Remote sensing of tree/grass fractional cover using phenological signal decomposition of MODIS time series data

Thesis submitted for the degree of

Doctor of Philosophy at the

University of Leicester

by

Sa'ad Ibrahim MSc

Department of Geography

University of Leicester

2018

Abstract

Remote sensing of tree/grass fractional cover using phenological signal decomposition of MODIS time series data

Reliable assessments of tree/grass fractional cover in savanna using remote sensing are challenging due to the heterogeneous mixture of the two plant functional types (PFTs) and soil backgrounds. This thesis reduces this knowledge gap in the remote sensing of tree/grass fractional cover. Tree/grass dynamics in heterogeneous savanna ecosystems are assessed using time-series decomposition of MODIS data acquired from 2002 to 2015. The decomposition method follows a harmonic analysis and tests the harmonic terms for significance. Several scales of spatial and temporal variability are considered for these PFTs (for each field plot against 14 years dataset as well as for the whole study area). In most harmonic cycles, the tree greening-up period started earlier than grasses. While changes in tree cover are more gradual, grasses have high variability over time. The phase ($R^2 = 0.60$, slope = 1, RMSE = 12.52%), cycles ($R^2 = 0.44$, slope = 1.2, RMSE = 17.64%) and amplitude ($R^2 = 0.36$, slope = 0.83, RMSE = 16.28%) of the strongest harmonic terms show good estimate of tree cover. The estimates of tree cover from the simple linear regression of field data and dry season $NDVI_{pixel}$/$SAVI_{pixel}$ images had good performance. The tree cover estimated using soil determining methods had an improved slope for NDVI and SAVI but yield slightly a high RMSE. A comparison of tree cover using Pearson's correlation indicated strong agreement with LiDAR/SAR and Bucini woody cover maps. The errors, uncertainties and the challenges in discriminating and estimating trees and grasses using signal decomposition methods are discussed. Tree cover maps will be helpful for vegetation monitoring, climate change impact assessment and vegetation model validation. Finally, the techniques employed for the assessment of tree-grass mixtures in this study would be useful for earth observation especially where end-members of the woody-herbaceous continuum are being considered.

Dedication

To my late father, Alhaji Ibrahim Barga. May Allah SWT shower His mercies upon him and continue to grant his gentle soul eternal rest in Jannatul Firdausi.

Acknowledgements

I am thankful to Almighty Allah, most Gracious, who in His infinite mercy has guided me to complete this study. May peace and blessings of Allah be upon His Prophet Muhammad (peace be upon him). I am highly indebted to my parents, brothers and sisters for their continuous support and encouragement that made every opportunity available to me throughout my life.

I would like to express my heartfelt thanks to my supervisors, Prof Heiko Balzter and Prof Kevin Tansey for their encouragements and support throughout the planning and completion of this study. I am very grateful to these people for introducing me into the field of remote sensing right from my MSc. Again, I am very grateful to Kevin Tansey who appointed me as a Project Assistant (part-time job) under the Copernicus global land services programme (Scientific Quality Evaluation of Burnt Area Products) while I was completing my corrections (to help complete the thesis) and at the very time that my official funding terminated. I remain ever grateful. I remained highly indebted to all academic and non-academic staff of the department of Geography (University of Leicester, UK) in general, and to the people of Centre for Landscape and Climate Research (CLCR) in particular. My special regards to Dr Virginia Nicolás Perea (Research Development Manager, CLCR), Dr Azad Rasul, Dr Sarah Cook, Dr Peshawa Najmaddin, Dr Chloe Barnes, Alex Cumming, Valentin Louis, Ana Maria Pacheco-Pascagaza, Maryam Pourshamsi, Dr James Wheeler and Dr Pedro Rodriguez-Veiga. I am very grateful to National Centre for Earth Observation (University of Leicester) for funding my fieldwork (through the Natural Environment Research Council's support for the National Centre for Earth Observation).

I would also like to express my profound gratitude to Tertiary Education Trust Fund (TETFUND) for sponsoring my MSc and PhD study. My special thanks to M. Dauda Mohammed Lawal, the Deputy Provost, Adamu Augie College of Education, Argungu, for his moral, academic and administrative support. Many thanks for all the motivation in seeing that I had secured the nomination for this scholarship. I will forever be grateful to him.

I would like to thank the management of the Kruger National Park (KNP) for granting me the permission to conduct fieldwork in their park. I 'm most grateful to people in the Scientific Services Department of the KNP such as Izak Smit for providing necessary information regarding my initial proposal and field work, Chenay Simms for her prompt reply to emails and supply of GIS data for KNP and Velly Ndloru (Game guard) who provided me with the necessary security while conducting the field work. Special thanks to Robert J Scholes (University of the Witwatersrand) for helping me in the field. Thanks to Renaud Mathieu (Natural resource and environment, CSIR) for sharing the LiDAR/SAR tree fractional cover product which I have used for validation and comparison with my products. To Narumasa Tsutsumida (Graduate School of Global Environmental Studies, Kyoto University, Kyoto, Japan) for his help in R programming.

I would also like to express my appreciation to Nigerian friends in the UK: Dr Idris Muhammad Jega, Malam Isah Hamisu, Malam Usman Lawal, Dr Bashir Adamu, Dr Yahaya Ibrahim Zayyana, Dr Aminu Umar Argungu, Dr Onojeghuo Ajoke and Onyia Nkeiruka for their very cordial relationship.

Finally, my special appreciation goes to my family: my beloved children Abdulsamad, Muhammad Faisal and Aisha and to my wife, Saratu Muhammad for her much-valued contribution to this programme.

Table of Contents

List of Figures

List of tables

List of acronyms and abbreviations

AGB - Above ground biomass

AVHRR - Advanced Very High-Resolution Radiometer

ANOVA - Analysis of variance

ANPP - Annual net primary productivity

CAO - Carnegie Airborne Observatory

CHM - Canopy Height Model

CSIR – Council for Scientific and Industrial Research

DSM - Digital Surface Model

DFT - Discrete Fourier Transform

DGVMs - Dynamic Global Vegetation Models

ERTS - Earth Resources Technology Satellite

EM - Electromagnetic spectrum.

EMD - Empirical Mode Decomposition

ETM+ - Enhanced Thematic Mapper Plus

EVI - Enhanced Difference Vegetation Index

FVC - Fractional Vegetation Cover

FAPAR - Fraction of Absorbed Photosynthetically Active Radiation

FIA - Forest Inventory and Analysis

GIS - Geographical Information Science

GDP - Gross Domestic Product

HWSD - Harmonized World Soil Database

ICESat - Ice Cloud and Land Elevation Satellite

IGBP - International Geosphere-Biosphere Program

IMFs - Intrinsic Mode Functions

ISRIC - International Soil Reference and Information Centre

JAROS - Japan Resources Observation System Organization

KNP - Kruger National Park

LAI - Leaf Area Index

LiDAR - Light Detection and Ranging

LPJ - Lund–Potsdam–Jena

MODIS - Moderate Resolution Imagery Spectroradiometer

NASA - National Aeronautics and Space Administration

NDVI - Normalised Difference Vegetation Index

PCNM - Principal Coordinates Neighbor Matrices

PALSAR - Phased Array Type L-band Synthetic Aperture Radar

PFTs - Plant functional types

RMSE - Root mean square error

SAR - Synthetic Aperture Systems

SAVI - Soil Adjusted Vegetation Index

SWReGAP - Southwest Regional GAP

SRTM - Shuttle Radar Topography Mission

SMA - Spectral Mixture Analysis

UNEP - United Nations Environment Programme

VCF - Vegetation Continuous Fields

WT - Wavelets transform

List of publications

Peer reviewed articles

Sa'ad Ibrahim, Heiko Balzter, Kevin Tansey, Narumasa Tsutsumida, Renaud Mathieu, (2018). Estimating fractional cover of plant functional types in African savanna from harmonic analysis of MODIS time-series, *International journal of remote sensing,* 39:9, 2718-2745

Sa'ad Ibrahim, Heiko Balzter, Kevin Tansey, Renaud Mathieu, Narumasa Tsutsumida., (under review). Estimating woody fractional cover in African savanna using MODIS time series data: *remote sensing*

Peer reviewed conference proceedings

Sa'ad Ibrahim, Heiko Balzter, Kevin Tansey., 2015. Mapping Savanna vegetation phenology using NDVI time series derived from MODIS. *13th Savanna Science Network Meeting*., organised by the South African National Park Scientific Services (2015), South Africa.

Sa'ad Ibrahim, Heiko Balzter, Kevin Tansey., 2016. A comparison of models for signal decomposition of satellite time-series data to estimate tree/grass fractional cover in African savanna. *RSPSoc Annual Conference 2016 (Earth Observation for Knowledge Discovery*), 5-8th September: University of Nottingham and British Geological Survey

Chapter 1

Introduction and thesis overview

1.1 Research context

Large-scale information on plant functional types (PFTs) composition of a landscape can reveal ecological processes leading to vegetation fluctuation and succession which may allow assessment of ecosystem vulnerability (Lu et al., 2003, Smith et al., 2014a, Smith et al., 2014c). PFTs uses structural, physiological and phenological features to group plant species in terms of their response to environmental conditions and determine their impact on the ecosystem function (Ustin and Gamon, 2010). The monitoring of long-term changes in the tree/grass cover in savannas is required for an assessment of the ecosystems processes, biosphere-atmosphere transfer (e.g. hydrology) and carbon budget (e.g. carbon sequestration potential) to understand a changing climate (Yang et al., 2012, Williams et al., 2007, Jin et al., 2013, Hoffmann and Jackson, 2000). Vegetation fraction datasets can be used for climate models and ecological models (Los et al., 2012). For example, information on tree/grass fractional cover is commonly used as an input to many ecological models in the context of ecosystem change for the assessment of fire, deforestation, degradation, urban extension and water management, etc., (Montandon and Small, 2008, Guan et al., 2012, Mathieu et al., 2013, Gessner et al., 2013, Villegas et al., 2015, Giglio et al., 2006, Li and Strahler, 1985, DeFries et al., 2007, Los et al., 2012).

Tree cover datasets exist at both coarse and moderate spatial resolutions (Los et al., 2012, Hansen et al., 2003b). The existing vegetation continuous fields (VCF) product from the Moderate Resolution Imaging Spectroradiometer (MODIS) do not adequately capture woody species (DeFries et al., 2000, Hansen et al., 2003a). Despite the fact that information on vegetation fractional cover (FVC) provides immense benefit to ecological modelling and promotes understanding of ecosystem function in savanna (Hill and Hanan, 2010b, Verger et al., 2009a), spatially explicit information on tree fractional cover, for example, is rarely available in savannas due to mixtures of PFTs (Gessner et al., 2008, Gessner et al., 2013, Cho et al., 2010). Research in this direction is useful since changes in the woody cover may have profound effects and unpredictable consequences for ecosystem function (Ustin and Gamon, 2010, Jiménez-Muñoz et al., 2009). Some of the key requirements for effective measurement accuracy include a proper understanding of vegetation structure and phenological characteristics (Ustin and Gamon, 2010).

Although several techniques are still being adopted for estimating tree/grass fractions, and their interannual variability, the complexity of tree/grass coexistence limits previous approaches to derive spatially explicit information about heterogeneous landscapes (House et al., 2003, Gill et al., 2009, Brandt et al., 2016, Bonan et al., 2003). This is because trees and grasses in savannas are not distributed uniformly across the landscape but instead show different degrees of 'clumping' or 'patchiness' (Scholes and Archer, 1997). Assessing the quantity and quality of PFTs could be better achieved by mapping tree/grass fractional cover as a better representation of these landscapes (Gessner et al., 2013, Mairota et al., 2015).

Many previous efforts have been made to quantify the tree/grass canopies using field-based methods (Beale, 1973, Walker et al., 1972, House et al., 2003, Scholes, 2003). Vegetation indices such as Leaf Area Index (LAI), biomass, basal area, fractional cover, density, etc., are commonly measured or estimated using different methodologies based on ecological theories and sampling protocols. Although some models have empirical support, their validity as general mechanisms of tree/grass coexistence have been questioned purely due to the subsets of datasets and the consideration to limiting factors of establishing and assessing PFTs (Sankaran et al., 2004, House et al., 2003, Scholes, 2003).

There are many challenges which make the assessment of fractional vegetation cover using a field-based method limited. Some of these include the difficulty in sampling the number of species to be investigated, the impact of the sampling design itself and the challenges in defining the population under concern (Rocchini et al., 2015). Challenges arise due to site-specific differences in vegetation distribution, probably a result of differences in species composition, soil types, changing climate and anthropogenic disturbances. Besides landscape factors, understanding characteristic differences in the biology of trees and grasses is very useful to ecosystem change modelling (Higgins et al., 2011). This is typically challenging in a savanna ecosystem because of the heterogeneity in the distribution of trees and grasses and their differences in leafing periods and water and nutrient requirements. While this is problematic, there are pressing needs for reliable tree/grass information for model validation (Boke-Olén et al., 2016). Subsequently, a robust technique which could capture the spatial heterogeneity of PFTs is needed (Boke-Olén et al., 2016).

Remote sensing is one of the most cost-effective approaches with which to identify and predict changes in PFTs (Rocchini et al., 2015, Mairota et al., 2015). Advances in remote sensing technology have increased our understanding of tree/grass structure, physiology and phenology, which yield insight into the concept of PFTs (Ustin and Gamon, 2010, Schmidtlein and Fassnacht, 2017). Remote sensing measurements in different spatial, radiometric and temporal resolutions offer a key source of updated, consistent and spatially explicit data on biophysical indices including Leaf Area Index, fraction of photosynthetic active radiation (fPAR), phytomass and canopy height, for assessing PFTs (Avitabile et al., 2012). Remote sensing data are acquired from airborne and space-borne sensors, from multispectral sensors to hyperspectral, at different wavelengths from visible to microwave, and at a range of spatial and temporal scales (Xie et al., 2008, Los et al., 1994). This provides a greater possibility for large area measurement, time series analysis to capture the spatial and temporal variability of the various PFTs and allow systematic observations at both local, regional and global scales (Ustin and Gamon, 2010, Zimmermann et al., 2007, Avitabile et al., 2012). Many efforts have been made previously using microwave (e.g. synthetic aperture radar such as Sentinel-1) and optical sensing (e.g. airborne LiDAR; passive multispectral such as Landsat) to characterise PFTs (Tucker et al., 1985, Ollinger, 2011, Balzter et al., 2007b, Khalefa et al., 2013, Tansey et al., 2004). Remote sensing measurements are therefore more advantageous than a field-based method which usually extrapolates tree/grass biophysical variables at limited subsamples (House et al., 2003, Roy et al., 2014).

Recent advances in remote sensing technology present opportunities to develop proper understanding and characterisation of PFTs. The technology is often linked to ecological theory considering structural, physiological and phenological traits of PFTs based on resource constraints. This idea is leading to an emerging hypothesis in remote sensing, referred to as *optical types* (Ustin and Gamon, 2010). The optical types relate to the assessment of PFTs based on optical principles, meaning PFTs distributed across spatial scales with variability in resource availability (e.g. moisture, nutrient, light, temperature) are detectable by remote sensing through radiative transfer and spectroscopy (Ustin and Gamon, 2010). This is because vegetation structure, biogeochemistry, physiology, and phenology have a strong influence on vegetation optical properties (Ustin and Gamon, 2010, Ollinger, 2011, Alton et al., 2005). Consequently, the reflectance properties of PFTs are not only influenced by the differences in fractional cover but also the nature and chemistry of the plants (Asner et al., 2011, Los et al., 2005).

Considerable efforts have been made using vegetation indices to characterise species based on their leaf area, biomass, fractional cover and physiological functioning (Myneni and Hall, 1995, Lotsch et al., 2003). From the space-borne optical sensors (e.g. NOAA AVHRR, NASA-MODIS), large measurements of satellite-derived parameters, such as biophysical and biochemical variables over the land surface, are being provided at very fine temporal frequencies. Satellite time series products relating to land surface phenology include indices of 'greenness', such as Normalised Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI) and Enhanced Difference Vegetation Index (EVI). These are currently being utilised to extract phenology metrics for understanding climate change and biogeochemical models (Ma et al., 2013). Land surface phenology is defined as the spatiotemporal development of the vegetated land surface (De Beurs and Henebry, 2004).

While phenology signals such as the NDVI are useful for vegetation characterisation, they do not separate trees and grasses, and they contain a bare soil contribution (Fuller et al., 1997, Boke-Olén et al., 2016, Ding et al., 2016: Montandon and Small, 2008). This means an increased uncertainty in the use of the vegetation index for image analysis and interpretation as trees and grasses can have different phenological cycles subject to differences in species types, locations and time (Boke-Olén et al., 2016). Another challenge is that remote sensing data itself can be affected by several factors such as signal contamination, sensor viewing geometry, and habitat type (Tucker et al., 1985, Yengoh et al., 2015, Cleland et al., 2007, Los et al., 2005). These pose significant challenges for PFTs characterisation using remote sensing. Hence, there is an increasing use of various approaches aiming to reduce uncertainty for PFTs characterisation using remote sensing. For example, signal decomposition techniques have been utilised to transform original satellite time series data to understand PFTs and ecosystem dynamics (Lu et al., 2003, Cleveland et al., 1990, Atkinson et al., 2012, Jakubauskas et al., 2001, Lhermitte et al., 2011, Kostadinov et al., 2017).

Fourier or harmonic analyses of a time series is the decomposition of time series data into the sum of the sinusoidal components, and the coefficients are the discrete transform of the series (Bloomfield, 2004). The Discrete Fourier Transform (DFT) is useful in quantifying phenological metrics from remote sensing data to characterise PFTs (Kostadinov et al., 2017). DFT presents an analytical way for transforming original data into sinusoids in order to partition observations into specific components (Kong et al., 2015, Thomson, 1982). The decomposed components derived from the phenology of vegetation provide important information on the PFT's growing cycles (e.g. seasonal, annual and biannual signals and the timing of greening). Seasonal cycles are a key property of ecosystems (Kostadinov et al., 2017). The decomposed components can reflect the interannual variability of vegetation growing cycles driven by weather, fires and human activities among others. Signal decomposition can be applied to PFTs characterisation in the savanna ecosystem where higher inter-annual climatic variability and disturbances are more prominent (Ma et al., 2013). While signal decomposition is useful for identification, mapping, and modelling PFTs, a proper understanding of the satellite time series characteristics and ecosystems are fundamental (Lhermitte et al., 2011, de Beurs and Henebry, 2010).

It is obvious that remote sensing is one of the viable means for assessing and estimating vegetation phenology. From the literature cited above, it appears that remote sensing methods for tree/grass characterisation are many and varied; each has its strengths and weaknesses. In general, it should be noted that from the review of field-based methods and remote sensing approaches presented above, much has been done for the identification, modelling and mapping of PFTs in savannas as well as in many ecosystems. Yet, there is a wide range of opportunities for improvement. The techniques that would efficiently answer some of the pressing questions, especially regarding tree/grass characterisation using passive data, are not fully in place yet.

This thesis aims to improve on existing harmonic analysis methods (Shatkay, 1995, Moody and Johnson, 2001) using satellite time series data on tree/grass phenology by estimating statistical significance using the Hartley test and correcting for multiple testing with the Bonferroni method (Hartley, 1949) and multitaper approach (Barbour and Parker, 2014), adding knowledge to the assessment of tree/grass dynamics, as well as estimation of their fractional cover in African savanna. The study estimated the seasonal cycle as amplitude and phase derived from the annual frequency over the entire time series data (14 years of MODIS NDVI). The use of dry season images from MODIS data (NDVI and SAVI) to estimate tree cover fractions is also investigated. Previous methods of determining soil from the NDVI/SAVI were employed for tree cover estimation due to uncertainty inherent in vegetation indices resulting from soil backgrounds (Ding et al., 2016: Montandon and Small, 2008). The study is presented as a novel remote sensing approach that can be in principle applied to the assessment of tree/grass phenology in savanna sites worldwide.

1.2 Thesis structure

This thesis comprises eight chapters. The following section provides a brief description of each chapter.

- Chapter one introduces the problem and provides background information about the need for tree/grass estimates, discusses the limitations of field methods and advantages being offered by the remote sensing approaches. It also discussed uncertainty surrounding remote sensing data (e.g. vegetation index) especially for tree cover estimates in savannas. The importance of signal decomposition in the context of retrieving phenological cycles of PFTs (e.g. Fourier analysis) and soil determining methods (e.g. soil contribution in the NDVI) have been discussed briefly.
- Chapter two reviews the literature on the concept of savanna and approaches to estimates of tree and grass fractions in savannas. The strength and weaknesses of field and remote sensing approaches were reviewed. A general review was made on the optical passive sensors (e.g. radiative transfer theory, MODIS VCF), LiDAR, and radar microwave remote sensing. Specifically, a review of remote sensing methods for signal decomposition of satellite time series data was considered in this chapter. Research gap, questions, and objectives were stated. This chapter formed a strong foundation upon which analyses in the subsequent chapters were based.
- Chapter three presents a description of the study area and general methods. The chapter described how the harmonic analyses were applied for the decomposition of tree/grass phenology using MODIS time series data. Methods for pixel as well as image analyses were explained. However, not all methods employed in this study are explained in this chapter. Some methods are described in each analysis chapter as they are more specific. All datasets used in this study were outlined and discussed.
- Chapter four presents pixel-based analysis for MODIS NDVI data (2002 to 2015) of twenty-eight field plot data collected from a field campaign in 2015. The

analyses estimated the tree/grass amplitude, number of cycles and phase values for each growing cycle that are statistically significant based on the Bonferroni-Hartley test. The interannual variability of trees and grasses were assessed based on the annual NDVI of a 14-year data aggregated as one as well as for each analysis year separately. Tree/grass cover was estimated using phase and cycles as derived from the harmonic analysis.

- Chapter five provides estimates of tree cover using amplitude, $NDVI_{pixel}$, $SAVI_{pixel}$ and a field data collected from the field campaign in 2015. Detail description of the methods such as soil determining and regression analyses were provided in this chapter. Tree cover maps estimated with regression equations derived from different models were presented and explained.
- Chapter six presents a validation of fractional tree cover estimated in chapter 5 using field data on percent tree cover. A comparison of accuracy was made between MODIS NDVI harmonic, $NDVI_{pixel}$, $SAVI_{pixel}$, NDVI/SAVI (soil determining methods) tree cover maps and MODIS VCF using LiDAR-SAR tree cover and Bucini woody cover maps. A wide range of statistical tests were used to assess the level of uncertainty in the estimated tree cover.
- Chapter seven presents an overarching discussion for the thesis and main thesis contribution, conclusion, limitations and future research direction. This section provides a summary of the overall analysis chapters and findings while referring to gaps, significance and novelty of the study.

Chapter 2

Approaches to tree/grass estimates in savannas (field and remote sensing methods)

2.1 Introduction

This chapter begins by defining the concept of savanna, explaining the ecological threat (e.g. fires) and potentials of savanna ecosystems. It also highlights the gradual development of remote sensing technology to modelling of PFTs through satellites and airborne observations. The difficulties and challenges associated with some remote sensing techniques especially in estimating tree/grass phenological characteristics were highlighted. The advantages and gaps associated with signal decomposition methods on harmonic analysis (Fourier analysis) were identified. The research questions and objectives identified in the course of reviewing the literature were stated and formed the basis for the analyses in the subsequent chapters.

2.2 The Savanna ecosystem

A savanna is broadly defined as a grassy biome composed mainly of shrubs, herbs, grasses and scattered trees (Bond and Parr, 2010). Savanna is also termed as mixed tree-grass communities consisting of mainly herbaceous systems and of a discontinuous woody cover (Sankaran et al., 2008). Dansereau (1957) defined savanna as "a mixed physiognomy of grasses and woody plants in any geographical area (Hill et al. 2011). These definitions imply that savanna vegetation cuts across different sets of global ecosystems. It represents one of the world's most important terrestrial ecosystems, comprising up to 20% of the total global land surface depending on the definition. Savanna occupies nearly 50% of the African continent and one-third of the South African land area (Shackleton and Scholes, 2011).

Savanna demarcation has however been a subject of disagreement among biogeographers and ecologists (Shackleton and Scholes, 2011). The extent or coverage of a savanna at a given location or region depends largely on the definition in use. Different classification systems use different parameters (e.g. tree/grass cover and density) for these definitions, and different thresholds for these parameters. This is due to savanna heterogeneity and diversity in species richness, and in the composition of both flora and fauna. This complexity makes it difficult to describe the global savanna biome. Figure 2.1 is a representation of global tropical and subtropical savannas constructed by the World Wildlife Fund (WWF) and based on terrestrial ecoregions of the world according to species richness, endemism, higher taxonomic uniqueness and global rarity of the major habitat type, among other parameters.

Figure 2. 1 global tropical and subtropical savannas by WWF (Hill et al., 2011)

The heterogeneity of the savanna ecosystem is the result of inherently diverse ecology owing to striking variability of environmental indices such as rainfall, temperature, plant composition, topography, soils, fire and herbivores, etc., in this environment.

The global distribution of tree/grass mixtures as revealed by the MODIS Vegetation Continuous Field (VCF) data reflects the global importance attached to savanna ecosystems (Figure 2.2). Savanna alone accounts for nearly 30% of the primary productivity of all terrestrial vegetation in the tropics (Ribeiro et al., 2013). Hence, savanna ecosystems have large carbon sequestration potential (Bombelli et al., 2009). Savanna ecosystems are also a major source of livelihood for many people across the landscape. For example, global tree/grass environments and savannas produce meat and milk to the value of $1.4 trillion annually. In the African savanna, cattle rearing is a major source of employment to a large number of labours which contribute substantially to revenue generation and Gross Domestic Product (GDP) (Thornton, 2010)

Globally, savannas are facing ecological threats which, if properly handled, could result in harnessing its myriad of potential. This is because, the regions are characterised by vulnerable people who are mostly subsistence farmers. For instance, rapid population growth coupled with increased poverty, often leads to a high demand for agricultural land and large-scale deforestation due to biomass burning for charcoal and firewood extraction. In combination, these pressures lead to general land fragmentation and degradation; the effects of which are yet to be fully realised. Savanna accounts for nearly 90% of the world global burnt area. In Africa, for instance, fire has caused carbon emissions of up to 1.03 ± 0.22 Pg C yr^{-1}, of which 90% is emanating from the burning of savanna and woodlands (Valentini et al., 2014). Savanna has a strong, inter-annual climatic variability which consequently affects livestock production, subsistence agriculture, tree/grass density, biomass, carbon and biogeochemical dynamics (Valentini et al., 2014).

Figure 2. 2: Global distribution of tree-grass mixtures based on classification according to MODIS Vegetation Continuous Field (VCF) data (Hansen et al., 2005)

2.3 Field method for tree/grass modelling

For estimating PFTs using the field method, vegetation indices such as Leaf Area Index (LAI), biomass, basal area, fractional cover and density etc., are commonly measured or estimated using different methodologies based on ecological theories and sampling protocols. Such methodologies are usually employed to assess tree/grass characteristics as a basis for model parameterisation. However, most of field method (models) have been largely developed for forest ecosystems rather than savannas. Previous studies show that many studies in African savannas do not use allometric equations for the retrieval of tree/grass variables. The few that applied such equations are often concentrated in the narrowest of geographical regions, with sampled trees and huge inconsistences in methods, which might make the comparison of data very difficult. This gap was identified by Gibbs et al., 2007 and was also well acknowledged in the IPCC 2006 (Kamelarczyk, 2009).

Therefore, for a comprehensive understanding of this process, information at large spatial and temporal scales is needed. While a field experiments are being considered as the most accurate method for estimating tree/grass variables, often served as reference datasets for other modelling approaches, limited and short-term observations constrained their ability to capture spatial heterogeneity and enable comprehensive tree/grass modelling (House et al., 2003). Therefore, combined field studies and other modelling approaches will go a long way in providing an enabling platform with which to assess a tree/grass system in savannas. Specifically, the use of satellite remote sensing together with field studies could provide better opportunities for understanding tree-grass systems at a wide range of spatial scales.

2.4 Remote sensing of tree/grass composition

Satellites and airborne observations rely on the spectral reflectance signatures of vegetation to distinguish various PFTs characteristics (Chuvieco and Huete, 2010). All objects with a temperature above absolute zero (0 K or -273°C) absorb and emit energy from and to the atmosphere. The range of radiant energy is called the electromagnetic

(EM) spectrum. The magnitude of absorption and reflectance varies with objects due to differences of the radiant energy in wavelength or frequency. Signals acquired by the remote sensor from terrestrial objects enable analysts to characterise these objects by their behaviour of emittance over the spectral regions of the EM spectrum. Figure. 2.4 shows the electromagnetic spectrum with the visible spectrum highlighted (Chuvieco and Huete, 2010, Lillesand et al., 2014).

The visible (VIS) region (0.4 – 0.7µm) of the electromagnetic spectrum corresponds to the segment of EM that can be seen by human eye. It consists of three bands, the blue (0.4 – 0.5 µm), green (0.5 – 0.6 µm) and the red (0.6 – 0.7 µm). The near-infrared (NIR) region (0.7 – 1.2 µm) extends beyond the eye's perception. It is one of the most useful regions for its ability to discriminate green vegetation. The mid-infrared (MIR) region (0.12 – 8 µm) is the transition between the NIR and the thermal infrared regions. It is known as the short infrared (SWIR). The thermal infrared (TIR) lies between 8 to 14 µm while the microwave region (>1mm) is the long wavelength which can penetrate cloud and dense forest canopies. Optical sensors operate within the visible, infrared and thermal portions of the EM spectrum, while radar wavelength is situated in the microwave region. LiDAR is an active sensor like radar, its pulses are in the range of visible and short infrared portions of the spectrum. Unlike the optical passive sensors, active sensors illuminate objects and detect the returns or backscatter (Lillesand et al., 2014, Chuvieco and Huete, 2010).

Figure 2. 3: The electromagnetic spectrum with visible spectrum indicated (source: Chuvieco and Huete 2010).

For a long time now, remote sensing studies have used measurements of spectral reflectance within visible and NIR regions to characterise vegetation (Chuvieco and Huete, 2010). This was made possible since spectral reflectance signatures of PFTs behave differently across the regions of electromagnetic spectrum. Many factors such as leaf type, leaf physiology, chlorophyll content, water content, plant stress and senescence influence spectral reflectance signatures of PFTs. The reflectance for both visible and SWIR spectrums is low due to high chlorophyll absorption by plants and high in the NIR due to individual leaves and whole plant canopies which strongly scatter NIR energy (Ollinger, 2011). Figure. 2.4 illustrates how the typical reflectance of green vegetation, dry vegetation and soil appear in different spectral wavelengths.

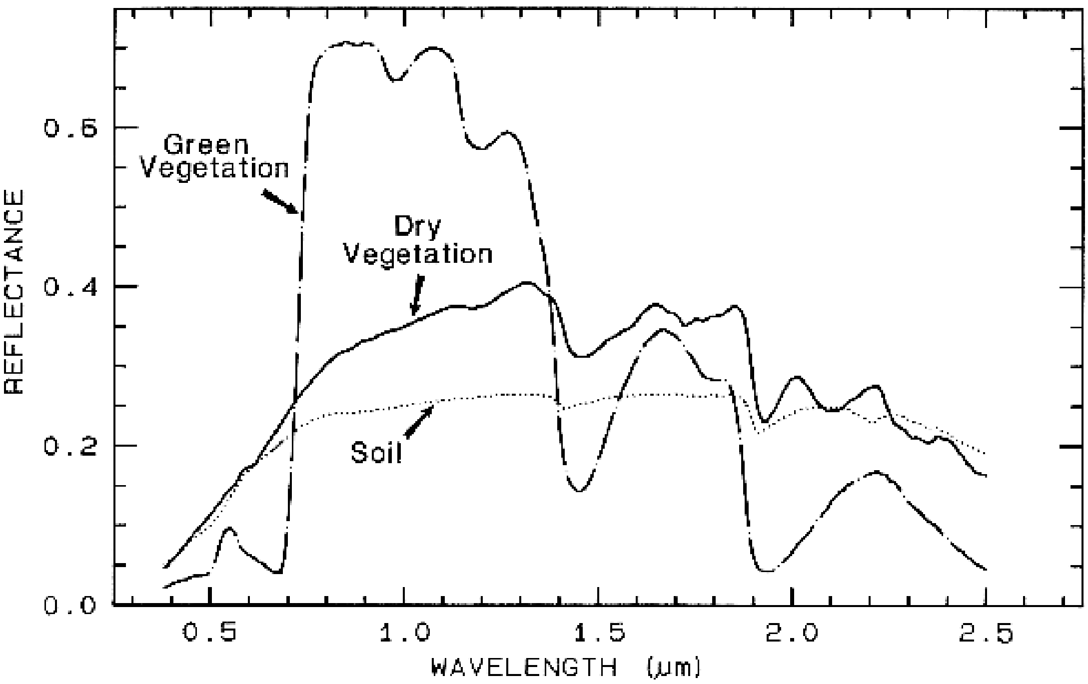


Figure 2. 4: typical spectral reflectance curves for green vegetation, dry vegetation, and soil (Source:(Clark, 1999) Adapted with the kind permission of Wiley and Sons inc)

There are several attempts to describe PFTs from single to combinations of bands using remote sensing. Through different bands combinations, a range of vegetation indices are derived which usually enable more information on plant behaviour to be better assessed than in a single band. In the context of the remote sensing of PFTs, Normalised Difference Vegetation Index (NDVI) is the most widely used (Archibald and Scholes, 2007b, Gill et al., 2009, Lu et al., 2003, Ustin and Gamon, 2010). This index is derived from the combination of NIR and red bands:

$$\text{NDVI} = \frac{NIR - Red}{NIR + Red}$$

However, different sensor platforms ranging from optical to microwave systems have provided wider opportunities for the characterisation of the structure and function of various PFTs under different conditions. Key challenges in the use of satellite data for modelling PFTs include data type probably due to data processing, atmosphere, energy sources or sensor characteristics. Thus, trade-offs exist between datasets in terms of their applicability and usage. To better understand the dynamic of PFTs, various users of this data have since identified the suitability of different sensors and for assessment of different phenomena (Lillesand et al., 2014). However, the full range of opportunities provided by these sensors are still being explored. The next section discusses previous research in remote sensing (the optical passive sensor and active such as radar and the LiDAR system) with specific consideration to the vegetation.

2.4.1 Optical remote sensing

Remote sensing data from optical passive sensors have been very useful for monitoring, analysing, and mapping temporal and spatial distribution of PFTs at both regional and global scales. The availability of remotely sensed time series satellite data covering more than three decades, provides a large data resource for PFTs characterization and measurement. Thus, the long-term history of the optical (passive) data archive has increased interest in the assessment physiological and biophysical characteristics of PFTs (Gitelson, 2004, Labrecque et al., 2006, Yang et al., 2012).

The Landsat satellite, or Earth Resources Technology Satellite (ERTS 1) as it was then called, was first used for the development of the land cover classification system (with particular reference to North America) based on plants physiognomy types which correspond to 16 growth forms defined by Von Humboldt in 1807 (Ustin and Gamon, 2010). This facilitates the rapid understanding of PFTs based on their biochemical composition and morphological structure. Since PFTs are composed of different growth forms, their signals as captured by sensors, would produce different characteristic reflectance patterns (e.g. very low reflectance in the NIR for developed forest compared to grasslands) making it possible to distinguish species types, to quantify fractional cover and estimate canopy structure. Since then, research on vegetation phenology using global weather satellites (e.g. AVHRR and MODIS) for monitoring climate responses has increased (Tucker et al., 1985, Tucker and Sellers, 1986). However, the discrimination of a mixture of several vegetation types within a pixel is a challenge in the application of remote sensing.

Remotely sensed time series data is relevant for studying PFTs, because of their consistency and repeatability at a large spatial scale (Verbesselt et al., 2010). Despite their limitations due to the phenology itself, inter-annual climate variability, disturbance, signal contamination and sensor conditions, studies on the application phenology in ecology, agriculture, modelling climate and human-induced change have been widely published. The analysis of the NDVI, Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) and Leaf Area Index (LAI) using time series satellite data from the Advanced Very High Resolution Radiometer (AVHRR), Moderate Resolution Imagery

Spectroradiometer (MODIS) and Landsat for detecting vegetation phenology has been applied to the estimation of net primary productivity (Tucker and Sellers, 1986, Reed et al., 1994), fire risk assessment (Hernandez-Leal et al., 2006), crop type discrimination (Mingwei et al., 2008), biosphere/climate feedbacks (Balzter et al., 2007a) and vegetation dynamics (Martínez and Gilabert, 2009, Verbesselt et al., 2010).

The existing land cover products (those on tree/grass cover) ranging from local, regioal to global scale (DeFries et al., 2000, Hansen et al., 2000, Hansen et al., 2003a, Hansen et al., 2005a, Herold et al., 2008, Sexton et al., 2013) are still being utilised for a number of purposes such as global estimation of burnt area (Giglio et al., 2006), in addition to the mapping crop cycles (Li et al., 2014), and greenhouse gas emissions and deforestation (DeFries et al., 2007). However, there are inherent limitations and challenges in the application of these products for different purposes. Most of these products were derived from different datasets, modelling approaches, (e.g. spatial resolution) and contained different thematic classes.

2.4.1.1 Radiative transfer theory

Radiative transfer theory provides understanding on how light interacts with canopies (Chandrasekhar, 2013) . Since the development of radiative transfer theory, studies on radiative transfer modelling have been widely published (North et al., 2010, Alton et al., 2005). The amount of radiant energy being transported in a specified frequency is affected by the nature of the surface configurations. The canopy architecture, soil scattering, and the effect of Bidirectional Reflectance Distribution Function (BRDF) are key elements being considered in remote sensing image interpretations.

The canopy architecture could be described by three main structural parameters (the vertical leaf area density, the leaf normal orientation and the function of leaf spatial dispersion). The photons striking a canopy are either absorbed or scattered (Myneni et al., 1989). For example, as explained earlier, the reflectance of vegetation in the visible and SWIR spectrums is low due to high chlorophyll absorption by plants and high in the NIR due to individual leaves and whole plant canopies which strongly scatter NIR energy (Ollinger, 2011).

The soil scattering is usually induced by the variability of the soil structure, soil moisture, colour, organic matter, and surface roughness. The behaviour of the reflected light from a soil surface lies on the nature of the surface and the amount of radiation incident upon the surface. Therefore, spectral behaviour of soil depends on the region of the electromagnetic spectrum (Fuller et al., 1997, Price, 1990, Baumgardner et al., 1986). Soil colour is one of the most useful characteristics for differentiating soil reflectance variation among soil types (Baumgardner et al., 1986). Although differences in soil reflectance depend on the soil types, the impact of soil moisture on reflectance could be higher than the soil categories. Soil moisture reflectance usually increases with decreasing soil moisture (Muller and Décamps, 2001). This means that mapping plant functional types using vegetation indices such as the NDVI need an account of soil contribution in each pixel.

The surface reflectance which many remote sensing systems rely upon is directional, therefore is influenced by the incident solar and receiving detector angles and is a function of wavelength (Shell and James, 2005). BRDF defines how light scatter when it contacts surface materials. The interaction of light depends basically on the physical characteristics of the light as well as the physical composition and characteristics of the matter. A BRDF is a function of incoming (light) direction and outgoing (view) direction relative to a local orientation at the light interaction point. The interaction of light with a surface is function of wavelengths and the nature of the surface being illuminated, certain portion of light may be absorbed, reflected and transmitted at varying degree (Wynn, 2000). All objects on the earth surface indicate degree of spectral reflectance anisotropy when illuminated by the sunlight due to BRDF (Su et al., 2009). Fassnacht et al. (2016) defined anisotropy as the property of a natural object or phenomena being directionally dependent (Fassnacht et al., 2016). Los et al. (2005) explained that bi-directional reflectance distribution function (BRDF) can alters the seasonal and inter-annual variations of plants exhibited in satellite data which consequently can limit an effective interpretation of temporal variations in land-surface vegetation. Specifically, Los et al. developed a method to assess the bi-directional effects in AVHRR NDVI using a MODIS BRDF kernels (NDVI) data. The method apply correction to the AVHRR NDVI to a standard illumination and viewing geometry. The techniques reduces BRDF effects in AVHRR NDVI observations by about 50 to 85% (Los et al., 2005).

Remote sensing methods for assessing PFTs must take account of these factors in savannas with complex vegetation structure (species composition and diversity), exposed surface with varying degree of soil types and high interannual variability due to fires and deforestation (Bombelli et al., 2009). Although radiative transfer modelling has contributed much to the understanding of canopy reflectance without the need for ground data (Alton et al., 2005), it should be noted that it is beyond the scope of this study. The principles behind radiative transfer theory are however acknowledged in the chosen techniques.

2.4.1.2 **MODIS Vegetation Continues Field (MODIS VCF)**

MODIS VCF is an Earth observation product of percent tree cover, percent non-tree vegetation (mostly herbaceous) and percent bare surface and is available at 250 m resolution provided by NASA, from the Land Processes Distribution Active Archive Centre (LP DAAC) available at http://e4ftl01.cr.usgs.gov/MOLT/. The variables, algorithms, relevance, and limitation of the product are briefly explained in the following sections;

- **Variables used for MODIS VCF products**

  MODIS vegetation continuous (MODIS VCF) field is also called MOD44B as a standard MODIS product (Townshend et al., 2011, Hansen et al., 2000). The collection 3 of the MODIS VCF is 500 m and provides information on sub-pixel estimates of the proportional tree, herbaceous and bare surface cover for the year 2001. Collection 5 (V005) is provided at 250 m resolution containing information on tree cover only, from 2000 to 2014 (Gessner et al., 2013). The version 0051 is the most recent version of this dataset (V0051) at the time of writing this thesis and provides estimates of percent tree cover, percent non-tree vegetation and percent bare surface of temporal coverage from 2000 to 2014 (on an annual basis). A global water mask is also included as an embedded ancillary layer. The primary target of MODIS VCF was to estimate tree cover greater than 5 m in height.

  MODIS VCF is produced from 16 days composite MODIS surface reflectance (bands 1-7) and brightness temperature (bands 20, 31, 32). MODIS band 1-7 are corrected for atmospheric effect with some algorithms that uses aerosol and water vapour information by the sensor.

There are 23, 16 days' composite for each year of data which were further composited to 8 days per year such that cloud effect could be minimized.

Phenology metrics were then derived from these MODIS surface reflectance data. The phenology metrics used for MODIS VCF are independent of specific timing of vegetation dynamics (Hansen et al., 2002). For instance, the metrics calculated from band 1-7 include amplitude of minimum and 5th darkest reflectance, Minimum reflectance, mean 3 darkest reflectance, reflectance at peak NDVI, maximum NDVI, mean of 3 highest NDVI values, mean of 5 highest NDVI values, Mean of 8 highest NDVI values. The approach for deriving the metrics and training data is fully described in Hansen et al. (Hansen et al., 2002)

- **The algorithms**

Regression tree algorithm was used to produce MODIS VCF product. Regression tree is one of the machine learning algorithms which is being applied to remote sensing data to characterise global land surface. It is based on the principles that create a model to predict the value of a target variable based on several input variables. It is robust to handling nonlinear relationships within remotely sensed datasets (Hansen et al., 2002).

The training datasets from previously VCF products were derived from Landsat 5, Thematic Mapper data of 1980's. In this version, as training dataset is very crucial to regression tree algorithm, efforts have been made which provide completely a new training data that matched the acquisition dates of the MODIS data (2000 to present). The training dataset is created by a discrete classification of Landsat as a training dataset verified using IKONOS, Quick-bird and Google Earth (where tree crowns can be distinguished clearly) into four classes (0, 20, 50, and 80+). The training data at 30 m were then averaged to MODIS resolution of 250 m.

The algorithm was run in three basic steps (sampling the inputs data under the training, creating the models and applying the models to the output). These processes were achieved using open-source software (Weka data mining software) as well as coding in C programming language. In step 1, 30 independent samples were created from the training dataset. In step 2, The Weka data mining software is used for model creation which provides

30 independent regression trees through bagging (M5' linear model regression tree algorithm). In step 3, the 30 independent models were applied to MODIS data which provides 30 independent results. These results were then averaged as one for every individual pixel. Detail explanations on MODIS continuous fields of vegetation cover algorithm is described in Hansen et al. (Hansen et al., 2002, Hansen et al., 2003).

- **The strength of the MODIS VCF product**

The MODIS product has been validated by the providers using limited amount of field data from two sites in Maryland, and three sites in Brazil, South America. Recent validation shows that the new (collection 5 VCF) product is significantly more accurate compared to ground based measurements of canopy cover with more than 50% improvement compared to the two old versions. The validation site in Maryland indicates a RMSE of 19.27% and 9.47%, and Mean Absolute Error of 14.37% and 7.87% for the old and new products respectively. The providers have since encouraged validation of these products (for overall improvement in the VCF tree cover products) with available ground based validation data across the globe.

Many previous studies have attempted to validate MODIS VCF product in different ecosystems. Sexton et al. (2012) who provide a global, 30-m resolution continuous fields of tree cover using Landsat and LiDAR to rescale MODIS vegetation continuous fields revealed that the Landsat-based estimates maintained consistency with the MODIS VCF. An independent validation with LiDAR measurement show an RMSE 16.8% and 17.4% for MODIS and Landsat-based estimates respectively. However, the RMSE of MODIS was 3.3% higher than that of the Landsat estimates data in an agricultural region. Los et al. (2012) provides an estimates of vegetation height and vegetation cover fraction between 60˚ S and 60˚ N at 0.5˚ spatial resolution using Geoscience Laser Altimeter aboard the Ice Cloud and Land Elevation Satellite (ICESat GLAS) data collected from 2003-2009. The application of filter to the GLAS vegetation height increases the correlation with aircraft data from $r = 0.33$ to $r = 0.78$. The GLAS product was compared globally with MODIS VCF data. The tree fractional cover calculated from the GLAS highly correlated with the MODIS VCF ($r = 0.76$). Similarly, the bare soil fraction from the GLAS was also strongly correlated with MODIS VCF estimates of bare soil fraction ($r = 0.65$).

i. The use of the multispectral bands, specifically designed for land cover monitoring, for the MODIS VCF is an improvement compared to the product derived from a single spectral reflectance. The use of multi-spectral bands also reduces background from adjacent pixels (Hansen et al., 2003b). The spectral signatures allow for a robust vegetation mapping. For example, the mean red reflectance of the five darkest red composite values performed well for tree cover discrimination. Specifically, the aggregate mean red reflectance value was found more robust in separating tree cover than any single composite red reflectance (Hansen et al., 2005b)

ii. The use of regression tree algorithm as a nonlinear, flexible model provides opportunity for large area mapping with limited training datasets (Hansen et al., 2003b).

iii. The validity of MODIS VCF product dataset for mapping percent tree cover has been proven in areas of dense tree cover in Zambia (Hansen et al., 2005b)

- **Limitations of MODIS VCF product**

Despite the strength of the MODIS VCF product in tree, non-tree, and bare surface characterisation as highlighted in the previous sections, there are associated shortcomings with the product in areas of mixed trees, shrub lands and herbaceous vegetation (e.g. savannas) (Gessner et al., 2013). The following are considered as limitations:

i. The multi-temporal metrics applied in the MODIS VCF algorithm are not specific to vegetation dynamics. The use of multi-temporal metrics, which are independent of the specific timing of vegetation dynamics, for mapping PFTs may be limited in an ecosystem with a mixture of upper canopy, under-story, and bare soil. Although MODIS VCF algorithm works in a non-linear fashion, mapping tree/grass system in some ecosystem require important consideration in the specific periodicity of the PFTs phenology due to influence of bare soil, understory and sensor's limitation (Los et al., 2005, Su et al., 2009).

The multi-temporal metrics used for MODIS VCF have their pros and cos in mapping savanna vegetation. For instance, the annual metrics such as maximum NDVI indicates a period when all PFTs type are in leaf. This metric may be more useful to forest ecosystem which are homogeneous and less complex. Hansen et al. (2005) noted that such composites perform well comparably to metrics in areas with a dominant phenological profile where common cover types share a common seasonal variation. Seasonal and inter-annual variability of the PFTs and sensor viewing geometry can posed significant challenges in discrimination PFTs (Los et al., 2000, Los et al., 2005). It has been observed that during the peak of the growing season grasslands are indistinguishable from woodlands, hence dry-season imagery may improve tree cover characterization in those regions (Hansen et al., 2002, Bucini et al., 2009, Bucini et al., 2010).

ii. Underestimation of the low tree cover and overestimation of the dense canopy. The MODIS VCF (500 m resolution) has been assessed in western USA, a region spanning semi-arid deserts, sparse dry woodlands, and cool mesic upland forests by White et al. (2005) using two independent ground-based tree cover databases. The results show overall RMSE at 24% for SWReGAP and 31% for ground-based Forest Inventory and Analysis (FIA) (1176 plots) and Southwest Regional GAP (SWReGAP) datasets (2778 plots) respectively. However, the RMSE for classes indicated a more positive values at > 10% cover than 15% for FIA and 12% for SWReGAP. At canopy cover >60% the error is high (49 for FIA and 44% for SWReGAP) (White et al., 2005).

iii. Cloud cover: the influence of cloud in the MODIS reflectance data can also be considered as another limitation of the product as it is the input data for the MODIS VCF. However, the providers have made a quality layer (at per pixel level) available to users for every corresponding pixel of the variables estimated (percent tree cover, percent non-tree vegetation (mostly herbaceous) and percent bare surface). Users are cautioned to take account of the pixel affected by the cloud.

2.4.1.2 Mapping PTFs in savannas

The measurement accuracy and precision with regard to mapping PFTs vary with the modelling approach, datasets and the ecosystems in which the study is being carried out. (Hansen et al., 2003a, House et al., 2003, Gitelson, 2013). Unike the estimate of fractional cover in a uniform land surfaces such as cropland or a dense forest, the measurement accuracy of the mixed vegetated (e.g. tree/grass fractional cover) landscape requires consideration of many factors.

PFTs in savannas are generally not distributed uniformly across the landscape, but rather show different degrees of clumping or patchiness. Also, the effects of soil moisture and inter-annual weather variability on vegetation phenology can affect the information content of satellite imagery, for example, Landsat imagery has revisit period of 16 days. This revisit rate means that Landsat can miss important phenological changes between two subsequent acquisitions (Hilker et al., 2009). It is therefore challenging to identify tree cover changes with a snap-shot of images. This means for an accurate assessment of tree cover, there is a need for not only a method that is robust to inter-annual variation due high inter-annual weather variability in savannas, but also remote sensing data with a frequent revisit period to capture phenology of PFTs adequately.

Tree/grass separation is one of the major challenges for estimating tree or grass cover due to a mixture of the two PFTs in savannas. Mapping tree cover for example, requires that the contribution from herbaceous layers and the soil background is removed (Gill et al., 2009, Tim Danaher et al., 2011, Guan et al., 2012, Gessner et al., 2013). Previous effort have been made to separate trees and grass (Chidumayo, 2001a, Jolly and Running, 2004). The need for such techniques in the analysis of multi-temporal satellite data is one of the most significant challenges facing remote sensing (Martínez and Gilabert, 2009, Boke-Olén et al., 2016). For accurate representations of PFTs, one important thing to considered is the differences in tree/grass growing cycles over the years.

Trees and grasses have distinct phenological cycles in savannas (Moore et al., 2016, Boke-Olén et al., 2016). Figure 2.7 explained the theoretical contributions of trees and grasses to total landscape and how tree green-up curve can be extracted from the satellite data (NDVI data). Tree normally green earlier and stay green for longer period, but grasses usually have high productivity during the peak of the gwowing season. In the end of the growing season, grass dries while tree greening remain for sometimes. Therefore, tree is mostly found to be green while grass is dry in autumn and spring, except where perenial grass exist (Archibald and Scholes, 2007b).

Figure 2. 5: (a) Theoretical contributions of trees and grasses to total landscape LAI over a growing season. Trees go green earlier, and stay green for longer, but grasses have a higher LAI at the height of the growing season (b) Schematic showing how to extract the tree green-up curve from satellite NDVI data (Archibald and Scholes, 2007a).

The phenology metrics for trees and grasses is however affected by weather variability, species composition and the geomorphology of the land surface in savannas (Hill et al., 2011, Cho et al., 2012, Higgins et al., 2011). Previous studies applied moving average to extract tree/grass green-up curve. In spite of the fact that the moving average method usually maintains the area and mean position of the seasonal peak in a time series of satellite data (Eklundha and Jönssonb, 2012), it is inappropriate to also assume that the moving average captures all phenology metrics including those due to noise, fire and land degradation. There are also no standard criteria for choosing the delay time in the moving window which consequently affects the degree of changes being observed in the data (de Beurs and Henebry, 2010, Verbesselt et al., 2010).

Many approaches to estimates tree/grass fractions using remote sensing are not robust to interannual variability and noise reduction (Lhermitte et al., 2011). The growing cycles of various PFTs are often not considered for their characterisation. Therefore, a reliable technique for tree/grass discrimination should be one that is capable of decomposing satellite signals into useful phenological characteristics. Signal decomposition methods are considered very efficient in performing this task since they can decomposed satellite signals into cycles, with amplitude and phase angle (timing of greening) and minimise the effect of noise in satellite data (Andres et al., 1994, Lhermitte et al., 2011). The next section reviews signal decomposition methods applied to remote sensing data for PFTs characterisation.

2.4.1.3 Signal decomposition and smoothing techniques for the analysis of PFTs.

Several approaches for spectral signal decomposition of remote sensing time series data have been proposed (Cleveland et al., 1990, Lu et al., 2003). Lu et al., (2003) developed a model which decomposes the tree/grass phenology signals from Advanced Very High-Resolution Radiometer (AVHRR) data based on assumptions that take account of vegetation biophysical characteristics and noise reduction. The model is an extension of the Seasonal-Trend Decomposition based on the Loess algorithm (Cleveland et al., 1990).

Although the model retrieves periodic information on regional and global vegetation, it has several limitations, including the effects of soil background colour. Another limitation of the model is the threshold used to determine the low and high varying components of woody and herbaceous cover. The tendency for over or underestimation of these components is very high and can result in false separation of the vegetation index components (Lu et al., 2003).

Several modern approaches have been proposed for signal decomposition of time series data such as empirical mode decomposition (EMD) (Kong et al., 2015), principal coordinates neighbor matrices (PCNM) (Pottier and Evette, 2011, Lewis et al., 2014, Dray et al., 2006), wavelets transform (WT) (Martínez and Gilabert, 2009) and harmonic analysis (Kostadinova et al. 2017).

EMD is one important technique which decomposes signals into its components adaptively without using a priori basis. It is very useful to ecological phenomena expecting to follow a non-stationary dynamic in the time domain (Attoh-Okine et al., 2008). The EMD decomposes data into Intrinsic Mode Functions (IMFs) and residue. The IMFs usually stand for the oscillatory mode of the original data while the residue stands for the overall trend. One of the challenges of using EMD is the lack of sufficient theoretical framework which often make characterization and evaluation of time series data difficult using this approach (Niang et al., 2010).

PCNM is developed for building a spatial matrix based on a spectral decomposition of space. Moran's Eigenvector Maps (MEMs), which are based on PCNM, uses the spatial or temporal coordinates of the observations to compute a series of sine waves similar to a Fourier decomposition (Legendre and Gauthier, 2014). Although it has proved very useful as it considers the spatial structure of an ecosystem through spatial weighting matrix, this method requires large samples, providing a large range of distances between plots (Lewis et al., 2014, Pottier and Evette, 2011).

WT is another technique that decomposes the signal into different frequencies. Although, wavelets transform has several advantages in the analysis of time series data which are commonly non-stationary dynamic, the standard WT is shift sensitive (which may lead to complete different transform coefficients) and lacks phase information (Muhammad et al., 2002, Strang and Nguyen, 1996).

Archibald and Scholes (2007) developed a method for separating tree and grass responses to environmental cues in African savanna. This assumed that trees and grasses exhibit different biological characteristics in their life-forms which can be separated using, for example, a moving average method to extract their phenological metrics. In savannas, trees green-up before grasses because they rely less on water from rainfall. Their maximum green-up rates, as well as the rate of greening, are generally constant between years (Archibald and Scholes, 2007). Results show that while productivity of grasses is higher than trees during the peak of the rainy season, trees have a less variable phenological cycle than grasses.

Most spatiotemporal statistical methods currently being used for time series analysis have their strength and weaknesses (de Beurs and Henebry, 2010). de Beurs and Henbry (2010) who discussed 12 existing spatiotemporal statistical methods being used for the analysis of land surface phenology (LSP) from time series of satellite data, emphasized a lack of consensus regarding nomenclature, model significance, uncertainty and error structure. It is therefore challenging to find a given method that is best for all vegetation types (de Beurs and Henebry, 2010, Atkinson et al., 2012). However, Atkinson et al. (2012) compared four smoothing algorithms: Fourier analysis, the asymmetric Gaussian model, the double logistic model and the Whittaker filter to assess the annual vegetation growth cycle and a reliable estimate of phenological parameters in four vegetation types. The Fourier analysis and the Whittaker filter outperformed the asymmetric Gaussian and double logistic models (Atkinson et al., 2012).

Fourier analysis decompose time series data into varying signals and can improving data quality through noise reduction (Jakubauskas et al. 2001). The method has since been recommended as one of the most efficient techniques for modelling and the change detection of ecosystem dynamics (Lhermitte et al., 2011). Lhermitte et al. (2011) explained that Fourier analysis is a suitable method for modelling and change detection of PFTs dynamics in consideration to time series satellite characteristics such as:

i. Serial correlation: This occurs when time series values are correlated between different temporal observations within one-time series. This is often caused mainly by the seasonal variation of vegetation (Zoffoli et al., 2008). Its implication is that where temporal observations are combined, for examples on trees and grasses variables, one PFTs may contribute in the information to yield a relationship. Harmonic analysis partition time series observations into annual and biannual signals. It is evident that in savannas, trees and grasses have different phenological cycles. The onset of grass is mainly supported by precipitation, soil moisture, and day length while tree growth is usually supported by the day temperature and day length (Archibald and Scholes, 2007b, Boke-Olén et al., 2016). The use of field data or prior information on the biophysical parameters could be useful in harmonic analysis. One of the advantages of harmonic analysis over other methods such as principle component analysis is the evaluation of time-series data on a per-pixel basis. Periodicity of a pixel in the time series data is usually evaluated independently of other pixels unlike in a principle component analysis that evaluates the variance of all pixels in all images of the time-series data to derive transformation coefficients which are later applied to each pixel (Jakubauskas et al. 2001).

ii. Time series stationarity - temporal observations with larger variance may influence the behavior of several similar measures when the time series variance is not constant over time. The influence of non-stationary dynamic phenomena such as tree/grass phenology is minimal with harmonic analysis approach, because of its requirement for a long time series data (de Beurs and Henebry, 2010). The comparison of means and application of statistically significant harmonics in Fourier analysis could provide a realistic manner with which to assess changes either at annual or within the interannual variability. In contrast, the use of threshold methods which set arbitrary level of amplitude to determine the start of season or end of season is disadvantageous since it is unlikely for the changes in the time series data to be stable through time and could change significantly, for example due to disturbance processes (de Beurs and Henebry, 2010).

iii. Temporal resolution - the variability in the time interval between consecutive observations is also an important factor to be considered. This relates with the sensor (e.g. MODIS and Landsat) revisit period. Both high and low temporal resolutions have their advantages. A high resolution is more appropriate if the phenomena are dynamic. Harmonic analysis requires a high temporal resolution and can be said to be appropriate since the variability of tree and grass phenological cycles may not easily lead to serial correlation of the temporal observations. And with the fact that changes in plant phenology could be seasonal (e.g. intra-annual changes due to differences in PFTs life cycle), gradual (e.g. because of climate variability) and abrupt (e.g. due to fires and drought) (Lhermitte et al., 2011, Verbesselt et al., 2010, de Beurs and Henebry, 2010). Harmonic analysis can decompose time series data into various cycles to assess phenological development of vegetation.

iv. Noise - Satellite-derived NDVI time series are necessary to the remote sensing of vegetation phenology, but their application is limited by the present of noise resulting mainly from varying sun sensor-surface viewing geometries and atmospheric conditions. Effective noise reduction is required for the analysis of time series data, especially those from passive optical sensors. The errors in the NDVI due to noise include negatively biased noise and the present of spurious drops and spikes (Hird and McDermid, 2009). Harmonic analysis employs the decomposition of noise-affected time series into periodic signals in the frequency domain and was found to maintain the integrity of the data in the process. The effectiveness of certain methods to reduce noise sometime depends on the individual metrics being extracted and the vegetation indices (Hird and McDermid, 2009, Roerink et al., 2000).

v. Unequally spaced observations or missing data - Harmonic analysis is good for interpolating missing values or the compositing of missing data (Roerink et al., 2000).

Because of the suitability of harmonic analysis for analysis ecological phenomena with time series data, the next section provides a more detailed review of its application to identify additional gaps in the literature with a focus on PFTs discrimination using a time series of remote sensing data.

2.4.1.4 Harmonic (Fourier) analysis for phenological signal decomposition of PFTs

Harmonic analysis is based on the Fourier transformation and is one of the most reliable techniques for land cover discrimination from decoupled vegetation phenological signals (Andres et al., 1994). Harmonic analysis enables a phenological time series to be expressed as a sum of cosine waves and an additive term. Each individual wave is called a harmonic term and is characterized by its amplitude (height of the maximum), frequency (number of cycles) and phase (delay from time zero). Figure. 2.7 illustrates harmonic terms (amplitude and phase).

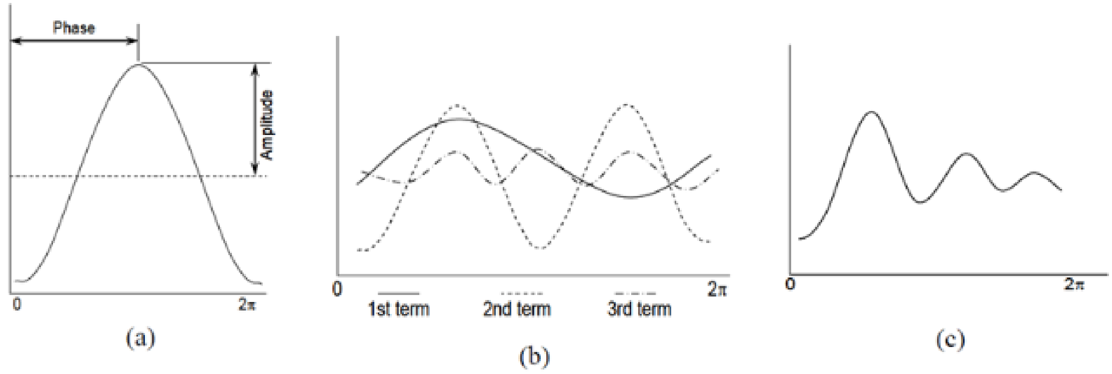


Figure 2. 6: (a) simple cosine curve showing amplitude and first harmonic (b) 1st, 2nd 3rd harmonic terms (curve produced from addition of curves in b) (Source :(Jakubauskas et al. 2002), adapted with kind permission of Elsevier).

Previous studies have applied this method to successfully characterize seasonal changes for natural land cover/land use types (Jakubauskas et al., 2001, Jakubauskas et al., 2002, Canisius et al., 2007, Westra and De Wulf, 2007). Jakubauskas & Legates (2000) implemented harmonic analysis to a nine-year time series of vegetation index from NOAA AVHRR to identify PFTs and analyse their changes across a given time span. In their study, the amplitude values were used to assess the variation in the temporal NDVI data. While high variance was captured by the first and additive terms on certain land cover types, a few crops exhibited bimodal NDVI periodicity. In another major study, Moody and Johnson (2001) applied discrete Fourier analysis to derive a mean-phase-amplitude space to separate six vegetation types from different geographical regions by classifying AVHRR data. Validation results indicated that grassland had the highest accuracy (77%) and the most common confusion in their classification accuracy was between grassland and savannah (23%). This was partly due to mixed pixels being influenced by annual grass understory variations. Geerken et al. (2005) investigated the application of Fourier analysis to classify rangeland vegetation type and coverage from NDVI values. The study compared the unsupervised classification of phenology metrics derived from Fourier analysis and original NDVI. The validation results for Fourier derived phenology metrics were strongly correlated with field data on homogeneous plant species fractional cover (*Noaea mucronata* (R^2 = 0.50) and *Cornulaca setifera* (R^2 = 0.75)). The unsupervised classification of the Fourier derived phenology metrics for vegetation types had the best accuracy (overall accuracy = 73% and kappa = 0.60). The harmonic analysis is therefore a good approach for assessing PFTs. Although the estimate of spatial variations in sub-grid PFTs has been considered more appropriate than by defining class boundary, it is far more challenging in a heterogenous ecosystem (Geerken et al., 2005).

Jakubauskas et al. (2001) demonstrated the use of a single year data to characterise different PFTs (Corn, Shortgrass prairie, Sandsage prairie, Alfalfa). Harmonic functions such as amplitude values and phase angle were the major components upon which the discrimination analysis of these crops were based. Although they have used variance to characterise the differences in the harmonic terms, a test of the confidence interval of the observation is not applied in the estimate of harmonic terms. This approach might have a negative implication for change analysis within the range of monthly variability as to whether those changes between harmonic terms have really occurred (Chen et al., 2011, de Beurs and Henebry, 2010).

Many previous studies from remote sensing using harmonic analysis have paid more to crop rather than tree/grass characterisation (Canisius et al., 2007, Mingwei et al., 2008). Mingwei et al. (2008) showed that harmonic analysis can identify double cropping systems. The results for the estimated crop areas have shown to be correlated with statistics derived from field surveys. The study demonstrated an improved method for discriminating PFTs over the traditional means of categorisation and classification in remote sensing analysis (Chen et al., 2011, de Beurs and Henebry, 2010). Similarly, Canicus et al., (2006) demonstrated the used NDVI time series 10-day composites derived from NOAA AVHRR (at scale of 0.1°) data to discriminate bimodal agricultural areas (where two seasons of cultivation occur per year) from other land cover types (e.g. forest, bushes, mixed rain-fed areas etc.) through the application of the Fourier approach and decision classifier to characterise harmonic signals. The amplitude and phase shift for all land cover types were estimated for the study area. The amplitude signals of the second harmonic term of where bimodal agriculture was predominant had a relationship (R^2 = 0.38) with previous statistics of the sample areas (irrigated areas). The relationship between the two variables was used to generate an estimate of the bimodal agriculture area. However, Some of the coarse satellite time series data, such as the NOAA AVHRR, are highly limited for PFTs discrimination due to their requirement of re-compositing, low resolution and quality assessment (Loveland et al., 2000, Roerink et al., 2000). The moderate resolution and good atmospheric correction for MODIS data (e.g. MOD1Q31) have made them more advantageous in land cover discrimination than the AVHRR.

The influence of low spatial resolution for some satellite data may increase the difficulty of PFTs discrimination as a pixel may contained a mixture of PFTs (as the land cover is spatially heterogeneous) and can result in confusion in defining the required signals using conventional classification methods. Another limitation relates to a selection of harmonic terms for the PFTs discrimination, for instance, where the analysis relies on second order harmonic to discriminate PFTs, a separation of these PFTs can be challenging as more than one PFTs may have bimodal characteristics. This means that a good knowledge of ecosystem dynamics, and the PFTs, can be a useful guide to this analysis. With field data on the PFTs perhaps based on fractional cover, an empirical analysis using moderate resolution data is more likely to improve tree/grass characterisation using harmonic analysis.

Despite limitations identified in the application of harmonic analysis, previous studies have demonstrated its usefulness for the analysis of time series data to characterise PFTs. From the foregoing, it is evident that this method has great potential for discriminating main PFTs with distinct phenological characteristics (e.g. trees and grasses). However, one of the most important questions in the application of Fourier (harmonic analysis) is the assessment of statistically significant harmonics in the measurement of land surface phenology observations which have not been adequately addressed in remote sensing studies (de Beurs and Henebry, 2010). Most previous studies that applied harmonic analysis to satellite data only assumed that noise is contained within the higher-order harmonic components (Moody and Johnson, 2001, Lhermitte et al., 2011, Gessner et al., 2013) without measure to their degree of significance. The need to incorporate more statistical techniques to test confidence intervals is important (Bloomfield, 2004).

2.4.2 Radar remote sensing

Radar is an acronym, meaning radio detection and ranging. Radar is best known as active sensor system. The active systems have the capacity to generate energy pulses and collect them after the surface target reflects them back (Chuvieco and Huete, 2010). Radar sensors are capable of imaging all parts of the globe regardless of cloud cover, day or night. Depending on wavelengths and polarisations, radar data can be affected by the interacting medium such as the vegetation (e.g. trunk, leaves, branches and surface roughness (Balzter et al., 2007b, Lillesand et al., 2014, Chuvieco and Huete, 2010). Radar data are powerful microwave data used for vegetation structure retrieval. It has been used extensively for biomass mapping and canopy characterisation (Lucas et al., 2007). The use of dual wavelength SAR interferometry for canopy height characterisation and carbon stock estimation has been previously demonstrated (Balzter et al., 2007b).

2.4.3 LiDAR remote sensing

The continued development of remote sensing technology has increased technical capabilities which stimulates and further facilitates the assessment of vegetation biophysical parameters using data from the active sensors. LiDAR is an acronym, meaning "Light Detection and Ranging". LiDAR systems work with the polarised light and operates within ultraviolet to near-infrared range of the spectrum. Like radar systems, it emits pulses to the observed surface and records the reflected energy from the target. LiDAR sensor calculates the distances from the platform to the observed target by measuring the precise time that the return signal reaches the sensor. LiDAR data are recognised as the most accurate data used for assessing vegetation structure by measuring the three-dimensional (3D) structure of a forest or savanna vegetation attributes (Chuvieco and Huete, 2010, Asner et al., 2011).

LiDAR data are often used to estimate tree canopy height, map individual tree species and fractional cover or the spacing of individual trees and shrubs. Most techniques require a Digital Elevation Model to be derived from a laser pulse with which to distinguish terrain and vegetation returns (Chen, 2007). A Digital Surface Model (DSM) is used to calculate tree canopy height, biomass or FVC. Popescu & Wynne 2004 used kriging while Persson et al. 2002 have used an active contour algorithm to create a Canopy Height Model (CHM) or DSM.

The use of space-borne LiDAR, specifically the Geoscience Laser Altimeter aboard the Ice Cloud and Land Elevation Satellite (ICESat) data is also receiving increased attention. ICESat data have been used for the estimates of tree canopy height and biomass (Lefsky et al., 2006, Lefsky et al., 2007, Gwenzi and Lefsky, 2014). Popescu et al. (2011) shows a comparison of satellite LiDAR with foot airborne LiDAR ground elevation and four vegetation variables. The results indicated a strong correlation for terrain elevation between GLAS and airborne LiDAR with R^2 of 0.98 and a root square error of 0.78 m.

The accessibility of GLAS-LiDAR from regional to continental and global extents makes it a useful resource to assist with the assessment of various PFTs. A previous study demonstrated a successful retrieval of canopy height which correlates (59–68%) with field estimates of tree height in three different ecosystems using a combination of ICESat and Shuttle Radar Topography Mission (SRTM) (Lefsky et al., 2006). The synergistic use of ICESat data with other datasets to overcome its patchy coverage has since been recommended (Popescu et al., 2011, Mitchard et al., 2012).

While the use of LiDAR data, the Airborne data, is quite novel, the high cost of obtaining the data makes it difficult for researchers to explore various environmental conditions. However, there is increased attention regarding the integration of different datasets to achieve maximum benefit and reduce uncertainty. Current remote sensing methods have indeed shifted to the novel application and complimentary use of the data through data fusion and integration (Smith et al., 2014b). Many studies using data from the passive optical sensors data now rely on data from the active sensors for model calibration and validation (Gill et al., 2009, Sexton et al., 2013, Naidoo et al., 2015).

2.6 Research gaps, questions and objectives

In the literature cited above, knowledge gaps are quite broad. Thus, in the context of this research, the following are considered most important. For each gap, research questions and objectives were outlined. In addition, the analysis chapter that addresses each research question is stated.

Research gap 1:

The uncertainty of the metrics of land surface phenology from Fourier analysis have not been adequately addressed in remote sensing studies (for estimating tree/grass fractions using satellite data).

Research question 1:

How can satellite time series data for tree/grass signals be decomposed into statistically significant harmonic terms?

Objective 1:

To decompose tree/grass signals into statistically significant harmonics. This was addressed in chapter four (4).

Research gap 2:

Due to high interannual variability in savannas, the assessment of changes in tree and grass composition is challenging**.**

Research question 2:

How does tree/grass phenology vary inter-annually?

Objective 2:

To assess spatiotemporal variability of tree/grass phenology over a 14-year record of MODIS NDVI time series data. This was addressed in chapter four (4).

Research gap 3:

The estimates of tree/grass fraction in savannas is challenging due the mixture of the two PFTs and bare soil. Moreover, some of the existing remote sensing products on tree cover acquired at sub pixel level do not capture woody species adequately.

Research question 3:

How well the satellite derived metrics from MODIS data be used to estimate tree/grass fractions?

Objective 3:

To estimate tree/grass cover using field data collected from a field campaign. This was addressed in chapter four (4 and 5)

Research gap 4:

The accuracy of remote sensing products is very vital for understanding land surface interactions, yet it is challenging to establish model to estimates tree cover at a considerable accuracy.

Research question 4:

How accurate can a tree fractional cover estimated using MODIS data derived satellite phenology metrics be?

Objective 4:

To test the performance of tree cover estimated using plot data from a field campaign in 2015 and compare it with previous satellite products (such as LiDAR and MODIS VCF). This was addressed in chapter six (6)

Chapter 3

Study area and general method

3.1 study area

KNP is located on the northeastern tip of South Africa, bordering Zimbabwe to the north and Mozambique to the east (Figure 3.1). It is one of the largest national parks in Africa. It is approximately 2 million hectares in size and extends 380 km from north to south and 60 km from east to west. Its elevation ranges from 260 m to 839 m above mean sea level. The mean annual precipitation ranges from 440 mm in the north to 750 mm in the south. The park has large perennial rivers as well as seasonally flooded dryland river channels.

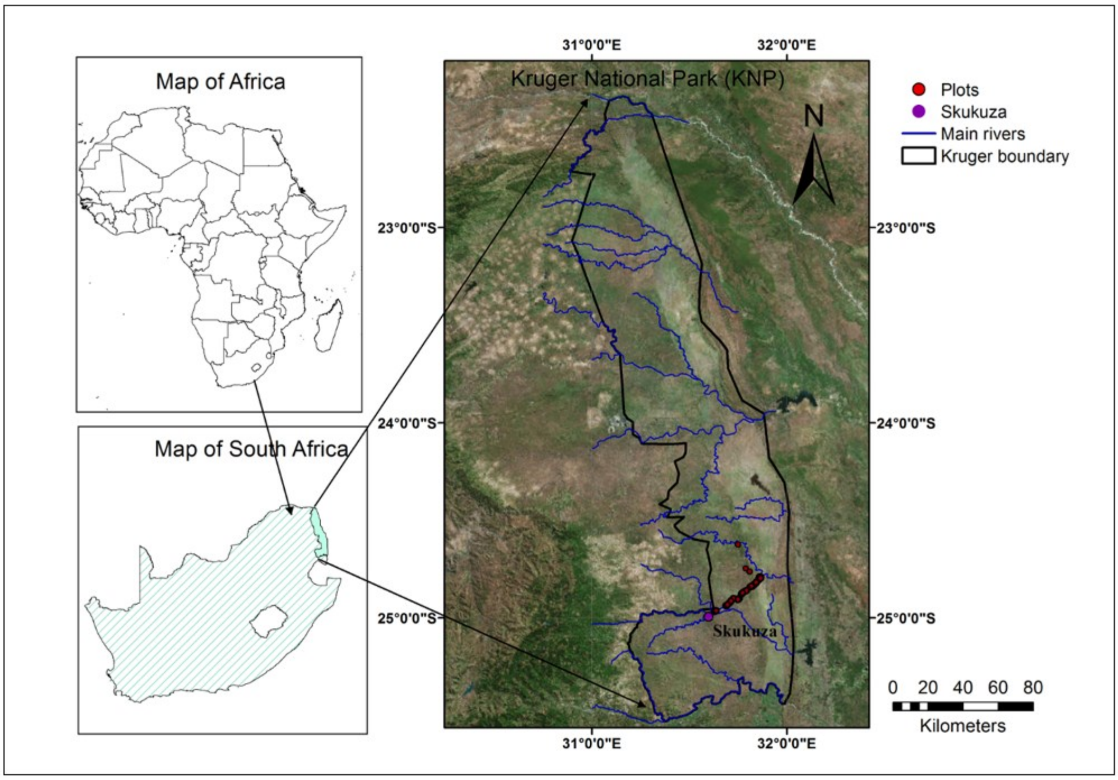


Figure 3. 1: Location of the study area of Kruger National Park in South Africa and its main river courses, an indication of the locations of the sample plots of the field data collection in 2015. The red circles indicate the field plot locations.

Geologically, the area is divided into two main zones. The western part is situated on granite and the east on basaltic bedrocks. Therefore, a strong influence of geological structures on soil formation processes and plant species distribution is observable in the park. Soils within KNP can be grouped into a more fertile basaltic zone in the east, and a less fertile granitic zone in the west (Venter et al., 2003a). The soil type (Figure 3.3) has a strong influence on the plant species. The granitic soils are dominated by wooded species like *Combretum apiculatum*, while the basaltic soils support finer-leaved species such as the *Acacia* (Codron et al., 2006). Figure. 3.2 shows the geology of the KNP.

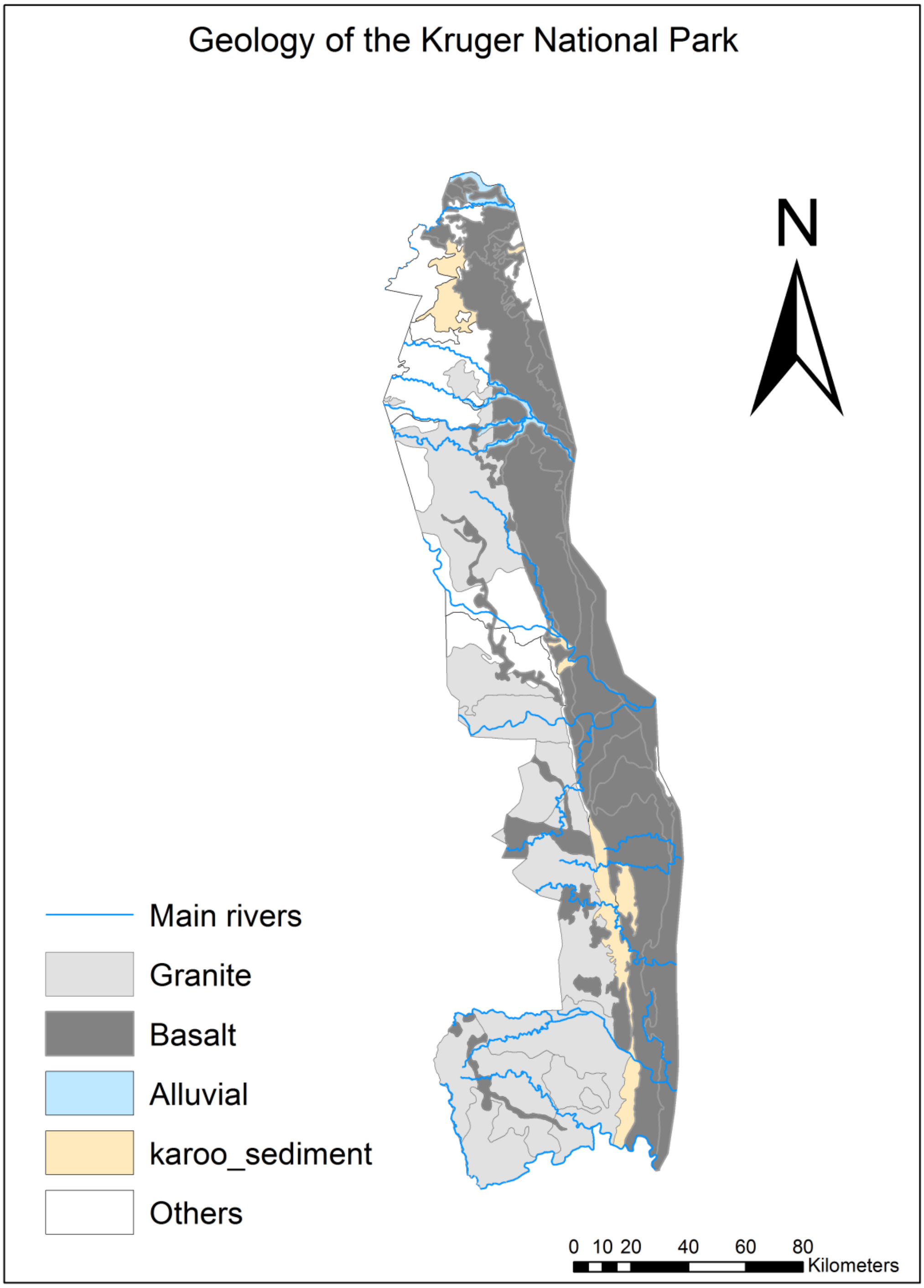


Figure 3. 2: Location of the study area of Kruger National Park in South Africa showing geology.

A global soil and terrain database at a scale of 1:1 million developed by the initiative of the International Union of Soil Sciences (IUSS), the United Nations Environment Programme (UNEP), the FAO, and the International Soil Reference and Information Centre (ISRIC) (van Engelen and Hartemink, 2000) available at http://www.isric.org/content/soilgrids is presented in Figure 3.3. Eutric regosols, Nitisols, and Luvisols are the main types of soil in our field plots locations.

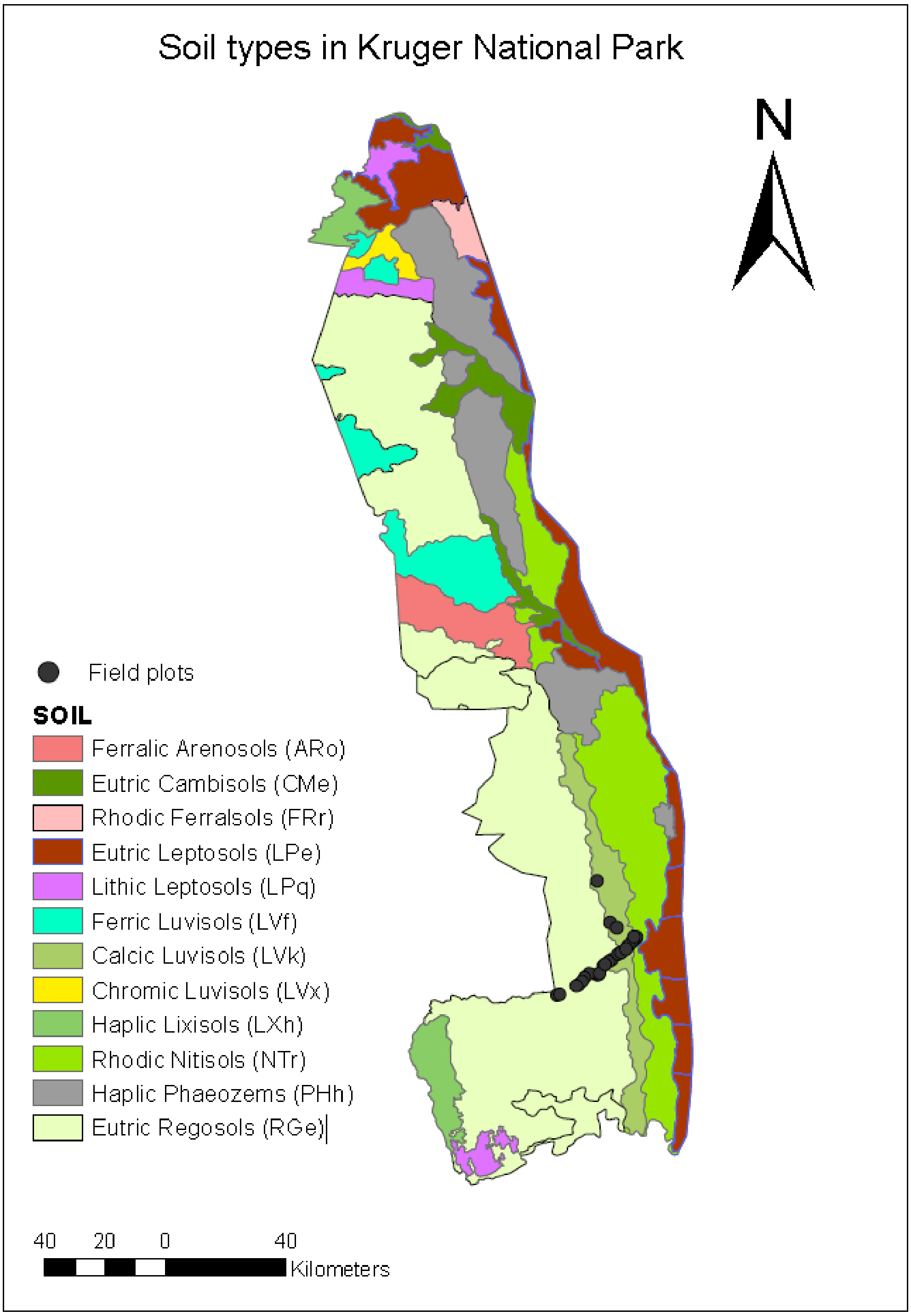


Figure 3. 3: Location of the study area of Kruger National Park in South Africa showing geology.

KNP is home to a gradient of more than 1,900 plant species including trees and grasses (Eckhardt et al., 2000). The status of each PFTs has been a focus of savanna ecology for many decades. The park has 20 ecotones (e.g. Skukuza thickets, open trees, dense trees, and bush savanna) based on a floristic classification (Eckhardt et al., 2000, Archibald and Scholes, 2007b, Khalefa et al., 2013). The dominant tree species in the southwestern part include *Combretum apiculatum, Acacia nigrescens, Sapirostachys Africana, Combretum hereroense, Sclerocarya birrea, Terminalia sericea, Combretum zeyheri* etc. The drier northeastern part is dominated by mopane savanna. Grass species include *Aristida congesta spp, Digitaria eriantha, Erasiantha, Uracholoa mosambicensis, Themeda triandra, Panicum colouratum* etc. (Eckhardt et al., 2000, Archibald and Scholes, 2007b, Khalefa et al., 2013). Figure. 3.4 is a map of KNP showing landscape and species types. KNP was chosen as a study site because of its size, diversity of vegetation formations at the MODIS resolution of 250 m, and absence of agro-pastoral systems. Harmonic analysis is more applicable in a diverse area with varied species composition, density, and environmental conditions (Moody and Johnson, 2001).

There are several minor subdivisions of landscape formations within the park, each with different species composition. Van et al. (2014) have provided a description of various tree species, as listed below, according to their main features and their habitats in southern Africa.

1. ***Acacia nigrescens*** is of varying size from a medium to large deciduous tree. Its trunk and thicker branches are usually persistent with thorns. Its leaves are usually 10-30 x 7-30mm. Flowers are usually pale cream to pale yellow appearing before or with the new leaves. It is found in the Bushveld and Rugged veld, commonly in heavy soils.
2. ***Colophospermun mopane*** is commonly shrub with medium to tall deciduous trees. It can be found as a single or multistemmed. It is dark grey and has thick bark. Its leaves are alternate compound with two leaflets (bifoliolate), petiole 20-40mm long. It is found in many landscapes such as mopanevveld, shrubveld on calcrete, gabbro and basaltic soils.

3. ***Combretum apiculatum*** is a small to medium-sized deciduous tree and is multistemmed. Its leaves are usually opposite, simple and elliptic to broadly obviate. It is usually hairless. Its flowers are usually small and creamy yellow. It is found in Bushveld, mostly in rocky places. It has dense greyish hair. It is found in Bushveld, especially floodplains.

4. ***Combretum hereroense*** this should not be confused with Combretum apiculatum. It is a shrub or small deciduous tree. It is commonly multistemmed with a roundish crown found in Bushveld. It is found mostly on sandy soils and on termitaria.
5. ***Schlerocarya birrea*** is also a medium-sized to large deciduous tree. It has a spreading crown and is rounded. Its leaflets are usually 3-7 pairs plus terminal one, dark green above and pale and bluish green below. The species usually flowers in unbranched parts before the new leaves. It is found in Bushveld and woodland.
6. ***Terminalia sericea*** is a small to medium-sized deciduous tree. Its branches are dark brown or purplish. Its leaves are usually clustered towards the end of branches. The flowers are usually small in auxiliary spike, pale cream to pale yellow. It is found in Bushveld on deep, sandy soils often in dense stands.
7. ***Salvadora australis*** is shrub to small evergreen tree, its trunk is short with round branches which usually drop to the ground.
8. ***Euclea divinorum*** is a small evergreen tree, usually multistemmed. Its leaves are dark green to greyish green. It is found in Bushveld, commonly on brackish floodplain along rivers and or termitaria.

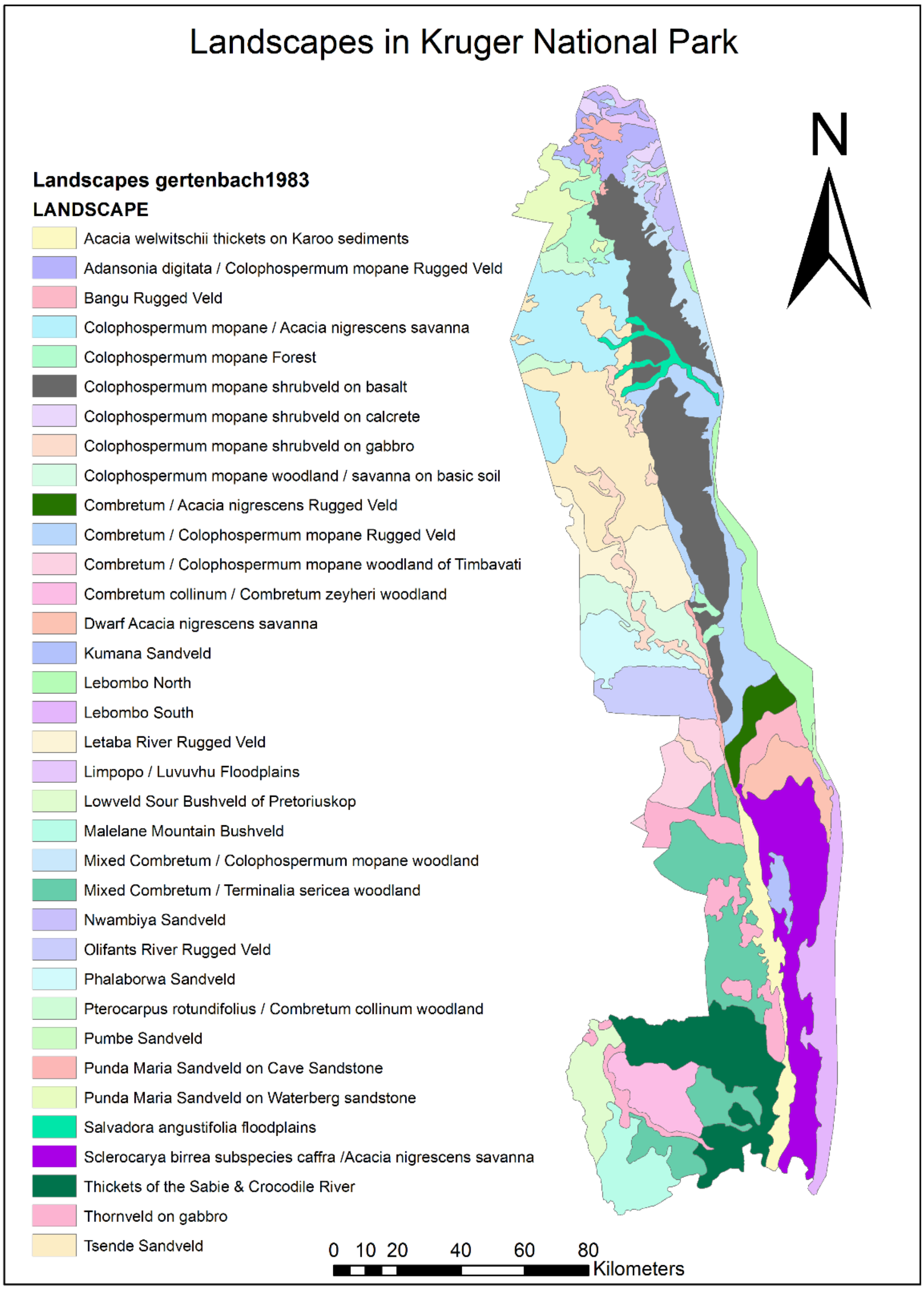


Figure 3. 4: Location of the study area of KNP showing landscape and tree compositions

3.3 Data

3.3.1 MODIS time series NDVI data

MODIS Normalized Difference Vegetation Index (NDVI) data (MOD1Q31) were obtained from the National Aeronautics and Space Administration (NASA) via http://reverb.echo.nasa.gov/rever. A total of 322 images (from July 2001 to June 2015) were used for harmonic analysis. For single year analysis, MODIS images from July to June (e.g. 2014/2015) were considered for each analysis year. The estimates of tree cover which were not based on harmonic analysis uses only dry season images for each growing season. A previous study developed a method based on MODIS Ross-Thick and Li-Sparse kernels to estimates the effects of BRDF in NOAA-AVHRR NDVI time series. The results indicated that in most cases uncorrected NDVI time series do not reflect actual seasonal and interannual variation in vegetation greenness. It was found out that the techniques reduces BRDF effects in AVHRR NDVI observations by about 50% to 85% (Los et al., 2005). MOD13Q1 used in this study, is a gridded level 3 product provided at 250 m spatial resolution every 16 days produced from atmospherically corrected bi-directional surface reflectance factors (BRFs) and masked for water, clouds, and cloud shadows (Strahler et al., 1999).

MODIS NDVI has been used widely for retrieving vegetation composition such as vegetation structure and annual net primary productivity (ANPP) dynamics in grassland-shrub land areas (Moreno-de las Heras et al., 2015), for tree cover change (Gill et al., 2009), and tree-grass separation/green-up dates (Archibald and Scholes, 2007b) and for Analysis of trends to assess the effects of CO_2 fertilization effect in global vegetation (Los, 2013). NDVI has also been used to examine the relationship between vegetation productivity and rainfall distribution along environmental gradients (Foody, 2003, Chamaille-Jammes and Fritz, 2009). Furthermore, Jung and Chang (2015) assessed land-cover change from harmonic analysis using NDVI data. Muñoz Peña and Navarro (2016) assessed the spatiotemporal variability of NDVI to study deforestation using harmonic analysis. NDVI is used here as the proxy of vegetation productivity as numerous studies have identified a strong relationship between the NDVI and NPP (Prince and Goward, 1995, Zhu and Southworth, 2013, Mbow et al., 2014).

3.3.2 MODIS SAVI data

The Soil Adjusted Vegetation Index (SAVI) was developed mainly to minimize soil brightness effects from spectral information involving red and near-infrared (NIR) wavelengths (Huete, 1988). In this study, the SAVI vegetation index was derived from the dry season images of the MODIS data. The SAVI index was calculated as thus:

$$SAVI = (NIR - R) / (NIR + R + L) * (1 + L) \quad (3.1)$$

where the NIR is the near-infrared band, R is the red band and L stands for a soil correction factor (ranges from 0 to 1). L = 0.5 was used in this study being appropriate for savanna ecosystems (Gilabert et al., 2002). The estimates of tree cover using SAVI index were based on individual growing season which normally starts from October to April. To capture a complete cycle, dry season images for each growing year from July to June were considered. So, for example, the estimates of tree cover for the year 2014/2015 uses SAVI index calculated from dry season images of this season.

3.3.3 MODIS vegetation continuous field (VCF)

MODIS VCF data have been described in the previous Chapter. It is an Earth observation product of percent tree cover, percent non-tree vegetation (mostly herbaceous) and percent bare surface and is available at 250 m resolution provided by NASA, from the Land Processes Distribution Active Archive Centre (LP DAAC) available at http://e4ftl01.cr.usgs.gov/MOLT/. It is called MOD44B as a standard MODIS product (Townshend et al., 2011, Hansen et al., 2000). This product used in this study, is the most recent version of this dataset at the time of writing this thesis (collection 0051). This version provides estimates of percent tree cover, percent non-tree vegetation and percent bare surface of temporal coverage from 2000 to 2014 (on an annual basis). Specifically, MODIS VCF for the years 2008 and 2014 were used in this study for validation and comparison purposes. None of the MODIS VCF (2014) pixel values was of bad quality within field plots used in this study.

3.3.4 LiDAR/SAR woody cover map

A woody cover map of KNP produced by the Ecosystem Earth Observation Research Group (Natural Resources and the Environment) of the Council for Scientific and Industrial Research (CSIR), South Africa, was used as independent validation dataset. This map was produced using 14 dual-polarized (HV, HH) 12.5m L-band ALOS

PALSAR images trained with a random forest algorithm and 25,000 ha of airborne LiDAR data (Wessels et al., 2011). The Phased Array Type L-band Synthetic Aperture Radar (PALSAR) is an active microwave sensor developed by the joint project between the JAXA and the Japan Resources Observation System Organization (JAROS). The L-band frequency could achieve a high cloud-free and day-and-night land imaging. The L-band SAR data were known for its advantages over other SAR data such as the C-band (Urbazaev et al., 2015, Li et al., 2012).

As explained in Mathieu et al. (Mathieu et al., 2013), the LiDAR data were obtained from the Carnegie Airborne Observatory (CAO). The Carnegie Airborne Observatory (CAO) Alpha system was flown over eight sites in April–May 2008. The CAO Alpha system has three integrated sub-systems: a high fidelity visible-to-near infrared imaging spectrometer (new design of CASI-1500), a waveform LiDAR (LiDAR) capable of operating simultaneously in discrete-return and waveform modes and a Global Positioning System-Inertial Measurement Unit (GPS-IMU) system which makes for an accurate registration and projection of the hyperspectral and LiDAR datasets. The woody cover was extrapolated from the two LiDAR datasets generated structural metrics. The Physical models of ground surfaces (Digital Elevation Model, DEM) and top-of-canopy surface models (CSM) were created by processing the raw LiDAR points (Mathieu et al., 2013).

The LiDAR data were acquired in April 2008 (end of wet season) when woody plants were leaf-on, and the SAR images in July-August 2008 (dry season, leaf-off) to avoid soil moisture effects on the radar signal (Mathieu et al., 2013). This was shown to be the best season to model woody cover (Mathieu et al., 2013). Woody plants of at least 1 m canopy height included, for details of the LiDAR and SAR datasets see Naidoo et al. (Naidoo et al. 2015). Validation of the SAR-map with independent LiDAR data produced an R^2= 0.8 and RMSE=7.7% (Naidoo et al., 2015). Since the MODIS data is 250 m while the LiDAR-SAR product is 12.5 m, one partial solution was to resample the LiDAR-SAR product to a larger pixel size using cubic convolution (interpolation). The cubic convolution approach assigns a weighted average of the 16 nearest cells.

3.3.5 Bucini woody cover map

The woody cover map by Bucini was provided by Scientific Services (GIS unit) of SANPark. The woody cover product is a woody vegetation map of the Kruger National Park for the year provided at 90 m spatial resolution. For this study, the map was resampled to 250 m through bilinear interpolation using the MODIS NDVI data as reference (using raster package in R). Bilinear interpolation takes a distance-weighted average of the values of the four nearest pixels (Lillesand et al. 2008). The woody cover map was produced through the calibration of remote sensing data and field measurements (Bucini et al., 2009, Bucini et al., 2010). The fusion of optical and radar data was performed to enhanced the remote sensing data. Specifically, Landsat ETM+ scenes (bands 1, 2, 3, 4, 5, 7) acquired between 2000 and 2001 and 11 JERS-1 Synthetic Aperture Radar (SAR) scenes (L-band, HH polarization) acquired between 1995 and 1996 were used. The Landsat scene were chosen for beginning of the dry season for the optical dataset to maximize discrimination of woody vegetation (still photosynthetically active) from the grass layer (dormant). The woody cover was extrapolated by multiple regression model developed between Landsat ETM+ and JERS-1 data and the field woody cover. The best predictive model was selected based on the Akaike information criterion (AIC). The variables JERS-1 backscatter and Landsat band 2 (green) being the most important variables were used to predict woody cover for KNP. The validation of the woody cover show an $R^2 = 061$, residual error = 0.89 and $P< 0.0001$(Bucini et al., 2009).

3.3.6 Precipitation

This study used precipitation observations from the weather stations of the KNP. The rainfall data from Skukuza, Pafuri, Mahlengeni and Satara weather stations for 14-year period were used. The annual rainfall values for each station were plotted against each pixel values of the second harmonic term (amplitude) as well as for the mean dry season NDVI (Chapter 4). The mean annual rainfall and corresponding pixel values for the estimated tree cover were also plotted for each station (Chapter 5).

3.3.7 GIS layer of landscape units

The GIS layer of landscape units was provided by Scientific Services (GIS unit) of SANParks where landscape types are classified based on dominant soil series, dominant woody plant species, hill slope units and landform characteristics (relief, soil, slope length, stream frequency). The landscape map is used for management planning and ecological studies in KNP. We used this map as reference to woody species distribution in KNP.

3.2 General methods

For the assessment of tree/grass phenology, ground-based cover estimation, signal decomposition (harmonic analysis) and regression analysis were used. A signal decomposition method was applied to estimate amplitude, phase, and cycles of NDVI values for all field plots extracted from MODIS data (2002 to 2015). The resultant amplitude values for all plots were used together with amplitude images (derived from the decomposition using harmonic analysis) to characterise the phenological characteristics of tree/grass phenology and to estimate fractional tree cover in KNP. A validation of tree fractional cover products and a comparison with MODIS VCF and the LiDAR-SAR fractional woody cover map were undertaken.

3.2.1 Harmonic analysis

A discrete Fourier analysis was applied to decompose the time series into harmonic terms that characterise phenology features of woody and herbaceous vegetation. With reference to satellite time series data, a given pixel can be expressed by a Fourier function as (Eq. 3.1):

$$f(t) = \overline{\boldsymbol{f(t)}} + \sum_{n=1}^{L/2} \left(A_n \cos \frac{2\pi nt}{L} - \emptyset_{\boldsymbol{n}} \right) \qquad \textbf{(3.1)}$$

such that $f(t)$ stands for the NDVI images and $\overline{f(t)}$ is the mean of $f(t)$; A_n denotes the amplitude A of the n^{th} harmonic term (number of the harmonic terms); $\emptyset_n$ represents the phase of the n^{th} harmonic term; and L is the number of observations within the study period (that is (322 observations (for each plot analysed or when the whole study area was considered) (de Beurs and Henebry, 2010).

In a satellite NDVI time series, the strength of the harmonic terms is expressed as their amplitude. The phase represents the delay of the wave relative to a standard cosine wave. The phase angle indicates the time delay of the greening cycles of PFTs. The Discrete Fourier Transform (DFT) was used (Shatkay, 1995). The sampling rate is the number of samples in the time series while the fundamental frequency is the number of time steps. DFT requires regular spacing of the samples. The maximum frequency is the content of the Nyquist frequency. Thus, DFT can separate noisy NDVI time series data into their individual cosine waves of different frequencies and filter the time series to reconstruct the complex waveform domain of the vegetation index data (Evrendilek and Gulbeyaz, 2008) as in Eq. (3.2):

$$y_{k} = \frac{1}{N} \sum_{k=0}^{N-1} c_{k} e^{-i2\pi k} / N \qquad (3.2)$$

where N is the number of samples in the time series, k is an index representing the current sample number, i is an imaginary number, and c is the k^{th} sample value (Shatkay, 1995, Moody and Johnson, 2001).

The extracted NDVI pixels data for each plot was prepared as a .txt file (322 observations for each plot) and the signal decomposition method was run in R. First, the model used here converts the satellite time series data from the time domain to the frequency domain and retrieves the spectral information in a periodogram. Schuster (1898) defined periodogram as a measure of the relative power of a time series as a function of frequency (Hernandez, 1999). For each harmonic term, the frequency, amplitude and phase angle were calculated. The harmonic analysis model here encompasses a linear detrending method to remove the gradient in the data. It then identifies the strongest harmonic terms based on their amplitudes and tests for significance using the test by Hartley (Hartley, 1949) and Fisher's F tests. An analysis of variance (ANOVA) was used to test the Null hypothesis that there are no significant harmonic terms within the time series. The alternative is that there are at least statistically significant peaks in the model. Therefore, whenever we reject the null hypothesis, we also determine the significant peaks (harmonics) that are in the models. The procedure was applied to trees or grasses data as a one-tailed test procedure. Previous estimates often leads to bias, large variance, and spectral leakage which might make one frequency spilling into the neighbouring frequencies, consequently improve higher frequency features but distort lower frequency spectra (Barbour and Parker, 2014). Two harmonic analysis methods were employed for the temporal analysis presented in chapter four (Bonferroni and multitaper methods). The test by Hartley was applied together with Bonferroni method which used to control the experiment-wise type I error at 5% for multiple testing (Köhl et al., 2006) to select the statistically significant harmonic. The multitaper approach (uses Fisher's F test) was initially developed by David J. Thomson to estimate the power spectrum of a stationary time series (Thomson, 1982). The multitaper method is one of the methods that reduces spectral bias and variance (Thomson, 1982). In this study, the multitaper package (in R) developed by Rahim et al. (Rahim et al., 2017) was used. The approach is considered as an improvement to previous harmonic analysis methods as it is designed to work with data that is nonstationary. Multitaper method applied in this study also uses 95% confidence interval to select statistically significant harmonic. The difference between the two methods is that the selection of significant harmonics by Bonferroni is through adjustment of the overall alpha level while the multitaper is performed using the tapering, which means the multiplication of the time series by the window function in the time plan based on discrete prolate spheroidal sequences (Thomson, 1982). The NDVI time series data were detrended for all harmonic analysis methods presented in chapter 4.

The extracted NDVI data for each plot was prepared for each field plot (322 observations for each plot), and the signals were decomposed. Noise is biased towards low frequencies (Allen and Smith, 1994). Red noise is being considered as a way of attaining less noise in the estimate of power spectrum. At certain threshold, some harmonics are likely to stand above the red noise spectrum. Therefore, to account for red noise, the Hartley and Fisher's tests were used to select statistically harmonic terms whose threshold is high. Most geophysical time series tend to have larger power at lower frequencies (Ghil et al., 2002). The threshold set the confidence intervals outside which the time series can be considered significantly different from a generic red noise simulation. By selecting the lower harmonics whose power is significantly different from the red noise, we can be confident to reject all other red noise processes (Allen and Smith, 1994, Sella et al., 2010).

The spatial analysis of MODIS 16 day NDVI composites comprising 322 images for the whole study area was presented in Chapter 5. The harmonic analysis presented in chapter was based on Bonferroni method and no detrending was applied in that case. The strongest harmonic term is the annual signal which can be caused by all PFTs. The second strongest harmonic term represents PFTs with bimodal characteristics, i.e. two peaks per year (Moody and Johnson, 2001). The following R packages were used:

• zoo: S3 Infrastructure for Regular and Irregular Time Series (Zeileis and Grothendieck, 2005).

• maptools: Tools for Reading and Handling Spatial Objects (Bivand and Lewin-Koh, 2013).

• raster: Geographic Data Analysis and Modelling (Hijmans and Van Etten, 2013).

• rasterVis: Visualization Methods for Raster Data (Lamigueiro et al., 2011).

3.2.2 Assessing interannual variability

Based on harmonic analysis, the interannual variability of trees and grasses were assessed with reference to amplitude, phase, and cycles using MODIS NDVI time series data (14 year). Much emphasis was placed to first and second harmonic terms. A comparison of annual rainfall data with the amplitude and the mean of the dry season NDVI images for each growing season over the 14-year period was provided. The details of the methods applied have been explained in chapter four where the analyses were carried out.

3.2.3 Tree cover estimations

Tree cover was estimated in Chapter 4 and 5. The estimates of percent tree cover presented in Chapter four were based on phase and cycles as derived from the first and second strongest harmonic terms respectively. The estimates of tree cover presented in chapter five were based on harmonic analysis (using amplitude), NDVI and SAVI vegetation indices with soil determining methods. Detail explanation of these methods has been provided in the relevant chapters.

3.2.3 Regression analysis

Regression models were used for data calibration and validations. Simple, logarithmic, polynomial and the multiple linear regression models are established between the vegetation indices and the plot data collected from the field. Research questions were carefully explained based on the chosen method in each relevant chapter where the analyses were carried out.

3.2.44 Field observation

A field campaign was carried out in March 2015 towards the end of the wet season when the photosynthetic activity of the vegetation was still high (Archibald and Scholes, 2007b). Fractional vegetation cover (FVC) of trees and grasses was estimated following the visual estimation procedure by Law et al. (Law et al., 2008). This procedure is usually applied to estimate percent canopy cover when for instance tree canopy cover is clearly less than or greater than 10 % live tree canopy cover (Figure 3.5) (Law et al., 2008, Riemann et al., 2016). A recent study compared four methods of estimating tree canopy cover (Riemann et al., 2016). These methods include: Stem-mapped canopy cover (SMCC) which is derived directly from ground inventory data using modeled relationships between tree diameter at breast height (DBH) and crown diameter for each tree species, the Field-collected percent canopy cover (FCC) derived through Ocular estimates, Photo-interpreted percent canopy cover data (PCC), (PCC collected from leaf-on, 1 m resolution, digital color-infrared National Agriculture Imagery Program (NAIP) imagery) and Geographic-Object-Based Image Analysis (GEOBIA) approach that uses both high resolution imagery and leaf-off LiDAR data. RMSE and coefficient of agreement (AC) were used to compare the accuracy of the methods. FCC (Ocular based method) show high agreement with PCC (AC: 0.73, RMSE = 16), GCC (AC = 0.7, 23%) and SMCC (AC = 0.78, RMSE = 14%). The FCC was also evaluated based on quality assurance (QA) data which is collected on 4 % of the plots nationwide of the original plots (revisit and re-inventory by a separate crew and all variables are re-measured). The results indicate better agreement (AC = 0.92, RMSE = 7.4) than what was observed in all the previous methods. Ocular estimate together with satellite data have promise for large area estimation of tree cover (Riemann et al., 2016).

In the field, tree and grass FVC was estimated visually in 25 plots along the main road from Skukuza to Tshokwane. Three additional plots were added based on the visual interpretation of Google Earth images to incorporate areas with more dominant tree cover than was found in the plots. The plots span different vegetation types, rainfall, geological conditions and soil types and cover a gradient of tree/grass mixtures that are very distinct regarding their structure, type, density and distribution. Specifically, the tree species for these plots include *Acacia gradicortuna* (e.g. plot 1, 11, 12), *Combretum (Zeyheri/apiculatum*) (e.g. plot 13, 14, 25), *Acacia nigrescens* (e.g. plot 5), *Dicchrostachys cinnerea, Scelerocarya birrea* (e.g. plot 28), *Euclea divinorum* (e.g. plot 4), *Combretum hereroense* (e.g. plot 5), *Albizia harveyi* (15) and *Terminalia sericea* (e.g. plot 3) etc. Grass species include *Aristida congesta spp., Digitaria eriantha, Erasiantha, Uracholoa mosambicensis, Themeda triandra, Panicum colouratum* etc. The cover scale adopted from Law et al., (2008) is presented on Figure. 3.5.

FVC of four structural vegetation types was estimated for each plot (trees ≥ 6m, shrubs 1-5 m, forbs and grasses). Trees and shrubs were merged into a single group in the analysis to represent overall woody cover while forbs and grasses were merged to describe overall herbaceous cover. Since, phenology is a function of species composition, time and maturity of the plant life-forms, only sites with healthy vegetation at the time of the field work were incorporated (Archibald and Scholes, 2007b, Ma et al., 2013). Figure 3.6 shows photographs of selected field plots and a map from OpenStreetMap. The field data was sampled about 50 m away from the road to avoid proximity effects (Smit and Asner, 2012). Considering the MODIS pixel size of 250 m, each plot was chosen as a 200 m x 200 m square along transect about 25 km in length. As the method adopted a rough approximation, for each plot we consider the cover scale used by Law et al. (Law et al., 2008). The sampling applied is random, but a distance of at least 1 km between plots was maintained to capture landscape variability to include samples that are representative of the ecosystem. A bounding circle (cover scale) adopted from Law et al. (2008) was used for measuring the tree/grass cover. The procedure started by establishing a plot boundary. It should be noted however that no subplot measurement was done in order to ease field data process due to nature of the study area (Riemann et al., 2016, Mairota et al., 2015). Previous studies highlighted the difficulty of relating field information to image data at a comparable scale due to product mismatch and insufficient field data if the plots are smaller than $0.1 km^2$ (Gill et al., 2009).

Figure 3. 5: Examples of percent cover (Law et al. 2008)

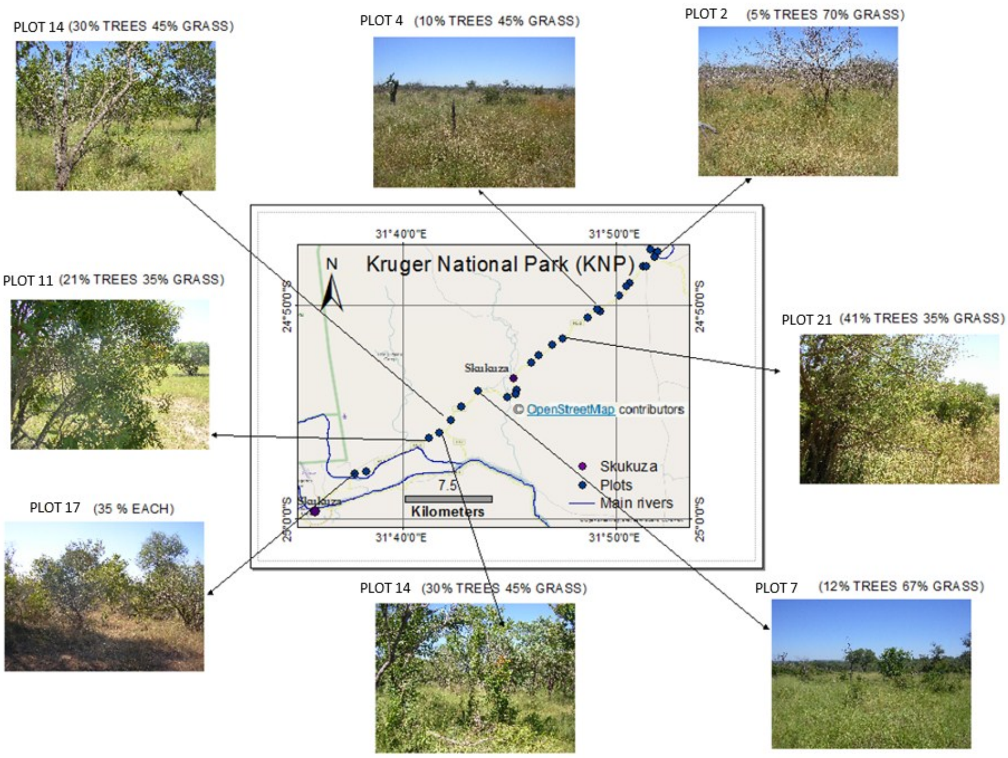


Figure 3. 6: Map of the study sites along the Skukuza/Tshokwane road, showing photographs of some selected field plots to illustrate the different tree/grass compositions.

Chapter 4

An assessment of tree/grass fractional cover using phenological signal decomposition of MODIS data

4.1 Introduction

Many previous studies have conceptualized the distribution of vegetation in savanna as a function of soil, climatic gradient and human activities (Foody, 2003, Chamaille-Jammes et al., 2006, Chamaille-Jammes and Fritz, 2009) without consideration to trees and grasses separately, even though they have different growing cycles and contribute differently to ecosystem function in savannas.

To date, the full potential of remote sensing of mixed tree/grass communities with time-series analysis has not been fully realized. PFTs mapping from time-series decomposition adds immense value to the long-term data archives of satellite imagery (Lu et al., 2003, Archibald and Scholes, 2007b). Challenges lie in the differences in tree/grass structure, physiology, phenology and seasonality. Because the signals from different PFTs are mixed within the same pixel in medium-resolution satellite images (Hill et al., 2011, Hill and Hanan, 2010a). The partitioning of tree and grass phenology could be possible since the two PFTs have different phenological cycles (Boke-Olén et al., 2016). In addition, mapping fractional tree/grass cover as a continuous variable is more applicable in African savannas; where landscapes are dominated by gradual transitions between open and closed shrub and grasslands rather than by distinct class boundaries (Gessner et al., 2008).

The use of uncertainty measures in land surface phenology is also important as they allow a statistically unusual event at a given probability level and an event within the normal range of variability to be distinguished (White and Nemani, 2006). White and Nemani (2006) presented a statistical measure of uncertainty using a confidence interval for real-time monitoring and short-time forecasting of LSP. Many previous studies that have used Fourier analysis do not assessed harmonic terms based on statistically significant harmonics (de Beurs and Henebry, 2010, Moody and Johnson, 2001). The need to incorporate more statistical techniques to estimate confidence intervals is important (Bloomfield, 2004). The aim of this study was to use satellite phenology data to assess the temporal dynamics of PFTs in the African savanna of Kruger National Park, South Africa. The harmonic analysis was applied to vegetation phenology observations of a heterogeneous savanna based on MODIS NDVI time series data. The decomposition model is tested for its ability to estimate separate tree/grass cover fractions as well as for their annual and interannual variability using plot data from a field campaign in 2015. The study seeks to answer the following questions:

1. How can satellite time series data on tree/grass signals be decomposed into statistically significant harmonic terms?
2. How does tree and grass phenology vary inter-annually?
3. How well are the satellite derived metrics estimated from the NDVI using Fourier analysis can estimate tree/grass fractions?

4. 2 Objectives

1. To decomposed MODIS NDVI data into significant strongest harmonic terms of tree and grass signals
2. To characterise inter-annual variability of the PFTs with reference to amplitude phase, cycles based on their temporal frequency domain.
3. To assess relationship between the tree and grass cover fractions as derived from the field.
4. To assess the accuracy of tree/grass cover estimates from phase and cycles using field data collected in 2015.

4.3 Methods

This section addresses how the NDVI and the derived amplitude, cycles and phase values were analyzed. The signal decomposition procedure for the estimation of statistically significant harmonic terms has been explained in chapter three.

4.3.1 Assessing changes in PFTs from NDVI time series

The changes in the MODIS NDVI time series data were assessed and compared between different field plots. The purpose was to determine the interannual variability of tree/grass fractional and investigate whether a distinction can be made among various PFTs as they are composed of varying density.

4.3.2 Identifying variability of amplitude and phase values using harmonic terms

Harmonic terms of the NDVI time-series data were used to characterize fractional tree/grass phenology changes. The analysis was first run for all field plots data to identify all statistically significant harmonic terms. The annual and interannual variability with corresponding amplitude and phase values for these plots were then assessed. However, it should be noted that two kinds of analyses were performed with the NDVI time series data. The first analysis considered the entire time series over the study period (322 observations) and this was applied to all plots. The second analysis was applied to few plots and for phenological year (23 observations for each year). The selection of the plots is based on a threshold of ≥ 35% and ≥ 10% of tree cover (TC) and grass cover (GC) respectively. Specifically, plot 25 (TC = 55%, GC = 15%), 28 (TC= 70%, GC = 20%), for dominated tree sites were used. For grass dominated sites, plot 1 (TC = 5%, GC= 85%), plot 10 (TC = 20%, G C= 70%). Others include plots where the proportion of trees and grasses are equal. These are plot 17 (TC = 35, GC = 35%) and plot 24 (TC = 45, GC = 45%). These plots cover parts of the thickets of the Sabie and Crocodile River, where trees were dominated by *Combretum/Terminalia serica* woodland, *Sclerocarya birrea*, *Euclea divinorum, Sapirostachys Africana, Acacia nigrescens* savanna and in the *Acacia welwitschii* thickets in the Karoo sediments landscape. The major species in the grass dominated plots include *Themeda triandra* and *Panicum colouratum*.

The assessments of these PFTs were made with reference to strongest harmonic terms. For example, first and second strongest harmonics are the most important as they have a clear ecological interpretation (Scharlemann et al., 2008, Moody and Johnson, 2001). The first and second strongest harmonic represent the annual and biannual signals respectively. The first strongest harmonic term can be caused by all PFTs while the second strongest is more likely to represent PFTs with bimodal characteristics (Moody and Johnson, 2001). The annual rainfall data for four stations have been compared using a bar and line plots with the corresponding pixels values of the second strongest harmonic term and the mean NDVI of the dry season for each year (2002/2015). The stations include Skukuza (in the extreme south), the Pafuri (in the extreme north), Satara (in the far south) and Mahlengeni in the north.

4.3.3 Regression analyses

Regression analysis is a statistical method of estimating the relationships among the variables. For example, a simple linear regression model can be expressed in the form:

$$y=\alpha+\beta x+\varepsilon \qquad (4.1)$$

In this model, the two variables to be related are y, the dependent variable, and x, the independent variable. The parameters, α represents the intercept and β expresses the slope of the relationship between the two variables, and an error term, ε. When there is more than one independent variable (multiple linear model), the regression model is typically expressed as:

$$y = \beta_0 + \beta_1 x_1 + \ldots\ldots \beta_n x_n + \varepsilon \qquad (4.2)$$

where β_0 is the intercept and $\beta_1 - \beta_n$ represent the slope coefficients for the independent variables $x_1 - x_n$, respectively. The relationship between tree and grass cover as obtained from the field was first assessed using simple linear regression. A simple linear regression was also used to estimates percent tree/grass cover using phase values of the first harmonic term. The number of cycles were also used to estimate the percent tree/grass cover. Both phase and cycles were used in a multiple linear regression to assess whether the synergy between the two variables would yield better accuracy in this estimation. The field data was divided into two: one half were used for calibration and the other for validation. The procedure for calibration and validation for the estimates of percent tree/grass cover are explained in the following sections. The phase values of the first harmonic term were used for tree cover estimation. For the Bonferroni method, the number of cycles of the second harmonic term were used, while the number of cycles used to estimate tree cover for multitaper method was the sum of five harmonic terms. The accuracies of tree cover estimated using Bonferroni and multitaper methods were compared.

4.3.3.1 Calibration of tree/grass cover using field data

In this Chapter, the estimates of tree/grass percent cover were made by establishing the relationship between the phase, cycles and the field data. The field data were used as the dependent variable while the phase and cycles as the independent. Only percent tree cover was estimated using the phase and cycles in the multiple regression.

4.4.3. 2 Validation of tree/grass cover using field data

The remaining half of the field data on tree/grass cover were used for model validation. To assess model performance for tree cover estimated in this study, the coefficient of determination (R^2) was used to measure the strength of the relationship between the predicted and the observed values. The predicted data for each model is taken as the independent variable while the observed as the dependent as explained in Piñeiro et al. (Piñeiro et al., 2008). In addition, the root mean square error (RMSE) was used to determine the goodness-of-fit.

4.4 Results

The time series analysis of NDVI for some selected plots is presented. The results from signal decomposition are used for the assessment of PTF dynamics based on the statistically significant signals. The first three strongest harmonic terms of the entire field plots are presented based on their amplitude and phase values. Annual and interannual variability of some tree/grass dominated plots are assessed based on their productivity represented by the amplitude. Finally, tree/grass cover were estimated and validated.

4.4.1 Time series of PFTs based on percent cover

Examples of a visual exploration of the time series analysis results based on selected field plots with different mixtures of FVC of trees and grasses are presented in Figure 4.1 (a-d). These time series plots do not clearly represent the FVC of individual PFTs at first sight except for years with an unusual event. Generally, a distinction can be made between tree and grass dominated sites regarding their minimum and maximum NDVI values. Most of the sites dominated by grasses show low and high varying signals (NDVI values 0.1-0.8), while the tree dominated plots show NDVI values around 0.3 and higher. The PFTs show interannual variability with the shape of the NDVI curve being differently between the years. Figure 4.1a and 41b distinguished between a tree and grass dominated sites. For instance, the difference between the two is more obvious in the very dry growing season of 2002/2003. Although it is challenging to differentiate the two when reference is made to the time series data, the drought year is exceptional for these PFTs. In this year, very little grass growth can be observed while tree species maintained high NDVI. Biologically, the trees are more resilient to water-constrained conditions than grasses. Trees usually have well-developed roots which enables them to tap water from the ground. They are less dependent on rainfall than the shallow-rooting grasses, which rely on short-lived rainfall (Whitecross, Witkowski, and Archibald 2017b). Trees are more resistant to fire and can recover more easily after a fire once they have grown to a certain height and escaped the 'fire trap'. In the NDVI time series data, little variation can be seen between them, although they (tree and grass) have different growing cycles. A spectral analysis through the harmonic model is applied in the next section which decomposed the time series data into harmonic cycles (cycles, amplitude and phase values).

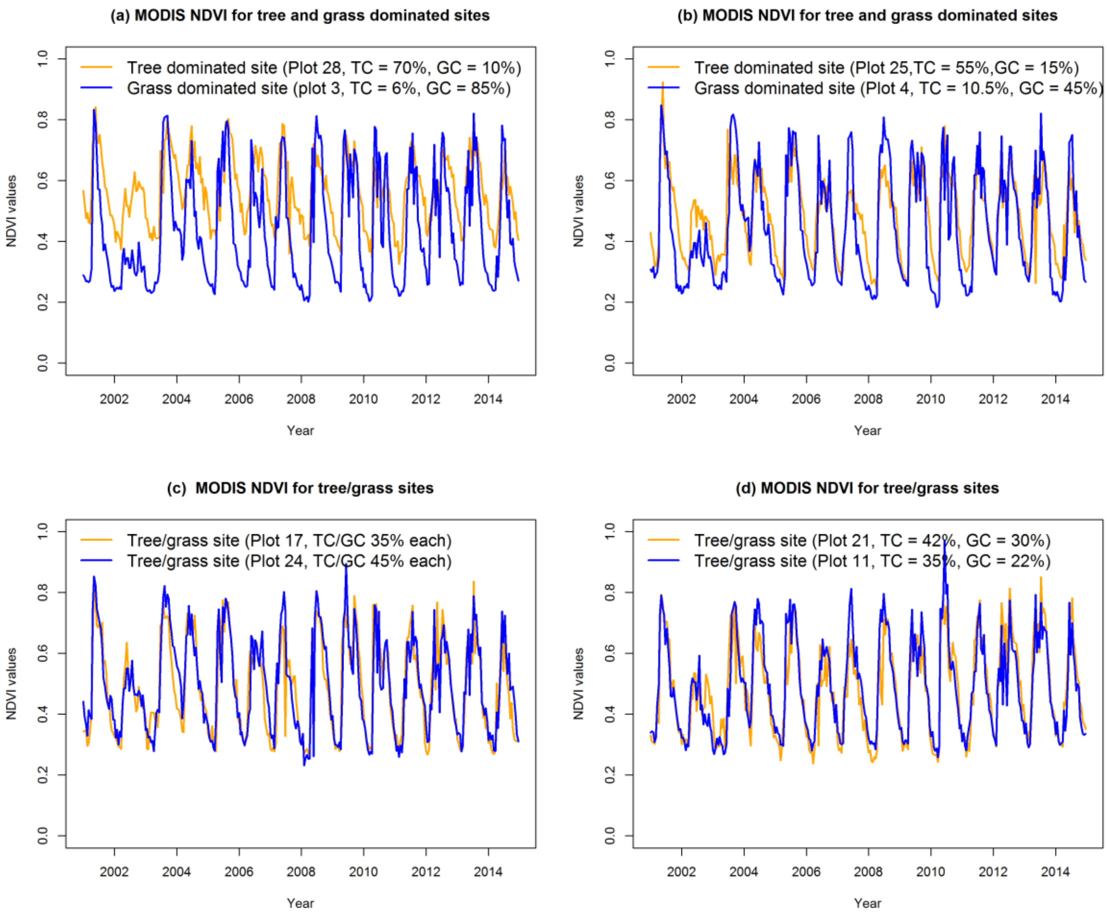


Figure 4. 1: Time series plots of selected field plots with different FVC of the tree and grass PFTs, (a) tree dominated site (plot 28) and grass dominated site (plot 3), (b) tree dominated site (plot 25) and grass dominated site (plot 4), (c) mixed tree/grass site (plot 17) and (plot 24), (c) mixed tree/grass site (plot 21) and (plot 11)

4.4.2 Signal decomposition of MODIS NDVI data

Given that NDVI for the PFTs varies temporally and cannot be easily understood as a mixed signal, numeric decomposition of these values into harmonic terms to estimate their amplitude and phase values for each cycle is applied to assess the interannual variability of PFTs and greening-up period for all field plots. The five plots, predominantly composed of trees and grasses and a mixture of the two, are presented. The signal decomposition results for the first and second strongest harmonic terms are presented for the all field plots in Table 4.3.

i. Mixed tree/grass (plot 21, TC = 41%, GC = 35%)

In Figure 4.2, the NDVI time series (a), the linear trend (b), the detrended time series (black) and the model composed of all significant harmonic terms (red) (c), all individual harmonic terms that are significant (d), and spectral density of the harmonics at different frequency (e) are presented for this field plot (plot 21). This plot shows the annual variability and changes associated with a site having a considerable tree/grass mixture. In Figure 4.2 (b), the black line indicates the trend whereas the red line shows the composite harmonic model. This harmonic model is a good fit as it reflects changes in the time series. The trend line shows that the rate of change was higher at the beginning driven probably by the drought year (2002/2003). In this plot, the early years had also low NDVI values. The spectrum plot shows the aggregate frequencies of the entire time series (Figure 4.2e). The statistically significant harmonic terms are usually characterized by increasing power at lower frequencies (Figure. 4.2 e).

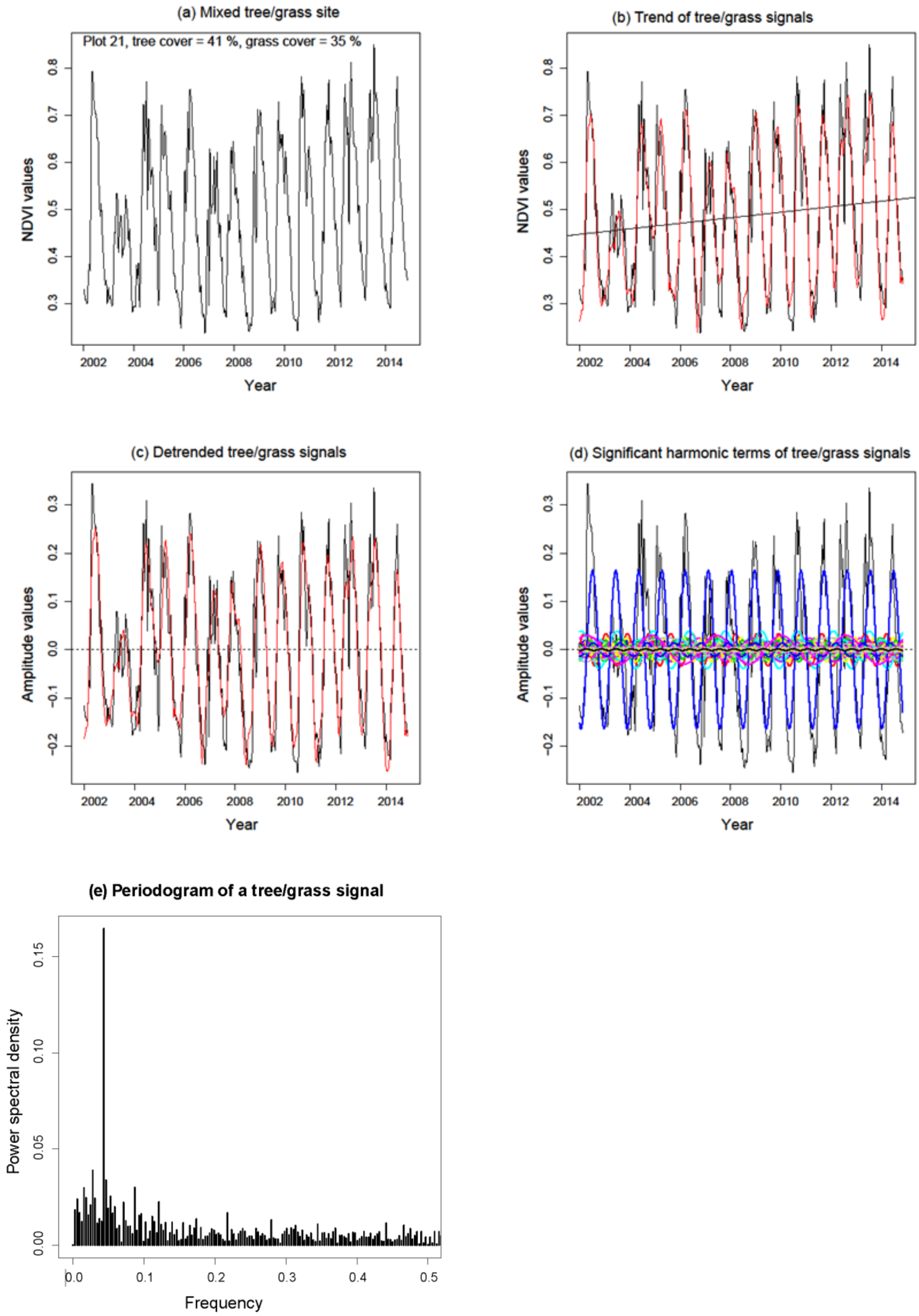


Figure 4. 2: 1 Harmonic decomposition of MODIS NDVI for a tree/grass field plot (21). (a) MODIS NDVI time series. (b) The red line indicates the composite harmonic model of all significant terms while the black line shows the trend line. (c) Detrending result, with the red line indicating the composite harmonic model. (d) Significant harmonic

terms of tree/grass signal based on the harmonic model. (e) Spectral density of the harmonics at different frequency.

ii. Grass dominated plot (plot 3, TC = 5%, GC= 85%)

The same analysis for a grass dominated field plot (plot 3) is shown in Figure 4.3a-e, with the MODIS NDVI time series (a), the linear trend (b), the detrended time series (black) and the model composed of all significant harmonic terms (red) (c), all individual harmonic terms that are significant (d), and spectral density of the harmonics at different frequency (e). The NDVI data indicated annual variability over the study period. Similarly, for most of the years, the NDVI values exceeds 0.25. Figure 4.3d presented a graphical expression of these significant terms. There is also a decreasing NDVI value at the beginning resulted probably due to manifestation effect of drought (e.g. 2002/2003).

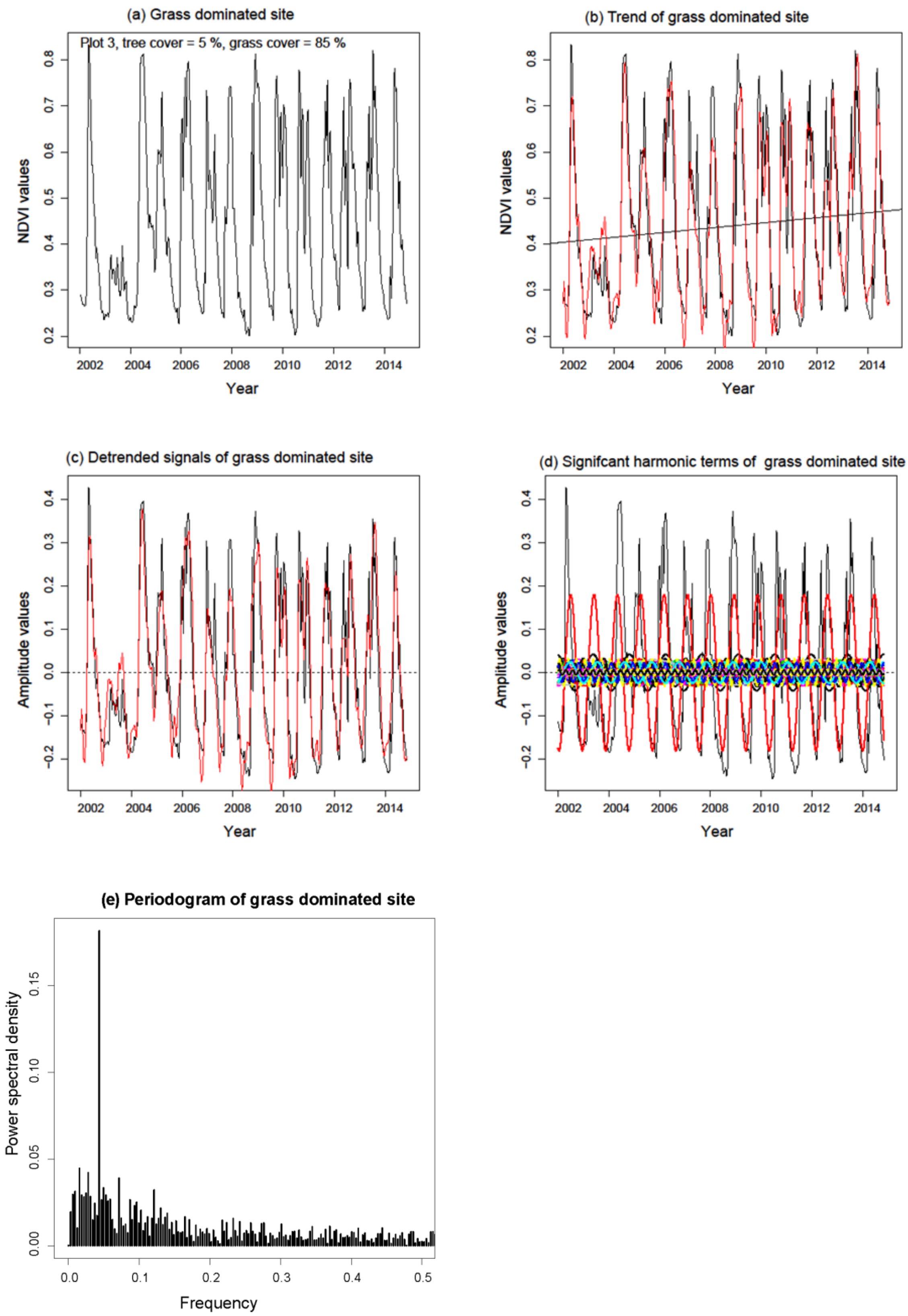
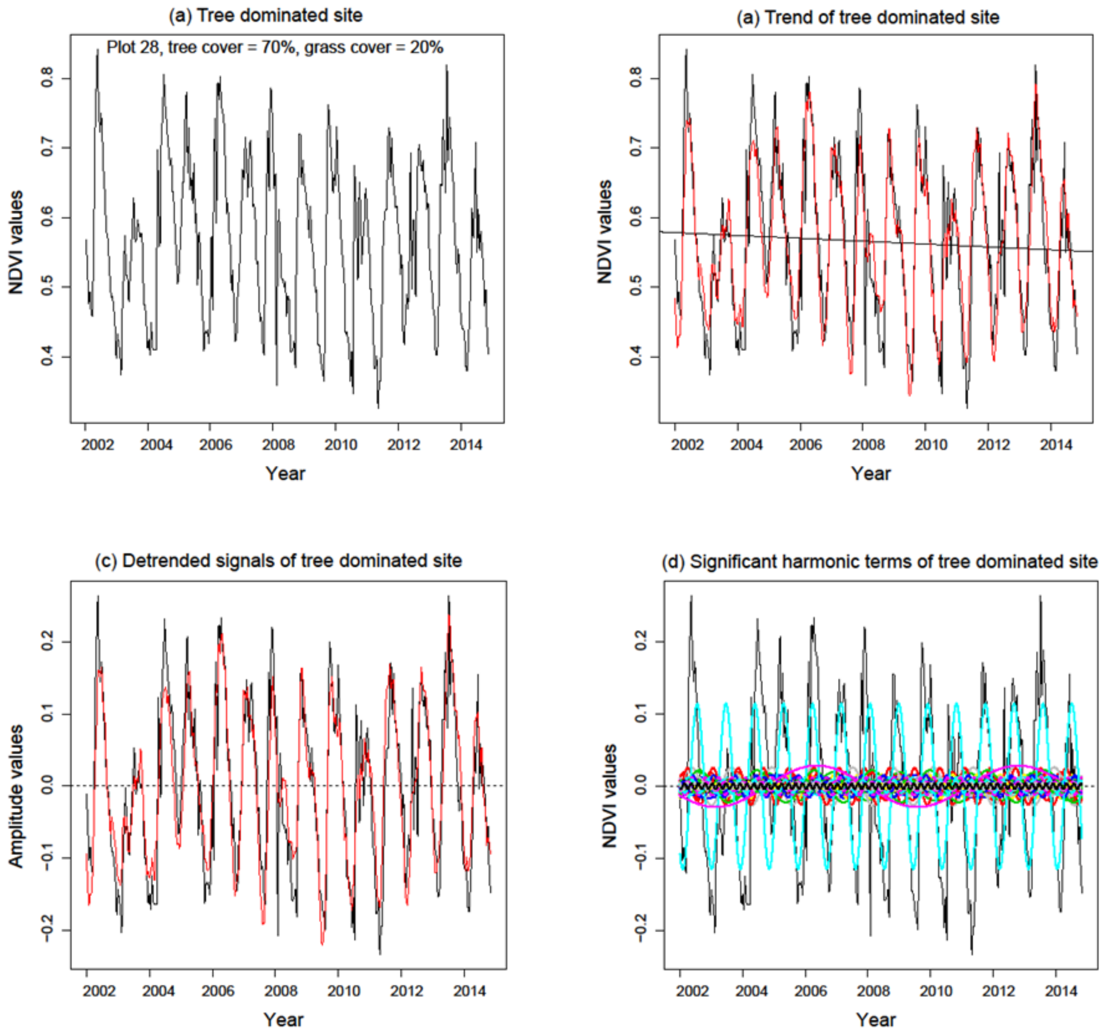


Figure 4. 3: Harmonic decomposition of MODIS NDVI for a grass dominated field plot (2). (a) MODIS NDVI time series. (b) The red line indicates the composite harmonic model of all significant terms while the black line shows the trend line. (c) Detrending result, with the red line indicating the composite harmonic model. (d) Significant

harmonic terms of tree dominated signals based on the harmonic model. (e) Spectral density of the harmonics at different frequency.

iii. Tree dominated plot (plot 28, TC = 70%, GC = 20%)

In Figure 4.4, the NDVI time series (a), the linear trend (b), the detrended time series (black) and the model composed of all significant harmonic terms (red) (c), and all individual harmonic terms that are significant (d), spectral density of the harmonics at different frequency (e) are presented for the tree dominated plot (plot 28). The NDVI data for this plot indicated the annual variability and changes associated with a predominantly tree dominated site. In Figure 4.4, the black line indicates the trend whereas the red line shows the composite harmonic model. This harmonic model is a good fit as it reflects the key changes in the time series. The trend line was relatively stable from the beginning and was higher towards the end. The strongest significant harmonic terms, based on the Bonferroni-Hartley test and their corresponding amplitude, phase values and cycles are presented in Table 4.3.



d

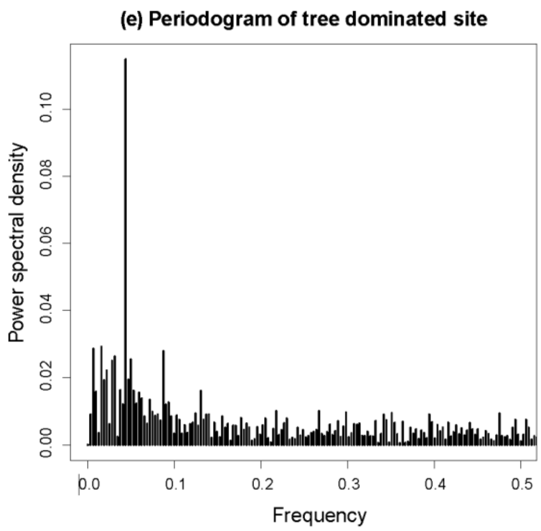


Figure 4. 4: Harmonic decomposition of MODIS NDVI for a tree dominated field plot (28). (a) MODIS NDVI time series. (b) The red line indicates the composite harmonic model of all significant terms while the black line shows the trend line. (c) Detrending result, with the red line indicating the composite harmonic model. (d) Significant

harmonic terms of tree dominated signals based on the harmonic model. (e) Spectral density of the harmonics at different frequency.

4.5.5 Comparison of strongest significant harmonic terms between tree and grass dominated plots

The amplitude values represent the productivity level of the PFTs (extent of the wave in y direction) whereas the phase angle shows the time delay of the wave term (shift along the time axis). The variability of the amplitude and phase values here only indicated the maximum productivity level of the PFTs of a given cycle.

Comparison of the strongest significant harmonic terms between tree and grass dominated field plots in Figure 4.5 was presented with reference to the ten most powerful harmonic terms (Table 4.1). The table shows distinct amplitude and phase angle values for trees and grasses. Both exhibit a large amplitude value of 14 cycles as the strongest harmonic term, which corresponds to the annual seasonality over the 14-year length of the time series. The PFTs (trees and grasses) usually attained their most active photosynthetic stage during this time. In the second strongest harmonic term, the tree phenology has 28 cycles, i.e. two cycles per year.

In contrast, the grass phenology (Table 4.1) has only 5 peaks over the 14-year period in the second strongest harmonic term. This implies that the time of the maximum amplitude of the second strongest term for grass does not follow an annual pattern or a multiple thereof. Instead, it is reached in different years. Thus, the second strongest term shows that a subtle bimodal phenological pattern was found for tree phenology, overlaying the annual cycle, while in contrast, and the grass phenology has a stronger second harmonic term that does not follow an annual pattern (cycle 5). There is no complete bimodal annual signal component found for grasses up to fourth strongest harmonic term which had only 23 cycles.

Table 4. 1: Significant harmonic terms of the main PFTs (tree and grass) selected with Hartley's ANOVA F-test at 5% ($p < 0.05$) analysed over the entire time series (14 years) to compare tree and grass dominated sites.

Strongest terms	Tree dominated site (plot 28)			Grass dominated site (plot 3)		
	Cycles	Amp	Phase	Cycles	Amp	Phase
1	14	0.115	144	14	0.182	164
2	28	0.029	-32.7	5	0.042	-21
3	10	0.028	116	9	0.040	15
4	16	0.028	64.2	23	0.036	-146
5	9	0.026	-101	16	0.035	-85.7
6	7	0.025	-112	8	0.032	-176
7	15	0.025	-14.9	17	0.031	-151
8	6	0.022	-121	39	0.030	-78.7
9	12	0.019	-52.6	7	0.029	-110
10	17	0.019	-99.6	6	0.028	-66.9

4.5.3 Inter-annual variability of selected tree/grass productivity with reference to amplitude and phase values estimated with Harley's test

The length of time series for the calculated harmonic terms in this case is shorter (based on each growing season separately). The amplitude values for the strongest harmonic term and its corresponding phase shifts for the six plots: tree dominated, grass dominated and mixed tree/grass plots (plot 28, 3, 17, 25, 2 and 24,) were presented in Table 4.2 and illustrated in Figure 4.5 a-d. Plot 28 and 25 are the tree dominated plots while plot 2 and 3 are grass dominated. Plot 24 and 17 contained a considerable mixture of both trees and grasses at almost equal proportion. The amplitude of the strongest harmonic term show changes in phenology of these PFTs over the years. The grass dominated plots have higher amplitude values with reference to this strongest harmonic term. The maximum amplitude values for grass are close to 0.29. Tree (plot 28) records maximum values at 0.18 for plot 28 while plot 25 has maximum value at 0.20. Plot 25 is dominated by the *Sapirostachys Africana* and *Euclea divinorum* tree species. The highest amplitude value recorded by the plots was in 2003/2004 phenological year which also had a high record of rainfall (Figure 4.6).

In general, these changes in tree/grass phenology could be seasonal (e.g. intra-annual changes due to differences in the PFTs life cycle), gradual (e.g. because of climate variability) and abrupt (e.g. due to fires and drought). With reference to phase values, it can be observed that the greening of PFTs is variable with species and time. There are inconsistencies in the amplitude and greening period of these PFTs over the years. Mostly, the tree dominated plot had an earlier greening than grasses. This might be connected to the fact that some tree species usually grow leaves (leaf flush) before the rainy season while most grass species depend on water availability during the rainy season. For instance, tree dominated site (Plot 28) had earlier greening period than grass throughout the study period. For the comparison of plot 25 and plot 2, the early greening period for tree dominated plots occur only in some years (Figure 4a/d, Table 4.2). Grass dominated plot appear to have an earlier greening period for six years (Figure 4a/d). The year 2014/2015 in which the field campaign was carried out show that the estimated tree dominated sites (plot 28 and 25) had earlier greening period than grass dominated sites (plot 3 and 2).

Table 4. 2: Parameters of the strongest significant harmonic term of the main PFTs (tree and grass). Plot 28, Plot 3, Plot 17, Plot 25, Plot 2 and Plot 24 analysed for each phenological year separately.

year	Plot 28 (TC = 70%, GC = 10%)		Plot 3 (TC = 6%, GC = 85%)		Plot 17 (TC = 35%, GC = 35%)		Plot 25 (TC = 55%, GC = 15%)		Plot 2 (TC = 5%, GC = 70%)		Plot 24 (TC = 45%, GC = 45%)	
2002	0.14	189	0.20	198	0.22	178	0.18	174	0.22	195	0.18	191
2003	0.08	115	0.04	156	0.11	168	0.08	148	0.06	69.5	0.08	157
2004	0.19	94.1	0.29	114	0.20	126	0.18	120	0.28	111	0.25	111
2005	0.09	142	0.12	191	0.16	169	0.12	150	0.11	189	0.11	181
2006	0.18	121	0.26	159	0.24	149	0.20	139	0.25	160	0.22	153
2007	0.11	132	0.16	150	0.16	165	0.12	146	0.16	132	0.16	137
2008	0.12	196	0.22	210	0.13	178	0.12	169	0.24	209	0.19	206
2009	0.12	153	0.28	157	0.22	157	0.17	149	0.26	147	0.24	143
2010	0.15	144	0.22	154	0.20	142	0.16	138	0.19	155	0.20	153
2011	0.10	148	0.21	163	0.19	165	0.18	155	0.17	161	0.15	155
2012	0.16	144	0.22	148	0.19	164	0.16	159	0.21	150	0.19	147
2013	0.10	142	0.16	181	0.18	181	0.13	174	0.15	183	0.13	173
2014	0.15	158	0.21	173	0.20	171	0.14	152	0.21	172	0.19	165
2015	0.10	153	0.20	168	0.16	180	0.13	166	0.19	169	0.17	161

In Figure. 4.5a-d, an example of the results derived from harmonic analysis of the inter-annual dataset indicating different harmonic terms for discriminating tree/grass phenology is presented. These are results already presented in Table 4.5. The Figure shows the variability of amplitude values between plots that are predominantly trees or grasses and those with a considerable mixture of the two. Figure 4.5a/b shows amplitude and phase values of tree and grass cover with up to 70% as well as the mixed tree/grass site consist of 35% for each PFTs. The grass dominated plot has peak values in most cycles while the mixed site follows in between the dominated tree/grass plots. The inter-annual variability in phenology is, however, site-specific.

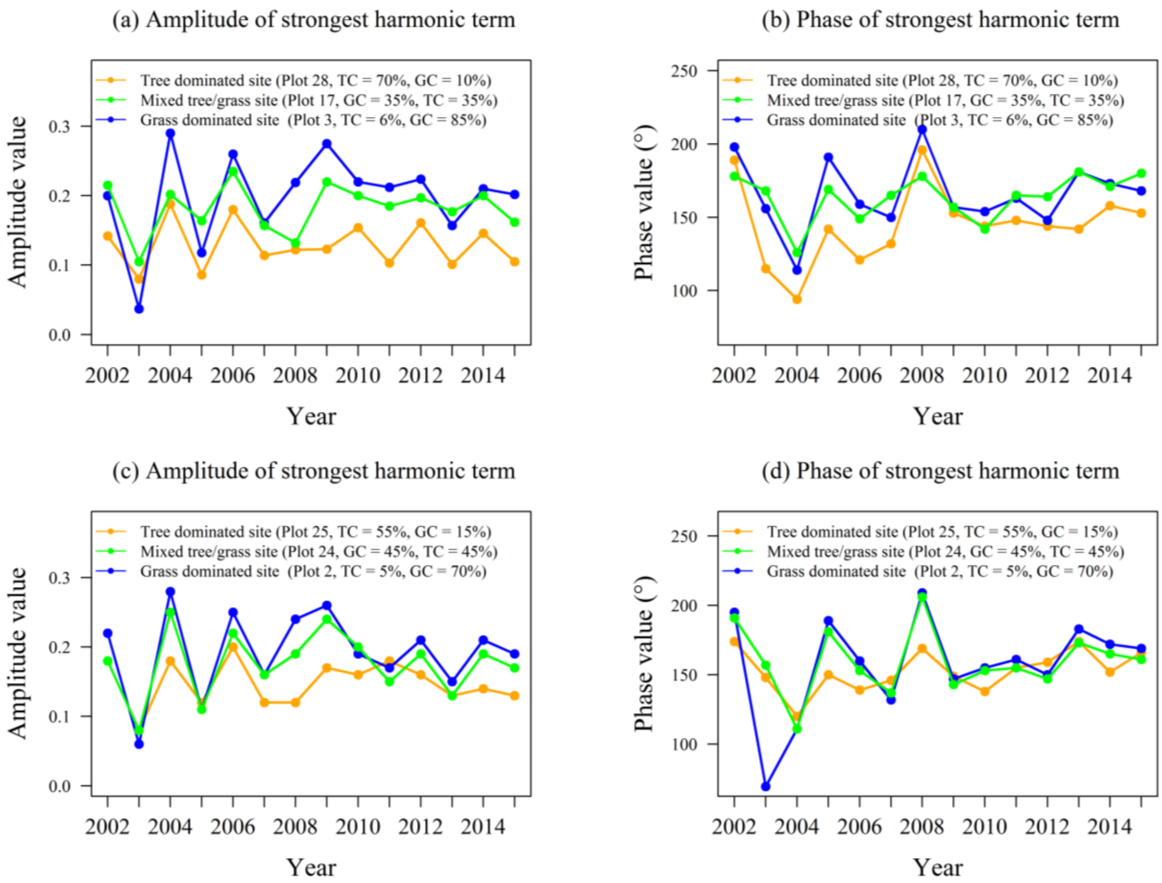


Figure 4. 5: Interannual variability of strongest harmonic terms of the decomposed tree/grass NDVI time series, estimated per year. (a) and (b) Annual amplitude of the first strongest term (c) and (d) Annual phase of the first strongest term.

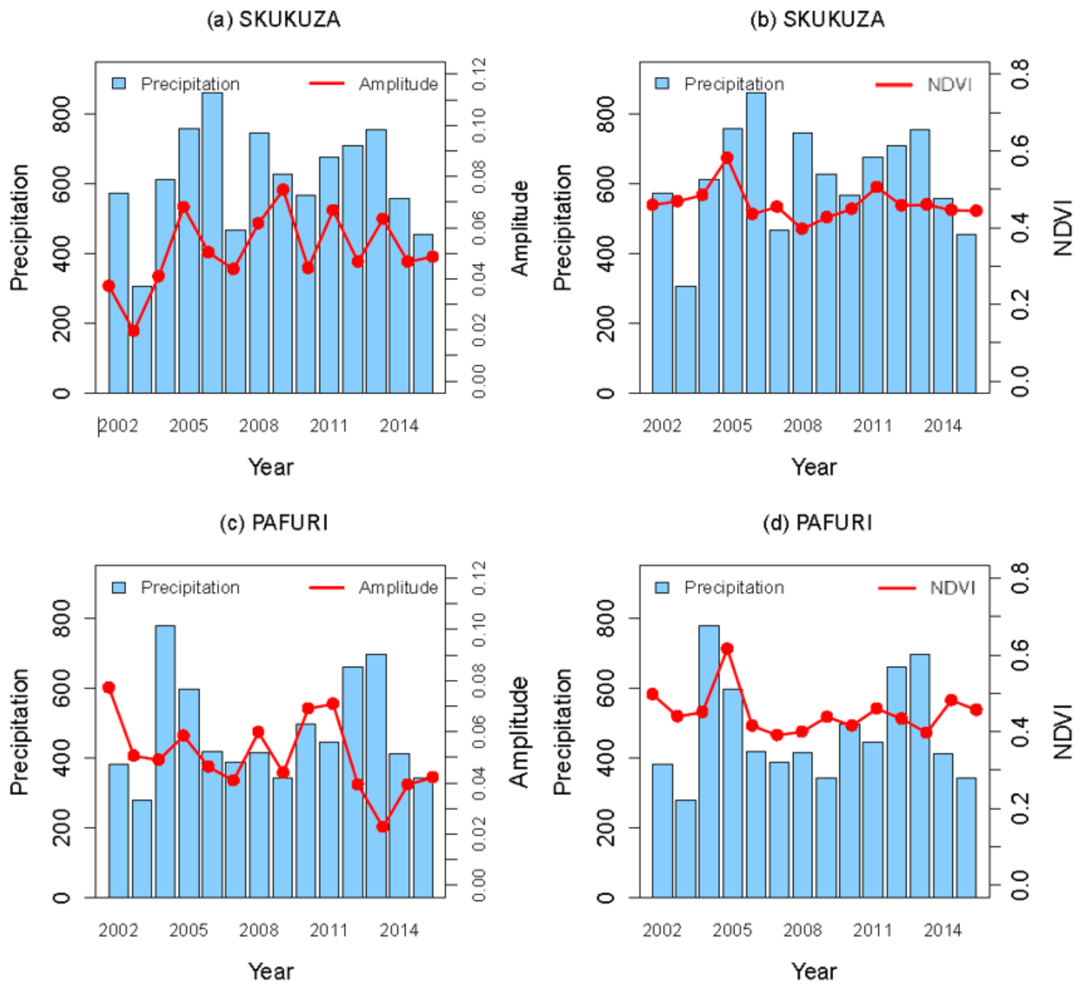
4.5.5 Statistically significant harmonics for tree and grass dominated sites assessed using Hartley-test and Bonferroni approach

Harmonic models were applied to all field plot data to assess their temporal changes using the amplitude and phase values of the first and second strongest harmonic terms (Table 4.3). The first and second harmonic terms presented here are statistically significant terms. There is, however, sites specific differences in the density of amplitude of the first strongest harmonic terms. In the second term, the amplitude values were quite lower. On the other hand, the phase values have shown a different scenario remarkably due to the dissimilarity of the PFTs (Table 4.6) because their greenness periods vary. Generally, earlier greening periods for these plots were found where tree cover is higher (e.g. plot 28 (143) and plot 27 (145) and later for grass cover (e.g. plot 1 (161) and plot 3 (164) as shown by the first harmonic term (annual). This is expected since signals from PFTs under various natural processes usually present immediate changes in their amplitude and frequency due their response to ecosystems environmental condition. For example, precipitation, fires, herbivore influence spatial and temporal variability of vegetation induce strong changes to annual and inter-annual amplitude and frequency. Figure 4.6 indicates a 14-year observed annual rainfall data for weather four stations and their corresponding amplitude values of the second strongest harmonic terms and mean NDVI. The bar plot shows the annual rainfall for each year and at each station while the red line shows amplitude or mean dry season MODIS NDVI. All phenology metrics have responded strongly to annual variation of rainfall. However, the amplitude values appear to be more sensitive to rainfall fluctuations than the dry season NDVI (Figure 4.6). With reference to a drought year (2002/2003), the differences between the two is more obvious. As the grass layer is usually non-photosynthetic in the dry season, the dry season NDVI further indicates the resistant of woody species to environmental harsh condition and less dependent on rainfall. The sensitivity of the second amplitude values might be the result of certain grass species being captured by this harmonic term.

Table 4. 3: Amplitude and phase values of the first and second strongest harmonic terms of field data selected with Hartley's ANOVA F-test at 5% using Bonferroni ($p < 0.05$) analysed over the entire time series (14 year) for all plots.

Plot	**TC (%)**	**GC (%)**	**Longitude**	**Latitude**	**Amplitude values**		**Cycles**		**Phase values**	
					1st	2nd	1st	2nd	1st	2nd
1	5	85	31.8632	-24.7952	0.166	0.035	14	9	161	13.3
2	5	70	31.8658	-24.7909	0.177	0.047	14	9	162	10.3
3	6	85	31.8568	-24.8024	0.182	0.042	14	5	164	-21.3
4	10.5	45	31.854	-24.8024	0.188	0.047	14	5	162	-14.3
5	11	85	31.7916	-24.8591	0.19	0.041	14	9	158	0.75
6	11	42	31.7483	-24.9045	0.172	0.032	14	9	159	-11.3
7	12	67	31.7832	-24.8642	0.184	0.033	14	17	158	173
8	12	70	31.8411	-24.8182	0.177	0.0408	14	9	161	8.95
9	17	78	31.7666	-24.8781	0.182	0.0365	14	23	159	-142
10	20	70	31.7544	-24.9025	0.179	0.028	14	15	159	-43.4
11	22	35	31.7256	-24.9003	0.171	0.039	14	9	163	163
12	30	45	31.7123	-24.9122	0.199	0.037	14	15	160	-51.6
13	30	40	31.8185	-24.8366	0.181	0.044	14	9	158	-18.3
14	30	45	31.7558	-24.899	0.177	0.029	14	15	158	-56.9
15	31	55	31.8433	-24.8155	0.174	0.042	14	9	161	4.22
16	32	35	31.8111	-24.8427	0.174	0.038	14	9	156	-16.9
17	35	35	31.6289	-24.9643	0.176	0.037	14	28	163	72.5
18	35	50	31.8206	-24.8376	0.181	0.044	14	9	157	-20.3
19	35	22	31.6948	-24.9326	0.184	0.033	14	28	161	66.6

20	35	57	31.7726	-24.8721	0.186	0.032	14	23	161	-148
21	41	35	31.8354	-24.8258	0.165	0.036	14	23	157	-15.4
22	42	30	31.7035	-24.9229	0.193	0.033	14	23	163	-160
23	45	50	31.6866	-24.9367	0.19	0.032	14	2	161	-157
24	45	45	31.8596	-24.7895	0.165	0.039	14	9	157	10.4
25	55	15	31.6374	-24.9625	0.146	0.036	14	28	152	67.6
26	65	10	31.7873	-24.7475	0.157	0.033	14	16	152	-110
27	69	20	31.8074	-24.762	0.151	0.036	14	28	145	74.3
28	70	10	31.7487	-24.6234	0.116	0.028	14	28	143	59.7



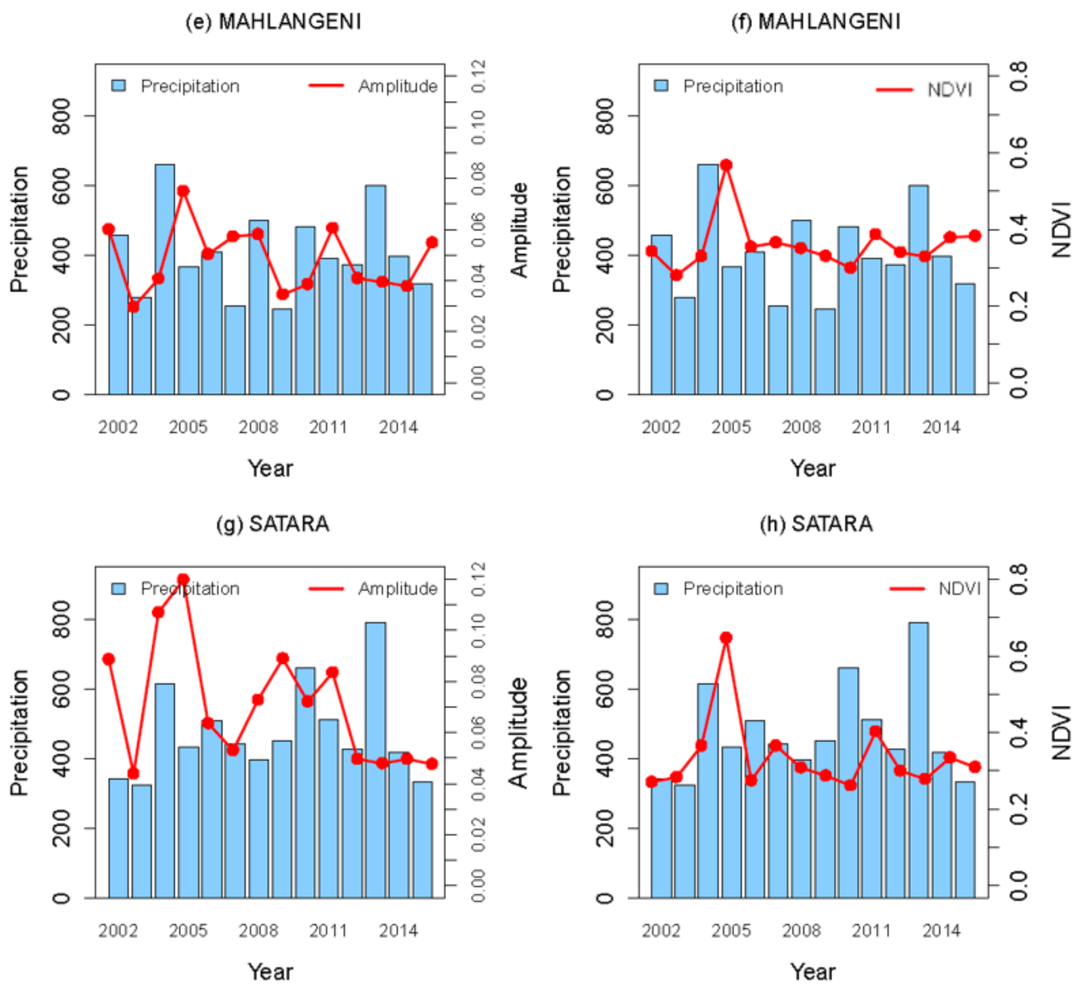


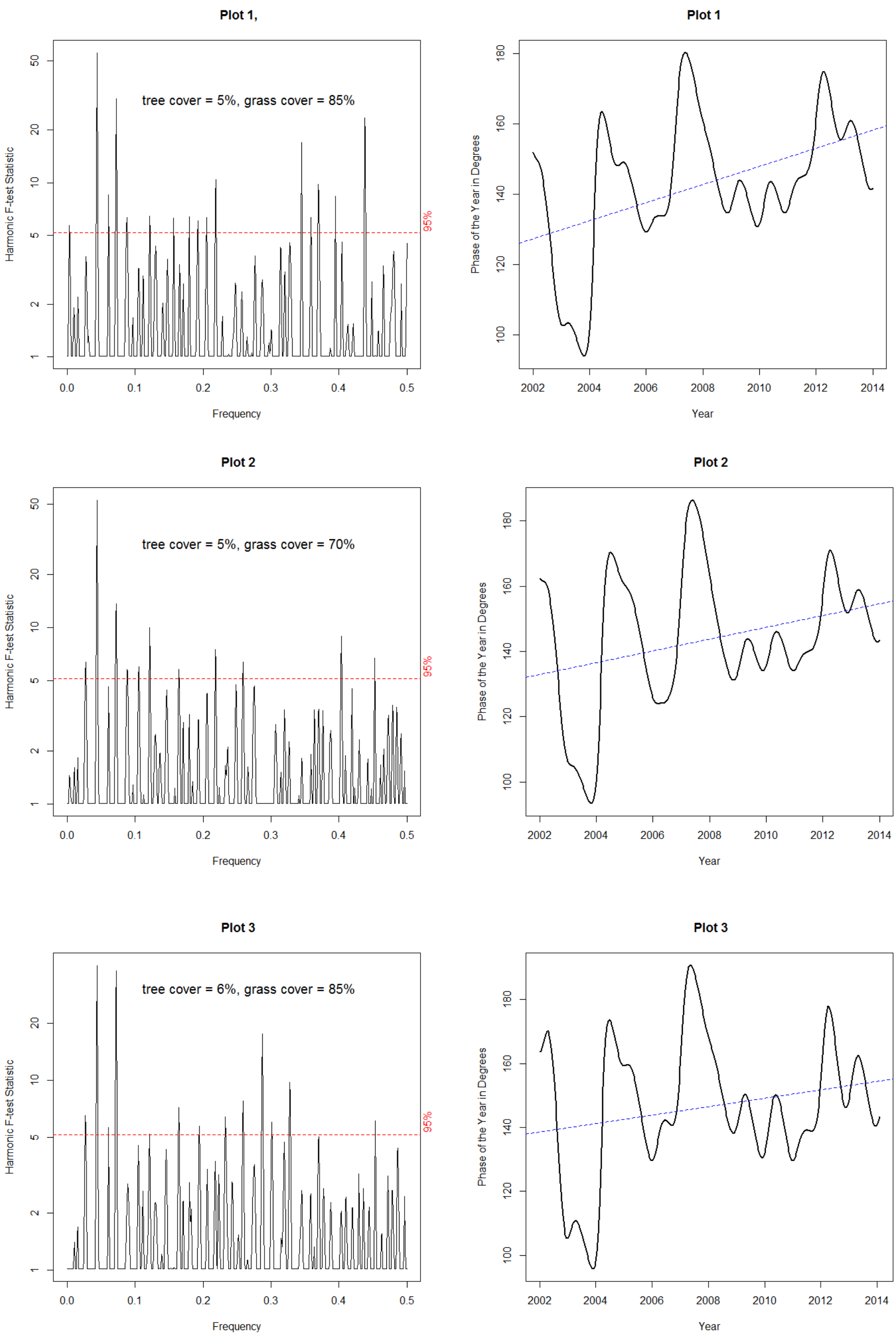
Figure 4. 6: Annual precipitation data from weather their corresponding phenology metrics from the MODIS data (2002-2015), showing Skukuza with amplitude (a) and mean dry season NDVI (b), Pafuri with amplitude (c), and mean dry season NDVI (d), Mahlengeni with amplitude (e) and mean dry season NDVI (f), SATARA with amplitude (g), and mean dry season NDVI (h)

4.5.6 Statistically significant peaks for tree and grass dominated sites assessed using F-test through multitaper method.

Figure 4.7 illustrates the periodogram of the field plot data (the tree and grass dominated sites) extracted from MODIS NDVI time series from 2001-2015 (June to July) showing significant peaks assessed using statistical confidence level of the power spectrum calculated by the F-test (at 95%) through multi-taper method. Similar to Bonferroni approach, the NDVI time series data were also detrended in this method. The multi-taper estimator detects peaks in the lower, middle and the higher frequency. The pattern of the cycles as derived from the multi-taper do not appear to be consistent for all sites because the phenomena itself is nonstationary dynamic. However, although, trees and grasses have distinct growing cycles, the tree/grass phenology could not be distinguished by cycles using the tapers. The Bonferroni approach presented earlier distinguished these PFTs as presented earlier. The multiplication of the time series in multi-taper approach usually reshaped the original time series by the window functions. The number of tapers as well as time band parameter being applied to the original time series to decrease the dynamic range of a dataset may change the distribution of the power spectrum.

It should be noted that the multitaper package (in R) used in this study offer possibility to computes complex demodulate of a given series around a given central frequency which returns amplitude and phase values based on certain parameters (Rahim et al., 2017). As explained previously in the method section, the amplitude and phase values (Figure 4.7 and Table 4.4) were estimated from the multitaper package based on time-bandwidth of 2 and 46 number of the length of sub-block (to be used in the time series) and the step-size of 1. Stepsize is a proposed option that sets the index step size between blocks. The number of cycles for the five harmonic terms, the average amplitude values and phase values of the 14 year MODIS NDVI data are presented in Table 4.4. The estimated phase values are presented (Figure 4.7) in their nonstationary form as computed from the multitaper techniques. For tree/grass cover estimate, the phase values were averaged. The amplitude values are lower for most tree dominated plots because only the peak of the annual cycle was considered (Figure 4.7). The amplitude values are therefore like the estimate from the Bonferroni approach presented earlier. The behavior of tree/grass phenology as assessed using Bonferroni with tree phenology having earlier greening period is consistent with multitaper method. Figure 4.7 shows tree dominated plots with a range of phase values from 100 to 140 while grass dominated sites mostly occur from

130 to 160 over a 14-year period. However, the two approaches also differ in phase estimation. The multitaper phase estimation follows a nonstationary dynamic of the original time series.



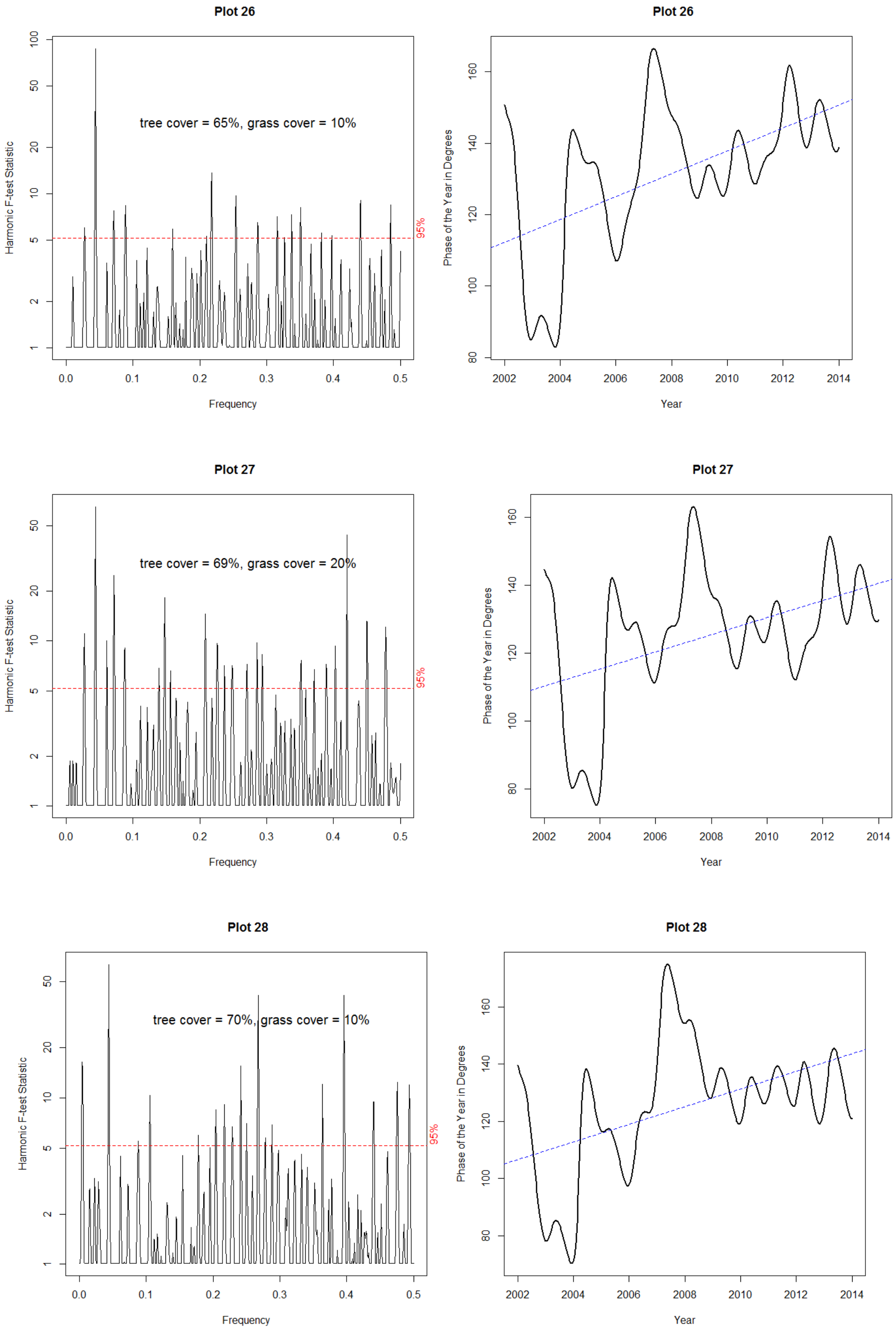


Figure 4. 7: Statistically significant peaks of tree and grass dominated sites assessed calculated by the F-test (at 95%) through multi-taper method: the red-dash lines indicate the boundary of the statistical confidence level of the power spectrum

Table 4. 4: Amplitude and phase values of the statistically significant harmonic term of field data selected with Hartley's ANOVA F-test at 5% using multi-taper method ($p < 0.05$) analysed for the entire time series (14 years) and for all plots.

Plot	**TC (%)**	**GC (%)**	**Longitude**	**Latitude**	**Amplitude values**	**Cycles**	**Phase values**
1	5	85	31.8632	-24.7952	0.18	16	143.45
2	5	70	31.8658	-24.7909	0.19	11	144.26
3	6	85	31.8568	-24.8024	0.20	12	147.00
4	10.5	45	31.854	-24.8024	0.20	14	144.60
5	11	85	31.7916	-24.8591	0.21	14	140.61
6	11	42	31.7483	-24.9045	0.18	14	143.15
7	12	67	31.7832	-24.8642	0.20	10	140.18
8	12	70	31.8411	-24.8182	0.19	12	142.30
9	17	78	31.7666	-24.8781	0.20	16	140.36
10	20	70	31.7544	-24.9025	0.19	20	141.86
11	22	35	31.7256	-24.9003	0.18	15	145.01
12	30	45	31.7123	-24.9122	0.21	16	141.50
13	30	40	31.8185	-24.8366	0.19	17	140.71
14	30	45	31.7558	-24.899	0.19	19	139.49
15	31	55	31.8433	-24.8155	0.18	16	141.75
16	32	35	31.8111	-24.8427	0.18	15	138.33
17	35	35	31.6289	-24.9643	0.18	8	144.48
18	35	50	31.8206	-24.8376	0.19	20	139.03
19	35	22	31.6948	-24.9326	0.20	16	143.60
20	35	57	31.7726	-24.8721	0.20	15	141.32
21	41	35	31.8354	-24.8258	0.17	17	137.63
22	42	30	31.7035	-24.9229	0.21	9	145.89
23	45	50	31.6866	-24.9367	0.20	27	143.93
24	45	45	31.8596	-24.7895	0.17	13	139.91
25	55	15	31.6374	-24.9625	0.15	11	134.76
26	65	10	31.7873	-24.7475	0.17	16	132.77
27	69	20	31.8074	-24.762	0.16	22	126.25
28	70	10	31.7487	-24.6234	0.12	18	127.07

4.5.7 Inter-annual variability of selected tree/grass productivity regarding amplitude and phase values estimated using Fisher's test (with multitaper)

Similar harmonic analysis of selected tree/grass plots demonstrated earlier using the Harley's ANOVA F test (with Bonferroni) is presented for Fisher's F test (with multitaper method) to assess tree/grass interannual variability. The amplitude values for the strongest harmonic term and its corresponding phase shifts for the six plots: tree dominated, grass dominated and mixed tree/grass plots (plot 28, 3, 17, 25, 2 and 24,) were presented in Figure 4.5 a-d. Plot 28 and 25 are the tree dominated plots while plot 2 and 3 are grass dominated. Plot 24 and 17 contained a considerable mixture of both trees and grasses. Unlike the statistically significant harmonic selected with Bonferroni using the Harley's ANOVA F test, the multitaper method allow to estimate amplitude in the nonstationary form of the original time series such that there are varying amplitude values throughout the year. The amplitude of the harmonic term show changes in phenology of these PFTs over the years. The maximum amplitude values for grass are close to 0.27 (plot 3). The maximum amplitude values recorded for tree vegetation 0.20 (plot 25). Plot 25 is dominated by the *Sapirostachys Africana* and *Euclea divinorum* tree species. Furthermore, the amplitude values estimated here are like the estimate using Bonferroni approach. With reference to phase values, it can be observed that the greening of PFTs is variable with tree species having early greening period than grasses. There are however few years with grass species having early greening period than trees. For example, in the 2009/2010 (Figure 4. 8 a) and 2003/2004 and 2011/12 (Figure 4.8 a).

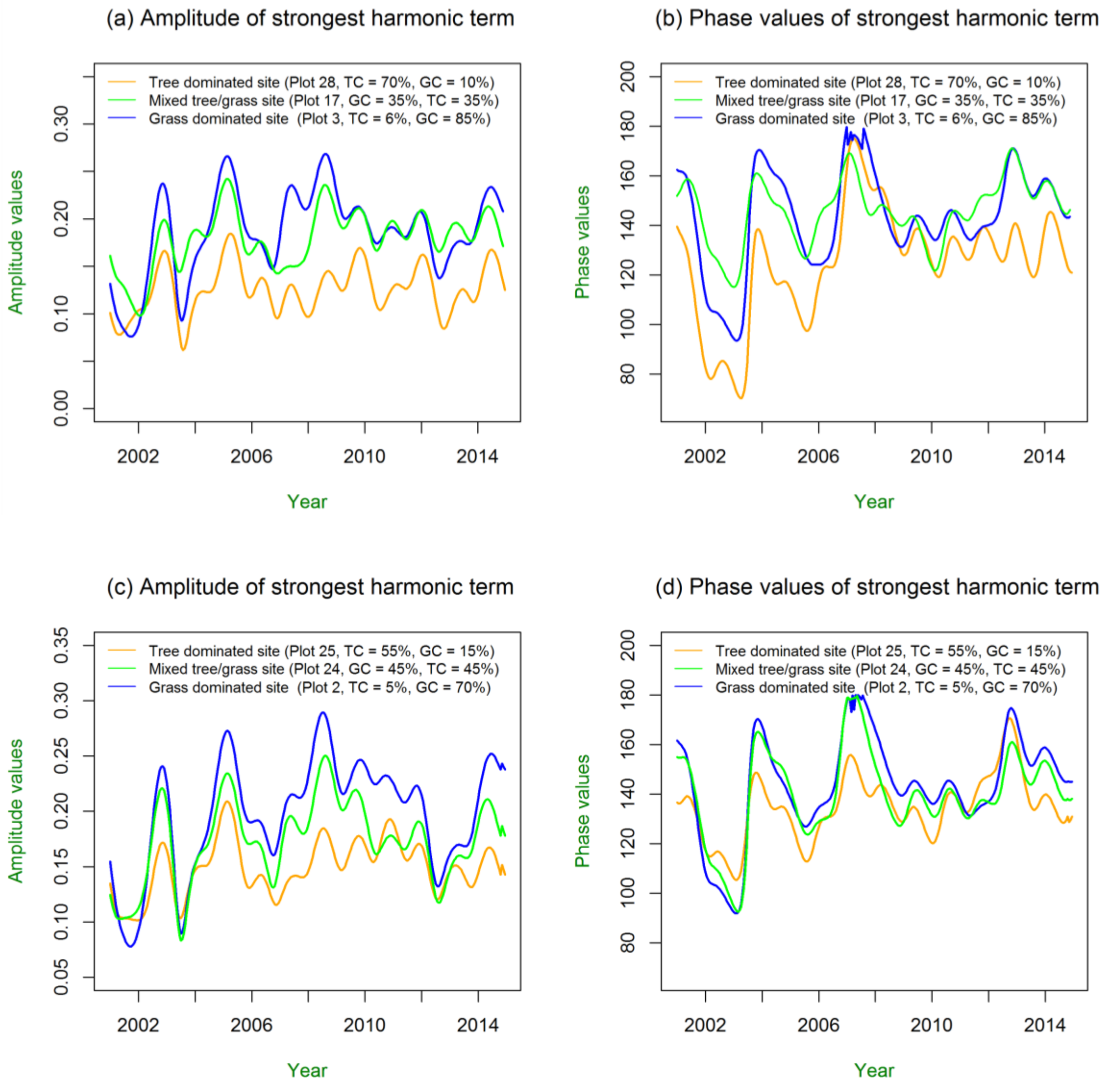


Figure 4.8: Interannual variability of strongest harmonic terms of the decomposed tree/grass NDVI time series, estimated per year (a) and (b) Annual amplitude of the first strongest term (c) and (d) Annual phase of the first strongest term.

4.5.8 Comparison of number of cycles for the first three harmonic terms estimated using Harley's ANOVA F test (with Bonferroni) and Fisher's F test (with multitaper method).

A comparison of the number of cycles estimated from the first three significant harmonic terms for tree/grass field plots (28 field plot data) using the Harley's ANOVA F test (with Bonferroni) and Fisher's F test (with multitaper method) is presented in Table 4.5. The table shows distinct number of cycles for trees and grasses with Harley's ANOVA F test and correction for multiple testing with Bonferroni. There are 14 cycles in the strongest harmonic term, which corresponds to the annual seasonality over the 14-year length of the time series. In the second strongest harmonic term, the tree phenology has 28 cycles, i.e. two cycles per year. The number of cycles estimated with Harley's ANOVA F test in the first and second harmonic terms, appear to be consistent with biological characteristics of tree vegetation in savannas.

In contrast, the power spectrum estimated by the Thomson multitaper method using the statistical confidence as evaluated by the F-test do not presents tree and grass phenology as two distinct PFTs. In Table 4.5, the 28 field plots have a range of number of cycles from 2-6 and 0-7 in the first and second harmonic terms respectively. Therefore, the cycles have less physical meaning as they do not follow an annual pattern of the tree/grass phenology. The number of cycles do not also differentiate between trees and grasses as was the case for the Harley's ANOVA F test which indicate a subtle bimodal phenological pattern for tree phenology,

Table 4. 5 : Number of cycles of the first three harmonic terms estimated using Harley's ANOVA F test (with Bonferroni) and Fisher's F test (with multitaper method).

Plot	TC (%)	GC (%)	Longitude	Latitude	Hartley's test (with Bonferroni)			Fisher's (with multitaper)		
					1st	2nd	3rd	1st	2nd	3rd
1	5	85	31.8632	-24.7952	14	9	5	6	5	0
2	5	70	31.8658	-24.7909	14	9	5	6	1	1
3	6	85	31.8568	-24.8024	14	5	9	4	3	4
4	10.5	45	31.854	-24.8024	14	5	9	3	5	3
5	11	85	31.7916	-24.8591	14	9	5	4	5	1
6	11	42	31.7483	-24.9045	14	9	23	4	3	3
7	12	67	31.7832	-24.8642	14	17	5	5	3	0
8	12	70	31.8411	-24.8182	14	9	5	5	2	2
9	17	78	31.7666	-24.8781	14	23	9	4	0	3
10	20	70	31.7544	-24.9025	14	15	23	3	4	3
11	22	35	31.7256	-24.9003	14	9	23	2	3	3
12	30	45	31.7123	-24.9122	14	15	15	5	2	2
13	30	40	31.8185	-24.8366	14	9	5	5	6	2
14	30	45	31.7558	-24.899	14	15	23	3	4	3
15	31	55	31.8433	-24.8155	14	9	5	5	3	2
16	32	35	31.8111	-24.8427	14	9	15	4	3	3
17	35	35	31.6289	-24.9643	14	28	2	4	1	0
18	35	50	31.8206	-24.8376	14	9	15	5	3	3
19	35	22	31.6948	-24.9326	14	28	9	5	3	2
20	35	57	31.7726	-24.8721	14	23	9	5	2	3
21	41	35	31.8354	-24.8258	14	23	15	6	4	2
22	42	30	31.7035	-24.9229	14	23	9	6	0	2

23	45	50	31.6866	-24.9367	14	2	9	4	7	6
24	45	45	31.8596	-24.7895	14	9	5	7	2	2
25	55	15	31.6374	-24.9625	14	28	10	4	3	1
26	65	10	31.7873	-24.7475	14	16	9	4	4	4
27	69	20	31.8074	-24.762	14	28	5	5	7	3
28	70	10	31.7487	-24.6234	14	28	2	4	2	7

4. 5.6 Relationship between field data on fractional cover of trees and grasses

A relationship between the field estimate of tree and grass cover is presented in Figure. 4.9a/b. The relationship indicated a strong negative correlation between tree and grass cover (R^2=0.69 (P<0.01) and R^2=0.71 (P<0.01)). Grass cover decreased linearly with increasing tree cover and vice-versa. In general, there are more plots with a high proportion of grasses than trees.

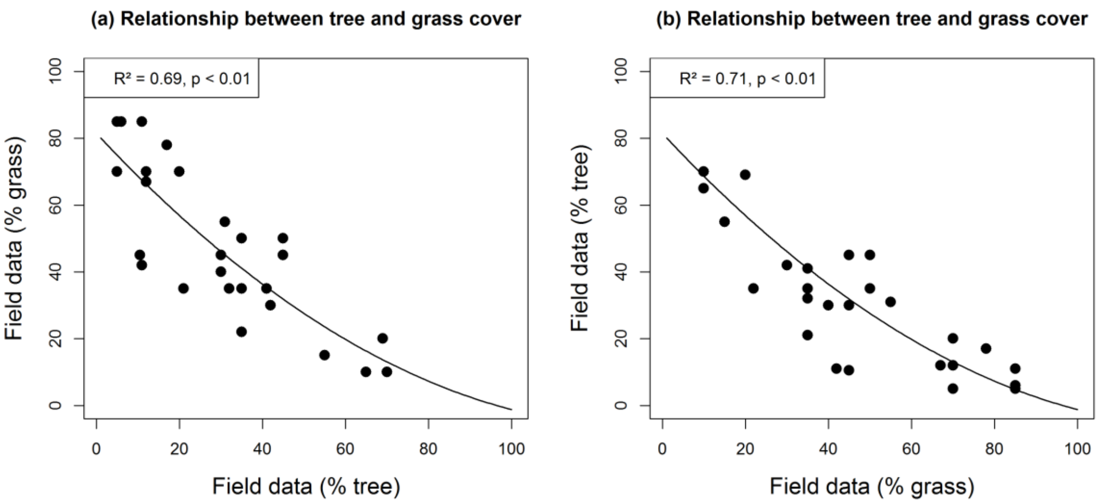


Figure 4. 9 Relationship between field data on fractional cover of trees and grasses (a) tree cover vs grass cover (b) grass cover vs tree cover

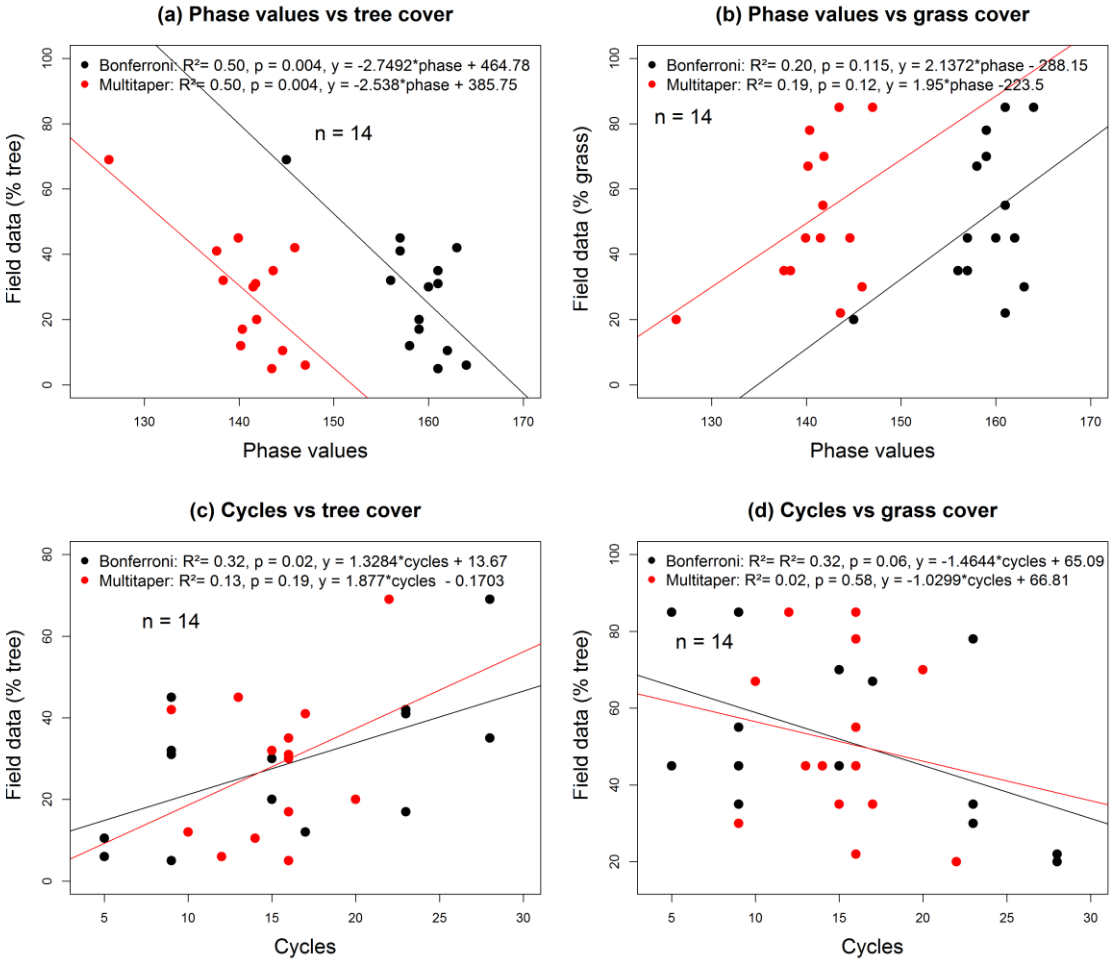
4.5.7 Model calibration using simple regression between the phase, cycles vs percent tree/grass cover (field data)

Table 4.6 indicates field plots, phase and cycles used for data calibration. The Figure 4.10 indicated the relationship between phase values as well the cycles (of the second harmonic terms for Bonferroni method) derived from the MODIS NDVI data over KNP (2002-2015) as calculated with the F-test using Bonferroni and multitaper methods. The phase values of the first strongest harmonic term had strong linear relationship with percent tree cover, R^2 = 0.50, p <0.01 and R^2 = 0.50, p <0.01 for Bonferroni and multitaper method respectively (Figure 4.10a). There is no strong relationship between the phase and percent grass cover for both methods (Bonferroni: $R^2 = 0.20$, $p = 0.11$, multitaper: $R^2 = 0.19$, $p = 0.12$, Figure 4.10b). There is a moderate relationship between the cycles and percent tree cover ($R^2 = 0.32$, $p = 0.02$) for Bonferroni but an insignificant relationship ($R^2 = 0.13$, $p = 0.19$) was observed with cycles derived from the multitaper method (Figure 4.10c). Again, the relationship between the cycles and the percent grass cover is weak in all methods (Bonferroni: $R^2 = 0.32$, $p = 0.06$ and Bonferroni: $R^2 = 0.02$, $p = 0.58$). The reason for the weak relationship of grass cover with cycles is not unconnected to the grass layer being more susceptible to temporal changes (e.g. drought) as previously demonstrated using the time series data. Overall, the positive correlation between the phase and percent tree or grass cover implies that it could be used as surrogates to percent cover in areas that are not well known.

Tree cover at less than 50% show an inconsistent greening and very unstable developmental growth (Table 4.5). Contrary to what was presented above for Figure 10a/c, Figure 10e/h show calibration results for phase and cycles when plot 27 (outlier) is removed. All models (for phase and cycles) have very weak relationship linear relationships with an observed tree/grass cover when plot 27 is removed (Figure 10e/h). Although the variability in the distribution of phase values with reference to plots used in this relationship is not very high, plot 27 has an important contribution in establishing this relationship being ecologically different compared to other plots. Only few plots had tree cover above 40%. Plot 27 which has 69% tree cover had the earliest greening period. The insufficient number of plots especially at high tree cover may constraint these models in the estimates of tree/grass cover overlarge area even where plot 27 was used. Therefore, only relationships that used plot 27 where used for tree/grass estimations and validation.

Table 4. 6: calibration of phase (1st harmonic) and cycle values (2nd harmonic) to estimates tree/grass cover.

Plot	**Tree cover (%)**	**Grass cover (%)**	**Longitude**	**Latitude**	**Bonferroni**		**Multitaper**	
					Cycles	Phase values	Cycles	Phase values
					2nd	1st	All peaks	1st
1	5	85	31.8632	-24.7952	9	161	16	143.45
3	6	85	31.8568	-24.8024	5	164	12	147
4	10.5	45	31.854	-24.8024	5	162	14	144.6
7	12	67	31.7832	-24.8642	17	158	10	140.18
9	17	78	31.7666	-24.8781	23	159	16	140.36
10	20	70	31.7544	-24.9025	15	159	20	141.86
12	30	45	31.7123	-24.9122	15	160	16	141.5
15	31	55	31.8433	-24.8155	9	161	16	141.75
16	32	35	31.8111	-24.8427	9	156	15	138.33
19	35	22	31.6948	-24.9326	28	161	16	143.6
21	41	35	31.8354	-24.8258	23	157	17	137.63
22	42	30	31.7035	-24.9229	23	163	9	145.89
24	45	45	31.8596	-24.7895	9	157	13	139.91
27	69	20	31.8074	-24.762	28	145	22	126.25


(a) Phase values vs tree cover
Bonferroni: R²= 0.50, p = 0.004, y = -2.7492*phase + 464.78
Multitaper: R²= 0.50, p = 0.004, y = -2.538*phase + 385.75
n = 14
Field data (% tree)
Phase values
(b) Phase values vs grass cover
Bonferroni: R²= 0.20, p = 0.115, y = 2.1372*phase - 288.15
Multitaper: R²= 0.19, p = 0.12, y = 1.95*phase -223.5
n = 14
Field data (% grass)
Phase values
(c) Cycles vs tree cover
Bonferroni: R²= 0.32, p = 0.02, y = 1.3284*cycles + 13.67
Multitaper: R²= 0.13, p = 0.19, y = 1.877*cycles - 0.1703
n = 14
Field data (% tree)
Cycles
(d) Cycles vs grass cover
Bonferroni: R²= R²= 0.32, p = 0.06, y = -1.4644*cycles + 65.09
Multitaper: R²= 0.02, p = 0.58, y = -1.0299*cycles + 66.81
n = 14
Field data (% tree)
Cycles

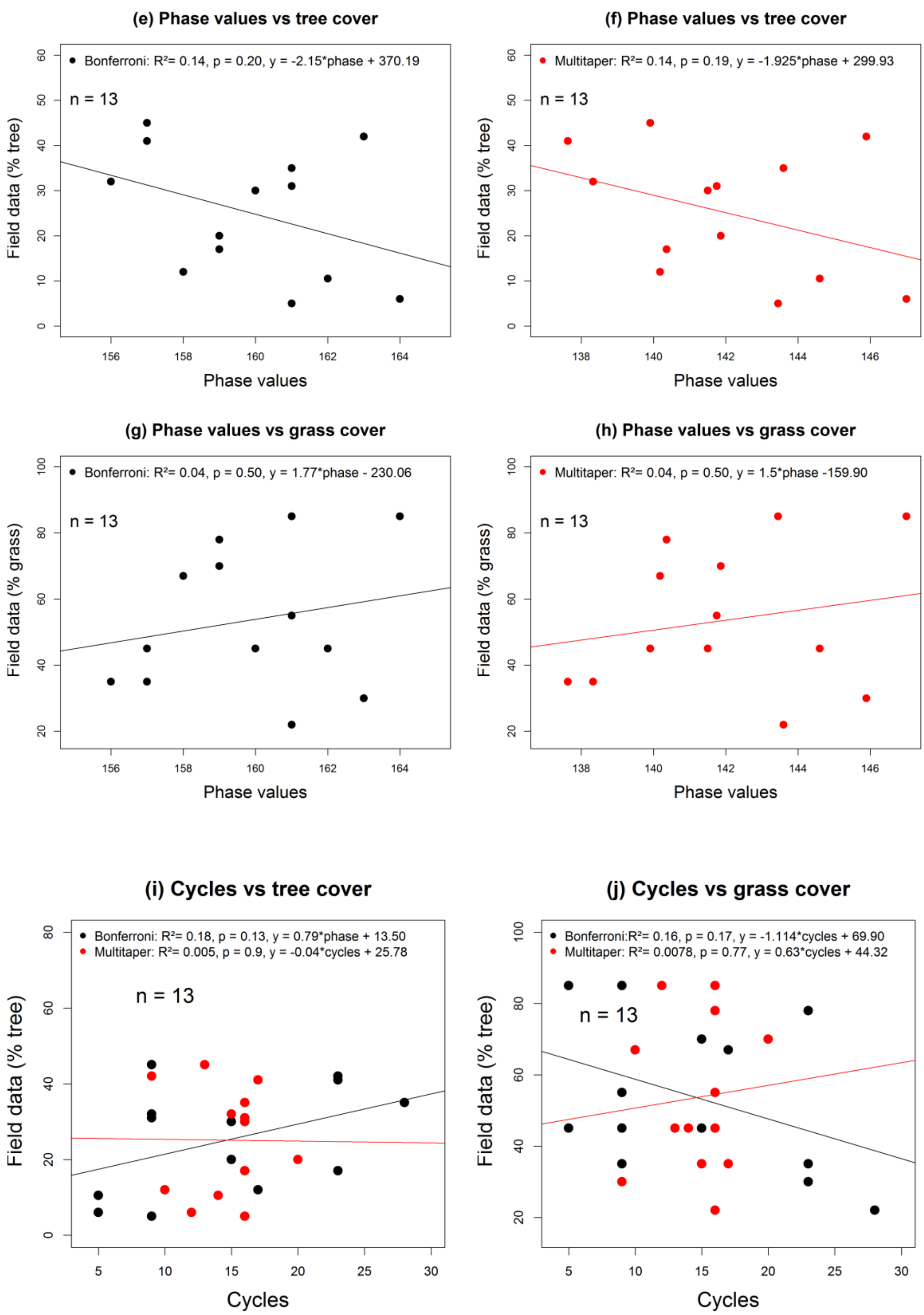


Figure 4. 10: calibration of phase (1st harmonic) and cycle values (2nd harmonic) to estimates tree/grass cover, (a) Phase values vs tree cover, (b) Phase values vs grass cover, (c) Cycles vs tree cover, (d) Cycles vs grass cover and (e-j) phase vs tree/grass cover without plot 27

4.5.8 Calibration using multiple regression between cycles, phase values and observed tree cover

The table 4.7 shows a strong linear relationship between the phase and cycles as the independent variables and the percent tree cover as the dependent for Bonferroni and multitaper methods. The relationships show strong linear correlation with percent tree cover, $R^2 = 0.61$, $p < 0.01$ and $R^2 = 0.51$, $p = 0.01$ for Bonferroni and multitaper method respectively (Table 4.5). The validation results presented in next section would confirm whether multiple regression model is more useful than when variables were tested individually.

Table 4. 7 Multiple linear regression coefficient for phase and cycles with percent tree cover

Regression coefficient		**Bonferroni**	**Multitaper**
intercept		363.22	444.34
Slope	Phase	-2.181	-2.873
	Cycles	0.732	-0.754
R^2		0.60	0.51
P value		0.007	0.01

4.5.9 **Validation of tree/grass cover estimates using phase and cycles**

4.5.9.1 Simple linear regression tree/grass cover estimates

Table 4.8 and 4.9 shows field validation plots and their corresponding phase values, cycles, as well as estimated tree cover for Bonferroni and multitaper methods respectively. Figure 4.11 presents an accuracy assessment of tree and grass cover estimated using field data. It should however be noted that each estimated phase and cycles values here resulted from MODIS NDVI data of a 14-year period (July 2001 to June 2015). The MODIS NDVI harmonic tree cover estimated from the phase using Bonferroni method has an $R^2 = 0.55$, $p < 0.01$, slope = 1, with RMSE = 13.13% and $R^2 = 0.44$, $p = 0.01$, slope = 1.2, with RMSE = 17.64% for tree cover and grass cover respectively. While the MODIS NDVI harmonic tree cover estimated from the phase using multitaper method has an $R^2 = 0.62$, $p < 0.01$, slope = 1.2, with RMSE = 12.52% and $R^2 = 0.41$, $p = 0.01$, slope = 1.4, with RMSE = 18.02% for tree cover and grass cover respectively. The estimates of grass cover had the highest error. The estimate of tree cover is better with cycles ($R^2 = 0.55$, $p = 0.03$, slope = 1, with RMSE = 16.07%) using Bonferroni method and otherwise for grasses ($R^2 = 0.32$, $p = 0.03$, slope = 1, with RMSE

= 17.91%). The estimates of tree (R^2 = 0.10, p = 0.27, slope = 0.74, RMSE = 19.11%) and grass cover using cycles (R^2 = 0.03, p = 0.84, slope = 0.26, RMSE = 23.91%) are not significant using the multitaper method. The level of accuracy in the estimates of grass cover as opposed to calibration could simply be explained as a function of strong inverse relationship of the two PFTs.

Table 4. 8: Validation of tree/grass cover estimates derived through simple linear regression using phase, cycles and field data (using Bonferroni)

Plot no	**Observe tree cover (%)**	**Observe grass cover (%)**	**Phase values**	**Cycles**	**Phase**		**Cycles**	
			1st	2nd	Estimated Tree cover (%)	Estimated Grass cover (%)	Estimated Tree cover (%)	Estimated Grass cover (%)
2	5	70	162	9	19.41	58.08	25.63	51.91
5	11	85	158	9	30.41	49.53	25.63	51.91
6	11	42	159	9	27.66	51.66	25.63	51.91
8	12	70	161	9	22.16	55.94	25.63	51.91
11	22	35	163	9	16.16	60.21	25.63	51.91
13	30	40	158	9	30.41	49.53	25.63	51.91
14	30	45	158	15	30.41	49.53	33.60	43.12
17	35	35	163	28	16.66	60.21	50.87	24.09
18	35	50	157	9	33.16	47.39	25.63	51.91
20	35	57	161	23	22.16	55.94	44.22	31.41
23	45	50	161	2	22.16	55.94	16.33	62.16
25	55	15	152	28	46.90	36.70	50.87	24.09
26	65	10	152	16	46.90	36.70	34.92	41.66
28	70	10	143	28	71.64	17.47	50.87	24.09

Table 4. 9: Validation of tree/grass cover estimates derived through simple linear regression using phase, cycles and field data (using multitaper)

Plot no	**Observe tree cover (%)**	**Observe grass cover (%)**	**Cycles**	**Phase values**	**Phase**		**Cycles**	
				1st	Estimated Tree cover (%)	Estimated Grass cover (%)	Estimated Tree cover (%)	Estimated Grass cover (%)
2	5	70	11	144.26	19.67	57.81	20.47	55.48
5	11	85	14	140.61	28.93	50.69	26.10	52.39
6	11	42	14	143.15	22.49	55.64	26.10	52.39
8	12	70	12	142.3	24.65	53.99	22.35	54.45
11	22	35	15	145.01	17.77	59.27	27.98	51.36
13	30	40	17	140.71	28.68	50.88	31.74	49.30
14	30	45	19	139.49	31.78	48.51	35.49	47.24
17	35	35	8	144.48	19.11	58.24	14.84	58.57
18	35	50	20	139.03	32.94	47.61	37.37	46.21
20	35	57	15	141.32	27.13	52.07	27.98	51.36
23	45	50	27	143.93	20.51	57.16	50.50	39.00
25	55	15	11	134.76	43.78	39.28	20.47	55.48
26	65	10	16	132.77	48.83	35.40	29.86	50.33
28	70	10	18	127.07	63.29	24.29	33.61	48.27

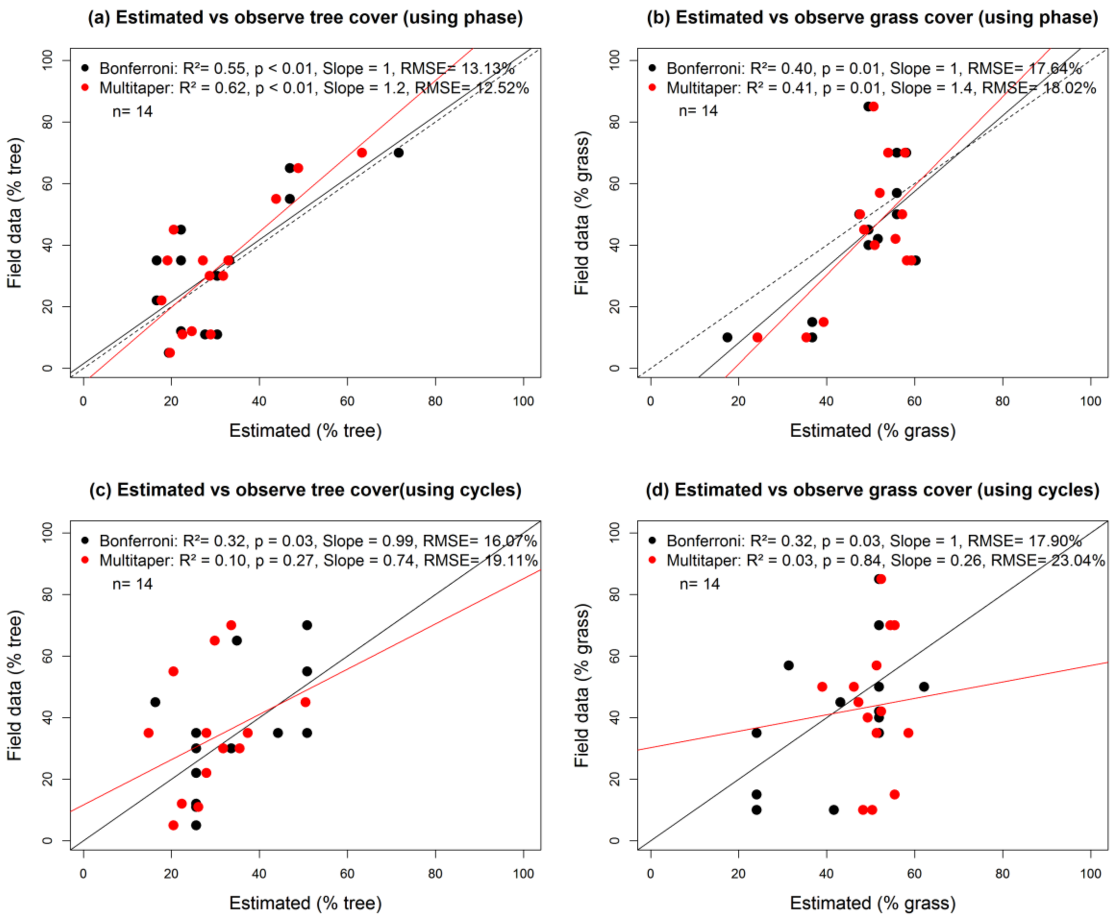


Figure 4. 11 Validation of tree/grass cover estimates derived through simple linear regression using phase, cycles and field data, (a) Validation of tree cover estimated with phase values, (b) validation of grass cover estimated with phase values, (c) validation of tree cover estimated with cycles, (d) validation of grass cover estimated with cycles

4.5.9.2 Multiple linear regression estimates

The Figure 4.11 shows the accuracy assessment for tree cover estimated from multiple variables. This model has an increased accuracy compared to the individual variables in a simple linear regression (Figure 4.12). The accuracy of multiple regression for Bonferroni estimate of phase and cycles shows an $R^2 = 0.61$, $p = 0.001$, and slope = 0.99 and the least RMSE = 12.54 %. The accuracy of the multiple regression model estimated between the phase, cycles (multitaper variables) and tree cover is lower than ($R^2 = 0.51$, $p < 0.01$, and slope = 0.98 and RMSE = 13.85 %, Figure 4.12) when phase alone was used ($R^2 = 0.62$, $p < 0.01$, slope = 1.2, with RMSE = 12.52%, Figure 4.11). This means that cycles estimated from the Bonferroni method appear to be physically more meaningful than the ones derived from the multitaper method.

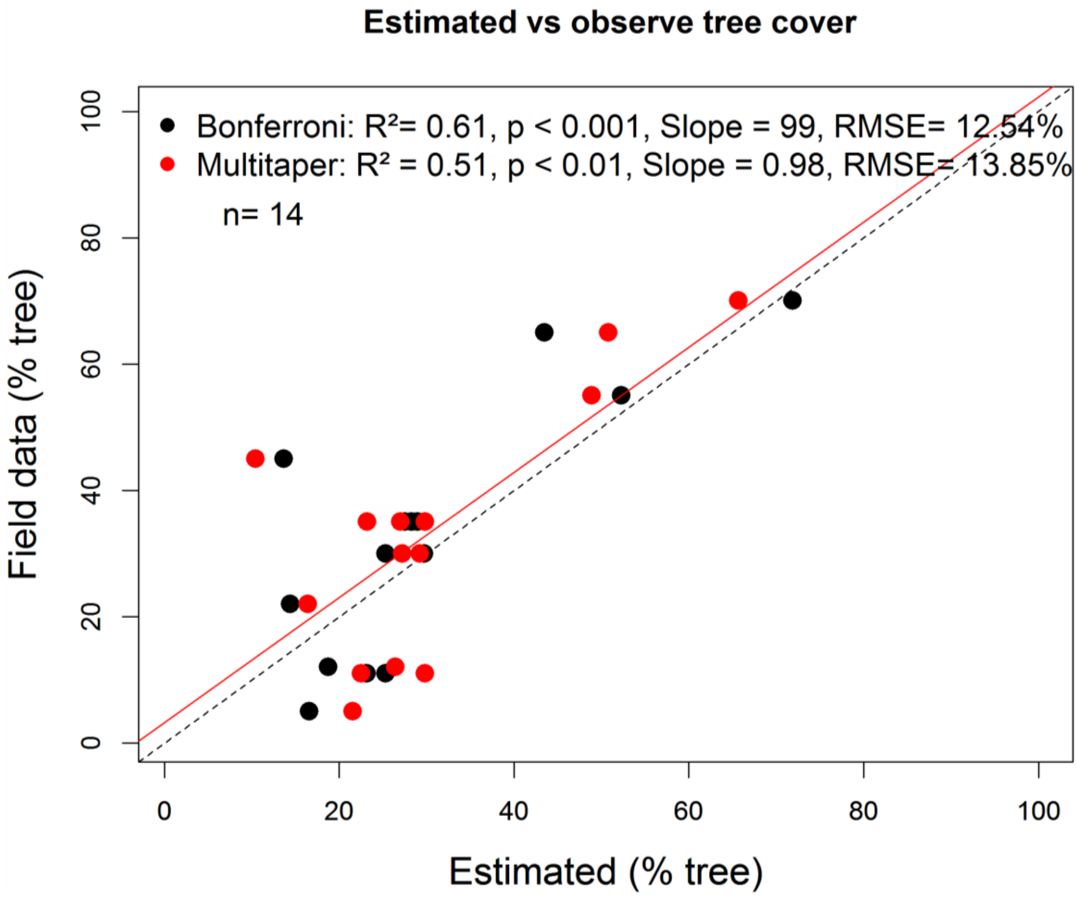


Figure 4. 12: Validation of tree cover estimates through multiple regression of phase and cycles for Bonferroni and Multitaper

4.6 Discussion

This study shows a harmonic decomposition of a 14-year time series of MODIS NDVI data over a savanna site in Kruger National Park, South Africa. The study has shown that the interannual variability tree/grass phenology can be derived from the amplitude values of the harmonic terms (Figure. 4.5, 4.6, 4.7, 4.8 and Table 4.3 and 4.4). The greening pattern (seasonal and interannual NDVI) of these two main PFTs varies with their relative composition. Trees green up earlier in the wet season than grasses, as observed using signal decomposition (Table 4.1, 4.2, 4.3 and Figure 4.5). In the study area, the rainy season normally starts around September and ends in April. The earliest period in these series was recorded in the growing cycles of the second harmonic term. Except for the harmonic term with the highest amplitude (14 cycles over 14 years, i.e. the annual signal), the number of cycles for PFTs for most harmonic terms differ (Table 4.1, 4.2). The harmonic analysis revealed that grass phenology has the maximum amplitude at the peak of the growing season (strongest harmonic term) than tree phenology. This is caused by grasses responding more strongly to the annual seasonality of wet/dry seasons than trees, which can tap into deeper water reservoirs through their deep roots (Whitecross, Witkowski, and Archibald 2017a).

Grasses usually have a high NDVI values in their most active photosynthetic stage during the rainy season, as was found in a previous study by Archibald and Scholes (Archibald and Scholes, 2007b). In their paper, however, there was a hidden periodicity for the timing of the maximum NDVI of trees that was not captured by the moving average method they applied (Archibald and Scholes, 2007b). An explanation of that issue is that the moving average method has no standard criteria for choosing the delay time in the moving window and that it is inappropriate to assume that the moving average can capture the phenology in its entirety in savannas where weather variability, fire frequency and herbivory are prevalent (Eklundha and Jönssonb, 2012, Bombelli et al., 2009, de Beurs and Henebry, 2010, Verbesselt et al., 2010).

Here, this study identified tree/grass dynamics in 28 plots of Kruger National Park from the amplitude, cycles and phase values of the strongest harmonic terms, excluding any terms that were not statistically significant when applying the Hartley test and correcting for multiple testing with the Bonferroni method. When applying harmonic analysis to sequences of one year of NDVI data, the amplitude of the two PFTs varies between years (Figure 4.5 and Table 4.2). The phase values show inconsistencies concerning the timing of the tree/grass phenology, especially in the terms that are weaker than the one with the highest amplitude (Fig. 4.5a-d). This might be due to an asynchronous start of the rainy season leading to grasses greening up while the trees are limited by the temperatures and photoperiod (Archibald and Scholes, 2007b). Although our study cannot conclusively attribute these changes to specific factors, it is known from the literature that the magnitude and consistency of the first and second strongest harmonic terms relate to secondary succession, weather anomalies and other land cover changes (Moody and Johnson, 2001). Similarly, Jakubauskas et al. (2001) explained that changes in harmonic parameters (amplitude and phase) can indicate changes in the natural vegetation, e.g. in terms of maximum greenness (due to onset of greenness), or changes in vegetation condition due to drought, flooding or overgrazing or land surface condition (changes arising from post-fire regeneration, natural or anthropogenic disturbance). The fluctuations in the amplitude values as well as the NDVI of the dry season have been observed using the annual rainfall data over four weather stations (Figure 4.6). The dry season NDVI appear to be more stable than the amplitude.

Furthermore, the time of greening and changes in the minimum and maximum NDVI/amplitude values of PFTs in different ecosystems and species can be influenced by the climatic conditions or anthropogenic disturbance. This finding can be further supported with the work of Sankaran *et al.,* 2005 who investigated the determinants of woody cover in over 854 savannah sites in African. The study highlights the influence of resource (such as water, nutrients, fire and herbivore etc.) availability and the distribution of plant species within savannah ecosystem. The analysis demonstrated that the tree cover is not simply associated to resource abundance, that for all sites with <650mm mean annul precipitation (MAP), tree cover is constrained linearly with moisture availability (and permits grasses to coexist), while for the sites having >650mm MAP canopy closure is possible. Disturbances such as fire can prevent canopy closure. Therefore, climate, soil nutrients, fire are some of the essential components that monitor savannah dynamics.

The phase and cycles have been consistent measures for discriminating tree and grasses as well as for estimating percent tree and grass cover using a field data collected in 2015 (Figure 4.5, 4.8, 4.9, 4.10, 4.11, 4.12). As the tree phenology had earlier greening period, the phase values had produced an estimates of tree cover with the least error compared to grass (Figure 4.11a/b). The estimate of tree cover is also better with the cycles due to its bimodal characteristics (Figure 4.10c/d) using Bonferroni. However, the relationship between tree and grass cover with cycles estimated with multitaper method is not significant (Figure 4.10c/d). The F-test with the multitaper estimator tends to find significant peaks in the low frequencies as well as more spurious peaks in the middle and high frequencies (Table 4.4, 4.5 and Figure 4.7). Similar behavior of cycles estimated from the multitaper method was observed by Pardo-Igúzquiza and Rodríguez-Tovar (2015) who compared the statistical significance of maximum entropy estimates and Thomson multitaper method using an F-test in the estimates of statistical significance of their power spectrum estimates for the cyclicity of past ocean/atmosphere dynamics (from decadal to millennial time scales). Their results show that the cycles identified as significant by maximum entropy have a clear physical interpretation (as presented with

Bonferroni in this study) while Thomson multitaper estimator had no significant peaks in the low frequencies and tends to give as significant more spurious peaks in the middle and high frequencies (Pardo-Igúzquiza and Rodríguez-Tovar, 2015).

The synergy between phase and cycles estimated with Bonferroni for the estimates has yielded an increased accuracy for tree cover estimate (Figure 4.12). These estimates are more accurate to tree cover than grasses and for all methods. In model calibration, grass cover had no significant relationship with either of the variables but appear to be accurate in the validation results for Bonferroni. The accuracy of grass cover in the validation result was due to presence of three plots (Table 4. 6) with predominant tree cover. This occurs due to the inverse relation between tree and grass cover (Figure 4.8). Grass cover decreased non-linearly with increasing tree cover as shown in Figure 4.8. Previous studies demonstrate an empirical relationship between the fraction of maximum tree cover and annual grass productivity (Aucamp et al., 1983, Beale, 1973, Walker et al., 1972), where grass density decreases as the fractional tree cover increases. The use of phase and cycles using empirical methods is an important contribution to remote sensing of tree/grass fractional cover estimations as the effects of soil backgrounds remain a significant challenge in the estimate tree/grass cover fractions especially when using vegetation indices.

Although raw NDVI values have been reported as being sub-optimal for FVC estimation, some methods which account for soil background contribution in the NDVI have shown good relationships with ground measurements (Moreno-de las Heras et al., 2015). However, signal contamination, soil background colour and saturation problems limited the NDVI-FVC relationship (Verger et al., 2009b). Although detrending the time series may be useful in spectral analysis even with phenomena that is nonstationary dynamic (Wu et al., 2007), the estimate of tree cover from the amplitude may be more appropriate without removing the trend (Hernandez, 1999). While detrending may provide information on the significant peak (e.g. annual), tree and grass have different phenological cycles. The implication of detrending could lead to depressing the amplitude of the lower frequency components and increases the amplitude of the higher frequency components of the original signal's spectrum (Hernandez, 1999). Detrending the NDVI signals in harmonic analysis amplitude may limit amplitude values in predicting tree cover because the detrended data do not represents full spectrum of the original data

(Hernandez, 1999). Therefore, the estimates of tree cover require that the full spectrum relative to whole time series data be used. The tree cover was estimated in the next chapter using amplitude derived from harmonic analysis without detrending. In addition, the dry season NDVI appear to be applicable to woody cover estimation as suggested by the tree phenological behaviour in the phase and cycles of the first and second strongest harmonic term respectively (Figure 4.10, 4.11, 4.12).

In general, the use of harmonic analysis has been considered limited due to its demand for prior ecological knowledge, long-term datasets and the need for effective interpretation of confidence intervals of the observations in the power spectrum (de Beurs and Henebry, 2010). Our study recognizes these limitations and adopts approaches to minimize them. A long-term dataset of over a decade was interpreted using field information from 28 plots collected in one year.

Despite careful experimental planning, a pixel-level analysis in remote sensing and geographical information science (GIS) is subject to some remaining uncertainties depending on the datasets and modelling approaches (Fisher, 1997). Here, the uncertainty in the phenology analysis can be identified as having three main components:
1) Uncertainty in aligning field data with satellite pixel areas when matching NDVI time-series to fractional cover data from plots;

2) Uncertainty how the tree/grass cover inside the field plots may have changed over the 14 years;

3) uncertainty inherent in the NDVI retrieval from the MODIS sensor, e.g. the impact of sensor viewing geometry and attenuation of the signal by tree canopies (McCoy, 2005, Gill et al., 2009, Los et al., 2005).

Firstly, since the field survey has considered the MODIS satellite pixel size by sampling plot of almost equal to the size of the pixel, the centers of the plots were chosen such that they are representative of a large surrounding area and that the point data were geo-referenced to the projection system of MODIS; the first uncertainty term is considered minimal. Second, the use of recent field data for mapping annual and interannual tree/grass phenology in savanna may not entail significant uncertainty. In the savanna, a tree cover change is more stable over time than changes in the grass layer (Guerschman et al., 2009). High variability is reflected by the grass contribution to the overall phenological signal of a plot and this can affect the tree/grass separation method (Archibald and Scholes, 2007b). Although there can be small changes in the amplitude mostly contributed by the grass layer (Scanlon et al., 2002), the use of a composite dataset (Holben, 1986, Hilker et al., 2009) is more promising than a single-date dataset (Mondal et al., 2014). Third, for MODIS data product MOD1Q31 the influence of viewing geometry on vegetation indices has been investigated and was found to be insignificant in ecosystems with less complex canopies structures (Gill et al., 2009)**.** MODIS NDVI is atmospherically corrected bi-directional surface reflectance factors (BRDFs) and masked for water, clouds, and cloud shadows (Strahler et al., 1999). NDVI is also less sensitive to BRDFs compared to individual bands (Los et al., 2005).

The phase values and cycles, particularly the first and second strongest terms, provide a robust method for estimating FVC of trees and grasses for the Skukuza study site (Table 4. 7, Figure 4. 10, 4.11, 4.12). The significance of this study is that it shows that harmonic analysis has a high discriminatory ability of trees and grasses in savannas. Tests in other savanna types could help show whether it is sufficiently robust to retrieve FVC of trees and grasses at the continental scale. The availability of robust tree/grass FVC datasets over time would enable new ecological studies of tree/grass coexistence to be carried out.

4.7 Summary

The signal decomposition method of harmonic analysis was applied to estimate fractional cover of tree and grass PFTs in Kruger National Park. MODIS NDVI time series data over 14-years were decomposed and statistically significant harmonic terms were estimated. The amplitude, cycles and phase values show distinct patterns for trees and for grasses. The cycles and phase of the strongest harmonic terms were a robust discriminator between tree and grass phenology because grasses respond more strongly to the annual seasonal cycle than trees. The phase values show that trees green up earlier than grasses. Tree/grass phenology from satellite remote sensing can be used to estimate their fractional covers as the phase has an $R^2 = 0.55$, $p = 0.002$, slope = 1, with RMSE = 13.13% and $R^2 = 0.44$, $p < 0.01$, slope = 1.2, with RMSE = 17.64% for tree cover and grass cover respectively (based on Bonferroni). The estimates of grass cover had the highest error. The estimate of tree cover is better with cycles ($R^2 = 0.55$, $p = 03$, slope = 1, with RMSE = 16.07%) than for grass ($R^2 = 0.32$, $p = 03$, slope = 1, with RMSE = 17.91%). The accuracy assessment of the multiple linear regression model for tree cover estimate shows an $R^2 = 0.61$, $p < 0.001$, and slope = 0.99 and the least RMSE = 12.54 % using Bonferroni method. The accuracy has increased greatly compared to when models were observed with individual variables based on Bonferroni method. The estimates of grass cover using harmonic is more challenging as grasses tend to be more susceptible to environmental changes than trees. The multitaper method is more accurate in the estimates of tree cover using phase information than cycles ($R^2 = 0.62$, $p < 0.01$, slope = 1.2, with RMSE = 12.52%). Generally, the estimates of tree/grass cover using harmonic analysis is limited to areas that are well-known.

Despite the successfully implementation of harmonic analysis using MODIS, it is necessary for further studies to characterize changes in a more diversified vegetation types, particularly beyond the sample field data to simulate another condition base on the field information and within the trend of MODIS data using signal decomposition. Although MODIS high temporal resolution is important for change analysis, validation of such woody fractional cover is valuable. Comparison of the estimates tree cover can be encourage using medium (e.g. Landsat), high (LiDAR) resolution dataset and other previous satellite products (such as MODIS VCF) derived from these sensors to examine to what extent can the methods develop for different datasets will reflect minimum reduction or addition in the quantitative estimates of FVC. In the next chapter (5), tree cover was estimated for a large part of the study area using the amplitude images, NDVI and SAVI (dry season images) vegetation indices based on the understanding of tree phenological behavior in KNP as observed and presented in this chapter.

Chapter 5

Estimating tree fractional cover in African savanna using MODIS time series data

5.1 Introduction

Woody and herbaceous vegetation as well as bare soil contribute differently to the NDVI resulting in a mixed signal at the sub pixel scale (Helman et al., 2015:Montandon and Small, 2008, Ding et al., 2016). To estimate tree cover, therefore, the contribution from grasses and the soil background need to be accounted for, especially where soil inventories are available with good temporal resolution (Smallman et al., 2017). In savanna, the mixed tree/grass system is controlled by climatic conditions, soil moisture and disturbances (Scholes and Archer, 1997, Sankaran et al., 2005). Grasses cannot use deeper groundwater reservoirs and depend mostly on instantaneous rainfall during the rainy season; therefore, they are limited by water availability. In contrast, many tree species flush their leaves before the first rain in response to photoperiodic and temperature triggers (Archibald and Scholes, 2007b). Thus, signal decomposition of MODIS NDVI (amplitude) time series data could be useful in tree cover estimation (Gessner et al., 2013).

Moreover, from the analyses presented in chapter four, the phase and cycles show that tree species had an earlier greening period than grasses in KNP. The analysis of MODIS NDVI (amplitude) time series data using DFT may provide an estimate of fractional tree cover over a given period when relevant frequency components of DFT are considered (Moody and Johnson, 2001). In addition, the dry season images from MODIS data for NDVI and SAVI could also be used to provide estimate of tree cover if soil determining methods are considered (Montandon and Small, 2008, Zeng et al., 2000, Ding et al., 2016). In this chapter, tree cover was estimated in three main ways: (1) by using the amplitude values as derived from a signal decomposition of NDVI data using harmonic analysis, (2) using the NDVI data without applying harmonic analysis (3) by the decomposition of NDVI and SAVI signals to determine the contribution of bare soil using two soil determining methods.

5. How well the satellite data derived metrics from MODIS can be used to estimate percent tree cover?

5.2 Objectives

5. To estimate tree fractional cover using amplitude (2004-2015) derived from the annual MODIS NDVI data as well as from dry season NDVI and SAVI data.

5.3 Methods

5.3.1 Tree cover estimation

Definitions of the variables and notations used for these methods are provided below:

f_t - is the tree fractional cover

f_s - fractional cover of bare soil

f_g - grass fractional cover

$NDVI_{pixel}$ - *is the mean dry season MODIS NDVI*

$SAVI_{pixel}$- *is the mean dry season MODIS SAVI*

$NDVI_{T}$ - *is the fraction of NDVI for tree cover*

$SAVI_{T}$- *is the fraction of SAVI for tree cover*

$NDVI_{veg}$ - *is the NDVI maximum value for vegetation*

$SAVI_{veg}$ - *is the SAVI maximum value for vegetation*

$NDVI_{s}$ - *is the soil threshold value for NDVI*

$SAVI_{s}$ - *is the soil threshold value for SAVI*

$NDVI_{si}$ - *is the soil type threshold value for NDVI (at certain location (i))*

$SAVI_{si}$ - *is the soil type threshold value for SAVI (at certain location (i))*

$NDVI_{soil}$ - *is the soil NDVI value for a pixel*

$SAVI_{soil}$ - *is the soil SAVI value for a pixel*

$NDVI_{soil i}$ - *is the soil NDVI value for a pixel based on the soil type at certain location*

5.3.2.1 Tree cover estimate using harmonic analysis

From the previous analyses presented in chapter four (4), phase and cycles values derived through a temporal analysis of each pixel have identified an inter-annual variability over the period of study which further suggests the possibility for tree cover estimation. In this chapter, the derived amplitude of the first harmonic term were used for tree cover estimates. It should be noted that unlike the previous analysis presented in chapter four, the estimated harmonics presented in this chapter are not based on detrended time series data. Since leaf-out period for plant species varies in this region, and that tree phenology has more growing cycles compared to grasses, the statistically significant harmonics based on Bonferroni method need to be assessed based on full spectrum (Hernandez, 1999). In this chapter, a spatial analysis of signal decomposition with Hartley-test using Bonferroni method for the MODIS NDVI data (July 2001- June 2015) was performed. The first strongest harmonic term was used for tree cover estimation.

5.3.2.2 Tree cover estimate using $NDVI_{pixel}$ and $SAVI_{pixel}$

The mean NDVI and SAVI over the dry period were calculated ($NDVI_{pixel}$/$SAVI_{pixel}$) and used to estimate percent tree cover based on certain assumptions. In this ecosystem, woody species have two growing cycles at the time when the herbaceous layer is dormant. Grass usually dries before woody species lose their leaves in autumn so that we have two small periods with dry grass and green woody canopy (before and after the wet seasons) (Archibald and Scholes, 2007b). These cycles occur before (May-August) and after the wet season (May-August). The wet season starts from September and ends in April. Therefore, growing season overlap each calendar year (Archibald and Scholes, 2007b). To reduce an overlap of tree and grass phenology occurring probably due to delayed start or end of season, dry season images were chosen for the months of July and August before the start of season and May and June after the end of season. This is useful in capturing the phenology of woody plants. There are reasons for these considerations: (1) Tree species are usually fully green before the first significant rains (e.g. *Sclerocarya birrea, Acacia nigrescens*) in KNP. (2) Some trees such as *Combretum Apiculatum* are usually late in their leafing but take a shorter period to present full leaf than early flushers. (3) Woody species in KNP usually take 8 weeks to reach full leaf from the date that first woody vegetation started leafing (Archibald and Scholes, 2007b).

Although these vegetation indices are sensitive to vegetation fractional cover, they are also sensitive to soil background (Montandon and Small, 2008, Ding et al., 2016). The assumption here does not completely argue that there is no bare soil influence in the NDVI. For the SAVI vegetation index, the soil correction factor usually applied to derive the vegetation index might reduce bare soil influence. While this approach may constitutes uncertainty in tree cover estimations, the effects of soil reflectance due the nature of soil type characteristics (e.g. such as soil brightness, moisture) in this region is likely to be negligible because of the extent of the spatial scale (smaller) being considered (Lillesand et al., 2014). In addition, the soil background reflectance values are lower than

the canopy reflectance due to high albedo in the tropics (Los et al., 2000). Jiang et al (2006) found that the nonlinearity of NDVI over partially vegetated surfaces is more prominent with darker soil backgrounds and shadow (Jiang et al., 2006).
The aim is to establish the fundamental relationships between $NDVI_{pixel}$ and field data on tree cover to develop a calibration technique to assess the extent and accuracy to which such relationships could estimate tree cover. Both linear and non-linear regression analyses methods have been used for data calibration as the relationship between NDVI and measurements of canopy structures vary with vegetation types and seasonality (Gamon et al., 1995, Sellers, 1987).

5.3.2.3 Tree cover estimate using $NDVI_{soil}$ and $SAVI_{soil}$ determining methods.

In this scenario, first, it is assumed that each pixel consists of three constant fractional covers: tree cover (T), bare soil cover (S), and grass cover (G). In savannas, during the dry period, most of grass fractions are occupied by the bare soil (or remain dried) while in the wet seasons, the grass fractional cover makes up most of the contribution (Scanlon et al., 2002, Archibald and Scholes, 2007b). In the dry season, the grass layer becomes non-photosynthetic and dries. So, the non-photosynthetic grass layer was merged with bare soil as the fractional cover for each pixel = 1. It has been reported that the grass layer (fraction of photosynthetic vegetation) in savannas may have changed from 85% to 8% in the dry season and fraction of non-photosynthetic vegetation of the same layer may increase from 7% to 79% in the wet season (Guerschman et al., 2009). Furthermore, an investigation of field reflectance measurements of bare soil, grass and tree indicated that dry grass had the lowest NDVI (Guerschman et al., 2009). $NDVI_{pixel}$ is attributed to only woody and bare soil as expressed in equation 5.1.

$$NDVI * f\_t + NDVI_{soil} * (f\_s+f\_g) = NDVI_{pixel} \quad (5.1)$$

f_t - is the tree fractional cover, f_s - fractional cover of bare soil, f_g - grass fractional cover, NDVIpixel -is the mean NDVI of the dry season from MODIS data. The contribution of from grass is assumed very low in the dry season. Therefore, equation 5.1 merged the fractional cover of bare soil and grass cover ($NDVI_{soil} * (f\_s+f\_g)$). This means that the influence of bare soil and grass cover on the vegetation index does not usually allow spectral signals of woody vegetation to vary. To estimate the spatial variability of woody cover, over a large area, the contribution of bare soil and grass cover need to be accounted for. Different techniques of vegetation fractional cover estimate have been

proposed previously (which accounted for bare soil contribution in the pixel). Some of these techniques are usually invariant to soil types and characteristics (Gutman and Ignatov, 1998, Sobrino and Raissouni, 2000, Los et al., 2000).These methods are usually based on the assumption that pixels with FVC = 1 and 0 exist in an image. These are donated as $NDVI_{veg}$ and $NDVI_{s}$ for maximum vegetation and bare soil respectively.
Gutman and Ignatov, (1998) who used low spatial resolution data (0.15˚x15˚) proposed $NDVI_{veg}$ at 0.52 ± 0.03 and NDV_{soil} at 0.05 ± 0.03. Similarly, Sobrino and Raissouni (2000) applied thresholds of 0.5 and 0.02 for $NDVI_{veg}$ and NDV_{soil} respectively. In this section, previous methods by Zeng et al. (2000) and Wu et al. (2014) are adopted with modification due to consideration of the relatively small study site, spatial data resolution, and lack of field soil spectral reflectance measurement. The same procedure was applied to SAVI to determine the contribution of soil as Ding et al. (2016) show that the impact of soil backgrounds in the SAVI still need to be accounted for.

i. The first method is invariant to soil characteristics for determining $NDVI_{soil}$. Zeng et al. (2000) determined $NDVI_{soil}$ by utilizing the percentile of vegetation types using the International Geosphere-Biosphere Program (IGBP) land cover classification with 1 km NOAA AVHRR NDVI data. They used the fifth percentile of the histogram of the maximum NDVI for the barren or sparsely vegetated category as the $NDVI_{s}$, which was 0.05, to estimate global FVC. Note however, that only tree cover is estimated in this study as opposed to Zeng et al. whose aim was to assess the statistically most likely FVC using the spectra of soil collected from different datasets. The procedure, however, requires that the histogram for each land cover to be computed. Considering the size of the study area, the histogram for the whole image was computed. Figure 5.1a shows the $NDVI_{pixel}$ histogram extracted from mean dry season NDVI for the 2014/2015. The graph shows a minimum and maximum values of NDVI at 0.12 and 0.65. Since, the maximum NDVI for this image is 0.65, this suggests lower $NDVI_{soil}$ than for barren and sparsely vegetated areas. In this case, $NDVI_{veg}$ and $NDVI_{soil}$ are approximated as 0.02 and 0.7 for bare soil and maximum NDVI respectively. The 0.7 is the maximum vegetation NDVI for the whole KNP. The 0.02 is threshold for $NDVI_{S}$ since the method is pixel dependent coupled with zero values observed in the NDVI histogram. For the SAVI, Figure 5.1b shows $SAVI_{pixel}$ histogram extracted from mean dry season NDVI for the 2014/2015. The $SAVI_s$ is threshold at 0.05 while maximum vegetation at 1.32 (the maximum for the whole KNP). The fractions of the NDVI/SAVI that is representing tree ($NDVI_T$/$SAVI_T$) was then calculated.

$$NDVI_T = \frac{NDVIpixel - NDVIsoil}{NDVIveg - NDVIsoil} \tag{5.2}$$

$$SAVI_T = \frac{SAVIpixel - SAVIsoil}{SAVIveg - SAVIsoil} \tag{5.3}$$

(a) $NDVI_{pixel}$ (2014/15) histogram

(b) $SAVI_{pixel}$ (2014/15) histogram

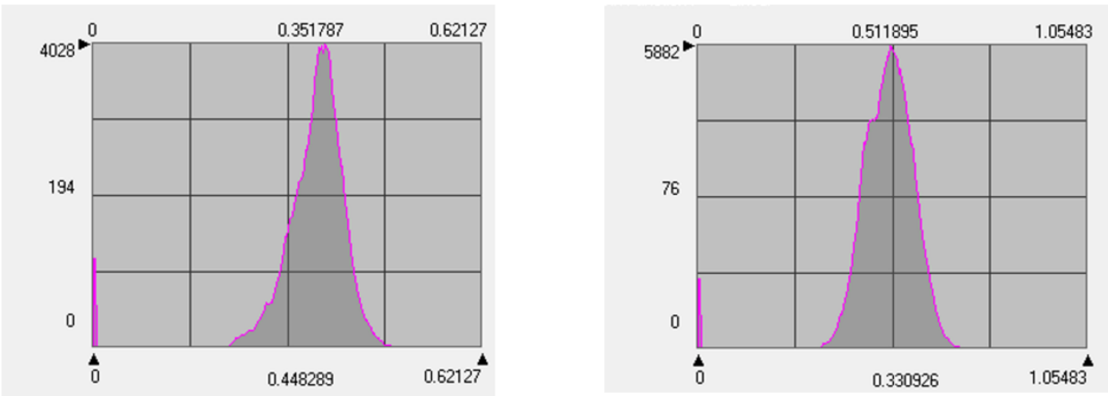


Figure 5. 1: NDVI histogram extracted from mean dry season image (2014/2015).

ii. Tree cover estimates using the $NDVI_{soil}$ determining method (Wu et al., (2014). The estimates of $NDVI_{soil}$ could be performed with considerable accuracy if available soil reflectance data from *in situ* measurements exist for the major types of soil in a study area (Montandon and Small, 2008, Wu et al., 2014, Ding et al., 2016). Unfortunately, it is always challenging to acquire this information (Montandon and Small, 2008). Consequently, many previous studies have relied on World Soil Database and the International Geosphere–Biosphere Program (IGBP) land cover classification to assign $NDVI_{soil}$ for each vegetation types especially when regional scale is being considered (Zeng et al., 2000, Wu et al., 2014).

Wu et al. (2014) used the Harmonized World Soil Database (HWSD) Version 1.21 which was produced by the International Institute for Applied Systems Analysis (IIASA) and the Food and Agriculture Organization of the United

Nations (FAO) to determine $NDVI_s$ for soil types. The Harmonized World Soil Database (HWSD) do not cover all the major soil types in KNP due to missing data from the map especially for the parts of our field data over KNP. On this basis, and for the purpose of this study, a global soil and terrain database at a scale of 1:1 million developed by the initiative of the International Union of Soil Sciences (IUSS), the United Nations Environment Programme (UNEP), the FAO, and the International Soil Reference and Information Centre (ISRIC) (van Engelen and Hartemink, 2000) has been used. Figure 5.2a/b show soil types for the KNP. Figure 5.2b indicates the major types of soil within our field plots data. Based on study by Wu et al. (2014), the $NDVI_{si}$ for the three types of soils in our plots locations which include Regosols, Luvisols, and Nitisols have been thresholded at 0.21, 0.24, and 0.32 respectively. The high $NDVI_{soil}$ in Wu et al (2011) might be because of the spatial resolution of datasets they have used (GIMMS NDVI dataset was derived at a 10 km) coupled with differences in the soil type and ecosystem conditions. Only three types of soils are covered in this study. Wu et al. (2014) investigated FVC at global scale. Even where soil types for this study are the same with Wu et al., the threshold used for Wu et al. is very high for this study considered minimum and maximum NDVI/SAVI for this study area. Another reason is seasonality. Wu et al. has the maximum NDVI between 0.6 to 0.94 for different vegetation types. This is much larger when compared to the NDVI used in this study (the dry season NDVI data) which has maximum value of only 0.65.

Moreover, though reference was made to the types of soils in determining the $NDVI_s$, the approach in this study differ from Wu et al. First a linear method is applied as opposed to Wu et al. Given the small size of the study area, only soil types are considered for each of the plots while a single value for maximum NDVI ($NDVI_{veg}$) was considered for all locations. For the NDVI, Regosols, Luvisols, and Nitisols have been thresholded at 0.015, 0.02, and 0.02 respectively. While, for SAVI Regosols, Luvisols, and Nitisols have been thresholded at 0.04, 0.05, and 0.06 respectively. This is in consideration to image histogram as the method is purely pixel dependent (Zeng et al., 2000). The fractions of the NDVI/SAVI that is representing trees ($NDVI_T$/$SAVI_T$) is also calculated as thus:

$$\mathrm{NDVI_T} = \frac{\mathrm{NDVIpixel} - \mathrm{NDVIsoil}i}{\mathrm{NDVIveg} - \mathrm{NDVIsoil}i} \quad (5.4)$$

$$\mathrm{SAVI_T} = \frac{\mathrm{SAVIpixel} - \mathrm{SAVIsoil}i}{\mathrm{SAVIveg} - \mathrm{SAVIsoil}i} \quad (5.5)$$

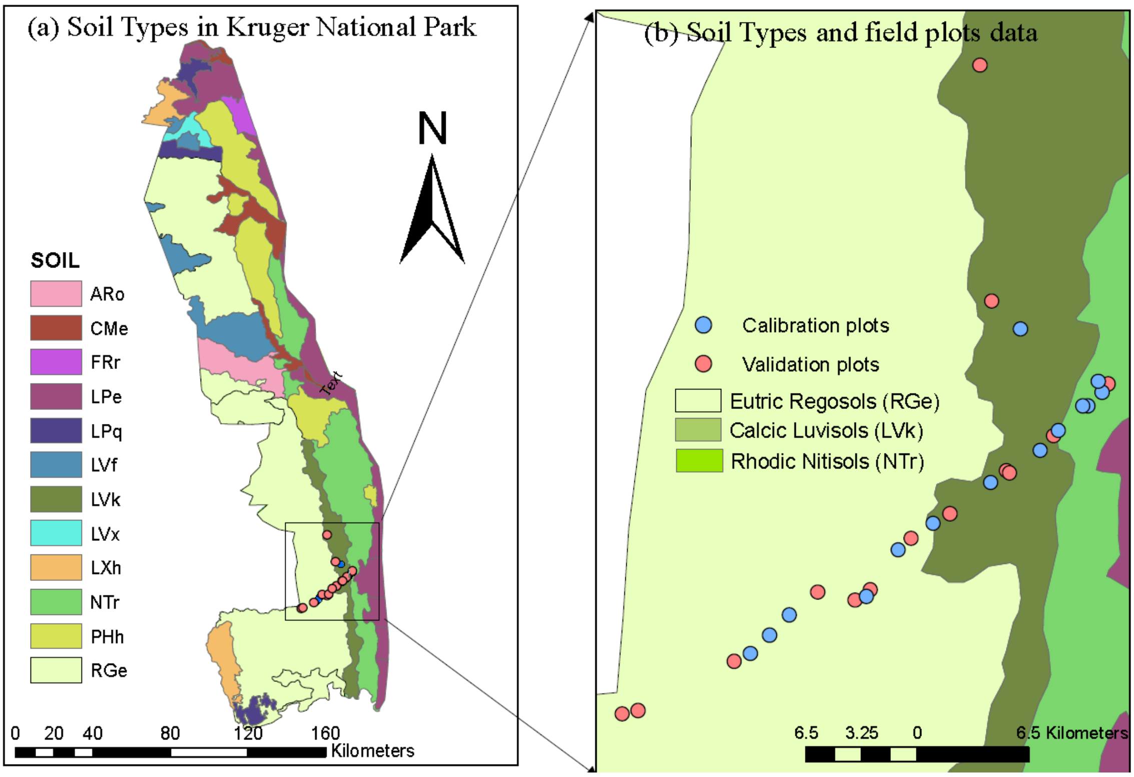


Figure 5. 2: (a) Soil types in KNP, while the blue and pink points in the box indicate calibration and validation plots respectively, (b) Main soil types in the field plot area, field plots data (the blue and pink points indicate calibration and validation plots.

5.3.3 Regression analyses

In this study, different types of regression models were established between the observed tree cover from a field campaign in 2015 and the independent variable(s). Only 50% of the field data on tree cover is used for (14 out of the 28 plots) model calibration while holding the remaining 50% for model validation. It should be noted that the type of regression analyses applied in this study depends on the nature of the phenology metric being used. The procedure is based on the assumption that NDVI-fractional vegetation cover relationship (or SAVI) is a function of vegetation type, the influence of understory and bare soil (Gamon et al., 1995, Sellers, 1987, Wang et al., 2016, Carlson and Ripley, 1997). The regression model applied for each phenology metric is explained below:

i. A nonlinear regression is applied to the amplitude data which resulted from a 14-year MODIS NDVI data to estimate percent tree cover. Specially, a logarithmic regression was used.

$$\% \text{ tree} = 228.12*\ln(\text{amplitude}) + 36.955 \qquad (5.6)$$

ii. Since the assumption to mean dry season images of MODIS data ($NDVI_{pixel}$ and $SAVI_{pixel}$) for the estimate of tree cover does not precludes the presence of bare soil, in this context, different regression models were tested in estimating tree cover. Although it has been previously reported that the relationship between vegetation indices (especially the NDVI) and percent cover depend largely on vegetation type (Gamon et al., 1995), or may even have strong linear relationship in a sparse vegetation (Sellers, 1987, Wang et al., 2016), it is not well-known how tree cover would be in KNP. Therefore, simple linear, polynomial and logarithmic equations were tested for both vegetation indices to find the best fit for percent tree cover estimation. The field data on percent tree as the dependent while the $NDVI_{pixel}$ or $SAVI_{pixel}$ as the independent. The equations are presented for linear, polynomial and logarithmic regression for the NDVI and SAVI respectively.

$$\% \text{ tree} = 284.86*NDVI_{pixel} - 71.158 \qquad (5.7)$$

$$\% \text{ tree} = 625.1*NDVI_{pixel^2} - 175*NDVI_{pixel} + 11.89 \qquad (5.8)$$

$$\% \text{ tree} = 99.02*\ln(NDVI_{pixel}) + 133.31 \qquad (5.9)$$

$$\% \text{ tree} = 240.69*SAVI_{pixel} - 97.692 \qquad (5.10)$$

$$\% \text{ tree} = 3.1994*SAVI_{pixel^2} + 236.99*NDVI_{pixel} - 96.597 \qquad (5.11)$$

$$\% \text{ tree} = 132.26*\ln(SAVI_{pixel}) + 114.73 \qquad (5.12)$$

iii. Only a simple linear regression is applied to $NDVI_{soil}$ and $SAVI_{soil}$ determining methods as the methods are themselves linear in this study. The fraction of $NDVI_T$ or $SAVI_T$ is used as the independent variable while the percent tree cover as the dependent variable. The model is the same for the two methods and for both $NDVI_T$ and $SAVI_T$. The regression equations for the variant and invariant methods are presented on equations 5.13 – 5.14 and 5.15 – 5.16 respectively.

$$\% \text{ tree} = 270.68*NDVI_T - 56.78 \qquad (5.13)$$

$$\% \text{ tree} = 177.74*SAVI_T - 57.845 \qquad (5.14)$$

$$\% \text{ tree} = 277.68*NDVI_T - 60.05 \qquad (5.15)$$

$$\%tree = 215.36 * SAVI_T - 71.28 \quad (5.16)$$

iv. A multiple linear regression was also established between the field data as a dependent variable and three variables with the least RMSE (when assessed individually). Therefore, not all variables were used in the multiple regression applied here to reduce multicollinearity effect. Specifically, the amplitude values calculated from the harmonic analysis, $NDVI_{pixel}$ and $SAVI_{pixel}$ were used for multiple linear regression.

$$\%tree = -139.67 + 75.69 * amplitude - 21.88 * NDVI_{pixel} + 196.17 * SAVI_{pixel} \quad (5.17)$$

v. Validation methods of all models explained here were presented in Chapter six where the actual validation results were presented.

5.4 Results

5.4. 1 Amplitude image of the first harmonic term, gauge stations and field plot data.

Figure 5.1a shows the amplitude data (14-year MODIS NDVI aggregated as one) of the first strongest harmonic term derived from the MODIS NDVI time series data (July, 2001 to June, 2015). On the image, the average annual rainfall (14 year) for the corresponding amplitude data was also presented (Figure 5.1). The amplitude image shows the spatial variability of vegetation distribution over KNP. The density of vegetation in KNP is usually influenced by environmental gradients of geological formations and annual rainfall (Smit and Asner 2012; Bucini et al 2010). The southern parts of KNP receive higher amounts of rainfall (e.g. Skukuza: 620.52 mm) and tend to be covered by dense thickets, while the northern part is mostly dominated by grassland because it receives insufficient rainfall to support dense vegetation cover (e.g. Mahlangeni: 410.71 mm). It is evident that the southern part of KNP has higher amplitude than the northern part in the first strongest harmonic term. It should also be noted that the woody species are more resistant to harsh environmental conditions (e.g. drought) than grasses as already discussed in Chapter four. The spatial distribution of field plots data on percent tree cover is shown on Figure 5.1b. As the next section dwells on model calibration (and validation in Chapter six), it is important to give the reader an insight into where these locations are.

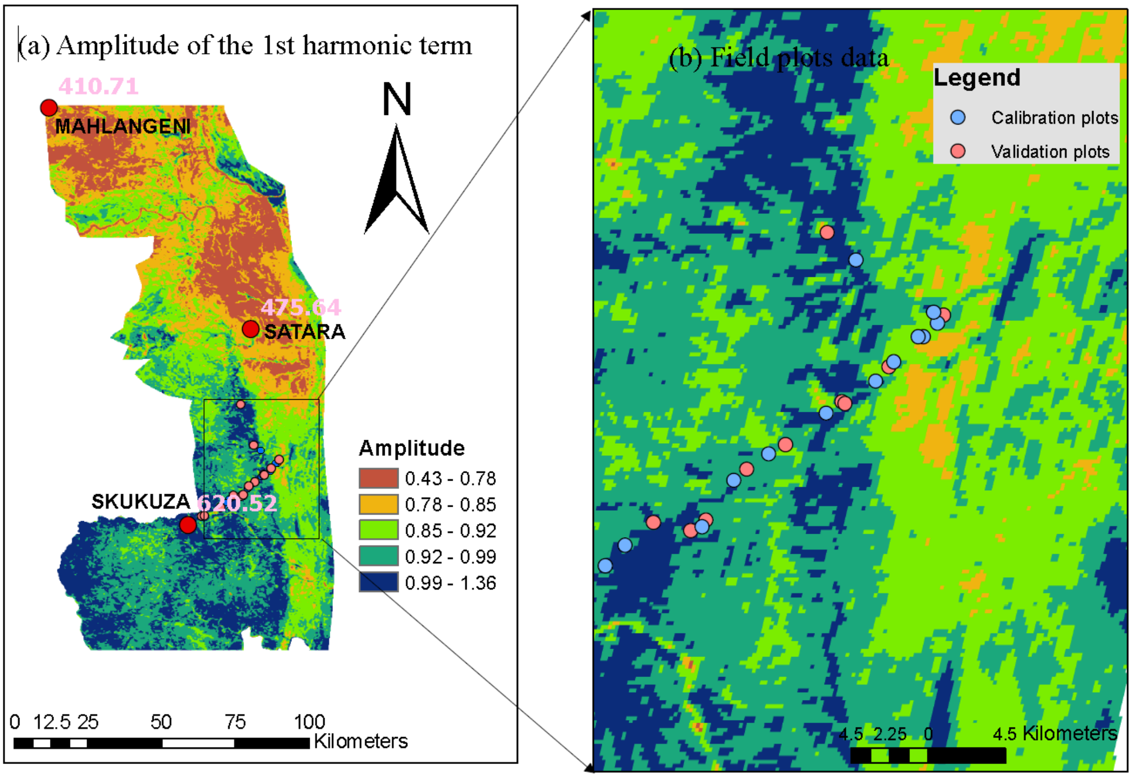


Figure 5. 3: (a) Amplitude image of the first harmonic term: red points show the gauge stations, the number in pink indicate the mean annual of rainfall for each location while the blue and red points in the box indicate calibration and validation plots respectively, (b) field plots data (the blue and pink points indicate calibration and validation plots

5.4. 2 Calibration of NDVI Harmonic tree cover estimate (Amplitude)

Table 5.2 indicates field plots and amplitude values used for data calibration. The Figure 5.4 displays the relationship between percent tree cover and amplitude as estimated from the MODIS NDVI data over KNP (2002-2015). The amplitude values of the first strongest harmonic term had strong correlation with percent tree cover (R^2 = 0.56, p <0.01). The amplitude values are fit to a nonlinear regression due to presence of high canopy, under-story and bare soil. Overall, the positive correlation between the amplitude and percent tree cover implies that it could be used as surrogates to percent tree cover at large spatial scales. However, the validation result with the remaining field plots data which is not used for calibration, will show whether this method is robustly enough to estimate percent tree cover over large area.

Table 5. 1: calibration of NDVI Harmonic tree cover estimate

Plot no	Tree cover	Amplitude (1st)
1	5	0.96
3	6	0.87
4	10.5	0.90
7	12	0.97
9	17	0.97
10	20	0.94
12	30	0.98
15	31	0.90
16	32	0.96
19	35	0.98
21	41	0.94
22	42	1.01
24	45	1.00
27	69	1.11

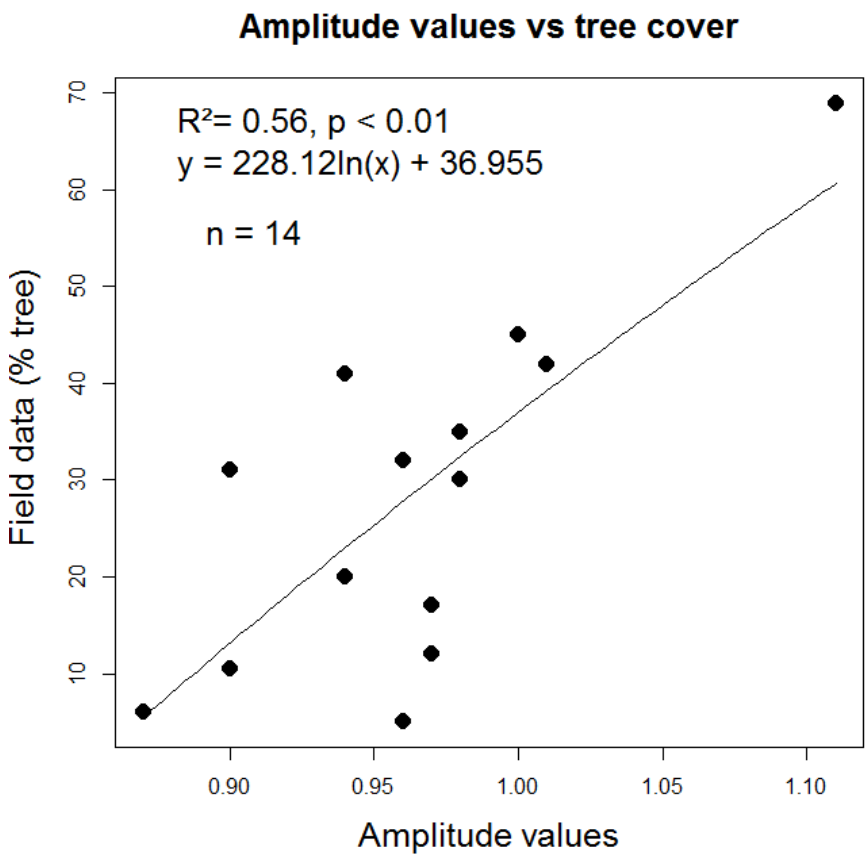


Figure 5. 4: Relationship between field data on tree percent cover and amplitude (1st harmonic term)

5.4. 3 Calibration of $NDVI_{pixel}$ and $SAVI_{pixel}$ for tree cover estimate

Figure 5.5a-f shows the relationship between $NDVI_{pixel}$ and $SAVI_{pixel}$ for the growing season of 2014/2015 and percent tree cover from a field campaign in 2015. The $NDVI_{pixel}$ and $SAVI_{pixel}$ relationships had stronger correlations with the percent tree cover estimated in the field. The $NDVI_{pixel}$ (Figure 5a/c/e) had strong correlation with the percent tree cover ($R^2 = 0.53–0.58$, $p < 0.01$). The relationships that yielded the largest accuracy ($R^2 = 0.56$, $p < 0.01$ and $R^2 = 0.58$, $p < 0.01$) are for linear (Figure 5a) and polynomial (Figure 5c) regression respectively. Although the differences between linear and nonlinear regression (Figure 5e: $R^2 = 0.53$, $p < 0.01$) is relatively small, the result for nonlinear regression reflects the nonlinearity of the NDVI (which normally increases with increasing species composition). In the dry season, certain PFTs especially the herbaceous plants are leaf-out thereby making NDVI more sensitive to woody vegetation. On the other hand, the accuracy of $SAVI_{pixel}$ appear to be better (Figure 5b/d/f) than the $NDVI_{pixel}$ with $R^2 = 0.56–0.67$, $p < 0.01$). There is no doubt since SAVI vegetation index has correction factor such that the effect of soil background reduces. Overall, the positive correlation for both $NDVI_{pixel}$ and $SAVI_{pixel}$ with percent tree cover implies that they could be used as surrogates to percent tree cover at some considerable spatial scales.

Table 5. 2: Calibration of NDVIpixel and SAVIpixel for tree cover estimate

Plot no	**Tree cover**	**Grass cover**	**Bare soil**	**Grass cover plus bare soil**	**Soil types**	**$NDVI_{pixel}$**	**$SAVI_{pixel}$**
1	5	85	10	95	Nitisols	0.336	0.485
3	6	85	9	94	Nitisols	0.289	0.454
4	10.5	45	44.5	89.5	Regosols	0.264	0.434
7	12	67	21	88	Luvisols	0.3492	0.544
9	17	78	5	83	Regosols	0.349	0.499
10	20	70	10	80	Regosols	0.3547	0.520
12	30	45	25	70	Regosols	0.352	0.511
15	31	55	14	69	Luvisols	0.335	0.498
16	32	35	33	68	Luvisols	0.375	0.557
19	35	22	43	65	Regosols	0.341	0.502
21	41	35	24	59	Luvisols	0.389	0.551
22	42	30	28	58	Regosols	0.346	0.515
24	45	45	10	55	Nitisols	0.330	0.562
27	69	20	11	31	Luvisols	0.471	0.691

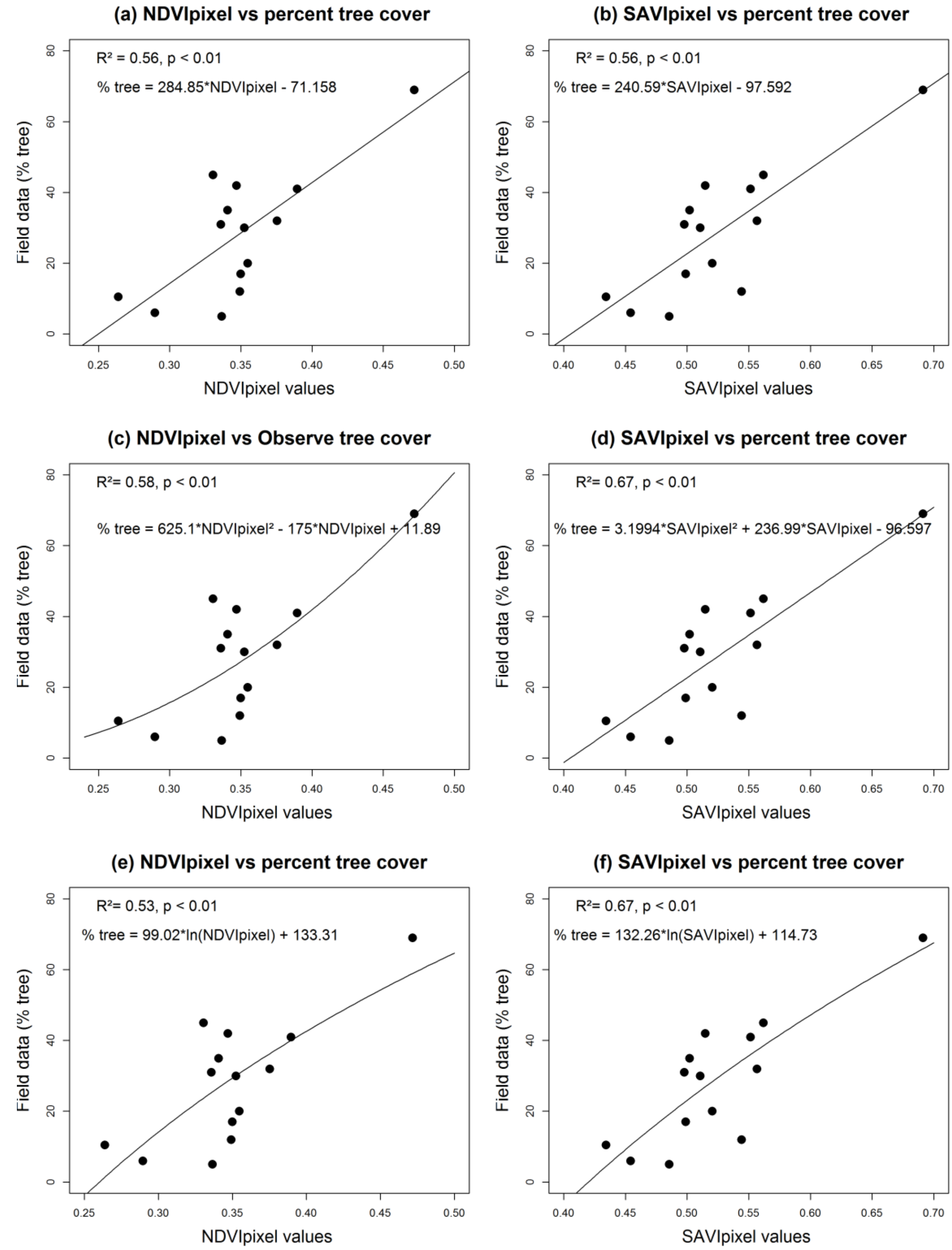


Figure 5. 5: The relationship between $NDVI_{pixel}$, $SAVI_{pixel}$ and field percent cover estimate with regression analyses, with simple linear (a and b), polynomial (c and d) and nonlinear regression (e and f).

5.4.4 Calibration of NDVI and SAVI for tree cover estimate with $NDVI_{soil}$ determining methods using a modified procedure by Zeng et al. and Wu et al.

Table 5.3 shows the field plots data for the percent trees, grasses, bare soil and the type of soil for each of the calibration plots. The table (5.3) also indicates the fraction of the $NDVI_T$ and $SAVI_T$ estimated using the two soil determining methods explained above. Figure 5.4a-d shows the calibration results for $NDVI_T$ and $SAVI_T$ and field data. NDVI estimates for both methods showed an increased accuracy which is much better than when $NDVI_{soil}$ is not removed. The invariant method (Zeng's et al. 2000) for which the threshold of $NDVI_s$ was 0.02 has a strong relationship with percent tree cover: $R^2 = 0.67$, $p < 0.01$ (Figure 5.4a) while the other approach (Wu et al., 2014) which considered the world soil database to determine the $NDVI_s$ for each soil type in our plot locations had also a strong relationship with the percent cover ($R^2 = 0.67$, $p < 0.01$, Figure 5.4c). There is slight difference between the two as only three types of soil were found in the plot locations. And the threshold $NDVI_s$ for Regosols is determined at 0.015 while Luvisols and Nitisols a value of 0.02 was used. This has however made an impact as the accuracy of the relationship has improved.

The invariant soil determining method for SAVI which threshold $NDVI_s$ at 0.05, is not very effective as its accuracy is slightly lower than ($R^2 = 0.50$, $p < 0.01$) the initial relationship for which the $NDVI_{soil}$ is not accounted for. This means that the invariant method applied here may be less accurate in inferring tree cover compared to other approaches although validation results might show otherwise. On the other hand, when soil types are considered in determining the $NDVI_s$, a strong relationship is observed between the SAVI and percent tree cover ($R^2 = 0.80$, $p < 0.01$). In this case, the $NDVI_s$ at 0.04 was threshold for Regosols while Luvisols and Nitisols at 0.05 and 0.06 respectively. Overall, all vegetation indices have shown a good relationship with the percent tree cover.

Table 5. 3: estimates of percent cover using NDVI and SAVI with soil determining methods using a modified procedure by Zeng et al. and Wu et al

Plot no	Tree cover	Grass cover	Bare soil	Grass & bare soil	Soil types	$NDVI_{pixel}$	$NDVI_T$ (Zeng)	$NDVI_T$ (Wu)	$SAVI_{pixel}$	$SAVI_T$ (Zeng)	$SAVI_T$ (Wu)
1	5	85	10	95	Nitisols	0.336	0.290	0.290	0.485	0.404	0.384
3	6	85	9	94	Nitisols	0.289	0.244	0.244	0.454	0.469	0.354
4	10.5	45	44.5	89.5	Regosols	0.263	0.220	0.231	0.434	0.504	0.371
7	12	67	21	88	Luvisols	0.349	0.306	0.306	0.544	0.426	0.466
9	17	78	5	83	Regosols	0.349	0.310	0.320	0.499	0.501	0.440
10	20	70	10	80	Regosols	0.354	0.316	0.326	0.520	0.447	0.463
12	30	45	25	70	Regosols	0.352	0.318	0.327	0.511	0.454	0.461
15	31	55	14	69	Luvisols	0.335	0.302	0.302	0.498	0.451	0.437
16	32	35	33	68	Luvisols	0.375	0.342	0.342	0.557	0.488	0.497
19	35	22	43	65	Regosols	0.340	0.309	0.317	0.502	0.447	0.456
21	41	35	24	59	Luvisols	0.389	0.361	0.361	0.551	0.438	0.499
22	42	30	28	58	Regosols	0.346	0.319	0.326	0.515	0.493	0.474
24	45	45	10	55	Nitisols	0.330	0.304	0.304	0.562	0.578	0.504
27	69	20	11	31	Luvisols	0.471	0.457	0.457	0.691	0.683	0.664

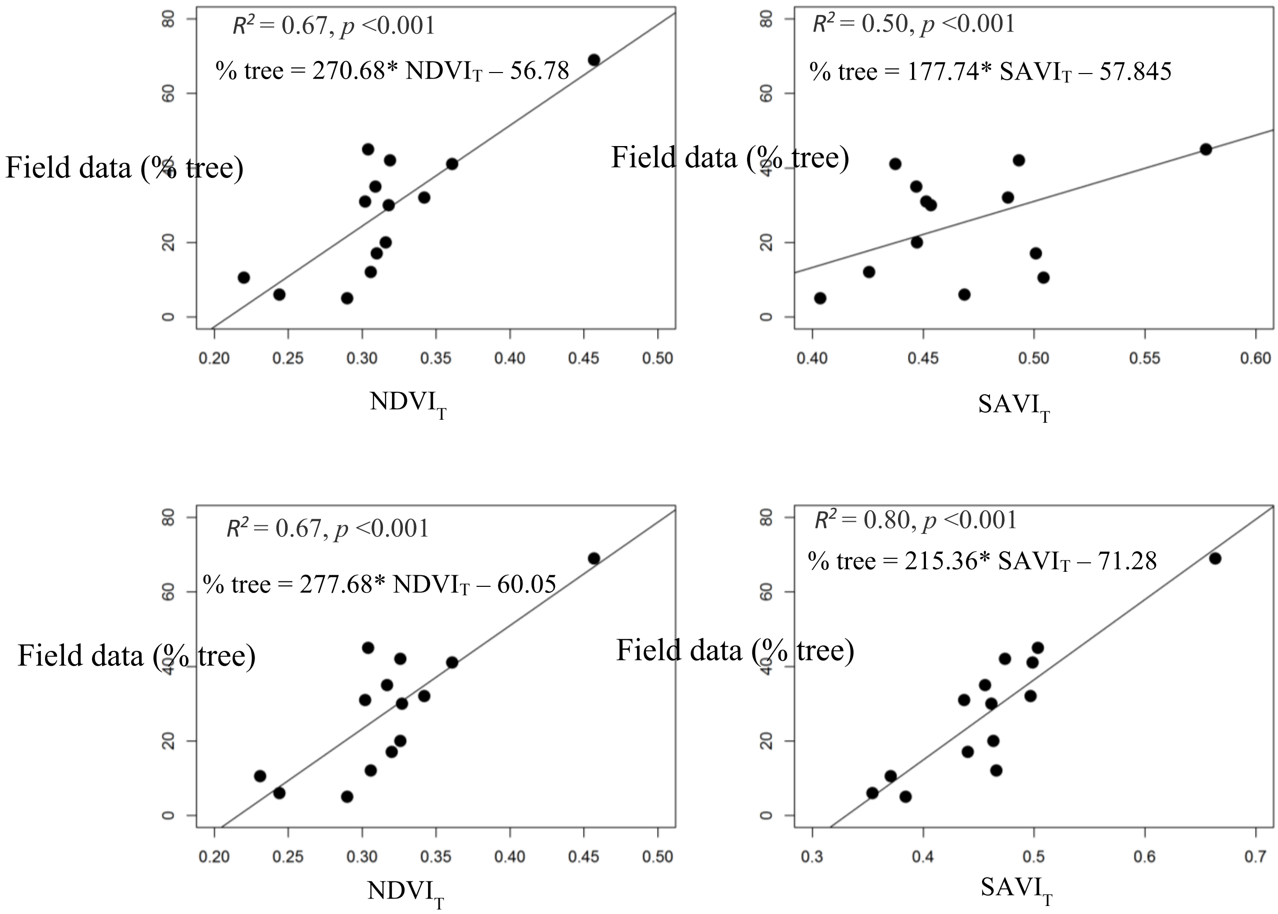


Figure 5. 6: Calibration of $NDVI_T$ and $SAVI_T$ for tree cover estimate with soil determining methods using a modified procedure by Zeng et al. and Wu et al: NDVI (a

and c for Zeng et al. and Wu et al respectively), SAVI (b and d for Zeng et al. and Wu et al respectively).

5.4.5 Calibration of amplitude, $NDVI_{pixel}$ and $SAVI_{pixel}$ for tree cover estimate using multiple regression

Table 5.4 shows that the relationship between the amplitude, $NDVI_{pixel}$ and $SAVI_{pixel}$ with a field data (percent tree cover) had also a high accuracy. Although multiple regression is expected to improve model accuracy due to contribution from many variables, the coefficient of determination obtained in this approach ($R^2 = 0.60$, $p < 0.01$) is slightly lower compared to other models with single variables (not all) such as the SAVI-percent cover relationship. This may be attributable to differences in phenology metrics derived from various indices. However, the amplitude data is useful to the multiple regression model due to interannual variability of tree phenology. While this relationship indicates that percent tree cover can be inferred from these variables, overall model performance evaluated and presented in chapter 6 would show whether this model would have a high/low accuracy than the model variables tested individually.

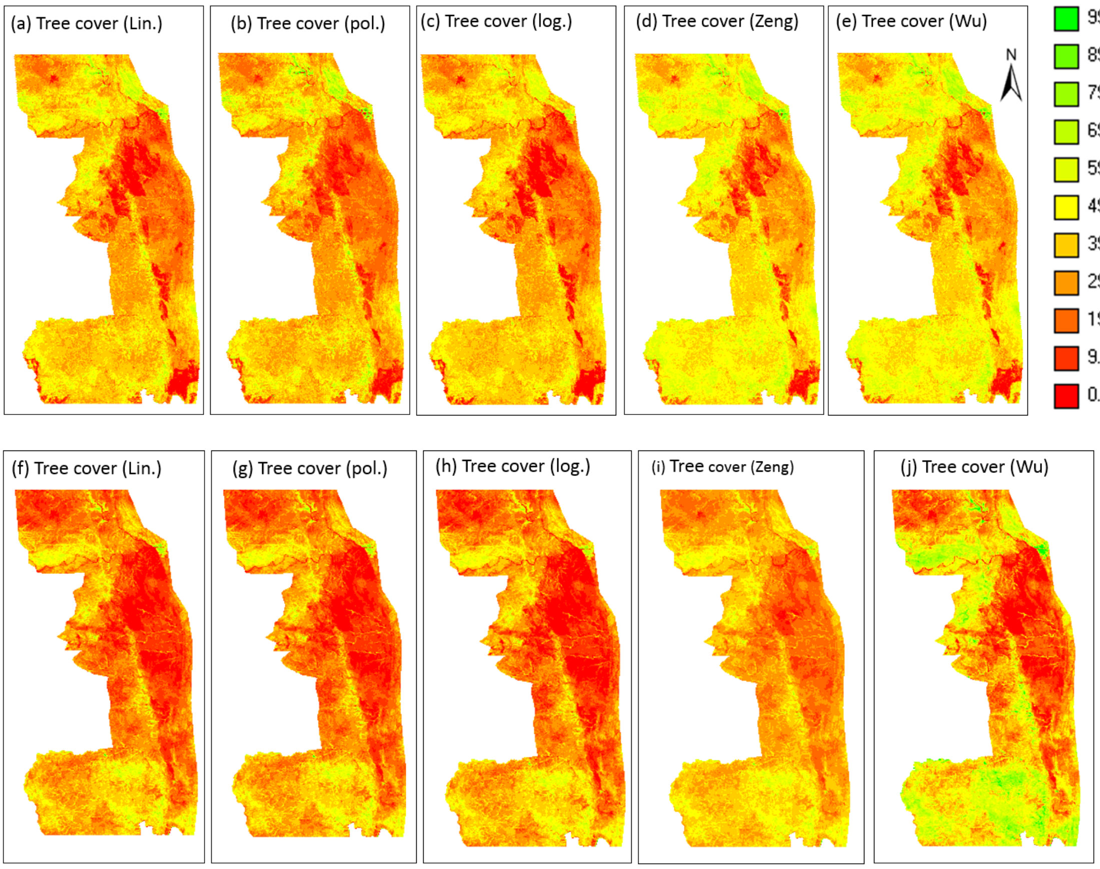
Table 5. 4: Multiple regression coefficient

Regression coefficient		**values**
intercept		-139.67
Slope	Amplitude	75.69
	$NDVI_{pixel}$	-21.88
	SAVIpixel	196.17
R^2		0.60
P value		0.007

5.4. 5 The tree cover maps

Figure 5.5a-l presents the NDVI, SAVI and NDVI harmonic tree cover maps derived from different model calibration presented above. Figure 5m-n shows the LiDAR/SAR (Figure 5m) and Bucini (Figure 5n) tree cover maps. It should be noted however that a comparison of tree cover maps (including the MODIS VCF products) with these previous products has been presented in the Chapter six. All tree cover estimated from our models have demonstrated a consistent pattern of tree cover distribution in KNP. Visually, the distribution of these tree cover maps is consistent with the geological formation of the region. The presence of high and low tree cover cut across various landscapes in the region. The difference between the east and west division is apparent with the western part supporting in general high tree cover. For example, the high tree cover in the Granitic plains with *Terminalia sericea* tree savanna landscape in the extreme south and very low tree cover in the Basaltic plains or Rhyolite Mountains with *Combretum apiculatim landscape* or *Colophospermum mopane* bush savanna in the east. There are however little pockets of high tree cover in the eastern part.

While estimates of tree cover from all models as well as for all vegetation indices indicated the known heterogeneity in the distribution of tree cover over KNP, there are differences between them. Tree cover derived from $NDVI_{soil}$ (Figure 5.5d/e) and $SAVI_{soil}$ (Figure 5.5i/j) determining methods appear to have dense tree cover compared to other maps. Generally, the estimates from the NDVI (Figure 5.5a/e) is high compared to SAVI (Figure 5.5f/j). The estimate from the amplitude (average of 14 year) (Figure 5.5k) is slightly different from the NDVI and SAVI which is due to differences in the phenology of tree species in 14 years as represented by the amplitude data.



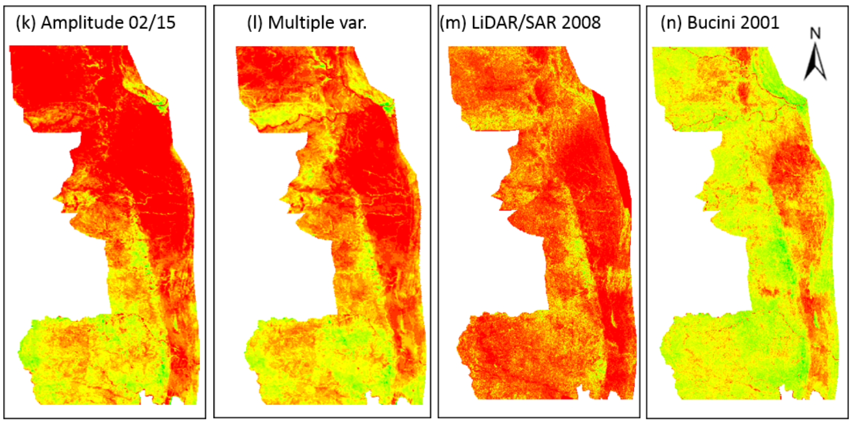


Figure 5.7: The tree maps, NDVI tree cover maps: with linear (a), polynomial (b), nonlinear regression (c), Zeng's et al. (d), Wu et al. (e), SAVI tree cover maps: linear (f), polynomial (g), nonlinear regression (h), Zeng's et al. (i), Wu et al. (j), amplitude (k), Mul.var. (l), LiDAR/SAR (m), Bucini (n). Tree cover maps are derived from the regression equations of each method presented in the data calibration above

5.5 Discussion

This chapter estimated the percent tree cover using phenology metrics derived from MODIS data, firstly, by using the amplitude (harmonic analysis), mean NDVI and SAVI data of the dry season ($NDVI_{pixel}$ and $SAVI_{pixel}$) and through the signal decomposition to account for $NDVI_{soil}$ and $SAVI_{soil}$ in KNP. The estimated tree cover maps were obtained by calibrating the remote sensing data and field measurements using regression analyses (simple linear, logarithmic and polynomial regression).

Although there are limited plots for model calibration, results in this study showed strong relationships between the vegetation indices and field data on percent tree cover collected in 2015 (Figure 5.4-5.6 and Table:5.4). It should be noted however that the accuracy of the models differs with respect to the type of regression analysis and vegetation indices used. Validation results for these models were presented in the next chapter (chapter six). Visually, the spatial pattern of tree covers as shown by most of these models, and to large extent, appear to be consistent with ecological condition in KNP (Figure 5.7).

The strong relationship between averaged tree phenology of a 14-year MODIS NDVI determined by Fourier analysis and field plot measurements indicated the usefulness of harmonic analysis in capturing the interannually variability of PFTs in savannas. The strong relationship found with the amplitude values of the first harmonic term to percent tree cover (Figure 5.4) may be attributable to tree species being more resistant to harsh environmental conditions. This may be because of the influence of site, climatic conditions and disturbances (e.g. fire, drought) on the productivity of each PFTs (House et al., 2003, Archibald and Scholes, 2007b, Moreno-de las Heras et al., 2015, Kaduk and Los, 2011). The results demonstrated that amplitude values can be used as surrogates for tree cover estimation more broadly to capture the spatial variability of heterogeneous savannas (Figure 5.3 and 5.4, 5.7).

The vegetation indices and percent tree cover relationships are strong for both $NDVI_{pixel}$ and $SAVI_{pixel}$ (Figure 5.5a-f). Even though logarithmic (nonlinear) model had the least accuracy (Figure 5.5e) for the $NDVI_{pixel}$, the differences between the models are not very significant, in essence, the spatial variability is well noticeable from the tree maps (Figure 5.7). The $SAVI_{pixel}$ appear to be better than the $NDVI_{pixel}$ in all models (Figure 5.5b/d/f). From the calibration results obtained for the linear model in soil determining methods, the trend of the linear regression is greatly improved for the NDVI and SAVI (Figure 5.6a/c) except for the invariant method (Figure 5.6b) in SAVI. This means that the NDVI vegetation index is more sensitive to correction than the SAVI except for an invariant method. Despite improvements, from the visual assessment, it is evident that the soil determining methods have a high estimates of tree cover in KNP for both NDVI and SAVI. The estimate of fractional vegetation cover usually underestimates where ever $NDVI_{soil}$ is overestimated and the exact opposite occurs when the $NDVI_{soil}$ is underestimated (Montandon and Small, 2008). It is difficult to conclude from the visual inspection as to whether soil determining methods overestimate tree cover or not. At plot level, in this case, the validation results may show if there is overestimation (Chapter six). Validation results presented in chapter six indicated the RMSE for each model.

5.6 Summary

In this study, tree cover was estimated from MODIS data phenology metrics (NDVI amplitude, mean NDVI, SAVI). Both spatial and temporal scales were considered for model development. The percent tree cover maps have reflected spatial and temporal variability in many sites, probably due the influences of biophysical processes such as climate, geomorphology, and disturbances (e.g. fires). Moreover, validation of these maps will show whether the approaches employed in this study are sufficiently robust for tree cover estimation in KNP. Validation of these maps with the remaining half of the field data was shown in the next chapter (six). In addition, the comparison of the estimated tree cover with previous woody cover maps derived from high-resolution data was also presented in the same chapter (six).

Chapter 6

Validation of tree fractional cover map derived from phenological signal decomposition of MODIS time series data

6.1 Introduction

All estimates of land cover types including plant functional types (PFTs) using remote sensing data have an associated error and uncertainty of an unknown magnitude (Rocchini et al., 2013). It is therefore, one of the ultimate goals of many remote sensing studies to provide information that is accurate. To achieve this, an assessment of errors and uncertainty needs to be made to ascertain the level of accuracy of a given remote sensing product.

Although approaches to tree cover modelling differ, careful considerations to the choice of efficient method, data, and field sampling protocols are needed to reduce uncertainty in the input parameters. For instance, mapping tree cover in savanna using Boolean logic has been considered inappropriate given conditions to savanna as gradual transitions between open and closed shrub and grasslands (Gessner et al., 2008, Rocchini et al., 2013). A sub-pixel analysis is being considered as an appropriate method especially in heterogeneous regions such as savannas (Foody and Cox, 1994: Gessner et al., 2008). Proper field data collection is an essential requirement since the accuracy of tree cover estimate is affected by the nature of the landscape formation (Herrmann et al., 2013).

Due to limited field measurements, and high costs, as well as the volume associated with high-resolution datasets, a variety of modelling approaches, have relied on the synergy between the field samples and optical imagery (multi-spectral and multi-temporal data). Thus, the usefulness of these varied modelling outputs lies closely with their accuracy. The essence of quantifying uncertainties resulting from those processes has never been more pertinent. Specifically, many previous studies have resorted to the use of the high-resolution data for validation (Gill et al., 2009, Hansen et al., 2000, Sexton et al., 2013).

While quantifying uncertainties resulting from these estimates might enhance proper assessment of the effectiveness of a given technique, the quality and robust nature of the procedure employed in the process are even more critical. One of the useful methods is the adoption of statistical observations that are unbiased and also, make comparisons with maps and other data sources (Strahler et al., 2006). Several statistical techniques are essential for efficient tests of correct agreement between different observations. In this Chapter, the accuracy of the tree cover maps produced from the NDVI, SAVI, and the MODIS VCF were evaluated. However, all datasets including the LiDAR/SAR product were also validated using a field data from a field campaign in 2015. The LiDAR/SAR product was previously validated and found to be highly accurate. However, it has been resampled for this study. This aggregation made to the product means that its accuracy should be evaluated. The study here asked the following questions:

6. How accurate can a signal decomposition model of tree fractional cover estimates using derived satellite phenology metrics be?

6.2 Objectives

- To compare MODIS NDVI, SAVI, and MODIS VCF tree cover maps with existing tree cover maps.
- To test the performance of signal decomposition for the estimates of tree cover maps (MODIS NDVI, SAVI tree cover maps) using an observation data from a field campaign in 2015.

6.3 Material and method

6.3.1 Assessment of model performance

The assessment of model performance for fractional tree cover from the MODIS NDVI/SAVI time series data uses the remaining half of the field observed data (14 plots) not used for calibration. The LiDAR/SAR-based tree cover map and the MODIS VCF datasets were first compared to the field observations to quantify their accuracies. The validation of MODIS VCF with field data uses the MODIS VCF data for the year 2014 since the field campaign was in 2015. To assess model performance for tree cover estimated in this study, the coefficient of determination (R^2) was used to measure the strength of the relationship between the predicted and the observed values. The predicted data for each model is taken as the independent variable while the observed as the dependent as explained in Piñeiro et al. (Piñeiro et al., 2008). In addition, the root mean square error (RMSE) was used to determine the goodness-of-fit.

6.3.2 Comparison of tree cover estimates with LiDAR/SAR and Bucini woody cover maps

Pearson correlation which is the measure of the linear dependence (correlation) between two variables X and Y, was used to assess whether tree cover estimates (from this study) as well as the MODIS VCF product (Y) are related to LiDAR/SAR and Bucini woody cover maps (X). The significant of the relationship was also assessed at alpha (α) value of 0.05.

6.4 Result

6.4.1 The tree cover maps

Figure 6.1 a-k presents a comparison of maps between the LiDAR/SAR, Bucini woody cover map, MODIS VCF, NDVI harmonic, $NDVI_{pixel}$, $SAVI_{pixel}$ and tree cover resulting from the multiple variables as explained in Chapter five. Visually, the spatial variability is present in NDVI harmonic, $NDVI_{pixel}$, $SAVI_{pixel}$ and tree cover resulting from the multiple variables and that the maps have consistent patterning with LiDAR/SAR product (Figure 6.1a) as well as for Bucini woody cover map (Figure 6.1b). The presence of high and low tree cover cut across all the landscape in the region. The difference between the east and west division is apparent with the western part supporting high tree cover. There are however little pockets of high tree cover in the east part. The Bucini woody cover (Figure 6.1b) (2001) depicts a dense woody cover in the KNP. The difference between Bucini and the LiDAR/SAR (2008) could be due to phenology changes which occur over the years (7-year period) as well as the differences in the datasets for which the two maps were produced. The Bucini woody cover is produced from the synergy between the optical (Landsat ETM+ scenes for 2000 and 2001) and SAR (1995 and 1996) datasets. The MODIS VCF (Figure 6.1c/d) estimates of tree cover is well below 30% resulting probably due to its model which limited tree cover at certain minimum height (> 5 m). It is therefore obvious since savannas are generally less dense with tree height greater than 5 m.

The estimates from the NDVI Harmonic (Figure 6.1e) especially the amplitude indicated the known heterogeneity in tree cover over KNP. It has clearly demarcated the differences between the west and eastern parts. Both $NDVI_{pixel}$ and $SAVI_{pixel}$ tree cover maps have shown almost a similar trend with LiDAR/SAR and Bucini woody cover maps. These estimates closely replicate the vegetation structure of the region as a combination of both dense and sparse tree coverage. The distribution of tree cover estimated with the amplitude might also result due to influence of precipitation in KNP. The southern parts have high rainfall than the extreme north parts of the study area (Figure 6.2). Figure 6.2

shows the comparison of pixel values for percent tree cover and mean annual rainfall for weather stations in KNP (14-year weather station data at: Mahlengeni = 410, mm, Satara = 475 mm, Skukuza = 620 mm) approximated to 400 - 600 mm of three weather stations in KNP with LiDAR/SAR, amplitude, Bucini, NDVI, SAVI, and MODIS VCF. The estimates from the $NDVI_{pixel}$ and $SAVI_{pixel}$ showed a clear distinction between the dense thickets and an open shrub to grassland. Since most of these tree cover maps are phenology based, it should be noted that apart from geology, several factors may influence the density of tree cover in this region. For example, fire, herbivore and weather variability can also have a strong impact on the density of tree cover in KNP. Moreover, differences between tree cover maps are also obvious.

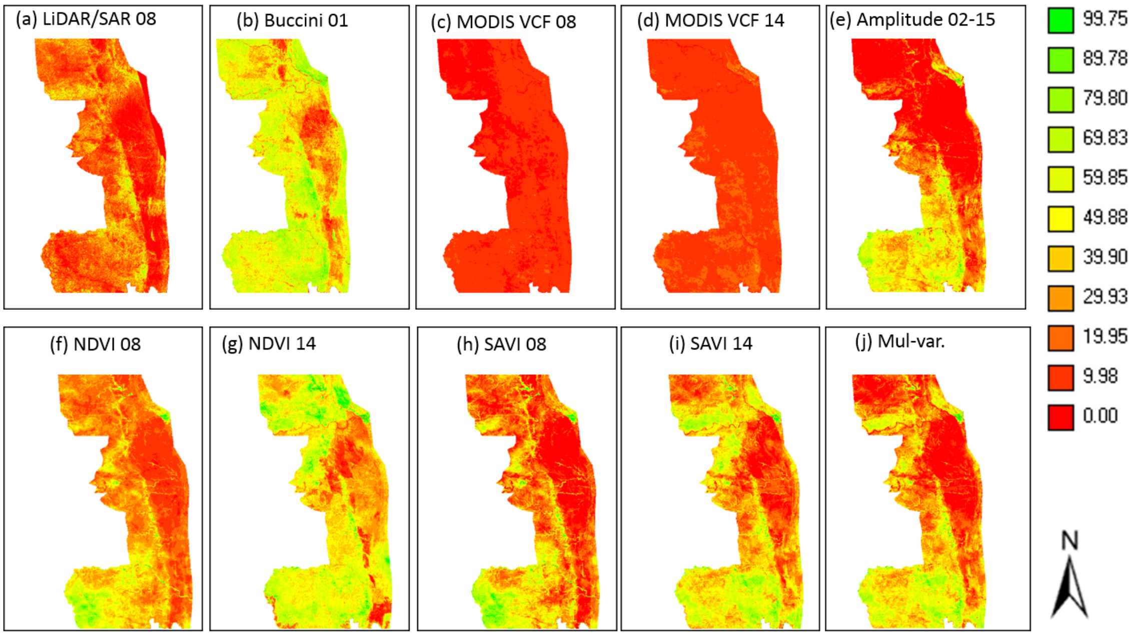


Figure 6. 1: A comparison of maps (a) LiDAR/SAR, (b) Bucini woody cover map, (c) MODIS VCF (2008), (d) MODIS VCF (2014), (e) Amplitude tree cover, (f) NDVI tree cover (2008), (g) NDVI tree cover (2014), (h) SAVI tree cover (2008), (i) SAVI tree cover (2014), (j) Mul-var. tree cover map.

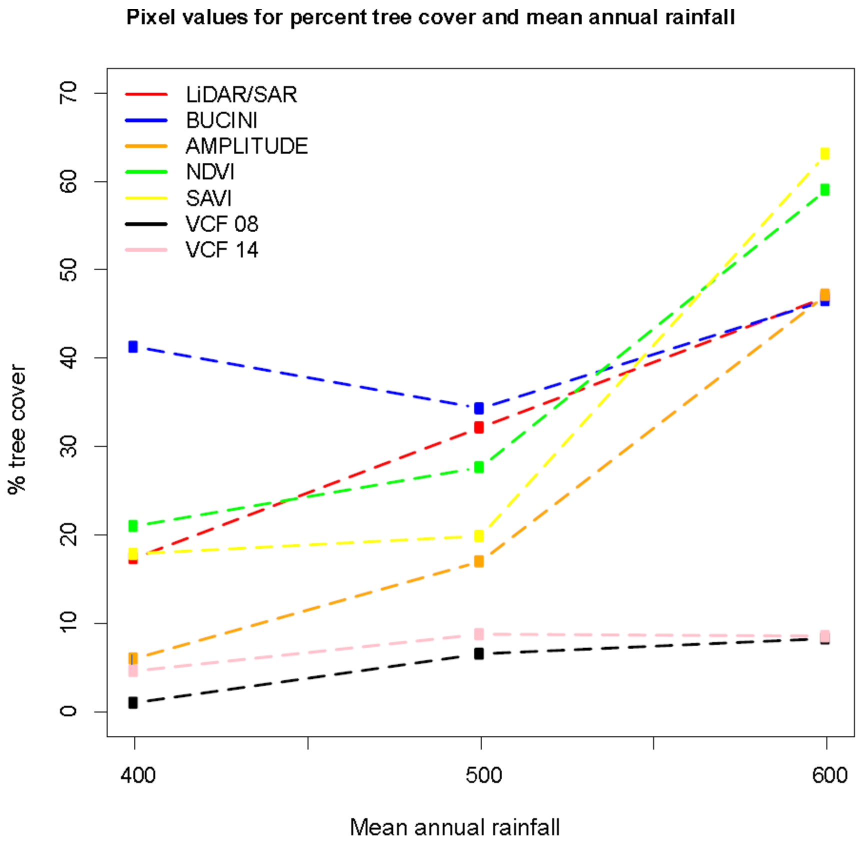


Figure 6. 2: comparison of pixel values for percent tree cover and mean annual rainfall (an average of a 14-year weather station data) data of three weather station in KNP with LiDAR/SAR, amplitude, Bucini, NDVI, SAVI, and MODIS VCF

6.4.2 Regression of percent tree against mean annual precipitation data (of a 14-year) obtained from three weather stations

Table 1 shows the relationship between the tree cover and mean annual precipitation (14 years) of the KNP. Overall, the tree cover has a positive relationship with precipitation in the study area but not significant ($p < 0.001$) for all tree cover datasets except the amplitude.

Table 6.1: Regression of percent tree against precipitation

	LiDAR/SAR	Bucini	Amplitude	NDVI	SAVI	VCF14	VCF08
R^2	0.95	0.37	0.99	0.97	0.92	0.50	0.50
***P* value**	0.13	0.57	0.02	0.09	0.17	0.49	0.49

6.4.3 LiDAR/SAR, Buccini and MODIS VCF tree cover with field data

The accuracy assessments for the LiDAR/SAR, Bucini and MODIS VCF tree cover maps with field data collected in 2015 (28 plots) are presented in Figure 6.3a/b. The assessment for MODIS VCF was made in two ways: first, the assessment was made with reference to all tree canopies regardless of height. Secondly, only tree cover greater than

5 m was considered as MODIS VCF was calibrated on tree of certain minimum height. The first assessment with field estimates indicated that the LiDAR/SAR tree cover map has the highest accuracy ($R^2 = 0.45$, $p < 0.001$, Slope = 0.5, RMSE=15.90%) followed by Bucini ($R^2 = 0.48$; $p < 0.001$, Slope = 0.5, RMSE=17.54%) compared to MODIS VCF ($R^2 = 0.53$, $p < 0.001$, Slope = 0.05, RMSE = 27.5%). The difference between these tree cover estimates is more obvious in the RMSE and Slope (Figure 6.3a). For field data on tree cover above 5 m, the MODIS VCF had better accuracy with the RMSE ($R^2 = 0.19$, $p < 0.02$, slope = 0.1, RMSE = 7.03) (Figure 6.3b). However, with reference to slope and intercept (slope =0.12, intercept = 8.9) the MODIS VCF significantly underestimated tree cover in all cases compared to the field data (Figure 6.3b).

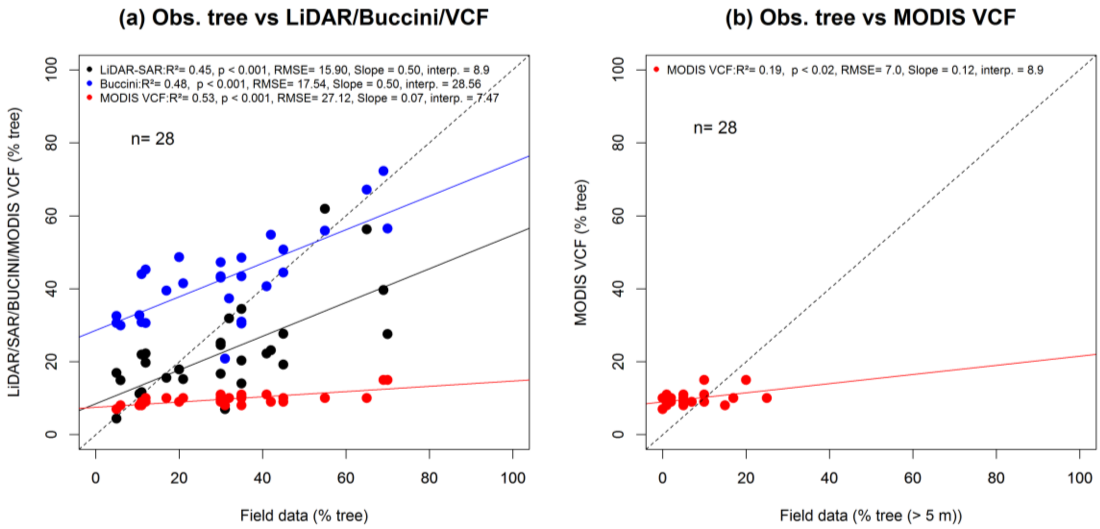


Figure 6. 3: (a) for the LiDAR/SAR tree cover map Bucini woody cover map (2001) and MODIS VCF using field plots (all canopies), (b) MODIS VCF using field plots (for tree cover > 5 m). The dashed line is the 1:1 line.

6.4.4 Validation of MODIS NDVI Harmonic Tree cover map

Table 6. 2 shows the field validation plots and their corresponding amplitude (1st harmonic) and estimated tree cover. Figure 6.4 presents an accuracy assessment of tree cover from amplitude using the field data from a field campaign in 2015. The MODIS NDVI Harmonic tree cover estimated from the amplitude has an $R^2 = 0.36$, $p = 0.03$, slope = 0.83, with RMSE = 16.28%. The challenges of estimating tree cover with amplitude increases with the complexity of species diversity over a large area.

Table 6. 2: Validation of MODIS NDVI harmonic tree cover map

plot	Tree cover	Amplitude	Estimated tree cover (amplitude)
1	5	0.945	24.14
3	11	0.935	21.64
4	11	1.032	44.21
7	12	0.905	14.30
9	21	1.017	40.77
10	30	0.920	17.84
12	30	0.939	22.59
15	35	1.007	38.60
16	35	0.953	26.00
19	35	0.934	21.49
21	45	0.980	32.24
22	55	0.958	27.08
24	65	1.050	48.16
27	70	1.130	64.75

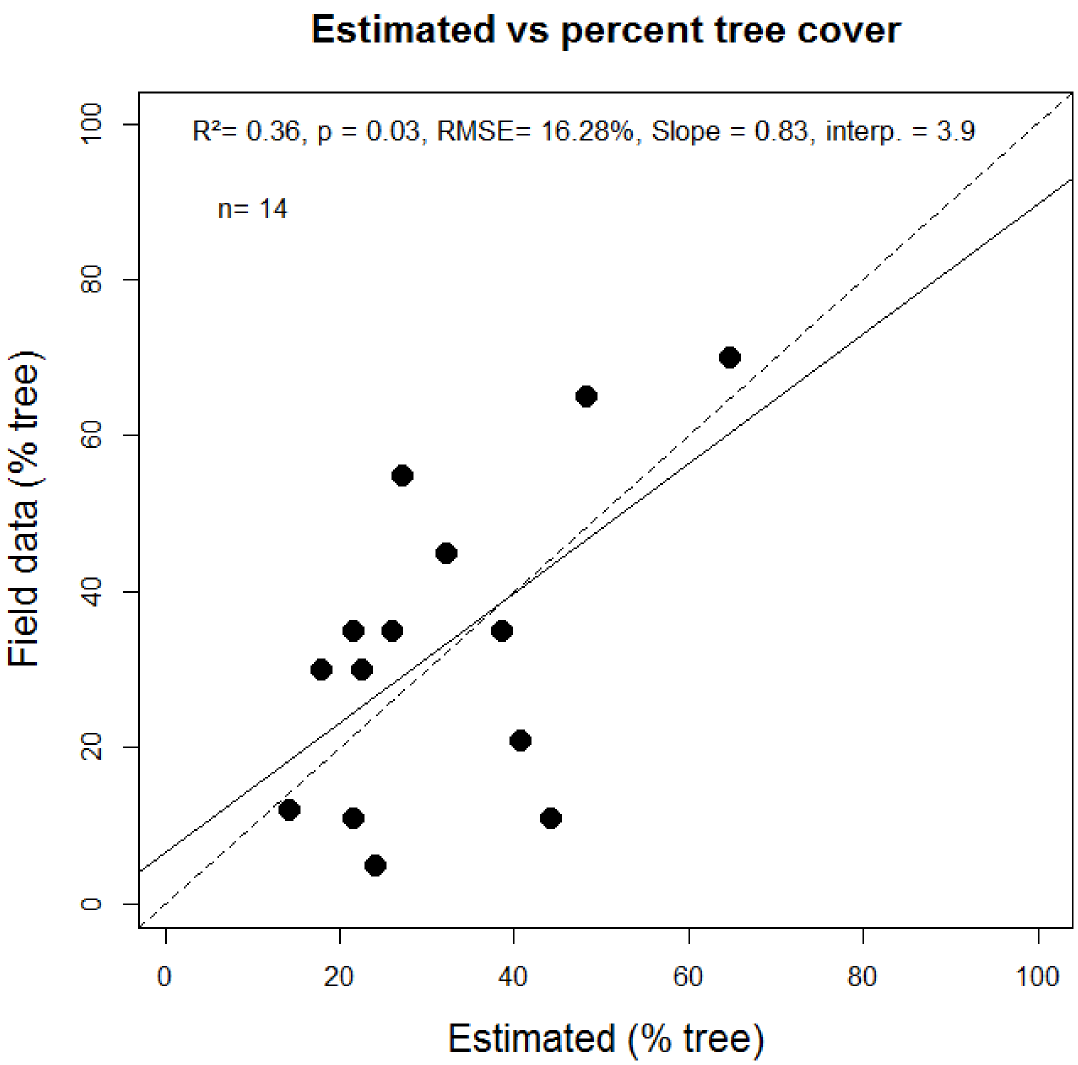


Figure 6. 4: Assessment of model performance of tree cover estimated from harmonic analysis using field data (14 plots). The dashed line is the 1:1 line and the solid line is the regression line

6.4.5 Validation of MODIS $NDVI_{pixel}$ and $SAVI_{pixel}$ tree cover maps

Table 6.3 shows field validation plots and $NDVI_{pixel}$, $SAVI_{pixel}$ as well their corresponding tree cover estimated from simple linear, polynomial and logarithmic regression equations. Figure 6.5a-c presents an accuracy assessments of tree cover from the $NDVI_{pixel}$ and $SAVI_{pixel}$ (mean of dry season images for 2014/2015) using the field data from a field campaign in 2015. The estimated tree cover using linear regression has an $R^2 = 0.40$, $p < 0.01$, slope = 1.01, with RMSE = 15.26% and $R^2 = 0.32$, $p < 0.03$, slope = 0.79, with RMSE = 16.39% for $NDVI_{pixel}$ and $SAVI_{pixel}$ respectively. The level of accuracy for $NDVI_{pixel}$ and $SAVI_{pixel}$ with polynomial regression is not far from the simple linear regression ($NDVI_{pixel}$: $R^2 = 0.40$, $p < 0.01$, slope = 0.89, with RMSE = 15.21%, $SAVI_{pixel}$: $R^2 = 0.32$, $p < 0.03$, slope = 0.78, with RMSE = 15.39%). The logarithmic is slightly less accurate with RMSE and slope for both vegetation indices ($NDVI_{pixel} = R^2 = 0.40$, $p < 0.01$, slope = 0.79, with RMSE = 15.44%, $SAVI_{pixel} = R^2 = 0.32$, $p < 0.03$, slope = 0.82, with RMSE = 16.51%). These results suggested that both $NDVI_{pixel}$ and $SAVI_{pixel}$ are sensitive to percent tree cover during this period and at this vegetation type.

Table 6. 3 Validation of MODIS $NDVI_{pixel}$ and $SAVI_{pixel}$ tree cover estimates

plot no.	**Tree cover (%)**	**$NDVI_{pixel}$**	**$SAVI_{pixel}$**	**$NDVI_{pixel}$ estimated tree cover**			**$SAVI_{pixel}$ estimated tree cover**		
				Linear	Polynomial	logarithmic	Linear	Polynomial	logarithmic
2	5	0.307	0.486	16.17	16.99	16.24	19.35	19.35	19.32
5	11	0.359	0.550	31.21	29.74	31.98	34.80	34.78	35.73
6	11	0.385	0.582	38.42	37.07	38.71	42.43	42.41	43.14
8	12	0.345	0.502	27.08	25.88	27.89	23.19	23.19	23.59
11	21	0.401	0.573	43.04	42.20	42.80	40.23	40.21	41.04
13	30	0.371	0.517	34.47	32.95	35.08	26.69	26.68	27.37
14	30	0.350	0.514	28.41	27.10	29.23	26.13	26.12	26.77
17	35	0.347	0.511	27.78	26.52	28.60	25.41	25.40	26.00
18	35	0.374	0.547	35.24	33.74	35.80	34.08	34.06	35.00
20	35	0.338	0.503	25.26	24.27	26.04	23.50	23.50	23.93
23	45	0.338	0.489	25.22	24.24	26.01	19.99	19.99	20.03
25	55	0.366	0.544	33.20	31.68	33.88	33.19	33.17	34.11
26	65	0.425	0.625	49.99	50.53	48.65	52.84	52.84	52.62
28	70	0.489	0.710	68.22	75.92	62.53	73.12	73.17	69.35

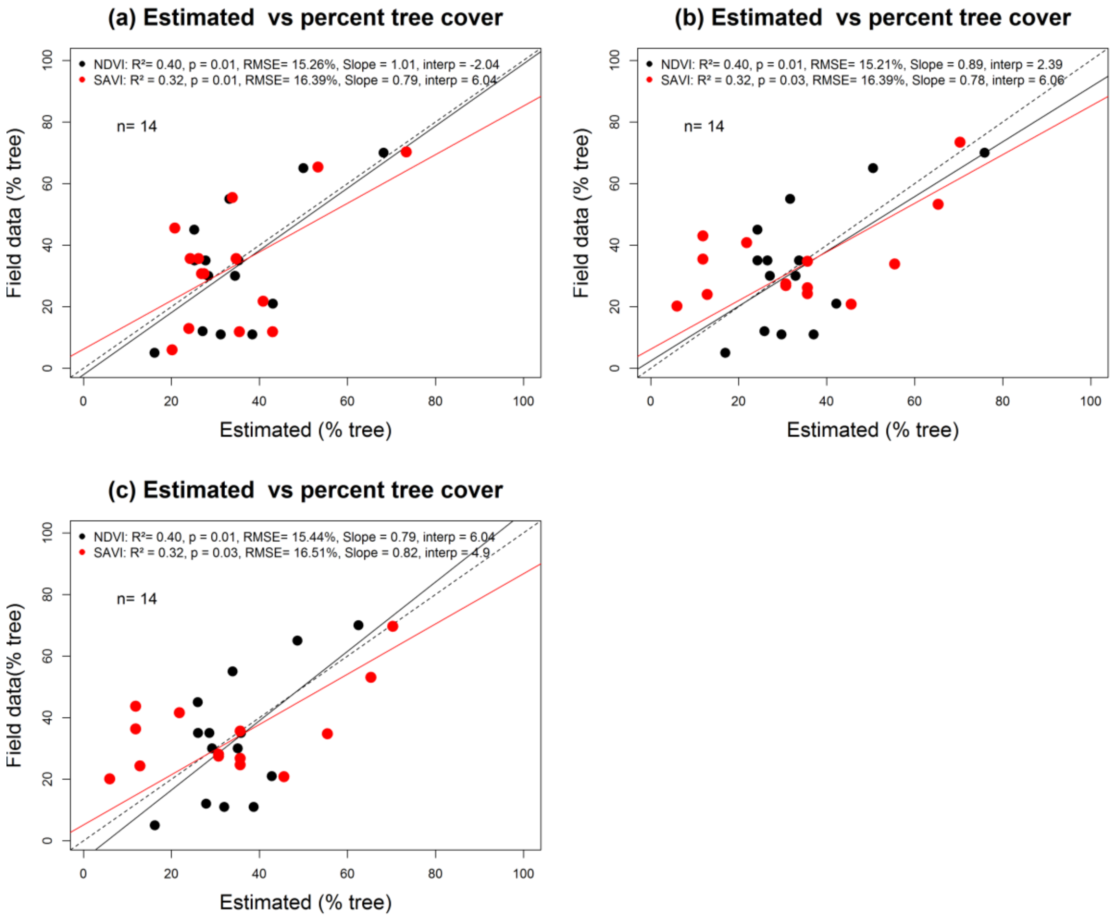


Figure 6. 5: Validation of tree cover estimates derived with NDVIpixel and SAVIpixel using regression analyses, (a) With simple linear, (b) polynomial (c) Logarithmic regressions

6.4.6 Validation of NDVI and SAVI tree cover estimate using $NDVI_{soil}$ determining methods

Table 6. 4 shows the field validation plots and their corresponding tree cover estimated using two approaches which account for $NDVI_{soil}$ and $SAVI_{soil}$ in the estimation. Figure 6.6a/b Shows validation of tree cover estimates from a modified procedure of vegetation fractional estimates by Zeng et al., for NDVI (Figure 6.6a) ($R^2 = 0.40$, $p < 0.01$, slope = 1.06; RMSE = 19.04%) as well as for SAVI (Figure 6.6b) ($R^2 = 0.32$, $p < 0.3$, slope = 1.06; RMSE = 17.34%) vegetation indices. The tree cover estimated for both vegetation indices using Zeng's procedure indicated that the approach can be used to infer tree fractional cover using dry season satellite data even though the accuracy of the estimated tree cover were slightly lower than when $NDVI_{soil}$ were not accounted for. In these methods, there is also an overestimation of tree cover in the lower percent cover (< 30%)

where the contribution of soil is higher demonstrating the implication of an invariant $NDVI_{soil}$ removal approach. Figure 6.6b shows validation of tree cover estimates from a modified procedure for vegetation fractional estimates using $NDVI_{soil}$ determining method by Wu et al., (NDVI: $R^2 = 0.40$, $p < 0.01$, slope = 0.98; RMSE = 18.28%, SAVI: $R^2 = 0.32$, $p < 0.02$, slope = 0.88; RMSE = 19.17%). The accuracy of this approach is slightly high above the previous method. The difference between the two methods is more obvious in the RMSE and slope demonstrating the importance of $NDVI_{soil}$ accounting method that considers soil type characteristics. However, from the graph, it could be seen that there is clustering of few points from 30-40% cover which might not be the influence of soil alone but also the possibility of field underestimation of tree cover or presence of grass layer being active in the dry season at those plots (Figure 6.6a/b). Although the accuracy for $NDVI_{pixel}$ and $SAVI_{pixel}$ is high, the tree cover maps estimated from these approaches as presented in chapter (Figure 5.6) have evidently indicated the well-known pattern of tree cover distribution over KNP.

Table 6. 4 Validation of MODIS NDVI tree cover maps estimated using two methods of soil removal

plot no.	**Observe tree cover (%)**	**$NDVI_{pixel}$**	**$SAVI_{pixel}$**	**Estimated tree cover (%) (Zeng et al)**	**Estimated tree cover (%) (Wu et al)**	**Estimated tree cover (%) (Zeng et al)**	**Estimated tree cover (%) (Wu et al)**
2.00	5.00	0.307	0.486	28.52	28.99	29.382	33.39
5.00	11.00	0.359	0.550	39.94	43.83	40.757	47.17
6.00	11.00	0.385	0.582	45.57	51.16	46.445	54.06
8.00	12.00	0.345	0.502	31.36	32.68	32.225	36.83
11.00	21.00	0.401	0.573	43.95	49.05	44.845	52.12
13.00	30.00	0.371	0.517	33.94	36.04	34.892	40.06
14.00	30.00	0.350	0.514	33.53	35.50	34.358	39.42
17.00	35.00	0.347	0.511	33.00	34.81	33.825	38.77
18.00	35.00	0.374	0.547	39.40	43.14	40.224	46.52
20.00	35.00	0.338	0.503	31.59	32.98	32.403	37.05
23.00	45.00	0.338	0.489	28.99	29.60	29.915	34.03
25.00	55.00	0.366	0.544	38.75	42.29	39.691	45.88
26.00	65.00	0.425	0.625	53.27	61.17	54.088	63.32
28.00	70.00	0.489	0.710	68.25	80.65	69.195	81.63

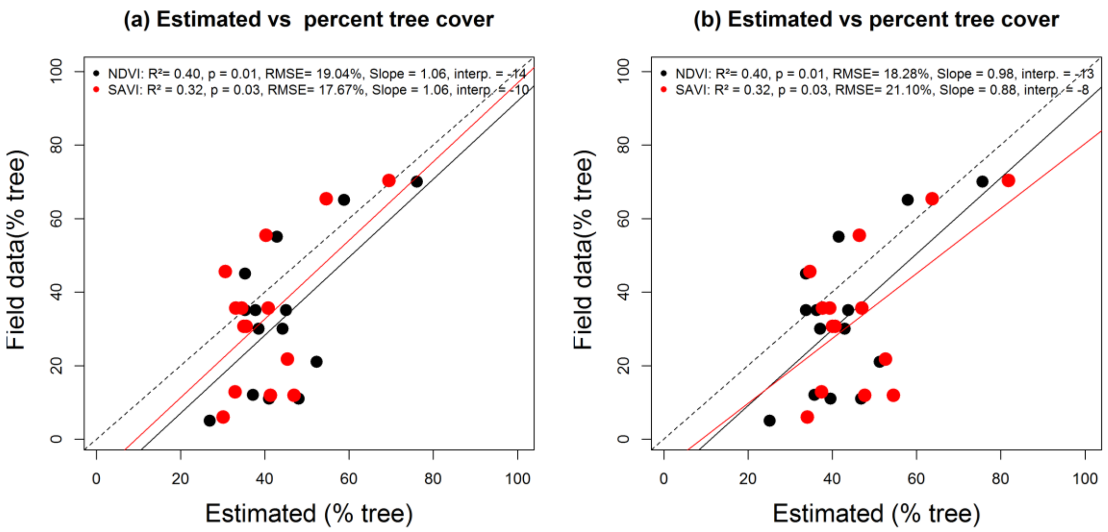


Figure 6. 6: Validation of NDV and SAVI tree Cover estimates (a) modified procedure by Zeng et al. (2000), (b) modified procedure by Wu et al. (2014)

6.4.7 Validation of Tree Cover estimates through multiple regression of Amplitude, $NDVI_{pixel}$ and $SAVI_{pixel,}$

The table 6.5 shows the observed tree cover and corresponding amplitude values, NDVI and SAVI which were used as synergy in a multiple regression model for improved tree cover estimate in KNP. Figure 6.7 shows the accuracy assessment for tree cover estimate from multiple variables. This model produced the highest $R^2 = 0.60$ and slope = 1.2 and had the least RMSE (13.10 %) compared variables investigated individually. The map estimated from these multiple variables as discussed above (Figure 6.1) has similar patterning with LiDAR/SAR and Bucini woody cover maps. The contribution from the NDVI, SAVI, and amplitude in the multiple model is more useful for tree cover estimation.

Table 6. 5: Validation of tree cover estimates through multiple regression of amplitude, $NDVI_{pixel,}$ and $SAVI_{pixel.}$

Plot no.	Observe tree cover (%)	Amplitude	$NDVI_{pixel}$	$SAVI_{pixel}$	Estimated tree cover (%)
2	5	0.95	0.307	0.485	20.66
5	11	0.94	0.359	0.454	12.68
6	11	1.03	0.385	0.434	15.00
8	12	0.91	0.345	0.544	28.37
11	21	1.02	0.401	0.499	26.64
13	30	0.92	0.371	0.520	23.85
14	30	0.94	0.350	0.511	24.06
17	35	1.01	0.347	0.498	26.87
18	35	0.95	0.374	0.557	33.31
20	35	0.93	0.338	0.502	21.80

23	45	0.98	0.338	0.551	35.20
25	55	0.96	0.366	0.515	26.01
26	65	1.05	0.425	0.562	40.75
28	70	1.13	0.489	0.691	70.71

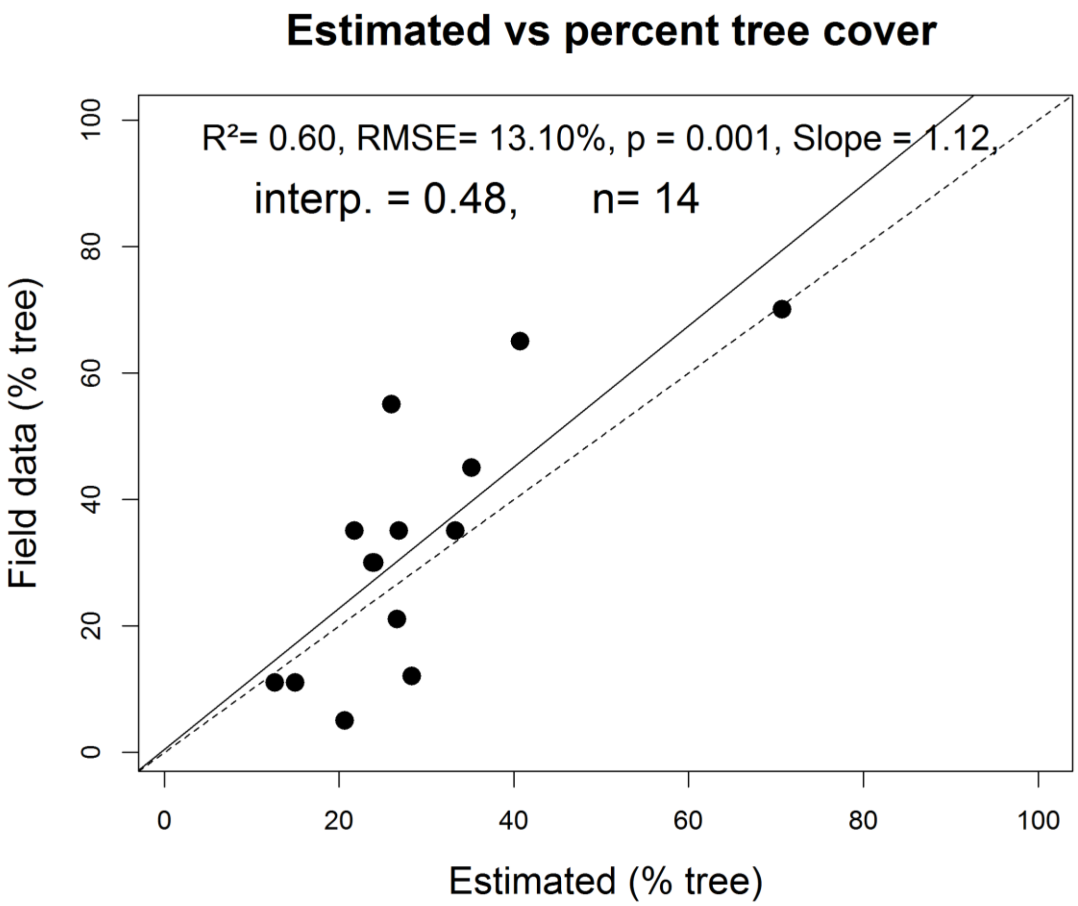


Figure 6. 7: Validation of tree cover estimates through multiple regression of amplitude, NDVIpixel, and SAVIpixel.

6.4.8 Comparison of estimated tree cover with LiDAR/SAR and Bucini woody cover maps using Pearson correlation

Table 6.6 presents a comparison of NDVI, SAVI, and MODIS CVF tree cover maps with LiDAR/SAR and Bucini using 14 validation plots datasets collected from a field campaign in 2015. All vegetation indices have a significant relationship with previous tree cover maps except the polynomial in the NDVIpixel. However, the linear model had a best correlation for both vegetation indices ($NDVI_{pixel}$: r = 0.52, p = 0.05 with LiDAR/SAR and r = 0.63, p = 0.014 with Bucini ($SAVI_{pixel}$: r = 0.53, p = 0.05 with LiDAR/SAR and r = 0.59, p = 0.02 with Bucini). The relationship between MODIS VCF with the previous tree cover maps is however not very significant ($r = 0.39$, $p = 0.16$ with LiDAR/SAR and $r = 0.40$, $p = 0.17$ with Bucini).

Table 6.6: Correlation of estimated tree cover maps with LiDAR/SAR and Bucini woody cover maps

Tree cover estimates	LiDAR/SAR 2008		Bucini 2001	
	r	*p* value	*r*	*p* value
$NDVI_{pixel}$ (Linear)	0.52	0.05	0.63	0.014
$NDVI_{pixel}$ (Polynomial)	0.49	0.07	0.62	0.016
$NDVI_{pixel}$ (Logarithmic)	0.52	0.05	0.63	0.014
NDVI (Zeng's et al.)	0.52	0.05	0.59	0.02
NDVI (Wu et al.)	0.52	0.05	0.59	0.02
$SAVI_{pixel}$ (Linear)	0.53	0.05	0.59	0.02
$SAVI_{pixel}$ (Polynomial)	0.53	0.05	0.59	0.02
$SAVI_{pixel}$ (Logarithmic)	0.53	0.05	0.58	0.02
SAVI (Zeng's et al.)	0.52	0.05	0.59	0.02
SAVI (Wu et al.)	0.52	0.05	0.59	0.02
MODIS VCF	0.39	0.16	0.40	0.17

6.5 Discussion

6.5.1 The LiDAR/SAR, Bucini and MODIS VCF:

The previous products on tree cover used in this study have been validated by the providers and were found relatively accurate (Naidoo et al., 2015, Bucini et al., 2009, Hansen et al., 2003b, Hansen et al., 2005b). The LiDAR/SAR and Bucini woody cover map have been found the most accurate using our field data (Figure 6.3a). Even though MODIS VCF was assessed based on tree cover greater than 5 m (Figure 6.2b), the accuracy from the validation carried out with the field observed tree cover in KNP (slope = 0.1). MODIS VCF showed a moderate correlation with LiDAR/SAR and Bucini woody cover maps (Table 6.5). Despite differences between the time of field campaign (2015) and remote sensing data, the LiDAR/SAR (2008) and Bucini (2001) tree cover maps are consistent with field measurements. (Figure 6.3a). LiDAR/SAR tree cover has the advantage because of its ability to measure vegetation in three dimension (Los et al., 2012, Khalefa et al., 2013). The Bucini woody which is produced from the synergy between optical (Landsat ETM+ and JER-S) and SAR data. The accuracy of the Bucini product could be due to consideration to acquire the SAR images in July-August 2008 (dry season, leaf-off) to avoid soil moisture effects on the radar signal (Mathieu et al., 2013) as well as by using the dry season images for the optical dataset to maximize the discrimination of woody vegetation. The study has also accounted for the effects of climate, soil characteristics, topography, fire frequency and herbivory in a regression analysis to estimate woody canopy (Smit and Asner 2012; Bucini et al 2010).

The accuracy of MODIS VCF datasets when trees height above 5 m are considered (without the shrubs) is higher than when all canopies are considered, especially when reference is made to RMSE error only. The RMSE (7%) found with MODIS VCF in this study is 2% lower than the recent validation of the product carried out by the providers at sites in Maryland (9.47%). Despite strength of MODIS VCF datasets as observed in many studies (Giglio et al., 2006, Los et al., 2012, Sexton et al., 2013, Hansen et al., 2005b), the accuracy of the product is less with savannas particularly when certain statistical observations are put into considerations. For instance, in this study, MODIS VCF had a high RMSE (28.56) and low slope (0.07) when all canopies were considered (Figure 6. 8a). At the same time, a low slope (0.12) and high intercept (8.9) was recorded for the product at tree height of 5 m. This simply means that there is an underestimation of tree or woody species from the MODIS VCF in this region. The underestimation of the MODIS VCF with *in situ* as observed in this study is similar to a recent study (Brandt et al., 2016) whose estimate of woody cover in the Sahel was nine times higher than the MODIS VCF. Consequently, the low accuracy for MODIS VCF has been reported in scientific literatures (White et al., 2005, Gessner et al., 2013, Herrmann et al., 2013).

The use of large number of phenology metrics acquired at different period regardless of vegetation dynamics (Hansen et al., 2002, Hansen et al., 2005b), the presence of bad pixels (cloud cover), the training datasets (regression tree usually require large samples), limitations inherent in the MODIS sensor viewing geometry (the effects is more with the individual bands than the vegetation index-NDVI) and cloud contamination may be responsible for the limitations of the MODIS VCF in savannas due to heterogeneity and the complexity in species diversity. In this study, none of the MODIS VCF (2014) pixel value used for our field plots is of bad quality (with reference to information of quality pixels by the providers). Consideration to the seasonal vegetation dynamics could aid global scale mapping of tree cover in savannas from space. This is due to large differences in vegetation phenology during the wet and dry season (Venter et al., 2003b) and consequent sensor's limitations such as the cloud cover and sensor view geometry which may affect the interannual and seasonal variation of tree/grass phenology (Los et al., 2005, Su et al., 2009).

6.5.2 The NDVI harmonic tree cover map

The harmonic analysis applied to MODIS NDVI data for a 14-year time series was found relatively accurate (Figure 6.4). The estimates of tree cover using phase and cycles from a temporal analysis presented in chapter 4 show the importance of harmonic analysis for tree/grass fractional cover estimates. The amplitude estimated from the MODIS NDVI achieved an overall RMSE of less than 16% (Figure 6.4). The accuracy of tree cover estimated from the amplitude is relatively high with Slope (slope = 0.83) (Figure 6.4). The amplitude of the first harmonic term from the NDVI time-series used in this study was extracted from the composite of images covering various stages of the growing season cycles that could sufficiently reflect phenology features of tree species in this area (Gessner et al., 2013). However, the error resulting from NDVI harmonic might be due to confusion between the woody and grass layer occurring probably due to environmental conditions (Smit et al., 2010). A previous study applied on discrete Fourier analysis to derive a mean-phase-amplitude space to separate six vegetation types into geographic regions using classification with AVHRR data reported confusion in the classification accuracy between grassland and savanna (achieved 23% accuracy). Fourier analysis such as that applied in this study is more robust than traditional methods of analysing single acquisition dates which identify very little information about the phenology of the PFTs. Though tree species usually flush their leaves before the first rain in KNP, there are differences in their leaf-out period. For example, *Acacia spp.* usually starts leaf-flushing earlier than *Combretum apiculatum*. Tree species in KNP usually take 8 weeks to reach full leaf from the date first trees started leafing (Archibald and Scholes, 2007b). Although there are limitations to some of the metrics that capture annual phenology in estimating tree cover because of non-linearity in the NDVI occurring due to presence of tree, grass, and bare soil (Gamon et al., 1995, Jiang et al., 2006, Verger et al., 2009b), such annual composites could perform well comparably to metrics in areas with a dominant phenological profile where common cover types share a common seasonal variation (Hansen et al., 2005b).

6.5.3 The $NDVI_{pixel}$ and $SAVI_{pixel}$

The composites dry season images used for vegetation indices to estimate tree cover in KNP were found to be moderately accurate (Figure 6.5a-c). The NDVI-tree fractional cover relationships had stronger accuracy with the linear and polynomial regression than the logarithmic relationship (Figure 6.5a-c), illustrating the strong dependence of NDVI on tree canopy structure during the dry season in KNP. This also demonstrated the presence of photosynthetic active tree layer in the dry season as previously observed (Bucini et al., 2009). The relationship between the NDVI and PFTs depend on the nature of the ecosystem in question and modelling technique (Los et al., 2000, Los et al., 2005, Gamon et al., 1995).

The relationship between the $NDVI_{pixel}$ and percent tree cover is relatively linear (Figure 6.4 and 6.5). The relationship between the $SAVI_{pixel}$ and percent tree cover is relatively linear (Figure 6.5 and 6.6). The nonlinearity in the NDVI-species relationships increases with increasing species diversity (Wang et al., 2016, Gamon et al., 1995) and darker soil. It also decreases with the sparse vegetation and soil brightness (Jiang et al., 2006). Most of the plots used to estimate tree cover are within granite site where soil types differ compared to the basalt in the east, hence, the results discussed in this section agree with previous studies (Gamon et al., 1995, Wang et al., 2016, Sellers, 1987). A recent study found a linear relationship between FVC and the EVI as well as the SAVI vegetation index. In the same study, the relationship between the NDVI and FVC was nonlinear due to saturation effects at high vegetation fractions due to presence of shadow as well as the influence of soil background (Sousa and Small, 2017).

The tree cover estimated for the KNP and for both vegetation indices show consistent patterning of KNP landscape formation (Bucini et al., 2009, Naidoo et al., 2015). However, the error rate between the $SAVI_{pixel}$ and $NDVI_{pixel}$ is almost similar (Figure 6.5a-c) which further demonstrates less soil variation in this region as $SAVI_{pixel}$ reduces soil effects on canopy reflectance (Huete, 1988). There is strong relationship between the $SAVI_{pixel}$ as well as the $NDVI_{pixel}$ with Bucini and LiDAR/SAR (Table 6.5). From the scatter plots in figure 5 a-c, it can be observed that at the lower percent tree cover, the relationship with field data was not very good (especially for NDVI). The influence of soil background on the signal is therefore evident. This demonstrated the influence of radiative transfer from the surface on canopy reflectance especially where there is mixed tree, grass and bare soil fractions (Fuller et al., 1997, Price, 1990, Baumgardner et al., 1986, Los et al., 2012). The influence of soil reflectance from the NDVI reduces with decreasing canopy gaps (Walter-Shea et al., 1992, Ding et al., 2016).

6.5.4 The tree cover maps estimated using two $NDVI_{soil}$ and $SAVI_{soil}$ determining methods

The assessment of tree cover maps from the soil determining methods indicated a moderate linear relationship between the predicted and the observed percent tree cover (Figure 6.6). Although the slope of the regression line for both vegetation had improved (Figure 6.6) compared with tree estimates with no soil remove (Figure 6.5a), the RMSE is high with these approaches. The SAVI vegetation index was found to be less sensitive to soil removal than the NDVI. SAVI index is one of the vegetation indices specifically developed to reduce soil backgrounds effects. Although soil colour is useful for differentiating soil reflectance (Baumgardner et al., 1986), soil moisture was considered the most important factor in influencing vegetation indices (Muller and Décamps, 2001). From the results in Figure 6.6 and Table 6.4, uncertainties in the estimates of percent tree cover at lower NDVI or SAVI are high for all methods. However, the second approach of soil determining methods had a high estimate in the lower percent cover. This can be explained by the sensitivity of soil background on the NDVI or SAVI or the result of changing canopy structure which might decrease in the NIR reflectance and increasing visible reflectance consequently leading to reduce NDVI. Furthermore, the sensitivity soil backgrounds to vegetation indices was found to be greatest in the lower tree cover than in dense canopy (Huete et al., 1985, Ding et al., 2016). Sometimes larger canopies can have more shadows thereby also reducing the NIR reflectance. These challenges can introduce a substantial uncertainty for tree cover estimation. For example, there can be overestimation of percent tree cover whenever soil contribution is underestimated. The exact opposite is often the case when the soil contribution is overestimated (Montandon and Small, 2008).

Overall, despite uncertainty suffered in the estimation of the percent tree cover with the soil removal approaches used in this study, the heterogeneity is present in tree cover maps. The heterogeneity in the spatial distribution of percent tree cover with regards to areas of high and lower tree cover. Visually, the maps from these approaches appear to be similar to Bucini woody cover maps at global scale. At local scale, using validation plots, the maps show strong and significant relationship with LiDAR/SAR and Bucini woody cover maps (Table 6.5).

6.5.5 Tree cover estimated from the multiple variables

The multilinear model using the amplitude, $NDVI_{pixel}$ and $SAVI_{pixel}$ demonstrated the usefulness of synergy between phenology metrics for tree cover estimation in savannas. This model produced the highest R^2 (0.60) and slope (1.2) and has the least RMSE (13.08 %) compared to all the variables investigated individually. The amplitude (an averaged of a 14 year MODIS NDVI) has been considered useful in this model because of the changes in tree cover due interannual variability (Brandt et al., 2016). $SAVI_{pixel}$ was specifically developed to reduce soil background effects (Huete et al., 1985) and the linearity of the NDVI with the dry season images due a reduced species richness in that season (dry season) (Gamon et al., 1995, Wang et al., 2016).

6.5.5 The uncertainties and sources of errors and proposed improvements

While our results demonstrated the potential of MODIS data to estimate percent tree cover from vegetation indices using signal decomposition, the estimated tree cover in this study has some limitations and remaining uncertainties that must be considered:

i. Phenology

Phenology of PFTs in savannas is usually influenced by many environmental factors (Prins, 1988, Archibald and Scholes, 2007a). Specifically, tree phenology is influenced mainly by temperature and day length (Chidumayo, 2001b), or precipitation and disturbance in certain condition (February et al., 2005). For these reasons, the estimates of percent tree cover from the passive sensor are less accurate compared to active sensor. While active sensor such as the LiDAR can determine the canopy cover by measuring its 3D structure, the estimates from the passive optical sensor mostly rely on the green density of the canopy cover within a pixel (Brandt et al., 2016). Therefore, changes in tree/grass phenology due to high interannual variability (seasonality, fires and drought) in savanna (Bombelli et al., 2009) may have important implication for tree cover estimates. For instance, it has been reported that the grass layer (fraction of photosynthetic vegetation) in savannas may change from 85% to 8% in the dry season and fraction of non-photosynthetic vegetation of the same layer may increase from 7% to 79% in the wet season (Guerschman et al., 2009). In this study, although attention has been paid to defining the dry season which is approximation for the whole study site, yet, tree/grass separation can, therefore, be affected by certain grass species that are supported by soil moisture and temperature (Archibald and Scholes, 2007b, Higgins et al., 2011). In savannas, certain environmental factors favor grass growth and influence its phenology, productivity and biomass allocation (Scholes and Archer, 1997, Scholes, 2003). This might contribute to the overestimation of tree cover in this study in a highly mixed tree/grass area. Furthermore, although the MODIS vegetation indices (e.g. NDVI) are less sensitive to the effect of illumination and viewing geometry than individual bands as previously reported in the literature, the estimates of tree cover in this study may have remaining uncertainties despite being specific to a particular season due to differences in tree/grass structures (Los et al., 2005, Su et al., 2009).

ii. The ground data (field plots data on percent tree/grass cover)

Although the LiDAR/SAR product showed a good slope (0.5) and RMSE of 15.90 as assessed with field data, the accuracy of the product is lower than the previous validation (R^2=0.8 and RMSE=7.7%) (Naidoo et al., 2015). The level of uncertainty in the estimates from the LiDAR/SAR and Bucini woody cover maps is probably due to time gap between these products and the field campaign. Furthermore, the Bucini woody cover map was produced from the Landsat acquired between 2000 and 2001 and the JERS-1 Synthetic Aperture Radar (SAR) scenes (L-band, HH polarization) were acquired between 1995 and 1996 while the field campaign for this study was 2015. The time gap means that there could be significant changes in tree cover in KNP over this period. Despite these differences, The Bucini woody cover appear to show a significant relationship with the field data (R^2=0.48, $p < 0.001$ and RMSE=17.54%).

Moreover, the estimates in this study using field data is also limited due to limited field plot data. It should be noted that the calibration data used in our models may not be the representative of all species over the KNP landscape. The field method for tree cover estimation is also a visual approach which may also constraints accuracy of our model due to remaining uncertainties in the field data collection. However, the results presented in this study demonstrated that percent tree cover can be estimated from vegetation indices in savannas, and that single regression model based on our field data has relatively high accuracy. The accuracy assessments indicated that the RMSE ranges from 15 to 21% for the individual models tested in this study. High uncertainty is attributed to percent tree cover less than 40% where spectral signatures are probably dominated by understorey (dry grass) and soil backgrounds. This arise due to presence of soil and dry grass, underestimation of percent tree cover in the field campaign or changes in tree phenology due to fire over the period (Smit et al., 2010). For both calibration and validation plots, there are only few plots with percent tree cover above 40%. This means that if the regression models are to be developed and applied consistently over the large area, it is important that they are established on a much larger sample than was presented in the current study. This may reduce uncertainty and increase model accuracy. It may also indicate high variability in the level of accuracies (variability) of various models tested in this study.

iii. The $NDVI_{soil}$

The use of *in situ* measurement of soil reflectance remains a crucial step for an effective determination of $NDVI_{soil}$ in a pixel to estimate vegetation cover fraction (Montandon and Small, 2008, Muller and Décamps, 2001, Stoner and Baumgardner, 1981, Smallman et al., 2017) in savannas where vegetation indices at MODIS resolution of 250 m is essential for capturing not only vegetation but also bare soil. One of the biggest challenges for tree cover estimation is the lack of ground measuremnts of soil reflectance since soil reflectance values vary with soil types and characteristics (e.g. soil moisture) in both spatial and temporal resolution.

In this study, as demonstrated from the validation results using different models, quantifying the influence of $NDVI_{soil}$ is challenging without in *situ* measurement (Smallman et al., 2017) of soil reflectance. Though the estimate of tree cover from the linear regression using soil determining methods had better slope (Figure 6.5a/b) for both vegetation indices, the $NDVI_{pixel}$ and $SAVI_{pixel}$ had the least RMSE (Figure 6.4a). The high error in tree cover estimation with soil determining methods can be explained by the fact that the thresholds used to determine the $NDVI_{soil}$ for each pixel were rough estimations. Therefore, tree cover estimation with these approaches might introduce errors especially where the in *situ* measurement of soil reflectance is lacking (Ding et al., 2016). Challenges in tree/grass or soil separation remain critical to model accuracy. However, the methods employed in this study, would have been more accurate if larger environment is considered as some of these approaches are insensitive to a particular land cover type (Zeng et al., 2000). The influence of spectral response pattern of both vegetation and soil can have a strong temporal and spatial effects (Lillesand et al., 2004). The spatial effects may be negligible if small area is being considered (Lillesand et al., 2004). The temporal effects for soil (Smallman et al., 2017) as well as for vegetation is important due to species changes throughout the growing season as well as due to sensor limitations (Los et al., 2005, Los et al., 2000, Kaduk and Los, 2011).

Smallman et al., (2017) who evaluated the critical role of repeated woody biomass estimates in constraining the dynamics of the major ecosystem carbon pools, highlighted the challenges with dead organic carbon stocks and soil using the Harmonised World soil database (HWSD) to account for bare soil. In their estimates of carbon stock, the in-*situ* soil carbon observations have lower uncertainty than the one which used the HWSD, and unlike the HWSD, the in-*situ* data they have used is well constrained in time. They stressed that the impact of the HWSD prior is reduced due to lack of a robust assessment of the uncertainty associated with the database and the lack of information on the time for which the priors are representative, necessitating a conservative use of the database. The in-*situ* measurements of soil reflectance if available would be more useful regardless of the extent of spatial scale being considered for tree cover estimation (Ding et al., 2016, Muller and Décamps, 2001, Smallman et al., 2017).

6.6 Summary

Remote sensing based models of tree cover in the Savanna were developed from vegetation indices (amplitude, NDVI and SAVI of the dry season images) derived from MODIS data and a field data (in situ) on percent tree cover measured at 28 sites in KNP. The models were developed on the understanding that during the dry season only woody species are photosynthetically active. Some of these models had however accounted for bare soil (as non-photosynthetic vegetation) in the estimation. A strong linear relationship was found between the phenology and tree cover observations from a field campaign in 2015. The MODIS NDVI harmonic tree cover estimated from the MODIS NDVI had an $R^2 = 0.36$, $p < 0.03$, slope = 0.83, with RMSE = 16.28%. While tree cover estimated from the dry season MODIS data had $R^2 = 0.40$, $p < 0.01$, slope = 1.01, RMSE = 15.26% and $R^2 = 0.32$, $p < 0.03$, slope = 0.79, RMSE = 16.39% for $NDVI_{pixel}$ and $SAVI_{pixel}$ respectively. The percent tree cover estimated from the soil determining methods had an improved slope for both NDVI and SAVI but yield slightly a high RMSE. The multiple regression model produced with amplitude $NDVI_{pixel}$, $SAVI_{pixel}$ had the highest accuracy:

$R^2 = 0.46$, $p < 0.001$ and slope = 1.2 and RMSE = 13.08 % compared to all variables investigated individually.

The tree cover estimated from all models had a high correlation and significant relationship with the LiDAR/SAR and Bucini woody cover maps. The linear model had a best correlation for both vegetation indices ($NDVI_{pixel}$: $r = 0.52$, $p = 0.05$ with LiDAR/SAR and $r = 0.63$, $p = 0.014$ with Bucini ($SAVI_{pixel}$: $r = 0.53$, $p = 0.05$ with LiDAR/SAR and $r = 0.59$, $p = 0.02$ with Bucini). The MODIS VCF tree cover datasets found to be less accurate compared to field percent tree cover with $R^2 = 0.53$, $p < 0.001$, Slope = 0.05, RMSE = 27.5% for all canopies as well as when certain minimum height (> 5 m) of tree was considered ($R^2 = 0.19$, $p < 0.02$, slope = 0.1, RMSE = 7.03) due to poor slope and low $R^2 = 0.19$. The relationship between MODIS VCF with the previous tree cover maps is not significant ($r = 0.39$, $p = 0.16$ with LiDAR/SAR and $r = 0.40$, $p = 0.17$ Bucini). This implies that MODIS VCF which is calibrated on trees at a certain minimum height (tree > 5 m in height) was only detecting a proportion of the woody cover in KNP. The results presented in this study suggest an improvement compared to previous phenology based maps. The maps of tree cover presented here will be useful in understanding tree/grass interactions in wooded savannas. Future work will have to ascertain the transferability of these methods to savanna sites globally.

Chapter 7

Discussion and Conclusions

7.1 Discussion

This study used harmonic analysis and soil determining methods from vegetation indices in African savanna. MODIS time series data were used to assess the interannual variability of tree and grasses. A range of spatial and temporal scales have been analysed. The pixel values of 28 field plots (collected in 2015) were extracted over 14-year MODIS NDVI time series data. The harmonic analysis presented in chapter 4 uses amplitude, phase and seasonal cycles of trees and grasses to assess annual and interannual variability. Tree and grass fractions were also estimated using phase and cycles as retrieved from the first and second strongest harmonic. The phase and frequency of the strongest harmonic terms are consistent measures for tree/grass discrimination. Tree and grass cover were also estimated from the NDVI and SAVI vegetation indices using soil determining methods. The accuracies and uncertainties of the various models tested in this study were presented.

The statistically significant harmonics were estimated from the decomposition of tree/grass phenology as an average of 14-year datasets and for each phenological year over the study period. The results from decomposition method presented in this study have been useful for change analysis, discrimination and mapping of PFTs in KNP. The statistically significant harmonics using harmonic analysis is important in many ways for reasons being that red noise and correlation of time series data have been reduced. The implication of phenology metrics from the harmonic analysis which are not based on assessment of significant harmonics is likely to provide unreliable result due to their statistical features of uncertainty arising due to presence of high frequencies or correlation of time series which consequetly makes signal separation capabilities challenging (Griffith and Chun, 2016, Ghil et al., 2002, Barbour and Parker, 2014). Previous studies which aimed to assess observations based on time series data using harmonic analysis do not offer a much analytical solution to the error structure of the statistics. The time series observations may be correlated and for example, can inflate type I or II error depending upon the sampling size and the nature of the time series data. The estimated harmonics in such case may often lead to bias, large variance, and spectral leakage which might make one frequency spilling into the neighbouring frequencies, consequently improve higher frequency but distort lower frequency spectra (Barbour and Parker, 2014). In such estimates, it is likely that estimated changes on PFTs occurring at a period do not have occurred at all. It is, therefore, challenging to understand the reliability of the observed changes in the phenological metrics within the expected range of interannual variability since little independent information exists in the observations (de Beurs and Henebry, 2010, Bence, 1995). In this study, the use of Hartley test and correcting for multiple testing with the Bonferroni method as well as the multi-taper approach for estimating phenology metrics have introduced a new and improved technique for analying changes in PFTs (trees and grasses) resulting from environmental conditions and disturbance in savannas.

The changes in PFTs are key indications of climate change and disturbances, but the assessment of interannual variability in tree/grass phenology in savannas, is one of the most significant challenges to facing remote sensing (Rusch et al., 2003, Bradstock and Kenny, 2003, Cleland et al., 2007). Therefore, although there is an increased development of the species-specific phenology models, remote sensing data is offering an immense contribution at regional and global scales. In this study, the assessment of tree/grass phenology using remote sensing data have promoted our understanding in the site-specific differences and peculiarities of varied PFTs. All PFTs assessed with reference to cycles in this study, have shown fourteen peaks over a 14-year data in the strongest harmonic term. While a comparison of tree and grass dominated plots indicated that only tree dominated plots have 28 cycles i.e. two cycles per year in the second strongest harmonic terms with Bonferroni approach. Trees are mostly influenced by the soil moisture resulting from the previous growing cycles. (Scholes and Archer, 1997).

In contrast, the grass phenology has a stronger second harmonic term that does not follow an annual pattern, Similar findings of unimodal phenological pattern and bimodal characteristic have been reported for wheat which span over nine months of growing period before harvesting (Jakubauskas et al., 2002). Canisius et al., (2007) identified 19 cycles of annual pattern of vegetation over 19 years and 38 cycles which cover two seasons far year over the same period. Canisius et al., (2007) stressed the importance of identifying bimodality (e.g. agricultural areas) using biannual signals derived from the harmonic analysis. The technique presented here is an improvement to the previous analysis of time series analysis of remote sensing data. The assessment of bimodality in tree cover distribution has been used as proof that savanna and forest are alternative stable states. The time series analysis of PFTs using harmonic analysis can, therefore, provide important grounds in the possibility of identifying interactions and causal *nexi* between drivers and state variables for various PFTs (De Michele and Accatino, 2014).

The assessment of tree/grass co-existence through the combined field studies and improved remote sensing methods to support empirical analysis could facilitate a better understanding of tree/grass system. This study shows how empirical method with harmonic analysis and field data can be used as a synergy for the estimate of tree cover. Much have been discussed in chapter 5 on the estimates of tree cover using MODIS NDVI and SAVI, primarily when the soil backgrounds and understory in a pixel are being considered. The use of dry season images from MODIS data for the estimate of tree cover through the soil determining methods has demonstrated the difficulties and challenges in dealing with remote sensing data (especially where the field measurement of soil spectral information is lacking). However, it is evident from the empirical relationship between the field estimate of tree/grass fractional cover and the satellite-derived phenology metrics of MODIS (NDVI and SAVI) that remote sensing presents a clear opportunity for assessing tree cover. The regression models used in this study appears to be crucial in the assessment of fractional cover in sparse vegetation type where the relationship between the field data on percent tree cover and vegetation indices is not well-understood due to present of bare soil. While a linear relationship was found between the field data on percent tree cover and for all vegetation indices, there are limitations in most of the models used in this study especially in the lower cover where the contribution of bare soil and understory is likely to be higher (radiative transfer effects- light interaction with canopies and bare soil) (Fuller et al., 1997). However, tree cover is estimated with a considerable accuracy using the field data and had a high correlation with LiDAR/SAR and Bucini woody cover maps. The MODIS VCF product has much lower accuracies when evaluated with field data and previous products. As the MODIS VCF do not consider woody species smaller than 5 m in height, it is therefore limited for landscape decisions in areas of predominantly woody species.

The usefulness of phenology metrics in the frequency domain such as the amplitude, phase and cycles, as they demonstrate sensitivity of PFTs to greening, a measure of separability, inter-annual variability and predictive capability for the wide area for tree fractional cover estimates. This has been confirmed by the use of field data and high-resolution reference dataset (LiDAR-SAR product). From the evidence gathered on tree phenology using Fourier analysis, the selection of dry season NDVI/SAVI from MODIS images to estimates tree fractional cover provides another insight to minimising uncertainty due to influence of understory and soil backgrounds. However, uncertainties in the estimates of tree cover carried out in this study were previously highlighted in the discussion section of each analysis chapter (4,5 and 6). Jointly, results presented in this study using signal decomposition method improves the understanding of time series data from remote sensing and for tree/grass characterisation in an African savanna.

7.2 Thesis conclusion

This study focused on the use of signal decomposition of satellite time series data to assess interannual variability of the main PFTs (trees and grasses). The phenological metrics retrieved from these methods (Fourier analysis and soil determining methods from vegetation indices) were used to characterized PFTs and estimate fractional tree cover in Kruger National Park using MODIS NDVI time series data over 14-year period. The estimate of tree cover was established based remote sensing model of tree cover in the savanna from vegetation indices metrics (NDVI amplitude, NDVI and SAVI of the dry season images) derived from MODIS data and a field data (*in situ*) on percent tree cover measured at 28 sites in KNP. The conclusions drawn from this study were as thus:

- Statistically significant harmonic terms estimated based signal decomposition have revealed a distinct pattern for trees and for grasses. The used of Bonferroni-Hartley tests in Fourier analysis and multitaper method are very useful for the estimate of significant harmonics from the time series.
- Interannual variability of these PFTs assessed from the amplitude, cycles and phase values of the strongest harmonic terms is robust to tree and grass phenology characterization since grasses respond more strongly to the annual seasonal cycle than trees. For the whole study area, estimates from Fourier analysis show changes in the distribution of PFTs in both temporal and spatial domains.
- The phase values indicated that in most cases, trees green up earlier than grasses. However, the study also notes an inconsistent condition in the greening of these PFT over the period especially at annual temporal scale. This relates to site-specific differences, species composition and the differences in the fractional cover of PFTs.

- Tree/grass phenology from Fourier analysis of satellite remote sensing can be used to estimate their fractional covers as the phase has an $R^2 = 0.60$, $p = 0.001$, slope = 1, with RMSE = 12.52% and $R^2 = 0.44$, p = 0.01, slope = 1.2, with RMSE = 17.64% for tree cover and grass cover respectively. The estimates of grass cover had the highest error. The estimates of tree cover are still better with cycles ($R^2 = 0.55$, $p = 0.03$, slope = 1, with RMSE = 16.07%) than for grass ($R^2 = 0.32$, $p = 0.03$, slope = 1, with RMSE = 17.91%). The accuracy assessment of multiple linear regression model for tree cover estimate shows an $R^2 = 0.61$, $p < 0.001$ and slope = 0.99 and had the least RMSE = 12.54 %. The accuracy has increased greatly compared to when models that were assessed with individual variables. The MODIS NDVI harmonic tree cover estimated from the NDVI dataset has an $R^2 = 0.36$, p = 0.03, slope = 0.83, with RMSE = 16.28%.
- While harmonic analysis is robust to estimates of tree/grass phenology, like in many other techniques, it is limited to areas that are not well-known. This is especially the case for the phase and cycles in inferring tree/grass cover. The behavior of PFTs as captured by the frequency domain of the original time series is linked to the influence of site-specific differences and time series characteristics. Thus, for a better use of signal decomposition such as that applied in this study, it is essential to have a priori knowledge of the ecosystem and proper understanding of time series data characteristics.
- Despite the limitations in the Fourier analysis, this study demonstrates how estimating statistically significant harmonics using the Hartley test and correcting for multiple testing with the Bonferroni method is a good option for any remote sensing study of time series, especially for assessing inter-annual variability and estimating fractional cover of PFTs in savannas.
- NDVI and SAVI ($NDVI_{pixel}$ and $SAVI_{pixel}$) vegetation indices from the dry season MODIS dataare found suitable for the estimates of tree cover in KNP due to reduced species richness as grass layer is usually non-photosynthetic in the dry season ($R^2 = 0.40$, $p = 0.01$, slope = 1.01, RMSE = 15.26% and $R^2 = 0.32$, $p = 0.03$, slope = 0.79, RMSE = 16.39% were found for $NDVI_{pixel}$ and $SAVI_{pixel}$ respectively). Tree cover estimated from these vegetation indices agree with previous products ($NDVI_{pixel}$: $r = 0.52$, $p = 0.05$ with LiDAR/SAR and $r = 0.63$,

p = 0.014 with Bucini ($SAVI_{pixel}$: $r = 0.53$, $p = 0.05$ with LiDAR/SAR and $r = 0.59$, $p = 0.02$ with Bucini).

- The percent tree cover estimated from the soil determining methods had an improved slope for the vegetation indices but yield slightly a high RMSE for both methods. The invariant method had an $R^2 = 0.40$, $p = 0.01$, slope = 1.06; RMSE = 19.04% for the NDVI and an $R^2 = 0.32$, $p = 0.03$, slope = 1.06; RMSE = 17.34% for SAVI. While method which take account of soil types had an $R^2 = 0.40$, $p = 0.01$, slope = 0.98; RMSE = 18.28% for NDVI and an $R^2 = 0.32$, $p = 0.03$, slope = 0.88; RMSE = 19.17% for SAVI. Tree cover estimates from these approaches agree with previous products (LiDAR/SAR and Bucini woody cover maps). The high RMSE in the estimate of tree cover based on soil determining $NDVI_{soil}$/$SAVI_{soil}$ removal were applied relied on based on the soil database, could be due to uncertainty in the spatial and temporal precision of the soil types and characteristics over KNP.
- The understanding of soil characteristics remains strong pivotal for an effective soil determination from the NDVI. Although an *in-situ* measurements of soil reflectance is key to successful estimates of tree cover in savannas especially when using vegetation indices, the use of global Harmonized World Soil Database provides cost effective remote sensing approach especially where field measurements of soil reflectance is lacking or impractical.

- The MODIS VCF tree cover datasets were found relatively accurate compared to field percent tree cover with $R^2 = 0.53$, $p < 0.001$, Slope = 0.05, RMSE = 27.5% for all canopies as well as when certain minimum heights were considered ($R^2 = 0.19$, $p < 0.02$, slope = 0.1, RMSE = 7.03). MODIS VCF has large error from the analysis conducted in this study (regardless of height). The low R^2 is found when tested with a data that set limit to tree height (as was the case for MODIS VCF's model calibration). This implies that MODIS VCF which is calibrated on trees at a certain minimum height (tree > 5 m in height) was only detecting a proportion of the woody cover in KNP.
- To reduce uncertainty in tree cover estimates using remote sensing, especially in savannas, requires multi-temporal metrics derived with reference to a specific vegetation phenology.

- This study demonstrated an improvement in tree cover estimates using different phenology metrics compared to previous phenology based maps as savanna site is dominated by the woody species (shrubs and trees). The techniques presented in this study demonstrated a significant improvement and an important contribution to future studies. The maps of tree cover will be useful in understanding tree/grass interactions in wooded savannas.

7.3 Original research contributions

There are several important areas where this study makes an original contribution to knowledge. The assessments of spectral response of PFTs from a time-series of remotely sensed data implemented in this study presents new methodological applications for multi-temporal change analysis in savannas. The techniques applied here present an interesting opportunity for the exploration of many ecological questions regarding the mixed tree/grass system. The harmonic analysis that employed confidence interval to estimate statistically significant harmonics in the spatiotemporal development of trees and grasses as revealed by satellite sensor was evaluated. The study also demonstrated that differences in PFTs (such as the time of greening) due to fractional changes, site differences or species composition could be quantified from remote sensing. The signal decomposition method applied in chapter 4, and 5, together with accuracy assessments presented in chapter 6 demonstrated the effectiveness and usefulness of this study.

The phase, amplitude and the number of cycles are robust discriminators of trees and grasses. The empirical relationships evaluated between the satellite-derived phenology metrics and field data on tree/grass cover provides another insight in savanna remote sensing. This finding is a clear indication that remote sensing technology is providing an important opportunity to advance the understanding of tree/grass structure and their phenological properties. Moreover, the study also signifies the role of remote sensing particularly in identifying PFTs whose phenological properties are defined by resource constraints. In this context, it is observed that the synergy between remote sensing and field observation is more likely to provide a good measure of functional types than is possible from field observations alone (Ustin and Gamon, 2010).

The partitioning of remotely sensed time series data into frequency domain such as the amplitude images provides a very useful parameter for ecological studies. The amplitude images especially the first and second harmonic terms are essential for the management of biodiversity in KNP. This is because the extracted harmonics terms have strong biological meaning and can, therefore, be used in addition to other physical variables to evaluate ecological status. This thesis produced the first KNP tree fractional cover maps at MODIS resolution of 250 m that capture woody species adequately. The availability of robust tree/grass FVC datasets over time would enable new ecological studies of tree/grass coexistence to be carried out. It is important therefore for monitoring woody vegetation trends in KNP. This will have significant ecological implication in changing landscape decisions.

This study has offered a strong insight for the assessments of the main PFTs using signal decomposition which can be further developed for tree/grass characterization using remote sensing application. It is evident from this study that phenology metrics estimated from signal decomposition require proper interpretation and therefore becomes necessary for modelling tree/grass phenology in an African savanna.

7.4 Future research directions

Future work will have to ascertain the transferability of the method to savanna sites worldwide. For instance, more efforts should go into defining the empirical relationships between field data on tree/grass cover and satellite phenology metrics as remote sensing is proven to be promising. Therefore, the study can be further improved by collecting more field samples on tree/grass cover. Research is therefore required to understand the underlying mechanisms and factors influencing these relationships. This is particularly useful as there are no reliable global datasets for trees (which capture woody species adequately) and grass fractions (Boke-Olén et al., 2016).

Specifically, even the estimation of woody cover in this study is limited by field data. The overestimations of woody cover especially in areas of high vegetation result due to lack of coincidence and sufficient field data. However, spatial variability is present in the tree cover maps. Future studies require large amount of field information (on woody cover and soil reflectance measurement) to improve woody cover estimation in KNP.

Although this study was carried out at a local scale, several inconsistencies were found relating to productivity and time of greening of PFTs. The question remains how this variability fit over a large area and what are the mechanisms for these differences. Specifically, there is the need to monitor changes in woody cover with a focus on species composition and their environmental conditions. There is need to test the techniques presented in this study to derive a global woody cover map for savannas. The synergy between remote sensing and field method could be given more attention.

Although MODIS data has proven promising in this study, this method is not specific to a dataset and could be applied to detect and characterize changes in PFTs within other remotely sensed image time series (e.g. Landsat). Future studies could test the possibility of using Landsat data for harmonic analysis because of its spatial and temporal resolution. Newly launched satellite data i.e., the Sentinel-2 can also be tested if the adequate time series data becomes available. Both datasets have advantages because of their pixel resolutions considerably finer than MODIS spatial resolution. This is suitable for identifying the non-stationary dynamic nature of PFTs in savanna ecosystems. These datasets are also applicable for studies at regional and global scales.

Bibliography

ACCATINO, F., DE MICHELE, C., VEZZOLI, R., DONZELLI, D. & SCHOLES, R. J. 2010. Tree-grass co-existence in savanna: Interactions of rain and fire. *Journal of theoretical biology,* 267**,** 235-242.

ALLEN, M. R. & SMITH, L. A. 1994. Investigating the origins and significance of low-frequency modes of climate variability. *Geophysical Research Letters,* 21**,** 883-886.

ALTON, P., NORTH, P., KADUK, J. & LOS, S. 2005. Radiative transfer modeling of direct and diffuse sunlight in a Siberian pine forest. *Journal of Geophysical Research: Atmospheres,* 110.

ANDRES, L., SALAS, W. A. & SKOLE, D. 1994. Fourier analysis of multi-temporal AVHRR data applied to a land cover classification. *TitleREMOTE SENSING,* 15**,** 1115-1121.

ARCHIBALD, S. & SCHOLES, R. 2007a. Leaf green-up in a semi-arid African savanna-separating tree and grass responses to environmental cues. *Journal of Vegetation Science,* 18**,** 583-594.

ARCHIBALD, S. & SCHOLES, R. J. 2007b. Leaf green-up in a semi-arid African savanna separating tree and grass responses to environmental cues. *Journal of Vegetation Science,* 18**,** 583-594.

ASNER, G. P., LEVICK, S. R., KENNEDY-BOWDOIN, T., KNAPP, D. E., EMERSON, R., JACOBSON, J., COLGAN, M. S. & MARTIN, R. E. 2009. Large-scale impacts of herbivores on the structural diversity of African savannas. *Proceedings of the National Academy of Sciences,* 106**,** 4947-4952.

ASNER, G. P., LEVICK, S. R. & SMIT, I. P. 2011. Remote sensing of fractional cover and biochemistry in savannas. *Ecosystem funcfion in savannas: Measurement and modelling at landscape to global scales***,** 195-218.

ATKINSON, P. M., JEGANATHAN, C., DASH, J. & ATZBERGER, C. 2012. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sensing of Environment,* 123**,** 400-417.

ATTOH-OKINE, N., BARNER, K., BENTIL, D. & ZHANG, R. 2008. The empirical mode decomposition and the Hilbert-Huang transform. *EURASIP Journal on Advances in Signal Processing,* 2008**,** 1-2.

AVITABILE, V., BACCINI, A., FRIEDL, M. A. & SCHMULLIUS, C. 2012. Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda. *Remote Sensing of Environment,* 117**,** 366-380.

BALZTER, H., GERARD, F., GEORGE, C., WEEDON, G., GREY, W., COMBAL, B., BARTHOLOMÉ, E., BARTALEV, S. & LOS, S. 2007a. Coupling of vegetation growing season anomalies and fire activity with hemispheric and regional-scale climate patterns in central and east Siberia. *Journal of Climate,* 20**,** 3713-3729.

BALZTER, H., ROWLAND, C. S. & SAICH, P. 2007b. Forest canopy height and carbon estimation at Monks Wood National Nature Reserve, UK, using dual-wavelength SAR interferometry. *Remote Sensing of Environment,* 108**,** 224-239.

BARBOUR, A. J. & PARKER, R. L. 2014. psd: Adaptive, sine multitaper power spectral density estimation for R. *Computers & Geosciences,* 63**,** 1-8.

BAUMGARDNER, M. F., SILVA, L. F., BIEHL, L. L. & STONER, E. R. 1986. Reflectance Properties of Soils. *In:* BRADY, N. C. (ed.) *Advances in Agronomy.* Academic Press.

BEALE, I. F. 1973. Tree density effects on yields of herbage and tree components in south west Queensland mulga (Acacia aneura F. Muell.) scrub. *Trop Grasslands*.

BENCE, J. R. 1995. Analysis of short time series: correcting for autocorrelation. *Ecology***,** 628-639.

BIVAND, R. & LEWIN-KOH, N. 2013. maptools: Tools for reading and handling spatial objects. *R package version 0.8–27*.

BLOOMFIELD, P. 2004. *Fourier analysis of time series: an introduction*, John Wiley & Sons.

BOKE-OLÉN, N., LEHSTEN, V., ARDÖ, J., BERINGER, J., EKLUNDH, L., HOLST, T., VEENENDAAL, E. & TAGESSON, T. 2016. Estimating and Analyzing Savannah Phenology with a Lagged Time Series Model. *PLoS ONE,* 11**,** e0154615.

BOMBELLI, A., HENRY, M., CASTALDI, S., ADU-BREDU, S., ARNETH, A., DE GRANDCOURT, A., GRIECO, E., KUTSCH, W. L., LEHSTEN, V., RASILE, A., REICHSTEIN, M., TANSEY, K., WEBER, U., VALENTINI, R., NATURVETENSKAP, SCIENCE, INSTITUTIONEN FÖR NATURGEOGRAFI OCH, E., GEOCENTRUM, II, GEOCENTRE, II, LUNDS, U., DEPT OF PHYSICAL, G., ECOSYSTEM, S., LUND, U., SECTION OF, B., EARTH, S. & BIOLOGISK-GEOVETENSKAPLIGA, V. 2009. An outlook on the Sub-Saharan Africa carbon balance. *Biogeosciences,* 6**,** 2193-2205.

BONAN, G. B., LEVIS, S., SITCH, S., VERTENSTEIN, M. & OLESON, K. W. 2003. A dynamic global vegetation model for use with climate models: concepts and description of simulated vegetation dynamics. *Global Change Biology,* 9**,** 1543-1566.

BOND, W. J. & KEELEY, J. E. 2005. Fire as a global 'herbivore': the ecology and evolution of flammable ecosystems. *Trends in ecology & evolution,* 20**,** 387-394.

BOND, W. J. & PARR, C. L. 2010. Beyond the forest edge: ecology, diversity and conservation of the grassy biomes. *Biological Conservation,* 143**,** 2395-2404.

BRADSTOCK, R. & KENNY, B. 2003. An application of plant functional types to fire management in a conservation reserve in southeastern Australia. *Journal of Vegetation Science,* 14**,** 345-354.

BRANDT, M., HIERNAUX, P., TAGESSON, T., VERGER, A., RASMUSSEN, K., DIOUF, A. A., MBOW, C., MOUGIN, E. & FENSHOLT, R. 2016. Woody plant cover estimation in drylands from Earth Observation based seasonal metrics. *Remote Sensing of Environment,* 172**,** 28-38.

BUCINI, G., HANAN, N., BOONE, R., SMIT, I., SAATCHI, S., LEFSKY, M. & ASNER, G. 2010. Woody fractional cover in Kruger National Park, South Africa: remote-sensing-based maps and ecological insights. *Ecosystem function in savannas: measurement and modeling at landscape to global scales***,** 219-238.

BUCINI, G., SAATCHI, S., HANAN, N., BOONE, R. B. & SMIT, I. Woody cover and heterogeneity in the savannas of the Kruger National Park, South Africa. Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009, 2009. IEEE, IV-334-IV-337.

CANISIUS, F., TURRAL, H. & MOLDEN, D. 2007. Fourier analysis of historical NOAA time series data to estimate bimodal agriculture. *International Journal of Remote Sensing,* 28**,** 5503-5522.

CARLSON, T. N. & RIPLEY, D. A. 1997. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote sensing of Environment,* 62**,** 241-252.

CHAMAILLE-JAMMES, S. & FRITZ, H. 2009. Precipitation-NDVI relationships in eastern and southern African savannas vary along a precipitation gradient. *International Journal of Remote Sensing,* 30**,** 3409-3422.
CHAMAILLE-JAMMES, S., FRITZ, H. & MURINDAGOMO, F. 2006. Spatial patterns of the NDVI-rainfall relationship at the seasonal and interannual time scales in an African savanna. *International Journal of Remote Sensing,* 27**,** 5185-5200.
CHANDRASEKHAR, S. 2013. *Radiative transfer*, Courier Corporation.
CHEN, C., SON, N., CHANG, L. & CHEN, C. 2011. Classification of rice cropping systems by empirical mode decomposition and linear mixture model for time-series MODIS 250 m NDVI data in the Mekong Delta, Vietnam. *International journal of remote sensing,* 32**,** 5115-5134.
CHEN, Q. 2007. Airborne lidar data processing and information extraction. *Photogrammetric Engineering and Remote Sensing,* 73**,** 109.
CHIDUMAYO, E. 2001a. Climate and phenology of savanna vegetation in southern Africa. *Journal of Vegetation Science***,** 347-354.
CHIDUMAYO, E. N. 2001b. Climate and Phenology of Savanna Vegetation in Southern Africa. *Journal of Vegetation Science,* 12**,** 347-354.
CHO, M. A., DEBBA, P., MATHIEU, R., NAIDOO, L., VAN AARDT, J. & ASNER, G. P. 2010. Improving discrimination of savanna tree species through a multiple-endmember spectral angle mapper approach: Canopy-level analysis. *Geoscience and Remote Sensing, IEEE Transactions on,* 48**,** 4133-4142.
CHO, M. A., MATHIEU, R., ASNER, G. P., NAIDOO, L., VAN AARDT, J., RAMOELO, A., DEBBA, P., WESSELS, K., MAIN, R., SMIT, I. P. J. & ERASMUS, B. 2012. Mapping tree species composition in South African savannas using an integrated airborne spectral and LiDAR system. *Remote Sensing of Environment,* 125**,** 214-226.
CHUVIECO, E. & HUETE, A. 2010. *Fundamentals of satellite remote sensing,* Boca Raton, Fla; London, CRC.
CLARK, R. N. 1999. Spectroscopy of rocks and minerals, and principles of spectroscopy. *Manual of remote sensing,* 3**,** 3-58.
CLELAND, E. E., CHUINE, I., MENZEL, A., MOONEY, H. A. & SCHWARTZ, M. D. 2007. Shifting plant phenology in response to global change. *Trends in ecology & evolution,* 22**,** 357-365.
CLEVELAND, R. B., CLEVELAND, W. S., MCRAE, J. E. & TERPENNING, I. 1990. STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics,* 6**,** 3-73.
CODRON, J., LEE-THORP, J. A., SPONHEIMER, M., CODRON, D., GRANT, R. C. & DE RUITER, D. J. 2006. Elephant (Loxodonta africana) diets in Kruger National Park, South Africa: spatial and landscape differences. *Journal of Mammalogy,* 87**,** 27-34.
DE BEURS, K. M. & HENEBRY, G. M. 2004. Land surface phenology, climatic variation, and institutional change: analyzing agricultural land cover change in Kazakhstan. *Remote Sensing of Environment,* 89**,** 497-509.
DE BEURS, K. M. & HENEBRY, G. M. 2010. Spatio-temporal statistical methods for modelling land surface phenology. *Phenological research.* Springer.
DE MICHELE, C. & ACCATINO, F. 2014. Tree cover bimodality in savannas and forests emerging from the switching between two fire dynamics. *PloS one,* 9**,** e91195.
DEFRIES, R., ACHARD, F., BROWN, S., HEROLD, M., MURDIYARSO, D., SCHLAMADINGER, B. & DE SOUZA, C. 2007. Earth observations for

estimating greenhouse gas emissions from deforestation in developing countries. *environmental science & policy,* 10**,** 385-394.
DEFRIES, R., HANSEN, M., TOWNSHEND, J., JANETOS, A. & LOVELAND, T. 2000. A new global 1-km dataset of percentage tree cover derived from remote sensing. *Global Change Biology,* 6**,** 247-254.
DING, Y., ZHENG, X., ZHAO, K., XIN, X. & LIU, H. 2016. Quantifying the Impact of NDVIsoil Determination Methods and NDVIsoil Variability on the Estimation of Fractional Vegetation Cover in Northeast China. *Remote Sensing,* 8**,** 29.
DRAY, S., LEGENDRE, P. & PERES-NETO, P. R. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *ecological modelling,* 196**,** 483-493.
ECKHARDT, H., VAN WILGEN, B. & BIGGS, H. 2000. Trends in woody vegetation cover in the Kruger National Park, South Africa, between 1940 and 1998.
EKLUNDHA, L. & JÖNSSONB, P. 2012. TIMESAT 3.1 software manual. *Lund University, Sweden*.
EVRENDILEK, F. & GULBEYAZ, O. 2008. Deriving vegetation dynamics of natural terrestrial ecosystems from MODIS NDVI/EVI data over Turkey. *Sensors,* 8**,** 5270-5302.
FASSNACHT, F. E., LATIFI, H., STEREŃCZAK, K., MODZELEWSKA, A., LEFSKY, M., WASER, L. T., STRAUB, C. & GHOSH, A. 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment,* 186**,** 64-87.
FEBRUARY, E. C., FROST, P. G. H., SANKARAN, M., DIOUF, A., FERAL, C. J., CADE, B. S., PRINS, H. H. T., BICINI, G., HRABAR, H., COUGHENOUR, M. B., TEWS, J., HIGGINS, S. I., CAYLOR, K. K., GIGNOUX, J., LUDWIG, F., HANAN, N. P., ROUX, L. X., AUGUSTINE, D. J., SEA, W., METZGER, K. L., BRONN, A., ZAMBATIS, N., WORDEN, J., ARDO, J., EKAYA, W., SCHOLES, R. J., BANYIKWA, F., RINGROSE, S., RATNAM, J., HIERNAUX, P., LUND, U., SECTION OF, B., EARTH, S., NATURVETENSKAP, DEPT OF PHYSICAL, G., ECOSYSTEM, S., BIOLOGISK-GEOVETENSKAPLIGA, V., LUNDS, U., GEOCENTRUM, II, GEOCENTRE, II, INSTITUTIONEN FÖR NATURGEOGRAFI OCH, E. & SCIENCE 2005. Determinants of woody cover in African savannas. *Nature,* 438**,** 846-849.
FISHER, P. 1997. The pixel: a snare and a delusion. *International Journal of Remote Sensing,* 18**,** 679-685.
FOODY, G. & COX, D. 1994. Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions. *Remote sensing,* 15**,** 619-631.
FOODY, G. M. 2003. Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI–rainfall relationship. *Remote Sensing of Environment,* 88**,** 283-293.
FULLER, D. O., PRINCE, S. D. & ASTLE, W. L. 1997. The influence of canopy strata on remotely sensed observations of savanna-woodlands. *International Journal of Remote Sensing,* 18**,** 2985-3009.
GAMON, J. A., FIELD, C. B., GOULDEN, M. L., GRIFFIN, K. L., HARTLEY, A. E., JOEL, G., PENUELAS, J. & VALENTINI, R. 1995. Relationships between NDVI, canopy structure, and photosynthesis in three Californian vegetation types. *Ecological Applications,* 5**,** 28-41.
GEERKEN, R., ZAITCHIK, B. & EVANS, J. 2005. Classifying rangeland vegetation type and coverage from NDVI time series using Fourier Filtered Cycle Similarity. *International Journal of Remote Sensing,* 26**,** 5535-5554.

GESSNER, U., CONRAD, C., HUTTICH, C., KEIL, M., SCHMIDT, M., SCHRAMM, M. & DECH, S. A Multi-Scale Approch for Retreiving Proportional Cover of Life Forms. Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International, 2008. IEEE, III-700-III-703.

GESSNER, U., MACHWITZ, M., CONRAD, C. & DECH, S. 2013. Estimating the fractional cover of growth forms and bare surface in savannas. A multi-resolution approach based on regression tree ensembles. *Remote Sensing of Environment,* 129**,** 90-102.

GHIL, M., ALLEN, M., DETTINGER, M., IDE, K., KONDRASHOV, D., MANN, M., ROBERTSON, A. W., SAUNDERS, A., TIAN, Y. & VARADI, F. 2002. Advanced spectral methods for climatic time series. *Reviews of geophysics,* 40.

GIGLIO, L., VAN DER WERF, G., RANDERSON, J., COLLATZ, G. & KASIBHATLA, P. 2006. Global estimation of burned area using MODIS active fire observations. *Atmospheric Chemistry and Physics,* 6**,** 957-974.

GILABERT, M., GONZÁLEZ-PIQUERAS, J., GARCıA-HARO, F. & MELIÁ, J. 2002. A generalized soil-adjusted vegetation index. *Remote Sensing of environment,* 82**,** 303-310.

GILL, T. K., PHINN, S. R., ARMSTON, J. D. & PAILTHORPE, B. A. 2009. Estimating tree-cover change in Australia: challenges of using the MODIS vegetation index product. *International Journal of Remote Sensing,* 30**,** 1547-1565.

GITELSON, A. A. 2004. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of Plant Physiology,* 161**,** 165-173.

GITELSON, A. A. 2013. Remote estimation of crop fractional vegetation cover: the use of noise equivalent as an indicator of performance of vegetation indices. *International journal of remote sensing,* 34**,** 6054-6066.

GRIFFITH, D. & CHUN, Y. 2016. Spatial Autocorrelation and Uncertainty Associated with Remotely-Sensed Data. *Remote Sensing,* 8**,** 535.

GUAN, K., WOOD, E. F. & CAYLOR, K. K. 2012. Multi-sensor derivation of regional vegetation fractional cover in Africa. *Remote Sensing of Environment,* 124**,** 653-665.

GUERSCHMAN, J. P., HILL, M. J., RENZULLO, L. J., BARRETT, D. J., MARKS, A. S. & BOTHA, E. J. 2009. Estimating fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil in the Australian tropical savanna region upscaling the EO-1 Hyperion and MODIS sensors. *Remote Sensing of Environment,* 113**,** 928-945.

GUTMAN, G. & IGNATOV, A. 1998. The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *International Journal of remote sensing,* 19**,** 1533-1543.

GWENZI, D. & LEFSKY, M. A. 2014. Modeling canopy height in a savanna ecosystem using spaceborne lidar waveforms. *Remote sensing of environment,* 154**,** 338-344.

HANSEN, M., DEFRIES, R., TOWNSHEND, J., CARROLL, M., DIMICELI, C. & SOHLBERG, R. 2003a. Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm. *Earth Interactions,* 7**,** 1-15.

HANSEN, M., DEFRIES, R., TOWNSHEND, J. R. & SOHLBERG, R. 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. *International journal of remote sensing,* 21**,** 1331-1364.

HANSEN, M. C., DEFRIES, R. S., TOWNSHEND, J. R. G., CARROLL, M., DIMICELI, C. & SOHLBERG, R. A. 2003b. Global percent tree cover at a spatial

resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm. *Earth Interactions,* 7**,** 1-15.
HANSEN, M. C., DEFRIES, R. S., TOWNSHEND, J. R. G., SOHLBERG, R., DIMICELI, C. & CARROLL, M. 2002. Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. *Remote Sensing of Environment,* 83**,** 303-319.
HANSEN, M. C., TOWNSHEND, J. R., DEFRIES, R. S. & CARROLL, M. 2005a. Estimation of tree cover using MODIS data at global, continental and regional/local scales. *International Journal of Remote Sensing,* 26**,** 4359-4380.
HANSEN, M. C., TOWNSHEND, J. R. G., DEFRIES, R. S. & CARROLL, M. 2005b. Estimation of tree cover using MODIS data at global, continental and regional/local scales. *International Journal of Remote Sensing,* 26**,** 4359-4380.
HARTLEY, H. 1949. Tests of significance in harmonic analysis. *Biometrika***,** 194-201.
HELMAN, D., LENSKY, I. M., TESSLER, N. & OSEM, Y. 2015. A phenology-based method for monitoring woody and herbaceous vegetation in Mediterranean forests from NDVI time series. *Remote Sensing,* 7**,** 12314-12335.
HERNANDEZ-LEAL, P., ARBELO, M. & GONZALEZ-CALVO, A. 2006. Fire risk assessment using satellite data. *Advances in Space research,* 37**,** 741-746.
HERNANDEZ, G. 1999. Time series, periodograms, and significance. *Journal of Geophysical Research: Space Physics,* 104**,** 10355-10368.
HEROLD, M., MAYAUX, P., WOODCOCK, C., BACCINI, A. & SCHMULLIUS, C. 2008. Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment,* 112**,** 2538-2556.
HERRMANN, S. M., WICKHORST, A. J. & MARSH, S. E. 2013. Estimation of tree cover in an agricultural parkland of Senegal using rule-based regression tree modeling. *Remote Sensing,* 5**,** 4900-4918.
HIGGINS, S. I., DELGADO‐CARTAY, M. D., FEBRUARY, E. C. & COMBRINK, H. J. 2011. Is there a temporal niche separation in the leaf phenology of savanna trees and grasses? *Journal of Biogeography,* 38**,** 2165-2175.
HIJMANS, R. J. & VAN ETTEN, J. 2013. Raster: geographic data analysis and modeling. R package version 2.1-49. See http://CRAN. R-project. org/package= raster.
HILKER, T., WULDER, M. A., COOPS, N. C., LINKE, J., MCDERMID, G., MASEK, J. G., GAO, F. & WHITE, J. C. 2009. A new data fusion model for high spatial-and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sensing of Environment,* 113**,** 1613-1627.
HILL, M. J. & HANAN, N. P. 2010a. *Ecosystem Function in Savannas: measurement and modeling at landscape to global scales*, CRC Press.
HILL, M. J. & HANAN, N. P. 2010b. *Ecosystem function in savannas: Measurement and modeling at landscape to global scales*, CRC Press.
HILL, M. J., ROMÁN, M. O., SCHAAF, C. B., HUTLEY, L., BRANNSTROM, C., ETTER, A. & HANAN, N. P. 2011. Characterizing vegetation cover in global savannas with an annual foliage clumping index derived from the MODIS BRDF product. *Remote Sensing of Environment,* 115**,** 2008-2024.
HIRD, J. N. & MCDERMID, G. J. 2009. Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment,* 113**,** 248-258.
HOFFMANN, W. A. & JACKSON, R. B. 2000. Vegetation-climate feedbacks in the conversion of tropical savanna to grassland. *Journal of Climate,* 13**,** 1593-1602.

HOLBEN, B. N. 1986. Characteristics of maximum-value composite images from temporal AVHRR data. *International journal of remote sensing,* 7**,** 1417-1434.

HOUSE, J. I., ARCHER, S., BRESHEARS, D. D. & SCHOLES, R. J. 2003. Conundrums in mixed woody–herbaceous plant systems. *Journal of biogeography,* 30**,** 1763-1777.

HUETE, A., JACKSON, R. & POST, D. 1985. Spectral response of a plant canopy with different soil backgrounds. *Remote sensing of environment,* 17**,** 37-53.

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

JAKUBAUSKAS, M. E., LEGATES, D. R. & KASTENS, J. H. 2001. Harmonic analysis of time-series AVHRR NDVI data. *Photogrammetric Engineering and Remote Sensing,* 67**,** 461-470.

JAKUBAUSKAS, M. E., LEGATES, D. R. & KASTENS, J. H. 2002. Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Computers and electronics in agriculture,* 37**,** 127-139.

JIANG, Z., HUETE, A. R., CHEN, J., CHEN, Y., LI, J., YAN, G. & ZHANG, X. 2006. Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. *Remote Sensing of Environment,* 101**,** 366-378.

JIMÉNEZ-MUÑOZ, J. C., SOBRINO, J. A., PLAZA, A., GUANTER, L., MORENO, J. & MARTÍNEZ, P. 2009. Comparison between fractional vegetation cover retrievals from vegetation indices and spectral mixture analysis: Case study of PROBA/CHRIS data over an agricultural area. *Sensors,* 9**,** 768-793.

JIN, C., XIAO, X., MERBOLD, L., ARNETH, A., VEENENDAAL, E. & KUTSCH, W. L. 2013. Phenology and gross primary production of two dominant savanna woodland ecosystems in Southern Africa. *Remote Sensing of Environment,* 135**,** 189-201.

JOLLY, W. M. & RUNNING, S. W. 2004. Effects of precipitation and soil water potential on drought deciduous phenology in the Kalahari. *Global Change Biology,* 10**,** 303-308.

KADUK, J. D. & LOS, S. O. 2011. Predicting the time of green up in temperate and boreal biomes. *Climatic Change,* 107**,** 277-304.

KAMELARCZYK, K. B. F. 2009. *Carbon stock assessment and modelling in Zambia: a UN-REDD programme study*, UNDP.

KHALEFA, E., SMIT, I. P. J., NICKLESS, A., ARCHIBALD, S., COMBER, A. & BALZTER, H. 2013. Retrieval of savanna vegetation canopy height from ICESat-GLAS spaceborne LiDAR with terrain correction.

KÖHL, M., MAGNUSSEN, S. S. & MARCHETTI, M. 2006. *Sampling methods, remote sensing and GIS multiresource forest inventory*, Springer Science & Business Media.

KONG, Y.-L., MENG, Y., LI, W., YUE, A.-Z. & YUAN, Y. 2015. Satellite Image Time Series Decomposition Based on EEMD. *Remote Sensing,* 7**,** 15583-15604.

KOSTADINOV, T. S., CABRÉ, A., VEDANTHAM, H., MARINOV, I., BRACHER, A., BREWIN, R. J. W., BRICAUD, A., HIRATA, T., HIRAWAKE, T., HARDMAN-MOUNTFORD, N. J., MOUW, C., ROY, S. & UITZ, J. 2017. Inter-comparison of phytoplankton functional type phenology metrics derived from ocean color algorithms and Earth System Models. *Remote Sensing of Environment,* 190**,** 162-177.

LABRECQUE, S., FOURNIER, R. A., LUTHER, J. E. & PIERCEY, D. 2006. A comparison of four methods to map biomass from Landsat-TM and inventory data in western Newfoundland. *Forest Ecology and Management,* 226**,** 129-144.

LAMIGUEIRO, O. P., HIJMANS, R. & LAMIGUEIRO, M. O. P. 2011. Package 'rasterVis'. *Visualization Methods for the Raster Package. URL: http://rastervis. r-forge. r-project. org*.
LAW, B. E., ARKEBAUER, T., CAMPBELL, J. L., CHEN, J., SUN, O., SCHWARTZ, M., VAN INGEN, C. & VERMA, S. 2008. Terrestrial carbon observations: Protocols for vegetation sampling and data submission. *FAO, Rome*.
LEFSKY, M. A., HARDING, D. J., KELLER, M., COHEN, W. B., CARABAJAL, C. C., DEL BOM ESPIRITO-SANTO, F., HUNTER, M. O., DE OLIVEIRA, R. & DE CAMARGO, P. B. 2006. Correction to "Estimates of forest canopy height and aboveground biomass using ICESat". *Geophysical Research Letters,* 33.
LEFSKY, M. A., KELLER, M., PANG, Y., DE CAMARGO, P. B. & HUNTER, M. O. 2007. Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. *Journal of Applied Remote Sensing,* 1**,** 013537-013537-18.
LEGENDRE, P. & GAUTHIER, O. 2014. Statistical methods for temporal and space–time analysis of community composition data. *Proceedings of the Royal Society of London B: Biological Sciences,* 281**,** 20132728.
LEWIS, R. J., PAKEMAN, R. J. & MARRS, R. H. 2014. Identifying the multi-scale spatial structure of plant community determinants of an important national resource. *Journal of Vegetation Science,* 25**,** 184-197.
LHERMITTE, S., VERBESSELT, J., VERSTRAETEN, W. & COPPIN, P. 2011. A comparison of time series similarity measures for classification and change detection of ecosystem dynamics. *Remote Sensing of Environment,* 115**,** 3129-3152.
LI, G., LU, D., MORAN, E., DUTRA, L. & BATISTELLA, M. 2012. A comparative analysis of ALOS PALSAR L-band and RADARSAT-2 C-band data for land-cover classification in a tropical moist region. *ISPRS Journal of Photogrammetry and Remote Sensing,* 70**,** 26-38.
LI, L., FRIEDL, M. A., XIN, Q., GRAY, J., PAN, Y. & FROLKING, S. 2014. Mapping Crop cycles in China using MODIS-EVI time series. *Remote Sensing,* 6**,** 2473-2493.
LI, X. & STRAHLER, A. H. 1985. *IEEE Transactions on Geoscience and Remote Sensing,* 23**,** pp. 705.
LILLESAND, T., KIEFER, R. W. & CHIPMAN, J. 2014. *Remote sensing and image interpretation*, John Wiley & Sons.
LILLESAND, T. M., KIEFER, R. W. & CHIPMAN, J. W. 2004. *Remote sensing and image interpretation*, John Wiley & Sons Ltd.
LOS, S. 2013. Analysis of trends in fused AVHRR and MODIS NDVI data for 1982–2006: Indication for a CO2 fertilization effect in global vegetation. *Global Biogeochemical Cycles,* 27**,** 318-330.
LOS, S., POLLACK, N., PARRIS, M., COLLATZ, G., TUCKER, C., SELLERS, P., MALMSTRÖM, C., DEFRIES, R., BOUNOUA, L. & DAZLICH, D. 2000. A global 9-yr biophysical land surface dataset from NOAA AVHRR data. *Journal of Hydrometeorology,* 1**,** 183-199.
LOS, S., ROSETTE, J., KLJUN, N., NORTH, P., CHASMER, L., SUÁREZ, J., HOPKINSON, C., HILL, R., GORSEL, E. V. & MAHONEY, C. 2012. Vegetation height and cover fraction between 60 S and 60 N from ICESat GLAS data. *Geoscientific Model Development,* 5**,** 413-432.

LOS, S. O., JUSTICE, C. & TUCKER, C. 1994. A global 1 by 1 NDVI data set for climate studies derived from the GIMMS continental NDVI data. *International Journal of Remote Sensing,* 15**,** 3493-3518.

LOS, S. O., NORTH, P. R. J., GREY, W. M. F. & BARNSLEY, M. J. 2005. A method to convert AVHRR Normalized Difference Vegetation Index time series to a standard viewing and illumination geometry. *Remote Sensing of Environment,* 99**,** 400-411.

LOTSCH, A., TIAN, Y., FRIEDL, M. & MYNENI, R. 2003. Land cover mapping in support of LAI and FPAR retrievals from EOS-MODIS and MISR: Classification methods and sensitivities to errors. *International Journal of Remote Sensing,* 24**,** 1997-2016.

LOVELAND, T., REED, B., BROWN, J., OHLEN, D., ZHU, Z., YANG, L. & MERCHANT, J. 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing,* 21**,** 1303-1330.

LU, H., RAUPACH, M. R., MCVICAR, T. R. & BARRETT, D. J. 2003. Decomposition of vegetation cover into woody and herbaceous components using AVHRR NDVI time series. *Remote Sensing of Environment,* 86**,** 1-18.

LUCAS, R. M., MITCHELL, A. L., ROSENQVIST, A., PROISY, C., MELIUS, A. & TICEHURST, C. 2007. The potential of L-band SAR for quantifying mangrove characteristics and change: case studies from the tropics. *Aquatic Conservation: Marine and Freshwater Ecosystems,* 17**,** 245-264.

MA, X., HUETE, A., YU, Q., COUPE, N. R., DAVIES, K., BROICH, M., RATANA, P., BERINGER, J., HUTLEY, L. B. & CLEVERLY, J. 2013. Spatial patterns and temporal dynamics in savanna vegetation phenology across the North Australian Tropical Transect. *Remote sensing of Environment,* 139**,** 97-115.

MAIROTA, P., CAFARELLI, B., DIDHAM, R. K., LOVERGINE, F. P., LUCAS, R. M., NAGENDRA, H., ROCCHINI, D. & TARANTINO, C. 2015. Challenges and opportunities in harnessing satellite remote-sensing for biodiversity monitoring. *Ecological Informatics,* 30**,** 207-214.

MARTÍNEZ, B. & GILABERT, M. A. 2009. Vegetation dynamics from NDVI time series analysis using the wavelet transform. *Remote Sensing of Environment,* 113**,** 1823-1842.

MATHIEU, R., NAIDOO, L., CHO, M. A., LEBLON, B., MAIN, R., WESSELS, K., ASNER, G. P., BUCKLEY, J., VAN AARDT, J. & ERASMUS, B. F. 2013. Toward structural assessment of semi-arid African savannahs and woodlands: The potential of multitemporal polarimetric RADARSAT-2 fine beam images. *Remote Sensing of Environment,* 138**,** 215-231.

MBOW, C., FENSHOLT, R., NIELSEN, T. T. & RASMUSSEN, K. 2014. Advances in monitoring vegetation and land use dynamics in the Sahel. *Geografisk Tidsskrift-Danish Journal of Geography,* 114**,** 84-91.

MCCOY, R. M. 2005. *Field methods in remote sensing*, Guilford Press.

MINGWEI, Z., QINGBO, Z., ZHONGXIN, C., JIA, L., YONG, Z. & CHONGFA, C. 2008. Crop discrimination in Northern China with double cropping systems using Fourier analysis of time-series MODIS data. *International Journal of Applied Earth Observation and Geoinformation,* 10**,** 476-485.

MITCHARD, E. T. A., SAATCHI, S. S., WHITE, L. J. T., ABERNETHY, K. A., JEFFERY, K. J., LEWIS, S. L., COLLINS, M., LEFSKY, M. A., LEAL, M. E. & WOODHOUSE, I. H. 2012. Mapping tropical forest biomass with radar and

spaceborne LiDAR in Lopé National Park, Gabon: overcoming problems of high biomass and persistent cloud. *Biogeosciences,* 9**,** 179-191.
MONDAL, S., JEGANATHAN, C., SINHA, N. K., RAJAN, H., ROY, T. & KUMAR, P. 2014. Extracting seasonal cropping patterns using multi-temporal vegetation indices from IRS LISS-III data in Muzaffarpur District of Bihar, India. *The Egyptian Journal of Remote Sensing and Space Science,* 17**,** 123-134.
MONTANDON, L. & SMALL, E. 2008. The impact of soil reflectance on the quantification of the green vegetation fraction from NDVI. *Remote Sensing of Environment,* 112**,** 1835-1845.
MOODY, A. & JOHNSON, D. M. 2001. Land-surface phenologies from AVHRR using the discrete Fourier transform. *Remote Sensing of Environment,* 75**,** 305-323.
MOORE, C. E., BERINGER, J., EVANS, B., HUTLEY, L. B. & TAPPER, N. J. 2016. Tree-grass phenology information improves light use efficiency modelling of gross primary productivity for an Australian tropical savanna. *Biogeosciences Discuss.,* 2016**,** 1-38.
MORENO-DE LAS HERAS, M., DÍAZ-SIERRA, R., TURNBULL, L. & WAINWRIGHT, J. 2015. Assessing vegetation structure and ANPP dynamics in a grassland-shrubland Chihuahuan ecotone using NDVI-rainfall relationships. *Biogeosciences Discussions,* 12**,** 51-92.
MUHAMMAD, S., WACHOWICZ, M. & DE CARVALHO, L. M. Evaluation of wavelet transform algorithms for multi-resolution image fusion. Information Fusion, 2002. Proceedings of the Fifth International Conference on, 2002. IEEE, 1573-1580.
MULLER, E. & DÉCAMPS, H. 2001. Modeling soil moisture–reflectance. *Remote Sensing of Environment,* 76**,** 173-180.
MYNENI, R. B. & HALL, F. G. 1995. The interpretation of spectral vegetation indexes. *Geoscience and Remote Sensing, IEEE Transactions on,* 33**,** 481-486.
MYNENI, R. B., ROSS, J. & ASRAR, G. 1989. A review on the theory of photon transport in leaf canopies. *Agricultural and Forest Meteorology,* 45**,** 1-153.
NAIDOO, L., MATHIEU, R., MAIN, R., KLEYNHANS, W., WESSELS, K., ASNER, G. & LEBLON, B. 2015. Savannah woody structure modelling and mapping using multi-frequency (X-, C-and L-band) Synthetic Aperture Radar data. *ISPRS Journal of Photogrammetry and Remote Sensing,* 105**,** 234-250.
NIANG, O., DELÉCHELLE, É. & LEMOINE, J. 2010. A spectral approach for sifting process in empirical mode decomposition. *Signal Processing, IEEE Transactions on,* 58**,** 5612-5623.
NORTH, P., ROSETTE, J., SUÁREZ, J. & LOS, S. 2010. A Monte Carlo radiative transfer model of satellite waveform LiDAR. *International Journal of Remote Sensing,* 31**,** 1343-1358.
OLLINGER, S. 2011. Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist,* 189**,** 375-394.
PARDO-IGÚZQUIZA, E. & RODRÍGUEZ-TOVAR, F. J. 2015. Comparison of the Performance of Two Advanced Spectral Methods for the Analysis of Times Series in Paleoceanography. *Journal of Marine Science and Engineering,* 3**,** 957-967.
PIÑEIRO, G., PERELMAN, S., GUERSCHMAN, J. P. & PARUELO, J. M. 2008. How to evaluate models: observed vs. predicted or predicted vs. observed? *Ecological Modelling,* 216**,** 316-322.
POPESCU, S. C., ZHAO, K., NEUENSCHWANDER, A. & LIN, C. 2011. Satellite lidar vs. small footprint airborne lidar: Comparing the accuracy of aboveground

biomass estimates and forest structure metrics at footprint level. *Remote Sensing of Environment,* 115**,** 2786-2797.
POTTIER, J. & EVETTE, A. 2011. Spatial pattern and process at the plant neighbourhood scale: insights from communities dominated by the clonal grass Elymus repens (L.) Gould. *Journal of Vegetation Science,* 22**,** 973-982.
PRICE, J. C. 1990. On the information content of soil reflectance spectra. *Remote sensing of environment,* 33**,** 113-121.
PRINCE, S. D. & GOWARD, S. N. 1995. Global primary production: a remote sensing approach. *Journal of biogeography***,** 815-835.
PRINS, H. 1988. Plant phenology patterns in Lake Manyara National Park, Tanzania. *Journal of biogeography***,** 465-480.
RAHIM, K., BURR, W. S. & RAHIM, M. K. 2017. Package 'multitaper'.
REED, B. C., BROWN, J. F., VANDERZEE, D., LOVELAND, T. R., MERCHANT, J. W. & OHLEN, D. O. 1994. Measuring phenological variability from satellite imagery. *Journal of vegetation science,* 5**,** 703-714.
RIBEIRO, N. S., MATOS, C. N., MOURA, I. R., WASHINGTON-ALLEN, R. A. & RIBEIRO, A. I. 2013. Monitoring vegetation dynamics and carbon stock density in miombo woodlands. *Carbon balance and management,* 8**,** 11.
RIEMANN, R., LIKNES, G., O'NEIL-DUNNE, J., TONEY, C. & LISTER, T. 2016. Comparative assessment of methods for estimating tree canopy cover across a rural-to-urban gradient in the mid-Atlantic region of the USA. *Environmental monitoring and assessment,* 188**,** 1-17.
ROCCHINI, D., BOYD, D. S., FÉRET, J. B., FOODY, G. M., HE, K. S., LAUSCH, A., NAGENDRA, H., WEGMANN, M. & PETTORELLI, N. 2015. Satellite remote sensing to monitor species diversity: potential and pitfalls. *Remote Sensing in Ecology and Conservation*.
ROCCHINI, D., FOODY, G. M., NAGENDRA, H., RICOTTA, C., ANAND, M., HE, K. S., AMICI, V., KLEINSCHMIT, B., FÖRSTER, M. & SCHMIDTLEIN, S. 2013. Uncertainty in ecosystem mapping by remote sensing. *Computers & geosciences,* 50**,** 128-135.
ROERINK, G., MENENTI, M. & VERHOEF, W. 2000. Reconstructing cloudfree NDVI composites using Fourier analysis of time series. *International Journal of Remote Sensing,* 21**,** 1911-1917.
ROY, D. P., WULDER, M. A., LOVELAND, T. R., C.E, W., ALLEN, R. G., ANDERSON, M. C., HELDER, D., IRONS, J. R., JOHNSON, D. M., KENNEDY, R., SCAMBOS, T. A., SCHAAF, C. B., SCHOTT, J. R., SHENG, Y., VERMOTE, E. F., BELWARD, A. S., BINDSCHADLER, R., COHEN, W. B., GAO, F., HIPPLE, J. D., HOSTERT, P., HUNTINGTON, J., JUSTICE, C. O., KILIC, A., KOVALSKYY, V., LEE, Z. P., LYMBURNER, L., MASEK, J. G., MCCORKEL, J., SHUAI, Y., TREZZA, R., VOGELMANN, J., WYNNE, R. H. & ZHU, Z. 2014. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment,* 145**,** 154-172.
RUSCH, G., PAUSAS, J. & LEPŠ, J. 2003. Plant functional types in relation to disturbance and land use: introduction. *Journal of Vegetation Science,* 14**,** 307-310.
SANKARAN, M., HANAN, N. P., SCHOLES, R. J., RATNAM, J., AUGUSTINE, D. J., CADE, B. S., GIGNOUX, J., HIGGINS, S. I., LE ROUX, X. & LUDWIG, F. 2005. Determinants of woody cover in African savannas. *Nature,* 438**,** 846-849.

SANKARAN, M., RATNAM, J. & HANAN, N. 2008. Woody cover in African savannas: the role of resources, fire and herbivory. *Global Ecology and Biogeography,* 17**,** 236-245.
SANKARAN, M., RATNAM, J. & HANAN, N. P. 2004. Tree–grass coexistence in savannas revisited–insights from an examination of assumptions and mechanisms invoked in existing models. *Ecology Letters,* 7**,** 480-490.
SCANLON, T. M., ALBERTSON, J. D., CAYLOR, K. K. & WILLIAMS, C. A. 2002. Determining land surface fractional cover from NDVI and rainfall time series for a savanna ecosystem. *Remote Sensing of Environment,* 82**,** 376-388.
SCHARLEMANN, J. P., BENZ, D., HAY, S. I., PURSE, B. V., TATEM, A. J., WINT, G. W. & ROGERS, D. J. 2008. Global data for ecology and epidemiology: a novel algorithm for temporal Fourier processing MODIS data. *PloS one,* 3**,** e1408.
SCHMIDTLEIN, S. & FASSNACHT, F. E. 2017. The spectral variability hypothesis does not hold across landscapes. *Remote Sensing of Environment,* 192**,** 114-125.
SCHOLES, R. 2003. Convex relationships in ecosystems containing mixtures of trees and grass. *Environmental and resource economics,* 26**,** 559-574.
SCHOLES, R. & ARCHER, S. 1997. Tree-grass interactions in savannas. *Annual review of Ecology and Systematics***,** 517-544.
SELLA, L., VIVALDO, G., GHIL, M. & GROTH, A. Economic cycles and their synchronization: Spectral analysis of macroeconomic series from Italy, The Netherlands, and the UK. EGU General Assembly Conference Abstracts, 2010. 11847.
SELLERS, P. 1987. Canopy reflectance, photosynthesis, and transpiration, II. The role of biophysics in the linearity of their interdependence. *Remote sensing of Environment,* 21**,** 143-183.
SEXTON, J. O., SONG, X.-P., FENG, M., NOOJIPADY, P., ANAND, A., HUANG, C., KIM, D.-H., COLLINS, K. M., CHANNAN, S. & DIMICELI, C. 2013. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth,* 6**,** 427-448.
SHACKLETON, C. M. & SCHOLES, R. J. 2011. Above ground woody community attributes, biomass and carbon stocks along a rainfall gradient in the savannas of the central lowveld, South Africa. *South African Journal of Botany,* 77**,** 184-192.
SHATKAY, H. 1995. The Fourier transform-A primer. *Brown University*.
SHELL, I. & JAMES, R. 2005. Polarimetric remote sensing in the visible to near infrared.
SMALLMAN, T., EXBRAYAT, J. F., MENCUCCINI, M., BLOOM, A. & WILLIAMS, M. 2017. Assimilation of repeated woody biomass observations constrains decadal ecosystem carbon cycle uncertainty in aggrading forests. *Journal of Geophysical Research: Biogeosciences*.
SMIT, I. & ASNER, G. 2012. Roads increase woody cover under varying geological, rainfall and fire regimes in African savanna. *Journal of Arid Environments,* 80**,** 74-80.
SMIT, I. P., ASNER, G. P., GOVENDER, N., KENNEDY-BOWDOIN, T., KNAPP, D. E. & JACOBSON, J. 2010. Effects of fire on woody vegetation structure in African savanna. *Ecological Applications,* 20**,** 1865-1875.
SMITH, A. M. S., FALKOWSKI, M. J., GREENBERG, J. A. & TINKHAM, W. T. 2014a. Remote sensing of vegetation structure, function, and condition: Special issue. *Remote Sensing of Environment,* 154**,** 319-321.

SMITH, A. M. S., FALKOWSKI, M. J., GREENBERG, J. A. & TINKHAM, W. T. 2014b. Remote sensing of vegetation structure, function, and condition: Special issue. *Remote Sensing of Environment,* 154**,** 319–321.

SMITH, A. M. S., KOLDEN, C. A., TINKHAM, W. T., TALHELM, A. F., MARSHALL, J. D., HUDAK, A. T., BOSCHETTI, L., FALKOWSKI, M. J., GREENBERG, J. A., ANDERSON, J. W., KLISKEY, A., ALESSA, L., KEEFE, R. F. & GOSZ, J. R. 2014c. Remote sensing the vulnerability of vegetation in natural terrestrial ecosystems. *Remote Sensing of Environment,* 154**,** 322-337.

SOBRINO, J. A. & RAISSOUNI, N. 2000. Toward remote sensing methods for land cover dynamic monitoring: application to Morocco. *International Journal of Remote Sensing,* 21**,** 353-366.

SOUSA, D. & SMALL, C. 2017. Global cross-calibration of Landsat spectral mixture models. *Remote Sensing of Environment,* 192**,** 139-149.

STONER, E. R. & BAUMGARDNER, M. 1981. Characteristic variations in reflectance of surface soils. *Soil Science Society of America Journal,* 45**,** 1161-1165.

STRAHLER, A. H., BOSCHETTI, L., FOODY, G. M., FRIEDL, M. A., HANSEN, M. C., HEROLD, M., MAYAUX, P., MORISETTE, J. T., STEHMAN, S. V. & WOODCOCK, C. E. 2006. Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps. *European Communities, Luxembourg,* 51.

STRAHLER, A. H., MULLER, J., LUCHT, W., SCHAAF, C., TSANG, T., GAO, F., LI, X., LEWIS, P. & BARNSLEY, M. J. 1999. MODIS BRDF/albedo product: algorithm theoretical basis document version 5.0. *MODIS documentation,* 23**,** 42-47.

STRANG, G. & NGUYEN, T. 1996. *Wavelets and filter banks*, SIAM.

SU, L., HUANG, Y., CHOPPING, M., RANGO, A. & MARTONCHIK, J. 2009. An empirical study on the utility of BRDF model parameters and topographic parameters for mapping vegetation in a semi-arid region with MISR imagery. *International Journal of Remote Sensing,* 30**,** 3463-3483.

TANSEY, K., GRÉGOIRE, J. M., STROPPIANA, D., SOUSA, A., SILVA, J., PEREIRA, J., BOSCHETTI, L., MAGGI, M., BRIVIO, P. A. & FRASER, R. 2004. Vegetation burning in the year 2000: Global burned area estimates from SPOT VEGETATION data. *Journal of Geophysical Research: Atmospheres,* 109.

THOMSON, D. J. 1982. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE,* 70**,** 1055-1096.

THORNTON, P. K. 2010. Livestock production: recent trends, future prospects. *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 365**,** 2853-2867.

TIM DANAHER, PETER SCARTH, JOHN ARMSTON, LISA COLLETT, JOANNA KITCHEN & GILLINGHAM, S. 2011. Remote Sensing of Tree–Grass Systems: The Eastern Australian Woodlands *In:* HILL, M. J. & HANAN, N. P. (eds.) *Ecosystem function in savannas: Measurement and modeling at landscape to global scales.* London: Taylor and Francis Group.

TOWNSHEND, J., CARROLL, M., DIMICELI, C., SOHLBERG, R., HANSEN, M. & DEFRIES, R. 2011. Vegetation continuous fields MOD44B, 2000 percent tree cover, collection 5. *University of Maryland, College Park, MD.*

TUCKER, C. & SELLERS, P. 1986. Satellite remote sensing of primary production. *International journal of remote sensing,* 7**,** 1395-1416.

TUCKER, C. J., TOWNSHEND, J. R. & GOFF, T. E. 1985. African land-cover classification using satellite data. *Science,* 227**,** 369-375.

URBAZAEV, M., THIEL, C., MATHIEU, R., NAIDOO, L., LEVICK, S. R., SMIT, I. P. J., ASNER, G. P. & SCHMULLIUS, C. 2015. Assessment of the mapping of fractional woody cover in southern African savannas using multi-temporal and polarimetric ALOS PALSAR L-band images. *Remote Sensing of Environment,* 166**,** 138-153.

USTIN, S. L. & GAMON, J. A. 2010. Remote sensing of plant functional types. *New Phytologist,* 186**,** 795-816.

VALENTINI, R., HENRY, M., HOUGHTON, R. A., JUNG, M., KUTSCH, W. L., MALHI, Y., MAYORGA, E., MERBOLD, L., MURRAY-TORTAROLO, G., PAPALE, D., PEYLIN, P., ARNETH, A., POULTER, B., RAYMOND, P. A., SANTINI, M., SITCH, S., VAGLIO LAURIN, G., VAN DER WERF, G. R., WILLIAMS, C. A., SCHOLES, R. J., BOMBELLI, A., CASTALDI, S., CAZZOLLA GATTI, R., CHEVALLIER, F., CIAIS, P., GRIECO, E. & HARTMANN, J. 2014. A full greenhouse gases budget of Africa: synthesis, uncertainties, and vulnerabilities. *Biogeosciences,* 11**,** 381-407.

VAN ENGELEN, V. & HARTEMINK, A. 2000. The global soils and terrain database (SOTER). *ACLEP Newsletter,* 9**,** 22-27.

VENTER, F. J., SCHOLES, R. J., ECKHARDT, H. C., DU TOIT, J., ROGERS, K. & BIGGS, H. 2003a. The abiotic template and its associated vegetation pattern. *The Kruger experience: Ecology and management of savanna heterogeneity,* 83**,** 129.

VENTER, F. J., SCHOLES, R. J., ECKHARDT, H. C., DU TOIT, J., ROGERS, K. & BIGGS, H. 2003b. The abiotic template and its associated vegetation pattern. *The Kruger experience: Ecology and management of savanna heterogeneity,* 21**,** 83-129.

VERBESSELT, J., HYNDMAN, R., NEWNHAM, G. & CULVENOR, D. 2010. Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment,* 114**,** 106-115.

VERGER, A., MARTÍNEZ, B., CAMACHO-DE COCA, F. & GARCÍA-HARO, F. 2009a. Accuracy assessment of fraction of vegetation cover and leaf area index estimates from pragmatic methods in a cropland area. *International Journal of Remote Sensing,* 30**,** 2685-2704.

VERGER, A., MARTÍNEZ, B., CAMACHO-DE COCA, F. & GARCÍA-HARO, F. J. 2009b. Accuracy assessment of fraction of vegetation cover and leaf area index estimates from pragmatic methods in a cropland area. *International Journal of Remote Sensing,* 30**,** 2685-2704.

VILLEGAS, J. C., DOMINGUEZ, F., BARRON-GAFFORD, G. A., ADAMS, H. D., GUARDIOLA-CLARAMONTE, M., SOMMER, E. D., SELVEY, A. W., ESPELETA, J. F., ZOU, C. B. & BRESHEARS, D. D. 2015. Sensitivity of regional evapotranspiration partitioning to variation in woody plant cover: insights from experimental dryland tree mosaics. *Global Ecology and Biogeography,* 24**,** 1040-1048.

WALKER, J., MOORE, R. M. & ROBERTSON, J. A. 1972. Herbage reponse to tree and shrub thinning in Eucalyptus populnea shrub woodlands. *Crop and Pasture Science,* 23**,** 405-410.

WALTER-SHEA, E., BLAD, B., HAYS, C., MESARCH, M., DEERING, D. & MIDDLETON, E. 1992. Biophysical properties affecting vegetative canopy reflectance and absorbed photosynthetically active radiation at the FIFE site. *Journal of Geophysical Research: Atmospheres,* 97**,** 18925-18934.

WANG, R., GAMON, J. A., MONTGOMERY, R. A., TOWNSEND, P. A., ZYGIELBAUM, A. I., BITAN, K., TILMAN, D. & CAVENDER-BARES, J.

2016. Seasonal variation in the NDVI–species richness relationship in a prairie grassland experiment (Cedar Creek). *Remote Sensing,* 8**,** 128.
WESSELS, K., MATHIEU, R., ERASMUS, B., ASNER, G., SMIT, I., VAN AARDT, J., MAIN, R. & FISHER, J. 2011. Impact of communal land use and conservation on woody vegetation structure in the Lowveld savannas of South Africa–LiDAR results.
WESTRA, T. & DE WULF, R. 2007. Monitoring Sahelian floodplains using Fourier analysis of MODIS time-series data and artificial neural networks. *International Journal of Remote Sensing,* 28**,** 1595-1610.
WHITECROSS, M. A., E. T. F. WITKOWSKI, AND S. ARCHIBALD. 2017a. "Assessing the Frequency and Drivers of Early-Greening in Broad-Leaved Woodlands along a Latitudinal Gradient in Southern Africa.".”Austral Ecology 42:: 341–353. doi:10.1111/aec.12448.
WHITECROSS, M. A., E. T. F. WITKOWSKI, AND S. ARCHIBALD. 2017b. "Savanna Tree-Grass Interactions: A Phenological Investigation of Green-Up in Relation to Water Availability over Three Seasons." South African Journal of Botany 108: 29–40. doi:10.1016/j.sajb.2016.09.003.
WHITE, M. A. & NEMANI, R. R. 2006. Real-time monitoring and short-term forecasting of land surface phenology. *Remote Sensing of Environment,* 104**,** 43-49.
WHITE, M. A., SHAW, J. D. & RAMSEY, R. D. 2005. Accuracy assessment of the vegetation continuous field tree cover product using 3954 ground plots in the south-western USA. *International Journal of Remote Sensing,* 26**,** 2699-2704.
WILLIAMS, C. A., HANAN, N. P., NEFF, J. C., SCHOLES, R. J., BERRY, J. A., DENNING, A. S. & BAKER, D. F. 2007. Africa and the global carbon cycle. *Carbon balance and management,* 2**,** 1.
WU, D., WU, H., ZHAO, X., ZHOU, T., TANG, B., ZHAO, W. & JIA, K. 2014. Evaluation of spatiotemporal variations of global fractional vegetation cover based on GIMMS NDVI data from 1982 to 2011. *Remote Sensing,* 6**,** 4217-4239.
WU, Z., HUANG, N. E., LONG, S. R. & PENG, C.-K. 2007. On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proceedings of the National Academy of Sciences,* 104**,** 14889-14894.
WYNN, C. 2000. An introduction to BRDF-based lighting. *Nvidia Corporation*.
XIE, Y., SHA, Z. & YU, M. 2008. Remote sensing imagery in vegetation mapping: a review. *Journal of plant ecology,* 1**,** 9-23.
YANG, J., WEISBERG, P. J. & BRISTOW, N. A. 2012. Landsat remote sensing approaches for monitoring long-term tree cover dynamics in semi-arid woodlands: Comparison of vegetation indices and spectral mixture analysis. *Remote Sensing of Environment,* 119**,** 62-71.
YENGOH, G. T., DENT, D., OLSSON, L., TENGBERG, A. E. & TUCKER III, C. J. 2015. *Use of the Normalized Difference Vegetation Index (NDVI) to Assess Land Degradation at Multiple Scales: Current Status, Future Trends, and Practical Considerations*, Springer.
ZEILEIS, A. & GROTHENDIECK, G. 2005. zoo: S3 infrastructure for regular and irregular time series. *arXiv preprint math/0505527*.
ZENG, X., DICKINSON, R. E., WALKER, A., SHAIKH, M., DEFRIES, R. S. & QI, J. 2000. Derivation and evaluation of global 1-km fractional vegetation cover data for land modeling. *Journal of Applied Meteorology,* 39**,** 826-839.
ZHU, L. & SOUTHWORTH, J. 2013. Disentangling the relationships between net primary production and precipitation in southern Africa savannas using satellite observations from 1982 to 2010. *Remote Sensing,* 5**,** 3803-3825.

ZIMMERMANN, N. E., T. C. EDWARDS, J., MOISEN, G. G., FRESCINO, T. S. & BLACKARD, J. A. 2007. Remote Sensing-Based Predictors Improve Distribution Models of Rare, Early Successional and Broadleaf Tree Species in Utah. *Journal of Applied Ecology,* 44**,** 1057-1067.

Appendix I: Contributions of co-authors

This thesis comprises three results chapters. The first result chapter has been published in the *international journal of remote sensing*, while the remaining two chapters are currently being prepared as one manuscript for submission into one of the remote sensing high impact journals. These parts of the thesis therefore contained a bit of other co-authors. The contribution of co-authors and the PhD student, are detailed below.

“Assessment of tree/grass fractional cover using phenological signal decomposition”

Parts of this work is under review with the *International journal of Remote Sensing*

Journal as:

Sa’ad Ibrahim, Heiko Balzter, Kevin Tansey, Narumasa Tsutsumida, Renaud Mathieu., (2018) Estimating fractional cover of plant functional types in African savanna from harmonic analysis of MODIS time-series. *International journal of Remote Sensing,* 39:9,

2718-2745

Respective co-author contributions

Sa’ad Ibrahim conceived the research idea, conduct the field work, analyses the data, created all tables, and figures, publication writing and revisions.

Heiko Balzter and Kevin Tansey provided feedback, edits throughout the planning and completion of the study. Heiko Balzter wrote the R code.

Tsutsumida, N. provide feedback and help with modification of the original R code written by Heiko for signal decomposition of MODIS images.

Mathieu, R. provides LiDAR-SAR woody fractional cover product and feedback, edits and review on the manuscripts

Ch. 5 & 6: “Estimating tree fractional cover in African savanna using MODIS time series” and “Validation of tree fractional cover map derived from phenological signal decomposition of MODIS time series data”

Parts of this work is being is prepared for submission to *Remote Sensing*

Sa’ad Ibrahim, Heiko Balzter, Kevin Tansey, Narumasa Tsutsumida, Renaud Mathieu (under review) Estimating woody fractional cover in African savanna using MODIS time series data. *Remote Sensing*

Respective co-author contributions

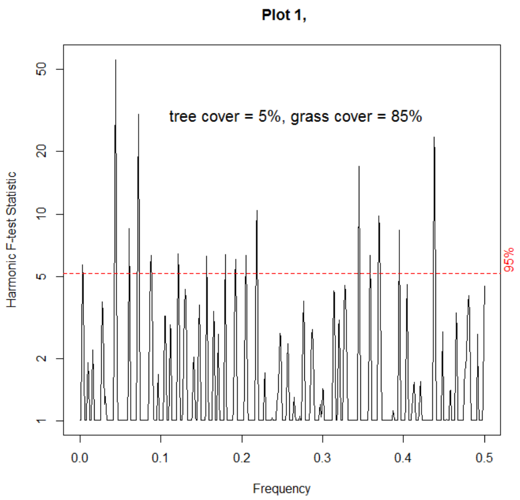
Sa’ad Ibrahim conceived the research idea, conduct the field work, analyses the data, created all tables, and figures, publication writing and revisions.

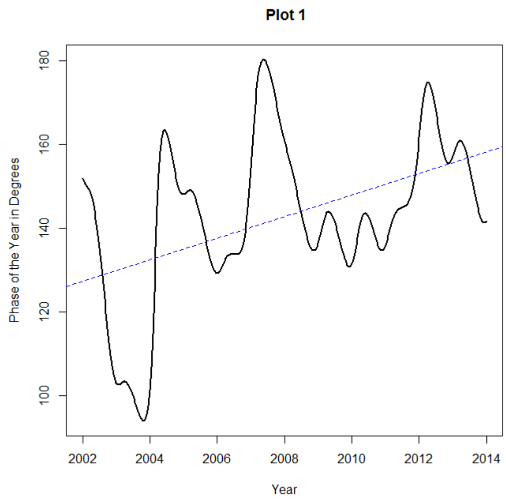
Heiko Balzter and Kevin Tansey provided feedback, edits throughout the planning and completion of the study. Heiko Balzter wrote the R code.

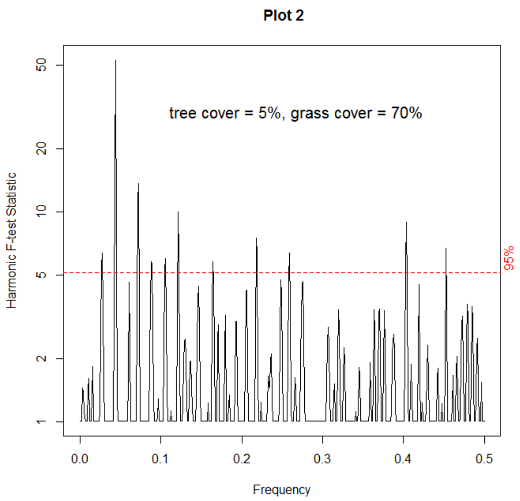
Mathieu, R. provides LiDAR-SAR woody fractional cover product and feedback, edits, and review on the manuscripts

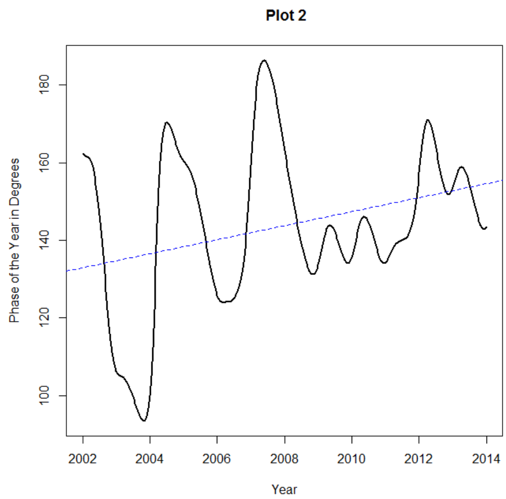
Tsutsumida, N. provide feedback and help with modification of the original R code written by Heiko for signal decomposition of MODIS images.

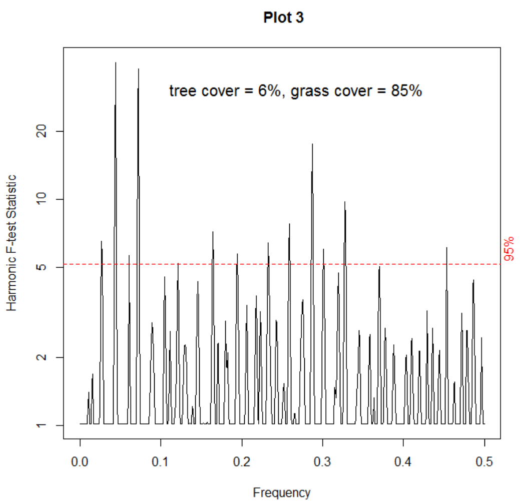
Appendix II: statistically significant peaks for all field data on tree/grass sites calculated with F-test using tapering method.

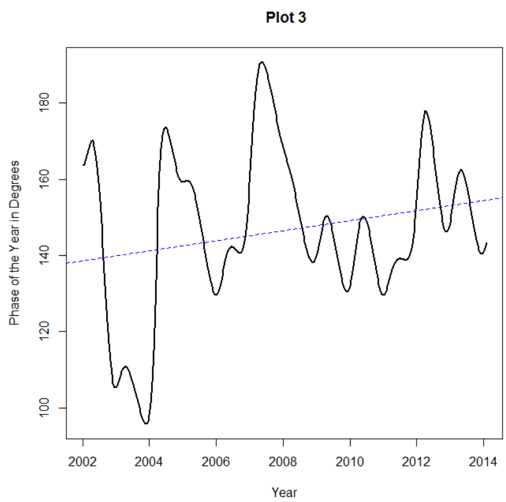


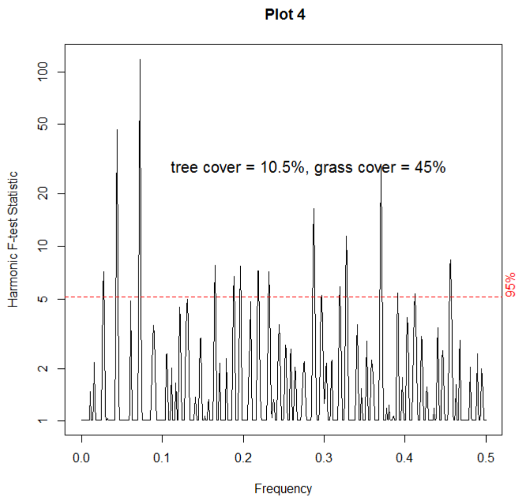


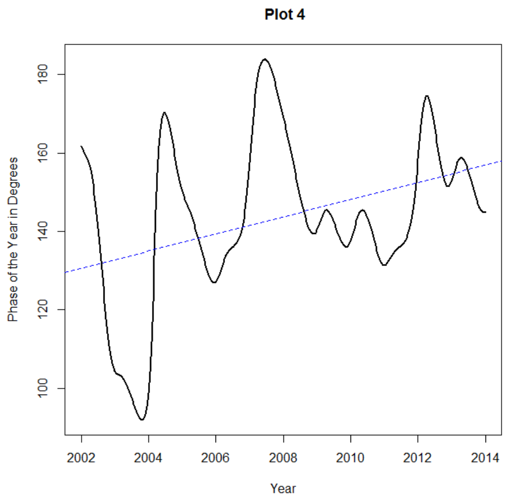


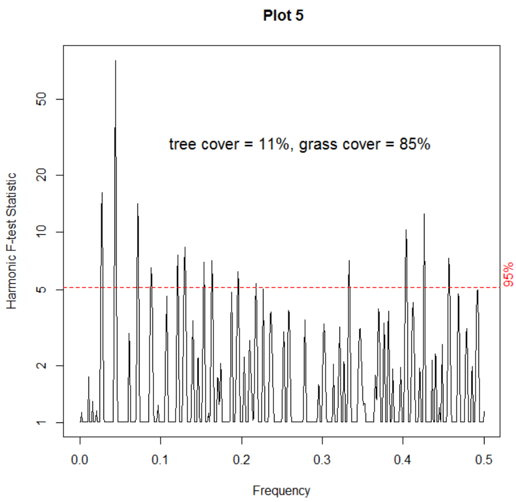


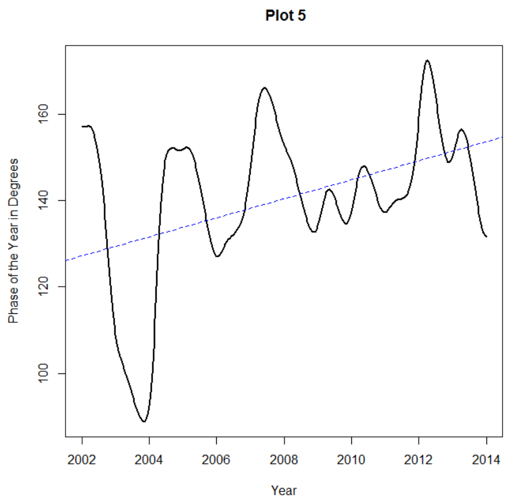


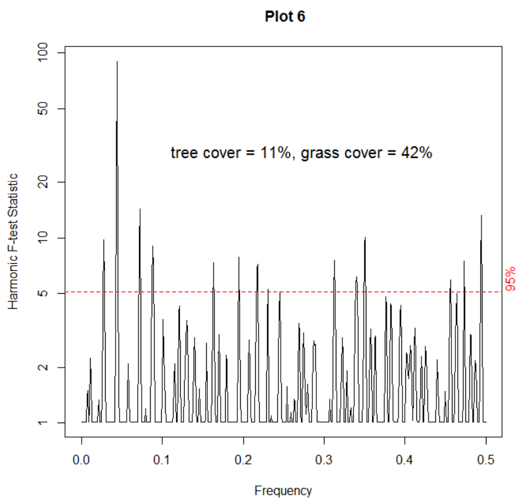


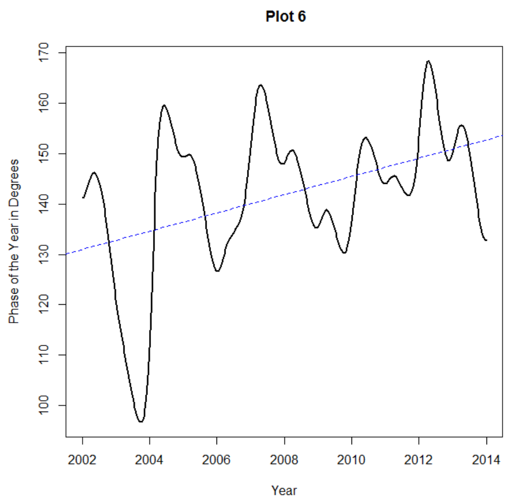

Plot 4
tree cover = 10.5%, grass cover = 45%
Harmonic F-test Statistic
1
2
5
10
20
50
100
95%
0.0
0.1
0.2
0.3
0.4
0.5
Frequency

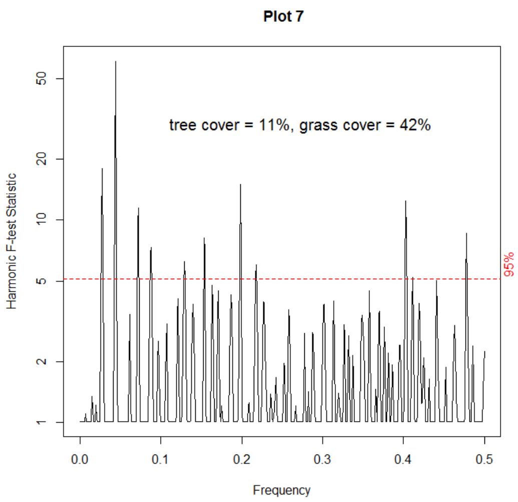

Plot 4
Phase of the Year in Degrees
100
120
140
160
180
2002
2004
2006
2008
2010
2012
2014
Year

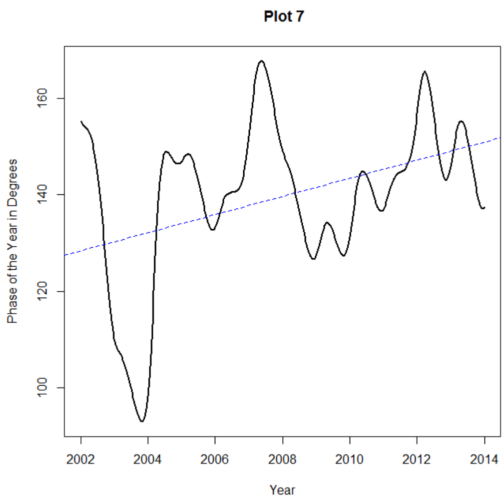

Plot 5
tree cover = 11%, grass cover = 85%
Harmonic F-test Statistic
1
2
5
10
20
50
95%
0.0
0.1
0.2
0.3
0.4
0.5
Frequency

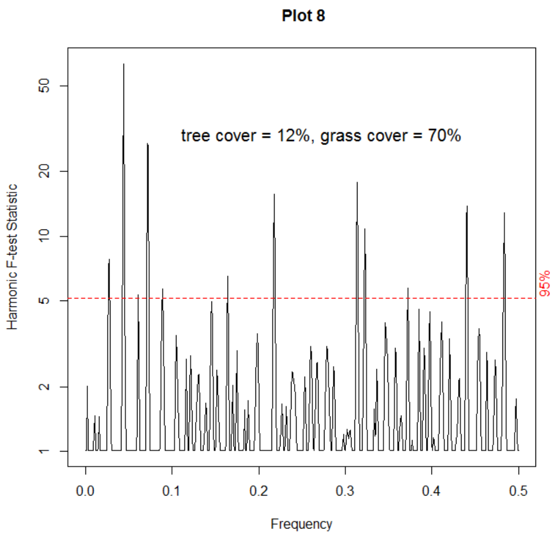

Plot 5
Phase of the Year in Degrees
100
120
140
160
2002
2004
2006
2008
2010
2012
2014
Year

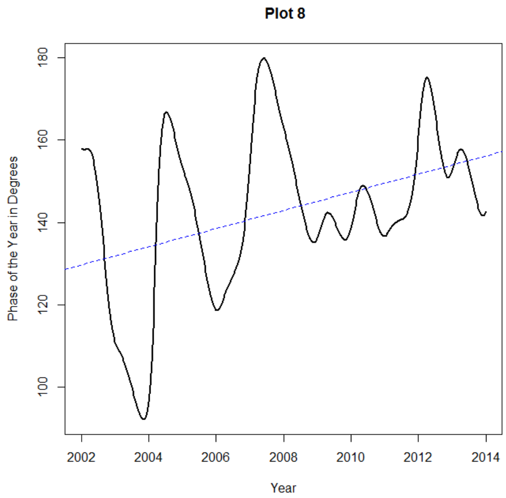

Plot 6
tree cover = 11%, grass cover = 42%
Harmonic F-test Statistic
1
2
5
10
20
50
100
95%
0.0
0.1
0.2
0.3
0.4
0.5
Frequency

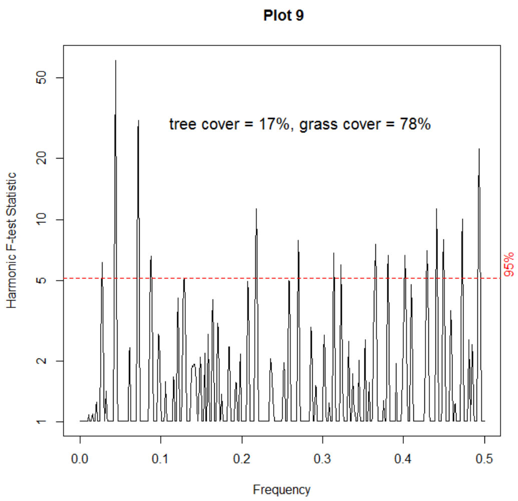

Plot 6
Phase of the Year in Degrees
100
110
120
130
140
150
160
170
2002
2004
2006
2008
2010
2012
2014
Year

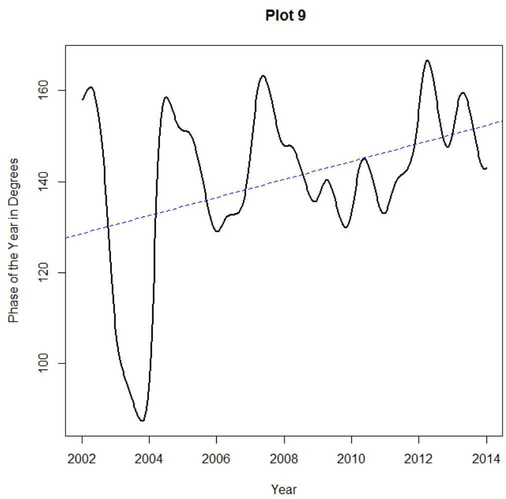

Plot 7
tree cover = 11%, grass cover = 42%
95%
Harmonic F-test Statistic
Frequency

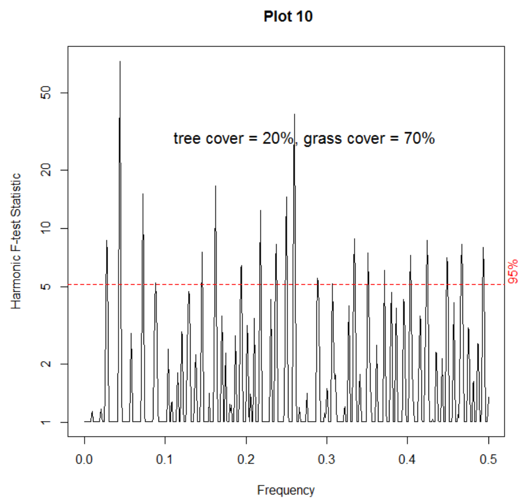

Plot 7
Phase of the Year in Degrees
Year

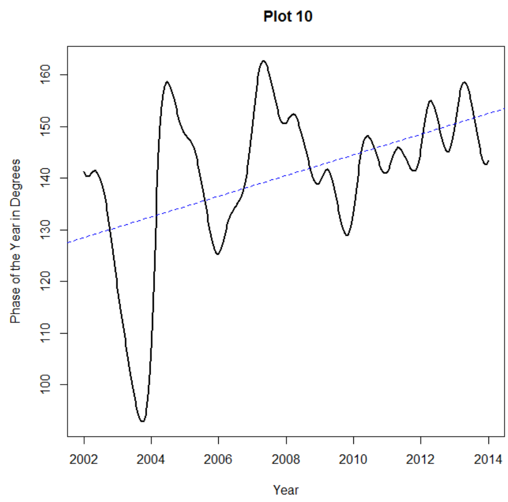

Plot 8
tree cover = 12%, grass cover = 70%
95%
Harmonic F-test Statistic
Frequency

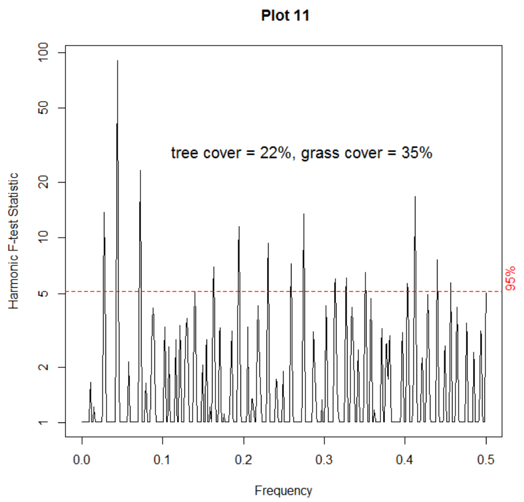

Plot 8
Phase of the Year in Degrees
Year

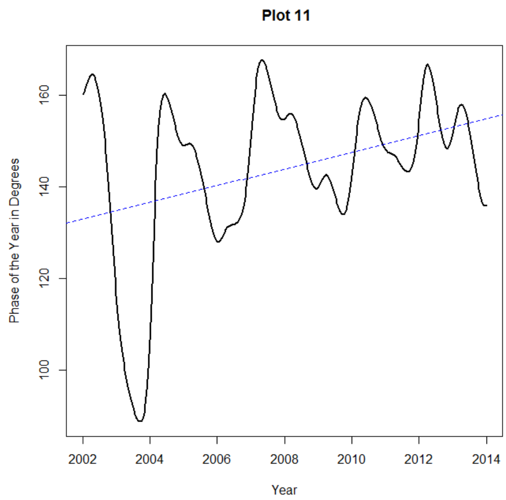

Plot 9
tree cover = 17%, grass cover = 78%
95%
Harmonic F-test Statistic
Frequency

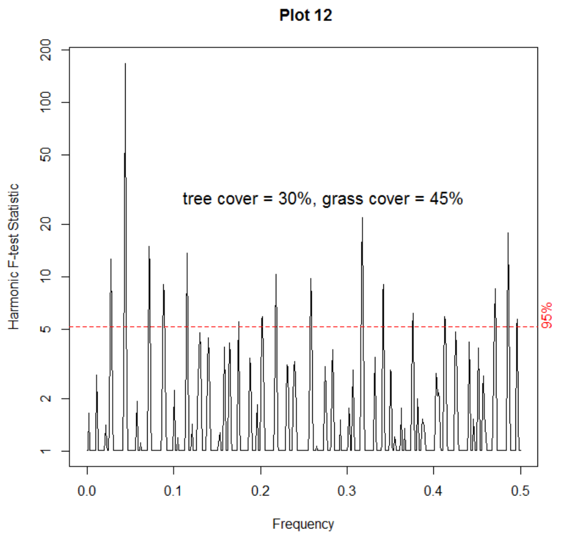

Plot 9
Phase of the Year in Degrees
Year

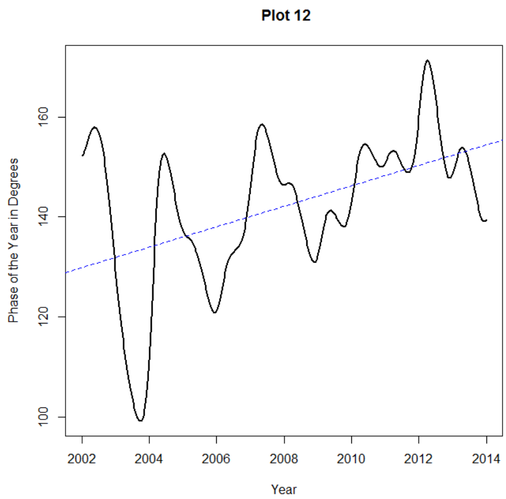

Plot 10
tree cover = 20%, grass cover = 70%
95%
Harmonic F-test Statistic
Frequency

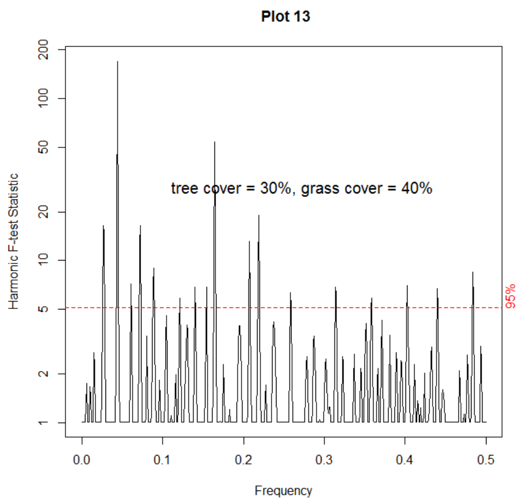

Plot 10
Phase of the Year in Degrees
Year

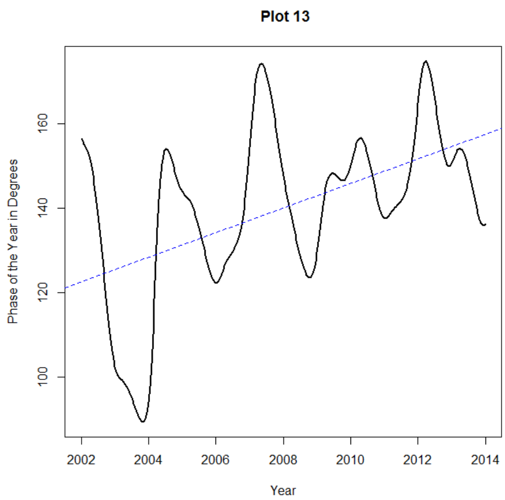

Plot 11
tree cover = 22%, grass cover = 35%
95%
Harmonic F-test Statistic
Frequency

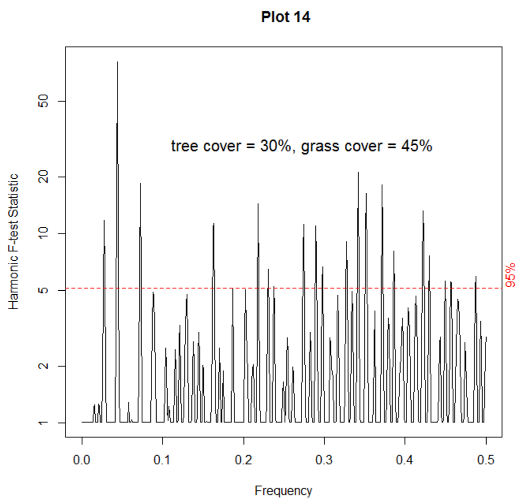

Plot 11
Phase of the Year in Degrees
Year

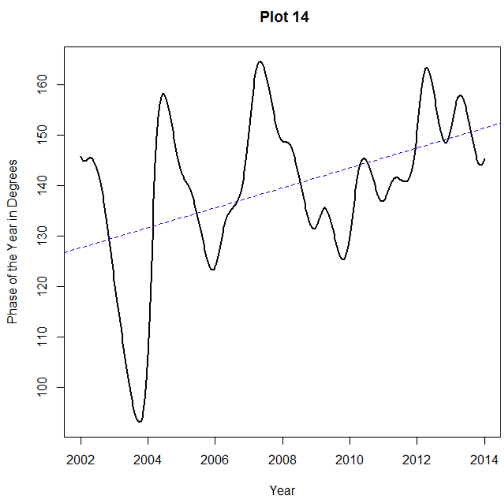

Plot 12
tree cover = 30%, grass cover = 45%
95%
Harmonic F-test Statistic
Frequency

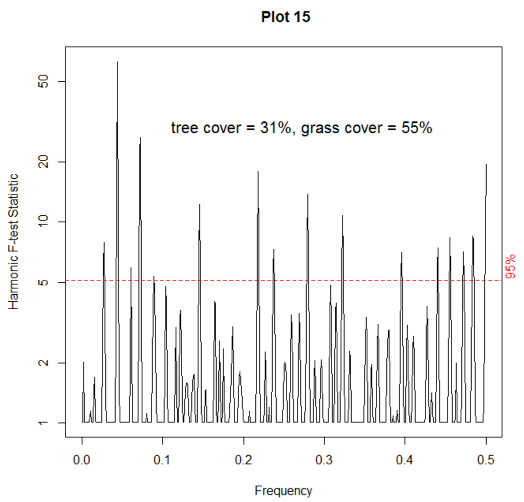

Plot 12
Phase of the Year in Degrees
Year

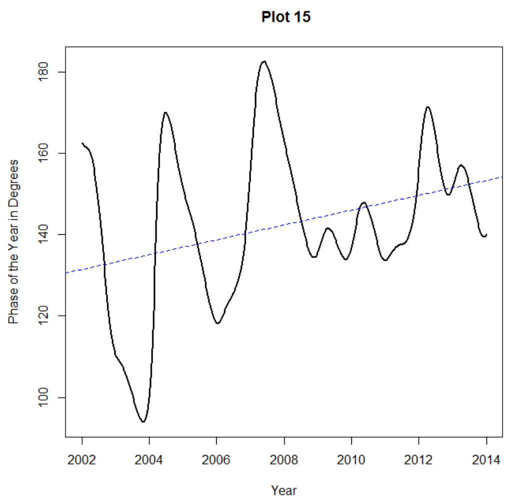

Plot 13
tree cover = 30%, grass cover = 40%
95%
Harmonic F-test Statistic
Frequency

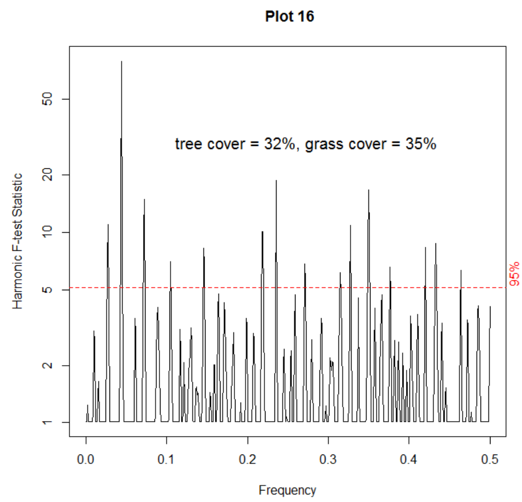

Plot 13
Phase of the Year in Degrees
Year

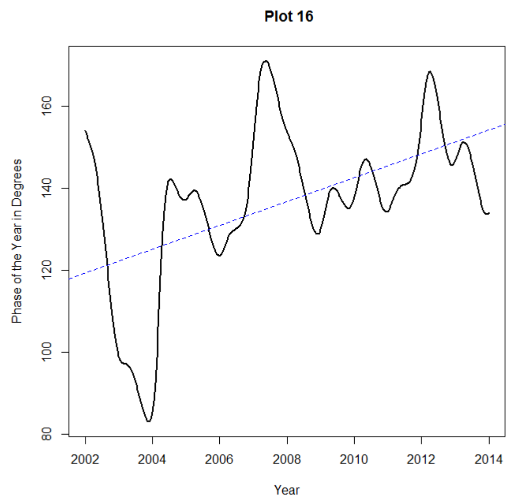

Plot 14
tree cover = 30%, grass cover = 45%
95%
Harmonic F-test Statistic
Frequency

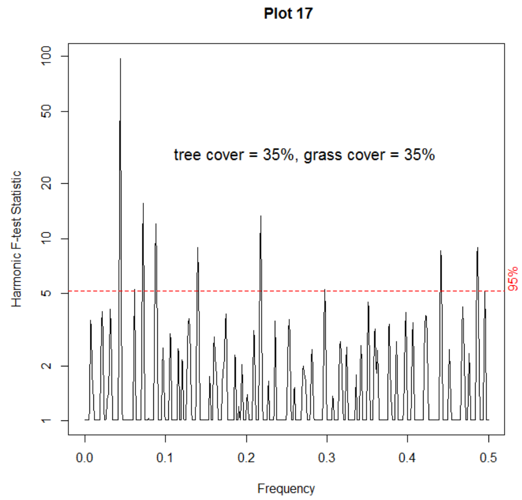

Plot 14
Phase of the Year in Degrees
Year

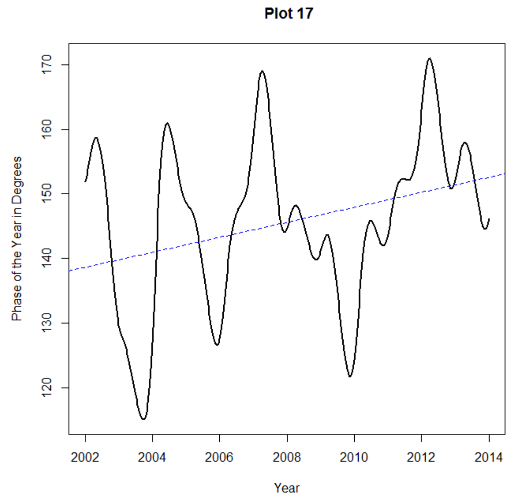

Plot 15
tree cover = 31%, grass cover = 55%
95%
Harmonic F-test Statistic
Frequency

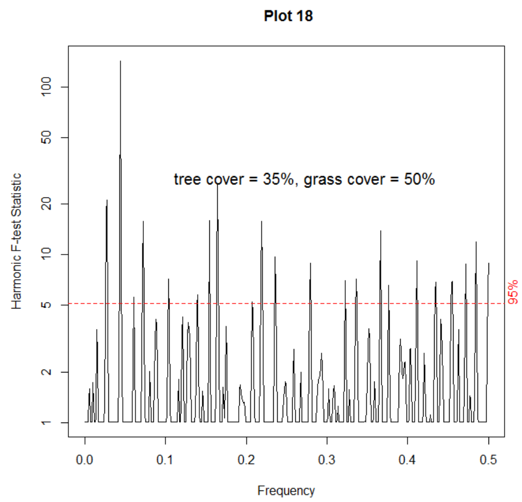

Plot 15
Phase of the Year in Degrees
Year

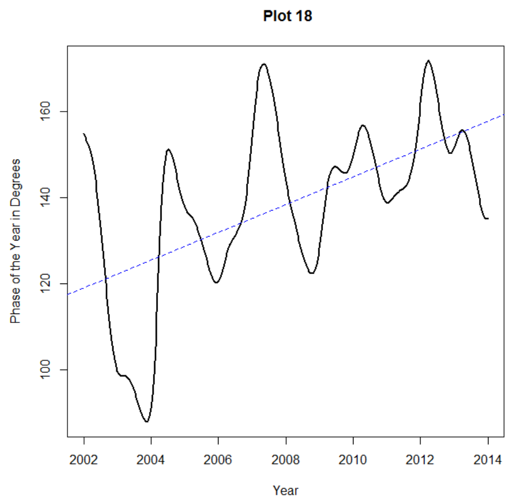

Plot 16
tree cover = 32%, grass cover = 35%
Harmonic F-test Statistic
Frequency
95%

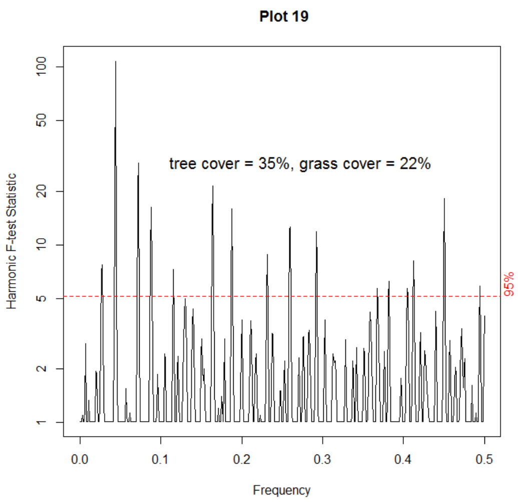

Plot 16
Phase of the Year in Degrees
Year

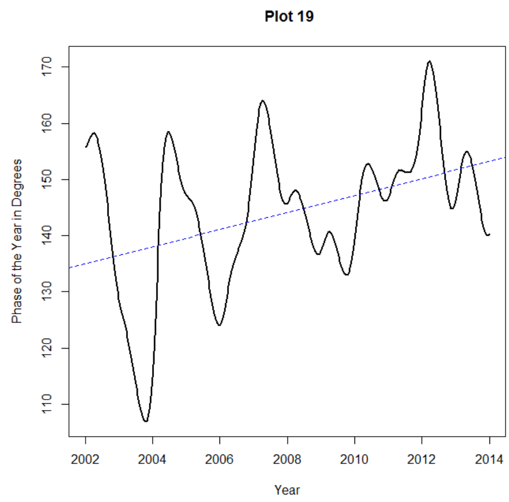

Plot 17
tree cover = 35%, grass cover = 35%
Harmonic F-test Statistic
Frequency
95%

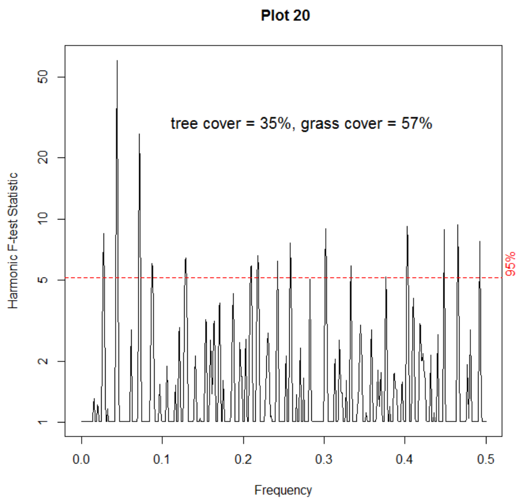

Plot 17
Phase of the Year in Degrees
Year

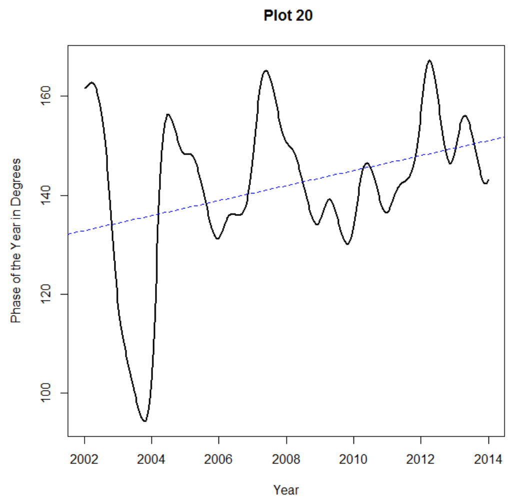

Plot 18
tree cover = 35%, grass cover = 50%
Harmonic F-test Statistic
Frequency
95%

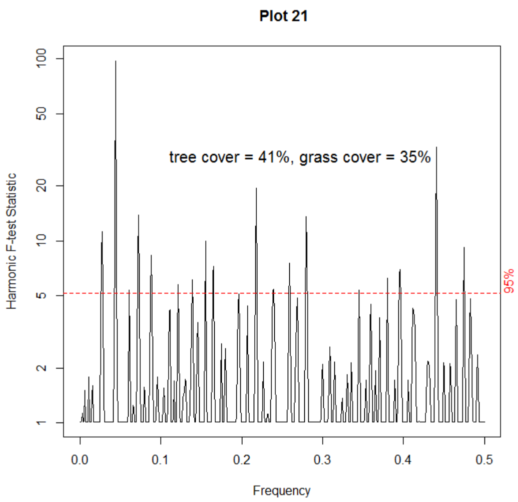

Plot 18
Phase of the Year in Degrees
Year

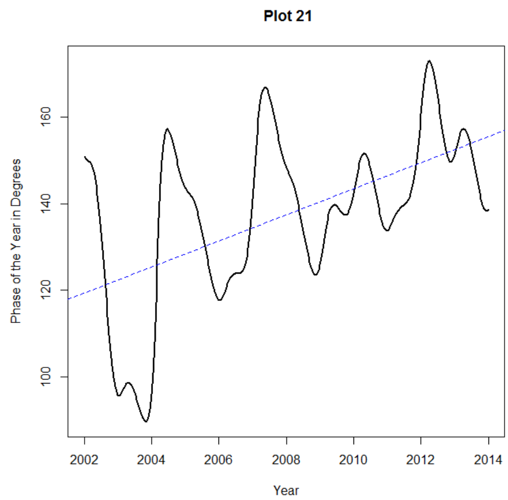

Plot 19
tree cover = 35%, grass cover = 22%
Harmonic F-test Statistic
Frequency
95%

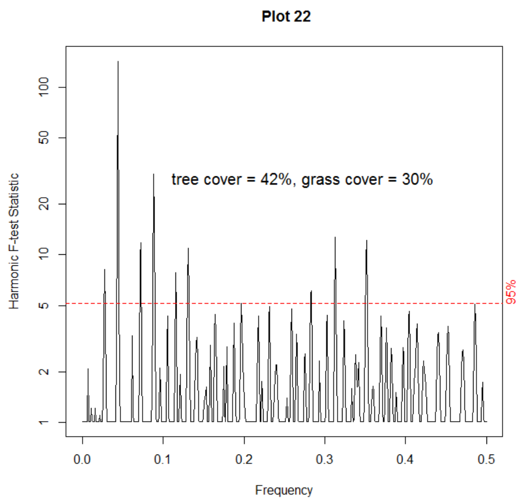

Plot 19
Phase of the Year in Degrees
Year

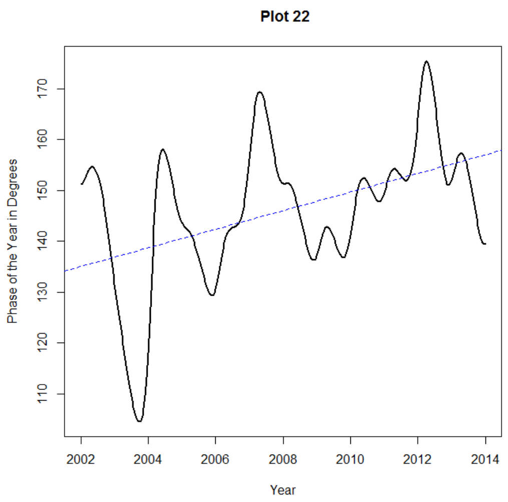

Plot 20
tree cover = 35%, grass cover = 57%
Harmonic F-test Statistic
Frequency
95%

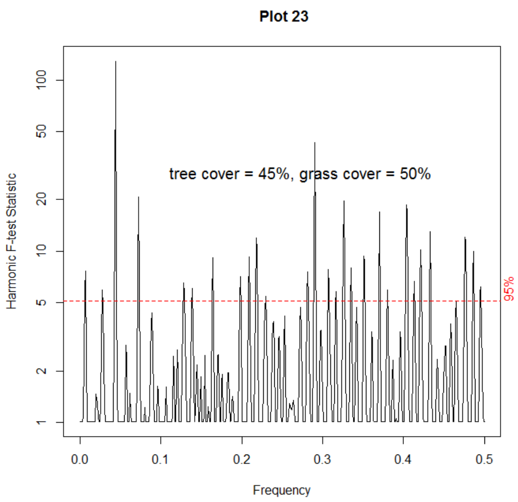

Plot 20
Phase of the Year in Degrees
Year

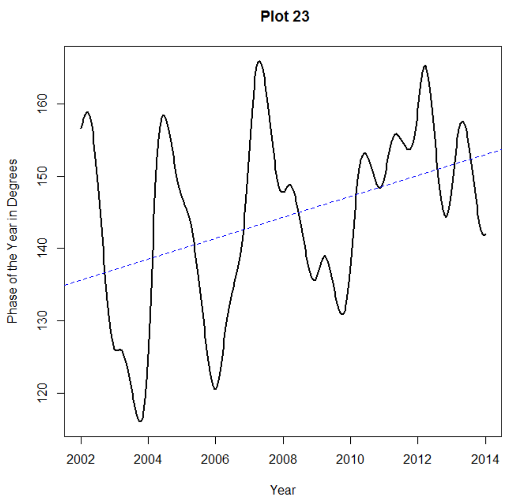

Plot 21
tree cover = 41%, grass cover = 35%
Harmonic F-test Statistic
Frequency
95%

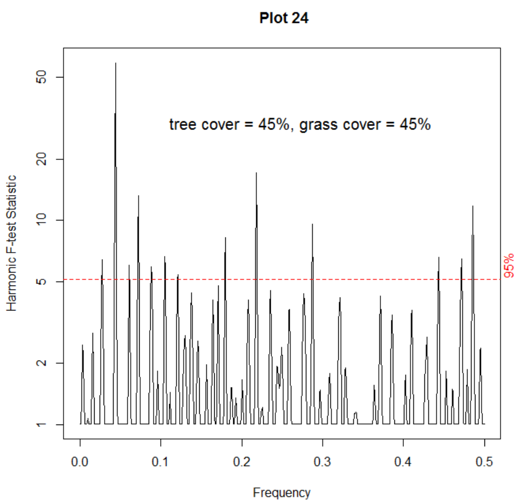

Plot 21
Phase of the Year in Degrees
Year

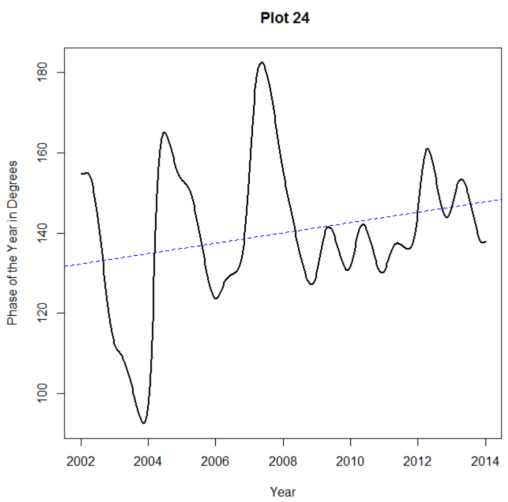

Plot 22
tree cover = 42%, grass cover = 30%
95%
Harmonic F-test Statistic
Frequency

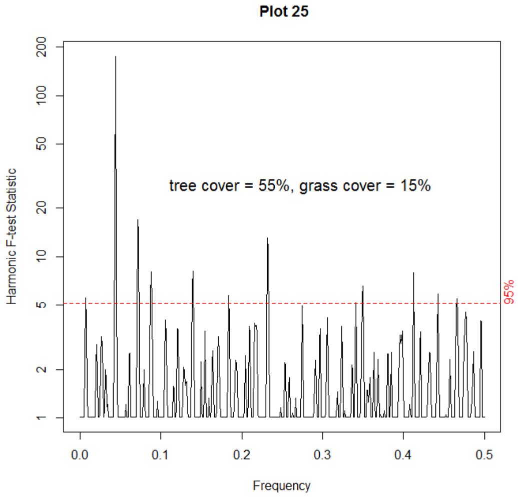

Plot 22
Phase of the Year in Degrees
Year

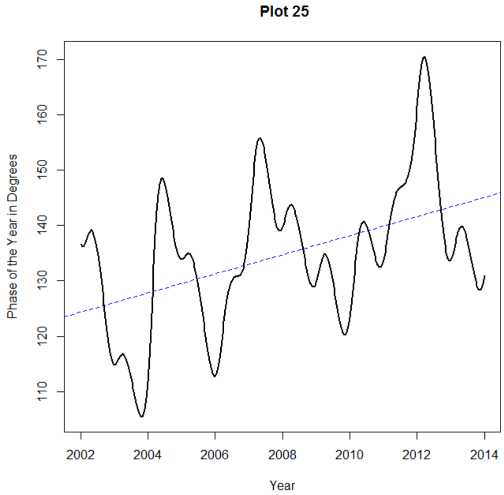

Plot 23
tree cover = 45%, grass cover = 50%
95%
Harmonic F-test Statistic
Frequency

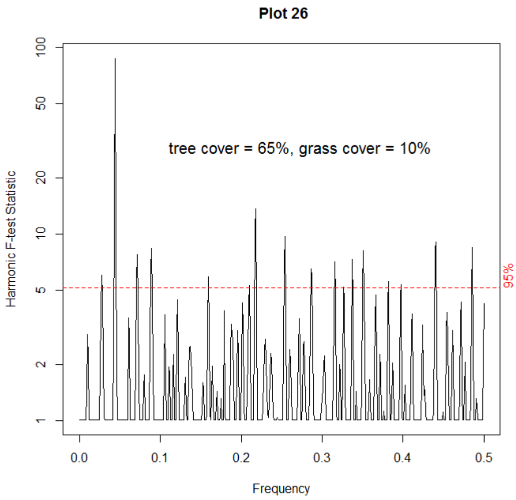

Plot 23
Phase of the Year in Degrees
Year

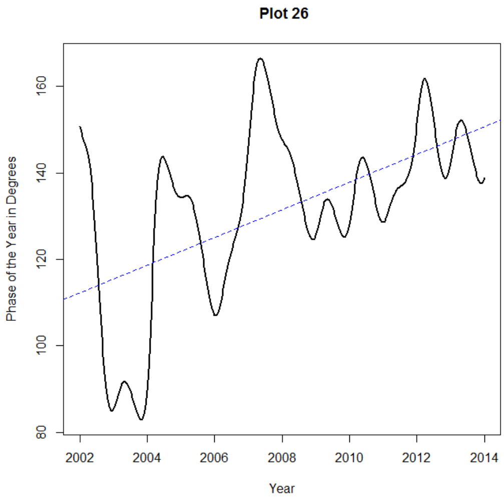

Plot 24
tree cover = 45%, grass cover = 45%
95%
Harmonic F-test Statistic
Frequency

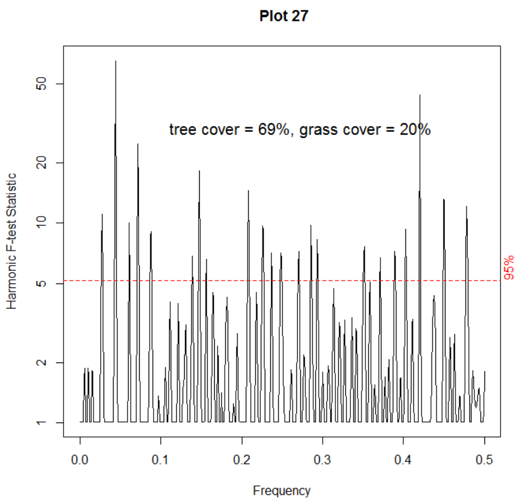

Plot 24
Phase of the Year in Degrees
Year

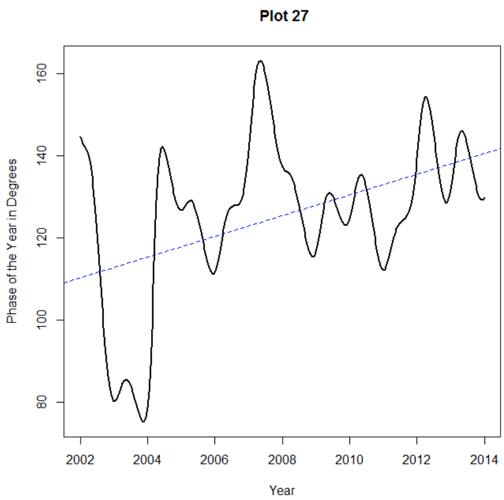

Plot 25
tree cover = 55%, grass cover = 15%
95%
Harmonic F-test Statistic
Frequency

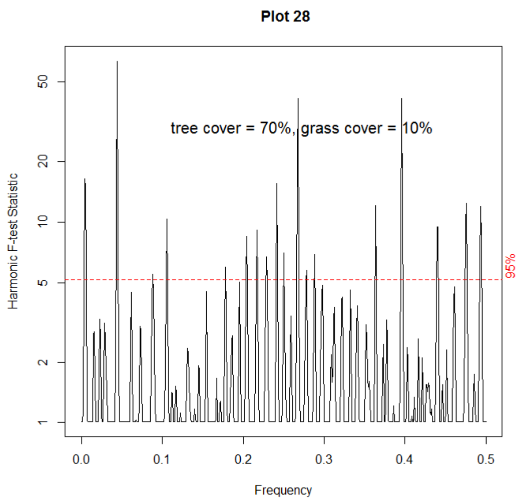

Plot 25
Phase of the Year in Degrees
Year

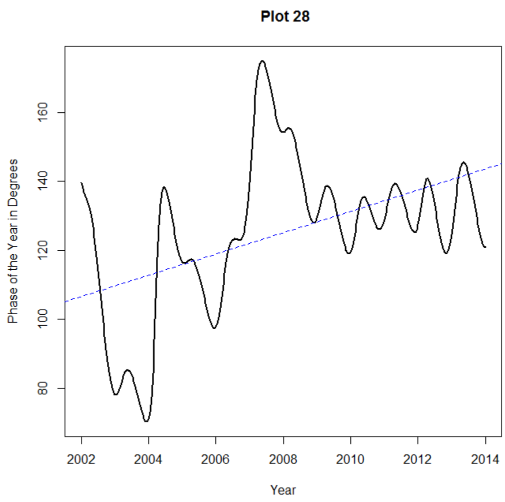

Plot 26
tree cover = 65%, grass cover = 10%
95%
Harmonic F-test Statistic
Frequency


Plot 26
Phase of the Year in Degrees
Year


Plot 27
tree cover = 69%, grass cover = 20%
95%
Harmonic F-test Statistic
Frequency


Plot 27
Phase of the Year in Degrees
Year


Plot 28
tree cover = 70%, grass cover = 10%
Harmonic F-test Statistic
Frequency
95%
1
2
5
10
20
50
0.0
0.1
0.2
0.3
0.4
0.5


Plot 28
Phase of the Year in Degrees
Year
80
100
120
140
160
2002
2004
2006
2008
2010
2012
2014

Appendix III. R codes for pixel-based signal decomposition of MODIS data

```
#########################################
# Generic package for harmonic analysis in R
#########################################
#
# written by Heiko Balzter, copyright 2014
# contact the author: hb91@le.ac.uk
#
# includes modified functions from source:
http://www.di.fc.ul.pt/~jpn/r/fourier/fourier.html
#
#########################################

# t is a time index from 0 to n-1 where n is the number of measurements in the time-
series
# xt is the vector of time points at which the measurements were taken
# The amplitude of a wave is defined as half the height from the maximum to the
minimum point.
# The phase of the wave is defined as the angle by which the sine wave is delayed to its
first peak.
# A harmonic term is defined by how many complete waves it has within the defined
time series, from start to end,
#    i.e. harmonic term 2 has two full waves (two maxima and two minima) within the
time series.
# The fundamental period is the period between the first sample and the last.
# The acquisition frequency is the number of measurements between two successive
units of time.
# The fundamental frequency f_0 is 1/N where N is the number of time steps.
# The frequencies of the wave components must be integer multiples of the fundamental
frequency.
# f_0 is called the first harmonic, the second harmonic is 2*f_0, the third is 3*f_0, etc.

get.trajectory <- function(X.k,xt,acq.freq) {
# Inverse Fourier Transform:
# Returns the x.n time series for a given time sequence (xt) and a vector with the
amount of frequencies k in the signal (X.k)
  n   <- length(xt)
  i   <- complex(real = 0, imaginary = 1)
  x.n <- rep(0,n)
  ks  <- 0:(length(X.k)-1)
  for(j in 0:(n-1)) { # compute each time point x_n based on freqs X.k
    x.n[j+1] <- sum(X.k * exp(i*2*pi*ks*j/n)) / n
  }
  x.n * n
}

plot.fourier <- function(fourier.series, f.0, xt, ...) {
# plot a Fourier series
```

```
# ***** This function has been verified. *****
  w <- 2*pi*f.0
  trajectory <- sapply(xt, function(t) fourier.series(t,w))
  plot(xt/length(xt), trajectory, type="l", xlab="time", ylab="f(t)"); abline(h=0,lty=3)
}

convert.fft <- function(x.k, acq.freq=1) {
# convert a FFT to amplitude and phase
# x.k is the vector of complex points to convert
  n <- length(Re(x.k)) # number of points
  x.k <- x.k / n # normalize
  distance.center <- function(c)signif( Mod(c),        4)
  angle           <- function(c)signif( 180*Arg(c)/pi, 3)
  df <- data.frame(cycle    = 0:(n-1),
                   freq     = 0:(n-1) / acq.freq,
                   t        = n / 0:(n-1) / acq.freq, # in time units, not sequential units
                   ampl     = sapply(x.k, distance.center) * 2 * n,
                   phase    = sapply(x.k, angle))
  df
}

plot.frequency.spectrum <- function(X.k, acq.freq=acq.freq, col = 1, lwd = 2, pch = "+",
...) {
# plot a frequency spectrum of a given X_k
  xax <- (0:(length(X.k)-1)) / length(X.k) * acq.freq
  xlimits <- c(0, max(xax)/2)
  plot.data  <- cbind( xax, 2 * Mod(X.k))
  plot(plot.data, t="h", main="Periodogram", xlab="Frequency", ylab="Power spectral
density",
      col = col, lwd = lwd, pch = pch, xlim=xlimits, ylim=c(0,max(Mod(plot.data[,2]))))
}

plot.harmonic <- function(xk, i, xt, acq.freq, mar=c(1,1,1,1),
  col = 3, lwd = 2, pch = "+", cex.lab = 1, cex.axis = 1, cex.main = 1, cex.sub = 1, ...) {
# plot.harmonic() plots the i-th harmonic on the current plot
# xk: the frequencies computed by the FFt
#  i: which harmonic
# xt: the sampling time points
# acq.freq: the acquisition rate
  xk.h <- rep(0,length(xk))
  xk.h[i+1] <- xk[i+1] # i-th harmonic
  harmonic.trajectory <- 2 * get.trajectory(xk.h, xt, acq.freq=acq.freq)
  points(xt, Re(harmonic.trajectory), type="l", mar=mar,
      col = col, lwd = lwd, pch = pch, cex.lab = cex.lab, cex.axis = cex.axis,
      cex.main = cex.main, cex.sub = cex.sub)
}

get.harmonic <- function(xk, i, xt, acq.freq) {
# Get the values that define the i-th harmonic term.
# xk: the frequencies computed by the FFt
```

```
#  i: which harmonic term(s)
# xt: the sampling time points
# acq.freq: the acquisition rate
  xk.h <- rep(0,length(xk))
  xk.h[i+1] <- xk[i+1] # i-th harmonic
  harmonic.trajectory <- 2 * get.trajectory(xk.h, xt, acq.freq=acq.freq)
  Re(harmonic.trajectory)
}

harmonic <- function(xt, x, acq.freq, N, alpha, detrend, which, test, ...) {
# core harmonic analysis function
# xt = a vector of time steps in units of s,min, hr or other time units, does not have to be integers
# x = a vector of time-series observations with the same length as xt
# N = number of the harmonic terms to be included, starting with largest amplitude
# acq.freq = number of measurements between two successive units of time.
# alpha = type I error probability for statistical significance testing (default 0.05 or 5%)
# detrend = TRUE or FALSE, if TRUE (default) then linear detrending is applied to the data
# which = "strongest": the N strongest harmonic terms are included in the model (default)
#       = "first": the first N terms are included, or
#       = "all": all harmonics are included.
# test = "bonferroni" adjusts type I error by the number of tests N; "holm" adjusts the type I error by N+1-k where k=1:N
#
# How to understand the harmonic terms:
# Cycle means the number of waves in the time series, i.e. cycle = 9 the wave fits 9 times into the length of the data
#      which is the annual cycle for a 9-year series.
# Freq is the position in the frequency domain (periodogram) from 0 to 0.5.
# t is the position in the time domain from 0 to n-1 where n is the number of measurements.
# Amplitude is the strength of the wave and
# phase is the delay of the wave in degrees (0-360).

  dig.aov <- 4 # number of significant digits for ANOVA table
  if (missing(acq.freq)) acq.freq <- 1
  if (missing(N)) N <- 20
  if (missing(alpha)) alpha <- 0.05
  if (missing(detrend)) detrend <- TRUE
  if (missing(which)) which <- "strongest"
  if (!(which %in% c("strongest", "first", "all"))) which <- "strongest"
  if (missing(test)) test <- "bonferroni"
  if (!(test %in% c("bonferroni", "holm"))) test <- "bonferroni"
  t <- 0:(n-1)
  # detrending
  n <- length(x)
  if (detrend) {
    trend <- lm(x~xt) # linear model
```

```
    cat("Linear detrending result:\n")
    print(summary(trend))
    detrended.x <- trend$residuals
    } else {
    cat("No detrending.\n\n")
    detrended.x <- x
    trend <- "No detrending"
    }

  detrended.x.k <- fft(detrended.x) / n
  windows()
  plot.frequency.spectrum(detrended.x.k, acq.freq=acq.freq)

  # Calculate the amplitude and phase angle for the N harmonic terms
  # Cycle 9 means that the harmonic wave repeats 9 times over the time series
  # Note that cycle is indexed from 0 and ndx from 1
  x.fft <- convert.fft(detrended.x.k, acq.freq)
  nx <- length(x.fft$cycle)
  # you can get the components of the table from:
  #   x.fft$cycle[ndx]
  #   x.fft$freq[ndx]
  #   x.fft$t[ndx]
  #   x.fft$ampl[ndx]
  #   x.fft$phase[ndx]

  # find the N strongest harmonics
  if (which=="strongest") {
    ndx <- order(x.fft$ampl[1:(nx/2)], decreasing = T)[1:N]
    #print(cbind(ndx, x.fft$ampl[ndx]))
    cat(paste(N, "strongest harmonic terms:\n", sep=" "))
    write.table(round(x.fft[ndx,],4), quote = F, sep = "\t", row.names=F)
    cat("\n")
    }
  if (which=="first") {
    ndx <- 2:(N+1)
    #print(cbind(ndx, x.fft$ampl[ndx]))
    cat(paste(N, "first harmonic terms:\n", sep=" "))
    write.table(round(x.fft[ndx,],4), quote = F, sep = "\t", row.names=F)
    cat("\n")
    }
  if (which=="all") {
    ndx <- 1:n
    cat(paste("All harmonic terms:\n", sep=" "))
    write.table(round(x.fft[ndx,],4), quote = F, sep = "\t", row.names=F)
    cat("\n")
    }

  # test for significance of the individual harmonic terms using the F test
  # Reference: Hartley, H. O. (1949): Tests of Significance in Harmonic Analysis.
Biometrika, 36, 194-201.
```

```
  # time dimension ts is in arbitrary units, with acq.freq measurements between two
successive units 1 and 2, say
  # without loss of generality, for the purpose of the significance testing we treat the
  #     time dimension as steps of 1, 2, ..., n
  # we do this by adjusting the time index tn <- xt * acq.freq

  if (which=="strongest" || which=="first") {
  bonferroni <- alpha / N   # adjusted alpha probability by the number of comparisons,
Bonferroni correction
  holm <- alpha / seq(N, 1, -1)   # adjusted alpha probability, Bonferroni/Holm
correction
  gamma = 2*pi/n # in Hartley's paper, but only for time steps of 1
  a0 <- mean(detrended.x)
  # work out the mean squares (MSQ) of each harmonic term for ANOVA table
  ssq <- rep(0, N) # SSQ components
  df <- rep(2, N) # degrees of freedom
  msq <- rep(0, N) # MSQ components = SSQ / df
  a <- rep(0, N) # ai
  b <- rep(0, N) # bi
  f <- rep(0, N) # F values for each harmonic term
  p <- rep(0, N) # p values for each harmonic term
  sig <- rep("n.s.", N) # significance
  for (i in 1:N) {
    a[i] <- 2/n * sum(detrended.x * cos(x.fft$cycle[ndx[i]] * t * gamma)) # note that we
use t here and not xt, see above
    b[i] <- 2/n * sum(detrended.x * sin(x.fft$cycle[ndx[i]] * t * gamma))
  }
  # calculate SSQ terms
  ssq <- n/2*(a^2+b^2)
  # calculate MSQ terms
  msq <- ssq/df
  # Calculate the residual MSQ variance component:
  rssq <- sum((detrended.x-a0)^2) - n/2 * sum(a^2+b^2)
  # The total df is n-1. The residual df is the total n – 2N - 1.
  rdf <- n-2*N-1 # residual df
  rmsq <- rssq/rdf
  # Work out the F values:
  f <- msq / rmsq
  # Each harmonic term has 2 degrees of freedom since it is characterised by 2
parameters ai and bi.
  # The F ratio is calculated by dividing the MSQ of each harmonic term by the residual
MSQ. It should be compared to the F distribution for 2,11 degrees of freedom for the
5%/m point, assuming type I error is controlled at 5%.
  p <- pf(f, df1=2, df2=n-N*2-1, lower=FALSE)
  # rounding for pretty printing:
  ssq <- round(ssq, 2)
  msq <- round(msq, 2)
  f <- round(f, 1)
  p <- round(p, 5)
  bonferroni <- round(bonferroni, 5)
```

```
  holm <- round(holm, dig.aov)
  rssq <- round(rssq, dig.aov)
  rmsq <- round(rmsq, dig.aov)
  # now merge all into a data.frame
  if (test=="bonferroni") {
    sig[p<bonferroni] <- "*"
    x.aov <- as.data.frame(cbind(x.fft$cycle[ndx], ssq, df, msq, f, p, bonferroni, sig))
    names(x.aov) <- c("cycle", "SSQ", "df", "MSQ", "F", "p", "pBonf", "Sig")
    }
  if (test=="holm") {
    sig[p<holm] <- "*"
    for (i in 1:(N-1)) if (p[i]>=holm[i]) sig[(i+1):N] <- rep("n.s.", N-i)
    x.aov <- as.data.frame(cbind(x.fft$cycle[ndx], ssq, df, msq, f, p, holm, sig))
    names(x.aov) <- c("cycle", "SSQ", "df", "MSQ", "F", "p", "pHolm", "Sig")
    }
  # print it
  cat("ANOVA table for the selected harmonic terms:\n")
  write.table(format(x.aov, trim = FALSE, justify = "right"), quote = F, sep = "\t",
row.names=F)
  write.table(format(cbind("Res.", rssq, rdf, rmsq), trim = FALSE, justify = "right"),
quote = F, sep = "\t", row.names=F, col.names=F)
  # cat("Res.", round(rssq, dig.aov), rdf, round(rmsq, dig.aov), "\n")
  cat("\n")
  # print(x.aov)
  x.aov <- merge.data.frame(x.aov, data.frame(c(NA, round(rssq,dig.aov), rdf,
round(rmsq,dig.aov), NA,NA,NA,NA), row.names = names(x.aov)))

  if (which=="strongest") { # select only the significant harmonic terms
    ndxs <- ndx[sig=="*"]
    N <- length(ndxs)
    cat("\nRetaining ", N, "significant harmonic terms.\n")
    }
  if (which=="first") { # select the first N harmonic terms
    ndxs <- ndx
    cat("\nRetaining the first", N, " harmonic terms.\n")
    }

  # plot detrended time series and overlay the individual N significant harmonics with
the largest amplitudes:
  # only plot up to 40 harmonics
  x.n <- get.trajectory(detrended.x.k, xt, acq.freq)  # create time wave from detrended
data (if detrending is switched on)
  windows()
  par(mfrow = c(1,1))
  plot(xt, Re(x.n), type="l", lwd=1, main="(d) Significant harmonic terms of tree/grass
signals")
  abline(h=0,lty=2)
  for (i in 1:min(40, N)) plot.harmonic(detrended.x.k, ndxs[i], xt, acq.freq, col=i+1)
```

```
  # Now plot detrended time series and the composite of the first significant N
harmonics:
  windows()
  plot(xt, Re(x.n), type="l", lwd=1, main="(c) Detrended tree/grass signals")
  abline(h=0, lty=2)
  wave <- get.harmonic(detrended.x.k, ndxs, xt, acq.freq)
  lines(xt, wave, col="red")

  # And now plot add the trend back on to the composite of the first N harmonics:
  if (detrend) {
    windows()
    plot(xt, x, type="l",lwd=1, main= "(b) Trend of tree/grass signals")
    abline(trend)
    wave <- wave + trend$coef[1] + trend$coef[2] * xt
    lines(xt, wave, col="red")
    }

  # plot residuals
    windows()
    plot(xt, x - wave, type="p", pch="+", main="Residuals of tree/grass signals")
  }
  else
  { # if which == "all"
  ndxs <- 2:n # in case all harmonics will be included except term 0
  N <- n
  x.aov <- "No ANOVA available. All harmonics are included."
  wave <- x
  }

  # return these components:
  harm <- list(xt)
  names(harm) <- "xt"
  harm$lm <- trend
  harm$detrended <- detrended.x
  harm$Nsig <- N
  harm$ndx <- ndx
  harm$frequency.spectrum <- detrended.x.k
  harm$Amp <- x.fft$ampl[ndx]
  harm$Ph <- x.fft$phase[ndx]
  harm$aov <- x.aov
  harm$fitted.values <- wave
  harm$residuals <- x - wave
  harm$call <- match.call()
  class(harm) <- "harmonic"

  # return results
  harm
}
##################
# end of functions
```

```
###################

###############
# verification:
###############


# file name and number of header lines in the file to skip
ndvi <- read.table("11_4168_2383.txt", header=T)
# column names of the file
ndvi <- ndvi[,1] # use first column
ndvi<- ndvi # extract biweekly ndvi measurements
time     <- length(ndvi)               # measuring time interval (in months)
acq.freq <- 1                         # data acquisition frequency (Hz), how many
measurements per month
ts <- seq(0,time-1/acq.freq,1/acq.freq)  # vector of sampling time-points in months
n <- length(ndvi)
f.0 <- 1/time
plot(ndvi, type="l", xlab="t", ylab="NDVI values",main="Tree/grass signals") # show
the data

# with Bonferroni adjusted alpha error
tmax.harm <- harmonic(ts, ndvi, N=100, alpha=0.05, detrend=TRUE,
test="bonferroni")

# with Holm adjusted alpha error
tmax.harm <- harmonic(ts, ndvi, N=100, alpha=0.05, detrend=TRUE, test="holm")

sum(ndvi.harm$Amp)
```

Appendix IV. R codes for image signal decomposition of MODIS data

Written by Heiko Balzter, copyright 2014, modified by Narumasa Tsutsumida

```
 harmonic for matrix to extend spatiotemporal data -------------------------------------------
-
##harmonic function for apply
harmonic4matrix <- function(x,xt, acq.freq,n, N, alpha, detrend, which, test, ...) {
  # core harmonic analysis function
  # xt = a vector of time steps in units of s,min, hr or other time units, does not have to
be integers
  # x = a vector of time-series observations with the same length as xt
  # N = number of the harmonic terms to be included, starting with largest
amplitude.ã€€"m" in paper (Harley 1949)ã€€
  # acq.freq = number of measurements between two successive units of time.
  # alpha = type I error probability for statistical significance testing (default 0.05 or 5%)
  # detrend = TRUE or FALSE, if TRUE (default) then linear detrending is applied to
the data
  # which = "strongest": the N strongest harmonic terms are included in the model
(default)
  #       = "first": the first N terms are included, or
  #       = "all": all harmonics are included.
  # test = "bonferroni" adjusts type I error by the number of tests N; "holm" adjusts the
type I error by N+1-k where k=1:N
  #
  # How to understand the harmonic terms:
  # Cycle means the number of waves in the time series, i.e. cycle = 9 the wave fits 9
times into the length of the data
  #      which is the annual cycle for a 9-year series.
  # Freq is the position in the frequency domain (periodogram) from 0 to 0.5.
  # t is the position in the time domain from 0 to n-1 where n is the number of
measurements.
  # Amplitude is the strength of the wave and
  # phase is the delay of the wave in degrees (0-360).

  dig.aov <- 4 # number of significant digits for ANOVA table
  if (missing(acq.freq)) acq.freq <- 1
  if (missing(N)) N <- 20
  if (missing(alpha)) alpha <- 0.05
  if (missing(detrend)) detrend <- FALSE
  if (missing(which)) which <- "strongest"
  if (!(which %in% c("strongest", "first", "all"))) which <- "strongest"
  if (missing(test)) test <- "holm"
  if (!(test %in% c("bonferroni", "holm"))) test <- "holm"
  t <- 0:(n-1)
  # detrending
 # n <- length(x)     #comment out
  if (detrend) {
```

```
  trend <- lm(x~xt) # linear model
  #cat("Linear detrending result:\n")
  #print(summary(trend))
  detrended.x <- trend$residuals
} else {
  #cat("No detrending.\n\n")
  detrended.x <- x
  trend <- "No detrending"
}

detrended.x.k <- fft(detrended.x) / n
#windows()
#plot.frequency.spectrum(detrended.x.k, acq.freq=acq.freq)

# Calculate the amplitude and phase angle for the N harmonic terms
# Cycle 9 means that the harmonic wave repeats 9 times over the time series
# Note that cycle is indexed from 0 and ndx from 1
x.fft <- convert.fft(detrended.x.k, acq.freq)
nx <- length(x.fft$cycle)
# you can get the components of the table from:
#   x.fft$cycle[ndx]
#   x.fft$freq[ndx]
#   x.fft$t[ndx]
#   x.fft$ampl[ndx]
#   x.fft$phase[ndx]

# find the N strongest harmonics
if (which=="strongest") {
  ndx <- order(x.fft$ampl[1:(nx/2)], decreasing = T)[1:N]
  #print(cbind(ndx, x.fft$ampl[ndx]))
  #cat(paste(N, "strongest harmonic terms:\n", sep=" "))
  #write.table(round(x.fft[ndx,],4), quote = F, sep = "\t", row.names=F)
  #cat("\n")
}
if (which=="first") {
  ndx <- 2:(N+1)
  #print(cbind(ndx, x.fft$ampl[ndx]))
  #cat(paste(N, "first harmonic terms:\n", sep=" "))
  #write.table(round(x.fft[ndx,],4), quote = F, sep = "\t", row.names=F)
  #cat("\n")
}
if (which=="all") {
  ndx <- 1:n
  #cat(paste("All harmonic terms:\n", sep=" "))
  #write.table(round(x.fft[ndx,],4), quote = F, sep = "\t", row.names=F)
  #cat("\n")
}

# test for significance of the individual harmonic terms using the F test
```

```
  # Reference: Hartley, H. O. (1949): Tests of Significance in Harmonic Analysis.
Biometrika, 36, 194-201.
  # time dimension ts is in arbitrary units, with acq.freq measurements between two
successive units 1 and 2, say
  # without loss of generality, for the purpose of the significance testing we treat the
  #     time dimension as steps of 1, 2, ..., n
  # we do this by adjusting the time index tn <- xt * acq.freq

  if (which=="strongest" || which=="first") {
    bonferroni <- alpha / N   # adjusted alpha probability by the number of comparisons,
Bonferroni correction
    holm <- alpha / seq(N, 1, -1)   # adjusted alpha probability, Bonferroni/Holm
correction
    gamma = 2*pi/n # in Hartley's paper, but only for time steps of 1
    a0 <- mean(detrended.x)
    # work out the mean squares (MSQ) of each harmonic term for ANOVA table
    ssq <- rep(0, N) # SSQ components
    df <- rep(2, N) # degrees of freedom
    msq <- rep(0, N) # MSQ components = SSQ / df
    a <- rep(0, N) # ai
    b <- rep(0, N) # bi
    f <- rep(0, N) # F values for each harmonic term
    p <- rep(0, N) # p values for each harmonic term
    sig <- rep("n.s.", N) # significance
    for (i in 1:N) {
      a[i] <- 2/n * sum(detrended.x * cos(x.fft$cycle[ndx[i]] * t * gamma)) # note that we
use t here and not xt, see above
      b[i] <- 2/n * sum(detrended.x * sin(x.fft$cycle[ndx[i]] * t * gamma))
    }
    # calculate SSQ terms
    ssq <- n/2*(a^2+b^2)    #a^2+b^2 is the observed intensities
    # calculate MSQ terms
    msq <- ssq/df     # m intensities are all independent kai-squere values, each based on
two degree of freedom
    # Calculate the residual MSQ variance component:
    rssq <- sum((detrended.x-a0)^2) - n/2 * sum(a^2+b^2)    #eq (3) in Harley's paper
    # The total df is n-1. The residual df is the total n ? 2N - 1.
    rdf <- n-2*N-1 # residual df : n - (N cos wave + N sin wave + a0)
    rmsq <- rssq/rdf
    # Work out the F values:
    f <- msq / rmsq    #eq (5) in Harley's paper
    # Each harmonic term has 2 degrees of freedom since it is characterised by 2
parameters ai and bi.
    # The F ratio is calculated by dividing the MSQ of each harmonic term by the
residual MSQ. It should be compared to the F distribution for 2,11 degrees of freedom
for the 5%/m point, assuming type I error is controlled at 5%.
    p <- pf(f, df1=2, df2=n-N*2-1, lower=FALSE)
    # rounding for pretty printing:
    ssq <- round(ssq, 2)
    msq <- round(msq, 2)
```

```
   f <- round(f, 1)
   p <- round(p, 5)
   bonferroni <- round(bonferroni, 5)
   holm <- round(holm, dig.aov)
   rssq <- round(rssq, dig.aov)
   rmsq <- round(rmsq, dig.aov)
   # now merge all into a data.frame
   if (test=="bonferroni") {
     sig[p<bonferroni] <- "*"
     x.aov <- as.data.frame(cbind(x.fft$cycle[ndx], ssq, df, msq, f, p, bonferroni, sig))
     names(x.aov) <- c("cycle", "SSQ", "df", "MSQ", "F", "p", "pBonf", "Sig")
   }
   if (test=="holm") {
     sig[p<holm] <- "*"
     for (i in 1:(N-1))
       if (p[i]>=holm[i]) sig[(i+1):N] <- rep("n.s.", N-i)
       x.aov <- as.data.frame(cbind(x.fft$cycle[ndx], ssq, df, msq, f, p, holm, sig))
     names(x.aov) <- c("cycle", "SSQ", "df", "MSQ", "F", "p", "pHolm", "Sig")
   }
   # print it
   #cat("ANOVA table for the selected harmonic terms:\n")
   #write.table(format(x.aov, trim = FALSE, justify = "right"), quote = F, sep = "\t",
row.names=F)
   #write.table(format(cbind("Res.", rssq, rdf, rmsq), trim = FALSE, justify = "right"),
quote = F, sep = "\t", row.names=F, col.names=F)
   # cat("Res.", round(rssq, dig.aov), rdf, round(rmsq, dig.aov), "\n")
   #cat("\n")
   # print(x.aov)
   x.aov <- merge.data.frame(x.aov, data.frame(c(NA, round(rssq,dig.aov), rdf,
round(rmsq,dig.aov), NA,NA,NA,NA), row.names = names(x.aov)))

   if (which=="strongest") { # select only the significant harmonic terms
     ndxs <- ndx
     #ndxs <- ndx[sig=="*"]
     N <- length(ndxs)
     #  cat("\nRetaining ", N, "significant harmonic terms.\n")
   }
   if (which=="first") { # select the first N harmonic terms
     ndxs <- ndx
     #  cat("\nRetaining the first", N, " harmonic terms.\n")
   }

   # plot detrended time series and overlay the individual N significant harmonics with
the largest amplitudes:
   # only plot up to 40 harmonics
   # x.n <- get.trajectory(detrended.x.k, xt, acq.freq)  # create time wave from detrended
data (if detrending is switched on)
   #windows()
   #par(mfrow = c(1,1))
   #plot(xt, Re(x.n), type="l", lwd=1)
```

```
    #abline(h=0,lty=2)
    #for (i in 1:min(40, N)) plot.harmonic(detrended.x.k, ndxs[i], xt, acq.freq, col=i+1)

    # Now plot detrended time series and the composite of the first significant N
harmonics:
    #windows()
    #plot(xt, Re(x.n), type="l", lwd=1)
    #abline(h=0, lty=2)
    wave <- get.harmonic(detrended.x.k, ndxs, xt, acq.freq)
    #lines(xt, wave, col="red")

    # And now plot add the trend back on to the composite of the first N harmonics:
    if (detrend) {
      #windows()
      #  plot(xt, x, type="l",lwd=1)
      #  abline(trend)
      wave <- wave + trend$coef[1] + trend$coef[2] * xt
      #  lines(xt, wave, col="red")
    }

    # plot residuals
    #windows()
    #  plot(xt, x - wave, type="p", pch="+")
  }
  else
  { # if which == "all"
    ndxs <- 2:n # in case all harmonics will be included except term 0
    N <- n
    x.aov <- "No ANOVA available. All harmonics are included."
    wave <- x
  }

  # return these components:
  #change
  #harm <- list(xt)                              #comment out
  harm <- list()                                #comment out
  #names(harm) <- "xt"                           #comment out
  #harm$lm <- trend                              #comment out
  #harm$detrended <- detrended.x                 #comment out
  harm$Nsig <- N
  #harm$ndx <- ndx                               #comment out
  #harm$frequency.spectrum <- detrended.x.k      #comment out
  #harm$Amp <- x.fft$ampl[ndx]                   #comment out
  harm$Amp <- x.fft$ampl[ndxs]                   #change ndx to ndxs
  harm$Ph <- x.fft$phase[ndx]
  #harm$aov <- x.aov                             #comment out
  #harm$fitted.values <- wave                    #comment out
  #harm$residuals <- x - wave                    #comment out
```

```
  harm$cycle <- x.fft$cycle[ndxs]                    #add
  harm$res <- trend$coefficients[1]               #add
  harm$slope <- trend$coefficients[2]               #add
  harm$lmPval <- summary(trend)$coefficients[2,4]     #add


  #harm$call <- match.call()
  #class(harm) <- "harmonic"

  # return results
  harm
}

##################
# end of functions
##################

library(zoo)
library(maptools)
library(raster)
library(rasterVis)
library(zoo)
library(foreach)
library(latticeExtra)
library(lattice)
library(RColorBrewer)
library(sp)
library(maptools)
library(MASS)
library(rgeos)
library(rgdal)
#library(GISTools)
library(raster)
library(maptools)
#library(rasterVis)


## harmonic function is amended not to consider statistically significant of waves.
  #source("harmonics2.0_verNT_Saad.R")
  source("harmonic_2.0_verNT6.R")

## For global regions analysis settings ##
  time.ndvi <- 1              #number of years
  acq.freq.ndvi <- 23          #frequency of observation in a year
  f.0.ndvi <- 1/time.ndvi
  ts.ndvi <- seq(0,time.ndvi-1/acq.freq.ndvi,1/acq.freq.ndvi)  # vector of sampling time-
points in months
  n.ndvi <- 10

harmonic4rast_func <- function(stk, na.rm=TRUE){
```

```
  ans <- harmonic4rast(stk,
                xt = ts.ndvi,
                acq.freq = acq.freq.ndvi,  #frequency of observation in a year
                n = time.ndvi*acq.freq.ndvi,   #number of images
                N = n.ndvi,   #considered number of waves
                alpha = 0.05,
                detrend = FALSE,
                which = "strongest",
                test = "holm",
                sig = FALSE,    ###extract statistically significant waves or not
                approx = FALSE, ### apply na.approx() to input time series
                slent = TRUE,
                na.rm = TRUE)

        return(ans)
}


## creating mask
  #bnd_shp <- readShapePoly("Export_Output.shp")

  tif <- stack("2001_stack1.tif")
# e <- extent(3300000,3310000,-2800000,-2790000)
 tif <- crop(tif, e) ##test for small area without NAs
  ans.harmonic <- calc(tif, harmonic4rast_func)
  names(ans.harmonic)

  ##if name of variables are not attached....
    detrend <- FALSE
    N <- 10
    time.ndvi <- 1              #number of years
    acq.freq.ndvi <- 23           #frequency of observation in a year
  if (detrend){
    names(ans.harmonic) <- c("Nsig", paste0("Amp_", 1:N), paste0("Ph_", 1:N),
paste0("Cycle_", 1:N), paste0("Residuals_", 1:(time.ndvi*acq.freq.ndvi)), "lmres",
"lmslope", "lmPval")
  } else names(ans.harmonic) <- c("Nsig", paste0("Amp_", 1:N), paste0("Ph_", 1:N),
paste0("Cycle_", 1:N),  paste0("Residuals_", 1:(time.ndvi*acq.freq.ndvi)))


    spplot(ans.harmonic, "Ph_1")
```