

MONITORING POPLARS USING GEOSPATIAL APPROACHES IN SOUTH AFRICA

MORENA JOHN MAPURU



Thesis submitted to the School of Agricultural, Earth and Environmental Sciences at the University of KwaZulu-Natal in fulfilment of the requirements for the degree of

Doctor of Philosophy in Geography

Supervisor: Dr. Sifiso Xulu

November 2024

Durban, South Africa

PREFACE

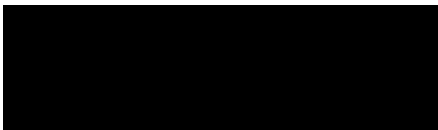
I hereby submit this thesis for the degree of Doctor of Philosophy at the University of KwaZulu-Natal. The research outlined in this document was carried out under the guidance of Dr. Sifiso Xulu in the School of Agricultural, Earth, and Environmental Sciences.

To the best of my knowledge, this work is original, except where I have acknowledged and referenced previous research. I confirm that this thesis, or any substantially similar thesis, has not been submitted for any other degree or qualification at any other University.

.....

Morena John Mapuru

I, as the supervisor of Morena John Mapuru, verify the aforementioned documents and have authorized the submission of this thesis.



.....

Dr. Sifiso Xulu

PUBLICATIONS

1. Mapuru, M.J., Xulu, S. and Gebreslasie, M., 2023. Remote sensing applications in monitoring poplars: A review. *Forests*, 14(12), p.2301. (IF = 2.4)
2. Mapuru, M., Xulu, S., Gebreslasie, M. and Sadiki, M. [Under review]. Sentinel-2 based classification and separability of land surface features in the poplar-invaded landscape of the eastern Free State. *Journal of Sensors*. (IF = 1.4)
3. Mapuru, M., Xulu, S. and Gebreslasie, M., 2024. Modelling possible habitats for poplar invasion in South Africa. *KOEDOE-African Protected Area Conservation and Science*, 66(1), p.1783. (IF = 1.5)
4. Mapuru, M.J., Xulu, S., Gebreslasie, M., and Daemane, E.M., 2024. Exploring environmental factors that influence the distribution of poplar trees. *African Journal of Ecology*, 62, e13310. <https://doi.org/10.1111/aje.13310> (IF = 1.1)

CONFERENCE PRESENTATIONS

1. Mapuru, M.J., Xulu, S. and Gebreslasie, M.T., 2023. Modelling the distribution of poplars in South Africa using Generalized Linear Model and Ensemble Models. *Society of South African Geographers (SSAG) student conference held at University of the Free State (UFS) QwaQwa campus from 2 to 4 October 2023*.
2. Mapuru, M., Xulu, S., Gebreslasie, M. and Sadiki, M., 2024. Sentinel-2 based classification and separability of land surface features in the poplar invaded landscape of the eastern Free State. *Kimberly Biodiversity Research Symposium (KBRS) held at Sol Plaatje University south campus on the 18th September 2024*.

ABSTRACT

The proliferation of aggressive invasive alien species (IAS) such as poplar (*Populus spp*) poses a significant threat to both natural and human-modified ecosystems, because it is highly adaptable and has the ability to outcompete and displace important native species, particularly in rangelands and remote areas. This has raised serious concerns among conservationists and land managers. Therefore, there is a need to implement effective measures to monitor and manage the spread of poplar in potentially invaded landscapes. In South Africa, poplar commonly occur in several geographic regions but is merely allowed under strict legal restrictions in specific areas. To understand the geographic distribution and patterns of poplar across landscapes, it is crucial to use approaches based on geographic information technology (GIT) involving remote sensing and Geographical Information Systems (GIS), as they have consistently shown the ability to uncover complex patterns. While these techniques have been used globally to study IAS, it is surprising that poplar research in the country has not yet made full use of them. In response, the current study is the first to address this research gap as it investigates, and review and test the GIT applications in studying poplar trees in South Africa. This study aims to explore geospatial techniques to map and classify, model, and predict poplar distribution patterns. To achieve this aim, a set of four specific objectives were developed.

In the first objective, the study synthesized and reviewed 148 peer-reviewed poplar studies from Scopus, Google Scholar, and Science Direct databases. The primary goal of this review was to determine the extent to which geospatial tools have been used to study poplar and to identify any existing gaps. The analysis revealed a significant increase in the use of Earth observation for poplar research between 1991 and 2022, using predominantly airborne platforms, especially LiDAR, then satellites and, to a lesser extent ground-based sensor platforms. More than two-thirds of the studies were conducted in the Global North, much of which took place in agricultural regions, while some covered forested and riparian zones and relatively less over urban areas and mountainous regions. Spectral indicators have been widely used to assess poplar health and productivity, with increasing use of machine learning.

In the second objective, Sentinel-2 imagery combined with Random Forests (RF) was used to map poplar in the northeastern Free State province of South Africa. The study aimed to address the current gap in mapping IAS in mountainous areas. The study successfully: (a) identified and distinguished poplar tree species from other land cover classes in the study region throughout the year, (b) analysed the usefulness of Sentinel-2 bands and their combinations in the mapping process, and (c) evaluated the mapping accuracy for the first and second half of the year. The findings revealed that images from

the first half of the year (Jan-Jun) had a higher classification accuracy (overall accuracy: 91% and Kappa: 0.89) compared to images from the second half of the year (Jul-Dec) (overall accuracy: 87% and Kappa: 0.87). Variables such as SWIR-1, NDVI, Blue, PI-1, MNDWI, NIR, and PI-3 were identified as crucial parameters in the classification of poplar trees in mountainous areas. Overall, the study showed that Sentinel-2 bands and indices with the RF classifier holds much potential for mapping poplar alien trees in mountainous environments.

In the third objective, the study used a generalized linear model (GLM), RF, and Support Vector Machines (SVM) to predict the potential habitat of poplar trees in South Africa under the current climatic conditions, thereby assessing the influence of different climatic parameters. The study found that all models performed well, with RF outperforming SVM and GLM. Geographically, the highest concentrations of suitable poplar habitat were found in the high-rainfall areas of South Africa. The study's analysis further revealed that annual precipitation, as well as precipitation during the driest and warmest quarters, significantly influenced the distribution patterns of poplar trees. Overall, the study demonstrates that regression and machine learning algorithms show promise in predicting suitable habitats for invasive alien species such as poplar.

In the fourth objective, the focus of the study was to investigate the local environmental factors that can potentially influence the presence and distribution of poplars in mountainous areas. The study used a GLM to identify these factors and then ranked them based on their influence on poplar distribution using a RF strategy. The GLM analysis indicated that soil composition (specifically soil organic carbon, sand, and silt), topographical features (slope and aspect), surface moisture content (NDMI), and distance from rivers are all significant factors affecting the presence of poplars. Upon ranking these factors, the study found that the most influential factors were aspect, slope, and distance from rivers, followed by sand, silt, NDMI, and soil organic carbon. The findings of the study align with previous research, suggesting that topographical features and water availability are the primary drivers behind the spread of alien invasive plants.

DEDICATION

I dedicate this PhD thesis to Mapuru family at large.

Ke phethile taelo

ACKNOWLEDGEMENTS

I wish to extend my profound gratitude to the following individuals, without whom the execution of this project would not have been possible:

1. My supervisor Dr. Sifiso Xulu for his unwavering efforts, guidance, and overall supervision of this PhD work.
2. Dr. Ernest Daemane, for proofreading, editing, and overall contributions to the Chapter 5 manuscript. I am truly grateful for your generosity in assisting me with the article processing charges for Chapter 4 paper. Among all the heroes I have encountered, you come first to mind. Thank you for everything.
3. My parents, Mr. Sekhukhune Mapuru and Mrs. Nthofela Mapuru, for their unyielding moral support and endless motivation throughout this study.
4. My kids, Atlehang and Itumeleng, for consistently bringing joy and purpose to my life.
5. My siblings, Khetheng, Seipati, Khopotso, Keletso, and Kananelo, for their constant love and support.
6. Miss Maleho Sadiki, for her invaluable contributions to the analysis and co-authoring of the Chapter 3 manuscript.
7. Mr. Lufuno Munyai, for his indispensable support and assistance during data collection, and for being a pillar and brother in Christ.
8. The entire SANParks scientific service Kimberly team, for their unwavering support and motivation during this study.
9. Finally, I would like to express my heartfelt gratitude to the "***God of Mount Zion***" for granting me the strength and resources to carry out this research study.

CONTENTS

PREFACE.....	II
PUBLICATIONS	III
ABSTRACT.....	IV
DEDICATION.....	VI
ACKNOWLEDGEMENTS.....	VII
CONTENTS.....	VIII
FIGURES.....	X
TABLES.....	XI
ACRONYMS AND ABBREVIATIONS.....	XII
CHAPTER 1 GENERAL INTRODUCTION	1
1.1 INTRODUCTION	1
1.2 SIGNIFICANCE OF THE STUDY	4
1.3 RESEARCH AIM AND OBJECTIVES	5
1.4 STUDY REGION	6
1.5 THESIS DESIGN AND OUTLINE.....	8
1.6 REFERENCES	10
CHAPTER 2 REMOTE SENSING APPLICATIONS IN MONITORING POPLARS: A REVIEW	17
ABSTRACT.....	18
2.1 INTRODUCTION	19
2.2 MATERIALS AND METHODS.....	20
2.2.1 Database search process.....	20
2.2.3 Data analysis methods	22
2.3 RESULTS AND DISCUSSION	22
2.3.1 Historical research trends	23
2.3.2 Geographical distribution of poplar remote sensing studies.....	27
2.3.4 Characteristics of methods and materials.....	30
2.3.5 Co-occurrence words/keywords	34
2.4 LIMITATIONS AND FUTURE RESEARCH	37
2.5 CONCLUSION	38
2.6 REFERENCES	39
CHAPTER 3 SENTINEL-2 BASED CLASSIFICATION AND SPECTRAL SEPARABILITY OF LAND SURFACE FEATURES IN THE POPLAR-INVADDED LANDSCAPE OF THE EASTERN FREE STATE, SOUTH AFRICA	48
ABSTRACT.....	49
3.1 INTRODUCTION	50
3.2 METHODS AND MATERIALS	52
3.2.1 Description of the study area.....	52
3.2.2 Acquisition and processing of the images.....	53
3.2.3 Dataset for training and validation	54
3.2.4 Time series analysis.....	55
3.2.5 Classification.....	56
3.2.6 Classification accuracy assessment.....	57
3.3 RESULTS	57
3.3.1 Spectral separability of classes.....	57
3.3.2 Influential importance.....	58
3.3.3 Classification algorithm performance	59
3.4 DISCUSSION	61
3.5 CONCLUSION	63
3.6 REFERENCES	64

CHAPTER 4 MODELLING POSSIBLE HABITATS FOR POPLAR INVASION IN SOUTH AFRICA	71
ABSTRACT	72
4.1 INTRODUCTION	73
4.2 MATERIALS AND METHODS	75
4.2.1 Study area	75
4.2.2 Data collection	76
4.2.3 Environmental predictors.....	77
4.2.4 Modelling methods	79
4.3. RESULTS	79
4.4 DISCUSSION	84
4.5 CONCLUSION	87
4.6 REFERENCES	87
CHAPTER 5 EXPLORING ENVIRONMENTAL FACTORS THAT INFLUENCE THE DISTRIBUTION OF POPLAR TREES	106
ABSTRACT	107
5.1 INTRODUCTION	108
5.2 MATERIALS AND METHODS	110
5.2.1 Study area	110
5.2.2 Field data collection	111
5.2.3 Environmental variables.....	112
5.2.4 Statistical analysis.....	113
5.3 RESULTS	114
5.3.1 Significant environmental predictors	114
5.3.2 Relative variable importance	115
5.4 DISCUSSION	116
5.5 STUDY LIMITATIONS.....	117
5.6 CONCLUSION	118
5.7 REFERENCES	118
CHAPTER 6 SYNTHESIS AND CONCLUSION	128
6.1 SYNTHESIS	128
6.1.1 Key findings and their integration	128
6.1.2 Broader implications	130
6.1.3 Recommendations	131
6.1.4 Limitations of the study	131
6.2 CONCLUDING REMARKS	132

FIGURES

Figure 1.1 Location of the study area in the eastern Free State Province of South Africa.....	5
Figure 1.2 Research design.....	6
Figure 2.1 A flowchart of the literature search process based on PRISMA.	22
Figure 2.2 Publication statistics in remote sensing of poplar species (1991–2022).	23
Figure 2.3 Remote sensing sensors used in the study of poplar species per year between 1991 and 2022.....	25
Figure 2.4 Countries that participated in poplar studies and their number of publication output.	27
Figure 2.5 Landscape categories where poplar studies were conducted.....	30
Figure 2.6 The network visualization of terms in the title/abstract of publications in the remote sensing of poplar. Circle size represents the number of publications derived from the database. Circle colors represent: yellow is vegetation mapping; green is forest ecology; blue is vegetation physiology; and red is machine learning lines represent the link between titles and abstracts.....	36
Figure 3.1 Study area map showing the location of the study site (marked with a red dot) in South Africa. An image in true colour displays Phuthaditjhaba, the current study site.	52
Figure 3.2 Spectral signatures for land cover classes ((a) PI-1, (b) SWIR-1, (c) Blue, (d) NDVI, (e) MNDWI, (f) NIR and (g) PI-3).	58
Figure 3.3 Variable importance for (a) Jan-Jun and (b) Jul-Dec as measured by the Random Forests algorithm indicating the usefulness of the different image bands and indices. The higher the value, the higher the predictor importance.	59
Figure 3.4 Random Forest classification maps for (a) Jan-Jun and (b) Jul-Dec Sentinel-2 imageries. True colour image (c) is included for visual inspection against classified maps.	61
Figure 3.5 Spectral signatures for land cover classes ((a) Green, (b) LAI, (c) PI-2, (d) PI-4, (e) Red, (f) Red edge-1, (g) Red edge-2, (h) Red edge-3 and (i) SWIR-2)	70
Figure 4.1 Geographic distribution of Poplars in South Africa.	76
Figure 4.2 Receiver operator characteristics (ROC) curve for (a) Random Forests, (b) Support Vector Machine, and (c) Generalized Linear Model	80
Figure 4.3 Modelled distribution of poplars in South Africa using (a) Random Forests, (b) Support Vector Machine, and (c) Generalized Linear Model. The legend/ramp on the y axis represent suitability levels, with green representing highly suitable site, while white-pinkish colour represents low suitability areas.....	81
Figure 4.4 a) RF prediction map and b) a June 2022 to June 2023 rainfall map with poplar points overlaid on both.	82
Figure 4.5 Random Forests variable importance.	84
Figure 5.1 A study map depicting South Africa (a), Free State (b), and Golden Gate Highlands National Parks (c).	111

TABLES

Table 2.1 Inclusion and exclusion criteria for the review.....	21
Table 2.2 Top 10 biophysical indicators (a), spectral indices (b), classifiers (c), and common software packages used in the reviewed studies.	33
Table 2.3 Journals where Popular studies were published and their funding agencies.	37
Table 3.1 Sentinel-2 imagery bands and vegetation indices used.....	54
Table 3.2 Land cover classes and corresponding training, validation, and total number of samples.	55
Table 3.3 Confusion matrix for (a) Jan-Jun and (b) Jul-Dec.	60
Table 4.1 Poplar species gathered from GBIF, including localities and counts.	77
Table 4.2 Bioclimatic variables and their descriptions	78
Table 4.3 Summary of model performance	80
Table 4.4 Evaluation of environmental factors for predicting the distributions of poplars.	83
Table 5.1 Results from the GLM on the effects of environmental predictors on poplar distribution.	115
Table 5.2 Variable ranking according to RF.....	116
Table 5.3 Descriptive statistics of all predictor variables used in the study.	127

ACRONYMS AND ABBREVIATIONS

AGB	Above Ground Biomass
AUC	Area Under the Curve
CART	Classification and Regression Tree
CVI	Chlorophyll Vegetation Index
DI	Diversity Index
DSM	Digital Surface Model
DT	Decision Tree
DTM	Digital Terrain Model
ESA	European Space Agency
GBIF	Global Biodiversity Information Facility
GGHNP	Golden Gate Highlands National Park
GLM	Generalized Linear Model
GPP	Gross Primary Productivity
GRNDVI	Green-Red Normalized Vegetation Index
GSV	Growing Stock Volume
IAS	Invasive Alien Species
IncMSE	Mean Decrease in Accuracy
LAI	Leaf Area Index
LCC	Leaf Chlorophyll Content
MaxEnt	Maximum Entropy
MERIS	Medium Resolution Imaging Spectrometer
ML	Machine Learning
MNDWI	Modified Normalized Difference Water Index
NDMI	Normalized Difference Moisture Index
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NEMBA	National Environment Management Biodiversity
NIR	Near-infrared
OA	Overall Accuracy
OBIA	Object Based Image Analysis
PA	Produce Accuracy
PRI	Photochemical Reflectance Index
	Preferred Reporting Items for Systematic Reviews and
PRISMA	Meta-Analysis
PTI	Poplar Tree Index
QGIS	Quantum Geographic Information Systems
RF	Random Forests
ROC	Receiver Operator Characteristics
SANParks	South African National Parks
SAVI	Soil-Adjusted Vegetation Index
SDM	Species Distribution Modelling
SSI	Stress Susceptibility Index
SVM	Support Vector Machine
SWIR	Shortwave-infrared
TCI	Terrestrial Chlorophyll Index
UA	User Accuracy

CHAPTER 1 GENERAL INTRODUCTION

1.1 INTRODUCTION

The Poplar tree species are native to Asia, Europe, and North America, comprising around 35 species of deciduous flowering plants in the Salicaceae family (Danise et al. 2021). With its significant genetic diversity, *Populus* can reach heights of 15 to 30 meters and have trunks up to 2.5 meters in diameter (Orlovic et al. 1998; Henderson, 2020). This species is valued for its wood production, rapid growth, and adaptability to various environments (Khayatinejad Aqgonbad et al. 2022). Its ecological benefits include erosion control, soil stability improvement, and nutrient cycle enhancement (Zhang et al. 2019). Poplar trees flourish in humid environments characterized by high rainfall and optimal temperatures (Mapuru et al. 2024). While they are native to most Global North countries, they are considered invasive alien species (IAS) in the Global South countries, including South Africa. In South Africa, strict conditions allow only certain species of poplar in specific areas due to their potential negative impact on local biodiversity (Henderson, 2007).

The presence of alien plants, such as poplar, is a significant problem in South Africa (Ntshidi et al. 2018). These plants harm the environment by altering groundwater systems, surface run-off, and intensifying floods, while also amplifying the potential occurrence of destructive wildfires (Görgens and Van Wilgen, 2004). Alien species outcompete native plants for water resources, directly impacting groundwater reserves and the flow of streams (WWF, 2016). According to WWF (2016), invasive plants use over 1 billion m³ of water yearly. This water is sufficient to irrigate more than 100000 hectares of cultivated land (WWF, 2016).

IAS often have deeper roots and tend to be taller than the native plants occurring in their immediate surroundings (Liu et al. 2023). These characteristics enable them to significantly increase their productivity, resulting in their dominance in invaded landscapes (Versveld and Van Wilgen, 1986). Their increase also results in greater fuel loads, intensifying wildfires and causing damage to soil and vegetation (Vega et al. 2022). The intense fires can create water-repellent layers in the soil, further contributing to soil erosion (Smith et al. 2011). The increase in ₁ sediment in water due to soil erosion compromises

water quality, impacting aquatic ecosystems. A study conducted by Richardson and Van Wilgen (2004) also highlighted how IAS disrupt rangeland quality, affecting livestock, wildlife, and biodiversity.

The impact of IAS goes beyond causing environmental damage and also leads to economic losses in South Africa. For instance, the Department of Environmental Affairs (DEA) (2018) indicates that the annual cost of clearing these species in the country is approximately R600 million. This money could be better utilized or invested in other critical areas to support life in South Africa. While a government body such as DEA have been crucial in the clearing, monitoring, and managing alien invasive species in the country, the Biodiversity Act 10 of 2024 includes the participation of private property owners in the management of alien invasive species.

According to the Biodiversity Act 10 of 2024, landowners must monitor and manage the presence of alien plants on their property. The first step is to identify these plants and then apply the necessary methods to either remove or control them (Wilson, 2017). There are three standardized approaches for controlling or removing alien plant species: mechanical, chemical, and biological methods (van Wilgen et al. 2001). The mechanical approach involves physically destroying or removing the plants (Núñez-González et al. 2021), while chemical methods include applying herbicides (van Wilgen et al. 2001). Biological control entails releasing natural enemies to reduce plant health and population (Clewley et al. 2012). Regular monitoring is essential for identifying emerging seedlings for ongoing removal. Regular monitoring using traditional survey can be costly and labour-intensive and may be difficult to apply in remote or inaccessible areas. In such cases, Remote Sensing and Geographical Information Systems can be valuable tools for mapping and identifying of alien invasive species of interest (Bley Dalouman et al. 2023). These tools are especially important for planning clearing and treatment, as well as monitoring cleared and invaded sites at lower costs (Adam et al. 2017).

Mapping invasive alien species through remote sensing can be achieved with imagery from sensors such as Sentinel (see Kattenborn et al. 2019; Zagajewski et al. 2024), Landsat (de Sa et al. 2017; Calleja et al. 2019), and Moderate Resolution Imaging Spectroradiometer (MODIS) (Wallace et al. 2016; Landmann et al. 2020), provided that the species has distinguishing characteristics. For example, Rajah et al. (2019)

used Sentinel 1 and 2 along with Support Vector Machine (SVM) to detect and map invasive alien species at Drakensberg Park, South Africa, achieving an accuracy of 80%. Valero-Jorge et al. (2024) studied the mapping and monitoring of *Dichrostachys cineria* using a combination of machine learning and Landsat imagery. Their results indicated that Random Forests outperformed other methods, achieving an overall performance of over 90%. Using the MODIS dataset, Wallace et al. (2016) accurately mapped approximately 49% to 55% of invasive Buffelgrass in Southern Arizona. Andrew and Ustin (2006) used an airborne hyperspectral imaging sensor to identify *Lepidium latifolium* in California, while Mitchell and Glenn (2009) identified *Euphorbia esula* in Southeast Idaho using HyMap. Data derived from remote sensing is typically used with traditional survey data to create invasive alien species maps, aiding in early detection and focusing control efforts.

One effective strategy for managing and understanding the distribution of invasive alien species is Species Distribution Modelling (SDM). SDMs use spatial occurrence data and a range of environmental factors to predict spatial patterns of environmental suitability for species (Elith and Leathwick, 2009). These models can forecast how species distribution patterns may change in response to global changes (Hill et al. 2015). Machine learning techniques have been used in SDMs to identify and prioritize the contribution of various predictor variables in determining the spread of invasive trees (Elias et al. 2019; Mapuru et al. 2024). These techniques have been successfully applied in plant studies. For example, Zhang et al. (2020) used Random Forests (RF) to predict the potential distribution of *Anredera cordifolia*, and temperature was found to be the most influential parameter impacting the species' distribution patterns. Additionally, Kaky et al. (2020) assessed the effectiveness of MaxEnt and ensemble modelling approaches in predicting the potential spread of Egyptian medicinal plants. Understanding the distribution trends of invasive alien plants is crucial, as it provides insight into species habitat preferences and the factors influencing their spread. This information is essential for planning clearances, monitoring, and control interventions for these species.

1.2 SIGNIFICANCE OF THE STUDY

The poplar tree is valuable for commercial purposes in many parts of the world, but it is an invasive alien plant in South Africa. Despite significant scholarly interest in studying poplar physiological traits, morphological characteristics, and growth patterns using remote sensing, there is a lack of comprehensive research reviews on how, where, and when remote sensing techniques have been used to study these trees. The current study aimed to fill this gap by analyzing articles that used remote sensing techniques to study poplar species across the world. This review focused on applications, methods, spatiotemporal trends, goals, and indicators used to study poplar. The study critically examined various application trends and identified important research gaps for subsequent studies. The study revealed a significant increase in research on poplar trees that utilizes remote sensing techniques. Among these, airborne LiDAR is the most commonly used platform, followed by satellite and ground-based remote sensing platforms, particularly in countries of the Global North. The findings also showed that only a few studies have focused on poplar trees in mountain and urban areas. While the mentioned remote sensing platforms data have been employed to study poplar trees globally, there is a notable lack of such research in the Global South, including South Africa. This highlights a significant research gap that this study aims to address.

The presence of poplar trees as invasive species in South Africa's delicate landscapes has raised concerns due to their high-water requirements and their ability to thrive near streams and other water sources. This poses a threat to riverine ecosystems by depleting water resources and impacting groundwater levels and surface runoff (Ntshidi et al. 2018). Additionally, they outcompete indigenous vegetation for essential resources (SANParks, 2020). To address this issue, the South African National Parks (SANParks) has been actively removing poplar trees and monitoring new species in its national parks to control their aggressive spread (SANParks, 2020). While the SANParks approach is effective, the extensive areas that need to be cleared make locating and removing the poplar trees costly and laborious. Geospatial techniques have the potential to accurately locate, map and predict the distribution of poplar species and their suitable habitats based on comprehensive environmental-climatic requirements. However, the application of geospatial techniques in locating and mapping poplar in SANParks has not been fully

explored. The current study used machine learning and regression techniques to bridge this research gap by predicting poplar invasions and identifying suitable habitats in South Africa. The study further identified sites that have the potential to host poplar species and Golden Gate Highlands National Park (GGHNP), an entity of SANParks, came out as one. To leverage the knowledge and understanding of poplar drivers in GGHNP, a study as part of this research has been conducted to investigate the environmental variables behind the distribution of poplar species.

An important finding from the current study review is the realization that there is very little research on poplar invasion in mountainous areas using remote sensing across the globe. Meanwhile, mountainous areas are considered hot spots for poplar invasion due to high rainfall. While there is a need to map and quantify poplar invasions and resource use in these areas, it is difficult to achieve accurate results with remote sensing due to shadows resulting from relief and energy differences, as this decreases mapping accuracy, especially when using passive sensor products. Using composite Sentinel-2 imagery and indices with machine learning techniques has shown the capacity to map alien invasive species in mountainous areas effectively. However, this methodological approach was yet to be tested in mountainous areas of South Africa. To address this gap, a combination of Random Forest machine-learning classifier and Sentinel-2 sensor images have been used to discriminate poplar trees against other land cover features in the mountainous landscapes of South Africa and achieved impressive results with above 85% overall accuracy across all phases of the year. The study also identified specific spectral bands and derived indices crucial for distinguishing poplar from other land cover types and highlighted important periods of the year for this study. The study's sound results imply that maps derived from this classification process can be utilized to update the inventory of alien trees in South African mountainous systems. Overall, the research demonstrated the potential of geospatial techniques in effectively discerning, identifying, and predicting the spatial distribution of poplar species.

1.3 RESEARCH AIM AND OBJECTIVES

The primary aim of this research was to explore the capacity of geospatial technologies involving remote sensing analysis for mapping and predicting the distribution and patterns of poplars in South Africa. To achieve this aim, the following specific objectives were formulated:

1. Review and synthesize remote sensing studies of poplars over a 30-year period (1991–2022).
2. Examine the capacity of the Random Forest algorithm and Sentinel-2 images for discriminating poplar species against other land cover types in mountainous regions.
3. Predict the potential suitable habitat of poplar trees using regression and machine learning techniques in South Africa.
4. Identify the local environmental variables which influence the occurrence and distribution of poplar in mountainous areas.

1.4 STUDY REGION

The study was carried out at both national and local spatial scales in South Africa (Figure 1.1). This was due to the availability and compatibility of datasets necessary to fulfil various study objectives. Chapter 4 of the study, which focused on predicting suitable habitats for poplar trees, was conducted at a national scale to complement the spatial resolution of bioclimatic variables. Chapters 3 and 5 were conducted at the local scale in the eastern Free State, where the poplar tree is a significant invasive species. The approach taken in this study involves initially building a model at a local scale and then scaling it up to a national level, with a focus on identifying suitable areas for poplar invasion. As a result, the study area description encompasses aspects of South Africa at large, which is built based on the detailed exploration of the eastern Free State.

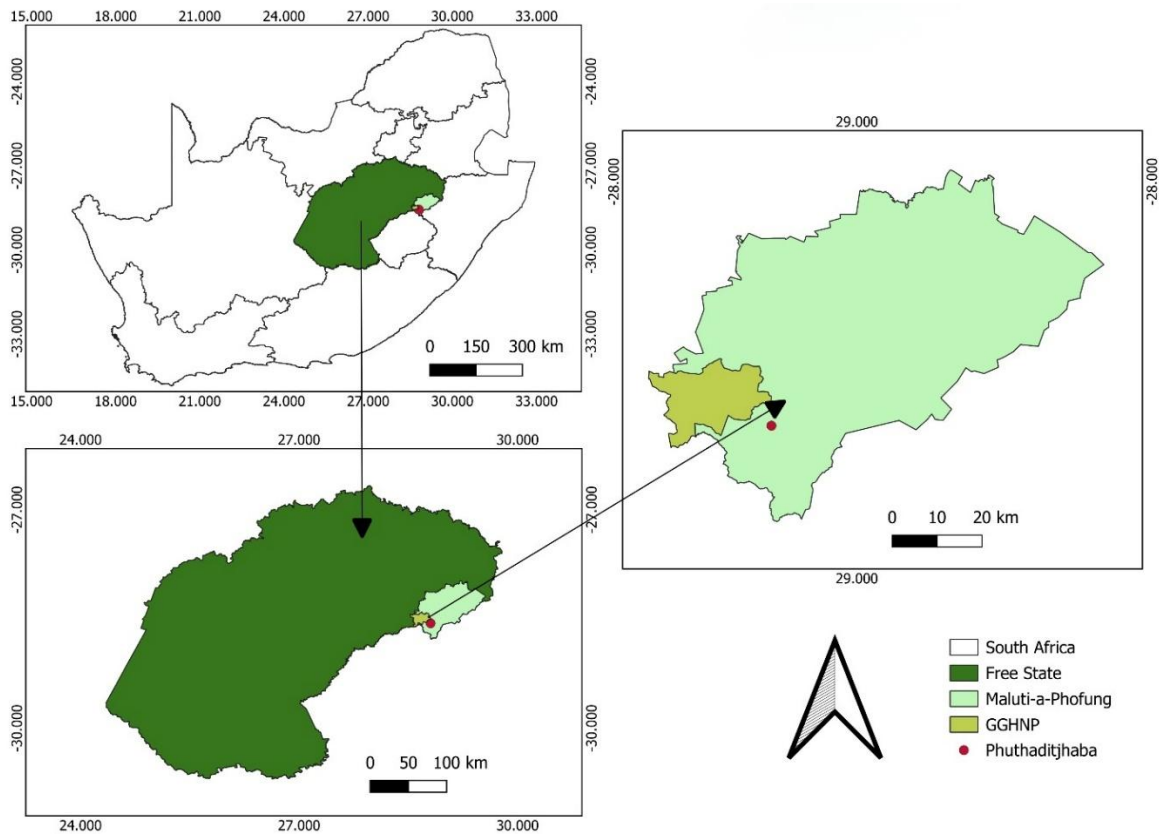


Figure 1.1 Location of the study area in the eastern Free State Province of South Africa

South Africa

South Africa is located in the southern part of the African continent. Its average height above sea level is approximately 1000 meters (Van Wilgen et al., 2020; King, 1942). The country experiences dry and cold winters and humid and warm summers (Blamey et al., 2017). Most South African provinces receive rainfall in summer, while the Western Cape experiences winter rainfall. On average, the country receives around 464 mm of rainfall annually (Roffe et al., 2019). South Africa has 9 vegetation zones, including Succulent Karoo, Albany thicket, Savanna, Forest, Desert, Fynbos, Karoo, Grassland, and Nama-Karoo (Dayaram et al., 2019; Mucina and Rutherford, 2006). The topography of South Africa includes the Great Escarpment and Cape Fold Belt mountains (Haughton, 1969).

Eastern Free State

The eastern Free State is located in the Free State province of South Africa and is characterized by its mountainous terrain, with elevations ranging from approximately 1500 to 3000 meters above sea level (Brand et al. 2013). The region experiences warm-to-hot summers and cool-to-cold winters, with an annual rainfall of 500 to 800 mm (Kay et al. 1993; Mapfumo et al. 2020). Winter brings frost and occasional snowfalls in the higher-lying areas, while summer temperatures average around 35°C and winter temperatures around 5°C. The eastern Free State falls under the Drakensberg and mesic grassland bioregion, featuring wild grasses like *Themeda triandra*, *Eragrostis* species, *Digitaria argyrograpta*, and *Tristachya leucothrix* (Mucina and Rutherford, 2006). Additionally, tree species such as *Acacia karroo*, *Olea europaea*, *Searsia lancea*, and *Populus* can be found in the area (SANParks, 2020). Geologically, the region is underlain by sedimentary rocks of the Beaufort and Karoo supergroups (Rutherford et al. 2015). The watershed and tributaries of the eastern Free State contribute significantly to the Vaal and Orange River systems, which in turn supply water to the economic hub of the country.

1.5 THESIS DESIGN AND OUTLINE

The research design presented in Figure 1.2 shows the framework within which this thesis is constructed. As shown in Figure 1.2, the design of this work divided the thesis into two interrelated geospatial segments. The first is the use of remote sensing applications on poplar, where the systematic review of peer-reviewed studies on this subject is provided (in Chapter 2), and the classification as well as spectral separability in a poplar-invaded landscape is presented (in Chapter 3). The second segment involves the use of machine learning-based models to evaluate possible habitats for invasion by poplar using presence data and environmental variables (Chapter 4), and to explore environmental variables that potentially influence poplar distribution (Chapter 5). The systematic review is undertaken globally to uncover patterns of poplar studies in terms of generally used satellite products and methods, after which the modelling is carried-out in South Africa, to paint a picture of potential areas for poplar invasion. Because of vast areas of possible invasion in the country, which also includes the eastern Free State, the test of influencing variables and the classification were carried out in this region. It is believed that the overall executions of individually interconnected workflows can provide background information on geospatial applications in the country and gain insights from around the world. These are hoped to help advance

conservation efforts of poplar as it is a species of concern (*listed as Category 1b species which essentially mean that it cannot be planted, regrown, or sold; they should be controlled*) in South Africa.

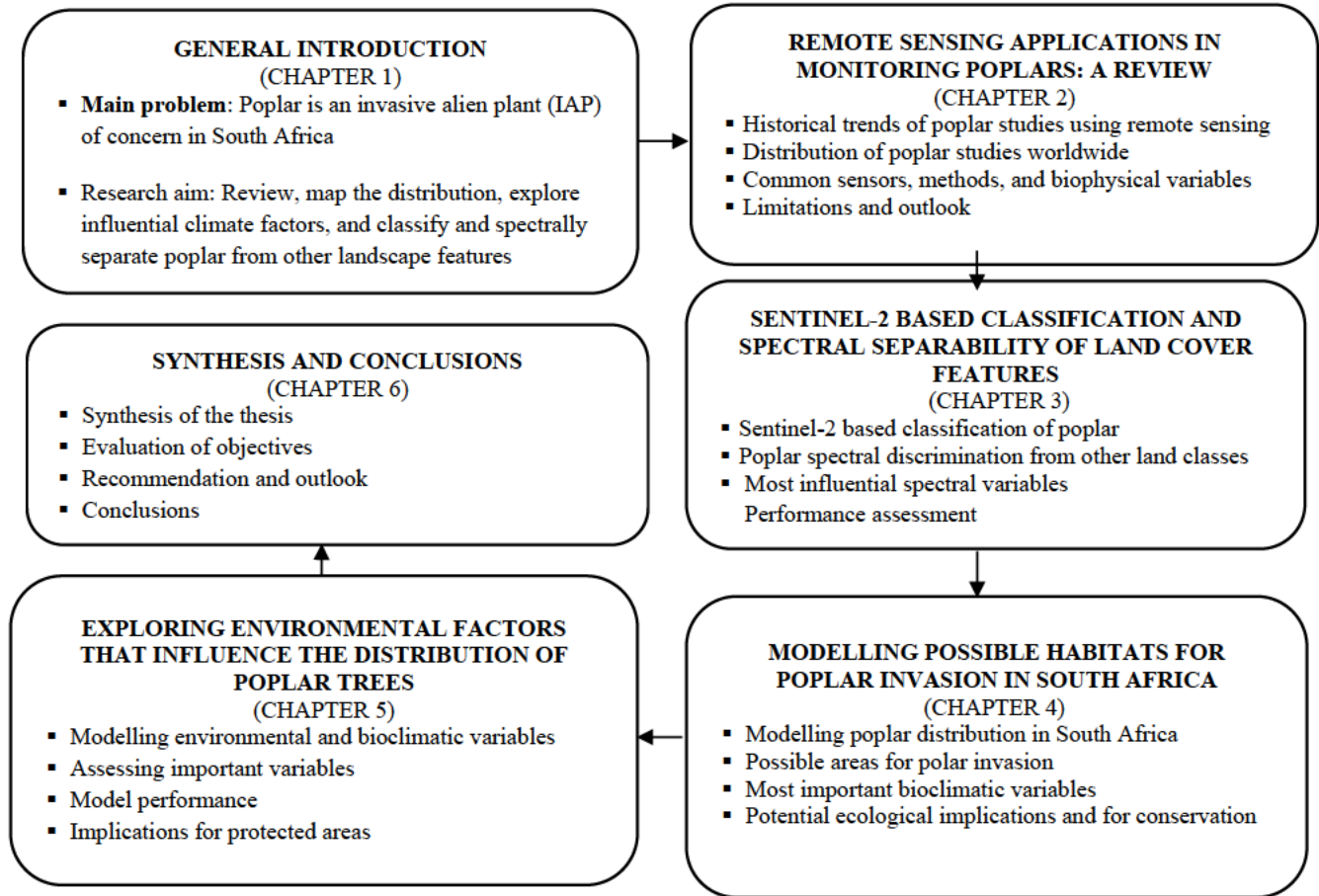


Figure 1.2 Research design

This research is organized into six chapters. Chapters 2 to 5 are independent research publications directly related to the research objectives outlined in section 1.3. Each chapter includes an introduction, methodology, results, discussion, and conclusions. Chapters 1 and 6 provide a general introduction and conclusions for this thesis.

Chapter 2 synthesized and reviewed scientific articles where remote sensing tools and procedures have been applied to quantify the existence and spread of poplar plants across the globe. This chapter offers insights into the methods, tools, spatiotemporal trends, and overall indicators used in these studies

Chapter 3 used Sentinel-2 imagery and the Random Forests algorithm to discern poplar species across the mountainous landscape. This involved distinguishing poplar amongst other land features at various times and identifying essential spectral bands and indices in the classification process. Additionally, the study assessed the accuracy of the classification throughout the study period.

Chapter 4 predicted the existence and distribution of poplar trees in the country using regression and machine learning algorithms. The study further examined the predictive power of each model while evaluating the significant climatic variables for poplar distribution.

Chapter 5 investigated and identified the local environmental parameters that influenced the existence and spread of poplar trees in Golden Gate Highlands National Park (GGHNP). The study further ranked the contributions of the identified variables in influencing the presence and distribution of poplar trees in GGHNP.

1.6 REFERENCES

Andrew, M.E. and Ustin, S.L., 2006. Spectral and physiological uniqueness of perennial pepperweed (*Lepidium latifolium*). *Weed Science*, 54(6), 1051-1062.

Adam, Y., Ngetar, N.S. and Ramdhani, S., 2017. The assessment of invasive alien plant species removal programs using remote sensing and GIS in two selected reserves in the eThekweni Municipality, KwaZulu-Natal. *South African Journal of Geomatics*, 6(1), 90-105.

Blamey, R.C., Middleton, C., Lennard, C. and Reason, C.J.C., 2017. A climatology of potential severe convective environments across South Africa, *Climate Dynamics*, 49, 2161-2178. <https://doi.org/10.1007/s00382-016-3434-7>

Brand, R.F., Du Preez, P.J. and Brown, L.R., 2013. High altitude montane wetland vegetation classification of the Eastern Free State, South Africa. *South African Journal of Botany*, 88, 223-236.

Bley Dalouman, H., Broust, F. and Tran, A., 2023. Combining remote sensing and a geographic information system to map and assess the accessibility of invasive alien species forest stands: Case of *Acacia mearnsii* on Reunion Island. *Forests*, 14(10), 2030.

Calleja, F., Ondiviela, B., Galván, C., Recio, M. and Juanes, J.A., 2019. Mapping estuarine vegetation using satellite imagery: The case of the invasive species *Baccharis halimifolia* at a Natura 2000 site. *Continental Shelf Research*, 174, 35-47.

Clewley, G.D., Eschen, R., Shaw, R.H. and Wright, D.J., 2012. The effectiveness of classical biological control of invasive plants. *Journal of Applied Ecology*, 49(6), 1287-1295.

Danise, T., Innangi, M., Curcio, E., Piccolella, S., Fioretto, A. and Pacifico, S., 2021. White poplar (*Populus alba* L.) leaf waste recovery and intercropping outcome on its polyphenols. *Industrial Crops and Products*, 171, 113866.

Dayaram, A., Harris, L.R., Grobler, B.A., Van Der Merwe, S., Rebelo, A.G., Ward Powrie, L., Vlok, J.H.J., Desmet, P.G., Qabaqaba, M., Hlahane, K.M. and Skowno, A.L., 2019, Vegetation map of South Africa, Lesotho and Swaziland 2018: A description of changes since 2006, *Bothalia-African Biodiversity & Conservation*, 49(1), 1–11. <https://doi.org/10.4102/abc.v49i1.2452>

de Sa, N.C., Carvalho, S., Castro, P., Marchante, E. and Marchante, H., 2017. Using Landsat time series to understand how management and disturbances influence the expansion of an invasive tree. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(7), 3243-3253.

Department of Environmental Affairs., 2018. [online] Available at: <https://www.environment.gov.za/projectsprogrammes/wfw/research#stats> [Accessed 3 May 2018].

Elias, F., Junior, B.H.M., de Oliveira, F.J.M., de Oliveira, J.C.A. and Marimon, B.S., 2019. Soil and topographic variation as a key factor driving the distribution of tree flora in the Amazonia/Cerrado transition. *Acta Oecologica*, 100, 103467.

Elith, J. and Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 677-697.

Görgens, AHM. and Van Wilgen, BW., 2004. Invasive alien plants and water resources: an assessment of current understanding, predictive ability and research challenges. *South African Journal of Science*, 100(1-2), 27-34.

Haughton, S.H., 1969, Geological society of South Africa, Johannesburg, South Africa.

Henderson, L., 2007. Invasive, naturalized and casual alien plants in southern Africa: A summary based on the Southern African Plant Invaders Atlas (SAPIA). *Bothalia*, 37(2), 215-248.

Henderson, L., 2020. Invasive Alien Plants in South Africa. Plant Protection Research Institute Handbook NO 21. Pretoria.

Hill, M.P., Thomson, L.J., Björkman, C. and Niemelä, P., 2015. Species distribution modelling in predicting response to climate change. *Climate Change and Insect Pests*, 16.

Kaky, E., Nolan, V., Alatawi, A. and 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, 101150.

Kattenborn, T., Lopatin, J., Förster, M., Braun, A.C. and Fassnacht, F.E., 2019. UAV data as alternative to field sampling to map woody invasive species based on combined Sentinel-1 and Sentinel-2 data. *Remote sensing of Environment*, 227, 61-73.

Kay, C., Bredenkamp, G.J. and Theron, G.K., 1993. The plant communities of the Golden Gate Highlands National Park in the north-eastern Orange Free State. *South African Journal of Botany*, 59(4), 442-449.

Khayatinejad Aqgonbad, S., Hossein Zadeh, O., Hajjarian, M. and Abdi, M.R., 2022. Impact of key stakeholders' participation in strategic decision-making regarding poplar tree farming development. *Small-scale Forestry*, 1-19.

King, L.C., 1942, South African scenery. A textbook of geomorphology, Oliver and Boyd, London, UK.

- Landmann, T., Dubovyk, O., Ghazaryan, G., Kimani, J. and Abdel-Rahman, E.M., 2020. Wide-area invasive species propagation mapping is possible using phenometric trends. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159, 1-12.
- Liu, C., Jia, X., Bai, X. and Shao, M.A., 2023. Analysis of soil water use by exotic and native vegetation in a semi-arid area and their associated interspecific competition. *Science of the Total Environment*, 905, 167066.
- Luo, J. and Zhou, J.J., 2019. Growth performance, photosynthesis, and root characteristics are associated with nitrogen use efficiency in six poplar species. *Environmental and Experimental Botany*, 164, 40-51.
- Mapfumo, P., Chagwiza, C. and Antwi, M., 2020. Impact of rainfall variability on maize yield in the KwaZulu-Natal, North-West and Free State provinces of South Africa (1987–2017). *Journal of Agribusiness and Rural Development*, 58(4), 359-367.
- Mapuru, M., Xulu, S. and Gebreslasie, M., 2024. Modelling possible habitats for poplar invasion in South Africa. *KOEDOE-African Protected Area Conservation and Science*, 66(1), 1783.
- Mitchell, J.J. and Glenn, N.F., 2009. Subpixel abundance estimates in mixture-tuned matched filtering classifications of leafy spurge (*Euphorbia esula L.*). *International Journal of Remote Sensing*, 30 (23), 6099-6119.
- Mucina, L. and Rutherford, M.C., 2006, The vegetation of South Africa, Lesotho and Swaziland, *Strelitzia* 19, 1-30.
- National Environmental Management: Biodiversity Act, Act 10 of 2004. Department of Environmental Affairs, Pretoria
- Ntshidi, Z., Gush, M.B., Dzikiti, S. and Le Maitre, D.C., 2018. Characterising the water use and hydraulic properties of riparian tree invasions: A case study of *Populus canescens* in South Africa. *Water SA*, 44(2), 328-337.

- Núñez-González, N., Rodríguez, J. and González, L., 2021. Managing the invasive plant *Carpobrotus edulis*: is mechanical control or specialized natural enemy more effective? *Journal of Environmental Management*, 298, 113554.
- Orlovic, S., Guzina, V., Krstic, B. and Merkulov, L., 1998. Genetic variability in anatomical, physiological and growth characteristics of hybrid poplar (*Populus x euramericana* Dode (Guinier)) and eastern cottonwood (*Populus deltoides* Bartr.) clones. *Silvae Genetica*, 47(4), 183-189.
- Rajah, P., Odindi, J., Mutanga, O. and Kiala, Z., 2019. The utility of Sentinel-2 vegetation indices (VIs) and Sentinel-1 synthetic aperture radar (SAR) for invasive alien species detection and mapping. *Nature Conservation*, 35, 41-61.
- Richardson, DM. and Van Wilgen, BW., 2004. Invasive alien plants in South Africa: How well do we understand the ecological impacts? *South African Journal of Science*, 100(1), 45-52.
- Roffe, S.J., Fitchett, J.M. and Curtis, C.J., 2019, Classifying and mapping rainfall seasonality in South Africa: A review, *South African Geographical Journal*, 101(2), 158-174. <https://doi.org/10.1080/03736245.2019.1573151>
- Rutherford, A.B., Rubidge, B.S. and Hancox, P.J., 2015. Sedimentology and palaeontology of the Beaufort Group in the Free State Province supports a reciprocal foreland basin model for the Karoo Supergroup, South Africa. *South African Journal of Geology*, 118(4), 355-372.
- Smith, H.G., Sheridan, G.J., Lane, P.N., Noske, P.J. and Heijnis, H., 2011. Changes to sediment sources following wildfire in a forested upland catchment, southeastern Australia. *Hydrological Processes*, 25(18), 2878-2889.
- South African National Parks (SANParks)., 2020. Golden Gate Highlands National Park Management Plan, SANParks, Pretoria.
- Valero-Jorge, A., González-De Zayas, R., Matos-Pupo, F., Becerra-González, A.L. and Álvarez-Taboada, F., 2024. Mapping and monitoring of the invasive species *Dichrostachys cinerea* (Marabú) in Central Cuba Using Landsat Imagery and Machine Learning (1994-2022). *Remote Sensing*, 16(5), 798.

- van Wilgen, B., Richardson, D. and Higgins, S.I., 2001. Integrated control of invasive alien plants in terrestrial ecosystems. *Land Use and Water Resources Research*, 1(1), 1-6.
- Van Wilgen, B.W., Measey, J., Richardson, D.M., Wilson, J.R. and Zengeya, T.A., 2020. *Biological invasions in South Africa*. Springer Nature, Knoxville, TN. 975.
- Vega, J.A., Arellano-Pérez, S., Álvarez-González, J.G., Fernández, C., Jiménez, E., Fernández-Alonso, J.M., Vega-Nieva, D.J., Briones-Herrera, C., Alonso-Rego, C., Fontúrbel, T. and Ruiz-González, A.D., 2022. Modelling aboveground biomass and fuel load components at stand level in shrub communities in NW Spain. *Forest Ecology and Management*, 505, 119926.
- Versveld, DB. and Van Wilgen BW., 1986. Impact of woody aliens on ecosystem properties. In MacDonals IAW., Kruger FJ and Ferrar AA (eds.) *The Ecology and Control of Biological Invasions in South Africa*. Oxford University Press, Cape Town, 239-246.
- Wallace, C.S., Walker, J.J., Skirvin, S.M., Patrick-Birdwell, C., Weltzin, J.F. and Raichle, H., 2016. Mapping presence and predicting phenological status of invasive buffelgrass in southern Arizona using MODIS, climate and citizen science observation data. *Remote Sensing*, 8(7), 524.
- Wilson, J.R., Panetta, F.D. and Lindgren, C., 2017. *Detecting and responding to alien plant incursions*. Cambridge University Press.
- World Wildlife Fund (WWF)., 2016. *Water: Rethinking South Africa's Water Future*. Cape Town, WWF-SA Report.
- Zagajewski, B., Kluczek, M., Zdunek, K.B. and Holland, D., 2024. Sentinel-2 versus PlanetScope Images for Goldenrod Invasive Plant Species Mapping. *Remote Sensing*, 16(4), 636.
- Zhang, X., Wei, H., Zhao, Z., Liu, J., Zhang, Q., Zhang, X. and Gu, W., 2020. The global potential distribution of invasive plants: *Anredera cordifolia* under climate change and human activity based on random forest models. *Sustainability*, 12(4), 1491.

Zhang, Y., Thenkabail, P.S. and Wang, P., 2019. A bibliometric profile of the remote sensing open access journal published by MDPI between 2009 and 2018. *Remote Sensing*, 11(1), 91.

CHAPTER 2 REMOTE SENSING APPLICATIONS IN MONITORING POPLARS: A REVIEW



Review

Remote Sensing Applications in Monitoring Poplars: A Review

Morena J. Mapuru ^{1,2,*}, Sifiso Xulu ¹ and Michael Gebreslasie ¹

¹ School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4000, South Africa; xulus4@ukzn.ac.za (S.X.); gebreslasie@ukzn.ac.za (M.G.)

² Scientific Services, South African National Parks, Kimberley 8301, South Africa

* Correspondence: johnmapuru97@gmail.com; Tel.: +27-031-260-7656

Link: <https://www.mdpi.com/1999-4907/14/12/2301>

This chapter is based on:

Mapuru, M.J., Xulu, S. and Gebreslasie, M., 2023. Remote Sensing Applications in Monitoring Poplars: A Review. *Forests*, 14(12), p.2301.

ABSTRACT

Given the ability of remote sensing to detect distinctive plant traits, it has emerged in recent decades as a useful and attractive research tool for forest trees such as poplars. Although poplars have been extensively studied using remote sensing over the past thirty years, no reviews have been conducted to understand the results of multiple applications. Here, we present a review and synthesis of poplar studies in this regard. We searched the Scopus, Google Scholar, and Science Direct databases and found 266 published articles, of which 148 were eligible and analysed. Our results show a rapid increase in remote sensing-based poplar publications over the period of 1991–2022, with airborne platforms, particularly LiDAR, being predominantly used, followed by satellite and ground-based sensors. Studies are widespread in the Global North, accounting for more than two-thirds of studies. The studies took place mainly in agricultural landscapes, followed by forest areas and riparian areas, with a few in mountain and urban areas. Commonly studied biophysical parameters were mostly obtained from LiDAR data. On the other hand, spectral indicators have been widely used to monitor the health and vitality of poplar trees, integrating various machine learning algorithms. Overall, remote sensing has been widely used in poplar studies, and the increasing use of free satellite data and processing platforms is expected to pave the way for data-poor countries to monitor poplar in the Global South, where resources are mainly limited.

Keywords: poplar; remote sensing; LiDAR; review; forest

2.1 INTRODUCTION

Ongoing developments in the field of Earth Observation have enabled the study of unique characteristics of various tree species, including the commercially important poplar (*Populus* spp.). Although this subject has attracted the attention of many scholars, reviews of the literature characterizing this research remain a gap that this study attempts to fill. Poplar is one of the most valuable trees for industrial production due to its desirable characteristics such as rapid growth, hybridization tendency, and adaptability to different environmental conditions (Khayatinejad Aqgonbad et al. 2022). Its high ecological value contributes toward the restoration of degraded landscapes through seed trapping (increasing surface roughness), soil stabilization, erosion control, and nutrient cycle improvements (Zhang et al. 2019). While poplar occurs in many countries as a native species, it has been established in many other places as an invasive alien species (IAS) (Kota et al. 2007; Sikorska et al. 2019). IAS are plants introduced mainly by human activities outside their natural environment that can colonize natural ecosystems, with varying ecological and socioeconomic consequences (Richardson and Rejmánek, 2011). Their presence and spread seriously threaten local biodiversity and are exacerbated by climate change and limited mechanisms to address it (González-Moreno et al. 2013). Typically, invasive alien species have competitive traits that allow them to colonize and displace native ecosystems (Caudullo and De Rigo, 2016). Poplar is native to regions such as Asia, Europe, and North America (Danise et al. 2021), and invasive in many African countries.

In countries like South Africa, poplar species are common in many landscapes and are only legally permitted in certain areas under controlled conditions (Henderson, 2007). Dense stands of cottonwood formed through rooting can narrow and block water channels, leading to flooding and increased siltation (Stettler, 2011). Their extensive stocks could lead to a significant reduction in the flow of the stream. Therefore, understanding the location and extent of poplar species is crucial for biodiversity conservation and land management. While poplar species are measurable in labor-intensive and costly field studies, their detection through remote sensing technology seems a practical option.

Remote sensing has emerged as a crucial data source for the retrieval and analysis of forest information because it offers a spatially continuous, highly consistent representation of the Earth's surface (Lechner

et al. 2020). With improvements in sensor technology and Open Data policies, as well as near real-time image acquisition, remote sensing can provide useful information about the Earth's surface at a lower cost than traditional field survey methods (Szuster et al. 2011). These advances, particularly the increasing availability of free satellite products, have expanded the options for affordable, efficient, and effective means of addressing many biodiversity monitoring challenges (Turner et al. 2003). Since its inception, the remote detection of poplar species from surrounding vegetation communities has been successful and has since become an active area of research (Tonbul et al. 2020). For example, Eslami and Zahedi (2011) identified poplar-cultivated landscapes in Iran using Indian remote sensing satellite data with an overall classification accuracy of 91.48%. Hamrouni et al. (2021) used Sentinel-2 time series to identify poplar plantations at the local scale with an accuracy of 89.5% to 99.3%. D'Amico et al. (2021) used multilayer perceptron and traditional logistic regression (deep learning approach) to map poplar plantations in Italy with Sentinel-2 and achieved an overall accuracy of 91%. Laslier et al. (2019) used LiDAR data to map poplars in riparian zones with an accuracy of 80%. These studies demonstrate high remote sensing capacity with growing interest in poplar research; however, their spatiotemporal patterns, commonly used remote sensing products, and methods remain poorly understood, and there is a need to collect evidence on the characteristics of all studies in this regard.

In this study, peer-reviewed articles related to remote sensing approaches aiming to detect the presence and distribution of poplar species in different regions in the world were analyzed. This study provides the first literature review of poplar remote sensing and sheds light on the spatiotemporal patterns, application goals, methods, and indicators used. Based on these results, we provide recommendations for future studies.

2.2 MATERIALS AND METHODS

2.2.1 Database search process

The study searched for peer-reviewed publications that have applied remote sensing approaches to monitoring poplar species using the keywords: Remote Sensing OR Earth Observation OR NDVI OR Landsat OR LiDAR OR Sentinel OR SAR OR MODIS AND poplar OR *Populus alba* OR *Populus nigra* OR White poplar OR Black poplar from the Scopus, Google Scholar, and Science Direct databases. In

this review, the study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines and refer the reader to Moher et al. (2015) for more information. The PRISMA was selected because of its popularity as an evidence-based and the most methodologically justified approach for generating systematic reports and meta-analyses (Javanmardi et al. 2020).

2.2.2 Inclusion/exclusion criteria

The search returned 266 articles, 148 of which met the criteria listed in Table 1. For emphasis, the study excluded all articles that used (did not use) remote sensing on non-poplar species (on poplar species), fugitive types of literature such as unpublished theses and reports from governments and institutions, and studies not published in English (Table 2.1). The study also provides a procedure for the article search, selection, and analysis process based on the PRISMA model in Figure 2.1.

Table 2.1 Inclusion and exclusion criteria for the review.

Inclusion Criteria	Exclusion Criteria
Studies conducted on poplar species	Studies conducted without poplar species
Remote sensing-based data sources	Studies without the use of remote sensing data
Peer-reviewed published articles	Fugitive literature and review articles
Articles published in English	Articles not written in English

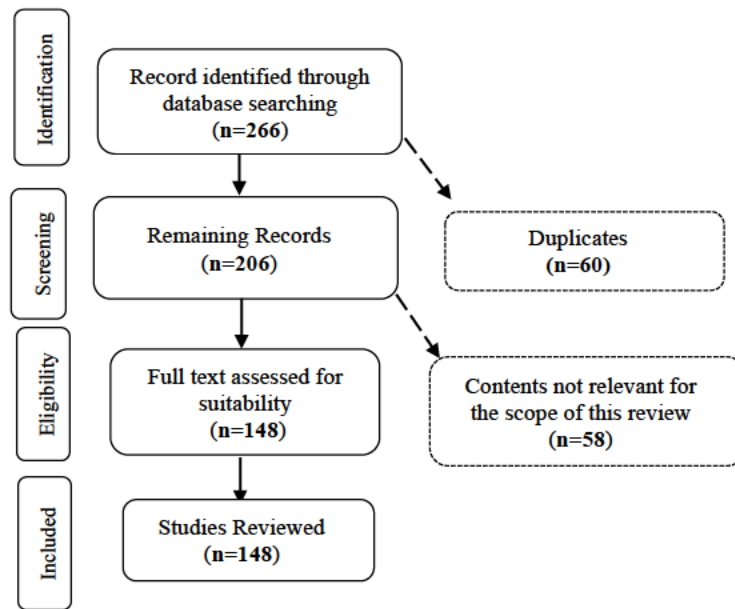


Figure 2.1 A flowchart of the literature search process based on PRISMA.

2.2.3 Data analysis methods

The study compiled and classified publications into categorical variables, including publication year, study area, location, sensor platform and type, application purpose, biophysical parameters, classification algorithms, software, first author, and the journal in which each study was published. For each publication, the study differentiated whether the poplar was studied as a native or non-native species. The study also used the VOS viewer platform (van Eck and Waltman, 2010) to create title/abstract network visualization maps, as well as ESRI ArcMap v 10.8.1 (ESRI, Redlands, CA, USA) and R studio v 4.1.3 (R Foundation for Statistical Computing, Vienna, Austria) (Team, 2010) for visualization and analysis.

2.3 RESULTS AND DISCUSSION

To begin the analysis of this research, the study starts off by describing the historical trend and patterns of eligible studies for this review, followed by a presentation of all sensor products used in the poplar studies. The study delves deeper and present the geographical distribution of the articles and their characteristics in terms of methods, materials, and journal selection.

2.3.1 Historical research trends

This section presents the temporal research pattern of 148 articles on poplar remote sensing published between 1991 and 2022 (Figure 2.2). Based on the literature search, the number of publications showed a wavy increase over this period, with most studies (67) using airborne sensors, followed by satellite (60) and terrestrial platforms (21). The first publication in 1991 was based on an airborne sensor, and since then, its use has steadily increased with satellites and, to a lesser extent, ground-based sensors. Ongoing advances in sensor technology continue to refine satellite-to-ground remote sensing products, and the increasing trend in research publications shown in Figure 2.2 is not surprising.

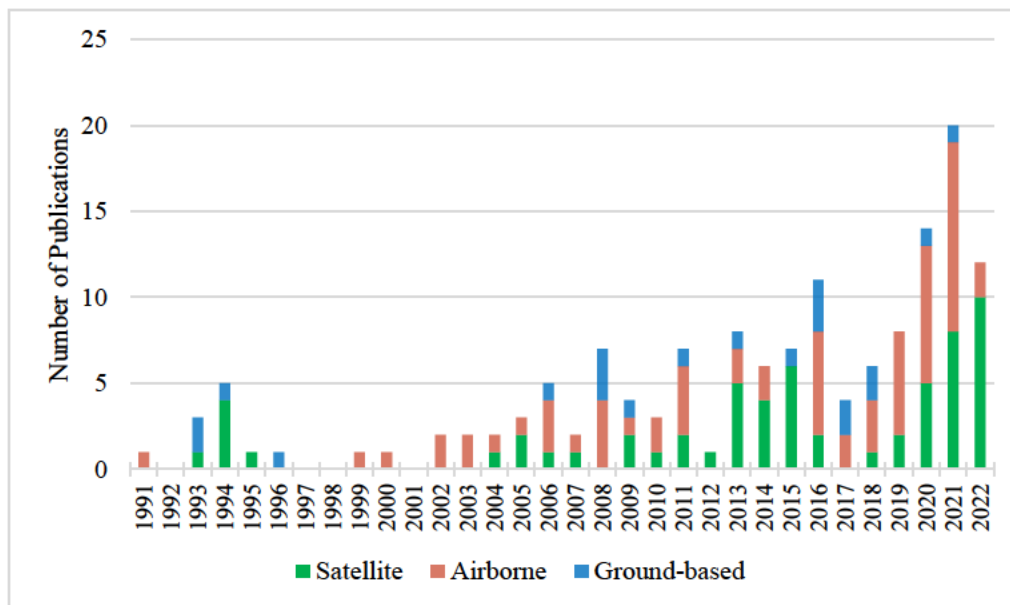


Figure 2.2 Publication statistics in remote sensing of poplar species (1991–2022).

Airborne remote sensing platforms, particularly LiDAR data, have been successfully used in many poplar studies (42 publications). Since the first publication in 1999, poplar studies using LiDAR data have shown an increasing trend, particularly since 2006 (Figure 2.3). The choice of LiDAR is mainly based on its ability to capture three-dimensional (3D) information on forest features, which is useful for characterizing species such as poplars (Man et al. 2020). In addition, this product has been used in studies to overcome saturation problems in dense forests because it transmits and receives laser pulses in a relatively short time and still provides extensive₂₃data from which biophysical information can be

obtained (Chen et al. 2021). Studies such as Xu et al. (2021) and Man et al. (2020) found LiDAR to be useful for discerning tree species in heterogeneous dense forests and inaccessible landscapes where poplars occur. These unique LiDAR properties make it more attractive than other high-resolution sensors for challenging forest stands (Wang et al. 2018; Marrs and Ni-Meister, 2019; Khorrami et al. 2022). For modeling forest properties, variables are typically extracted from airborne LiDAR data using two common methods (Lu et al. 2015): (a) extracting height metrics, density metrics, and forest profile features directly from point clouds (Pirotti, 2010; Duncanson et al. 2015; Cao et al. 2019; Gao and Zhang, 2021) and (b) extracting statistics from canopy height model (CHM) data, which is the distinction between a digital surface model (DSM) and a digital terrain model (DTM) derived from LiDAR data (Leite et al., 2020; Lin et al., 2022).

Despite these contributions, LiDAR data are cost-prohibitive, thus limiting its use in nations with limited resources. This could justify greater application in developed countries where resources are readily available. Unmanned aerial vehicles (UAVs) or drones have also increasingly been used to monitor poplar species since 2017 (Capolupo et al. 2018). While UAVs provide accurate information, the timing of data acquisition is crucial for analysis since the acquisition is usually made upon request; the data may only be useful if collected when the targeted plant species is distinct from its background and neighboring areas (Huang and Asner, 2009). Therefore, to ensure accuracy in the identification and detection of poplar, the phenology of the species needs to be taken into consideration in the surveys conducted when there is a clear distinction between the target species and other species features.

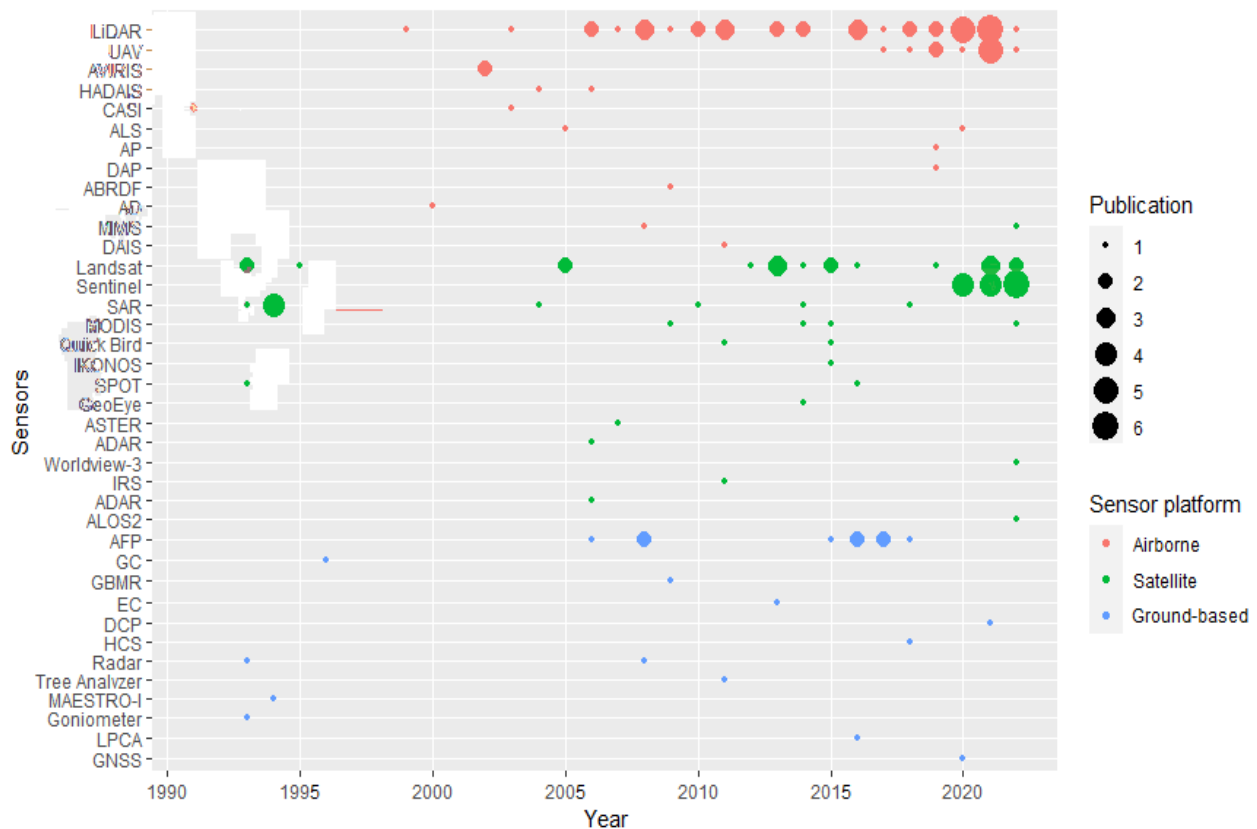


Figure 2.3 Remote sensing sensors used in the study of poplar species per year between 1991 and 2022.

Satellite products are arguably the most valuable remote sensing systems due to their consistent data collection, wider coverage, and lower cost than any other means, and are increasingly being used in poplar studies. In particular, Landsat data is the most widely used satellite product in poplar studies, with 23 publications, most of which were published after the 2008 Open Data directive. The Landsat dataset was primarily applied in agricultural and natural forested spaces, where there was a need for a sensor that has a high spatial, temporal, and spectral resolution (Wang et al. 2018). Sentinel-1 (5) and 2 (9) sensor applications also made essential contributions to mapping and quantifying poplar plantations through supervised (object-based classifier) and unsupervised (vegetation indices) remote sensing approaches (Tonbul et al. 2020; D’Amico et al. 2021). Typically, Sentinel sensors possess better data-capturing characteristics compared to Landsat, i.e., a spatial resolution of 10 m (visible and near-infrared bands), 20 m (red-edge and shortwave infrared bands), and 60 m (atmospheric correction bands) with a five-day temporal resolution (Tonbul et al. 2020). However, two things could have slowed their application over time; firstly, Sentinel sensor data became freely available to the public between 2014 and 2017,

which is the period after USGS established the Landsat open policy; secondly, Sentinel does not capture data in the thermal infrared band, which is the electromagnetic section that assists in analyzing vegetation stress and health. Be that as it may, Landsat and Sentinel sensors seem the most preferred due to their spatial coverage, making them efficient in scanning large areas without saturation.

This is important for updating forest inventory and quantifying poplar production rates in agricultural sectors. Commercial satellite products such as QuickBird, WorldView-3, GeoEye, SPOT, and IKONOS have also been successfully used in poplar research. The introduction of the Google Earth Engine has also enabled easy access to various sensor datasets, thereby increasing the possibility of conducting more vegetation studies through remotely sensed data. In particular, Landsat data are the most widely used satellite product in poplar studies, with 23 publications, most of which were published after the 2008 Open Data directive. The entire Landsat archive has since become largely open, allowing users to perform location-specific image analysis and high-density time-series analysis at different time scales since the early 1970s. Sentinel had 14 publications (Sentinel-1 (5) and Sentinel-2 (9)), and the latter provides better spatial (10 m), spectral (13 bands), and temporal (3–5 days) resolution data compared to Landsat. The Sentinel-1 satellites provide a combination of high spatial and temporal resolution of dual-polarization SAR data. Its dense time series enables the derivation and analysis of temporally filtered annual backscatter signals. The proliferation of free optical satellite imagery (FOSI) has opened possibilities for all users to monitor poplar at different scales. The introduction of the Google Earth Engine also enabled easy access to various sensor datasets, thereby increasing the possibility of conducting more vegetation studies through remotely sensed data.

Ground-based sensors (including in situ and hand-held), which are similar to airborne systems, are mostly operated under human supervision. Relatively few studies (21) have used these products for poplar research. These products are well known for having better spectral and spatial resolution than most space- and airborne sensors; however, most are costly, which limits the scope of their application in vegetation studies. Most of these imaging spectrometers (also known as hyperspectral imagers) acquire a continuous spectrum over a wide range of visible and/or short wavelengths, often using hundreds of spectral bands with narrow spectral intervals. This makes them better options for poplar studies as they can correctly

delineate floral features with greater fidelity, but their applications are still inadequate because of high costs associated with their access for use. Some of the hyperspectral sensors used in poplar studies include, but are not limited to, tree analyzer, Goniometer, and MAESTRO-1. The main benefit of using ground-based hyperspectral sensors is that detailed spectral profiles can be developed for native and non-native plants and spectral regions most sensitive to an abundance of the species of interest (Huang and Asner, 2009).

2.3.2 Geographical distribution of poplar remote sensing studies

As previously stated, the study search results yielded 148 peer-reviewed articles, confirming that scientists are finding remote sensing data for monitoring poplars. These studies were geographically distributed in over 20 countries, where poplar is either native or invasive, and their wide distribution is shown in Figure 2.4. Most studies were conducted in China (34), followed by the United States of America (22), Italy (21), France (13), and Canada (12), accounting for over two-thirds of the studies. The remaining publications came from Sweden (6), Australia (5), Belgium (5), Turkey (5), Germany (4), the Netherlands (4), Argentina (3), Hungary (3), Spain (3), India (2), and one each from the United Kingdom, Greece, Iran, Japan, and Chile (Figure 2.4). In the United States, Australia, India, the United Kingdom, and Iran, poplar is classified as invasive.

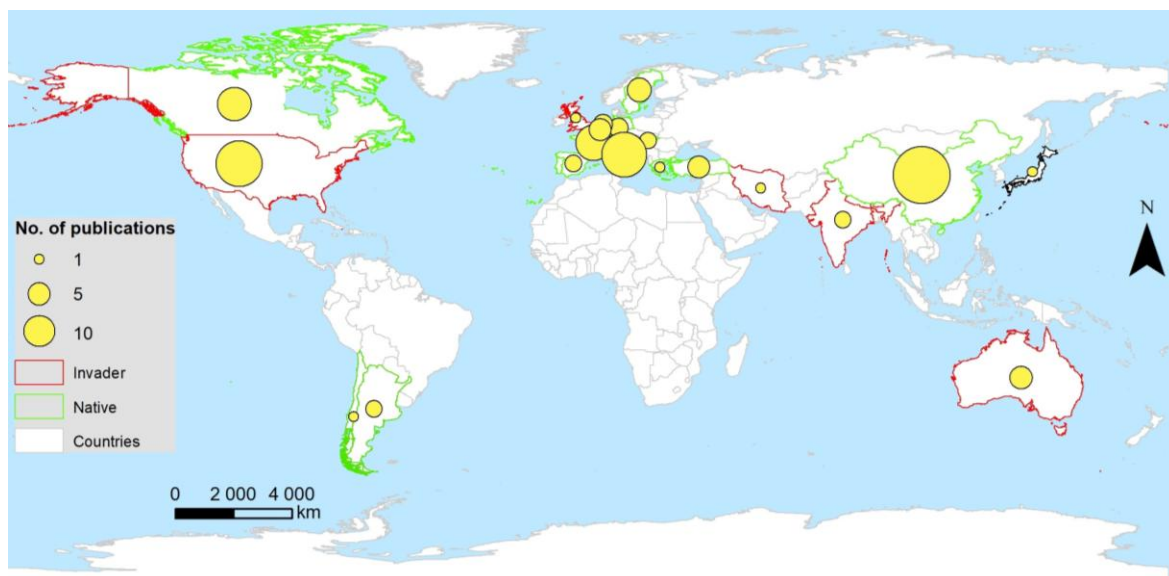


Figure 2.4 Countries that participated in poplar studies and their number of publication output.

For emphasis, note that despite reported cases in countries like South Africa, where it is classified as a Category 2 (occurring throughout/in part of South Africa and allowed only in specified areas under controlled conditions) weed under the National Environmental Management and Biodiversity Act (NEMBA) (No. 10 of 2004) and the Conservation of Agricultural Resources Act (CARA) (No. 43 of 1983) of South Africa's Invasive Species Legislation (Henderson, 2007), no studies have been conducted on the African continent. This underscores the need for rigorous research, particularly when the studied species is invading (e.g., South Africa and Africa). It is believed that perhaps the reason for this is that African countries are somewhat behind in terms of the use of advanced technologies (remote sensing), and insufficient funds to purchase the necessary equipment to conduct such research could delay progress. Be that as it may, the content of this section is intended to help researchers from different walks of life (including Africa) to know in which countries the literature should be consulted to conduct higher-quality studies on the use of remote sensing on poplar species.

Here, the results of the study seem to suggest that the applications of remote sensing for poplar species were more frequent in countries where they are native species than alien invaders. A similar trend was also observed in release rates, with most releases considering the species surveyed as native rather than invasive species. While poplar has a total of 35 species, the study does not consider their classifications.

The study results indicate that poplar remote sensing studies primarily occurred in agricultural landscapes (58), followed by forested areas (47), riparian zones (39), and to a lesser extent, in mountain and urban landscapes (2 each) (Figure 2.5). In agricultural spaces, poplar was mainly studied to understand its phenological changes, carbon sequestration capacity, gross primary productivity rate, spatial growth, and canopy and leaf water content, through various remote sensing algorithms and indices (Colombo et al. 2008; Lu et al. 2015; Barnes et al. 2017; Huang et al. 2021; Colkesen et al. 2022). For instance, Maleki et al. (2020) investigated the ideal vegetation index (VIs) for describing the phenology and interannual variability of the gross primary productivity (GPP) of poplar plantations. The results of the study alluded to the medium-resolution imaging spectrometer (MERIS) terrestrial chlorophyll index (TCI) being the best fit with GPP phenology. To quantify the holistic growth of poplar trees in agricultural space, indicators such as tree volume, height, stem, and crown cover were estimated (Chianucci et al. 2020).

The determination of poplar cultivated fields and mapping of their spatial sites play a vital role in enabling decision-makers and planners to enhance poplar trees' economic and ecological value. This review discovered that poplar plantations were mapped and classified using classification techniques such as object-based classifier (Tonbul et al. 2020). Colombo et al. (2008) investigated the applicability of empirical and radiative transfer models to estimate water content at leaf and landscape levels. The results arising from this study indicated that hyperspectral regression indices are good for estimating water content at both leaf and landscape levels. Remote sensing mapping is a vital activity in forest plantations as it helps to quantify timber stock procurement potential, estimate volume and profit, as well as update forest inventories. This is important as it helps in meeting supply demands.

In natural forests, poplar species were mapped and characterised as part of a forest with other tree species (87%) and individual trees (13%). For instance, studies by (Bonan, 1993; Zande et al. 2008; Santi et al. 2009; Ko et al. 2016) used remote sensing and machine learning techniques to enhance the description and identification of poplar with other trees, while studies by Moffiet et al. (2005), Warner et al. (2006), and Brandtberg (2007) studied poplar individualized as a tree. Like how poplar plantations were studied in agricultural landscapes, poplars in natural forest together with other trees were also investigated for their capacity with regard to carbon intake, water use, seasonal pollen dynamics, and seasonal fire severity using various remote sensing techniques (Bonan, 1993; Rupasinghe and Chow-Fraser, 2021; Li et al. 2022).

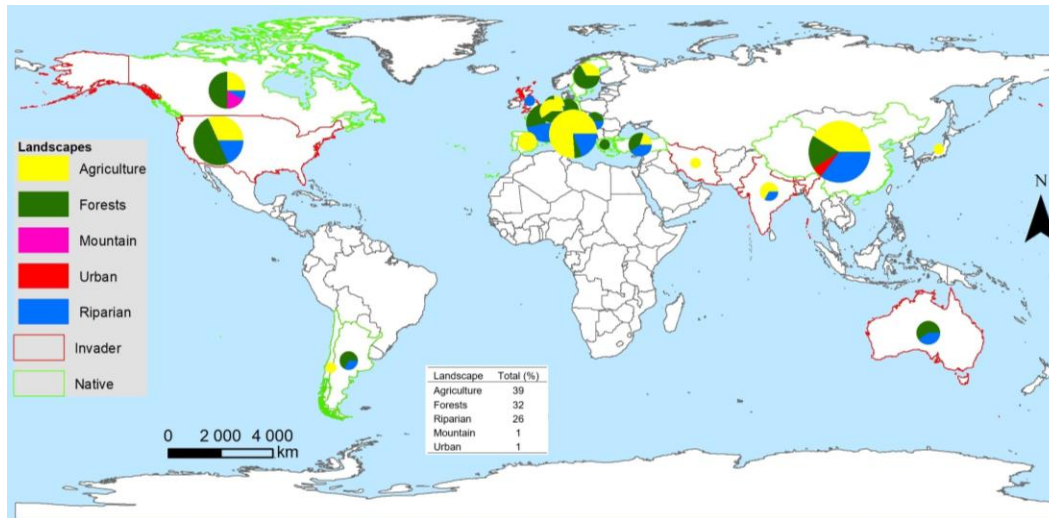


Figure 2.5 Landscape categories where poplar studies were conducted.

In mountainous regions, poplar was quantified using leaf area index (LAI) as an important agent for canopy structure, as well as biomass, carbon, and energy exchange estimation indices (Hall et al. 2003), while in urban landscapes, it was mostly studied to quantitatively relate leaf chlorophyll content (LCC) with other vegetation indices (Schweier et al. 2016; Hamrouni et al. 2022). The reason for the low number of studies in the mountainous landscape could be due to the nature of mountains coupled with inconsistent seasonal species changes, which affect the efficacy of some remote sensing sensors to accurately detect and discriminate floral species (Adagbasa et al. 2022). Steep mountainous terrain can influence the accuracy of the information derived from remote sensing data (Adepoju and Adelabu, 2020). These data are influenced by differences in relief and radiance energy, resulting in shadows, which decrease mapping accuracy. While this suggests an important niche to be filled in science, it is important to utilize proper remote sensing techniques and sensors to accurately delineate the heterogeneity associated with mountainous biota.

2.3.4 Characteristics of methods and materials

One of the most important tasks in remote sensing data analysis is the use of classification algorithms and specific indicators to improve the accuracy of feature detection and characterization, and numerous software programs have been developed for this purpose. The study results are shown for the top 10 dominant biophysical indicators (a), spectral indices (b), classifiers (c), and common software packages

(d) used in remote sensing studies on poplars (see Table 2). Based on the results, the most studied biophysical parameters are related to forest aboveground biomass (AGB) and growing stock volume (GSV) and include poplar growth (17), biomass (14), crown cover (7), and height (4) (Table 2.2), all of which are retrieved through the use of LiDAR data. These proved to be essential features used in various studies to discern species distribution, cover, diversity, and composition with higher accuracy. Traditionally, all these parameters were estimated through allometric equations using tree attributes collected from the field. The development and application of this equation based on destructive sampling are highly time consuming and expensive. Plantations with fast-growing species such as poplar require frequent monitoring over time without having to resort to destructive approaches, and this is possible with LiDAR applications. LiDAR is an active remote sensing product that accurately measures the energy returned from an object of interest within a short span of time at both the stand and tree scale.

On the other hand, indicators such as leaf area index (LAI; 22), normalized difference vegetation index (NDVI; 30), diversity index (DI; 15), green-red normalized difference vegetation index (GRNDVI; 5), and poplar tree index (PTI; 1) were mostly retrieved from spectral bands and their combinations in the electromagnetic spectrum (Table 2). These unsupervised remote sensing techniques were applied primarily through a method testing approach in poplar studies, i.e., testing different vegetation indices in discriminating poplars at different levels or against other vegetation features (Lu et al. 2015; Cañete-Salinas et al. 2020; Colkesen et al. 2022). For instance, Simovic et al. (2022) compared NDVI, NDRE, and EVI for detecting poplar phenology, and NDVI and NDRE were more successful than EVI in distinguishing the latter. However, we also found that NDVI is the most widely used unsupervised classification technique in many vegetation studies; however, we have found that it cannot be used in isolation to classify poplar occurrences. Therefore, the inclusion of an in situ dataset proved that there is a need for a supervised methodology to improve discrimination standards (Huang and Asner, 2009; Baird et al. 2012). On the other hand, in poplar plantations where LAI is extremely high, the blue wavelength can be used to improve the accuracy of NDVI, as this corrects for soil background signals and atmospheric correction. Equivocally, the newly invented PTI also proved to be more effective than NDVI in characterizing poplar from other trees; however, it has only been used in poplar plantations and not in

other landscapes where poplar occurs. We therefore recommend that it is tested in natural forests, especially occurring in mountainous landscapes where poplar is under-studied using remote sensing.

Table 2.2 Top 10 biophysical indicators (a), spectral indices (b), classifiers (c), and common software packages used in the reviewed studies.

Biophysical Indicators	N	Spectral Indices	N
Canopy cover	24	Normalized difference vegetation index (NDVI)	30
Poplar growth	20	Leaf area index (LAI)	22
Poplar biomass	18	Diversity index	15
Poplar distribution	15	Soil-adjusted vegetation index (SAVI)	7
Poplar cover	10	Chlorophyll vegetation index (CVI)	6
Poplar volume	8	Normalized difference water index (NDWI)	6
Breast diameter	7	Difference normalized burn ratio index (dNBRI)	5
Crown cover	7	Green-red normalized difference vegetation index (GR-NDVI)	5
Crown diameter	6	Photochemical reflectance index (PRI)	3
Canopy structure	5	Stress susceptibility index (SSI)	2
Classifiers	N	Software	N
Random forest	24	MATLAB v 5.3.1, 6.0. 7.1, 7.9, 8, 9.2 and 9.6 (MathWorks, Natick, MA, USA)	17
Support vector machine	15	R studio v 2.7.1, 3.0.2, 3.5.0 and 4.1.0 (R Foundation for Statistical Computing, Vienna, Austria)	14
Maximum likelihood	10	Python v 3.3.3, 3.6.7, 3.5.8 and 3.10.9 (Altoros labs, California, USA)	13
Decision tree	7	ArcMap v 10.6.1 (ESRI, Redlands, CA, USA)	9
Classification and regression Tree	6	ENVI v 5.0 (NV5 Geospatial, Hollywood, USA)	9
Object-based	6	LiDAR 360 v 4.0, 5.0 and 6.0 (Green valley International, Berkeley, California, USA)	8
Neural network	5	ENVI v 5.0 and 5.1 (NV5 Geospatial, Hollywood, USA)	8
k-nearest neighbor	4	ArcGIS 10.7 (ESRI, Redlands, CA, USA)	7
Density-based clustering	3	QGIS v 3.14, 3.18 and 3.30 (GeoApt LLC, Alaska, USA)	7
Rotation forest	3	SPSS v 22 and 25 (IBM Corp., Armonk, NY, USA)	6

The study discovered an increase in studies that used classifiers, such as Maximum likelihood (ML; 10), random forest (RF; 24), support vector machine (SVM; 15), decision tree (DT; 7), and object-based image analysis (OBIA; 6) (Table 2.2), to discriminate poplar plants from other traits (Mansfield et al. 2011; Linhui et al. 2021; Gao et al. 2022). Random forest has many exciting properties, such as high accuracy, robustness against overfitting of training data, and integrated measures of variable importance (Adam et al. 2014). Hamrouni et al. (2021) and Zhu et al. (2019) mapped poplar plantations and achieved an overall accuracy of 89.5% and 85%, respectively, using random forests. Similarly, Laslier et al. (2019) worked on identifying riparian tree genera based on leaf-on/leaf-out using a random forest and support vector machine and achieved an overall accuracy of 83.15%.

Decision tree and maximum likelihood in poplar studies were mainly used to classify the population dynamics of poplar species in different ecosystems. For example, Gao et al. (2022) applied the decision tree algorithm to study poplar clones and achieved more than 93% overall accuracy. Petráš et al. (2010) estimated the diameter of poplar clones using the maximum likelihood technique. The rate and quality of the outputs of object-based classification (Tonbul et al. 2020), neural network (Ceylan, 2008) and k-nearest neighbor (Franco-Lopez et al. 2001) signify great impact application in studying poplar. Although not yet used in poplar studies, spectral angle mapper classifiers are also associated with better discrimination of floral invasive species (Shafri et al. 2007); hence, they are also recommended for future use. Many of these computations were performed using Matlab (17), Python (13), ArcMap (9), RStudio (14), ENVI (version 5.0) (9), and other software (Table 2.2).

2.3.5 Co-occurrence words/keywords

From the 148 published journal articles reviewed/synthesized in this study, four main research areas were discovered that produced poplar studies, and these included forest ecology (green), vegetation mapping (yellow-green), machine learning (ML) (red), and vegetation physiology (blue). All of these research areas are interlinked. Mapping the frequency of common occurrences in the title/abstract fields of remote sensing application documents in the poplar literature with a minimum occurrence of five yielded 57 items distributed in four clusters representing four main research themes (Figure 2.6). The first cluster (green) focused on forest ecology and included the following elements: forestry, hyperspectral,

regression, reflection, mean square error, land use, ecology, plants (botany), poplar, China, normalized difference vegetation, NDVI, ecosystem, groundwater, and crops. The second cluster (yellowish-green) focused on vegetation mapping and included the following items: Sentinel-2, biomass, plantation, sentinel, wetland, deciduous tree, algorithm, sampling, image classification, and satellite data. The third cluster (red) focused on remote sensing techniques and included the following elements: riparian forest, rivers, France, radiative transmission, canopy, remote sensing, deciduous tree, United States, tree, leaf area index, optical radar, image analysis, phenology, *Liriodendra Tulipiferal*, LiDAR, and image processing. The fourth cluster (blue) focused on the overall vegetation physiology studies and included the following items: populus, satellite imagery, vegetation, plants, articles, *Populus deltoides*, comparative study, photosynthesis, chlorophyll, and physiology.

Based on the study results, all 148 poplar remote sensing studies were published in 70 different journals, with fifteen journals publishing three or more studies (*International Geoscience and Remote Sensing Symposium* (14), *Remote Sensing of Environment* (8), *International Journal of Remote Sensing* (7), *Remote Sensing* (7), *Canadian Journal of Forest Research* (5), *IEEE Transactions on Geoscience and Remote Sensing* (5), *Proceedings of SPIE—The International Society for Optical Engineering* (5), *ISPRS Archives* (4), the *ISPRS Journal of Photogrammetry and Remote Sensing* (4), the *Journal of Geophysical Research: Atmospheres* (4)), and the *International Conference on Remote Sensing, Environment and Transportation Engineering*; *Annals of Silvicultural Research*, *Forest Ecology and Management*; *Forests*; *IForest*; and *IOP Conference Series: Earth and Environmental Science* having published 3 articles each (Table 2.3). Table 2.3 presents only journals that published > 3 papers each.

Table 2.3 Journals where Poplar studies were published and their funding agencies.

Journals	Funder
International Geosciences and Remote sensing Symposium (14)	G (14)
Remote Sensing (8)	G (8)
Remote Sensing of the Environment (8)	G (8)
IEEE Transactions on Geoscience and Remote Sensing (7)	G (7)
International Journal of Remote Sensing (7)	G (7)
Proceedings of SPIE—the International Society for Optical Engineering (7)	G (7)
Canadian Journal of Remote Sensing (4)	G (3); G and P (1)
ISPRS Journal of Photogrammetry and Remote Sensing (4)	G (4)
International Archives of the Photogrammetry Remote Sensing and Spatial Information Science (4)	G (3); P (1)
Journal of Geophysical Research: Atmospheres (4)	G (3) and P (1)

G = government and P = private.

2.4 LIMITATIONS AND FUTURE RESEARCH

This review revealed the trends in the application of different remote sensing products in mapping, monitoring, and detecting poplar trees in different locations. Poplars occupy a more significant area in agriculture and natural forest areas. While remote sensing techniques have proven effective in delineating and extracting poplar tree features in both landscapes using LiDAR sensors, there are limitations related to spatial coverage. The combination of high-spatial-resolution sensors such as LiDAR and multispectral satellite sensors covering larger regions proved to be an excellent alternative to offset the mentioned limitations. The integration of these images will continue the improvements in the time-series mapping and monitoring needed to regularly update forest inventories. Good performance was also achieved by combining multitemporal and high-spectral sensors such as Sentinel 2 and machine (random forests) and deep learning techniques (connected neural network). Because of the robustness that results from combining these remote sensing techniques, researchers strongly advise using the latter in subsequent research.

In addition, the creation and application of poplar-specific indices such as the poplar tree index (PTI) proved to be more effective than the general vegetation index (NDVI) in mapping poplar plantations. Therefore, creating landscape-specific poplar indices and combining them with multitemporal and high-resolution imagery would improve the phenological analysis of poplars. While many multitemporal, multispectral, and high-resolution remote sensing images are now freely available on many platforms (e.g., GEE, USGS, Copernicus, etc.), other important remote sensing datasets that are suitable for mapping poplar trees have high costs, making their use limited in resource-constrained countries. Collaboration between researchers from developed countries—where resources are abundant—and researchers from developing countries is therefore thought to be advantageous. This is because it would allow for the sharing of resources and insights, raising the bar for remote sensing applications in poplars around the world.

A notable constraint of this research is that it just examined and synthesized peer-reviewed publications; it did not include papers like technical reports, theses, dissertations, conservation management plans, etc. While a thorough study of the materials that were excluded based on our process is advised, we also implore the scientific community to make sure that the documents in question are from reliable sources and have undergone a thorough review process before being published.

2.5 CONCLUSION

The current study provided an overview of peer-reviewed poplar studies using remote sensing approaches. The study results show an increase in the number of poplar remote sensing studies from 1991 to 2022, with a total of 148 during that period. Most studies used airborne sensors, particularly LiDAR, satellites (mostly using Landsat and Sentinel-2), and ground-based sensors. Over two thirds of the publications were widespread in the Global North, including China, where poplars were predominantly considered native rather than an invasive species. Poplar species have been studied primarily in agriculture, forestry, mountains, and urban environments. Commonly studied biophysical indicators used in the reviewed studies were crown diameter, poplar growth/height and volume, canopy cover, and poplar biomass. Spectral-based analysis included machine learning classifiers such as RF, SVM, and others. The study, therefore, concludes that remote sensing products are becoming an

essential data source for understanding poplar species dynamics, as they provide more consistent and low-cost information than traditional approaches. As high-resolution satellite data become accessible through various partnerships (e.g., PlanetScope with Google Earth Engine) and deep learning algorithms continue to evolve, the remote sensing of poplars is expected to incorporate these resources in the future. This may benefit countries that have been less studied, including those in Africa, where poplars occur.

This information should encourage cooperative research interests between developed and developing countries and allow land managers and conservationists with an opportunity to learn which remote sensing techniques work best when researching poplar in various regions. The study anticipates that more countries will be inspired to consider funding such projects, particularly in cases where poplar is considered an alien invasive species that invades delicate landscapes like riparian zones, as countries that published a higher number of peer-reviewed articles were the ones that received funding.

2.6 REFERENCES

- Adagbasa, E., Adelabu, S. and Okello, T., 2019. Application of deep learning with stratified K-fold for vegetation species discrimination in a protected mountainous region using Sentinel-2 image. *Geocarto International*, 37, 1-21. <https://doi.org/10.1080/10106049.2019.1704070>
- Adam, E., Mutanga, O., Odindi, J. and Abdel-Rahman, E.M., 2014. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing*, 35, 3440-3458. <https://doi.org/10.1080/01431161.2014.903435>
- Adepoju, K. and Adelabu, S., 2020. Improving accuracy evaluation of Landsat-8 OLI using image composite and multisource data with Google Earth Engine. *Remote Sensing Letters*, 11, 107-116. <https://doi.org/10.1080/2150704X.2019.1690792>
- Baird, R.A., Verbyla, D. and Hollingsworth, T.N., 2012. Browning of the landscape of interior Alaska based on 1986-2009 Landsat sensor NDVI. *Canadian Journal of Forest Research*, 42, 1371-1382. <https://doi.org/10.1139/X2012-088>

- Barnes, M.L., Breshears, D.D., Law, D.J., van Leeuwen, W.J.D., Monson, R.K., Fojtik, A.C., Barron-Gafford, G.A. and Moore, D.J.P., 2017. Beyond greenness: Detecting temporal changes in photosynthetic capacity with hyperspectral reflectance data. PLoS One, 12. <https://doi.org/10.1371/journal.pone.0189539>
- Bonan, G.B., 1993. Importance of leaf area index and forest type when estimating photosynthesis in boreal forests. Remote sensing of Environment, 43(3), 303-314. [https://doi.org/10.1016/0034-4257\(93\)90072-6](https://doi.org/10.1016/0034-4257(93)90072-6)
- Brandtberg, T., 2007. Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. ISPRS Journal of Photogrammetry and Remote Sensing, 61(5), 325-340.
- Cañete-Salinas, P., Zamudio, F., Yáñez, M., Gajardo, J., Valdés, H., Espinosa, C., Venegas, J., Retamal, L., Ortega-Farias, S. and Acevedo-Opazo, C., 2020. Evaluation of models to determine LAI on poplar stands using spectral indices from Sentinel-2 satellite images. Ecological modelling, 428, 109058. <https://doi.org/10.1016/j.ecolmodel.2020.109058>
- Cao, L., Liu, H., Fu, X., Zhang, Z., Shen, X. and Ruan, H., 2019. Comparison of UAV LiDAR and digital aerial photogrammetry point clouds for estimating forest structural attributes in subtropical planted forests. Forests, 10(2). 145. <https://doi.org/10.3390/f10020145>
- Capolupo, A., Nasta, P., Palladino, M., Cervelli, E., Boccia, L. and Romano, N., 2018. Assessing the ability of hybrid poplar for in-situ phytoextraction of cadmium by using UAV-photogrammetry and 3D flow simulator. International journal of remote sensing, 39(15-16), 5175-5194. <https://doi.org/10.1080/01431161.2017.1422876>
- Caudullo, G. and De Rigo, D., 2016. *Populus alba* in Europe: distribution, habitat, usage and threats. European Atlas of Forest Tree Species.
- Ceylan, I., 2008. Determination of drying characteristics of timber by using artificial neural networks and mathematical models. Drying Technology, 26(12). 1469-1476 <https://doi.org/10.1080/07373930802412132>
- Chen, L., Ren, C., Zhang, B., Wang, Z., Liu, M., Man, W. and Liu, J., 2021. Improved estimation of forest stand volume by the integration of GEDI LiDAR data and multi-sensor imagery in the Changbai Mountains

Mixed forests Ecoregion (CMMFE), northeast China. *International Journal of Applied Earth Observation and Geoinformation*, 100, 102326. <https://doi.org/10.1016/j.jag.2021.102326>

Chianucci, F., Puletti, N., Grotti, M., Ferrara, C., Giorcelli, A., Coaloa, D. and Tattoni, C., 2020. Nondestructive tree stem and crown volume allometry in hybrid poplar plantations derived from terrestrial laser scanning. *Forest Science*, 66(6), 737-746. <https://doi.org/10.1093/forsci/fxaa021>

Colkesen, I., Kavzoglu, T., Tonbul, H., Ozturk, M.Y. and Yilmaz, E.O., 2022. Poplar Tree Index (PTI): A New Vegetation Index for Monitoring Poplar Cultivated Areas. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 5567-5570. <https://doi.org/10.1109/IGARSS46834.2022.9883085>

Colombo, R., Meroni, M., Marchesi, A., Busetto, L., Rossini, M., Giardino, C. and Panigada, C., 2008. Estimation of leaf and canopy water content in poplar plantations by means of hyperspectral indices and inverse modeling. *Remote Sensing Environment*, 112(4), 1820-1834. <https://doi.org/10.1016/j.rse.2007.09.005>

D'Amico, G., Francini, S., Giannetti, F., Vangi, E., Travaglini, D., Chianucci, F., Mattioli, W., Grotti, M., Puletti, N., Corona, P. and Chirici, G., 2021. A deep learning approach for automatic mapping of poplar plantations using Sentinel-2 imagery. *GIScience & Remote Sensing*, 58(8), 1352-1368. <https://doi.org/10.1080/15481603.2021.1988427>

Danise, T., Innangi, M., Curcio, E., Piccolella, S., Fioretto, A. and Pacifico, S., 2021. White poplar (*Populus alba* L.) leaf waste recovery and intercropping outcome on its polyphenols. *Industrial Crops and Products*, 171, 113866. <https://doi.org/10.1016/j.indcrop.2021.113866>

der Zande, D., Jonckheere, I., Stuckens, J., Verstraeten, W. and Coppin, P., 2008. Sampling design of ground-based lidar measurements of forest canopy structure and its effect on shadowing. *Canadian Journal of Remote Sensing*, 34, 526-538. <https://doi.org/10.5589/m08-070>

Duncanson, L., Dubayah, R., Cook, B., Rosette, J. and Parker, G., 2015. The importance of spatial detail: Assessing the utility of individual crown information and scaling approaches for lidar-based biomass density estimation. *Remote Sensing of Environment*, 168, 102-112. <https://doi.org/10.1016/j.rse.2015.06.021>

Eslami, A. and Zahedi, S.S., 2011. Providing poplar plantation map by Indian remote sensing (IRS) satellite imagery in Northern Iran. *African Journal of Agricultural Research*, 6(20), 4769-4774.

Franco-Lopez, H., Ek, A. and Bauer, M., 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote sensing of Environment*, 77(3), 251-274. [https://doi.org/10.1016/S0034-4257\(01\)00209-7](https://doi.org/10.1016/S0034-4257(01)00209-7)

Gao, L. and Zhang, X., 2021. Above-ground biomass estimation of plantation with complex forest stand structure using multiple features from airborne laser scanning point cloud data. *Forests*, 12(12), 1713. <https://doi.org/10.3390/f12121713>

Gao, W., Zhou, L., Liu, K., Guan, Y., Gao, H. and Hu, J., 2022. Machine learning algorithms for rapid estimation of holocellulose content of poplar clones based on Raman spectroscopy. *Carbohydrate Polymers*, 292, 119635. <https://doi.org/10.1016/j.carbpol.2022.119635>

González-Moreno, P., Pino, J., Carreras, D., Basnou, C., Fernández-Rebollar, I. and Vilà, M., 2013. Quantifying the landscape influence on plant invasions in Mediterranean coastal habitats. *Landscape ecology*, 28, 891-903. <https://doi.org/10.1007/s10980-013-9857-1>

Hall, R.J., Davidson, D. and Peddle, D., 2003. Ground and Remote Estimation of Leaf Area Index in Rocky Mountain Forest Stands, Kananaskis, Alberta. *Canadian Journal of Remote Sensing*, 29, 411-427. <https://doi.org/10.5589/m03-012>

Hamrouni, Y., Paillassa, E., Chéret, V., Monteil, C. and Sheeren, D., 2022. Sentinel-2 Poplar Index for Operational Mapping of Poplar Plantations over Large Areas. *Remote Sensing*, 14(16), 3975. <https://doi.org/10.3390/rs14163975>

Hamrouni, Y., Paillassa, E., Chéret, V., Monteil, C. and Sheeren, D., 2021. From local to global: A transfer learning-based approach for mapping poplar plantations at national scale using Sentinel-2. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171, 76-100. <https://doi.org/10.1016/j.isprsjprs.2020.10.018>

- Henderson, L., 2007. Invasive, naturalized and casual alien plants in southern Africa: A summary based on the Southern African Plant Invaders Atlas (SAPIA). *Bothalia*, 37(2), 215-248. <https://doi.org/10.4102/abc.v37i2.322>
- Huang, C.Y. and Asner, G.P., 2009. Applications of remote sensing to alien invasive plant studies. *Sensors (Switzerland)*, 9(6), 4869-4889. <https://doi.org/10.3390/s90604869>
- Huang, T., Ding, X., Zhu, X., Chen, S., Chen, M., Jia, X., Lai, F., Zhang, X., 2021. Assessment of Poplar Looper (*Apocheima cinerarius* Erschoff) Infestation on Euphrates (*Populus euphratica*) Using Time-Series MODIS NDVI Data Based on the Wavelet Transform and Discriminant Analysis. *Remote Sensing*, 13(12), 2345. <https://doi.org/10.3390/rs13122345>
- Javanmardi, E., Liu, S. and Xie, N., 2020. Exploring grey systems theory-based methods and applications in sustainability studies: A systematic review approach. *Sustainability (Switzerland)*, 12(11), 4437. <https://doi.org/10.3390/su12114437>
- Khayatinejad Aqgonbad, S., Hossein Zadeh, O., Hajjarian, M. and Abdi, M.R., 2022. Impact of Key Stakeholders' Participation in Strategic Decision-Making Regarding Poplar Tree Farming Development. *Small-scale Forestry*, 21, 1-19. <https://doi.org/10.1007/s11842-021-09496-3>
- Khorrami, R.A., Naeimi, Z., Wing, M., Ahani, H. and Bateni, S.M., 2022. A New Multistep Approach to Identify Leaf-Off Poplar Plantations Using Airborne Imagery. *Journal of Geographic Information System*, 14(6), 634-651. <https://doi.org/10.4236/jgis.2022.146036>
- Ko, C., Sohn, G., Rimmel, T.K. and Miller, J.R., 2016. Maximizing the diversity of ensemble random forests for tree genera classification using high density LiDAR data. *Remote Sensing*, 8(8), 646. <https://doi.org/10.3390/rs8080646>
- Kota, N.L., Landenberger, R.E. and McGraw, J.B., 2007. Germination and early growth of *Ailanthus* and tulip poplar in three levels of forest disturbance. *Biological Invasions*, 9, 197-211. <https://doi.org/10.1007/s10530-006-9026-4>

- Laslier, M., Hubert-Moy, L. and Dufour, S., 2019. Mapping riparian vegetation functions using 3D bispectral LiDAR data. *Water (Switzerland)*, 11(3), 483. <https://doi.org/10.3390/w11030483>
- Lechner, A.M., Foody, G.M. and Boyd, D.S., 2020. Applications in Remote Sensing to Forest Ecology and Management. *One Earth*, 2(5), 405-412. <https://doi.org/10.1016/j.oneear.2020.05.001>
- Leite, R.V., Silva, C.A., Mohan, M., Cardil, A., de Almeida, D.R.A., e Carvalho, S. de P.C., Jaafar, W.S.W.M., Hernández, J.G., Weiskittel, A., Hudak, A.T., Broadbent, E.N., Prata, G., Valbuena, R., Leite, H.G., Taquetti, M.F., Soares, A.A.V., Scolforo, H.F., Do Amaral, C.H., Corte, A.P.D. and Klauberg, C., 2020. Individual tree attribute estimation and uniformity assessment in fast-growing eucalyptus spp. Forest plantations using lidar and linear mixed-effects models. *Remote Sensing*, 12(21), 3599. <https://doi.org/10.3390/rs12213599>
- Li, L., Hao, D., Xuecao, L., Chen, M., Zhou, Y., Jurgens, D., Asrar, G. and Sapkota, A., 2021. Satellite-based phenology products and in-situ pollen dynamics: A comparative assessment. *Environmental research*, 204, 111937. <https://doi.org/10.1016/j.envres.2021.111937>
- Lin, W., Lu, Y., Li, G., Jiang, X. and Lu, D., 2022. A comparative analysis of modeling approaches and canopy height-based data sources for mapping forest growing stock volume in a northern subtropical ecosystem of China. *GIScience & Remote Sensing*, 59 (1), 568-589. <https://doi.org/10.1080/15481603.2022.2044139>
- Linhui, L., Weipeng, J. and Huihui, W., 2021. Extracting the Forest Type from Remote Sensing Images by Random Forest. *IEEE Sensors Journal*, 21(16), 17447-17454. <https://doi.org/10.1109/JSEN.2020.3045501>
- Lu, S., Lu, X., Zhao, W., Liu, Y., Wang, Z. and Omasa, K., 2015. Comparing vegetation indices for remote chlorophyll measurement of white poplar and Chinese elm leaves with different adaxial and abaxial surfaces. *Journal of Experimental Botany*, 66(18), 5625-5637. <https://doi.org/10.1093/jxb/erv270>
- Maleki, M., Arriga, N., Barrios, J.M., Wieneke, S., Liu, Q., Peñuelas, J., Janssens, I.A. and Balzarolo, M., 2020. Estimation of Gross Primary Productivity (GPP) phenology of a short-rotation plantation using remotely sensed indices derived from Sentinel-2 images. *Remote Sensing*, 12(13), 2104. <https://doi.org/10.3390/rs12132104>

- Man, Q., Dong, P., Yang, X., Wu, Q. and Han, R., 2020. Automatic Extraction of grasses and individual trees in urban areas based on airborne hyperspectral and LiDAR data. *Remote Sensing*, 12(17), 2725. <https://doi.org/10.3390/RS12172725>
- Mansfield, S.D., Kang, K.Y., Iliadis, L., Tachos, S. and Avramidis, S., 2011. Predicting the strength of *Populus* spp. clones using artificial neural networks and ϵ -regression support vector machines (ϵ -rSVM). *Holzforschung*, 65. <https://doi.org/10.1515/HF.2011.107>
- Marrs, J. and Ni-Meister, W., 2019. Machine learning techniques for tree species classification using co-registered LiDAR and hyperspectral data. *Remote Sensing*, 11(7), 819. <https://doi.org/10.3390/rs11070819>
- Moffiet, T., Mengersen, K., Witte, C., King, R. and Denham, R., 2005. Airborne laser scanning: Exploratory data analysis indicates potential variables for classification of individual trees or forest stands according to species. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(5), 289-309.
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A., Estarli, M., Barrera, E.S.A., Martínez-Rodríguez, R., Baladia, E., Agüero, S.D., Camacho, S., Buhring, K., Herrero-López, A., Gil-González, D.M., Altman, D.G., Booth, A., Chan, A.W., Chang, S., Clifford, T., Dickersin, K., Egger, M., Gøtzsche, P.C., Grimshaw, J.M., Groves, T., Helfand, M., Higgins, J., Lasserson, T., Lau, J., Lohr, K., McGowan, J., Mulrow, C., Norton, M., Page, M., Sampson, M., Schünemann, H., Simera, I., Summerskill, W., Tetzlaff, J., Trikalinos, T.A., Tovey, D., Turner, L. and Whitlock, E., 2016. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Revista Espanola de Nutricion Humana y Dietetica*, 20. <https://doi.org/10.1186/2046-4053-4-1>
- Petráš, R., Mecko, J. and Nociar, V., 2010. Diameter structure of the stands of poplar clones. *Journal of Forest Science*, 56(4), 165-170. <https://doi.org/10.17221/65/2009-jfs>
- Pirotti, F., 2010. Assessing a template matching approach for tree height and position extraction from lidar-derived canopy height models of *Pinus pinaster* stands. *Forests*, 1(4), 194-208. <https://doi.org/10.3390/f1040194>
- R Core, T., 2004. R: a Language and Environment for Statistical Computing., <http://www.R-project.org/>.

Richardson, D.M. and Rejmánek, M., 2011. Trees and shrubs as invasive alien species - a global review. *Diversity and distributions*, 17(5), 788-809. <https://doi.org/10.1111/j.1472-4642.2011.00782.x>

Rupasinghe, P.A. and Chow-Fraser, P., 2021. Relating pre-fire canopy species, fire season, and proximity to surface waters to burn severity of boreal wildfires in Alberta, Canada. *Forest Ecology and Management*, 496, 119386. <https://doi.org/10.1016/j.foreco.2021.119386>

Santi, E., Paloscia, S., Pampaloni, P. and Pettinato, S., 2009. Ground-based microwave investigations of forest plots in Italy. *IEEE Transactions on Geoscience and Remote Sensing*, 47. <https://doi.org/10.1109/TGRS.2009.2021613>

Schweier, J., Schnitzler, J.P. and Becker, G., 2016. Selected environmental impacts of the technical production of wood chips from poplar short rotation coppice on marginal land. *Biomass and Bioenergy*, 85, 235-242. <https://doi.org/10.1016/j.biombioe.2015.12.018>

Shafri, H.Z.M., Suhaili, A. and Mansor, S., 2007. The Performance of Maximum Likelihood, Spectral Angle Mapper, Neural Network and Decision Tree Classifiers in Hyperspectral Image Analysis. *Journal of Computer Science*, 3, 419-423. <https://doi.org/10.3844/jcssp.2007.419.423>

Sikorska, D., Sikorski, P., Archiciński, P., Chormański, J. and Hopkins, R.J., 2019. You can't see the woods for the trees: Invasive *Acer negundo* L. in Urban riparian forests harms biodiversity and limits recreation activity. *Sustainability (Switzerland)*, 11. <https://doi.org/10.3390/su11205838>

Simović, I., Šikoparija, B., Panić, M., Radulović, M. and Lugonja, P., 2022. Remote Sensing of Poplar Phenophase and Leaf Miner Attack in Urban Forests. *Remote Sensing*, 14(24), 6331. <https://doi.org/10.3390/rs14246331>

Szuster, B.W., Chen, Q. and Borger, M., 2011. A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. *Applied Geography*, 31(2), 525-532. <https://doi.org/10.1016/j.apgeog.2010.11.007>

- Tonbul, H., Colkesen, I. and Kavzoglu, T., 2020. Classification of poplar trees with object-based ensemble learning algorithms using Sentinel-2A imagery. *Journal of Geodetic Science*, 10(1), 14-22. <https://doi.org/10.1515/jogs-2020-0003>
- van Eck, N.J. and Waltman, L., 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84, 523-538. <https://doi.org/10.1007/s11192-009-0146-3>
- Wang, K., Wang, T. and Liu, X., 2018. A review: Individual tree species classification using integrated airborne LiDAR and optical imagery with a focus on the urban environment. *Forests*, 10(1), 1. <https://doi.org/10.3390/f10010001>
- Warner, T.A., McGraw, J.B. and Landenberger, R., 2006. Segmentation and classification of high resolution imagery for mapping individual species in a closed canopy, deciduous forest. *Science in China Series E: Technological Sciences*, 49, 128-139.
- Xu, D., Wang, H., Xu, W., Luan, Z. and Xu, X., 2021. LiDAR applications to estimate forest biomass at individual tree scale: Opportunities, challenges and future perspectives. *Forests*, 12(5), 550. <https://doi.org/10.3390/f12050550>
- Zhang, Y.Y., Thenkabail, P.S. and Wang, P., 2019. A bibliometric profile of the Remote Sensing Open Access Journal published by MDPI between 2009 and 2018. *Remote Sensing*, 11(1), 91. <https://doi.org/10.3390/rs11010091>
- Zhu, L., Liu, X., Wu, L., Tang, Y. and Meng, Y., 2019. Long-term monitoring of cropland change near Dongting Lake, China, using the landtrendr algorithm with Landsat imagery. *Remote Sensing*, 11(10), 1234. <https://doi.org/10.3390/rs11101234>

CHAPTER 3 SENTINEL-2 BASED CLASSIFICATION AND SPECTRAL SEPARABILITY OF LAND SURFACE FEATURES IN THE POPLAR- INVADED LANDSCAPE OF THE EASTERN FREE STATE, SOUTH AFRICA

This Chapter is based on:

Mapuru, M., Xulu, S., Gebreslasie, M. and Sadiki, M. [*Under review*]. Sentinel-2 based classification and separability of land surface features in the poplar-invaded landscape of the eastern Free State. *Journal of Sensors*.

ABSTRACT

Mapping invasive alien plants (IAPs) has become an essential practice for land and biodiversity conservation authorities as these species are capable of transforming invaded areas. Fortunately, advances in remote sensing with publicly available products such as Sentinel-2, have enhanced this process, particularly in difficult to access mountainous regions. In South Africa, poplar (*Populus alba*) is among IAPs of concern and occurs in the eastern Free State and other parts of the country, but remote sensing has not yet been explored to map this species. Using Sentinel-2 imagery and Random Forests (RF) algorithm, this study enabled us to: (a) map and discern poplar tree species from other land classes throughout the year in the mountainous region, (b) evaluate influential bands and their combinations in the mapping process, and (c) assess the mapping accuracy for the first and second half of the year. We found that imageries from the first half of the year (Jan-Jun) had higher classification accuracy (91%) compared to images from the second half of the year (Jul-Dec) (87%). Poplar and all other classes were clearly separable, with poplar shown mostly in the riparian areas. We found that variables such as SWIR-1, NDVI, Blue, PI-1, MNDWI, NIR, and PI-3 are crucial parameters in the classification of poplar trees in mountainous areas. Overall, our study demonstrated that Sentinel-2 images with RF classifier hold much potential for mapping poplar alien trees in mountainous ecosystems.

Keywords: random forest, poplar, invasive alien plant, Sentinel-2

3.1 INTRODUCTION

Invasive alien plants (IAPs) pose a significant threat to natural and human-modified environments worldwide, reportedly affecting biodiversity, ecosystem services, the economy, and society (Gallardo et al. 2019; Shah et al. 2020; Haubrock et al. 2021). IAS refers to species that are introduced from their natural habitats to new habitats, intentionally or unintentionally, by humans or natural influences such as wind (Kannan et al. 2013; Robinson et al. 2020). Most of these species are adaptable and naturalize in new ecosystems, competing with native species for limited resources (Zhang and van Kleunen, 2019; Bradley et al. 2019). IAS have been documented to spread and replace native important species in many locations (Ngondya and Munishi, 2021; Poland et al. 2021), particularly in grazing and remote areas. This negatively impacts the agricultural and economic sectors, especially when these impacts affect the quality of production (Haubrock et al. 2021). Differences between IAS and native species in their requirements and modes of resource use and acquisition often alter soil structure (Zhang et al. 2019), decomposition, soil nutrient content (Linden et al. 2019), and moisture availability (Langmaier and Lapin, 2020). One IAS species that has become a problem in many parts of the world is the poplar (Bradley et al. 2010).

Poplar is a fast-growing and highly adaptable species that has been introduced to many regions outside of its natural range, including South Africa (Mbedzi et al. 2019). Due to its occurrence and impact in fragile landscapes such as riparian zones, it has been classified as a species of high concern by the South African National Parks (SANParks) (Le Maitre et al. 2016; SANParks 2020). Poplar species are often found near water bodies and consume more water, leaving native plants with less water (Dzikiti et al. 2017; Bekele et al. 2018). This signifies the need for mapping and monitoring of its distribution and abundance for effective management and conservation of natural resources. Mapuru et al. (2024) modelled the potential habitat for poplar invasion in the country, which includes the eastern Free State province. In the same region, manually the presence of poplar trees in the Golden Gate Highlands National Park using traditional survey methods. Here the study presented a first attempt to demonstrate the remote sensing ability to monitor poplar in South Africa.

A recent remote sensing review on poplar studies showed a growth trajectory worldwide, demonstrating with varying degrees of success that ground-to-satellite sensors are highly effective for mapping poplar (Mapuru et al. 2023). Despite these encouraging patterns and the poplar invasion in South Africa, no study using remote sensing approaches has yet been conducted in the country. Publicly available products such as Landsat-8 and Sentinel-2 have shown satisfactory results in discerning poplar in complex landscapes (Zhu et al. 2019; Marbán et al. 2019; D'Amico et al. 2021). For this purpose, Sentinel-2 imagery appear as a preferred choice over Landsat (Wang et al. 2022), especially in regions affected by frequent cloud cover. This is partly due to their relatively shorter acquisition cycle of three to five days (increasing the probability of cloud-free scenes) and thirteen spectral bands. These properties have facilitated the detailed characterization of poplar, including the biophysical structure and biomass from tree to landscape level (Cartisano et al. 2013; Liu et al. 2019). Other commonly used products such as LiDAR are limited by higher costs, particularly in developing countries with limited financial resources (Mapuru et al. 2023). Therefore, the use of Sentinel-2 sensor images has become a practical solution (D'Amico et al. 2021; Hamrouni et al. 2022), especially in mountainous areas (Hemmerling et al. 2021).

Studies using a combination of machine learning and Sentinel-2 images to map poplar have achieved optimal classification accuracies (Tonbul et al. 2020; D'Amico et al. 2021). For instance, Ozturk and Colkesen (2020) used Random Forests (RF), Support Vector Machine (SVM), and Adaboost along with Sentinel-2 imagery to map poplar tree growing areas in the Akyazi district of Turkey. The study yielded remarkable results, with RF achieving an accuracy of 99%, outperforming SVM (98%) and Adaboost (97%). Recent studies have improved accuracy by computing various indices suitable for poplar classification in addition to using bands of the images (see Hamrouni et al. 2022). While there have been successful remote sensing and machine learning techniques for mapping poplar in agriculture and natural forested areas, no study has been conducted to quantify the distribution of poplar in mountainous landscapes of South Africa. To our knowledge, our study will be the first to spectrally map poplar distribution against other land cover features in South Africa.

The objective of this study was to discern poplar (*Populus alba*) species over a mountainous landscape using Sentinel-2 imagery and the Random Forests algorithm. Specifically, our study allowed us to: (1)

distinguish poplar from other types of land cover classes in the first and second half of the year, (2) identify the key spectral bands and indices for the mapping procedure, and (3) evaluate the mapping across a whole the study period.

3.2 METHODS AND MATERIALS

3.2.1 Description of the study area

The research study was conducted at Phuthaditjhaba, an area situated in the northeastern Free State (refer to Figure 3.1). The area's elevation ranges from 1675 to over 2200 meters above sea level. Phuthaditjhaba is intersected by streams such as Metsi-Matso, Mphukojwane, and Namahadi, which converge to form the Elands River. According to the Agricultural Research Council Institute for Soil Climate and Water (2016), Phuthaditjhaba has moist and warm summers and dry, and cold winters. The region receives an average annual rainfall of about 629 mm, with most of it falling between October and March. The Karoo Supergroup is the primary geological formation in this area (Hobbs et al., 2013). Glenrosa, lithosols, and Mispah are the main soil types in Phuthaditjhaba (CEM, 2003). From a vegetation perspective, the study area is part of the grassland biome, with some forest patches.



Figure 3.1 Study area map showing the location of the study site (marked with a red dot) in South Africa. An image in true colour displays Phuthaditjhaba, the current study site.

3.2.2 Acquisition and processing of the images

The study utilized composite imagery from the Sentinel-2 dataset for the first and second half of 2023. This timeframe was chosen to cover all phenological phases (summer, autumn, winter and spring). Composite images (computed with median method) were preferred over individual ones because they minimize the impact of noisy data caused by elements such as clouds, cloud shadows, and haze. By merging multiple images into one, it becomes easier to compare images over time. The data was obtained from the Digital Earth Africa Datacube and analysed using the DE Africa Sandbox, which is a cloud-based computational platform operating within a Jupyter Lab environment. To ensure high data quality, only observations with at least 90% good data (i.e., less than 15% cloud cover and 100% coverage of the study area) were retained. Bands 1 (Coastal aerosol), 9 (water vapor), and 10 (SWIR Cirrus), which are intended for atmospheric information, were excluded from the analysis.

The Sentinel-2 imagery was utilized to compute vegetation indices selected for their ability to differentiate between vegetation and other land cover features while addressing soil noise effects. Indices incorporating red-edge and NIR wavelengths were developed to effectively map vegetation (Zwiggelaar, 1998; Xie et al. 2008). These indices include the Normalized Difference Vegetation Index (NDVI) (Rouse et al. 1973), Modified Normalized Difference Water Index (MNDWI) (Xu, 2006), Leaf Area Index (LAI) (Boegh et al. 2002), and Poplar Index (PI1, 2, 3, 4) (Hamrouni et al. 2022).

Table 3.1 Sentinel-2 imagery bands and vegetation indices used.

Band/index	Description
B2	Blue
B3	Green
B4	Red
B5	Red Edge 1
B6	Red Edge 2
B7	Red Edge 3
B8	Near Infrared (NIR) wide
B11	Short Wave Infrared (SWIR) 1
B12	Short Wave Infrared (SWIR) 2
NDVI	$\frac{NIR - RED}{NIR + RED}$
MNDWI	$\frac{Green - SWIR}{Green + SWIR}$
LAI	$3.618 * EVI - 0.118$
Poplar Index 1	$SWIR1 - SWIR2$
Poplar Index 2	$Red\ Edge\ 1 - (SWIR1 + SWIR2)$
Poplar Index 3	$\frac{Red\ Edge\ 1 - (SWIR1 + SWIR2)}{Red\ Edge\ 1 + (SWIR1 + SWIR2)}$
Poplar Index 4	$SWIR1 + SWIR2$

3.2.3 Dataset for training and validation

To create an accurate land cover map, the first step is to gather precise reference points to train the models (Holloway et al. 2018). To collect reference data for Poplar presence samples, we used a Garmin Global Positioning System set to 2-meter accuracy. Additionally, we supplemented this data with samples collected from high-resolution images integrated into the Google Earth Pro platform. The Poplar presence samples were collected between January 5th and 15th, 2023, as this is when the species is spectrally distinct from other land cover features. We collected sample points for Poplar in locations where it was dominant, mainly along riparian zones and other wet surfaces with sufficient water to support its growth. Sample points from various land cover classes such as bare land, grassland, settlement, roads, and water from poplar-infested areas were also collected to complete the mapping

process (see Table 3.2). For validation purposes, 70% of the data was allocated for training and 30% for validation using the Python programming language (Congalton, 2001; Abdel-Rahman et al. 2014; Adam et al. 2017).

Table 3.2 Land cover classes and corresponding training, validation, and total number of samples.

Classes	Training points	Validation points	Total
Bare land	139	59	198
Grassland	138	59	197
Settlement	106	45	151
Poplar	141	61	202
Road	71	30	101
Water	143	61	204

3.2.4 Time series analysis

The time series analysis was conducted to determine the optimal months of the year to include in the machine learning model for mapping Poplar species. The primary goal was to identify the periods when spectral characteristics of different land use/land cover (LULC) types, including Poplar species, were most distinctive and informative for classification purposes. By understanding these temporal variations, we aimed to enhance the accuracy and reliability of the machine learning models. For each LULC class, median values of the spectral indices were calculated across the time series. This step aimed to identify the months when the spectral characteristics of Poplar species were most distinguishable from other LULC types. The time series analysis revealed critical periods during the year when the spectral signatures of different LULC classes, including Poplar species, exhibited maximum distinctiveness. These insights guided the selection of specific months for inclusion in the machine learning model, ensuring that the temporal variability captured in the training data would enhance the classifier's performance. Based on the insights gained from the time series plots, it was decided to create semi-annual medians for the periods January to June and July to December.

3.2.5 Classification

Feature selection is the process of choosing a subset of variables that are important for mapping processes from the original set of data (Liu and Motoda, 1998; Pal and Foody, 2010). This process helps based on their contributions (Breiman, 2001). One method of feature selection involves identifying and removing highly correlated features from a dataset to address multicollinearity issues that can impact machine learning model performance. Initially, the correlation matrix of the dataset's features is computed and visualized using a heatmap to highlight correlations. A removal threshold of 0.9 is set, meaning any feature pair with a correlation coefficient equal to or greater than 0.9 is considered highly correlated. This reduction in data dimension simplifies processing and improves the classifier's performance (Poona et al. 2016). In this study, the Random Forest algorithm was used to rank the variables from a set of Sentinel-2 bands, and indices that were found to be most useful for mapping procedures.

This study used a Random Forest classifier for classifying various invasive plant species, as it has been successful in similar studies (Martin et al. 2018; Royimani et al. 2019; Paz-Kagan et al. 2019). The classifier evaluates the importance of different variables during classification. Variables such as image bands and vegetation indices with higher importance values contribute more to the accuracy of classification. The significance of variables is affected by the Random Forest method parameters, including the number of trees and variables per split. Optimal parameter selection for the Random Forest classifier can enhance the stability of variable importance scores and help differentiate between important and unimportant variables. In this study, the number of trees was specified, while other parameters defaulted to standard values if not specified. The number of trees was tested incrementally from 10 to 1000 to find the optimal number resulting in the highest overall accuracy. The model was built from 50 iterations. The algorithm also provides a measure of variable importance based on the Mean Decrease in Gini (MDG) value, which measures how much a variable reduces the Gini Impurity within a specific class. A higher MDG value indicates a purer and more influential variable (Han et al. 2016; Belgiu et al. 2016). The study used the MDG to calculate variable importance in the analysis.

3.2.6 Classification accuracy assessment

Classified maps from Random Forest were created for images from January to June and July to December. To evaluate the accuracy of the classification, we used a confusion matrix. The overall accuracy, user accuracy, producer accuracy and Kappa coefficient were used to measure the reliability of the classification process. Producer's accuracy (PA) indicates the likelihood that a randomly selected point in the field has the same class value as indicated on the map, while User's Accuracy (UA) indicates the likelihood that a randomly selected point on the map matches the class value in the field (Sammut and Webb, 2017). The overall Accuracy (OA) is determined by dividing the total number of correctly classified values by the total number of values (Sammut and Webb, 2017). Kappa coefficient commonly used to estimate interrater reliability, can be employed in the context of test – retest (Sammut and Webb, 2017).

3.3 RESULTS

3.3.1 Spectral separability of classes

The study begins this analysis with a detailed consideration of the spectral separability of all classes with a particular focus on poplars. As already mentioned, temporal signatures of different Sentinel-2 derived bands and their combinations for all classes over the year 2023 were plotted. Here, only plots of the most influential variables are shown in Figure 3.2, the remaining variables are listed in Appendix A. The study results show a clear pattern of class separability in some variables at certain times of the year while, spectral overlap is relatively small but noticeable. In particular, the blue band, shortwave infrared 1 (SWIR-1), near-infrared (NIR), Poplar Index 1 (PI-1), and Normalized Difference Vegetation Index (NDVI) show a high degree of separability between classes in the summer and autumn months (Figure 2). Poplar is more distinct than other classes with NDVI, blue band, NIR, and SWIR-1 during this period. In autumn, the blue band showed moderate separability like SWIR-1 in spring. Low separability between winter and spring was observed for the blue band, NIR, SWIR-1, NDVI, PI-1, and PI-3 with notable overlaps of spectral curves. The Modified Normalized Difference Water Index (MNDWI) consistently shows low separability throughout the year. Overall, our results suggest that there is high separability between classes in the summer and autumn months, while moderate to low separability is observed in the winter and spring months.

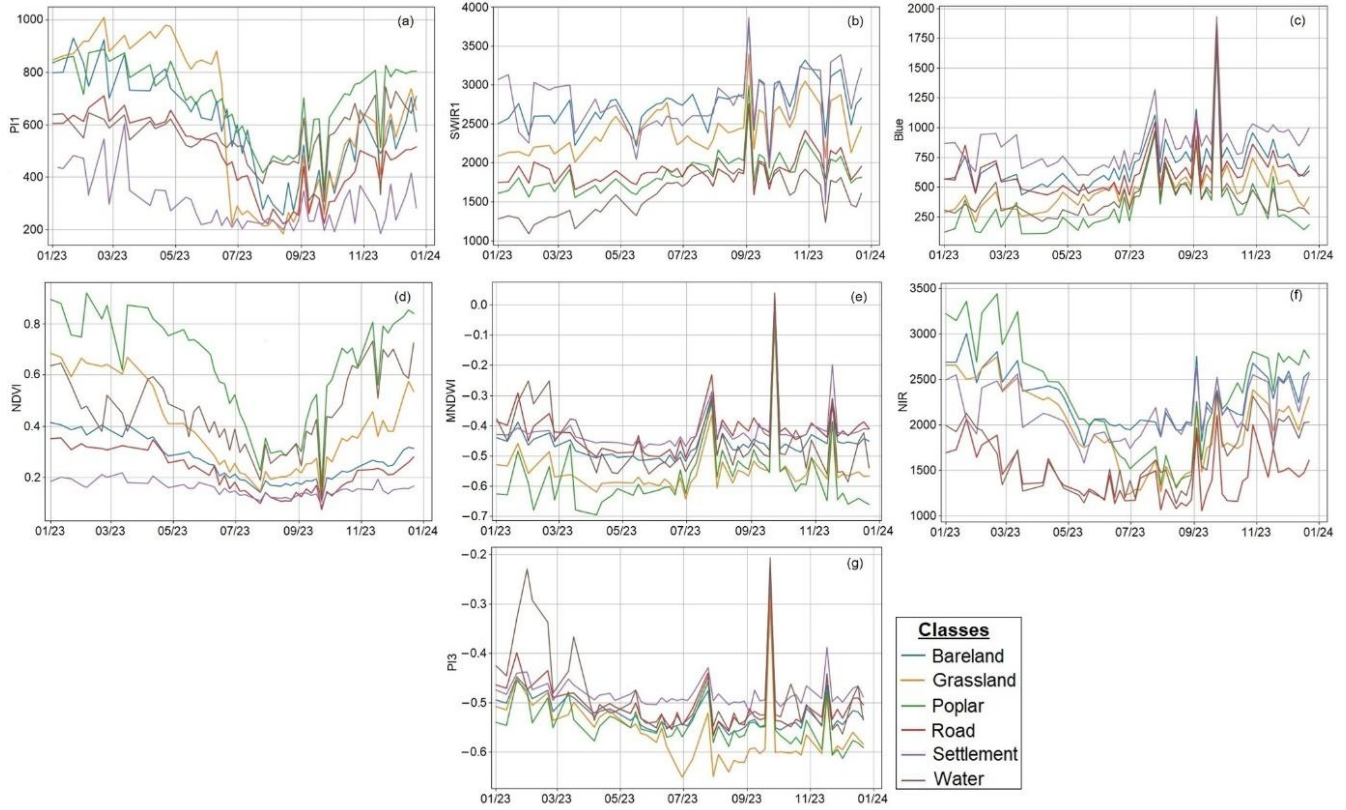


Figure 3.2 Spectral signatures for land cover classes ((a) PI-1, (b) SWIR-1, (c) Blue, (d) NDVI, (e) MNDWI, (f) NIR and (g) PI-3).

3.3.2 Influential importance

In Figure 3.3, the study shows the most important spectral bands and indices used to distinguish poplar from other types of land cover in mountainous regions for (a) Jan-Jun and (b) Jul-Dec 2023. Out of 16 spectral variables, only 7 was identified as important for the classification, however, their order of importance slightly varied in the first and second half of 2023. As Figure 3.3 (a) illustrates, the first half had SWIR-1 as highly ranked followed by the NDVI, blue band, PI-1, MNDWI, NIR, and PI-3. In the second, the NDVI was the most important in the classification process followed by the blue band, PI-3, MNDWI, SWIR-1, PI-1, and NIR Figure 3.3 (b).

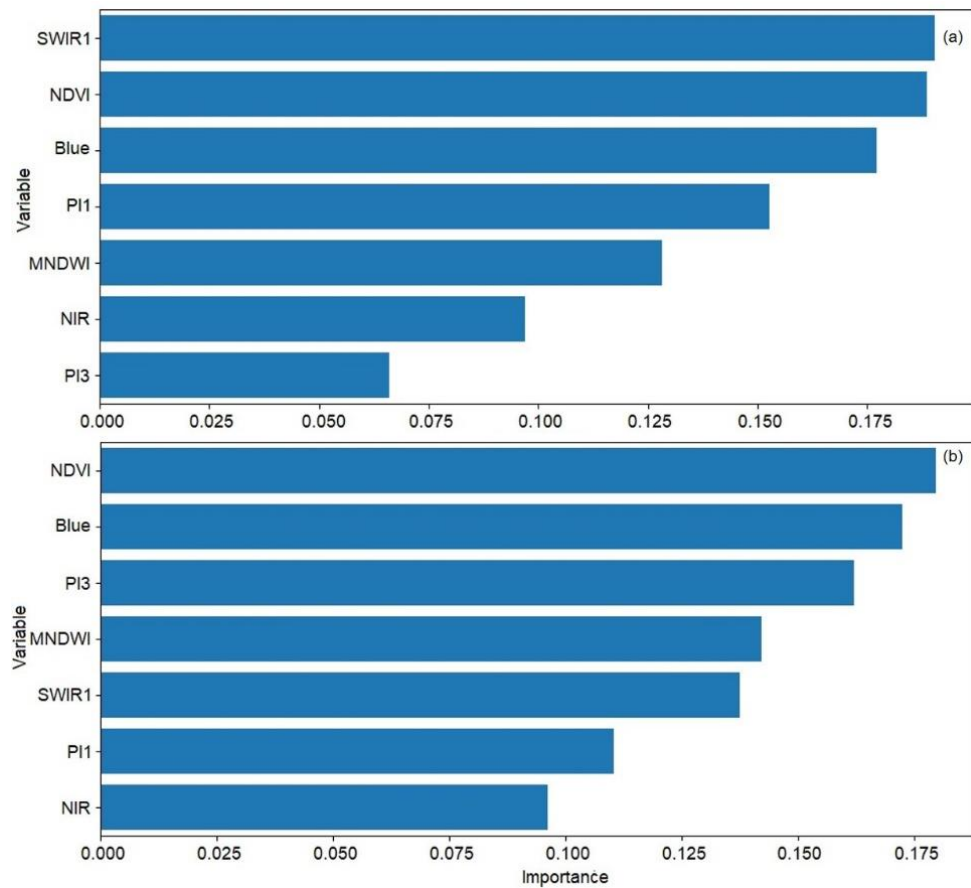


Figure 3.3 Variable importance for (a) Jan-Jun and (b) Jul-Dec as measured by the Random Forests algorithm indicating the usefulness of the different image bands and indices. The higher the value, the higher the predictor importance.

3.3.3 Classification algorithm performance

The study evaluated the classification results based on the combination of Sentinel-2 bands and indices, which resulted in an overall classification accuracy of over 80% for the Jan-Jun and Jul-Dec periods (Table 3.3). The accuracy for Jan-Jun classification was higher (UA (86–96%), PA (86–96%), OA (91%), and Kappa: 0.89) than for Jul-Dec (UA (73–93%), PA (79–94%), OA (87%), and Kappa: 0.87). These results suggest that the RF classifier and images acquired during the Jan-Jun period are more optimal for the classification and separability of poplar from other land cover types.

Table 3.3 Confusion matrix for (a) Jan-Jun and (b) Jul-Dec.

(a)	Bare land	Grassland	Settlement	Poplar	Roads	Water	UA
Bare land	181	6	2	5	3	1	91
Grassland	4	190	0	1	0	2	96
Settlement	7	1	141	0	2	0	93
Poplar	0	1	0	173	5	22	86
Road	2	0	4	3	92	0	91
Water	0	0	0	19	2	183	90
PA	93	96	96	86	88	88	-
OA				91			

(b)	Bare land	Grassland	Settlement	Poplar	Roads	Water	UA
Bare land	174	8	4	1	3	8	88
Grassland	8	184	0	2	0	3	93
Settlement	14	1	133	0	3	0	88
Poplar	4	2	0	147	9	39	73
Road	2	0	7	3	89	0	88
Water	1	0	0	18	0	185	91
PA	86	94	92	86	86	79	-
OA				87			

Figure 3.4 depict classification maps created from (a) Jan-Jun and (b) Jul-Dec Sentinel-2 imageries. Both maps exhibit similar spatial distribution patterns for different classes, with a slight overestimation of water, bare land, settlement, and poplar in the July-December images. Specifically, poplar appears to be concentrated along the riparian zones, which aligns with findings from previous studies on poplar distribution trends.

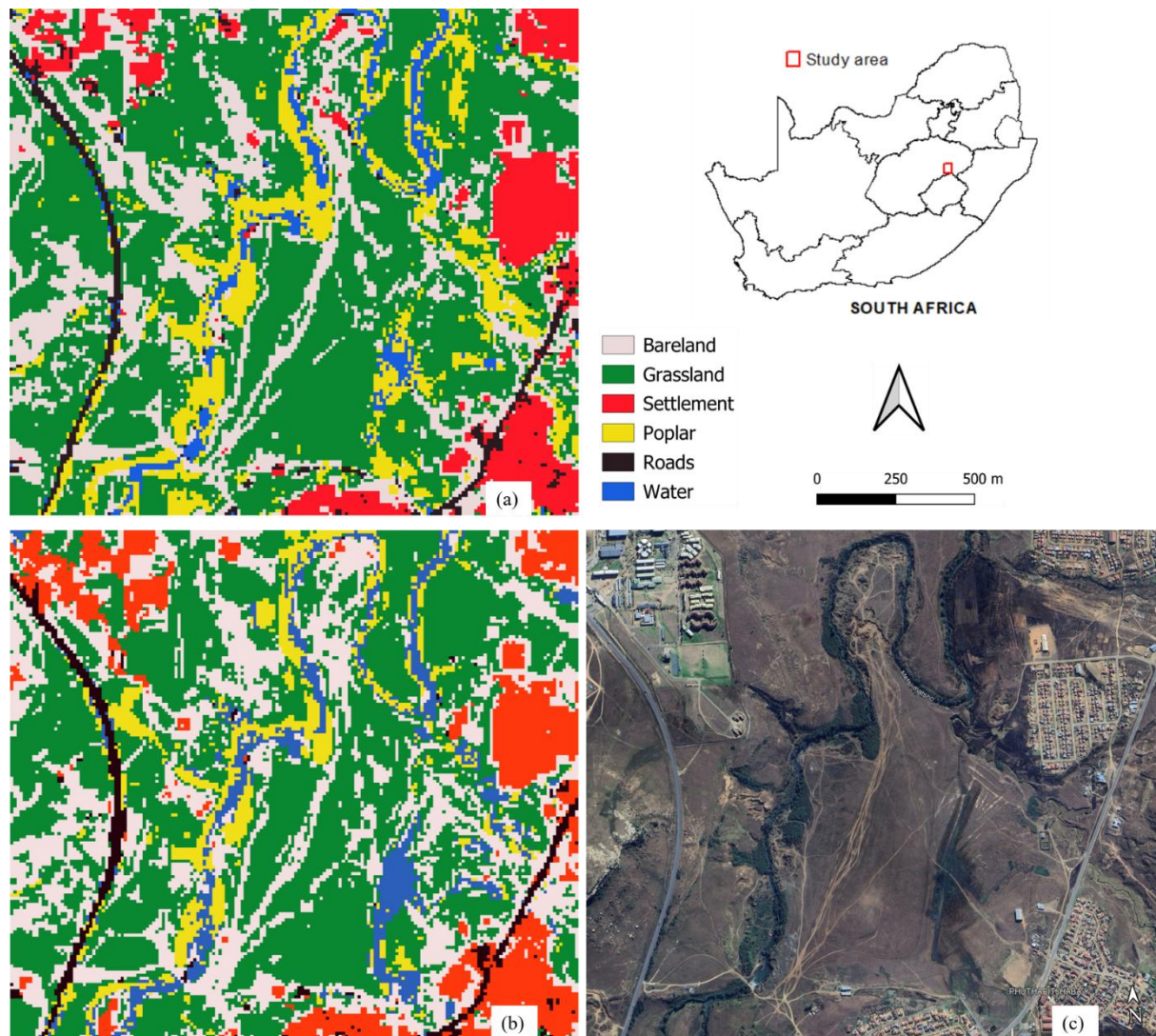


Figure 3.4 Random Forest classification maps for (a) Jan-Jun and (b) Jul-Dec Sentinel-2 imagery. True colour image (c) is included for visual inspection against classified maps.

3.4 DISCUSSION

In this study, Sentinel-2 satellite imagery in combination with the Random Forest classification algorithm was used to map the existence of poplar species in the mountainous landscapes of Phuthaditjhaba. The classification process involved imagery from the first and second half of the year, with the goal of assessing the mapping accuracy of both sets of images. Firstly, the study evaluated the distinguishability of poplar from other land cover types across the 61 different phases of the year through time series

analysis. Secondly, RF classifier was employed to identify the most contributing Sentinel-2 image bands and indices in the classification process for both halves of the year. Finally, the study computed and produced a classification map depicting the spatial distribution of poplar against other land cover features.

Time series analysis using indices and bands of Sentinel-2 imagery showed that there are specific months when poplar and other land cover types can be distinguished from each other, and there are also months when they are not easily separable. The study found that during summer and autumn, poplar and other land cover types have distinct spectral characteristics, while during most of the winter and spring, they exhibit somewhat similar spectral characteristics rendering them difficult to distinguish with high accuracy. These findings align with Cho et al. (2017), which observed that *Acacia spp* is distinct from surrounding features during the growing season (Nov-Jan) and peak period of vegetation productivity and less separable during winter period. The current study results provide detailed information on the bands and indices that are associated with high, medium, and low separability at different times of the year. This information is important for selecting the appropriate bands and indices for models that use images from periods where they are particularly distinctive and efficient for classification. It is therefore anticipated that this information will help develop robust models that can produce highly accurate results.

The current study assessed the important variables required for a robust model using Sentinel-2 bands and indices with a RF classifier to classify poplars against other land cover classes. We found that SWIR-1, NDVI, Blue, PI-1, MNDWI, NIR, and PI-3 are crucial parameters for this classification. These findings are consistent with a previous study by Kaila et al. (2020), which identified the Blue, SWIR, and NIR bands as important parameters for classifying *parthenium* weed. SWIR bands are effective for distinguishing classes (Atkinson et al. 2014), while NIR bands are sensitive to vegetation health and density (Addabbo et al. 2016). Additionally, the NIR band, combined with the red band, is used to calculate NDVI, which is utilized to measure vegetation greenness, biomass, and diversity (Michella et al. 2019; Puletti et al. 2019). Incorporating NDVI in a model enables discrimination of vegetation and non-vegetation features based on chlorophyll content sensitivity (Singh et al. 2016). PI-1 and -3 help highlight the difference between poplar and other landscape features (Hamrouni et al. 2022), making

them important variables for poplar spatial distribution modelling. The inclusion of MNDWI allows for the explicit differentiation of water features from other land cover types (Szabo et al. 2016). Since poplar is a water-dependent species found in riparian environments, the inclusion of MNDWI parameter is imperative to enhance discrimination efforts. The identification of these parameters has implications for the implementation of the methodology in various landscape settings as well as for the monitoring and management of poplar species, especially in mountainous areas.

Overall, the current study produced encouraging results with overall classification accuracy exceeding 80% in both study periods, where the performance from Jan to Jun was higher than Jul to Dec. As such, poplars were clearly separable from other landscape features, as bare patches and buildings/settlement were highly distinctive, leading to improved classification accuracy. However, during the winter and early spring periods, when vegetation loses its green colour and appears brownish like bare land and settlements, the classification accuracy is compromised. The high overall accuracy during Jan-Jun suggests that the combination of Sentinel-2 bands, derived indices, and RF can effectively map poplar in mountainous regions. These findings align with previous studies (see D'Amico et al. 2021; Kluczek et al. 2023; Wang et al. 2022). The produced maps could help fill the gap in updating local atlas inventories in South Africa where deficiencies exist in alien invasive plants land cover maps. While this study was conducted at a local spatial scale, the same approach and dataset (with more additions) could be used to classify other alien invasive plant species in mountainous and various other environmental settings at much larger spatial scale.

3.5 CONCLUSION

The current study aimed to assess the capability of Sentinel-2 spectral bands and derived indices, with RF classifier, to distinguish and map poplar trees in mountainous terrain from their surroundings. The study also aimed to identify the optimal period for effective and accurate poplar detection and mapping, while also determining the crucial spectral bands and indices to include in a model to maximize its performance. The study achieved results an overall accuracy of 91% and suggested that the optimal period for poplar detection and mapping as January-June than July-December. This study found SWIR, blue, NDVI, MNDWI, NIR, PI-1 and PI-3 to have high influence in the classification process. Poplar

was found to be more distinct from other classes with NDVI, blue, NIR and SWIR-1. These remarkable results strongly indicate that Sentinel-2 bands and indices with RF demonstrates significant practical and operational potential for poplar classification. Besides these practical benefits, the easy accessibility, wide coverage, and frequent revisits of Sentinel-2, combined with the reliability of RF, render them especially suitable for identifying invasive alien species in mountainous areas

3.6 REFERENCES

- Abdel-Rahman, E.M., Mutanga, O., Adam, E. and Ismail, R., 2014. Detecting *Sirex noctilio* grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 88, 48-59.
- Adam, E., Mureriwa, N. and Newete, S., 2017. Mapping *Prosopis glandulosa* (mesquite) in the semi-arid environment of South Africa using high-resolution WorldView-2 imagery and machine learning classifiers. *Journal of Arid Environments*, 145,43-51.
- Addabbo, P., Focareta, M., Marcuccio, S., Votto, C. and Ullo, S.L., 2016. Contribution of Sentinel-2 data for applications in vegetation monitoring. *Acta Imeko*, 5(2), 44-54.
- Agricultural Research Council Institute for Soil Climate and Water, South Africa., 2016. Meteorological data. Nearest Station: QwaQwa: UNIQWA; Station ID: 30426; Latitude: -28.48287; Longitude: 28.82522; Altitude: 1699 m.a.s.l. Date: 01/01/1986-31/08/2016.
- Andújar, D., Escolà i Agustí, A., Rosell Polo, J.R., Sanz Cortiella, R., Rueda-Ayala, V., Fernandez Quintanilla, C., Ribeiro, A. and Dorado, J., 2016. A LiDAR-based system to assess poplar biomass. *Gesunde Pflanzen*, 68, 155-162.
- Atkinson, J.T., Ismail, R. and Robertson, M., 2013. Mapping bugweed (*solanum mauritianum*) infestations in pinus patula plantations using hyperspectral imagery and support vector machines. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(1),17-28.
- Belgiu, M. and Drăguț, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of Photogrammetry and Remote Sensing*, 114, 24-31.
- Boegh, E., Soegaard, H., Broge, N., Hasager, C.B., Jensen, N.O., Schelde, K. and Thomsen, A., 2002. Airborne multispectral data for quantifying leaf area index, nitrogen concentration, and photosynthetic efficiency in agriculture. *Remote Sensing of Environment*, 81(2-3),179-193.

Bradley, B.A., Laginhas, B.B., Whitlock, R., Allen, J.M., Bates, A.E., Bernatchez, G., Diez, J.M., Early, R., Lenoir, J., Vilà, M. and Sorte, C.J., 2019. Disentangling the abundance–impact relationship for invasive species. *Proceedings of the National Academy of Sciences*, 116(20), 9919-9924.

Breiman, L., 2001. Random forests. *Machine Learning*, 45, 5-32.

Cartisano, R., Mattioli, W., Corona, P., Mugnozza, G.S., Sabatti, M., Ferrari, B., Cimini, D. and Giuliarelli, D., 2013. Assessing and mapping biomass potential productivity from poplar-dominated riparian forests: A case study. *Biomass and bioenergy*, 54, 293-302.

Centre for Environmental Management., 2003. State of the Namahadi and Harts rivers (October 2002 to March 2003): a biomonitoring report. Report prepared for Sedibeng Water, CEM, University of the Free State, Bloemfontein, South Africa.

Chianucci, F., Puletti, N., Grotti, M., Ferrara, C., Giorcelli, A., Coaloa, D. and Tattoni, C., 2020. Nondestructive tree stem and crown volume allometry in hybrid poplar plantations derived from terrestrial laser scanning. *Forest Science*, 66(6), 737-746.

Cho, M.A., Ramoelo, A. and Dziba, L., 2017. Response of land surface phenology to variation in tree cover during green-up and senescence periods in the semi-arid savanna of Southern Africa. *Remote Sensing*, 9(7), 689.

Congalton, R.G., 2001. Accuracy assessment and validation of remotely sensed and other spatial information. *International Journal of Wildland Fire*, 10(4), 321-328.

D'Amico, G., Francini, S., Giannetti, F., Vangi, E., Travaglini, D., Chianucci, F., Mattioli, W., Grotti, M., Puletti, N., Corona, P. and Chirici, G., 2021. A deep learning approach for automatic mapping of poplar plantations using Sentinel-2 imagery. *GIScience & Remote Sensing*, 58(8), 1352-1368.

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

Dongre, G. and Verma, R., 2021. Identification and estimation of poplar species area in Haryana and Punjab State Using Geo-Informatics. *Journal Remote Sensing and GIS*, 10(5), 287.

ESA. Sentinel Online. 2021. Available online: <https://sentinel.esa.int/web/sentinel/home> (accessed on 31 January 2021).

Gallardo, B., Bacher, S., Bradley, B., Comín, F.A., Gallien, L., Jeschke, J.M., Sorte, C.J. and Vilà, M., 2019. InvasiBES: Understanding and managing the impacts of Invasive alien species on Biodiversity and Ecosystem Services. *NeoBiota*, 50, 109-122.

- Hamrouni, Y., Paillassa, E., Chéret, V., Monteil, C. and Sheeren, D., 2022. Sentinel-2 poplar index for operational mapping of poplar plantations over large areas. *Remote Sensing*, 14(16), 3975.
- Han, H., Guo, X. and Yu, H., 2016. Variable selection using mean decrease accuracy and mean decrease gini based on random forest. *ICESSE*, 219-224.
- Haubrock, P.J., Cuthbert, R.N., Sundermann, A., Diagne, C., Golivets, M. and Courchamp, F., 2021. Economic costs of invasive species in Germany. *NeoBiota*, 67, 225-246.
- Haubrock, P.J., Cuthbert, R.N., Tricarico, E., Diagne, C., Courchamp, F. and Gozlan, R.E., 2021. The recorded economic costs of alien invasive species in Italy. *NeoBiota*, 67, 247-266.
- Hemmerling, J., Pflugmacher, D. and Hostert, P., 2021. Mapping temperate forest tree species using dense Sentinel-2 time series. *Remote Sensing of Environment*, 267, 112743.
- Hobbs, P., Colvin, C., Maherry, A., Bagan, R., Clarke, S., Petersen, C., Rose, R. and Carstens, M., 2013. Groundwater Reserve Determination for the Middle Vaal Water Management Area. *Water Research Commission*, 312, 13.
- Holloway, J. and Mengersen, K., 2018. Statistical machine learning methods and remote sensing for sustainable development goals: A review. *Remote Sensing*, 10(9), 1365.
- Kannan, R., Shackleton, C.M. and Uma Shaanker, R., 2013. Reconstructing the history of introduction and spread of the invasive species, *Lantana*, at three spatial scales in India. *Biological Invasions*, 15, 1287-1302.
- Kiala, Z., Mutanga, O., Odindi, J., Peerbhay, K.Y. and Slotow, R., 2020. Automated classification of a tropical landscape infested by *Parthenium* weed (*Parthenium hyterophorus*). *International Journal of Remote Sensing*, 41(22), 8497-8519.
- Kluczek, M., Zagajewski, B. and Zwijacz-Kozica, T., 2023. Mountain tree species mapping using Sentinel-2, PlanetScope, and airborne hypsypex hyperspectral imagery. *Remote Sensing*, 15(3), 844.
- Kovačević, J., Cvijetinović, Ž., Lakušić, D., Kuzmanović, N., Šinžar-Sekulić, J., Mitrović, M., Stančić, N., Brodić, N. and Mihajlović, D., 2020. Spatio-temporal classification framework for mapping Woody vegetation from multi-temporal Sentinel-2 imagery. *Remote Sensing*, 12(17), 2845.
- Labonté, J., Drolet, G., Sylvain, J.D., Thiffault, N., Hébert, F. and Girard, F., 2020. Phenology-based mapping of an alien invasive species using time series of multispectral satellite data: A case-study with glossy buckthorn in Québec, Canada. *Remote Sensing*, 12(6), 922.
- Langmaier, M. and Lapin, K., 2020. A systematic review of the impact of invasive alien plants on forest regeneration in European temperate forests. *Frontiers in Plant Science*, 11, 524969.

- Le Maitre, D.C., Forsyth, G.G., Dzikiti, S. and Gush, M.B., 2016. Estimates of the impacts of invasive alien plants on water flows in South Africa. *Water SA*, 42(4), 659-672.
- Linders, T.E.W., Schaffner, U., Eschen, R., Abebe, A., Choge, S.K., Nigatu, L., Mbaabu, P.R., Shiferaw, H. and Allan, E., 2019. Direct and indirect effects of invasive species: Biodiversity loss is a major mechanism by which an invasive tree affects ecosystem functioning. *Journal of Ecology*, 107(6), 2660-2672.
- Liu, H. and Motoda, H., 1998. *Feature Extraction, Construction and Selection: A Data Mining Perspective*; Kluwer Academic Publishers: Vancouver, BC, Canada, ISBN 0792381963.
- Liu, M., Cao, C., Chen, W. and Wang, X., 2019. Mapping canopy heights of poplar plantations in plain areas using ZY3-02 stereo and multispectral data. *ISPRS International Journal of Geo-Information*, 8(3), 106.
- Marbán, P.R., Mullinax, J.M., Resop, J.P. and Prosser, D.J., 2019. Assessing beach and island habitat loss in the Chesapeake Bay and Delmarva coastal bay region, USA, through processing of Landsat imagery: A case study. *Remote Sensing Applications: Society and Environment*, 16, 100265.
- Martin, F.M., Müllerová, J., Borgniet, L., Dommange, F., Breton, V. and Evette, A., 2018. Using single-and multi-date UAV and satellite imagery to accurately monitor invasive knotweed species. *Remote Sensing*, 10(10), 1662.
- Masemola, C., Cho, M.A. and Ramoelo, A., 2020. Towards a semi-automated mapping of Australia native invasive alien *Acacia* trees using Sentinel-2 and radiative transfer models in South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166, 153-168.
- Mbedzi, M., Tshisikhawe, P. and Sinthumule, I., 2019. The rate of recolonization by native plant species after the Eradication of the invasive alien plant species (*Populus alba* L.) in the Limpopo Province, South Africa.
- Michella, A., Vescovi, F.D. and Mustafee, I., 2019. Sentinel-2 Vegetation Health Index and Early-Onset Drought Detection in Kenya. *Airbus Defence and Space*, 7.
- Mucina, L. and Rutherford, M. C., 2006. The vegetation of South Africa, Lesotho and Swaziland. *Strelitzia*, 19, 1-30.
- Ngondya, I.B. and Munishi, L.K., 2021. Impact of invasive alien plants *Gutierrezia cordifolia* and *Tagetes minuta* on native taxa in the Ngorongoro crater, Tanzania. *Scientific African*, 13, e00946.
- Pal, M. and Foody, G.M., 2010. Feature selection for classification of hyperspectral data by SVM. *IEEE Transactions on Geoscience and Remote Sensing*, 48(5), 2297-2307.
- Paz-Kagan, T., Silver, M., Panov, N. and Karnieli, A., 2019. Multispectral approach for identifying invasive plant species based on flowering phenology characteristics. *Remote Sensing*, 11(8), 953.

- Peña, J.M., de Castro, A.I., Torres-Sánchez, J., Andújar, D., San Martín, C., Dorado, J., Fernández-Quintanilla, C. and López-Granados, F., 2018. Estimating tree height and biomass of a poplar plantation with image-based UAV technology. *AIMS Agric. Food*, 3(3), 313-326.
- Poland, T.M., Patel-Weynand, T., Finch, D.M., Miniati, C.F., Hayes, D.C. and Lopez, V.M., 2021. Invasive species in forests and rangelands of the United States: A comprehensive science synthesis for the United States forest sector. *Springer Nature*, 455.
- Poona, N.K., Van Niekerk, A., Nadel, R.L. and Ismail, R., 2016. Random forest (RF) wrappers for waveband selection and classification of hyperspectral data. *Applied spectroscopy*, 70(2), 322-333.
- Poynton, R.J., 2009. Tree planting in Southern Africa. Volume 3: other genera.
- Puletti, N., Mattioli, W., Bussotti, F. and Pollastrini, M., 2019. Monitoring the effects of extreme drought events on forest health by Sentinel-2 imagery. *Journal of Applied Remote Sensing*, 13(2), 020501-020501.
- Robinson, T.B., Peters, K. and Brooker, B., 2020. Coastal invasions: the South African context. In *Biological Invasions in South Africa*. Springer International, 229-247.
- Rouse Jr, J.W., Haas, R.H., Schell, J.A. and Deering, D.W., 1973. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation (No. NASA-CR-132982).
- Royimani, L., Mutanga, O., Odindi, J., Zolo, K.S., Sibanda, M. and Dube, T., 2019. Distribution of *Parthenium hysterophoru* L. with variation in rainfall using multi-year SPOT data and random forest classification. *Remote Sensing Applications: Society and Environment*, 13, 215-223.
- Sammut, C. and Webb, G.I., 2017. *Encyclopedia of machine learning and data mining*. Springer Publishing Company, Incorporated.
- SANParks, 2020. Golden Gate Highlands National Park Management Plan. SANParks. Pretoria.
- Shah, K.K., Tiwari, I., Tripathi, S., Subedi, S. and Shrestha, J., 2020. Invasive alien plant species: A threat to biodiversity and agriculture in Nepal. *Agriways*, 8(1), 62-73.
- Singh, R.P., Singh, N., Singh, S. and Mukherjee, S., 2016. Normalized difference vegetation index (NDVI) based classification to assess the change in land use/land cover (LULC) in Lower Assam, India. *International Journal of Advanced Remote Sensing and GIS*, ISSN 2320 – 0243, 5(10), 1963-1970.
- Slater, R., 2002. Between a rock and a hard place: contested livelihoods in Qwaqwa National Park, South Africa. *Geographical Journal*, 168(2), 116-129.

- Steenhuisen, S., Clark, V.R., Martin, G. and Cole, N., 2022. The current status of invasive alien plants in build-up areas within Golden Gate Highlands National Park. University of the Free State (Department of Plant Sciences), 1-18.
- Szabo, S., Gácsi, Z. and Balazs, B., 2016. Specific features of NDVI, NDWI and MNDWI as reflected in land cover categories. *Landscape & Environment*, 10(3/4), 194-202.
- Taylor, S.J., 2023. Exploring the of establishment of an urban forest in Phuthaditjhaba to Create a More Sustainable Future Urban Environment. In *Sustainable Futures in Southern Africa's Mountains: Multiple Perspectives on an Emerging City*. Springer International, 153-169.
- Wang, M., Li, M., Wang, F. and Ji, X., 2022. Exploring the optimal feature combination of tree species classification by fusing multi-feature and multi-temporal sentinel-2 data in Changbai mountain. *Forests*, 13(7), 1058.
- Wang, M., Zheng, Y., Huang, C., Meng, R., Pang, Y., Jia, W., Zhou, J., Huang, Z., Fang, L. and Zhao, F., 2022. Assessing Landsat-8 and Sentinel-2 spectral-temporal features for mapping tree species of northern plantation forests in Heilongjiang Province, China. *Forest Ecosystems*, 9, 100032.
- Wang, R., Li, G., Lu, Y. and Lu, D., 2023. A comparative analysis of grid-based and object-based modeling approaches for poplar forest growing stock volume estimation in plain regions using airborne LIDAR data. *Geospatial Information Science*, 1-19.
- White, J.C., Penner, M. and Woods, M., 2021. Assessing single photon LiDAR for operational implementation of an enhanced forest inventory in diverse mixedwood forests. *Forestry Chronicle*, 97(1), 78-96.
- Xie, Y., Sha, Z. and Yu, M., 2008. Remote sensing imagery in vegetation mapping: a review. *Journal of plant ecology*, 1(1), 9-23.
- Xu, D., Wang, H., Xu, W., Luan, Z. and Xu, X., 2021. LiDAR applications to estimate forest biomass at individual tree scale: Opportunities, challenges and future perspectives. *Forests*, 12(5), 550.
- 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), 3025-3033.
- Zhang, P., Li, B., Wu, J. and Hu, S., 2019. Invasive plants differentially affect soil biota through litter and rhizosphere pathways: a meta-analysis. *Ecology Letters*. 22(1), 200-210.
- Zhang, Z. and van Kleunen, M., 2019. Common alien plants are more competitive than rare natives but not than common natives. *Ecology Letters*. 22(9), 1378-1386.

Zhu, L., Liu, X., Wu, L., Tang, Y. and Meng, Y., 2019. Long-term monitoring of cropland change near Dongting Lake, China, using the LandTrendr algorithm with Landsat imagery. *Remote Sensing*, 11(10), 1234.

Zwiggelaar, R., 1998. A review of spectral properties of plants and their potential use for crop/weed discrimination in row-crops. *Crop protection*. 17(3), 189-206.

Appendix A

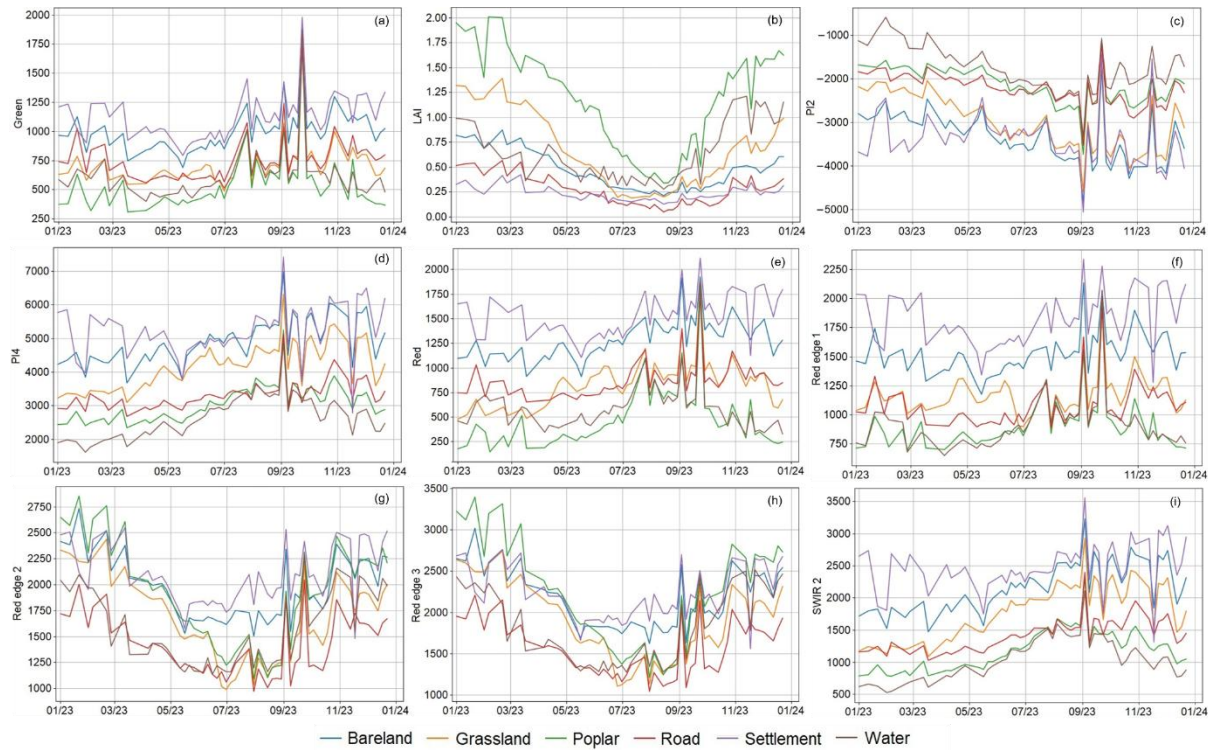


Figure 3.5 Spectral signatures for land cover classes ((a) Green, (b) LAI, (c) PI-2, (d) PI-4, (e) Red, (f) Red edge-1, (g) Red edge-2, (h) Red edge-3 and (i) SWIR-2)

CHAPTER 4 MODELLING POSSIBLE HABITATS FOR POPLAR INVASION IN SOUTH AFRICA

KOEDOE - African Protected Area Conservation and Science

ISSN: (Online) 2071-0771, (Print) 0075-6458



Page 1 of 14

Original Research

Modelling possible habitats for poplar invasion in South Africa



Authors:
Morena Mapuru^{1,2}
Sifiso Xulu¹
Michael Gebreslasie¹

Affiliations:
¹Discipline of Geography,
School of Agriculture, Earth,
and Environmental Sciences,
University of KwaZulu-Natal,
Durban, South Africa

Poplars (*Populus alba*, *Populus canadensis*, *Populus canescens*, *Populus deltoides*, *Populus fremontii*, *Populus nigra* and *Populus simonii*) are found throughout the world and are invasive in South Africa, where they are spatially permitted in certain areas under controlled conditions, as specified in the country's invasive species legislation. To better trace their geographic distribution, this study predicts the potentially suitable habitat of poplar trees in South Africa based on generalised linear model (GLM), Random Forests (RF) and Support Vector Machines (SVM) models and also assesses the climatic variables with the greatest impact on prediction performance. The results show excellent performance for all models (Area Under the Receiver Operation Characteristics Curve [AUC] > 0.9) in predicting the poplar distribution, with RF achieving the best performance ($r = 0.83$ and $AUC = 0.965$), followed by SVM ($r = 0.72$ and

Link: <https://journals.co.za/doi/full/10.4102/koedoe.v66i1.1783>

This Chapter is based on:

Mapuru, M., Xulu, S. and Gebreslasie, M., 2024. Modelling possible habitats for poplar invasion in South Africa. *KOEDOE-African Protected Area Conservation and Science*, 66(1), p.1783.

ABSTRACT

Poplars (*Populus alba*, *Populus canadensis*, *Populus canescens*, *Populus deltoides*, *Populus fremontii*, *Populus nigra* and *Populus simonii*) are found throughout the world and are invasive in South Africa, where they are spatially permitted in certain areas under controlled conditions, as specified in the country's invasive species legislation. To better trace their geographic distribution, this study predicts the potentially suitable habitat of poplar trees in South Africa based on generalised linear model (GLM), Random Forests (RF) and Support Vector Machines (SVM) models and also assesses the climatic variables with the greatest impact on prediction performance. The results show excellent performance for all models (Area Under the Receiver Operation Characteristics Curve [AUC] > 0.9) in predicting the poplar distribution, with RF achieving the best performance ($r = 0.83$ and $AUC = 0.965$), followed by SVM ($r = 0.72$ and $AUC = 0.959$) and then GLM ($r = 0.65$ and $AUC = 0.937$). In a geographical perspective, all models show a similar pattern, with the highest concentrations being in the south-western parts of the Western Cape, the Southern Cape on the Garden Route, the central-eastern Free State, Mpumalanga, and the southern parts of Limpopo. The evaluation of the relative importance of the bioclimates used showed that the warmest and driest quarter's precipitation and annual precipitation significantly contribute to the poplar population. These results demonstrate the power of machine learning and regression models for predicting suitable habitats and extracting valuable environmental-climatic knowledge for monitoring and managing invasive tree species such as poplars.

Conservation implications: Poplars are among the most aggressive invasive plant species in South Africa. The results of this study are expected to help conservation authorities understand the current climatic factors affecting the species distribution, as well as potential sites.

Keywords: species distribution model; poplars; random forest; support vector machine; generalized linear model; bioclimates; alien invasive species.

4.1 INTRODUCTION

Species distribution modelling (SDM) has emerged as a powerful tool in ecological research (Peterson et al. 2011), greatly facilitating the understanding of geo-ecology and predicting the potentially suitable environment for species (Stewart et al. 2022). In principle, SDMs link defined species locations to predictor variables to evaluate patterns of species occurrence and habitat suitability (Elith et al. 2010; Guisan et al. 2017; Wolmarans et al. 2010). The SDMs may even be used to predict species survival in response to future environmental shifts (Peterson et al. 2011). While SDMs often integrate climatic variables such as temperature and precipitation (Zhang et al. 2020), they also include other abiotic variables such as soil, land use, land cover and topography (Buri et al. 2017; Dubuis et al. 2013). Because of their fine spatial resolution for better fluctuations, the latter variables are included when the study aims to create SDM at local scale. Given their unique modelling ability, SDMs have become a popular strategy and a most widely used tool for characterizing species distribution and predicting habitat suitability (Araújo et al. 2019).

Advances in computing resources and machine learning algorithms, as well as the availability of fine-scale remote sensing products and climate data are an important step forward in the use of SDMs for various applications (Guisan et al. 2017; Kass et al. 2021; Thuiller et al. 2009). These developments offer the opportunity to implement SDMs with great effect because the resolution at which they are computed greatly influences their overall performance (Lee-Yaw et al. 2022). Besides, the growing utility of SDMs lies in their relatively simple application with publicly available software packages, as well as low data requirements (Zurell, 2019) and guidelines (Guisan et al. 2017).

Several studies have shown the high model performance in SDMs applications. In this regard, Zhang et al. (2020) predicted the possible distribution of *Anredera cordifolia* using Random Forest (RF) and identified the temperature as the most important factor influencing the predictive performance of this species. Yudaputra et al. (2019) predicted the potential current distribution of *Guettarda speciosa* in Indonesia using the Maximum Entropy (MaxEnt), Support Vector Machine (SVM), RF, Generalized Linear Model (GLM), Domain, and Bioclim. Their study showed that MaxEnt outperformed other competitors with the highest area under the receiver operation characteristics curve (AUC) value of 0.89.

Kaky et al. (2020) evaluated the performance of ensemble and MaxEnt modelling approaches in predicting the potential spread of Egyptian medicinal plants, where they found RF and MaxEnt performing better than SVM and Classification and Regression Tree (CART) in predicting the potential distribution of target species. However, the low performance of some models does not make them weak modelling approaches. Some models will perform poorly if all competency requirements are not met. For example, SVM sometimes performs better than MaxEnt and RF, possibly because the data structure is conducive to optimal performance. Therefore, ensuring that the model requirements are fully met before application and evaluation is essential. In general, studies have demonstrated high potential of machine learning and regression techniques for predicting species distributions in landscapes and ranking explanatory factors affecting distribution trends (Ahmed et al. 2021; Shabani et al. 2016). While these have been widely and successfully used to quantify the dispersal of many (native and non-native) floral species, they have not been applied in the study of Poplars in the Global South regions.

In South Africa, *Populus alba* and *Populus canescens* are classified as a Category 2 weed under the *National Environmental Management and Biodiversity Act (NEMBA) (No. 10 of 2004)* and *Conservation of Agricultural Resources Act (CARA) (No. 43 of 1983)* of South Africa's Invasive Species Legislation (Henderson, 2007). This means that poplars (Category 2) plants may not occur on land or inland water surfaces other than a demarcated area or a biological control reserve (Roy et al. 2023). For example, in most conservation areas under South African National Parks (SANParks), they are classified as a species of very high concern, encroaching on ecologically sensitive landscapes such as riparian zones; they are therefore cleared under continuous surveillance (SANParks, 2020). Essentially, poplar species consume a lot of water in streams, thereby worsening the conditions for aquatic species in the streams (Thérout Rancourt et al. 2015). Like any woody species, poplar competes with native species for resources, with a better chance of consuming more and more because of its growth structure and crown, which intercepts rain and sunlight that would otherwise be used by native species (Kumschick et al. 2020; Zhang et al. 2021). For these reasons, it is important to understand their spatial distribution and climatic drivers so that they can be appropriately monitored and managed in the country. This study was conducted on a national scale because at smaller spatial scale the bioclimatic variables do not give better variations (Fournier et al. 2017).

To achieve the aim of this study, the performance of GLM, RF and SVM models in predicting the existence and distribution of poplar trees in South Africa was tested and compared. The predictive power of each model under current different climatic conditions and evaluation of the most influential climatic parameters for Poplar distribution in the country was examined. The results of this study could provide valuable insights to support alien invasive species management actions and decision-making processes.

4.2 MATERIALS AND METHODS

4.2.1 Study area

This study was conducted in South Africa, the southernmost country on the African continent (Figure 4.1). The country lies between the 22–35° S and the 16–33° E (Van Wilgen et al. 2020). The country's interior consists of a large, almost flat plateau with an altitude of about 1000 m in most places above sea level (King, 1942). South Africa has a temperate climate with warm and humid summers and cold and dry winters (Blamey et al. 2017). The country's annual average rainfall is 464 mm, with the Western Cape receiving winter rainfall (June to August) and the rest of the country receiving summer rainfall (December to February) (Roffe et al. 2019). South Africa consists of nine vegetation biomes or units: Forests, Savannah, Fynbos, Grassland, Karoo, Desert, Nama-Karoo, Succulent Karoo, Indian Ocean Coastal Belts and Albany Thicket (Dayaram et al. 2019; Mucina and Rutherford, 2006). Large impact craters, orogenic belts, granite belts and cratons are only a few examples of South Africa's extremely diverse geology (Du Toit, 1940). South Africa's topography is made up of the craggy mountains of the Cape Fold Belt and the great escarpment, which encircle the country to the west, south and southeast. There is a narrow coastal plain strip beyond this (Haughton, 1969).

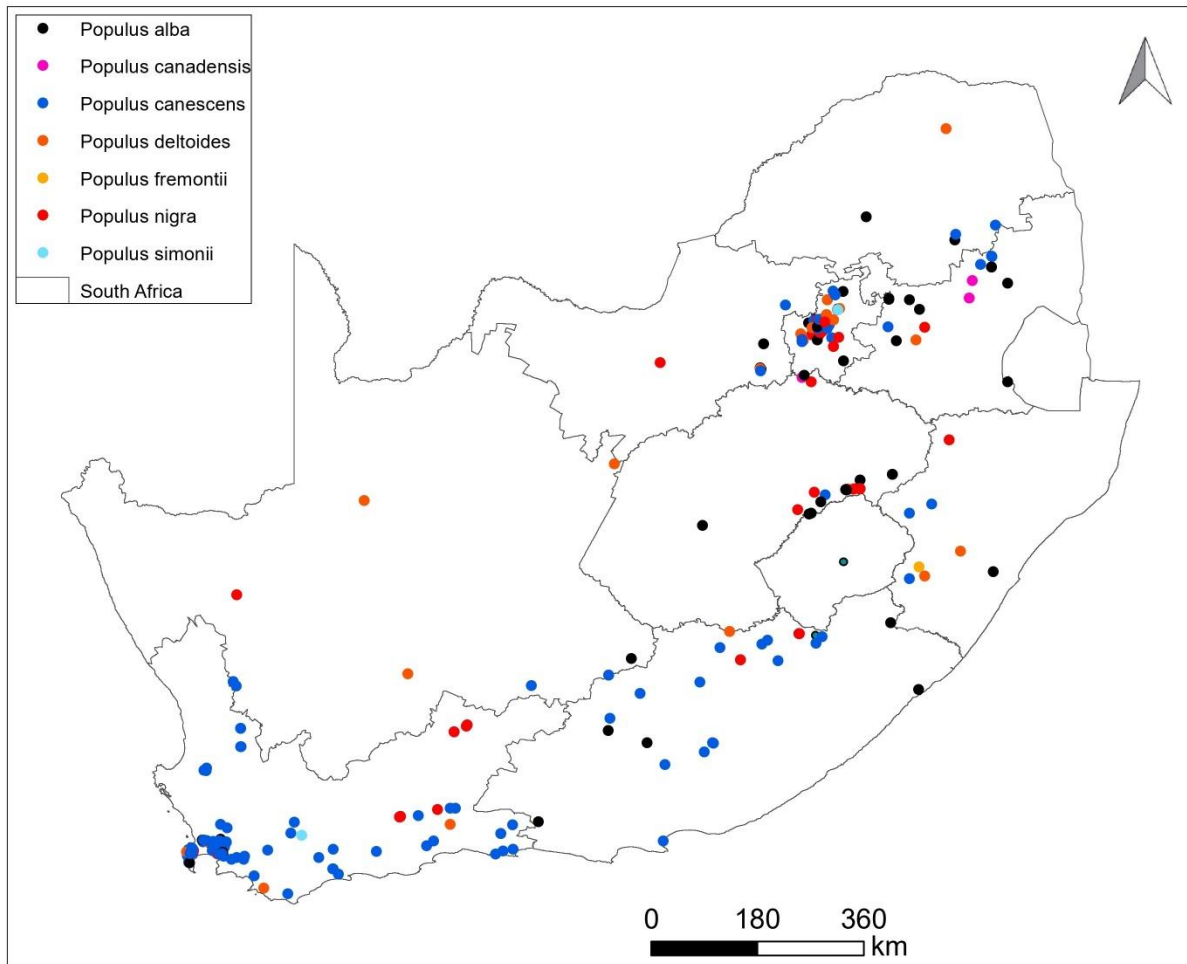


Figure 4.1 Geographic distribution of Poplars in South Africa.

4.2.2 Data collection

To quantitatively predict poplar (*Populus nigra*, *Populus simonii*, *Populus fremontii*, *Populus deltoides*, *P. canescens*, *Populus canadensis* and *P. alba*) distribution in South Africa, spatially referenced poplar occurrence sample points were obtained from the Global Biodiversity Information Facility (GBIF) (<https://www.gbif.org/> accessed on 12 June 2023 and updated on 20 December 2023). After filtering the data by South African boarder and removing duplicates, 467 poplar occurrence points were finally obtained and eligible for analysis. The species breakdown in terms of their locations is shown in Table 4.1. About 600 absence points were generated using the random point function in *R* environment.

This study also attempted to identify the South African protected areas most likely to be affected by poplars by superimposing the best-modelled poplar distribution results on the protected area shapefiles and extracted protected areas where poplar occur using intersection function in Quantum Geographic Information System (QGIS). Only mostly affected protected areas are reported; however, the wider distribution is provided.

Table 4.1 Poplar species gathered from GBIF, including localities and counts.

Poplar species	Province	No.
<i>Populus simonii</i>	Western Cape	2
<i>Populus nigra</i>	Northern Cape, North West, Free State, KwaZulu-Natal, Western Cape, Eastern Cape, Mpumalanga, Gauteng and Limpopo	58
<i>Populus fremontii</i>	KwaZulu-Natal	1
<i>Populus deltoides</i>	Northern Cape, KwaZulu-Natal, Western Cape, Eastern Cape, Mpumalanga, Gauteng and Limpopo	36
<i>Populus canescens</i>	Northern Cape, North West, Free State, KwaZulu-Natal, Western Cape, Eastern Cape, Mpumalanga, Gauteng and Limpopo	309
<i>Populus canadensis</i>	Free State and Mpumalanga	3
<i>Populus alba</i>	Northern Cape, North West, Free State, KwaZulu-Natal, Western Cape, Eastern Cape, Mpumalanga, Gauteng and Limpopo	58
Total		467

4.2.3 Environmental predictors

The Bioclim algorithm was used in this study; this is a commonly used climate-envelope-model (Booth et al. 2014) integrated into the R environment using a raster package, for environmental prediction from the Worldclim database (Table 4.2). Soil, land-use/cover, and topography have been documented as

important environmental variables for plant species distribution (Huang et al. 2021; Krauss et al. 2008). However, in this study, the focus was only on the current climate variables affecting the poplar species' range distribution in South Africa.

Table 4.2 Bioclimatic variables and their descriptions

Bioclimatic variables	Description
Bio 1	Annual Mean Temperature
Bio 2	Mean Diurnal Range
Bio 3	Isothermality
Bio 4	Temperature Seasonality
Bio 5	Maximum Temperature of the Warmest Month
Bio 6	Minimum Temperatures of the Coldest Month
Bio 7	Temperature Annual Range
Bio 8	Mean Temperature Wettest Quarter
Bio 9	Mean Temperature Driest Quarter
Bio 10	Mean Temperature of the Warmest Quarter
Bio 11	Mean Temperature of the Coldest Quarter
Bio 12	Annual Precipitation
Bio 13	Precipitation of Wettest Month
Bio 14	Precipitation of Driest Month
Bio 15	Precipitation Seasonality
Bio 16	Precipitation of Wettest Quarter
Bio 17	Precipitation of Driest Quarter
Bio 18	Precipitation of Warmest Quarter
Bio 19	Precipitation of Coldest Quarter

All Bioclim variables were spatially scaled to a 1 km grid and cropped to the extent of the South African border. Temperature Bioclim in the Worldclim database is usually multiplied by 10 for storage purposes to reduce size. Temperature units were converted to 78 degrees Celsius.

4.2.4 Modelling methods

This study was conducted using all available poplar species data as indicated in section ‘Distribution of poplar in South Africa’. Some species (e.g. *Populus fremontii*) had insufficient occurrence data and were unable to meet the modelling requirements for separate analyses. The commonly applied models for species distribution that include the GLMs (Melo-Merino et al. 2020), RFs (Breiman, 2001) and SVM (Vapnik, 1998) were used in the R statistical environment. Stepwise logistic regression was used to select the best model combination. This process was set to be repeated 10 times in a row to ensure that robust modelling output is achieved. The predictors from the model that yielded the lowest akaike information criterion (AIC) score was selected for modelling. This process is important as it allows for proper fitting of the model, which is a model that does not possess multicollinearity and data redundancy.

For all models, the same initial set of inputs was included, that is, poplar occurrence and absence points as response variable and 13 selected Bioclims as environmental predictors. Species presence was given value 1 and absence was given value 0 for modelling. Distribution data were divided into two parts, 75% for training and 25% for testing. The GLM was implemented using the ‘GLM’ function, link function and error distribution are given with the family arguments. Random Forests was implemented with ‘randomForest’ and the SVMs in R using the ‘ksvm’ in package ‘kernlab’ because it contains many different SVM formulations and ‘kernels’ and provides useful options and functions such as a method for plotting. Next, the RF was used to evaluate the most important variables (Genauer et al. 2010) in determining the occurrence and distribution of poplar in the South African landscape. All the models were set to repeat 10 times to ensure that our modelling results are consistent and accurate. The accuracy of all models was assessed by the area under the curve (AUC).

4.3. RESULTS

4.3.1 Model performance and validation

The geographic distribution of poplar trees (*P. alba*, *P. canadensis*, *P. canescens*, *Populus deltoides*, *P. fremontii*, *Populus nigra* and *Populus simonii*) were modelled in South Africa using GLM, RFs and

SVM models, and their performance is shown in Table 3. The results show excellent performance for all models, with RF achieving the highest correlation of 0.83 and 0.965 AUC. This was followed by SVM and then GLM with correlation and AUC of 0.72 and 0.959, and 0.65 and 0.937, respectively (Table 4.3). A key feature of these results is the superior performance of machine learning methods than a regression model. In Figure 4.2, the comparison of the model accuracies with the receiver operator characteristics (ROC) curve is shown. Here, ROC shows the proportion of the true positive (presence) rate and the false positive (absence) rate for poplars. Higher RF AUC values indicate the ability to identify the presence and absence of poplar samples (Figure 4.2a), followed by SVM (Figure 4.2b) and then GLM (Figure 4.2c).

Table 4.3 Summary of model performance

Model	Correlation	AUC
Random Forests	0.83	0.965
Support Vector Machine	0.72	0.959
Generalized Linear Model	0.65	0.937

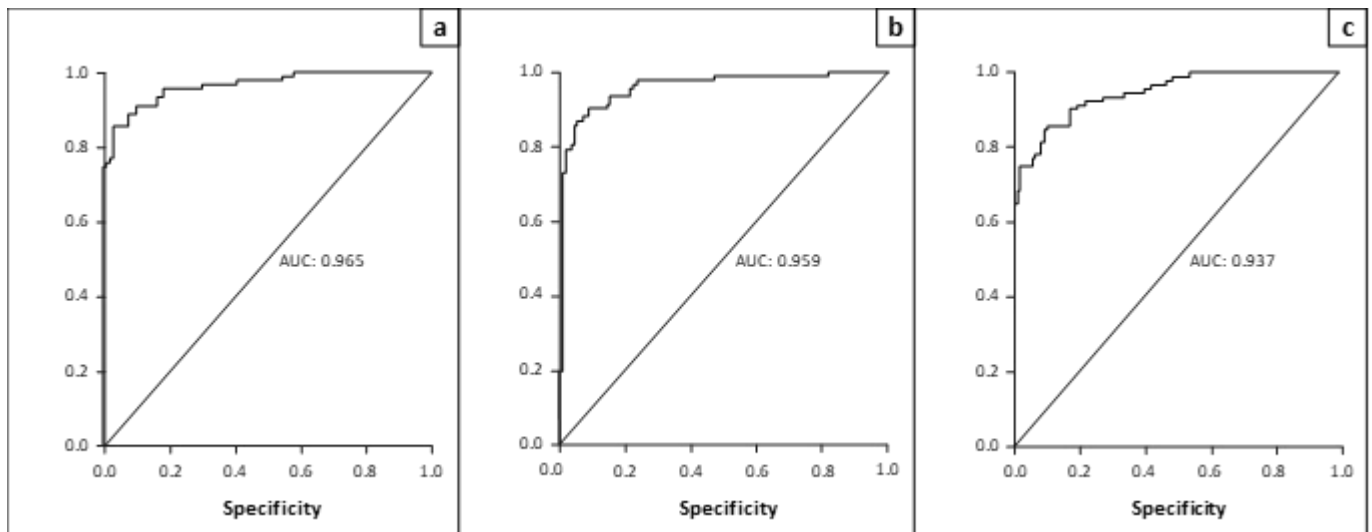


Figure 4.2 Receiver operator characteristics (ROC) curve for (a) Random Forests, (b) Support Vector Machine, and (c) Generalized Linear Model

4.3.2 Distribution of poplar in South Africa

In this section, the results of the modelled poplar distribution are presented for South Africa using RF, SVM and GLM, and the results are shown in Figure 4.3. All models showed a similar pattern of poplar distribution, with higher concentration in the south-western part of the Western Cape, the southern Cape on the Garden Route, the central-eastern Free State, and the north-eastern part of the country in Mpumalanga, even in Limpopo (Figure 4.3). However, RF had a better prediction performance than its counterparts (see Figure 4.2).

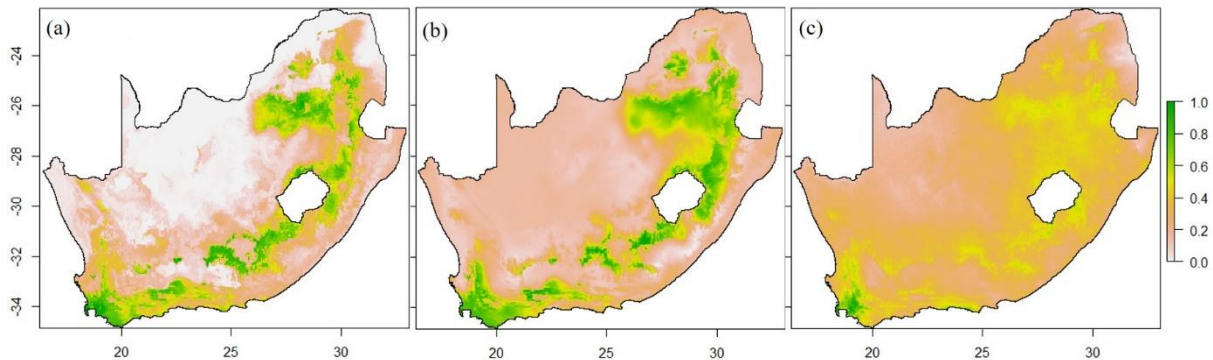


Figure 4.3 Modelled distribution of poplars in South Africa using (a) Random Forests, (b) Support Vector Machine, and (c) Generalized Linear Model. The legend/ramp on the y axis represent suitability levels, with green representing highly suitable site, while white-pinkish colour represents low suitability areas.

Populus nigra, *P. alba*, and *P. canescens* appear to have suitable habitats throughout the rest of the country, except for the extremely hot and arid northern parts of the Northern Cape (Figure 4.4a). This conclusion was reached after overlaying all poplar species points on the results of a random forests modelling application. Many provinces have populations of *Populus deltoides*, however it has not been recorded in provinces such as the Northwest and the Free State. Nevertheless, there are just a few records of *Populus simonii* (Western Cape), *P. fremontii* (Kwazulu Natal), and *P. canadensis* (Free State and Mpumalanga) in South Africa (see Figure 4.4a below). Overall, poplar species appear to be highly distributed in temperate regions with adequate rainfall patterns and less in arid deserted areas (see

Figure 4.4b and Appendix C). This is not surprising since majority of poplar species require an inordinate amount of water to flourish and survive.

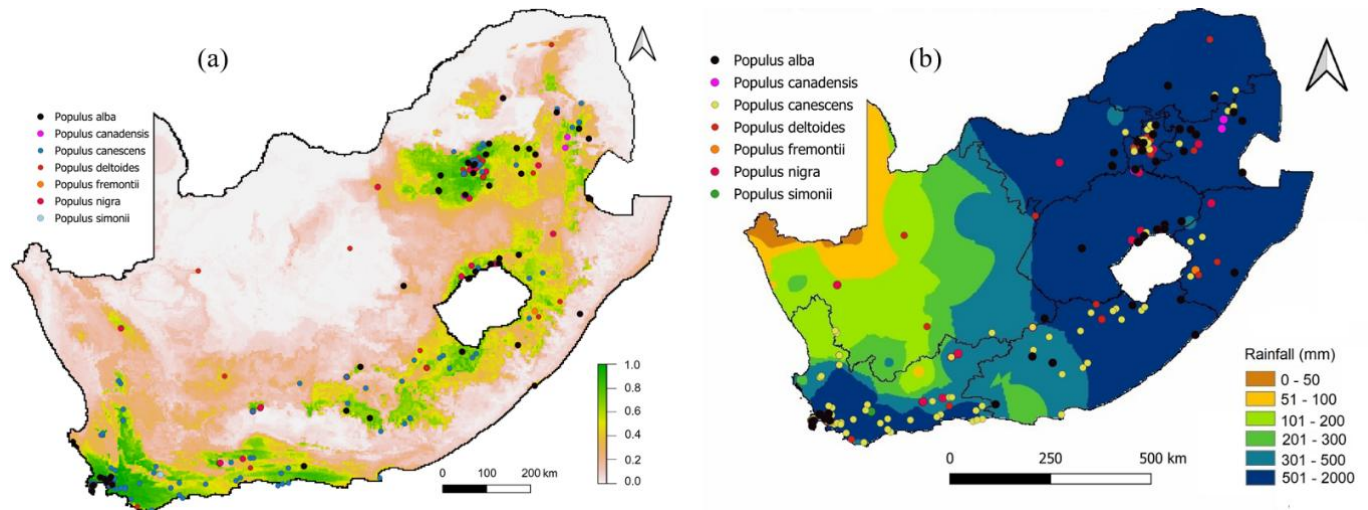


Figure 4.4 a) RF prediction map and b) a June 2022 to June 2023 rainfall map with poplar points overlaid on both.

In assessing the occurrence status of poplars in South African protected areas, this study found that poplars are common in few protected areas, where conditions are more favourable. These include, Golden Gate Highlands National Park, Table Mountain National Park, Cape Peninsula Nature Reserve, Tweefontein Reserve, and the Platberg Private Nature Reserve to name but a few (see appendix A and B).

Of the 13 predictors used, Bio 4 (temperature seasonality), Bio 10 (mean temperature of the warmest quarter), Bio 11 (mean temperature of the coldest quarter), Bio 12 (annual precipitation), Bio 13 (Precipitation of wettest month), Bio 14 (precipitation of the driest month), Bio 17 (precipitation of the driest quarter), and Bio 18 (precipitation of the warmest quarter) were found to be significant in determining the occurrence status of poplars (Table 4.4). The results seem to indicate that seasonal variation in temperatures and average temperatures of the warmest and coldest quarter are potential influencers of poplar distribution as far as temperature is concerned. From rainfall perspective, precipitation experienced annually, during the wettest month, driest month, driest quarter and warmest quarter are all integral in contributing towards the existence and invasion of poplar species in South

Africa. Overall, the results suggest that regions that receive sufficient rainfall in all phases of the year and experience optimum temperatures during warm and coldest months are likely to host poplar occurrences.

Table 4.4 Evaluation of environmental factors for predicting the distributions of poplars.

Exploratory variable	Bioclim	P value	Significance
Mean_Diurnal_Range	2	0.271961	
Temp_Seasonality	4	0.000045	***
Max_Temp_warmest.Month	5	0.531755	
Min_Temp_Coldest_Month	6	0.28809	
Mean_Temp_Driest_Quarter	9	0.467704	
Mean_Temp_Warmest_Quarter	10	0.0000528	***
Mean_Temp_Coldest_Quarter	11	0.0000974	***
Annual_Precip	12	0.000367	***
Precip_Wettest_Month	13	0.001182	**
Precip_Driest_Month	14	0.00000431	***
Precip_Wettest_Quarter	16	0.076492	.
Precip_Driest_Quarter	17	0.000157	***
Precip_Warmest_Quarter	18	0.022961	*
Precip_Coldest_Quarter	19	0.471873	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The study ranked the contribution of all 13 environmental variables in determining poplar occurrence and distribution, and found the warmest and driest quarter precipitation, and annual precipitation to be the top three contributing factors (refer to Figure 4.5). The findings highlight the significance of wet conditions and water availability for poplar presence in South Africa.

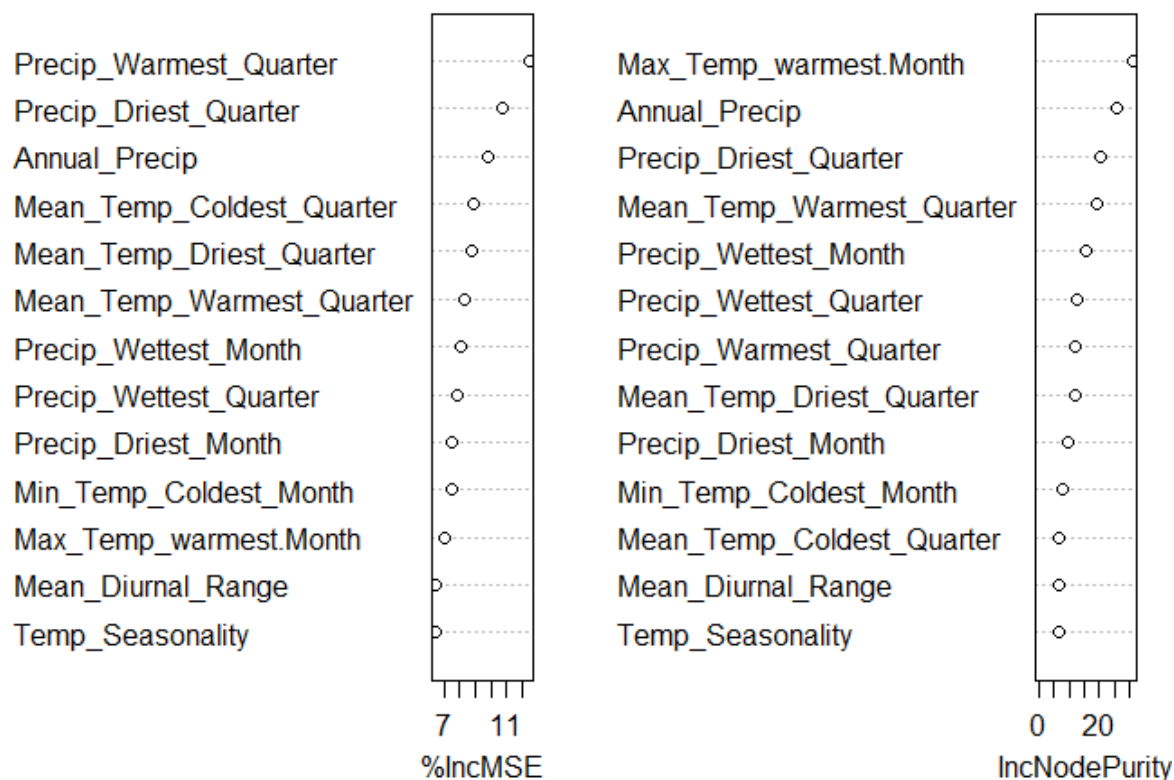


Figure 4.5 Random forests variable importance.

4.4 DISCUSSION

This study provided a successful prediction of poplar species distribution in South African landscapes and allowed the: (a) understanding of the geographic distribution and pattern of poplars in the country, (b) identification of environmental conditions that most affect poplar occurrences and dispersion, and (c) testing of the superior performance of Machine Learning models (RF and SVM), for modelling species distribution over a regression model.

The results showed that poplars are mainly found in regions of warm temperature and high rainfall in South Africa, including the south-western parts of the Western Cape and Northern Cape, the southern Cape on the Garden Route, the central-eastern Free State, western parts of KwaZulu-Natal, Gauteng, eastern parts of Northwest, the north-eastern part of the country in Mpumalanga and Limpopo. These

also include protected areas found in the aforementioned regions such as Golden Gate Highlands National Park (Daemane et al. 2010), Table Mountain National Park, Cape Peninsula Nature Reserve, Tweefontein Reserve, and Platberg Private Nature Reserve, to name but few. The findings also seem to suggest that species such as *P. nigra*, *P. alba*, *P. canescens*, and *P. deltoides*, are mostly found throughout the country, in the wetter and warmer provinces. Their ability to thrive in semi-drier to wetter conditions may cause this adaptation (Caudullo and De Rigo, 2016; De Rigo et al. 2016). The majority develop themselves organically and thrive when enough water and sunlight are available. For windbreaking and decorative purposes, poplar species including *P. simonii*, *P. fremontii*, and *P. canadensis* are planted in homes and streets and are carefully tended to with frequent irrigation. Overall, the results are consistent with Ntshidi et al. (2018) and Mtengwana et al. (2021), because, from a geographic perspective, invasive species predominate in the wetter parts of the country, which are important water sources for the country's major rivers.

Poplar needs enough water and warmer temperatures for better growth and survival (Fischer et al. 2018; Kalcsits et al. 2009). This justifies their occurrence in riparian zones, and warmer and wetter parts of the country (Sperandio et al. 2022; Xi et al. 2021). While its invasiveness has not been extensively studied, like any other invasive species, it competes with native important species for resources, and at most outcompete them (Poudel et al. 2019; Zhang et al. 2021). Therefore, this study results are important as they enable the understanding of the patterns of their spread across the country and selected protected areas to know what interventions to put in place to monitor and manage them. In protected areas under the South African National Parks (SANParks), poplar is categorized as species of special concern, which requires constant monitoring and management (SANParks, 2020). However, knowledge about what constitutes to its spread is not yet available. Therefore, it is hoped that this gap will be filled as an extension of this study.

It is well known that the geospatial distribution of plants is largely influenced by an interplay of multiple mechanisms, including climate and other environmental characteristics (Chen et al. 2014). Although some of these parameters are not part of this modelling approach, they are likely to help determine poplar distribution. To some extent, the variations in topography and land cover are related to climatic variations

(Huang et al. 2021; Krauss et al. 2008), hence it is preferred to focus solely on the bioclimatic variables. Therefore, it is assumed that the observed distribution of the poplars depends solely on the current climate. For this reason, the RF was used to evaluate the most influential climatic variables. In doing so, this study found that precipitation in the coldest quarter and maximum temperatures in the warmest month are the most important variables in the prediction of poplar trees. In fact, warmer and humid conditions are favourable for poplars, while cold and dry conditions appear to be unsuitable. Likewise, Yudaputra et al. (2019) using RF, examined the variables essential to the dispersal of *Eusideroxylon zwageri* in Indonesia and found that precipitation in the coldest months, precipitation seasonality, and isothermality are the most important variables affecting the dispersal of the target species. This information is important for predicting the habitat suitability of landscapes to allow for the existence of species and thus their future spread (Burns et al. 2020; Zhang et al. 2019).

The results of this study are consistent with previous studies in which machine learning methods (RF and SVM) outperformed a regression model (GLM) (Ahmed et al. 2021). The superior performance of RF was also confirmed by Jensen et al. (2020) in their prediction of an invasive Kudzu vine in the USA using Sentinel-2. RF uses only a random subset of the predictor variables for each subset as each tree grows. This creates decorrelated trees and reduces the variance of the final model (Hastie et al. 2009). Additionally, RF avoids overfitting, although they can have quite a complex response, making RF the best model for modelling species distribution. As the second-best model, SVM works by constructing a series of hyperplanes to separate data points based on their class (Ahmed et al. 2021; Kass et al. 2021). Performance is often poor when the number of background samples significantly exceeds the occurrence records. This is largely due to class overlap (García-Roselló et al. 2019).

This issue was combated by applying weights, which significantly increases the cost of misclassifying presence points (Duan et al., 2014). SVM has acknowledged for its performance in SDMs in many studies. For example, Bedia et al. (2011) used SVM and other models to predict the distribution of plant species in an alpine rangeland in northern Spain. Hailu et al. (2017) reported that SVM performed well together with MaxEnt in assessing the spatial distribution of *Coffea arabica* in the Ethiopian highlands. It is said that SVM essentially performs better than RF, especially when the number of observed

occurrences is small, since it can be trained with few meaningful pixels (Pouteau et al. 2012). GLM is the least performing model in this study and is commonly used to determine the distribution pattern of species using the iterative weighted linear regression technique to obtain the estimated maximum likelihood of the parameters (Fukuda et al. 2013; Godsoe, 2014). Its relative performance to RF and SVM is consistent with other similar studies. However, in some applications, the GLM has been found to perform better than the Boosted Regression Tree in predicting the distribution of eight different species in Australia (Shabani et al. 2016). Overall, all three models performed well as they achieved an AUC > 0.9, confirming their value to model species distributions, especially machine learning-based techniques.

4.5 CONCLUSION

This study predicted poplar distribution in South Africa using machine learning (RF and SVM) and regression (GLM) models. The results showed that poplars are mostly distributed in the warmer regions that receive the above average rainfall of the country, from the southwest of the Western Cape and Northern Cape, the central Free State, western parts of KwaZulu-Natal, eastern parts of Northwest, Mpumalanga, Gauteng, to the southern parts of Limpopo. This distribution appears to affect several protected areas from a conservation perspective, including the Golden Gate Highlands National Park, Table Mountain National Park, Cape Peninsula Nature Reserve, Tweefontein Reserve, and Platberg Private Nature Reserve to name but few. The results of this study showed that machine learning methods (RF; 0.965 and SVM; 0.959) outperformed GLM (0.937) in predicting the occurrence and distribution of poplar trees under the current climatic conditions, with RF performing the best. Based on RF, the coldest quarter precipitation and warmest-month maximum temperature were the most influential climatic parameters for poplar distribution in South Africa. Continued evolution of SDMs, more poplar distribution data and availability are expected to further improve the understanding of species distribution and environmental variables that influence predictive performance.

4.6 REFERENCES

Ahmed, N., Atzberger, C. and Zewdie, W., 2021, Species distribution modelling performance and its implication for Sentinel-2-based prediction of invasive *Prosopis juliflora* in lower Awash River basin, Ethiopia. *Ecological Processes*, 10(1), 1-16.

Araújo, M.B., Anderson, R.P., Márcia Barbosa, A., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., Naimi, B. and O'Hara, R.B., 2019. Standards for distribution models in biodiversity assessments. *Science Advances*, 5(1), 4858.

Astuti, I.P. and Cropper, W.P., 2019, Comparing six different species distribution models with several subsets of environmental variables: Predicting the potential current distribution of *Guettarda speciosa* in Indonesia. *Biodiversitas Journal of Biological Diversity*, 20(8).

Bedia, J., Busqué, J. and Gutiérrez, J.M., 2011, Predicting plant species distribution across an alpine rangeland in northern Spain. A comparison of probabilistic methods. *Applied Vegetation Science*, 14(3), 415-432.

Blamey, R.C., Middleton, C., Lennard, C. and Reason, C.J.C., 2017, A climatology of potential severe convective environments across South Africa. *Climate Dynamics*, 49, 2161-2178.

Booth, T.H., Nix, H.A, Busby, J.R. and Hutchinson, M.F, 2014, BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. *Diversity and Distributions*, 20, 1-9.

Breiman, L., 2001, Random forests. *Machine Learning*, 45, 5-32.

Buri, A., Cianfrani, C., Pinto-Figueroa, E., Yashiro, E., Spangenberg, J.E., Adate, T., Verrecchia, E., Guisan, A. and Pradervand, J.N., 2017. Soil factors improve predictions of plant species distribution in a mountain environment. *Progress in Physical Geography*, 41(6), 703-722.

Burns, P.A., Clemann, N. and White, M., 2020, Testing the utility of species distribution modelling using Random Forests for a species in decline. *Austral Ecology*, 45(6), 706-716.

Caudullo, G. and De Rigo, D., 2016. *Populus alba* in Europe: Distribution, habitat, usage and threats. *European Atlas of Forest Tree Species*; San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, A., Eds, 134-135.

- Chen, Y., Yang, X., Yang, Q., Li, D., Long, W., and Luo, W., 2014, Factors affecting the distribution pattern of wild plants with extremely small populations in Hainan Island, China. *PLoS one*, 9, 97751.
- Daemane, M.E., Van Wyk, B.E. and Moteetee, A., 2010. Checklist of ferns and seed plants of the Golden Gate Highlands National Park, South Africa. *Bothalia*, 40(2), 205-218.
- Dayaram, A., Harris, L.R., Grobler, B.A., Van der Merwe, S., Rebelo, A.G., Ward Powrie, L., Vlok, J.H., Desmet, P.G., Qabaqaba, M., Hlahane, K.M. and Skowno, A.L., 2019, Vegetation map of South Africa, Lesotho and Swaziland 2018: a description of changes since 2006. *Bothalia-African Biodiversity & Conservation*, 49(1), 1-11.
- De Rigo, D., Enescu, C.M., Houston Durrant, T. and Caudullo, G., 2016. *Populus nigra* in Europe: distribution, habitat, usage and threats. *European atlas of forest tree species*, 136-137.
- du Toit, A.L., 1940, 'The geology of South Africa'. Oliver and Boyd, 95, 389.
- Duan, R.Y., Kong, X.Q., Huang, M.Y., Fan, W.Y. and Wang, Z.G., 2014. The predictive performance and stability of six species distribution models. *PloS one*, 9(11), 112764.
- Dubuis, A., Giovanettina, S., Pellissier, L., Pottier, J., Vittoz, P. and Guisan, A., 2013. Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables. *Journal of Vegetation Science*, 24(4), 593-606.
- Elith, J., Kearney, M. and Phillips, S., 2010, The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1(4), 330-342.
- Fischer, M., Zenone, T., Trnka, M., Orság, M., Montagnani, L., Ward, E.J., Tripathi, A.M., Hlavinka, P., Seufert, G., Žalud, Z. and King, J.S., 2018, Water requirements of short rotation poplar coppice: Experimental and modelling analyses across Europe. *Agricultural and Forest Meteorology*, 250, 343-360.
- Fournier, A., Barbet-Massin, M., Rome, Q. and Courchamp, F., 2017. Predicting species distribution combining multi-scale drivers. *Global Ecology and Conservation*, 12, 215-226.

- Fukuda, S., De Baets, B., Waegeman, W., Verwaeren, J. and Mouton, A.M., 2013, Habitat prediction and knowledge extraction for spawning European grayling (*Thymallus thymallus* L.) using a broad range of species distribution models. *Environmental Modelling & Software*, 47, 1-6.
- García-Roselló, E., Guisande, C., González-Vilas, L., González-Dacosta, J., Heine, J., Pérez-Costas, E. and Lobo, J.M., 2019, A simple method to estimate the probable distribution of species. *Ecography*, 42(9), 1613-1622.
- Godsoe, W., 2014, Inferring the similarity of species distributions using Species' Distribution Models. *Ecography*, 37(2), 130-136.
- Guisan, A., Thuiller, W. and Zimmermann, N.E., 2017, *Habitat Suitability and Distribution Models with Applications in R*. Cambridge University Press.
- Hailu, B.T., Siljander, M., Maeda, E.E. and Pellikka, P., 2017, Assessing spatial distribution of *Coffea arabica* L. in Ethiopia's highlands using species distribution models and geospatial analysis methods. *Ecological Informatics*, 42, 79-89.
- Hastie, T., Tibshirani, R. and Friedman, J., 2009, *The elements of statistical learning: data mining, inference, and prediction*. Second edition. Springer Series in Statistics. Springer, New York, New York, USA
- Haughton, S.H., 1969, *Geological History of Southern Africa*. Geological Society of South Africa, 535.
- Henderson, L., 2007, Invasive, naturalized and casual alien plants in southern Africa: A summary based on the Southern African Plant Invaders Atlas (SAPIA). *Bothalia*, 37(2), 215-248. <https://doi.org/10.4102/abc.v37i2.322>
- Huang, E., Chen, Y., Fang, M., Zheng, Y. and Yu, S., 2021, Environmental drivers of plant distributions at global and regional scales. *Global Ecology and Biogeography*, 30(3), 697-709.
- Jensen, T., Seerup Hass, F., Seam Akbar, M., Holm Petersen, P. and Jokar Arsanjani, J., 2020. Employing machine learning for detection of invasive species using sentinel-2 and aviris data: The case of Kudzu in the United States. *Sustainability*, 12(9), 3544.

- Kaky, E., Nolan, V., Alatawi, A. and 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, 101150.
- Kalcsits, L.A., Silim, S. and Tanino, K., 2009, Warm temperature accelerates short photoperiod-induced growth cessation and dormancy induction in hybrid poplar (*Populus*× spp.). *Trees*, 23, 971-979.
- Kass, J.M., Meenan, S.I., Tinoco, N., Burneo, S.F. and Anderson, R.P., 2021, Improving area of occupancy estimates for parapatric species using distribution models and support vector machines. *Ecological Applications*, 31(1), 02228.
- King, L.C., 1942, South African scenery. A textbook of geomorphology. South African scenery. A textbook of geomorphology. Edinburgh. Oliver and Boyd, Ltd., Tweeddale Court, 364.
- Krauss, K.W., Lovelock, C.E., McKee, K.L., López-Hoffman, L., Ewe, S.M. and Sousa, W.P., 2008, Environmental drivers in mangrove establishment and early development: a review. *Aquatic botany*, 89(2), 105-127.
- Kumschick, S., Bacher, S., Bertolino, S., Blackburn, T.M., Evans, T., Roy, H.E. and Smith, K., 2020. Appropriate uses of EICAT protocol, data and classifications. *NeoBiota*, 193-212.
- Lee-Yaw, A.J., McCune, L.J., Pironon, S. and Sheth, N.S., 2022, Species distribution models rarely predict the biology of real populations. *Ecography*, 6, 05877.
- Melo-Merino, S.M., Reyes-Bonilla, H. and Lira-Noriega, A., 2020, Ecological niche models and species distribution models in marine environments: A literature review and spatial analysis of evidence. *Ecological Modelling*, 415, 108837.
- Mtengwana, B., Dube, T., Mudereri, B.T. and Shoko, C., 2021, Modeling the geographic spread and proliferation of invasive alien plants (IAPs) into new ecosystems using multi-source data and multiple predictive models in the Heuningnes catchment, South Africa. *GIScience & Remote Sensing*, 58(4), 483-500.

Mucina, L. and Rutherford, M.C., 2006, The vegetation of South Africa, Lesotho and Swaziland, *Strelitzia* 19(December), 1-30, viewed 23 February 2022, from <http://ebooks.cambridge.org/ref/id/CBO9781107415324A009>.

Ntshidi, Z., Gush, M.B., Dzikiti, S. and Le Maitre, D.C., 2018, Characterising the water use and hydraulic properties of riparian tree invasions: A case study of *Populus canescens* in South Africa. *Water SA*, 44, 328-337.

Peterson, A. T., Soberon, J., Pearson, R.G., Anderson, R.P., Martinez-Meyer, E., Nakamura, M., and Araujo, M.B., 2011, *Ecological Niches and Geographic Distributions*. Princeton University Press.

Poudel, D.R., Chen, H.Y., KC, M., Ge, Z., Bown, H.E. and Ruan, H., 2019. Understory vegetation dynamics across a poplar plantation chronosequence in reclaimed coastal saline soil. *Forests*, 10(9), 764.

Pouteau, R., Meyer, J.Y., Taputuarai, R. and Stoll, B., 2012. Support vector machines to map rare and endangered native plants in Pacific islands forests. *Ecological Informatics*, 9, 37-46.

Roffe, S.J., Fitchett, J.M. and Curtis, C.J., 2019, Classifying and mapping rainfall seasonality in South Africa: A review. *South African Geographical Journal*, 101(2), 158-174.

Roy, H., Pauchard, A. and Stoett, P., 2023. Summary for policymakers of the thematic assessment of invasive alien species and their control of the Intergovernmental Platform on Biodiversity and Ecosystem Services.

SANParks, 2020, Golden Gate Highlands National Park Management Plan. SANParks. Pretoria.

Shabani, F., Kumar, L. and Ahmadi, M., 2016, 'A comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area'. *Ecology and Evolution* 6(16), 5973-5986.

Sperandio, G., Pagano, M., Acampora, A., Civitarese, V., Cedrola, C., Mattei, P. and Tomasone, R., 2022, Deficit Irrigation for Efficiency and Water Saving in Poplar Plantations. *Sustainability*, 14(21), 13991.

- Stewart, S.B., Fedrigo, M., Kasel, S., Roxburgh, S.H., Choden, K., Tenzin, K., Allen, K. and Nitschke, C.R., 2022, Predicting plant species distributions using climate-based model ensembles with corresponding measures of congruence and uncertainty. *Diversity and Distributions*, 28(5), 1105-1122.
- Théroux Rancourt, G., Éthier, G. and Pepin, S., 2015. Greater efficiency of water use in poplar clones having a delayed response of mesophyll conductance to drought. *Tree Physiology*, 35(2), 172-184.
- Thuiller, W., Lafourcade, B., Engler, R. and Araújo, M.B., 2009. BIOMOD—a platform for ensemble forecasting of species distributions. *Ecography*, 32, 369-373.
- Van Wilgen, B.W., Measey, J., Richardson, D.M., Wilson, J.R. and Zengeya, T.A., 2020. *Biological Invasions in South Africa*. Springer Nature, 975.
- Vapnik, V., 1998, *Statistical Learning Theory*. Wiley, New York.
- Wolmarans, R., Robertson, M.P. and van Rensburg, B.J., 2010. Predicting invasive alien plant distributions: how geographical bias in occurrence records influences model performance. *Journal of Biogeography*, 37(9), 1797-1810.
- Xi, B., Clothier, B., Coleman, M., Duan, J., Hu, W., Li, D., Di, N., Liu, Y., Fu, J., Li, J. and Jia, L., 2021. Irrigation management in poplar (*Populus spp.*) plantations: A review. *Forest Ecology and Management*, 494, 119330.
- Yudaputra, A., Robiansyah, I. and Rinandio, D.S., 2019, May. The Implementation of Artificial Neural Network and Random Forest in Ecological Research: Species Distribution Modelling with Presence and Absence Dataset. In *Proceedings The SATREPS Conference.*, 2 (1), 38-49.
- Zhang, J., Qin, G., Zhai, Z., Zhou, S., Tang, L. and Tian, Y., 2021. Diverse understory vegetation alleviates nitrogen competition with crop trees in poplar plantations. *Forests*, 12(6), 705.
- Zhang, K., Zhang, Y., Jia, D. and Tao, J., 2020. Species distribution modeling of *Sassafras tzumu* and implications for forest management. *Sustainability*, 12(10), 4132.

Zhang, L., Huettmann, F., Liu, S., Sun, P., Yu, Z., Zhang, X. and Mi, C., 2019. Classification and regression with random forests as a standard method for presence-only data SDMs: a future conservation example using China tree species. *Ecological Informatics*, 52, 46-56.

Zurell, D., 2019. Introduction to species distribution modelling (SDM) in R. Available from <https://damarizsurell.github.io/SDM-Intro/>. Accessed 10 Mar 2023.

Appendix: A

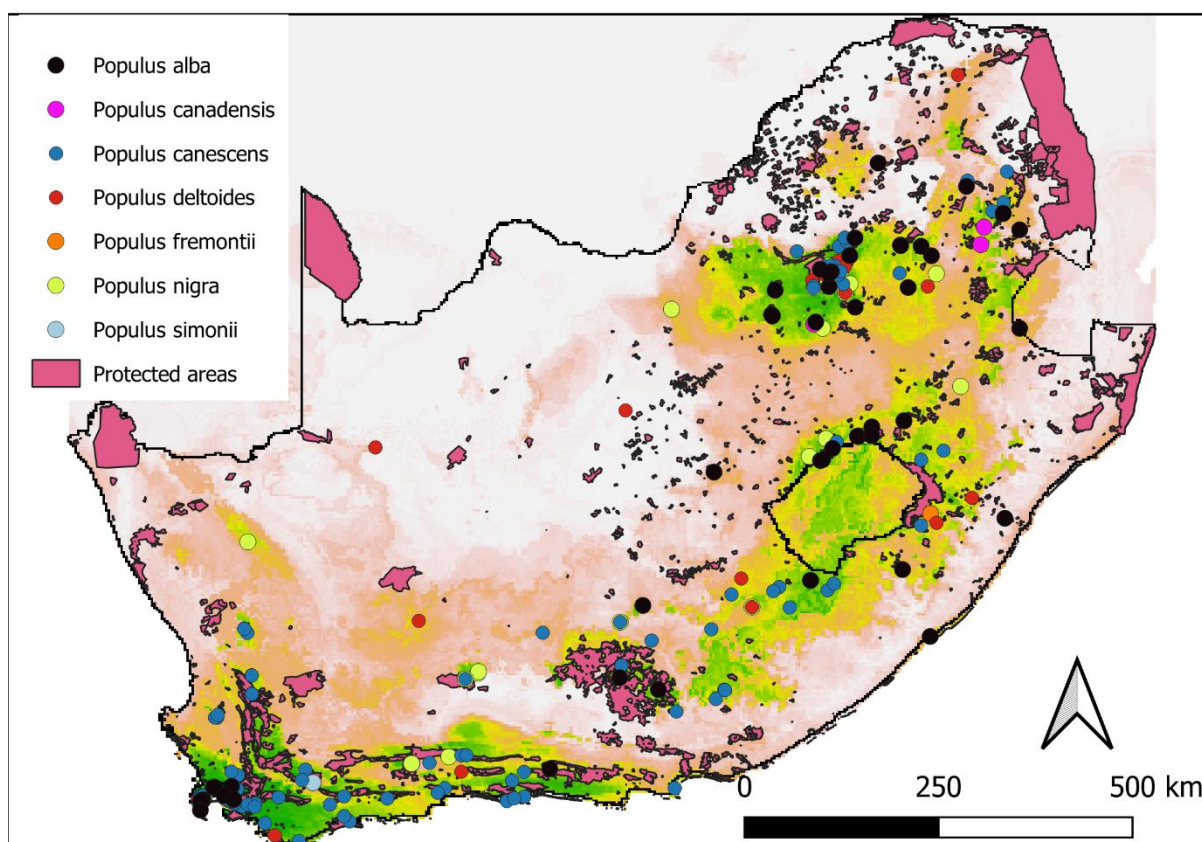


Figure 6: Protected areas overlaid on the poplar habitat suitability map to represent protected areas best suited for poplar growth.

Appendix: B

Table: A list of protected areas where poplars occur.

Species name	Latitude	Longitude	Protected area	Type of protected area
<i>Populus alba</i>	-30.711154	27.689445	Balloch Protected Environment	Protected Environment
<i>Populus alba</i>	-25.625	29.375	Buks Private Nature Reserve	Nature Reserve
<i>Populus alba</i>	-34.053869	18.430926	Cape Floral Region Protected Areas	World Heritage Site
<i>Populus alba</i>	-34.079385	18.406036	Cape Floral Region Protected Areas	World Heritage Site
<i>Populus alba</i>	-34.208938	18.401856	Cape Floral Region Protected Areas	World Heritage Site
<i>Populus alba</i>	-34.053869	18.430926	Cape Peninsula Nature Area	Nature Reserve
<i>Populus alba</i>	-28.52354	28.413926	Clarens Private Nature Reserve	Nature Reserve
<i>Populus alba</i>	-28.516535	28.404204	Clarens Private Nature Reserve	Nature Reserve
<i>Populus alba</i>	-25.97609	27.83956	Fossil Hominid Sites of SA	World Heritage Site
<i>Populus alba</i>	-28.505556	28.618333	Golden Gate Highlands National Park	National Park
<i>Populus alba</i>	-28.50488	28.61942	Golden Gate Highlands National Park	National Park
<i>Populus alba</i>	-34.048175	18.91107	Lourens River Protected Natural Environment	Protected Environment
<i>Populus alba</i>	-32.375	25.375	Mountain Zebra-Camdeboo Protected Environment	Protected Environment

<i>Populus alba</i>	-32.193655	24.78448	Mountain Zebra-Camdeboo Protected Environment	Protected Environment
<i>Populus alba</i>	-28.283675	29.113439	Platberg Private Nature Reserve	Nature Reserve
<i>Populus alba</i>	-28.283675	29.113439	Platberg Private Nature Reserve	Nature Reserve
<i>Populus alba</i>	-34.079385	18.406036	Silvermine Nature Reserve	Nature Reserve
<i>Populus alba</i>	-34.053869	18.430926	Table Mountain National Park	National Park
<i>Populus alba</i>	-34.079385	18.406036	Table Mountain National Park	National Park
<i>Populus alba</i>	-34.208938	18.401856	Table Mountain National Park	National Park
<i>Populus alba</i>	-25.125	30.625	Twefontein Reserve	Forest Nature Reserve
<i>Populus alba</i>	-25.125	30.625	Twefontein Reserve	Forest Nature Reserve
<i>Populus alba</i>	-33.860669	18.599146	Tygerberg Nature Reserve	Nature Reserve
<i>Populus canescens</i>	-33.966542	18.92672	Assegaaibos Provincial Nature Reserve	Nature Reserve
<i>Populus canescens</i>	-33.997081	18.4229	Cape Floral Region Protected Areas	World Heritage Site
<i>Populus canescens</i>	-34.026757	18.413942	Cape Floral Region Protected Areas	World Heritage Site
<i>Populus canescens</i>	-34.006993	18.387344	Cape Floral Region Protected Areas	World Heritage Site
<i>Populus canescens</i>	-34.006336	18.411158	Cape Floral Region Protected Areas	World Heritage Site

<i>Populus canescens</i>	-34.025989	18.413635	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.027215	18.414266	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.0537	18.4309	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.107	18.4095	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.031178	18.39395	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.053984	18.431027	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.053728	18.430565	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.062271	18.411171	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.0552	18.4306	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-33.9837	18.43437	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.031726	18.393571	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.054034	18.430863	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.000756	18.412271	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.006164	18.414939	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.062021	18.409423	Cape Areas	Floral	Region	Protected	World Site	Heritage

<i>Populus canescens</i>	-33.987648	18.428804	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-33.875	22.125	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.125	19.125	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.05395	18.430772	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.201709	18.390701	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-33.966542	18.92672	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-33.969362	18.934983	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-33.997283	18.415584	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-33.996747	18.41566	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-33.997462	18.415873	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.054055	18.43079	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.062137	18.411015	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-34.065174	18.412399	Cape Areas	Floral	Region	Protected	World Site	Heritage
<i>Populus canescens</i>	-33.997081	18.4229	Cape Peninsula Nature Area				Nature Reserve	
<i>Populus canescens</i>	-34.006336	18.411158	Cape Peninsula Nature Area				Nature Reserve	

<i>Populus canescens</i>	-34.0537	18.4309	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-34.031178	18.39395	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-34.053984	18.431027	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-34.053728	18.430565	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-34.062271	18.411171	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-34.0552	18.4306	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-33.9837	18.43437	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-34.031726	18.393571	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-34.054034	18.430863	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-34.000756	18.412271	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-34.006164	18.414939	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-34.062021	18.409423	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-33.987648	18.428804	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-34.05395	18.430772	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-33.997283	18.415584	Cape Peninsula Nature Area	Nature Reserve

<i>Populus canescens</i>	-33.996747	18.41566	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-33.997462	18.415873	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-34.054055	18.43079	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-34.062137	18.411015	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-34.065174	18.412399	Cape Peninsula Nature Area	Nature Reserve
<i>Populus canescens</i>	-28.516643	28.40519	Clarens Private Nature Reserve	Nature Reserve
<i>Populus canescens</i>	-25.625	28.125	De Onderstepoort Nature Reserve	Nature Reserve
<i>Populus canescens</i>	-34.108132	19.240476	Diepklowe Private Nature Reserve	Nature Reserve
<i>Populus canescens</i>	-34.014633	19.592889	Greyton Natuurpark	Nature Reserve
<i>Populus canescens</i>	-33.375	22.375	Groot Swartberg Mountain Catchment Area	Mountain Catchment Area
<i>Populus canescens</i>	-34.071052	23.068163	Knysna National Lake Area	Protected Environment
<i>Populus canescens</i>	-34.08293	18.869153	Lourens River Protected Natural Environment	Protected Environment
<i>Populus canescens</i>	-34.082997	18.870622	Lourens River Protected Natural Environment	Protected Environment
<i>Populus canescens</i>	-34.082436	18.87118	Lourens River Protected Natural Environment	Protected Environment
<i>Populus canescens</i>	-34.066775	18.896596	Lourens River Protected Natural Environment	Protected Environment

<i>Populus</i>	-32.006647	24.814122	Mountain	Zebra-Camdeboo	Protected
<i>canescens</i>			Protected Environment		Environment
<i>Populus</i>	-24.963083	30.631032	Ohrigstad Dam Nature Reserve		Nature Reserve
<i>canescens</i>					
<i>Populus</i>	-24.957372	30.627905	Ohrigstad Dam Nature Reserve		Nature Reserve
<i>canescens</i>					
<i>Populus</i>	-31.445146	19.070025	Oorlogskloof Nature Reserve		Nature Reserve
<i>canescens</i>					
<i>Populus</i>	-31.511203	19.116789	Oorlogskloof Nature Reserve		Nature Reserve
<i>canescens</i>					
<i>Populus</i>	-31.448824	19.068098	Oorlogskloof Nature Reserve		Nature Reserve
<i>canescens</i>					
<i>Populus</i>	-33.92787	18.850667	Papegaaiberg Nature Reserve		Nature Reserve
<i>canescens</i>					
<i>Populus</i>	-33.94088	18.844733	Papegaaiberg Nature Reserve		Nature Reserve
<i>canescens</i>					
<i>Populus</i>	-33.927813	18.850595	Papegaaiberg Nature Reserve		Nature Reserve
<i>canescens</i>					
<i>Populus</i>	-33.931747	18.851838	Papegaaiberg Nature Reserve		Nature Reserve
<i>canescens</i>					
<i>Populus</i>	-34.014633	19.592889	Riviersonderend	Mountain	Mountain
<i>canescens</i>			Catchment Area		Catchment Area
<i>Populus</i>	-33.627432	23.330077	Sustersdal Private Nature Reserve		Nature Reserve
<i>canescens</i>					
<i>Populus</i>	-33.997081	18.4229	Table Mountain National Park		National Park
<i>canescens</i>					
<i>Populus</i>	-34.026757	18.413942	Table Mountain National Park		National Park
<i>canescens</i>					
<i>Populus</i>	-34.006993	18.387344	Table Mountain National Park		National Park
<i>canescens</i>					

<i>Populus canescens</i>	-34.006336	18.411158	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.025989	18.413635	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.027215	18.414266	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.0537	18.4309	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.107	18.4095	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.031178	18.39395	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.053984	18.431027	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.053728	18.430565	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.062271	18.411171	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.0552	18.4306	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.031726	18.393571	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.054034	18.430863	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.000756	18.412271	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.006164	18.414939	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.062021	18.409423	Table Mountain National Park	National Park

<i>Populus canescens</i>	-34.05395	18.430772	Table Mountain National Park	National Park
<i>Populus canescens</i>	-33.997283	18.415584	Table Mountain National Park	National Park
<i>Populus canescens</i>	-33.996747	18.41566	Table Mountain National Park	National Park
<i>Populus canescens</i>	-33.997462	18.415873	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.054055	18.43079	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.062137	18.411015	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.065174	18.412399	Table Mountain National Park	National Park
<i>Populus canescens</i>	-34.006993	18.387344	Table Mountain Nature Reserve	Nature Reserve
<i>Populus deltoides</i>	-34.010574	18.404726	Cape Floral Region Protected Areas	World Heritage Site
<i>Populus deltoides</i>	-25.625	28.125	De Onderstepoort Nature Reserve	Nature Reserve
<i>Populus deltoides</i>	-34.010574	18.404726	Table Mountain National Park	National Park
<i>Populus deltoides</i>	-34.010574	18.404726	Table Mountain Nature Reserve	Nature Reserve
<i>Populus nigra</i>	-30.715312	27.697108	Balloch Protected Environment	Protected Environment
<i>Populus nigra</i>	-26.5839	25.574	Barberspan Nature Reserve	Nature Reserve
<i>Populus nigra</i>	-25.922158	27.974828	Diepsloot Nature Reserve	Nature Reserve

<i>Populus</i>	-28.505786	28.618008	Golden Gate Highlands National	National Park
<i>nigra</i>			Park	
<i>Populus</i>	-28.504375	28.612449	Golden Gate Highlands National	National Park
<i>nigra</i>			Park	
<i>Populus</i>	-28.505313	28.620811	Golden Gate Highlands National	National Park
<i>nigra</i>			Park	
<i>Populus</i>	-28.508797	28.622981	Golden Gate Highlands National	National Park
<i>nigra</i>			Park	
<i>Populus</i>	-28.505086	28.616564	Golden Gate Highlands National	National Park
<i>nigra</i>			Park	
<i>Populus</i>	-28.503198	28.62032	Golden Gate Highlands National	National Park
<i>nigra</i>			Park	
<i>Populus</i>	-28.504928	28.616906	Golden Gate Highlands National	National Park
<i>nigra</i>			Park	
<i>Populus</i>	-28.505086	28.616564	Golden Gate Highlands National	National Park
<i>nigra</i>			Park	

Appendix: C

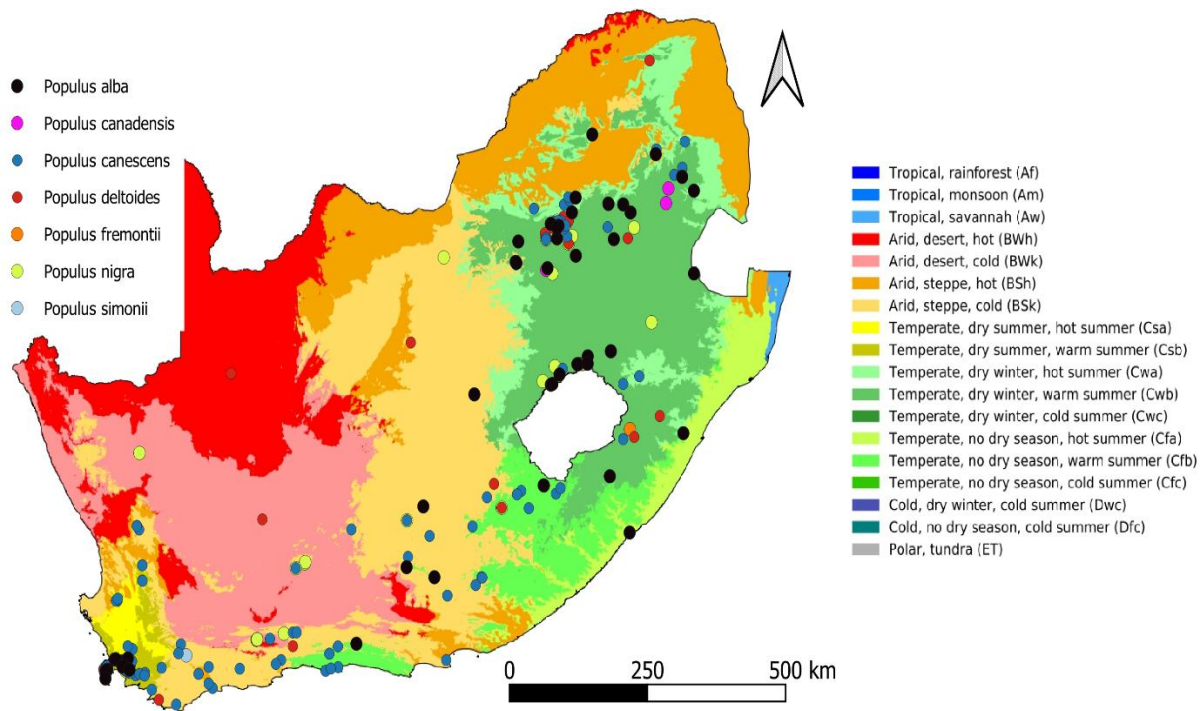


Figure 7: South African Climatic Zones with poplar species overlaid.

CHAPTER 5 EXPLORING ENVIRONMENTAL FACTORS THAT INFLUENCE THE DISTRIBUTION OF POPLAR TREES



Received: 10 February 2024 | Revised: 21 June 2024 | Accepted: 28 June 2024

DOI: 10.1111/aje.13310

SHORT COMMUNICATION

African Journal of Ecology WILEY

Exploring environmental factors that influence the distribution of poplar trees

Morena J. Mapuru^{1,2} | Sifiso Xulu¹ | Michael Gebreslasie¹ | Ernest M. Daemane²

¹School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Durban, South Africa

²Scientific Services, Conservation Services Division, South African National Parks, Kimberly, South Africa

Correspondence

Morena J. Mapuru, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4000, South Africa.
Email: johnmapuru97@gmail.com

Abstract

Poplar is an invasive alien plant (IAP) that encroaches on fragile landscapes in South Africa, but the environmental factors that contribute to its spread are not well understood. To address this gap, we conducted a study using generalised linear models (GLM) to identify the most significant local environmental factors that drive poplar establishment and dispersion in Golden Gate Highlands National Park (GGHNP). We then ranked these variables according to their contribution to poplar occurrence using a random forests regression approach. Our GLM analysis revealed that soil properties such as soil organic carbon, sand and silt, topographical variables such as aspect and

Link: <https://doi.org/10.1111/aje.13310>

This Chapter is based on:

Mapuru, M. J., Xulu, S., Gebreslasie, M., & Daemane, E. M. (2024). Exploring environmental factors that influence the distribution of poplar trees. *African Journal of Ecology*, 62, e13310.

ABSTRACT

Poplar is an invasive alien plant (IAP) that encroaches on fragile landscapes in South Africa, but the environmental factors that contribute to its spread are not well understood. To address this gap, the study was conducted using generalized linear models (GLM) to identify the most significant local environmental factors that drive poplar establishment and dispersion in Golden Gate Highlands National Park (GGHNP). The study then ranked these variables according to their contribution to poplar occurrence using a random forests regression approach. The GLM analysis revealed that soil properties such as soil organic carbon, sand and silt, topographical variables such as aspect and slope, normalized difference moisture index (NDMI) and poplar distance from rivers are significant factors affecting poplar survival. After ranking the variables, the study found that the three most important factors influencing poplar occurrence were aspect, slope and poplar distance to rivers, followed by the sand, silt, NDMI, and soil organic carbon. The current study results are consistent with previous studies, indicating that topographical variables and the availability of sufficient water supply are the primary drivers of the spread of IAPs. This study provides useful insights in managing and monitoring poplars in delicate landscapes.

Keywords: poplar; random forests; generalized linear model; species distribution model; *Populus*

5.1 INTRODUCTION

It has been observed that poplar (*Populus canescens* (grey) and *Populus alba* (silver)), an alien invasive species, has been spreading rapidly in the landscapes of GGHNP (SANParks, 2020). Unchecked growth of invasive plant species, such as poplar, can cause significant damage to ecosystems and result in higher restoration costs (Vilà et al. 2011). The presence of invasive tree species can have a substantial impact on forest management by affecting nutrient cycling and moisture regimes, among other ecosystem processes (Marchante et al. 2008). The extent and duration of the invasion, as well as the interactions between biotic and abiotic factors (McGill, 2010), can all influence the effects of IAPs, particularly trees (Martínez et al. 2010; Herrero-Jáuregui et al. 2012).

It is essential to comprehend the causes and patterns of invasion to put into practice effective management measures (Brundu and Richardson, 2016). To support various forms of intervention, from early detection to management of established invaders (Van Wilgen et al. 2011), it is becoming increasingly crucial to predict invasions at smaller spatial scales (Fernandes et al. 2014). Gaining an understanding of the local environmental characteristics of an area suitable for the species can provide better information and more precise forecasts, ultimately aiding in the management of invasions. Factors such as physical and chemical soil characteristics (Huang et al. 2021), moisture availability (Darby et al. 2016; Moeslund et al. 2013), topographical variables (Elias et al. 2019), and the density of surrounding plant species (Fernandes et al. 2014) are the biggest drivers of invasion at a local scale. Topography directly and indirectly affects plant distribution by inducing variation in environmental conditions, such as microclimate and soil stability (Barry, 2008). In addition, topography represents a geographical barrier for plant dispersal across landscapes (Lomolino, 2001). Soil provides plants with essential nutrients, oxygen, and water required for growth and development (Onwuka and Mang, 2018). Alien invasive plants such as poplar tend to thrive in regions with ample water supply and often establish themselves along streams and artificial canals (Ntshidi et al. 2018). These environmental variables are usually used as predictor variables in Species Distribution Models (SDMs) to model their influence in determining the distribution and occurrence of delicate alien plant species (Elias et al. 2019; Huang et al. 2021; Moeslund et al. 2013). SDM techniques, such as Random Forests (RF) machine learning and Generalized Linear Model (GLM) regression model, have achieved commendable results in modelling the distribution of

plant species (Hill et al. 2017; Mapuru et al. 2024; Williams et al. 2009). The GLM significantly extends linear regression by incorporating a link function to relate the linear model to the response variable and decisively identifies the variables influencing the occurrence of species of interest (Mohammady et al. 2021), while the RF regression technique is mostly used in ranking the contribution predictor variables towards the establishment of plant species (Evans et al. 2010). Although the use of the SDM techniques mentioned above in conjunction with the local environmental variables can inform what drives poplar distribution in GGHNP, this option has not been sufficiently explored. Therefore, the current study aims to fill this research gap.

The environmental factors that affect the occurrence and distribution of poplar species in Golden Gate Highlands National Park have not been extensively researched. There have been few studies on the invasive potential of some plant species in GGHNP, and still fewer attempts have been made to understand what could be driving their existence and distribution in delicate landscapes (Canavan et al. 2022; Masole et al. 2022). There are two types of poplar found in GGHNP: grey and silver poplar. The park management has identified this plant as a species of special concern, and it has been removed while being closely monitored throughout the park since 2019 (SANParks, 2020). The clearing of poplar and other IAPs is one of the strategic objectives for landscape conservation and management in the GGHNP (SANParks, 2020). These poplar species consume large amounts of water from streams, leading to potential ecological imbalances (Ntshidi et al. 2019). Additionally, they are known to outcompete native grass species for resources and are often found in sensitive areas like riparian zones (SANParks, 2020). It is therefore crucial for GGHNP management to have a comprehensive understanding of their habitat requirements and local environmental drivers to effectively manage their distribution pattern in the park.

The main aim of this study was to investigate the local environmental factors that affect the occurrence and spread of poplar trees in the GGHNP. To achieve this goal, two objectives were set to: (1) identify local environmental factors that influence the distribution of poplar trees and (2) rank the contribution of the identified variables in determining the presence and spread of poplar trees across the landscape. Here, we hypothesize that topographical variables and water availability are the primary determinants of poplar distribution.

5.2 MATERIALS AND METHODS

5.2.1 Study area

The current study was conducted in the Golden Gate Highlands National Park ([28°30'22"S, 28°37'0"E](#)), situated in the eastern Free State Province of South Africa (Figure 5.1). The park spans an area of approximately 340 km² and is located at the foot of the Maloti Mountains. The highest peak in the park is approximately 2,829 m a.s.l (SANParks, 2020). The park experiences dry climate from June to August and variable rainfall between October and April. The annual rainfall in the region is around 800 mm (Kay et al., 1993). The park's geology is divided into five formations, consisting of four sedimentary and one igneous (SANParks, 2020). The park falls within the grassland biome of South Africa, representing the Drakensberg grassland bioregion and the mesic highland grassland bioregion (SANParks, 2020). The two largest plant families in the park are Poaceae and Asteraceae. The park's grassland is slightly undulating with streams and rivers, dominated by *Eragrostis sp.*, *Tristachya leucothrix*, and *Themeda triandra*, and an herb layer dominated by Asteraceae (SANParks, 2020). Unfortunately, there are various invasive alien floral species present in the park, including poplars, which are of particular concern. This is due to the various ecological damages they inflict on the environment and other species (SANParks, 2020). The park is located on the watershed between the Vaal and the Orange Rivers systems. It contributes quality water to the Gauteng region through the Lesotho Highlands water project and the Tugela-Vaal transfer scheme. The Little Caledon River and Klerkspruit River, which rise in the park, contribute approximately 30% of the total water supply in Southern Africa.

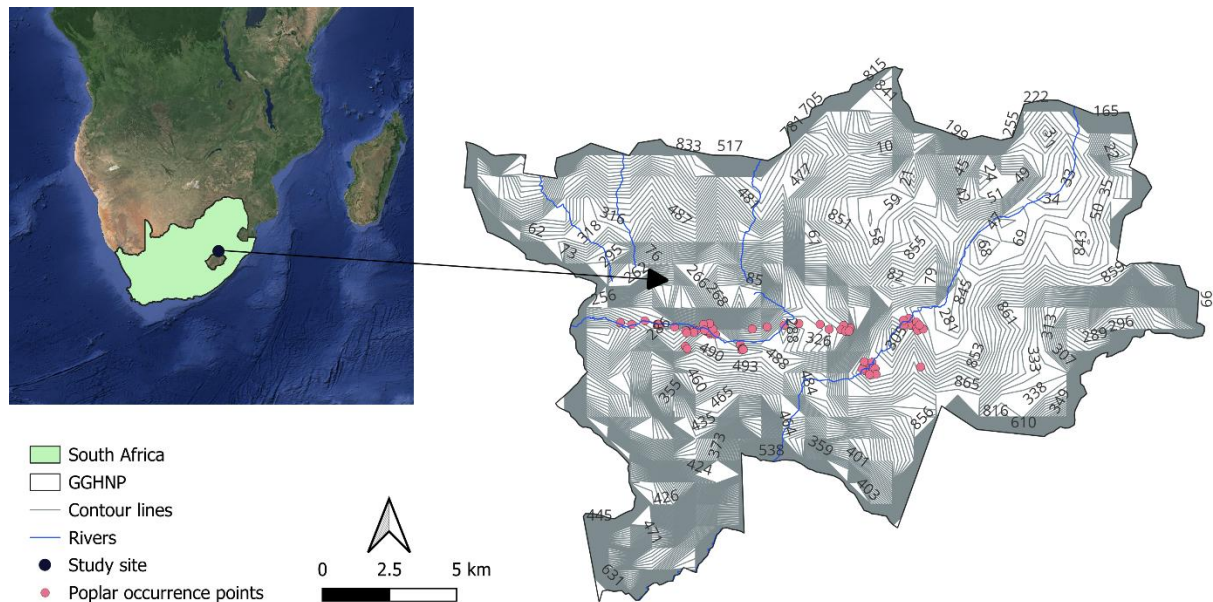


Figure 5.1 A study map depicting South Africa (a), Free State (b), and Golden Gate Highlands National Parks (c).

5.2.2 Field data collection

A total of 100 poplar occurrence points were mapped using GPS throughout the GGHNP; this included sites that had been cleared but with evidence of historical poplar occurrence. Points were randomly collected from poplar trees. In the remote areas that were not easy to access, such as riverbanks, high spatial resolution image integrated in Google Earth Pro was used to supplement the poplar presence points. One hundred absence points were created in R environment using a random point function to characterize areas where poplar was not occurring in GGHNP. The number of absence points was deliberately set to match the number of presence points to prevent any modelling biases (Barbet-Massin et al. 2012; Lobo et al. 2010; VanDerWal et al. 2009). Both presence and absence points were combined and used as response variable in models, and were given the value of 1 and 0, respectively.

5.2.3 Environmental variables

The physical and chemical properties of the soil were acquired from the International Soil Reference and Information Center (<https://soilgrids.org>) at 1 km spatial resolution raster format (Hengl et al. 2014). This soil data was validated using ground data from across the globe (Hengl et al. 2017; Shangguan et al. 2017). Data included soil organic carbon (dg/kg), bulk density (cg/cm³), silt fragments (g/kg), soil pH, topsoil nitrogen (cg/kg), and sand content (g/kg) (Mashiane et al. 2023). We used the Digital Elevation Model (10 m) from the WorldClim database (<https://www.worldclim.org/data/worldclim21.html>) to calculate topographic features such as elevation (m.a.s.l), slope (angle°), and aspect (°) in QGIS. The 2021 to 2022 Sentinel-2 imagery (10 m) captured by the Copernicus Data Space Ecosystem was used to calculate and derive the Normalized Difference Moisture Index (NDMI) (Ohman et al. 2022), Normalized Difference Vegetation Index (NDVI) (Labonté et al. 2020) and Normalized Difference Water Index (NDWI) (Gerardo and de Lima, 2022) using the following formulas:

$$NDMI = \frac{B8 - B11}{B8 + B11} \quad (1)$$

$$NDVI = \frac{B8 - B4}{B8 + B4} \quad (2)$$

$$NDWI = \frac{B8 - B12}{B8 + B12} \quad (3)$$

where $B4$ = Red band; $B8$ = Near Infrared band, and $B11$ and 12 = Shortwave Infrared bands

The "distance to the nearest hub" function in QGIS was utilized to determine the distance between river courses and poplar occurrence points. The study resampled soil physical and chemical variables to 10 m spatial resolution to account for raster variable resolution discrepancies and guarantee that the results are consistent with topographical (slope, aspect, and elevation) and remote sensing (NDMI, NDWI, and NDVI) data. Appendix A provides descriptive statistics of all the environmental predictors.

Using QGIS's "sample raster values" function, the study additionally extracted raster predictor variable values linked to poplar points. This function¹¹² retrieves raster values at specific point locations

and populates the values in corresponding columns of an attribute table (Treglia, 2011). The results from each of these datasets were put together into an Excel spreadsheet, which was then used later to compute the statistical analysis in the R programming language.

5.2.4 Statistical analysis

To ensure the accuracy of the analysis, the study dataset underwent a normality test using Q-Q plot and Shapiro-Wilk test approach but found that it was not normally distributed. To address this issue, the current study used Generalised Linear Models (GLM) designed to work with non-normally distributed data (Buckley, 2014; Hammouri et al. 2020). GLM was used with statistical significance set at $p < 0.05$. We used the Akaike Corrected Information Criterion (AICc) and the AICc weight, through the "dredge" function in the "MuMIn" package in R (Barton, 2009; Johnson and Omland, 2004), to select the optimum model combination (Mapuru et al. 2021). Dredge function automates model selection by sub-setting the best model, allowing optional constraints for model inclusion (Barton, 2009). This helped to avoid model overestimations and multicollinearity. The study used only the environmental predictors that were part of the best model combination estimate to perform GLM analysis. Before running any modelling analysis, all predictors were standardized using the scale function in R (VanDerWal et al. 2009) and also the study used Bonferroni correction procedure to reduce the probability of incurring the type 1 error (VanDerWal et al. 2009). The "Dsquared" function from the "modEvA" package in R was used to calculate the model's performance (Barbosa et al. 2020).

The study employed a Random Forests (RF), a method linked to Classification and Regression Trees (CART) and a bagging procedure (Breiman, 2001), to rank the significance of environmental predictors. RF constructs multiple trees based on random samples of cases by bootstrapping technique, instead of building a single tree. Each split is created from a random sample of predictors (Breiman, 2001). This technique ensures that 1/3 of the observations are excluded during the fitting process and used to calculate the prediction error. RF was conducted in a regression setting with poplar presence (100) and absence (100) treated as response variables in a model (Liaw and Wiener, 2002). Only significant environmental variables returned by "dredge" function were used as predictors. The distribution data was split into 25%

for testing and 75% for training (Mapuru et al. 2024). To implement RF, the "randomForest" function in the R statistical environment was used. The predictors' importance was rated based on specified changes of mean decrease in accuracy (%IncMSE), i.e. predictors were removed one by one from the model, thus revealing their adequacy (Berk, 2008). The RF regression required the definition of parameters such as ntree (number of trees), node-size, and importance (variable importance). Model values were set as follows: ntree = 500, node-size = 5 (default value), and importance = T (TRUE) (Vincenzi et al. 2011).

5.3 RESULTS

5.3.1 Significant environmental predictors

The analysis using the GLM revealed that soil physical and chemical properties such as silt fragments, sand, and soil organic carbon, along with topographical variables such as slope and aspect, had a significant impact on poplar occurrence (Table 5.1). Surface moisture content (NDMI) and distance to rivers also play a crucial role in determining poplar occurrence and spread (Table 5.1). Although factors like pH, nitrogen, and bulk density can impact poplar occurrence, their effects were not significant enough to support poplar establishment. These results suggests that the primary drivers of poplar occurrence and distribution in the GGHNP are the topographical characteristics of the region and water availability.

Table 5.1 Results from the GLM on the effects of environmental predictors on poplar distribution.

Response	Predictors	t value	p value	
Poplar presence and absence	pH	0.919	0.36209	
	Nitrogen	0.934	0.35419	
	Silt	-3.509	0.03677	*
	Slope	16.363	< 2e-16	***
	Aspect	14.066	< 2e-16	***
	Sand	-3.866	0.0002	***
	Bulk density	-1.057	0.967	
	Soil organic carbon	-4.682	1.67e-05	***
	NDMI	6.753	6.59e-09	***
	Distance to the rivers	-6.133	7.36e-08	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

5.3.2 Relative variable importance

Table 5.2 presents the relative significance of environmental predictors that were considered. This analysis revealed that variables such as slope, aspect, and distance to the rivers were the three most important predictors. This was followed by soil organic carbon, sand, silt fragments, and NDMI in decreasing order of importance.

Table 5.2 Variable ranking according to RF

Predictors	%INcMSE	IncNodePurity
Slope	19.419	0.020
Aspect	17.820	0.02
Distance to the rivers	14.630	0.01
Soil organic carbon	13.086	0.01
Sand	11.934	0.003
Silt	9.142	0.002
NDMI	6.12	0.002

5.4 DISCUSSION

This study aimed to investigate the environmental factors that affect the presence and distribution of poplar trees in the GGHNP using Generalised Linear Models and Random Forests. The results showed that soil properties such as silt fragments, sand and organic carbon content, and topographical variables such as slope and aspect as well as soil moisture content, and the distance of poplar trees from streams were the main environmental factors that influenced the occurrence and distribution of poplars in GGHNP. Poplars thrive in riparian areas where soil moisture is high (Laslier et al. 2019; Bahn and An, 2020). This was confirmed as *distance to rivers* and *soil moisture content* were among the main influencing factors of poplar occurrence. Riparian soils are characterized by a mixture of sand, silt, and clay, and high organic carbon content that results from the deposition and decomposition of vegetation fragments (Franklin et al. 2020; Gomes et al. 2019). Such soils provide high aeration, easy water percolation, and good water retention, providing plant roots with easy access to water.

The current study identified slope and aspect as the most important parameters of poplar distribution. The study findings are consistent with those of Lan et al. (2011), who also found that topographic variables are crucial predictors of vegetation distribution. Topography affects climatic conditions (Grzyl et al. 2014), soil formation (Ridolfi et al. 2008), infiltration, runoff, and seed migration (Jiao et al., 2009; Li et al., 2021). Altitude determines the elevational zonation of soil (Fu et al. 2018; Tsui et al. 2013),

slope controls the speed of runoff (Liu et al. 2015), and aspect influences flow direction, amount of solar radiation received, and evaporation intensity (Moore et al. 1993). The combination of these factors determines the conditions for growth and, consequently, the distribution of vegetation.

Numerous studies have investigated the impact of IAPs on the abundance of understory species (Avalos et al. 2006; Dyderski and Jagodziński, 2020; Fu et al. 2018; Jo et al. 2018). However, only a few studies have examined whether the density of understory species affects their occurrence and distribution. The current study tested whether the density of the understory species, as determined from the NDVI, influenced the presence of poplar trees. These results suggest that the density of understory species has less influence on the presence of poplar species.

Plant species typically require specific soil pH, nitrogen content, and soil density to thrive in particular locations (Weidenhamer et al. 2010; Reinhart and Callaway, 2006). However, this current research on poplar trees in GGHNP showed that none of these variables have played a clear role in explaining the distribution or occurrence of the species in the area. One potential reason for this outcome could be that the soil data utilized in this study was not directly measured at the site, but instead taken from a global database at a 1 km grid. While a previous study conducted at GGHNP by Mashiane et al. (2023) yielded favourable results using similar datasets, the study suggests that future research studies consider incorporating a combination of direct measurements and data from global soil databases. This approach will help enhance the precision of the results and illustrate the accuracy of data within global soil databases.

5.5 STUDY LIMITATIONS

The current study has limitations that should be recognized. First, the study did not consider the biological interaction between poplar and other co-occurring species (both fauna and flora), which could affect the adaptability and occurrence of poplar in certain sites. In the future, it would be essential to investigate the interaction between poplar and other plant species as one of the limiting factors for dispersal. Second, the study did not explore the impact of fire incidents on the distribution and occurrence of poplar species as it has the potential to return nutrients to the soil,¹¹⁷ which could stimulate the germination of seeds

beneath the ground. The study recommends future research to examine the effect of fire on poplar distribution. In South Africa, there are about six species of poplar, but the current study only focused on two located in GGHNP. As all these poplar species pose problems, we encourage similar investigations on other species occurring elsewhere in the country.

5.6 CONCLUSION

This research was conducted to investigate the local environmental factors that influence the existence and distribution of poplar trees in Golden Gate Highlands National Park. The study aimed to explore two objectives: (1) To determine which fine-scale environmental variables significantly affect the distribution of poplar trees in GGHNP, and (2) To rank the contribution of the identified fine-scale environmental variables in determining the presence and spread of poplar trees across the park's landscape. The study revealed that soil (soil organic carbon, sand, and silt fragments), topographical variables (slope and aspect), surface moisture content (NDMI), and poplar distance from rivers are significant variables informing poplar existence in the park. The study analysed the contributions of the identified significant predictors in explaining the existence of poplar and discovered that topographical variables and water resource availability are the most important factors for poplar occurrence and distribution. With the discovery of these important study findings, it is anticipated that this information will be useful to park management, especially in making decisions around the management of poplar-prone places.

5.7 REFERENCES

- Ahmed, N., Atzberger, C. and Zewdie, W., 2021. Species distribution modelling performance and its implication for Sentinel-2-based prediction of invasive *Prosopis juliflora* in lower Awash River basin, Ethiopia. *Ecological Processes*, 10(1). <https://doi.org/10.1186/s13717-021-00285-6>
- Araújo, M. B., and Peterson, A. T., 2012. Uses and misuses of bioclimatic envelope modeling. *Ecology*, 93(7). <https://doi.org/10.1890/11-1930.1>
- Avalos, G., Hoell, K., Gardner, J., Anderson, S. and Lee, C., 2006. Impact of the invasive plant *Syzigium jambos* (Myrtaceae) on patterns of understory seedling abundance in a Tropical Premontane Forest, Costa Rica. *Revista de Biología Tropical*, 54(2). <https://doi.org/10.15517/rbt.v54i2.13883>

- Bahn, G. S., and An, B. C., 2020. Analysis of environmental purification effect of riparian forest with poplar trees for ecological watershed management: A case study in the floodplain of the dam reservoir in Korea. Sustainability (Switzerland), 12(17). <https://doi.org/10.3390/SU12176871>
- Barbet-Massin, M., Jiguet, F., Albert, C. H., and Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: How, where and how many? Methods in ecology and evolution, 3(2), 327-338.
- Barry, R. G., 2008. Mountain weather and climate. Cambridge University Press.
- Berk, R.A., 2008. Statistical learning from a regression perspective. New York: Springer, 14.
- Breiman, L., 2001. Random forests. Machine learning, 45, 5-32.
- Brundu, G., and Richardson, D. M., 2016. Planted forests and invasive alien trees in Europe: A code for managing existing and future plantings to mitigate the risk of negative impacts from invasions. NeoBiota, 30, 5-47. <https://doi.org/10.3897/neobiota.30.7015>
- Buckley, Y.M., 2014. Generalised linear models. Ecological Statistics. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199672547.003.7>.
- Canavan, K., Canavan, S., Clark, V.R., Gwate, O., Mapaura, A., Richardson, D.M., Steenhuisen, S., Martin, G.D., 2022. Invasive alien plants in the Montane areas of South Africa: Impacts and management options. In Exploring the Multiple Values of Nature: Connecting Ecosystems and People Across Landscapes; Springer: Cham, Switzerland.
- Darby, A., Draguljić, D., Glunk, A. & Gotsch, S.G., 2016. Habitat moisture is an important driver of patterns of sap flow and water balance in tropical montane cloud forest epiphytes. Oecologia, 182, 357-371.
- Dimitrakopoulos, P.G., Koukoulas, S., Michelaki, C. and Galanidis, A., 2022. Anthropogenic and environmental determinants of alien plant species spatial distribution on an island scale. Science of the Total Environment, 805, 150314.

Domisch, S., Kuemmerlen, M., Jähnig, S. C. and Haase, P., 2013. Choice of study area and predictors affect habitat suitability projections, but not the performance of species distribution models of stream biota. *Ecological Modelling*, 257, 1-10.

Dumitrascu, M., Grigorescu, I., Nastase, M., Dragota, C. and Kucsicsa, G., 2010. The main environmental driving forces of the invasive plant species in the Romanian protected areas. In *Proceedings of BALWOIS Conference on Water Observation and Information Systems for Decision Support*, 1-12.

Dyderski, M. K., and Jagodziński, A. M., 2020. Impacts of alien tree species on the abundance and diversity of terricolous bryophytes. *Folia Geobotanica*, 55(4). <https://doi.org/10.1007/s12224-020-09384-2>

Elias, F., Junior, B.H.M., de Oliveira, F.J.M., de Oliveira, J.C.A. and Marimon, B.S., 2019. Soil and topographic variation as a key factor driving the distribution of tree flora in the Amazonia/Cerrado transition. *Acta Oecologica*, 100, 103467.

Evans, J.S., Murphy, M.A., Holden, Z.A. and Cushman, S.A., 2010. Modeling species distribution and change using random forest. In *Predictive species and habitat modeling in landscape ecology: Concepts and applications*. New York, NY: Springer New York, 139-159.

Fernandes, R. F., Vicente, J. R., Georges, D., Alves, P., Thuiller, W., and Honrado, J. P., 2014. A novel downscaling approach to predict plant invasions and improve local conservation actions. *Biological Invasions*, 16, 2577-2590.

Franklin, H. M., Chen, C., Carroll, A. R., Saeck, E., Fisher, P., and Burford, M. A., 2020. Leaf litter of two riparian tree species has contrasting effects on nutrients leaching from soil during large rainfall events. *Plant and Soil*, 457(1-2), 389-406. <https://doi.org/10.1007/s11104-020-04721-y>

Fu, D., Wu, X., Huang, N., and Duan, C., 2018. Effects of the invasive herb *Ageratina adenophora* on understory plant communities and tree seedling growth in *Pinus yunnanensis* forests in Yunnan, China. *Journal of Forest Research*, 23(2). <https://doi.org/10.1080/13416979.2018.1429202>

Fukuda, S., De Baets, B., Waegeman, W., Verwaeren, J., and Mouton, A. M., 2013. Habitat prediction and knowledge extraction for spawning European grayling (*Thymallus thymallus* L.) using a broad range of species distribution models. *Environmental Modelling & Software*, 47, 1-6.

Gomes, T. F., Van de Broek, M., Govers, G., Silva, R. W. C., Moraes, J. M., Camargo, P. B., Mazzi, E. A. and Martinelli, L. A., 2019. Runoff, soil loss, and sources of particulate organic carbon delivered to streams by sugarcane and riparian areas: An isotopic approach. *Catena*, 181. <https://doi.org/10.1016/j.catena.2019.104083>

Grzyl, A., Kiedrzyński, M., Zielińska, K. M., and Rewicz, A., 2014. The relationship between climatic conditions and generative reproduction of a lowland population of *Pulsatilla vernalis*: The last breath of a relict plant or a fluctuating cycle of regeneration? *Plant Ecology*, 215(4), 457-466. <https://doi.org/10.1007/s11258-014-0316-0>

Hammouri, H.M., Sabo, R.T., Alsaadawi, R. and Kheirallah, K.A., 2020. Handling skewed data: A comparison of two popular methods. *Applied Sciences*, 10(18), 6247.

Hengl, T., De Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G., Walsh, M.G. and Gonzalez, M.R., 2014. SoilGrids1km—global soil information based on automated mapping. *PloS one*, 9(8), 105992.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B. and Guevara, M.A., 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS one*, 12(2), 0169748.

Herrero-Jáuregui, C., Sist, P., and Casado, M. A., 2012. Population structure of two low-density neotropical tree species under different management systems. *Forest Ecology and Management*, 280. <https://doi.org/10.1016/j.foreco.2012.06.006>

Hirzel, A. H., Hausser, J., Chessel, D. and Perrin, N., 2002. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology*, 83(7). [https://doi.org/10.1890/0012-9658\(2002\)083\[2027:ENFAHT\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[2027:ENFAHT]2.0.CO;2)

Huang, E., Chen, Y., Fang, M., Zheng, Y. and Yu, S., 2021. Environmental drivers of plant distributions at global and regional scales. *Global Ecology and Biogeography*, 30(3), 697-709.

Jiao, J., Zou, H., Jia, Y. and Wang, N., 2009. Research progress on the effects of soil erosion on vegetation. *Shengtai Xuebao/ Acta Ecologica Sinica*, 29(2), 85-91. <https://doi.org/10.1016/j.chnaes.2009.05.001>

Jo, I., Potter, K. M., Domke, G. M. and Fei, S., 2018. Dominant forest tree mycorrhizal type mediates understory plant invasions. *Ecology Letters*, 21(2). <https://doi.org/10.1111/ele.12884>

Kaky, E., Nolan, V., Alatawi, A. and 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. <https://doi.org/10.1016/j.ecoinf.2020.101150>

Kay, C., Bredenkamp, G. J. and Theron, G. K., 1993. The plant communities of the Golden Gate Highlands National Park in the north-eastern Orange Free State. *South African Journal of Botany*, 59(4), 442-449. [https://doi.org/10.1016/S0254-6299\(16\)30717-7](https://doi.org/10.1016/S0254-6299(16)30717-7)

Kesleman, H.J., Othman, A.R. and Wilcox, R.R., 2016. Generalized linear model analyses for treatment group equality when data are non-normal. *Journal of Modern Applied Statistical Methods*, 15(1), 4.

Kosicki, J.Z., 2020. Generalised additive models and random forest approach as effective methods for predictive species density and functional species richness. *Environmental and Ecological Statistics*, 27(2), 273-292.

Lan, G., Hu, Y., Cao, M. and Zhu, H., 2011. Topography related spatial distribution of dominant tree species in a tropical seasonal rain forest in China. *Forest Ecology and Management*, 262(8). <https://doi.org/10.1016/j.foreco.2011.06.052>

Laslier, M., Hubert-Moy, L. and Dufour, S., 2019. Mapping riparian vegetation functions using 3D bispectral LiDAR data. *Water (Switzerland)*, 11(3). <https://doi.org/10.3390/w11030483>

Li, S. Y., Ma, H. Y., Zhao, D. D., Ma, M. Y. and Qi, W. W., 2021. Research progress on the effects of fire-related cues on seed germination. *Chinese Journal of Plant Ecology*, 45(11). <https://doi.org/10.17521/cjpe.2021.0123>

Liaw, A. and Wiener, M., 2002. Classification and regression by randomForest. *R news*, 2(3), 18-22.

Liu, Q. J., An, J., Wang, L. Z., Wu, Y. Z. and Zhang, H. Y., 2015. Influence of ridge height, row grade, and field slope on soil erosion in contour ridging systems under seepage conditions. *Soil and Tillage Research*, 147. <https://doi.org/10.1016/j.still.2014.11.008>

Liu, Z., Zhang, L., Yan, Z., Ren, Z., Han, F., Tan, X., Xiang, Z., Dong, F., Yang, Z., Liu, G., Wang, Z., Zhang, J., Que, T., Tang, C., Li, Y., Wang, S., Wu, J., Li, L., Huang, C., Roos, C. and Li, M., 2020. Genomic mechanisms of physiological and morphological adaptations of Limestone Langurs to Karst Habitats. *Molecular Biology and Evolution*, 37(4), 952-968. <https://doi.org/10.1093/molbev/msz301>

Lobo, J. M., Jiménez-Valverde, A. and Hortal, J., 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, 33(1), 103-114.

Lomolino, M. V., 2001. Elevation gradients of species-density: Historical and prospective views. *Global Ecology and Biogeography*, 10(1), 3-13. <https://doi.org/10.1046/j.1466-822x.2001.00229.x>

Luan, J., Zhang, C., Xu, B., Xue, Y. and Ren, Y., 2020. The predictive performances of random forest models with limited sample size and different species traits. *Fisheries Research*, 227, 105534.

Mallinis, G., Chrysafis, I., Korakis, G., Pana, E. and Kyriazopoulos, A.P., 2020. A random forest modelling procedure for a multi-sensor assessment of tree species diversity. *Remote Sensing*, 12(7), 1210.

Mapuru, M., Hansen, C. and Haussmann, N., 2021. Exploring determinants of aardvark burrow locations at Rietvlei Nature Reserve, South Africa. *Society of South African Geographers and the Southern African Association of Geomorphologists*, 62.

Mapuru, M., Xulu, S. and Gebreslasie, M., 2024. Modelling possible habitats for poplar invasion in South Africa, *Koedoe*, 66(1), 1783. <https://doi.org/10.4102/koedoe.v66i1.1783>

Marchante, E., Kjølner, A., Struwe, S. and Freitas, H., 2008. Short- and long-term impacts of *Acacia longifolia* invasion on the belowground processes of a Mediterranean coastal dune ecosystem. *Applied Soil Ecology*, 40(2). <https://doi.org/10.1016/j.apsoil.2008.04.004>

Martínez, I., Wiegand, T., González-Taboada, F. and Obeso, J. R., 2010. Spatial associations among tree species in a temperate forest community in North-western Spain. *Forest Ecology and Management*, 260(4), 456-465.

Mashiane, K. K., Ramoelo, A., Adelebu, S., and Daemane, E., 2023. Assessing environmental factors contributing to plant species richness in mountainous mesic grasslands. *Koedoe*, 65(1), 1-7.

Masole, P., Steenhuisen, S.L. and Martin, G.D., 2022. Current status of the invasive shrub *Berberis julianae* CK Schneid (Berberidaceae) in Golden Gate Highlands National Park (Free State Province, South Africa). *South African Journal of Botany*, 150, 99-105.

McGill, B. J., 2010. Matters of scale. *Science*, 328(5978), 575-576.

Moeslund, J.E., Arge, L., Bøcher, P.K., Dalgaard, T., Ejrnæs, R., Odgaard, M.V. and Svenning, J.C., 2013. Topographically controlled soil moisture drives plant diversity patterns within grasslands. *Biodiversity and conservation*, 22, 2151-2166.

Mohammady, M., Pourghasemi, H.R., Yousefi, S., Dastres, E., Edalat, M., Pouyan, S. and Eskandari, S., 2021. Modeling and prediction of habitat suitability for *Ferula gummosa* medicinal plant in a mountainous area. *Natural Resources Research*, 30, 4861-4884.

Moore, I. D., Gessler, P. E., Nielsen, G. A. and Peterson, G. A., 1993. Soil attribute prediction using terrain Analysis. *Soil Science Society of America Journal*, 57(2). <https://doi.org/10.2136/sssaj1993.03615995005700020058x>

Ntshidi, Z., Gush, M.B., Dzikiti, S. and Le Maitre, D.C., 2018. Characterising the water use and hydraulic properties of riparian tree invasions: A case study of *Populus canescens* in South Africa, *Water SA*, 44, 328-337. <https://doi.org/10.4314/wsa.v44i2.18>

Onwuka, B. and Mang, B., 2018. Effects of soil temperature on some soil properties and plant growth. *Plants & Agriculture Research*, 8(1), 34-37. <https://doi.org/10.15406/apar.2018.08.00288>

R Development Core Team., 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rehman, H. U., Tariq, A., Ashraf, I., Ahmed, M., Muscolo, A., Basra, S. M. A. and Reynolds, M., 2021. Evaluation of physiological and morphological traits for improving spring wheat adaptation to terminal heat stress. *Plants*, 10(3), 1-15. <https://doi.org/10.3390/plants10030455>

Reinhart, K.O. and Callaway, R.M., 2006. Soil biota and invasive plants. *New phytologist*, 170(3), 445-457.

Reisz, R.R., Scott, D., Sues, H., Evans, D.C. and Raath, M.A., 2005. Embryos of an early Jurassic Prosauropod dinosaur and their evolutionary significance. *Science*, 309, 761-764.

Ridolfi, L., Laio, F. and Odorico, P. D., 2008. Fertility island formation and evolution in dryland *Ecology and Society*, 13(1). <http://www.ecologyandsociety.org/vol13/iss1/art5/>

South African National Parks (SANParks)., 2020. Golden Gate Highlands National Park Management Plan, SANParks, Pretoria.

Shangguan, W., Hengl, T., Mendes de Jesus, J., Yuan, H. and Dai, Y., 2017. Mapping the global depth to bedrock for land surface modeling. *Journal of Advances in Modeling Earth Systems*, 9(1), 65-88.

Tsui, C. C., Tsai, C. C. and Chen, Z. S., 2013. Soil organic carbon stocks in relation to elevation gradients in volcanic ash soils of Taiwan. *Geoderma*, 119-127. <https://doi.org/10.1016/j.geoderma.2013.06.013>

Van Wilgen, B. W., Dyer, C., Hoffmann, J. H., Ivey, P., Le Maitre, D. C., Moore, J. L., Richardson, D. M., Rouget, M., Wannenburgh, A. and Wilson, J. R. U., 2011. National-scale strategic approaches for managing introduced plants: Insights from Australian acacias in South Africa. *Diversity and Distributions*, 17(5), 1060-1075. <https://doi.org/10.1111/j.1472-4642.2011.00785.x>

- VanDerWal, J., Shoo, L. P., Graham, C. and Williams, S. E., 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecological Modelling*, 220(4), 589-594.
- Vicente, J., Randin, C. F., Gonçalves, J., Metzger, M. J., Lomba, Â., Honrado, J. and Guisan, A., 2011. Where will conflicts between alien and rare species occur after climate and land-use change? A test with a novel combined modelling approach. *Biological Invasions*, 13, 1209-1227.
- Vicente, J.R., Kueffer, C., Richardson, D.M., Vaz, A.S., Cabral, J.A., Hui, C., Araújo, M.B., Kühn, I., Kull, C.A., Verburg, P.H. and Marchante, E., 2019. Different environmental drivers of alien tree invasion affect different life-stages and operate at different spatial scales. *Forest Ecology and Management*, 433, 263-275.
- Vilà, M., Espinar, J. L., Hejda, M., Hulme, P. E., Jarošík, V., Maron, J. L., Pergl, J., Schaffner, U., Sun, Y. and Pyšek, P., 2011. Ecological impacts of invasive alien plants: a meta-analysis of their effects on species, communities and ecosystems. *Ecology Letters*, 14(7), 702-708.
- Vincenzi, S., Zucchetta, M., Franzoi, P., Pellizzato, M., Pranovi, F., De Leo, G.A. and Torricelli, P., 2011. Application of a Random Forest algorithm to predict spatial distribution of the potential yield of *Ruditapes philippinarum* in the Venice lagoon, Italy. *Ecological Modelling*, 222(8), 1471-1478.
- Weidenhamer, J.D. and Callaway, R.M., 2010. Direct and indirect effects of invasive plants on soil chemistry and ecosystem function. *Journal of Chemical Ecology*, 36, 59-69.
- Yudaputra, A., Pujiastuti, I. and Cropper, W. P., 2019. Comparing six different species distribution models with several subsets of environmental variables: Predicting the potential current distribution of *Guettarda speciosa* in Indonesia. *Biodiversitas*, 20(8), 2321-2328. <https://doi.org/10.13057/biodiv/d200830>

Appendix A.

Table 5.3 Descriptive statistics of all predictor variables used in the study.

Predictors	Units	Min	Max	Mean	SD	Skew
Silt	g/kg	195	246	232.5735294	14.53489489	-0.871272728
Slope	angle°	89.97338867	89.98472595	89.97759696	0.005397273	0.505260403
Aspect		49.80130768	350.1307678	235.4042661	146.8992046	-0.495182267
Sand	g/kg	494	541	509.9705882	15.36711996	0.800088171
Soil organic carbon	dg/kg	46	56	50.94117647	2.184629843	0.192387591
Distance to the rivers.	m	489.4346539	4350.189629	1801.332064	1155.43566	1.113651901
Moisture content		-0.186813191	0.086956523	-0.02522498	0.07836258	-0.061289992
pH	pH ¹⁰	58	62	59	0.622944381	1.908374528
Nitrogen of topsoil	cg/kg	221	270	245.4411765	16.8358026	0.056126784
Elevation	masl	1840	2057	1995.852941	94.22619098	-1.089447137
Bulk density	cg/cm ³	125	133	129.2647059	2.696695861	0.194323211
NDVI		-0.002375297	0.281997025	0.126055723	0.048336962	0.261934663
NDWI		-0.291122288	0.056886226	-0.165957023	0.062532357	0.820349642

CHAPTER 6 SYNTHESIS AND CONCLUSION

6.1 SYNTHESIS

The proliferation of invasive alien species (IAS), such as poplar (*Populus spp.*), poses a significant threat to both natural and human-modified ecosystems due to their adaptability and competitive dominance over native species. In response to these challenges, this study explored the use of geospatial technologies, including Geographic Information Systems (GIS) and remote sensing, to monitor, map, and predict poplar distribution in South Africa. This chapter synthesizes the major findings from the four research objectives and discusses their broader implications for conservation, land management, and geospatial applications. This section further integrates the recommendations and limitations of the overall research.

6.1.1 Key findings and their integration

Review and synthesize remote sensing studies of poplars over a 30-year period (1991–2022).

The first objective provided a comprehensive review of 148 peer-reviewed studies, revealing an increasing reliance on Earth observation techniques for poplar research between 1991 and 2022. The findings highlight the predominant use of airborne LiDAR, followed by satellite imagery and, to a lesser extent, ground-based sensors. The geographic focus of these studies was primarily in the Global North, with limited research conducted in African landscapes. Furthermore, the review underscores the growing use of spectral indicators and machine learning in assessing poplar health and productivity.

Synthesis: The widespread adoption of remote sensing and machine learning techniques underscores their effectiveness in IAS research. However, the limited focus on South African landscapes necessitates targeted research efforts to bridge this knowledge gap.

Examine the capacity of Random Forest algorithm and Sentinel-2 images for discriminating poplar species against other land cover types in mountainous regions.

The second objective successfully demonstrated the capability of Sentinel-2 imagery combined with Random Forests (RF) for mapping poplar in mountainous regions of South Africa. The

classification accuracy was notably higher (91%) in the first half of the year compared to the second half (87%), indicating seasonal variations in spectral responses. Key spectral variables, including SWIR-1, NDVI, Blue, PI-1, MNDWI, NIR, and PI-3, played a crucial role in distinguishing poplar trees from other land cover types.

Synthesis: These findings validate the applicability of Sentinel-2 imagery and RF classification in IAS mapping, particularly in complex terrains. The seasonal differences in classification accuracy highlight the importance of temporal considerations in remote sensing studies.

Predict the potential suitable habitat of poplar trees using regression and machine learning techniques in South Africa.

The third objective applied generalized linear models (GLM), RF, and Support Vector Machines (SVM) to predict the potential distribution of poplar trees under current climatic conditions. Among these models, RF exhibited the highest predictive accuracy. The study found that areas with high rainfall, particularly those experiencing significant precipitation in the driest and warmest quarters, were the most suitable habitats for poplar.

Synthesis: These findings align with global studies that emphasize the role of climatic factors in determining IAS distribution. The demonstrated effectiveness of machine learning models in habitat prediction reinforces their utility for proactive management and conservation planning.

Identify the local environmental variables which influence the occurrence and distribution of poplar in mountainous areas.

The fourth objective identified key environmental factors influencing poplar distribution in mountainous regions. Using a GLM approach, the study ranked these factors based on their significance, revealing that topographical features (aspect and slope), soil composition (sand, silt, and organic carbon), and surface moisture (NDMI) played crucial roles. Distance from rivers emerged as a particularly influential factor, further emphasizing the importance of water availability in poplar spread.

Synthesis: The findings corroborate existing literature on IAS ecology, demonstrating that water availability and terrain characteristics are primary determinants of invasive species distribution. This understanding can inform targeted management strategies aimed at controlling poplar expansion in sensitive ecosystems.

6.1.2 Broader implications

Implications for conservation and land management

The study's findings have significant implications for conservationists and land managers in South Africa. The ability to map, predict, and understand the environmental drivers of poplar distribution enables the development of more effective monitoring and control strategies. The identification of key spectral indices and machine learning models provides a robust framework for future IAS studies, potentially extending to other invasive species.

Advancements in geospatial technology application

This study highlights the growing importance of GIS and remote sensing in ecological research. The demonstrated effectiveness of Sentinel-2 imagery, coupled with machine learning classifiers, presents a scalable and cost-effective approach for IAS monitoring. Moreover, the study's integration of various predictive modelling techniques showcases the potential of geospatial analytics in ecological forecasting.

Addressing research gaps in South Africa

By being the first to systematically apply geospatial techniques to study poplar distribution in South Africa, this research fills a critical knowledge gap. The findings emphasize the need for continued research efforts in African landscapes, particularly in leveraging Earth observation data for biodiversity conservation.

6.1.3 Recommendations

Future studies should integrate LiDAR and multispectral satellite sensor images to leverage LiDAR's high spatial resolution alongside the extensive coverage of satellite imagery. The combination of plant-specific indices, multispectral-temporal image data, and machine learning algorithms has been shown to improve mapping accuracy, and the study strongly recommends using the poplar tree index (PTI) in conjunction with these techniques for mapping poplar in sensitive landscapes. More research should be conducted on poplar in mountainous regions using geospatial approaches, with international collaboration between developed and underdeveloped countries encouraged to facilitate knowledge exchange and resource sharing. Expanding the methodology to map other invasive alien tree species in mountainous regions would further enhance the understanding of IAS dynamics. Additionally, future studies should evaluate tree mapping accuracy using different input combinations, such as bands-only images, indices-only data, and a mixture of both, while integrating machine learning techniques to refine classification accuracy. While this study focused on present poplar distribution, future research should forecast the species' potential distribution under changing climatic conditions by incorporating future bioclimatic variables. Furthermore, biological interactions between poplar and coexisting species, as well as the impact of fire on its spread and adaptation, should be investigated. Finally, since all poplar species pose risks as invasive alien plants in South Africa, similar studies should be conducted across their inhabited regions to ensure comprehensive monitoring and management strategies.

6.1.4 Limitations of the study

The literature survey paper focused exclusively on studies published in peer-reviewed journals. It did not analyze documents such as thesis, dissertations, scientific reports, or conservation management plans. While synthesizing these types of documents is encouraged, it is essential to ensure their content is of high quality and validity. Although South Africa has six species of poplar, Chapter 5 of this research concentrated only on two of these species. Future research is encouraged for the remaining species, particularly using similar methodological approaches.

6.2 CONCLUDING REMARKS

This study successfully demonstrates the power of geospatial technologies in mapping, predicting, and understanding the spread of poplar in South Africa. By synthesizing the major findings from the four research objectives, it becomes evident that remote sensing, machine learning, and environmental analysis collectively contribute to a more comprehensive understanding of IAS dynamics. The insights gained from this research not only advance scientific knowledge but also provide practical guidance for conservation efforts and policy formulation in managing invasive species.