


Figure 4-3: Workflow of 4-category slope position classification in ArcGIS. The ModelBuilder code is provided to calculate TPI (1), standardise TPI (2) and classify slope positions (3). Method based on that of Becker et al. (2014).

1. Calculate TPI

Calculation of the TPI grids (given by equation (*3*)), is based on Weiss (2001), where scale denotes the outer radius in map units, irad denotes the inner radius of the annulus in cells, and orad denotes the

outer radius of the annulus in cells. TPI values are standardised using the code given on pg. 31. Slopes are classified as per the code given on pg. 31.

$$TPI\ (scalefactor) = int\ ((dem - focalmean\ (dem, annulus, irad, orad)) + 0.5) \quad (3)$$

ModelBuilder code

```
TPI (Scale Factor) = (Input_DEM – FocalStatistics (Input_DEM, NbrAnnulus((Radius/Cell_Size) – 5, Radius/Cell_Size, CELL), MEAN, DATA)) + 0.5
```

2. Standardise TPI

ModelBuilder code

```
TPI_STDI = (((TPI – TPI_Mean) / TPI_STDDEV) *100) + 0.5
```

3. Classify slope positions

ModelBuilder code

```
Con ((TPI_STDI >= 100),1,
Con ((TPI_STDI > -100) & (TPI_STDI < 100) & (Slope >= 6),2,
Con ((TPI_STDI > -100) & (TPI_STDI < 100) & (Slope < 6),3,
Con ((TPI_STDI <= -100),4,))))
```

Two neighbourhood sizes (small and large) to create an annulus (ring-shaped object) neighbourhood were selected based on the accuracy assessment of the slope position classification. The workflow of the landform classification is given by Figure 4-4. Code as used in the ModelBuilder is provided below as well.

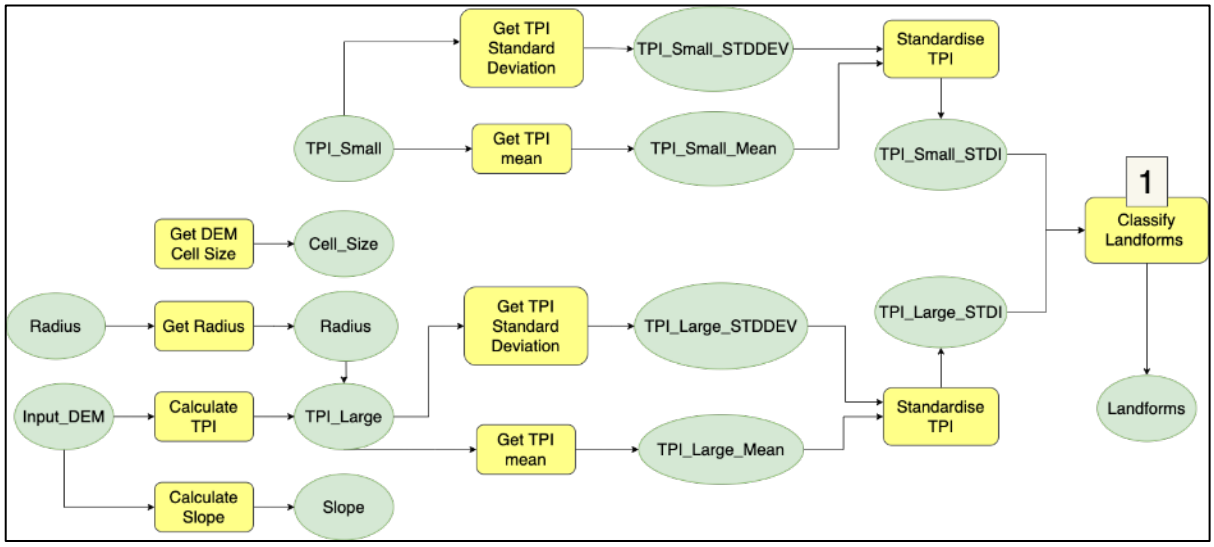


Figure 4-4: Workflow of landform classification in ArcGIS. ModelBuilder code is provided to classify landforms (1). Method based on that of Becker et al. (2014).

Up until the classification of the landforms, all the steps are like those used in the slope position classification (pg. 30 onwards). The classification of landforms was achieved using code below.

ModelBuilder code

```
Con ((TPI_Small_STDI > -100) & (TPI_Small_STDI < 100) & (TPI_Large_STDI > -100) & (TPI_Large_STDI < 100) & (Slope <= 5),5,
Con ((TPI_Small_STDI > -100) & (TPI_Small_STDI < 100) & (TPI_Large_STDI > -100) & (TPI_Large_STDI < 100) & (Slope >= 6),5,
Con ((TPI_Small_STDI > -100) & (TPI_Small_STDI < 100) & (TPI_Large_STDI >= 100),7,
Con ((TPI_Small_STDI > -100) & (TPI_Small_STDI < 100) & (TPI_Large_STDI <= - 100),4,
Con ((TPI_Small_STDI <= 100) & (TPI_Large_STDI > -100) & (TPI_Large_STDI < 100), 2,
Con ((TPI_Small_STDI >= 100) & (TPI_Large_STDI > -100) & (TPI_Large_STDI < 100), 9,
Con ((TPI_Small_STDI <= -100) & (TPI_Large_STDI >= 100), 3,
Con ((TPI_Small_STDI <= -100) & (TPI_Large_STDI <= -100), 1,
Con ((TPI_Small_STDI >= 100) & (TPI_Large_STDI >= 100), 10,
Con ((TPI_Small_STDI >= 100) & (TPI_Large_STDI >= -100), 8,))))))))))
```

The Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Water Index (NDWI) were employed as proxies for local vegetation cover and surface wetness (i.e., surface water presence), respectively. They were extracted from the freely available Sentinel-2 user product from the European Space Agency (ESA). Sentinel-2 products are geometric and radiometrically corrected images. Sentinel-2 multispectral imagery has a fine spatial resolution (i.e., 10 m for four bands) and a high revisit rate (10 days) at free cost, making it good for the mapping of water bodies (Du et al., 2016). A Sentinel-2 image has a total of 13 bands, with different spatial resolutions (as shown in Table 4-1). Due to persistent cloud cover on the PEIs, it was not possible to acquire a cloud free image for Marion Island, therefore, three images (two from October 5, 2020, and one from October 10, 2020) were used to create a mosaic for Marion Island, while the cloud free image for Prince Edward Island is from November 10, 2017. The resulting Marion Island mosaic is depicted in Figure 4-6 (pg. 34). The mean monthly total rainfall and mean temperature on Marion vary slightly in a year, as seen in Figure 4-5 (pg. 33) (Sadiki, 2019). As a result, there are no seasons to consider while gathering images, therefore comparing the influence of imagery obtained during different months of the year was unnecessary. The Semi-Automatic Classification plugin in Quantum GIS (QGIS) was used to download the data, mask the clouds, and mosaic the images to create a mosaic of the images. The Normalized Difference Vegetation Index (NDVI) was used as a proxy for local vegetation cover on the Prince Edward Islands (PEIs). The NDVI is an index intended to assess surface biomass and productivity using the near infrared (NIR) and red bands to enhance the presence of vegetation, while it has no effect on water features (McFeeters, 1996). It is defined by equation (*4*) below and has already been described in detail in Wetland Mapping (pg. 8 onward). In this index the near-infrared (NIR) band is used primarily for imaging vegetation and the red (Red) band is mainly used to image man-made objects, soils, and vegetation.

$$\mathbf{NDVI} = \frac{NIR - Red}{NIR + Red} \qquad (4)$$

The Normalized Difference Water Index (NDWI), already described in detail in Wetland Mapping (pg. 8 onward), was used as a proxy for surface wetness over the Prince Edward Islands. McFeeters (1996) suggested the NDWI, which is an adaptation of the NDVI, where the red band in the NDVI is replaced by the green band, used for imaging vegetation and deep-water structures. Thus, McFeeters (1996) was able to introduce a new method for delineating open water features from satellite imagery, while simultaneously removing soil and vegetation features. The NDWI is based on the idea that water has the highest absorption capabilities and vegetation has the highest reflectance in the near infrared (NIR) (Huang et al., 2018a). The NDWI is defined by equation (*5*).

$$\mathbf{NDWI} = \frac{Green - NIR}{Green + NIR} \qquad (5)$$

Table 4-1: Sentinel-2 image spectral band specifications.

Band	Spectral Region	Central wavelength (µm)	Spatial Resolution (m)
B1	Coastal aerosol	0.443	60
B2	Blue	0.490	10
B3	Green	0.560	10
B4	Red	0.665	10
B5	Vegetation red edge	0.705	20
B6	Vegetation red edge	0.740	20
B7	Vegetation red edge	0.783	20
B8	Near infrared	0.842	10
B8A	Vegetation red edge	0.865	20
B9	Water vapour	0.945	60
B10	Shortwave infrared band - Cirrus	1.375	60
B11	Shortwave infrared band	1.610	20
B12	Shortwave infrared band	2.190	20

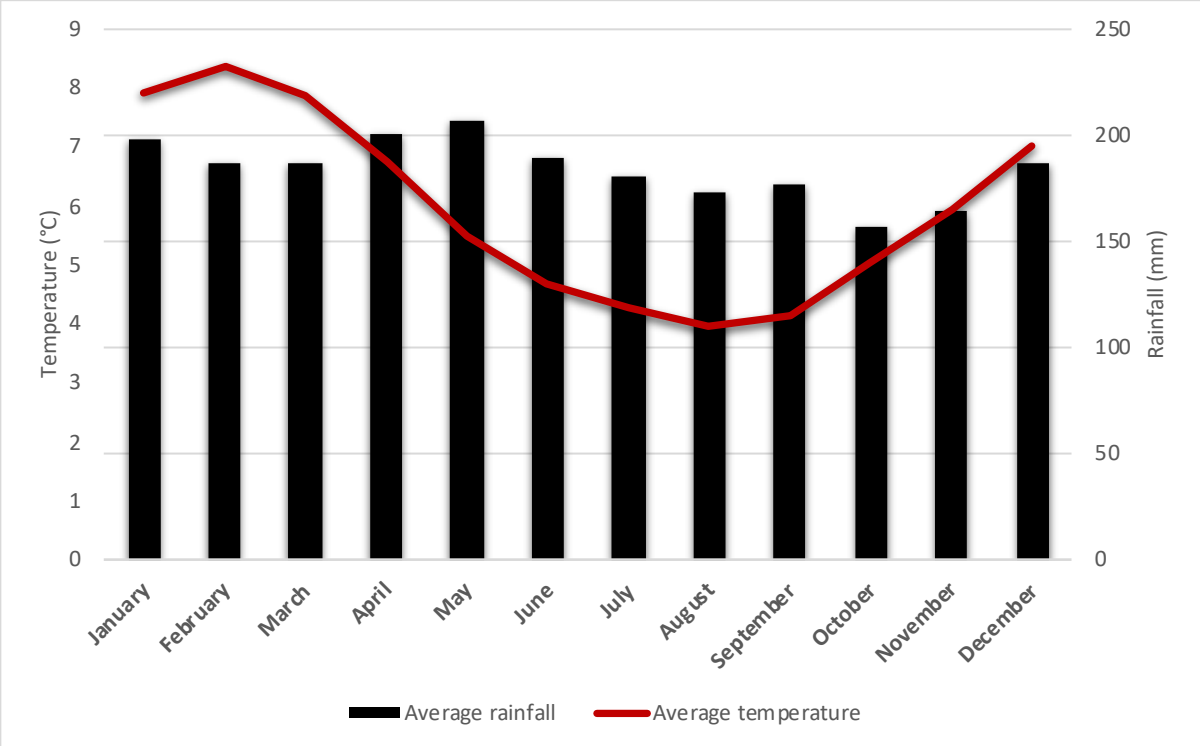


Figure 4-5: Monthly mean total rainfall and mean temperature on Marion Island (Sadiki, 2019).

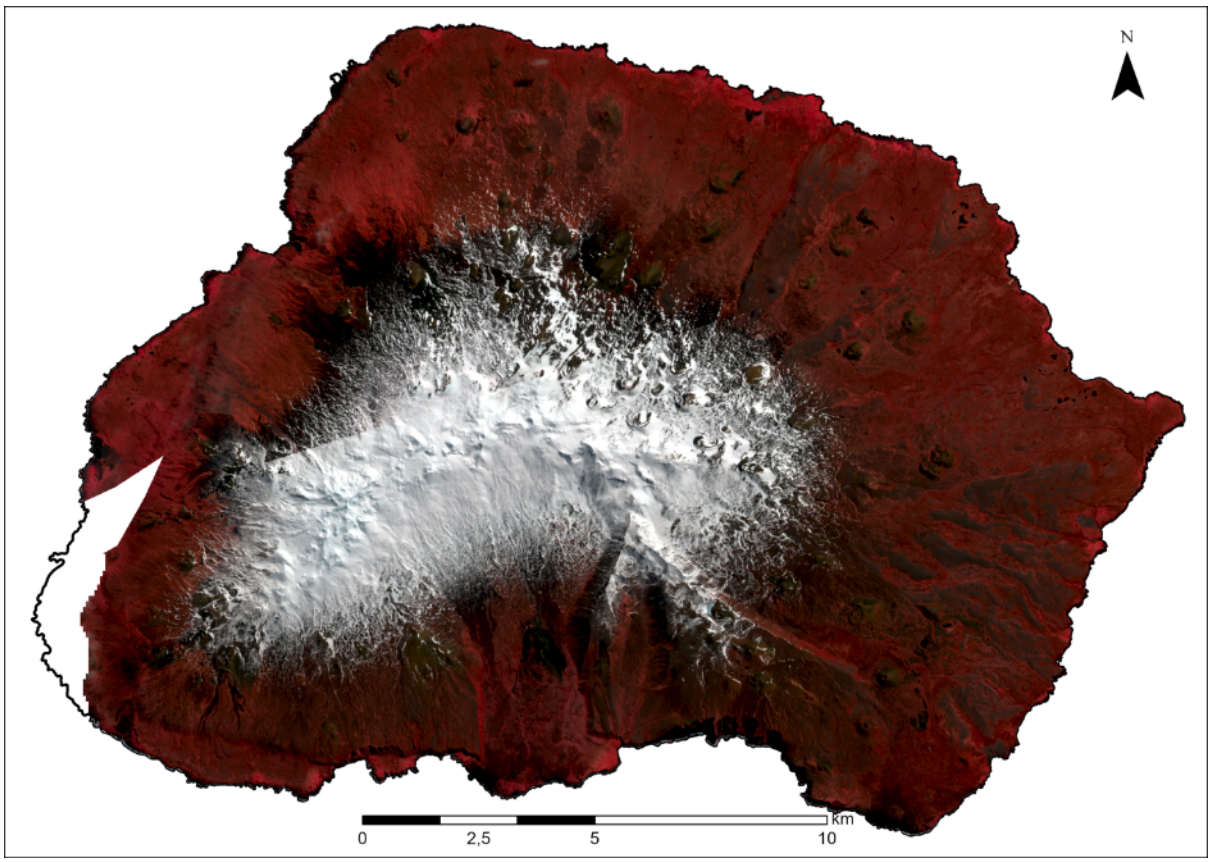


Figure 4-6: Marion Island Sentinel-2 mosaic. Three images (two from October 5th, 2020, and one from October 10th, 2020) to create a mosaic for Marion Island.

Climate data were acquired, at a spatial resolution of 30 arc seconds (~1 km²), from the WorldClim database, a database of high spatial resolution global weather and climate data for use in research and related activities. The data represent annual trends, seasonality and limiting environmental factors for the years between 1970 and 2000. Current climate data were downloaded from WorldClim Version 2 (WorldClim2). The data include 19 bioclimatic variables (Table 4-2) averaged between 1970 and 2000 at a spatial resolution of 30 arc second (~1 km (900 m) at the equator) (Fick and Hijmans, 2017). The bioclimatic datasets are known to possess high levels of multicollinearity as they are derived from mean maximum and minimum temperature, and precipitation values (Hijmans et al., 2005). Therefore, only variables thought to affect the occurrence of mires were considered in this study. These variables were annual mean temperature (BIO01), temperature seasonality (BIO04), annual precipitation (BIO12) and precipitation seasonality (BIO15) (Essl et al., 2012). The standard deviation of the 12 mean monthly temperature values was used to compute temperature seasonality. To maintain significant digits, the resulting value in degrees Celsius was multiplied by 100 (O'Donnell and Ignizio, 2012). The annual precipitation was computed by adding the precipitation values from each of the year's 12 months. To calculate precipitation seasonality, the standard deviation was calculated for each of the 12 monthly precipitation totals, then divided by the mean monthly precipitation value, and for cases where the mean rainfall was less than one, one was added to the denominator to avoid odd coefficient of variation values, and finally, the value was multiplied by 100 to give precipitation seasonality as a percentage (O'Donnell and Ignizio, 2012). The WorldClim data do not overlap perfectly with the coastal zone of the PEIs, which include some of the occurrence points. As such, Inverse Distance Weighting (IDW) was used to interpolate the values of the climatic variables in these areas. To determine cell values, inverse distance

weighted (IDW) interpolation employs a linearly weighted combination of a set of sample points, with the inverse distance determining the weight. To perform IDW, the bioclimatic variable raster was transformed to points (at 10 m), using the centre point of each cell, and the IDW tool in ArcGIS Pro 2.8 was used to interpolate the surfaces to include the missing value regions.

Table 4-2: Description of the 19 WorldClim bioclimatic variables (O'Donnell and Ignizio, 2012).

Variables	Description	Unit	Equation
BIO01	Annual Mean Temperature	°C	$(\sum_{i=1}^{12} t_i)/12$
BIO02	Mean Diurnal Range	°C	$(\sum_{i=1}^{12} (t_{max} - t_{min}))/12$
BIO03	Isothermality	%	$(t * (t_{max} - t_{min})/t_{max} - t_{min}) * 100$
BIO04	Temperature Seasonality (Coefficient of variation)	°C	$(\sqrt{\frac{1}{11}\sum_{i=1}^{12}(t_i - (\sum_{i=1}^{12} t_i/12))^2}) * 100$
BIO05	Max Temperature of Warmest Month	°C	
BIO06	Min Temperature of Coldest Month	°C	
BIO07	Temperature Annual Range	°C	$t_{max} - t_{min}$
BIO08	Mean Temperature of Wettest Quarter	°C	
BIO09	Mean Temperature of Driest Quarter	°C	
BIO10	Mean Temperature of Warmest Quarter	°C	
BIO11	Mean Temperature of Coldest Quarter	°C	
BIO12	Annual Precipitation	mm	$\sum_{i=1}^{12} p_i$
BIO13	Precipitation of Wettest Month	mm	
BIO14	Precipitation of Driest Month	mm	
BIO15	Precipitation Seasonality (Coefficient of variation)	%	$(\frac{\sqrt{\frac{1}{11}\sum_{i=1}^{12}(p_i - (\sum_{i=1}^{12} p_i/12))^2}}{\frac{\sum_{i=1}^{12} p_i}{12}})$
BIO16	Precipitation of Wettest Quarter	mm	
BIO17	Precipitation of Driest Quarter	mm	
BIO18	Precipitation of Warmest Quarter	mm	
BIO19	Precipitation of Coldest Quarter	mm	

Additional data on the geology and soils of the Islands was acquired. There are five classes in both the geology and soil layer. See Table 4-3.

Table 4-3: Prince Edward Islands geology and soil classes.

Class Number	Geology Classes	Soil Classes
1	Black lava (incl. recent: post 1980)	Histosols
2	Grey lava	Andosols
3	Grey-bedded ash	Regosols
4	Scoria	Histic andosols
5	Wind-blown ash	Solid rock

Table 4-4 (pg. 36) summarizes the environmental predictor variables that were collected for the study. All predictor variables were resampled to the extent of the DEMs for the PEIs using ArcGIS Pro 2.8. As a result, the variables were standardised in terms of data format (GeoTiff), spatial resolution (10 m), projection (WGS 84 UTM Zone 37S), and extent. The spatial resolution of 10 m was based on the spatial resolution of the Sentinel-2 imagery. Although the finer spatial resolution does not improve the

resolution of the predictor variables with lower spatial resolution, it was essential as the software utilised (RStudio) requires that all predictor variables have the same extent and projection.

Table 4-4: Predictor variables used in the study.

Variable Type	Predictor Variables	Data type	Source
Topographic (DEM derived)	Elevation (msl)	Continuous	DEM
	Slope (°)	Continuous	DEM
	Distance from coast (m)	Continuous	DEM
	Topographic wetness (TWI)	Continuous	DEM
	Landforms (TPI)	Categorical	DEM
Geology and Soils	Geology	Categorical	(Rudolph et al., 2020)
	Soils	Categorical	(Lubbe, 2010)
Satellite imagery	Vegetation density (NDVI)	Continuous	ESA
	Surface wetness (NDWI)	Continuous	ESA
Bioclimatic	BIO01: Annual Mean Temperature	Continuous	WorldClim
	BIO04: Temperature Seasonality	Continuous	WorldClim
	BIO12: Annual Precipitation	Continuous	WorldClim
	BIO15: Precipitation Seasonality	Continuous	WorldClim

4.2. Species Distribution Models

A workflow depicting the steps followed (model data, model development, and evaluation) to construct and evaluate the SDMs is shown in Figure 4-7.

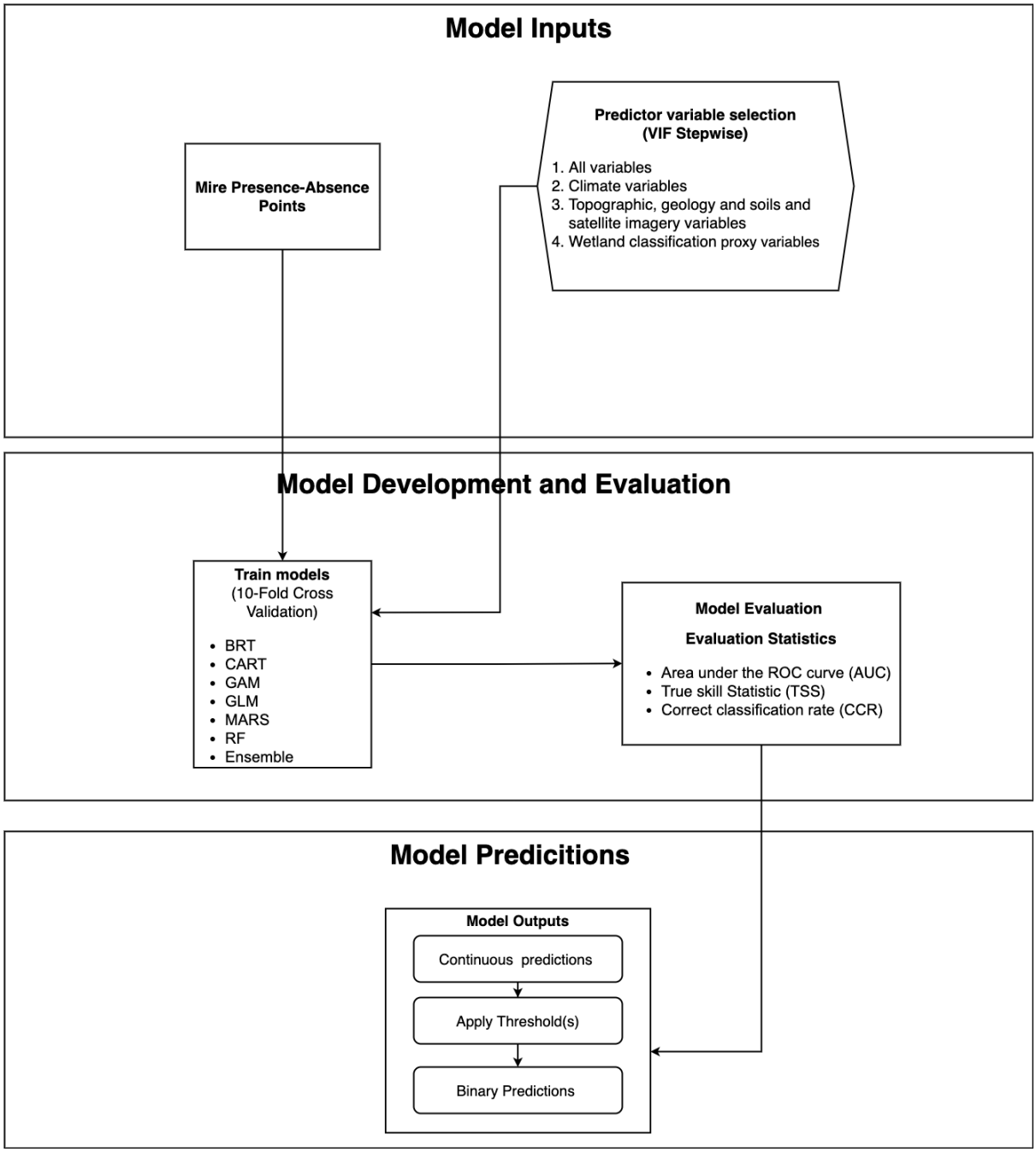


Figure 4-7: Workflow of the Species Distribution Modelling (SDM) for mire occurrence.

In this study, SDMs were developed using a presence-absence approach in the 'sdm' package in R (Naimi and Araújo, 2016). As the mire occurrences dataset (Figure 4-2, pg. 29) has presences and true absences, presence-absence algorithms were implemented. Models as listed in Table 4-5, are either regression-based and machine learning algorithms or an Ensemble of methods; models were implemented using the default model settings (Naimi and Araújo, 2016).

Table 4-5: The regression based, and machine learning presence-absence algorithm implemented in this study.

Model	Description	Reference
BRT	Boosted Regression Trees	(Elith et al., 2008)
CART	Classification and Regression Trees	(Breiman et al., 1984)
GAM	Generalised Additive Models	(Guisan et al., 2002)
GLM	Generalised Linear Models	(McCullagh and Nelder, 1989)
MARS	Multivariate Adaptive Regression Splines	(Friedman, 1991)
RF	Random Forests (RF)	(Breiman, 2001)

4.2.1. Model Inputs

Due to the cloud cover issue on Marion Island (Figure 4-6, pg. 34), some mire presence-absence observations were excluded from the study, leaving 1375 observations (248 presence, 1127 absences). SDMs are often built utilising predictor data in the hopes of determining which of those characteristics are important to the species' occurrence; however, some believe that only ecologically relevant predictor data should be utilized instead (Elith and Leathwick, 2009). Peatland development areas have distinct hydrologic regimes, climates, chemistry, landforms, substrates, and flora (Bourgeau-Chavez et al., 2018; Minasny et al., 2019). As it is difficult to accurately identify which environmental variables drive the distribution of mires on the Prince Edward Islands (PEIs), there is a need to test a combination of predictors to identify the best ones. Models were created using various variable combination groups to compare them. The variable combination groups were based on six variable scenarios. The variable scenarios included modelling:

1. All variables;
2. Climate variables;
3. Topographic, geology and soils and satellite imagery variables; and
4. Wetland classification proxy variables.
    a. The Ramsar convention classification system
    b. The Hydrogeomorphic (HGM) classification system classification
    c. The International Union for Conservation of Nature (IUCN) Global Ecosystem Typology 2.0

Collinearity among predictor variables can lead to spurious results, therefore, correlation between the predictor variables was assessed as variables that are not independent may result in multicollinearity issues in the models (Dormann et al., 2013; Naimi and Araújo, 2016), which reduces SDM efficiency and increases model uncertainty (De Marco and Nóbrega, 2018). Leihy et al. (2018) stressed that variable collinearity is of concern on the sub-Antarctic islands. By assessing intercorrelation between the predictors, redundant predictors can be identified and consequently removed, to avoid using unreliable variables (Hiestermann and Rivers-Moore, 2015). There are several ways to detect collinearity and the variance inflation factor (VIF), which measures how strongly the rest of the predictor variables can explain each predictor, is often used (Naimi and Araújo, 2016). To avoid overfitting, the collinear variables in this study were identified and excluded using the VIF (vifstep) approach in RStudio (Naimi and Araújo, 2016). The VIF stepwise technique calculates VIF measures using equation (*6*), with values larger than 10 indicating strong correlation between variables, and hence removing them

owing to collinearity issues. As a stepwise process, the VIF values were recalculated until all values were below the threshold.

$$\mathbf{VIF} = \frac{1}{1 - R^2}; \textit{ where } R^2 \textit{ is the regression equation's coefficient of determination} \quad (6)$$

The study employs the VIF stepwise technique to exclude highly collinear variables. For visualisation purposes, a correlation matrix presenting correlation among the predictor variables for all predictor variable scenarios was computed based on Pearson's correlation coefficients for all predictor variable scenarios and presented in Table A-1 (Appendix 1, pg. 75). If two variables have a Pearson (r) correlation coefficient larger than a threshold, for instance 0.7, this might indicate that they are correlated, and using both variables in the model development may produce collinearity problems. (Naimi and Araújo, 2016). Removing predictor variables with multicollinearity issues necessitates caution when interpreting the effect of the remaining variables on the distribution of the species. Whatever influence the remaining variable has on the distribution, the variable removed for being correlated to that variable likely has a similar effect.

Scenario 1: all variables

As suggested by Table A-1 (Appendix 1, pg. 75), Elevation, NDVI and Annual Precipitation (BIO12) have collinearity issues, therefore, they were removed from the list of predictor variables for this scenario using the VIF stepwise approach.

Scenario 2: bioclimatic variables

In the list of identified predictor variables (Table 4-4, pg. 36), four bioclimatic variables were included. Collinearity among WorldClim bioclimatic variables is known to be high and depicted by correlation matrix presented in Table A-1 (Appendix 1, pg. 75), Annual Precipitation (BIO12) is highly correlated with Annual Mean Temperature (BIO01) and Precipitation Seasonality (BIO15). Therefore, it was removed from the list of bioclimatic variables in this scenario, leaving three bioclimatic variables.

Scenario 3: topographic, geology and soils and satellite imagery variables

One of the nine variables in this scenario (NDVI) exhibited multicollinearity issues with NDWI and was removed from this scenario.

Scenario 4: wetland classification proxy variables

Wetland predictors were identified based on some common wetland classification methods used in South Africa, namely, the 1) Ramsar classification, the 2) Hydrogeomorphic System, and the 3) International Union for Conservation of Nature's (IUCN's) typology for wetland ecosystem types.

a. ***The Ramsar convention classification system*** divides wetlands into three major categories: marine and coastal; inland; and human-made; it also divides wetlands into subcategories based on their location, water permanence, soils, substates, and flora (Ramsar Convention Secretariat, 2010; Finlayson, 2018). As such, the topographic wetness (TWI) was selected as a proxy for water permanence, and vegetation density (NDVI) for vegetation, in addition to soils. None of the three variables have collinearity issues.

b. ***The Hydrogeomorphic (HGM) classification system*** is based on the premise that the landscape influences the flow and accumulation of water, and hydrology and geomorphic setting (landforms) are the most obvious characteristics used to characterise wetlands in this system (Semeniuk and Semeniuk, 1995; Ollis et al., 2013; Sieben et al., 2018). The system categorises wetlands into classes based on geomorphic, water supply, and hydrodynamic properties (Ollis et al., 2013). Therefore, landforms (modelled using the TPI), surface wetness (NDWI) and topographic wetness (modelled using the TWI) were selected as proxy predictor variables for the occurrence of mires. None of the three variables have collinearity issues.

c. ***The International Union for Conservation of Nature (IUCN)*** **Global Ecosystem Typology 2.0** describes the profiles of biomes and ecosystem functional groups, providing key ecological traits of functionally different ecosystems and their drivers (Keith et al., 2020). Marion Island is structurally and functionally characteristic of the most climatically harsh variety of tundra, high Arctic polar deserts (Smith, 2008). Therefore, the PEIs fit the description of the ecological traits of polar tundra and deserts as described in the IUCN Global Ecosystem Typology 2.0 (Keith et al., 2020). As such, the vegetation density proxy (NDVI) and the four bioclimatic variables were selected as proxy predictor variables for the occurrence of mires. The VIF stepwise approach removed Annual Precipitation (BIO12) due to correlation issues with Annual Mean Temperature (BIO 1) and Precipitation Seasonality (BIO15).

The predictor variables, with multicollinearity accounted for among all the variables in each variable scenario, are presented in Table 4-6.

Table 4-6: Predictor variables per variables scenario with multicollinearity accounted for. A check mark (✓) indicates that a variable was included in a particular variable scenario.

Variable Type	Predictor Variable	Variable Scenario					
		1	2	3	4a	4b	4c
Topographic (DEM derived)	Elevation			✓			
	Slope	✓		✓			
	Distance from coast	✓		✓			
	Topographic wetness (TWI)	✓		✓	✓	✓	
	Landforms (TPI)	✓		✓		✓	
Geology and Soils	Geology	✓		✓			
	Soils	✓		✓	✓		
Satellite imagery	Vegetation density (NDVI)				✓		✓
	Surface wetness (NDWI)	✓		✓		✓	
Bioclimatic	BIO01: Annual Mean Temperature	✓	✓				✓
	BIO04: Temperature Seasonality	✓	✓				✓
	BIO15: Precipitation Seasonality	✓	✓				✓

4.2.2. Model Development and Evaluation

As previously discussed, SDMs were developed using a presence-absence approach in the 'sdm' package in R (Naimi and Araújo, 2016). The models used to predict the distribution of mires on the PEIs (see Table 4-5, pg. 38) were evaluated to find the method that would provide the most accurate prediction. The R 4.1 software "sdm" package developed by Naimi and Araújo (2016) was used to validate the models using the K-fold cross-validation with 10-folds technique, which is the most common way of measuring model accuracy.

The models' performance was assessed using a multi-metric method that included the metrics listed in Table 4-7. It is important to utilise different metrics to assess model performance since each measures a distinct component of predictive performance (Elith and Graham, 2009). To assess model predictive performance in this study, three approaches were used: 1) area under the curve (AUC) of a receiver operating characteristic (ROC) plot, 2) the True Test Statistic (TSS) and 3) the Correct Classification Rate (CCR). The ROC curve is the most popular metric to compare SDM performance (Hijmans and Elith, 2013; Yu et al., 2020). By summarizing total model performance over all potential thresholds, it eliminates subjectivity in the threshold selection process, when continuous probability generated scores are transformed to a binary presence–absence variable (Lobo et al., 2008; Peterson et al., 2008). Area under the curve (AUC) values vary from 0 to 1, and an AUC of 0.5 or less indicates that the model is as good as a random guess (i.e., imply that the model is no better than randomly predicting presences against absences), and values larger than 0.7 indicate that the model is sufficient for modelling species distributions (Swets, 1988). The TSS metric compares the proportion of correct predictions to the proportion of hypothetical predictions, disregarding any predictions that result when the model is no better than predicting presences versus absences at random (Allouche et al., 2006). TSS values range between -1 and +1, with larger positive values indicating a better model. By crossing all presences-absences with the binary (presence/absence) predictions, the correct classification rate (CCR) was used to quantify the overall model prediction accuracy (Fielding and Bell, 1997; Martínez-Freiría et al., 2016).

All models with AUC values sufficient for species distribution modelling (0.70 or more) and TSS values indicative of at least fair model performance (0.2 or more) were used to create an ensemble model for each scenario. The ensemble models were created using the 'ensemble' function based on weighted TSS values, with higher statistical weights assigned to models with higher TSS values (Naimi and Araújo, 2016).

The relative relevance for each predictor variable in explaining the current distribution of a mire ecosystem was also assessed. High values indicate that the predictor variable is more important.

Table 4-7: Metric ranges and thresholds for evaluating model predictive performance.

Metric	Range	Threshold	Source
AUC	0 to 1	> 0.7	(Swets, 1988; González-Ferreras et al., 2016)
TSS	-1 to +1	> 0.2	(González-Ferreras et al., 2016)
CCR	0 to 1	Larger value indicates a better model	(Fielding and Bell, 1997)

Model predictions were used to assess mire distribution predictions. The outcome was a series of continuous surfaces that reflect the probability of finding a mire across the region. By applying a statistically determined threshold to the continuous surface and selecting the value that maximized the sum of sensitivity and specificity (i.e., sensitivity + specificity/ 2) (comparable to optimizing the TSS), ArcGIS Pro 2.8 was used to construct binary presence or absence maps (Liu et al., 2005).

Chapter 5: Results

5.1. Derived Topographic and Environmental Parameters

The topographic and environmental parameters required for species distribution modelling derived from digital elevation models (DEMs) and satellite imagery for Marion Island and Prince Edward Island are presented by Figure 5-1 (pg. 43), and Figure 5-2 (pg. 44), respectively. The parameters include elevation, distance from coast, slope, Normalised Difference Vegetation Index (NDVI) as a proxy for vegetation density, Normalised Difference Water Index (NDWI) as a proxy for surface wetness, Topographic Wetness Index (TWI), and landforms (TPI based landform classification). No soils dataset was acquired for Prince Edward Island, therefore, there were 13 predictor variable datasets for Marion Island and 12 for Prince Edward Island. The range of values for each predictor variable are presented by Figure 5-3 (pg. 45), and Figure 5-4 (pg. 46).

The results of the landform classification process is also described in this section. The results of the accuracy of the slope position classification is presented in Table 5-1. As it was impossible to gather ground-truth points through fieldwork, 50 random points were generated in ArcGIS pro and referenced using a Marion Island satellite image with a spatial resolution of approximately 0.5 m. All 5 neighbourhoods managed to classify the slope positions with overall accuracy greater than 60%. The Kappa statistic, which expresses the agreement between the reference ('ground-truth') and classification data, gave poor agreement (less than 0.4) for the 10 m neighbourhood and moderate agreement (between 0.4 and 0.8) for all other neighbourhoods. Therefore, it can be said that the 50 m, 100 m, 150 m and 200 m neighbourhoods were able to fairly classify the slope positions.

Table 5-1: Marion Island slope position classifications accuracy assessment.

Classification Neighbourhood	Overall Accuracy	Kappa Statistic
10 m	64%	0.35
50 m	74%	0.57
100 m	70%	0.52
150 m	72%	0.55
200 m	72%	0.55

In this study, the 10 m and 50 m neighbourhoods were considered small neighbourhood sizes, while 100 m, 150 m and 200 m are considered larger neighbourhood sizes. The 50 m slope position classification achieved a higher overall accuracy as well as agreement between the reference and classification data (Kappa statistic) than the 10 m slope position classification. Therefore, the 50 m slope position classification neighbourhood was selected as the small neighbourhood (radius) for the TPI based landform classification. The 150 m and 200 m neighbourhoods performed the same; the larger of the two (200 m) was selected as the preferred larger neighbourhood size. As such, an annulus (ring-shaped object) neighbourhood (R1 = 50 m & R2 = 200 m) was used to classify landforms for the Prince Edward Islands (PEIs).

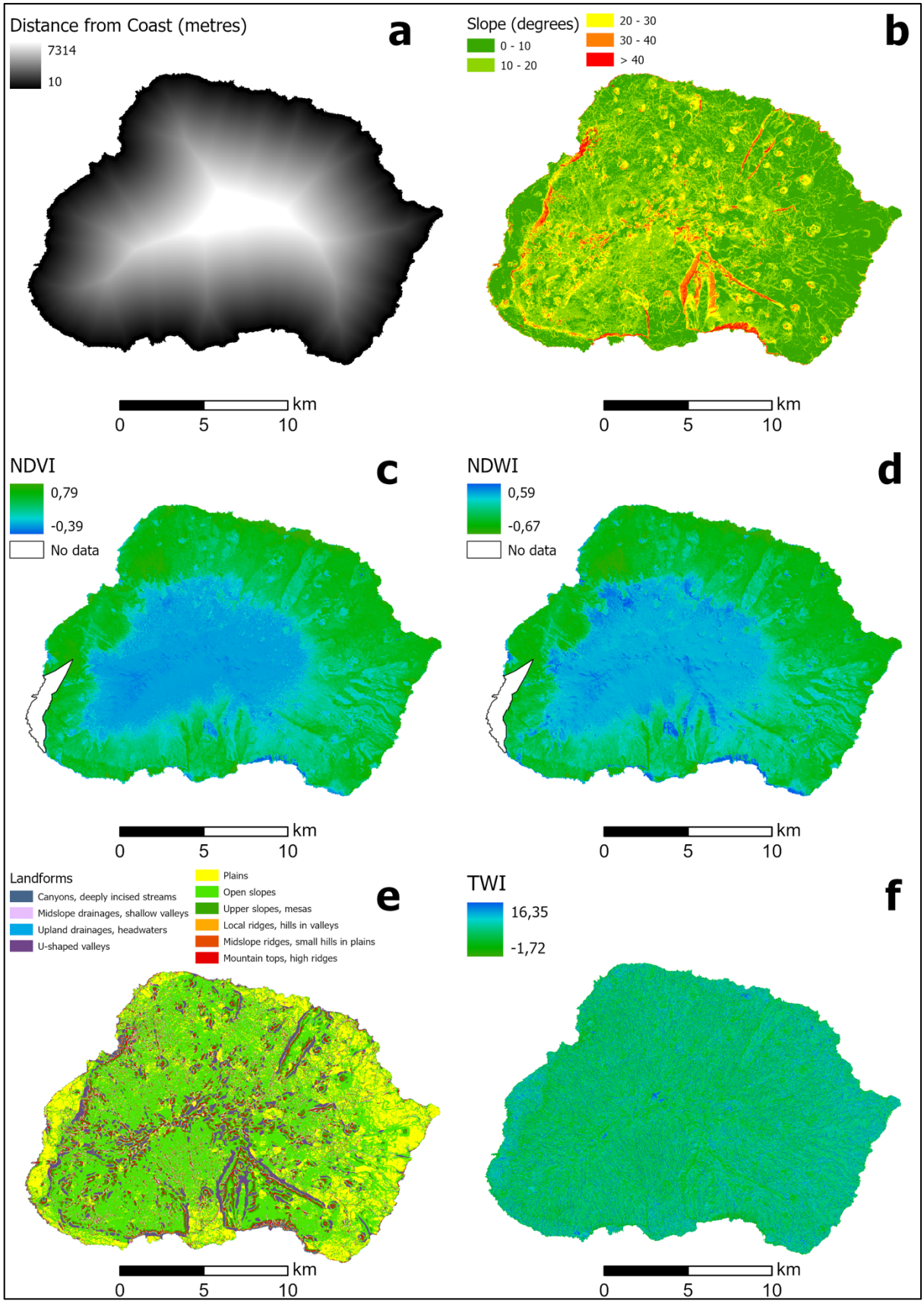


Figure 5-1: Derived topographic and environmental parameters for Marion Island required as predictor variables for species distribution modelling. Distance from Coast (a), Slope (b), NDVI (c), NDWI (d), Landforms (e), TWI (f).

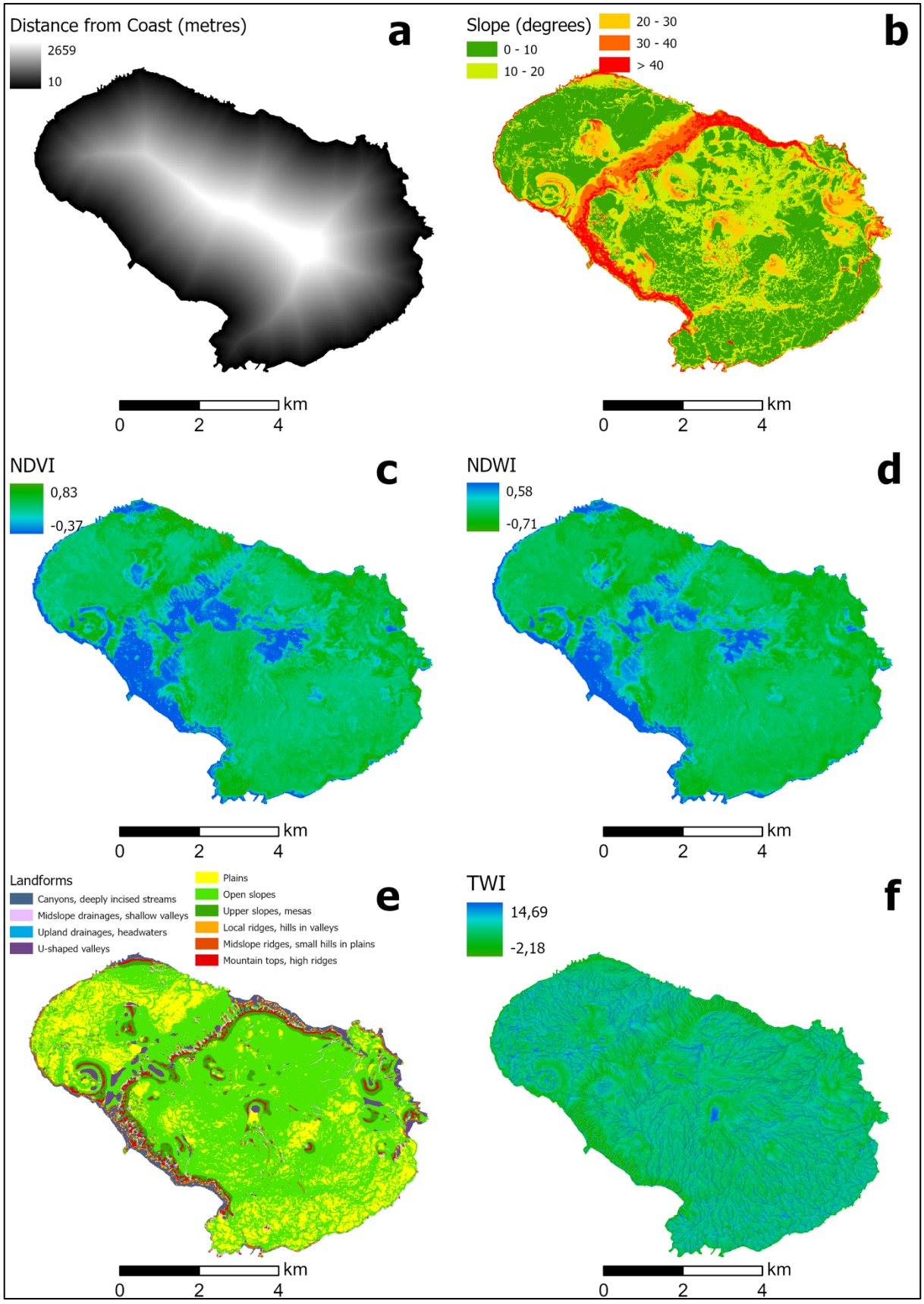


Figure 5-2: Derived topographic and environmental parameters for Marion Island required as predictor variables for species distribution modelling. Distance from Coast (a), Slope (b), NDVI (c), NDWI (d), Landforms (e), TWI (f).

mir

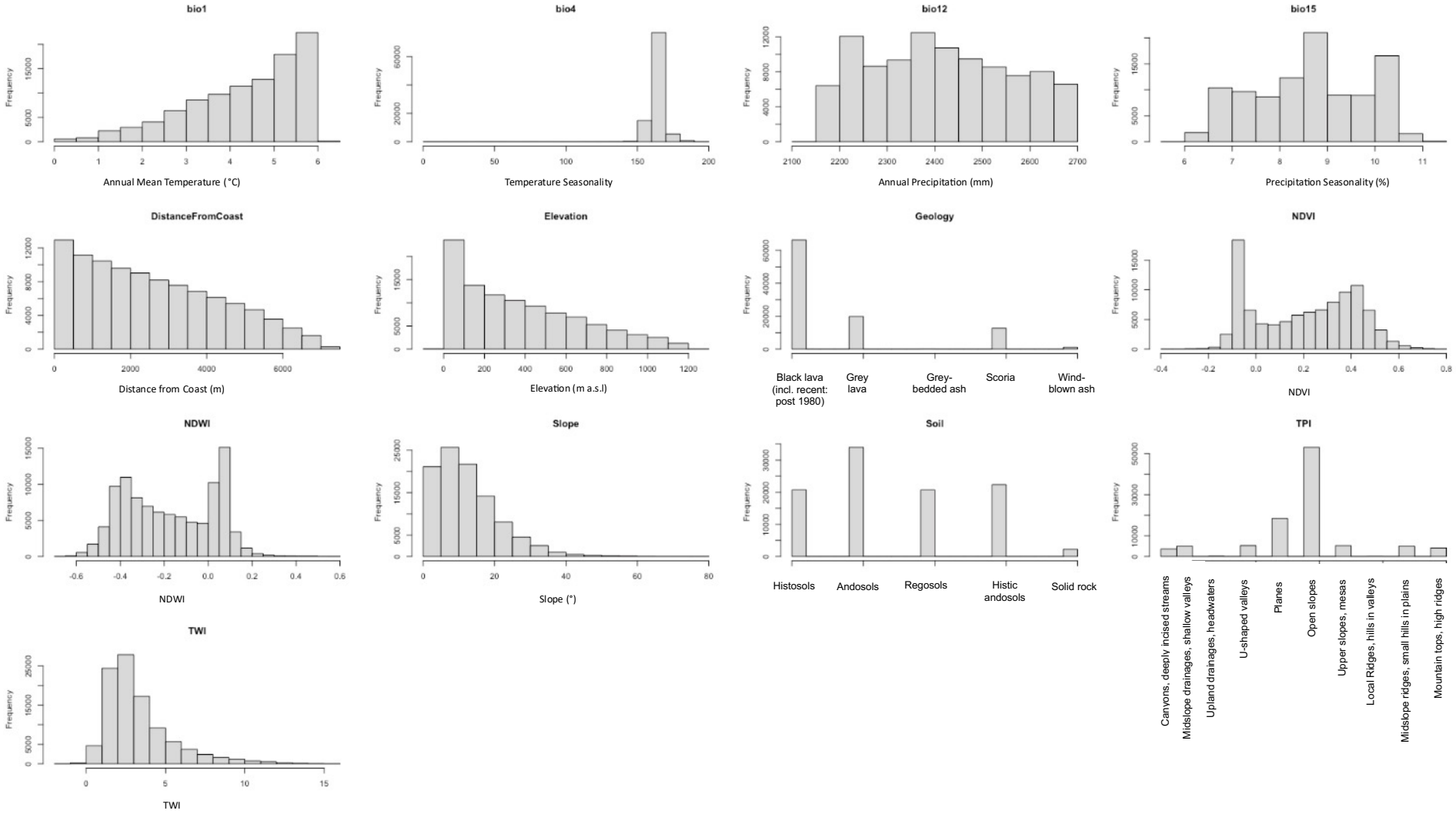


Figure 5-3: Range of values for each of Marion Island's predictor variables. The bioclimatic variables' abbreviations are as follows: bio1 = Annual Mean Temperature , bio 4 = Temperature Seasonality , bio12 = Annual Precipitation, bio15 = Precipitation Seasonality.

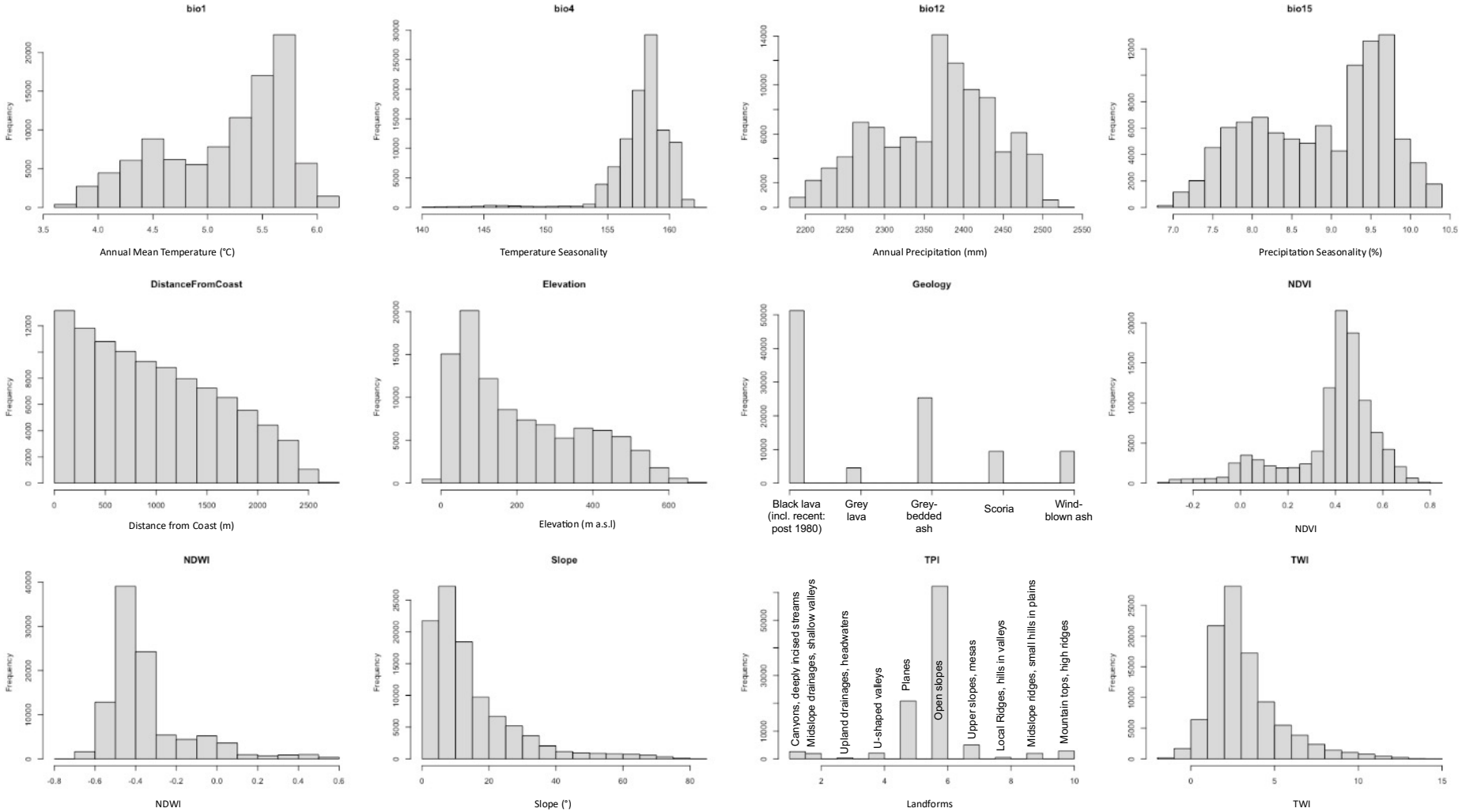


Figure 5-4: Range of values for each of Prince Edward Island's predictor variables. The bioclimatic variables' abbreviations are as follows: bio1 = Annual Mean Temperature , bio 4 = Temperature Seasonality , bio12 = Annual Precipitation, bio15 = Precipitation Seasonality.

5.2. Individual Model Performance

Six models were considered and tested against variables based on the following predictor variable scenarios (Table 4-6):

1. All variables;
2. Climate variables;
3. Topographic, geology and soils and satellite imagery variables; and
4. Wetland classification proxy variables.
    a. The Ramsar convention classification system
    b. The Hydrogeomorphic (HGM) classification system classification
    c. The International Union for Conservation of Nature (IUCN) Global Ecosystem Typology 2.0

In order to select the best model to be used to predict the distribution of mires on the PEIs, the models' performance was evaluated using the Areas under the Curve (AUC), True Skill Statistics (TSS) and correct classification rate (CCR). The models' performance in the AUC and TSS metrics are presented in Table 5-2.

Table 5-2: The Areas under the Curve (AUC) and True Skill Statistics (TSS) associated 10-fold cross validation. Model abbreviations are as follows: BRT = Boosted Regression Tree, CART = Classification and Regression Trees, GAM = Generalised Additive Models, GLM = Generalized Linear Model, MARS = Multivariate Adaptive Regression Splines and RF = Random Forest. Variable scenarios are summarised in Table 4-6.

Scenario	BRT		CART		GAM		GLM		MARS		RF	
	AUC	TSS	AUC	TSS	AUC	TSS	AUC	TSS	AUC	TSS	AUC	TSS
1	0.72	0.40	0.68	0.32	0.74	0.42	0.72	0.41	0.70	0.37	0.74	0.42
2	0.68	0.34	0.64	0.24	0.67	0.33	0.67	0.32	0.67	0.34	0.70	0.37
3	0.71	0.39	0.68	0.32	0.72	0.40	0.71	0.39	0.68	0.35	0.72	0.38
4 (a)	0.67	0.36	0.66	0.30	0.68	0.36	0.65	0.31	0.67	0.32	0.65	0.30
4 (b)	0.69	0.38	0.65	0.29	0.69	0.39	0.69	0.38	0.70	0.37	0.66	0.33
4 (c)	0.69	0.34	0.66	0.27	0.67	0.31	0.66	0.30	0.68	0.33	0.72	0.39

A good model should have an AUC value more than 0.9, although a score greater than 0.7 is adequate for species distribution modelling (Swets, 1988; González-Ferreras et al., 2016). As depicted in Figure 5-5 (pg. 48), the models in this study displayed consistent performance in the AUC metric. With values between 0.60 and 0.80, the models' performance can be said to indicate poor to fair model performance. The CART models performed slightly poorer compared to the other models in terms of the AUC, with all CART models failing to achieve AUC values sufficient for species distribution modelling. Models trained with Scenario 1 and 3 variables reported higher AUC values compared to the other variable scenarios, with the average AUC values greater than 0.70 (Figure 5-6, pg. 49).

Similar to AUC, TSS values suggest the model accuracies were fair to moderate with values ranging between 0.24 and 0.42 (Figure 5-5 [pg. 48], Figure 5-7 [pg. 49]). Once again, the CART model performed poorly for the TSS metric reporting the lowest TSS values of 0.24 under variable Scenario 2. Both CART and MARS models failed to achieve a mean value of 0.30 for all models. Overall, the best models scored AUC and TSS values greater than 0.70 and 0.40, respectively.

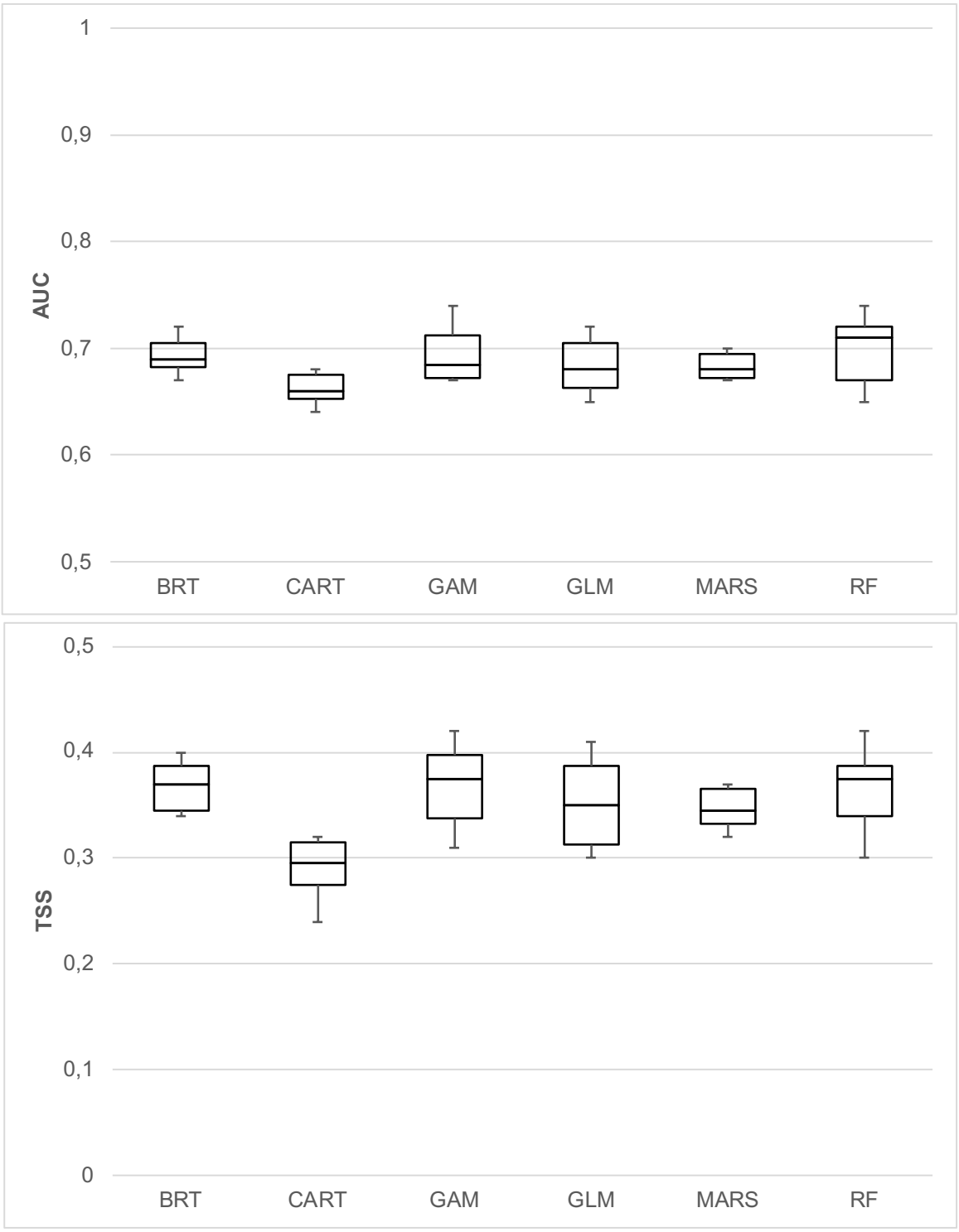


Figure 5-5: Boxplot depicting mean AUC and TSS for each SDM algorithm. Model abbreviations are as follows: BRT = Boosted Regression Tree, CART = Classification and Regression Trees, GAM = Generalised Additive Models, GLM = Generalized Linear Model, MARS = Multivariate Adaptive Regression Splines and RF = Random Forest.

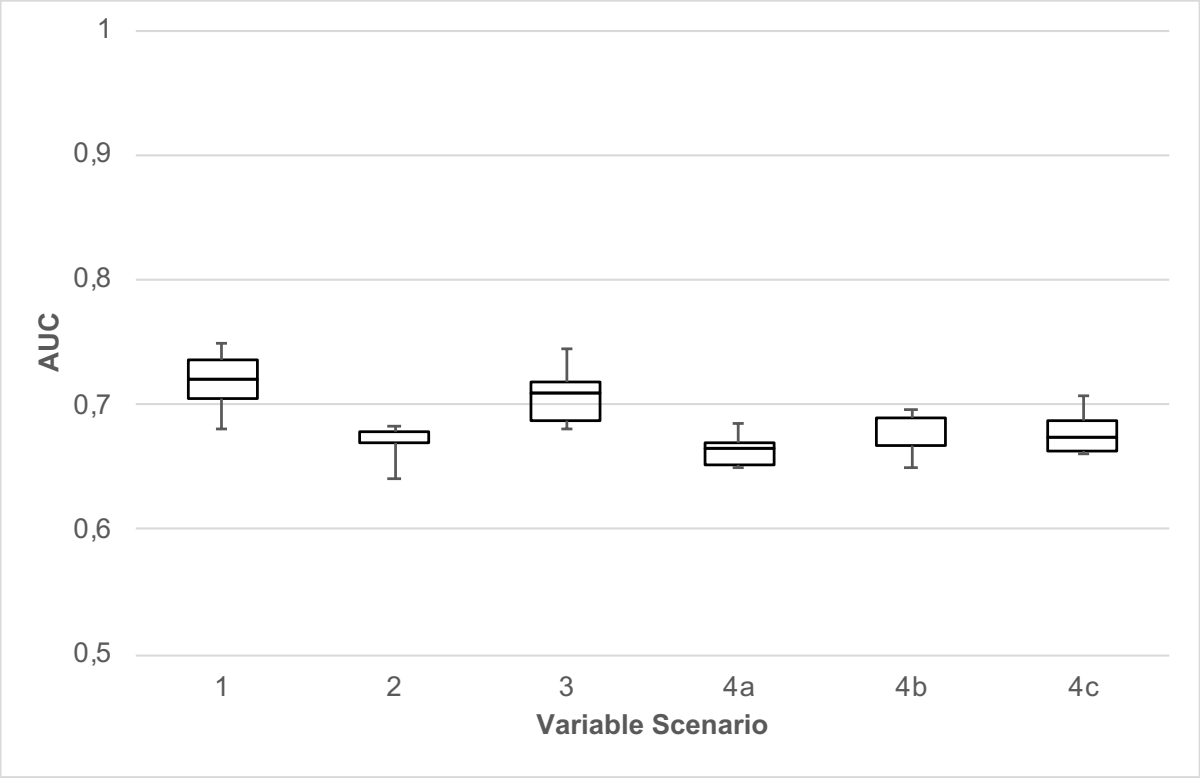


Figure 5-6: Boxplot depicting mean AUC of all models under each variable scenario. Variable scenarios are summarised in Table 4-6.

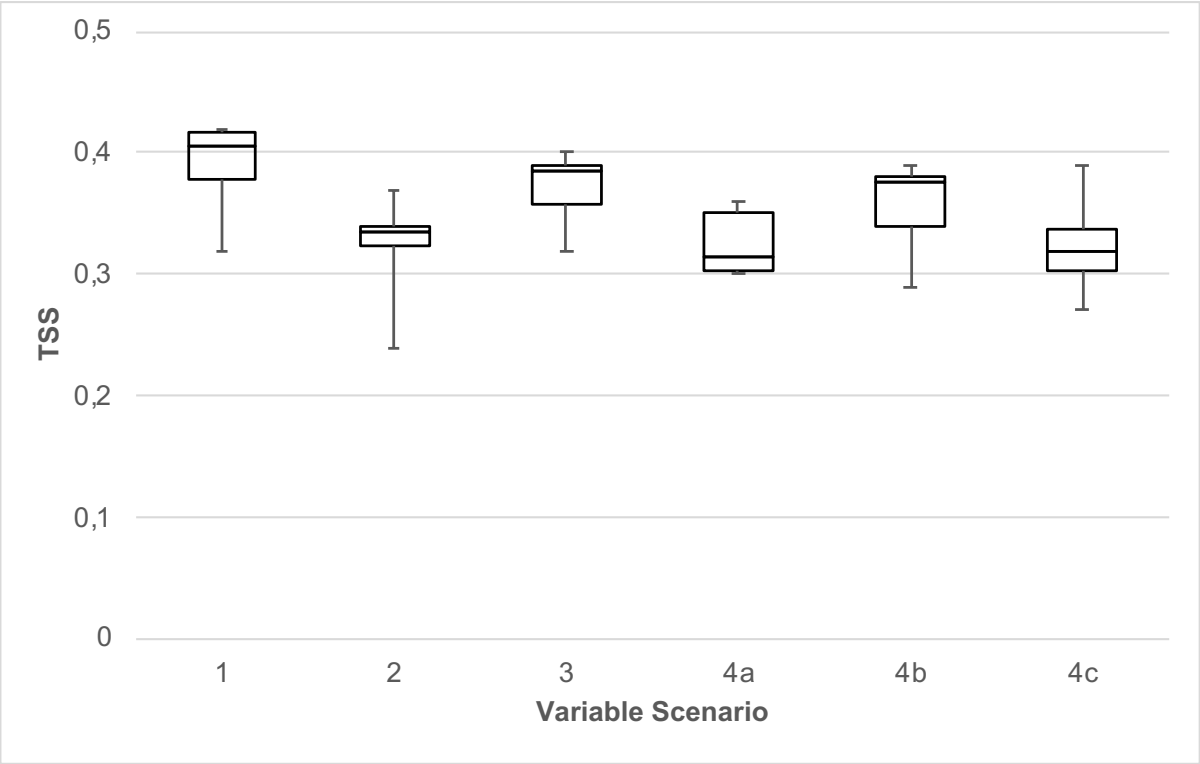


Figure 5-7: Boxplot depicting mean TSS of all models under each variable scenario. Variable scenarios are summarised in Table 4-6.

Although the AUC and TSS scores are not as high as would be desired (> 0.9 and closer to 1, respectively), they do provide insight into models' predictive ability. Based on their performance in the

AUC and TSS metrics, most models were ruled out as unsuitable for application in this study. These were models that scored AUC values less than 0.70 (Swets, 1988) or TSS values less than 0.20 (González-Ferreras et al., 2016). The individual model performance did not provide an obvious best model and combination of variables for the prediction of mire distribution, but rather provided five possible algorithms (BRT, GAM, GLM, MARS, RF) from two variable scenarios, Scenarios 1 and 3.

However, two models, a GAM and RF model based on Scenario 1 variables performed the same (AUC = 0.74, TSS = 0.42). As different models make different assumptions and extrapolate differently to new environments, their predictions, binary maps were constructed by applying a statistically determined threshold by selecting the value that maximized the sum of sensitivity and specificity (i.e., sensitivity + specificity/ 2), as based on Liu et al. (2005). The thresholds for each model in this study are presented in Table 5-3 and the predictions for each model are presented in Appendix 2 (pg. 76).

Table 5-3: Thresholds based on maximized sum of sensitivity and specificity (i.e., sensitivity + specificity/ 2) for models in the study. Model abbreviations are as follows: BRT = Boosted Regression Tree, CART = Classification and Regression Trees, GAM = Generalised Additive Models, GLM = Generalized Linear Model, MARS = Multivariate Adaptive Regression Splines and RF = Random Forest. Variable scenarios are summarised in Table 4-6.

Scenario	BRT	CART	GAM	GLM	MARS	RF
1	0.1706	0.2162	0.1790	0.1931	0.1986	0.4881
2	0.1874	0.2175	0.1748	0.2129	0.1882	0.4477
3	0.1768	0.2285	0.1842	0.2403	0.1925	0.4711
4 (a)	0.2080	0.2974	0.1917	0.1957	0.2044	0.2587
4 (b)	0.1665	0.2958	0.1890	0.1873	0.1750	0.2618
4 (c)	0.1867	0.2682	0.1588	0.1683	0.1684	0.4537

The binary maps for the best performing Scenario 1 models GAM and RF models are presented by Figure 5-8 (pg. 51). Additionally, an ensemble model was constructed for Scenario 1, with all the models deemed suitable for mire distribution modelling. That is, all models except for the CART model in the two scenarios. The correct classification rate (CCR) of each model was calculated by crossing all presences-absences with the binary predictions in order to determine the proportion of more presence-absence observations each model classified correctly. The confusion matrices for the GAM, RF and Ensemble for Scenario 1 variables are presented Appendix 2 (pg. 82).

The RF model had the best CCR with 99% compared to the 72% and 85% of the GAM and Ensemble, respectively. As evidenced by their individual confusion matrices in Appendix 2 (pg. 82), the GAM and Ensemble models overestimated the distribution of mires, misclassifying more mire absence locations as mire presences. Although the RF model achieved an almost perfect CCR, based on the description provided by Smith and Mucina (2006) as described in detail in Wetlands (pg. 6), the model underestimates the distribution of mires on the western side of the Island, where a large mire is said to exist (as depicted in Figure 2-1).

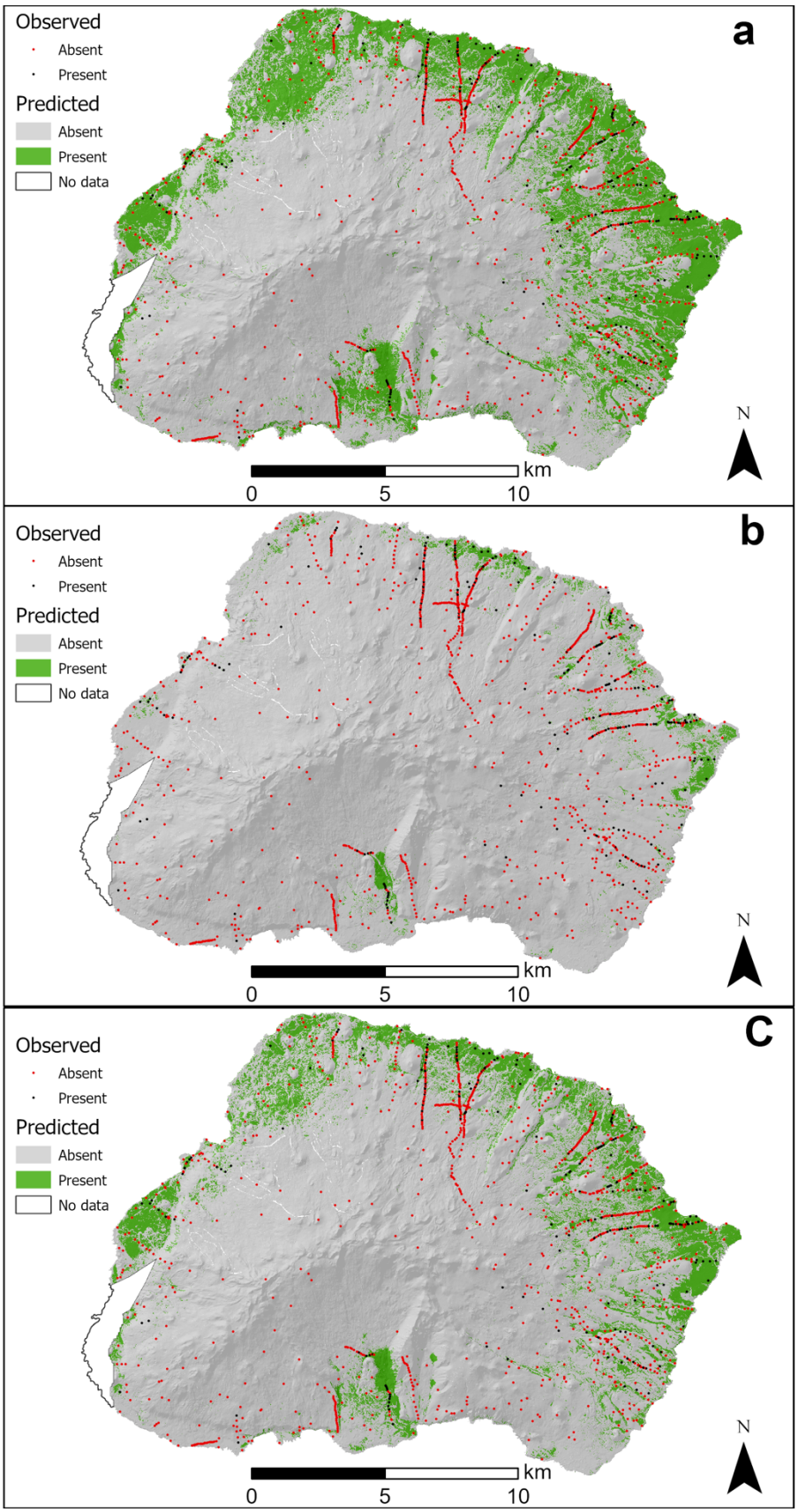


Figure 5-8: Possible mire distribution on Marion Island according to Scenario 1's suitable models: Boosted Regression Tree (a), Random Forest (b), Ensemble (c). Threshold based on the maximized the sum of sensitivity and specificity (i.e., sensitivity + specificity/ 2) Liu et al. (2005).

Consequently, the RF model trained with Scenario 1 variables was chosen as the model to predict the distribution of mires on the PEIs. However, even though soils are one of the variables in Scenario 1, a soils layer is only available for Marion Island. As a result, because the model trained on Marion Island was used for Prince Edward Island to predict the distribution of mires there, the model predictor variables should be the same for both Islands. Prior to eliminating the soils variable from the model, the contribution of each Scenario 1 variable in the RF model was evaluated. This allowed determining the effect the soils variable has on model performance, and, as a result, understanding how removing it may change the prediction of mire distribution.

According to Figure 5-9 below, soils and geology were the two least influential variables in both models. With a contribution of 2%, the soils variable contributed 30% less than the most important variable (surface wetness [NDWI]). As the variable did not offer much influence in the model, it was removed from the predictor variables used to train the RF model to predict the distribution of mires on the PEIs.

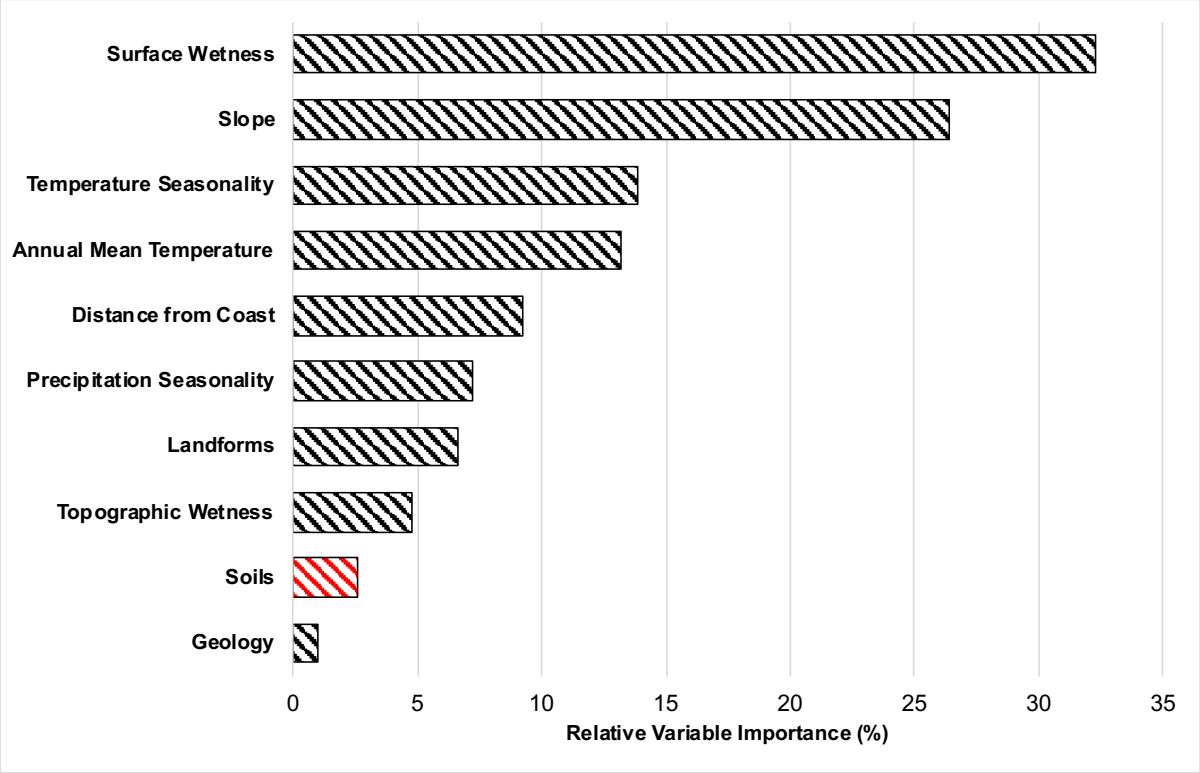


Figure 5-9: Relative importance (%) of Scenario 1 variables in Random Forest (RF) model. Higher relative variable importance values indicate that the variable is more significant in affecting model classification accuracy.

Therefore, the optimal model for predicting the distribution of mires on the PEIs was the RF model based on the variables presented in Table 5-4 (pg. 53).

Table 5-4: Predictor variables used to predict the distribution of mires on the Prince Edward Islands.

Variable Type	Predictor Variable
Topographic (DEM derived)	Slope
	Distance from coast
	Topographic wetness (TWI)
	Landforms (TPI)
Geology and Soils	Geology
Satellite imagery	Vegetation density (NDVI)
	Surface wetness (NDWI)
Bioclimatic	BIO01: Annual Mean Temperature
	BIO04: Temperature Seasonality
	BIO15: Precipitation Seasonality

5.3. Current Mire Distribution

According to the predictions presented in Figure 5-11 (pg. 55), mires currently occupy 8.7 km² (of ~290 km²) of Marion Island and 2.63 km² (of ~ 45 km²) of Prince Edward Island. Two of the three large mires identified by Smith and Mucina (2006, p. 716) were predicted by the model. These large mires were found on the eastern side of the Island on the coastal plain between Repetto's Hill and Long Ridge, as well as inland from East Cape, Macaroni Bay (Figure 2-1, pg. 7). The one on the western coastal plain between Kleinkoppie and Kampkoppie, was only partially predicted. Another noticeable mire prediction on Marion Island is on the southern part of the Island. According to predictions, Prince Edward Island's mires are prevalent in the north-western and south-eastern parts of the Island, where plains are the dominant landform (Figure 5-2 (e), pg. 44).

Surface wetness (26.4%) and slope (23.2%) were by far the most important predictor factors of mire distribution, as shown in Figure 5-10 (pg. 54). Following that were the two temperature variables, temperature seasonality (10.7%) and annual mean temperature (9.4%). The distance from coast (8.6%) was less essential, but still played a role in predicting mire occurrence. Precipitation seasonality (5.5%), landforms (5%), topographic wetness (4.3%), and geology (1.1%) all played a minor influence in mire prediction, each contributing less than 6%; however, geology was by far the least important variable in the model.

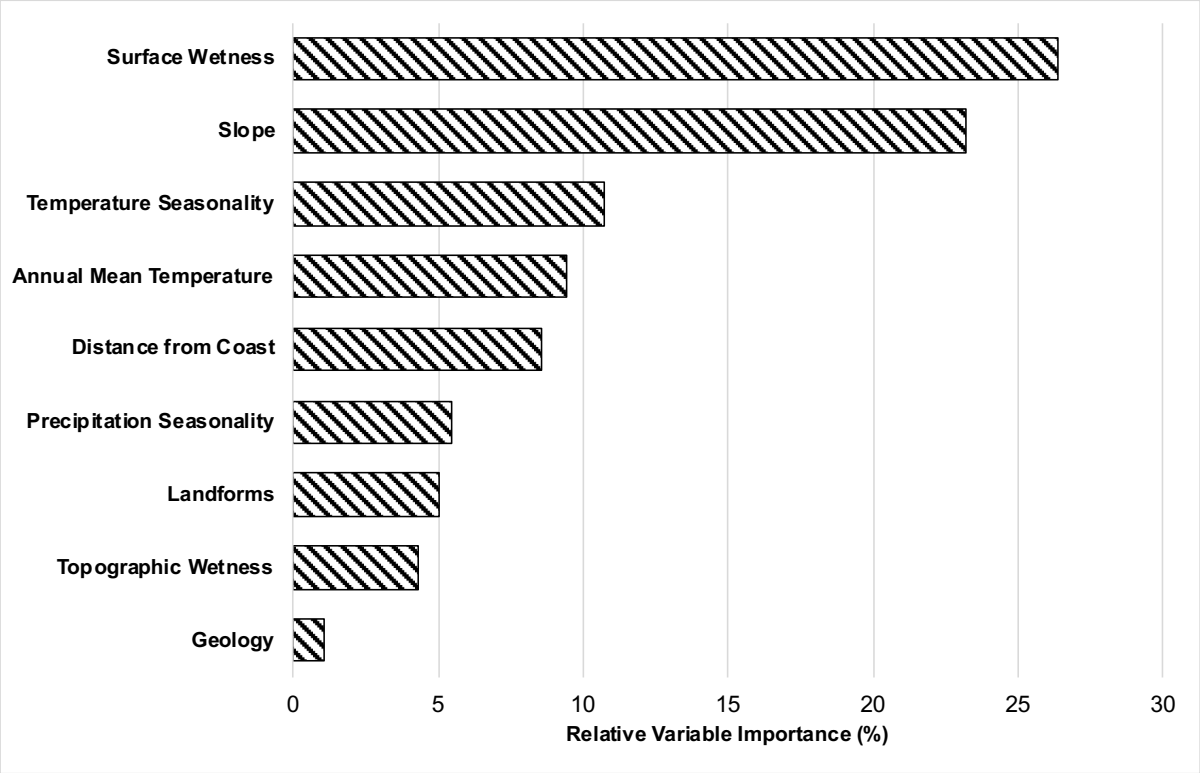


Figure 5-10: Relative importance (%) of variables selected for final Random Forest (RF) model. Higher relative variable importance values indicate that the variable is more significant in affecting model classification accuracy.

Response curves, which represent the association between a species' probability of occurring and each predictor variable, were used to examine the influence of each predictor variable on the probability of mire occurrence and are provided in Table 5-5 (pg. 56).

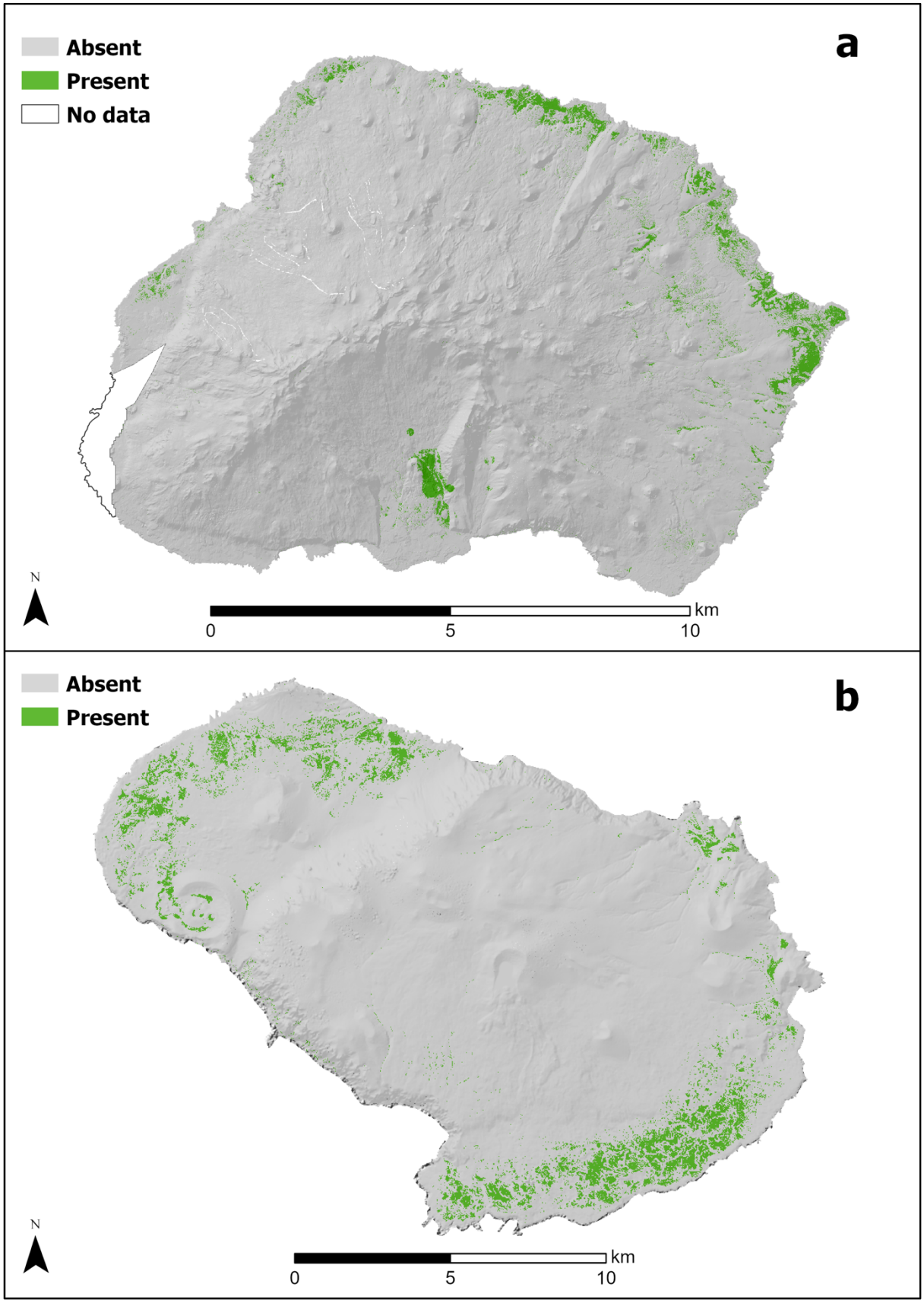
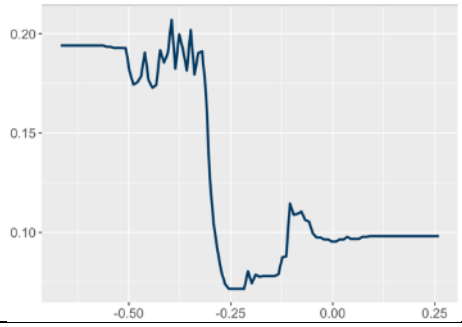
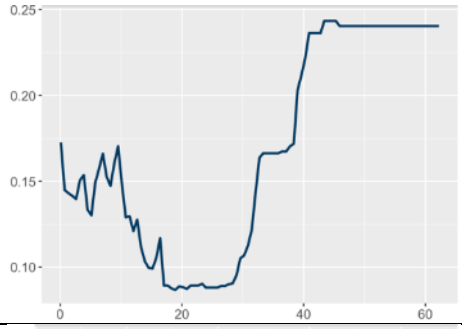
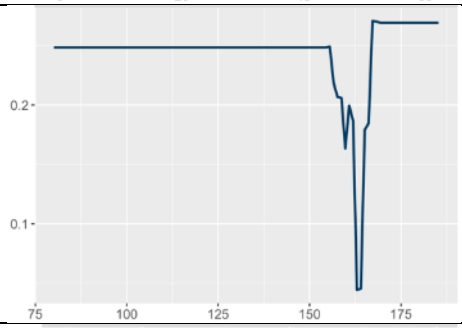
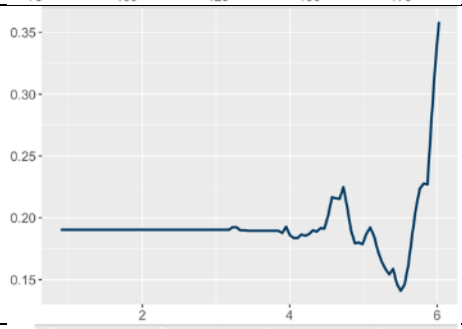
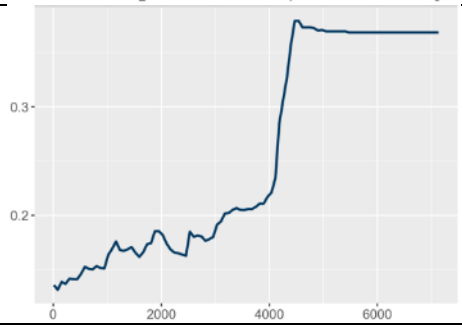
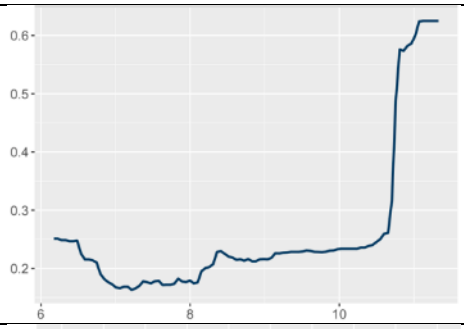
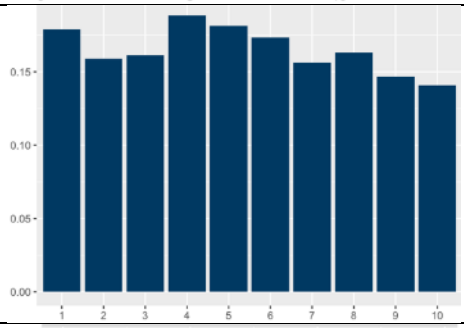
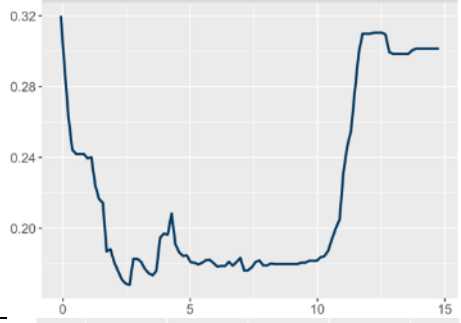
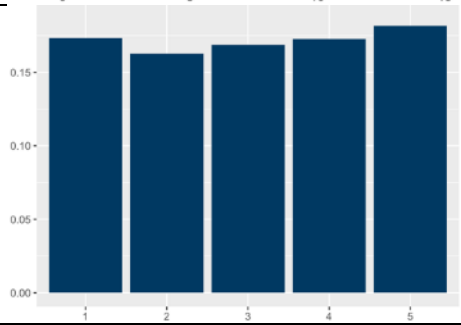


Figure 5-11: Predicted distribution of mires on the Prince Edward Islands (PEIs) based on Scenario 1 predictor variables. Marion Island (a), Prince Edward Island (b).

Table 5-5: Response curves for each variable, indicating the effect of a predictor variable on the probability of the response variable. Values closer to 1 on the y axis indicate high probability of occurrence at a range of predictor variable values on the x-axis. The variables are listed by order of importance.

Predictor Variable	Response Curve	Interpretation
Surface Wetness		The curve shows a steep decline with increasing surface wetness (NDWI) values. Higher NDWI values are associated with open water or surface water presence and mires are known to occur where there is high soil moisture. Therefore, as mires occur where values are between -0.75 and -0.25, with some preferring -0.1 to -0.25, this is an indication that mires do not occur in areas where the water is visible at the surface (open surface water)
Slope (º)		The curve shows an increase with increasing slope from about 30º and 62º, indicating mires on Marion Island likely prefer steeper slope, while some prefer slopes between 0º and 10º.
Temperature Seasonality		The WorldClim temperature seasonality values, as can be seen in Table 4-2 (pg. 35), are multiplied by 100 to preserve significant digits (O'Donnell and Ignizio, 2012). Therefore, mire occurrences are likely when the temperature seasonality is no more than 1.87 ºC, peaking between 1.60 ºC and 1.87 ºC.
Annual Mean Temperature (ºC)		Mires are more likely to occur where annual mean temperatures are around 6 ºC
Distance from Coast (m)		Mires are likely to occur on the lowland between the coast and 4.2 km, but mires can be found up to 7.2 km away from coast.

Precipitation Seasonality (%)		Mires prefer areas where precipitation seasonality is between 8 and 11.5%.
Landforms		Landforms are low on the variable importance list and have no discernible relationship with mire distribution.
Topographic Wetness		Mires occur when the Topographic Wetness Index (TWI) value is equal to zero, then the probability of their occurrence falls drastically until it rises between 10 and 15.
Geology		Geology is at the bottom of the variable importance list and has no discernible relationship with mire distribution.

The Normalised Difference Vegetation Index (NDVI) (Table A-1, pg. 75), which was utilized as a proxy for vegetation density, was strongly correlated with surface wetness. As a result, vegetation density must have a comparable impact on predicting mire distribution on the PEIs and may thus be used instead of the surface wetness proxy. As seen in Figure 5-12, when surface wetness is replaced with vegetation density, the latter has a significant impact on prediction; however, it is not the most important factor as surface wetness was, but rather the second most important after slope. The remaining variables are still ranked in the same order. Figure 5-13 (pg. 58) depicts the response curve for the vegetation density proxy, suggesting mires peak where NDVI values are around 0.375. A similar situation exists between annual mean temperature (BIO01) and annual precipitation (BIO12). When annual temperature is used over annual mean temperature, the former is the third most important variable in the model (Figure 5-15, pg. 59), with mires being slightly negatively impacted by annual temperature and mires preferring lower annual precipitation amounts (Figure 5-14, pg. 58).

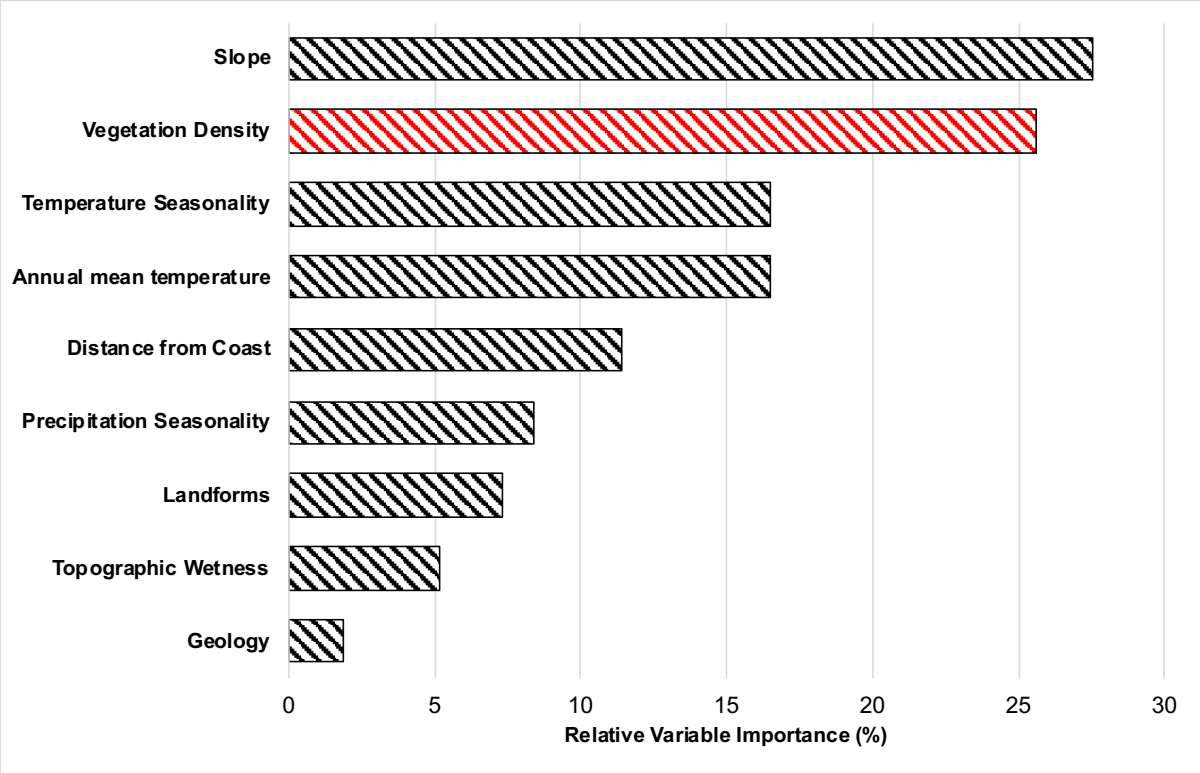


Figure 5-12: Relative importance (%) of variables selected for final Random Forest (RF) model with surface wetness (NDWI) replaced with vegetation density (NDVI). Higher relative variable importance values indicate that the variable is more important in affecting model classification accuracy.

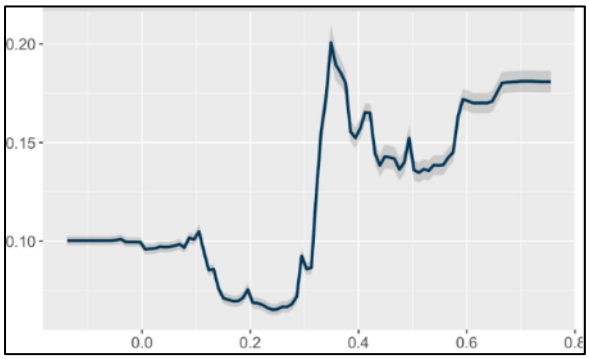


Figure 5-13: Response curve for vegetation density proxy (NDVI) when it is used over surface wetness proxy (NDWI) in mire distribution prediction.

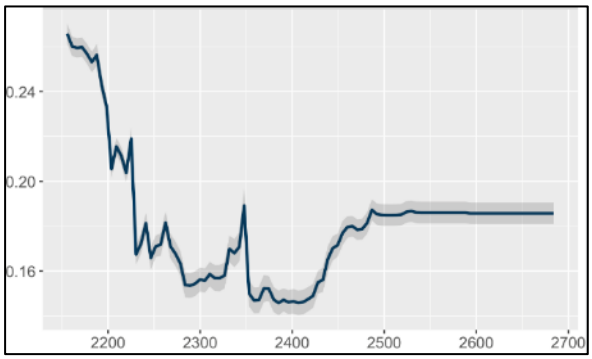


Figure 5-14: Response curve for annual precipitation (BIO12). when it is used over mean annual temperature (BIO01) in mire distribution prediction.

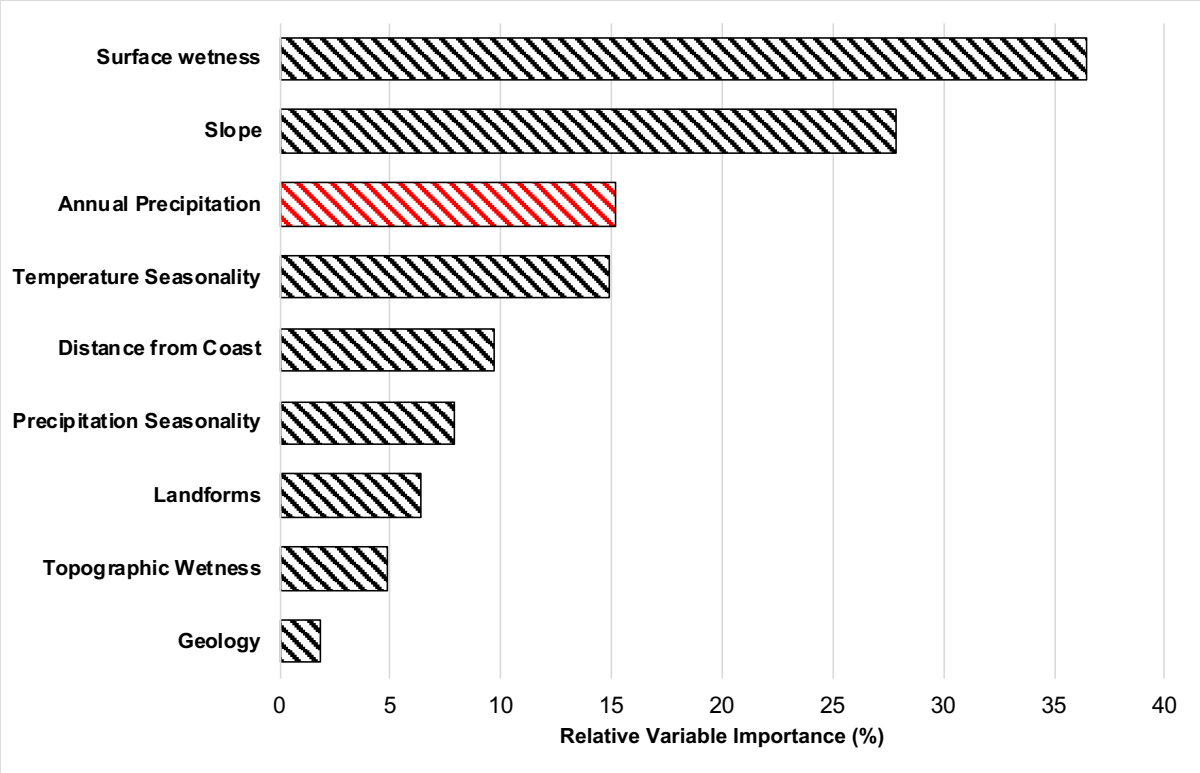


Figure 5-15: Relative importance (%) of variables selected for final Random Forest (RF) model with mean annual temperature (BIO01) replaced with annual precipitation (BIO12). Higher relative variable importance values indicate that the variable is more significant.

Chapter 6: Discussion

The ability of multiple regression-based and machine learning species distribution modelling algorithms (Table 4-5, pg. 38) to predict the distribution of mires on the Prince Edward Islands using several combinations of predictor variables (Scenarios) (Table 4-6, pg. 40) was assessed in this study.

The best performing model in this study performed only fairly (AUC = 0.74, TSS = 0.42) (González-Ferreras et al., 2016). Therefore, it is clear that the predictive power of the models in this study was poor and just only sufficient for species distribution modelling. The Random Forest model trained with a combination of digital elevation model (DEM) derived topographic, WorldClim bioclimatic variables, satellite imagery derived, and geology variables was selected as the optimal model to predict the distribution of mires on the PEIs.

6.1. Marion Island

The model's predictions suggested mires occupy 8.7 km² (of ~290 km²) of Marion Island (Figure 5-11, pg. 55). Smith and Mucina (2006, p. 716) provided a description of the distribution of mires on Marion Island (Figure 2-1, pg. 7). Based on this description, the model was able to predict two of the large mires known to exist on the eastern coast of Marion Island, on the coastal plain between Repetto's Hill and Long Ridge, inland of East Cape, Macaroni Bay. A third mire was said to exist on the western coastal plain between Kleinkoppie and Kampkoppie. The model, however, anticipated a minimal extent of mires in this area. The underestimation suggests that there may be some critical environmental factors that were not included owing to unavailability. It is also probable that some of the environmental factors differ between the eastern and western parts of the island due to minor variances in climate on both sides of the island, implying that the model did not possess the same predictive ability on both sides. Furthermore, the underestimation in that area is also possibly due to the uneven distribution of presence-absence observations. There were 1415 presence-absence observations (255 presences, 1415 absences) in total, which was sufficient to train the model. However, the majority of the observations were located on the island's eastern rather than western side. And given the presumed minor differences in the conditions under which mires occur as a result of differing climate conditions between the western and eastern sides of the Island, the model prediction on the former side of the island may have been less precise, as the conditions of the predictor variables on the latter side carried more weight (as there is more of them) than those on the former side, resulting in mire underestimation on the former (western) side of the island.

According to the model predictions, the distribution of the mires on Marion Islands is primarily influenced by surface wetness (derived using the Normalised Difference Water Index (NDWI)) and the slope, contributing 15% and 12% more than the next best variable, respectively (Figure 5-10, pg. 54). The derived model variable response curves (Table 5-5, pg. 56) suggest mires are likely to occur where NDWI values are between -0.75 and -0.25. This range of values represent the absence of open surface water and is typically associated with land cover classifications such as vegetation, bare soil or rock. This suggests that although mires are habitats that occur in wet areas (high soil moisture), the water is

rarely visible at the surface. As vegetation density (derived using the Normalised Difference Vegetation Index (NDVI)) was highly correlated to surface wetness, it can be used as a substitute in the model. With mires favouring areas with NDVI values around 0.375, mires are modelled to occur for moderate NDVI values (approximately 0.2 to 0.5) associated with sparse vegetation, such as shrubs and grasslands (USGS, 2021). As mires are characterised by a layer of peat at the surface, this range of values is plausible. In terms of the slope, the model suggests slopes between 30º and 62º, with some mires preferring gentler slopes between 0º and 10º. Accumulation (of water or peat or soil) is expected to occur on gentler slopes and the largest mires on Marion Island are known to occur on undulating landscapes associated with gentle slopes; yet, the model predicts the mires on Marion Island occur on steeper slopes. This may be due to the fact that mires on Marion Island often have exposed ridges and plateaus around their edges (Yeloff et al., 2007), and the Marion Island landscape is highly variable. As such, the scale at which the study was conducted (10 m) could possibly have generalised slope, ridges and plateaus, influencing the modelling of mire occurrence. In addition, this result may be as a result of inaccuracies in the digital elevation model (DEM) which may have occurred in the data collection or DEM generation process. As the explanation for the finding is unclear, additional investigation into the model's suggestion of mires on steep slopes is warranted.

Temperature variables (temperature seasonality (BIO04), and annual mean temperature (BIO01), in that order), were the third and fourth most important variables. Mires thrive in cool climatic conditions (Essl et al., 2012), hence the proposed temperature seasonality of no more than 1.87 ºC and annual mean temperatures of roughly 6 ºC give optimal conditions for peat formation, which is necessary for mires existence. The PEIs are considered 'thermally stable' (Smith, 2002), therefore, small temperature variations are expected. Because of its relationship to the annual mean temperature (BIO01), the annual precipitation (BIO12) variable was not included in the model. As a result, at annual precipitation amounts of approximately 2150 mm, a similar impact as supplied by annual mean temperature can be expected from annual precipitation.

With the presence of mires, the distance from the coast is also a significant element to consider. The model suggests that mires can be found everywhere on Marion Island, from the shore up to around 7 km inland. This, however, contradicts what is known about mires on the Island, which are generally lowland features located closer to the coast as depicted in Figure 3-7 (Smith et al., 2001; Smith and Mucina, 2006). As a result, the model's recommendation with regards to the distance from coast may be considered unexpected and requires further investigation. The other climate variable (precipitation seasonality; BIO15), was slightly less important compared to the top five variables, but still played a role in predicting mire occurrence. As with temperature seasonality, low variability in monthly precipitation total (8% to 11.5%) is ideal for mires on Marion Island. According to their response curves, the third and final least important variables (landforms, and geology), have no discernible relationship with mire distribution. These two elements are known to have an effect on the existence of mires, particularly when using Hydrogeomorphic (HGM) classification system, although the model revealed on Marion Island they had only a little effect. There is a close relation between topography and surface water because of its fluid characteristics. The shape and permeability of a landscape controls the movement of surface water (Wolock et al., 2004; MacMillan and Shary, 2009; Huang et al., 2018b). In

terms of geology on Marion Island, peats are substantially thicker (up to 3 m) on older, impermeable grey lava flows, whereas they are shallower (usually less than 1 m) on newer, more porous black lava flows and scoria deposits (Gremmen, 1981; Smith and Mucina, 2006). However, the response curve (Table 5-5, pg. 56) implies that, in addition to the geology variable being negligible in terms of contribution in the model, all five geology classes (Table 4-3, 35) have a comparable influence on the occurrence of mires. The geology of the PEIs was mapped by Verwoerd and Langenegger (1968), who employed interpolation methods to map the geology over the terrain and were recently updated by Rudolph et al. (2020). As a result of the dataset's high degree of generalisation, geology did not give any insight into the distribution of mires. As with the slope, landforms were generalised due to the working spatial resolution for this study (10 m), which could have resulted in the loss of information on which landforms are appropriate for mire occurrence. Thus, it is also suggested that all 10 classes of landforms (Table 2-3, pg. 14) offer comparable influence on the occurrence of mires. Topographic wetness (as measured by the Topographic Wetness Index (TWI)) was the second least important variable, providing only a minor contribution to the model, along with landforms and geology. According to the response curve, mires are more likely to exist in two places: areas prone to water accumulation (low slope angle high values between 10 and 15), and well-drained dry areas (steep slopes) linked with low TWI values (close to zero). As such, the TWI does not meaningfully contribute to the occurrence of mires on Marion Island. Since it is modelled using the DEM, which was resampled to a spatial resolution of 10 m, the same limitations as apply to the TPI (landform classification), apply here.

6.2. Prince Edward Island

Due to the high levels of protection on Prince Edward Island, there exists no data that can be used to train the model and consequently predict the distribution of mires. However, the geographical location, climate, terrain, geology, hydrology, and flora of the PEIs are recognised to be similar (Gremmen, 1971; Gremmen, 1981; Smith and Steenkamp, 1990; Smith, 2002; Pakhomov and Chown, 2003; Pendlebury and Barnes-Keoghan, 2007). Therefore, the variables influencing mire distribution on Marion Island are likely to be similar to those on Prince Edward Island. As a result, the model trained on Marion Island was used to model mire occurrence on Prince Edward Island. The model's predictions suggested mires occupy 2.63 km² (of ~ 45 km²) of Prince Edward Island (Figure 5-11, pg. 55). According to this model, mires are common in the north-western and south-eastern sections of Prince Edward Island, where plains are the major landform and slopes are gentler (Figure 5-2, pg. 44). Without mire presence-absence observations for the Island, there is no way of validating the results of the prediction. Therefore, without a field survey, the distribution of mires on Prince Edward Island will remain unverified.

6.3. Conclusion

Considering the variables available for use on the PEIs, the importance of each variable and the thresholds suggested for each variable, six of the nine variables, namely, surface wetness, slope, temperature seasonality, annual mean temperature, and distance from coast, can effectively be used to predict mire distribution on the PEIs a spatial resolution of 10 m (Table 6-1).

Table 6-1: Predictor variables and thresholds that can be used to predict the distribution of mires on the Prince Edward Islands (PEIs). The italicised variables can replace the variables they are correlated with (which are directly above them).

Variable Type	Predictor Variable	Threshold(s)
Topographic (DEM derived)	Slope	30º to 62º
	Distance from coast	0 to 7.2 km
Satellite imagery	Surface wetness (NDWI)	-0.75 to -0.25.
	Vegetation density (NDVI)	*~3*
Bioclimatic	BIO01: Annual Mean Temperature	6 ºC
	BIO12: Annual precipitation	*~2150 mm*
	BIO04: Temperature Seasonality	0 ºC to 1.87 ºC
	BIO15: Precipitation Seasonality	8% to 11.5%

7. : Conclusion and Recommendations

The aim of this study was to use species distribution modelling to predict and understand the distribution of mires on the Prince Edward Islands (PEIs). In order to do so, the ability of multiple regression-based and machine learning species distribution modelling algorithms (Table 4-5, pg. 38) to predict the distribution of mires on the Prince Edward Islands using several combinations of predictor variables (Scenarios) (Table 4-6, pg. 40) was assessed.

The Areas under the Curve (AUC) and True Skill Statistics (TSS) values for the models in this study suggest no more than fair model performance. The Random Forest (RF) model built using the default parameters in 'sdm' package in R (Naimi and Araújo, 2016) trained with a combination of digital elevation model (DEM) derived topographic, WorldClim bioclimatic variables, satellite imagery derived, and geology variables (Table 5-4, pg. 53) was able to predict the distribution of mires with a correct classification rate (CCR) of 99%. Overall, this model had a fair performance when predicting mires for Marion Island. The Model predicts mires for 8.7 km² (3%) of Marion Island and 2.63 km² (5.8%) of Prince Edward Island. Furthermore, the predicted mire distribution on Marion Island was realistic when compared to the description of mires provided by Smith and Mucina (2006, p. 716). The distribution estimates on Marion Island were, however, less than expected, with underestimation on the western side of Marion Island. The underestimation can be ascribed to a combination of unavailable or differing environmental factors across the landscape, persistent cloud cover (affecting satellite imagery derived indices), or a change of mire occurrence since the evaluation by Smith and Mucina (2006). As Prince Edward Island is similar to Marion Island in terms of its geographical location, climate, terrain, geology, hydrology, and flora, the model trained on Marion Island was used to predict on Prince Edward Island where no presence-absence data for mire occurrence exist. However, the accuracy of the prediction on Prince Edward Island is unknown as there are no data that can be used to validate the results.

The goal of this study was to ascertain what factors influence mire distribution on the PEIs so that future mire distribution can be predicted in light of climate change. With the present climatic trajectory on the PEIs, the Islands can be expected to become warmer and dryer, which will dry out mires and potentially reduce their distribution over the terrain. However, of the six potential factors influencing mire distribution on the PEIs (slope, distance from the coast, surface wetness, and climate variables (annual mean temperature, temperature, and precipitation seasonality)), the surface wetness proxy and slope were the main drivers of current mire distribution (Figure 5-15), while the three climatic variables were less important. Given that the models in this study only performed fairly and the influence of climate factors in the RF model, this model should be used with caution when predicting mire distribution on the PEIs under future climate scenarios. Regardless, projecting this model onto future climate conditions may provide some insight into the impact of climate change on future mire distribution.

There were four main limitations to this study. The explanatory variables for the SDMs were chosen based on their availability and prediction potential for mires on the PEIs; however, given the limited available data, the selected environmental variables may have excluded some important variables that could be useful for mire occurrence prediction. Furthermore, the climate conditions on Marion Island

are known to differ slightly, with the western side of the island experiencing more humidity, cloudiness, and precipitation than the eastern side (Rouault, 2005; le Roux, 2008), which may cause minor differences in environmental variables on both sides of the Island. This may have in turn resulted in the RF model employed in this study underestimating the distribution of mires on the western side of the Island. Therefore, the inclusion of more environmental variables may improve the predictive ability of the models.

The second limitation was encountered in the mire presence- absence observations. Although there were enough to train models in the study, their distribution differs on the eastern and western side of Marion Island (Figure 4-2), with the distribution on the latter side lesser than on the former side of the island. As one of the three large mires on Marion Island is known to be located on the former side, more mire presence points are expected in that region than what is suggested by the observations recorded during the survey of the area. Therefore, this was likely due to poor surveying on this side of the Island. As a result, the RF model employed in this study underestimated the distribution of mires on the western side of the Island. Therefore, improved surveying on this side of the Island may improve the predictive ability of the models, especially on the western side of the island.

The third limitation related to the resolution of the data used in the study. Geology and soils are known to be influencers of mire occurrence. However, the resolution of the data in this study was too coarse to offer any useful information about their impact on the distribution of mires on the PEIs. Additionally, the climate data downloaded from the WorldClim version 2.1 database has a coarse spatial resolution. Due to the limited number of meteorological stations on oceanic islands, these datasets exhibit substantial precision errors (Fick and Hijmans, 2017; Leihy et al., 2018). This is due to the fact that the data is created from models that smooth data from available meteorological stations, which are lacking for oceanic islands (Leihy et al., 2018). Therefore, the PEIs are no different, with their climate datasets interpolated from measurements acquired from only one meteorological station located on Marion Island. Furthermore, climate interpolation in topographically complex and steep areas is fraught with error (Fick and Hijmans, 2017). Considering the data was available at a spatial resolution of approximately 900 m at the equator and was interpolated from a single meteorological station for Islands that are considered topologically complex with steep areas, the climate data is not completely reliable (Leihy et al., 2018). However, according to Treasure et al. (2019), rainfall on Marion Island increased with height, particularly on the eastern side, which is consistent with WorldClim2 data. In addition, the availability of cloud free, high resolution satellite imagery was also an issue. Although Sentinel-2 imagery is available for the Islands, there is persistent cloud cover, which results in a reduced number of imagery available for use on the Islands. As a result, the imagery for the two Islands were from two different years (2017 and 2020). However, this limitation did not significantly affect the presence of mires, since a 3-year period is unlikely to reflect a drastic change in mire occurrence. Regardless, the presence of cloud cover required that multiple image tiles were used to create one cloud-free mosaic of Marion Island. Although the images used for the mosaic were collected a few days apart (5 days), their reflectance values were not the same, affecting resultant analyses. Furthermore, the mosaic had a section of the Island cut off because of cloud cover (refer to Figure 4-6, pg. 34). Due to this limitation, this small section of Marion Island has no predictions. Therefore, obtaining higher-resolution geology

and soil datasets, improved climate variable datasets for the Islands, and a high-resolution cloud-free satellite image from a similar time frame to base remote sensing indices on for inclusion in the models for future research can improve the predictive performance of the models.

Finally, the inability to conduct field surveying to ground truth the results of the study was an additional issue. Although the occurrence points used to train the SDMs in this study are a true reflection of mire occurrence on Marion Island, the accuracy of the prediction on Prince Edward Island could not be evaluated. Therefore, the research can be improved by verifying modelled mire locations on Prince Edward Island, and subsequently using the data to improve the models; and apply those results to other sub-Antarctic islands to determine if they can be used elsewhere in the Southern Ocean.

Regardless of the limitations presented here, this study found that mires can be predicted with fair accuracy using species distribution models (SDMs) in conjunction with remote sensing and Geographic Information Systems (GIS) derived products. Mires represent crucial ecosystems, and their loss, or threat by climate change and other factors, will have an effect on the environment they are found in. As such, being able to model their distribution with any measure of accuracy for areas where data are lacking allows for better management and monitoring of these sensitive ecosystems.

References

Allouche, O., Tsoar, A. & Kadmon, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of applied ecology,* 43(6), pp.1223-1232. doi:10.1111/j.1365-2664.2006.01214.x.

Alsdorf, D. E., Rodríguez, E. & Lettenmaier, D. P. 2007. Measuring surface water from space. *Reviews of Geophysics,* 45(2). doi:10.1029/2006rg000197.

Amler, E., Schmidt, M. & Menz, G. 2015. Definitions and Mapping of East African Wetlands: A Review. *Remote Sensing,* 7(5), pp.5256-5282. doi:10.3390/rs70505256.

Araujo, M. B. & New, M. 2007. Ensemble forecasting of species distributions. *Trends in ecology & evolution,* 22(1), pp.42-47. doi:10.1016/j.tree.2006.09.010.

Austin, M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological modelling,* 200(1-2), pp.1-19. doi:10.1016/j.ecolmodel.2006.07.005.

Baker, C., Lawrence, R. L., Montagne, C. & Patten, D. 2007. Change detection of wetland ecosystems using Landsat imagery and change vector analysis. *Wetlands,* 27(3), pp.610-619.

Ballerine, C. 2017. Topographic Wetness Index Urban Flooding Awareness Act Action Support. Will & DuPage Counties, Illinois: University of Illinois.

Bassi, N., Kumar, M. D., Sharma, A. & Pardha-Saradhi, P. 2014. Status of wetlands in India: A review of extent, ecosystem benefits, threats and management strategies. *Journal of Hydrology: Regional Studies,* 2, pp.1-19. doi:10.1016/j.ejrh.2014.07.001.

Bates, B., Kundzewicz, Z. & Wu, S. (eds) 2008. *Climate Change and Water. Technical Paper of the Intergovernmental Panel on Climate Change,* Geneva: IPCC Secretariat.

Becker, D., De Andrés-Herrero, M., Willmes, C. & Weniger, G.-C. 2014. GIS-Based Automated Landform Classification for Analysis of Achieological sites. University of Cologne.

Bergstrom, D. M. & Chown, S. L. 1999. Life at the front: History, ecology and change on southern ocean islands. *Trends in Ecology and Evolution,* 14(12), pp.472-477. doi:10.1016/S0169-5347(99)01688-2.

Beven, K. J. & Kirkby, M. J. 1979. A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Journal,* 24(1), pp.43-69. doi:10.1080/02626667909491834.

Birsrat, E. & Berhanu, B. 2018. Identification of surface water storing sites using Topographic Wetness Index (TWI) and Normalized Difference Vegetation Index (NDVI). *Journal of Natural Resources and Development,* 8, pp.91-100. doi:10.5027/jnrd.v8i0.09.

Boelhouwers, J. C., Meiklejohn, K. I. & Hedding, D. W. 2008. Geology, geomorphology and climate change. In: Chown, S. & Froneman, P. W. (eds), The Prince Edward Islands: Land-sea interactions in a chaniging ecosystem, 1st ed. Stellenbosch: African Sun Media, pp.65-96.

Bourgeau-Chavez, L., Endres, S. L., Graham, J., Hribljan, J. A., Chimner, R., Lillieskov, E. & Battaglia, M. 2018. Mapping peatlands in boreal and tropical ecoregions. In: Liang, S. (eds), Comprehensive Remote Sensing. United Kingdom: Elsevier, pp.24-44.

Breiman, L. 2001. Random forests. *Machine learning,* 45(1), pp.5-32.

Breiman, L., Friedman, J., Olshen, R. & Stone, C. 1984. *Classification And Regression Trees.*

Brown, J. L. 2014. SDM toolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods in Ecology and Evolution,* 5(7), pp.694-700. doi:doi.org/10.1111/2041-210X.12200.

Budd, G. Changes in Heard Island glaciers, king penguins and fur seals since 1947. Papers and Proceedings of the Royal Society of Tasmania,

Chown, S., Davies, S. & Joubert, L. 2006. Draft Prince Edward Islands Environmental Management Plan. Version 0.1. Stellenbosch, South Africa: DST-NRF Centre of Excellence for Invasion Biology, University of Stellenbosch.

Chown, S. & Froneman, P. W. 2008. The Prince Edward Islands in a global context. In: Chown, S. & Froneman, P. W. (eds), The Prince Edward Islands: Land-sea interactions in a chaniging ecosystem. Stellenbosch: African Sun Media, pp.1-16.

Chown, S. & Smith, V. 1993. Climate change and the short-term impact of feral house mice at the sub-Antarctic Prince Edward Islands. *Oecologia,* 96(4), pp.508-516.

Chown, S. L., Marshall, D. J. & Pakhomov, E. A. 2008. Regional membership: biogeography. In: Chown, S. & Froneman, P. W. (eds), The Prince Edward Islands: Land-sea interactions in a chaniging ecosystem. Stellenbosch: African Sun Media, pp.277-300.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement,* 20(1), pp.37-46. doi:10.1177/001316446002000104.

Cooper, J. 2008. Human history. In: Chown, S. L. & Froneman, P. W. (eds), The Prince Edward Islands: Land-sea interactions in a chaniging ecosystem. Stellenbosch: African Sun Media, pp.331-350.
Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B. & Lambin, E. 2004. Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing,* 25(9), pp.1565-1596. doi:10.1080/0143116031000101675.
Crafford, J., Scholtz, C. & Chown, S. 1986. The insects of sub-Antarctic Marion and Prince Edward Islands; with a bibliography of entomology of the Kerguelen Biogeographical Province. *South African Journal of Antarctic Research,* 16(2), pp.42-84.
Dartnall, H. J. & Smith, V. R. 2012. Freshwater invertebrates of sub-Antarctic Marion Island. *African Zoology,* 47(2), pp.203-215. doi:10.1080/15627020.2012.11407548.
De Marco, P. & Nóbrega, C. C. 2018. Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation. *PloS one,* 13(9). doi:10.1371/journal.pone.0202403.
Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B. & Leitão, P. J. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography,* 36(1), pp.27-46. doi:10.1111/j.1600-0587.2012.07348.x.
Driessen, P., Deckers, J., Spaargaren, O. & Nachtergaele, F. 2001. *Lecture notes on the major soils of the world.* Rome: Food and Agriculture Organization.
Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W. & Li, X. 2016. Water bodies' mapping from Sentinel-2 imagery with Modified Normalized Difference Water Index at 10-m spatial resolution produced by sharpening the swir band. *Remote Sensing,* 8(4). doi:10.3390/rs8040354.
Elith, J. & Graham, C. H. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography,* 32(1), pp.66-77. doi:10.1111/j.1600-0587.2008.05505.x.
Elith, J. & Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual review of ecology, evolution, and systematics,* 40, pp.677-697. doi:10.1146/annurev.ecolsys.110308.120159.
Elith, J., Leathwick, J. R. & Hastie, T. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology,* 77(4), pp.802-813. doi:10.1111/j.1365-2656.2008.01390.x.
Erwin, K. L. 2009. Wetlands and global climate change: the role of wetland restoration in a changing world. *Wetlands Ecology and management,* 17(1), pp.71-84. doi:10.1007/s11273-008-9119-1.
Essl, F., Dullinger, S., Moser, D., Rabitsch, W. & Kleinbauer, I. 2012. Vulnerability of mires under climate change: implications for nature conservation and climate change adaptation. *Biodiversity and Conservation,* 21, pp.655-669. doi:10.1007/s10531-011-0206-x.
Feyisa, G. L., Meilby, H., Fensholt, R. & Proud, S. R. 2014. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment,* 140, pp.23-35. doi:10.1016/j.rse.2013.08.029.
Fick, S. E. & Hijmans, R. J. 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology,* 37(12), pp.4302-4315. doi:10.1002/joc.5086.
Fielding, A. H. & Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental conservation,* 24(1), pp.38-49.
Finlayson, C. 2018. Ramsar convention typology of wetlands. In: Finlayson, C., Everard, M., Irvine, K., McInnes, R. J., Middleton, B. A., van Dam, A. A. & Davidson, N. C. (eds), The wetland book I: Structure and function, management and methods. Springer, pp.1529-1532.
Finlayson, C., Capon, S., Rissik, D., Pittock, J., Fisk, G., Davidson, N., Bodmin, K., Papas, P., Robertson, H. & Schallenberg, M. 2017. Policy considerations for managing wetlands under a changing climate. *Marine and Freshwater Research,* 68(10), pp.1803-1815.
Food and Agriculture Organization of the United Nations 2020. Peatland mapping and monitoring-Recommendations and technical overview. Rome. doi:10.4060/ca8200en.
Franklin, J. 2009. *Mapping species distributions: spatial inference and prediction.* Cambridge University Press.
Frenot, Y., Gloaguen, J. C. & Tréhen, P. 1997. Climate change in Kerguelen Islands and colonization of recently deglaciated areas by Poa kerguelensis and P. annua. In: Battaglia, B., Valencia, J. & Walton, D. (eds), Antarctic communities: species, structure and survival. pp.358-366.
Friedman, J. H. 1991. Multivariate adaptive regression splines. *The Annals of Statistics,* 19(1), pp.1-67.
Früh, L., Kampen, H., Kerkow, A., Schaub, G. A., Walther, D. & Wieland, R. 2018. Modelling the potential distribution of an invasive mosquito species: comparative evaluation of four machine learning methods and their combinations. *Ecological Modelling,* 388, pp.136-144. doi:10.1016/j.ecolmodel.2018.08.011.
Gallant, A. L. 2015. The challenges of remote monitoring of wetlands. *Remote Sensing,* 7(8), pp.10938-10950. doi:10.3390/rs70810938

Gautam, V. K., Gaurav, P. K., Murugan, P. & Annadurai, M. 2015. Assessment of Surface Water Dynamicsin Bangalore Using WRI, NDWI, MNDWI, Supervised Classification and K-T Transformation. *Aquatic Procedia,* 4, pp.739-746. doi:10.1016/j.aqpro.2015.02.095.

Giles, P. T. & Franklin, S. E. 1998. An automated approach to the classification of the slope units using digital data. *Geomorphology,* 21(3-4), pp.251-264. doi:10.1016/S0169-555X(97)00064-0.

Gomes, V. H. F., Ijff, S. D., Raes, N., Amaral, I. L., Salomao, R. P., De Souza Coelho, L., De Almeida Matos, F. D., Castilho, C. V., De Andrade Lima Filho, D., Lopez, D. C. & Guevara, J. E. 2018. Species Distribution Modelling: Contrasting presence-only models with plot abundance data. *Scientific reports,* 8(1), pp.1003. doi:10.1038/s41598-017-18927-1.

González-Ferreras, A., Barquín, J. & Peñas, F. 2016. Integration of habitat models to predict fish distributions in several watersheds of N orthern S pain. *Journal of Applied Ichthyology,* 32(1), pp.204-216. doi:10.1111/jai.13024.

Gremmen, N. J. M. 1971. The distribution of alien vascular plants on Marion and Prince Edward Islands. *South African Journal of Antarctic Research,* 4(5).

Gremmen, N. J. M. 1981. *The vegetation of the subantarctic islands, Marion and Prince Edward.* Doctrate, Radboud University Nijmegen,

Gremmen, N. J. M. & Smith, V. R. 2008. Terrestrial vegetation and dynamics. In: Chown, S. L. & Froneman, P. W. (eds), The Prince Edward Islands: Land-sea interactions in a chaniging ecosystem. Stellenbosch: African Sun Media, pp.215-244.

Greve, M., Von Der Meden, C. E. O. & Janion-Scheepers, C. 2020. Biological invasions in South Africa's offshore sub-Antarctic territories. In: Van Wilgen, B. W., Measey, J., Richardson, D. M., Wilson, J. R. & Zengeya, T. A. (eds), Biological Invasions in South Africa. Springer Nature, pp.207-227.

Grimm, N. B., Chapin Iii, F. S., Bierwagen, B., Gonzalez, P., Groffman, P. M., Luo, Y., Melton, F., Nadelhoffer, K., Pairis, A. & Raymond, P. A. 2013. The impacts of climate change on ecosystem structure and function. *Frontiers in Ecology and the Environment,* 11(9), pp.474-482. doi:10.1890/120282.

Grundling, P., Grundling, A. & Pretorius, L. 2017. *South African peatlands: ecohydrological characteristics and socio-economic value: report to the Water Research Commission.* Water Research Commission.

Guisan, A., Edwards Jr, T. C. & Hastie, T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological modelling,* 157(2-3), pp.89-100. doi:10.1016/S0304-3800(02)00204-1.

Guisan, A. & Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. *Ecological modelling,* 135(2-3), pp.147-186. doi:10.1016/S0304-3800(00)00354-9.

Hallstan, S. 2011. *Species distribution models.*

Harenda, K. M., Lamentowicz, M., Samson, M. & Chojnicki, B. H. 2018. The role of peatlands and their carbon storage function in the context of climate change. In: Zielinski, T., Sagan, I. & Surosz, W. (eds), Interdisciplinary approaches for sustainable development goals. Springer, pp.169-187.

Hariri, S., Gustedt, J., Weill, S. & Charpentier, I. 2021. A Hybrid Breaching-Filling method for sink removal adapted to parallel hydrological simulations. *EGU General Assembly 2021*. doi:10.5194/egusphere-egu21-7849.

Hedding, D. W. 2006. *Geomorphology and geomorphological responses to climate change in the interior of sub-Antarctic Marion Island.* Doctoral dissertation, University of Pretoria,

Hedding, D. W. 2020. Comment on "Clinopyroxene megacrysts from Marion Island, Antarctic Ocean: evidence for a late stage shallow origin"(2019) by RJ Roberts, KD Lehong, AEJ Botha, G. Costin, FC De Beer, WJ Hoffman and CJ Hetherington. *Mineralogy and Petrology,* 114(4), pp.357-359. doi:10.1007/s00710-020-00713-z.

Hedding, D. W. & Greve, M. 2018. Letters. *Weather,* 73(6), pp.203-203. doi:10.1002/wea.3245.

Heymann, G., Erasmus, T., Huntley, B. J. & Liebenberg, C. A. 1987. An environmental impact assessment of a proposed emergency landing facility on Marion Island - 1987. Pretoria: National Scientific Programmes Unit: CSIR.

Hiestermann, J. & Rivers-Moore, N. 2015. Predictive modelling of wetland occurrence in KwaZulu-Natal, South Africa. *South African Journal of Science,* 111(7-8), pp.1-10. doi:10.17159/SAJS.2015/20140179

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology: A Journal of the Royal Meteorological Society,* 25(15), pp.1965-1978. doi:10.1002/joc.1276.

Hijmans, R. J. & Elith, J. 2013. Species distribution modeling with R. *R CRAN Project*.

Hjerdt, K., Mcdonnell, J., Seibert, J. & Rodhe, A. 2004. A new topographic index to quantify downslope controls on local drainage. *Water Resources Research,* 40(5). doi:10.1029/2004WR003130.

Hu, S., Niu, Z., Chen, Y., Li, L. & Zhang, H. 2017. Global wetlands: Potential distribution, wetland loss, and status. *Science of the Total Environment,* 586, pp.319-327. doi:10.1016/j.scitotenv.2017.02.001.

Huang, C., Chen, Y., Zhang, S. & Wu, J. 2018a. Detecting, extracting, and monitoring surface water from space using optical sensors: A review. *Reviews of Geophysics,* 56(2), pp.333-360. doi:10.1029/2018RG000598.

Huang, C., Nguyen, B. D., Zhang, S., Cao, S. & Wagner, W. 2017. A comparison of terrain indices toward their ability in assisting surface water mapping from Sentinel-1 data. *ISPRS International Journal of Geo-Information,* 6(5). doi:10.3390/ijgi6050140

Huang, W., Devries, B., Huang, C., Lang, M., Jones, J., Creed, I. & Carroll, M. 2018b. Automated extraction of surface water extent from Sentinel-1 Data. *Remote Sensing,* 10(5). doi:10.3390/rs10050797.

Hunter, E. A., Raney, P. A., Gibbs, J. P. & Leopold, D. J. 2012. Improving wetland mitigation site identification through community distribution modeling and a patch-based ranking scheme. *Wetlands,* 32(5), pp.841-850.

Intergovernmental Panel on Climate Change 2021. Climate Change 2021: The physical science basis. contribution of working group I to the sixth assessment report of the Intergovernmental Panel on Climate Change.

Jenness, J. & Enterprises, J. 2006. Topographic Position Index (TPI) v. 1.2.

Ji, L., Zhang, L. & Wylie, B. 2009. Analysis of dynamic thresholds for the Normalized Difference Water Index. *Photogrammetric Engineering & Remote Sensing,* 75(11), pp.1307-1317. doi:10.14358/PERS.75.11.1307.

Jiang, H., Feng, M., Zhu, Y., Lu, N., Huang, J. & Xiao, T. 2014. An automated method for extracting rivers and lakes from Landsat imagery. *Remote Sensing,* 6(6), pp.5067-5089. doi:10.3390/rs6065067.

Jiménez-Valverde, A. 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography,* 21(4), pp.498-507. doi:10.1111/j.1466-8238.2011.00683.x.

Jiménez-Valverde, A., Lobo, J. M. & Hortal, J. 2008. Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and distributions,* 14(6), pp.885-890. doi:10.1111/j.1472-4642.2008.00496.x.

Joosten, H. 2012. Status and prospects of global peatlands. *Natur und Landschaft,* 87(2), pp.50.

Joosten, H. & Clarke, D. 2002. *Wise use of mires and peatlands.* International Mire Conservation Group and International Peat Society.

Kaky, E., Nolan, V., Alatawi, A. & Gilbert, F. 2020. A comparison between Ensemble and MaxEnt species distribution modelling approaches for conservation: A case study with Egyptian medicinal plants. *Ecological Informatics,* 60. doi:10.1016/j.ecoinf.2020.101150.

Kameyama, S., Yamagata, Y., Nakamura, F. & Kaneko, M. 2001. Development of WTI and turbidity estimation model using SMA—application to Kushiro Mire, eastern Hokkaido, Japan. *Remote Sensing of Environment,* 77(1), pp.1-9. doi:10.1016/S0034-4257(01)00189-4.

Kaplan, G. & Avdan, U. 2017. Mapping and monitoring wetlands using Sentinel-2 satellite imagery. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences,* 4, pp.271-277. doi:10.5194/isprs-annals-IV-4-W4-271-2017.

Keith, D. A., Essl, F., Young, K. & Körner, C. 2020. Polar tundra and deserts. In: Keith, D. A., Ferrer-Paris, J. R., Nicholson, E. & Kingsford, R. T. (eds), The IUCN Global Ecosystem Typology 2.0: Descriptive profiles for biomes and ecosystem functional groups. Gland, Switzerland: IUCN,

Kirkpatrick, J. & Scott, J. 2002. Change in undisturbed vegetation on the coastal slopes of subantarctic Macquarie Island, 1980–1995. *Arctic, Antarctic, and Alpine Research,* 34(3), pp.300-307. doi:10.1080/15230430.2002.12003498.

Klemas, V. & Pieterse, A. 2015. Using remote sensing to map and monitor water resources in arid and semiarid regions. In: Younos, T. & Parece, T. E. (eds), Advances in watershed science and assessment, The Handbook of Environmental Chemistry Cham: Springer, pp.33-60.

Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai,* 14(2), pp.1137-1145.

Landis, J. R. & Koch, G. G. 1977. The measurement of observer agreement for categorical data. *Biometrics,* 33(1), pp.159-174. doi:10.2307/2529310.

Le Roux, P. C. 2008. Climate and climate change. In: Chown, S. & Froneman, P. W. (eds), The Prince Edward Islands: Land-sea interactions in a chaniging ecosystem. Stellenbosch: African Sun Media, pp.39-64.

Le Roux, P. C. & Mcgeoch, M. A. 2008. Changes in climate extremes, variability and signature on sub-Antarctic Marion Island. *Climatic Change,* 86(3-4), pp.309-329. doi:10.1007/s10584-007-9259-y.

Leihy, R. I., Duffy, G. A., Nortje, E. & Chown, S. L. 2018. High resolution temperature data for ecological research and management on the Southern Ocean Islands. *Scientific Data,* 5(1). doi:10.1038/sdata.2018.177.

Leroy, B., Delsol, R., Hugueny, B., Meynard, C. N., Barhoumi, C., Barbet-Massin, M. & Bellard, C. 2018. Without quality presence–absence data, discrimination metrics such as TSS can be misleading measures of model performance. *Journal of Biogeography,* 45(9), pp.1994-2002. doi:10.1111/jbi.13402.

Li, S., Macmillan, R., Lobb, D. A., Mcconkey, B. G., Moulin, A. & Fraser, W. R. 2011. Lidar DEM error analyses and topographic depression identification in a hummocky landscape in the prairie region of Canada. *Geomorphology,* 129(3-4), pp.263-275. doi:10.1016/j.geomorph.2011.02.020.

Li, X. & Wang, Y. 2013. Applying various algorithms for species distribution modelling. *Integrative Zoology,* 8(2), pp.124-135. doi:10.1111/1749-4877.12000.

Lindsay, J. B. 2016. Efficient hybrid breaching-filling sink removal methods for flow path enforcement in digital elevation models. *Hydrological Processes,* 30(6), pp.846-857. doi:10.1002/hyp.10648.

Liu, C., Berry, P. M., Dawson, T. P. & Pearson, R. G. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography,* 28(3), pp.385-393.

Lobo, J. M., Jiménez-Valverde, A. & Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography,* 17(2), pp.145-151. doi:10.1111/j.1466-8238.2007.00358.x.

Lubbe, N. R. 2010. *Soil characteristics and pedogenesis on sub-Antarctic Marion Island.* Doctoral dissertation, University of Pretoria,

Macmillan, R. & Shary, P. 2009. Landforms and landform elements in geomorphometry. In: Hengl, T. & Reuter, H. I. (eds), Geomorphometry: Concepts, Software, Applications, Developments in Soil Science. Newnes, pp.227-254.

Mainali, K. P., Warren, D. L., Dhileepan, K., Mcconnachie, A., Strathie, L., Hassan, G., Karki, D., Shrestha, B. B. & Parmesan, C. 2015. Projecting future expansion of invasive species: comparing and improving methodologies for species distribution modeling. *Global Change Biology,* 21(12), pp.4464-4480. doi:10.1111/gcb.13038.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K. & Thuiller, W. 2009. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and distributions,* 15(1), pp.59-69. doi:10.1111/j.1472-4642.2008.00491.x.

Martínez-Freiría, F., Tarroso, P., Rebelo, H. & Brito, J. C. 2016. Contemporary niche contraction affects climate change predictions for elephants and giraffes. *Diversity and distributions,* 22(4), pp.432-444. doi:10.1111/ddi.12406.

Matsuoka, K., Skoglund, A., Roth, G., De Pomereu, J., Griffiths, H., Headland, R., Herried, B., Katsumata, K., Le Brocq, A. & Licht, K. 2021. Quantarctica, an integrated mapping environment for Antarctica, the Southern Ocean, and sub-Antarctic islands. *Environmental Modelling & Software,* 140. doi:10.1016/j.envsoft.2021.105015.

Mattivi, P., Franci, F., Lambertini, A. & Bitelli, G. 2019. TWI computation: a comparison of different open source GISs. *Open Geospatial Data, Software and Standards,* 4(1), pp.1-12. doi:10.1186/s40965-019-0066-y.

Mccullagh, P. & Nelder, J. 1989. Generalized Linear Models II, Advanced School and Conference on Statistics and Applied Probability in Life Sciences,

Mcfeeters, S. K. 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing,* 17(7), pp.1425-1432. doi:10.1080/01431169608948714.

Mcginnis, D. & Rango, A. 1975. Earth resources satellite systems for flood monitoring. *Geophysical Research Letters,* 2(4), pp.132-135. doi:10.1029/GL002i004p00132.

Mcpherson, J. M., Jetz, W. & Rogers, D. J. 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of applied ecology,* 41(5), pp.811-823. doi:10.1111/j.0021-8901.2004.00943.x.

Milevski, I. 2007. Morphometric elements of terrain morphology in the Republic of Macedonia and their influence on soil erosion, International conference erosion and torrent control as a factor in sustainable river basin management, Belgrade, Serbia

Millennium Ecosystem Assessment 2005. *Ecosystems and human well-being.* Island Press Washington, DC.

Miller, J. 2010. Species distribution modeling. *Geography Compass,* 4(6), pp.490-509. doi:10.1111/j.1749-8198.2010.00351.x.

Minasny, B., Berglund, Ö., Connolly, J., Hedley, C., De Vries, F., Gimona, A., Kempen, B., Kidd, D., Lilja, H. & Malone, B. 2019. Digital mapping of peatlands–A critical review. *Earth-Science Reviews,* 196. doi:10.1016/j.earscirev.2019.05.014.

Moore, I. D., Grayson, R. & Ladson, A. 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrological processes,* 5(1), pp.3-30. doi:10.1002/hyp.3360050103.
Naimi, B. & Araújo, M. B. 2016. sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography,* 39(4), pp.368-375. doi:10.1111/ecog.01881.
Naimi, B., Skidmore, A. K., Groen, T. A. & Hamm, N. A. 2011. Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modelling. *Journal of Biogeography,* 38(8), pp.1497-1509. doi:10.1111/j.1365-2699.2011.02523.x.
O'donnell, M. S. & Ignizio, D. A. 2012. Bioclimatic predictors for supporting ecological applications in the conterminous United States. *US geological survey data series,* 691(10), pp.4-9.
Ollis, D., Snaddon, C. & Job, N. 2013. Classification system for wetlands and other aquatic ecosystems in South Africa. *Water SA,* 41(5), pp.727-745. doi:10.4314/wsa.v41i5.16.
Øvstedal, D. & Gremmen, N. 2001. The lichens of Marion and Prince Edward Islands. *South African Journal of Botany,* 67(4), pp.552-572. doi:10.1016/S0254-6299(15)31187-X.
Ozesmi, S. L. & Bauer, M. E. 2002. Satellite remote sensing of wetlands. *Wetlands Ecology and Management,* 10, pp.381-402.
Pakhomov, E. A. & Chown, S. L. 2003. The Prince Edward islands: southern ocean oasis. *Ocean Yearbook Online,* 17(1), pp.348-379.
Pearson, R. G. 2007. Species' distribution modeling for conservation educators and practitioners. *Lessons in Conservation,* 3, pp.54-89.
Pekel, J.-F., Cottam, A., Gorelick, N. & Belward, A. S. 2016. High-resolution mapping of global surface water and its long-term changes. *Nature,* 540, pp.418-422. doi:10.1038/nature20584.
Pendlebury, S. & Barnes-Keoghan, I. P. 2007. Climate and climate change in the sub-Antarctic. *Papers and Proceedings of the Royal Society of Tasmania,* 141(1), pp.67-81. doi:10.26749/rstpp.141.1.67.
Peterson, A. T., Papeş, M. & Soberón, J. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological modelling,* 213(1), pp.63-72. doi:10.1016/j.ecolmodel.2007.11.008.
Phillips, S. J., Anderson, R. P. & Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological modelling,* 190(3-4), pp.231-259. doi:10.1016/j.ecolmodel.2005.03.026.
Pontius Jr, R. G. & Millones, M. 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing,* 32(15), pp.4407-4429. doi:10.1080/01431161.2011.552923.
Qin, C.-Z., Zhu, A.-X., Pei, T., Li, B.-L., Scholten, T., Behrens, T. & Zhou, C.-H. 2011. An approach to computing topographic wetness index based on maximum downslope gradient. *Precision agriculture,* 12(1), pp.32-43.
Raes, N. & Ter Steege, H. 2007. A null-model for significance testing of presence-only species distribution models. *Ecography,* 30(5), pp.727-736. doi:10.1111/j.2007.0906-7590.05041.x.
Raeymaekers, G., Sundseth, K. & Gazenbeek, A. 2000. *Conserving mires in the European Union.* Office for Official Publications of the European Communities.
Ramsar Convention on Wetlands 2018. Global wetland outlook: state of the World's wetlands and their services to people, Ramsar Convention Secretariat, Gland, Switzerland,
Ramsar Convention Secretariat 2010. Designating ramsar sites: strategic framework and guidelines for the future development of the list of wetland for international importance, Ramsar handbooks for the wise use of wetlands, Gland, Switzerland,
Rebelo, A. J., Scheunders, P., Esler, K. J. & Meire, P. 2017. Detecting, mapping and classifying wetland fragments at a landscape scale. *Remote Sensing Applications: Society and Environment,* 8, pp.212-223. doi:10.1016/j.rsase.2017.09.005.
Rokni, K., Ahmad, A., Selamat, A. & Hazini, S. 2014. Water feature extraction and change detection using multitemporal landsat imagery. *Remote Sensing,* 6(5), pp.4173-4189. doi:10.3390/rs6054173.
Rouault, M. 2005. Climate variability at Marion Island, Southern Ocean, since 1960. *Journal of Geophysical Research,* 110(C5), pp.1-9. doi:10.1029/2004jc002492.
Rudolph, E., Hedding, D. & Nel, W. 2020. The surface geology of the Prince Edward Islands: refined spatial data and call for geoconservation. *South African Journal of Geology,* 124(3). doi:10.25131/sajg.124.0014.
Rydin, H., Jeglum, J. K. & Bennett, K. D. 2013. *The biology of peatlands, 2e.* Oxford university press.
Rydin, H., Sjörs, H. & Löfroth, M. 1999. 7. Mires. *Acta Phytogeographica Suecica,* 84, pp.91-112.
Sadiki, M. 2019. *Quantifying changes in surface water bodies in response to climate change on Marion Island.* Unpublished Honours project, University of Pretoria,

Sarp, G. & Ozcelik, M. 2018. Water body extraction and change detection using time series: A case study of Lake Burdur, Turkey. *Journal of Taibah University for Science,* 11(3), pp.381-391. doi:10.1016/j.jtusci.2016.04.005.
Selkirk, P. 2007. The nature and importance of the sub-Antarctic. *Papers and Proceedings of the Royal Society of Tasmania,* 141(1), pp.1-6. doi:10.26749/rstpp.141.1.1.
Semeniuk, C. & Semeniuk, V. 1995. A geomorphic approach to global classification for inland wetlands. In: Finlayson, C. & van der Valk, A. (eds), Classification and Inventory of the World's Wetlands, Advances in Vegetation Science. Dordrecht: Springer, pp.103-124.
Sieben, E. J., Khubeka, S. P., Sithole, S., Job, N. M. & Kotze, D. C. 2018. The classification of wetlands: integration of top-down and bottom-up approaches and their significance for ecosystem service determination. *Wetlands Ecology and Management,* 26(3), pp.441-458.
Skentos, A. 2017. Topographic Position Index based landform analysis of Messaria (Ikaria Island, Greece). *Acta Geobalcanica,* 4(1), pp.7-15. doi:10.18509/agb.2018.01.
Skentos, A. & Ourania, A. 2017. Landform analysis using terrain attributes. A GIS application on the island of Ikaria (Aegean Sea, Greece). *Annals of Valahia University of Targoviste, Geographical Series,* 17(1), pp.90-97. doi:10.1515/avutgs-2017-0009.
Smith, V. R. 2002. Climate change in the Sub-Antarctic: an illustration from Marion Island. *Climatic Change,* 52(3), pp.345-357.
Smith, V. R. 2008. Terrestrial and freshwater primary production and nutrient cycling. In: Chown, S. L. & Froneman, P. W. (eds), The Prince Edward Islands. Land-Sea Interactions in a Changing Ecosystem. Stellenbosch: African Sun Media, pp.181-214.
Smith, V. R. & Lewis Smith, R. I. 1987. The biota and conservation status of sub-Antarctic islands. *Environment International,* 13(1), pp.95-104. doi:10.1016/0160-4120(87)90047-X.
Smith, V. R. & Mucina, L. 2006. Vegetation of Subantarctic Marion and Prince Edward Islands. In: Mucina, L. & Rutherford, M. (eds), The Vegetation of South Africa, Lesotho and Swaziland. South African National Biodiversity Institute, Pretoria: pp.699-723.
Smith, V. R. & Steenkamp, M. 1990. Climatic change and its ecological implications at a subantarctic island. *Oecologia,* 85(1), pp.14-24. doi:10.1007/BF00317338.
Smith, V. R., Steenkamp, M. & Gremmen, N. J. M. 2001. Terrestrial habitats on sub-Antarctic Marion Island: their vegetation, edaphic attributes, distribution and response to climate change. *South African Journal of Botany,* 67(4), pp.641-654. doi:10.1016/s0254-6299(15)31195-9.
Stagl, J., Mayr, E., Koch, H., Hattermann, F. F. & Huang, S. 2014. Effects of Climate Change on the Hydrological Cycle in Central and Eastern Europe. In: Rannow, S. & Neubert, M. (eds), Managing Protected Areas in Central and Eastern Europe Under Climate Change, Advances in Global Change Research 58 ed. Dordrecht: Springer, pp.31-43.
Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. *Science,* 240(4857), pp.1285-1293. doi:10.1126/science.3287615.
Tağıl, Ş. & Jenness, J. 2008. GIS-based automated landform classification and topographic, landcover and geologic attributes of landforms around the Yazoren Polje, Turkey. *Journal of Applied Sciences,* 8(6), pp.910-921.
Termansen, M., Mcclean, C. J. & Preston, C. D. 2006. The use of genetic algorithms and Bayesian classification to model species distributions. *Ecological modelling,* 192(3-4), pp.410-424. doi:10.1016/j.ecolmodel.2005.07.009.
The Editors of Encyclopaedia Britannica. 2016a. *Andosol* [Online]. Encyclopædia Britannica, inc. Available: https://www.britannica.com/science/Andosol [Accessed 30 July 2020].
The Editors of Encyclopaedia Britannica. 2016b. *Histosol* [Online]. Encyclopædia Britannica, inc. Available: https://www.britannica.com/science/Histosol-soil [Accessed 30 July 2020].
The Editors of Encyclopaedia Britannica. 2016c. *Regosol* [Online]. Encyclopædia Britannica, inc. Available: https://www.britannica.com/science/Regosol [Accessed 30 July 2020].
Thomas, R. F., Kingsford, R. T., Lu, Y., Cox, S. J., Sims, N. C. & Hunter, S. J. 2015. Mapping inundation in the heterogeneous floodplain wetlands of the Macquarie Marshes, using Landsat Thematic Mapper. *Journal of Hydrology,* 524, pp.194-213. doi:10.1016/j.jhydrol.2015.02.029.
Torbick, N., Lawrence, P. & Czajkowski, K. 2007. Application and assessment of a GIScience model for jurisdictional wetlands identification in Northwestern Ohio. In: Ji, W. (eds), Wetland and Water Resource Modeling and Assessment: A Watershed Perspective. pp.2-12.
Treasure, A. M., Le Roux, P. C., Mashau, M. H. & Chown, S. L. 2019. Species-energy relationships of indigenous and invasive species may arise in different ways–a demonstration using springtails. *Scientific reports,* 9(1), pp.1-12.
U.S. Geological Survey. 2021. *NDVI, the Foundation for Remote Sensing Phenology* [Online]. Available: https://www.usgs.gov/core-science-systems/eros/phenology/science/ndvi-foundation-remote-sensing-phenology?cv=1&qt-science_center_objects=0#qt-science_center_objects [Accessed 25 October 2021].

Uuemaa, E., Hughes, A. O. & Tanner, C. C. 2018. Identifying feasible locations for wetland creation or restoration in catchments by suitability modelling using light detection and ranging (LiDAR) digital elevation model (DEM). *Water,* 10(4). doi:10.3390/w10040464.

Van Zinderen Bakker Sr, E. 1970. Marion and Prince Edward Islands.

Verwoerd, W. & Langenegger, O. 1968. Marion and Prince Edward Islands: geological studies. *Nature,* 213(5073), pp.231-232.

Wang, L. & Liu, H. 2006. An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. *International Journal of Geographical Information Science,* 20(2), pp.193-213. doi:10.1080/13658810500433453.

Wang, Y., Colby, J. & Mulcahy, K. 2002. An efficient method for mapping flood extent in a coastal floodplain using Landsat TM and DEM data. *International Journal of Remote Sensing,* 23(18), pp.3681-3696. doi:10.1080/01431160110114484.

Weiss, A. 2001. Topographic position and landforms analysis, Poster presentation, ESRI user conference, San Diego, California, 9 July 2001.

Wolf, J. & Fricker, A. 2013. *Topographic-Wetness-Index: A revised version of the TWI written for arcpython.*

Wolock, D. M., Winter, T. C. & Mcmahon, G. 2004. Delineation and evaluation of hydrologic-landscape regions in the United States using geographic information system tools and multivariate statistical analyses. *Environmental management,* 34(1), pp.S71-S88. doi:10.1007/s00267-003-5077-9.

Work, E. A., Arbor, A. & Gilmer, D. S. 1976. Utilization of satellite data for inventorying prairie ponds and lakes. *Photogrammetric Engineering and Remote Sensing,* 42(5), pp.685-694.

Xu, H. 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing,* 27(14), pp.3025-3033. doi:10.1080/01431160600589179.

Yeloff, D., Mauquoy, D., Barber, K., Way, S., Van Geel, B. & Turney, C. S. 2007. Volcanic ash deposition and long-term vegetation change on subantarctic Marion Island. *Arctic, Antarctic, and Alpine Research,* 39(3), pp.500-511.

Yu, H., Cooper, A. R. & Infante, D. M. 2020. Improving species distribution model predictive accuracy using species abundance: Application with boosted regression trees. *Ecological Modelling,* 432. doi:10.1016/j.ecolmodel.2020.109202.

Yu, Z., Beilman, D. W. & Jones, M. C. 2009. Sensitivity of northern peatland carbon dynamics to Holocene climate change. *Carbon cycling in northern peatlands,* 184, pp.55-69.

Zhai, K., Wu, X., Qin, Y. & Du, P. 2015. Comparison of surface water extraction performances of different classic water indices using OLI and TM imageries in different situations. *Geo-Spatial Information Science,* 18(1), pp.32-42. doi:10.1080/10095020.2015.1017911.

Zhong, Y., Xue, Z., Jiang, M., Liu, B. & Wang, G. 2021. The application of species distribution modeling in wetland restoration: A case study in the Songnen Plain, Northeast China. *Ecological Indicators,* 121. doi:10.1016/j.ecolind.2020.107137.

Zimmermann, N. E., Edwards Jr, T. C., Graham, C. H., Pearman, P. B. & Svenning, J. C. 2010. New trends in species distribution modelling. *Ecography,* 33(6), pp.985-989. doi:10.1111/j.1600-0587.2010.06953.x.

Appendices

Appendix 1

Table A-1: Correlation matrix showing Pearson's correlation coefficient for all 28 predictor variables. If two variables have a Pearson (r) correlation coefficient larger than a threshold, they are correlated (Naimi and Araújo, 2016). Correlation greater or equal to 0.9 (highlighted red) suggests multicollinearity issues.

	Distance From Coast	Elevation	Geology	NDVI	NDWI	Slope	Soil	TPI	TWI	BIO01	BIO04	BIO12	BIO15
Distance From Coast	1												
Elevation	0,84	1											
Geology	0,17	0,25	1										
NDVI	-0,71	-0,8	-0,27	1									
NDWI	0,64	0,73	0,23	-0,97	1								
Slope	0,17	0,28	0,25	-0,24	0,26	1							
Soil	0,25	0,3	0,08	-0,25	0,24	0,15	1						
TPI	0,05	0,11	0,07	-0,11	0,11	0,11	0,04	1					
TWI	-0,07	-0,11	-0,09	0,11	-0,12	-0,37	-0,06	-0,33	1				
BIO01	-0,77	-0,97	-0,22	0,79	-0,74	-0,3	-0,29	-0,07	0,11	1			
BIO04	-0,02	-0,04	-0,11	-0,06	0,07	0	0,04	0,01	0	0,07	1		
BIO12	0,68	0,9	0,15	-0,72	0,68	0,33	0,26	0,07	-0,11	-0,94	0,02	1	
BIO15	0,58	0,75	0,12	-0,63	0,6	0,3	0,27	0,06	-0,1	-0,8	0,05	0,92	1

Appendix 2

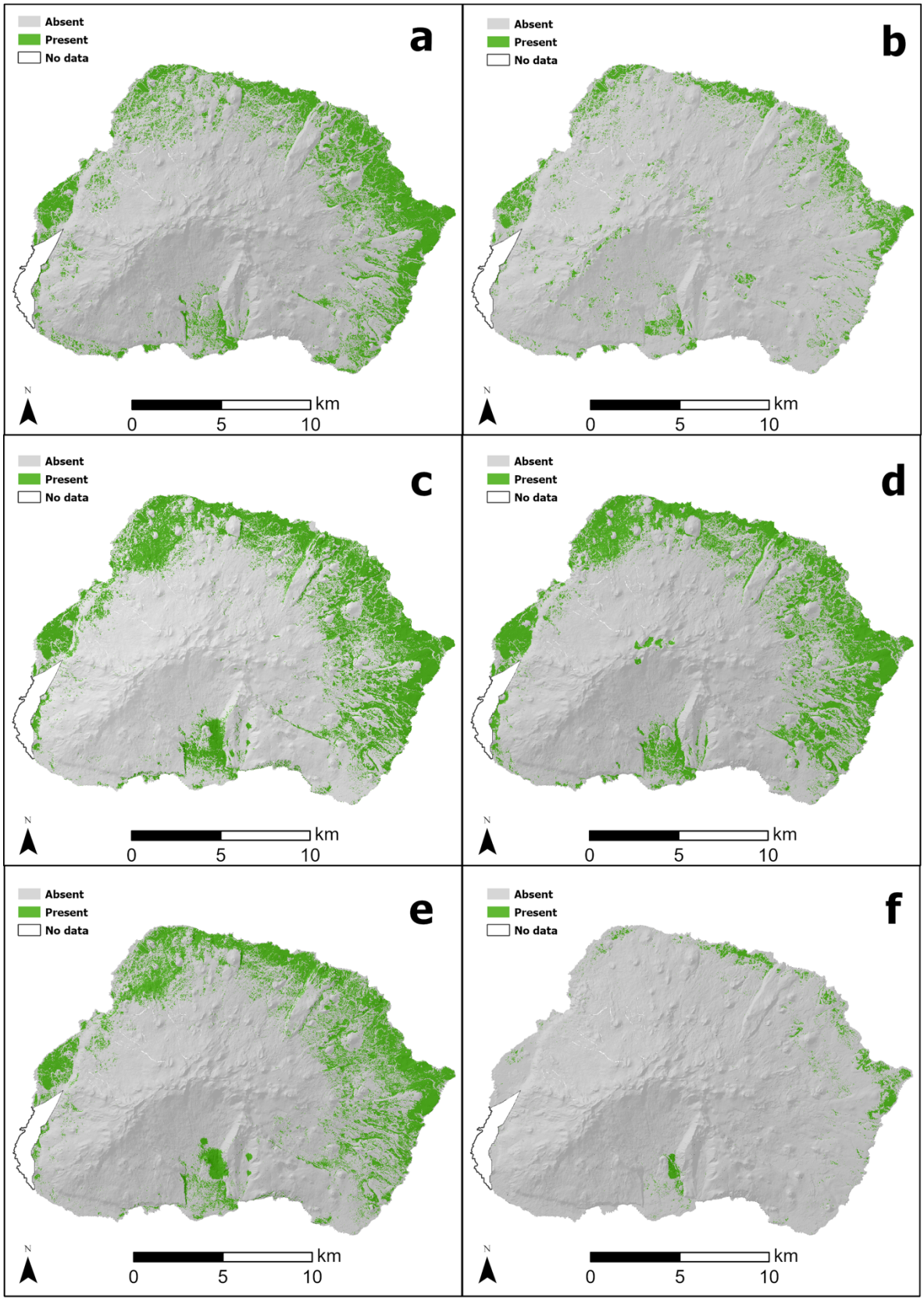


Figure A-1: Scenario 1 model predictions based on the mean of 10 fold cross-validation predictions per model. BRT (a), CART (b), GAM (c), GLM (d), MARS (e), RF (f).

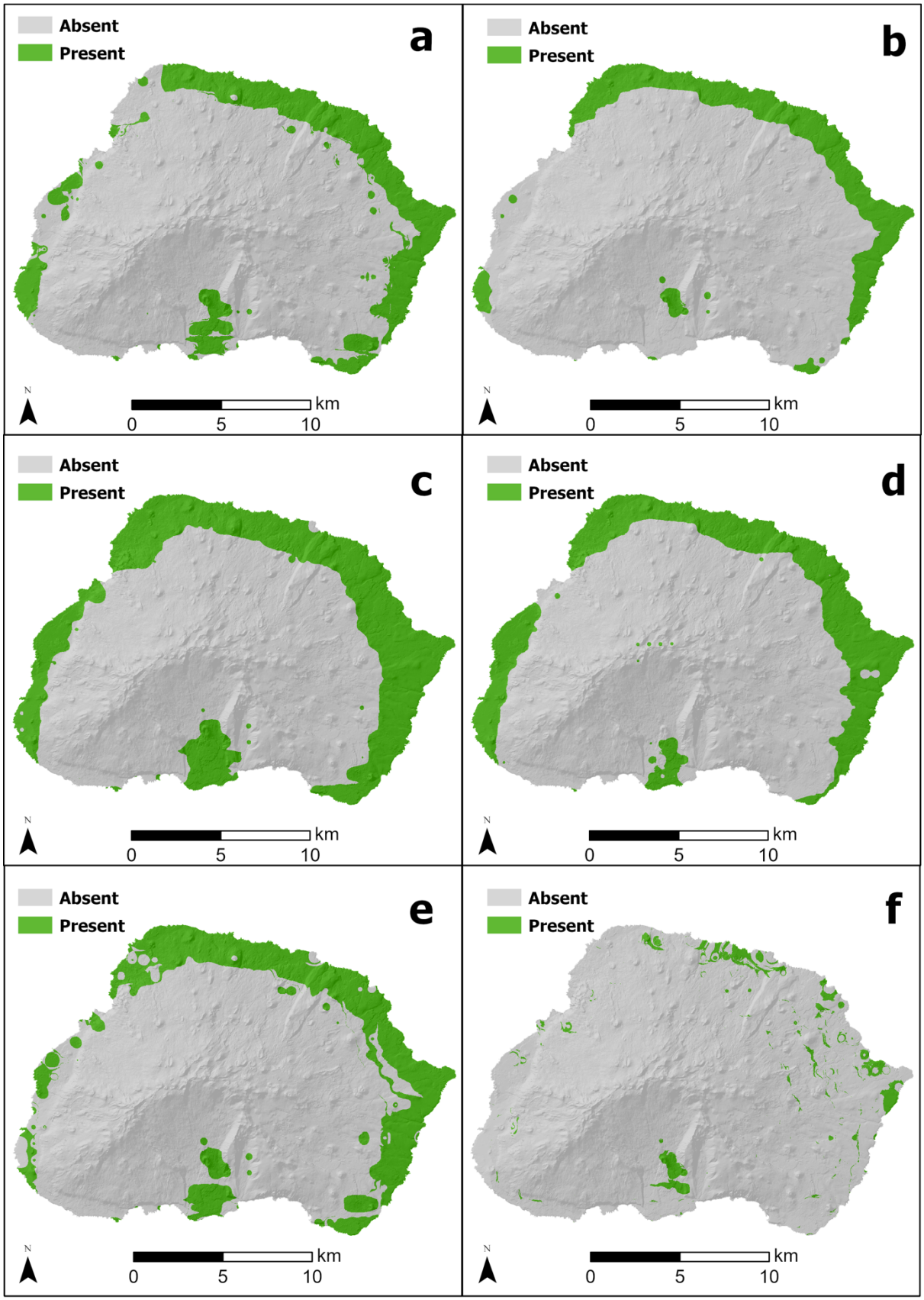


Figure A-2: Scenario 2 model predictions based on the mean of 10 fold cross-validation predictions per model. BRT (a), CART (b), GAM (c), GLM (d), MARS (e), RF (f).

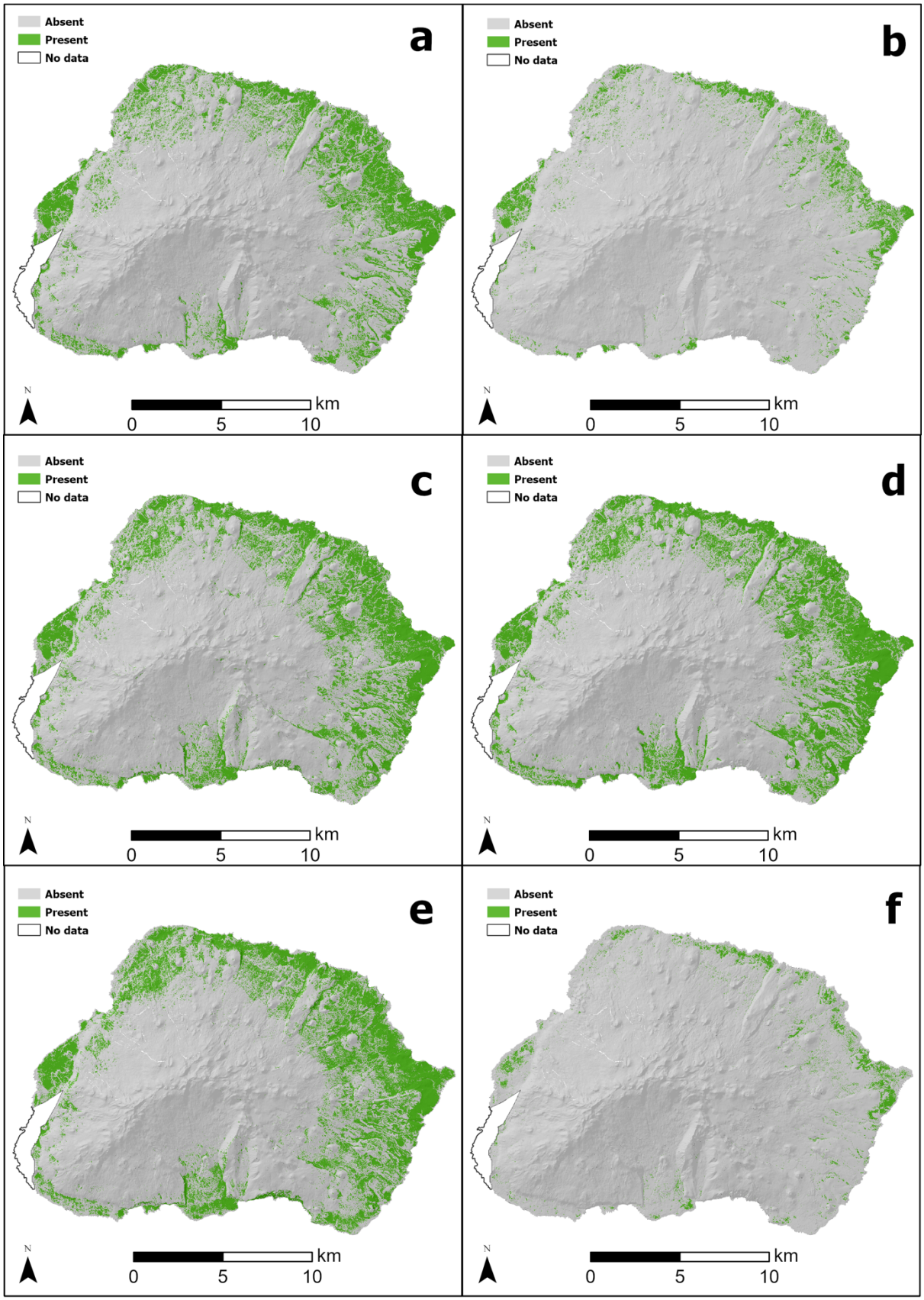


Figure A-3: Scenario 3 model predictions based on the mean of 10 fold cross-validation predictions per model. BRT (a), CART (b), GAM (c), GLM (d), MARS (e), RF (f).

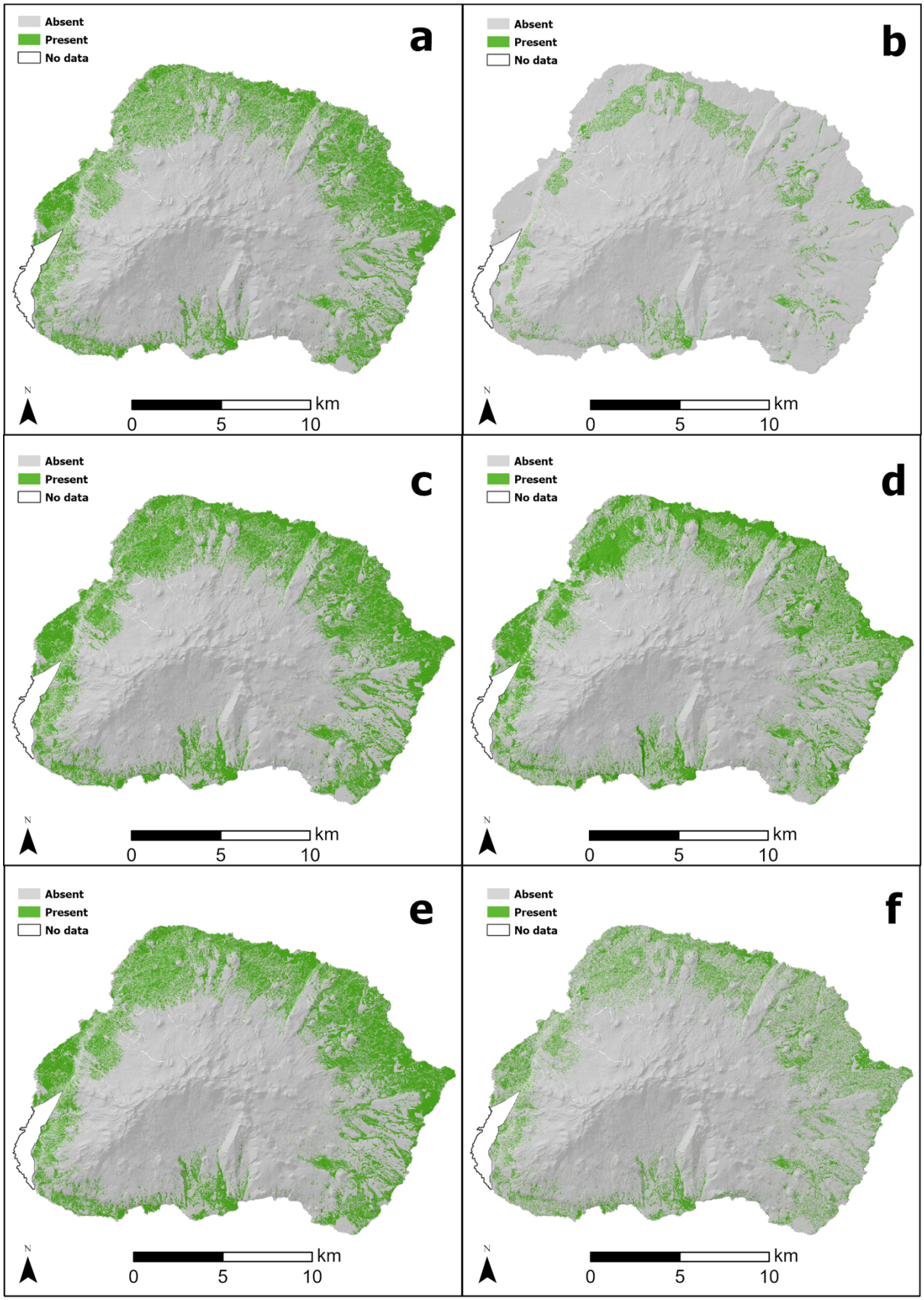


Figure A-4: Scenario 4(a) model predictions based on the mean of 10 fold cross-validation predictions per model. BRT (a), CART (b), GAM (c), GLM (d), MARS (e), RF (f).

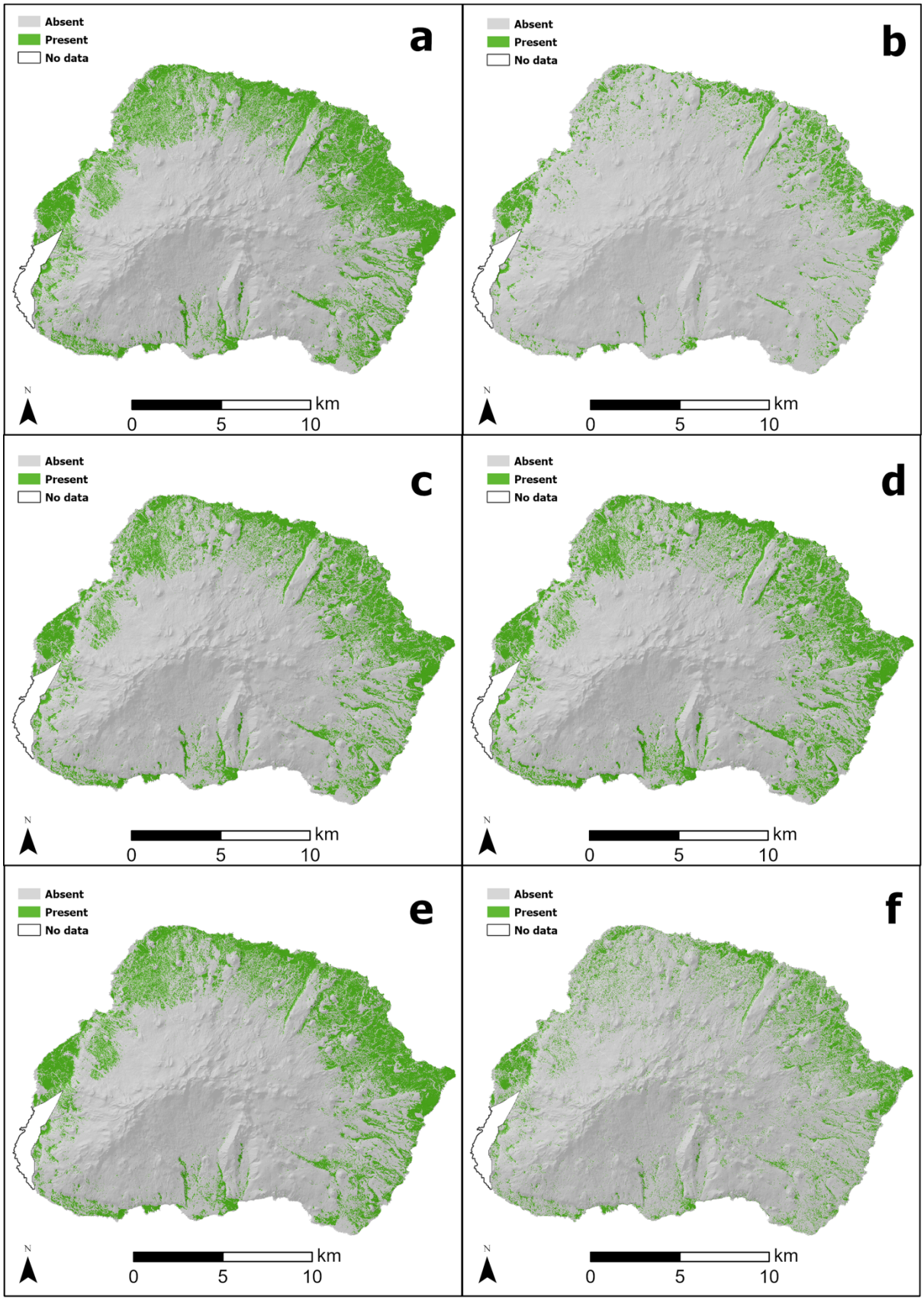


Figure A-5: Scenario 4(b) model predictions based on the mean of 10 fold cross-validation predictions per model. BRT (a), CART (b), GAM (c), GLM (d), MARS (e), RF (f).

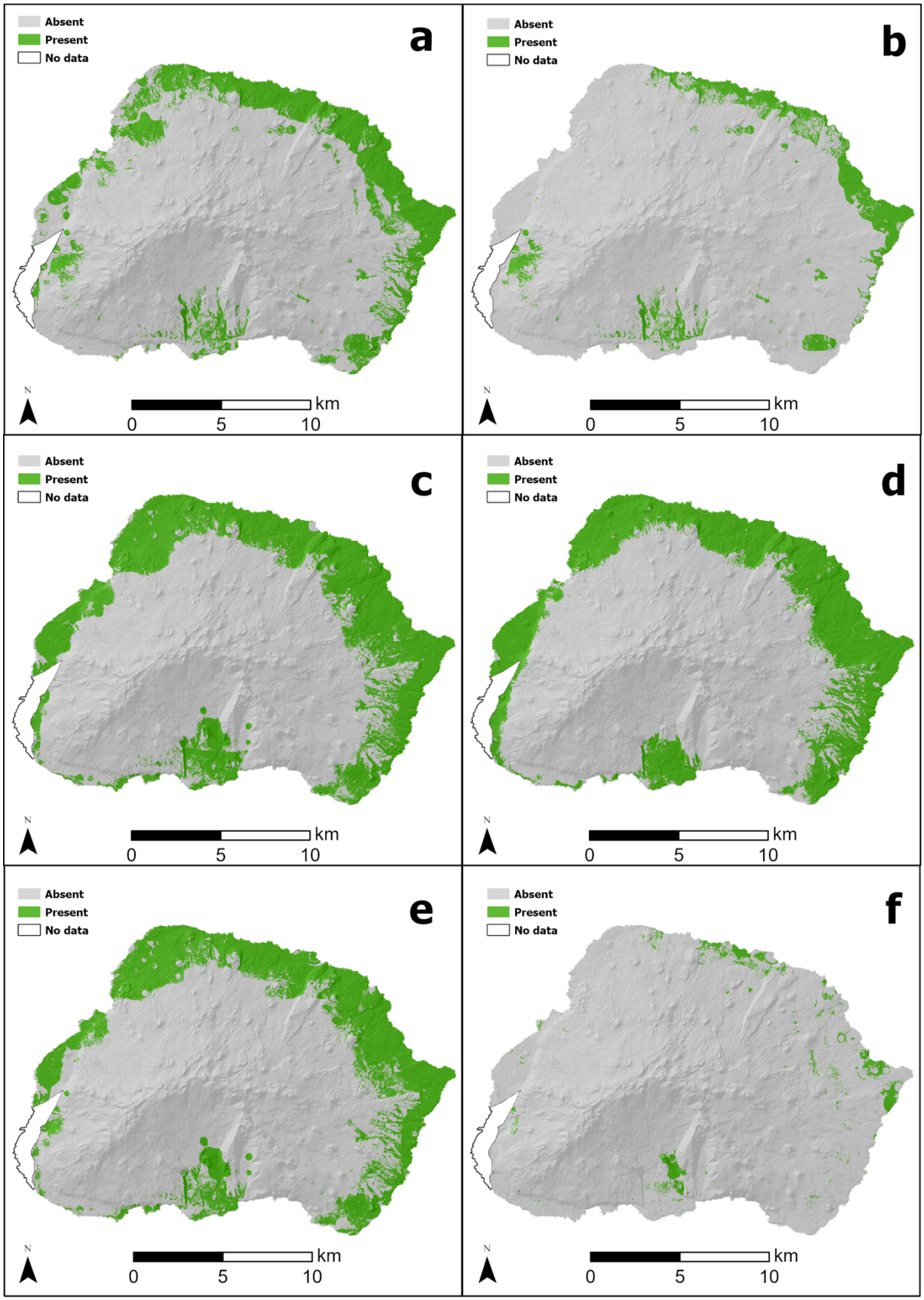


Figure A-6: Scenario 4(c) model predictions based on the mean of 10 fold cross-validation predictions per model. BRT (a), CART (b), GAM (c), GLM (d), MARS (e), RF (f).

Table A-2: Confusion matrix for Scenario 1's GAM model.

		Reference Data			
		Absent	**Present**	**Row Total**	**User Accuracy**
Classified data	**Absent**	813	71	885	0,918644
	Present	314	176	490	0,359184
	Column Total	1127	248	1375	0
	Producer Accuracy	0,721384	0,709677	0	**0,719273**

Overall Accuracy = 72%, KAPPA = 0.31

Table A-3: Confusion matrix for Scenario 1's RF model.

		Reference Data			
		Absent	**Present**	**Row Total**	**User Accuracy**
Classified data	**Absent**	1126	11	1137	0,990325
	Present	1	237	238	0,995798
	Column Total	1127	248	1375	0
	Producer Accuracy	0,999113	0,754032258	0	**0,991278**

Overall Accuracy = 99%, KAPPA = 0.97

Table A-4: Confusion matrix for Scenario 1's Ensemble model.

		Reference Data			
		Absent	**Present**	**Row Total**	**User Accuracy**
Classified data	**Absent**	973	59	1032	0,942829
	Present	154	189	343	0,55102
	Column Total	1127	248	1375	0
	Producer Accuracy	0,863354	0,762097	0	**0,8450918**

Overall Accuracy = 85%, KAPPA = 0.54