OBJECT-BASED GEOMORPHOMETRY AND SOIL SPECTROSCOPY IN A FLUVIAL TERRACE DOMINATED LANDSCAPE

By GERRIT JACOBUS LOUW

Dissertation presented for the degree of Doctor of Philosophy
in the Faculty of Science at Stellenbosch University

Promotors: Prof. A van Niekerk

Dr. A Rozanov

March 2021

DECLARATION

By submitting this research dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

With regard to Chapters 3, 4, 5 and 6 the nature and scope of my contribution were as follows:

Chapter	Nature of contribution	Extent of contribution (%)
Chapter 3	This chapter was co-authored by my supervisor who helped in the writing of the manuscript. Additionally, Andrei Rozanov helped with the field reconnaissance work and conceptualisation and writing of the manuscript, Willem de Clercq also helped with the field reconnaissance work and Garry Paterson helped in the writing of the manuscript.	G Louw 85% A Van Niekerk 5% A Rozanov 5% G Paterson 2.5% W De Clercq 2.5%
Chapter 4	This chapter was published as a journal article (Geomorphology) and was co-authored by my supervisor who helped in the conceptualisation and writing of the manuscript.	G Louw 85% A Van Niekerk 15%
Chapter 5	This chapter was co-authored by my supervisor, who helped in the conceptualisation and writing of the manuscript, and Andrei Rozanov who helped in the conceptualisation of the manuscript and the collection of reference data.	G Louw 85% A Van Niekerk 10% A Rozanov 5%
Chapter 6	This chapter was published as a journal article (Geomorphology) and was co-authored by my supervisor, who helped in the conceptualisation and writing of the manuscript, and Andrei Rozanov who contributed to the conceptualisation of the manuscript and, together with Liesl Wiese, helped with the chemometric data analyses.	G Louw 85% A Van Niekerk 6.5% A Rozanov 6.5% L Wiese 2%

Signature of candidate: Declaration with signature in possession of candidate and Supervisor

Signature of supervisor: Declaration with signature in possession of candidate and Supervisor

Date: March 2021

SUMMARY

The adaptation and mitigation of adverse global change necessitate better understanding and management of natural resources. Fluvial terrace staircases are a paleo-fluvial phenomenon critical to mining, agriculture, ecosystem services and the deciphering of paleo-tectonic and -climatic conditions. This study aims to develop methodologies for generating fine-scale object-based fluvial terrace maps. A brief treatise contextualises relevant concepts, techniques and methodologies related to object-based geomorphometry and soil spectroscopy.

The paleo-fluvial landscape of the Great Letaba River Catchment was investigated. An overview of the formation of the Great Escarpment and related tectonic uplift events provides Lowveld paleo-fluvial context. Moreover, the limited available evidence detailing the existence of fluvial terraces and planation surfaces along the Great Letaba River is reviewed. The combination of field reconnaissance and terrain analysis results with 1:250 000 scale soil data enabled the characterisation of extensive fluvial terrace and planation surface landforms and the collection of critical geomorphometric reference data. The findings expand existing morphological knowledge of the Great Letaba River Catchment fluvial terraces and planation surfaces and identify a distinct landscape, soil, climate and lithology paradigm that drives a complex cycle of landscape evolutionary processes.

Specific geomorphometric land surface segmentation approaches delineate predefined landform units using largely scale-independent techniques that do not incorporate hierarchical landscape representations. This study details a new approach that delineates predefined landforms at multiple scale levels using discrete geomorphometric principles and local variance statistical techniques. First, land surface segmentation scale optimisation was exposed as an ill-structured problem, and a methodology for defining well-structured landform conditions was outlined. Next, an ensemble of scale optimisation techniques was constructed and applied to evaluate each well-structured condition at incrementally increased scale increments. Agreement between the scale optimisation techniques indicates scale levels at which the predefined landform conditions are met. The results indicate that ensemble scale optimisation techniques, unlike existing single technique approaches, produce refined scale selections that minimises analyst involvement in the scale optimisation process.

The bulk of geomorphometric fluvial terrace mapping strategies employ per-pixel strategies that are either based on deterministic approaches that require extensive user parameterisation or black-box supervised classifiers that exclude expert knowledge from the classification process. A new

methodology that couples object-based landform units with white-box decision tree (DT) classifiers was envisioned as an intermediary between knowledge-based and black-box classifiers. To evaluate the new methodology, an experiment was designed where both binary classes (terrace and non-terrace) and multiple fluvial terraces (non-terrace and ten fluvial terrace levels) were classified using per-pixel and object-based approaches. Through inductive DT classifier analysis, the object-based rulesets were shown to produce the most intuitively interpretable fluvial terrace rulesets. Moreover, qualitative and quantitative accuracy assessments reaffirm the superiority of object-based fluvial terrace classifications.

The differentiation and correlation of fluvial terrace remnants in eroded landscapes often require resource intensive in situ stratigraphic and physiographic interpretations. Using discrete geomorphometric techniques and soil spectroscopy, the utility of the soil-landscape paradigm for fluvial terrace level mapping was illustrated. First, a chemometric principle component analysis (PCA) and partial least-squares regression (PLSR) approach was used to establish a good linear relationship between 88 spectral signatures and the relative height of fluvial terraces. Next, random forest (RF) was used to classify two sets of fluvial terrace level classes and to isolate the most important spectral wavelengths. Subsequent RF models indicated that individual wavelengths associated with smectite and kaolinite fractions provide sufficient variance to accurately classify fluvial terrace levels. The methodology introduced outlines a low-cost, semi-automated spectroscopic alternative to in situ fluvial terrace mapping approaches.

Fluvial terrace maps are crucial for facilitating – among other things – the production of fine-scale resource inventories, land use planning and global change adaptation and mitigation strategies. Yet, most southern African fluvial terraces remain largely unexplored. This research contributes new knowledge to the fields of southern African paleo-fluvial geomorphology, geomorphometry and fluvial terrace mapping through landscape characterisations of the Great Letaba River Catchment and the development of novel geomorphometric and soil spectroscopic methodologies. These contributions provide a sound foundation for further Lowveld paleo-fluvial investigations, the regional mapping of fluvial terraces and the development of transferable geomorphometric landform mapping methodologies.

KEYWORDS.

Fluvial terraces, Letaba River, Lowveld, discrete geomorphometry, land surface segmentation, landform classification, multi-resolution segmentation, machine learning, soil spectroscopy

OPSOMMING

Die aanpassing en versagting van nadelige globale verandering noodsaak beter kennis en bestuur van natuurlike hulpbronne. Rivierterrastrappe is 'n wêreldwye paleo-fluviale verskynsel wat van kritieke belang vir mynbou, landbou, ekostelseldienste en die ontsyfering van paleo-tektoniese en klimaatstoestande is. Die doel van die studie is om nuwe metodologieë vir die skep van fynskaalse voorwerpgebaseerde kaarte van fluviale terrasse te ontwikkel. 'n Kort verhandeling kontekstualiseer relevante konsepte, tegnieke en metodologieë wat met voorwerpgebaseerde geomorfometrie en grondspektroskopie verband hou.

Die paleo-fluviale landskap van die Groot Letabarivier opvanggebied is ondersoek. 'n Oorsig van die Groot Platorand formasie en verwante tektoniese opheffingsgebeurtenisse verskaf konteks oor die Laeveldse paleo-fluviale prosesse. Daarbenewens word die beperkte beskikbare bewyse wat die bestaan van fluviale terrasse en planasieoppervlaktes langs die Groot Letabarivier hersien. Die kombinasie van veldverkennings- en terreinanalise-resultate met 1: 250 000 skaalgronddata het die karakterisering van wydverspreide fluviale-terras- en planasie-oppervlaklandvorms en die inwinning van kritieke geomorfometriese verwysingsdata moontlik gemaak. Die bevindings brei die bestaande morfologiese kennis van die Groot Letabarivier opvansgebied fluviale terrasse en planasieoppervlaktes aansienlik uit en identifiseer 'n duidelike paradigma vir landskap, grond, klimaat en litologie wat 'n komplekse kringloop van evolusieprosesse in die landskap dryf.

Spesifieke geomorfometriese landoppervlak-segmenteringsbenaderings definieer vooraf-gedefinieerde landvorm-eenhede met grootliks skaal-onafhanklike tegnieke wat nie hiërargiese landskapvoorstellings inkorporeer nie. Hierdie studie sit 'n nuwe benadering uiteen wat vooraf-gedefinieerde landvorms op meervoudige skaalvlakke met behulp van diskrete geomorfometriese beginsels en statistiese tegnieke vir plaaslike variansie afbaken. Eerstens is die skaaloptimering van landoppervlak segmentering as 'n ongestruktureerde probleem blootgelê, en 'n metodologie vir die definisie van goed gestruktureerde landvormtoestande is geïllustreer. Vervolgens is 'n ensemble van skaaloptimaliseringstegnieke saamgestel en toegepas om elke goed gestruktureerde toestand met toenemende skaalstygings te evalueer. Ooreenkoms tussen die skaaloptimaliseringstegnieke dui skaalvlakke aan waar daar aan die vooraf-gedefinieerde landvormvoorwaardes voldoen word. Die resultate dui aan dat ensemble-skaaloptimaliseringstegnieke, in teenstelling met bestaande enkele tegniek benaderings, verfynde skaalkeuses oplewer wat ontleder betrokkenheid by die skaaloptimaliseringsproses beperk.

Die meeste geomorfometriese fluviale-terraskaartstrategieë gebruik per-piksel strategieë wat óf gebaseer is op deterministiese benaderings wat omvattende gebruikersparameterisering vereis, óf swart-boks gekontroleerde-klassifikasies wat kundige kennis van die klassifikasieproses uitsluit. ’n Nuwe metodologie wat voorwerpgebaseerde landvorm-eenhede met witboks-beslissingsboomklasseerders koppel, is as ’n tussenganger tussen kennisgebaseerde- en swartboks-klassifiseerders beskou. ’n Eksperiment, wat beide binêre klasse (terras en nie-terras) en meervoudige fluviale terrasse (nie-terras en tien fluviale-terrasvlakke) met behulp van per-piksel en voorwerpgebaseerde benaderings klassifiseer, is ontwerp om die nuwe metodologie te evalueer. Deur middel van induktiewe beslissingsboomklassifiseerder-analise is getoon dat die voorwerpgebaseerde reëlstelle die mees instinktiewe, interpreteerbare reëlstelle vir fluviale terrasse lewer. Daarbenewens bevestig kwalitatiewe en kwantitatiewe akkuraatheidsbeoordelings die superioriteit van voorwerpgebaseerde fluviale-terrasklassifikasies.

Die differensiasie en korrelasie van oorblyfsels van fluviale terrasse in geërodeerde landskappe vereis dikwels hulpbronintensiewe in situ stratigrafiese en fisiografiese interpretasies. Met behulp van diskrete geomorfometriese tegnieke en grondspektroskopie, word die nut van die grondlandskapparadigma vir die kartering van fluviale-terrasvlakke geïllustreer. Eerstens is ’n chemometriese beginsel-komponent-analise en gedeeltelike regressie van die kleinste vierkante benadering gebruik om ’n goeie lineêre verband tussen 88 spektrale tekens en die relatiewe hoogte van fluviale terrasse vas te stel. Vervolgens is ewekansige woud (random forest) gebruik om twee stelle fluviale-terrasvlakklasse te klassifiseer en die belangrikste spektrale golflengtes te isoleer. Daaropvolgende ewekansige-woud-modelle het aangedui dat individuele golflengtes wat verband hou met smektiet- en kaolinietfraksies voldoende variansie bied om fluviale-terrasvlakke akkuraat te klassifiseer. Die aangevoerde metodologie sit ’n goedkoop, semi-outomatiese spektroskopiese alternatief vir in situ fluviale-terras-karteringsbenaderinge uiteen.

Fluviale-terraskaarte is van kardinale belang in, onder andere, die fasilitering van die produksie van fynskaalse hulpbronvoorrade, grondgebruikbeplanning en globale veranderingsaanpassings- en versagtingsstrategieë. Tog bly die meeste fluviale terrasse in suidelike Afrika grotendeels onverken. Hierdie navorsing dra nuwe kennis by tot die velde van suidelike Afrika paleo-fluviale geomorfologie, geomorfometrie en fluviale-terraskartering deur landskapkarakterisering van die Groot Letaba-rivieropvangsgebied en die ontwikkeling van nuwe geomorfometriese en grondspektroskopiese metodologieë. Hierdie bydraes vorm ’n goeie grondslag vir verdere ondersoek na die Laeveld-landskap, die plaaslike kartering van terrasse en die ontwikkeling van oordraagbare geomorfometriese landvormkarteringmetodologieë.

TREFWOORDE

Fluviale terrasse, Letaba-rivier, Laeveld, diskrete geomorfometrie, landoppervlak segmentering, landvorm-klassifikasie, multi-resolusie segmentering, masjienleer, grondspektroskopie

ACKNOWLEDGEMENTS

I sincerely thank:

- My partner, family and friends for their infinite support.
- My supervisors, Prof. Adriaan van Niekerk and Dr. Andrei Rozanov, for their guidance and patient encouragement.
- The Centre for Geographical Analysis and Stellenbosch University Geoinformatics staff for their technical support.
- The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF.

CONTENTS

TABLES

FIGURES

ACRONYMS AND ABBREVIATIONS

ALCoM	Automated land-components mapper
ANN	Artificial neural network
ANUDEM	Australian National University digital elevation model
ASTER	Advanced spaceborne thermal emission and reflection radiometer
DEM	Digital elevation model
DRS	Diffuse reflectance spectroscopy
DSM	Digital soil mapping
DT	Decision trees
ESP	Estimation of scale parameter
GEOBIA	Geographic object-based image analysis
GIS	Geographic information system
GISc	Geographic information science
GLR	Great Letaba River
GLWC	Great Letaba Water Catchment
HARC	Height above river channel
ISODATA	Iterative self-organising data analysis technique algorithm
k-NN	k-nearest neighbour
KNP	Kruger National Park
LE	Letaba Estates
LLR	Little Letaba River

LSP	Land surface parameter
LTS	Land type survey
LV	Local variance
LVR	Local variance ratio
MAE	Mean absolute error
Mid-IR	Mid-infrared
MLR	Middle Letaba River
MRS	Multi-resolution segmentation
MSC	Multiplicative scatter correction
NIR	Near infra-red
OBLV	Object boundary local variance
OOB	Out-of-bag error
OOO	Object-orientated ontology
OOP	Object-orientated programming
PCA	Principal component analysis
PLSR	Partial least-squares regression
R^2	Coefficient of determination
RF	Random Forest
RMSE	Root mean square error
ROC	Rate of change
RPD	Residual prediction deviation

SP	Scale parameter
SRTM	Shuttle Radar Topography Mission
SUDEM	Stellenbosch University digital elevation mode
SVM	Support vector machine
VDTCN	Vertical distance to channel network

CHAPTER 1: INTRODUCTION

The dawn of the Anthropocene marks significant global change in climate, ecosystems and species behaviour (Lewis & Maslin 2005; Steffen et al. 2011). Mitigation and adaptation strategies that serve to limit the adverse effects of global change necessitate better understanding and management of natural resources (Holm et al. 2013). The need for accurate and fine-resolution geomorphological and soil resource inventories are increasing (Sanchez et al. 2009), and cost-effective data-driven methodologies are required to replace resource intensive conventional mapping approaches (McBratney, Mendonca Santos & Minasny 2003; Sanchez et al. 2009). This dissertation's primary foci are centred around the development of geomorphometric methodologies for mapping fluvial terraces – dynamic landscape features critical for placer mineral mining, agriculture, ecosystem services and paleo-fluvial interpretations of tectonic uplift, climate and even paleontological settlements (Bridgland & Westaway 2008; Leshchinskiy et al. 2006; Pazzaglia 2013; Rozanov et al. 2017; Vandenberghe 2015). This introductory chapter succinctly overviews South African fluvial terraces, the principals of object-based geomorphometry and spectroscopy assisted geomorphometry. Moreover, the research problem, aim and objectives are formulated and the research methodology, design and dissertation structure are outlined.

1.1 FLUVIAL TERRACES OF THE SOUTH AFRICAN PALEO-FLUVIAL LANDSCAPE

Fluvial terraces are landforms found where variations in the vertical incision rate of rivers result in the deposition of fluvial sediments in step-like sequences called staircases. The rate of vertical incision of rivers is largely controlled by environmental (vegetation, geomorphic and hydrologic) responses to climate change, tectonic uplift and base-level changes (Bridgland & Westaway 2008; Pazzaglia 2013). Each level of a fluvial terrace staircase represents the remnants of an old floodplain that was abandoned during the incision of a river into underlying bedrock, visible as elongated flat surfaces that run parallel to a river's course (Leopold, Wolman & Miller 1964; Pazzaglia 2013). Fluvial terraces are formed globally. Northern hemisphere terrace formation during the Quaternary has been attributed to glacial-interglacial-driven geomorphic and hydrologic fluctuations linked to solar isolation changes and the 100 ky Milankovitch cycle (Bridgland & Westaway 2008). The southern African continent, on the other hand, has not been subjected to

continental glaciation events, and fluvial terrace formation has been largely attributed to tectonic uplift events linked to the breakup of Gondwana (Partridge & Maud 1987).

The paleo-fluvial landscape of South Africa has been extensively studied, with fluvial terraces being a key research interest for over a hundred years (Dollar 1998). The earliest documented observations of South African rivers were by European colonists, travellers and missionaries (Lewis-Williams 1981). Scientific research efforts commenced after the first discovery of South African diamonds along the Orange River in 1867 (De Wit 1996). Subsequent discoveries of diamonds along the Vaal (Shaw 1872; Du Toit 1933) and Limpopo rivers (Trevor, Mellor & Kynaston 1908) and in the Schweizer Reneke district (De Wit 1996) focussed concerted research efforts largely on diamondiferous terraces (Du Toit 1910). By the 1940s, scholars were characterising regional drainage patterns (Wellington 1945; Wellington 1929), studying river superimposition (Wellington 1941) and reconstructing paleo-drainage lines (King 1944; Taljaard 1944). Factors driving fluvial terrace formation were attributed to tectonic uplift (Wellington 1955) and climatic fluctuations (Maufe 1930; Rogers 1922). Subsequently, the search and excavation of diamond-bearing terraces and gravels led to a range of publications relating to the geomorphology of three major diamond-bearing drainage basins, namely the Vaal River (Butzer et al. 1974; Helgren & Butzer 1974; Van Riet Lowe 1952; De Wit 2004), Orange River (De Wit 1996) and the Molopo River (Marshall & Baxter-Brown 1995) basins. Research motivated by academic objectives were granted less attention; yet, notable geomorphological surveys include investigations of the Eerste Rivier terraces near Stellenbosch (Krige 1927; Shand 1913) and the geologic setting of the Gouritz River (Rogers 1903). Moreover, extensive investigations of the Sundays River Valley, conducted between 1945 and 1968 (Ruddock 1957; Ruddock 1968; Ruddock 1945) and again between 1994 and 1996 (Hattingh 1994; Hattingh 1996), showed that fluvial terrace formation along coastal rivers can be attributed to base-level variations (sea-level changes) brought about by marine transgressions.

Despite the extensive paleo-fluvial geomorphology research conducted in southern Africa during the last century (Dollar 1998) and seminal research that characterises the southern African fluvial landscapes (Partridge et al. 2010; Partridge & Maud 1987), many fluvial terrace staircases remain unexplored. Recent fluvial geomorphology research of the Letaba Valley, an arid and semi-arid landscape located in Limpopo (South Africa), has largely been limited to the confines of the Kruger National Park (Heritage, Moon & Large 2001; Moon & Heritage 2001) and the production of the 1:250 000 scale Land Type Survey (LTS; Land Type Survey Staff 2006). Forestry and

agricultural activities (Moon & Heritage 2001) in the upper reaches of the catchment are largely restricted to the Great Letaba River (GLR) shaped topography and its associated soils. Renewed interest in the geomorphology of the Great Letaba Water Catchment (GLWC) was stimulated by floods in the years 2000 and 2012 – which resulted in infrastructure damage, mudslides, erosion and loss of life – and increased awareness of the importance of the river for agriculture and conservation (Heritage, Moon & Large 2001; Rozanov et al. 2017). While the existence of fluvial terraces along the GLR are known (Botha & De Wit 1996; Moon & Heritage 2001; Rozanov et al. 2017), characterisations and maps of the fluvial terraces are severely lacking. The main reason for this knowledge gap is that most higher-lying terraces of the GLWC have been removed from the landscape during the African and post-African I planation events (Partridge & Maud 1987) and are preserved only as remnants in the landscape.

1.2 OBJECT-BASED LAND SURFACE MAPPING

Geographic object-based image analysis (GEOBIA) generally involves two steps (Blaschke 2010). Image segmentation, the first step, aims to create objects that spatially delineate the target features (e.g. fluvial terraces). The second step is classification, which involves assigning each object to a categorical class representative of the target feature (e.g. different fluvial terrace levels). The following two subsections briefly overview concepts and challenges concerning object-based land surface segmentation and classification.

1.2.1 Land surface segmentation

Progressive computing developments and the increasing availability of high-resolution digital elevation models (DEMs) stimulated the mainstream adoption of segmentation techniques and GEOBIA applications (Drăgut, Eisank & Strasser 2011; Smith & Morton 2008). The increasing popularity of segmentation techniques for geomorphometric purposes can be attributed to the ability of segmentation algorithms to incorporate scale when delineating land surface features. Initially, only coarse resolution digital elevation data were available to analysts, and the issue of scale was largely ignored when performing per-pixel analyses (Drăgut, Eisank & Strasser 2011). When employing high-resolution digital elevation data, on the other hand, the prevalence of spatial autocorrelation and the scale-dependency of land surface parameters (LSPs) can impede interpolation computational efficiency and classification accuracy (Drăgut, Eisank & Strasser 2011). Although some progress has been made in raster-based scale optimisation (Hengl 2006; Jasiewicz & Stepinski 2013; Woodcock & Strahler 1987), approaches that utilise object-based

segmentation algorithms to upscale high-resolution terrain data are frequently lauded as superior (Drăgut, Eisank & Strasser 2011; Drăguţ, Tiede & Levick 2010).

Multi-resolution segmentation (MRS) is a region-merging segmentation algorithm that has been shown to delineate landform (morphological) boundaries accurately (Van Niekerk 2010). The algorithm is parameterised with the use of three user-defined settings. Shape and compactness (the first two) determine the homogeneity of the shape of the resultant objects. The third MRS setting, scale parameter (SP), determines the maximum standard deviation of the homogeneity criterion and controls both the shape and spectral (morphometric) homogeneity (Baatz & Schäpe 2000; Trimble 2014b). In other words, SP controls the object internal homogeneity and, subsequently, the size of the resultant objects.

SP is the most important MRS parameter when performing land surface segmentation, while the shape homogeneity criteria are often excluded from the analysis to ensure that the segmentation procedure only considers morphometric homogeneity (Drăguţ & Eisank 2012; Eisank, Smith & Hillier 2014). Two SP optimisation approaches are typically employed for land surface segmentation applications. The first utilises analyst interpretation to qualitatively compare and evaluate the delineated objects against predefined criteria (Anders, Seijmonsbergen & Bouten 2015; Van Asselen & Seijmonsbergen 2006; Mashimbye, De Clercq & Van Niekerk 2014; Seijmonsbergen, Hengl & Anders 2011). The accurate selection of appropriate SP values using analyst interpretation is subject to analyst experience and intuition. The second SP optimisation approach aims to mitigate analyst bias by employing quantitative data-driven methodologies (Anders, Seijmonsbergen & Bouten 2011; Drăguţ, Tiede & Levick 2010). A popular data-driven SP optimisation approach is the estimation of SP (ESP) tool, which uses local variance (LV) to detect characteristic levels of scale by employing standard deviation to exploit LSPs spatial autocorrelation (Dornik, Drăguţ & Urdea 2017; Drăguţ & Dornik 2013; Drăgut, Eisank & Strasser 2011; Drăguţ, Tiede & Levick 2010; Eisank, Drăguț & Blaschke 2011; Gerçek, Toprak & Strobl 2011). However, data-driven SP optimisation techniques are only useful when detecting scales inherent to the dataset and are not necessarily appropriate when aiming to delineate specific landforms (Anders et al. 2011). Despite recent innovations in SP optimisation techniques, a data-driven SP optimisation methodology that detects SP levels at which predefined morphological conditions are met has not yet been developed.

1.2.2 Landform classification

Landform classification is necessary for assigning meaningful labels (e.g. categorical classes) to objects created using land surface segmentation procedures. Knowledge-based rulesets and supervised classification algorithms are both proven landform classification approaches. In geomorphometric studies, knowledge-based systems typically employ bodies of knowledge – represented as either expert knowledge or structured semantic models – to construct if-then ruleset classifiers based on multi-layered LSP datasets (Dehn, Gärtner & Dikau 2001; Eisank & Drăguţ 2010; Eisank, Drăguț & Blaschke 2011). Landform expert knowledge systems are primarily based on the academic background, experience and/or personal preferences of the programmer and have been shown to map geomorphological features accurately. Van Asselen & Seijmonsbergen (2006) delineated and classified eight landforms with the use of the MRS algorithm and expert knowledge-based rulesets, with fluvial terraces being classified with an accuracy of sixty-nine per cent (69%). Anders et al. (2011) employed MRS and a stratified rule-based approach to classify fluvial terraces and floodplains with an accuracy of sixty-seven per cent (67%). Despite these notable successes, expert system rulesets are often location- and data-specific with poor transferability potential (Eisank, Drăguț & Blaschke 2011). Structured semantic models aim to address some of the transferability limitations associated with expert knowledge systems by formalising existing knowledge and concepts (Dehn, Gärtner & Dikau 2001; Eisank, Drăguț & Blaschke 2011). Diverse geomorphometric definitions and incompatible nomenclature are just some of the challenges that hinder semantic model development (Gerçek, Toprak & Strobl 2011), and little progress towards establishing operational solutions has been made.

The adoption of supervised machine learning and deep learning algorithms for geomorphometric applications have received attention in recent literature, e.g. Florin (2019), Li et al. (2020), Li and Chen (2020) and Mithan et al. (2019). Swan (2017) evaluated five supervised machine learners – Mahalanobis distance, winner-take-all, normal Bayes, random forest (RF) and support vector machine (SVM) – to classify fluvial terraces and reported that, while SVM produced the best results when assessed qualitatively, the Bayesian and RF classifiers consistently produced better quantitative accuracies. A noteworthy difference between knowledge-based systems and supervised machine learners is that most machine learners are black-box and cannot be scrutinised by analysts (Rudin 2019), which limits expert involvement in the classification process, scientific understanding and the generation of new knowledge. White-box machine learners, on the other hand, are interpretable and can be studied by analysts to identify relationships among predictor

variables and their contributions to the resulting model. Decision trees (DTs) are white-box machine learners that are well-established in the fields of remote sensing (Laliberte, Fredrickson & Rango 2007) and geomorphometry (Bou Kheir et al. 2010) and can be used for both inductive and predictive purposes (Quinlan 1987). To date, the capability of DTs to produce interpretable and meaningful fluvial terrace rulesets remains unexamined.

1.3 SOIL SPECTROSCOPY ASSISTED GEOMORPHOMETRY

Digital soil mapping (DSM), a term often synonymously used with soil-landscape modelling, involves the spatial and temporal prediction of soil parameters and classes. The development of DSM methodologies has been primarily driven by advances in remote sensing, computing, geographical information systems (GIS) and statistical modelling (McBratney, Mendonca Santos & Minasny 2003; Minasny & Mcbratney 2015). The conceptual framework which underpins DSM is based on the premise that soil formation is driven by environmental factors, which was first formalised by Jenny (1941) into the well-known *corpt* model (climate, organisms, parent material, relief and time) and later adapted into the *scorpan* model (which includes age, position and soil attributes) by McBratney et al. (2003). In general, DSM approaches aim to empirically model relationships between known soil parameters and various environmental covariates that present soil forming factors.

McBratney et al. (2003) surveyed DSM literature and found that the bulk (80%) of DSM studies employed DEM-derived LSPs as environmental covariates, which indicates a high reliance on geomorphometrical data and the soil-landscape paradigm. It was found that many of these methodologies are area- and application-specific and have low transferability potential. Mulder (2013), consequently, evaluated the potential of diffuse reflectance spectroscopy (DRS) supported DSM and found that a combination of remotely sensed data, soil DRS and geostatistical analyses is effective for characterising soil mineral composition at regional scales. Rizzo et al. (2016), Vasat et al. (2017) and Wadoux (2015) illustrated the potential of integrating soil DRS data and LSPs to define soil mapping units, improve the prediction accuracy of soil organic carbon and establish correlations between soil spectra, soil parameters and landscape covariates.

Although the relationship between soils and landscape processes has been extensively exploited for soil mapping purposes, very few initiatives have taken advantage of the soil-landscape paradigm to map geomorphological features and improve our understanding of landscape processes. Wysocki et al. (2000) identify the understanding of landscape geomorphic histories

through investigation of soil-landscape dynamics as one of the primary goals of soil-geomorphologic studies. Accordingly, Masseroli et al. (2020) demonstrated the use of soil to decipher the geomorphological mechanisms responsible for the Late Holocene evolution of a typical alpine catchment. Rozanov et al. (2017), similarly, explored the soil-landscape paradigm as an indicator of landscape evolutionary processes by analysing the clay mineralogy of fluvial terrace soils to investigate fluvial sediment parent material sources. The authors concluded that the soil clay fractions reflect the underlying lithology and not that of the Drakensberg Escarpment, thus challenging the King (1963) theory that pediplanation was the main driving force behind the landscape evolution of the Lowveld. The potential of integrating DRS with LSP datasets for DSM applications and the utility of soil analytical techniques for landscape evolutionary studies illustrate a potential for DRS investigations of landform geomorphology that have not yet been addressed.

1.4 RESEARCH PROBLEM

Paleo-fluvial landforms make up a significant proportion of the southern African landscape (Dollar 1998; Partridge et al. 2010; Partridge & Maud 1987) and fluvial terraces play an intricate role in agriculture, ecosystem services and mining (Dollar 1998; Heritage, Moon & Large 2001; Moon & Heritage 2001; Paterson, Nell & Seabi 2011; Rozanov et al. 2017). Fluvial terrace research in South Africa has been predominantly driven by the search for diamondiferous gravels (Dollar 1998). However, recent efforts to characterise and map fluvial terraces in South Africa have dwindled, regardless of an immediate need for fine-scale and quantitative resource inventories to facilitate land use planning, global change adaptation and mitigation strategies (Holm et al. 2013; Lewis & Maslin 2005; Steffen et al. 2011). Moreover, geomorphological and soil mapping initiatives continue to rely on costly and time-consuming conventional techniques, and the 1:250 000 scale of the South African LTS database (Land Type Survey Staff 2006) – which represents the bulk of soil and geomorphological information available in the public domain – is too small for fine-scale environmental modelling.

Semi-automated geomorphometric methodologies can facilitate the production of large-scale geomorphological maps of fluvial terraces and provide much needed information for environmental models. Nevertheless, geomorphometric methodologies require expert knowledge or reference datasets to construct meaningful rulesets (models) that delineate landforms accurately. Given the severe lack of quantitative geomorphological data in South Africa, extensive work is needed to characterise fluvial terrace staircases to facilitate geomorphometric mapping initiatives.

Most of the existing geomorphometric methodologies for delineating fluvial terrace landforms are pixel-based (Clubb et al. 2017; Demoulin et al. 2007; Li et al. 2019; Stout & Belmont 2014; Swan 2017) and do not take the scale specificity of landforms or scale-dependency of LSPs into account (Drăgut, Eisank & Strasser 2011; Evans 2012). The object-based MRS algorithm has been shown to delineate morphological boundaries accurately (Van Niekerk 2010) and forms the basis of ESP 2, a popular data-driven scale optimisation tool that detects inherent scale patterns in spatial data (Drăguţ, Tiede & Levick 2010). However, available data-driven techniques are not always useful for detecting appropriate scale levels with which to delineate specific landforms (e.g. fluvial terraces), and analysts are frequently required to make subjective decisions when performing land surface segmentations (Anders, Seijmonsbergen & Bouten 2011). A data-driven SP optimisation methodology to evaluate MRS outputs against predefined morphological conditions may enable more accurate landform delineation.

Fluvial terrace classification methodologies are predominantly based on some form of expert knowledge (Clubb et al. 2017; Demoulin et al. 2007; Li et al. 2019; Stout & Belmont 2014) and are therefore subject to analyst bias. Despite the advent of geomorphometric machine learning applications that limit analyst influence, most geomorphologists still favour conventional geomorphological techniques (Oguchi 2019). Most machine learning classifiers are opaque (black-box) algorithms that exclude expert knowledge from the classification process. However, white-box machine learners – e.g. DTs – offer a way to bridge the gap between knowledge-based and machine learning classification approaches (Breiman et al. 1984; Quinlan 1986) as the resulting trees can be interpreted to identify the rules (variables and thresholds) that contributed to the thematic outputs. In addition, expert knowledge can be used to modify the rules and/or introduce new rules.

Although some research has been done on the development of expert knowledge and supervised machine learning geomorphometric approaches for fluvial terrace delineation, the differentiation between fluvial terrace levels has received little attention. Fluvial terrace surfaces located in denudated land surfaces, such as those found in the GLWC, are poorly preserved, and costly stratigraphic and physiographic interpretations are usually required to assess fluvial terrace continuity (Leopold, Wolman & Miller 1964). The close link between soils and land surface processes is well-known and extensively exploited in soil mapping (Jenny 1941; McBratney, Mendonca Santos & Minasny 2003). Yet, little work has been done on relating the spectral

characteristics of soils to their position in the landscape, and the usefulness of soil DRS for fluvial terrace mapping remains unexplored. Given these research gaps, the following questions arise:

1. *To what extent do fluvial terraces dominate the GLWC landscape, and how do the local climate and underlying geology influence their morphology and preservation?*
2. *How can data-driven object-based land surface segmentation scale optimisation approaches be employed to delineate objects that conform to pre-defined morphological conditions?*
3. *What is the value of supervised white-box machine learners for constructing fluvial terrace rulesets that are both accurate and comparable to existing expert-knowledge?*
4. *Is DRS a viable soil analytical technique with which to differentiate between various fluvial terrace levels, and are there any specific wavelengths that can be employed to develop transferable methodologies?*

1.5 RESEARCH AIM AND OBJECTIVES

The aim of this study was to develop methodologies for generating fine-scale object-based fluvial terrace maps. Geomorphometric techniques, object-based land surface segmentation approaches, machine learning algorithms and soil DRS analyses were combined and applied in the GLWC to map fluvial terraces to better understand the landscape evolution of the South African Lowveld.

To achieve this aim, the following objectives were set:

1. Review literature on fluvial geomorphology, geomorphometry, GEOBIA, machine learning and soil spectroscopy.
2. Characterise the geomorphology and soils of the GLWC fluvial terraces.
3. Develop a data-driven land surface segmentation scale optimisation approach that evaluates segmentation results against predefined morphological conditions.
4. Assess the viability of interpretable DT-derived fluvial terrace rulesets as a supervised object-based classification methodology.
5. Evaluate soil DRS analyses as a means with which to differentiate between fluvial terrace levels.

The above objectives relate directly to the four research questions formulated in Section 1.4. Objective 1 serves to provide the theoretical background for the dissertation and to support the investigation into all four research questions. Each of the following objectives relates to a single research question, e.g. Objective 2 aims to answer Research Question 1, etc.

1.6 RESEARCH METHODOLOGY AND DESIGN

The research was investigative and experimental in nature and considered fluvial terrace mapping and object-based geomorphometry from a synoptic point of view. It is inter-disciplinary as it draws from multiple fields; including geomorphology, geomorphometry, GEOBIA, machine learning, soil science and problem-solving theories. This dissertation utilises a broad range of technologies, methodologies and techniques within the quantitative research paradigm.

The research agenda is shown in Figure 1-1. Chapter 2 presents a brief treatise on object-based geomorphometry and soil spectroscopy in partial fulfilment of Objective 1. The following four chapters (Chapters 3 – 6) were prepared as journal articles and therefore summarise relevant literature in further fulfilment of Objective 1. Chapter 3, titled *The Great Letaba fluvial terraces: an investigation into the landscape and soil evolution of the South African Lowveld*, employed geomorphometric and conventional geomorphological mapping techniques to characterise fluvial terrace and planation surface morphology in an effort to better understand the landscape evolution of the South African Lowveld (Objective 2). Objective 3 is addressed in Chapter 4, in which MRS and scale optimisation techniques were investigated within the context of ill-structured problem-solving, and an ensemble SP optimisation approach was developed to delineate predefined morphological primitives. The chapter was published as a scientific article in the *Geomorphology* journal under the title *Object-based land surface segmentation scale optimisation: an ill-structured problem* (Louw & Van Niekerk 2019). Objective 4 necessitated the use of MRS and DT machine learners to produce transparent object-based rulesets with which to classify multiple levels of fluvial terrace staircases. Resultantly, the potential of using white-box techniques as an intermediate between expert systems and machine learning classifiers is illustrated in Chapter 5, titled *An object-based machine learning approach for classifying fluvial terraces*. Chapter 6, considering Objective 5, details a conceptual framework for comparing soil DRS-spectra to landscape characteristics is outlined, and specific wavelengths relating to clay mineralogical content are isolated and used in a RF machine learning approach to model fluvial terrace levels. It was published as a scientific paper in the *Geomorphology* journal under the title *Soil NIR-spectroscopy and object-based land surface segmentation for fluvial terrace level differentiation*

(Louw et al. 2021). Chapter 7 summarises the research findings, reviews the contributions to knowledge and makes recommendations for future research.

Given that standalone journal articles presented in a paper-based dissertation may be misconstrued as independent case studies, a logical progression and flow between the chapters needs to be emphasised to highlight the primary narrative of the dissertation. Chapters 3 – 6 follow the object-based geomorphometric workflow highlighted by the structure of this Chapter. First, Chapter 3 provides the geomorphological context and reference data needed to produce and evaluate the geomorphometric models developed in Chapters 5 and 6. Next, Chapter 4 relates to the ill-structured nature land surface segmentation scale parameter optimisation, a fundamental step in object-based geomorphometric approaches. Finally, Chapters 5 and 6 utilises machine learning algorithms to perform object-based classifications of fluvial terraces and to complete the final step of an object-based geomorphometric workflow.

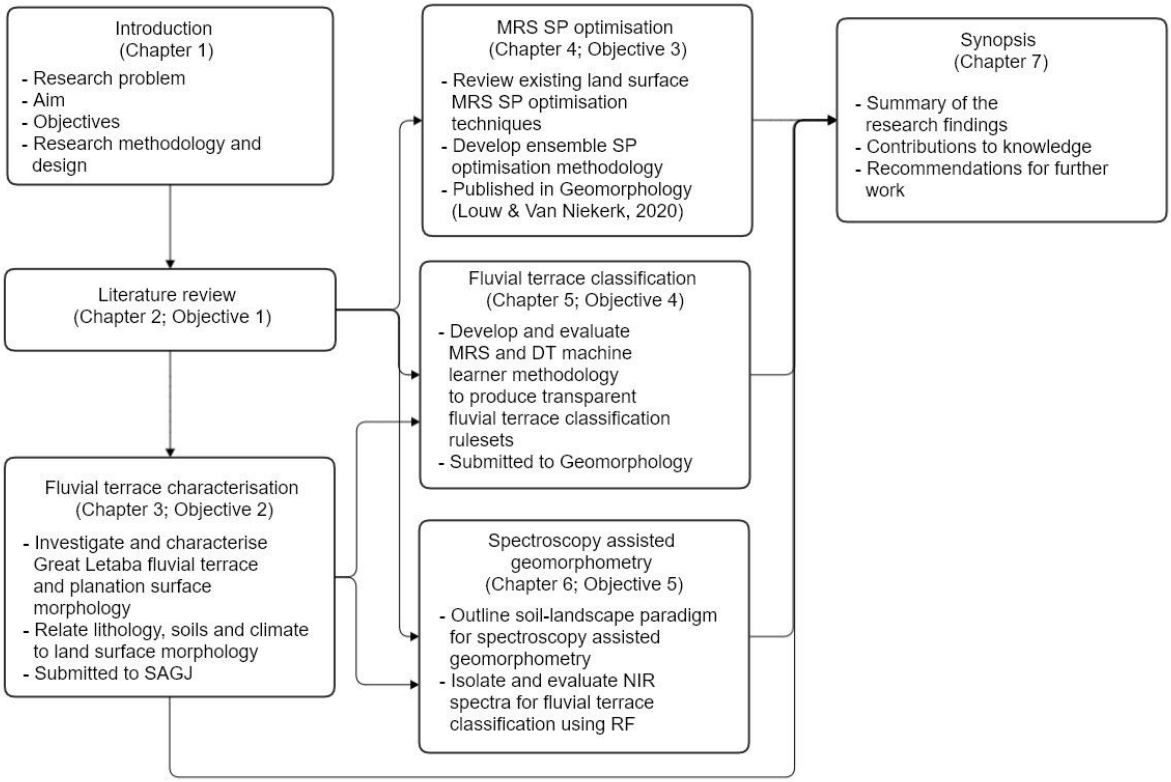


Figure 1-1 Research design and dissertation structure

CHAPTER 2: A BRIEF TREATISE ON OBJECT-BASED GEOMORPHOMETRY AND SOIL SPECTROSCOPY

This study draws from existing knowledge in fluvial geomorphology, defined as the study of "interactions between river channel forms and processes at a range of space and time scales" by Charlton (2008: 1) and the geomorphometric delineation and classification of landforms. Accordingly, the following subsections review literature related to object-orientated disciplines, object-based geomorphometry and soil spectroscopy. Moreover, basic terminology and fundamental principles related to geomorphometric inputs, land surface segmentation, landform classification and accuracy assessments are introduced. For a comprehensive overview of the southern African paleo-fluvial landscape, fluvial terrace geomorphology and the soil-landscape paradigm, the reader is referred to Chapters 1, 3, 5 and 6.

2.1 ON OBJECT-ORIENTATION

It is necessary to conceptually distinguish between metaphysical object-orientated ontology (OOO), object-orientated programming (OOP) and GEOBIA to avoid confusion regarding overlapping principles and terminology. Although there may be conceptual overlap between OOO, OOP and GEOBIA, any semantic intersections should be considered inconsequential for the purpose of this dissertation. In order to delineate clear conceptual boundaries, a short description of each is given. OOO is a philosophical school that rejects anthropocentrism and favours equal treatment of all objects, whether human or non-human. OOO gained popularity in the early 2000s when Graham Harman's 1999 doctoral dissertation called *Tool-Being: Elements in a Theory of Objects* was published. The two types of objects recognised by Harman, i.e. real and sensual, are considered to be significantly related, but not identical, to their properties. Real objects can only inter-relate through a sensual intermediary. Consequently, OOO is philosophically closer to aesthetics – a philosophical branch primarily concerned with the appreciation of nature and art – than to mathematics or natural science (Harman 2018).

In computer science, OOP employs objects as separate units of variables (properties) or computer code (methods). It is one of the most widely adopted programming structures and provides the foundation of most major programming languages, including: C++, C#, Java and Python. OOP is based on four basic concepts: encapsulation (isolation of objects), abstraction (partial isolation of object properties and methods), inheritance (functionality extension to mitigate redundant

repetition) and polymorphism (redefinition of object functionality by overriding or overloading). Benefits of OOP include faster execution times, clearer program structures, elimination of repetitive code and easier reuse of applications (Budd 2001; Stefik & Bobrow 1985).

GEOBIA is a sub-discipline of geographical information science (GISc) that bridges geographic information systems (GIS), remote sensing and image processing (Blaschke et al. 2014). Image processing software capable of partitioning remotely sensed data into objects first became commercially available around the year 2000. Subsequent research initiatives revealed clear advantages of GEOBIA over per-pixel classification approaches (Blaschke et al. 2014; Hay et al. 2006). Pixels, the smallest entity of remote sensing images, cannot be considered geographically representative as their size and shape are determined by the sensor, irrespective of the target real-world features (Blaschke & Strobl 2015; Fisher 1997; Hengl 2006). Low-resolution pixels often represent a mixture of real-world features' spectral properties, which can confuse classifiers and negatively affect performance. High-resolution imagery is less susceptible to spectral mixing as real-world objects are more likely to be represented by several smaller pixels with similar spectral properties. However, the performance of per-pixel classifications of high-resolution imagery are susceptible to intra-class variability, spatial autocorrelation and the modifiable aerial unit problem. Conceptually, objects created using image segmentation procedures are geographically representative as they are an amalgamation of pixels that delineate real-world features at predefined levels of homogeneity (Baatz & Schäpe 2000; Drăguţ et al. 2014).

GEOBIA applications are mainly applied to high-resolution imagery and are able to mitigate the adverse effects of intra-class variability, spatial autocorrelation and the modifiable aerial unit problem by calculating intra-object statistical derivatives (e.g. mean and standard deviation) of pixel values. Moreover, the delimitation of images into objects conceptually relate to how humans perceive and organise real-world features. As a result, classifiers implemented within an object-based paradigm are able to consider image parameters that were previously only exploited through analyst interpretations, i.e. texture, shape and context. While per-pixel moving kernels and local variance filters can be used to mimic feature texture and shape, object-based approaches are able to comprehensively incorporate texture, shape and context by taking advantage of interactions between spatially adjacent objects, object geometry and nuanced multi-scale interpretations of real-world patterns. As a result, object-based classifiers are able to facilitate more holistic classification methodologies that are more akin to human intuition (Blaschke et al. 2014; Hay et al. 2006).

2.2 OBJECT-BASED GEOMORPHOMETRY

2.2.1 Defining and categorising geomorphometric approaches

Geomorphology is the "science of scenery" or, more specifically, the study of landforms and landscapes (Fairbridge 1968). It is an empirical science that is primarily concerned with the classification of landforms; the geometry, topology and structure of landforms; and the understanding of processes that drive landform evolution and landform relationships with other processes, e.g. hydrologic, climatic and anthropogenic (Goudie 2004). Geomorphometry is a direct development of theoretical and applied geomorphology and deals with the quantitative analysis of land surfaces (Pike & Dikau 1995). It draws from research and development in mathematics, Earth sciences and computer science to model both landforms and land surfaces. The difference between geomorphometry – also called terrain analysis and terrain modelling – and quantitative geomorphology is that the latter studies the morphology of landforms without the aid of digital data (Pike, Evans & Hengl 2009). Evans (1972) recognised two geomorphometric approaches, namely: general and specific geomorphometry. General geomorphometry is not particularly relevant to this study; however, a brief definition is given and a basic overview of surface derivatives are detailed to illustrate an intersection between general and specific geomorphometry.

General geomorphometry analyses the continuous land surface that covers the entire Earth and other planetary bodies (Evans 2012) using – among other approaches – fractal analysis and surface derivatives (Evans 1980; Evans 1972; Goudie 2004). The basic principles of general geomorphometry are based on the relationship between elevation (altitude) and land surface processes. Hypsometry (the measurement of land elevation above sea level) has been extensively used to characterise land surfaces with the use of integrals, statistical derivatives (e.g. mean and standard deviation) and dimensional indices. Slope gradient, a derivative of elevation, is crucial for the understanding of hydrologic and fluvial geomorphology as it can be used to infer mass and energy movements across surfaces (e.g. fluvial systems or landslides). When combined with slope, aspect is particularly valuable to model local climates as a function of sun and wind exposure. Further derivatives of slope and aspect can be used to construct curvature types (e.g. plan, profile, etc.) and surface roughness, both of which are useful when examining surface runoff rates (Goudie 2004). Derivatives of elevation are of general importance to geomorphometrists and, as such, are also central to specific geomorphometric approaches (Evans 1987).

Specific geomorphometry deals with the delimitation and classification of individual landforms and the multivariate characterisation of their properties. A prerequisite for specific geomorphometry is the complete delimitation of land surface features according to predefined definitions that incorporate (among other morphometric properties) size, shape and scale (Jasiewicz & Stepinski 2013; Minár et al. 2018; Minár & Evans 2008; Pike, Evans & Hengl 2009). The delimitation of discrete landform elements and landforms is therefore fundamental to specific geomorphometric approaches. Landform elements (also called elementary forms or morphometric primitives) are hierarchical sub-components of landforms that are bounded by morphological discontinuities and are uniform (homogenous) in morphology. They represent the smallest and simplest landscape feature – identifiable at specific scale levels or spatial resolutions – that cannot be subdivided into meaningful features.

Landforms (also called landform components or terrain units), an integral part of land surfaces, are formed through natural or anthropocentric processes, e.g. fluvial terraces or canals. They consist of sequential combinations of landform elements and represent a second level of landscape complexity. Land systems comprise landforms and offer a third level of landscape complexity (Dikau 1992; Minár & Evans 2008). The delimitation of continuous land surfaces into specific landform unit divisions (landform elements, landforms or land systems) is typically achieved through per-pixel classification and post-classification aggregation approaches (Sections 2.2.3 and 2.2.4). Despite the widespread adoption of specific geomorphometric principles, the modelling of certain landform types still requires the use of general geomorphometric approaches. Examples include extensively pedimented or fluvial plains where the standard deviation of elevation and slope is very low and profile curvature is near zero (Evans 2012). Conversely, the widespread occurrence of hillslope and drainage network landforms gives a general importance to their specific geomorphometry (Evans 1987).

Drawing from both general and specific geomorphometry principles, Drăguţ & Eisank (2011) introduced “discrete geomorphometry” for the multi-scale, object-based delimitation of landform elements and landforms. Discrete geomorphometry pertains specifically to the data-driven delimitation of discrete land surface features defined exclusively by scale (object homogeneity) criteria, calculated as a function of LSP data (Drăguţ & Eisank 2011; Drăguţ, Tiede & Levick 2010; Louw & Van Niekerk 2019). While specific geomorphometric approaches delimit land surfaces through per-pixel classifications, followed by post-classification aggregation, discrete geomorphometry approaches employ land surface segmentation techniques to create homogenous

objects that delimit morphological discontinuities. By partitioning land surfaces into spatially intact landforms prior to classification, discrete geomorphometry enables the transition from the field-model (i.e. per-pixel classifications) to the object model (Brändli 1996) and from general geomorphometry to specific geomorphometry (Drăguţ & Eisank 2011; Evans 1972; Minár & Evans 2008).

Two main land surface segmentation strategies are presented by Drăguţ & Eisank (2011) as viable discrete geomorphometric approaches. The first considers the delimitation of "elementary forms" through a specific geomorphometric land surface classification approach developed by Minár & Evans (2008). While the approach does ensure a basic level of object homogeneity, it is conceptually a scale-independent approach that requires further object aggregation to facilitate multi-scale analyses (the concept and delimitation of elementary forms are detailed in Section 2.2.3). The second land surface segmentation strategy involves employing image segmentation algorithms to delimit homogenous land surface features. This approach offers four key benefits (Eisank, Drăguț & Blaschke 2011). First, the objects derived are geomorphometric units (as opposed to non-geomorphometrically representative pixels) that relate better to real-world landforms and human intuition. Second, image segmentation algorithms (e.g. MRS) can facilitate multi-scale hierarchical delimitations of land surface features. Third, landform spatial characteristics (e.g. topology and context) can be utilised for both land surface segmentation and landform classification purposes. Finally, GEOBIA approaches are not limited to specific data types or formats (e.g. raster and vector). Discrete geomorphometric land surface segmentation approaches are elaborated on in Section 2.2.3 and in Chapter 4.

2.2.2 Geomorphometric inputs (DEMs and its derivatives)

Digital elevation models (DEMs) are the most commonly employed source of terrain data for geomorphometric analyses (Pike, Evans & Hengl 2009). DEMs are square-grid (raster) representations of land surface elevation, structured in a Cartesian space. Point (e.g. triangular arrays) and line (e.g. contours) representations of elevation data are not considered DEMs, as the native data-structure of DEMs dictate the continuous representation of elevation values at a given pixel (cell) resolution. Resultant disadvantages of DEMs include the representation of elevation data without reference to land surface morphology, extensive data storage requirements and the simultaneous under- and over-sampling of complex heterogeneous landscapes. Nevertheless, the popularity of DEMs can be attributed to its simple data-structure. The raster format of DEMs

enables automatic derivation of most technical raster properties from only the pixel size (Hengl 2006) and the simple calculation of morphologically meaningful derivatives such as land surface parameters (LSPs) and land surface objects (Pike, Evans & Hengl 2009). LSPs are derived with the use of geomorphometric algorithms that essentially apply mathematical operations to DEMs. Geomorphometric algorithms typically employ neighbourhood operations, which are regular matrix filter windows, to calculate values for each pixel, based on the values of neighbouring pixels. Primary LSPs (e.g. slope) are derived directly from DEMs, whereas secondary LSPs (e.g. soil wetness indices) are derivatives of primary LSPs and require additional processing steps (Wilson & Gallant 2000). In addition to LSPs and landform objects, DEMs can be used to derive cross-profiles at user-defined intervals and sampling schemes (Olaya & Conrad 2009). The use of cross-profiles for landscape mapping, in particular in fluvial geomorphology, has been well-established for over a century (Leopold, Wolman & Miller 1964). For additional technical details on geomorphometric algorithms the reader is referred to Pike, Evans & Hengl (2009).

Minár & Evans (2008) state that the accurate delimitation of specific landform units is dependent on appropriate input data. Input data are often selected through qualitative assessments of candidate DEMs and LSPs (Eisank, Smith & Hillier 2014) or according to semantic definitions of target landforms (Clubb et al. 2017). Nevertheless, the suitability of DEMs employed for land surface segmentation procedures is often not questioned. Mashimbye, De Clercq & Van Niekerk (2014) evaluated five DEMs as MRS input to determine whether they delimit landform components accurately. A 5 m GEOEYE DEM (derived from GeoEye stereo-images) and the 5 m Stellenbosch University DEM (SUDEM) L2 produced the most internally homogenous landform components and delimited morphological boundaries most accurately. Segmentations performed using 90 m shuttle radar topography mission (SRTM) and 5 m SUDEM L1 also produced internally homogenous landform components, but delimited morphological discontinuities less accurately than the GEOEYE DEM and SUDEM L2 products. The 30 m resolution advanced spaceborne thermal emission and reflection radiometer global DEM (ASTER GDEM 2) often failed to delimit morphological discontinuities. The authors note that high-resolution DEMs, such as GEOEYE DEM and SUDEM L2, are needed to successfully delimit meaningful landform components.

2.2.3 Land surface segmentation

Land surface segmentation approaches aim to delineate specific landform divisions (landform elements, landforms, etc.) – genetically defined and bordered by "natural" morphological discontinuities – by employing well-defined segmentation algorithms that minimise analyst subjectivity. The presence of morphological boundaries in the landscape can be attributed to morphogenetic land surface processes that are largely driven by gravity. The delimited landform units should have stronger internal than external associations. In other words, delimited landform units must be internally homogenous and distinct from neighbouring landforms (Minár & Evans 2008).

Minár & Evans (2008) differentiate between graph-based and classification-based land surface segmentation approaches. Graph-based land surface segmentation algorithms are primarily concerned with the identification of landform boundaries and less with the properties of landform interiors. By adapting morphological mapping practices (Savigear 1965; Waters 1958) to consider form instead of process, graph-based segmentation procedures perceive the landscape as a mosaic of planes (also called facets or segments) that are joined at discontinuities. Dalrymple, Blong & Conacher (1968) introduced a hypothetical land surface model that delimits nine landform units from slope profiles based on a catena sequence. The model deviates from morphological mapping practices by considering process transitions across discontinuities, along with position, slope, profile curvature, microforms and other land surface properties such as lithology, soils and flora. Lastoczkin (1987) introduced a conceptual graph-based approach that utilises structural lines and characteristic points to construct elementary surfaces. Structural lines represent discontinuity types that are classified based on the profile shapes (linear, concave and convex) of neighbouring surfaces. Characteristic points represent land surface features such as peaks and structural line end-points or intersections.

Unlike graph-based approaches, classification-based land surface segmentation approaches consider the interior properties of landforms and, thereafter, derive discontinuity boundaries (Minár & Evans 2008). First generation classification approaches interpret profile and plan curvature signs to identify landform divisions, which can be used to understand various gravity-related land surface processes (Troeh 1965). More complex procedures involve, for example, the classification of landform units with the use of map overlays of curvature isolines and interpretations of morphometric criteria such as elevation, slope and aspect (Minár & Evans 2008).

The homogeneity of the classified landform units can be defined as a function of scale by incorporating morphometric thresholds (Bolongaro-Crevenna et al. 2005). The resultant maps are consequently scale-dependent (Reuter, Wendroth & Kersebaum 2006).

Following the specific geomorphometric principals of Evans (1972), Minár & Evans (2008: 244) introduced a signature approach to classify predefined elementary forms based on statistical representations of morphometric properties. The authors define elementary forms as "landform elements with a constant value of altitude, or of two or more readily interpretable morphometric variables, bounded by lines of discontinuity". Elementary forms are indivisible landform units that maximise internal homogeneity and external heterogeneity and are classified into typological systems based on specific geometry types defined by generalised polynomial functions. The system also incorporates various discontinuity types, connections and transformations. Elementary forms are inherently scale-independent and are not suitable for multi-scale analyses. Jasiewicz & Stepinski (2013) introduced the geomorphon pattern recognition algorithm that assigns one of 498 unique, predefined geomorphons (simple patterns composed of three components that represent a specific terrain morphology) to each raster pixel and, thereafter, generalises the geomorphons into a simplified set of 10 landform elements (peak, ridge, shoulder, spur, slope, hollow, foot-slope, valley, pit and flat). The algorithm incorporates scale by making use of the line-of-sight principle that self-adapts to the local topography. While landform units derived with the use signature classification approaches have been shown to facilitate soil mapping strategies (Flynn et al. 2019) and landslides susceptibility maps (Luo & Liu 2018), systems for aggregating elementary forms and geomorphons into specific landforms (e.g. fluvial terraces) do not yet exist.

The most frequently employed classification approach to land surface segmentation involves knowledge-based or supervised categorical classifications of individual pixels, followed by the aggregation of pixels according to the thematic classes to which the pixels were assigned (Drăguţ & Eisank 2011; Pike, Evans & Hengl 2009). Several shortcomings are associated with the categorical classification of pixels (Drăguţ & Eisank 2011). First, per-pixel classifiers tend to treat pixels as spatially independent. As a result, they typically produce spatially scattered thematic classifications (also called the "salt-and-pepper" effect) and are consequently prone to delimiting objects that do not represent homogenous landforms (Blaschke 2010; Gao et al. 2011). Second, post-classification pixel aggregation is needed to delimit landform boundaries, which often do not accurately reflect morphological discontinuities (Drăguţ & Eisank 2011). A third limitation of per-pixel categorical classifications is that they are incapable of considering certain spatial landform

attributes (e.g. topology and context) as they merely apply statistical rules on a pixel-by-pixel basis without consideration of relative location.

Object-based segmentation approaches, applied within the discrete geomorphometry paradigm, evade the limitations associated with categorical per-pixel classifiers by clustering (regionalising) pixels into objects prior to performing thematic classifications. In so doing, homogenous and coherent landform units that are less prone to the "salt-and-pepper" effect are delimited (Blaschke 2010; Drăguţ & Eisank 2011; Liu & Xia 2010). Analyst control over the internal homogeneity of the objects enables the incorporation of scale in the land surface segmentation process and, given that the exhibition of morphometric discontinuities are a function of scale (Minár & Evans 2008), morphological discontinuities are more accurately delineated. Furthermore, the ability to optimise the internal homogeneity of the resultant objects enables analyst control over the resultant objects' size and eliminates the need for object aggregation. Factors that hinder object-based segmentation algorithms to accurately delimit real-world landforms include the selection of appropriate segmentation algorithms, setting suitable algorithm parameters, deciding on the best input data (see Section 2.2.2) and the interpretation of results (Minár & Evans 2008).

Van Niekerk (2010) statistically evaluated three segmentation algorithms to determine which performs best when considering the homogeneity of the objects created and whether the resultant objects' boundaries follow morphological discontinuities. First, the popular iterative self-organising data analysis technique algorithm (ISODATA; see Section 2.2.4.1) was selected as it is a clustering technique capable of segmenting large multi-layer datasets into homogenous areas (Hall & Khanna 1977). The second evaluated algorithm, namely automated land-components mapper (ALCoM), is a graph-based approach that statistically identifies natural breaks in slope gradient (Van Niekerk & Schloms 2002). Finally, MRS was selected as it regionalises pixels into homogenous objects according to predefined homogeneity criteria (Baatz & Schäpe 2000). The results showed that ISODATA managed to create relatively homogenous objects but generally failed to produce land components that follow morphological discontinuities. Compared to ISODATA, ALCoM produced objects with greater inter-heterogeneity and land components that are more meaningful. However, of the three algorithms, MRS produced the most useful objects as it was the most sensitive to morphological discontinuities and produced the most internally homogeneous land components. Van Niekerk (2010) further states that the analysts' level of control over the outcome (due to configurable parameters such as shape, compactness and scale) sets MRS apart from other segmentation algorithms.

Van Asselen & Seijmonsbergen (2006) and Drăguţ & Blaschke (2006) introduced MRS for land surface segmentation by demonstrating an expert-driven multi-scale classification of eight landforms and an automated multi-scale classification of nine landform elements. Lately, MRS has increasingly been applied to delimit land surfaces into discrete objects. The rising popularity of MRS can be attributed to its ability to hierarchically delineate scale-dependent morphological discontinuities (Anders, Seijmonsbergen & Bouten 2015; Anders, Seijmonsbergen & Bouten 2011; Dornik, Drăguţ & Urdea 2017; Drăguţ et al. 2014; Drăguţ & Dornik 2013; Eisank, Smith & Hillier 2014; Mashimbye, De Clercq & Van Niekerk 2014). Minár & Evans (2008) warn that intricate combinations of segmentation algorithm input parameters can complicate geomorphological interpretations of delimited landforms and their association with soils or land cover. Yet, MRS landform delimitation approaches are typically concerned with only a single parameter, i.e. the scale parameter (SP). All other parameters, i.e. shape and compactness, are excluded from the segmentation process to ensure that only morphological homogeneity is considered in the segmentation process (Drăguţ & Eisank 2012; Eisank, Smith & Hillier 2014). The development of models investigating the influence of scale on land surface calculations has been identified as a critical research area in geomorphology, hydrology, pedology and geomorphometry. As a result, scale analysis has become a crucial component of terrain analysis (Drăguţ & Eisank 2011).

Drăguţ et al. (2010) and Drăguţ et al. (2014) adapted the local variance (LV) principles introduced by Woodcock & Strahler (1987) to develop a data-driven technique called the estimation of scale parameter 2 (ESP), which investigates the influence of scale on land surface objects to detect "signature" or "characteristic" scales. LV takes advantage of the spatial autocorrelation present in high-resolution images to model the relationship between the size of the real-world target features and the landform units delimited. In a per-pixel paradigm, LV uses a 3x3 standard deviation moving window to calculate LV at sequentially upscaled pixel resolutions. ESP, on the other hand, calculates LV as a function of object interior standard deviation and plots LV against sequentially increased MRS SP values. A conceptual concern with the data-driven principles that support ESP relates to the discrepancy between the "characteristic" scales detected and the scales at which real-world target landforms are accurately represented (Anders, Seijmonsbergen & Bouten 2011). This discrepancy can be attributed to "blind" segmentation approaches that fail to consider the morphometric properties of the target landform units. As a result, the relationship between the delimited landform units and genesis geomorphic processes is lost (Brown, Lusch & Duda 1998;

Minár & Evans 2008). The reader is referred to Minár & Evans (2008) and Drăguţ & Eisank (2011) for supplementary content on land surface segmentation and the relevance of scale for landform delimitation. Moreover, Chapter 4 expands on the discrete geomorphometry, MRS and SP optimisation concepts introduced in this section.

2.2.4 Landform classification

2.2.4.1 Unsupervised classification

Unsupervised classifiers require only minimal interaction from analysts to identify and group similar pixels or objects (Campbell & Wynne 2011). K-means clustering, for example, groups data based on sets of equal variances by minimising an inertia criterion based on the intra-cluster sum-of-squares. In other words, it creates clusters of pixels or objects that are internally coherent. First, it randomly selects centroids for a user-defined number of clusters. Thereafter, it assigns each pixel or object to the nearest centroid and plots a new position for each centroid based on the mean of the samples assigned to each centroid. The algorithm iteratively calculates the difference between the old centroids and new centroids and plots new centroid positions until the difference calculated is smaller than a threshold (Pedregosa, Weiss & Brucher 2011).

ISODATA is operationally the same as k-means clustering, except that it can be implemented without predefined cluster numbers, and additional processing steps enable the merging and splitting of clusters. Clusters are merged when the distance between the centres of two clusters are less than a certain threshold or if the number of data points within a cluster is lower than a specified minimum. Clusters are split when the standard deviation of a cluster is more than a specified value or if the number of data points within a cluster is two times more than the specified minimum (Tou & Gonzalez 1994). Another distinguishing feature of ISODATA is its ability to incorporate training data as a supervised approach to estimate cluster means before initialising the first operational step (Campbell & Wynne 2011).

The mathematical principles that define k-means clustering and ISODATA enable a broad range of applications, particularly in geospatial analyses. Specialised unsupervised approaches, on the other hand, can be based on statistical procedures programmed to classify specific features. For instance, Clubb et al. (2017) developed an automated unsupervised approach to classify floodplains and fluvial terraces by deriving topographic thresholds with the use of quantile-quantile plots (see Section 5.1).

2.2.4.2 Knowledge-based systems and semantic models

Knowledge-based, often referred to as rule-based, classifiers are based on the principles of expert systems that consist of a knowledge base, an inference engine and a user interface. The knowledge base contains facts and rules used by the inference engine to infer logical conclusions about the data used as output by the user interface. The rules in the knowledge base are often defined by an expert or through experimentation (Skidmore et al. 1996). In landform classification studies, rules defined by experts are frequently based on an individual's experience and intuition, which results in study area specific and scale-dependent rulesets. Moreover, incompatible landform definitions employed by analysts of different disciplines result in ontological ambiguity and require frequent revisions (Dehn, Gärtner & Dikau 2001). Hierarchical semantic modelling integrates structured knowledge to overcome limitations associated with transferability and landform scale specificity (Eisank, Drăguț & Blaschke 2011). In a GIS environment, semantic modelling essentially links qualitative landform ontologies with quantitative computer-based modelling of DEMs by translating landform definitions into applicable topographic (LSP) thresholds.

A central objective of geomorphometric semantic modelling is to define ontologically correct definitions of specific landforms that can be agreed upon and used by all disciplines, thereby enabling the exchange of ideas and data between disciplines (Dehn, Gärtner & Dikau 2001). Accordingly, Dehn, Ga & Dikau (2001) introduced a semantic framework that considers landform geometry as a fundamental property. This framework can be extended by landform topology and semantic definitions. They illustrated the approach using hillslopes as a case study.

Eisank, Drăguț & Blaschke (2011) presented a conceptual framework for a generic object-based semantic approach to landform classification, demonstrated by the glacial landforms. The framework introduces a four-step approach. First, interoperable glacial landform concepts are defined. Next, multi-scale characteristic landscape patterns are derived with the use of MRS and ESP. Third, landform concepts are formalised using semantic modelling. Lastly, quantitative and relational rules are extracted to facilitate hierarchical classification. The authors note that their semantic framework captures landform semantic signatures that encapsulate morphometric, morphologic and contextual information. Gerçek, Toprak & Strobl (2011) performed landform classifications by coupling MRS and ESP delimited multi-scale landform units with the semantic import model. The semantic import model enables fuzzy landform classifications by defining conceptual landform elements using imprecise semantics (Dehn, Gärtner & Dikau 2001;

MacMillan et al. 2000). The methodology operates within the general geomorphometry paradigm as it incorporates landform geometry by employing “relative terrain position” and “terrain network” within a basic morphometric LSP context. The landform classes defined were intuitively understandable by analysts, relevant to various disciplines and remained reasonably consistent when considering the level of ambiguity inherent to landforms.

Building and evaluating semantic models are an exhaustive and laborious process (Gerçek, Toprak & Strobl 2011). Consequently, landform classification rulesets are often constructed using inductive statistical analyses to derive topographic thresholds. Van Asselen & Seijmonsbergen (2006), for example, utilised zonal statistics to derive topographic thresholds to define eight landforms, and Anders et al. (2011) used a stratified approach to semi-automatically extract topographic thresholds of nine landforms. Both studies achieved reasonable success. While inductive approaches have been shown to be useful for extracting topographic thresholds with minimal analyst involvement, the semantic knowledge gained using primitive statistical techniques (Van Asselen & Seijmonsbergen 2006; Anders et al. 2011) are often limited.

2.2.4.3 Supervised classification

In geospatial studies, supervised classification can be defined as the process of using training data (descriptions) of known features to classify unknown pixels or objects. Classifiers compare each target pixel or object to training data and either classify it as one of the features in the training set or as unknown (Campbell & Wynne 2011; Lillesand, Kiefer & Chipman 2008). The collection of training data typically involves analyst interpretations of maps and aerial photographs of the study area. Field reconnaissance is essential for becoming familiar with study areas and specific target features or phenomena and to make in situ observations. The objective of training data collection is to spatially define areas representative of the spectral or morphometric variation that is contained by each thematic class or informational region (Campbell & Wynne 2011).

While a multitude of different supervised strategies have been developed, only a selection of relevant supervised classifiers is described here. For a comprehensive overview of supervised classification and machine learning, the reader is referred to Maxwell et al. (2018) and Lary et al. (2016).

Partial least-squares regression (PLSR) uses training data to construct a multivariate model that is capable of analysing noisy, collinear and incomplete data. In essence, PLSR establishes linear

regressions by projecting multiple dependent and independent variables into a new space and optimising the variables for maximum covariance using the least number of dimensions (Wold & Sjostrom 2001). K-nearest neighbour (KNN), on the other hand, employs training data and a distance-based approach to measure a k-number (typically a small integer) of nearest neighbours of each pixel in a multispectral data space. The target pixel is then allocated to the class that most neighbours belong to (Campbell & Wynne 2011).

Artificial neural networks (ANNs) differ from the aforementioned approaches as they simulate human learning process by establishing and reinforcing pathways between dependent and independent datasets. ANNs perform best when applied to large multivariate datasets with a relatively small number of classes in relation to the number of input layers. Training data are needed to facilitate pathways between the input data and classes defined by the analyst and to establish the weights of input layers in hidden ANN layers, which are then applied to classify unknown data (Campbell & Wynne 2011). DTs and RFs are non-distance-based classification algorithms that derive rulesets by recursively partitioning training datasets to define binary divisions (nodes representing spectral or topographic thresholds) of the dependent variables and assign independent variables to classes (Breiman et al. 1984; Campbell & Wynne 2011). The key difference between DT and RF is that DTs produce a single "tree" or ruleset and RF produces an ensemble of decision trees (Breiman 2001). DTs and RF classifiers are detailed in Chapters 5 and 6, respectively.

Supervised classification algorithms can be categorised as either white-box or black-box. The term "white-box" is often ambiguously used to refer to both "interpretable" classifiers and "explainable" models. This study will henceforth use "white-box" to refer to interpretable machine learners (e.g. DTs) and "explainable models" to refer to ad-hoc algorithms that decipher black-box machine learners. Black-box approaches are characterised by complex mathematical procedures (e.g. PLSR, ANN and RF) or representations of distance or space (e.g. kNN) that are unrecognisable or incomprehensible to experts in practical scenarios. Recently, an upsurge in the use of white-box algorithms in industries such as healthcare, military and finances have been driven by a need for, and often the mandatory use of, models that can be scrutinised to validate and defend classification (decision) outcomes (Rudin 2019). Yet, the adoption of machine learning approaches in remote sensing, geomorphometry and soil-landscape modelling have been largely limited to black-box algorithms (Heung et al. 2016; Lary et al. 2016; Li & Chen 2020; Maxwell et al. 2018; Swan 2017).

The extensive prevalence of black-box approaches may be attributed to the tendency of experts to apply machine learning without prioritising insight into the underlying mathematical principles, as they only require an understandable output (Loyola-Gonzalez 2019). However, the reluctance of experts to adopt machine learning approaches is often fuelled by an instinctive distrust of black-box machine learners, which can be attributed to a disjoint between the knowledge transfer mechanisms used by experts (e.g. logical reasoning) and the somewhat abstract approaches of black-box algorithms (Loyola-Gonzalez 2019). This reluctance is often exacerbated by the notion that black-box approaches perform better than white-box approaches, despite there being no discernible performance difference between the two (Letham, Rudin & McCormick 2012; Loyola-Gonzalez 2019; Rudin 2019). Given that the adoption of machine learning in geomorphometry is driven by environmental decision-making support and its facilitation of expert understanding of the land surface phenomena being investigated, it seems logical that white-box approaches should be prioritised. The use of supervised classifiers for landform classification and soil-landscape modelling are explored further in Chapter 5 and Chapter 6, respectively.

2.2.4.4 Accuracy assessment

Statistical measures are needed to assess the success and representative accuracy of classifier outputs and to optimise hyperparameters. While a wide range of statistical techniques has been developed for this purpose, this section overviews techniques used to assess unsupervised and supervised classification results. The selection of techniques was limited to those available in popular scientific Scikit-learn machine learning library (Pedregosa, Weiss & Brucher 2011). Many of the techniques described here are also applicable to knowledge-based classification approaches. The reader is referred to Campbell & Wynne (2011) for definitions of basic accuracy assessment terminology (e.g. accuracy, precision and significance) and information about different sources and characteristics of classification error.

Scikit-Learn employs both classifier scoring methods and scoring parameters (Pedregosa, Weiss & Brucher 2011). Classifier scoring methods are computed by the classifier itself and typically only provide basic accuracy estimates such as model fit and out-of-bag scores (average prediction error of each training data point). Scoring parameters, on the other hand, use model-evaluation tools (e.g. cross-validations) and metric functions to implement scoring strategies. Cross-validation is a resampling procedure that divides reference data into a k-number of sets and iteratively uses one as a testing set to evaluate the accuracy of the classifier that is trained on the remaining (k-1) sets. Metric functions are combined with cross-validation schemes to assess

classification error for specific purposes. Two of the simplest examples of metric functions are validation and training scores. Validation scores quantify how accurately each of the testing datasets was predicted and training scores quantify how well the model is generalised or fitted to the training data. A high training score and low validation score indicate that the classifier is well fitted to the data, but fails to construct a generalised model, which is a sign of overfitting and low transferability potential. Generalisation error can be described as a function of bias, variance and noise. Bias is quantified by calculating the mean error of different training datasets, and variance indicates how sensitive a classifier is to various training datasets. Bias and variance are inherent properties of classifiers and noise is a property of data. Classifier selection and hyperparameter optimisation are therefore crucial to minimise bias and variation.

Validation curve tools are useful for assessing classifier generalisation by evaluating whether a model under- or over-fits given specific hyperparameters. Such tools iterate a range of values for a single hyperparameter and plot the resultant training and validation scores on a graph. If both scores are low, then the classifier is under-fitting and not representing the training data accurately. Overfitting, on the other hand, is prevalent when the training score is high and the validation score is low. A general rule-of-thumb is to select hyperparameters at which both scores approximate each other. However, hyperparameter optimisation using validation curves introduces bias and undermines the validity of the validation curves itself. This supports the use of grid search functions for optimising hyperparameters.

Grid search functions iterate k-fold cross-validation procedures for each value within a predefined set of multi-hyperparameter values. A grid search cross-validation procedure evaluating two hyperparameters using two candidate values for each and a 10-fold cross-validator will produce an array with 40 validation scores. Grid search cross-validation algorithms consider every combination of selected hyperparameters by using a predefined set of candidate values to determine the optimal hyperparameter set based on the highest validation score achieved.

While validation and training scores are useful for assessing classifier accuracy and optimising hyperparameters, they only give an overview of the validity of trained classifier models and do not provide comprehensive information regarding the classification accuracy of each class. A confusion matrix provides both overall classification accuracies and information about individual class misclassification rates. The matrix compares reference data with single classification outputs to determine how each of the reference data classes are represented in the classification. The matrix

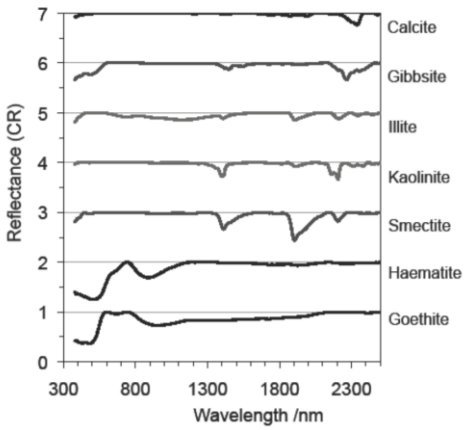
tabulates the number of times the reference points of each class is assigned to specific classes by the classifier. Both the error of omission (or false negatives) and the error of commission (or false positives) can then be calculated for every possible combination of reference and classified classes. Subsequently, user's and producer's accuracies can be computed. User's accuracy (calculated by subtracting the commission error from 100%) takes the point of view of the end-user to quantify how often the classifier output accurately represents the reference data. Producer's accuracy (calculated by subtracting the omission error from 100%) takes the view of the analyst to indicate how often the reference data is accurately represented by the classification output (Campbell & Wynne 2011; Pedregosa, Weiss & Brucher 2011). For examples of confusion matrices, the reader is referred to Section 5.3.2.

2.3 SOIL SPECTROSCOPY

Soil is a natural resource critical for ecosystem services. It is fundamental to food production, water regulation and the filtering of metals, nutrients and contaminants. Moreover, it is a biological habitat and an important carbon sink for the mitigation of climate change. Soils are complex aggregates of inorganic matter, organic matter, water and air (Stenberg et al. 2010). Analytical techniques capable of quantifying soil properties are crucial to facilitate soil mapping and environmental modelling strategies.

Conventional soil analysis is costly and time-consuming as it relies on extensive field surveys and laboratory analyses. DRS has been suggested to reduce the time spent in the field and laboratory (Paterson et al. 2015). DRS measures the spectral response of matter that is generated when matter interacts with the electromagnetic spectrum. Electromagnetic radiation containing all frequencies (either directed from the sun or a light source) forces the molecules of matter to bend and stretch as they vibrate. Subsequently, radiation is absorbed by the matter and a specific energy quantum (smallest quantity of radiant energy) is reflected. Given that the energy quantum is inversely proportional to wavelength, a characteristic absorption spectrum is created (Miller 2001). The shape of the absorption spectrum is dictated by soil properties and environmental factors, thereby enabling the measurement of various molecules that may contain similar bonds (Stenberg et al. 2010). Mid-IR (2 500 – 25 000 nm), NIR (780 – 2 500 nm) and visible (400 – 780 nm) spectra can be combined with multivariate statistics to estimate various soil properties. Molecular interactions of soil components occur predominantly in the mid-IR, whereas overtones and combinations of fundamental mid-IR vibrations dominate the NIR. Subsequently, the NIR exhibits predominantly weak superimposed vibrations that manifest as broad absorption characteristics.

By analysing spectral signatures and the presence or absence of certain absorption features, Stoner & Baumgardner (1981) produced five characteristic soil spectral reflectance signatures. More recently, studies uncovered strong correlations between spectral characteristics and soil properties (Stenberg et al. 2010) such as common occurring soil minerals (Figure 2-1). For example, certain absorptions in the visible range are characteristic of dark organic matter that may contain chromophores and iron minerals such as haematite and goethite (Mortimore et al. 2004; Sherman & Waite 1985). NIR absorption characteristics result from overtones caused by the presence of OH, SO_4 and CO_3 molecules and combination features associated with water and CO_2 (Clark 1999). Clay minerals such as Illite, Kaolinite and Smectite cause metal-OH molecule bending and OH molecule stretching in the visible and NIR range (Viscarra Rossel, McGlynn & McBratney 2006). Many other soil properties have also been successfully linked to characteristics absorption features, some of which include: texture, salinity, carbonates, non-photosynthetic vegetation and lichens (Mulder 2013).



Source: Stenberg et al. (2010)

Figure 2-1 Spectra (continuum-removed) illustrating the absorption characteristics of various soil minerals

DRS of soil is complex and non-specific with regards to soil properties. Interferences related to overlapping spectra of soil properties, weak overtones and molecular vibration combinations obscure distinct relationships between specific soil properties and absorption characteristics. Moreover, measurement bias caused by instrument noise, light scatter and pathway variations are also inherent to spectral signatures. Multivariate calibrations performed using chemometric techniques have long been the de facto standard for deciphering soil spectral data and for correlating soil properties with absorption characteristics. Chemometrics involve the statistical and mathematical design of measurement strategies used to accurately estimate chemical properties. Multivariate calibrations allow for the empirical estimation of one variable (e.g. soil moisture) using a dataset of multiple measurements (e.g. a spectral signature). Soil analysts prefer the use of

PLSR (see Section 2.2.4.2 and Chapter 6) over other multivariate calibration techniques such as multiple linear regression and principal components regression, as it is not hindered by large, highly collinear datasets of predictor variables and can model the variance of response variables in an interpretable number of components (Rossel & McBratney 2008).

However, the accuracy of PLSR tends to decrease as a function of non-linearity between soil properties and soil spectra (Araújo et al. 2014). Non-linear approaches, such as ANN (Mouazen et al. 2010), SVM (Vohland et al. 2011) and RF (Viscarra Rossel & Behrens 2010), have been shown to outperform PLSR accuracy when employed for spectroscopic analysis (Nawar & Mouazen 2019). See Section 2.2.4.3 and Chapter 6 for information on supervised classifiers and soil spectroscopic analyses, respectively.

2.4 SYNTHESIS AND LITERATURE EVALUATION

Geomorphological and soil mapping have been practiced for over half a millennium (Dollar 1998) and land surface segmentation was conceptually conceived more than 50 years ago (Savigear 1965; Troeh 1965; Waters 1958). It would be an immense task to collect, disseminate and review even a small portion of the large body of available research. As such, this chapter briefly overviewed methodologies and techniques pertaining only to object-based geomorphometry and soil spectroscopy. Nevertheless, context was provided by digressing to the domains of adjacent subject matter (e.g. specific geomorphometric principles).

This chapter defined several object-orientated disciplines for the sake of conceptual comparison and terminological differentiation. GEOBIA was defined as a sub-discipline of GISc that utilises image segmentation approaches to delimit discrete spatial objects. Unlike per-pixel analyses, GEOBIA approaches delimit objects that are geographically representative and analogous to human intuition. This characteristic makes GEOBIA particularly suitable for land surface segmentation, given that the delimitation of specific landforms is an inherently subjective practice based on ontological perspectives of analysts (Eisank & Drăguţ 2010; Eisank, Drăguț & Blaschke 2011).

Prior to the advent of GIS and DEMs, the implementation of land surface segmentation approaches was paper-based, qualitative, difficult to implement and time-consuming (Dalrymple, Blong & Conacher 1968; Lastoczkin 1987; Savigear 1965; Waters 1958). Inevitably, the fundamental quantitative geomorphologic principles developed through paper-based analyses informed various

geomorphometric techniques and systems. Most of these principles are based on the per-pixel classification of specific landforms without regard for the scale specificity of LSPs and landforms (Minár & Evans 2008). While progress towards objective and scale-dependent per-pixel land surface segmentation has been made (Jasiewicz & Stepinski 2013), state-of-the-art discrete geomorphometry approaches offer the ability to hierarchically delimit specific landforms at multiple scale representations using data-driven approaches (Drăguţ & Eisank 2011).

ESP 2, an automated data-driven MRS SP optimisation technique (Drăguţ et al. 2014; Drăguț, Tiede & Levick 2010), gained significant popularity in the land surface segmentation field. However, it only detects characteristic levels of scale present in the landscape and does not incorporate the morphological parameters of target landforms in the scale optimisation process (Anders, Seijmonsbergen & Bouten 2011). Data-driven discrete geomorphometric segmentation approaches that detect appropriate scales based on specific geomorphometric criteria defined prior to segmentation are needed to eliminate the discrepancy between characteristic levels of scale present in the landscape and the scales at which specific landforms are observable.

Although no literature detailing the intersection between the specific and discrete geomorphometric paradigms (within the context of land surface segmentation) exists, discrete geomorphometric land surface segmentation principles have been integrated with specific geomorphometric semantic models to unify landform ontologies and derive transferable rulesets that consider morphometric, morphologic and contextual criteria (Eisank, Drăguț & Blaschke 2011; Gerçek, Toprak & Strobl 2011). The development of landform semantic models is, however, increasingly overshadowed by the increasing popularity of black-box supervised machine learning approaches in geomorphometry (Florin 2019; Li et al. 2020; Li & Chen 2020; Mithan, Hales & Cleall 2019; Swan 2017). Black-box classifiers are typically characterised by complex mathematical principles or distance representations that are inaccessible to human understanding (Loyola-Gonzalez 2019; Rudin 2019). As such, studies employing black-box machine learners are unable to incorporate expert knowledge in the classification process and are restricted to grid search procedures that tweak hyperparameters based only on classifier accuracy (Pedregosa, Weiss & Brucher 2011). White-box machine learners, on the other hand, are intuitive, easy to understand and can be inductively employed to compare trained models with expert knowledge to optimise hyperparameters (Quinlan 1987; Quinlan 1986). Literature detailing semantic modelling of landforms and supervised landform classification approaches have been found to be mutually exclusive. Moreover, white-box classifiers are underutilised according to geomorphometry

literature and no studies could be found that used white-box machine learners to inductively develop landform rulesets.

DSM studies rely extensively on geomorphometric techniques to model the inter-relations between topography – one of the Jenny (1941) soil formation factors – and soil parameters (McBratney, Mendonca Santos & Minasny 2003; Minasny & Mcbratney 2015). Over the last couple of decades, soil DRS coupled with chemometric techniques have also been established as a reliable way of estimating various soil parameters (Stenberg et al. 2010; Stoner & Baumgardner 1981). The integration of soil DRS and DSM methodologies is a logical progression in pursuit of low-cost, fine-scale soil maps (Rizzo et al. 2016; Vasat et al. 2017; Wadoux 2015). Moreover, the advent of machine learning techniques enables more accurate modelling of non-linear soil-landscape and soil spectra relationships (Minasny & Mcbratney 2015; Nawar & Mouazen 2019). Although an extensive body of literature is available on DSM and DRS, the soil-landscape paradigm is unrepresented in specific geomorphometry literature, despite the evident potential of soil spectroscopy for geomorphological mapping.

Against this background, this study investigates object-based geomorphometric and soil spectroscopy techniques and aims to develop methodologies for the mapping of fluvial terraces. This chapter serves to facilitate a working understanding of the techniques and methodologies employed henceforth, albeit without delving too deep into the fundamental basics. As such, the literature covered in the introductory chapter (Chapter 1), as well as in this chapter, is supplementary and does not introduce all topics relevant to this dissertation. Each of the subsequent chapters, excluding the synopsis (Chapter 7), has been prepared as freestanding journal articles and contains additional literature reviews relating to their specific subject matter.

CHAPTER 3: THE GREAT LETABA FLUVIAL TERRACES: AN INVESTIGATION INTO THE LANDSCAPE AND SOILS OF THE SOUTH AFRICAN LOWVELD

3.1 ABSTRACT

Understanding the soil-geomorphology relationships within the context of landscape evolution is one of the fundamental tasks of soil science. The soils in the valleys of African rivers are strongly influenced by the long history of weathering and fluvial sediment transport, and, unlike the valleys of the northern temperate regions, these soils have not been affected by recent sediments of continental glaciations.

Land and water use, land degradation and desertification in the sensitive ecosystems of the Kruger National Park (South Africa) and the adjacent areas can be better understood within the context of the recent and former natural history of the area. Here we focus on the semi-arid/dry sub-humid section of the Great Letaba River – one of the third-level tributaries of the larger Limpopo River – to understand the soil pattern of what we deem to be a typical river-shaped landscape of the northern Lowveld. Field reconnaissance and terrain analysis results are coupled with 1:250 000 scale soil data to characterise fluvial terraces and planation surfaces to investigate the terrain, soil, climate and bedrock paradigm. Our findings expand the current understanding of the fluvial terrace morphology and distribution, and a clear relationship between fluvial terrace morphology, soil classes, local climate and underlying geology is identified. Finally, our investigation of the present-day planation surface morphology highlights complex landscape evolutionary cycles driven by river incision, fluvial terrace formation and land surface planation.

3.2 INTRODUCTION

Fluvial (river) terraces are remnants of old floodplains abandoned during the incision of a river into underlying bedrock and frequently occur in step-like sequences called terrace staircases. The formation of fluvial terrace staircases is driven by alternating events of sediment aggradation and river incision: during each consecutive flood event, sediment is deposited on the floodplain, which is later abandoned when the river cuts into the underlying bedrock. A new, lower lying floodplain is consequently established and the formation of a new fluvial terrace step initiated (Bridgland & Westaway 2008; Knight, Mitchell & Rose 2011; Pazzaglia 2013; Vandenberghe 2015). Terrace staircases may constitute the primary landform in a catchment area and can extend all the way to

the catchment boundary; however, in most settings older terraces are removed by erosion (Pazzaglia 2013).

Two types of fluvial terraces exist, namely: fill and strath (sometimes called erosional) terraces. Fill terraces consist of abandoned alluvial valleys and floodplains, whereas strath terraces are erosional in nature and cut into underlying bedrock (Leopold, Wolman & Miller 1964). Vandenberghe (2015) distinguishes between strath and erosional terraces by reserving the term "strath terraces" for terraces that cut exclusively into underlying bedrock and "erosional terraces" for terraces that also cut into unconsolidated sediments. The term "strath" is used to refer to the erosional base of a terrace. Fluvial terraces, or remnants thereof, that occur at the same elevation on both sides of the valley are referred to as paired fluvial terraces, and those that occur only occasionally at one side of the valley and further downstream at the other side are called non-paired fluvial terraces (Bucher 1932; Campbell 1929; Pazzaglia 2013).

In the last century, fluvial terraces received continuous attention from researchers due to the presence of economic placer minerals (Du Toit 1922; Du Toit 1910) and the suitability of large flat terrace surfaces for agriculture (Pazzaglia 2013). Furthermore, paleo climatic and tectonic conditions can be deduced by determining the ages of fluvial terraces, since base-level drop, climatic fluctuation and tectonic uplift all contribute to the cyclical formation of fluvial terraces (Blum & Törnqvist 2000; Bridgland & Westaway 2008; Pazzaglia 2013; Vandenberghe 2015). In regions far from the coast, for example in crustal provinces where the effect of sea-level rise is minimal, tectonic uplift is an essential requirement, and differences in uplift histories often result in variations in terrace records between different areas and particularly in different crustal provinces (Bridgland & Westaway 2008; Maddy 1997; Maddy, Bridgland & Green 2000).

Accordingly, the tectonic uplift events that followed the dismantling of Gondwana contributed to the formation of a high marginal escarpment (known as the Great Escarpment) and established late Pliocene to Holocene fluvial terraces along rivers in southern Africa (Partridge & Maud 1987). Three distinct uplift events are associated with the formation of the African, post-African I and the post-African II surfaces. The African surface resulted from post-rifting polycyclic planation, driven by the newly formed Atlantic and Indian Ocean base levels and a tropical Cretaceous climate, and lasted from the late Jurassic or early Cretaceous until late Cretaceous or early Miocene. Tectonic uplift of up to 300 m resulted in denudation that lasted from the early Miocene until the late Pliocene and produced the post-African I surface. The post-African II surface can be

attributed to further denudation brought about by asymmetrical uplift of up to 900 m during the late Pliocene and lasted well into the Holocene. Concurrently, climatic fluctuations and subsequent glacio-eustatic sea-level changes further facilitated the formation of fluvial terraces in the southern African paleo-fluvial landscape (Dollar 1998; Partridge et al. 2010; Partridge & Maud 1987).

The Great Letaba Water Catchment (GLWC), which borders the Kruger National Park (KNP) and the Letaba Water Catchment, is a good example of a modern-day fluvial environment which was shaped by the geomorphological template set by the dismantling of Gondwana. The GLWC, which forms part of the Lowveld Geomorphic Province and is located within the Kaapvaal craton (Nguuri et al. 2001), is characterised by African planation surfaces in the west and post-African I planation surfaces in the east (Partridge & Maud 1987). The Great Letaba River (GLR) originates along the boundary of the Great Escarpment and, together with the Middle Letaba River (MLR) and Little Letaba River (LLR), is one of three major tributaries of the Letaba River. The MLR is a tributary of the LLR, which meets the GLR at the border of the KNP. After the confluence of the LLR and GLR, the river name changes to only the Letaba River. Further downstream the Letaba River flows into the Olifants River, which is a tributary of the Limpopo River. Despite the Letaba River remaining largely untouched within the confines of the KNP, upstream the GLR has been severely affected by anthropogenic activity, e.g. water extraction for irrigation and the construction of dams (Moon & Heritage 2001).

The influence of the rivers on the shaping of the topography and the formation of the dominant landforms in the Lowveld directed Lester King to formulate the pediplain theory. King (1963) hypothesised that the presence of extensively pedimented planes in the landscape of the Lowveld can be attributed to the gradual retreat of the Great Escarpment. In other words, the erosion of the Great Escarpment by the local rivers provided the sediment that aggregated along the major rivers (including the GLR) to produce fluvial terraces. More recently, Rozanov et al. (2017) resumed discussion on the geomorphology of the Lowveld by identifying three steps of fluvial terraces along the GLR. The first was located between 1 and 6 m above the river channel and would therefore correspond to the terraces identified next to the present-day floodplain by Moon & Heritage (2001). The second fluvial terrace identified by Rozanov et al. (2017) occurs between 17 and 36 m above the river channel and, lastly, the third between 40 and 41 m above the river channel. Using clay mineralogy, Rozanov et al. (2017) showed that the fluvial terrace sediments along the LLR and the GLR are locally sourced as the clay content reflected the underlying

bedrock and not the lithologies associated with the Great Escarpment. This contradicts King's pediplain theory and is indicative of localised planation activity throughout the catchment.

While the findings of Rozanov et al. (2017) contribute significantly towards understanding fluvial terrace soil formation in the Letaba valley, little is known about the soils associated with the fluvial terraces. The only catchment-level soil database available is the National Land Type database (Land Type Survey Staff 2006), which is a soil association map that indicates portions of the landscape associated with various soil types at reconnaissance scale (Flynn et al. 2019).

Furthermore, while fluvial terraces such as those located within the Vaal basin have been extensively studied and excavated in search for diamondiferous gravels (Bridgland & Westaway 2008; Helgren & Butzer 1974; De Wit 2004), the GLWC fluvial terraces have received little attention from scholars due to a lack of economic placer deposits. Recently, the association of the GLR with the KNP and increased awareness of the importance of the river for the agricultural production and water supply to municipalities and industry have drawn the attention of both agricultural and conservation scientists. Better understanding of the recent and former natural history of the GLWC is critical for land use optimisation in support of sustainable development of land resources and for the mitigation of land degradation and desertification in the sensitive ecosystems of the KNP.

Although past works mention the existence of fluvial terraces immediately next to the modern-day floodplain (Moon & Heritage 2001), as well as flights of terraces up to 41 m above the river channel (Botha & De Wit 1996; Rozanov et al. 2017), no investigation into the geomorphology and soils of the fluvial terraces along the GLR has been made. This study aims to characterise the geomorphology and soils of the fluvial terraces located along the GLR and to evaluate whether the GLWC land surface morphology supports the Rozanov et al. (2017) hypothesis of localised erosion.

3.3 MATERIALS AND METHODS

3.3.1 Study area

The GLWC (Figure 3-1) is located within the northern Lowveld Geomorphic Province, which consists predominantly of undulating plains that are severely pedimented and eroded. The underlying bedrock of the GLWC mostly comprises granite and gneiss bedrock. Lesser amounts of diorite, greenstone, quartzite, shale and other sedimentary rocks occur in the extreme south. Granite koppies (inselbergs), dolerite (mafic) dikes and lava outcrops are also scattered throughout the catchment (Partridge et al. 2010). The region can be characterised as semi-arid/dry sub-humid with rainfall that occurs mainly in the summer months. The annual precipitation rates of the catchment vary from 400 mm yr^{-1} in the east of the catchment to 900 mm yr^{-1} in the west along the Drakensberg escarpment (Land Type Survey Staff 2006). The GLR originates along the escarpment at an elevation of roughly 1 400 m above sea level and the confluence with the LLR is at the border of the KNP.

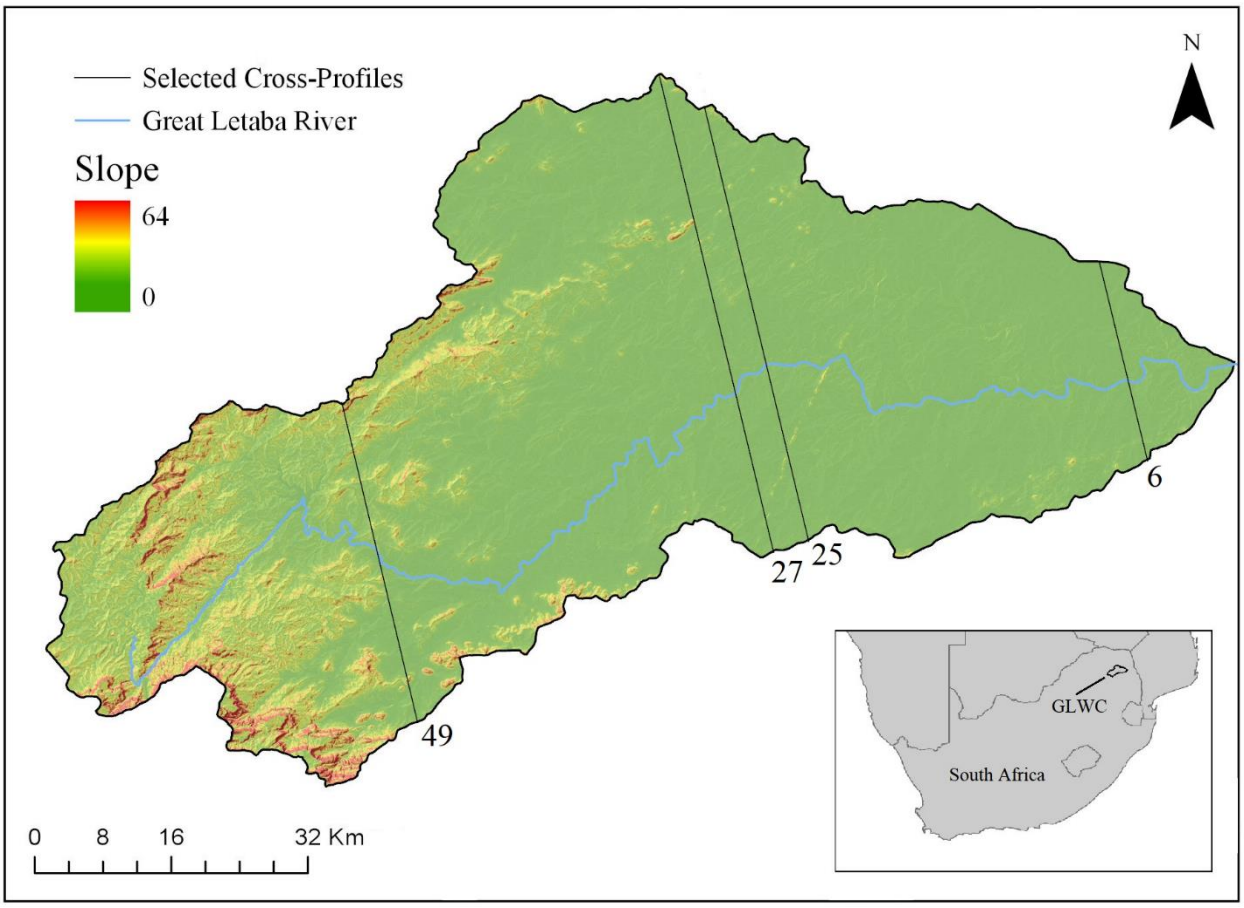


Figure 3-1 Location map of the demarcated study area, showing the location of the GLWC within southern Africa and the selected cross-profiles visually inspected

3.3.2 Field reconnaissance

Two field reconnaissance excursions were carried out to investigate the fluvial terraces and planation surfaces of the GLWC. The first excursion, detailed by Rozanov et al. (2017), investigated the fluvial terraces of the GLR and LLR and included soil analysis of 37 soil profiles located on various fluvial terraces. The second excursion focussed exclusively on the terraces of the GLR and served to collect field notes and photographs to record geomorphological descriptions of various fluvial landforms.

3.3.3 Geomorphometry

3.3.3.1 Digital elevation model

The Stellenbosch University DEM (SUDEM) was selected for the creation of the cross-profiles. The SUDEM, which has a resolution of 5 m and a mean absolute error of 2.1 m, was interpolated from large-scale contours and spot height data and fused with the 30 m shuttle radar topography mission DEM (Van Niekerk 2015a).

3.3.3.2 Terrain analysis

The System for Automated Geographical Analysis (SAGA) software package was used to create cross-profiles along the GLR. Elevation values were sampled every two pixels (or 10 m) starting from the river channel and terminating at the catchment boundary on both sides of the river. A total of 50 cross-profiles, with a two-kilometre spacing between each, was created between the confluence of the GLR and LLR and the Tzaneen dam (Figure 3-2).

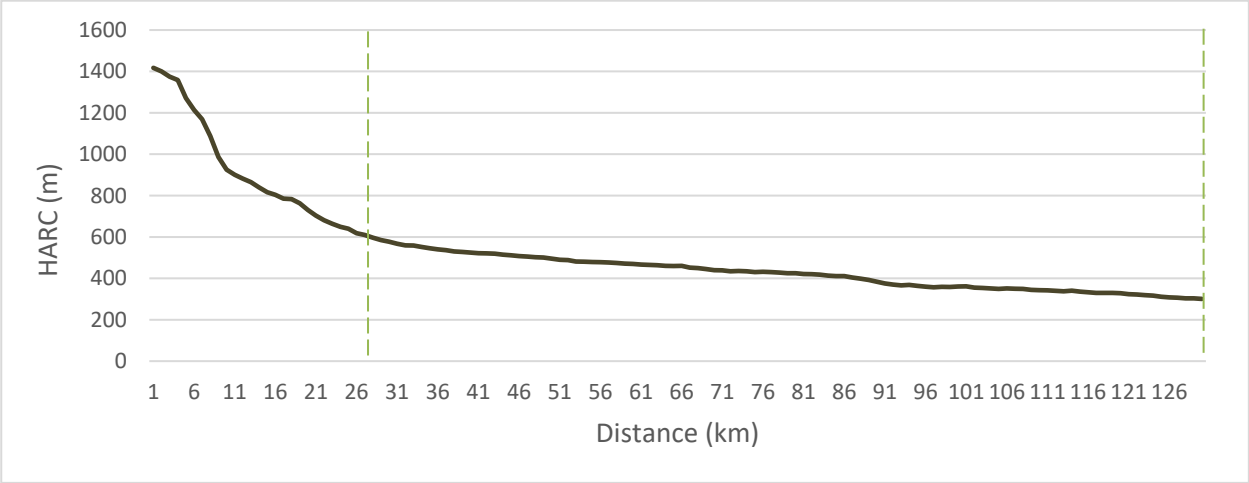


Figure 3-2 Great Letaba River profile, cross-profiles between the elevations of 303 m and 609 m above sea level (demarcated by two vertical dashed lines) were produced for visual assessment

Further analysis of the cross-profiles included identifying the position of the river channel in each cross-profile and separating each profile by splitting the north bank of the river from the south bank. Additionally, for each cross-profile, the lowest point of the river channel was selected and the height above the river channel (HARC) calculated.

Comprehensive visual assessment of every cross-profile was carried out by comparing each selected cross-profile with its neighbours, high-resolution aerial photographs and detailed field reconnaissance notes. The HARC, terrace width and comments on the surface condition were recorded for each terrace level.

3.3.3.3 Soil data

The South African Land Type Survey (LTS) is a 1:250 000 scale land resource map that regionalises the landscape into unique polygons based on climate, terrain and soil (Land Type Survey Staff 2006). The LTS does not delineate specific soil properties, instead, it outlines soil class percentages, soil depth ranges, soil depth class percentages, root limiting layers, clay percentages and soil textures associated with up to five landscape positions defined on the basis of generalised places of convergence and divergence of water movement, i.e. crest, scarp, mid-slope, foot-slope and valley (Flynn et al. 2019). Whereas specific soil types are given in the land type inventories, soil classes (Table 3-1) were utilised by the LTS staff to group soil types for spatial representation. In addition to the LTS polygons, the individual soil profile data used to produce the GLWC LTS polygons were kindly provided by the ARC-ISWC (Pretoria). A boxplot was used to investigate the slope gradient values associated with each soil type in the study area and to compare soil type slope gradient values.

Table 3-1 Classification of soil types according to the South African soil classification system devised by the Soil Classification Working Group (1991)

Soil class	Soil types	Description
1	Ia, Ma, Kp, No	Soils with humic topsoil horizons
2	Hu, Cv, Gf, Sd, Oa	Freely drained, structureless soils
3	Av, Gc, Bv, Pn	Red or yellow structureless soils with a plinthic horizon
4	Sp, Ct, Vf, Du, Fw	Excessively drained sandy soils
5	Ar	Dark clay soils which are not strongly swelling
6	Bo, Ik, Tk	Swelling clay soils
7	Va, Sw	Soils with a pedocutanic (blocky structured) horizon
8	We, Cf, Lo, Wa	Imperfectly drained soils, often shallow and often with a plinthic horizon
9	Lt, Hh	Podzols
10	Wo	Poorly drained dark clay soils which are not strongly swelling
11	Rg	Poorly drained swelling clay soils
12	My, Mw	Dark clay soils, often shallow, on hard or weathering rock
13	Ms, Gs	Lithosols (shallow soils on hard of weathering rock)
14	Se, Ss, Kd	Duplex soils, often poorly drained
15	Ch, Ka, Fw	Wetlands
16	Non-soil land classes	Pans, rivers, stream beds, erosion, marshes, reclaimed land, dunes and gravels
17	Rock	

Source: MacVicar (1977)

3.4 RESULTS

3.4.1 Field reconnaissance

The present-day morphology of the GLR channel and the adjacent landscape change significantly as one travels from the escarpment (west) to the KNP (east). Near the escarpment, the river is constrained by narrow valleys, which hinders development of laterally extensive fluvial terraces. Nonetheless, the subtropical climate and high rainfall in this area of the catchment promote regular flooding and subsequent formation of fill fluvial terraces next to the modern-day river channel. These fluvial terraces have deep soils and are covered by either riparian vegetation or agricultural fields. In some places, the GLR meanders and erodes lower terraces. Consequently, large portions of terraces are removed on one side of the river while the terraces on the other side of the river remain intact, which results in the appearance of unpaired terraces. Older terrace surfaces are present at higher elevations. However, some are incised by tributaries or otherwise denuded and no longer exhibit well-preserved fluvial terrace morphology. Commercial agriculture is widespread in the western parts of the catchment, and related activities, such as ploughing, have

had a significant impact on the morphology of the landscape. Consequently, terrace boundaries are less pronounced and soil mixing is inevitable in these areas.

Further downstream from the escarpment, the landscape opens into a considerably wider and flatter valley. The fill fluvial terraces that occur extensively near the escarpment are less prevalent downstream, infrequently inundated and typically occur in close proximity to the present-day river channel. Evidence of planation activity dominate the higher-lying areas of the valley, covered only by a thin layer of alluvium.

3.4.2 Terrain analysis

In total, 50 cross-profiles were scrutinised and visually interpreted to corroborate and refine the observations made during the field reconnaissance expedition. The fluvial terraces and planation surfaces of the GLWC are described and detailed in the following subsections. All elevation values are given as the HARC derived from the SUDEM.

3.4.2.1 Fluvial terraces

Interpretations of the cross-profiles indicate that the morphology of the GLWC fluvial terraces are not uniform across the catchment. Near the escarpment, the incision of the GLR into the underlying bedrock resulted in a narrow valley, where steep slopes and mountainous terrain correspond to underlying leucocratic biotite granite. Further east, near the KNP, gently-sloped wide valleys match the presence of Goudplaas gneiss. Frequent inundation of presently abandoned floodplains resulted in numerous steps of fluvial terraces on both sides of the river channel. For the purpose of identifying fluvial terrace surfaces produced by the GLR, only the first 5 km of each cross-profile was examined, thereby excluding most tributary inundated fluvial surfaces. Any prominent tributary fluvial terraces present on the cross-profiles are marked with a grey ellipse. Figure 3-3 (cross-profile 49) shows seven fluvial terrace steps, located at 3, 7, 16, 22, 32, 47 and 58 m on the north side of the GLR. The fluvial terraces located at 2 and 7 m above the river channel are flat; however, all the fluvial terraces located between 16 and 58 m are sloping towards the river channel. No morphological features representative of fluvial terraces remain above 58 m, but comparisons with other profiles show that the peak located at 62 m correlates with fluvial terrace surfaces further downstream, indicating the presence of an incised fluvial terrace. A fluvial terrace produced by a tributary is present at 92 m.

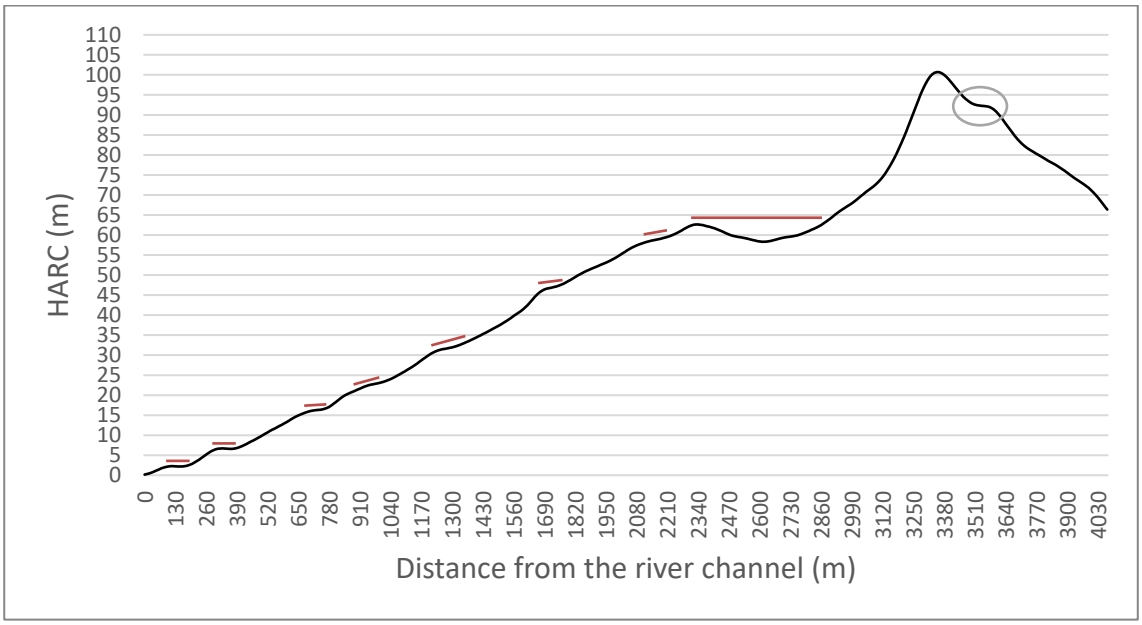


Figure 3-3 Cross-profile 49 (northern bank of the GLR, located near the escarpment), where the red lines demarcate fluvial terraces and the grey ellipses indicate tributary fluvial terraces

Figure 3-4 shows the south side of cross-profile 49. Key differences between the two sides of the river channel are the lack of fluvial terraces at 3 and 7 m and a very large erosional surface located at roughly 21 m. Unlike the northern side of the GLR, a small fluvial terrace is present at 37 m. Similar to the northern side of the GLR, fluvial terraces are present at 16, 32 and 46 m. At 64 m, a peak indicates the elevation of a heavily dissected surface, incised by a tributary, that produced at least three fluvial terrace steps, the lowest of which also shows signs of incision.

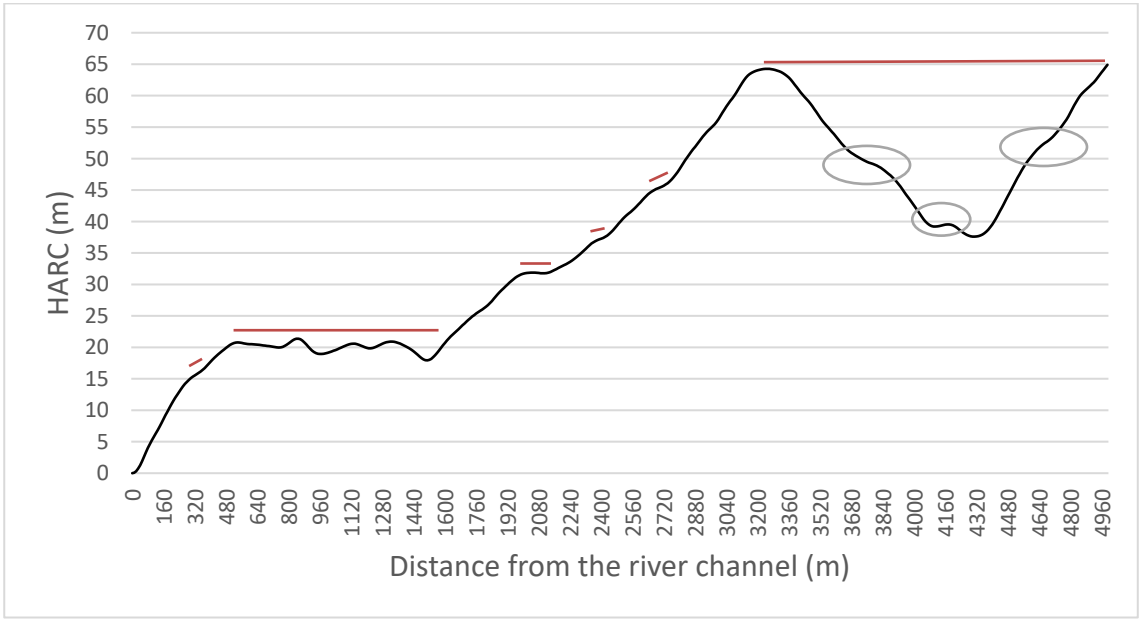


Figure 3-4 Cross-profile 49 (southern bank of the GLR, located near the escarpment), where the red lines demarcate fluvial terraces and the grey ellipses indicate tributary fluvial terraces

Further downstream, the landscape morphology changes to a flatter and wider valley, underlain almost exclusively by Goudplaas gneiss. Cross-profile 54 (Figures 3-5 and 3-6) was chosen to illustrate the landscape morphology roughly halfway between the position of cross-profile 96 and the confluence of the GLR and the LLR. The northern side of cross-profile 54 shows distinct fluvial terraces present immediately next to the river channel at 3 and 5 m. Two peaks (at 13 m and 16 m) specify the elevation of two heavily dissected fluvial terraces, both of which have small surfaces, which is evident of further fluvial terrace development by incising tributaries. Furthermore, at 18 m and 21 m, distinct small surfaces resemble fluvial terrace morphological features. A large peak at 34 m demarcates the boundary of a large planation surface which extends ~4.8 km north and a minimum of 1 km to the east and correlates with fluvial terraces to the west. This surface has been significantly incised, and fluvial terrace development contributed by tributary river is extensive.

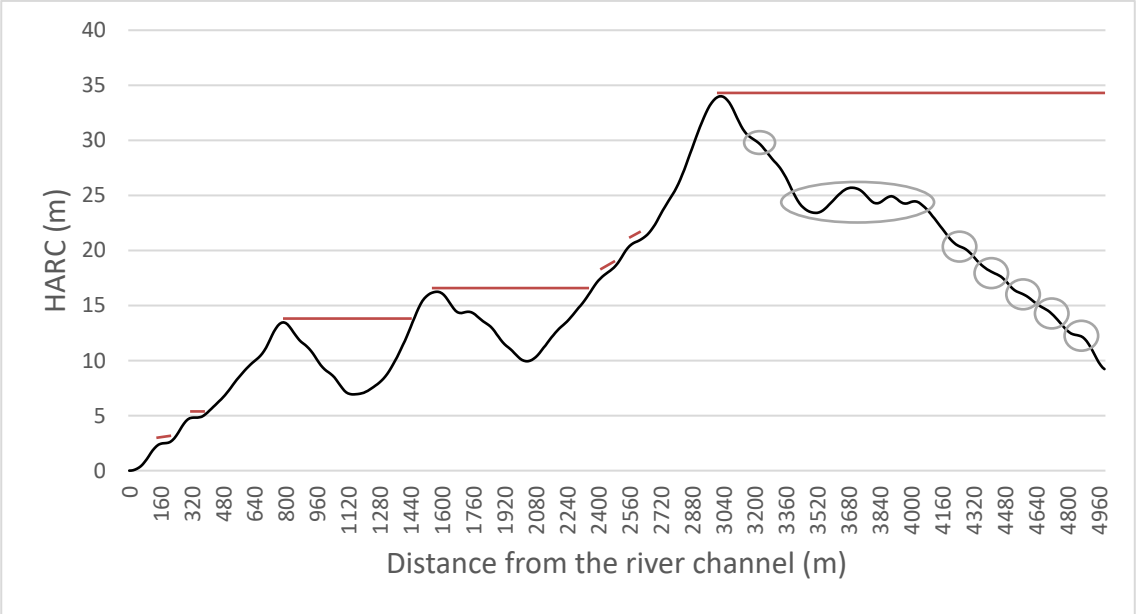


Figure 3-5 Cross-profile 27 (northern bank of the GLR, located roughly halfway between the escarpment and the border of the KNP), where the red lines demarcate fluvial terraces and the grey ellipses indicate tributary fluvial terraces

Figure 3-6 shows the southern bank of the GLR at cross-profile 27. Fluvial terraces are present at 3, 7, 15, 19, 27 and 30 m. The fluvial terraces at 3, 15, 19 and 30 m exhibit surfaces sloping toward the river channel, whereas at 7 and 27 m, the surfaces are mostly flat and only slightly incised. At 24 and 35 m, large and extremely dissected surfaces occur, the appearance of which are more representative of planation surfaces than fluvial terraces; however, comparisons with neighbouring cross-profiles correlate these surfaces with existing fluvial terraces. Both surfaces have at least one set of fluvial terraces produced by the incising tributary. At 40 m, a very large flat surface is

present, while at 43 and 45 m, two small, flat surfaces are representative of the occurrence of fluvial terrace morphology.

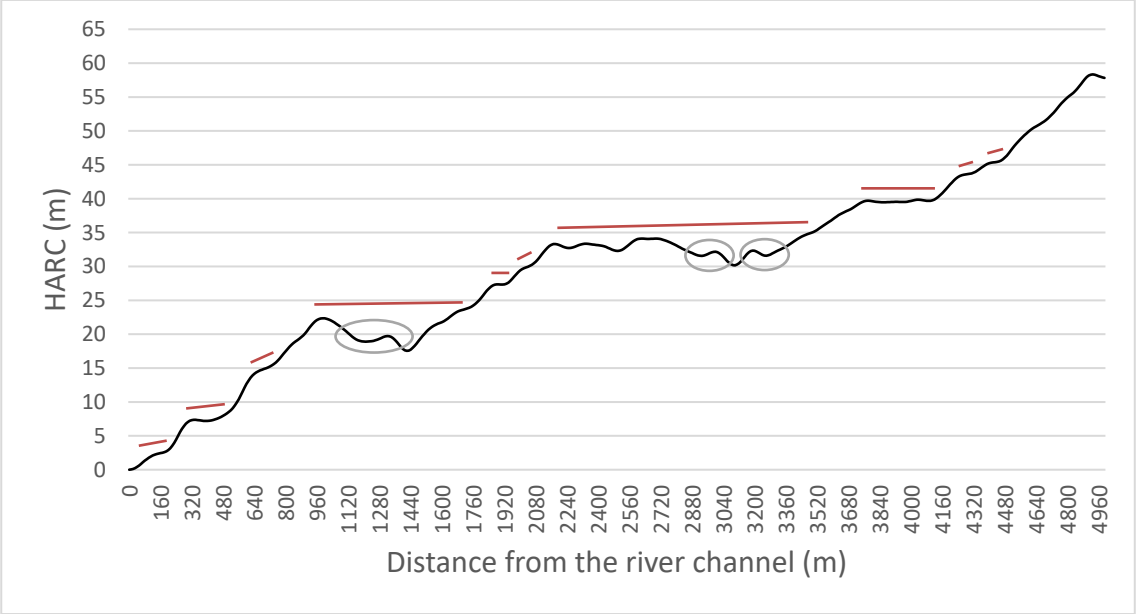


Figure 3-6 Cross-profile 27 (southern bank, located roughly halfway between the escarpment and the border of the KNP), where the red lines demarcate fluvial terraces and the grey ellipses indicate tributary fluvial terraces

Near the KNP, where the GLR meets the LLR at the border of the GLWC, cross-profile 6 was selected to describe the fluvial terraces in the most eastern part of the GLWC. Figure 3-7, illustrating the north side of the GLR, shows that fluvial terraces that occur next to the river channel are smaller than in the western part of the GLWC. Starting at the river channel, the first three fluvial terraces are located at 3, 5 and 9 m and show surfaces sloping towards the river channel. Higher up, two peaks indicate the elevation of heavily dissected terrace surfaces at 17 and 22 m. Both of these surfaces were incised by a fluvial terrace forming tributary. Two small terrace surfaces are present at 27 and 28 m, above which the upper landscape is too denudated to recognise any fluvial terrace morphological features. By comparing the peaks above 28 m with peaks and preserved flat surfaces of nearby cross-profiles, it was determined that fluvial terrace surfaces are present both upstream and downstream at 32, 36, 42 and 45 m. At 47 m, however, a planation surface extends up to 8 km inland and contains fluvial terraces formed by the incising tributary.

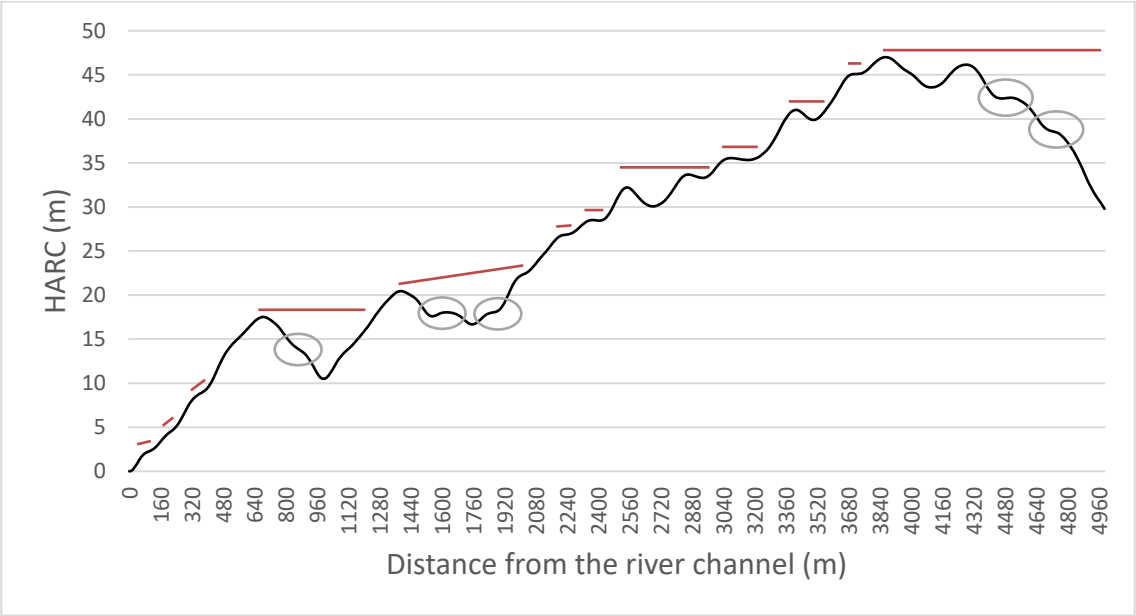


Figure 3-7 Cross-profile 6 (northern bank of the GLR, located near the border of the KNP), where the red lines demarcate fluvial terraces and the grey ellipses indicate tributary fluvial terraces

Similar to the north bank of the GLR at cross-profile 6, the south bank (Figure 3-8) is heavily denudated and difficult to interpret. Consequently, comparisons with adjacent cross-profiles were needed to decipher the landscape structure. Next to the river channel, fluvial terraces are present at 6 and 8 m. At 6 m, the surface is incised and do not reflect the flat morphometry of fluvial terraces at the same elevation further upstream. The second fluvial terrace, located at 8 m, exhibits a flat surface that is sloping towards the river channel at a shallow angle; however, the back slope of the surface is incised to almost the elevation of the GLR channel and shows two steps of tributary fluvial terraces. Similarly, local peaks at 11 and 13 m specify the elevations of two fluvial terraces surfaces that have been almost completely removed, the latter of which also has one step of fluvial terrace produced during the incision of the original surface. At 18 m, a large planation surface, which correlates well with fluvial terraces to the west, extends for roughly 1.3 km to the north and contains evidence of more recent fluvial terrace formation by the second order streams. Finally, fluvial terrace surfaces identified at 20, 24 and 31 m are significantly incised and may contain fluvial terraces produced by tributaries; at 22 m a well-preserved fluvial terrace surface is visible.

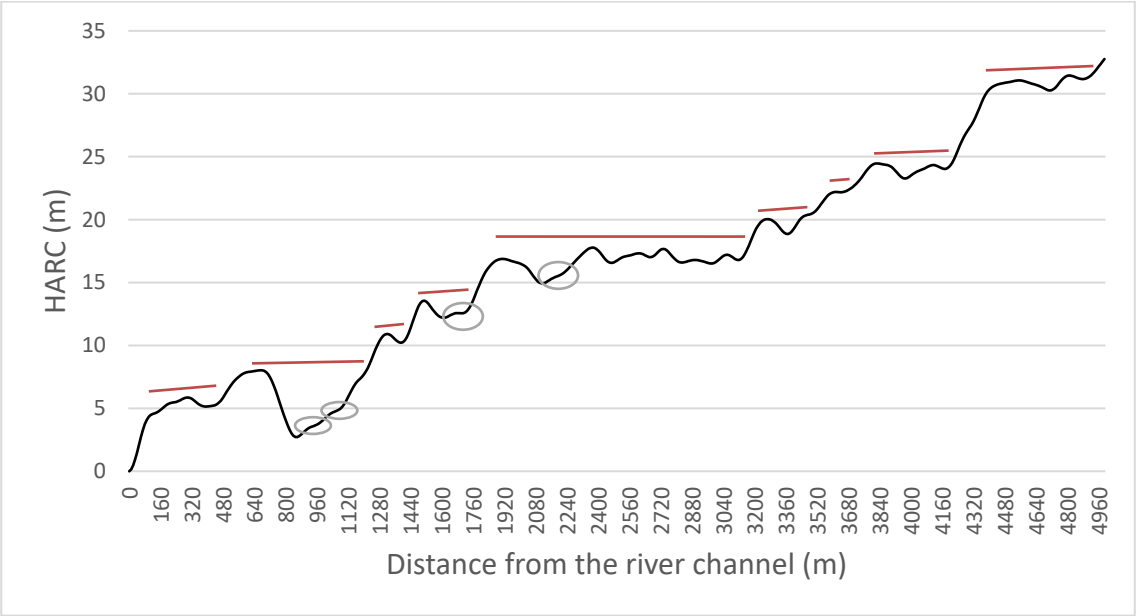


Figure 3-8 Cross-profile 6 (southern bank of the GLR, located near the KNP), where the red lines demarcate fluvial terraces and the grey ellipses indicate tributary fluvial terraces

Figures 3-3 to 3-8 clearly show that well-preserved fluvial terraces are mostly located in close proximity to the present-day river channel and that, as the distance from the river channel increases, the fluvial terraces become heavily denudated, incised or almost completely removed. In some cases, however, well-preserved fluvial terrace surfaces, or remnants thereof, are found significant distances from the present-day river channel. Figure 3-9 plots cross-profile 25 on the south side of the GLR, starting at 10 km from the GLR channel and terminating at the catchment boundary. Therefore, only the land surface above 70 m is shown, allowing investigations of only the higher-lying landforms. Flat surfaces representative of fluvial terrace morphology can be identified at 108, 124 and 134 m. This exemplifies that fluvial terraces may occur up to the catchment boundary.

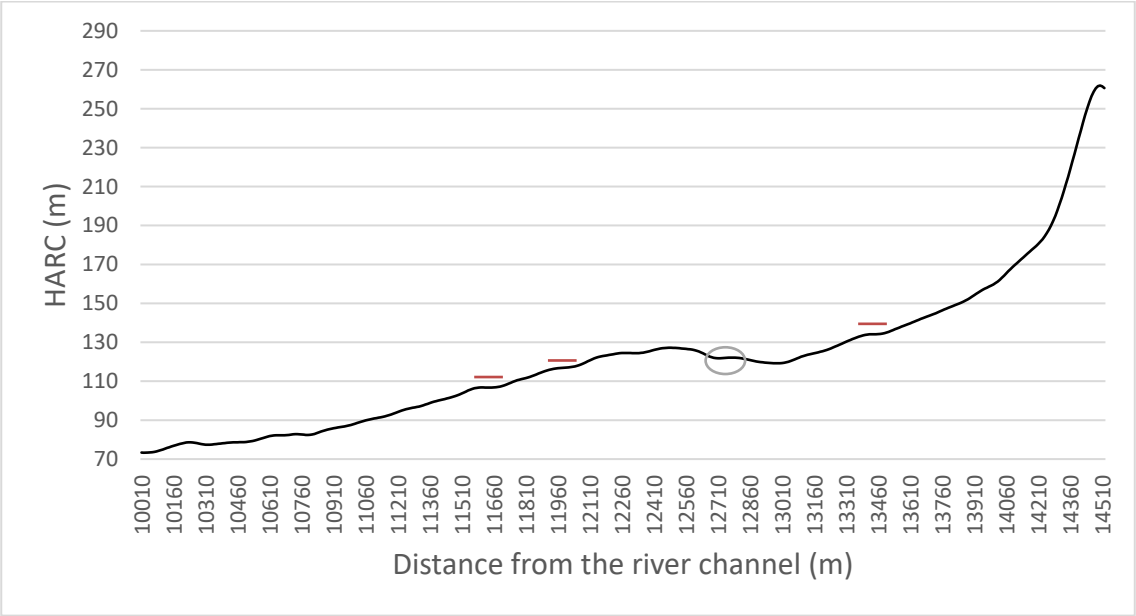


Figure 3-9 Cross-profile 25 (southern bank of the GLR, located roughly halfway between the escarpment and the border of the KNP), where the red lines demarcate fluvial terraces and the grey ellipses indicate tributary fluvial terraces

The HARC of the first five fluvial terrace steps, identified from each of the interpreted cross-profiles, was plotted (Figure 3-10 and Figure 3-11) and visually analysed. From Figure 3-10 it is clear that all fluvial terraces on the north side of the GLR have been partially removed or dissected at some point. The first terrace surface, for example, is only present 23 times out of 50 instances. Furthermore, the first terrace surface is not present at a uniform HARC; instead its HARC varies significantly. The second fluvial terrace is much more widespread and occurs at more uniform HARC values. Terraces three to five are also widespread and not often absent from the analysed cross-profiles. Nevertheless, the HARC values measured for all of these terraces vary significantly.

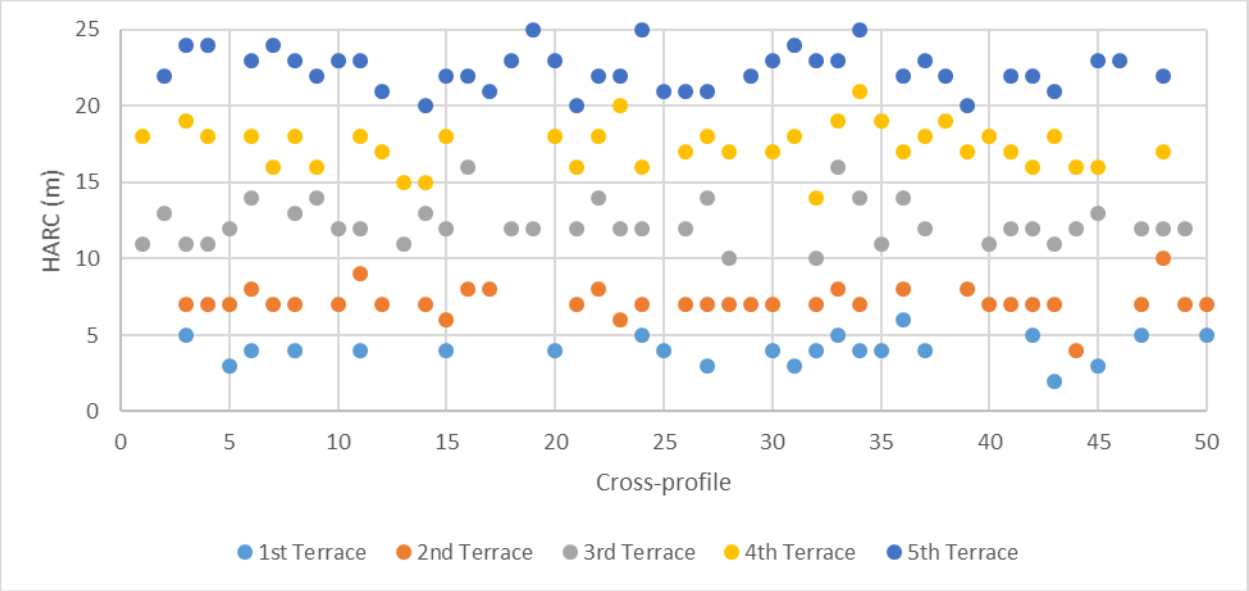


Figure 3-10 Height above channel of the first five fluvial terrace steps identified from cross-profiles on the northern bank of the GLR, located between the Tzaneen dam and the border of the KNP

The first five fluvial terrace steps on the southern side of the GLR (Figure 3-11) are more dissected compared to that of the northern side. In some places, the first two terraces have been completely removed (cross-profiles 17 to 19 and 36 to 37), and near the escarpment, all five terraces have been eroded from two cross-profiles (cross-profiles 46 to 47). None of the fluvial terrace steps present on the southern side of the GLR occur at uniform HARC values. Significant levels of denudation and erosion of the fluvial terraces are therefore apparent on both sides of the GLR.

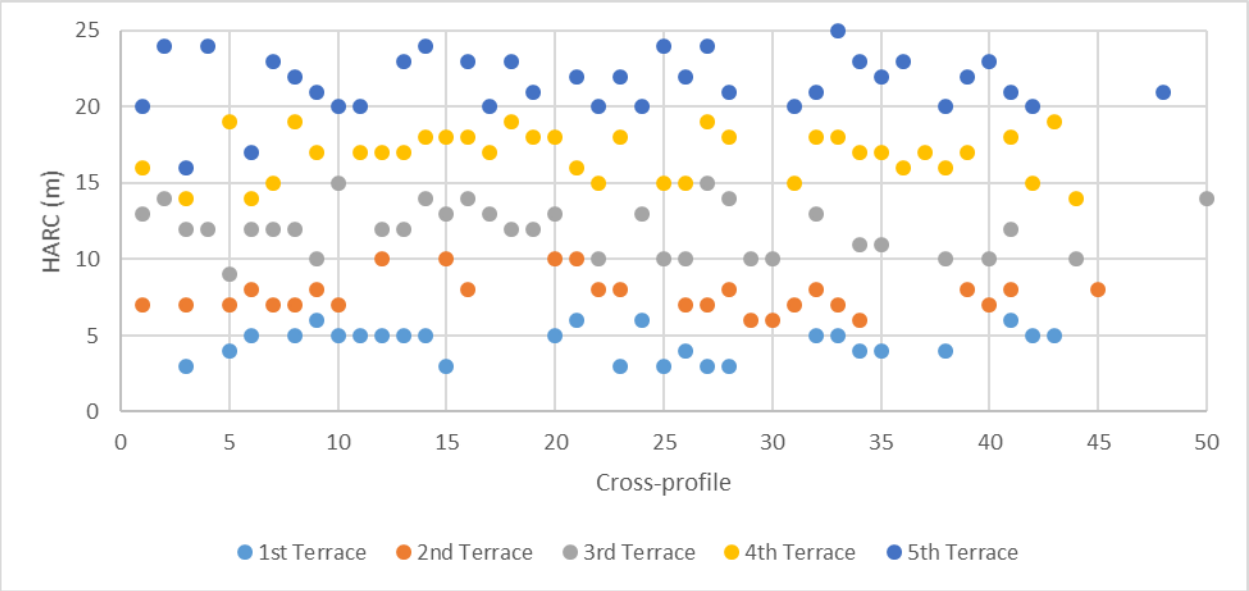


Figure 3-11 Height above channel of the first five fluvial terrace steps identified from cross-profiles on the southern bank of the GLR, located between the Tzaneen dam and the border of the KNP

In summation, the first terrace level occurs next to the modern-day GLR channel and does not exhibit uniform distribution throughout. In some sections along the river, the first terrace level is almost non-existent, e.g. in particularly straight river sections or on the inside of a river bend. The first terrace is more laterally extensive where the river is meandering significantly or on the outside of a river bend. Terrace levels two to five are generally more widespread compared to the first terrace. Figure 3-10 and Figure 3-11 illustrate that higher-lying terraces have been eroded down to the lower floodplain levels, thereby creating “islands” of higher-lying terrace remnants.

3.4.2.2 Planation surfaces

Planation surfaces have been mentioned thus far, but comprehensive descriptions are still needed. The presence of planation surfaces in the Lowveld have been attributed to the African, post-African I and post-African II erosional cycles. In the northern territories of the Lowveld, the GLWC is subject only to the African surface in the west and the post-African I surface in the east (Partridge et al. 2010; Partridge & Maud 1987). The cross-profiles examined show evidence of extensive planation activity throughout the catchment, particularly in the east near the KNP and less frequently in the west near the escarpment. While the northern side of the catchment is dominated by surfaces that often extend up 20 km inland, south of the GLR the planation surfaces are smaller and occur less frequently. All the planation surfaces described are heavily dissected and very little of the original surfaces remain. To extrapolate the heights of the original planation surfaces, points representing the boundaries of the planation surfaces were selected by identifying the highest cross-profile HARC.

Figure 3-12 and Figure 3-13 plot HARC for the north and south sides of cross-profile 50 starting at the GLR and ending at the catchment boundary. The north side of the GLR (Figure 3-12) is dominated by a single planation surface at 95 m. This particular surface starts 9.2 km from the GLR and extends 18.6 km inland. Unlike the scenario illustrated in Figure 3-12, the northern side of the GLR is not typically dominated by a single planation surface but by several planation levels that often correlate with the HARC of neighbouring fluvial terrace remnants or strath surfaces.

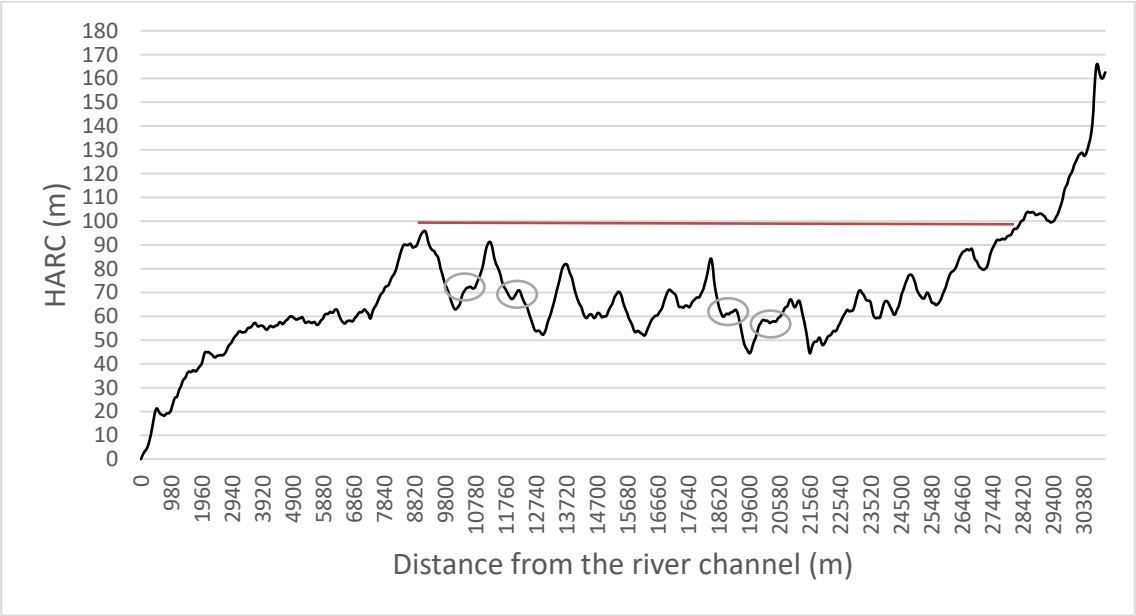


Figure 3-12 Cross-profile 50 (northern bank of the GLR, located near the escarpment), where the red lines demarcate planation surfaces and the grey ellipses indicate tributary fluvial terraces

The south side of the GLR at cross-profile 50 demonstrates the multilevel planation surface scenario. Figure 3-13 indicates four steps of extensive planation surfaces at 36, 54, 58 and 72 m. The first and fourth surfaces are the smallest and extends inland for 1.5 and 1.6 km, respectively. The second and third planation surfaces, however, are more extensive as they extend inland for 3.6 and 9.7 km, respectively. All the planation surfaces located in the GLWC are heavily dissected by tributaries that produced multiple steps of fluvial terraces as they continued to incise into the underlying bedrock. In some cases, it is clear that even the fluvial terraces produced by the tributaries have been incised by possible third-order fluvial terrace-producing streams; a setting that can be described as a nested-hierarchy of fluvial terraces. A small portion of the tributary inundated fluvial terraces are marked with grey ellipses on Figure 3-12 and Figure 3-13.

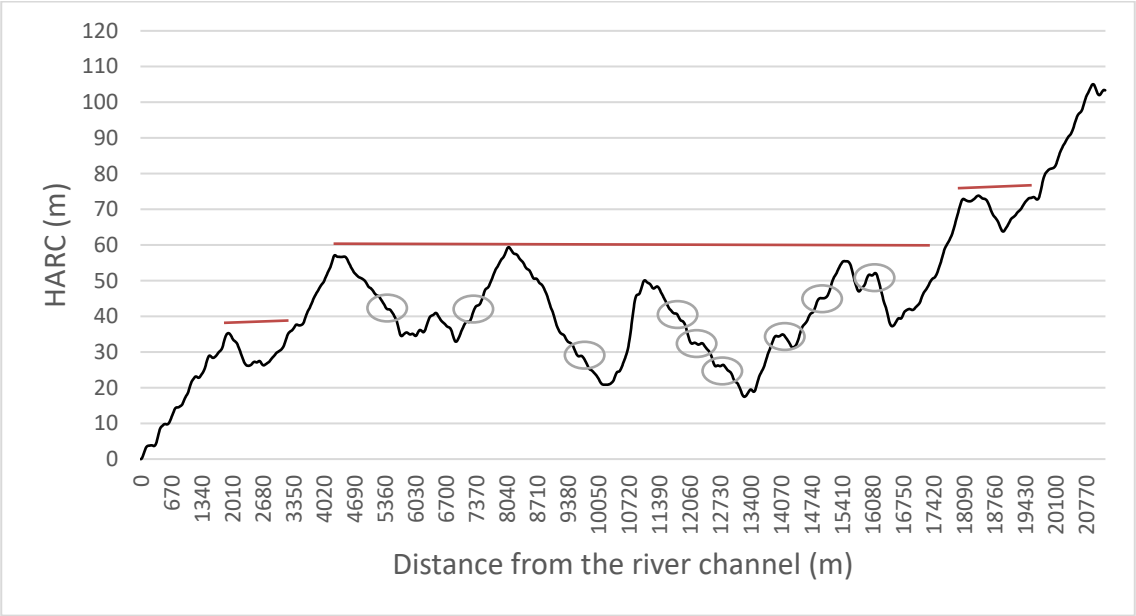


Figure 3-13 Cross-profile 50 (southern bank of the GLR, located near the escarpment), where the red lines demarcate planation surfaces and the grey ellipses indicate tributary fluvial terraces

3.4.3 Dominant soils

The South African LTS was used to characterise the various soils in the GLWC, as well as their association with the dominant landforms. A total of 86 unique polygons were delineated by the LTS staff (Figure 3-14b–d). First, a distinct difference between the land types located in the western and eastern parts of the catchment is noted. Due to large variations in terrain and soil morphology in the west, the land types are small and plentiful. The land types in the east, on the other hand, are much larger and demarcate more homogenous land- and soilscapes. The transition between these two regions can be attributed to both the underlying bedrock and climate. The land types in the west are underlain by a variety of lithologies, including leucocratic biotite granite, Goudplaas gneiss, Turfloop granite, Rooiwater diorite and Gravelotte lavas. The lithologies associated with the larger land types in the east are almost exclusively Goudplaas gneiss; however, some lavas, quartzite and granite are present in the north-west and extreme south-east (Figure 3-14a).

These patterns are also discernible when looking at the distribution of depth classes captured in the LTS (Figure 3-14b). In the extreme west of the catchment, the soils surrounding the Tzaneen dam are very deep (>1 200 mm), and as one moves to the east, the soils become progressively shallower. The LTS polygons also indicate that the most abundant soil classes (Figure 3-14c) that correspond to land types with deeper soils (>601 mm) are freely drained, structureless soils (soil class 2), and in one instance the soil was imperfectly drained and often shallow and had a plinthic

horizon (soil class 8). The second most abundant soil type (Figure 3-14d), occurring where deeper soils (>601 mm) are more prevalent, can be classified into 11 different soil classes. Shallower soils (< 601 mm), on the other hand, are dominated by lithosols (soil class 13) and lesser amounts of soils with pedocrutanic horizons (soil class 7). The second most abundant soils occurring in LTS polygons with overall shallower soils (<601 mm) are almost exclusively freely drained, structureless soils (soil class 2).

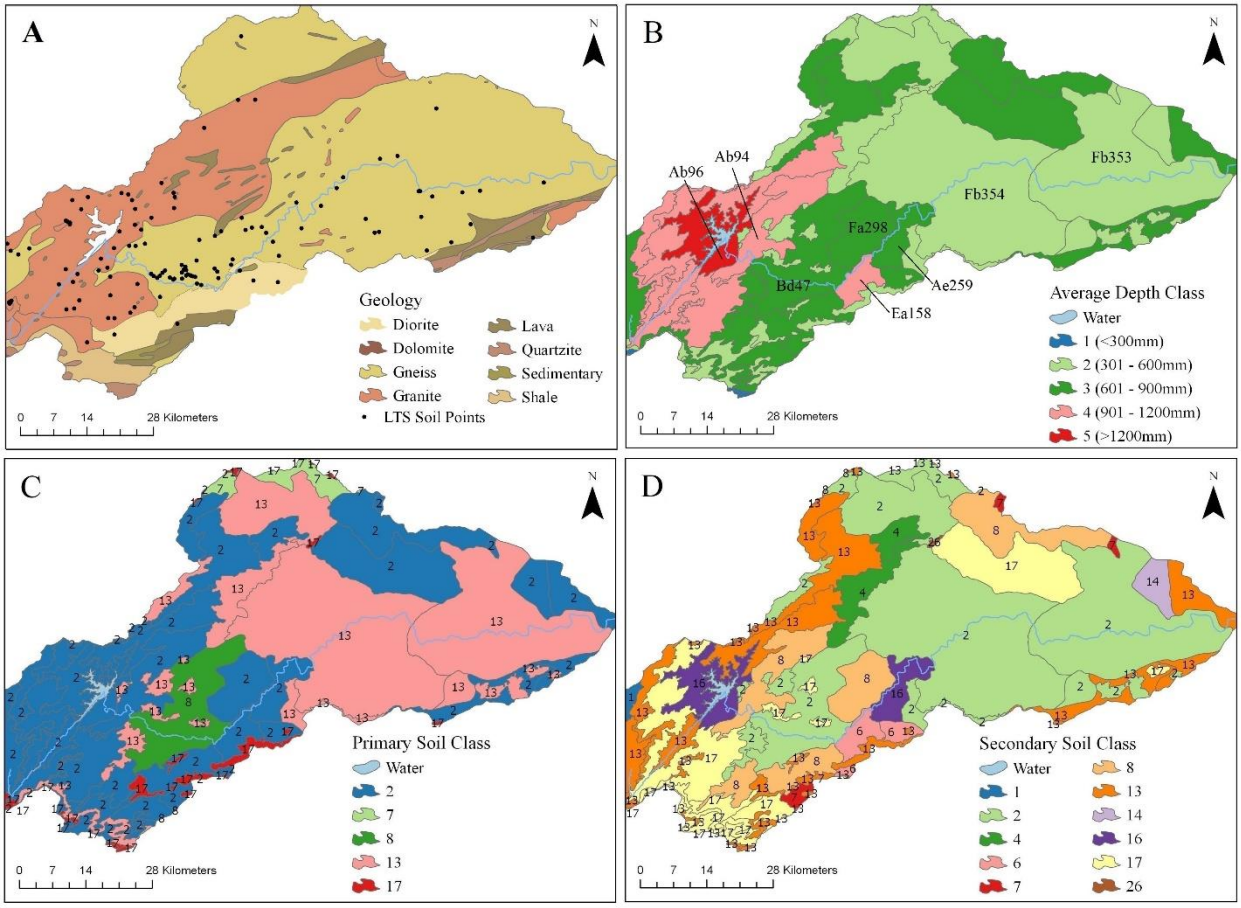


Figure 3-14 a) Geological map showing all dominant lithologies and the positions of the LTS soil profiles; b) average depth classes of all the GLWC land type polygons; c) primary soil classes of the all the GLWC land type polygons; and d) secondary soil classes associated with all the GLWC land type polygons

The LTS also estimates the probability of specific soil types occurring at each of the five landscape positions (i.e. crest, scarp, mid-slope, foot-slope and valley). Table 3-2 lists the dominant soil types estimated at each the five landscape positions for six land types selected along the GLR (Figure 3-14b). Particular attention is given to the mid-slope and foot-slope positions, as this is where terraces and planation surfaces most likely occur. Immediately next to the escarpment, the land types Ab96 and Ab94 both encompass substantially deep (>1 200 mm) Hu soils (class 2) on every landscape position. The land types Bd47, Ea158 and Ae259 cover significant portions of the western and south-western part of the GLR and are predominantly covered by mid-slope and foot-

slope landforms (between 70% and 90% of the land type surface). Bd47 covers a large proportion of the landscape to the south and north of the GLR and consists mostly of imperfectly drained Cf and Lo soils (class 8). Ea158 and Ae259 delineate almost exclusively class 2 Hu and Sd soils, respectively. Fa298 is located on the northern bank of the GLR, next to Ea158 and Ae259, and contains a mixture of class 2 (Hu and Oa), class 8 (Cf) and class 13 (Gs) soils. 75% of the Fa298 land type surface is covered by mid-slope and foot-slope surfaces. Finally, Fb 354 and Fb353 cover most of the eastern part of the catchment and are identical in soil composition and distribution. Significant proportions of class 13 soils (Ms and Gs) are found on crests and mid-slopes of the land surface, which cover around 26% and 38% of the land type surface, respectively. Both the mid-slopes and foot-slopes, covering 24% of the land type surface, also contain substantial amounts of class 2 Hu, Sd and Oa soils.

Table 3-2 Dominant soil types occurring at each landscape position for six land types selected within the GLWC

Land type	Crest	Scarp	Mid-slope	Foot-slope	Valley
Ab94	Hu, Sd		Hu, Sd	Hu, Sd	Ka, Hu
Ab96	Hu		Hu	Hu	Streambed, Hu
Bd47	Cf, Lo		Cf, Lo	Cf, Lo	Streambed, Va
Ea158				Sd, Bo	Streambed
Fa298	Cf, Hu, Gs		Cf, Hu, Gs	Oa, Cf	Oa, Streambed
Ae259	Hu		Hu	Hu	Streambed, Hu
Fb354	Ms, Gs, Rock	Rock, Ms	Gs, Hu, Sd	Gs, Sd	Oa, Streambed
Fb353	Ms, Gs, Rock	Rock, Ms	Gs, Hu, Sd	Gs, Sd	Oa, Streambed

The relationship between the land surface and soil types/classes was further investigated by evaluating the relationship between slope gradient and soil type/class. Figure 3-15 describes the slope values associated with each soil type and class. It is clear that the first soil class (Ia) is mostly associated with higher slope gradients, whereas the second soil class (Cv, Gf, Hu, Oa and Sd) occurs on terrain with both very low and very high slope gradient values. This indicates that class 1 and 2 soils are associated with steep slope gradient landforms located along the escarpment; however, class 2 soils are also found on shallow slope landforms, e.g. fluvial terraces and floodplains. Soil classes 3 to 15 correspond mostly to low slope gradient landforms associated with the eastern GLWC.

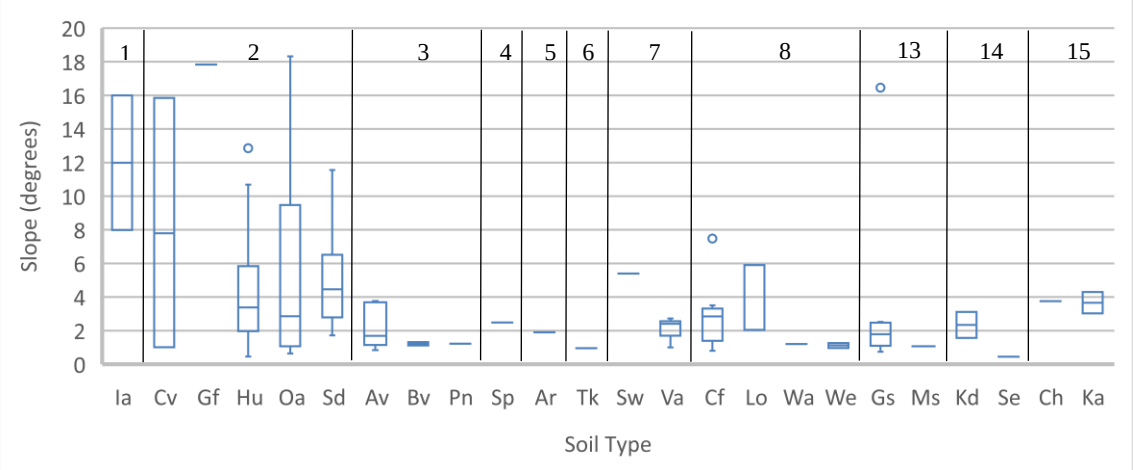


Figure 3-15 Boxplot indicating the relationship between slope gradient and various soil types occurring within the GLWC. The soil types are grouped and labelled according to their respective soil classes by the black vertical lines and corresponding small black numbers on the top of the graph.

3.5 DISCUSSION

Although the Kaapvaal craton is characterised by present-day tectonic stability, the presence of fluvial terraces in the Lowveld is indicative of previous tectonic uplift, river rejuvenation and increased river incision (Dollar 1998; Partridge et al. 2010; Venter & Bristow 1986). The GLWC was selected as a study area to characterise the geomorphology and soils of fluvial terraces and planation surfaces occurring in the northern Lowveld. Synthesis and discussion of the above results are detailed below and serves to: 1) demonstrate the extensive distribution of fluvial terraces and planation surfaces; 2) identify a clear relationship between fluvial terrace morphology, soil classes, local climate and underlying geology; and 3) highlight the complex landscape evolutionary cycles driven by river incision, fluvial terrace formation and land surface planation.

The interpreted cross-profiles positioned near the escarpment (Figure 3-3 and Figure 3-4) show several steps of unpaired well-preserved fill terraces present immediately next to the GLR and higher-lying terrace surfaces that are often heavily dissected by fluvial terrace-producing tributaries.

Very deep (>901 mm) Hu and Sd soils (soil class 2) are widespread in the western-extreme of the catchment surrounding the Tzaneen dam, where substantial annual rainfall coincides with landscapes underlain by mostly granite (Ab96 and Ab94) and diorite (Ea158). Land types that are both near the escarpment and underlain by mostly gneiss (Bd47, Fa298 and Ae259) delineate comparatively shallower but still deep (601 – 900 mm) Cf and Lo (soil class 8) and Hu and Sd (soil class 2) soils. The transition to shallower soils may be ascribed to both the transition from

granite to gneiss bedrock and to the progressive eastward decline in annual rainfall. Furthermore, the narrower valleys associated with the escarpment constrain fluvial terrace size laterally and drive the formation of very deep to deep soils.

Further east, wider and flatter valleys associated with the underlying gneiss bedrock produce more laterally extensive terraces with shallower soils. Two land types (Fb354 and FB 353) were selected to investigate the fluvial terraces of the eastern GLWC. Both land types are underlain by gneiss bedrock and covered by soil of moderate average depth (301 – 600 mm). The LTS also shows significant amounts of class 13 soils (Ms and Gs) and bare rock present on the crests, scarps, mid-slopes and toe-slopes of Fb354 and Fb353. This indicates that shallow soils are present throughout most of the landscape and validates the premise that fill terraces are less prominent near the KNP. Nevertheless, the well-preserved fluvial terrace surfaces identified (Figure 3-5, Figure 3-6, Figure 3-7 and Figure 3-8), together with the Hu and Sd (class 2) soils found on the mid-slopes and foot-slopes of Fb354 and Fb353, show that fill terraces are present immediately next to the GLR. Terraces located above the first three or four terrace surfaces are seldom preserved and are in most cases eroded beyond their original strath surfaces. Still, flat surfaces representative of fluvial terrace morphology have been identified significant distances from the river and even near the catchment boundary (Figure 3-9). The result of widespread denudation is evident throughout the entire GLWC. Along the GLR, fluvial terraces of all ages have been eroded by tributaries and are often incised beyond their strath surfaces. A history of sequential tributary incision events is apparent, given the presence of fluvial terrace staircases flanking tributaries. Furthermore, the development of tributary drainage networks resulted in the incision of most fluvial terrace surfaces, producing an “island” effect where only remnants of terrace surfaces are still visible.

This “island” phenomenon is also apparent when interpreting planation surfaces. Partridge & Maud (1987) postulated that planation activity in the northern Lowveld occurred primarily during the African and post-African I erosional cycles and that these surfaces have, since then, been heavily dissected, leaving only remnants of the original surfaces preserved as interfluves (peaks between two river valleys). Figure 3-12 and Figure 3-13 reaffirms the findings of Partridge & Maud (1987), as peaks in the terrain have been found to correlate at specific HARC along several adjacent cross-profiles, thereby enabling potential future extrapolation of the original planation surfaces. Furthermore, it was found that the African and post-African I cycles may have produced planation surfaces in a staircase-fashion, where individual surfaces often correlate with and may have been influenced by fluvial terraces in the landscape at the time of the planation surface

formation. The presence of fluvial terraces in the landscape at the time when the African and post-African I cycles were actively eroding is corroborated by evidence of fluvial terraces occurring up to catchment boundary and well above the elevation of any identified planation surface remnants.

Each of the planation surfaces identified by this study (Figure 3-12 and Figure 3-13) is heavily incised by several tributaries, each producing their own staircase of fluvial terraces. Some of the terrace surfaces positioned next to the tributaries have also been incised by third-order fluvial terrace-producing tributaries. It is likely that the dissection of the GLWC planation surfaces and consequent formation of nested fluvial terrace staircases were driven by the same tectonic and climatic factors that continually result in the formation of fluvial terraces next to the present-day GLR. This process of reworking older fluvial terraces into new second- and third-order fluvial terraces, coupled with the active incision of the GLR and its tributaries into underlying bedrock, support the Rozanov et al. (2017) hypothesis that localised planation activity dominates the evolution of the GLWC.

3.6 CONCLUSION

This study reviewed the discourse on the paleo-fluvial landscape evolution of the northern Lowveld and contributes new knowledge about the soil-geomorphology relationship of the GLWC, a typical river-shaped landscape of the northern Lowveld. The complex morphological setting of the GLWC, shaped by a sequence of erosion cycles, was characterised by describing the extensive fluvial terrace staircases of the GLR and its tributaries. Fluvial terraces were found to fall within clear elevation ranges with distinct sediment/soil characteristics relating to the sequence of land types associated with the river course. While there is no direct correspondence between the terrace level and unique soil type, the soils associated with fill and strath terraces are quite unique. The soils associated with individual terrace levels differ between the land type, climatic and geomorphic subdivisions of the Letaba valley.

All the terraces were found to be dissected to various extents, and most of the older terraces have been eroded up to and beyond their strath surfaces. As a result, most terraces and planation surfaces are preserved as remnants that are visible “islands”, or interfluves, in the topography. Field reconnaissance, geomorphometrical terrain study and analysis of the South African LTS database demonstrated the significant influence of localised climatic differences (driven by higher rainfall near the escarpment) and changes in underlying geology on the formation of fluvial terraces and their soils along the GLR. Evidence of nested fluvial terrace hierarchies produced by incising

tributaries confirms that the evolution of the GLWC landscape and soils is dominated by cyclical events of fluvial terrace formation and denudation. Further work is needed to create fine-scale geomorphological maps delineating fluvial terraces to facilitate land use optimisation and to support sustainable development of land resources in order to mitigate land degradation and desertification.

CHAPTER 4: OBJECT-BASED LAND SURFACE SEGMENTATION: AN ILL-STRUCTURED PROBLEM

4.1 ABSTRACT

The segmentation of continuous land surfaces into morphologically representative objects has received increasing attention in recent years. Multi-resolution segmentation (MRS) has been shown to delimit morphological boundaries accurately, and the use of unsupervised data-driven local variance (LV) based methods for detecting characteristic levels of scale parameter (SP) in land surfaces has been established. However, whether the detected SPs accurately delimit target morphological features is unclear. This study illustrates that multi-resolution land surface segmentation SP optimisation is an ill-structured problem that can be divided into subsets of well-structured problems by defining "conceptual" morphometric primitive (henceforth referred to as morphometric primitive) conditions.

A new methodology is proposed where an ensemble of unsupervised data-driven LV-based SP optimisation techniques are implemented to evaluate objects against each of the morphometric primitive conditions. To construct an ensemble of SP optimisation techniques, an established method, estimation of scale parameter 2 (ESP 2), is reviewed and existing LV concepts expanded to include two new SP optimisation techniques, namely: object boundary local variance (OBLV) and local variance ratio (LVR). Agreement between the different SP optimisation techniques are indicative of SPs where the probability that morphometric primitives are delimited is higher than when applying SPs selected by single SP optimisation approaches.

4.2 INTRODUCTION

Land surface segmentation involves delimiting genetically and morphologically homogenous geomorphic features from continuous terrain (Minár & Evans 2008). Two strategies have been suggested to achieve this (Drăguţ & Eisank 2011). The first relates to specific geomorphology as it concerns the delimitation of predefined elementary forms based on fitted polynomial functions and varying degrees of homogeneity (Minár & Evans 2008). Elementary forms are defined by Minár & Evans (2008, p244) as features "with a constant value of altitude, or of two or more readily interpretable morphometric variables, bounded by lines of discontinuity". These forms cannot be divided into smaller components. Landforms consist of sequential combinations of elementary forms that can be combined in a deterministic manner to form land systems (Minár &

Evans 2008). The second strategy is data-driven and relates to discrete geomorphometry (Drăguţ & Eisank 2011) as it utilises local variance (LV) as a measure of homogeneity to detect characteristic levels of scale in the landscape (Drăguţ et al. 2009; Drăgut, Eisank & Strasser 2011; Drăguţ, Tiede & Levick 2010). In discrete geomorphometry, the term morphometric primitives are often employed to refer to objects that are "carriers of information on land-surface parameters" at multiple levels of scale and objects that can be classified into elementary forms (Drăgut, Eisank & Strasser 2011: 163). The identification of characteristic landscape scale levels is therefore based on the principle that the landscape can be represented as a complex structure of nested hierarchies made up of morphometric primitives.

The separation of the intrinsic hierarchical structure of the landscape into discrete levels of organisation and morphologically representative objects can be achieved by performing multi-scale land surface segmentations. Van Niekerk (2010) evaluated segmentation algorithms for land surface segmentation and concluded that MRS significantly outperformed the other algorithms as it generated the most morphologically representative features. MRS is a region-merging algorithm that creates homogenous objects based on a user-defined scale parameter (SP) (Baatz & Schäpe 2000). SP is primarily used to control the mean size of the objects by minimising the average heterogeneity of the objects, normalised by the number of pixels included in each object. As a result, SP can be considered the primary MRS parameter in geomorphometric studies (Drăgut, Eisank & Strasser 2011), while other MRS parameters (such as shape and compactness) are often set to zero (Drăguţ & Eisank 2012; Eisank, Smith & Hillier 2014).

Estimation of scale parameter (ESP) is a popular data-driven unsupervised SP optimisation technique that investigates LV at different SP intervals (Drăguţ et al. 2014; Drăguţ et al. 2009; Drăguţ, Tiede & Levick 2010) and has been implemented in various geomorphometric (Drăguţ et al. 2009; Drăguţ & Eisank 2012; Drăgut, Eisank & Strasser 2011; Gerçek, Toprak & Strobl 2011) and digital soil mapping studies (Dornik, Drăguţ & Urdea 2017; Drăguţ & Dornik 2013). Woodcock & Strahler (1987) first introduced LV for the investigation of scale dependencies in the spatial structure of remotely sensed imagery, and Kim, Madden & Warner (2008) introduced the concept to geographic object-based image analysis (GEOBIA). Woodcock & Strahler (1987) quantified LV by calculating the mean standard deviation for all pixels in the image as the pixel size was systematically increased. By taking advantage of the inherent spatial autocorrelation, a high LV value indicates that the pixel size approximates the size of the real-world features.

ESP is based on the premise that a systematic increase in SP will result in objects that will, at some point, approximate real-world features, and that object boundaries will be preserved and object standard deviation will remain constant. The collective effect of the constant object standard deviations at increasing SP values will reflect on the LV of the image (Drăguţ, Tiede & Levick 2010), which can then be used to select SPs at which the probability that the generated objects will match real-world features will be higher than at the unselected SPs (Drăgut, Eisank & Strasser 2011). A second version of the tool (ESP 2) allows for multi-layered inputs (Drăguţ et al. 2014).

Mashimbye, De Clercq & Van Niekerk (2014) extended the concept of LV to quantify how well object boundaries follow morphological discontinuities in slope gradient datasets. Morphological discontinuities are natural boundaries of geomorphic features that may result from morphogenic processes or contrasts in regional geology (Minár & Evans 2008). By assuming that terrain variation will be higher at the location of a morphological discontinuity (compared to object interiors), the ratio between the object boundary LV (OBLV) and the LV of the interior of each object is calculated. Using this approach, a high LV ratio (LVR) indicates where the internal homogeneity of the object is maximised and the homogeneity at the edge is minimised, resulting in homogenous objects bordered by morphological discontinuities. Although OBLV and LVR have been applied for evaluating land surface segmentations at specific SPs (Mashimbye, De Clercq & Van Niekerk 2014), they have, to our knowledge, not yet been applied for land surface segmentation SP optimisation.

The most significant difference between the specific and data-driven land surface segmentation strategies is that in the former approach the characteristics of elementary forms are known prior to segmentation, whereas the characteristics of morphometric primitives at meaningful levels of scale are unknown in data-driven approaches. Morphometric primitives, delimited at scales derived from data-driven SP optimisation techniques, are typically assigned to analyst-defined thematic classes using either deterministic (rule-based) or machine learning (supervised) classifiers (Anders, Seijmonsbergen & Bouten 2011). Despite the reported favourable accuracies of these classification approaches, it is not known whether the SPs (implemented during the segmentation process) accurately delimit the morphological features represented by the user-defined classes. Accordingly, the data-driven selection of SPs is only useful for determining the scales present in a dataset and not necessarily for identifying the scales that are most appropriate for the application (Anders, Seijmonsbergen & Bouten 2011). This discrepancy can be ascribed to SP optimisation being an ill-structured problem. Ill-structured problems are problems with many possible

solutions, solution paths or parameters (Kitchener & King 1981; Van Niekerk et al. 2016). The concept of ill-structured problems relating to other MRS parameters has also been noted in remote sensing applications (Gilbertson & Niekerk 2017).

Voss & Post (1988), citing (Reitmann 1965), argued that well-structured and ill-structured problems represent different points on a spectrum. By decomposing ill-structured problems into a set of well-structured problems, and by specifying the information relevant to the solution, ill-structured problems may become well-structured during the solving process (Simon 1973). Solving ill-structured problems often involves two steps, namely: the representation phase and the solution phase. The solution phase typically follows the representation phase and, accordingly, the representation largely dictates the solution (Voss & Post 1988). Reitmann (1965: 151) further states that "a problem situation (which) evokes a high level of agreement over a specified community of problem solvers … may be termed unambiguous or well-defined with respect to that community". In other words, the representation phase involves deconstructing the ill-structured problem into a series of well-structured problems (or conditions) which can each be evaluated by a problem solver. During the solution phase, a "community of problem solvers" is employed to evaluate the results (or outputs) against each well-structured condition in order to determine, based on agreement between the community of problem solvers, which solution best satisfies the well-structured conditions.

This study aims to demonstrate how the ill-structured nature of the data-driven SP optimisation problems can be addressed by applying an ensemble of SP optimisation techniques to evaluate MRS results against predefined morphometric primitive conditions. To illustrate the concept and to serve as a case study, this article describes the representation and solution phase to delimit "conceptual" morphometric primitives (henceforth referred to as only morphometric primitives). The conditions of our morphometric primitives are defined according to the criteria proposed by Mashimbye, De Clercq & Van Niekerk (2014), namely: 1) the object must be homogenous; and 2) it must be bordered by a discontinuity. In doing so, the ill-structured SP optimisation problem (i.e. which SP produces the most appropriate set of morphometric primitives) is decomposed into two well-structured problems (i.e. the two morphometric primitive conditions) that can be evaluated by an ensemble of SP optimisation techniques (i.e. community of problem solvers). ESP 2 will be used to evaluate each segmentation's output to determine at which SPs the first morphometric primitive condition is met. The second morphometric primitive condition will be evaluated by adapting OBLV and LVR for SP optimisation. Therefore, each SP optimisation

technique acts as a problem solver that aims to solve a well-structured subset of the ill-structured SP optimisation problem.

4.3 MATERIALS AND METHODS

4.3.1 Study area

A 119 km^2 (7 km x 17 km) area in the Cape Winelands District municipality in South Africa (Figure 4-1), approximately 40 km east of Cape Town, was selected for this study. The elevation of the chosen site ranges from 64 to 1 476 metres. The western part of the study area is dominated by flat and undulating hills consisting mostly of alluvial deposits. The land uses in this area include agriculture, urban developments and conservation. The eastern part of the study area, used for forestry and nature conservation, is dominated by mountainous terrain. The steep slopes and prominent cliffs of this area are made up of sandstone and quartzite, which form part of the Cape Fold Belt.

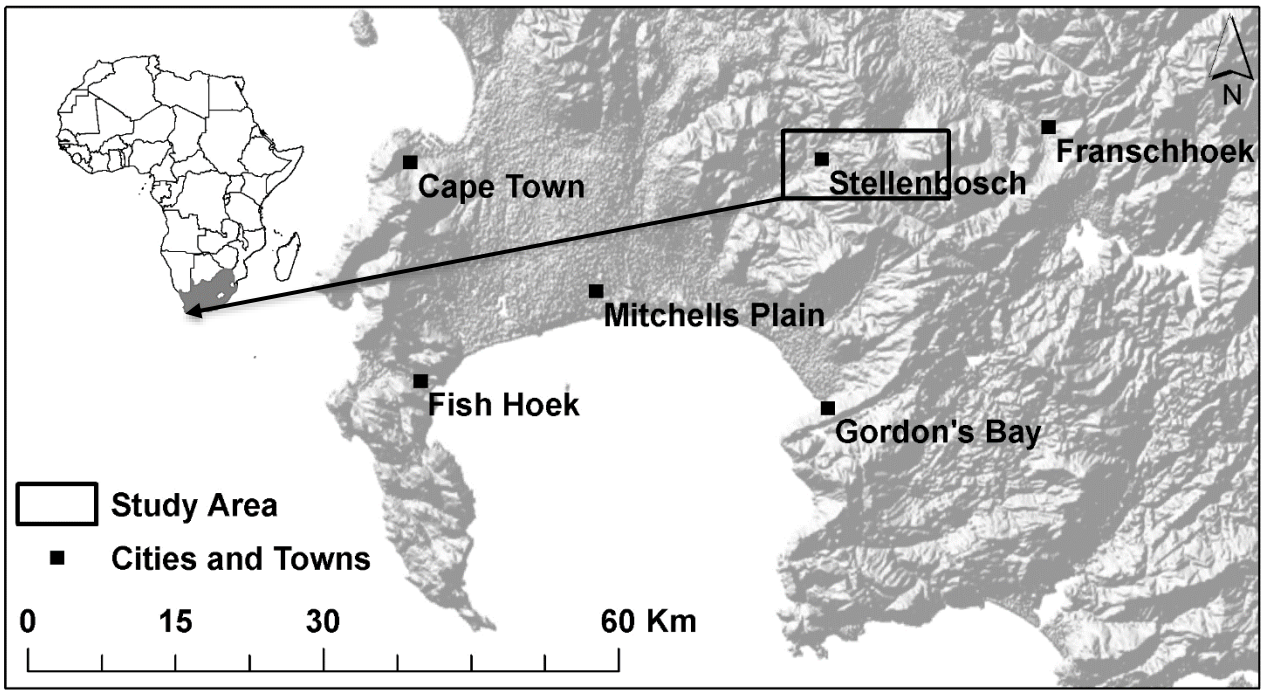


Figure 4-1 Location map showing the study area east of Cape Town in South Africa

4.3.2 Source DEM

The five-metre-resolution Stellenbosch University DEM (SUDEM) Level 2 was selected as input data for this study. The SUDEM was interpolated from large scale (1:10 000 and 1:50 000) contours and spot height data, supplemented with the 30 m SRTM DEM (Van Niekerk 2015a).

The Australian National University DEM (ANUDEM) algorithm was used for interpolating the contour and spot height data (Hutchinson 1989). The interpolated DEM was used to fill the voids in the SRTM DEM. Afterwards, the void-filled SRTM DEM and interpolated DEM were fused using a weighting scheme that favours the SRTM DEM in areas of moderate terrain (i.e. where contour and spot heights densities are low). This resulted in a DEM with a mean absolute error (MAE) of 2.1 metres and a root mean square error (RMSE) of 10.1 metres (Van Niekerk 2015a).

4.3.3 Land surface parameter generation

When delimiting complex land surfaces, morphological mapping necessitates a multi-land surface parameter (LSP) approach. However, for demonstration purposes and for the sake of simplicity, a single-LSP approach was used in this study. Slope gradient was selected as it is frequently used in object-based geomorphometric studies (Drăgut, Eisank & Strasser 2011; Mashimbye, De Clercq & Van Niekerk 2014). Slope gradient was calculated in ArcGIS 10.3 using a 3x3 kernel in which the maximum difference in elevation between the target cell and the eight neighbouring cells is calculated. Slope gradient was calculated using the maximum averaging technique (Burroughs & McDonnel 1998; ESRI 2016).

4.3.4 Techniques and algorithms

4.3.4.1 MRS

Land surface segmentation involves identifying morphological discontinuities that can be used to regionalise continuous terrain into homogenous morphometric primitives at specific scales (Drăgut, Eisank & Strasser 2011). This can be achieved with MRS, a bottom-up pairwise region-merging process to cluster pixels together based on decision heuristics and local homogeneity criteria (Baatz & Schäpe 2000; Trimble 2014a). Decision heuristics are used to determine which objects will merge at each segmentation step, and the homogeneity criterion is used to assign a degree of fitting (or merging cost) to each possible merge (Baatz & Schäpe 2000). The homogeneity criterion considers both spectral homogeneity and shape homogeneity. The analyst can optimise the homogeneity criterion with two parameters, namely shape and compactness.

The spectral homogeneity of the objects is controlled by the colour criterion, and the shape homogeneity corresponds to an object's deviation from a compact or smooth shape (compactness and smoothness criteria). The colour and smoothness criteria are both automatically derived from

the user-defined shape and compactness weights, where colour is equal to one minus the weight of shape, and smoothness is equal to one minus the weight of compactness (Trimble 2014b).

SP, the third user-defined MRS parameter, is used to set the maximum allowed standard deviation of the homogeneity criterion. SP thus acts as a homogeneity criterion threshold. The merge between the objects will be successful if the homogeneity criterion of two objects (defined by both the shape and colour criteria) is smaller than the user-defined SP. A high SP value will therefore result in a higher homogeneity threshold, bringing forth more merges and subsequent larger objects (Trimble 2014b). In other words, SP controls the object size and the ability of the segmentation procedure to accurately delimit different hierarchical levels in the landscape structure.

In this study, the shape and compactness MRS parameters were both set to zero, thereby specifying that the segmentation procedure must only consider colour when merging pixels (Drăguţ & Eisank 2012; Eisank, Smith & Hillier 2014). In land surface segmentation studies, the MRS colour criterion refers to the morphometric value (e.g. slope gradient values). Shape and compactness values above zero will attempt to generate smooth or compact objects, which may reduce its ability to delimit objects along subtle morphological discontinuities. The objects created with SPs larger than 100 were visually determined to be too large to be representative of the terrain and consequently SPs were only iterated for every integer from one to 100.

4.3.4.2 ESP 2

ESP 2 was implemented using default values throughout, except for the hierarchy option which was turned off so that the segmentations produced by ESP 2 matched those assessed by the other SP optimisation techniques. When LV is calculated within an object-based paradigm, it increases with an increase in SP in a semi-variogram like fashion (Kim, Madden & Warner 2008). According to Drăguţ, Tiede & Levick (2010), candidate SPs can be selected at positions where LV levels off (represented by a plateau in the LV graph). However, the LV graph can be difficult to interpret as the curve has a smooth shape (similar to that of a semi-variogram) with very limited variation. To mitigate this, Drăguţ, Tiede & Levick (2010) suggested using ROC (Equation 4-1) as it highlights LV dynamics across SP. ROC is a concept borrowed from the field of economics where it is used for evaluating changes in stock price over time (Bauer & Dahlquist 1999). Drăguţ, Tiede & Levick (2010) expressed ROC as:

$$ROC = \left[\frac{L-(L-1)}{(L-1)}\right] * 100$$ Equation 4-1

where L LV at the chosen SP; and
L - 1 LV at the preceding SP.

Peaks in LV-ROC indicate SPs at which object homogeneity matches those of real-world features and can thus be used to segment land surfaces at appropriate levels. This approach differs from the pixel-based LV approach where peaks in LV are selected (Drăgut, Eisank & Strasser 2011; Woodcock & Strahler 1987).

4.3.4.3 Adapting OBLV and LVR for land surface segmentation SP optimisation

OBLV aims to quantify whether object boundaries accurately delimit morphological discontinuities (Mashimbye, De Clercq & Van Niekerk 2014). The underlying principal of this method is centred on the pixel-based calculation of LV (standard deviation calculated within a 3x3 window) from LSPs, as first introduced by Woodcock & Strahler (1987). The pixel-based calculation of LV involves calculating the standard deviation of slope gradient within a 3x3 kernel. To compute OBLV, the LSP is segmented using MRS, after which the resultant objects are used to calculate mean LV values along both the inner and outer edges of each object (Equation 4-2). Finally, OBLV is calculated for each segmentation output (i.e. set of objects produced by a single set of segmentation parameters) as the mean of the OBLV of all the objects (Equation 4-3). A high OBLV value indicate a segmentation that delimits morphological discontinuities more accurately than segmentations that produce low OBLV values (Mashimbye, De Clercq & Van Niekerk 2014).

$$OBLV_{Object} = \frac{\left[\frac{\sum Std\ Dev\ (inner\ border)}{Number\ of\ pixels\ (inner\ border)}\right] + \left[\frac{\sum Std\ Dev\ (outer\ border)}{Number\ of\ pixels\ (outer\ border)}\right]}{2}$$ Equation 4-2

where Std Dev Standard deviation value calculated for each pixel using a 3x3 kernel

$$OBLV_{Segmentation} = \frac{\sum(OBLV_{Object})}{Number\ of\ objects}$$ Equation 4-3

A LVR for each object is calculated by dividing the OBLV of each object by the object interior LV (Mashimbye, De Clercq & Van Niekerk 2014), as illustrated in Equation 4-4. A morphometric primitive will have a low object interior LV and a high object boundary LV, resulting in a high LVR. Finally, the mean of all the object LVRs is obtained to calculate a LVR value for each segmentation output (Equation 4-5). High LVR values are the result of objects that have homogenous interiors and follow morphological discontinuities (Mashimbye, De Clercq & Van Niekerk 2014). Like LV graphs, the identification of suitable SPs from the OBLV and LVR graphs can also be facilitated using ROC.

$$LVR_{Object} = \frac{\sum(OBLV_{Object})}{\left[\frac{\sum Std\ Dev\ (object\ interior\ pixels)}{Number\ of\ pixels}\right]}$$

Equation 4-4

where *Std Dev* Standard deviation value calculated for each pixel using a 3x3 kernel

$$LVR_{Segmentation} = \frac{\sum(LVR_{Object})}{Number\ of\ objects}$$

Equation 4-5

4.4 RESULTS

4.4.1 LV

Following Drăguţ, Tiede & Levick (2010), LV values were obtained and used to calculate the LV-ROC for each successive segmentation output. To select candidate SPs from the LV graph (Figure 4-2) proved difficult, as the smooth appearance of the graph obscures relevant information. Consequently, only peaks in the LV-ROC were considered. This resulted in 29 candidate SP values, ranging between 11 – 97.

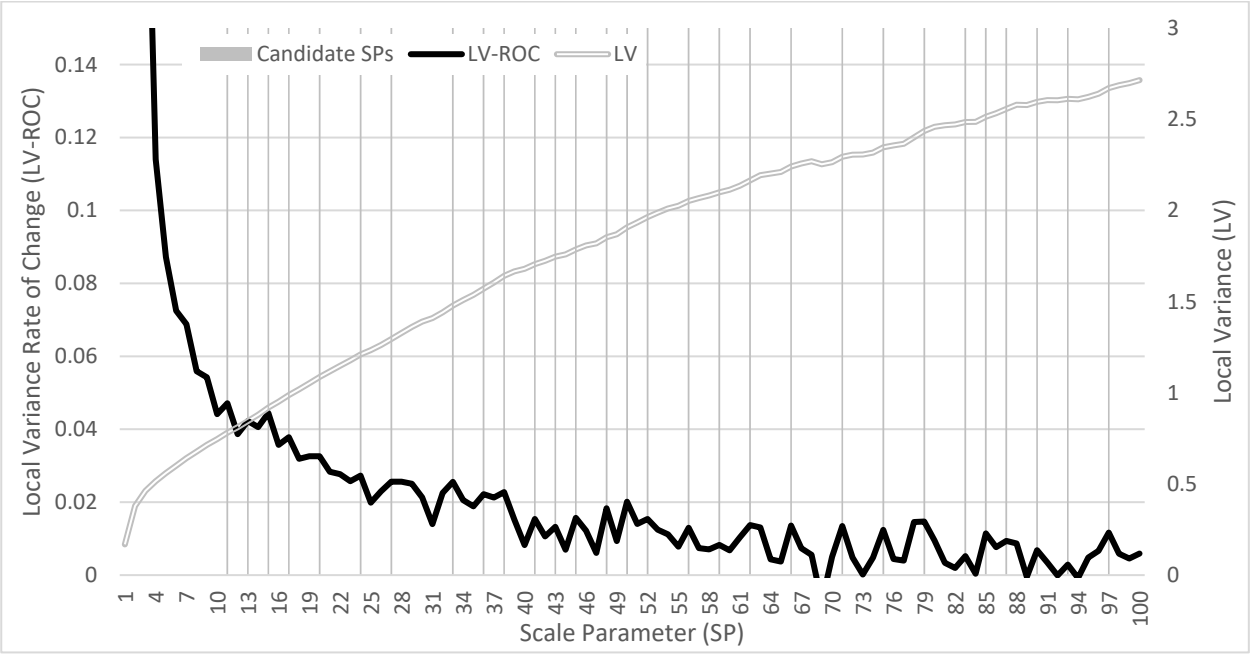


Figure 4-2 LV and LV-ROC values produced using the ESP 2 toolbox. The vertical lines represent the candidate SPs selected by identifying peaks in LV-ROC.

4.4.2 OBLV

The calculation of LV along the generated object boundaries (resulting in OBLV) serves to quantify terrain variation to determine whether the object boundaries fall on morphological discontinuities. Peaks in OBLV – when OBLV is plotted across increasing SP values – indicate SPs where morphological boundaries are delimited. Two approaches were followed to select candidate SPs from OBLV values. First, peaks in OBLV (Figure 4-3) were identified with the aid of a conditional (if-then) statement, which yielded a total of 16 candidate SPs. Next, OBLV-ROC (Figure 4-4) was used to select another 31 candidate SPs. Eight SPs (7, 39, 41, 43, 46, 64, 83 and 91) were duplicated in using these two approaches.

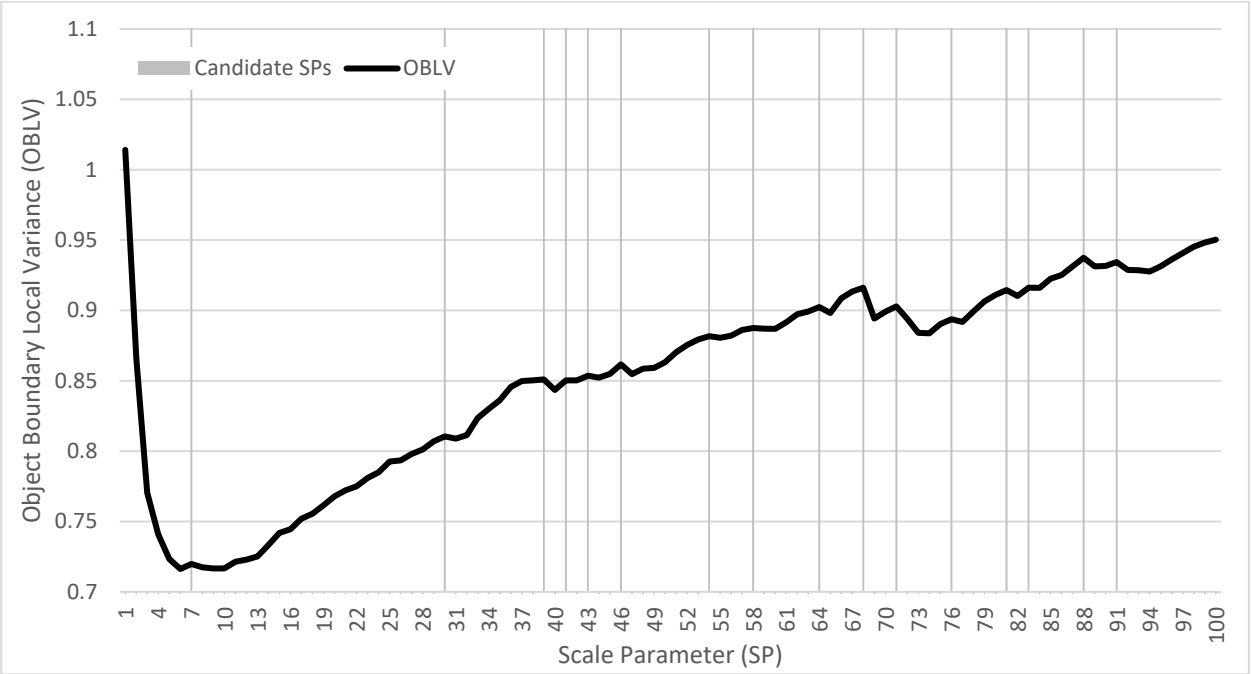


Figure 4-3 OBLV of each segmentation output plotted against increasing values of SP. The vertical lines represent the SPs selected using OBLV peaks.

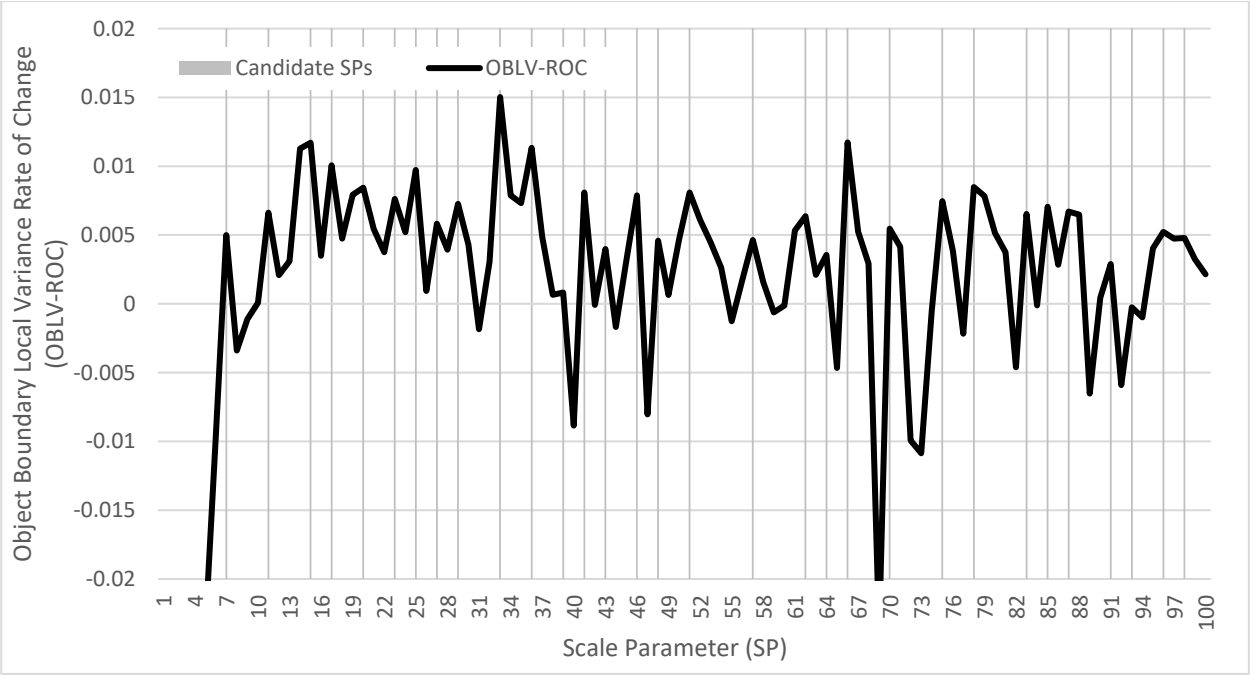


Figure 4-4 OBLV-ROC plotted against increasing values of SP. The vertical lines represent the SPs selected using OBLV-ROC peaks.

4.4.3 LVR

The calculation of LVR assumes that the variation in terrain along morphological boundaries will be greater than that of landform component interiors. Peaks in LVR are indicative of SPs where both conditions of morphometric primitives are satisfied. As with LV (Figure 4-2), LVR forms a smooth semi-variogram-like curve (Figure 4-5), which can make SP optimisation difficult. Nevertheless, with the help of a conditional (if-then) statement, eight candidate SPs were selected. Peaks in the LVR-ROC were more pronounced and 29 candidate SPs were identified (Figure 4-6). Two of these SPs (81 and 95), identified with the use of both approaches, matched.

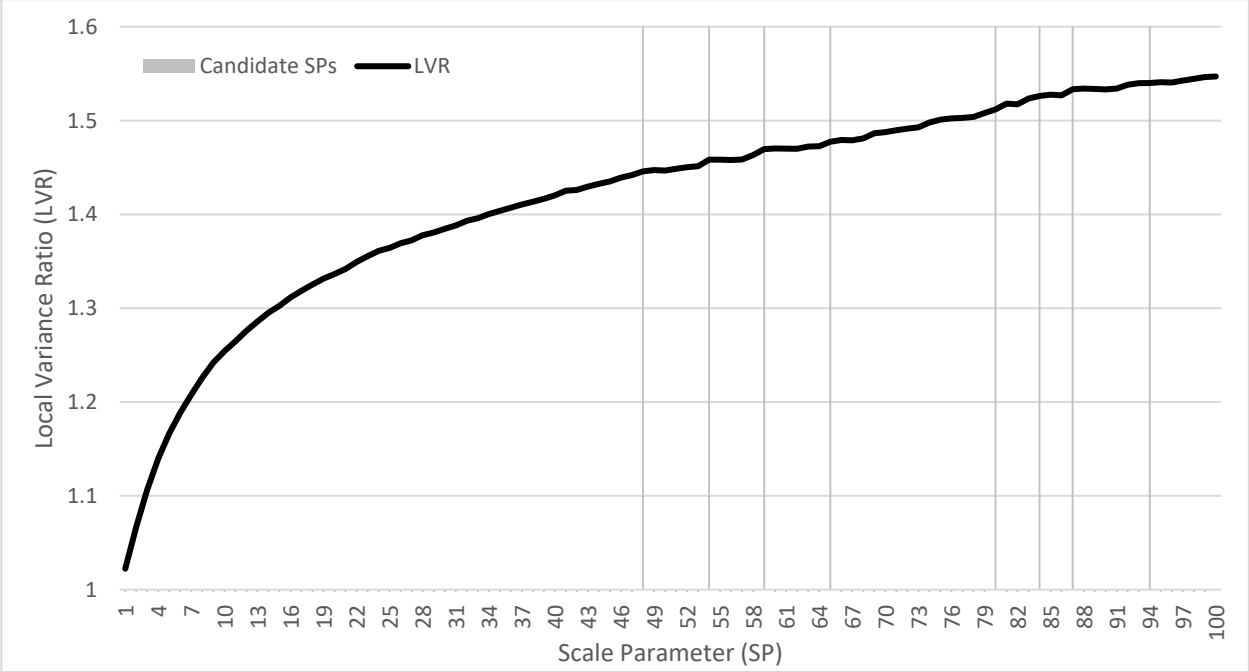


Figure 4-5 LVR of each segmentation output plotted against increasing values of SP. The vertical lines represent the candidate SPs selected using LVR peaks.

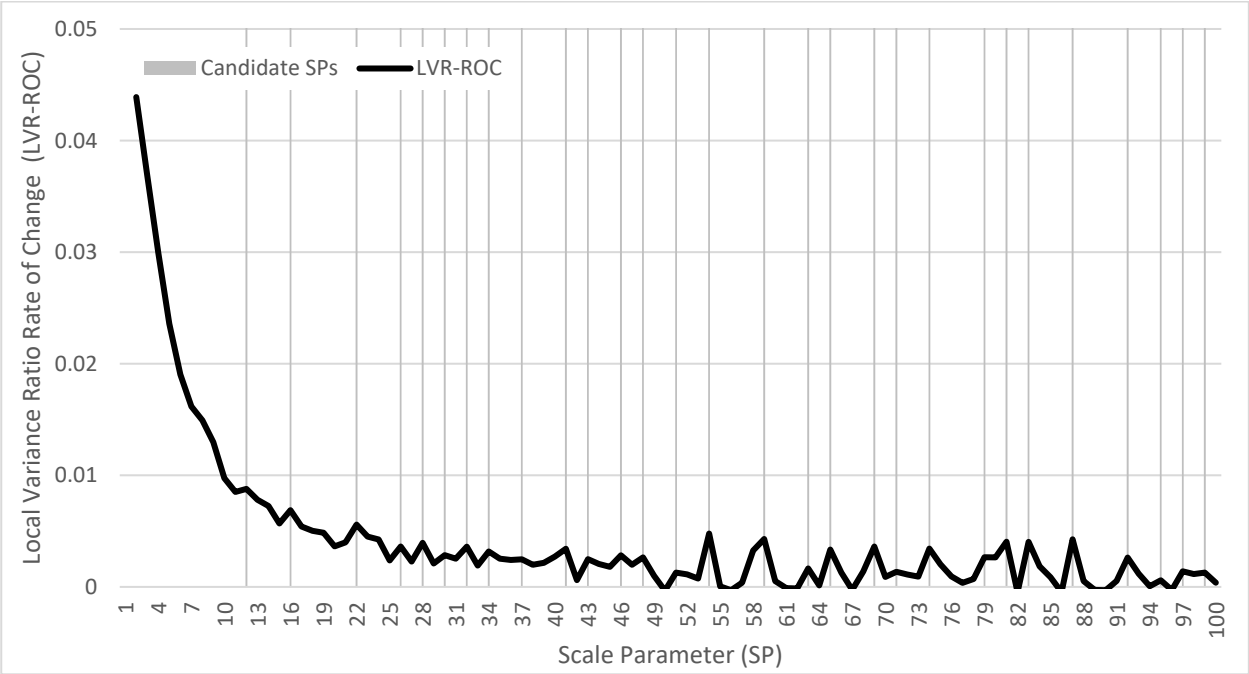


Figure 4-6 LVR-ROC plotted against increasing values of SP. The vertical lines represent the candidate SPs selected using LVR-ROC peaks.

4.4.4 Synthesis of results

To serve as a case study, two morphometric primitive conditions were defined. The condition of homogeneity was assessed with ESP 2 by calculating object LV for each consecutive SP. As a measure of the second condition, LV along object boundaries was calculated to compute OBLV and LVR to determine how well object boundaries follow morphological discontinuities. Agreement between the techniques are indicative of SPs where both morphometric primitive conditions are met (in this case study, morphometric primitives were thus defined as being internally homogenous in terms of slope gradient, with terrain discontinuities occurring at their edges).

By comparing the results generated by the three SP optimisation techniques, the extent to which consistent candidate SPs were generated can be evaluated. This was achieved by selecting the candidate SPs that were identified by the largest number of SP optimisation technique outputs (absolute and ROC values considered). The absolute value of LV was not considered, since candidate SPs are selected at positions where LV levels off (represented by a plateau), which makes the selection of candidate SPs from LV graphs difficult (Drăguţ, Tiede & Levick 2010). When selecting candidate SPs using OBLV and LVR, however, both the absolute value and the ROC are useful, since SPs can easily be selected by identifying peaks in both variables.

All candidate SPs were compared through visual assessment of Figure 4-7 to evaluate SP optimisation technique agreement. Three candidate SPs (41, 43 and 83) were selected by four SP optimisation technique outputs, namely: LV-ROC, OBLV, OBLV-ROC and LVR-ROC. Conversely, 13 SPs were selected three times, 23 SPs twice, 39 SPs once and the remaining 22 SPs were not selected by any of the SP optimisation techniques. When comparing candidate SPs of only LV-ROC, OBLV-ROC and LVR-ROC, the number of selected candidate SPs increases from three to five (41, 43, 48, 83 and 87).

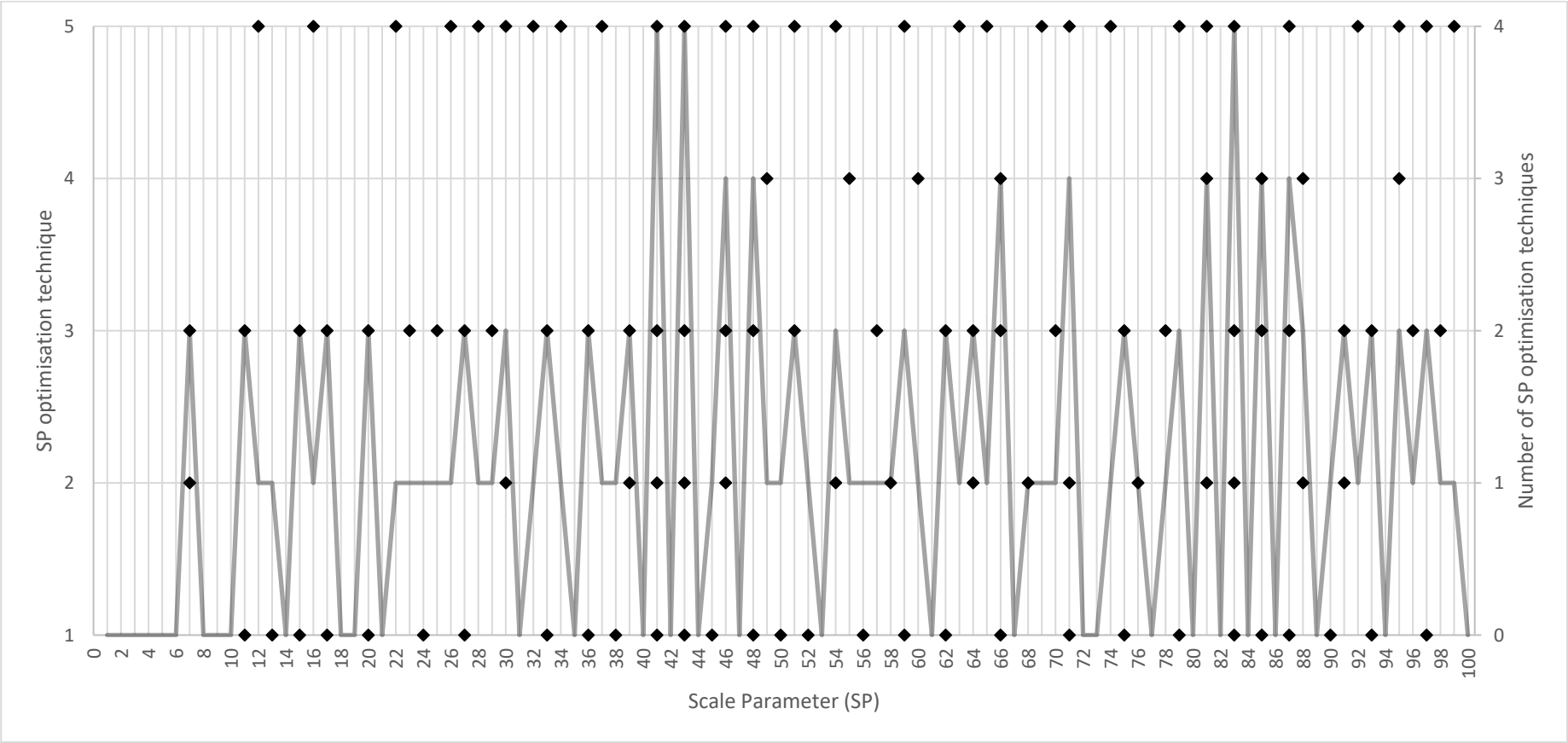


Figure 4-7 Compound graph showing: 1) candidate SPs selected by each SP optimisation technique (points) and 2) the number of SP optimisation techniques that selected each SP as a candidate (line). The primary y-axis (left) shows each SP optimisation technique, where 1 = LV-ROC (ESP 2); 2 = OBLV; 3 = OBLV-ROC; 4 = LVR; and 5 = LVR-ROC. The secondary y-axis (right) shows the number of SP optimisation techniques that selected each SP as a candidate.

4.5 DISCUSSION

The ill-structured nature of data-driven SP optimisation has been highlighted by conceptually juxtaposing the specific and data-driven land surface segmentation strategies. In essence, the problem is that data-driven SP optimisation techniques do not take the targeted land surface features into consideration. Instead, characteristic levels of scale inherent to the landscape structure are sought. Consequently, morphometric primitives delimited (and later classified) often do not adequately represent the morphological features represented by user-defined classes. We propose a new approach through which the ill-structured nature of SP optimisation is broken down into well-structured problems by defining morphometric primitive conditions prior to segmentation. In this manner, the domains of the data-driven and specific strategies intersect. In other words, by approaching data-driven SP optimisation as an ill-structured problem, morphometric primitives with predefined properties can be delimited.

The proposed approach is similar to selecting SPs based on the visual inspection of segmentation outputs as analysts tend to intuitively create conditions or scenarios that evaluate segmentation outputs against real-world land surface features. Instead of relying on an analyst's experience and judgement, a set of conditions is defined based on the specific characteristics of the targeted morphometric primitives. The segmentation outputs are then quantitatively assessed against these conditions to find suitable SPs, thereby eliminating human bias and error.

The results show that there was agreement between two or more SP optimisation techniques (or their ROC representations) for 39 of the 100 candidate SPs. Furthermore, after accounting for duplications between OBLV and OBLV-ROC (eight candidate SPs), and between LVR and LVR-ROC (two candidate SP), a remaining 31 candidate SPs were in agreement when considering all three implemented SP optimisation techniques. A meaningful relationship between LV, OBLV and LVR is therefore evident. However, the agreement between the different SP optimisation techniques are by no means universal, as 39 out of the 100 candidate SPs were selected by only one SP optimisation technique. Evidently, not all SPs at which objects are delimited with significant levels of either object interior homogeneity or object boundary heterogeneity result in objects that satisfy both morphometric primitive conditions.

The decrease in the number of agreeing SPs as the number of techniques increases shows that SP selection becomes more refined as more optimisation techniques are considered, which can be

ascribed to the solving process of ill-structured problems. This concept is further substantiated by the fact that two additional SPs (48 and 87) – additional to the three SPs (41, 43 and 83) selected by LV-ROC, OBLV, OBLV-ROC and LVR-ROC – were selected when absolute value graphs were not employed. In essence, for each SP optimisation technique that is added, more information relevant to the problem is specified, which results in a well-structured problem during the solving process (Simon 1973; Voss & Post 1988). This reaffirms that the solution is dependent on and dictated by the representation of the ill-structured problem (Voss & Post 1988) and points out the importance of the representation phase in the solving process.

Accordingly, when selecting SPs from the different sets of candidate SPs, the SPs selected by most of the SP optimisation techniques will most probably, compared to other candidate SPs, result in objects that conform to the well-structured conditions set out in the representation phase. Three candidate SPs (41, 43 and 83) were identified as most suitable, as each of the three was selected by LV-ROC, OBLV, OBLV-ROC and LVR-ROC.

Most studies (Drăgut et al. 2009; Drăgut, Eisank & Strasser 2011; Drăguţ, Tiede & Levick 2010; Mashimbye, De Clercq & Van Niekerk 2014) make use of a single SP optimisation technique, but this often results in large sets of candidate SPs from which the selection of a few SPs – adequately representing the morphometric primitives and often defined by thematic classification classes – is difficult. To illustrate, consider that a total of 29 SPs were selected from the ESP 2 ROC graph of which 15 SPs were separated by only a single SP. Through the implementation of the ensemble approach, this was reduced to three SPs, where two SPs (41 and 43) are separated by a single SP. These results show that by employing an ensemble of SP optimisation techniques, the number of selected SPs (selected based on agreement) decreases, which allows for the identification of a smaller selection of more suitable SPs (or at least discard many SPs from further consideration). The occurrence of selected SPs that are “very close”, i.e. separated by only one SP, in the refined selection still necessitates some user discretion, as these SPs do not represent significantly different outputs. Elimination of all occurrences of “very close” candidate SPs could possibly be achieved by further decomposing the SP optimisation problem into more suitable well-structured conditions against which each SP can be iteratively evaluated. New overarching landform definitions that can be used as well-structured conditions for ensemble SP optimisation should stem from ontological work in specific geomorphometry. The dependence of data-driven ensemble SP optimisation on specific geomorphometry (i.e. the dependence of the well-structured conditions on the

representation of the ill-structured problem) emphasises the convergence of the specific and discrete geomorphometry domains.

In this study, morphometric primitives were defined as being internally homogenous in terms of slope gradient, with their boundaries following terrain discontinuities. However, for some object-based geomorphological mapping tasks, the targeted land surface feature classes may be more complex and the conditions of our morphometric primitives defined here might be insufficient. Furthermore, whether the morphometric primitive conditions defined in this study, as well as the ensemble of SP optimisation techniques employed, delimit real-world land surface features (e.g. river terraces or drumlins) is still to be determined. More research is thus needed to investigate how the proposed ensemble SP optimisation approach can be used to delimit specific land surface features. Lastly, the automation of LV-based SP optimisation has been documented in past work (Drăguţ et al. 2014) and, since the ensemble SP optimisation approach detailed here is an extension of previous work (Drăgut et al. 2009; Drăgut, Eisank & Strasser 2011; Drăguţ, Tiede & Levick 2010; Mashimbye, De Clercq & Van Niekerk 2014), the process of automating the techniques and algorithms used in this study should be similar to that detailed by Drăguţ et al. (2014). The automation of ensemble SP optimisation techniques would provide immense value and support to the wider adoption of ensemble SP optimisation for land surface segmentation.

4.6 CONCLUSION

The accurate mapping of land surfaces involves segmenting continuous terrain into discrete, morphologically representative objects. Both specific and data-driven strategies have been developed (Drăguţ, Tiede & Levick 2010; Minár & Evans 2008). The specific strategy concerns the identification of predefined elementary landforms in the landscape, whereas the data-driven strategy aims to delimit morphometric primitives at characteristic scale levels in the landscape. Compared to other data-driven segmentation methods, MRS has been shown to generate object boundaries that represent morphological discontinuities (Van Niekerk 2010). The success of MRS to produce morphologically accurate objects depend on the selection of suitable SPs. The optimisation of SP, however, is an ill-structured problem with many possible solutions and solution paths.

By unpacking the complexities of land surface segmentation SP optimisation, this study investigated the ill-structured nature of SP optimisation and proposed a new methodology, based on an ensemble of SP optimisation techniques, to select candidate SPs. First, established methods

and techniques (MRS and ESP 2) were reviewed as well as existing concepts (OBLV and LVR) expanded. Next, as an example, the delimitation of morphologically representative morphometric primitives was decomposed into a set of well-structured problems by defining a set of conditions for morphometric primitives. Each of the SP optimisation techniques was implemented and agreement between candidate SPs evaluated. The results show that the implementation of multiple SP optimisation techniques refines the selection of candidate SPs; however, whether the identified SPs accurately delimit real-world land surface features is still to be determined. The main contributions of this study were to expose data-driven SP optimisation as an ill-structured problem and to propose that it is subdivided it into well-structured problems. The study provides a good foundation for further work that investigates the delimitation of predefined morphometric primitives.

CHAPTER 5: AN OBJECT-BASED MACHINE LEARNING APPROACH FOR DELIMITING FLUVIAL TERRACES

5.1 ABSTRACT

The bulk of available fluvial terrace delineation techniques are based on per-pixel geomorphometric principles that are insensitive to the scale-dependency of land surface parameters (LSPs) and scale specificity of landforms. Object-based discrete geomorphometry strategies, on the other hand, make use of land surface segmentation techniques to produce homogenous scale-dependent morphometric primitives and have been shown to delineate fluvial terraces more accurately.

This study employs white-box decision tree (DT) machine learners to: 1) investigate the potential of object-based discrete geomorphometry for fluvial terrace classification; and 2) evaluate the interpretability of DTs to construct transparent supervised rulesets. DTs are trained on reference datasets to classify binary (terrace and non-terrace) and multiple (10 ordinal terrace levels and non-terrace features) fluvial terrace classes. Interpretation of the resulting DT rulesets revealed that per-pixel approaches (validation scores of 0.83 and 0.77) are more conducive to overfitting and construct sets of convoluted rules that are difficult to interpret. In comparison, the object-based approaches (validation scores of 0.88 and 0.83) produced much simpler rulesets that correspond to existing fluvial terrace expert knowledge. We conclude that an object-based DT approach offers a promising methodology with which to accurately classify fluvial terraces and to potentially build rulesets that can be applied over large areas to produce fluvial terrace maps.

5.2 INTRODUCTION

Fluvial terraces, a global phenomenon, are formed when a river periodically abandons its floodplain by incising into underlying sediment or bedrock. Each abandoned floodplain is preserved as a terrace step in a sequence of terrace levels (Leopold, Wolman & Miller 1964). The importance of fluvial terraces relates to the deciphering of past tectonic uplift and climatic fluctuation events, which contribute significantly to their formation and to the mapping of arable soils (Bridgland & Westaway, 2008; Pazzaglia, 2013; Vandenberghe, 2015). The delineation of fluvial terrace surfaces can be achieved by employing any of numerous field mapping and land surface modelling strategies. Traditionally, field campaigns employ semantic definitions of fluvial terraces, e.g. continuous landforms with low gradients that occur next to a river channel (Clubb et

al. 2017), in situ surveying techniques to sketch cross-profiles of fluvial terraces at frequent intervals along a river course and aerial photographs to delineate potential terrace boundaries (Demoulin et al. 2007; Moon & Heritage 2001; Van Riet Lowe 1952). Traditional techniques are often subjective, influenced by analyst experience and dependent on additional information regarding stratigraphic discontinuities and physiographic criteria, as well as assumptions regarding terrace continuity and relative elevation (Leopold, Wolman & Miller 1964).

Geomorphometric land surface delineation techniques based on digital elevation model (DEM) analysis offer an alternative to traditional landform delineation approaches. Geomorphometric strategies include general geomorphometry, which aims to mathematically model land surfaces as a continuous scalar field, and specific geomorphometry, which involves the regionalisation of landscapes into discrete landforms. Another distinction between these strategies is that general geomorphometry is concerned with how landforms affect the flow of energy and mass, whereas specific geomorphometry aims to delineate discrete landforms according to the processes that formed them (Evans 2012; Swan 2017). Both semi-automated and automated methods for delineating fluvial terraces have been developed based on specific geomorphometry principles. The majority of these methods rely on per-pixel geomorphometric techniques.

For example, Demoulin et al. (2007) used a per-pixel deterministic fluvial terrace mapping approach, DEM segmentation and smoothed bivariate scatterplots – considering local slope and relative elevation – to identify low slope areas that are bounded by relatively higher sloped landforms. Although their methodology correctly identified 78% of the reference terrace remnants, the implementation of this methodology is dependent on user-facilitated parameterisation.

The TerEx toolbox, another per-pixel deterministic approach, was developed by Stout and Belmont (2014). The toolbox was implemented in nine different study areas and evaluated using a performance rubric. The results were deemed “excellent” for three of the study areas, “good” for five of the study areas and “poor” for one study area. The toolbox requires significant user parameterisation of local relief, minimum area and maximum distance from the channel, and iterative computation is needed. A visual comparison of the extracted fluvial terraces with field data is required to edit the produced vectors. Hopkins (2014) evaluated the TerEx toolbox along the Sheepscot River in Maine together with three other fluvial terrace delineation techniques, namely edge-detection using MATLAB, land surface feature classification (Wood 1996) and the Rahnis method (Walter et al. 2007). TerEx, land surface feature classification and the Rahnis

method were found to overestimate terrace surfaces, while the most accurate results were obtained in areas within confined river valleys with steep slopes. Edge-detection using MATLAB failed to delineate terrace boundaries in dissected topographies.

An automated per-pixel fluvial terrace and floodplain delineation methodology that selects slope and elevation above nearest channel thresholds with the use of quantile-quantile plots was developed and demonstrated by Clubb et al. (2017). By comparing the delineated fluvial terraces to field data from seven sites, an overall quality between 0.39 and 0.68 was obtained (quality was calculated as a ratio value between zero and one using the number of true positives, false positives and false negatives as a measure of "goodness"). Using the same method, Li et al. (2019) evaluated low-cost structure-from-motion photogrammetry-derived digital surface model data for fluvial terrace delineation by identifying three terrace steps. The authors noted that their results correlated well with terrace remnants identified in the field and suggested that future research focus on the object-based implementation of the automated methodology.

Swan (2017) evaluated the ability of five black-box machine learners (mahalanobis distance, winner-take-all, normal Bayes, random forest and support vector machine) to delineate fluvial terraces from DEM derivatives along the Buffalo River in Arkansas. The study showed that significant increases in accuracy were observed when both regional and local LSPs were used as input data, compared to when only local LSP input data are used, and that non-distance-based learners outperformed the distance-based classifiers.

Although the aforementioned per-pixel strategies have been shown to delineate fluvial terraces, they fail to incorporate the scale-dependency of LSPs and scale specificity of specific landforms (Drăgut, Eisank & Strasser 2011; Evans 2012). Progressive developments in object-based image analysis include the development of the multi-resolution segmentation (MRS) algorithm (Baatz & Schäpe 2000) and the subsequent introduction of various land surface segmentation scale optimisation techniques (Drăgut, Eisank & Strasser 2011; Drăguţ, Tiede & Levick 2010; Louw & Van Niekerk 2019). Unlike per-pixel approaches, object-based land surface segmentation allows for the calculation of LSP statistics at specific scales that can be used for fluvial terrace identification.

The term 'discrete geomorphometry' was introduced to nomenclatively classify methodologies that aim to delineate morphometrically representative objects (morphometric primitives) defined exclusively by scale thresholds (or homogeneity criteria) applied to sets of LSPs (Drăguţ & Eisank

2011). Various object-based landform mapping studies have evaluated discrete geomorphometric approaches, two of which include the delineation of fluvial terraces. Van Asselen & Seijmonsbergen (2006) delineated eight landform classes, including fluvial terraces, by segmenting LSPs derived from a one metre resolution LiDAR-based DEM into two scale levels using the MRS algorithm (Baatz & Schäpe 2000). Classification was achieved using fuzzy membership features based on zonal statistics derived from manually created training data. Sixty-nine per cent (69%) of the fluvial terraces were correctly classified with this approach.

Anders et al. (2011) delineated fluvial terraces, among seven other landforms, with a stratified land surface segmentation methodology. Fluvial terraces and floodplains were grouped into a single class and delineated using slope and topographic openness derived from a one metre resolution LiDAR-based DEM. A MRS scale parameter (SP) value of 15 was used and classification was performed with an expert knowledge defined ruleset. A user's accuracy of 67% was obtained by evaluating the classification with a limited number (nine) of samples.

Evidently, a diverse set of geomorphometric fluvial terrace delineation methodologies has been developed and applied, the bulk of which were performed in a per-pixel paradigm and are dependent on expert knowledge classifications. Although it is recognised that supervised machine learning may provide an alternative to knowledge-based classification approaches (Swan 2017), the adoption of supervised machine learners for geomorphometric mapping purposes has been slow when compared to related geospatial technologies such as remote sensing (Lary et al. 2016; Maxwell et al. 2018). The limited use of machine learners could be attributed to the black-box nature of most algorithms and the consequent exclusion of expert knowledge from the classification process. "Explainable" machine learning models, separate models that attempt to decipher black-box machine learners, have been suggested to address the need for transparent and accountable models. However, such models are often not reliable and seldom replicate black-box machine learners accurately. The scientific community habitually favours complex back-box machine learners for their reported high predictive accuracies, despite there being no real compromise between predictive accuracy and interpretability (Rudin 2019). In contrast, interpretable (white-box) machine learners may offer a way to combine expert knowledge and machine learner classifiers to construct rulesets that are both effective and transparent.

A DT, a well-understood white-box, is a non-parametric- and non-distance-based machine learning algorithm that can be used for both inductive and predicative purposes (Breiman et al.

1984; Quinlan 1986). In inductive studies, DTs are used to compute feature importances and construct rulesets that can be readily compared and evaluated against expert knowledge-based rules (Bou Kheir et al. 2010; Quinlan 1987). For predictive purposes, DT-derived rulesets can be employed to assign features to categorical classes based on inferred relationships between explanatory variables and categorical response variables. The predictive use of DTs is well-established in the fields of remote sensing (Laliberte, Fredrickson & Rango 2007) and geomorphometry (Bou Kheir et al. 2010; Stepinski, Ghosh & Vilalta 2006), but, to our knowledge, DTs have not yet been employed to evaluate and compare per-pixel or object-based fluvial terrace delineation approaches.

This study aims to 1) investigate the efficacy of object-based land surface delineation for classifying fluvial terraces; and 2) evaluate the interpretability of DT-derived rulesets to construct rulesets that can potentially be applied to map fluvial terraces over large areas. Accordingly, the feature importance of various LSPs for fluvial terrace classification are assessed, the qualitative and quantitative performance of per-pixel and object-based fluvial terrace classifications are compared and DT-derived rulesets are evaluated against expert knowledge and reference data.

5.3 MATERIALS AND METHODS

5.3.1 Study area

Fluvial terrace formation in the GLWC (Figure 5-1), located within the Lowveld Geomorphic Province in South Africa, was driven by early Miocene to late Pliocene tectonic uplift events, coupled with cycles of climatic change during the Pleistocene period, and continued well into the Holocene (Dollar 1998; Partridge et al. 2010; Partridge & Maud 1987). At present, GLWC fluvial terraces occur immediately next to the Great Letaba River (GLR) and extend all the way to the catchment boundary. Significant planation activity eroded most fluvial surfaces up to and beyond their original strath surfaces, leaving most remaining fluvial terrace surfaces preserved as low-lying interfluves (peaks between two valleys). Given the degree of planation in the catchment, the slope gradient of the terraced topography (including terrace and non-terrace surfaces) rarely exceed 7.3 degrees.

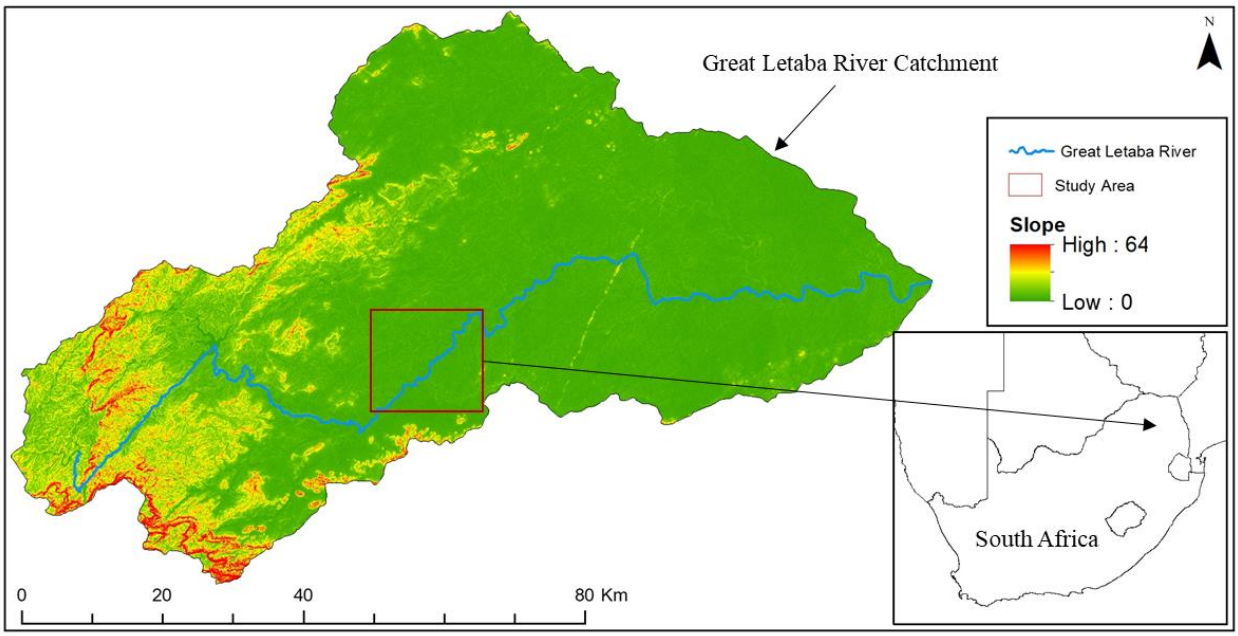


Figure 5-1 Location of the GLWC and the demarcated study area in South Africa

5.3.2 Source DEM and LSP generation

The Stellenbosch University DEM (SUDEM) was selected for generating the LSPs as it has a resolution of 5 metres, a mean absolute error of 1.77 metres and a standard deviation of error of 1.21 metres. The SUDEM was interpolated from both large scale contours and spot height data and was supplemented with the 30 m Shuttle Radar Topography Mission DEM (Van Niekerk 2015b).

Swan (2017) suggests that the combination of regional and local LSPs yields better classification accuracies when compared to classifiers based on only local LSPs. With this in mind, the System for Automated Geographical Analysis (SAGA) software package was used to generate 32 LSPs from the source DEM (Table 5-1).

Table 5-1 LSPs generated from the source DEM

1	Analytical hill shade	9	Longitudinal curvature	17	Negative terrain openness 250	25	Slope
2	Aspect	10	LS factor	18	Plan curvature	26	Elevation
3	Convergence index	11	Maximum curvature	19	Profile curvature	27	Terrain ruggedness index
4	Cross-sectional curvature	12	Minimum curvature	20	Positive terrain openness 25	28	Terrain surface convexity
5	Curvature classification	13	Morphometric features	21	Positive terrain openness 250	29	Terrain surface texture
6	Distance to river channel	14	Morphometric protection index	22	Relative slope position	30	Topographic wetness index
7	Flow accumulation	15	Multi-scale topographic position index	23	Vertical distance to channel network	31	TPI landform-based classification
8	Fuzzy landform element classification	16	Negative terrain openness 25	24	SAGA wetness index	32	Valley depth

5.3.3 Land surface segmentation

MRS has been shown to delineate morphological boundaries accurately (Van Niekerk, 2010). MRS is a bottom-up pairwise merging algorithm that groups pixels together based on decision heuristics and a homogeneity criterion (Baatz & Schäpe 2000). The homogeneity criterion incorporates both shape homogeneity and spectral homogeneity. Shape homogeneity is defined by the analyst-controlled shape and compactness parameters. Shape and compactness were both assigned a zero value to enable MRS to consider only morphometric values (spectral homogeneity) during the segmentation procedure (Drăguţ and Eisank, 2012; Eisank et al., 2014). SP, on the other hand, acts as a homogeneity criterion threshold by limiting the maximum standard deviation of the homogeneity criterion (Baatz & Schäpe 2000), thereby controlling the size of the resultant objects.

Using the eCognition software package, MRS was used to regionalise the GLWC landscape into homogenous morphometric primitives. For the purpose of this study, object homogeneity was prioritised over border heterogeneity (i.e. objects should rather be internally homogenous than follow morphological boundaries), and therefore SP was set to a value of five to ensure an over-segmented land surface. Clubb et al. (2017) showed that accurate DEM-based mapping of fluvial terraces can be achieved by extending the premise of field-based fluvial terrace mapping to identify fluvial terraces with slope gradient and vertical distance to channel network (VDTCN) thresholds. By the same logic, slope gradient and VDTCN were selected as MRS input data to delimit homogenous morphometric primitives representative of fluvial terraces. The eCognition chessboard segmentation algorithm was employed to create objects that are an exact match to pixels of the input LSPs in order to emulate per-pixel classification.

5.3.4 Field reconnaissance and reference data collection

Two field excursions provided the necessary reconnaissance data to support the collection of fluvial terrace reference data. The first excursion investigated the GLR and Little Letaba River fluvial terrace soil clay mineralogy to provide a better understanding of the formation of fluvial terraces and planation activity (Rozanov et al. 2017). The second excursion focussed exclusively on the geomorphology of the fluvial terraces along the GLR (Chapter 3).

A two-step process was followed to facilitate accurate selection of fluvial terrace reference data in the GLWC. First, MRS was implemented with a SP value of five to create over-segmented and homogenous morphometric primitives (Chapter 5.3.3). Next, cross-profiles with a 500 m spacing were created and visually scrutinised to manually classify the homogenous morphometric primitives into 11 classes, i.e. fluvial terrace steps one to 10 and non-terrace features. Per-pixel reference data were generated by selecting all the pixels contained within each of the manually selected homogenous morphometric primitives.

A total of 2 793 objects were manually selected (Figure 5-2) by drawing from the knowledge gathered in the field and by comparing the VDTCN and the distance from the river channel of each cross-profile identified fluvial terrace with that of homogenous morphometric primitives. In total, 881 morphometric primitives were assigned to the fluvial terrace class, and their corresponding level between one and 10 was indicated. A further 1 912 morphometric primitives were labelled as non-terrace landforms. For per-pixel analyses, the objects selected were converted into a raster matching the resolution and coordinate system of the source DEM. This resulted in 808 428 reference pixels of which 551 012 represent non-terrace features and 257 416 represent terrace features.

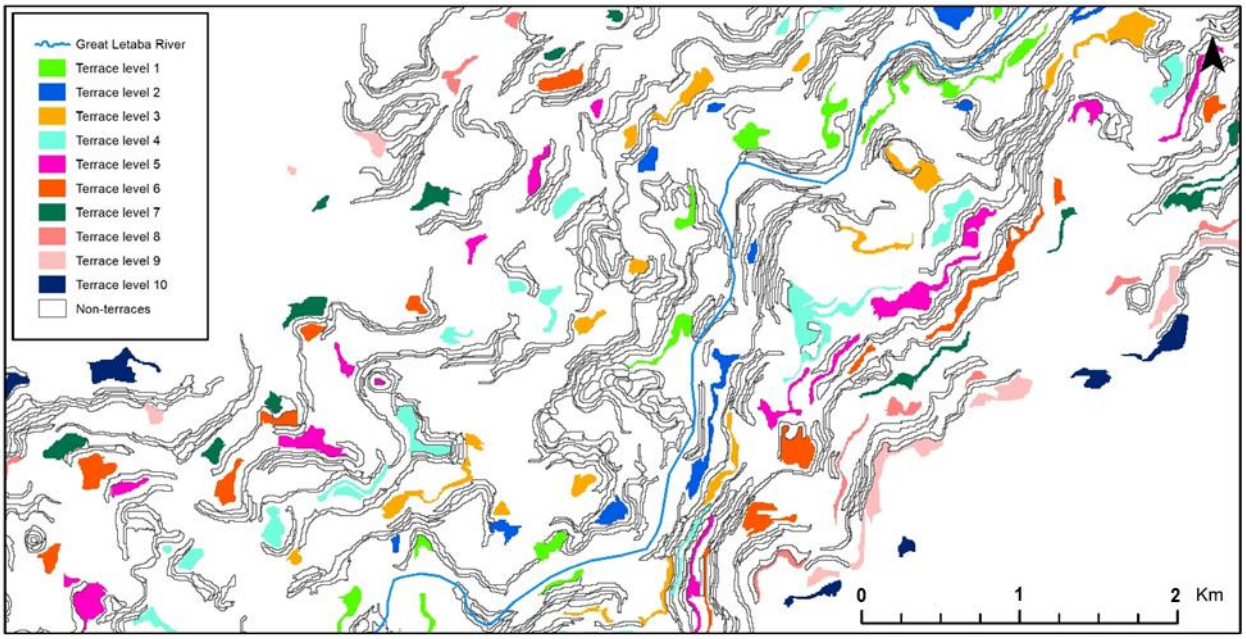


Figure 5-2 Subset of the reference data collected using cross-profiles

5.3.5 Preparation of LSP data

A correlation matrix was produced to evaluate whether any LSP combination pairs were highly correlated. One LSP of each combination pair with a correlation coefficient larger than 0.75 was discarded based on a qualitative (visual) assessment. Pre-processing of the LSPs included calculating the mean and standard deviation of all pixels contained within each morphometric primitive generated by the land surface segmentation procedure. For per-pixel analyses, however, no additional statistics were calculated (i.e. only the pixel values of each LSP were used).

5.3.6 Decision tree building and application

DT analysis was performed with the use of the Scikit-learn machine learning library (Pedregosa, Weiss & Brucher 2011). The library is free and operates within the Python programming environment. Scikit-learn was selected for this study based on its user-friendliness and extensive toolset of ancillary algorithms, e.g. cross-validation, grid search parameter optimisation and validation curves. The Scikit-learn DT library uses an optimised version of the classification and regression trees algorithm introduced by Breiman et al. (1984). It recursively partitions a dataset and assigns a simple prediction model (if-then statement) to each partition, which can be graphically visualised. DTs are particularly suitable for classifying fluvial terraces from LSPs as they are non-distance-based (Swan 2017), non-parametric machine learning algorithms that are insensitive to outliers, collinearities and missing data and has been shown to have excellent data

reduction and feature selection capabilities (Breiman et al. 1984; Laliberte, Fredrickson & Rango 2007).

For the purpose of this study, DTs were implemented using both default parameters and refined parameters selected through a hyper-parameterisation. Ideally, the Scikit-learn grid search parameterisation function should be used to iteratively evaluate every possible DT parameter combination within defined ranges to define an optimal parameter set, but the processing time for this approach is prohibitive. Instead, validation curves were used to evaluate the impact of each parameter on the DT validation, and training scores were used to eliminate redundant parameters from the grid search function, thereby reducing the processing time required. In other words, overfitting was mitigated by including only the most impactful DT parameters in the hyper-parameterisation process.

5.3.7 Experimental design

Eight experiments (Table 5-2) were carried out to evaluate the importance of each generated LSP and to select DT-derived topographic thresholds.

Table 5-2 Experimental design

Paradigm	Classes	Parameters	Experiment name	Transcribed
Per-pixel (PP)	2	Default (D)	PP-2-D	No
	2	Parameterised (P)	PP-2-P	Yes
	11	Default (D)	PP-11-D	No
	11	Parameterised (P)	PP-11-P	Yes
Object-based (OB)	2	Default (D)	OB-2-D	No
	2	Parameterised (P)	OB-2-P	Yes
	11	Default (D)	OB-11-D	No
	11	Parameterised (P)	OB-11-P	Yes

The experiments were designed to evaluate per-pixel and object-based approaches for differentiating either two terrace classes (i.e. terrace and non-terrace) or 11 fluvial terrace classes (i.e. the first ten ordinal fluvial terrace steps along the GLR and non-terrace landforms). Furthermore, each experiment was performed using either default DT parameters or hyper-parameterisation to mitigate overfitting, reduce the tree size and to select refined topographic thresholds.

The experiments were carried out as follow:

1. All available LSP data were used as input, and feature importances were computed to identify the most informative LSPs;
2. DT models were generated using both the default parameters and those found by the hyper-parameterisation process;
3. A ten-fold cross-validation was used to compute the training and validation scores of each experiment;
4. The Python package Graphviz was used to plot each of the eight DT models produced;
5. The hyper-parameterised DT models were manually transcribed into rulesets and implemented in eCognition software to map the results (DT models produced using default parameters were found to be too prohibitively complex to effectively transcribe into rulesets).

5.4 RESULTS

Table 5-3 – which summarises the training and validation scores of each experiment – shows that the experiments that classify binary classes performed better than those that classify 11 landform classes and the object-based experiments performed better than the per-pixel experiments. Respective validation scores of up to 0.88 and 0.83 were achieved for separating terraces from non-terraces and for differentiating various ordinal terrace levels and non-terraces in a very planate landscape. An interpretation of the feature importances, DTs, confusion matrices and resultant fluvial terrace maps is given in the following subsections. For the sake of brevity, the generated DT models for PP-11-P and OB-11-P and all feature importance graphs are not included in the text but are supplied in Appendix A.

Table 5-3 Quantitative results of each experiment

Experiment	Training score	Validation score
PP-2-D	1.00 +/- 0.00	0.75 +/- 0.04
PP-2-P	0.83 +/- 0.00	0.83 +/- 0.04
PP-11-D	1.00 +/- 0.00	0.70 +/- 0.03
PP-11-P	0.82 +/- 0.00	0.77 +/- 0.03
OB-2-D	1.00 +/- 0.00	0.83 +/- 0.03
OB-2-P	0.89 +/- 0.03	0.88 +/- 0.03
OB-11-D	1.00 +/- 0.00	0.78 +/- 0.03
OB-11-P	0.87 +/- 0.00	0.83 +/- 0.03

5.4.1 Hyper-parameterisation and LSP importance

A comparison of the PP-2-D and PP-2-P feature importances shows that the hyper-parameterisation process had a significant effect on the resulting DTs. By limiting the DT size to a maximum depth of three and the minimum number of sample leaf parameter to one, the importance of slope gradient was increased from 42% to 90% and the importance of VDTCN and elevation was reduced from 17% and 12% to 9% and 0.4%, respectively. Visual analysis (Figure 5-3) of the hyper-parameterised DT model (PP-2-P) indicates that a single slope threshold was used to distinguish between terrace and non-terrace features. Elevation and VDTCN thresholds were used to discriminate between terraces and non-terrace features representative of the floodplain. VDTCN was further used to inconsequentially subdivide non-terrace features (an indication of overfitting).

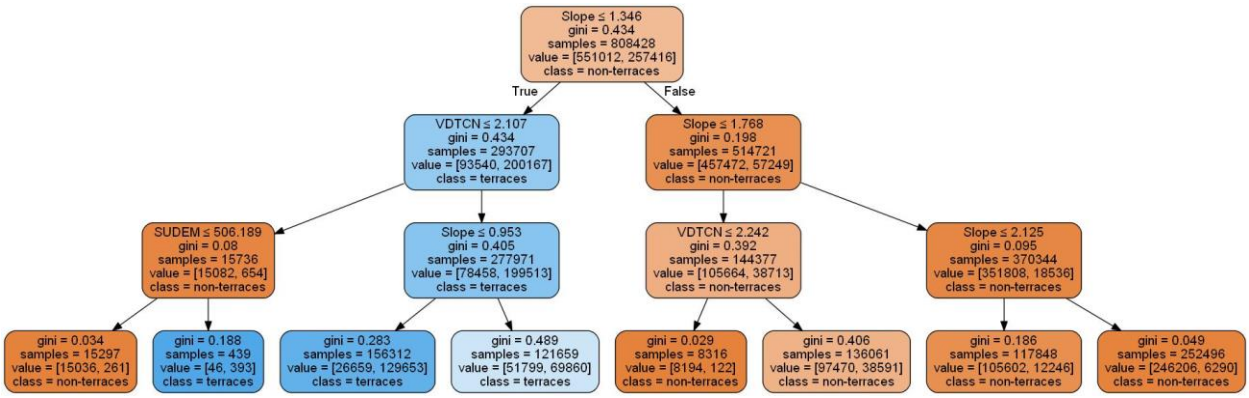


Figure 5-3 Hyper-parameterised DT model of PP-2-P

Similarly, PP-11-D and PP-11-P feature importances note an increase in the two most significant features' importance. VDTCN feature importance increased from 36% to 55%, slope feature importance increased from 20% to 37% and the feature importance of SUDEM decreased from 20% to 18% by limiting the DT size to a maximum depth of seven. Visual analysis of the PP-11-P DT revealed severe overfitting and an overly complex tree. While slope and VDTCN thresholds remain the main factors for distinguishing between terraces and non-terrace features and for identifying various terrace levels, additional explanatory variables, including elevation, valley depth, aspect, profile curvature and flow accumulation, are critical to further subdivide non-terrace and terrace features at lower tree levels.

The feature importance list for OB-2-D was dominated by slope (mean), which contributed 57%, and to a lesser degree by VDTCN (mean), aspect (standard deviation) and SAGA wetness index (mean), which altogether contributed 9%. By limiting the DT size to a maximum depth of two, the

minimum samples leaf to one and the maximum leaf nodes parameters to three (OB-2-P), the feature importances of slope (mean) and VDTCN (mean) were increased to 93% and 6%, respectively. Furthermore, OB-2-P (Figure 5-4) exhibited no signs of overfitting as it used only a single slope threshold to split the features into terrace and non-terrace features and a single VDTCN threshold to split the terrace features into terraces and non-terrace features representative of the floodplain.

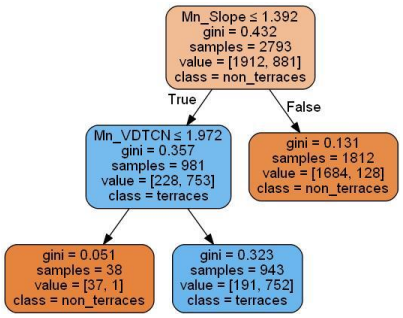


Figure 5-4 Hyper-parameterised DT model of OB-2-P

By limiting the DT size to a maximum depth of seven and the minimum samples leaf parameter to 17 (OB-11-P), the importance of VDTCN (mean) and slope (mean) increased from 34% to 41% and 27% to 49%, respectively. Visual inspection of the OB-11-P model indicated that a single slope (mean) threshold was used to split the dataset into terrace and non-terrace features. Thereafter, several other explanatory variables were used to subdivide the non-terrace features further, but no topographic thresholds were used for distinguishing between non-terrace and terrace features. VDTCN (mean) was principally used to subdivide the terrace feature set into subsets representative of a number of terrace levels. A DT depth of seven was needed to describe all terrace levels. Additional explanatory variables were used to subdivide terraces DT nodes to identify further non-terrace landforms; however, further experimentation concluded that the addition of these explanatory variables mostly result in overfitting.

5.4.2 Comparison of per-pixel and object-based approaches

5.4.2.1 Binary (terrace and non-terrace) classification

The validation score of 0.83 (Table 5-3) indicate that the binary per-pixel DT model (PP-2-P) successfully distinguished between terraces and non-terraces. The confusion matrix (Table 5-4) shows that non-terraces were more often classified correctly (producer and user accuracies of 0.89 and 0.86) than terraces (producer and user accuracies of 0.72 and 0.77).

Table 5-4 Confusion matrix for PP-2-P, where NT = non-terrace; T = terrace; OE = omission error; UA = user accuracy; CE = commission error; and PA = producer accuracy

	Predicted				
	NT	T	Total	OE	UA
NT	471713	79299	551 012	0.14	0.86
T	58131	199285	257 416	0.23	0.77
Total	529 844	278 584	808 428	-	-
CE	0.11	0.28	-	-	-
PA	0.89	0.72	-	-	-

Figure 5-5, which maps the features classified by the PP-2-P transcribed ruleset, illustrates the spatial pattern of error highlighted by the confusion matrix. Most of the landscape features were classified correctly and seem to follow the morphological discontinuities delineated by the reference data well, and little evidence of the salt-and-pepper effect, which is commonly associated with per-pixel classifications, is present. Misclassification of non-terrace reference features appears geographically grouped in areas comprising several neighbouring reference data polygons and is especially noticeable on the northern bank of the GLR. Terrace reference features, on the other hand, are often misclassified as a mixture of terrace and non-terrace features. Elongated morphometric primitives selected as reference terraces were more frequently misclassified than those appearing as rounded land surface features.

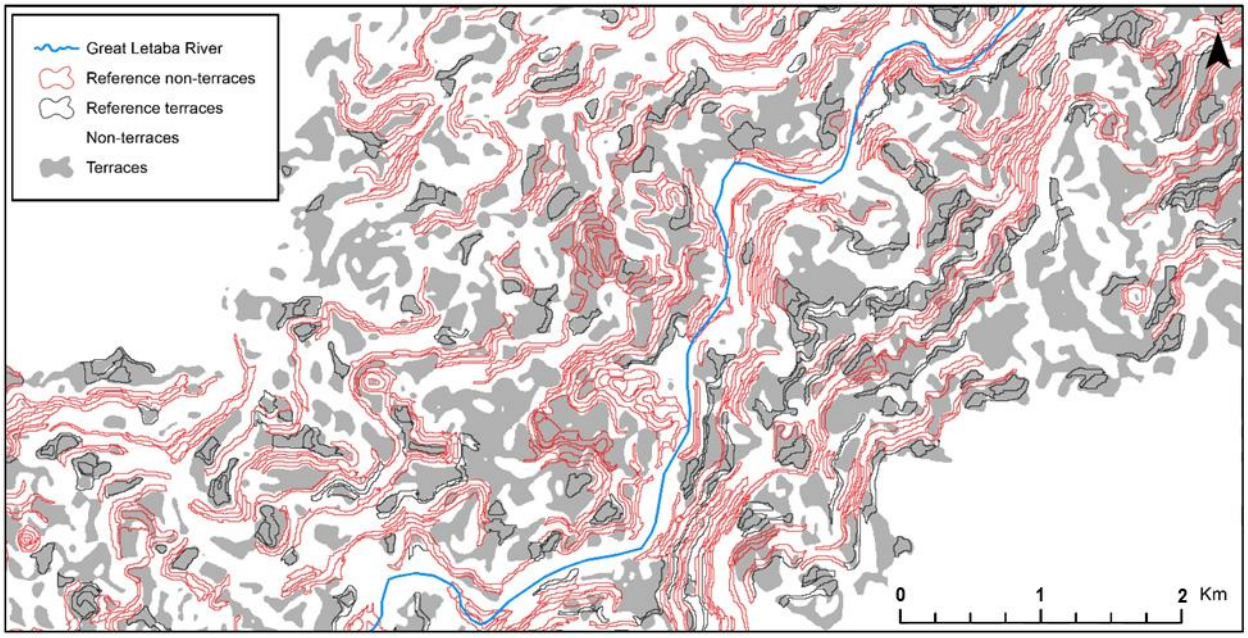


Figure 5-5 Comparison of per-pixel binary classification results (fluvial terraces indicated as grey and non-terraces as white) with reference data (reference terraces demarcated by black lines and non-terrace features by red lines)

The binary object-based DT model (OB-2-P), on the other hand, outperformed the PP-2-P model and achieved a relatively high validation score of 0.88 (Table 5-3). The resultant confusion matrix (Table 5-5) shows that non-terraces (producer and user accuracies of 0.90 and 0.93) were more often classified correctly than terraces (producer and user accuracies of 0.84 and 0.79).

Table 5-5 Confusion matrix for experiment OB-2-P, where NT = non-terrace; T = terrace; OE = omission error; UA = user accuracy; CE = commission error; and PA = producer accuracy

	Predicted				
	NT	T	Total	OE	UA
NT	1714	198	1 912	0.10	0.90
T	137	744	881	0.16	0.84
Total	1 851	942	2 793	-	-
CE	0.07	0.21	-	-	-
PA	0.93	0.79	-	-	-

As with experiment PP-2-P, the landscape features misclassified in experiment OB-2-P (Figure 5-6) are mostly those associated with large or elongated terrace surfaces. Potential misalignments between landscape feature boundaries and morphological discontinuities are avoided by utilising the same morphometric primitives used during the reference data collection process.

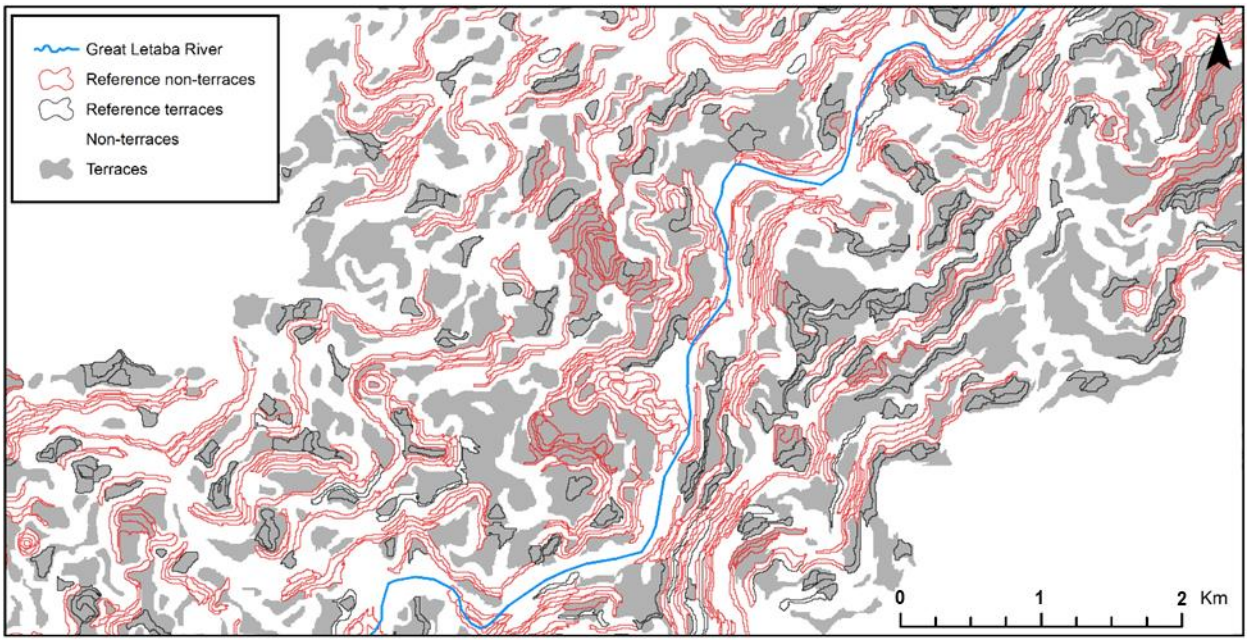


Figure 5-6 Comparison of object-based binary classification results (fluvial terraces indicated as grey and non-terraces as white) with reference data (reference terraces demarcated by black lines and non-terrace features by red lines)

5.4.2.2 Ordinal terrace level and non-terrace classification

The validation score of PP-11-P (0.77) is lower than that of PP-2-P (0.83) (Table 5-3). Inspection of the PP-11-P confusion matrix (Table 5-6) indicates that the poorer classification performance can, at least partially, be explained by the failure of PP-11-P to accurately classify non-terraces, as is evident from the high non-terrace omission (0.62) and commission (0.63) errors (PP-2-P produced non-terrace omission and commission errors of 0.14 and 0.11, respectively). Confusion between terrace levels was mostly confined to neighbouring classes (sequential levels). Terrace levels three (omission and commission errors of 0.51) and seven (omission error of 0.57) were most frequently misclassified.

Table 5-6 Confusion matrix for PP-11-P, where NT = non-terrace; T = terrace; OE = omission error; UA = user accuracy; CE = commission error; and PA = producer accuracy

	Predicted													
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	NT	Total	OE	UA
T1	11780	1453									4871	18104	0.35	0.65
T2	883	15683	1061								6057	23684	0.34	0.66
T3		1780	15138	1194							12567	30679	0.51	0.49
T4			801	18528	1512						11340	32181	0.42	0.58
T5				2602	14979	900					8728	27209	0.45	0.55
T6					2825	19525	1262				11047	34659	0.44	0.56
T7						1481	10360	714			11593	24148	0.57	0.43
T8							931	14843	277		6783	22834	0.35	0.65
T9								1526	16988	1042	6923	26479	0.36	0.64
T10									2205	12537	2697	17439	0.28	0.72
NT	7112	6968	14076	11024	7710	7507	5626	5075	8274	2596	47504	123472	0.62	0.38
Total	19775	25884	31076	33348	27026	29413	18179	22158	27744	16175	130110	380888		
CE	0.40	0.39	0.51	0.44	0.45	0.34	0.43	0.33	0.39	0.22	0.63			
PA	0.60	0.61	0.49	0.56	0.55	0.66	0.57	0.67	0.61	0.78	0.37			

Figure 5-7 shows that higher-lying terrace remnants are often surrounded by younger surfaces, and there are several examples of different terrace levels occurring immediately next to each other. The appearance of severely dissected and non-continuous terraces on the north-western bank of the GLR and some higher-lying terraces (e.g. levels six and seven) provides further evidence of the failure of PP-11-P to accurately classify non-terrace features.

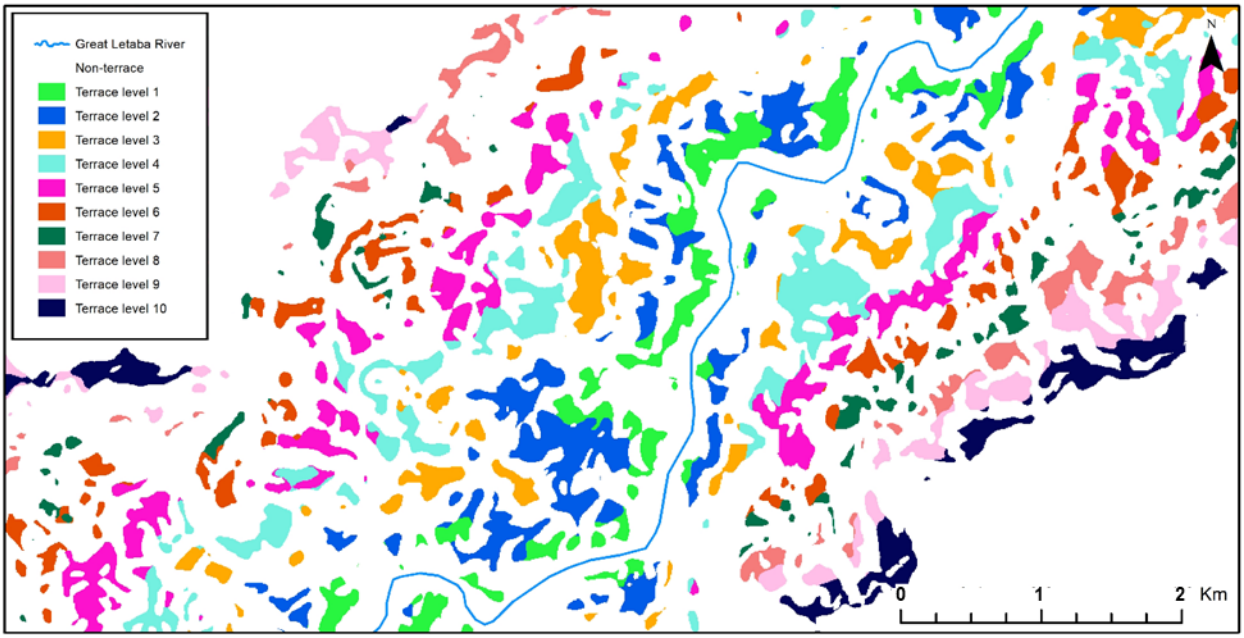


Figure 5-7 Per-pixel classification of various terrace levels and non-terrace features

The validation score produced by OB-11-P (0.83) is higher than that of PP-11-P (0.77), but lower than OB-2-P (0.88). Based on confusion matrix analysis (Table 5-7), the higher-lying terraces (terraces 6, 8, 9 and 10) were more frequently classified correctly (user or producer accuracies above 0.7) than terraces nearer to the GLR, while non-terraces were most accurately classified (user and producer accuracies of 0.91 and 0.89, respectively). Only one instance of misclassification between non-neighbouring terrace levels was identified and is considered negligible.

Table 5-7 Confusion matrix for OB-11-P, where NT = non-terrace; T = terrace; OE = omission error; UA = user accuracy; CE = commission error; and PA = producer accuracy

	Predicted													
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	NT	Total	EO	UA
T1	31	6									27	64	0.52	0.48
T2	5	40	4								26	75	0.47	0.53
T3		10	57	4			1				29	101	0.43	0.57
T4			1	69	8						29	107	0.36	0.64
T5			1	15	45	8	1				21	91	0.48	0.52
T6					9	78	4				23	114	0.32	0.68
T7		1	1			7	51	7			21	88	0.40	0.60
T8		1					8	63	6		12	90	0.29	0.71
T9								3	59	5	28	95	0.38	0.62
T10									6	45	5	56	0.20	0.80
NT	21	16	22	22	12	16	23	15	11	7	1747	1912	0.09	0.91
Total	57	74	86	110	74	109	88	88	82	57	1968	2793		
EC	0.46	0.43	0.31	0.37	0.39	0.28	0.40	0.28	0.28	0.21	0.11			
PA	0.54	0.57	0.69	0.63	0.61	0.72	0.60	0.72	0.72	0.79	0.89			

Figure 5-8 shows that the geographic distribution of the terrace levels and non-terrace features, classified with the OB-11-P transcribed ruleset, deviates significantly from the PP-11-P classification (Figure 5-6). While both demonstrate terraces that are partially surrounded by and/or occurring immediately next to younger terraces, the OB-11-PP terraces are more morphologically continuous and less frequently misclassified.

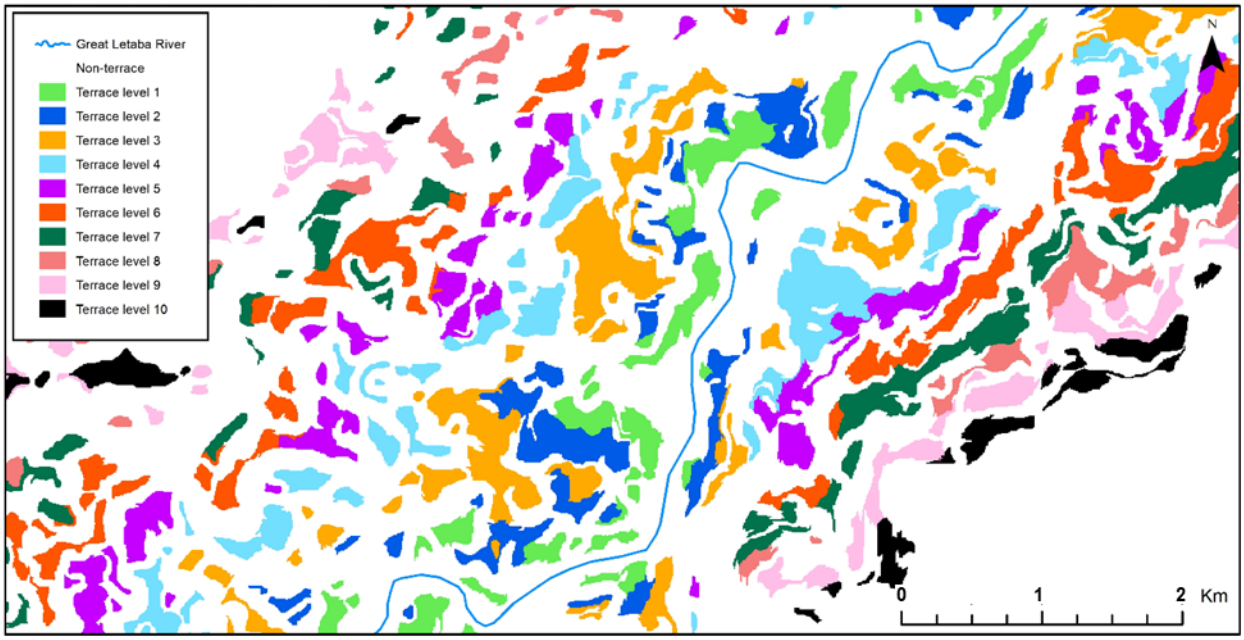


Figure 5-8 Object-based classification of various terrace levels and non-terrace features

5.5 DISCUSSION

All applications of DT were implemented using default parameters, after which overfitting was moderated through a hyper-parameterisation process using validation curves and grid search functions. The results outline key differences between per-pixel and object-based fluvial terrace delineation approaches.

The first difference relates to how LSP topographic thresholds are employed to differentiate between terrace and non-terrace landforms. A comparison of the results of the PP-2-P and OB-2-P models shows that both the per-pixel and object-based delineation approaches use only a single slope gradient threshold to effectively differentiate between terrace and non-terrace features. However, the PP-2-P model requires additional slope gradient, VDTCN and elevation thresholds to identify non-terrace features representative of the floodplain, while the OB-2-P model produced a shallower and simpler tree structure. The complexity of the per-pixel DTs is also apparent when PP-11-P and OB-11-P are compared. While both employ numerous LSPs and were found to be overfitted, the object-based DT graph is easier to understand and can be more intuitively transcribed into a simple ruleset. Visual analysis of the object-based DT graph determined that slope gradient and VDTCN contributed the most significant thresholds and all other LSPs thresholds were used to discriminate small quantities of terrace levels from non-terrace features. The simplicity of using only slope gradient and VDTCN thresholds in the DT generated by the

object-based experiments relate strongly to the fluvial terrace definition and field-based mapping approach (expert knowledge) adapted by Clubb et al. (2017).

The second key difference between the per-pixel and object-based fluvial terrace delineation approaches relates to their classification performance. Even though the hyper-parameterisation process increased the validation scores of the per-pixel models by a larger margin than the object-based models, the hyper-parameterised object-based experiments still produced consistently higher validation scores compared to the per-pixel experiments (Table 5-3). Visual assessment of the generated terrace maps revealed that both PP-2-P (Figure 5-5) and OB-2-P (Figure 5-6) struggled to identify large groupings of non-terraces and isolated terraces of elongated morphology, whereas confusion matrix analysis (Table 5-4 and Table 5-5) showed that both approaches' classification of non-terraces was more accurate than their classification of terraces.

Interpretation of the PP-11-P (Figure 5-7) and OB-11-P (Figure 5-8) generated terrace maps, on the other hand, reveals a complex land surface morphology, where older terraces are partially surrounded by younger terraces and terraces of different levels are located immediately next each other. Our field work shows that such instances of terrace mixing can be either indicative of paleo-fluvial processes that involve the denudation of terrace surfaces between depositional events or an indication of misclassification between terrace levels or non-terrace features (Chapter 3; Rozanov et al. 2017). In the case of the PP-11-P terrace map, the presence of severe terrace mixing is credited to the inability of the model to accurately classify non-terraces, as supported by omission and commission errors of 0.62 and 0.63, respectively (Table 5-6). The OB-11-P model classified non-terraces much more accurately and produced respective omission and commission errors of 0.09 and 0.11 (Table 5-7), which indicates that the limited presence of terrace mixing in the OB-11-P terrace map (Figure 5-8) can be reliably explained by the paleo-fluvial processes that dominate the Letaba valley.

In essence, the two main differences between per-pixel and object-based fluvial terrace delineation approaches discussed above highlight a potential inverse relationship between geomorphometric ruleset complexity and predictive accuracy. The superior predictive performance of the object-based models is attributed to the tendency of the object-based DTs to produce much simpler rules compared to per-pixel DTs. This is evident from the OB-2-P model's use of a single slope gradient threshold to differentiate between terraces and non-terrace features, compared to the PP-2-P model's use of multiple slope gradient thresholds. Also, the OB-11-P model defines each

categorical class using a much simpler set of topographic thresholds and relies on comparatively few criteria to distinguish between subtle geomorphometric differences, while PP-11-P relies heavily on intricate rules to distinguish between ordinal terrace levels and non-terrace features. This supports the premise that complex machine learners do not necessarily produce the most accurate results (Rudin 2019) and illustrates the potential of integrating expert systems and white-box machine learning approaches for fluvial terrace classification.

Comparing our results to previous work highlights two key advantages of an object-based DT-facilitated fluvial terrace mapping methodology. First, most available fluvial terrace mapping techniques are deterministic, based on expert knowledge or dependent on user parameterisation to fine-tune the method for a specific study area (Anders, Seijmonsbergen & Bouten 2011; Clubb et al. 2017; Demoulin et al. 2007; Stout & Belmont 2014). We have shown that the DTs' hyper-parameterisation without analyst involvement can be combined with object-based land surface segmentation to develop supervised fluvial terrace rulesets that are intuitively interpretable and comparable to expert knowledge. Second, various fluvial terrace extraction methods (Clubb et al. 2017; Hopkins 2014; Stout & Belmont 2014) perform best in mountainous areas, as an absence of contrast between terrace and non-terrace slope gradient and elevation above channel values in a flat topography often result in terrace surface misidentification. This problem was not encountered in this study, as is evident from the high accuracies reported and the fact that the GLWC is dominated by extensive planation activity that significantly levelled the topography. This insensitivity to flat topographies may be attributed to the combination of DT selection of topographic thresholds, morphometric primitive reference data collection and object-based fluvial terrace classification. This study supports the Li et al. (2019) suggestion that the automated mapping of floodplains and fluvial terraces may benefit from an object-based approach, given that the object-based derived rulesets relate strongly to a field-model mapping approach (Clubb et al. 2017). Further work should expand on the integration of expert system and white-box machine learning approaches, with the aim of developing and evaluating transparent and robust landform mapping rulesets that can be applied over large areas.

5.6 CONCLUSION

The bulk of available fluvial terrace delineation techniques employ per-pixel approaches, founded on geomorphometric principles, that do not take the scale-dependency of LSPs and scale specificity of particular landforms into account (Drăgut, Eisank & Strasser 2011; Evans 2012). In contrast, discrete geomorphometry- and object-based land surface segmentation principles are scale-dependent and have been shown to delineate fluvial terraces accurately (Anders, Seijmonsbergen & Bouten 2011; Van Asselen & Seijmonsbergen 2006).

This study employed a white-box machine learning methodology to demonstrate the benefits of object-based land surface delineation techniques for classifying fluvial terraces. With the use of inductive DT procedures, it was determined that rulesets derived from per-pixel LSP data are more susceptible to overfitting and produce overly complex rules. Rulesets derived from object-based LSP data were found to be less susceptible to overfitting and produced much simpler rules that are more intuitive and understandable and can be readily compared to expert knowledge, which illustrates the potential of integrating expert systems and white-box machine learning approaches for fluvial terrace classification. Qualitative and quantitative predictive performance evaluation of both approaches revealed that the object-based approach, which utilises scale-specific LSP statistics derived from homogenous morphometric primitives, consistently outperformed the scale-independent per-pixel methodologies. Based on these findings, it is clear that an object-based approach is a viable methodology for delineating fluvial terraces over large areas.

CHAPTER 6: SOIL NIR SPECTROSCOPY AND OBJECT-BASED LANDSURFACE SEGMENTATION FOR FLUVIAL TERRACE LEVEL DIFFERENTIATION

6.1 ABSTRACT

Establishing continuity of fluvial terrace remnants in eroded landscapes is often limited to resource intensive field interpretations of in situ stratigraphic and physiographic features. Soil NIR spectral reflectance is an integrative property of soil that have been successfully combined with digital terrain data for soil-landscape modelling purposes. This study assesses the viability of soil NIR spectroscopy as a rapid and cost-effective way to differentiate between various fluvial terrace levels. First, the correlation between the NIR spectra of 88 soil samples and the height above river channel (HARC) of various fluvial terrace levels were investigated using pre-processed spectra, principle component analysis (PCA) and partial least square regression (PLSR). A strong correlation (R^2 = 0.78, RPD = 2.11) was achieved using NIR spectra located between 7500-5446 cm^{-1}, which remained strong (R^2 = 0.55, RPD = 1.49) after cross-validation. Next, the Scikit-learn random forest (RF) machine learning library was used to classify two sets of fluvial terrace level classes.

The classification of two and three fluvial terrace classes produced validation scores of 0.73 and 0.76, respectively, when using the entire unprocessed spectral dataset as input. Interpretation of the subsequent feature importances indicated that spectral wavelength bands associated with the absorption characteristics of smectite, kaolinite, carbonates and talc were most important. Final validation scores of 0.74 (two fluvial terrace classes) and 0.76 (three fluvial terrace classes) were achieved by isolating specific wavelength bands associated with smectite (5334 cm^{-1} and 5269 cm^{-1}) and kaolinite (3664 cm^{-1} and 3699 cm^{-1}) fractions. This study demonstrates the clear potential of NIR spectroscopy as a soil analytical technique with which to differentiate between fluvial terrace levels.

6.2 INTRODUCTION

Fluvial terraces, a global geomorphic phenomenon, have been shown to support ecosystem services (Biondi et al. 2009; Saldana & Ibanez 2004) and are instrumental in deciphering paleo climatic conditions (Bridgland & Westaway 2008; Pazzaglia 2013; Vandenberghe 2015). Consecutive floodplain depositional and river incision events driven by tectonic uplift, base-level

drop and climate fluctuation often produce extensive fluvial terrace staircases. As a result, fluvial terrace levels of the same staircase often exhibit varying characteristics (Leopold, Wolman & Miller 1964; Pazzaglia 2013; Rozanov et al. 2017). Interpretation of in situ stratigraphic and physiographic features has long formed the basis for establishing terrace continuity in geomorphological settings where only terrace remnants are preserved and terrace continuity cannot be inferred from the terrace height above the river channel (Leopold, Wolman & Miller 1964). While the advent of geographical information systems (GIS) and digital elevation models (DEMs) introduced quantitative fluvial terrace mapping strategies (Clubb et al. 2017; Demoulin et al. 2007; Stout & Belmont 2014), resource intensive expert interpretation and field work are still frequently needed to evaluate terrace continuity.

Soil diffuse reflectance spectroscopy (DRS) may offer a rapid, inexpensive, non-destructive, reproducible and repeatable technique to differentiate between various fluvial terrace levels in support of terrace continuity assessments. The use of DRS in pedology is well-established and based on the principle that soil spectral reflectance is an integrative property of soil that can be used to estimate other soil properties (Nocita et al. 2015; Rossel et al. 2006; Rossel & McBratney 2008; Stenberg et al. 2010). Multivariate statistical and machine learning methodologies have been employed to predict various soil properties from soil DRS (for further explanation and review see Nocita et al. (2015) and Stenberg et al. (2010)). DRS has been shown to be a promising alternative to resource intensive conventional (chemical and physical) soil analytical techniques used in most soil-landscape modelling (or digital soil mapping) strategies.

Soil-landscape models are primarily based on the premise that similar soils correspond to similar environmental conditions (McBratney, Mendonca Santos & Minasny 2003; Minasny & Mcbratney 2015). As a result, DEMs and geomorphometrical techniques have been extensively used to produce land surface parameters (LSPs) representing various landscape characteristics that can be used to statistically facilitate soil parameter estimation. Vasily Dokuchaev (1846 – 1903) was the first to postulate that landscape characteristics such as climate, organisms, relief (topography), parent material and time contribute to soil formation. Jenny (1941) formalised the aforementioned soil formation factors into the seminal *corpt (c*limate, *o*rganisms, *r*elief, *p*arent material, *t*ime*)* model, which McBratney, Mendonca Santos & Minasny (2003) extended to introduce the *scorpan* (soil, *c*limate, *o*rganisms, *r*elief, *p*arent material, *a*ge and spatial positio*n*) model for the quantitative estimation of soil parameters. Subsequently, investigations employing the *corpt* or *scorpan* models depended primarily on correlations between point-sampled soil parameters and

LSP datasets. Various combinations of feature selection, data mining, geostatistical and machine learning techniques have been employed to produce successful wall-to-wall soil-landscape models (Heung et al. 2016; McBratney, Mendonca Santos & Minasny 2003; Minasny & Mcbratney 2015).

Given that soil spectral reflectance is a fundamental property of soil (Stenberg et al. 2010), methodologies that exploit the relationship between soil spectra and soil forming landscape characteristics is a logical progression for soil-landscape modelling strategies. Mulder (2013), for example, characterised mineral composition at a regional scale by using a combination of soil spectra and remotely sensed datasets, while Rizzo et al. (2016) used soil spectra to help define mapping units, select representative samples and classify soil types. Vasat et al. (2017) improved the prediction accuracy of soil organic carbon by including LSPs as auxiliary data in spectroscopic models. Wadoux (2015) established relationships between soil spectral bands, soil parameters and landscape covariates by employing regression algorithms (partial least-squares regression (PLSR), support vector machine, Cubist) and a variety of landscape characteristics (represented by 34 LSPs, land use and geology).

Nevertheless, attempts to establish a link between soil spectral signatures and landscape characteristics have been limited to the use of per-pixel LSP data, which fail to take into account the scale specificity of LSPs and discrete landforms. Discrete geomorphometry offers an alternative to per-pixel approaches as it involves the delineation of morphometric primitives that are exclusively defined by homogeneity criteria used as a proxy for scale (Drăguţ & Eisank 2011; Drăguţ, Tiede & Levick 2010; Louw & Van Niekerk 2019). Morphometric primitives are basic land surface building blocks used to transition from per-pixel LSP data to object-based landform features. Multi-resolution segmentation (MRS) has been shown to delineate morphological discontinuities accurately (Van Niekerk 2010) and, consequently, has been widely adopted to produce morphometric primitives to create fine-scale geomorphological maps in fluvial terraced landscapes (Anders, Seijmonsbergen & Bouten 2011; D'Oleire-Oltmanns et al. 2013) and for soil mapping purposes (Dornik, Drăguţ & Urdea 2017; Drăguţ & Dornik 2013).

Given the established relationship between soil spectra and landscape characteristics, this study aims to evaluate whether NIR spectra of soil can be employed to differentiate between fluvial terrace levels with varying stratigraphic characteristics. First, MRS is used to produce homogenous morphometric primitives. Next, PLSR is employed to investigate the relationship between the vertical distance above channel network (VDTCN) of the various terrace levels and pre-processed

soil NIR spectra. Finally, the entire unprocessed soil NIR spectra dataset is used as RF explanatory data to classify fluvial terrace levels classes and to identify the most significant NIR wavelengths for differentiating between fluvial terrace levels. Our results show that soil NIR spectroscopy provides a viable platform for accurate, cost-effective and rapid fluvial terrace mapping strategies.

6.3 MATERIALS AND METHODS

6.3.1 Study area

The Letaba Estates (LE), a ~2700 ha commercial citrus farm located at the foot of the Drakensberg Escarpment in the northern Lowveld of South Africa (Figure 6-1), is dominated by a sequence of fluvial terraces and distinct soil patterns. The geomorphologic evolution of the Lowveld was influenced primarily by three separate tectonic uplift events that occurred between the late Jurassic and the Holocene. Interspersed cycles of fluvial terrace formation and planation activity preserved most older fluvial terraces as interfluves in the landscape. Nevertheless, younger and better-preserved fluvial terraces, such as those of the LE, are still present immediately next to the present-day Great Letaba River (Dollar 1998; Partridge et al. 2010; Partridge & Maud 1987; Rozanov et al. 2017).

Due to its close proximity to the escarpment, the LE has a subtropical climate and receives most of its ~700 mm yr-1 rainfall in summer (Paterson, Nell & Seabi 2011). The soils were classified into soil forms according to the South African soil classification (Soil Classification Working Group, 1991) as part of the contractual soil survey of the Letaba Estates performed by the Institute for Soil, Climate and Water of the Agricultural Research Council (ARC-ISCW). The survey results are presented with permission from the Letaba Estates. Soils (see Table 3-1 for soil type descriptions) found immediately next to the GLR are characterised by deep, red, weakly structured, sandy clay loams (Hu in Figure 6-1). Higher up the slope and further from the GLR, the soils transition into moderately deep, structureless, sandy yellow-brown (Av and Gc/Av in Figure 6-1) or grey (Lo/Kd and Lo/Wa in Figure 6-1) loams. The higher areas of the LE are covered by moderately shallow to shallow soils that are often gravelly and underlain by bedrock or hard plinthite (Cf, We/Dr and Gs in Figure 6-1). Finally, shallow soils with greyed clay or soft plinthite (Ka/Kd and W in Figure 6-1) subsoils occur in places that are associated with non-perennial drainage networks.

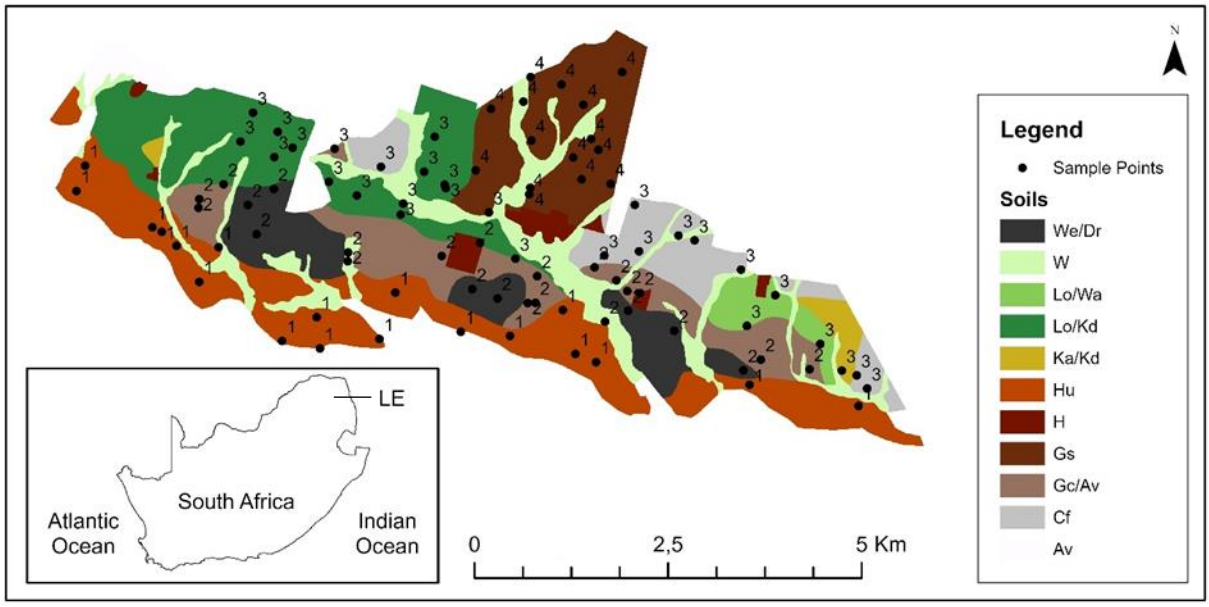


Figure 6-1 Location map showing the location of the LE and its soil types. Each point-sample site has been assigned to one of four analyst defined groupings and labelled accordingly.

(Adapted from Paterson, Nell & Seabi (2011))

6.3.2 Terrain analysis and land surface segmentation

The Stellenbosch University DEM (SUDEM) was chosen for all LSP calculations and to facilitate the delineation of morphometric primitives. The SUDEM was created by fusing interpolated spot height data and large-scale contours with the 30 m shuttle radar topography mission DEM. It has a pixel size of 5 m, a mean error of 1.77 m and a standard deviation of error of 1.21 m (Van Niekerk 2015b). Slope gradient and VDTCN were selected to consider the two most prominent morphological properties of fluvial terraces: 1) the low gradient nature of fluvial terrace surfaces; and 2) the varying distance above channel of different fluvial terrace steps (Clubb et al. 2017). Using the System for Automated Geographical Analysis (SAGA) software package, slope gradient and VDTCN were computed using the SUDEM as input, while cross-profiles were created using the VDTCN as input (Böhner & Selige 2006; Olaya & Conrad 2009).

The eCognition (version 9.5) MRS algorithm was used to delineate morphometric primitives with slope gradient as input. MRS is a pairwise region-merging algorithm that aggregates pixels into objects based on decision heuristics and a user-controlled homogeneity criterion that considers both spectral and shape homogeneity (Baatz & Schäpe 2000). As with many geomorphometric applications, the spectral homogeneity (calculated from the LSP pixel values) and the shape homogeneity (defined by the MRS parameters shape and compactness) parameters were

disregarded so as to favour only spectral (LSP) homogeneity in the segmentation process. The third MRS user-defined parameter is the scale parameter (SP), which acts as a homogeneity criterion threshold and determines the size of the resultant objects (Drăguţ et al. 2009; Drăguţ & Eisank 2012; Trimble 2014b). Over-segmented morphometric primitives were delineated using a SP value of five and shape and compactness parameters of zero to ensure internal object homogeneity.

6.3.3 Acquisition of spectra

Soil samples were collected at 88 predetermined locations selected using a random stratified sampling scheme, where the strata represent the broad soils classifications defined by Paterson, Nell & Seabi (2011) and the number of observations per strata were calculated based on percentage surface area of each soil grouping. The soil samples were scanned once to acquire the NIR reflectance spectra with a Bruker MPA (multi-purpose analyser) equipped with a quartz beam splitter and RT-PbS detector. The reflectance of the samples was measured from 12 500 to 3 600 cm^{-1} (800 – 2 778 nm) at 1 cm^{-1} using a rotating macro sample cup with an integrating sphere at 128 scans per sample (Figure 6-2). The OPUS (ver. 7.2.139.1294) software supplied with the Bruker MPA was used for spectral data collection.

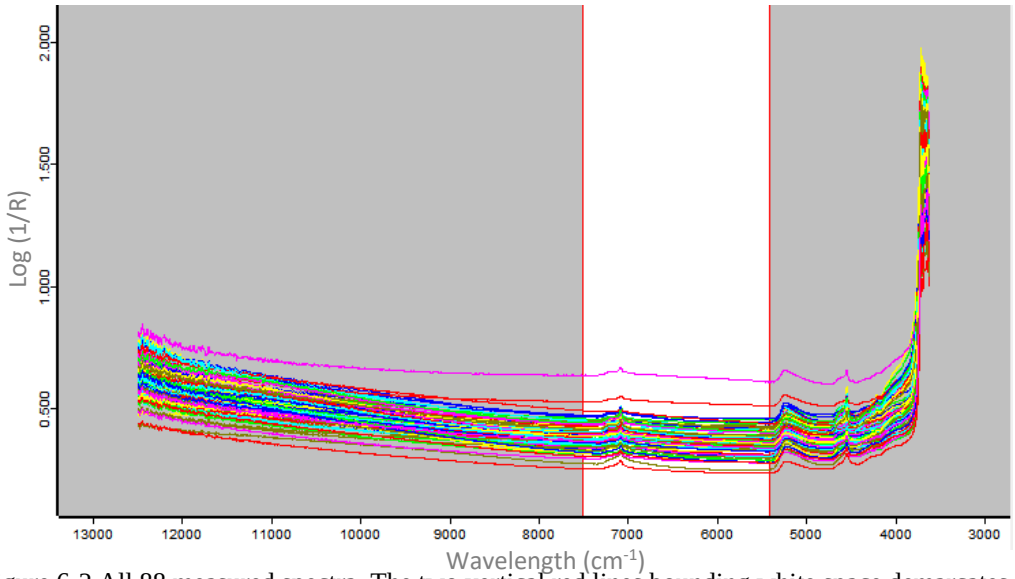


Figure 6-2 All 88 measured spectra. The two vertical red lines bounding white space demarcates the spectral region (7500-5446 cm^{-1}) selected through interpretation of PCA score coefficients.

6.3.4 PLSR

The combined use of spectrum pre-processing techniques, PCA and PLSR has been established for soil chemometric analyses (Stenberg et al. 2010). The Quant2 module of the OPUS software was employed to investigate the relationship between fluvial terrace VDTCN and soil NIR spectra. All NIR spectra were pre-processed with a three-step process. First, the 1st derivative of the spectra was calculated to emphasise small pronounced features. Next, a 17-point smoothing filter was applied to reduce noise. Finally, multiplicative scatter correction (MSC) was employed to perform a linear transformation of the spectra (Bruker 2006).

PCA was used to select suitable spectra for PLSR. PCA performs a variance-based factor analysis during which the spectra matrix is transformed into orthogonal factors (loadings) and corresponding scores. The first factor describes the largest portion of the variance, after which the portion of variance described by each following factor decreases as the factor number increases. The scores indicate how the original data are being represented by the factors and which spectra are most suitable for input to PLSR (Bruker 2006).

PLSR uses a linear multivariate two-block approach to model the relationship between two matrices. It can analyse strongly correlated and noisy data and model numerous predictor and response variables simultaneously (Wold & Sjostrom 2001). PLSR was performed using the VDTCN – calculated as the mean of each morphometric primitive – as response variables and the corresponding NIR spectra as predictor variables. The Quant2 module evaluates the PLSR fit by calculating coefficient of determination (R^2), root mean square error (RMSE) and residual prediction deviation (RPD) values. The regression is further evaluated through the calculation of cross-validated R^2, RMSE, RPD and bias values (Bruker 2006).

6.3.5 Categorising fluvial terrace levels

The classification of the fluvial terrace levels using soil spectra as independent data requires categorical classes that are represented by a sufficient number of soil spectral samples. Due to the high number of fluvial terrace levels present within the LE and the limited number of soil spectral samples, the terrace levels sampled were categorised into representative classes using a combination of qualitative (analyst interpretation) and quantitative (k-means clustering) approaches. First, the corresponding fluvial terrace level of each soil sample was identified using the cross-profiles created, and the dominant soil types were manually placed into four groupings

through analyst interpretation of the Paterson, Nell & Seabi (2011) LE soil map; Figure 6-1 illustrates the location of each soil sample and their respective analyst assigned group.

Figure 6-3 illustrates a clear soil distribution and a distinct association of certain soil groupings with specific terrace levels. Group one soils (Hu) only occur on terrace levels one to six. Group two soils (Gc, Av, We, Dr and W), on the other hand, are found on terrace levels three to nine and on terrace level 12. The rest of the soils – group three (Ka, Kd, Wa, Lo, Cf) and group four (Gs) – can be found on most terraces above the sixth level.

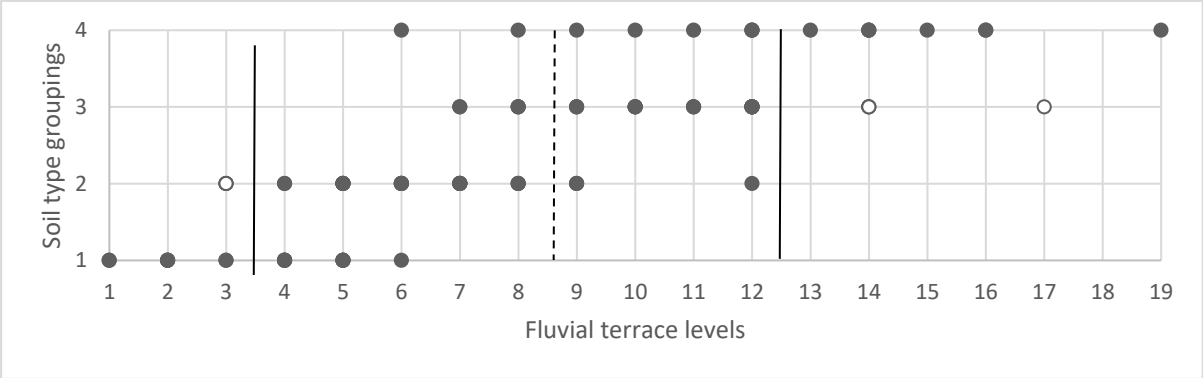


Figure 6-3 Distribution of analyst defined soil type groupings and fluvial terrace levels; where the soil groups: 1 = Hu, 2 = Gc/Av/We/Dr/W, 3 = Ka/Kd/Wa/Lo/Cf and 4 = Gs. The dashed vertical line demarcates the boundaries of the two fluvial terrace classes, solid vertical lines demarcate the boundaries of the three fluvial terrace classes and the white points indicate data excluded from the analysis.

To further aid the defining of the terrace classes, the Scikit-learn machine learning library (Pedregosa, Weiss & Brucher 2011) was used to execute a k-means unsupervised classification using the identified fluvial terrace levels as dependent data and all the unprocessed spectra as independent data. The number of target clusters was iterated between two and nine and the classification that produced the most visually meaningful clusters was selected. The most meaningful classification was achieved using three clusters (Figure 6-4), of which clusters one and three overlap substantially. The second cluster, however, group terraces nine to 12 and 14.

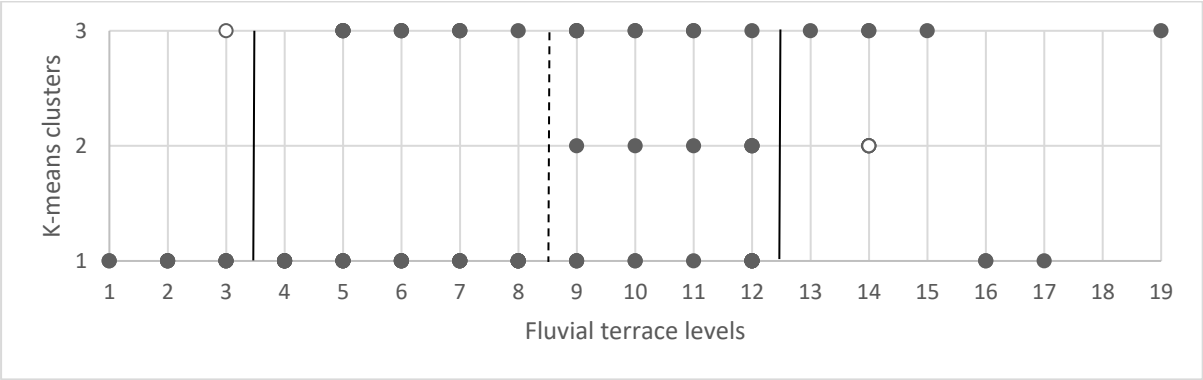


Figure 6-4 Results of the k-means unsupervised classification using 3 clusters. The dashed vertical line demarcates the boundaries of the two fluvial terrace classes, solid vertical lines demarcate the boundaries of the three fluvial terrace classes and the white points indicate data excluded from the analysis.

Finally, a combined interpretation of the soil groupings (Figure 6-3) and k-means clusters (Figure 6-4) enabled the identification of two sets of fluvial terrace classes. The first delineates two classes by splitting the dataset between terrace levels eight and nine (demarcated by dashed lines in Figure 6-3 and Figure 6-4), thereby grouping the bulk of the terraces underlying soil groupings one and two into the first class and soil groupings three to four and the second k-means cluster into the second class. The second set delineates three fluvial terrace classes (demarcated by solid lines in Figure 6-3 and Figure 6-4). The first group terraces one to three, which are primarily associated with the first soil grouping (Hu soils), and includes only one sample located on terrace three that is associated with the second soil grouping. The second class constitutes terrace levels four to 12 and represents the bulk of the second and third soil groupings. The upper boundary of the second class was selected to coincide with the upper limit of the second k-means cluster and the third soil grouping. The final class demarcates terraces 13 to 19 and coincides primarily with the fourth soil grouping. Table 6-1 summarises the fluvial terrace class boundaries.

6.3.6 RF classification

RF was selected as an appropriate supervised classifier as it has been shown to classify object-based soil types accurately (Dornik, Drăguţ & Urdea 2017), and its use in soil spectroscopy is well-established (De Santana, De Souza & Poppi 2018). RF is an ensemble decision tree machine learner capable of performing both regression analysis and categorical classification (Breiman 2001; Ho 1995). Decision trees work by recursively partitioning a dataset and by fitting a simple predictive model to each partition (Breiman et al. 1984). It is non-parametric, insensitive to outliers, collinearities and missing data and can be used for data reduction and feature selection purposes (Loh 2011; Loh 2014). RF employs bootstrap aggregation (bagging) to mitigate

overfitting by reducing the variance of individual trees by averaging a random selection of trees from the dataset (Breiman 2001).

The Scikit-learn machine learning library (Pedregosa, Weiss & Brucher 2011) was used to perform various RF classifications of the fluvial terrace levels, set out as four experiments. The first serves to classify two terrace classes by utilising all the available spectra (Experiment 1a), and, thereafter, a second classification is based only on the most important spectra of the first classification (Experiment 1b). The third and fourth experiments follow the same methodology to classify three terrace classes (Experiments 2a and 2b). The number of estimators parameter was set to 500 for all classifications to ensure sufficient out-of-bag (OOB) error stabilisation (Breiman 2001; De Santana, De Souza & Poppi 2018), and validation curves were used to identify RF parameters that had the most influence over the resultant training and validation scores. All RF classifications employed only unprocessed NIR spectra.

6.4 RESULTS AND DISCUSSION

6.4.1 PLSR

All pre-processed NIR spectra were analysed using the Bruker OPUS software PCA module. Interpretation of the score coefficients of the first factor indicated that spectral bands within the region of 7 500 – 5 446 cm^{-1} contributed the most variance (Figure 6-2). Subsequently, PLSR was performed using all pre-processed spectral bands between 7500-5446 cm^{-1} as predictor variables and VDTCN as the response variable. Fit of the PLSR produced a R^2 value of 0.78, a RMSE of 9.63 m and an RPD value of 2.11. Correspondence between the spectra and VDTCN was further evaluated through cross-validation, which produced a R^2 value of 0.55, a RMSE of 12.9 m, an RPD value of 1.49 and a bias of -0.03.

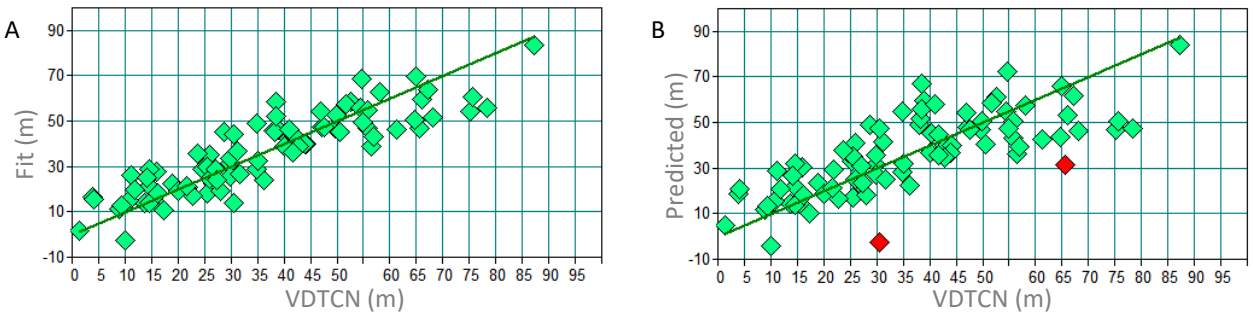


Figure 6-5 a) PLSR fit values vs. true VDTCN values; b) PLSR predicted values vs. true VDTCN values (the red markers indicate outliers)

6.4.2 RF classification

Validation curve analyses indicated that max depth is the most impactful RF parameter, which was iterated in increments of 10 between 10 and 300 with the sci-kit learn hyper-parameterisation grid search cross-validation tool. Resultantly, a max depth value of 10 produced a validation score of 0.76 and a training score of 0.99 (Table 6-1) when classifying two terrace classes using all the available spectra (Experiment 1a). Visual interpretation of the produced feature importances indicate that four distinct peaks, visible at 7 163 cm^{-1}, 5 334 cm^{-1}, 4 983 cm^{-1} and 3 664 cm^{-1}, are the most important wavelengths (Figure 6-6). Bivariate correlation calculations indicated that the reflectance values measured at 7 163 cm^{-1}, 5 334 cm^{-1} and 4 983 cm^{-1} are significantly correlated (Pearson correlation coefficients > 0.9). Consequently, only 5 334 cm^{-1} and 3 664 cm^{-1} were selected as explanatory variables for further RF classifications (Experiment 1b). This produced a validation score of 0.76 and a training score of 1.0 (Table 6-1). Feature importance of 5.68 and 4.32 were computed for 5 334 cm^{-1} and 3 664 cm^{-1}, respectively.

Table 6-1 Summary of fluvial terrace classes, optimal max depth values and RF accuracies

Number of classes	Terrace levels	Experiment	Spectra (cm^{-1})	Max depth	Training score	Validation score
2	1-8; 9-19	1a	12500 – 3600	10	0.99	0.76
		1b	3664; 5334	30	1.00	0.76
3	1-3; 4-12; 13-19	2a	12500 – 3600	100	1.00	0.73
		2b	3699; 5269	100	1.00	0.74

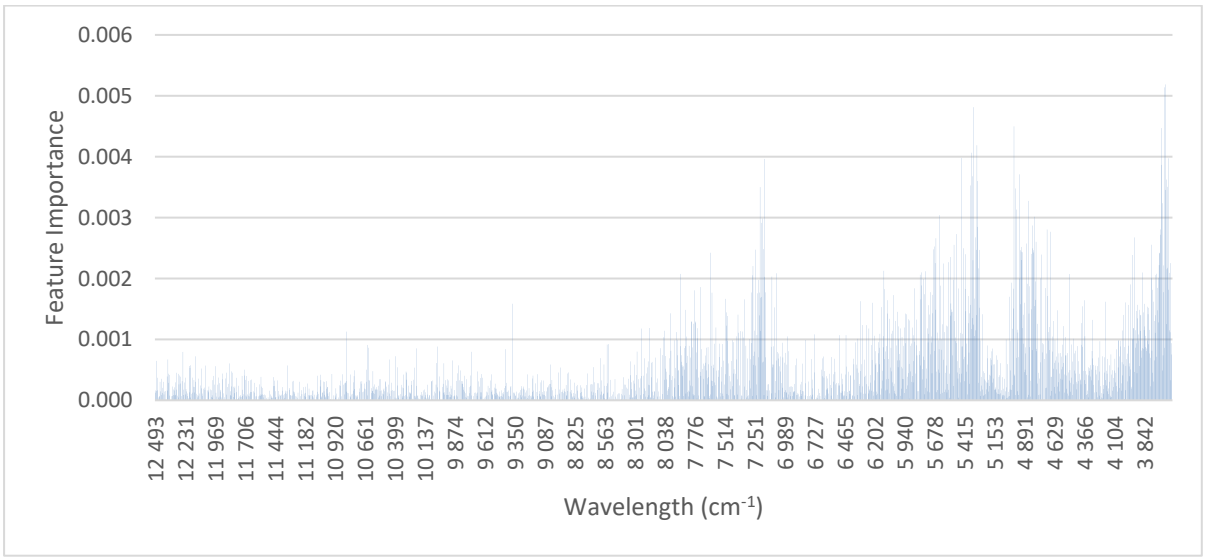


Figure 6-6 RF feature importance calculated while classifying two terrace classes

A max depth value of 120 was selected by employing the sci-kit learn hyper-parameterisation grid search cross-validation tool. This depth produced a validation score of 0.73 and a training score of 1.0 (Table 6-1) when classifying three terrace classes with all the available spectra (Experiment 2a). Six distinct feature importance peaks were identified with visual interpretation of Figure 6-7. These peaks are located at 6 106 cm^{-1}, 5 944 cm^{-1}, 5 604 cm^{-1}, 5 269 cm^{-1} and 3 699 cm^{-1}. Reflectance values 6 106 cm^{-1}, 5 944 cm^{-1}, 5 604 cm^{-1}, 5 269 cm^{-1} were found to be highly correlated (Pearson correlation coefficients > 0.9), therefore only 3 699 cm^{-1} and 5 269 cm^{-1} were retained for further RF classifications (Experiment 2b). When using these bands as input to RF, a validation score of 0.74 and a training score of 1.0 were achieved (Table 6-1), while feature importance of 0.51 and 4.86 were calculated for 5 269 cm^{-1} and 3 699 cm^{-1}, respectively.

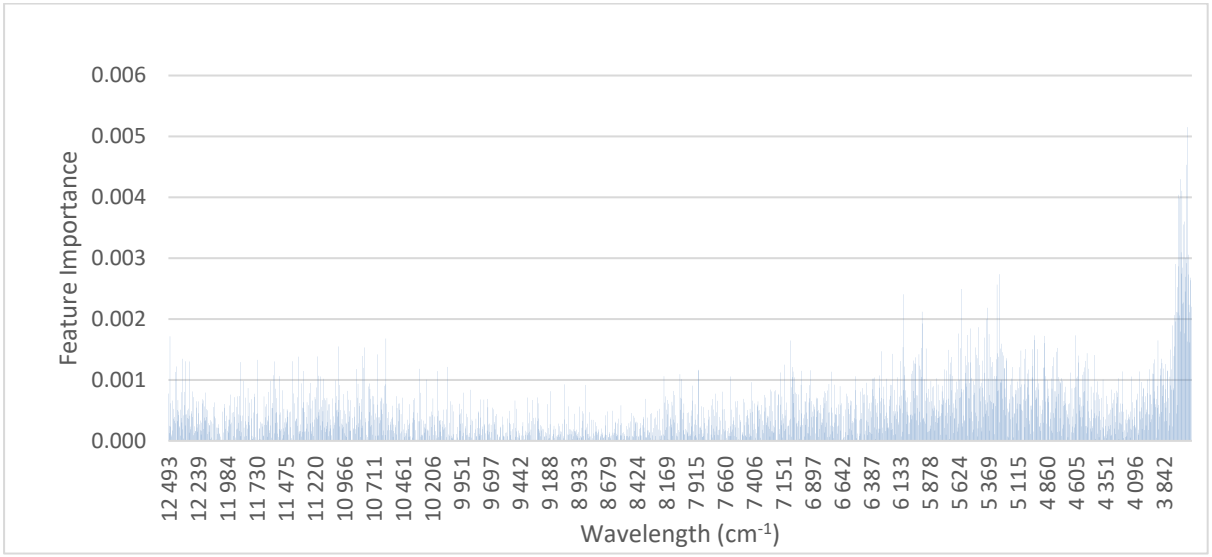


Figure 6-7 RF feature importance calculated while classifying three terrace classes

6.5 DISCUSSION

Two approaches were considered for evaluating the use of soil NIR spectra to differentiate between various fluvial terrace levels. The first employed an established chemometric methodology that uses PCA and PLSR (Bruker 2006) to investigate correlations between soil NIR spectra and mean VDTCN values derived from morphometric primitives. A strong correlation (R^2 = 0.78) was achieved using pre-processed NIR spectra (located between 7 500 – 5 446 cm^{-1}), which remained strong (R^2 = 0.55) after cross-validation. The initial RPD value of 2.11 decreased to 1.49 with cross-validation. While RPD values smaller than 1.5 are typically considered to indicate undesirable calibrations, analysis of heterogeneous materials such as soils necessitate further interpretation of the broader context and purpose when evaluating calibration RPD values (Fearn 2002). Cross-validated RPD values between 1.4 and 2.0 are often considered reasonable and RPD

values larger than 2.0 are considered excellent (Camacho-tamayo, Rubiano & Del Pilar Hurtado 2014; Chang et al. 2001; Minasny et al. 2009). PLSR fit and cross-validated RMSE values of 9.63 m and 13 m indicate that regression is not quite suitable for detecting differences between terrace levels where the vertical difference is less than 10 m. It is likely that the performance of the PLSR has been hindered by the small dataset. A larger data set will therefore produce a stronger correlation.

The second approach utilised unprocessed soil NIR spectra, machine learning (RF) and semi-automated hyper-parameterisation techniques (Pedregosa, Weiss & Brucher 2011). All four experiments, designed to classify fluvial terrace levels and refine the available NIR spectra, produced good validation scores of between 0.73 and 0.76 (Table 6-1). Training scores near 1.0 were recorded for every experiment, which suggests overfitting.

Extensive hyper-parameterisation had a negligible effect on the resultant training scores, which suggests that overfitting may be attributed to data limitations. Nevertheless, specific wavelengths were successfully isolated (Experiments 1b and 2b) through interpretation of feature importance (Experiments 1a and 2a). All four distinct feature importance peaks produced when classifying two fluvial terrace classes (experiment 1a) can be related to established soil NIR spectra characteristics. The peak located at 7 163 cm^{-1} (1 396 nm) can be attributed to the presence of kaolinite given overtones near 7139 cm^{-1} (1400 nm) caused by O-H stretch vibrations near 3 600 cm^{-1} (2 778 nm), which may also explain the peak at 3 664 cm^{-1} (2 729 nm) (Stenberg et al. 2010). Smectite also has characteristic absorptions near 7 139 cm^{-1} (1 400 nm) due to O-H stretch vibrations in its octahedral layer and near 7 139 cm^{-1} (1 400 nm) and 5 261 cm^{-1} (1 900 nm) due to combination vibrations caused by the presence of water in its interlayer lattices, thus explaining the peak located at 5 334 cm^{-1} (1 875 nm) (Bishop, Pieters & Edwards 1994).

The presence of kaolinite is explained by the weathering of the extensive underlying quartzofeldspathic Goudplass gneiss that dominates the upper reaches of the GLR (Anhaeusser 1992; Dippenaar & Van Rooy 2014; Rozanov et al. 2017). The presence of smectite, on the other hand, is attributed to the weathering of mafic dikes that occur among the higher-lying terraces of the GLR, from which material is periodically sourced (Rozanov et al. 2017).

The feature importance peak located at 4 983 cm^{-1} (2 007 nm) could be attributed to the presence of carbonates that present weak absorption features near 4 628 cm^{-1} (2 160 nm) and 5 261 cm^{-1} (1 990 nm) (Clark 1999; Stenberg et al. 2010). The Giyani greenstone outcrops of the GLWC is a

common source of talc, which may explain the presence of carbonates in the fluvial terrace soils (Anhaeusser 1992; Rozanov et al. 2017). Another potential explanation for the peak at 4 983 cm^{-1} (2 007 nm) considers kaolinite or smectite influence at 4 983 cm^{-1}, given that the wavelengths located at 7 163 cm^{-1}, 5 334 cm^{-1} and 4 983 cm^{-1} are significantly correlated (Pearson correlation coefficients > 0.9). Due to these high correlations, only the two most significant wavelengths were retained for further classifications: 3 664 cm^{-1}, representing the kaolinite fraction, and 5 334 cm^{-1}, representing the smectite fraction (Experiment 1b). The refining of the spectra dataset to two wavelengths marked no significant classification accuracy change as both approaches produced validation scores of 0.76 (Table 6-1).

Five wavelengths associated with distinct feature importance peaks were selected when classifying three fluvial terrace classes (Experiment 2a). Three of these wavelengths – located at 6 106 cm^{-1} (1 637 nm), 5 944 cm^{-1} (1 682 nm) and 5 604 cm^{-1} (1 784 nm) – do not, to our knowledge, match NIR spectra characteristics that have been linked to any specific soil properties. The remaining two, 5 269 cm^{-1} (1 897 nm) and 3 699 cm^{-1} (2 703 nm), can be ascribed to the presence of smectite and kaolinite, respectively (Stenberg et al. 2010). Moreover, given that 6 106 cm^{-1} (1 637 nm), 5 944 cm^{-1} (1 682 nm), 5 604 cm^{-1} (1 784 nm) and 5 269 cm^{-1} are significantly correlated (Pearson correlation coefficients > 0.9), the two most significant wavelengths, located at 5 269 cm^{-1} (1 897 nm) and 3 699 cm^{-1} (2 703 nm), were selected for further classification (Experiment 2b). The use of the refined spectra selection increased the validation score of the RF classifier to 0.74, compared to the 0.73 achieved when using the entire NIR spectral dataset (Table 6-1).

Soil clay mineral content appears to be a good predictor of fluvial terrace levels in the LE, given that the RF classification of both two and three terrace classes favoured wavelengths associated with kaolinite and smectite. The formation of clay minerals, including kaolinite and smectite, is heavily influenced by the same environmental factors that influence fluvial terrace formation, i.e. climate and tectonic regimes (Tebbens, Veldkamp & Kroonenberg 1998). Temperature and rainfall significantly influence chemical and mechanical weathering mechanisms and therefore drive weathering intensity and influence mineral composition (Curtis 1990; Irion 1991). Tectonic settings control local topography and, thereby, the rates and products of chemical weathering (Curtis 1990). Since a key factor for the formation of different fluvial terrace levels is varying climate and tectonic conditions, the presence of different clay mineral assemblages on various fluvial terrace levels is to be expected. Clay minerals are among some of the most common chemical weathering products, and kandite (kaolinite) and smectite clays are some of the most

commonly occurring clay mineral groups (Nichols 2009). Therefore, future studies should aim to develop transferable and operational spectroscopy methodologies that exploit clay mineral fraction variations to differentiate between fluvial terrace levels. Moreover, the findings of this study demonstrate how the feature ranking capability of RF allows for the selection and interpretation of significant wavelengths, which is a key advantage of tree-based machine learning approaches (RF) over chemometric analyses.

6.6 CONCLUSION

This study demonstrated the potential of NIR DRS as a soil analytical technique to differentiate between fluvial terrace levels with varying stratigraphic characteristics. First, the relationship between soil spectra to landscape characteristics was outlined based on the soil-landscape paradigm (Jenny 1941; McBratney, Mendonca Santos & Minasny 2003) and the proven integrative nature of soil spectra (Stenberg et al. 2010). Next, using an established chemometric methodology (Bruker 2006), spectra pre-processing, PCA and PLSR were employed to establish a strong correlation (R^2 = 0.78, RPD = 2.11) between a selected spectra subset and VDTCN values representing the various terrace levels. The correlation remained relatively strong after cross-validation (R^2 = 0.55, RPD = 1.49). Thirdly, the Sci-kit Learn RF algorithm (Pedregosa, Weiss & Brucher 2011) was implemented – as both a supervised classification and feature selection technique – to establish a good relationship (validation scores between 0.73 and 0.76) between representative fluvial terrace level classes and specific NIR spectra wavelengths associated with kaolinite and smectite clays. Despite the limitations of the dataset used, the findings provide support for the development of an automated and transferable supervised machine learning tool to differentiate between various terrace levels. Given the strong association of clay minerals with fluvial terrace sediments and the low operational cost of NIR spectroscopy, such a tool will likely facilitate cost-effective and rapid fluvial terrace mapping over large and complex areas.

CHAPTER 7: SYNOPSIS

Mitigation and adaptation strategies are needed to limit adverse effects of global climate, ecosystem and species behaviour changes brought about by anthropogenic activity (Holm et al. 2013; Lewis & Maslin 2005; Steffen et al. 2011). The unavailability of fine-scale geomorphological and soil resource inventories is limiting the sustainable management of natural resources and hindering mitigation and adaptation strategies. This is particularly problematic in the developing countries of southern Africa, where optimal use of soil resources is needed to support economic development and reduce climate-related impacts on agricultural production. In the last century, studies investigating fluvial terraces provided mainly qualitative descriptions of diamondiferous fluvial terraces (Dollar 1998). Despite increasing awareness of the importance of fluvial terraces for climate change vulnerable agriculture and ecosystem services, many fluvial terrace landscapes in southern Africa remain unexplored.

The aim of this study was to develop methodologies for generating fine-scale, object-based fluvial terrace maps. Geomorphometric techniques, object-based land surface segmentation approaches, machine learning algorithms and soil NIR spectroscopy analyses were combined and applied in the GLWC to map fluvial terraces and to better understand the landscape evolution of the South African Lowveld. Four research questions (Chapter 1) were formulated and individually addressed in the preceding chapters (Chapters 3 to 6). Each of the four chapters were prepared as freestanding articles. The findings of these chapters are summarised in the next section. The subsequent sections overview the study's contribution to new knowledge and outline its limitations, recommendations for future research and concluding remarks.

7.1 SUMMARY OF RESEARCH FINDINGS

7.1.1 The Great Letaba fluvial terraces: an investigation into the landscape and soil evolution of the South African Lowveld

Research question 1: To what extent do fluvial terraces dominate the GLWC landscape and how does the local climate and underlying geology influence their morphology and preservation?

This research question relates to the importance of the GLR and its surrounding landscape for agriculture, water management and ecosystem services. Very little is known about the land surface

morphology of GLWC and a better understanding of the natural history of the area is needed to facilitate quantitative environmental and landscape modelling, land use planning, sustainable management of natural resources and the mitigation of land and ecosystem degradation. In response, field reconnaissance, geomorphometric analysis and 1:250 000 scale soil data were employed in Chapter 3 to characterise the GLWC fluvial terraces and planation surfaces and to investigate the terrain, soil, climate and bedrock paradigm (Objective 2).

Interpretation of cross-profiles and two field reconnaissance expeditions demonstrated the widespread distribution and complex morphology of the GLWC fluvial terraces and planation surfaces that dominate the landscape. The fluvial terraces next to the GLR were found to occur within specific relative elevation (HARC) ranges and exhibit distinct soil/sediment characteristics associated with different land type, climate and geomorphic subdivisions of the valley. Fill fluvial terraces located near the river channel in the western parts of the GLWC (near the escarpment) are mostly well-preserved and associated with deep soils. The formation of the deep soils can be attributed to the presence of granite and diorite bedrock, high rainfall conditions and the narrowness of the GLR valley. A downstream transition to gneiss bedrock produces slightly shallower soils under similar climatic conditions. Further East (near the KNP), a drier climate and extensive gneiss bedrock resulted in a much wider valley that is associated with fewer fill fluvial terraces near the GLR channel and much shallower soils. Higher laying fluvial terraces, present as high as the GLWC boundaries, were found to be extensively eroded and mostly removed. Fluvial terrace remnants are consequently mainly visible as interfluves in the landscape. Tributaries play an active role in the dissection of fluvial terrace surfaces and have eroded portions of every fluvial terrace level identified. Most tributaries that cut into higher laying fluvial terraces also produce fluvial terraces, which are also incised by tributaries producing third-order fluvial terraces.

Investigations into the morphology of the planation surfaces failed to locate any intact surfaces. Similar to the higher laying fluvial terraces, only remnants of planation surfaces are visible as interfluves in the landscape that formed during the incision of fluvial terrace-producing tributaries. These remnants were found to stretch along significant portions of the GLR. In some cases, planation surface remnants correlated well with fluvial terrace remnants, which may indicate that the presence of fluvial terraces in the landscape influenced the formation of the African and post-African I planation surfaces of the northern Lowveld.

7.1.2 Object-based land surface segmentation scale optimisation: an ill-structured problem

Research question 2: How can data-driven, object-based land surface segmentation scale optimisation approaches be employed to delineate objects that conform to pre-defined morphological conditions?

Specific geomorphometric approaches segment land surfaces into elementary forms that conform to predefined definitions. But the characteristics of morphometric primitives delimited with discrete geomorphometric approaches are unknown during the segmentation process. Consequently, data-driven SP optimisation techniques based on discrete geomorphometric principals only detect scales inherent to the data and do not necessarily identify SPs that are most appropriate for delineating specific landforms. The failure of data-driven SP optimisation techniques to delimit specific landforms highlights a discrepancy between the morphometric characteristics of the target landforms and the criteria evaluated by SP optimisation techniques. This discrepancy can be attributed to land surface segmentation SP optimisation being an ill-structured problem. Chapter 4 demonstrates how this ill-structured problem can be deconstructed into a set of well-structured problems so that specific morphological conditions can be defined and used to support segmentation parameter decisions (Objective 3).

To serve as a case study, two morphological conditions – i.e. objects must be 1) internally homogenous; and 2) bordered by a discontinuity – were selected to define a "conceptual morphometric primitive." The first condition was evaluated with the use of ESP 2, an established SP optimisation technique that calculates object LV as a measure of object interior homogeneity. OBLV and LVR were adapted for SP optimisation to determine whether the generated objects' boundaries follow morphological discontinuities (the second condition). It was shown that agreement between the SPs selected using ESP2, OBLV and LVR are indicative of SPs that meet both morphometric conditions. This approach is comparable to the visual assessment of segmentation outputs where analysts qualitatively compare MRS results to various conditions or scenarios. The proposed ensemble data-driven SP optimisation approach was useful to quantitatively compare segmentation outputs against predefined conditions and select appropriate SPs that are more objective (i.e. not biased by analyst involvement). Synthesis of the SPs selected by the ensemble showed that it is useful to combine multiple SP optimisation techniques when refining SP selections and that more robust results (compared to when using an individual SP

optimization technique) are obtained. It was also shown that the number of SPs identified are inversely proportional to the number of techniques in the ensemble, which is advantageous because smaller sets of SPs are easier to implement and produce less complex hierarchies of objects.

7.1.3 An object-based methodology for classifying fluvial terraces

Research question 3: What is the value of supervised white box machine learners for constructing fluvial terrace rulesets that are both accurate and comparable to existing expert-knowledge?

Most available geomorphometric fluvial terrace mapping methodologies are implemented within the per-pixel paradigm and use knowledge-based approaches that are deterministic, parametric and dependent on expert knowledge. Non-parametric supervised machine learning approaches are not subject to these limitations. However, the black-box nature of most machine learners excludes expert knowledge from the classification process entirely. Chapter 5 presents a novel object-based fluvial terrace mapping methodology that utilises white-box supervised DT machine learners to produce transparent, interpretable and modifiable rulesets (Objective 4).

The methodology was evaluated in the GLWC where binary classes (terrace and non-terrace) and multiple fluvial terrace levels (non-terraces and the first 10 ordinal terrace levels) were classified using both per-pixel and object-based approaches. Quantitative accuracy assessment and qualitative (visual) analysis of the resultant fluvial terrace maps revealed that the object-based classifiers (validation scores of up to 0.88) performed consistently better than the per-pixel counterparts (validation scores of up to 0.83). Interpretation of the generated DT rulesets shows that the inferior accuracy of the per-pixel classifiers can be attributed to a tendency to produce complex, convoluted and overfitted rulesets. Rulesets produced with object-based classifiers, on the other hand, were found to be less prone to overfitting and produced much simpler rules that are more intuitive and comparable to expert knowledge. The ability of object-based DT rulesets to accurately classify fluvial terraces in a planate landscape, a feat not accomplished by earlier fluvial terrace mapping methodologies, is also demonstrated.

7.1.4 Soil NIR spectroscopy for fluvial terrace level differentiation

Research question 4: Is NIR DRS a viable soil analytical technique with which to differentiate between various fluvial terrace levels and are there any specific wavelengths that can be employed to develop transferable methodologies?

Establishing continuity between fluvial terrace remnants in eroded landscapes has long been a challenge for geomorphologists and soil scientists. While various geomorphometric fluvial terrace mapping techniques have been introduced, many still require extensive user parameterisation, and accurate mapping is often limited to well-preserved morphological settings. As a result, analysts still depend on resource intensive conventional in situ stratigraphic and physiographic interpretations when differentiating between fluvial terrace levels. Chapter 6 outlines a conceptual framework based on the soil-landscape paradigm and soil NIR spectroscopy and illustrates the potential of soil NIR spectroscopy for fluvial terrace level differentiation (Objective 5).

An established chemometric methodology was used to establish a strong cross-validated PLSR correlation (R2 = 0.55, RMSE = 12.9 m and RPD = 1.49) between pre-processed spectra selected using PCA (7 500 – 5 446 cm^{-1}) and the VDTCN of various fluvial terrace levels. Using the unprocessed spectra as input, RF was applied as both a classifier and feature selection tool. Validation scores of 0.76 and 0.73 were achieved when using all the available spectra to classify two and three terrace classes, respectively. The feature importances calculated for both classifications showed that wavelengths associated with clay minerals, carbonates and talc contributed most towards the RF model. Additional classifications were performed using only wavelengths associated with smectite and kaolinite fractions. A validation score of 0.76 was achieved when classifying two terrace classes with the 3 664 cm^{-1} and 5 334 cm^{-1} bands, and a validation score of 0.74 was obtained when classifying three terrace classes with the 3 699 cm^{-1} and 5 269 cm^{-1} bands. The exhibited link between clay mineralogical fraction variations and different fluvial terraces levels can be explained by climatic and tectonic regime changes that heavily influence both clay mineral and fluvial terrace formation.

7.2 CONTRIBUTIONS TO KNOWLEDGE

While many authors explored the fluvial terraces of the southern African paleo-fluvial landscape (Dollar 1998), descriptions of GLWC fluvial terraces and planation surfaces are superficial. Moon & Heritage (2001) mentioned the existence of one fluvial terrace level along the GLR, and Rozanov et al. (2017) considered only three fluvial terrace levels. Likewise, Partridge & Maud (1987) note the preservation of African and post-African I planation surface remnants as interfluves in a broad Lowveld context, leaving the reader to decipher landscape characterisations from 1:2 500 000 scale maps. This study provides new insight into the evolution of the GLWC, and by extension the Lowveld, through comprehensive fluvial terrace descriptions and characterisations of high laying fluvial terrace interfluves that occur up to the GWLC boundary. Moreover, the extensive nature of hierarchically-nested tributary fluvial terrace staircases and scattered planation surface and fluvial terrace interfluves were uncovered. These new insights led to the collection of critical reference data needed to facilitate the development, implementation and evaluation of novel geomorphometric methodologies (Chapter 5 and 6). This new information significantly expands on existing knowledge of GLWC fluvial terrace and planation surface morphologies (Objective 2) and provides a sound foundation for further paleo-fluvial investigations into the Lowveld landscape. This component of the dissertation was submitted as a scientific article to *The South African Geographical Journal*.

Chapter 4 addresses the limitations of existing MRS SP optimisation approaches. Prior to the advent of data-driven SP optimisation techniques, analysts were limited to subjective specific geomorphometric techniques to assess segmentation results and to select SPs that accurately delimit land surfaces (Van Asselen & Seijmonsbergen 2006). Subsequently, Drăguţ, Tiede & Levick (2010) and Drăguţ et al. (2014) conceived the popular data-driven ESP 2 technique within the discrete geomorphometric paradigm (Drăguţ & Eisank 2011) to assist with SP selection. However, ESP 2 only detects data-inherent scales and does not necessarily detect scales at which targeted landforms are accurately delimited (Anders, Seijmonsbergen & Bouten 2011). Moreover, analyst involvement is often needed to select suitable SPs from an extensive set of candidate SPs (Louw, Van Niekerk & Rozanov 2016). This study exposes data-driven SP optimisation as an ill-structured problem and introduces a conceptual framework for a novel ensemble SP optimisation methodology aimed at delimiting predefined landforms, e.g. fluvial terraces (Objective 3). The new methodology was demonstrated by employing ESP 2 along with two newly developed LV-based SP optimisation (OBLV and LVR) techniques for supporting SP identification and selection.

The conceptual framework introduced by this study provides a proof-of-concept and can be used to develop new ensembles of data-driven SP optimisation techniques that can delimit targeted landforms. Chapter 4 was published in the G*eomorphology* journal (Louw & Van Niekerk 2019).

Existing geomorphometric fluvial terrace mapping techniques are mainly based on per-pixel techniques that utilise deterministic rules (Demoulin et al. 2007; Hopkins 2014; Stout & Belmont 2014), statistical slope threshold estimations (Clubb et al. 2017; Li et al. 2019) or machine learners (Swan 2017). As a result, extensive user parameterisation is often needed and the black-box nature of most machine learners exclude expert knowledge from the classification process entirely. Although previous research (Anders, Seijmonsbergen & Bouten 2011; Van Asselen & Seijmonsbergen 2006) have shown that object-based approaches can classify fluvial terrace accurately, no semi-automated object-based machine learning methodology exists. Chapter 5 showcases a fluvial terrace mapping methodology (Objective 4) that couples object-based land surface segmentation approaches, white-box (DT) machine learners and hyper-parameterisation to produce transparent and intuitively interpretable rulesets without the need for analyst-parameterisation. Qualitative and quantitative evaluations showed that the new methodology consistently outperformed per-pixel DT rulesets and confirmed the benefits of object-based approaches for fluvial terrace mapping. Moreover, the utility of inductive machine learning approaches for geomorphometric ruleset development and the potential of white-box machine learning for semantic landform modelling were demonstrated. Chapter 5 has been submitted to the *Computers and Geoscience* journal.

Chapter 6 details a soil NIR spectroscopy and object-based geomorphometry methodology to differentiate between fluvial terrace levels (Objective 5). Eroded fluvial landscapes, such as the GLWC, often only exhibit fluvial terrace remnants that are difficult to differentiate and match to specific terrace staircase levels. Expensive in situ stratigraphic and physiographic interpretations (Leopold, Wolman & Miller 1964) are the mainstay of fluvial terrace level mapping, while geomorphometric techniques (NS Anders, Seijmonsbergen & Bouten 2011; Van Asselen & Seijmonsbergen 2006; Clubb et al. 2017; Demoulin et al. 2007; Stout & Belmont 2014; Swan 2017) rely on expert knowledge and/or reference data. The potential of the soil-landscape paradigm and soil NIR spectroscopy for fluvial terrace level differentiation was illustrated in Chapter 5 by combining an established chemometric approach, a novel methodology based on object-based geomorphometric principles and RF machine learning. Spectral wavelengths associated with clay mineral fractions were uncovered and the presented NIR spectroscopy, object-

based, RF methodology provides a promising framework for a low-cost, semi-automated spectroscopic alternative to resource intensive in situ field studies.

7.3 RESEARCH LIMITATIONS AND RECOMMENDATIONS

No research is without limitations. The lack of fine-scale geomorphological data prevented comprehensive validation of results, and the reliance on desktop methods to collect morphological reference data (Chapters 3 and 5) was not ideal. Nevertheless, two field excursions carried out as part of this research provided invaluable geomorphological context, and the affordability of the desktop methods enabled the collection of extensive reference data that are comparable with expert knowledge derived from literature (Chapter 3) and inductive machine learning approaches (Chapter 5). Moreover, only a limited number of soil samples (88) were available for NIR spectroscopy analyses (Chapter 6). This study would have benefitted from a large set of in situ data for quantitative validations.

Analyst input was needed to successfully group the collected soil samples into representative fluvial terrace level classes, which would be difficult to replicate in study areas where soil maps are not available. Substantial research initiatives are therefore needed to characterise previously unstudied landscapes to enable the collection of reference datasets and large-scale geomorphometric mapping. The collection and open-access hosting of reference datasets would aid in the development and evaluation of transferable geomorphometric rulesets and should be prioritised.

The conceptualisation and development of the SP optimisation ensemble (Chapter 4) were limited by the availability of appropriate LV-based techniques. As such, only a single case study was carried out and "conceptual" morphometric primitives were delimited. Whether the morphometric primitive conditions applied actually delimited real-world landforms is a source of uncertainty that must be taken into consideration. The discrepancy between the criteria evaluated by data-driven SP optimisation techniques and the properties of real-world landforms is related to landform semantics (Eisank, Drăguț & Blaschke 2011) and object ontology (Drăguţ & Eisank 2011). As such, ontological work in specific geomorphometry should aim to produce well-structured landform definitions that can be evaluated by data-driven SP optimisation tools. While Van Niekerk (2010) showed that MRS delimits morphological discontinuities accurately, further work is needed to relate "conceptual" landforms delimited at optimised SPs to real-world landform boundaries.

The evaluation of the conceptualised object-based fluvial terrace classification methodology (Chapter 5) is limited to comparisons with per-pixel approaches and assessments against reference data in a relatively small study site, which may not be representative of the entire catchment area. Further work should include conceptual and performance comparisons against other fluvial terrace mapping strategies (Clubb et al. 2017; Stout & Belmont 2014) in a range of fluvial landscapes that have been subjected to varying degrees of erosion.

The development of an in situ soil NIR spectrometer capable of differentiating between various fluvial terrace levels is a logical progression of this study (Chapter 6), given the strong association of clay minerals with fluvial terrace sediments and the low operational cost of NIR spectroscopy and machine learning software (Pedregosa, Weiss & Brucher 2011). Nonetheless, more work is needed to corroborate the utility of the soil-landscape paradigm for fluvial terrace differentiation and to assess methodological transferability to other study areas.

7.4 CONCLUSIONARY REMARKS

Fluvial terraces are common occurring landforms in paleo-fluvial landscapes world-wide (Bridgland & Westaway 2008). Diamondiferous fluvial terraces along southern African rivers have been extensively characterised in the last century (Dollar 1998), while the bulk of fluvial terraces have not received the same level of research attention. An increasing awareness of the importance of fluvial terraces for ecosystem services and agriculture, as well as a need for fine-scale resource inventories – critical for land use planning and global change adaptation and mitigation strategies, necessitate the development of fine-scale fluvial terrace maps and transferable geomorphometric landform mapping methodologies.

This dissertation developed novel object-based fluvial terrace mapping methodologies by combining and applying geomorphometric techniques, object-based land surface segmentation approaches, machine learning algorithms and soil NIR spectroscopic analyses to investigate the fluvial terraces and landscape evolution of the GLWC. Existing knowledge about the morphology and landscape evolution of the GLWC, a typical Lowveld landscape, was expanded through characterisation of fluvial terrace and planation surface landforms. Land surface MRS SP optimisation was exposed as an ill-structured problem and a novel conceptual methodology – that draws from the convergence of specific and discrete geomorphometry principles – was introduced to facilitate ensemble data-driven delimitation of predefined morphological features. An object-based white-box (DT) machine learner methodology for the construction of intuitive rulesets that

are comparable to existing (expert-based) field-based models was developed and demonstrated. Qualitative and quantitative comparisons with per-pixel rulesets confirmed the benefits of object-based classification approaches. The resulting methodology is universally applicable and can potentially be modified for mapping other landforms. The efficacy of soil NIR spectroscopy for fluvial terrace level differentiation was demonstrated and specific wavelengths associated with clay mineral fractions were successfully isolated, thereby providing a proof-of-concept for the development of a low-cost spectrometer that can be used for in-field application.

A resurgence in southern African landscape investigations are needed to support mitigation and adaptation strategies that aim to limit the adverse effects of global change during the Anthropocene. The geomorphological contributions and geomorphometrical methodologies presented in this study provide a basis for regional mapping of fluvial terraces in support of critically-needed fine-scale environmental resource inventories.

REFERENCES

Anders N, Seijmonsbergen H, Bouten W & Smith M 2011. *Optimizing object-based image analysis for semi-automated geomorphological mapping*. Proceedings of the Geomorphometry conference held 8 September 2011. Redlands: USA.

Anders N, Seijmonsbergen AC & Bouten W 2011. Segmentation optimization and stratified object-based analysis for semi-automated geomorphological mapping. *Remote Sensing of Environmen*t 115, 12: 2976–2985.

Anders N, Seijmonsbergen AC & Bouten W 2015. Rule Set Transferability for Object-Based Feature Extraction: An Example for Cirque Mapping. *Photogrammetric Engineering & Remote Sensing* 81, 6: 507–514.

Anhaeusser CR 1992. Structures in granitoid gneisses and associated migmatites close to the granulite boundary of the Limpopo Belt. South Africa. *Precambrian Research* 55: 81–92.

Araújo SR, Wetterlind J, Demattê JAM & Stenberg B 2014. Improving the prediction performance of a large tropical vis-NIR spectroscopic soil library from Brazil by clustering into smaller subsets or use of data mining calibration techniques. *European Journal of Soil Science* 65: 718–729.

Baatz M & Schäpe A 2000. *Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation*. In Strobl J, Blaschke, T & Griesebner, G (eds). Angewandte Geographische Informationsverarbeitung 2000. Herbert Wichmann Verlag, Salzburg: 12–23.

Bauer RJ & Dahlquist JR 1999. *Technical market indicators: Analysis & performance*. New York: Wiley.

Biondi E, Zivkovic L, Esposito L & Pesaresi S 2009. Vegetation, plant landscape and habitat analyses of a fluvial ecosystem in central Italy. *Acta Botanica Gallica* 156, 4: 571–587.

Bishop JL, Pieters CM & Edwards JO 1994. Infrared spectroscopic analyses on the nature of water in montmorillonite. *Clays and Clay Minerals* 42: 702–716.

Blaschke T & Strobl J 2015. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *Zeitschrift für Geoinformationssysteme*, Issue 6: 12–17.

Blaschke T 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 65, 1: 2–16.

Blaschke T, Hay GJ, Kelly M, Lang S, Hofmann P, Addink E, Feitosa RQ, Van Der Meer F, Van Der Werff H, Van Coillie F & Tiede D 2014. Geographic Object-Based Image Analysis – Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing* 87: 180–191.

Blum MD & Törnqvist TE 2000. Fluvial responses to climate and sea-level change: a review and look forward. *Sedimentology* 47: 2–48.

Böhner J & Selige T 2006. Spatial prediction of soil attributes using terrain analysis and climate regionalisation. *Göttinger Aeographische Abhandlungen* 115: 13–28.

Bolongaro-Crevenna A, Torres-Rodríguez V, Sorani V, Frame D & Ortiz MA 2005. Geomorphometric analysis for characterizing landforms in Morelos State, Mexico. *Geomorphology* 67: 407–422.

Botha GA & De Wit MCJ 1996. Post-gondwanan continental sedimentation, Limpopo region, southeastern Africa. *Journal of African Earth Sciences* 23, 2: 163–187.

Bou Kheir R, Bøcher PK, Greve MB & Greve MH 2010. The application of GIS based decision-tree models for generating the spatial distribution of hydromorphic organic landscapes in relation to digital terrain data. *Hydrology and Earth System Sciences* 14, 6: 847–857.

Brändli M 1996. *Hierarchical models for the definition and extraction of terrain features*. In Burrough PA & Frank AU (eds). Geographic Objects with Indeterminate Boundaries, 257–270. London: Taylor & Francis.

Breiman L 2001. Random Forests. *Machine Learning* 45: 5–32.

Breiman L, Friedman J, Olshen R & Stone C 1984. *Classification and regression trees*. Boca Raton: Chapman and Hall/CRC.

Bridgland D & Westaway R 2008. Climatically controlled river terrace staircases: A worldwide Quaternary phenomenon. *Geomorphology* 98: 285–315.

Brown DG, Lusch DP & Duda KA 1998. Supervised classification of types of glaciated landscapes using digital elevation data. *Geomorphology* 21: 233–250.

Bruker 2006. Opus Spectroscopy Software: Quant. Bruker. [Online Help]. Online available: http://files.nocnt.ru/hardware/science/senterra/opus65-doc-en/quant.pdf

Bucher WH 1932. "Strath" as a geomorphic term. *Science* 75: 130–131.

Budd TA 2001. *An Introduction to Object-Oriented Programming*. 3rd ed. Corvallis: Oregon State University.

Burroughs PP & McDonnel R 1998. *Principles of GIS*. London: Oxford Univeristy Press.

Butzer K, Helgren D, Fock G & Stuckenrath R 1974. The Alluvial Terraces of the Lower Vaal River, South Africa: A Reappraisal and Reinvestigation. *The Journal of Geology* 82, 5: 665–667.

Camacho-tamayo JH, Rubiano Y & Del Pilar Hurtado M 2014. Near-infrared (NIR) diffuse reflectance spectroscopy for the prediction of carbon and nitrogen in an Oxisol. *Agronomía Colombiana* 32, 1: 86–94.

Campbell J & Wynne R 2011. *Introduction to remote sensing*. 5th ed. London: The Guilford Press.

Campbell MR 1929. The river system, a study in the use of technical geographic terms. *Journal of Geography* 28: 123–128.

Chang C, Laird D, Mausbach MJ & Hurburgh C 2001. Near-Infrared Reflectance Spectroscopy – Principal Components Regression Analyses of Soil Properties. *Soil Science Society of America Journal* 65, 2: 480–490.

Charlton R 2008. *Fundamentals of fluvial geomorphology*. London: Taylor & Francis.

Clark RN 1999. *Spectroscopy of rocks and minerals and principles of spectroscopy*. In Rencz AN (ed) Remote Sensing for the Earth Sciences, 3–58. Chichester, UK: John Wiley & Sons.

Clubb FJ, Mudd SM, Milodowski DT, Valters DA, Slater LJ, Hurst MD & Limaye AB 2017. Geomorphometric delineation of floodplains and terraces from objectively defined topographic thresholds. *Earth Surface Dynamics* 5: 369–385.

Curtis CD 1990. Aspects of climatic influence on the clay mineralogy and geochemistry of soils, palaeosols and clastic sedimentary rocks. *Journal of the Geological Society of London* 147: 351–357.

D'Oleire-Oltmanns S, Eisank C, Drăgut L & Blaschke T 2013. An object-based workflow to extract landforms at multiple scales from two distinct data types. *IEEE Geoscience and Remote Sensing Letters* 10, 4: 947–951.

Dalrymple JB, Blong RJ & Conacher AJ 1968. A hypothetical nine unit landsurface model. *Zeitschrift für Geomorphologie* 12: 60–76.

De Santana F, De Souza A & Poppi R 2018. Visible and near infrared spectroscopy coupled to random forest to quantify some soil quality parameters. *Molecular and Biomolecular Spectroscopy* 191: 454–462.

De Wit MC. 1996. The distribution and stratigraphy of inland alluvial diamond deposits in South Africa. *African Geoscience Review* 3: 175–189.

De Wit MC. 2004. The diamondiferous sediments on the farm Nooitgedacht (66), Kimberley South Africa. *South African Journal of Geology* 107, 4: 477–488.

Dehn M, Gärtner H & Dikau R 2001. Principles of semantic modeling of landform structures. *Computers & Geosciences* 27, 8: 1005–1010.

Demoulin A, Bovy B, Rixhon G & Cornet Y 2007. An automated method to extract fluvial terraces from digital elevation models: The Vesdre valley, a case study in eastern Belgium. *Geomorphology* 91, 1–2: 51–64.

Dikau R 1992. Aspects of constructing a digital geomorphological base map. *Geologisches Jahrbuch A*, 122: 357–370.

Dippenaar MA & Van Rooy JL 2014. Review of engineering, hydrogeological and vadose zone hydrological aspects of the Lanseria Gneiss, Goudplaats - Hout River Gneiss and Nelspruit Suite Granite (South Africa). *Journal of African Earth Sciences* 91: 12–31.

Dollar ESJ 1998. Palaeofluvial geomorphology in southern Africa: a review. *Progress in Physical Geography* 22, 3: 325–349.

Dornik A, Drăguţ L & Urdea P 2017. Classification of soil types using geographic object-based image analysis and Random Forest. *Pedosphere* 28, 6: 913-925.

Drăguţ L & Blaschke T 2006. Automated classification of landform elements using object-based image analysis. *Geomorphology* 81: 330–344.

Drăguţ L & Dornik A 2013. Land-Surface Segmentation as sampling framework for soil mapping [online]. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.644.4666&rep=rep1&type=pdf [Accessed 5 August 2020].

Drăguţ L & Eisank C 2011. Object representations at multiple scales from digital elevation models. *Geomorphology* 129: 183–189.

Drăguţ L & Eisank C 2012. Automated object-based classification of topography from SRTM data. *Geomorphology* 141: 21–33.

Drăguţ L, Csillik O, Eisank C & Tiede D 2014. Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS Journal of Photogrammetry and Remote Sensing* 88: 119–127.

Drăgut L, Eisank C & Strasser T 2011. Local variance for multi-scale analysis in geomorphometry. *Geomorphology* 130: 162–172.

Drăguţ L, Eisank C, Strasser T & Blaschke T 2009. A comparison of methods to incorporate scale in Geomorphometry. Proceedings of the Geomorphometry 2009 conference held 31 August – 2 September 2009. Zurich: Switzerland.

Drăguţ L, Schauppenlehner T, Muhar A, Strobl J & Blaschke T 2009. Optimization of scale and parametrization for terrain segmentation: An application to soil-landscape modelling. *Computers and Geosciences* 35, 9: 1875–1883.

Drăguţ L, Tiede D & Levick SR 2010. ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science* 24, 6: 859–871.

Du Toit A. 1910. The evolution of the river system of Griqualand West. *Transactions of the Royal Society of South Africa* 1: 347-362.

Du Toit A. 1922. A physiographic sketch of the Vaal River. *South African Irrigation Department Magazine*.

Du Toit A. 1933. Crustal movement as a factor in the geographical evolution of South Africa. *South African Geographical Journal* 16: 3–20.

Eisank C & Drăguţ L 2010. Developing a semantic model of glacial landforms for object-based terrain classification–the example of glacial cirques. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38: 1–6.

Eisank C, Drăguț LD & Blaschke T 2011. A generic procedure for semantics-oriented landform classification using object-based image analysis. Proceedings of the Geomorphometry conference held 8 September 2011. Redlands: USA.

Eisank C, Smith M & Hillier J 2014. Assessment of multiresolution segmentation for delimiting drumlins in digital elevation models. *Geomorphology* 214: 452–464.

ESRI 2016. ArcGIS – The Complete Geographic Information System. [Online Help]. Online available: https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/how-slope-works.htm.

Evans IS 1972. *General geomorphometry, derivatives of altitude, and descriptive statistics*. In Chorley RJ (ed) Spatial Analysis in Geomorphology, 17–91. London: Methuen.

Evans IS 1980. An integrated system of terrain analysis and slope mapping. *Zeitschrift für Geomorphologie, Supplementband* 101, 36: 274–295.

Evans IS 1987. *The morphometry of specific landforms*. In Gardiner V (ed) International Geomorphology Part 2, 105–124. John Wiley.

Evans IS 2012. Geomorphometry and landform mapping: What is a landform? *Geomorphology* 137, 1: 94–106.

Fairbridge RW (ed) 1968. *Encyclopaedia of Geomorphology*. New York: Reinhold.

Fearn T 2002. Assessing calibrations: SEP, RPD, RER and R2. *NIR News* 13, 6: 12–14.

Fisher P 1997. The pixel: A snare and a delusion. *International Journal of Remote Sensing* 18, 3: 679–685.

Florin V 2019. Landform identification from surface networks. Master's thesis. Politecnico di Milano School.

Flynn T, Rozanov A, Clercq W De, Warr B & Clarke C 2019. Semi-automatic disaggregation of a national resource inventory into a farm- scale soil depth class map. *Geoderma* 337: 1136–1145.

Gao Y, Mas JF, Kerle N & Navarrete Pacheco JA 2011. Optimal region growing segmentation and its effect on classification accuracy. *International Journal of Remote Sensing* 32, 13: 3747–3763.

Gerçek D, Toprak V & Strobl J 2011. Object-based classification of landforms based on their local geometry and geomorphometric context. *International Journal of Geographical Information Science* 25, 6: 1011–1023.

Gilbertson JK & Niekerk A Van 2017. Value of dimensionality reduction for crop differentiation with multi- temporal imagery and machine learning. *Computers and Electronics in Agriculture* 142: 50–58.

Goudie AG (ed) 2004. *Encyclopaedia of Geomorphology*. 1st ed. London: Taylor & Francis Group.

Hall DJ & Khanna DK 1977. *The ISODATA method computation for the relative perception of similarities and differences in complex and real data*. In Enslein K, Ralston A & Wilf HS (eds) Statistical Methods for Digital Computers, 340–373. New York: John Wiley & Sons.

Harman G 2018. *Object-orientated ontology*. New York: Penguin.

Hattingh J 1994. Depositional environment of some gravel terraces in the Sundays River valley, Eastern Cape. *South African Journal of Geology* 97: 156–166.

Hattingh J 1996. Fluval response to allocycle influences during the development of the lower Sundays River, Eastern Cape, South Africa. *Quaternary International* 33: 3–10.

Hay GJ, Castilla G, Vi C & Vi WG 2006. Object-based image analysis: strengths, weaknesses, opportuinities and threats (SWOT). Proceedings of the first International Conference on OBIA held 4-5 July 2006, Salzburg University, Austria.

Helgren DM & Butzer KW 1974. Alluvial Terraces of the Lower Vaal River, South Africa: A Reappraisal and Reinvestigation. *The Journal of Geology* 82, 5: 665–667.

Hengl T 2006. Finding the right pixel size. *Computers & Geosciences* 32, 9: 1283–1298.

Heritage GL, Moon BP & Large AR. 2001. The February 2000 floods on the Letaba River, South Africa: an examination of magnitude and frequency. *Koedoe* 44, 2: 1–6.

Heung B, Chak H, Zhang J, Knudby A, Bulmer CE & Schmidt MG 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma* 265: 62–77.

Ho TK 1995. Random Decision Forests. Proceedings of the third International Conference on Document Analysis and Recognition, Montreal, Canada.

Holm P, Goodsite M, Cloetingh S, Agnoletti M, Moldan B, Lang DJ, Leemans R, Moeller JO, Buendia MP, Pohl W, Scholz RW, Sors A, Vanheusden B, Yusoff K & Zondervan R 2013. Collaboration between the natural, social and human sciences in Global Change Research. *Environmental Science & Policy* 28: 25–35.

Hopkins A 2014. A Comparison of DEM-based methods for fluvial terrace mapping and sediment volume calculation: Application to the Sheepscot River Watershed, Maine. Master's thesis. Department of Earth and Environmental Sciences, The Graduate School of Arts and Sciences, Boston College.

Hutchinson MF 1989. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology* 106: 211–232.

Irion G 1991. *Minerals in rivers*. In Degens ET, Kempe S & Richey JE (eds) Biogeochemistry of Major World Rivers. Hamburg: Wiley.

Jasiewicz J & Stepinski T 2013. Geomorphons — a pattern recognition approach to classi fi cation and mapping of landforms. *Geomorphology* 182: 147–156.

Jenny H 1941. *Factors of soil formation*. New York: McGraw-Hill.

Kim M, Madden M & Warner T 2008. *Estimation of optimal image object size for the segmentation of forest stands with multispectral IKONOS imagery*. In Blaschke T, Lang S & Hay G (eds). Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications, 291–307.

King LC 1944. On palaegeography. *South African Geographical Journal* 26: 1–13.

King LC 1963. *South African Scenery, a Textbook of Geomorphology*. 3rd ed. Edinburgh: Oliver and Boyd.

Kitchener KS & King PM 1981. Reflective judgment: concepts of justification and their relationship to age and education. *Journal of Applied Developmental Psychology* 2: 89–116.

Knight J, Mitchell WA & Rose J 2011. Geomorphological field mapping. In Developments in earth surface processes, Volume 15. Elsevier.

Krige AV 1927. An examination of the Tertiary and Quaternary changes of sea level in South Africa with special stress on the evidence in favour of a recent world-wide sinking of ocean level. *Annals of the University of Stellenbosch* 5.

Laliberte AS, Fredrickson EL & Rango A 2007. Combining decision trees with hierarchical object-oriented image analysis for mapping arid rangelands. *Photogrammetric Engineering and Remote Sensing* 73, 2: 197–207.

Land Type Survey Staff 2006. *Land types of South Africa: Digital map (1:250 000 scale) and soil inventory datasets*. Pretoria: Agriculture Research Council, Institute for Soil, Climate and Water.

Lary DJ, Alavi AH, Gandomi AH & Walker AL 2016. Machine learning in geosciences and remote sensing. *Geoscience Frontiers* 7, 1: 3–10.

Lastoczkin AN 1987. *Morfodynamicheskiy Analiz*. Leningrad: Nedra.

Leopold LB, Wolman MG & Miller JP 1964. *Fluvial processes in geomorphology*. New York: Dover publications, Inc.

Leshchinskiy S, Mashchenko E, Ponomareva E, Orlova L, Burkanova E, Konovalova V, Teterina I & Gevlya K 2006. Multidisciplinary paleontological and stratigraphic studies at Lugovskoe (2002-2004). *Archaeology Ethnology and Anthropology of Eurasia* 25: 54–69.

Letham B, Rudin C & McCormick TH 2012. Building Interpretable Classifiers with Rules using Bayesian Analysis. Publication No tr609. University of Washington: Department of Statistics.

Lewis SL & Maslin MA 2005. Defining the Anthropocene. *Nature* 519: 171–180.

Lewis-Williams J 1981. *Believing and seeing: symbolic meanings in southern San rock paintings*. London: Academic Press.

Li H, Chen L, Wang Z & Yu Z 2019. Mapping of River Terraces with Low-Cost UAS Based Structure-from-Motion Photogrammetry in a Complex Terrain Setting. *Remote Sensing* 11, 464.

Li S, Xiong L, Tang G & Strobl J 2020. Deep learning-based approach for landform classification from integrated data sources of digital elevation model and imagery. *Geomorphology* 354, 2–14.

Li Y & Chen W 2020. Landslide Susceptibility Evaluation Using Hybrid Integration of Evidential Belief Function and Machine Learning Techniques. *Water* 12, 113: 1–29.

Lillesand T, Kiefer R & Chipman J 2008. *Remote sensing and image interpretation*. Hoboken.

Liu D & Xia F 2010. Assessing object-based classification: advantages and limitations. *Remote Sensing Letters* 1, 4: 187–194.

Loh W 2011. Classification and regression trees. *Mining and Knowledge Discovery* 1, 1: 14–23.

Loh W 2014. Fifty Years of Classification and Regression Trees. *International Statistical Review* 82, 3: 329–348.

Louw G, Rozanov A, Wiese L & Van Niekerk A 2021. Soil NIR-spectroscopy and object-based land surface segmentation for fluvial terrace level differentiation. *Geomorphology*: 107668.

Louw G & Van Niekerk A 2019. Object-based land surface segmentation scale optimisation: An ill-structured problem. *Geomorphology* 327: 377–384.

Louw G., Van Niekerk A & Rozanov A 2016. *Object-based symmetric difference for land surface segmentation scale parameter optimisation*. Proceedings of the GEOBIA 2016: Solutions and Synergies conference held 14-16 September 2016. Enschede: Faculty of Geo-Information and Earth Observation (ITC).

Loyola-Gonzalez O 2019. Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses From a Practical Point of View. *IEEE Access* 7: 154096–154113.

Luo W & Liu C 2018. Innovative landslide susceptibility mapping supported by geomorphon and geographical detector methods. *Landslides* 15: 465–474.

MacMillan RA, Pettapiece WW, Nolan SC & Goddard TW 2000. A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. *Fuzzy sets and systems* 113, 1: 81–109.

MacVicar CN 1977. *Soil classification: A binomial system for South Africa*. Pretoria: Department of Agricultural Technical Services.

Maddy D 1997. Uplift-driven valley incision and river terrace formation in southern England. *Journal of Quaternary Science* 12: 539–545.

Maddy D, Bridgland D & Green C 2000. Crustal uplift in southern England; evidence from the river terrace records. *Geomorphology* 33: 167–181.

Marshall TR & Baxter-Brown R 1995. Basic principles of alluvial diamond exploration. *Journal of Geochemical Exploration* 53: 277–292.

Mashimbye ZE, De Clercq WP & Van Niekerk A 2014. An evaluation of digital elevation models (DEMs) for delineating land components. *Geoderma* 213: 312–319.

Masseroli A, Bollati IM, Proverbio SS, Pelfini M & Trombino L 2020. Soils as a useful tool for reconstructing geomorphic dynamics in high mountain environments: The case of the Buscagna stream hydrographic basin (Lepontine Alps). *Geomorphology* 28:107442.

Maufe HB 1930. Changes of climate in Southern Rhodesia during geological times. *South African Geographical Journal* 8: 12–16.

Maxwell AE, Warner TA & Fang F 2018. Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing* 39, 9: 2784–2817.

McBratney A, Mendonca Santos M & Minasny B 2003. On digital soil mapping. *Geoderma* 117: 3–52.

Miller CE 2001. *Chemical principles of near-infrared technology*. In Williams P & Norris K (eds) Near-Infrared Technology in the Agricultural and Food Industries, 19–37. St. Paul, The American Association of Cereal Chemists Inc.

Minár J & Evans IS 2008. Elementary forms for land surface segmentation: The theoretical basis of terrain analysis and geomorphological mapping. *Geomorphology* 95: 236–259.

Minár J, Bandura P, Holec J, Popov A & Evans IS 2018. Physically-based land surface segmentation: Theoretical background and outline of interpretations. *PeerJ Preprints* 6: 1–4.

Minasny B & Mcbratney AB 2015. Digital soil mapping: A brief history and some lessons. *Geoderma* 264: 301–311.

Minasny B, Tranter G, Mcbratney AB, Brough DM & Murphy BW 2009. Regional transferability of mid-infrared diffuse reflectance spectroscopic prediction for soil chemical properties. *Geoderma* 153: 155–162.

Mithan HT, Hales TC & Cleall PJ 2019. Supervised classification of landforms in Arctic mountains. *Permafrost and Periglac Process* 30: 131–145.

Moon BP & Heritage GL 2001. The contemporary geomorphology of the Letaba River in the Kruger National Park. *Koedoe* 44, 1: 45–56.

Mortimore JL, Marshall LJR, Almond MJ, Hollins P & Matthews W 2004. Analysis of red and yellow ochre samples from Clearwell Caves and Çatalhöyük by vibrational spectroscopy and other techniques. *Molecular and Biomolecular Spectroscopy* 60, 5: 1179–1188.

Mouazen A, Kuang M, De Baerdemaeker B. & Ramon H 2010. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma* 158: 23–31.

Mulder V 2013. Spectroscopy-supported digital soil mapping. Doctoral dissertation. Wageningen University: C.T. de Wit Graduate School of Production Ecology & Resource Conservation.

Nawar S & Mouazen AM 2019. Soil & Tillage Research On-line vis-NIR spectroscopy prediction of soil organic carbon using machine learning. *Soil & Tillage Research* 190: 120–127.

Nguuri T, Gore J, James D, Webb S, Wright C, Zengeni T, Gwavava O, Snoke J & Kaapvaal Seismic Group 2001. Crustal structure beneath southern Africa and its implications for the formation and evolution of the Kaapvaal and Zimbabwe cratons. *Geophysical Research Letters* 28: 2501 – 2504.

Nichols G 2009. *Sedimentology and stratigraphy*. 2nd edition. Chichester: John Wiley & Sons.

Nocita M, Stevens A, Van Wesemael B, Aitkenhead M, Bachmann M, Barth B, Csorba A, Dardenne P, Demattê JAM & Genot V 2015. Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. *Advances in Agronomy* 132: 139-159.

Oguchi T 2019. Geomorphological mapping based on DEMs and GIS: A review. Proceedings of the 29th International Cartographic Conference held 15-20 July 2019. Tokyo, Japan.

Olaya V & Conrad O 2009. Geomorphometry in SAGA. *Developments in soil science* 33, 08: 293–308.

Partridge TC & Maud RR 1987. Geomorphic evolution of southern Africa since the Mesozoic. *South African Journal of Geology* 90, 2: 179–208.

Partridge TC, Dollar ES., Moolman J & Dollar LH 2010. The geomorphic provinces of South Africa, Lesotho and Swaziland: A physiographic subdivision for earth and environmental scientists. *Transactions of the Royal Society of South Africa* 65, 1: 1–47.

Paterson DG, Nell JP & Seabi FT 2011. Soil investigation of existing orchard areas, Letaba Estates, near Tzaneen. Office report. Pretoria: Agriculture research council, Institute for Soil, Climate and Water.

Paterson G, Turner D, Wiese L, Van Zijl G, Clarke C & Van Tol J 2015. Spatial soil information in South Africa: Situational analysis, limitations and challenges. *South African Journal of Science* 111: 1–7.

Pazzaglia FJ 2013. *Fluvial terraces*. In Shroder J & Wohl E (eds) Treatise on Geomorphology, 379–412. San Diego: Academic Press.

Pedregosa F, Weiss R & Brucher M 2011. Scikit-learn: Machine Learning in Python. *The Journal of machine Learning research* 12: 2825–2830.

Pike RJ & Dikau R (eds) 1995. Advances in geomorphometry. *Zeitschrift für Geomorphologie, Supplementband* 101: 221–238.

Pike RJ, Evans IS & Hengl T 2009. *Geomorphometry: A Brief Guide*. In Hengl T & Reuter H (eds) Developments in Soil Science. Amsterdam: Elsevier.

Quinlan JR 1986. Induction of decision trees. *Machine Learning* 1, 1: 81–106.

Quinlan JR 1987. Generating production rules from decision trees. Proceedings of the Tenth International Joint Conference on Artificial Intelligence held 23-28 August 1987.

Reitmann W 1965. *Cognition and thought.* New York: Wiley.

Reuter HI, Wendroth O & Kersebaum KC 2006. Optimisation of relief classification for different levels of generalisation. *Geomorphology* 77: 79–89.

Rizzo R, Demattê JAM, Lepsch IF, Gallo BC & Fongaro CT 2016. Digital soil mapping at local scale using a multi-depth Vis – NIR spectral library and terrain attributes. *Geoderma* 274: 18–27.

Rogers AW 1903. The geological history of the Gouritz River system. *Transactions of the South African Philosophical Society* 2: 375-385.

Rogers AW 1922. The post Cretaceous climates of South Africa. *South African Journal of Science* 19: 233–239.

Rossel RAV & Mcbratney AB 2008. *Diffuse Reflectance Spectroscopy as a Tool for Digital Soil Mapping*. In Ahrens, R. J. (ed) Digital Soil Mapping with Limited Data, 165–172. Netherlands: Springer.

Rossel RAV, Walvoort DJJ, Mcbratney AB, Janik LJ & Skjemstad JO 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131: 59–75.

Rozanov A, Lessovaia S, Louw G, Polekhovsky Y & de Clercq W 2017. Soil clay mineralogy as a key to understanding planation and formation of fluvial terraces in the South African Lowveld. *Catena* 156: 375–382.

Ruddock A 1945. Terraces in the lower part of the Sundays River valley, Cape Province. *Transactions of the Royal Society of South Africa* 31: 347–370.

Ruddock A 1957. A note on the relation between Chelles-Archeul implements and Quaternary river terraces in the valleys of the Coega and Sundays Rivers, Cape Province. *South African Journal of Science* 53: 373–380.

Ruddock A 1968. Cenozoic sea-levels and diastrophism in a region bordering Algoa Bay. *Transactions of the Geological Society of South Africa* 71: 209–233.

Rudin C 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1: 206–215.

Saldana A & Ibanez J 2004. Pedodiversity analysis at large scales: an example of three fluvial terraces of the Henares River (central Spain). *Geomorphology* 62: 123–138.

Sanchez PA, Ahamed S, Carré F, Hartemink AE, Hempel J, Huising J, Lagacherie P, Mcbratney AB, Mckenzie NJ, De M, Mendonça-santos L, Minasny B, Montanarella L, Okoth P, Palm CA, Sachs JD, Shepherd KD, Vågen T, Vanlauwe B, Walsh MG, Winowiecki LA & Zhang G 2009. Digital soil map of the world. *Science* 325: 680–681.

Savigear RAG 1965. A technique of morphological mapping. *Annals of the Association of American Geographers* 55: 514–538.

Seijmonsbergen AC, Hengl T & Anders NS 2011. *Semi-Automated Identification and Extraction of Geomorphological Features Using Digital Elevation Data*. In Developments in earth surface processes, Volume 15: 297-335. Elsevier.

Shand SJ 1913. The terraces of the Eerste River at Stellenbosch. *Transactions of the Geological Society of South Africa* 16: 147–155.

Shaw GW 1872. Vaal River diamond gravels. *Quarterly Journal of the Geological Society of London* 38.

Sherman DM & Waite TD 1985. Electronic spectra of Fe3+ oxides and oxide hydroxides in the near IR to near UV. *American Mineralogist* 70, 11: 1262–1269.

Simon HA 1973. The structure of ill-structured problems. *Artificial Intelligence* 4: 181–201.

Skidmore A, Watford F, Luckananurug P & Ryan P 1996. An operational GIS expert system for mapping forest soils. *Photogrammetric Engineering & Remote Sensing* 62: 501–511.

Smith G & Morton D 2008. Segmentation: The Achilles' heel of object–based image analysis? *International archives of the photogrammetry, remote sensing and spatial information sciences* 38: 1–6.

Soil Classification Working Group 1991. Soil classification: a taxonomic system for South Africa. Pretoria: Department of Agricultural Development

Steffen W, Grinevald J, Crutzen P & McNeill J 2011. The Anthropocene: conceptual and historical. *Philosophical transactions of the Royal Society A* 369: 842–867.

Stefik M & Bobrow DG 1985. Object-Oriented Programming: Themes and Variations. *AI Magazine* 6, 4: 40–62.

Stenberg B, Rossel RAV, Mouazen AM & Wetterlind J 2010. *Visible and near infrared spectroscopy in soil science*. In Sparks D (ed) Advances in Agronomy, 1–44. Burlington: Academic Press.

Stepinski T, Ghosh S & Vilalta R 2006. Automatic Recognition of Landforms on Mars Using Terrain Segmentation and Classification. Proceedings of the ninth International Conference on Discovery Science (pp. 255-266) held 7 – 10 October 2006. Barcelona: Spain.

Stoner E & Baumgardner M 1981. Characteristic variations in reflectance of surface soils. *Soil Science Society of America Journal* 45, 6: 5.

Stout JC & Belmont P 2014. TerEx Toolbox for semi-automated selection of fluvial terrace and floodplain features from lidar. *Earth Surface Processes and Landforms* 39, 5: 569–580.

Swan BT 2017. Adaptation and Comparison of Machine Learning Methods for Geomorphological Mapping and Terrace Prediction: A Case Study of the Buffalo River, Arkansas. Master's thesis. Auburn University.

Taljaard MS 1944. On induced maturity in streams. *South African Geographical Journal* 26: 14–17.

Tebbens L, Veldkamp A & Kroonenberg S 1998. The impact of climate change on the bulk and clay geochemistry of fluvial residual channel infillings: the Late Weichselian and Early Holocene River Meuse sediments (The Netherlands). *Journal of Quaternary Science* 13: 345–356.

Tou JT & Gonzalez RC 1994. *Pattern Recognition Principles*. Massachusetts: Addison-Wesley Publishing Company.

Trevor TG, Mellor ET & Kynaston H 1908. Reconnaissance of the north west Zoutpansberg District. Office report. Transvaal Mines Department.

Trimble 2014a. eCognition Developer 9.0 Reference Book. [Online Help]. Online available: https://docs.ecognition.com/v9.5.0/Page%20collection/eCognition%20Suite%20Dev%20RB.htm

Trimble 2014b. eCognition Developer 9.0 User Guide. [Online Help] Online available: https://docs.ecognition.com/v9.5.0/Page%20collection/eCognition%20Suite%20Dev%20UG.htm

Troeh FR 1965. Landform equations fitted to contour maps. *American Journal of Science* 263: 616–627.

Van Asselen S & Seijmonsbergen AC 2006. Expert-driven semi-automated geomorphological mapping for a mountainous area using a laser DTM. *Geomorphology* 78, 3–4: 309–320.

Van Niekerk A & Schloms BHA 2002. A comparison of automatically mapped land components with large-scale soil maps. Proceedings of the Regional Conference of the International Geographical Union. Durban.

Van Niekerk A 2010. A comparison of land unit delineation techniques for land evaluation in the Western Cape, South Africa. *Land Use Policy* 27, 3: 937–945.

Van Niekerk A 2015a. Stellenbosch University Digital Elevation Model (SUDEM) 2015 Edition. March.

Van Niekerk A 2015b. Stellenbosch University Digital Elevation Model (SUDEM) Product Description, September.

Van Niekerk A, Boonzaaier I, Spocter M, Ferreira S, Loots L & Donaldson R 2016. Land Use Policy Development of a multi-criteria spatial planning support system for growth potential modelling in the Western Cape, South Africa. *Land UsePpolicy* 50: 179–193.

Van Riet Lowe C 1952. The Vaal River Chronology: An Up-to-Date Summary. *The South African Archaeological Bulletin* 7, 28: 135–149.

Vandenberghe J 2015. River terraces as a response to climatic forcing: Formation processes, sedimentary characteristics and sites for human occupation. *Quaternary International* 370: 3–11.

Vasat R, Kodesova R, Boruvka L, Jaksik O, Klement A & Brodsky L 2017. Combining reflectance spectroscopy and the digital elevation model for soil oxidizable carbon estimation. *Geoderma* 303: 133–142.

Venter F & Bristow J 1986. An account of the geomorphology and drainage of the Kruger National Park. *Keodoe* 29: 117–124.

Viscarra Rossel RA & Behrens T 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* 158: 46–54.

Viscarra Rossel RA, McGlynn RN & McBratney AB 2006. Determining the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy. *Geoderma* 137: 70–82.

Vohland M, Besold J, Hill J & Fründ HC 2011. Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy. *Geoderma* 166: 198–205.

Voss J & Post TA 1988. *On solving Ill Structured Problems*. In Chi M, Glaser R & Farr M (eds) The nature of expertise, 261–285.

Wadoux A 2015. Mid-Infrared spectroscopy for soil and terrain analysis. Master's thesis. Eberhard Karls University.

Waters RS 1958. Morphological mapping. *Geography* 43: 10–17.

Wellington JH 1929. The Vaal - Limpopo watershed. *South African Geographical Journal* 12: 36–45.

Wellington JH 1941. Stages in the process of river-super- imposition in the southern Transvaal. *South Africa Journal of Science* 37: 78–96.

Wellington JH 1945. Notes on the drainage of the western Free State sandveld. *South African Geographical Journal* 27: 73–77.

Wellington JH 1955. *Southern Africa: a geographical study*. Cambridge: Cambridge University Press.

Wilson JP & Gallant JC 2000. Digital terrain analysis. *Terrain analysis: Principles and applications* 6, 12: 1–27.

Wold S & Sjostrom M 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58: 109–130.

Woodcock CE & Strahler AH 1987. The factor of scale in remote sensing. *Remote Sensing of Environment* 21, 3: 311–332.

Wysocki DA, Schoeneberger PJ, Hirmas DR & LaGarry HE 2000. *Geomorphology of Soil Landscapes*. In Sumner ME (ed) Handbook of soil science, 315–321. CRC Press.

APPENDICES

APPENDIX: SUPPLEMENTARY MATERIAL FOR CHAPTER 5

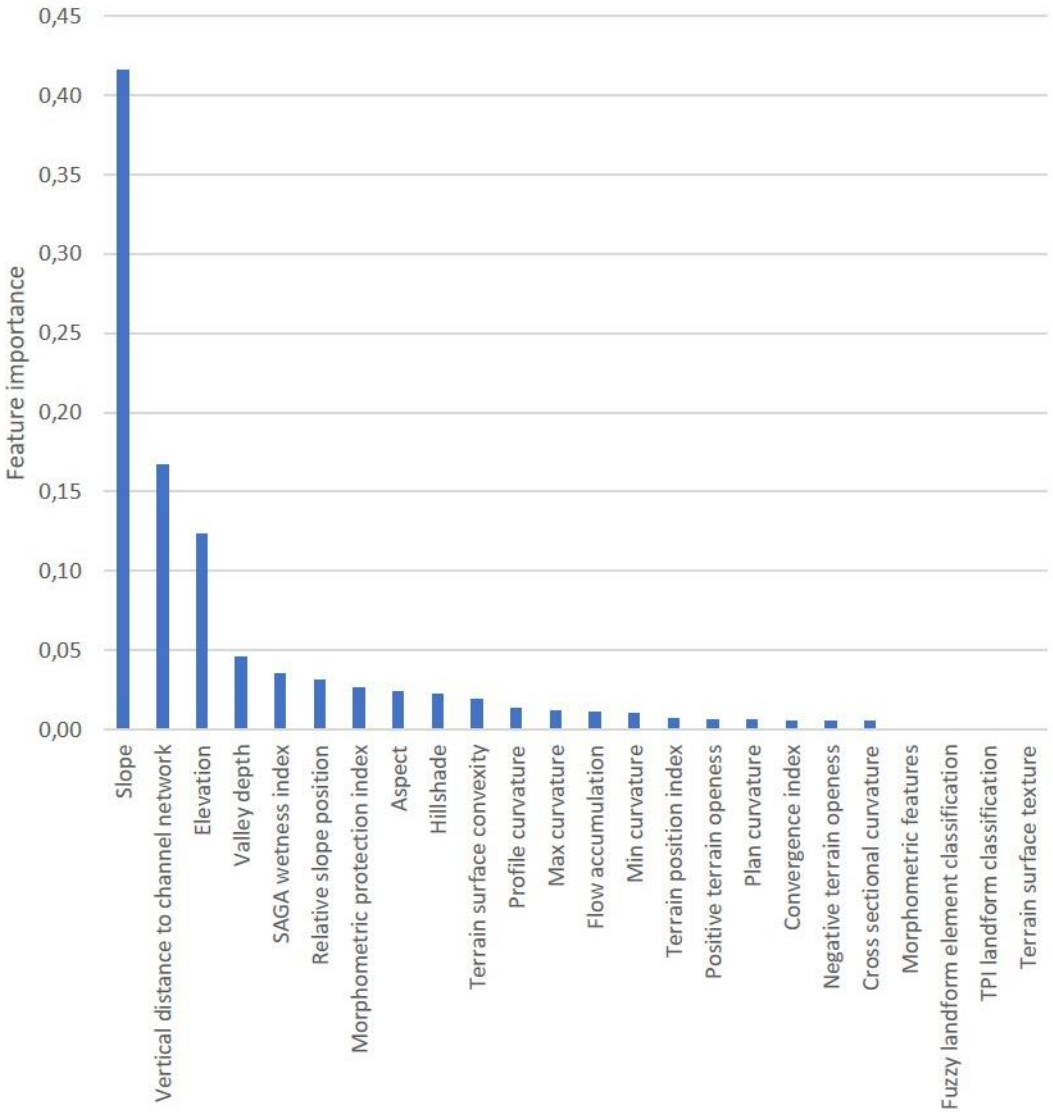


Figure 1 Feature importances of PP-2-D

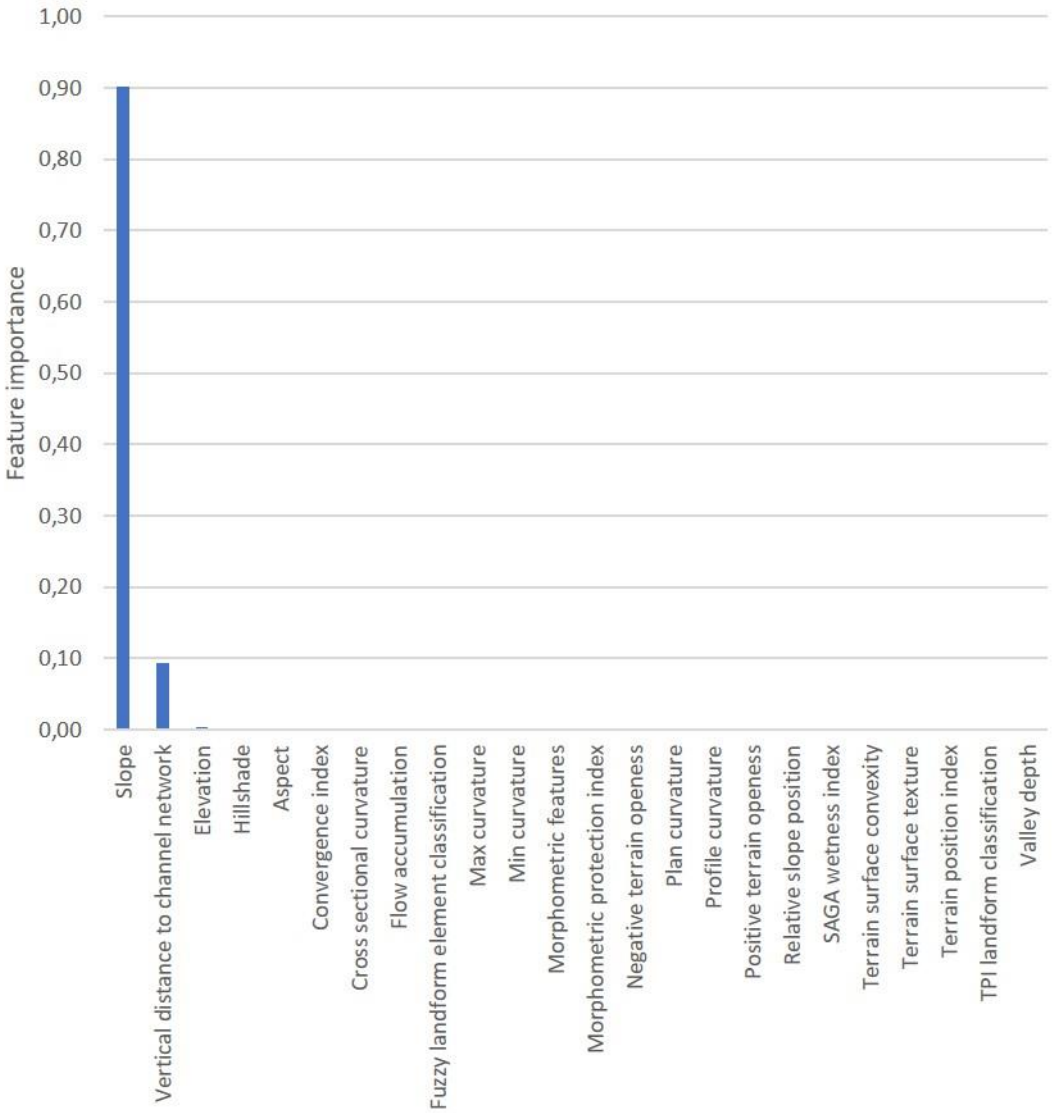


Figure 2 Feature importances of PP-2-P

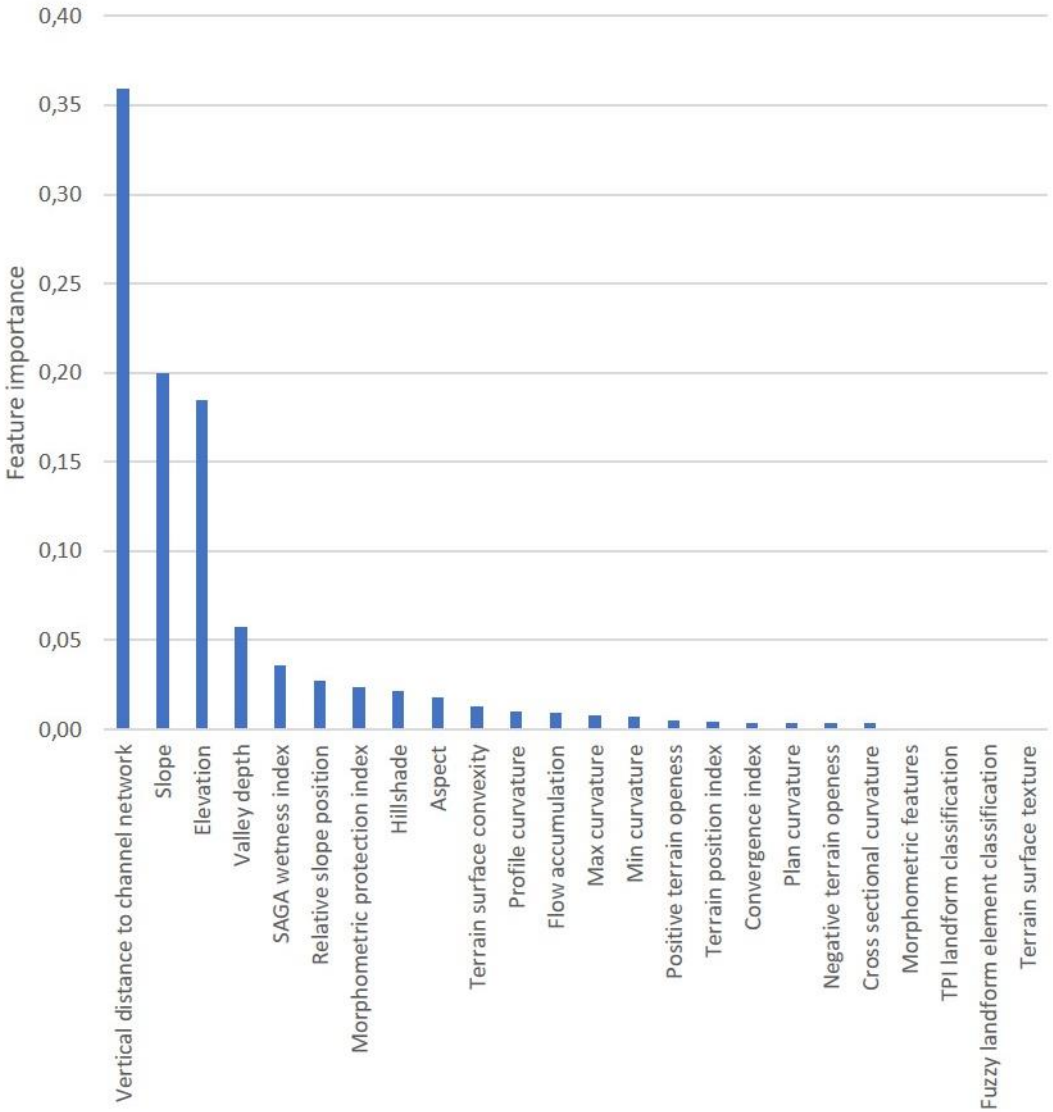


Figure 3 Feature importances of PP-11-D

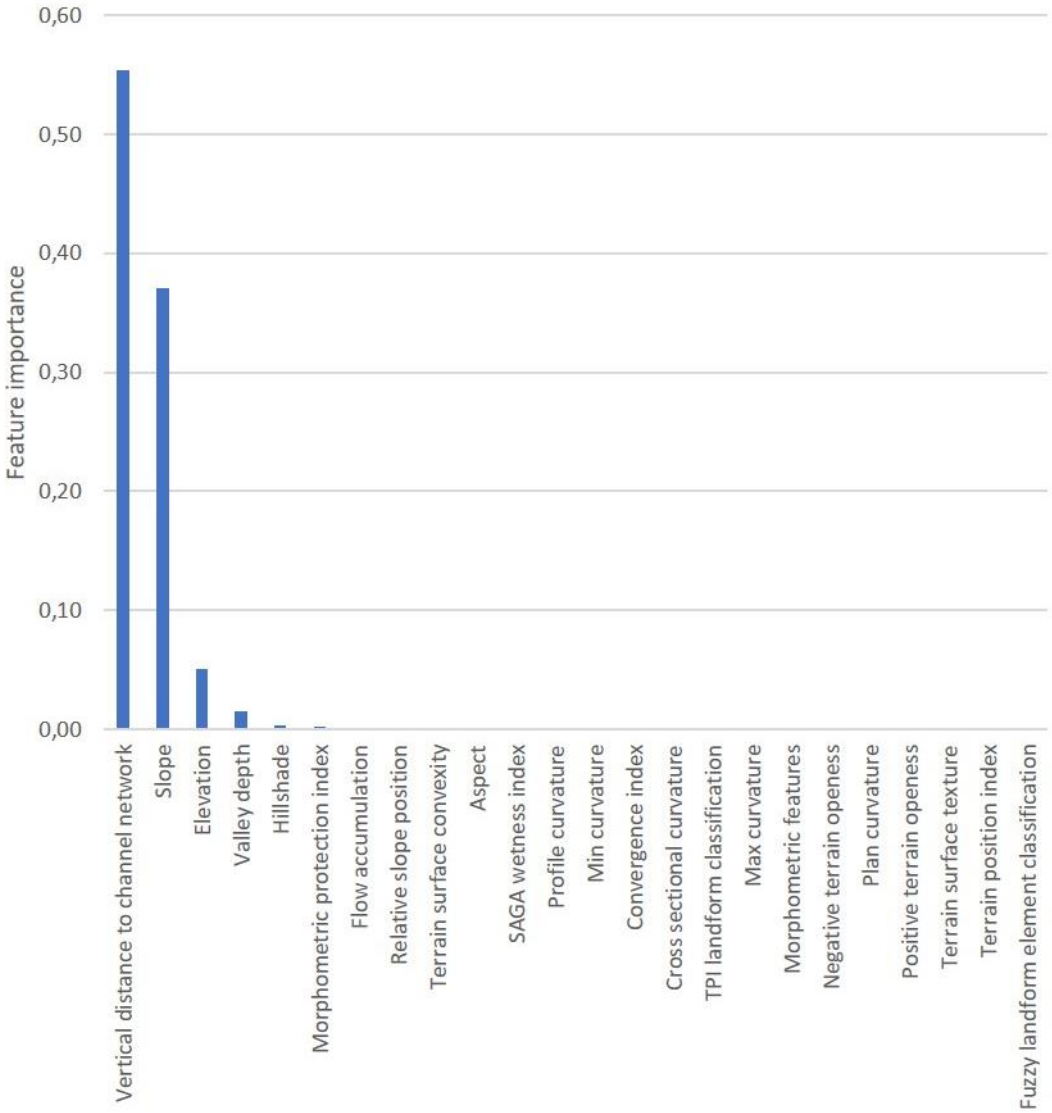


Figure 4 Feature importances of PP-11-P

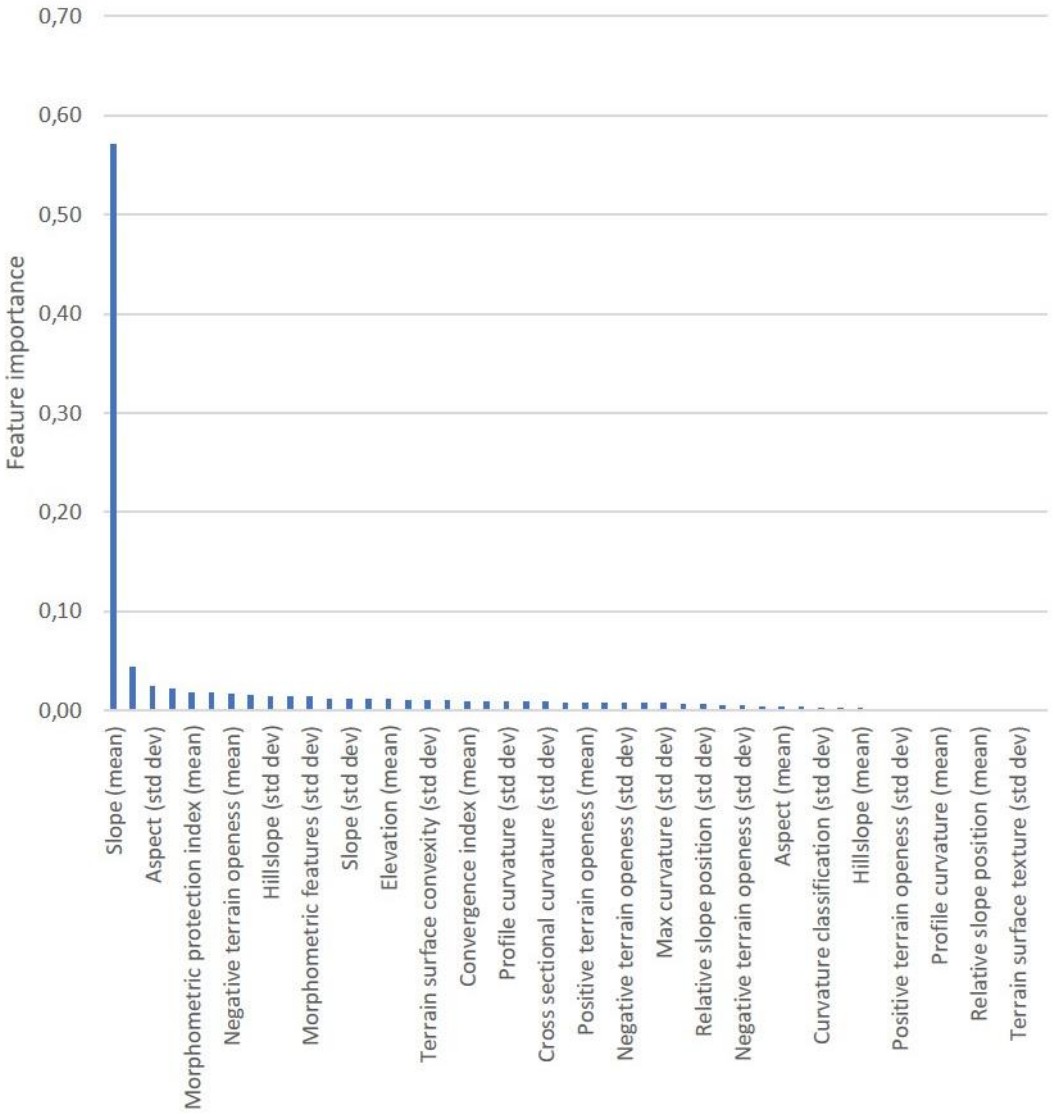


Figure 5 Feature importances of OB-2-D

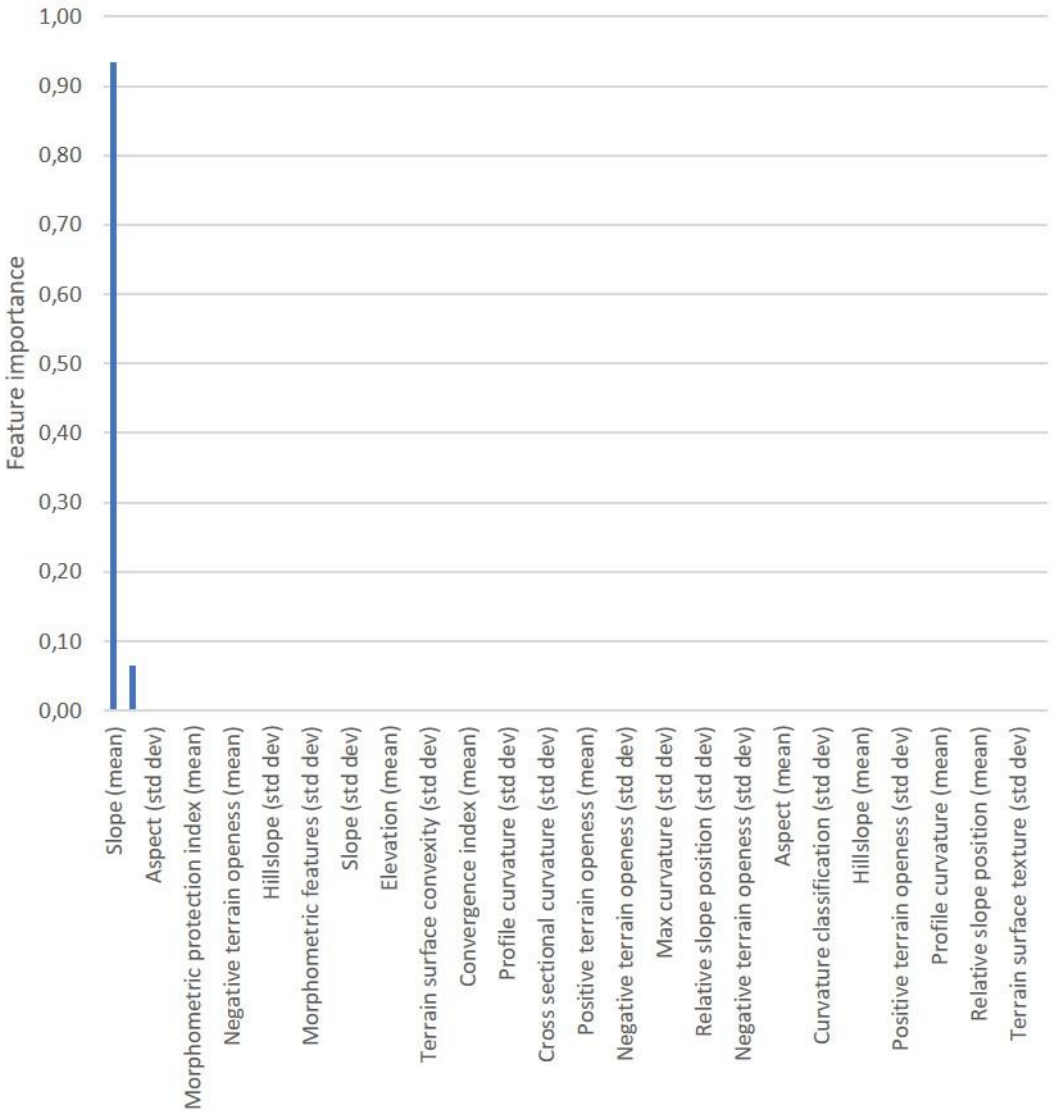


Figure 6 Feature importances of OB-2-P

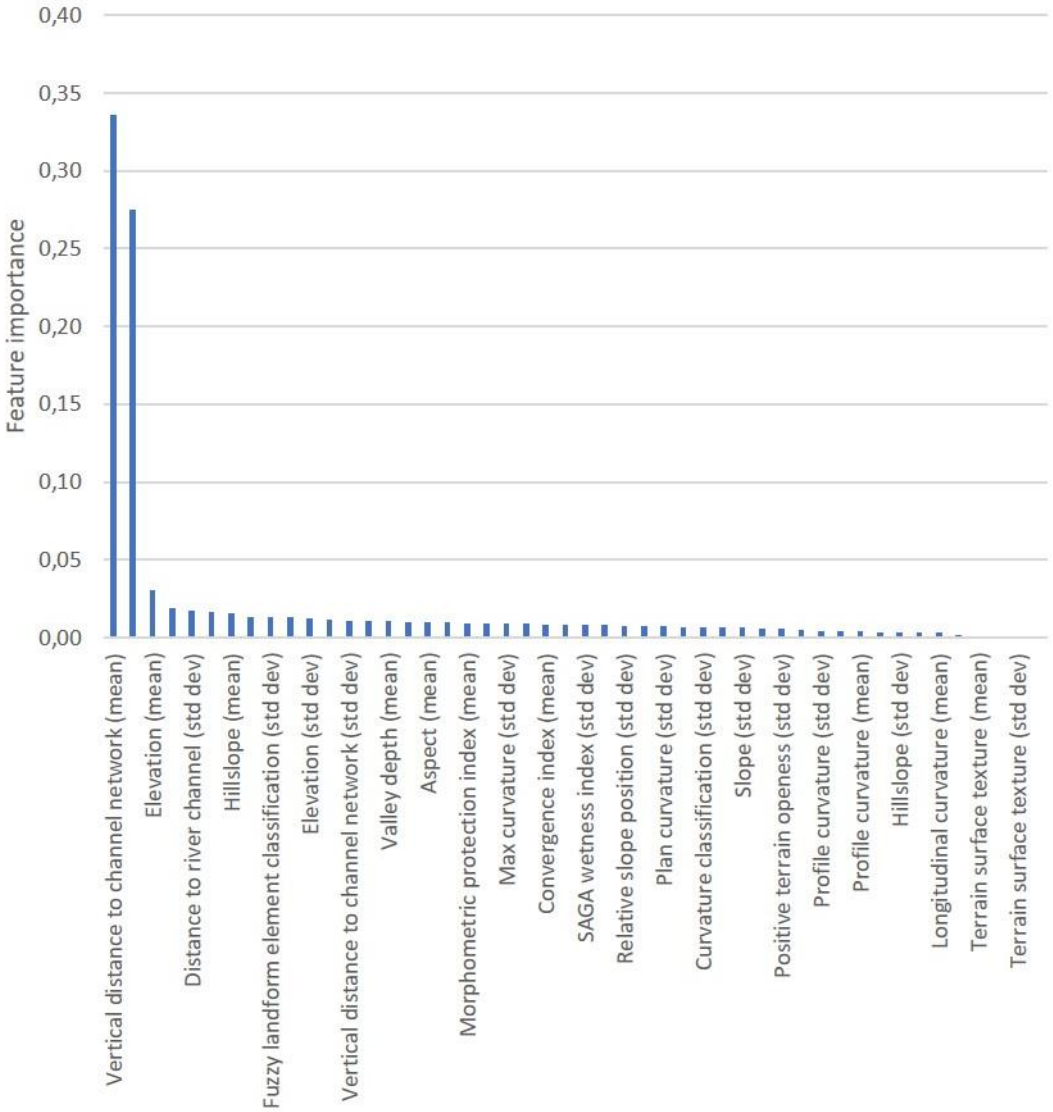


Figure 7 Feature importances of OB-11-D

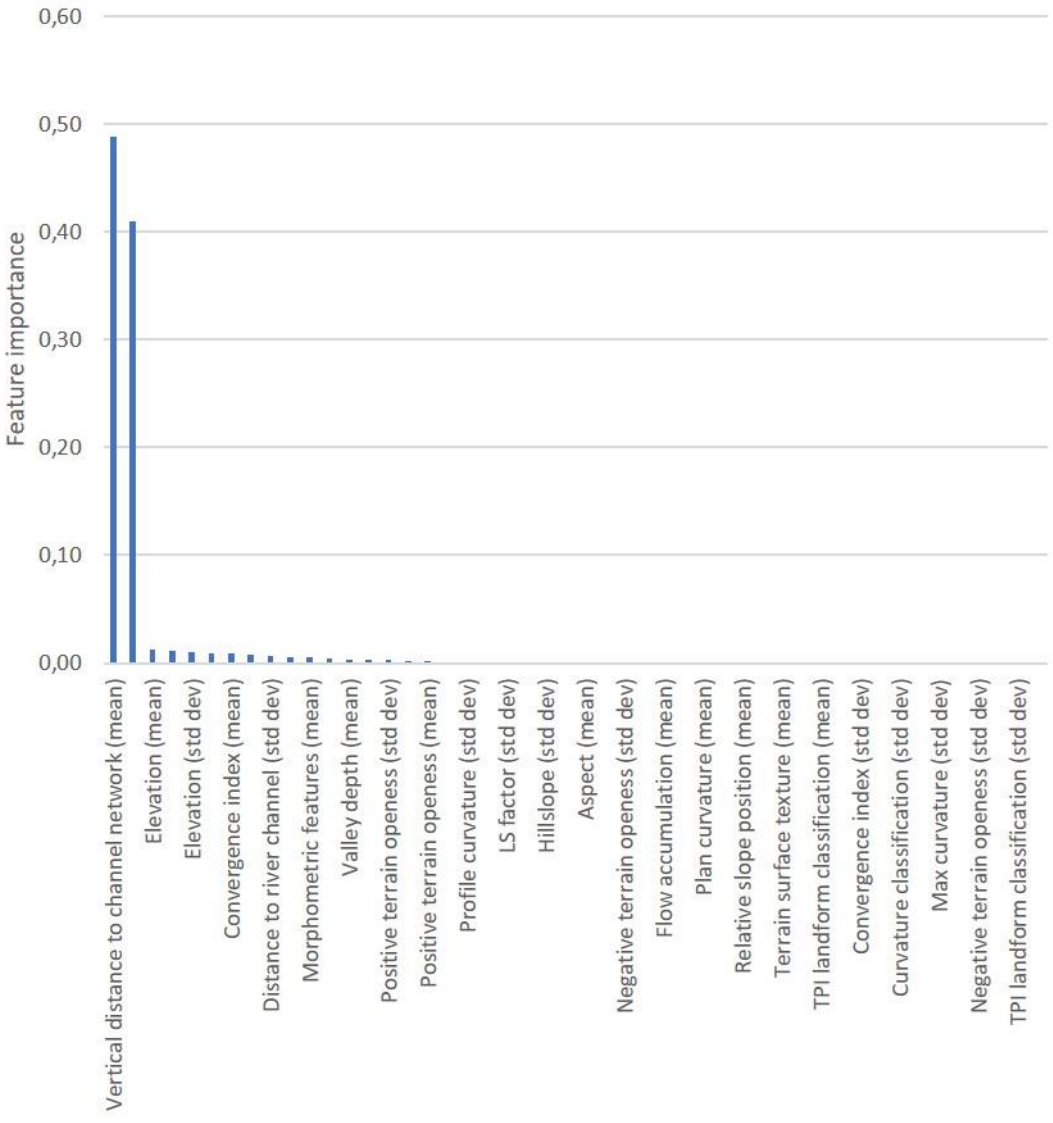


Figure 8 Feature importances of OB-11-P

https://drive.google.com/file/d/11uYE8Vc3avHEDKJHHMwFx0h3tBUq44nV/view?usp=sharing

Hyperlink 1 DT model of PP-11-P

https://drive.google.com/file/d/1-RlgmkXxP-WN2cV2ff9NBvMJl2iynYUy/view?usp=sharing

Hyperlink 2 DT model of OB-11-P