

ENVIRONMENTAL CONTROLS ON PHENOREGIONS ACROSS AN EAST AFRICAN MEGATRANSECT

By

Gloria Desanker

A THESIS

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Forestry-Master of Science

2019

## ABSTRACT

### ENVIRONMENTAL CONTROLS ON PHENOREGIONS ACROSS AN EASTERN AFRICA MEGATRANSECT

By

Gloria Desanker

Semi-arid and savanna-type (SAST) systems in East Africa have unique plant species compositions and characteristics that make quantifying this area's seasonality and inter-annual variability difficult. Phenoregion classification offers a way to use seasonality of vegetation growth dynamics to help understand the phenology of complex landscapes. Here, we used Normalized Difference Vegetation Index (NDVI) time series from the Landsat 8 imagery to map phenoregions in scenes centered around national parks from Mt. Kenya National Park (Kenya) to Limpopo National Park (Mozambique) to assess whether landscape-scale controls on phenology are consistent across the region or if they differ on a latitudinal gradient. We used MODIS Land Cover to assess land cover composition in each phenoregion, and discriminant analysis to determine the role that elevation, slope and aspect play in driving phenological differences. There was no clear latitudinal pattern seen in land cover or geologic composition. Most of the site's phenoregions showed no unique composition of either of the variables, meaning that land cover or geology type did not help in differentiating phenoregions. The discriminant analysis showed that topography was a strong predictor of many of the phenoregions, however, these also did not reveal any clear latitudinal pattern. Using seasonality of the NDVI time series to generate phenoregions provides different and even in some cases more ecologically relevant information, compared to past studies that use only land cover to generate ecoregions. With a significant population of humans and animals that live in and depend on SAST ecosystems, it is important to better understand vegetation processes and the factors that affect them as climate change becomes an increasingly pertinent issue in dry systems.

Copyright by  
GLORIA DESANKER  
2018

## **ACKNOWLEDGEMENTS**

This research would not have been possible without the advice and encouragement of my advisors: Dr. Kyla Dahlin and Dr. Andrew Finley. I am especially grateful for their continued enthusiasm and support throughout my time at Michigan State University. I am also grateful to the rest of my committee, Dr. Kim Hall, for professional advice and support. Thank you to my family and all the amazing friends I made over my two years who were always there for me and encouraged me to keep going when the going got tough. I really appreciate all the great friendships, conversations, inspiration and funny moments that have made my time at MSU so great!

## TABLE OF CONTENTS

<b>LIST OF TABLES.....</b>	<b>vi</b>
<b>LIST OF FIGURES.....</b>	<b>vii</b>
<b>1.INTRODUCTION.....</b>	<b>1</b>
<b>2. METHODS .....</b>	<b>6</b>
Study Sites.....	6
Analysis Workflow.....	11
Satellite Imagery .....	13
Google Earth Engine .....	13
Gapfilling.....	14
Principal Component Analysis .....	15
Determining Number of Clusters.....	15
K-means Clustering .....	17
Geology.....	18
Landcover.....	18
Discriminant Analysis.....	18
Phenoregions .....	20
Mapping Phenoregions and Landcover .....	20
<b>3. RESULTS.....</b>	<b>22</b>
Phenoregion Numbers and Composition.....	22
Spatial Trends.....	27
Temporal Trends.....	29
Discriminant Analysis.....	32
<b>4. DISCUSSION .....</b>	<b>41</b>
Spatial Patterns .....	41
Temporal Patterns.....	41
Environmental Controls.....	42
Management Implications.....	43
Caveats and Future Work.....	44
<b>5. CONCLUSION.....</b>	<b>46</b>
<b>APPENDIX.....</b>	<b>47</b>
<b>BIBLIOGRAPHY.....</b>	<b>100</b>

## LIST OF TABLES

Table 1: Geographic Information on Study Sites	8
Table 2: Kappa and Overall Accuracy	33
Table 3: Whisker Box Plot Summary	37

## LIST OF FIGURES

Figure 1: Study site megatransect	7
Figure 2: Phenoregion analysis flow chart	12
Figure 3: Land cover stacked bar plots	24
Figure 4: Geology stacked bar plots	25
Figure 5: Phenoregion maps	28
Figure 6: Average NDVI time series	31
Figure 7: Box and whisker plots	36
Figure S1a: Geology maps	48
Figure S1b: Land cover maps	49
Figure S1c: NDVI maps	50
Figure S1d: Transformed aspect maps	51
Figure S1e: Slope maps	52
Figure S1f: Elevation maps	53
Figure S2a: Mt. Kenya scree plot	54
Figure S2b: Mt. Kenya elbow method	55
Figure S2c: Mt. Kenya silhouette method	56
Figure S2d: Mt. Kenya phenoregion 1 NDVI profile	57
Figure S2e: Mt. Kenya phenoregion 2 NDVI profile	58
Figure S2f: Mt. Kenya phenoregion 3 NDVI profile	59
Figure S2g: Mt. Kenya phenoregion 4 NDVI profile	60
Figure S2h: Mt. Kenya phenoregion 5 NDVI profile	61
Figure S3a: Serengeti scree plot	62
Figure S3b: Serengeti elbow method	63
Figure S3c: Serengeti silhouette method	64
Figure S3d: Serengeti phenoregion 1 NDVI profile	65

Figure S3e: Serengeti phenoregion 2 NDVI profile	66
Figure S3f: Serengeti phenoregion 3 NDVI profile	67
Figure S3g: Serengeti phenoregion 4 NDVI profile	68
Figure S3h: Serengeti phenoregion 5 NDVI profile	69
Figure S3i: Serengeti phenoregion 6 NDVI profile	70
Figure S4a: Ruaha scree plot	71
Figure S4b: Ruaha elbow method	72
Figure S4c: Ruaha silhouette method	73
Figure S4d: Ruaha phenoregion 1 NDVI profile	74
Figure S4e: Ruaha phenoregion 2 NDVI profile	75
Figure S4f: Ruaha phenoregion 3 NDVI profile	76
Figure S4g: Ruaha phenoregion 4 NDVI profile	77
Figure S4h: Ruaha phenoregion 5 NDVI profile	78
Figure S4i: Ruaha phenoregion 6 NDVI profile	79
Figure S5a: Luangwa scree plot	80
Figure S5b: Luangwa elbow method	81
Figure S5c: Luangwa silhouette method	82
Figure S5d: Luangwa phenoregion 1 NDVI profile	83
Figure S5e: Luangwa phenoregion 2 NDVI profile	84
Figure S5f: Luangwa phenoregion 3 NDVI profile	85
Figure S5g: Luangwa phenoregion 4 NDVI profile	86
Figure S5h: Luangwa phenoregion 5 NDVI profile	87
Figure S5i: Luangwa phenoregion 6 NDVI profile	88
Figure S5j: Luangwa phenoregion 7 NDVI profile	89
Figure S5k: Luangwa phenoregion 8 NDVI profile	90
Figure S5m: Luangwa phenoregion 9 NDVI profile	91
Figure S5n: Luangwa phenoregion 10 NDVI profile	92
Figure S6a: Limpopo scree plot	93



Figure S6b: Limpopo elbow method	94
Figure S6c: Limpopo silhouette method	95
Figure S6d: Limpopo phenoregion 1 NDVI profile	96
Figure S6e: Limpopo phenoregion 2 NDVI profile	97
Figure S6f: Limpopo phenoregion 3 NDVI profile	98
Figure S6g: Limpopo phenoregion 4 NDVI profile	99

## 1. INTRODUCTION

Landscape scale vegetation seasonality reveals information about ecosystem processes and responses to changing environments and climates (Wang et al., 2016). As the impacts of climate change become increasingly indisputable, the need to identify vulnerable ecosystems has resulted in increased scientific research in environmental controls of vegetation dynamics (Archibald et al., 2007). Patterns in vegetation growth are influenced by processes involving water, energy exchanges, and land and atmospheric systems (Ma, 2013). The evidence found from studying the short and long-term vegetation dynamics across time and space confirms the effects of climate change on ecosystems (Cleland et al., 2007). Past studies have shown correlations between rising temperatures, shifting precipitation, rising CO<sub>2</sub> concentrations and other aspects of global change, to changes in timing of species and ecosystem level phenology, species-specific phenological shifts, and shifts in species ranges (Cleland et al., 2006; Dunne et al., 2011; Penuelas et al., 2004; Chuine, 2000; Chuine et al., 2001; Chuine et al., 2004). For many systems, small changes in temperature, precipitation and climatic patterns lead to large changes in the carbon cycle via changing timing and duration of leaf display (Archibald et al., 2007).

The study of seasonal changes in plant leaf growth is known as land surface phenology, though 'phenology' can also refer to other life cycle events like flowering and fruiting (Verbesselt et al., 2010). Ecosystem-scale phenology provides information on seasonality and inter-annual variability of vegetation over time and space (Bajocco et al., 2015). Other phenological measures include green-up (the onset photosynthetic activity), green-down or senescence (the decrease or end of photosynthetic activity), and peak greenness when leaf area is at its maximum (Zhang et al., 2003). Many field-based and remote studies have demonstrated the connectedness of phenological patterns to climatic patterns (e.g. Zhang et al., 2003).

Studies have shown correlation between plant shifting phenology and climate change or changing weather patterns such as rising temperatures and shifting precipitation (Cleland et al., 2007;

Ma et al., 2013; White et al., 2009; Keenan et al., 2014). Past research has focused on climatic factors such as temperature and aridity (Rial et al., 2017), rainfall (Ogutu et al., 2008; Archibald et al., 2007), and snow melt onset date (White et al., 2009) because there is an abundance of temporal and spatial data for these variables. Even though climate change is a dominating factor in many cases, there is little research showing how geophysical factors impact observed phenology (White et al., 2005).

Understanding ecosystem processes in geophysically heterogeneous landscapes is not only helpful for plant conservation, but also humans (Dahlin et al., 2016).

Globally there are more than three billion people living in semi-arid and savanna type (SAST) ecosystems (Dahlin et al., 2016), which cover more than 60% of the African continent (Guan et al., 2014). Savannas typically consist of woodland with varying mixtures of trees, shrubs and graminoids (Walker & Gillison, 1982). These SAST systems are also home to some of the largest populations of grazing animals in the world (Pimm et al., 2014). Fire disturbances due to drought and agriculture are common (Giglio et al., 2013). Both people and animals are reliant on SAST systems so it is important to understand baseline patterns of variation so you can detect change.

SAST systems in East Africa have unique characteristics that make studying this area challenging. The relationship between regional rainy seasons and growing seasons across much of this region is driven primarily by the Intertropical Convergence Zone (ITCZ) (Jonsson & Eklundh, 2002, 2004; Guan et al., 2014). Northern Hemisphere Africa shows latitudinal correspondence in onset dates for rainy and growing season, whereas Southern Hemisphere Africa shows relatively lower correspondence and has a more complex weather system including east-west orient of the ITCZ (Guan et al., 2014; Ziegler et al., 2013). In a study done by Potter et al. (2017), annual mean precipitation, mean annual temperature, and cloud cover amount differences are mapped across the globe. This complexity makes the spatial pattern of rainfall onset and offset complicated and makes savannas one of the most temporally and spatially dynamic global biomes (Guan et al., 2014; Ma et al., 2013).

This complexity likely contributes to poor performance of phenological models that perform well in other landscapes. Land cover or plant functional type is typically used as the main determinant of phenological classification. For example, the Community Land Model (CLM) groups plant functional types into either seasonally deciduous or stress deciduous algorithms (Dahlin et al., 2015). These models are used for local and global decision making that affects wildlife, ecotourism, climate change mitigation and more (e.g. Kay et al., 2015; Poulter et al., 2015; Reid et al., 2016; Bluwstein 2017). Carbon and vegetation models like the CLM are based off of coarsely derived phenoclusters for homogenous landscapes that experience more predictable seasonality but do not capture the seasonal variation that can be seen in the NDVI time series for SAST ecosystems. Other reasons that models fail to perform in SAST regions are because the models themselves are substantially influenced by the chosen study design and statistical methodologies. There is no established standard for phenological data collection or methodological framework for analyzing the data and there are assumptions and errors that need to be considered when using a model on any landscape (Parmesan, 2007). The climate in SAST regions that drives temporal vegetation growth would require the use of non-parametric approaches to fit a model (Ma et al., 2013).

An analytical tool that has the potential to improve our understanding of SAST phenology is the identification of phenoregions. Phenoregions are discrete groupings in a landscape based on similarities in phenological cycles (White et al., 1997). The majority of land surface phenology studies focus on extracting critical points in the seasonal Normalized Difference Vegetation Index (NDVI) trajectory that correspond with, for example, peak greenness during the growing season (Verbesselt et al., 2010) or categorize the NDVI signals by plant functional type (Geerken, 2009). These analyses consider information contained in the NDVI signal but fail to take into account seasonal and interannual growth dynamics that can improve vegetation classification (Geerken, 2009). Phenoregions have been used to

show the heterogeneity of seemingly similar or consistent areas reflecting spatial and temporal characteristics that other methods may leave out (White et al., 2005).

The NDVI is a commonly used metric that differentiates green vegetation from abiotic objects in satellite imagery (e.g. Verbesselt et al., 2010; Liu et al., 2017; Zhou et al., 2015; Braget et al., 2014). NDVI has been widely used to monitor and assess vegetation changes in savanna systems (e.g. Chamaille-Jammes et al., 2006; Guan et al., 2014) as well as to study a variety of global and local scale processes (Tucker et al., 2005). NDVI time series can be used to bridge the gap of knowledge about savanna phenology in East Africa, especially with the use of high spatial resolution satellite imagery.

Remote sensing, measuring the Earth's surface from a distance, is the only feasible means of cross-scale phenology monitoring as it can be used at a local, region or global scale (Ma et al., 2013). Space-borne instruments like the Landsat 8 satellite collect high resolution images of the Earth's surface that can be used to produce a finer scale representation of NDVI phenological patterns in SAST systems. Vegetation indices such as the NDVI calculated in Landsat 8 data are used in time series analyses of greenness patterns over multiple year time periods (see, e.g., Chen et al., 2005; Ma et al., 2013; Scrivani, 2015). These Earth observation systems examine broader scale phenomena that allow retrievals of whole-system phenological metrics, such as the timing and magnitudes of greening, peak activity, and drying phases of the growing season (Tucker et al., 2005). When calculated at the Landsat 8 30 m x 30 m resolution, NDVI can accurately represent vegetation coverage and be used to understand the links between changing vegetation greenness and the environment.

The primary objective of this study is to determine which abiotic variables impact phenoregion patterns in SAST regions. We are also interested in the number of phenoregions per site, their land cover and geological composition, and any visible patterns in their NDVI time series. The following list of sub-objectives are used to accomplish our overarching goal:

1. Understand if the environmental controls, elevation, slope and aspect, explain variation in phenology in SAST regions.
2. Identify phenoregions using only the NDVI signals as opposed to using land cover.
3. Evaluate the relationship between the spatial and temporal patterns, and the environmental controls, and what information can be extracted from these relationships to inform us about SAST vegetation dynamics?

## 2. METHODS

### Study Sites

The study sites follow a latitudinal transect along East Africa from Kenya to Mozambique (Figure 1). This 'megatransect' was selected because it captures the diversity of climate, geography, topography and composition of SAST ecosystems in East Africa. The Landsat scenes used are completely or nearly completely overlapping with national parks, and for convenience we refer to the scene by the name of the dominant national park it contains (geographic information listed in Table 1). These scenes were selected because they have minimal anthropogenic disturbance on vegetation cover (conversion to agriculture or deforestation), relative to more populated areas. The NDVI, environmental controls, land cover and geology maps can be found in Appendix (Figures S1a- S1f).

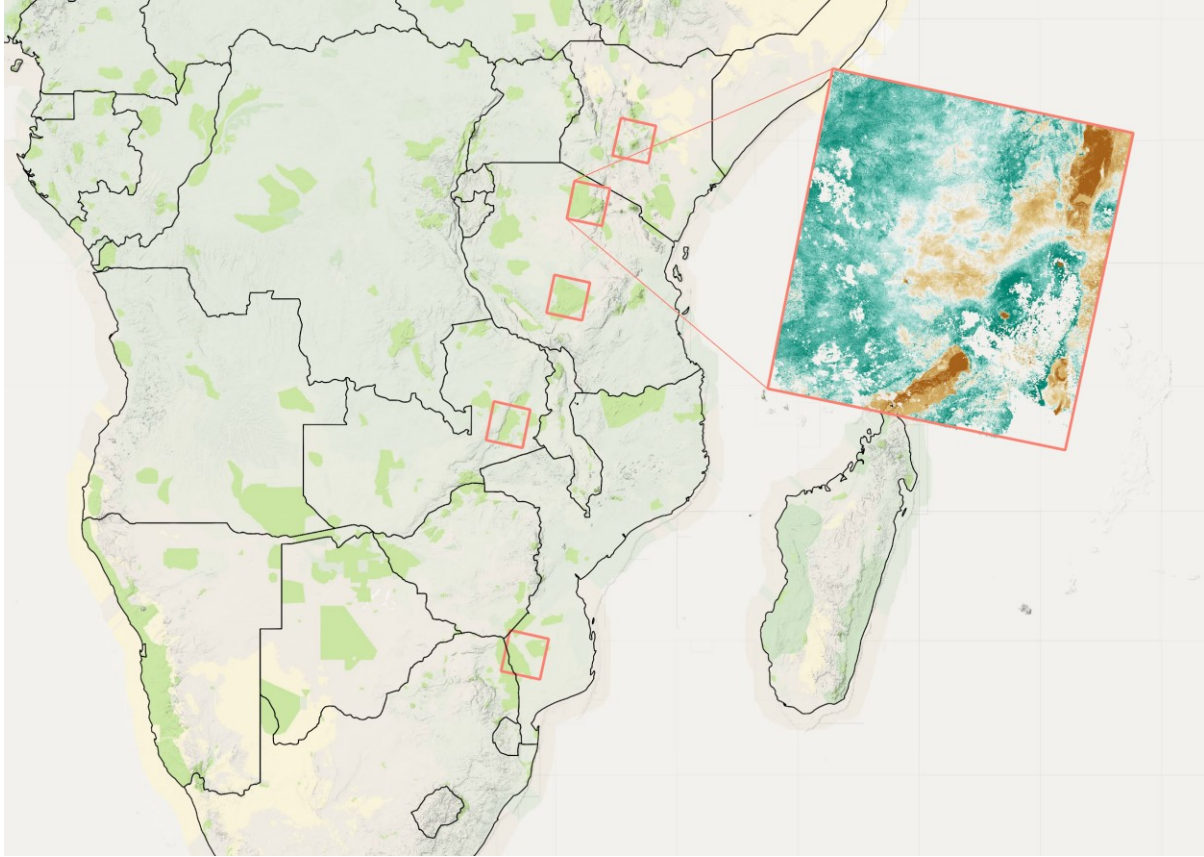


Figure 1: **Study site megatranssect.** Geographic Information on Study Sites. The five sites used in this study (red outlines) which run along a megatranssect of East Africa. From south to north, Limpopo, Luangwa, Ruaha, Serengeti and Mt. Kenya. The inset map is mid rainy season NDVI for Serengeti. The background is the Google Maps (April 5, 2017).



Table 1: **Geographic Information on Study Sites.** In order of south most study site to north most study site. The latitude (lat) and longitude (lon) coordinates are taken from the top left corner of Landsat 8 scenes used in study. We included environmental control ranges and the top two land cover and geology types found in each of the study site locations.

<b>Study site</b>	<b>Mt. Kenya</b>	<b>Serengeti</b>	<b>Ruaha</b>	<b>Luangwa</b>	<b>Limpopo</b>
<b>Country</b>	Kenya	Tanzania	Tanzania	Zambia	South Africa/ Mozambique
<b>Lat</b>	31.6	30.85	33.66	34.6	36.75
<b>Lon</b>	-22.06	-11.95	-6.18	-1.85	1.06
<b>Landsat Path, Row</b>	168,60	169, 62	169,65	179,69	168,76
<b># scenes analyzed</b>	64	57	56	55	59
<b># phenoregions determined</b>	5	6	6	10	4
<b>Precipitation range (mm) <sup>1</sup></b>	490 to 900	500 to 1000	450 to 750	800 to 950	100 to 600
<b>Temperature range (degrees C) <sup>2</sup></b>	10.5 to 28.9	12.7 to 27.5	18 to 28	10.1 to 33.4	13 to 33
<b>Rainy season <sup>3</sup></b>	April to June and October to December	June to October	October to May	December to February	October to March
<b>Max elevation (m) <sup>4</sup></b>	5054	3618	1915	1728	505
<b>Max slope (%) <sup>4</sup></b>	76.69	73.24	45.65	51.61	31.24

Table 1 (cont'd)

<b>Major geologies</b> <sup>5</sup>	precambrian, quaternary igneous, tertiary igneous	precambrian, tertiary, tertiary igneous	precambrian, quaternary (undivided)	precambrian, jurassic-carboniferous, paleozoic-precambrian	mesozoic igneous, pleistocene, tertiary
<b>Land cover composition</b> <sup>6</sup>	grassland, open shrubland, woody savannn	woody savanna, croplands	savanna, woody savanna	woody savanna, barren ground	savanna, woody savanna

<sup>1</sup> Precipitation derived from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) for 1980 to 2009.

<sup>2</sup> Temperature from Hijmans et al., 2005; Hijmans et al., 2005; Mbungu et al., 2017; Hijmans et al., 2005; Cambula et al., 2012.

<sup>3</sup> Rainy season from Recha et al., 2012; Byrom et al., 2015; Yang et al., 2018; Chabwela et al., 2017; Reason et al., 2005.

<sup>4</sup> Elevation and slope derived from the Shuttle Radar Topography Mission (SRTM).

<sup>5</sup> Geology derived from The Geology, Oil and Gas Map showing geology published by the U.S. Geological Survey in 2002 was obtained from the Atlas of Cultural and Environmental Change in Arid Africa.

<sup>6</sup> Land cover derived from the MODIS Land Cover Type product (2005).

Mt. Kenya, in the northernmost scene in this study, is the second tallest mountain in Africa and serves as a critical water source for the surrounding agricultural fields, natural areas and pastures (Notter et al., 2007). This area in East Africa generally has two rainy seasons from April to June and October to December. Across the scene, the average total annual precipitation is around 900 mm, with rains in the first rainy season of the year being much heavier (Recha et al., 2012). The average temperature is 19.2°C, with an average annual maximum of 28.9°C and minimum of 10.5°C (Hijmans et al., 2005). The Landsat scene includes Mt. Kenya and land just north of Kenya's capital, Nairobi. There are two main highways that pass through some smaller cities and villages. Vegetation around Mt. Kenya is predominantly grass plains and savanna woodlands (Ogotu et al., 2008).

The Serengeti National Park is at the center of the Serengeti ecosystem, with elevation ranging from 1140 m to 2000 m (Byrom et al., 2015). The dry season occurs from June to October when precipitation is around 50 mm; the wet season is from November to May when precipitation is around 1200 mm (Byrom et al., 2015). The annual average precipitation in the southern Serengeti is around 500

mm and around 1,000 mm in the northern portion (Morrison et al., 2016). Temperatures range from 12.7 to 27.5°C with an annual average of 20.4°C across the scene (Hijmans et al., 2005). Serengeti National Park is composed of grasslands, savanna, woodlands and dense woodland/forest (Byrom et al., 2015). The Landsat scene falls just under the northern border of Tanzania, includes Lake Natron, Lake Manyara and Lake Eyasi, and a highway cutting east-west.

Ruaha is the largest national park in Tanzania and the Great Ruaha River is the only source of water for wildlife during the dry season (Yang et al., 2018). Ruaha National Park experiences its dry season from June to September and wet season from October to May. In the past few decades, decreased streamflow and an increasing number of days with “zero-flow” during the dry season have been observed and is a threat to biodiversity in the park (Yang et al., 2018). Ruaha receives an average of 450 to 750 mm of rainfall per rainy season (Kangalawe et al., 2011). The Landsat scene includes the majority of the national park, including the adjacent game reserves. The eastern half of Ruaha National Park is dominated by miombo woodland, while the western half is dominated by Acacia, Commiphora, and Combretum species (Barnes 1983). The average annual precipitation is around 685 mm (Hijmans et al., 2005). The mean annual temperature in the scene ranges from 18°C at higher altitudes to 28°C (Mbungu et al., 2017). Elevation ranges from 698 m to over 2300 m (Mbungu et al., 2017)

South Luangwa is one of the three national parks in the Luangwa River Valley. South Luangwa is primarily comprised of a mixture of Miombo woodlands and Mopane woodlands, corresponding to changes in topographic position (Astle, 1969). Mean annual precipitation is around 800 mm with peak rainfall in the summer months of December through February (Chabwela et al., 2017). The average annual precipitation is 950 mm and the average temperatures range from 10.1 to 33.4°C (Hijmans et al., 2005). The Landsat scene also includes surrounding reserves and parks that are mostly a mix of seasonally inundated wetlands and agriculture. The Luangwa River runs through the scene and South Luangwa National Park.

Limpopo National Park forms part of the trans-frontier park with South Africa, Zimbabwe and Mozambique (Cambule et al., 2014; Cambula et al. 2012). The Landsat scene also includes Banhine National Park, Mozambique and some of South Gonarezhou National Park, Zimbabwe. There are a few major highways that criss-cross across the scene, but it is mostly conserved land. Altitudes in Limpopo range from 50 m to 500 m above sea level (Stalmans et al. 2004). The mean annual precipitation is around 400 mm and ranges from 100 to 600 mm during the rainy season (van Vegten et al., 1983). Rainy season for Limpopo occurs from October to March (Reason et al., 2005). The mean maximum temperature is around 33°C and minimum temperature is 13°C, with peak rainfall from October to March (Cambula et al. 2012). Limpopo is composed of floodplains, mixed woodlands, and shrublands, with some villages practicing low-impact farming (Cambula et al. 2012).

### **Analysis Workflow**

Phenoregion classification considers the variability within different phenology types. This is especially important for SAST regions which consist of mixed landscapes and varying vegetation growth. In this study, we use imagery from Landsat 8 to analyze regional scale vegetation on a North-South gradient to capture the diversity of SAST regions across East Africa. NDVI time series from the Landsat 8 imagery are used to identify and map phenoregions in these study sites. The phenoregions are compared to land cover and geology maps to assess differences in vegetation composition. Multivariate statistics are used to determine which topographically derived environmental controls contribute to these distinct classes of phenology. Results are compared across the five study sites to assess whether landscape-scale controls on phenology are consistent across the region or if they differ along latitudinal gradient.

The objectives of this study were to map phenoregions in each Landsat scene, compare these phenoregions to an existing map of land cover, then ask which abiotic gradients contribute to the

phenoregion patterns. Figure 2 shows the process used to define and analyze each site’s phenoregions. The initial raw inputs are the Landsat 8 NDVI satellite imagery, Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) (Farr et al. 2007) imagery and The Geology, Oil and Gas Map showing geology published by the U.S. Geological Survey in 2002. Statistical analyses were conducted on gap-filled NDVI time series and included Principal Component Analysis (PCA), K-means clustering, and Discriminant Analysis (DA). Outputs from the statistical methods are phenoregion NDVI profiles, phenoregion maps, and inference about the impact of environmental controls on phenology. Subsequent sections detail data and methods applied within each of the steps illustrated in Figure 2.

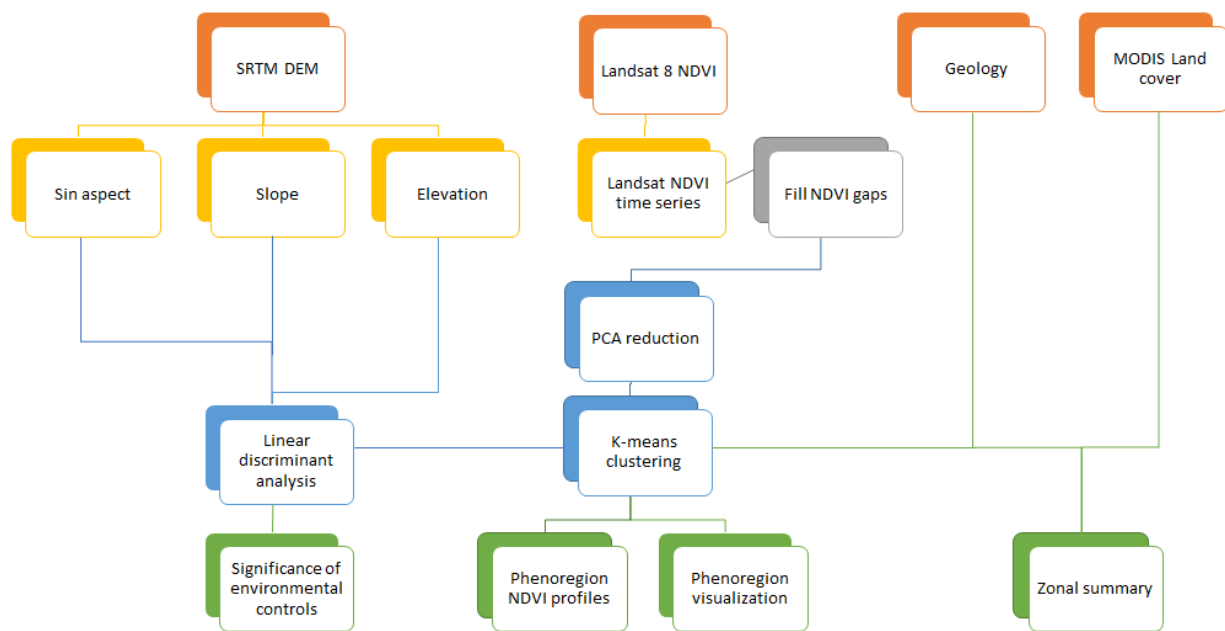


Figure 2: **Phenoregion analysis flow chart.** Flow chart showing the process used to define and analyze phenoregions for each site. The initial inputs are Landsat 8 NDVI, SRTM DEM imagery, and geology. Three environmental controls, aspect, slope, and elevation, are extracted from the SRTM DEM. We used Principal Components Analysis (PCA) to reduce the dimension of the gap-filled NDVI time series prior to applying a K-means clustering algorithm to identify phenoregions. Inference about the association between environmental controls and the phoregion clusters was then assessed via Discriminant Analysis (DA), dimension reduced phenoregion NDVI profiles, and map-based visualization.

## **Satellite Imagery**

Open source Landsat data provided by USGS/EROS (Wulder et al., 2012) is utilized globally to map small- and large-scale terrestrial ecosystem dynamics. Since remotely sensed data started being used to understand ecosystems over 40 years ago, the technology and its applications have drastically changed. Remote sensing techniques have improved the way data collection is performed. Traditionally data is collected manually in the field, which is time and resource consuming. Now, open source satellite imagery is available to the public, as well as GIS methodologies that provide simple and time effective solutions for data analytics (Mosleh et al., 2016). Satellite imagery is available at multiple spatial and temporal resolutions. Many previous studies in SAST regions use MODIS, which has a minimum spatial resolution of 250 m, or other coarse resolution remotely sensed data (e.g., Guan et al., 2014). Climate and phenology are highly variable in SAST regions, so the higher spatial resolution of Landsat images (relative to MODIS) should more accurately capture regional ecosystem changes.

## **Google Earth Engine**

Google Earth Engine (GEE) is a GIS tool that performs large scale environmental data driven analysis (Gorelick et al., 2017). GEE is unique because it provides a suite of highly efficient geospatial data processing algorithms (and facilities for using user-defined algorithms) able to run on massive cloud-based remote sensing datasets such as Landsat time series. This open source tool is helpful for large scale analysis and processing such as done in this study. GEE makes calculating indices and downloading the resulting image data products quick and efficient. After calculating NDVI for around 60 Landsat 8 images, we downloaded them from GEE for each study site (see Table 1 for exact number for each site). Elevation, slope, and aspect were calculated from the SRTM DEM images and downloaded for each site.

Also, in GEE, we calculated NDVI for each of the Landsat 8 images that were available at the time of download for a 3 to 4 year period. We used a mask based on the C Function of Mask (CFMask) algorithm that uses decision trees to estimate pixels that contain snow, clouds or cloud shadows (Foga et al., 2017). Slope and aspect were calculated using SRTM 30 m images. Given that aspect was computed in degrees (i.e., aspect of 0° and 360° are non-unique) the sine of aspect was calculated and used in the subsequent analysis (Burkhardt et al., 2012; Coulston et al., 2013; Kröber et al., 2012). We then exported the 300+ images out of GEE and further processing and analysis was done in R version 3.4.1 (R Core Team, 2018) using the packages ‘raster’ (Hijmans, 2017), ‘sp’ (Pebesma et al., 2005; Roger et al., 2013), ‘stats’ (R Core Team, 2017), ‘dplyr’ (Wickham et al., 2017), ‘knitr’ (Xie, 2018; Xie, 2015; Xie, 2014), ‘rgdal’ (Bivand et al., 2018), and ‘data.table’ (Dowle & Srinivasan, 2018). Additional packages for each method are listed in corresponding sections.

## **Gapfilling**

Satellite imagery has many benefits for landscape scale geospatial analysis’ but also comes with some caveats. Satellite sensors detect reflected wavelengths of light from the Earth’s surface, but atmospheric interference, like clouds, can cause incorrect measurements or gaps in observations of NDVI (Jonsson & Eklundh, 2002). Many different approaches have been developed for addressing and overcoming flaws in data, for example geospatial curve fitting or smoothing methods have been used for phenological time series (Filippa et al., 2016, Zhang et al., 2003, Beck et al., 2006, Gu et al., 2009).

In this study, we used the ‘greenbrown’ R package to linearly interpolate missing values in the NDVI time series (Forkel et al., 2015). Greenbrown has been used to gap fill pixels of snow, water and cloud in NDVI data and phenological studies such as Forkel et al., 2013 and 2015, and Filippa et al., 2016. Here we used greenbrown to fill masked pixels in the NDVI images. There are still some permanent gaps in the Landsat scenes of pixels that did not fit the criteria for the interpolation chosen; the gapfilling

linear interpolation only worked for pixels that had an NDVI value for the first and last date in the time series. As a result, we had to vary the end dates for each scene to use the most cloud free ones available, leading to slight differences in the overall number of images used per site (Table 1).

### **Principal Component Analysis**

Principal Component Analysis (PCA) is one of the most commonly used data reduction techniques in spatial analysis (Parveen et al., 2016), and commonly applied to high-dimensional remotely sensed data, including phenology data (Potter et al., 2017; Parveen et al., 2016; Verbesselt et al., 2010; Liu et al., 2017). PCA is a linear transformation that reduces the dimensionality of a dataset by extracting the most important information into Principal Components (PC) (Parveen et al., 2016; Jolliffe 2002; MacGarigal et al. 2000).

In this study, we used PCA to reduce the amount of information in the NDVI time series that was inputted in the K-means algorithm. For each site, a stack of ~60 rasters with upwards of 30 million pixels per layer is a very large dataset that is slow to work with, so reducing the data size ensured that the rest of the analysis was feasible. To prepare the data for PCA, the NDVI rasters for the three years were stacked and converted to the necessary format, accomplished mostly through using the 'raster', 'readr' and 'dplyr' packages. A scree plot of the PCs against the percent variance explained was used to help select the PC subset size (i.e., number of PCs to retain). It was decided to use a cut off of 90% variance explained to select the number of PCs that would enter the K-means clustering algorithm.

### **Determining Number of Clusters**

The number of clusters used in the K-means clustering algorithm determines the number of phenoregions for each study site. There are many unsupervised statistical methods that have been used to determine an "ideal" number of clusters, such as Bayesian statistics (Senthilnath et al., 2017) and the



Hierarchical method (Corstanje et al. 2016; Grafius et al., 2018; Chen et al., 2005). The elbow and silhouette methods (Scharsich et al., 2017; Rial et al., 2017; Subbalakshmi et al., 2015; Wang et al., 2017; Wang et al., 2017; Pascucci et al., 2018) were used in this study and chosen because of their ease of interpretation and reasonable processing time.

The elbow method considers within group sum of squares against possible number of clusters, and an ideal number of clusters is visually selected from a resulting “elbow” plot. This method calculates the sum of squares (SS) or within cluster dissimilarities for each cluster centroid, which decreases as number of centroids increases (Scharsich et al., 2017). SS is largest between dissimilar clusters and smaller between similar clusters. The optimal number of clusters corresponds to where SS is minimum, which creates an “elbow” on the resulting plot creates an elbow (Rial et al., 2017). The ‘gclus’ (Hurley, 2012) packages were used to perform the elbow method.

The silhouette method provides a graphical representation of how well observations belong to “natural” clusters in the data (Subbalakshmi et al., 2015). This method calculates the difference between the minimum Euclidean distance between cluster centroid and observation, and the average distance within the cluster (Wang et al., 2017). A hierarchical and nested clustering tree is constructed by calculating the similarity between the different clusters (Wang et al., 2017). An overall clustering quality is given between 1 and -1, good clustering and poor clustering solutions, respectively (Subbalakshmi et al., 2015). The clustering quality is based on similarity (shorter Euclidean distances between observation and cluster centroid) and dissimilarity (larger nearest neighbor distances between clusters) (Pascucci et al., 2018). The ‘fpc’ (Hennig, 2018) package was used to perform the silhouette method.

Our goal was to use the clustering method to delineate meaningful and practical phenoregions. Some of these methods returned a very large number of clusters so we followed the process outlined in White et al., (2005) to reduce the number of clusters. The following clusters were removed: clusters that

had 100 pixels or fewer, and cluster categories with the highest percentage of non-natural land cover such as agriculture or urban area.

In the case of Serengeti, there were some data processing errors with the gapfilling that needs to be addressed. There were many unreasonably high NDVI values for Serengeti (potentially due to processing errors) which were clustered into the same phenoregion. The NDVI values in this phenoregion were not replaced because of the magnitude of pixels that would need to have been replaced before gapfilling; rather, the resulting cluster was simply removed from the analysis.

### **K-means Clustering**

We used a clustering algorithm to delineate phenoregions based on phenological patterns (Mills et al., 2011). K-means partitions observations into intra-similar clusters while maximizing inter-cluster dissimilarity (Tzortzis et al., 2014). Centroids are established randomly, and the Euclidean distance is found from the centroids to each observation; all observations are assigned to the cluster centroid it is closest to (Hoffman et al., 2008). New centroids are calculated for each iteration and continues until the number of observations that change cluster assignment converges (Hoffman et al., 2008).

The simplicity and efficiency of the K-means clustering algorithm has resulted in its application across many disciplines (Bradley & Bradley, 1998; Pascucci et al., 2018; Kumar et al., 2011; Mills et al., 2011; Wang et al., 2017; Hoffman et al., 2008; Senthilnath et al., 2017). Because K-means is independent of location, the algorithm can categorize pixels in a Landsat scene that are not spatially close but belong to the same phenoregion (Kumar et al., 2011). K-means requires a user defined number of clusters to be specified (Tzortzis et al., 2014). Exhaustively searching for an optimal number of clusters is computationally infeasible with large datasets, so to avoid arbitrary specification of cluster number we used a few statistical based selection methods detailed in the subsequent section. The 'clusterR' package was used to perform the K-means clustering algorithm.

## **Geology**

The Geology, Oil and Gas Map showing geology published by the U.S. Geological Survey in 2002 was obtained from the Atlas of Cultural and Environmental Change in Arid Africa (available for download at [http://www.uni-koeln.de/sfb389/e/e1/e1\\_download\\_e.htm#geology](http://www.uni-koeln.de/sfb389/e/e1/e1_download_e.htm#geology)). The shapefile of geological classes was rasterized in ArcMap 10.4 (ESRI, Redlands, California, USA) at 30 m resolution and stacked bar plots were made to visualize the geological composition of each phenoregion. Since SAST regions are subject to irregular weather regimes, the ability of the geological strata to retain water plays a role in vegetation growth (Rodriguez-Moreno et al., 2015). For plants with short roots such as grasses, soil moisture can be the only source of water outside of the rainy season, whereas larger plants like trees can access groundwater (Rodriguez-Moreno et al., 2015). The hydrogeological heterogeneity of a landscape affects the movement or storage of water belowground, which affects the distribution of vegetation on the surface (Maxwell, 2010).

## **Landcover**

The MODIS Land Cover Type product (2005) at a resolution of 500m was downloaded from GEE which classified 17 global vegetation classification types by the International Geosphere Biosphere Programme (IGBP) (Friedl et al., 2015). The MODIS image was resampled in ArcMap to a spatial resolution of 30 m and stacked bar plots were made to visualize the land cover composition of each phenoregion.

## **Discriminant Analysis**

The general climate across each of the Landsat scenes is relatively constant due to the scenes' small size relative to large scale climate patterns like the ITCZ. As such, many discernible variations in NDVI time series within a scene will likely be due to local environmental factors like topography and

geology. NDVI profiles from each of the sites should have the same overall shape due to similarities in seasonality, but differ within a season. We expected that environmental factors, elevation, slope and sine(aspect) will influence NDVI seasonal profiles, but we also sought to understand which of these drivers has the greatest influence.

We used discriminant analysis (DA) to better understand how environmental drivers influence on the interannual variation among phenoregions. DA is a commonly used statistical method that tests the ability of variables to discriminate among group membership and then explain differences among the groups (Chambers et al., 2014; Hoagland et al., 2018; MacGarigal et al., 2000). The predictor variables: elevation, slope and sine(aspect) are evaluated for their ability to inform phenoregion classification of NDVI. A set of training samples were generated from the phenoregion assignment and predictor variables to determine probability of a pixel belonging to one of the phenoregion classifications (Jakubauskas et al., 2003).

The training sample was also used to measure accuracy of the DA model. Cohen's Kappa statistic (Kappa) and the overall accuracy were calculated for each combination of predictor variables in the DA models. Kappa is a measure of 'change-corrected classification' that represents the ratio of correctly classified observations to the total number of observations (MacGarigal et al., 2000). A Kappa of zero indicates no correct classifications of observations and a kappa of one indicates all observations are correctly classified (MacGarigal et al., 2000). The 'MASS' (Venables & Ripley, 2002) package was used to perform the DA.

Although geology influences NDVI seasonal profiles, this variable does not work with DA because this analysis uses linear combinations of the predictor variables to maximize the difference among group membership (MacGarigal et al., 2000). Geology is a categorical variable and a linear combination of categorical variables is not possible because there is no logical ordering of geological

values. Instead, stacked bar plots were made to visualize the geologic composition of each phenoregion, similar to land cover.

## **Phenoregions**

Phenologically and climatically self-similar clusters can be used to monitor climate change impacts (White et al., 2005). Clusters are pixels in a scene that have a distinguishable similar seasonal NDVI signal and therefore share similar vegetation phenology (White et al., 2005). Termed phenoregions, these can provide a basis for effectively classifying areas with high heterogeneity and phenological variability. Phenoregions is a relatively new concept that has been used to create a global (White et al., 2005) and national (Kumar et al., 2011) ecosystem classification scheme.

In this study, we expect the phenoregions to represent the different plant functional types in each study site. Overall, we expect to see the same general trend in the NDVI time series representing the growing seasons. Landcover, geology and the environmental controls should be main drivers of phenoregion classification since these environmental factors influence vegetation growth and distribution. Study sites with higher land cover heterogeneity are expected to have more phenoregions (White et al., 2005). Interannual variability within a growing season is what differentiates a phenoregion's profile from others.

## **Mapping Phenoregions and Landcover**

Once maps of phenoregion clusters were developed, we intersected them with the MODIS land cover classification to assess cluster land cover type composition. The MODIS land cover rasters were resampled from 500m x 500m to match that of the Landsat derived 30m x 30m phenoregion product in QGIS (Development Team, 2017) . In R, an empty raster of the same resolution and extent as the phenoregion raster was used to extract values from the land cover raster so that pixel values matched

up. Once resampled, the count of each land cover type was calculated for each phenoregion. Due to cloud cover in many of the scenes, even with gapfilling we were unable to produce wall-to-wall phenoregion maps. As a result, we present the land cover composition as a percentage of each phenoregion, and we do not report the overall area of each phenoregion class in each scene.

### 3. RESULTS

Geospatial and temporal analysis were conducted using PCA to reduce the NDVI time series dimensionality. Given the time series principal components, K-means algorithm assigned pixels to phenoregions. The Elbow and Silhouette methods were used to determine the number of phenoregions. Phenoregion assignment for each pixel in the scene was mapped. Given pixel assignments, DA was applied to determine the extent to which environmental controls explained variation among phenoregions.

#### **Phenoregion Numbers and Composition**

Numbers of phenoregions in each site varied from a low of 4 at the Limpopo site to a high of 10 at the Luangwa site (Table 1). Phenoregions at each site varied in composition to different extents. For example, the woody savanna land cover type was found in all four phenoregions in the Limpopo, whereas for Mt. Kenya, the same land cover type was found in at most two of the phenoregions (Figure 3).

The six phenoregions in Ruaha are composed of croplands, woody savannas, cropland/natural vegetation mosaic and are late season. Phenoregion composition for Serengeti and Mt. Kenya are more heterogeneous than the other sites. Serengeti phenoregions are composed of either savanna, cropland/natural vegetation mosaic, woody savannas and grasslands, though to varying extents. Whereas the phenoregions for Mt. Kenya vary in composition and percentage of land cover type.

Land cover composition was very similar for Limpopo, Luangwa and Ruaha. The phenoregions were mostly composed of woody savanna which makes up at least 50% of the mapped area in each of these three sites (Figure 3). The remaining land cover types made up less than 10% of a phenoregion's composition. Phenoregions in Mt. Kenya are mostly comprised urban built-up area, grasslands, barren

or sparsely vegetated land, and open shrubland cover classes. Among these phenoregions, the percentage varies from around 5% to 60%. Mt. Kenya phenoregions are composed of grasslands, cropland/natural vegetation mosaic, savannas and woody savannas.



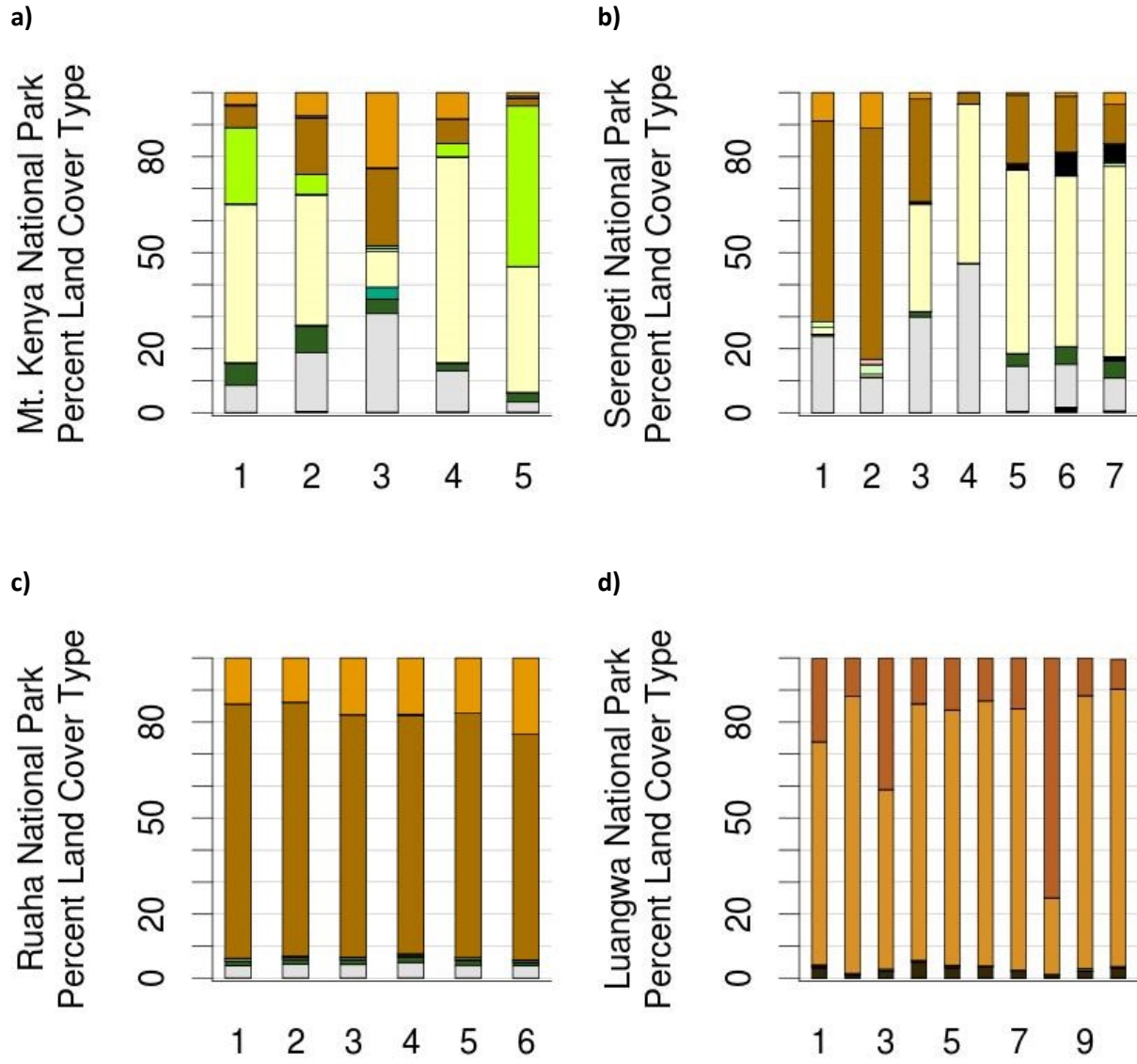


Figure 3: **Land cover stacked bar plots.** Stacked bar plots of percent land cover type for each study site's phenoregions. a) Mt. Kenya, b) Serengeti, c) Ruaha, d) Luangwa and e) Limpopo.

Figure 3 (cont'd)

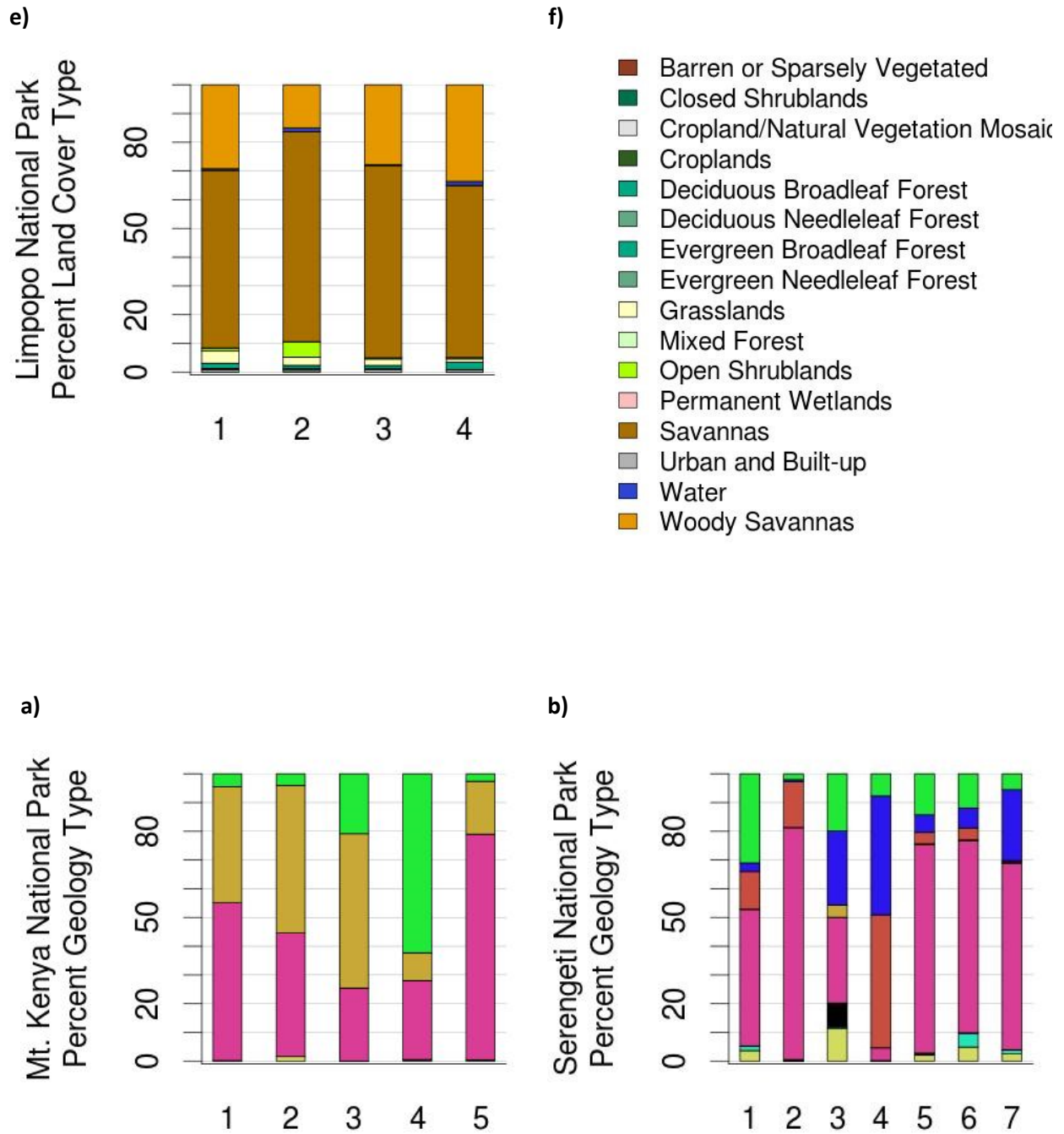
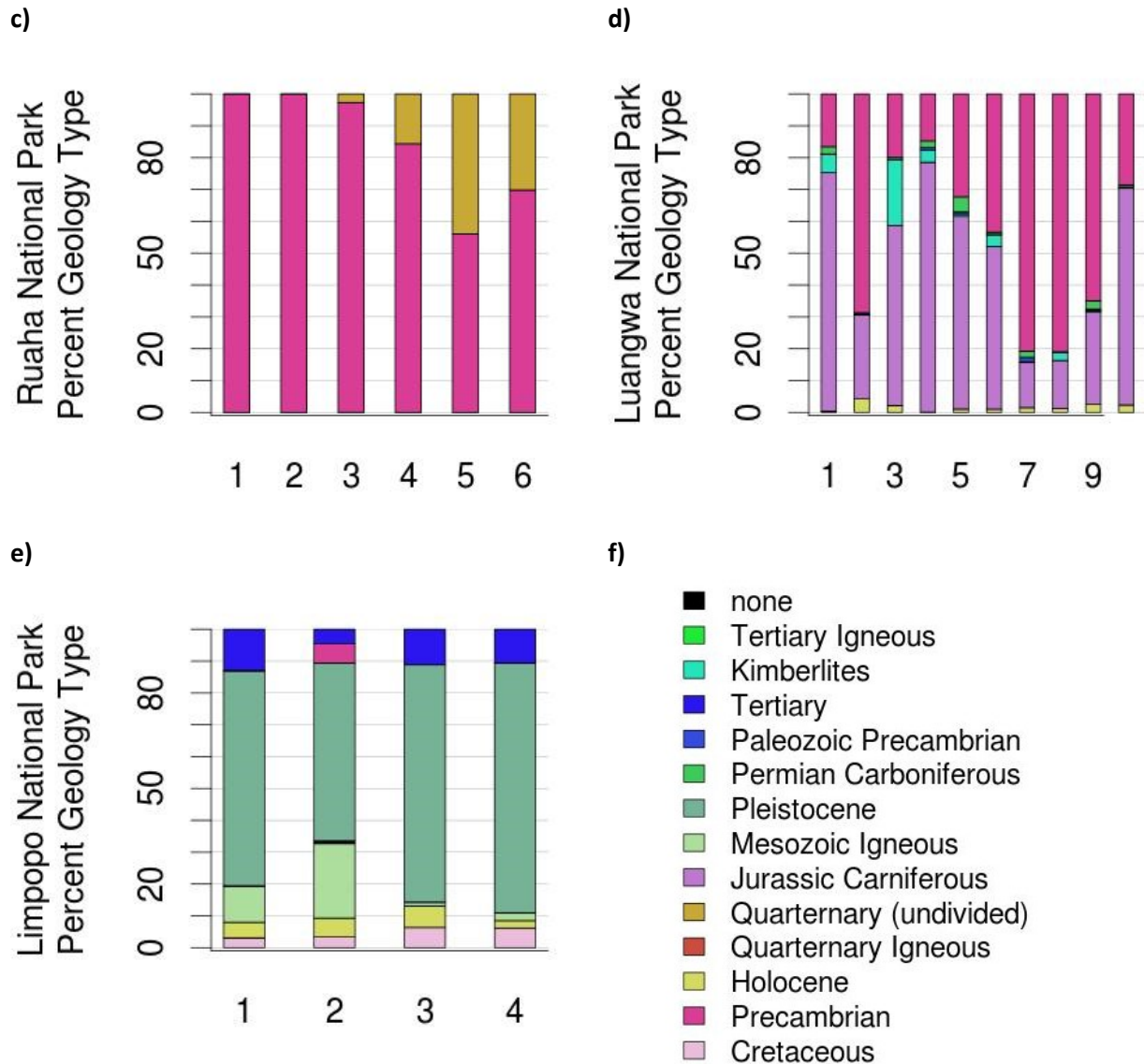


Figure 4: **Geology stacked bar plots.** Stacked bar plots of percent geology type for each study site's phenoregions. a) Mt. Kenya, b) Serengeti, c) Ruaha, d) Luangwa and e) Limpopo.

Figure 4 (cont'd)



Geology composition for Ruaha and Limpopo are relatively homogeneous (Figure 4). There are only two geologic types in the Ruaha scene and three of the six phenoregions are all or mostly Precambrian. Phenoregions four through six have around 18 to 45% Holocene. Limpopo is composed of mostly Pleistocene and varying amounts of Tertiary, Mesozoic Igneous, Holocene and Cretaceous. Phenoregion 2 is about 5% Precambrian. Geologic composition is more heterogeneous for Mt. Kenya, Serengeti and

Luangwa. Mt. Kenya has only three geology types whose compositions vary across phenoregions. Most of Serengeti's phenoregions have a large composition of Precambrian geology with varying amounts of Tertiary Igneous, Quaternary Igneous, Holocene and Tertiary. Similarly, in Luangwa, most of the phenoregions were composed of Precambrian and Jurassic Carboniferous and varying amounts of Kimberlites, Permian Carboniferous and Holocene.

### **Spatial Trends**

The number of phenoregions per study site ranged from 4 to 10, Luangwa National Park having the most and Limpopo National Park having the least. Phenoregion assignment for each site can be seen in Figure 5.

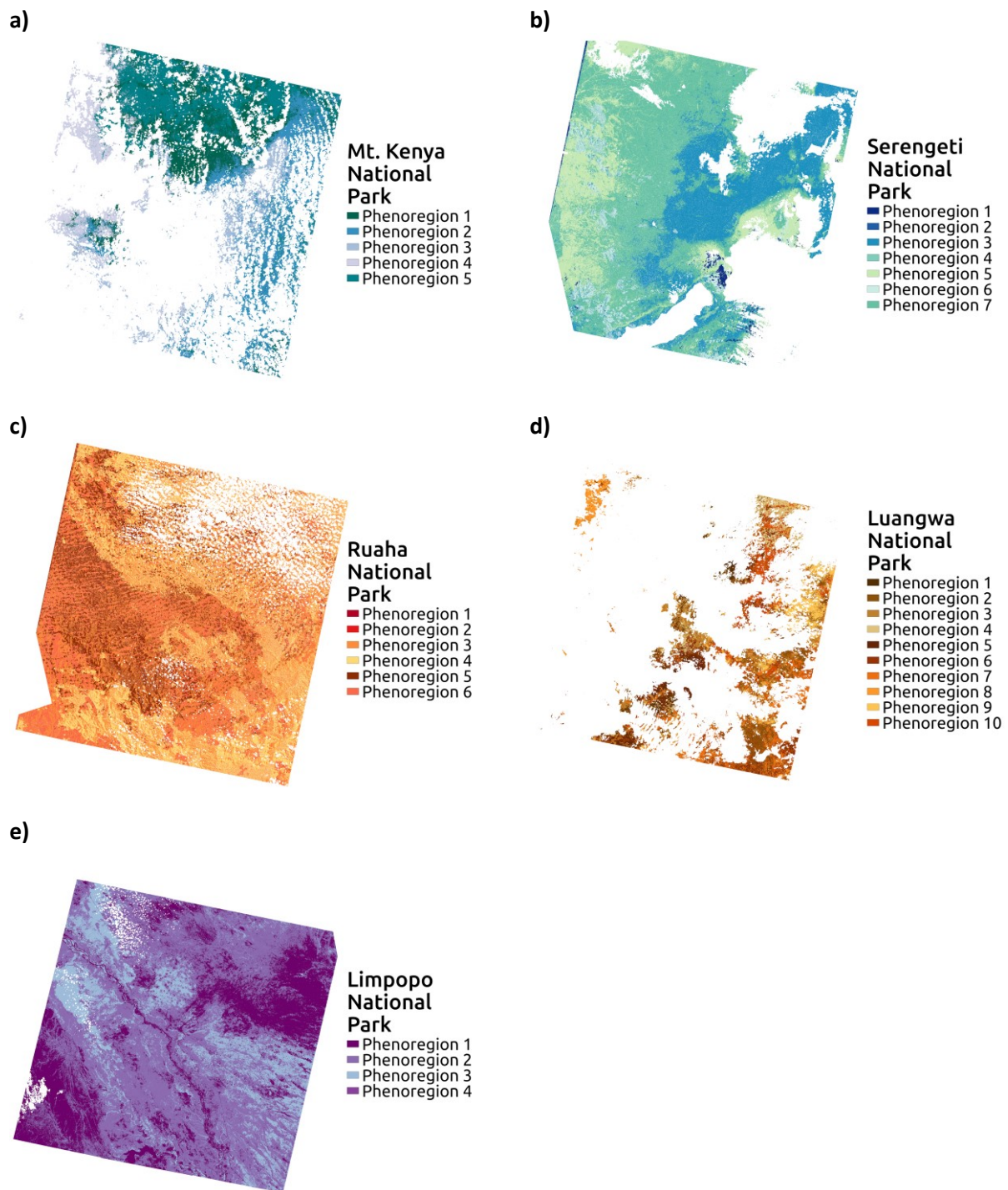


Figure 5: **Phenoregion maps.** Phenoregion maps for each study site derived from the clustered Landsat 8 NDVI time series. a) Mt. Kenya has 5 phenoregions, b) Serengeti has 7 phenoregions, c) Ruaha has 6 phenoregions, d) Luangwa has 10 phenoregions and e) Limpopo has 4 phenoregions.

## Temporal Trends

Average NDVI profiles for Limpopo National Park indicate that phenoregions 1, 3 and 4 had the highest values across the time period (Figure 6e). Whereas the average NDVI profile for phenoregion 2 is always below the three others. This phenoregion has a slightly higher composition of open shrublands than the others. All four phenoregions reach maximum NDVI during the same month that precipitation is at maximum. Interannual variation can be seen in the phenoregions after the first peak in NDVI and precipitation between January and April 2015. As NDVI drops, another peak NDVI value is reached before decreasing during the dry season. Phenoregions 1 and 4 show more variation between July 2015 and January 2016, just before the next peak in NDVI.

Phenoregions in Serengeti also showed some interannual variation, though not as strongly. Between July and October 2014, average NDVI values for phenoregions 1, 3, 5, and 7 do not increase at the same rate. Phenoregion 2 does not follow the same pattern as the rest of the phenoregions in Serengeti. It experiences a peak in NDVI a month or two before the rest, then dips when the others increase. Phenoregion 2 also starts increasing and reaches a peak about four months before the rest.

Phenoregions in Ruaha, Luangwa and Mt. Kenya have relatively smooth average NDVI profiles with subtle distinctions between the phenoregion patterns. Mt. Kenya's phenoregions are very similar in pattern but not in value; phenoregion 4 having the highest values of NDVI and phenoregion 2 having the lowest. The phenoregions in Ruaha have the same or very similar average NDVI values across the three years. There is a slight bump in NDVI between November and December 2014 and then the six profiles split and increase at different rates after August 2016. The average NDVI values for Luangwa's phenoregions are also very similar but a little more spread out. They all reach their maximum average NDVI about a month after maximum precipitation. Phenoregions 3 and 8 do not decrease in NDVI as far as the rest, and phenoregion 6 does not reach the same peak NDVI as the others.

Each phenoregion is also classified as either early or late season, specifying when in the rainy season NDVI peaks or starts to first increase (Figure 3). Mt. Kenya phenoregions are all early season though their land cover compositions are relatively heterogeneous. Serengeti has one late season phenoregion which is composed of grasslands and cropland/natural mosaic vegetation type. Whereas the other phenoregions are composed mostly of grasslands and woody savanna. Ruaha phenoregions are all late season with homogeneous land cover compositions. Luangwa has two early season phenoregions though they have the same land cover composition as the late season phenoregions. The one late season phenoregion in Limpopo is composed more of closed/open shrublands, woody savanna and mixed forest. Whereas the early season phenoregions are composed of open shrublands, woody savanna and mixed forest.

Figure 6 shows the average NDVI profiles for each phenoregion. All phenoregions' profiles have the same green up and green down pattern, and interannual variability can be seen for Mt. Kenya's, Serengeti's and Limpopo's profiles. Phenoregions in Ruaha and Luangwa have relatively smooth and similar profiles. Ruaha's profiles fall on top of each other for most of the three years. Luangwa's profiles differ a little more than Ruaha's but are still very similar.

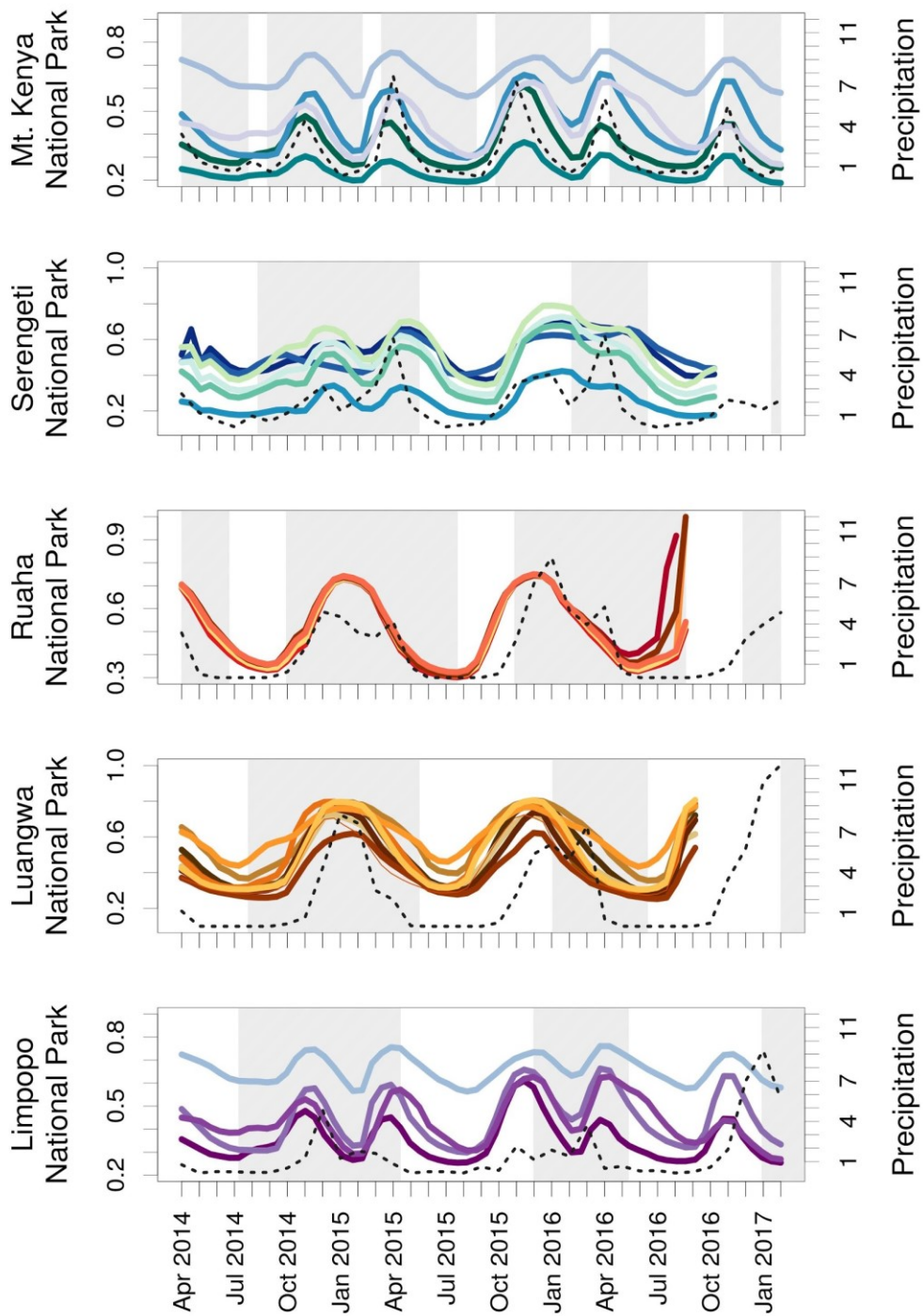


Figure 6: **Average NDVI time series.** Landsat 8 NDVI time series for phenoregions in each site (solid colored lines) and annual site precipitation (mm) (dotted line); a) Kenya National Park, b) Serengeti National Park, c) Ruaha National Park, d) Luangwa National Park, e) Limpopo National Park.



## **Discriminant Analysis**

In this study, we used DA to determine how well environmental controls explain variability among pheoregions. The coefficients of linear discriminants (LD) represent the ability of an environmental control to discriminate; the magnitude of the LD represents the size of the effect and the sign of the LD shows you the direction of the effect. In this study, we are interested in the size of the effect which tells us how an environmental control ranks relative to the others. Table 2 shows the Kappa and overall accuracy of the DA using all combinations of predictor values in the model. High Kappa and overall accuracy values indicate higher accuracy or more correctly classified pixels. 'klaR' (Weihs et al., 2005) and 'greenbrown' were used to calculate Kappa and the overall accuracy. The magnitude of Kappa and overall accuracy represent the size of effect for each environmental control across the seven model combinations.

Table 2: **Kappa and Overall Accuracy**. Linear discriminant analysis accuracy results. Kappa statistic and overall accuracy are used to evaluate discriminant analysis performance.

	<b>Predictor variables</b>	<b>Kappa</b>	<b>Overall Accuracy</b>
<b>Mt. Kenya</b>	<b>Elevation</b>	0.1874237	0.3880505
	<b>Slope</b>	0.06143426	0.3118948
	<b>Sine(aspect)</b>	0.00	0.2790556
	<b>Elevation, slope</b>	0.2509996	0.4280711
	<b>Elevation, sine(aspect)</b>	0.2067745	0.397117
	<b>Slope, sine(aspect)</b>	0.06744764	0.312914
	<b>Elevation, slope, sine(aspect)</b>	0.2605014	0.4334669
<b>Serengeti</b>	<b>Elevation</b>	-0.04334203	0.4499683
	<b>Slope</b>	0	0.4783009
	<b>Sine(aspect)</b>	0	0.4783009

Table 2 (cont'd)

	<b>Elevation, slope</b>	0	0.4783009
	<b>Elevation, sine(aspect)</b>	-0.02968986	0.4589568
	<b>Slope, sine(aspect)</b>	0	0.4783009
	<b>Elevation, slope, sine(aspect)</b>	-0.0154618	0.4682615
<b>Ruaha</b>	<b>Elevation</b>	0.4688608	0.6241138
	<b>Slope</b>	0	0.4002127
	<b>Sine(aspect)</b>	0	0.4002127
	<b>Elevation, slope</b>	0.4620553	0.6191012
	<b>Elevation, sine(aspect)</b>	0.4683218	0.6236376
	<b>Slope, sine(aspect)</b>	0	0.4002127
	<b>Elevation, slope, sine(aspect)</b>	0.4616824	0.6188615
<b>Luangwa</b>	<b>Elevation</b>	0.1507137	0.2573611

Table 2 (cont'd)

	<b>Slope</b>	0.01659345	0.156719
	<b>Sine(aspect)</b>	0.1298471	0.2381282
	<b>Elevation, slope</b>	0.143551	0.2514325
	<b>Elevation, sine(aspect)</b>	0.2307524	0.3258883
	<b>Slope, sine(aspect)</b>	0.1564591	0.2685392
	<b>Elevation, slope, sine(aspect)</b>	0.2307858	0.3263145
<b>Limpopo</b>	<b>Elevation</b>	0.1719816	0.4012388
	<b>Slope</b>	0.001396537	0.2949705
	<b>Sine(aspect)</b>	7.871149e-17	0.2947516
	<b>Elevation, slope</b>	0.178772	0.4064999
	<b>Elevation, sine(aspect)</b>	0.1724915	0.4029027
	<b>Slope, sine(aspect)</b>	0.00174670	0.2950897

Table 2 (cont'd)

	<b>Elevation, slope, sine(aspect)</b>	0.1768206	0.4054055
--	---	-----------	-----------

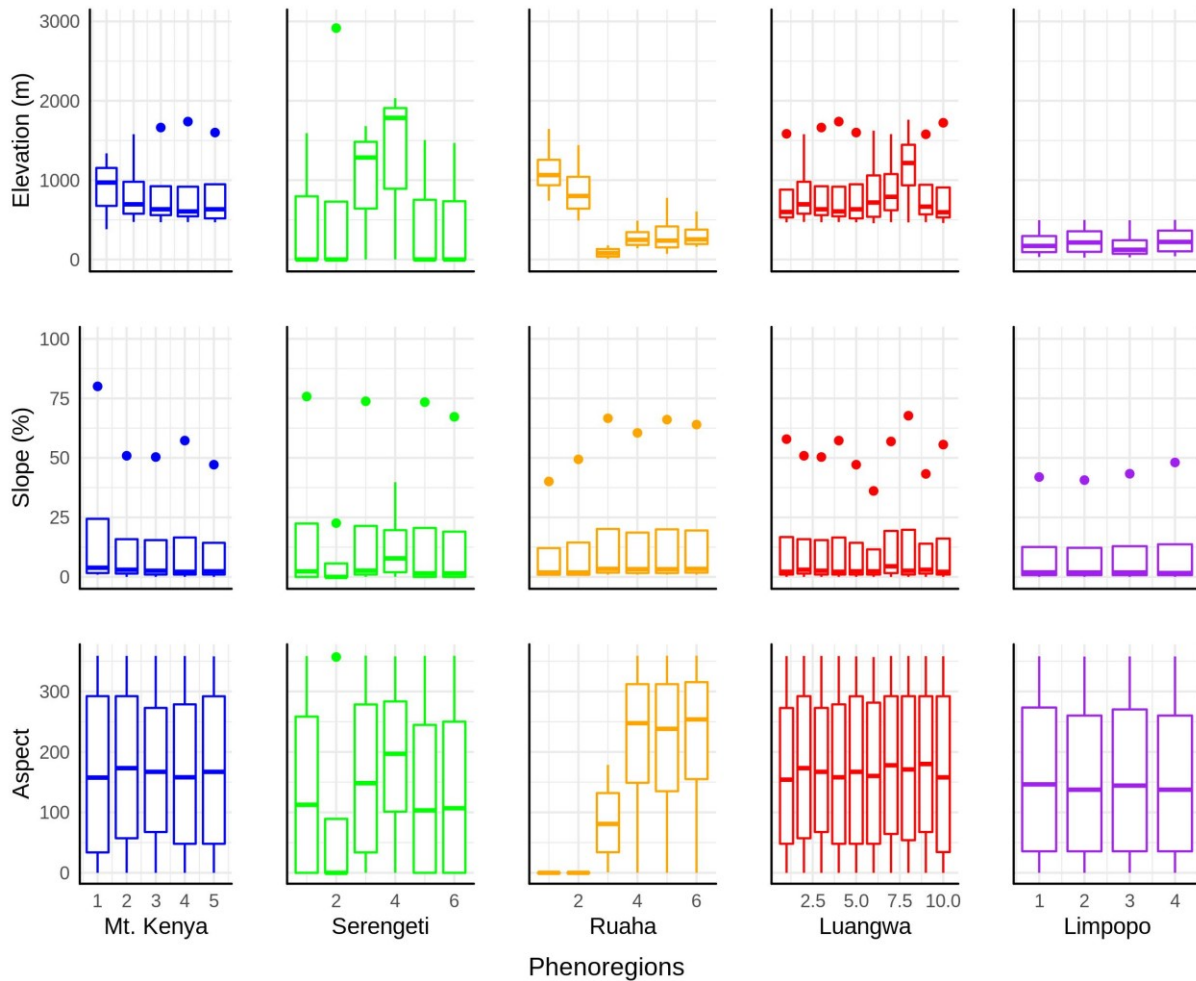


Figure 7: **Box and whisker plots.** Box and whisker plots showing variability of environmental controls in each phenoregion. The ends of the box are the upper and lower quartiles and the ends of the whiskers/points show the maximum and minimum observations. The horizontal line in the boxes represents the mean. The box and whisker colors correspond to phenoregion colors used in Figure 5.

Figure 7 and Table 3 show the ranges of elevation, slope and sine(aspect) for each phenoregion. We can see a lot more variability in elevation for each phenoregion in Mt. Kenya, Serengeti, Ruaha and

Luangwa. Slope is more variable for Serengeti and Luangwa. Aspect is most variable for Serengeti, Ruaha and Limpopo.

Table 3: **Whisker Box Plot Summary.** This table summarizes the box plots in Figure 7 and phenoregion characteristics. We use low, medium or high to describe elevation and slope relative to other phenoregions in a site. We use the direction that the slope is facing for aspect. And we list the top two land cover and geology types and the phenoregion’s seasonality.

	Phenoregion	Elevation	Slope	Aspect	Land cover	Geology	Season
<b>Mt. Kenya</b>	<b>1</b>	low	medium	southeast	grassland, open shrubland	Holocene, precambrian	early
	<b>2</b>	low	medium	southeast	grassland, cropland	Holocene, precambrian	early
	<b>3</b>	medium	high	southeast	woody savanna, cropland	Holocene, precambrian	early
	<b>4</b>	medium	medium	southeast	grassland, cropland	Tertiary igneous, precambrian	early
	<b>5</b>	high	low	southeast	open shrubland, grassland	Precambrian, holocene	early
<b>Serengeti</b>	<b>1</b>	low	medium	medium	croplands, woody savannas	Precambrian, tertiary igneous	early
	<b>2</b>	low	low	low	croplands, woody savannas	Precambrian, quaternary igneous	early
	<b>3</b>	medium	medium	medium	croplands, woody savannas	Precambrian, paleozoic precambrian	late

Table 3 (cont'd)

	4	high	high	high	croplands, woody savannas	Quaternary igneous, paleozoic precambrian	early
	5	low	low	low	croplands, woody savannas	Precambrian, tertiary igneous	early
	6	low	low	low	croplands, woody savannas	Precambrian, paleozoic precambrian	early
<b>Ruaha</b>	1	high	low	north	savanna, woody savanna	Precambrian	late
	2	high	low	north	savanna, woody savanna	Precambrian	late
	3	low	high	east	savanna, woody savanna	Precambrian, holocene	late
	4	low	high	west	savanna, woody savanna	Precambrian, holocene	late
	5	medium	medium	southeast	savanna, woody savanna	Precambrian, holocene	late
	6	medium	medium	west	savanna, woody savanna	Precambrian, holocene	late
<b>Luangwa</b>	1	low	low	southeast	woody savanna, barren	Jurassic carniferous, precambrian	late
	2	medium	medium	south	woody savanna, barren	Precambrian, jurassic carniferous	late

Table 3 (cont'd)

	3	low	medium	south	woody savanna, barren	Jurassic carniferous, precambrian	late
	4	low	low	south	woody savanna, barren	Jurassic carniferous, precambrian	early
	5	low	low	south	woody savanna, barren	Jurassic carniferous, precambrian	late
	6	medium	low	southeast	woody savanna, barren	Jurassic carniferous, precambrian	late
	7	medium	high	south	woody savanna, barren	Precambrian, jurassic carniferous	late
	8	high	medium	south	barren, savanna	Precambrian, jurassic carniferous	early
	9	low	medium	south	woody savanna, barren	Precambrian, jurassic carniferous	late
	10	low	low	southeast	woody savanna, barren	Jurassic carniferous, precambrian	late
<b>Limpopo</b>	1	medium	high	medium	savanna, woody savanna	Pleistocene, tertiary	early
	2	high	medium	low	savanna, woody savanna	Pleistocene, mesozoic igneous	late
	3	low	medium	high	savanna, woody savanna	Pleistocene, tertiary	early



Table 3 (cont'd)

	4	high	low	low	savanna, woody savanna	Pleistocene, tertiary	early
--	---	------	-----	-----	------------------------------	--------------------------	-------

## 4. DISCUSSION

### **Spatial Patterns**

Overall, the distribution of mapped phenoregions show the heterogeneity of vegetation patterns within the study sites. There was a visible gradient in phenoregion assignment and average NDVI. The ends of the transect, Mt. Kenya, Serengeti and Limpopo, were more different than the middle sites. The spacing between the phenoregion's average NDVI profiles were greater for the end sites and showed more interannual variability. Where as the average NDVI profiles were mostly identical or showed little to no interannual variability for Ruaha and Luangwa. The mapped phenoregions were also relatively good representations of the topographic heterogeneity seen in the environmental control figures (Appendix, Figures S1a- S1f).

### **Temporal Patterns**

We found that the central site, Ruaha, has the least variation among phenoregion classes and average NDVI, while the extreme ends, Limpopo and Mt. Kenya, are much more variable. This result highlights the importance within-land cover type variation in SAST ecosystems that is not captured by coarse land cover classifications. Serengeti and Mt. Kenya had the most heterogeneous phenoregions in terms of both land cover composition and NDVI profiles. Ruaha National Park was the only site that had almost the same land cover composition for all phenoregions and very similar NDVI profiles. Even though the phenoregions are the same in terms of land cover composition, the K-means algorithm still grouped NDVI-based PCs into six different clusters, suggesting subtle but still significant phenological differences across this landscape. This could be better understood with a longer time series. The average NDVI profiles start to branch out at the end of the time series around July 2016, which could be a result of a drought response.

While beyond the scope of this study, these differences suggest that broad-scale climatological patterns may drive the number and type of phenological strategies in these systems. The site with the largest number of phenoregions, Luangwa, is also the furthest inland site and one of the most topographically variable in terms of aspect which could influence the phenological variations here.

### **Environmental Controls**

Land cover and geology did not reveal very much about the variability seen in the average NDVI curves. The land cover barplots for Ruaha, Luangwa and Kruger show that there were no unique combinations of composition showing that 'ecoregions' which only use land cover, does not represent phenological differences in a landscape. Geology did not reveal anything about phenoregion assignment for Ruaha, Luangwa and Limpopo either. The land cover and geology compositions for Mt. Kenya and Serengeti show more heterogeneity in type and percent of each type. As seen in Serengeti, phenoregions one and two are composed mostly of savanna, woody savanna and cropland, where as the rest of the phenoregions have a lot more grassland and cropland. This could mean that land cover contributes to phenoregion assignment for Mt. Kenya and Serengeti, though not strongly.

Topography did, however, show a significant influence on phenoregion patterns. The Kappa and overall accuracy values show the probability of correctly classified pixels if one were to randomly assign classes; we want higher Kappa and overall accuracy values. Kappas that are less than zero, mean that the classification is worse than random. For Mt. Kenya, elevation did the best at explaining phenoregion assignment, having the highest Kappa and overall accuracy values. The values improved with elevation and slope in the discriminant model, and the addition of sine(aspect), though having a Kappa of zero by itself, resulted in the strongest combination of controls. In Figure 7 and Table 3, elevation has the most variation compared to slope and aspect. Though the distribution of all three environmental controls is relatively variable for all three. Elevation was the only control that explained phenoregion assignment

for Serengeti and adding either of the other environmental controls to the model weakened it. This is not seen in the whisker box plots; we see variability across all three controls even though most of the phenoregions are all under the low category. Elevation also explains phenoregion assignment the best for Ruaha, and the addition of other controls weakens the discriminant model although we see variation in the distribution of means in Figure 7. Elevation and sine(aspect) together explained the most for Luangwa though the whisker box plots for aspect are all relatively the same. Elevation and slope for Limpopo explained phenoregion assignment the best though all three controls show little variation. Even though we do or do not see much variability in the distribution of means, DA is telling us whether that variation is significant.

### **Management Implications**

The results from this study are applicable for on-the-ground management of SAST regions. Knowing and understanding how vegetation growth patterns change over time can be used to inform decisions on ecological conservation, agriculture and climate change mitigation strategies. The fine scale variability seen in this study's phenoregions, reflects the different ways that phenology influences environmental processes that humans and animals rely on.

Land cover or plant functional type is typically used as the main determinant of phenological classification. For example, the Community Land Model groups plant functional types into either seasonally deciduous or stress deciduous algorithms (Dahlin et al., 2015). Carbon and vegetation models are based off of coarsely derived phenoclusters but do not capture the seasonal variation that can be seen in the NDVI time series. These models are used for local and global decision making that affects wildlife, ecotourism, climate change mitigation and more (e.g. Kay et al., 2015; Poulter et al., 2015; Reid et al., 2016; Bluwstein, 2017). Our results show that even if there is shared land cover composition,

there are still differences between phenoregions that land cover misses. This is accounted for by the seasonality seen among phenoregion profiles as well as the environmental controls that are present.

### **Caveats and Future Work**

Apart from the interannual and annual NDVI variations, and environmental control variables used in this study, other factors that could influence the vegetation patterns include human and animal interactions. Cropland and urban built-up areas composed of around 5% for Ruaha phenoregions and 0% to 30% of phenoregions in Serengeti and Mt. Kenya. We chose to focus on scenes that overlapped with national parks to reduce the amount of human interaction on the vegetation, but there will always be some level of human disturbance in a landscape. In addition, many of these landscapes are inhabited by wildlife that impact vegetation patterns in ways that may or may not correspond to environmental gradients. Nevertheless, understanding local-scale phenological variation is an important part of understanding ecosystem patterns in process across space and through time.

The phenoregion maps for Luangwa and Mt. Kenya National Park (Figure 3) have a lot of cloudy pixels or pixels that did not work with the gapfilling algorithm and where hence masked out. This resulted in what looks like relatively sparse maps. Even so, both sites retained millions of pixels (compared to the tens of million pixels for other study sites) that were enough to continue the statistical analysis, given the assumption that masked pixels were not conditional on phenoregion. Even though the maps are not visually appealing, we are more interested in the relationship between phenoregion assignment and the environmental controls.

This study could be expanded by using another gapfilling algorithm to fill gaps in the NDVI time series. There are studies that have created phenology smoothing NDVI models that not only fill gaps but also smooth out the NDVI curves (Jönsson et al., 2004; Jönsson et al., 2002; Braget, 2017). NDVI curve smoothing is also possible using the 'greenbrown' package in R. Though it is possible to fill gaps in NDVI

curves, the smoothing would not have been ideal for this study because it would have removed the interannual variability that we were analyzing.

Future work could be to determine a discriminant method that can take both continuous and categorical variables. Since geological heterogeneity influences vegetation (Cleverly et al., 2016; Githumbi et al., 2018; Muvengwi et al., 2018), being able to statistically show the impact of geology on phenoregion assignment could improve inference. We could also include other environmental control variables or metrics such as the Compound Topographic Index (Marthews et al., 2015) or a combination of variables that can be used to indicate moisture availability or other soil characteristics. This study could also be expanded by continuing the analysis on other national parks in East Africa along the megatransect. And as more Landsat images are taken and made available, a longer time series of NDVI could lead to an even more accurate representation of vegetation seasonality in SAST regions.

## 5. CONCLUSION

The latitudinal gradient of phenoregion assignment and composition is apparent in the chosen Landsat scenes. Across the latitudinal gradient, there was more phenoregion heterogeneity at the ends of the megatransect with Mt. Kenya, Serengeti and Limpopo, than in the middle study sites, Ruaha and Luangwa. Variability in phenoregion composition and interannual variation seen in their NDVI time series indicate the importance of seasonality on phenoregion classification, especially since the environmental controls, land cover and geology compositions did not show any clear pattern in explaining phenoregion assignment. This study suggests the addition of phenoregion classification can improve ecological understanding about SAST regions.

## APPENDIX



## Geology

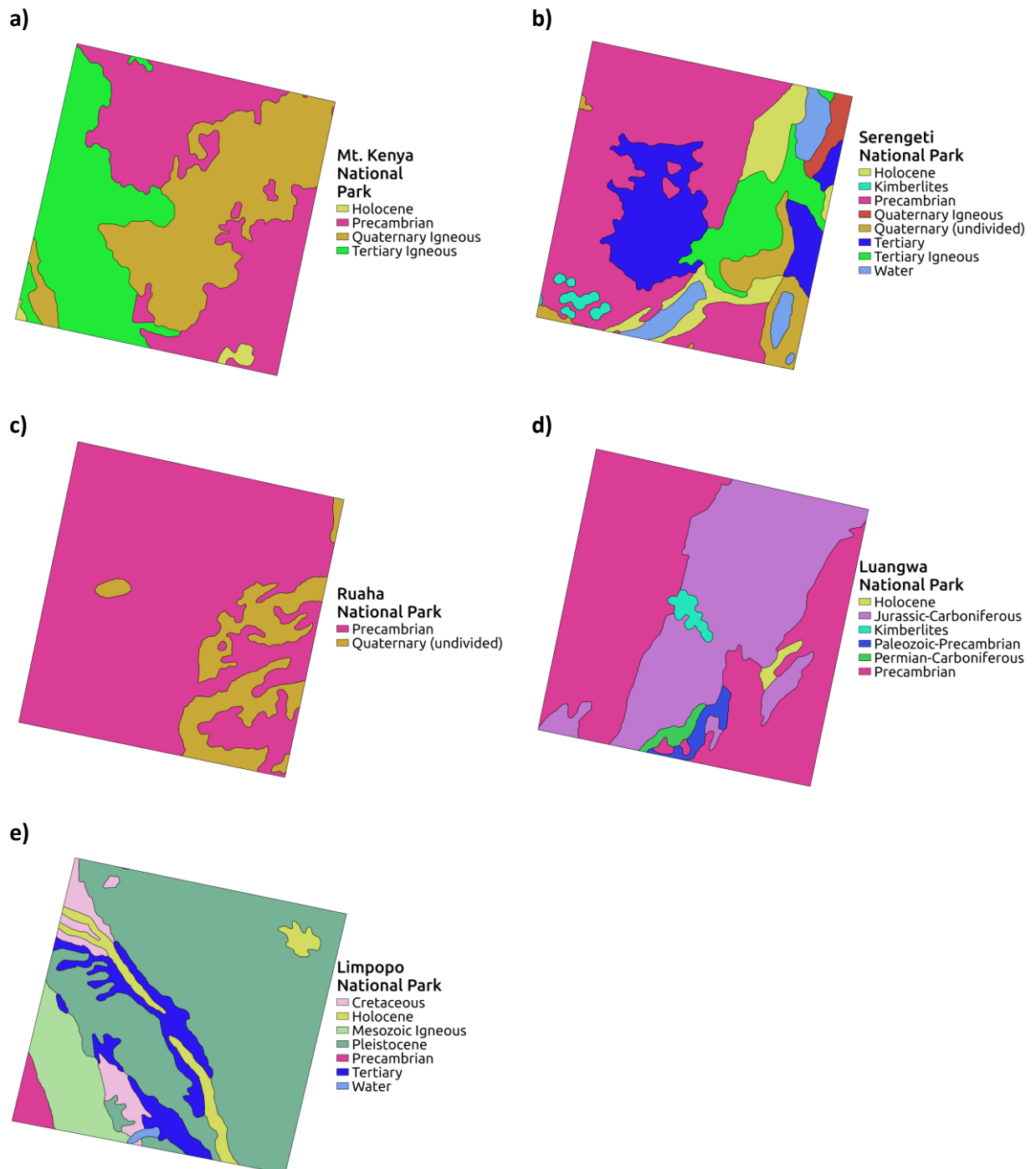


Figure S1a: **Geology maps.** Geology maps of the five study sites. a) Kenya National Park, b) Serengeti National Park, c) Ruaha National Park, d) Luangwa National Park, e) Limpopo National Park.

## Landcover

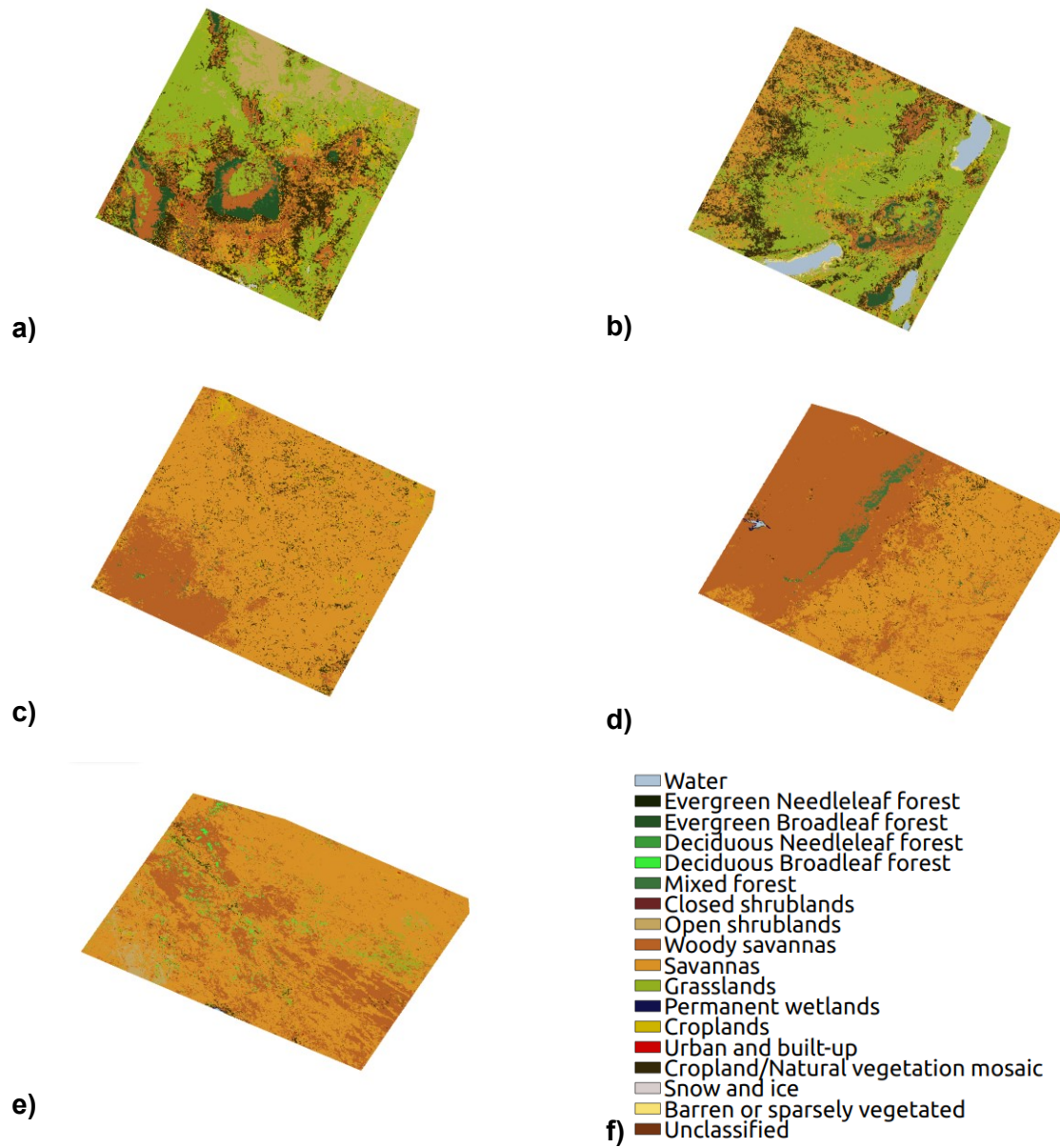
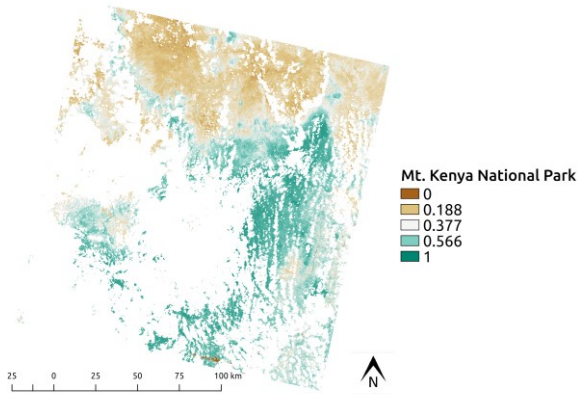


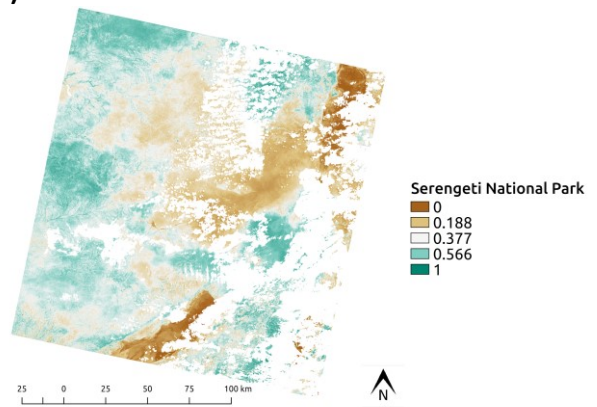
Figure S1b: **Land cover maps.** Land cover maps of the five study sites. a) Kenya National Park, b) Serengeti National Park, c) Ruaha National Park, d) Luangwa National Park, e) Limpopo National Park.

## NDVI

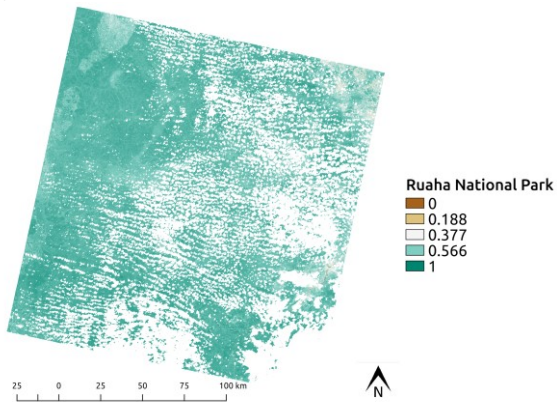
a)



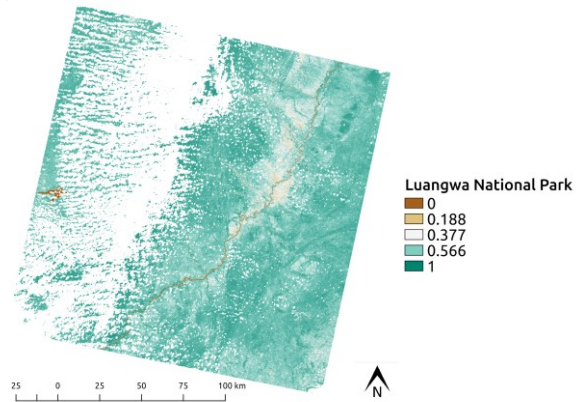
b)



c)



d)



e)

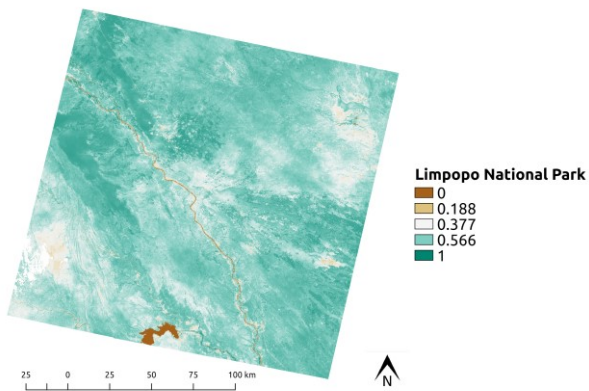


Figure S1c: **NDVI maps**. NDVI calculated from the first date of each site. a) Kenya National Park, b) Serengeti National Park, c) Ruaha National Park, d) Luangwa National Park, e) Limpopo National Park.

## Aspect

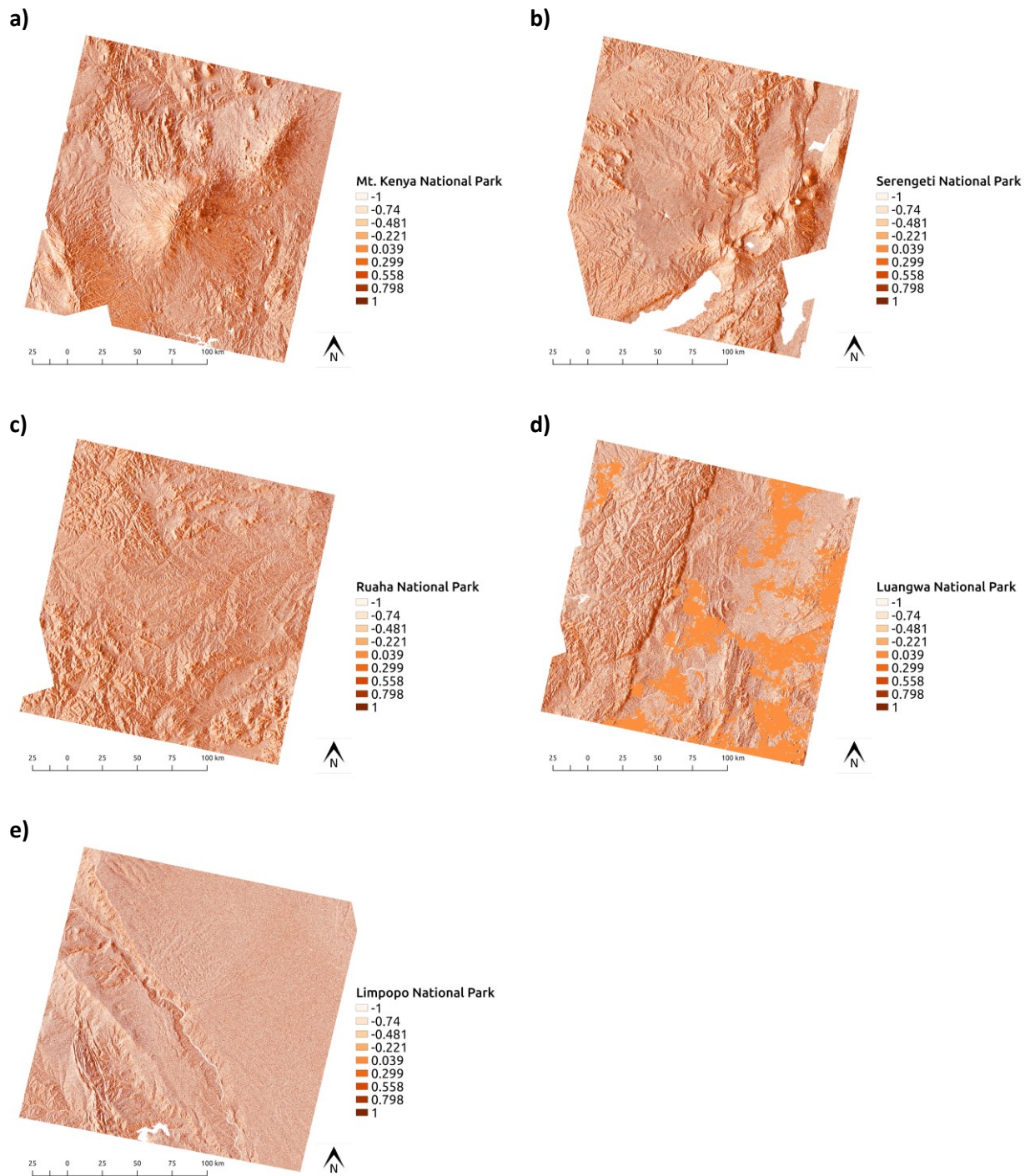


Figure S1d: **Transformed aspect maps.** Transformed aspect map for each of the five study sites. a) Kenya National Park, b) Serengeti National Park, c) Ruaha National Park, d) Luangwa National Park, e) Limpopo National Park.

## Slope

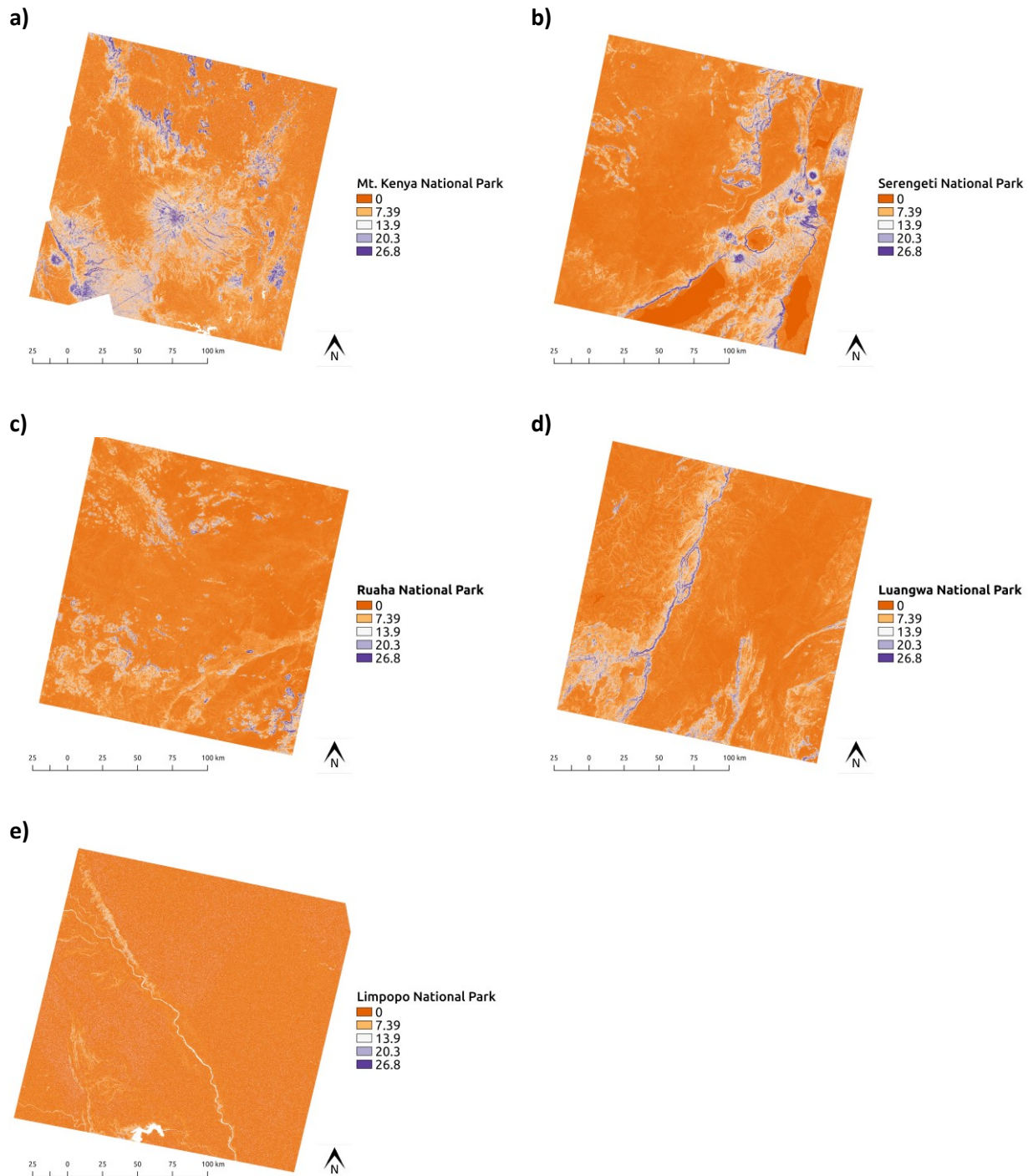


Figure S1e: **Slope maps.** Slope (degrees) maps for each of the five study sites. a) Kenya National Park, b) Serengeti National Park, c) Ruaha National Park, d) Luangwa National Park, e) Limpopo National Park.

## Elevation

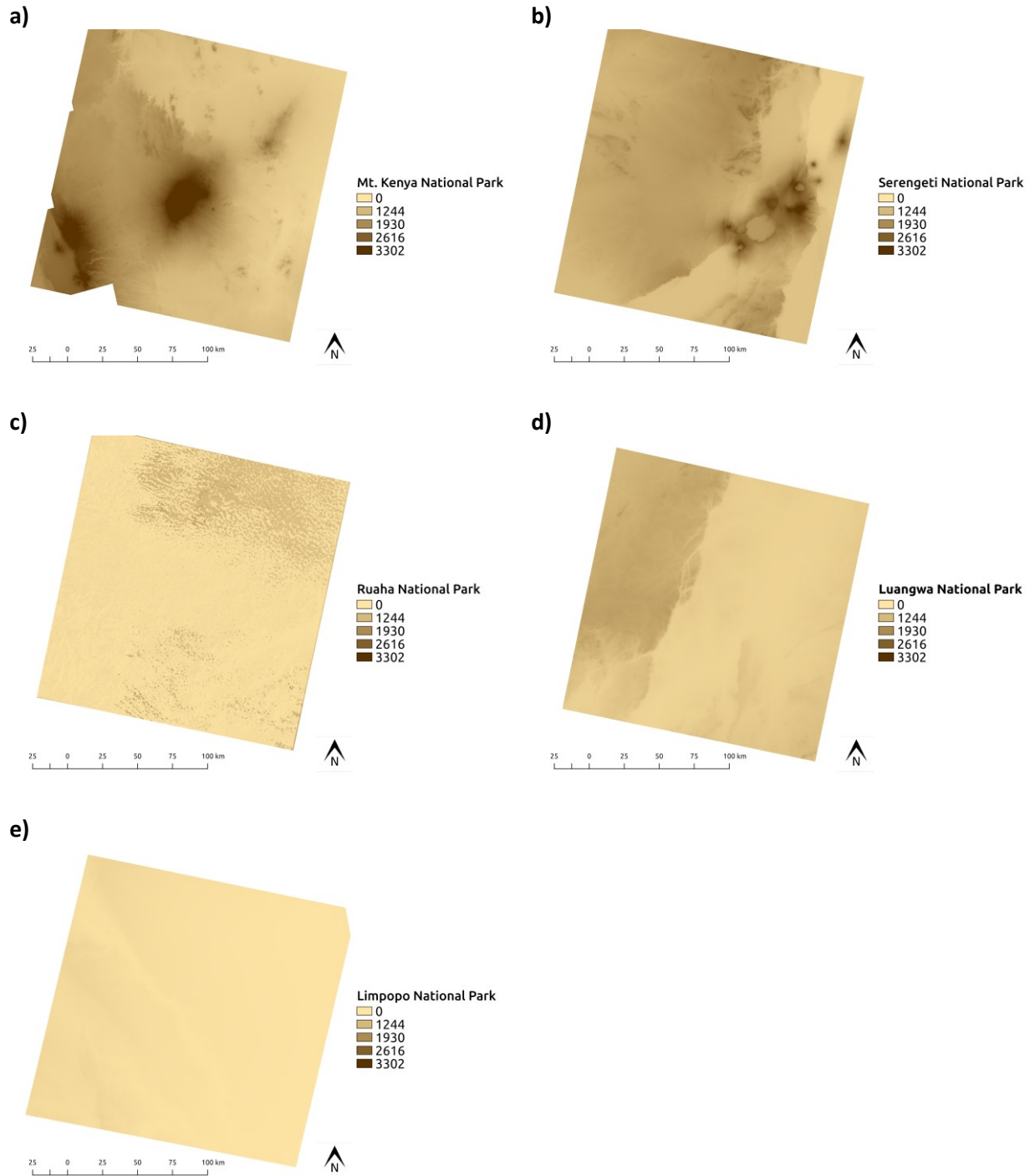


Figure S1f: **Elevation maps.** Elevation maps for each of the five study sites. a) Kenya National Park, b) Serengeti National Park, c) Ruaha National Park, d) Luangwa National Park, e) Limpopo National Park.

## Kenya

### PCA

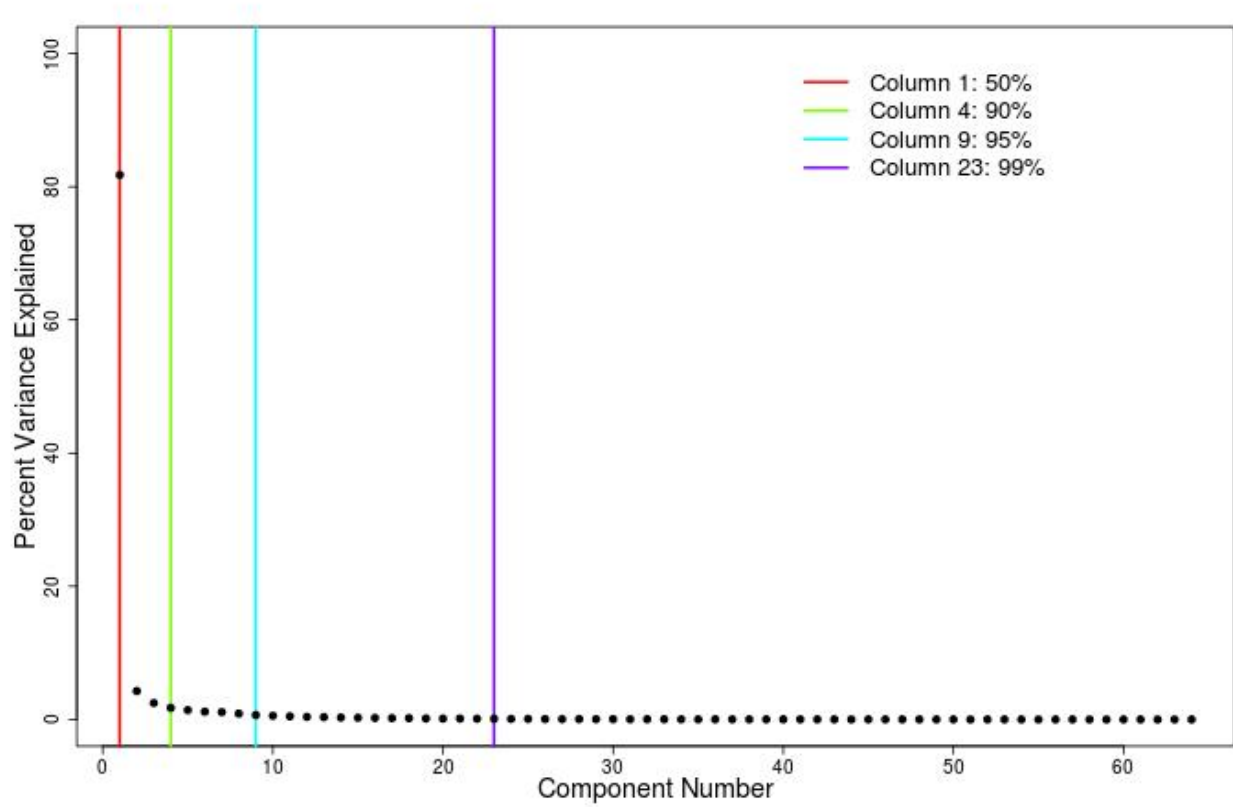


Figure S2a: **Mt. Kenya scree plot.** Scree plot of principal components (PCs) showing which PCs capture 50%, 90%, 95% and 99% of the variance in the NDVI time series. The first 4 PCs explain 90% of variability so are used in the rest of the analysis.

## Determining number of clusters

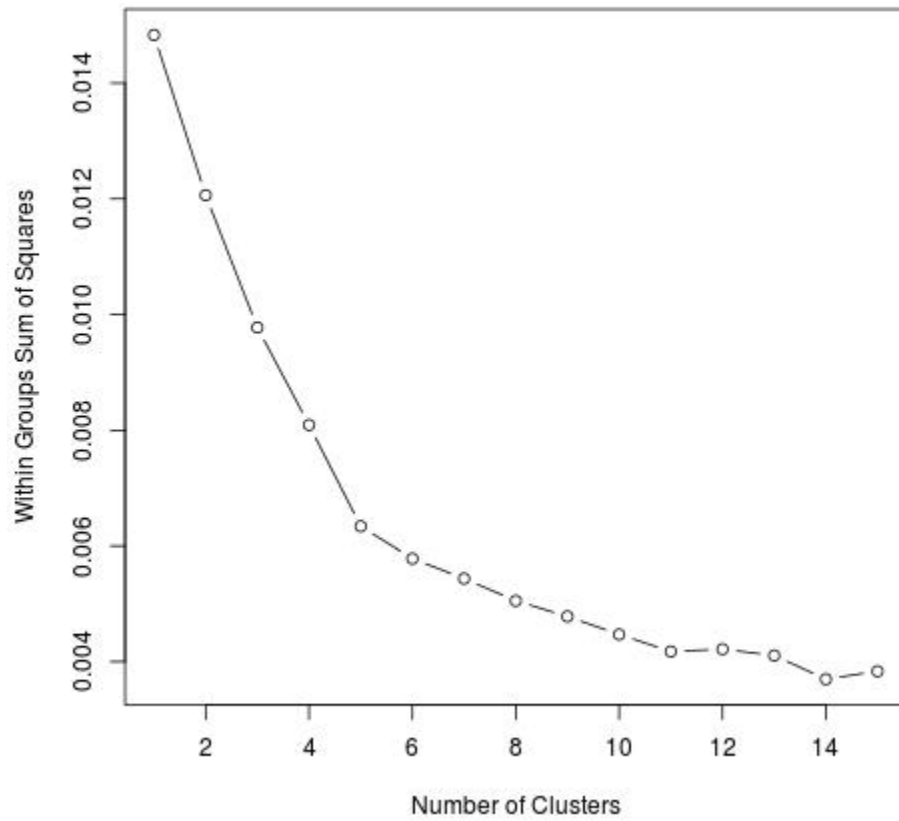


Figure S2b: **Mt. Kenya elbow method.** Within groups sum of squares vs. number of clusters. The inflection point at 5 clusters represents the optimal number of clusters to be obtained.



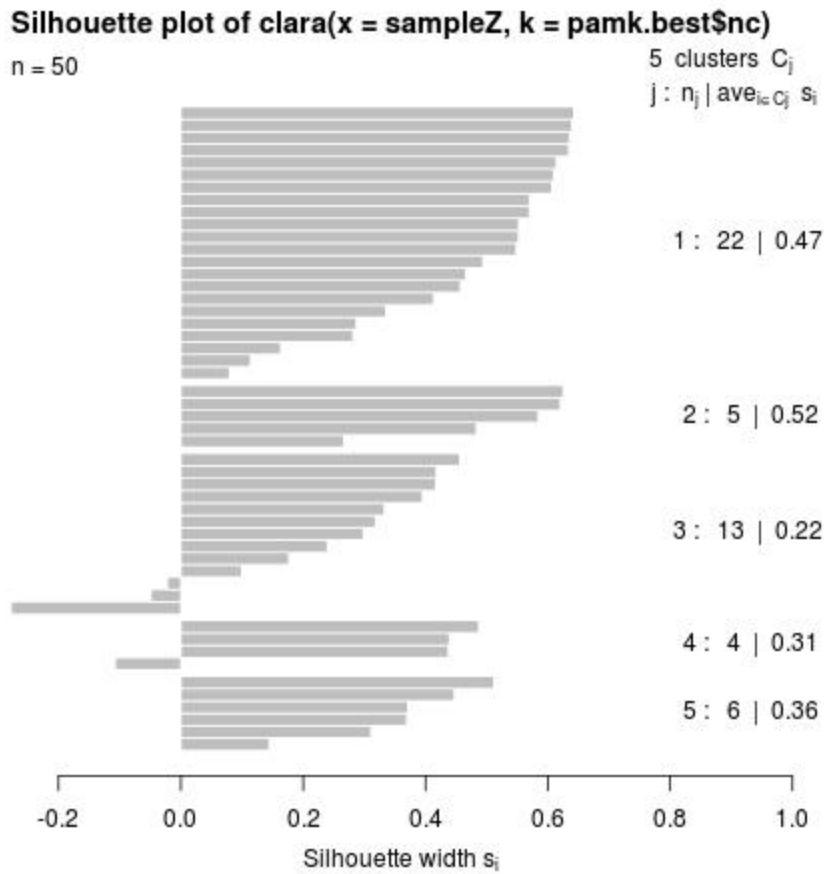


Figure S2c: **Mt. Kenya silhouette method.** Silhouette method plot. There are 5 silhouette widths greater than or equal to 0 which represent the number of optimal clusters to be obtained.

## Phenoregion profiles

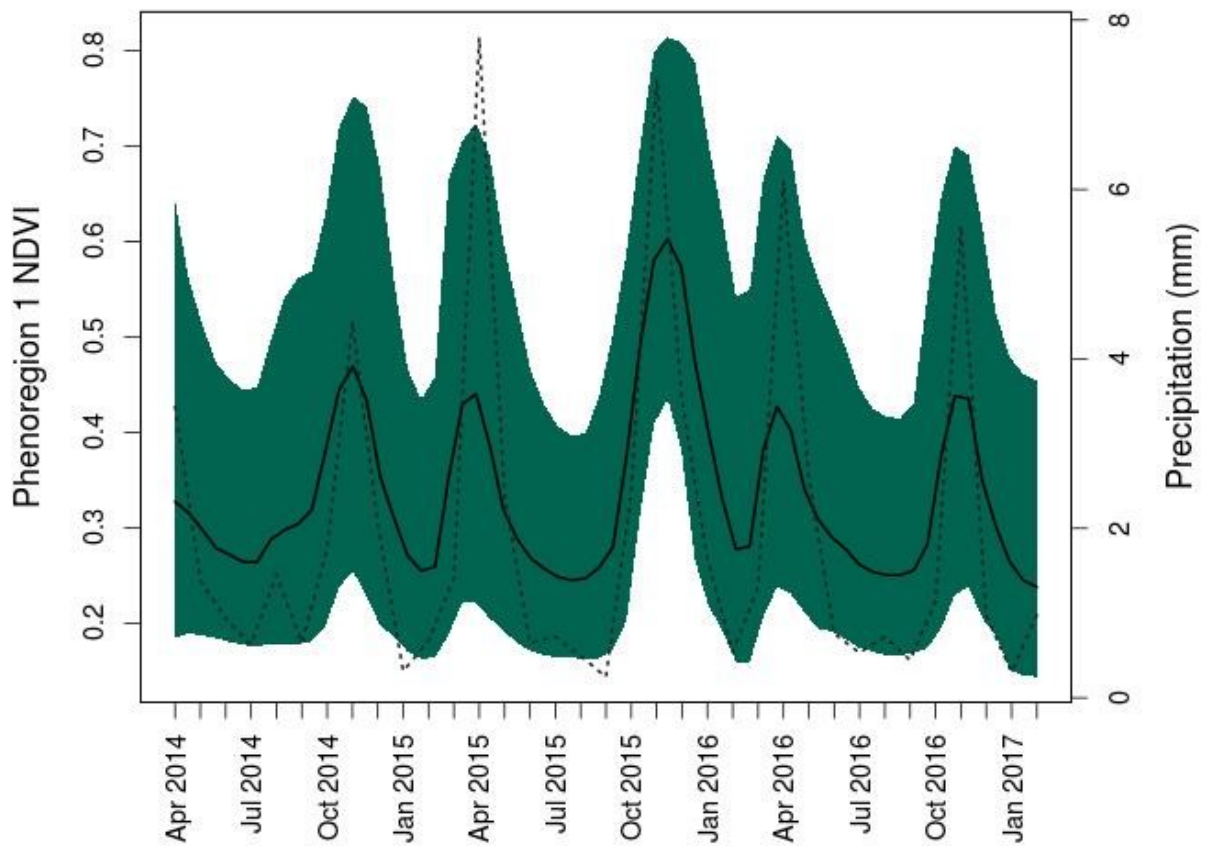


Figure S2d: **Mt. Kenya phenoregion 1 NDVI profile.** Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

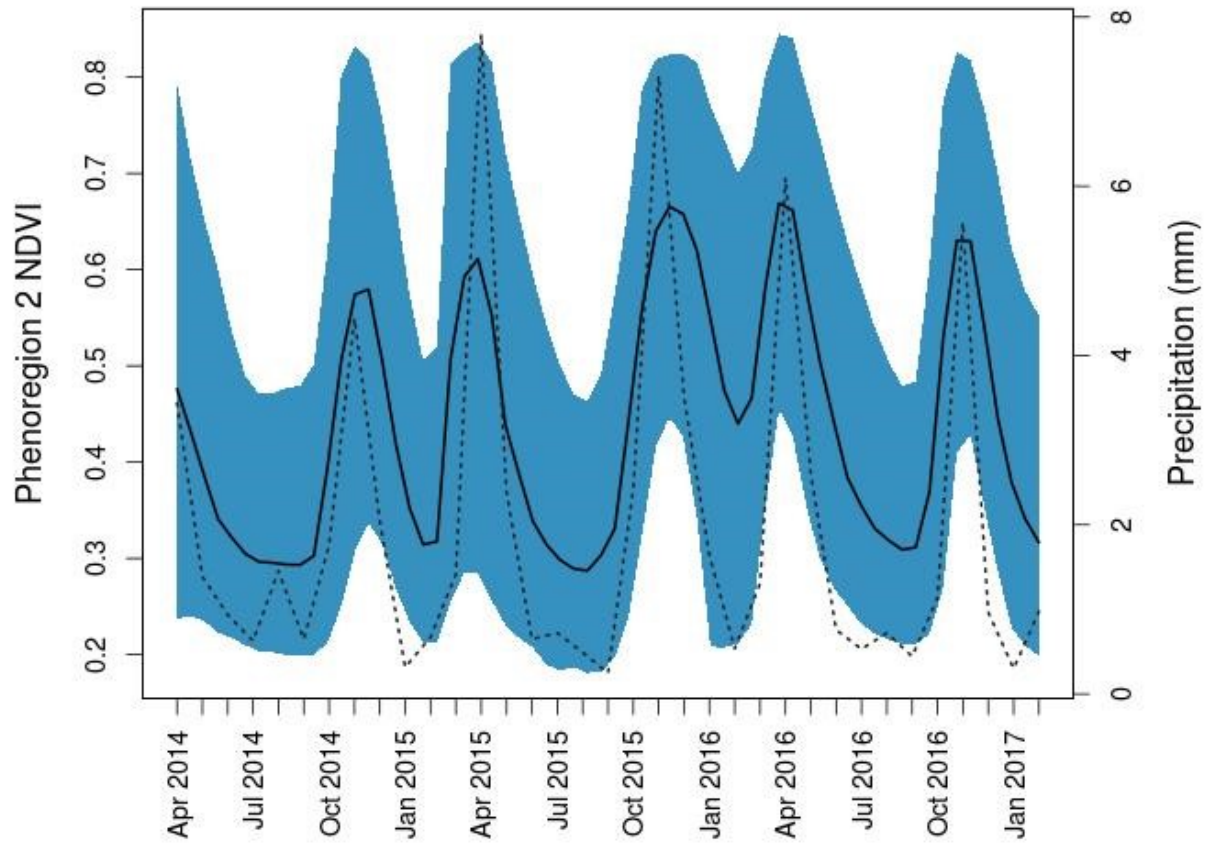


Figure S2e: **Mt. Kenya phenoregion 2 NDVI profile.** Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

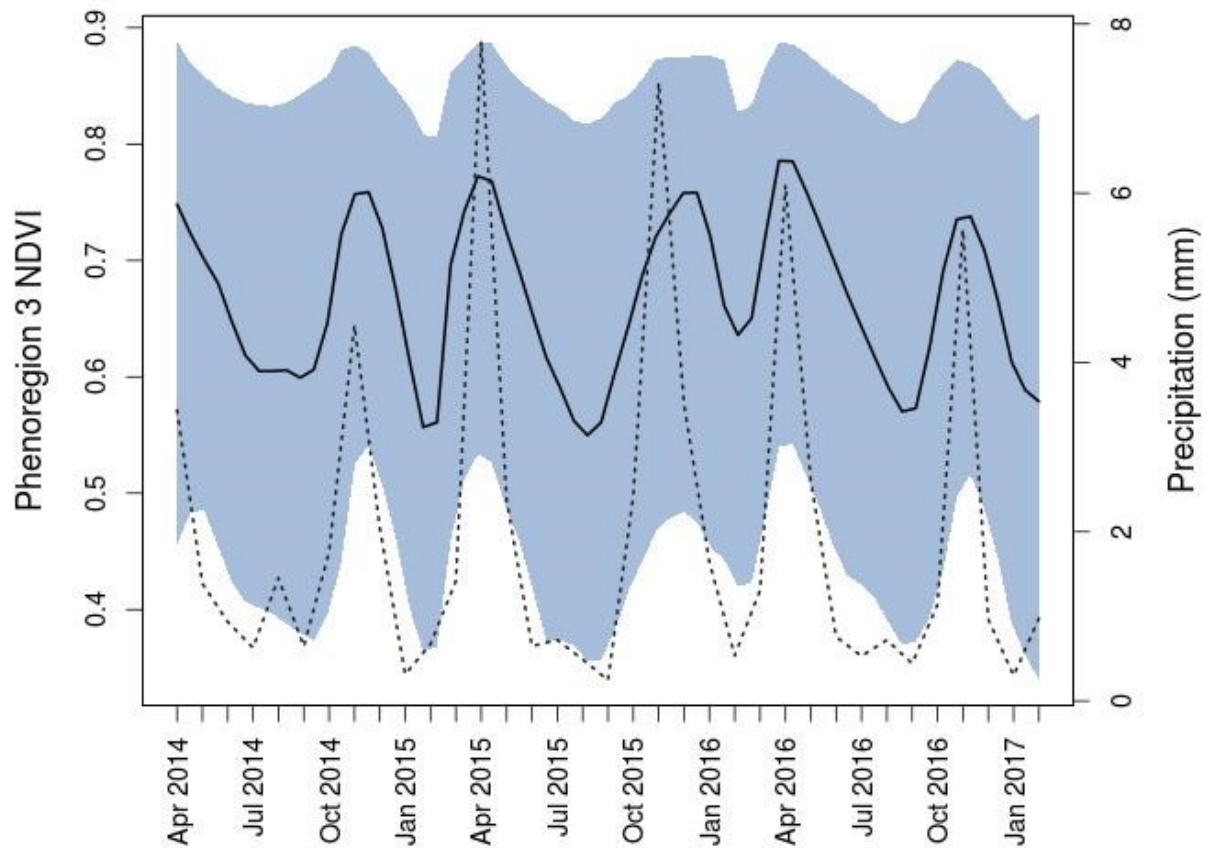


Figure S2f: **Mt. Kenya phenoregion 3 NDVI profile.** Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

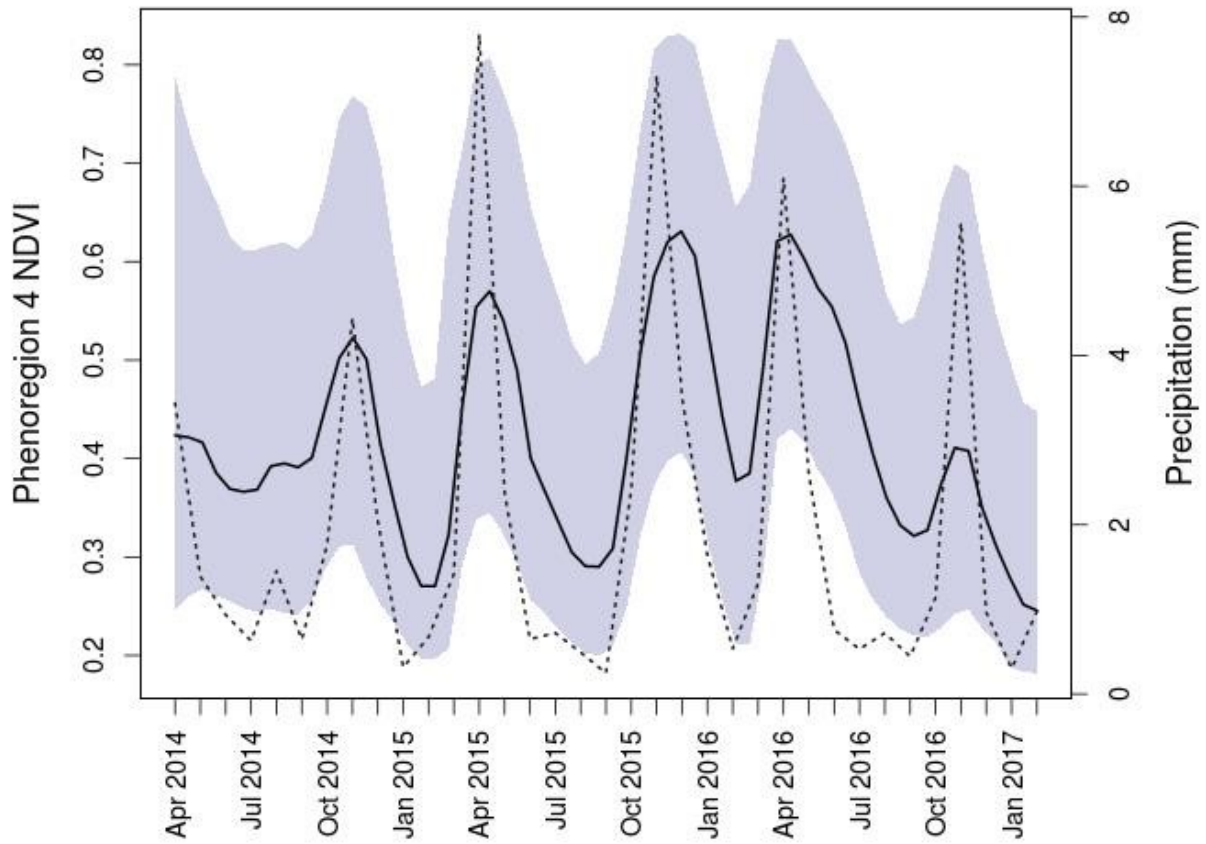


Figure S2g: **Mt. Kenya phenoregion 4 NDVI profile.** Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

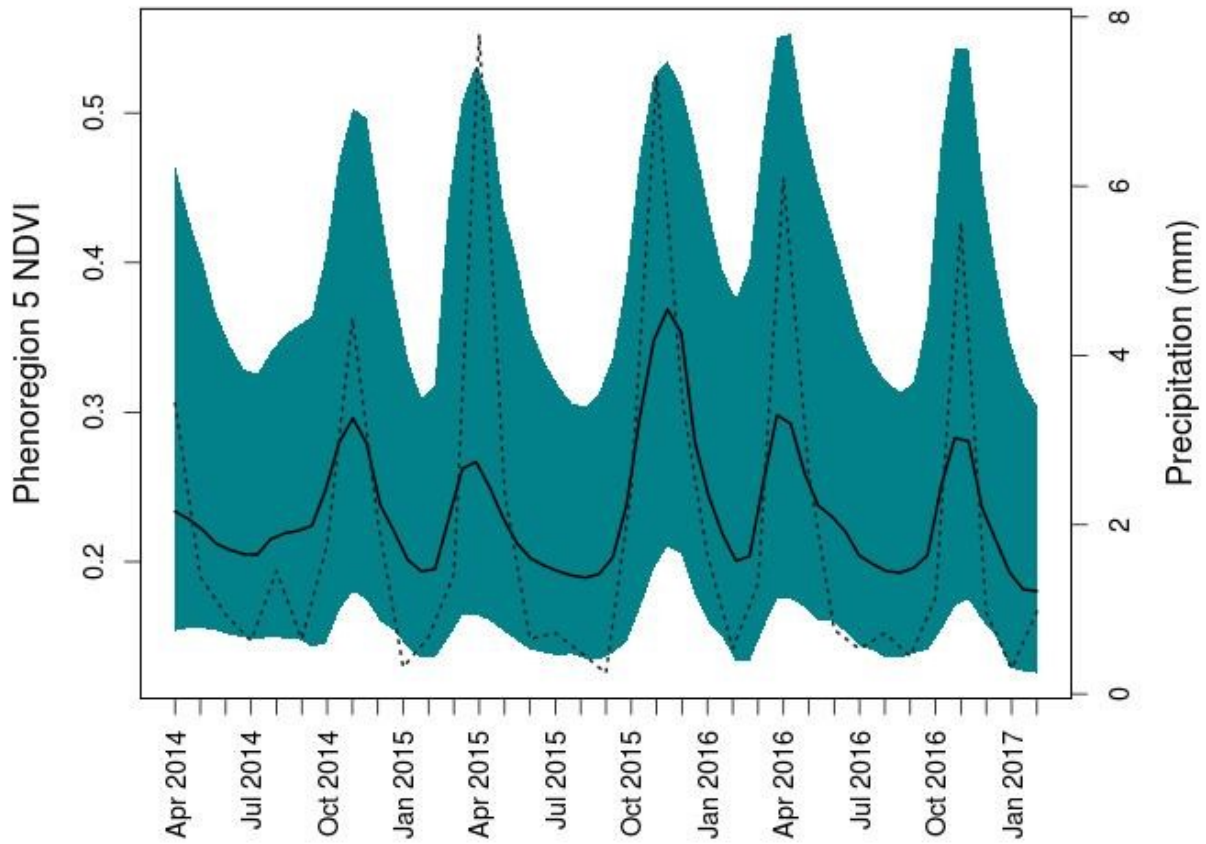


Figure S2h: **Mt. Kenya phenoregion 5 NDVI profile.** Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

## Serengeti

### PCA

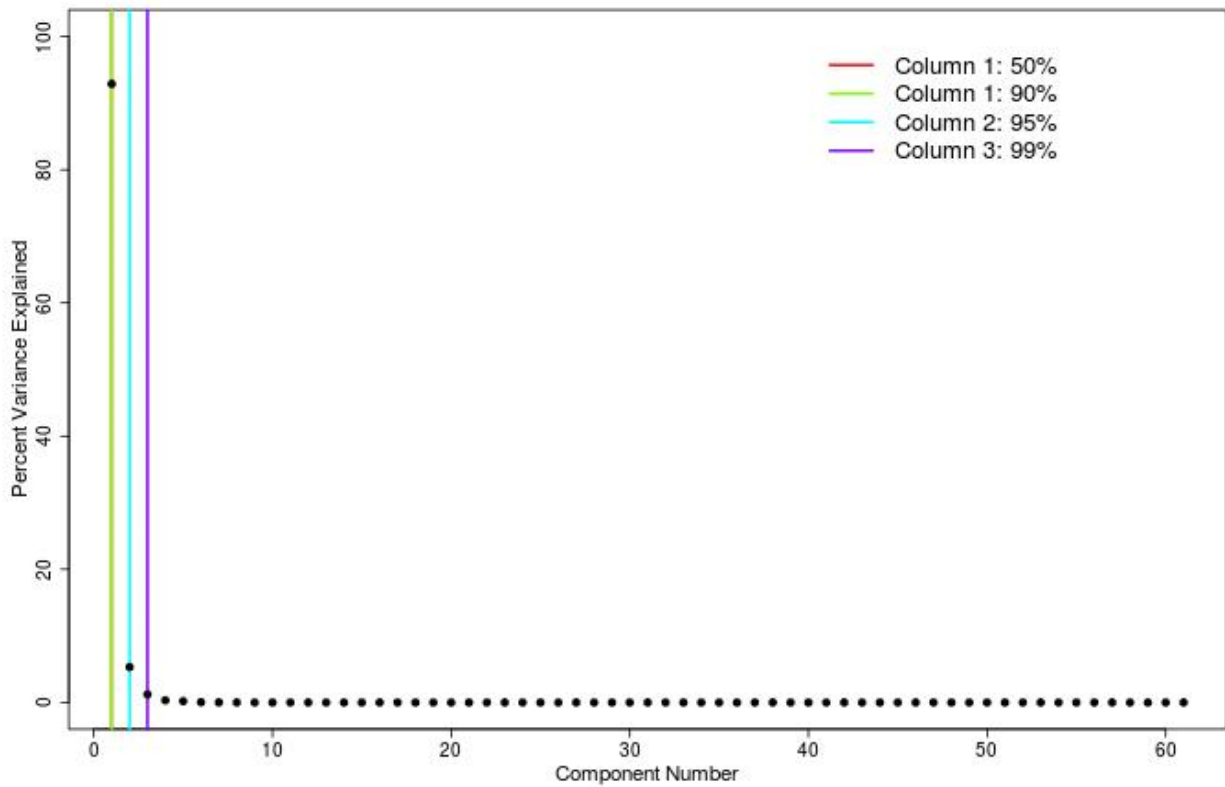


Figure S3a: **Serengeti scree plot**. Scree plot of principal components (PCs) showing which PCs capture 50%, 90%, 95% and 99% of the variance in the NDVI time series. The first PC explains 90% of variability so is used in the rest of the analysis.

Determining number of clusters

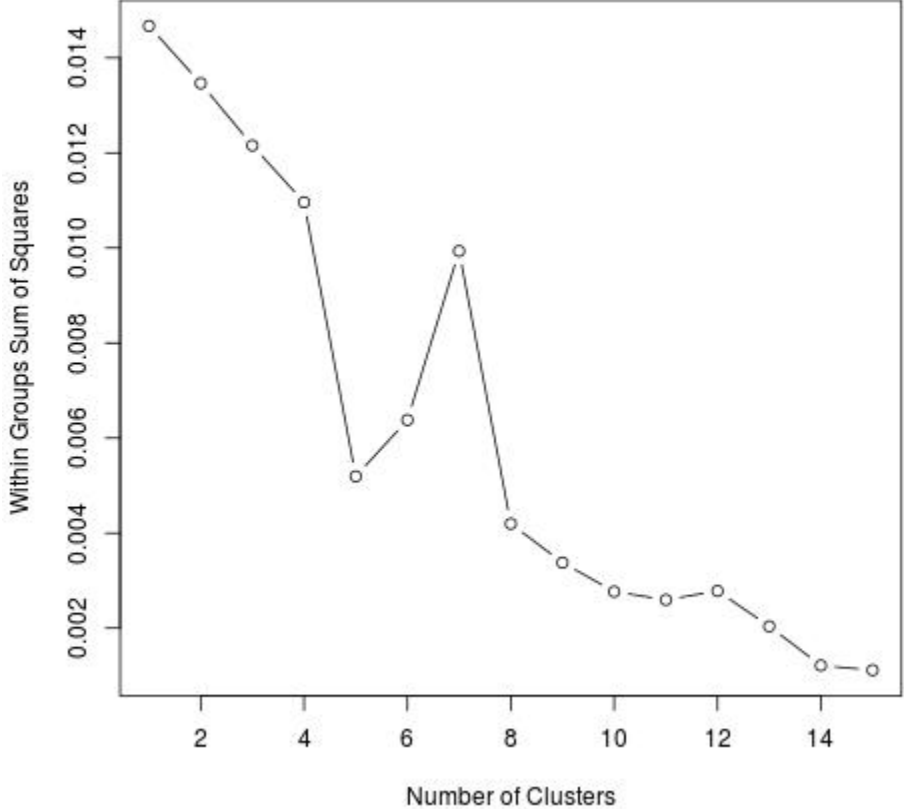
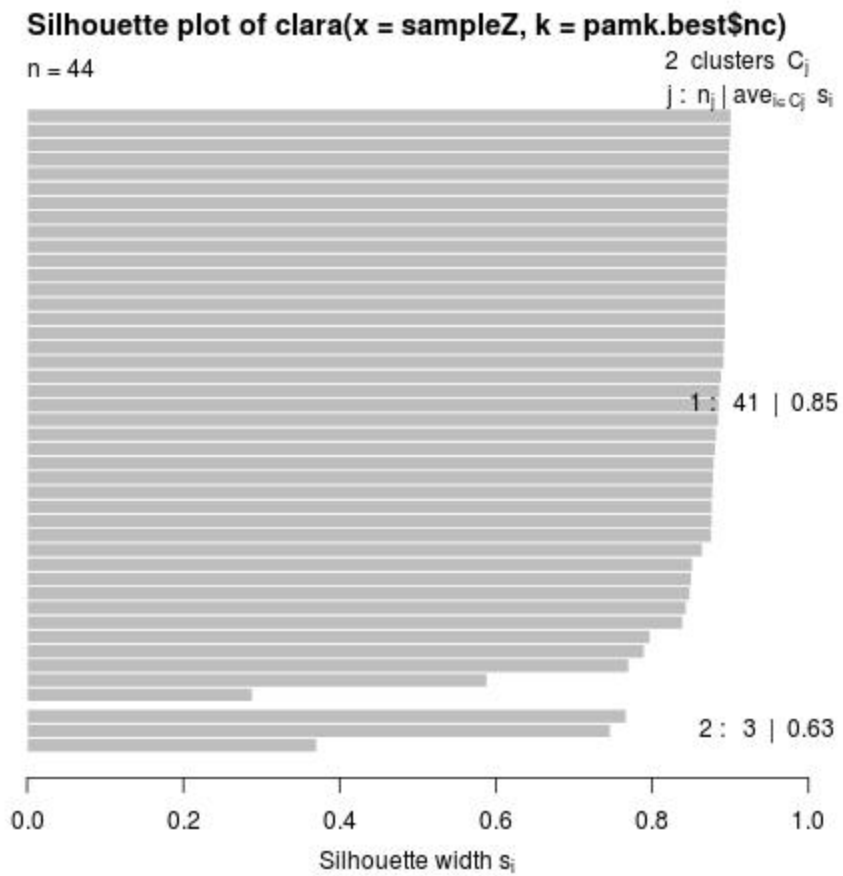


Figure S3b: **Serengeti elbow method**. Within groups sum of squares vs. number of clusters. The inflection point at 7 clusters represents the optimal number of clusters to be obtained.





Average silhouette width : 0.84

Figure S3c: **Serengeti silhouette method**. There are 2 silhouette widths greater than or equal to 0 which represent the number of optimal clusters to be obtained.

## Phenoregion profiles

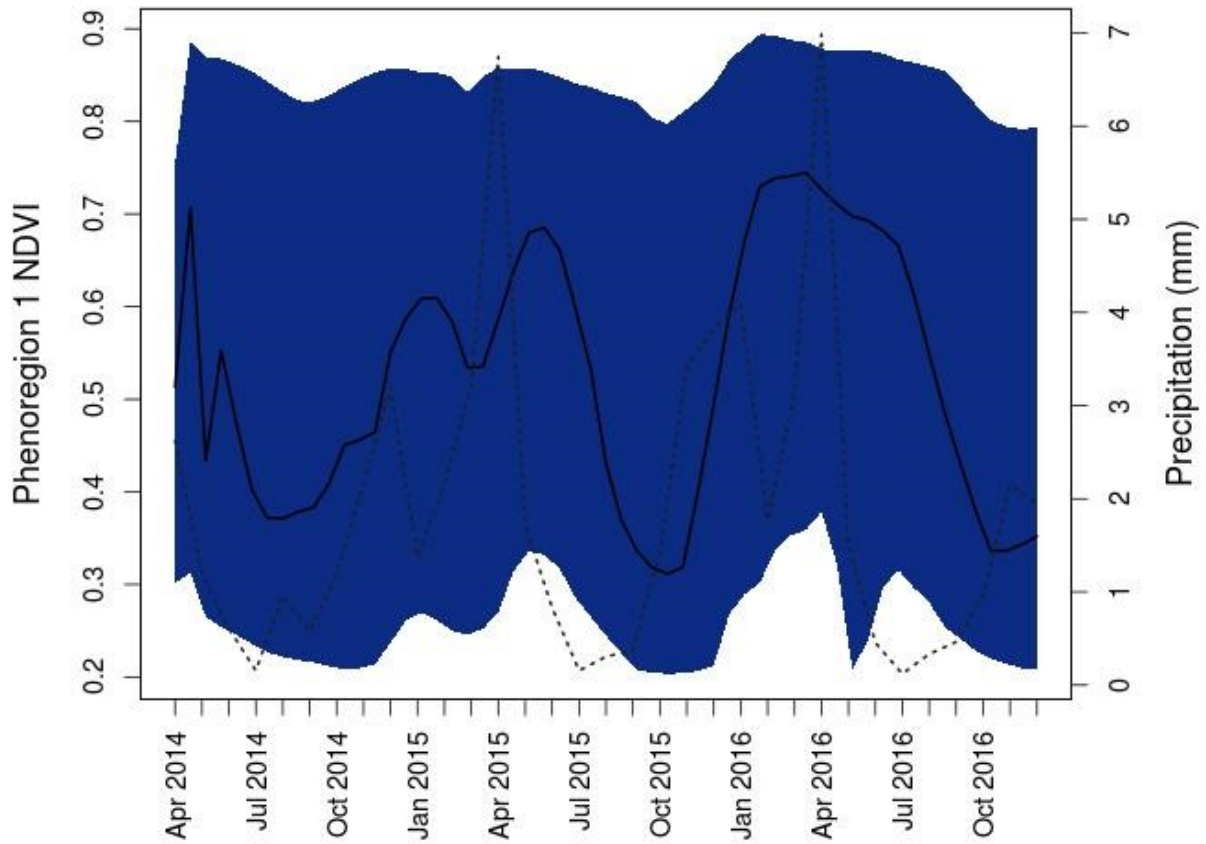


Figure S3d: **Serengeti phenoregion 1 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

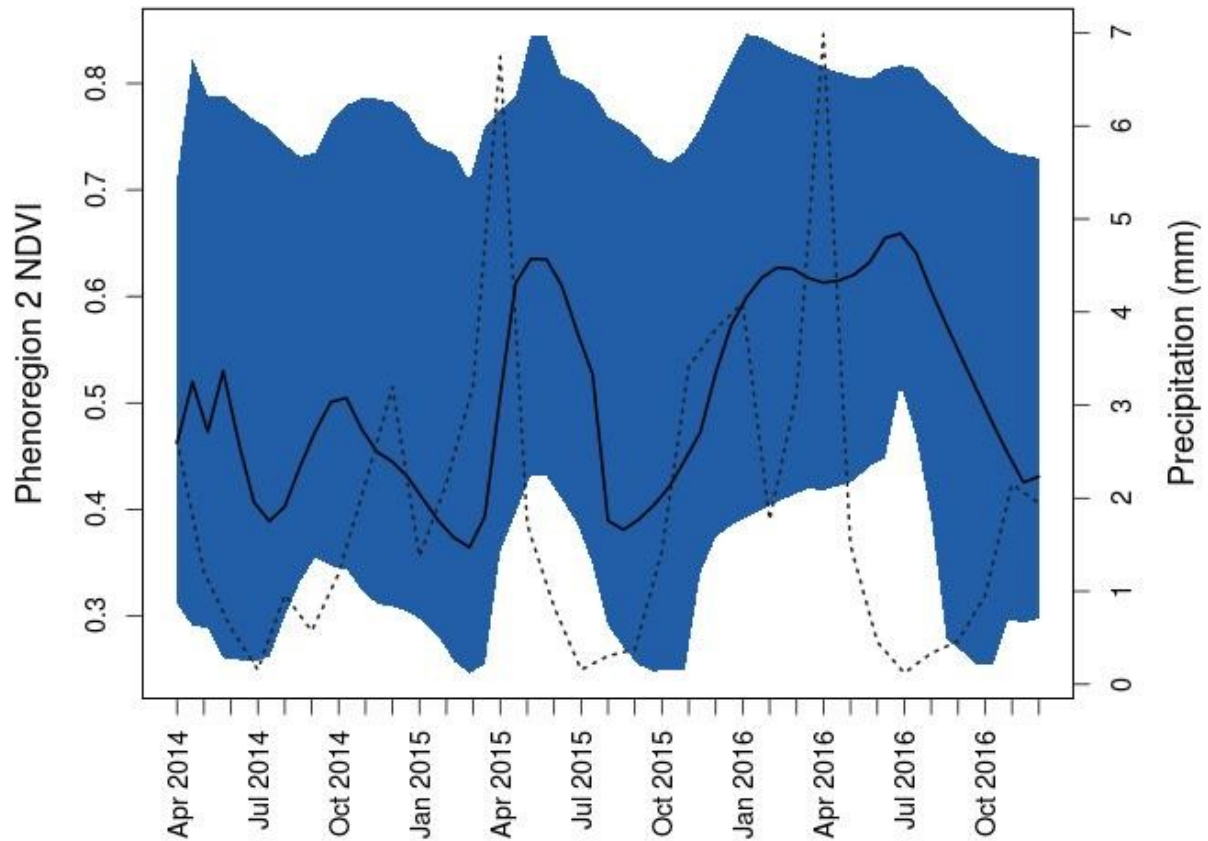


Figure S3e: **Serengeti phenoregion 2 NDVI profile.** Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

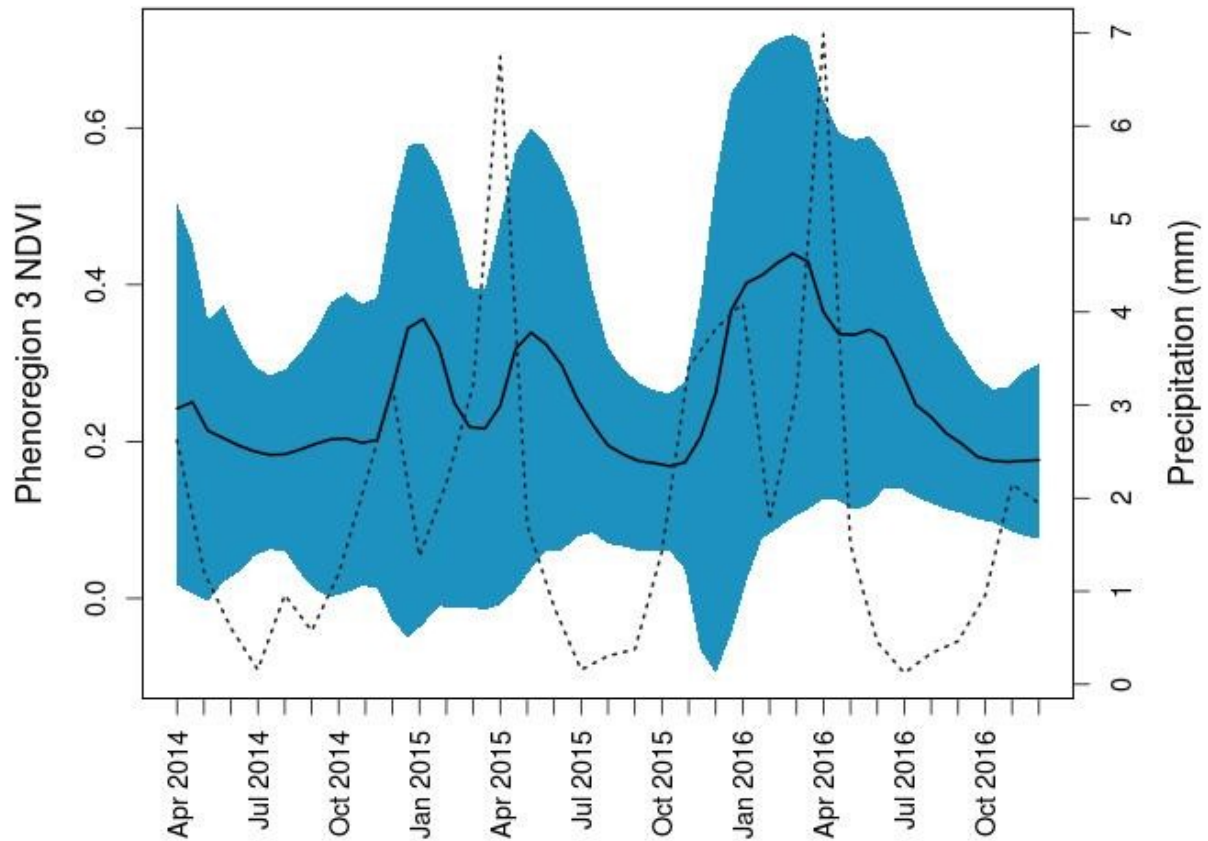


Figure S3f: **Serengeti phenoregion 3 NDVI profile.** Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

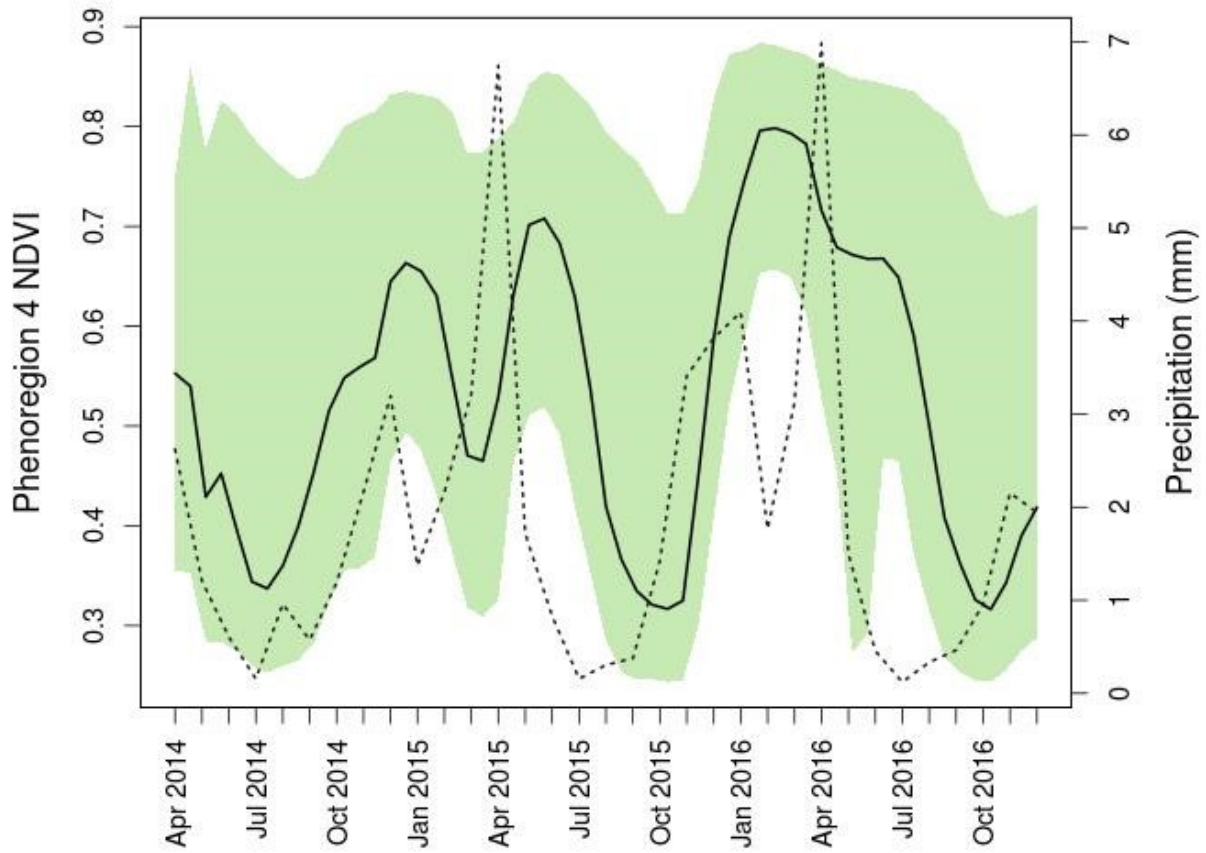


Figure S3g: **Serengeti phenoregion 4 NDVI profile.** Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

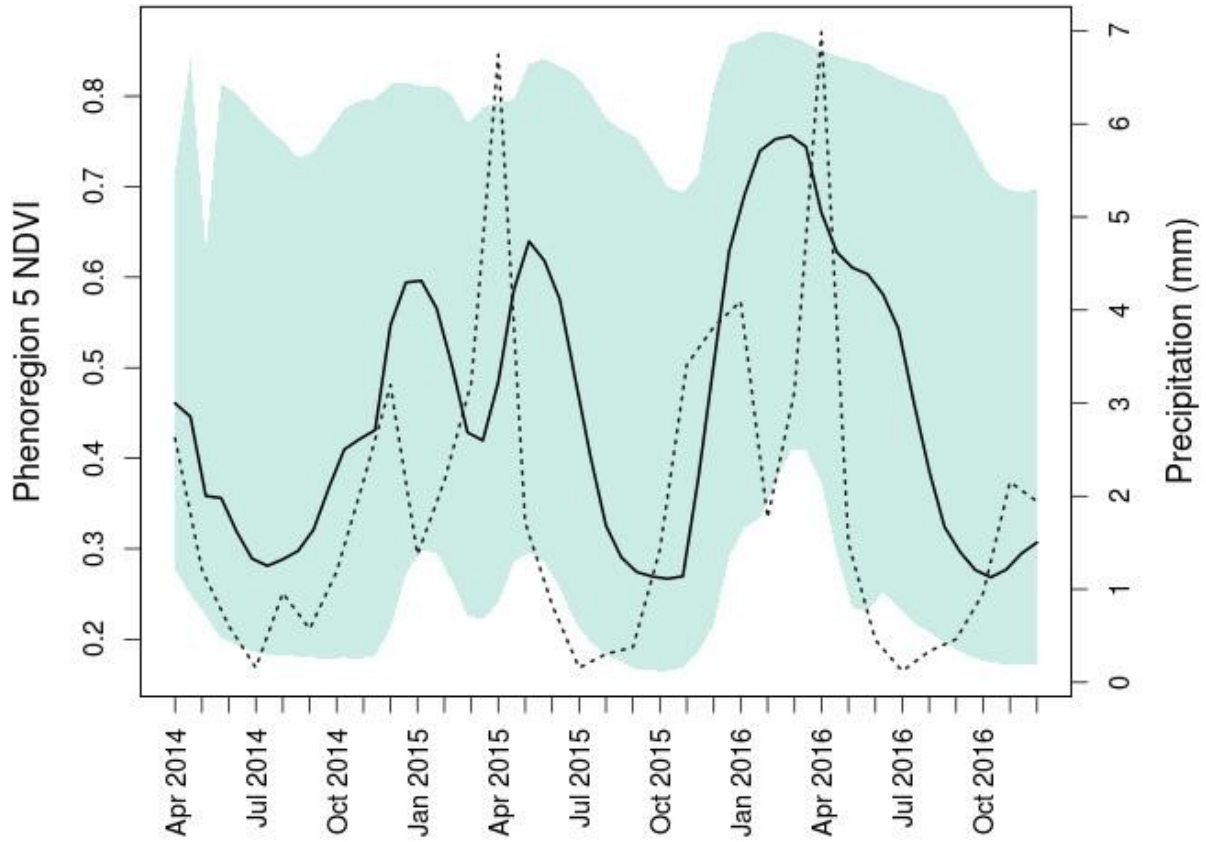


Figure S3h: **Serengeti phenoregion 5 NDVI profile.** Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

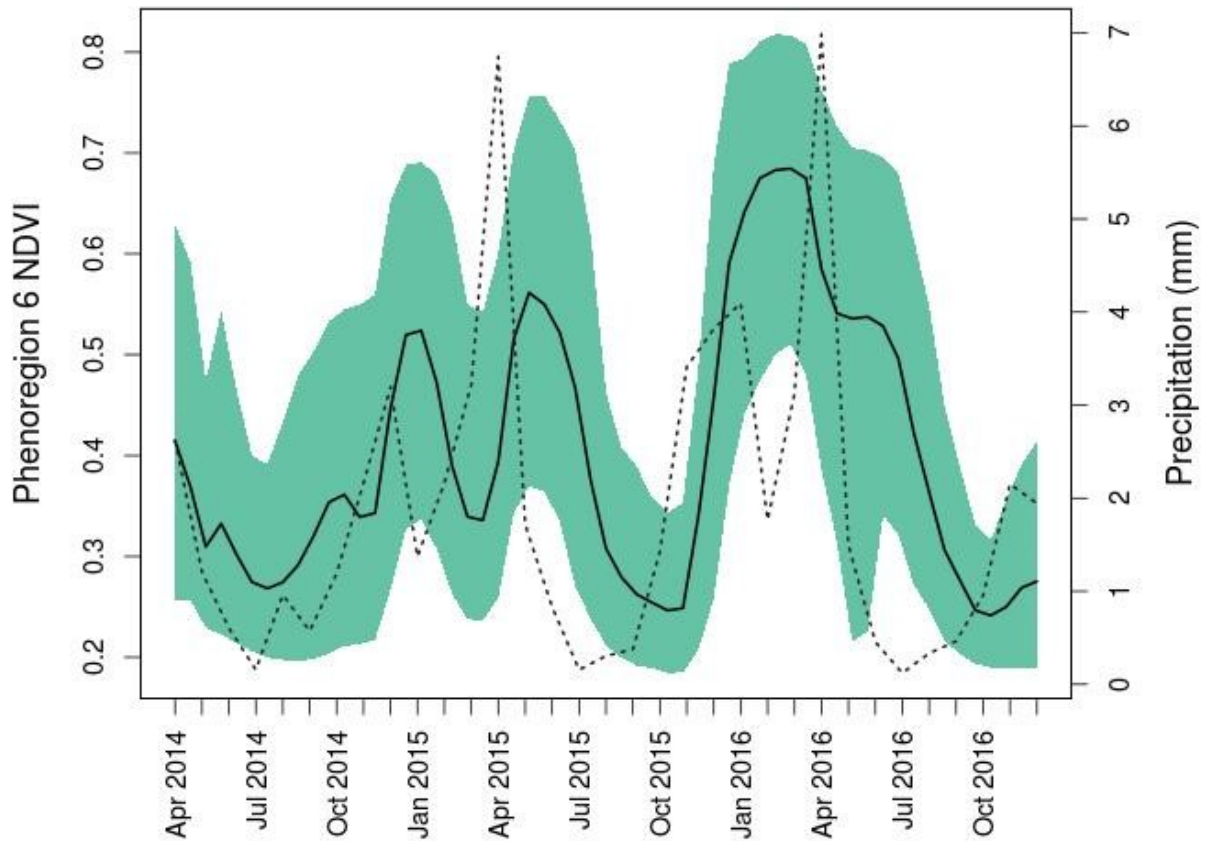


Figure S3i: **Serengeti phenoregion 6 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

# Ruaha

PCA

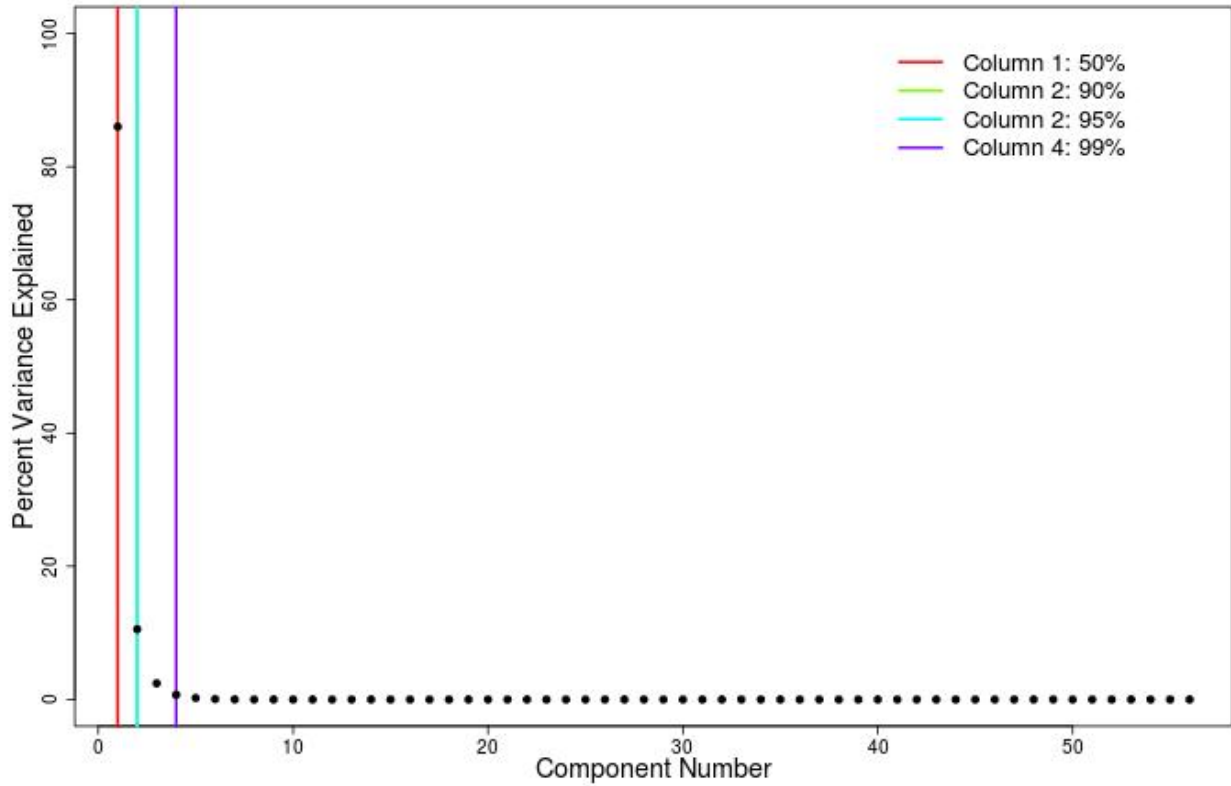


Figure S4a: **Ruaha scree plot.** Scree plot of principal components (PCs) showing which PCs capture 50%, 90%, 95% and 99% of the variance in the NDVI time series. The first 2 PCs explain 90% of variability so are used in the rest of the analysis.



Determining number of clusters

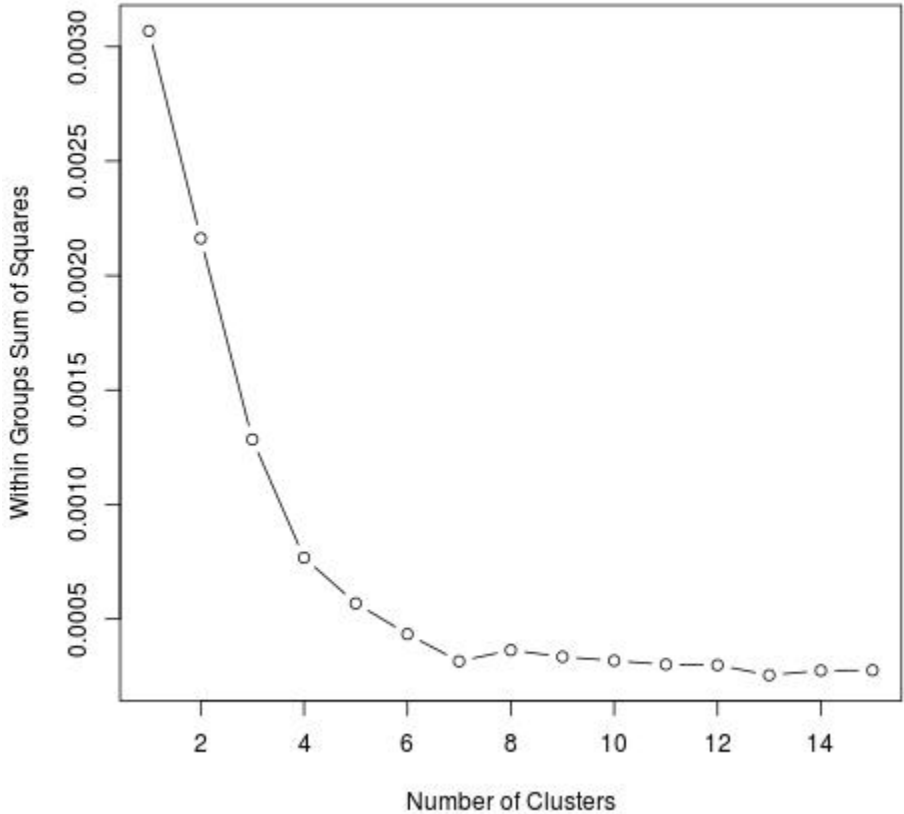
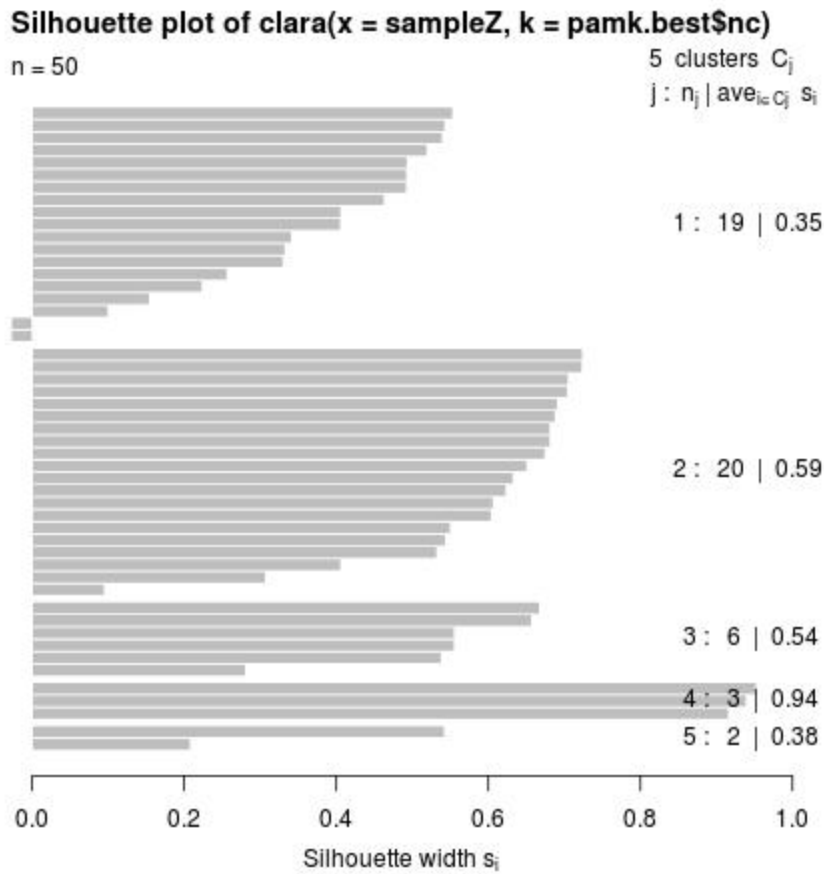


Figure S4b: **Ruaha elbow method**. Within groups sum of squares vs. number of clusters. The inflection point at 7 clusters represents the optimal number of clusters to be obtained.



Average silhouette width : 0.5

Figure S4c: **Ruaha silhouette method**. There are 5 silhouette widths greater than or equal to 0 which represent the number of optimal clusters to be obtained.

## Phenoregion profiles

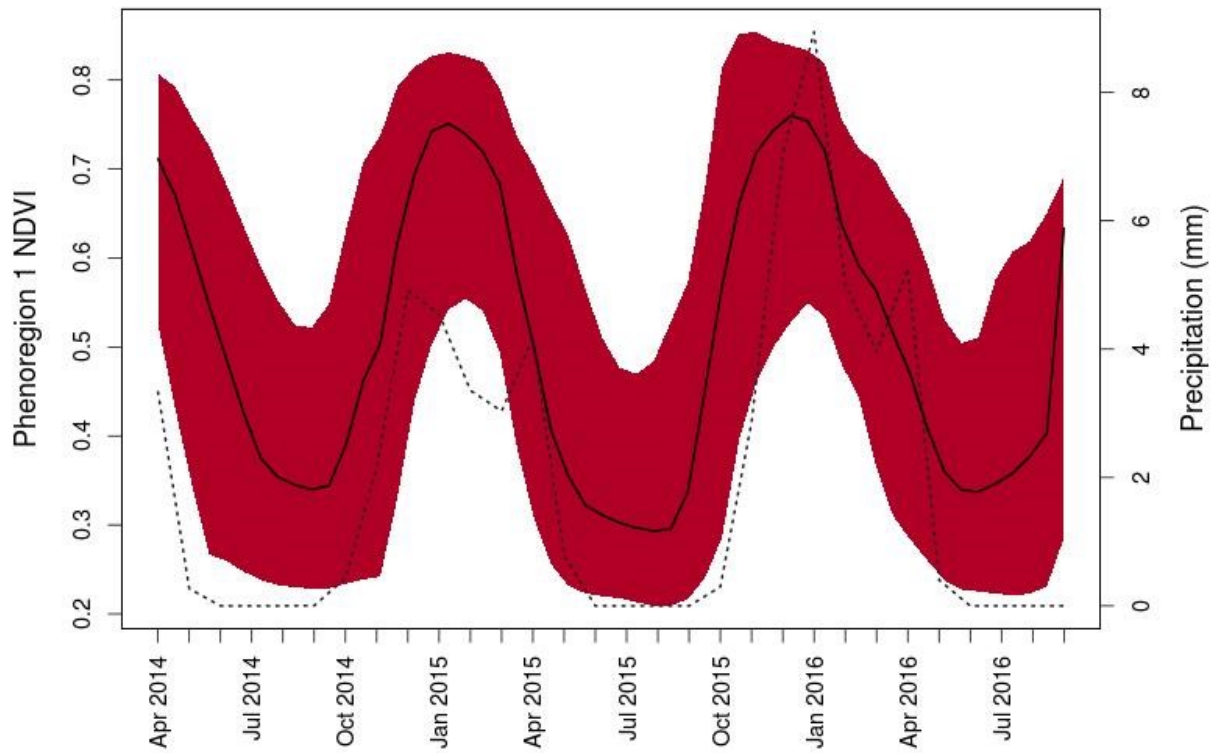


Figure S4d: **Ruaha phenoregion 1 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

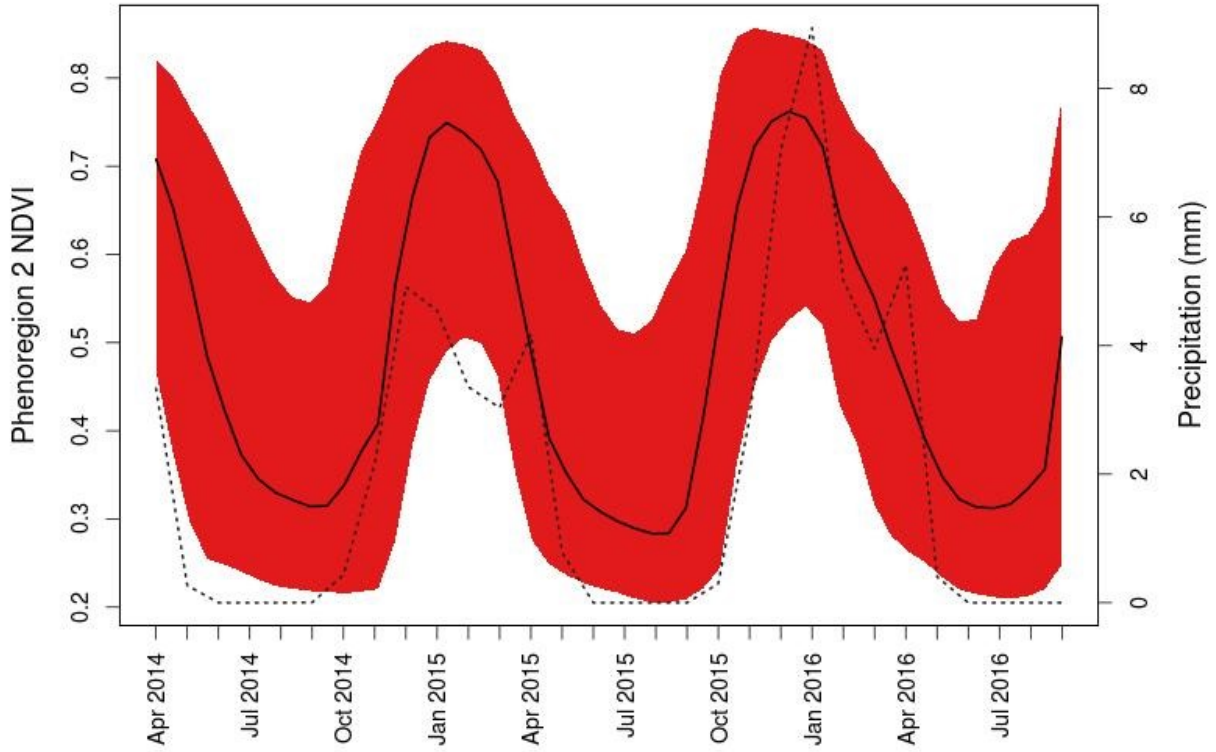


Figure S4e: **Ruaha phenoregion 2 NDVI profile.** Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

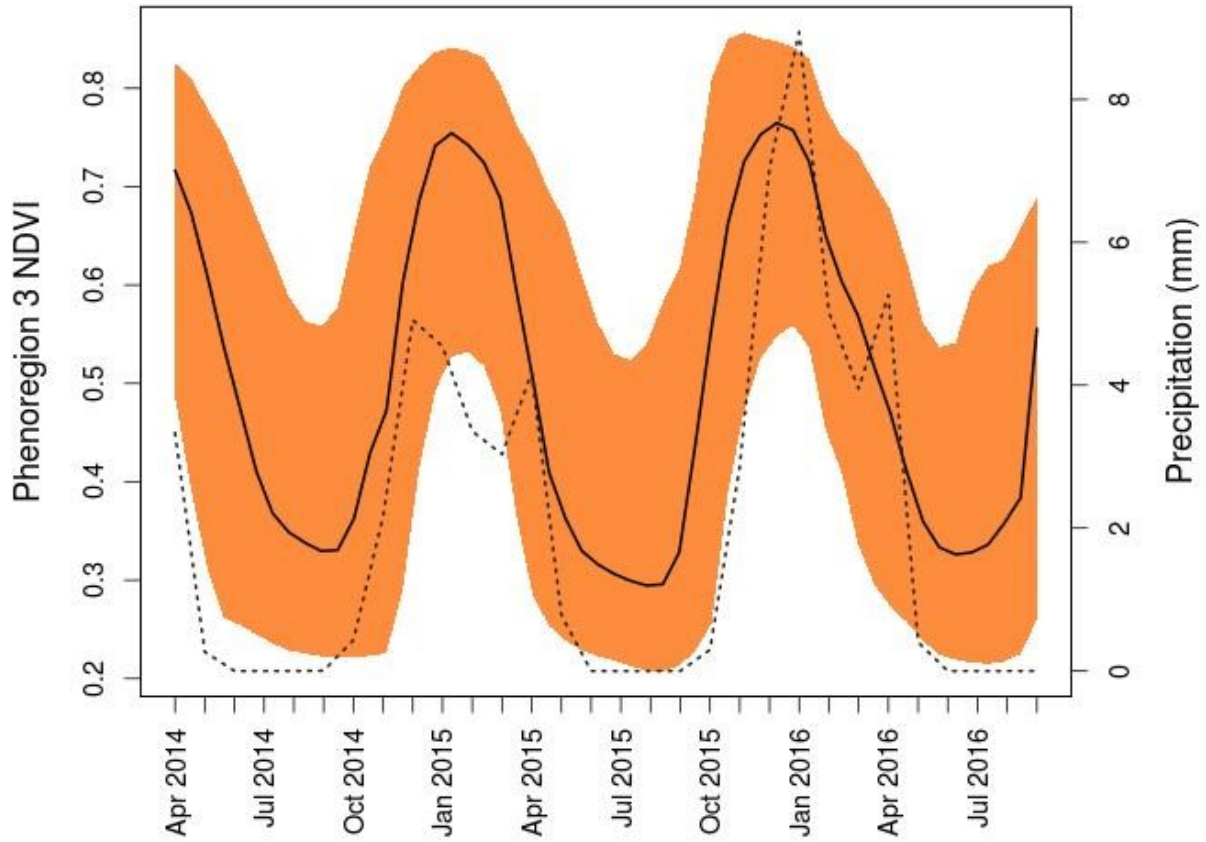


Figure S4f: **Ruaha phenoregion 3 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

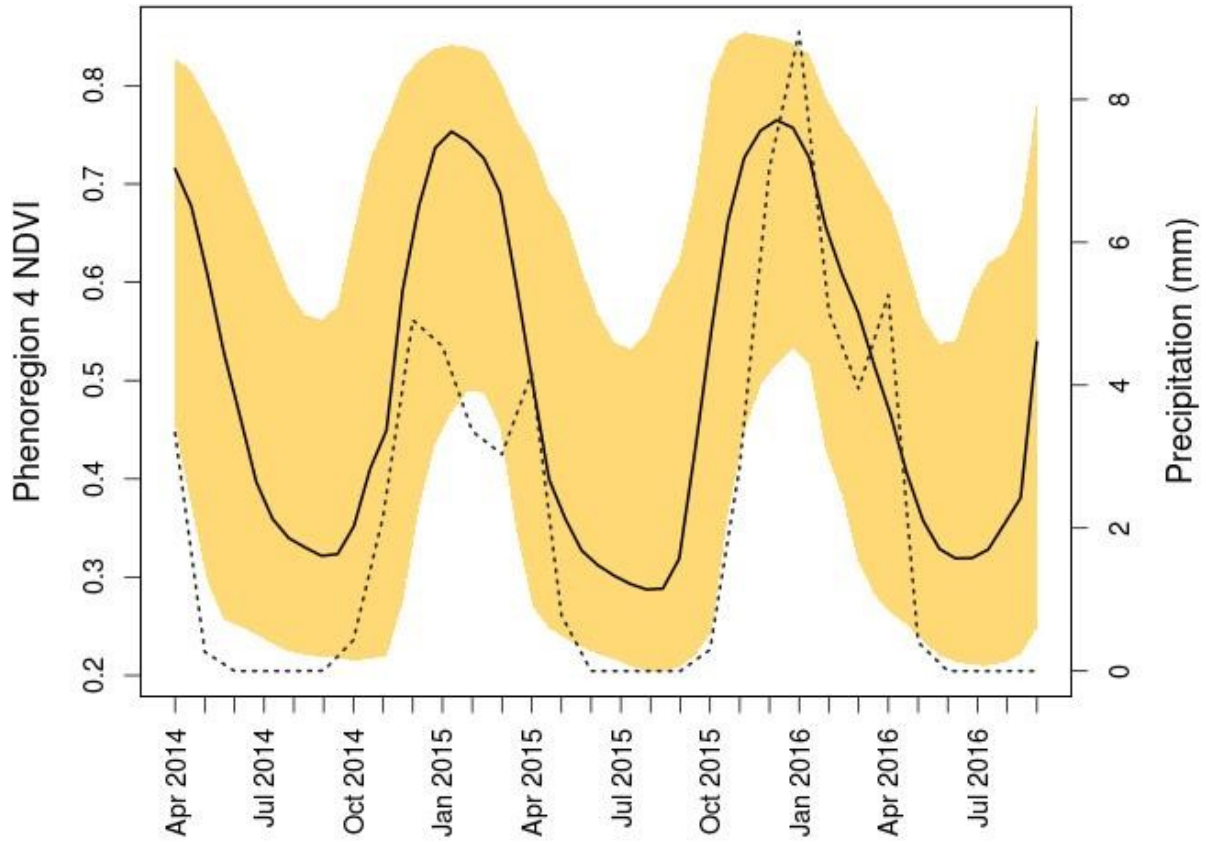


Figure S4g: **Ruaha phenoregion 4 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

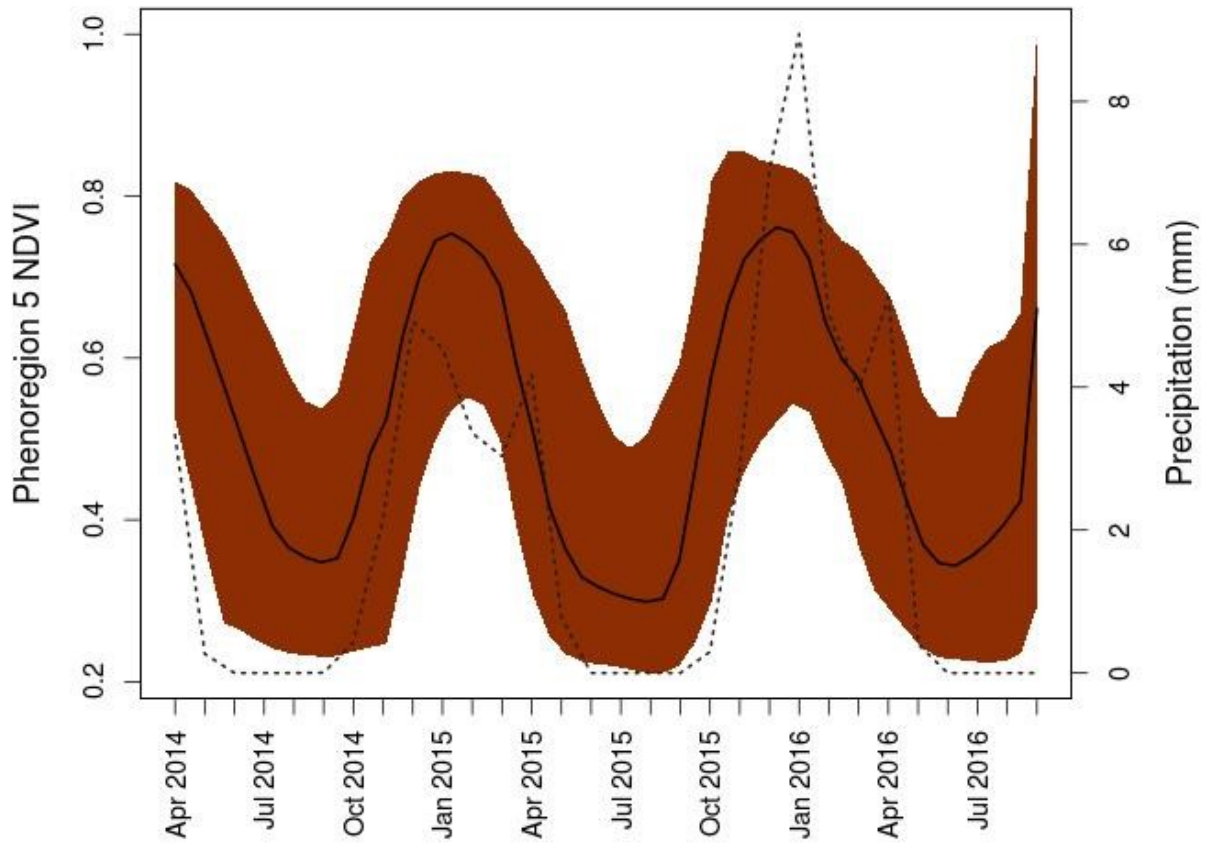


Figure S4h: **Ruaha phenoregion 5 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

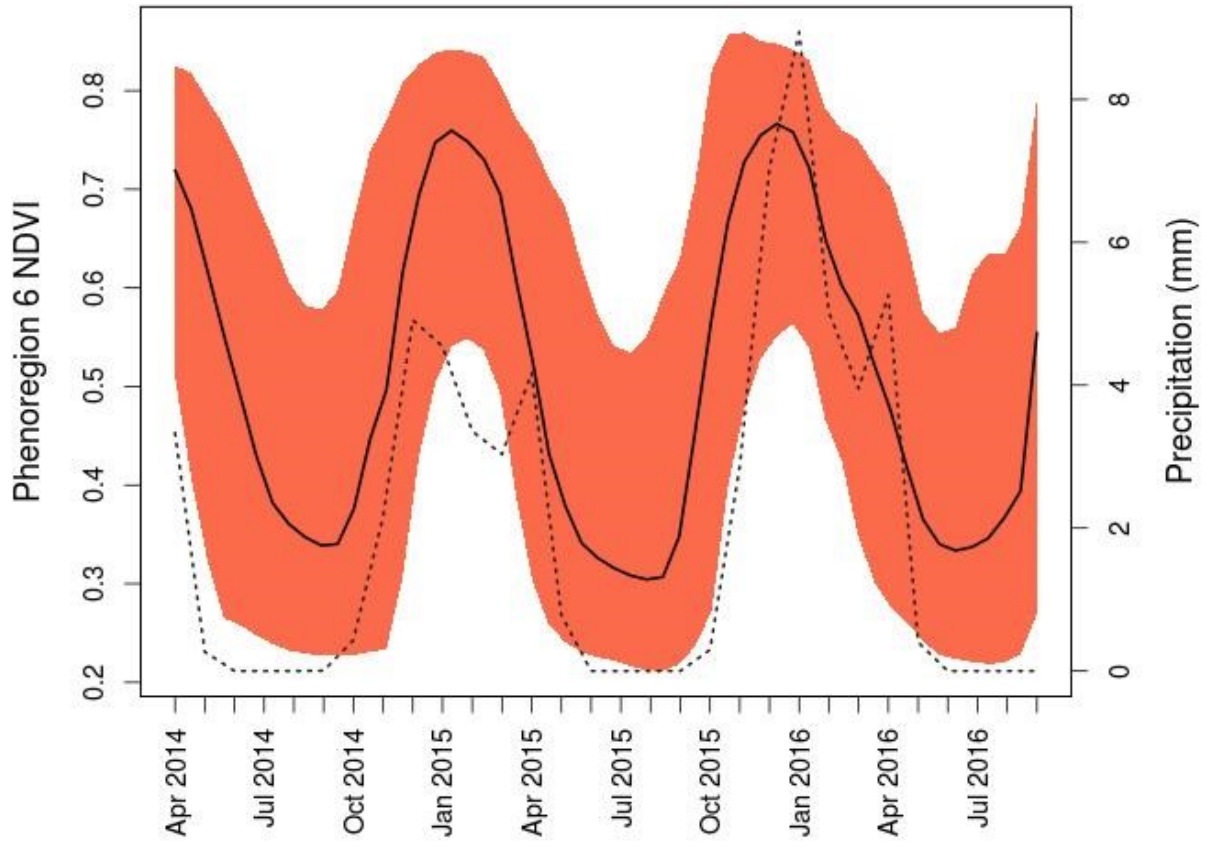


Figure S4i: **Ruaha phenoregion 6 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).



## Luangwa

### PCA

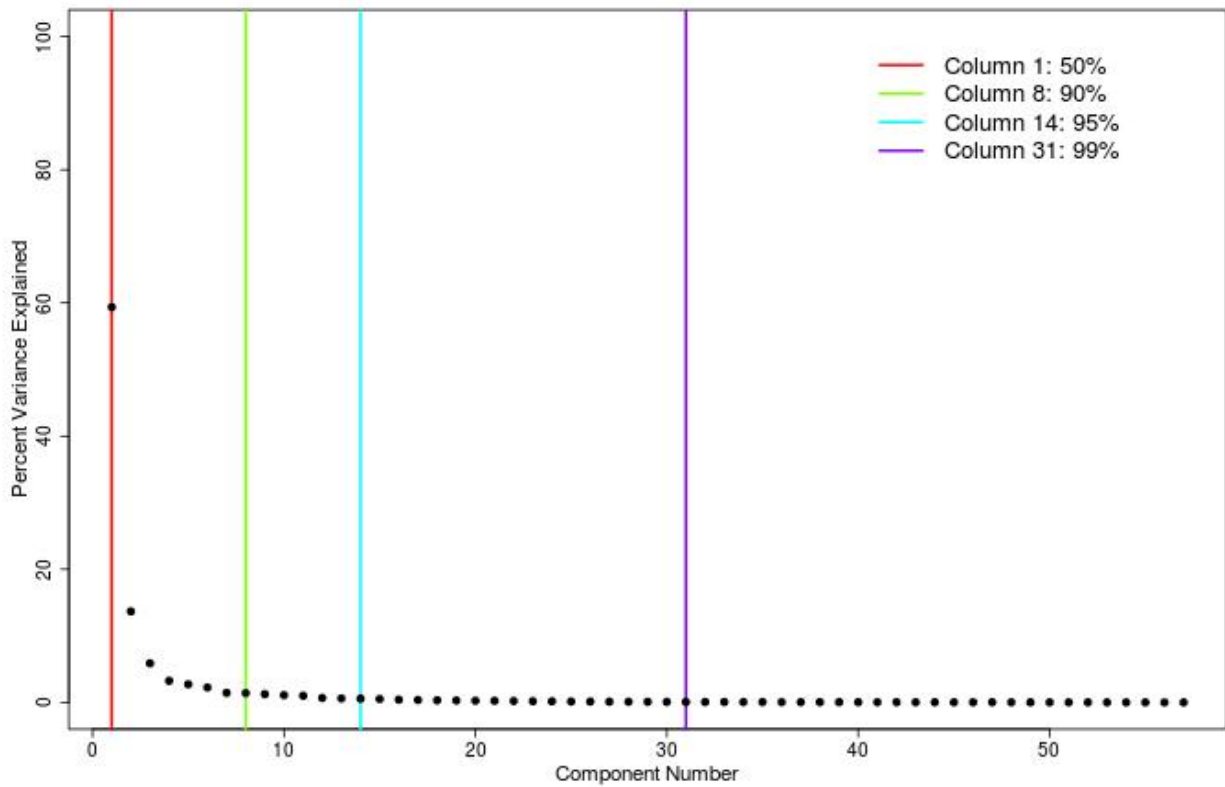


Figure S5a: **Luangwa scree plot**. Scree plot of principal components (PCs) showing which PCs capture 50%, 90%, 95% and 99% of the variance in the NDVI time series. The first 8 PCs explain 90% of variability so are used in the rest of the analysis.

Determining Number of clusters

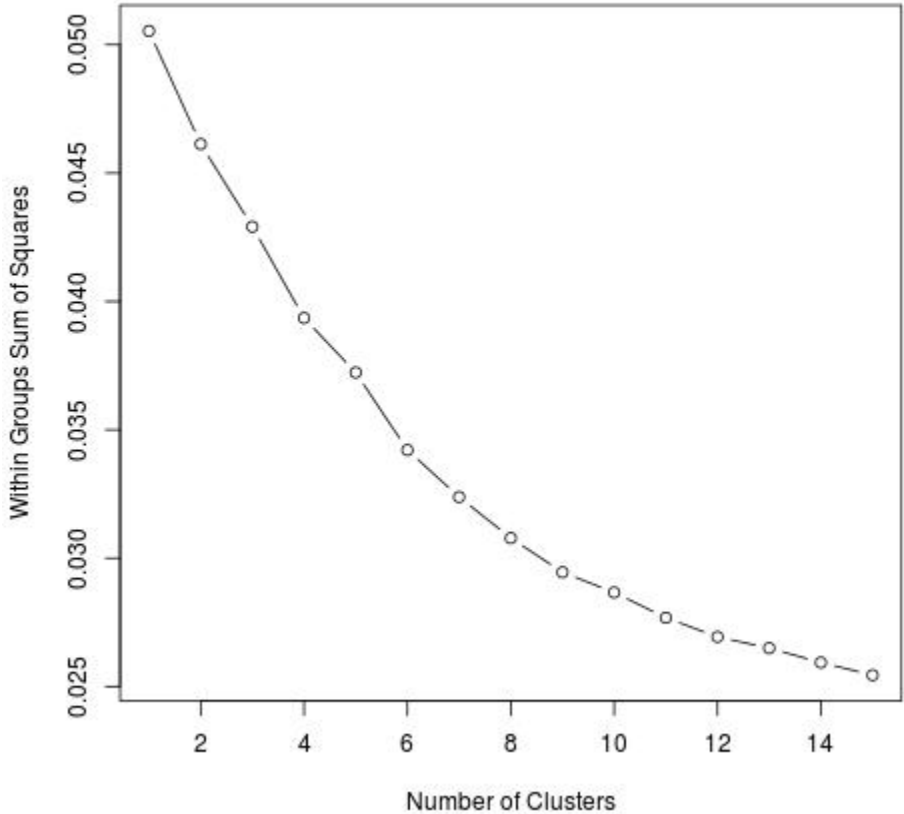


Figure S5b: **Luangwa elbow method**. Within groups sum of squares vs. number of clusters. The inflection point at 10 clusters represents the optimal number of clusters to be obtained.

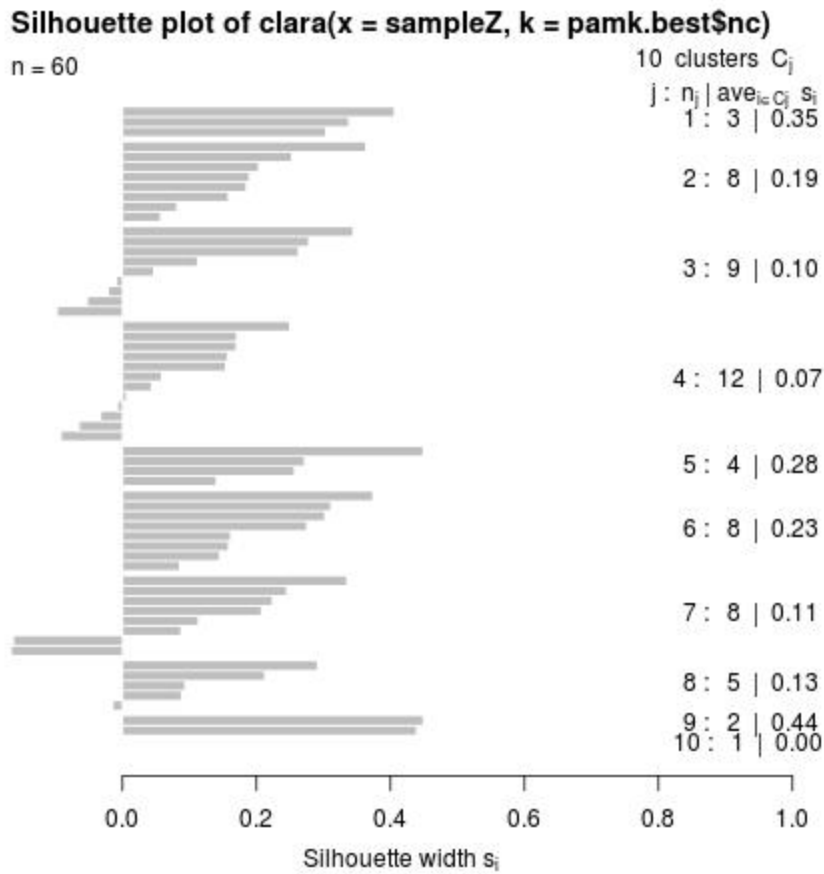


Figure S5c: **Luangwa silhouette method**. There are 10 silhouette widths greater than or equal to 0 which represent the number of optimal clusters to be obtained.

## Phenoregion profiles

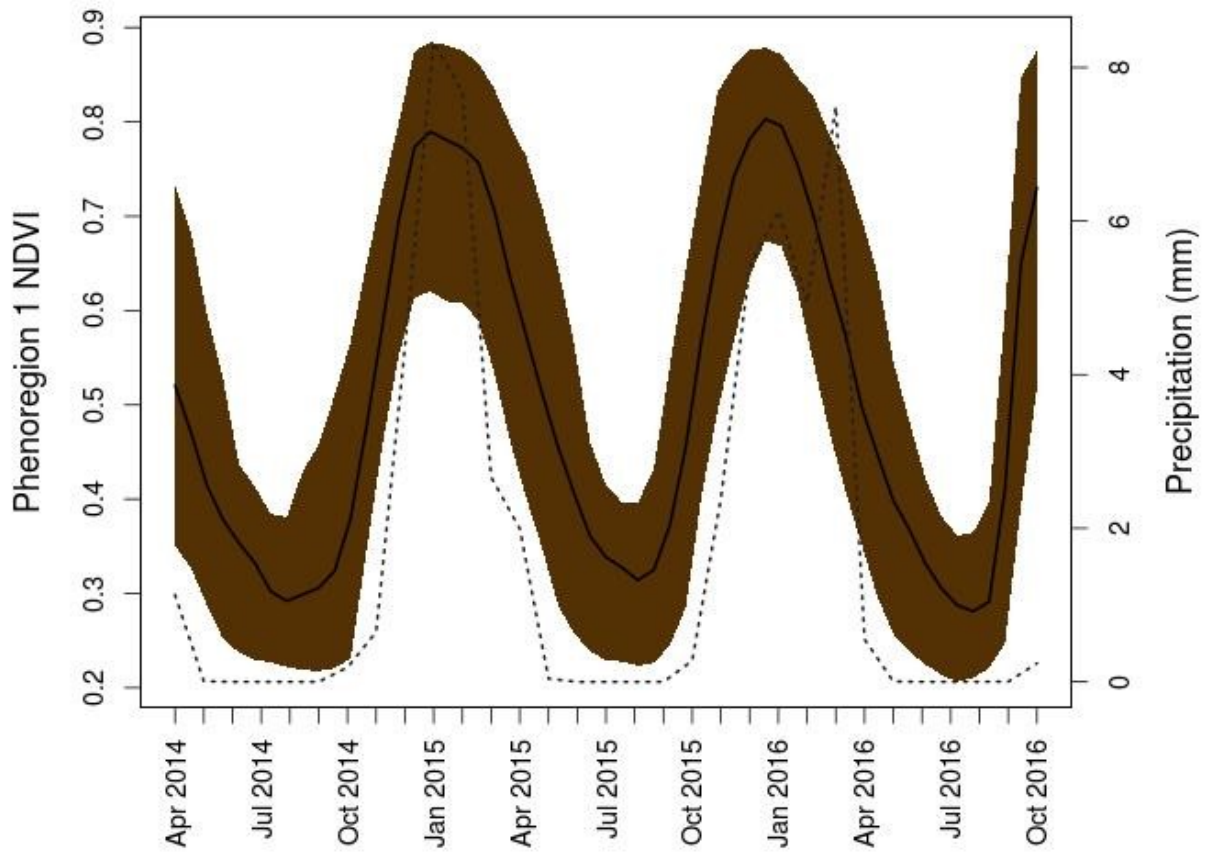


Figure S5d: **Luangwa phenoregion 1 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

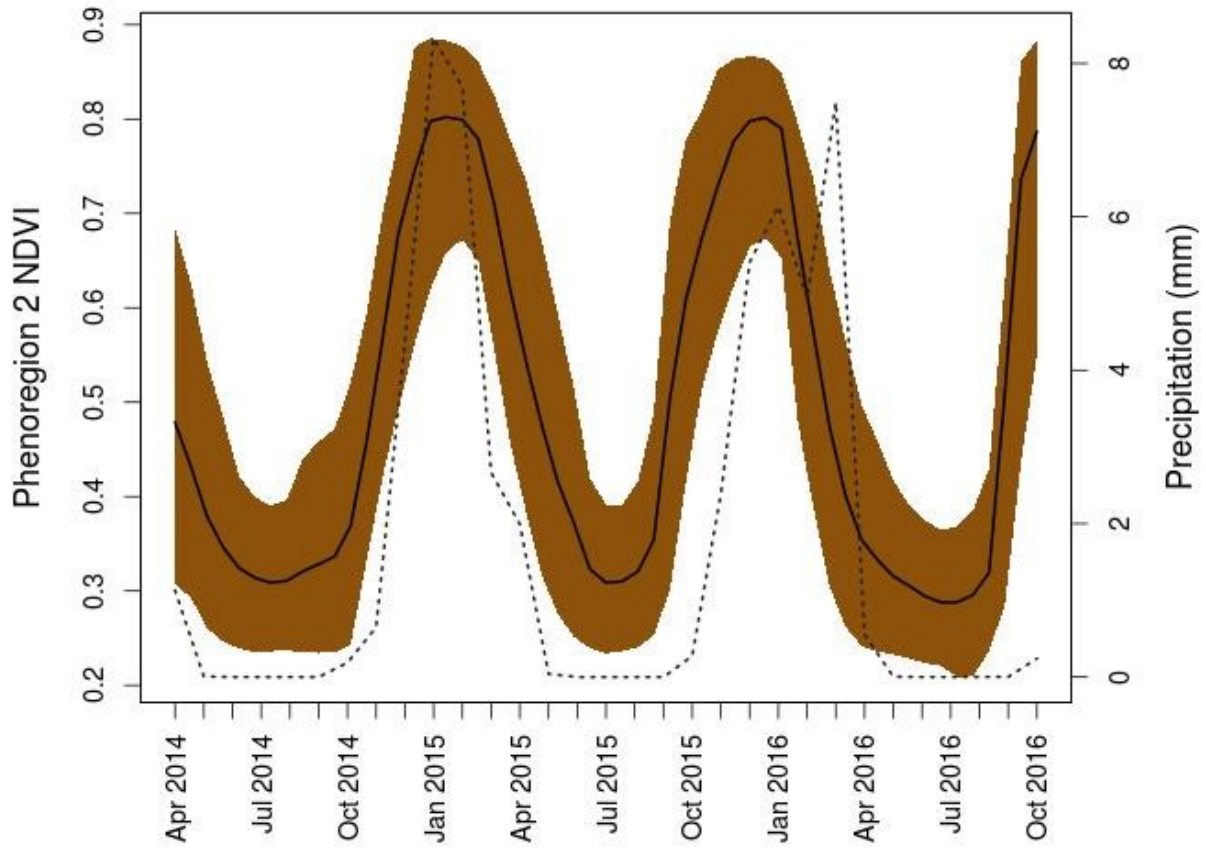


Figure S5e: **Luangwa phenoregion 2 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

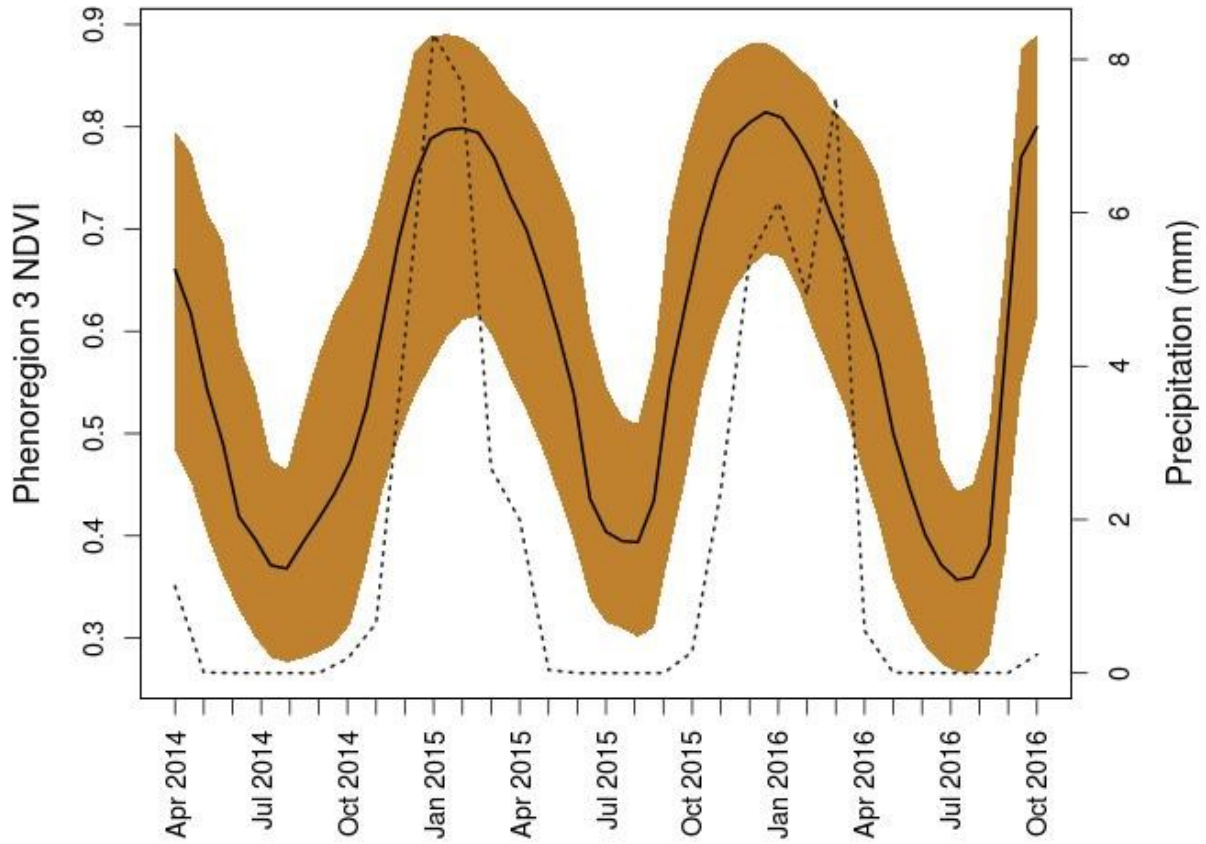


Figure S5f: **Luangwa phenoregion 3 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

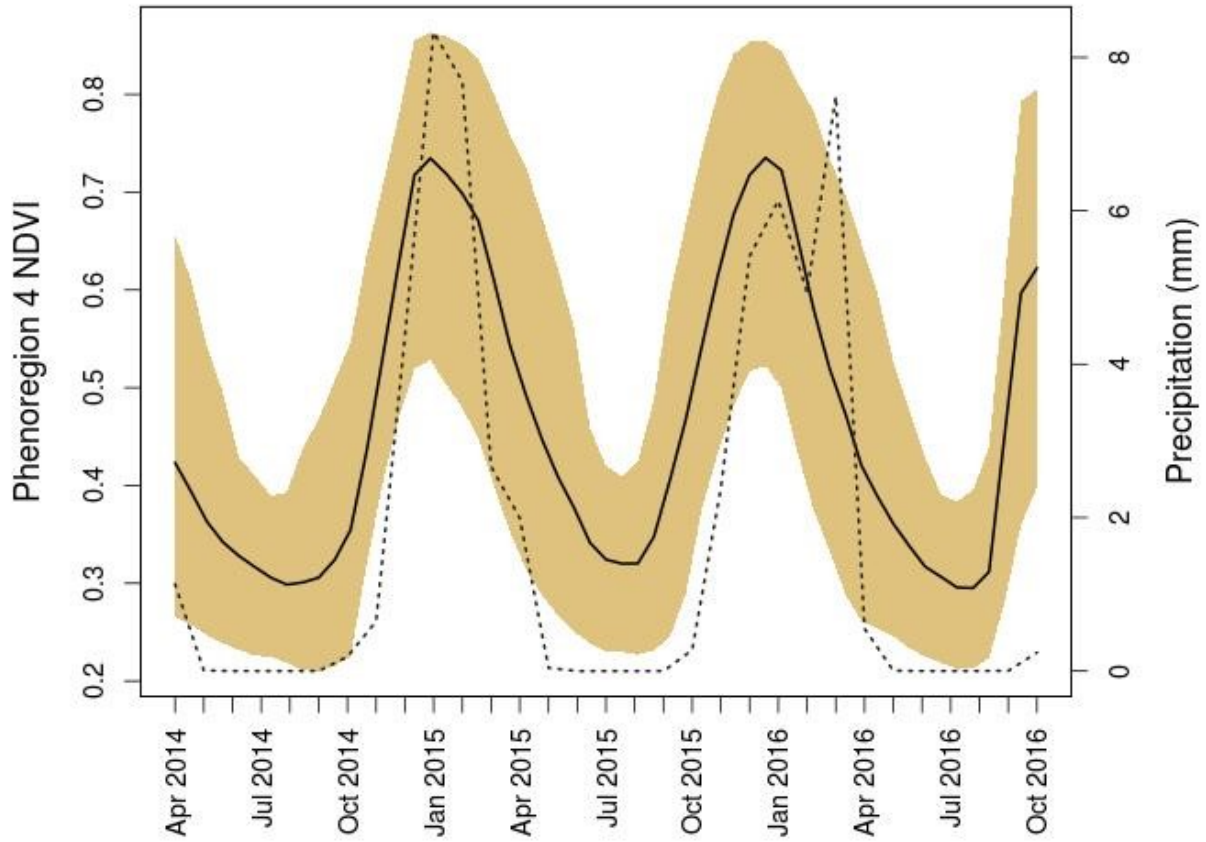


Figure S5g: **Luangwa phenoregion 4 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

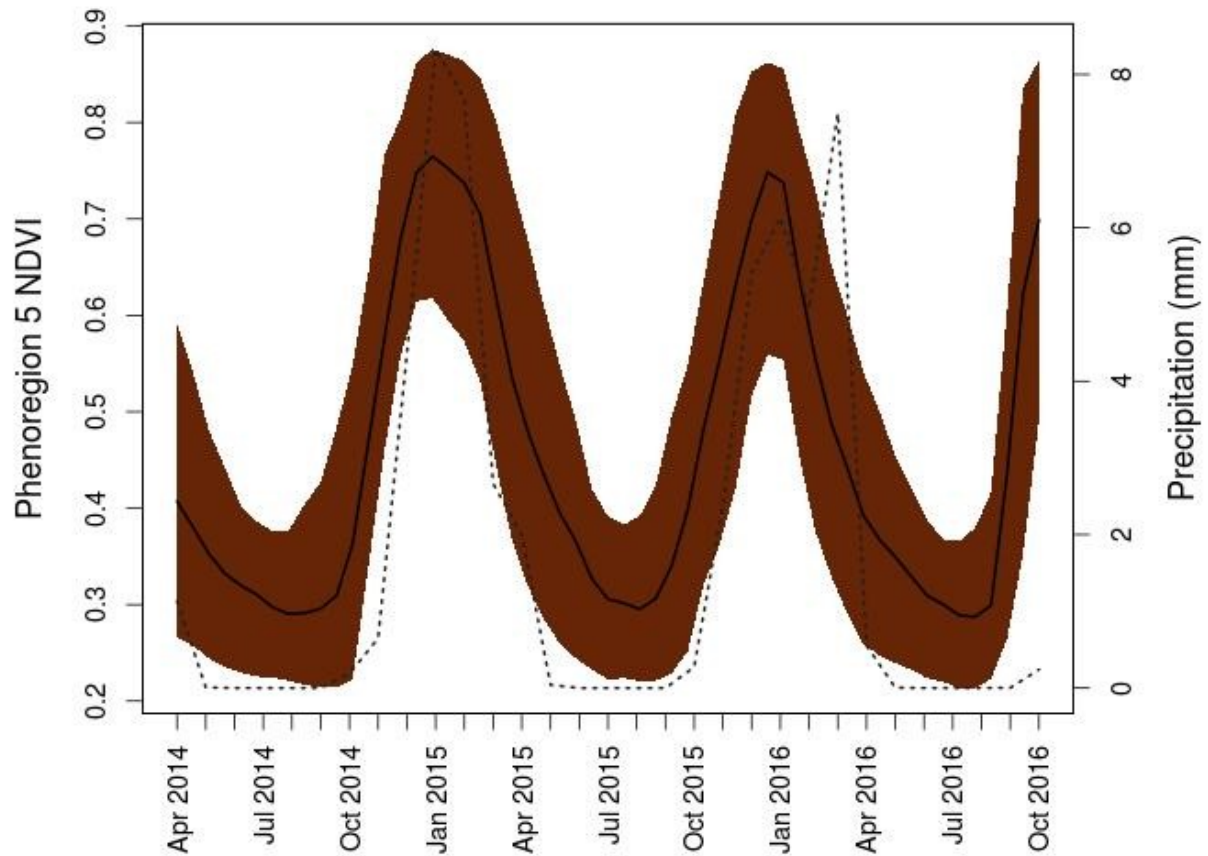


Figure S5h: **Luangwa phenoregion 5 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).



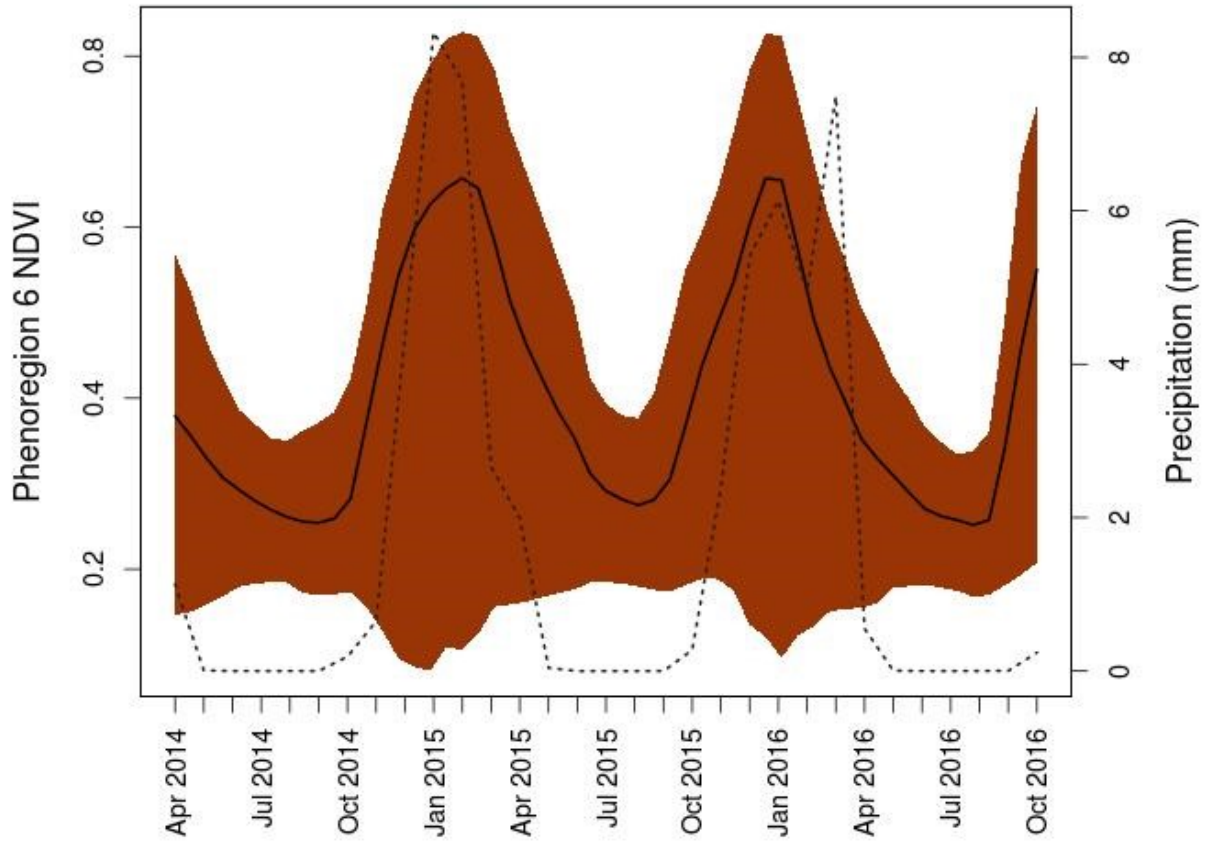


Figure S5i: **Luangwa phenoregion 6 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

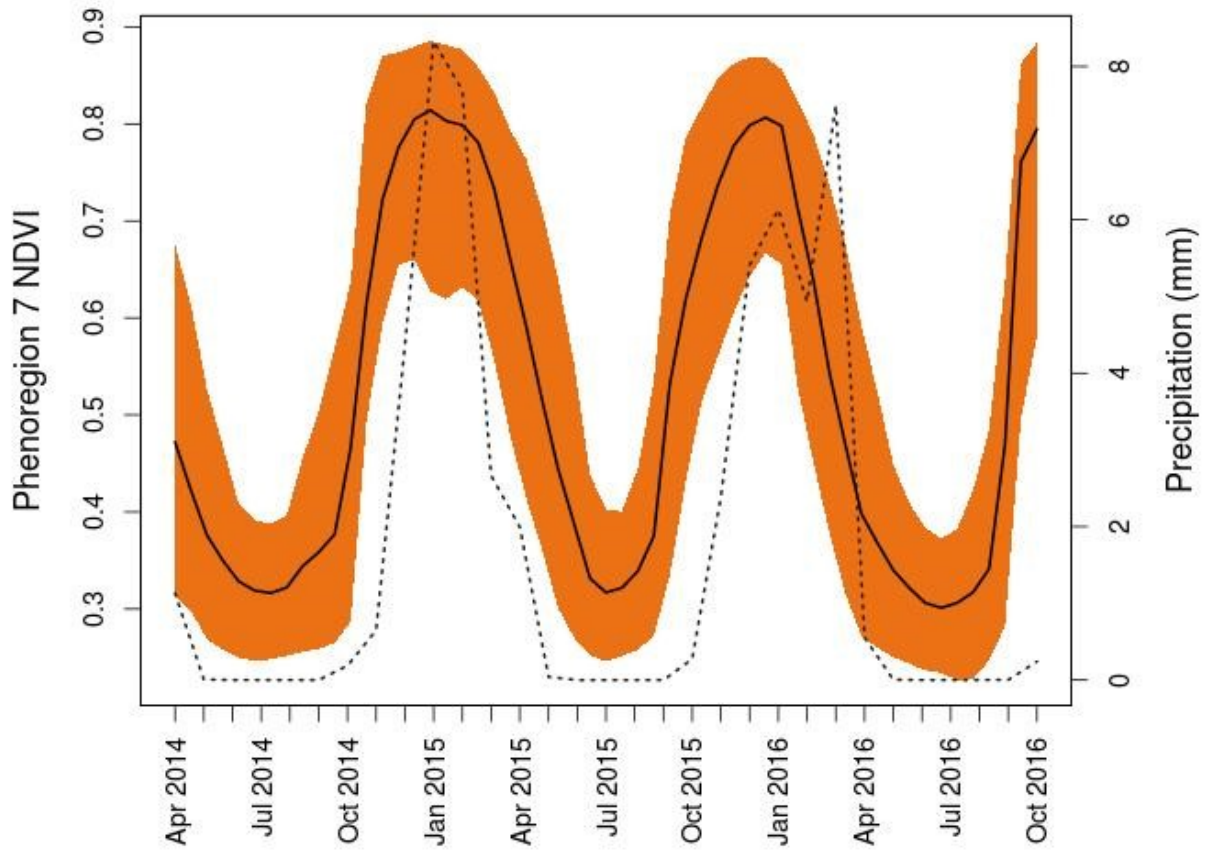


Figure S5j: **Luangwa phenoregion 7 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

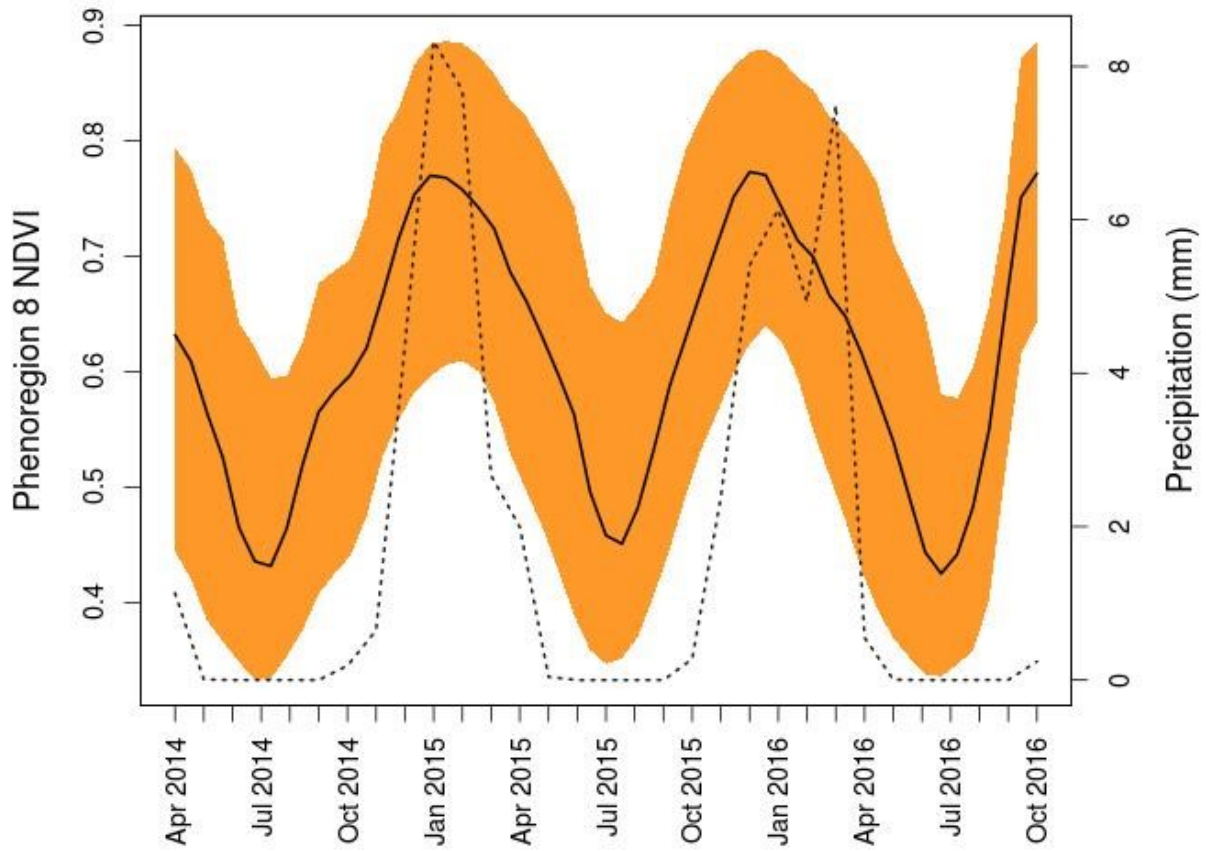


Figure S5k: **Luangwa phenoregion 8 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

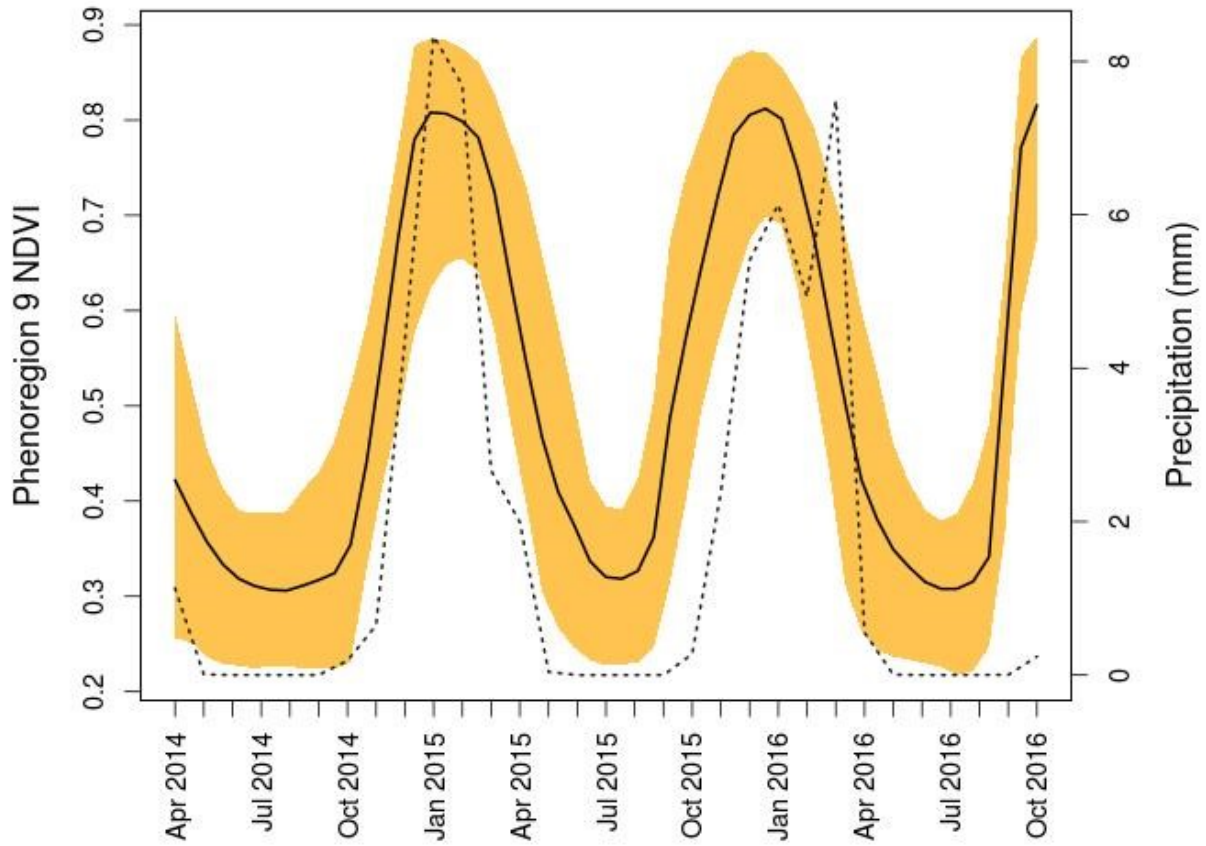


Figure S5m: **Luangwa phenoregion 9 NDVI profile.** Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

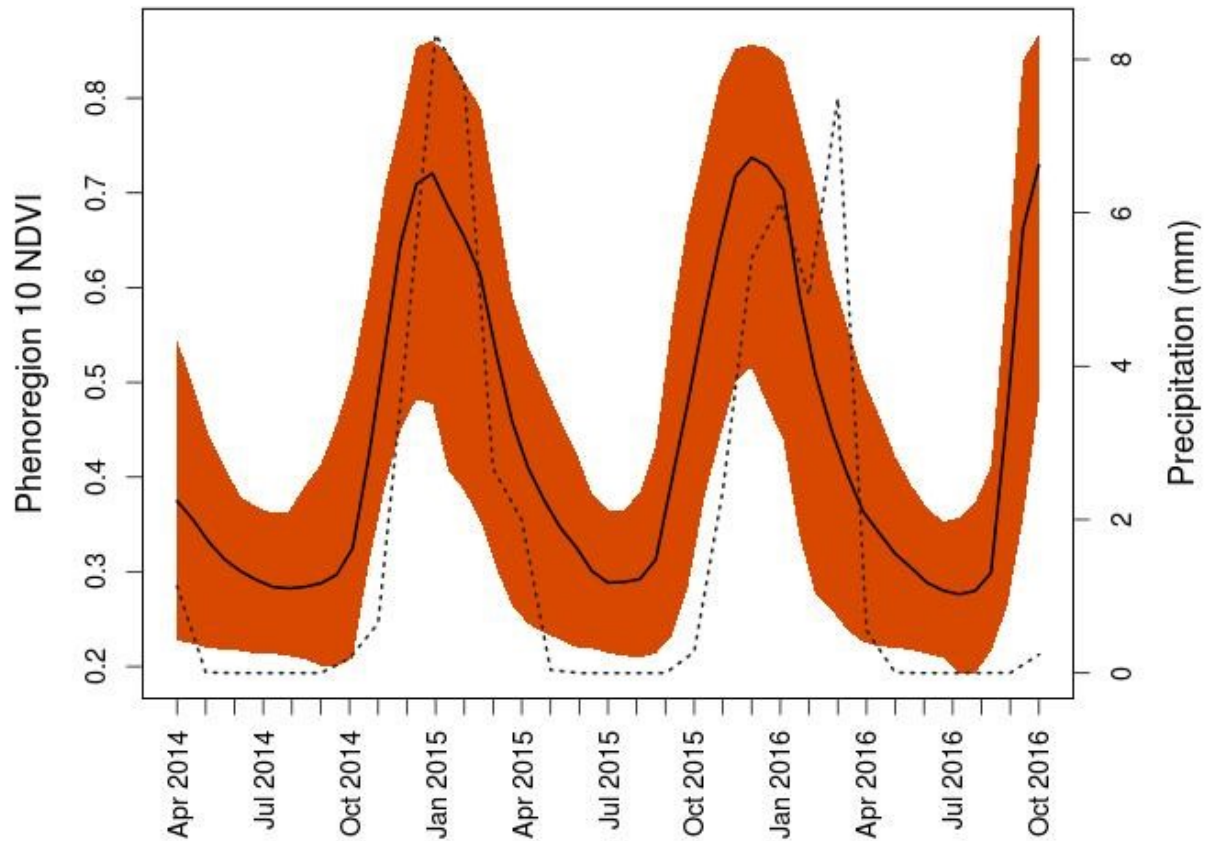


Figure S5n: **Luangwa phenoregion 10 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

# Limpopo

## PCA

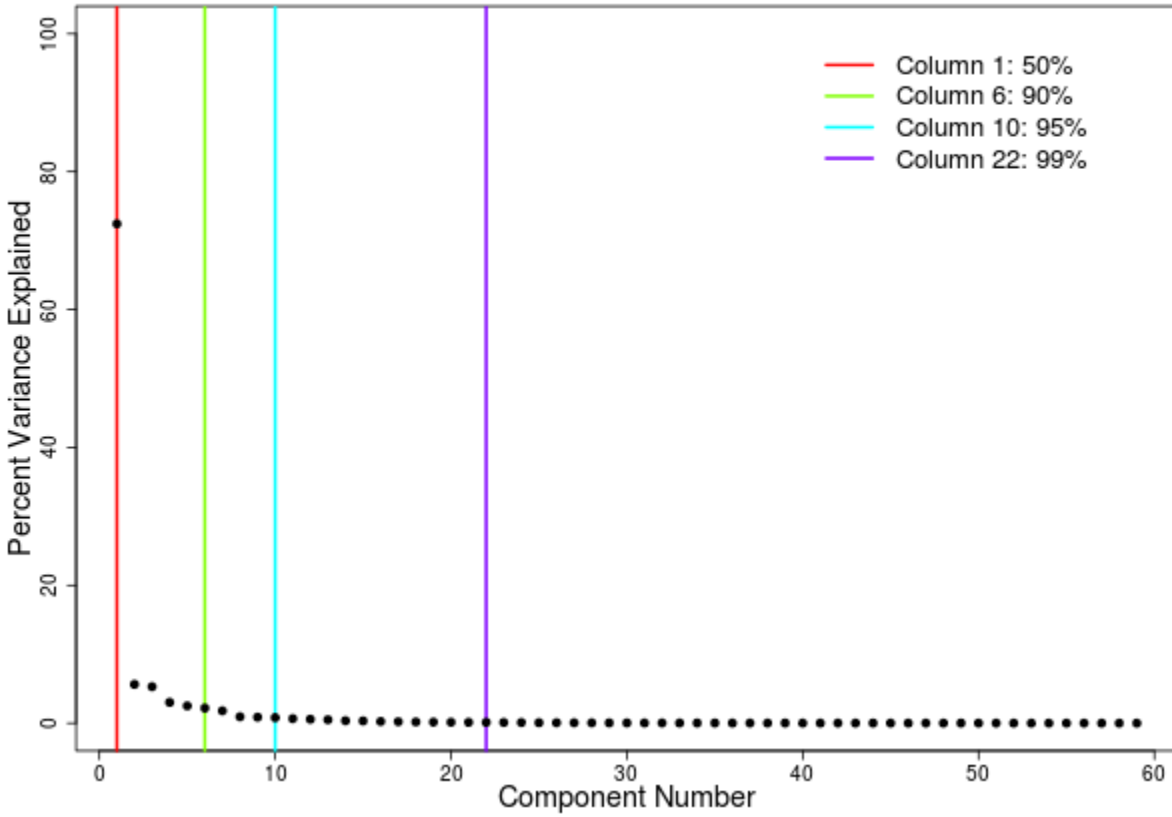


Figure S6a: **Limpopo scree plot**. Scree plot of principal components (PCs) showing which PCs capture 50%, 90%, 95% and 99% of the variance in the Limpopo NDVI time series. The first 6 PCs explain 90% of variability so are used in the rest of the analysis.

Determining number of clusters

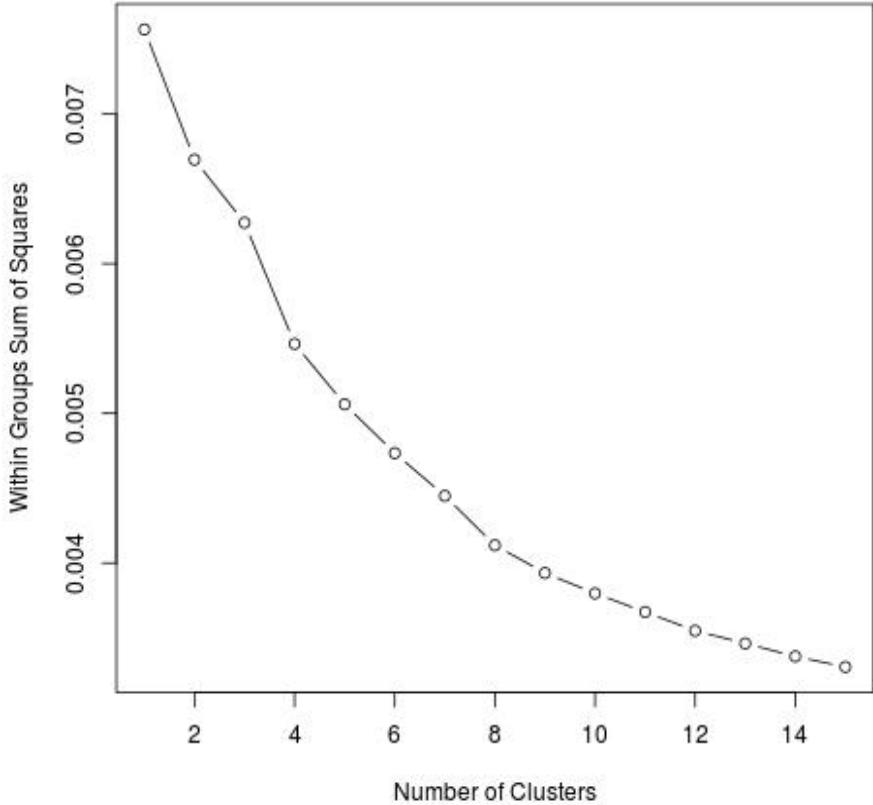


Figure S6b: **Limpopo elbow method**. Within groups sum of squares vs. number of clusters. The inflection point at 4 clusters represents the optimal number of clusters to be obtained.

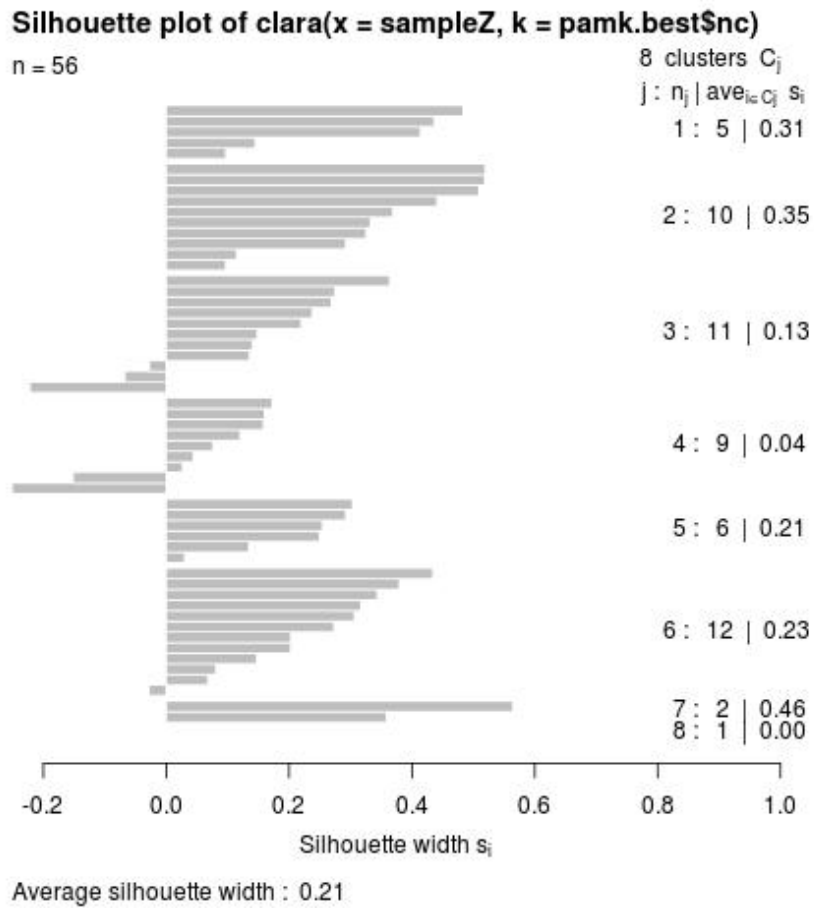


Figure S6c: **Limpopo silhouette method**. There are 8 silhouette widths greater than or equal to 0 which represent the number of optimal clusters to be obtained.



## Phenoregion profiles

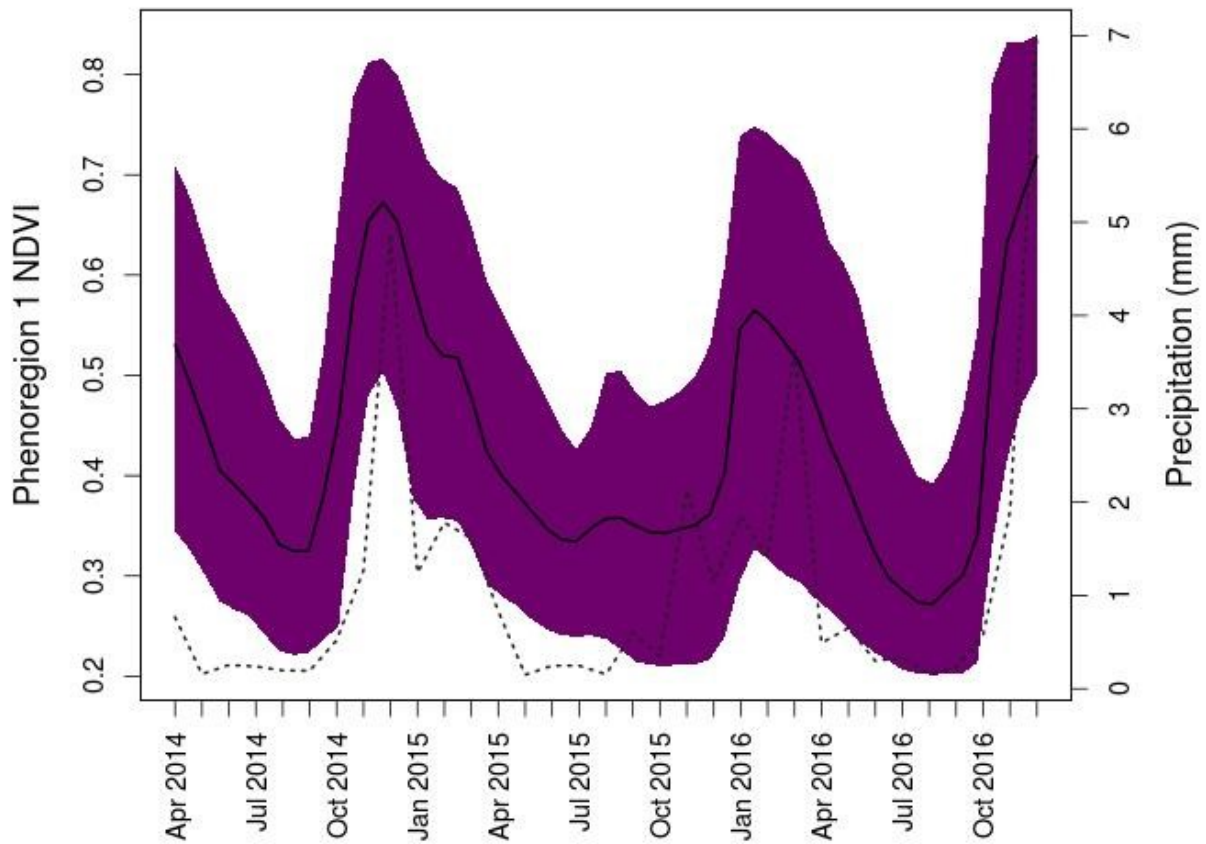


Figure S6d: **Limpopo phenoregion 1 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

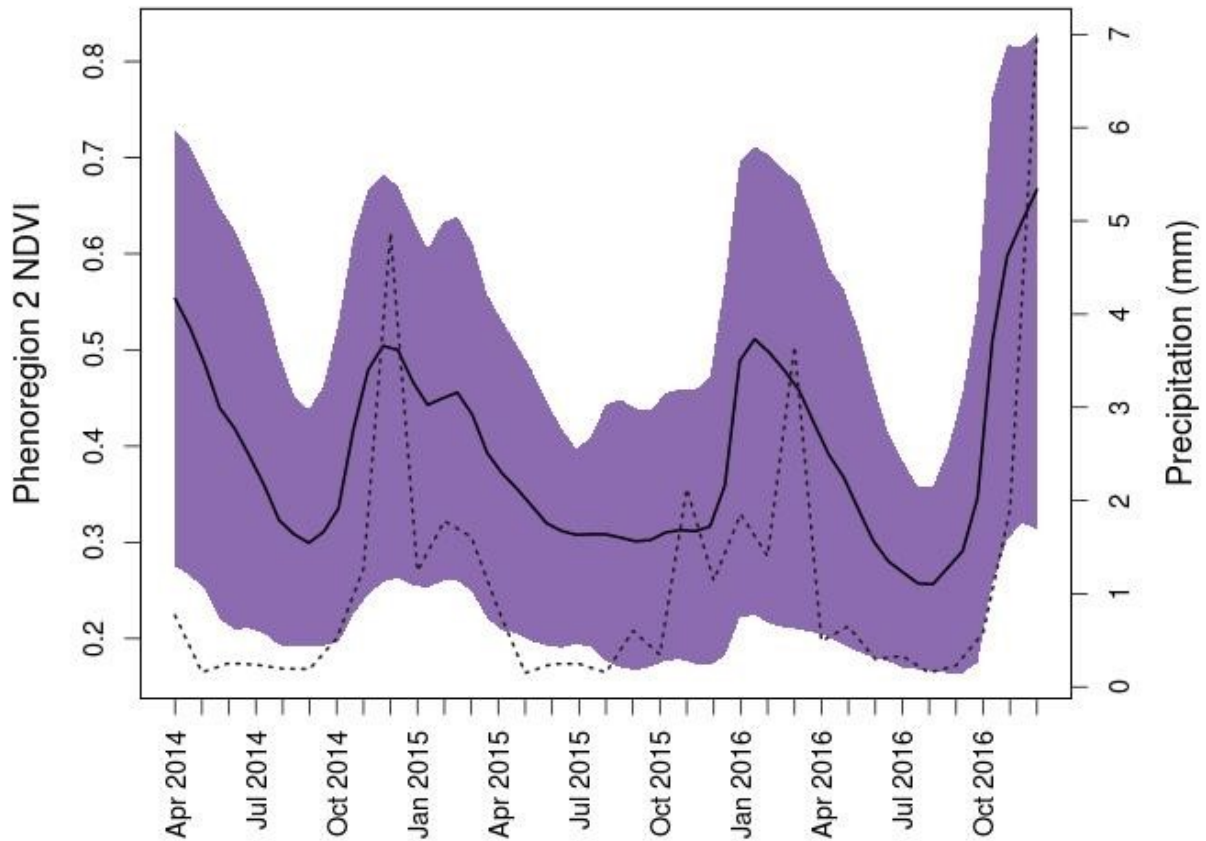


Figure S6e: **Limpopo phenoregion 2 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

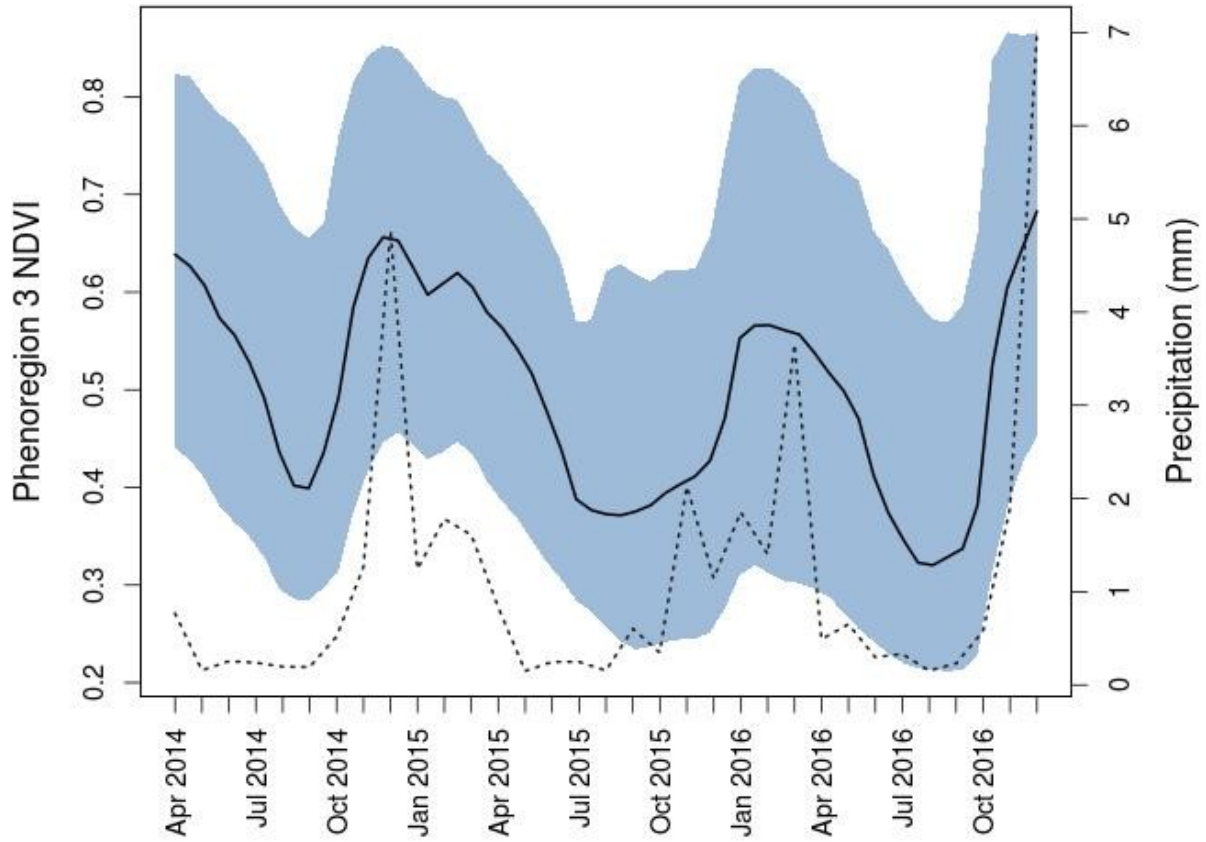


Figure S6f: **Limpopo phenoregion 2 NDVI profile**. Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

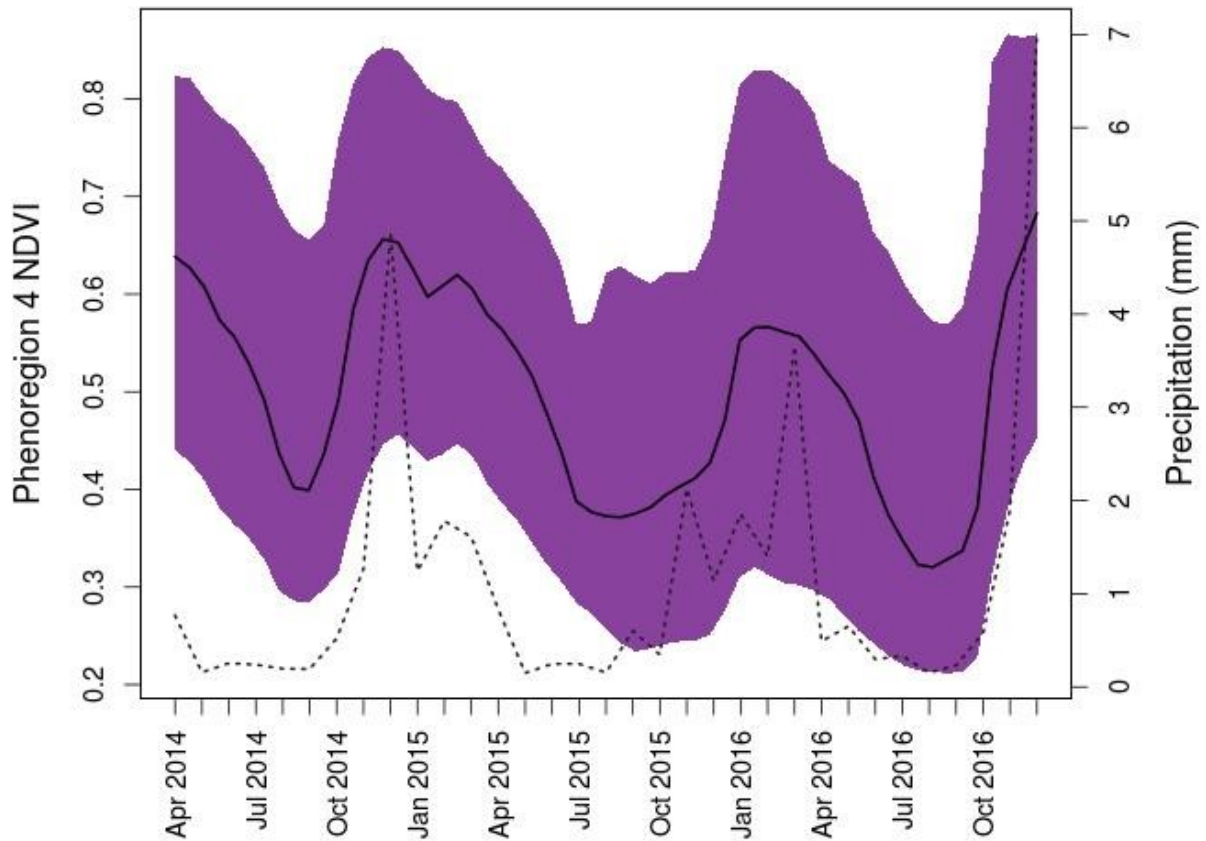


Figure S6g: **Limpopo phenoregion 4 NDVI profile.** Phenoregion NDVI profile across three years. The solid black line is the average NDVI profile for the entire scene. The dotted gray line is monthly precipitation (mm).

## **BIBLIOGRAPHY**

## BIBLIOGRAPHY

- Archibald, S., & Scholes, R. J. (2007). Leaf green-up in a semi-arid African savanna – separating tree and grass responses to environmental cues. *Journal of Vegetation Science*, 18(4), 583–594. <https://doi.org/10.1111/j.1654-1103.2007.tb02572.x>
- Astle, W.L., (1969). The vegetation and soils of Chishinga Ranch, Luapula Province, Zambia. *Kirkia* 7, 73–102.
- Bajocco, S., Dragoz, E., Gitas, I., Smiraglia, D., Salvati, L., & Ricotta, C. (2015). Mapping forest fuels through vegetation phenology: The role of coarse-resolution satellite time-series. *PLoS ONE*, 10(3), 1–14. <https://doi.org/10.1371/journal.pone.0119811>
- Barnes, R. F. W. (1983). The elephant problem in Ruaha National Park, Tanzania. *Biological Conservation*, 26(2), 127–148. [https://doi.org/10.1016/0006-3207\(83\)90062-9](https://doi.org/10.1016/0006-3207(83)90062-9)
- Beck, P.S.A., Atzberger, C., Høgda, K.A., Johansen, B., Skidmore, A.K. (2006). Improved monitoring of vegetation dynamics at very high latitudes: a new method using MODIS NDVI. *Remote Sens. Environ.* 100 (3), 321–334. <http://doi.org/10.1016/j.rse.2005.10.021>
- Bivand, R., Keitt, T., Rowlingson, B. (2018). rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.2-20. <https://CRAN.R-project.org/package=rgdal>
- Bluwstein, J. (2017). Creating ecotourism territories: Environmentalities in Tanzania’s community-based conservation. *Geoforum*, 83, 101–113. <https://doi.org/10.1016/j.geoforum.2017.04.009>
- Bradley, P. S., & Bradley, U. M. (1998). Refining initial points for K-means clustering. *Microsoft Research*, 91–99. <https://doi.org/10.1.1.44.5872>
- Braget, A. (2017). *Time Series Analysis of Phenometrics and Long-Term Vegetation Trends for the Flint Hills Ecoregion using Moderate Resolution Satellite Imagery*. Kansas State University.
- Burkhart, J. J., Peterman, W. E., Brocato, E. R., Romine, K. M., Willis, M. M. S., Ousterhout, B. H., ... Eggert, L. S. (2017). The influence of breeding phenology on the genetic structure of four pond-breeding salamanders. *Ecology and Evolution*, 7(13), 4670–4681. <https://doi.org/10.1002/ece3.3060>
- Byrom, A. E., Nkwabi, A. J. K., Metzger, K., Mduma, S. A. R., Forrester, G. J., Ruscoe, W. A., ... Sinclair, A. R. E. (2015). Anthropogenic stressors influence small mammal communities in tropical East African savanna at multiple spatial scales. *Wildlife Research*, 42(2), 119–131. <https://doi.org/10.1071/WR14223>
- Cambule, A. H., Rossiter, D. G., Stoorvogel, J. J., & Smaling, E. M. A. (2012). Building a near infrared spectral library for soil organic carbon estimation in the Limpopo National Park, Mozambique. *Geoderma*, 183–184, 41–48. <https://doi.org/10.1016/j.geoderma.2012.03.011>

- Cambule, A. H., Rossiter, D. G., Stoorvogel, J. J., & Smaling, E. M. A. (2014). Soil organic carbon stocks in the Limpopo National Park, Mozambique: Amount, spatial distribution and uncertainty. *Geoderma*, 213, 46–56. <https://doi.org/10.1016/j.geoderma.2013.07.015>
- Chabwela, H., Chomba, C., Kaweche, G., & Mwenya, A. (2017). Habitat selection by large mammals in South Luangwa National Park, Zambia. *Open Journal of Ecology*, 07(03), 179–192. <https://doi.org/10.4236/oje.2017.73013>
- 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(23), 5185–5200. <https://doi.org/10.1080/01431160600702392>
- Chambers, L. E., Beaumont, L. J., & Hudson, I. L. (2014). Continental scale analysis of bird migration timing: Influences of climate and life history traits - a generalized mixture model clustering and discriminant approach. *International Journal of Biometeorology*, 58, 1147–1162. <https://doi.org/10.1007/s00484-013-0707-2>
- Chen, X., Hu, B., & Yu, R. (2005). Spatial and temporal variation of phenological growing season and climate change impacts in temperate eastern China. *Global Change Biology*, 11(7), 1118–1130. <https://doi.org/10.1111/j.1365-2486.2005.00974.x>
- Chuine, I. (2000) Scaling phenology from the local to the regional level: advances from species-specific phenological models. *Global Change Biol.* 6, 943–952
- Chuine, I. and Beaubien, E.G. (2001) Phenology is a major determinant of tree species range. *Ecol. Lett.* 4, 500–510
- Chuine, I., Yiou P., Viovy, N. (2004) Historical phenology: grape ripening as a past climate indicator. *Nature* 432, 289–290
- Cleland, E.E., Chiariello, N.R., Loarie, S.R., Mooney, H.A., Field, C.B. (2006). Diverse responses of phenology to global changes in a grassland ecosystem. *Proc. Natl. Acad. Sci. U. S. A.* 103, 13740–13744.
- 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 and Evolution*, 22(7), 357–365. <https://doi.org/10.1016/j.tree.2007.04.003>
- Cleverly, J., Eamus, D., Restrepo Coupe, N., Chen, C., Maes, W., Li, L., ... Huete, A. (2016). Soil moisture controls on phenology and productivity in a semi-arid critical zone. *Science of the Total Environment*, 568, 1227–1237. <https://doi.org/10.1016/j.scitotenv.2016.05.142>
- Corstanje, R., Grafius, D. R., Zawadzka, J., Moreira Barradas, J., Vince, G., Ivanoff, D., & Pietro, K. (2016). A data mining approach to identifying spatial patterns of phosphorus forms in the Stormwater Treatment Areas in the Everglades, US. *Ecological Engineering*, 97, 567–576. <https://doi.org/10.1016/j.ecoleng.2016.10.003>

- Coulston, J. W., Jacobs, D. M., King, C. R., & Elmore, I. C. (2013). The influence of multi-season imagery on models of canopy cover: A case study. *Photogrammetric Engineering & Remote Sensing*, 79(5), 469–477. <https://doi.org/10.14358/PERS.79.5.469>
- Dahlin, K. M., Fisher, R. A., & Lawrence, P. J. (2015). Environmental drivers of drought deciduous phenology in the Community Land Model. *Biogeosciences*, 12(16), 5061–5074. <https://doi.org/10.5194/bg-12-5061-2015>
- Dowle, M., & Srinivasan, A., (2018). data.table: Extension of `data.frame`. R package version 1.11.4. <https://CRAN.R-project.org/package=data.table>
- Dunne, J. A., Harte, J., Taylor, K. J. (2003). Composition and structure flowering phenology responses to climate change: integrating experimental and gradient methods. *Ecological Monographs*, 73(1), 69–86.
- ESRI (2011). ArcGIS Desktop: Release 10.4 Redlands, CA: Environmental Systems Research Institute.
- Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., ... Alsdorf, D.E. (2007). The shuttle radar topography mission: Reviews of Geophysics, v. 45, no. 2, RG2004, at <http://dx.doi.org/10.1029/2005RG000183>
- Filippa, G., Cremonese, E., Migliavacca, M., Galvagno, M., Forkel, M., Wingate, L., ... Richardson, A. D. (2016). Phenopix: A R package for image-based vegetation phenology. *Agricultural and Forest Meteorology*, 220, 141–150. <https://doi.org/10.1016/j.agrformet.2016.01.006>
- Foga, S., Scaramuzza, P.L., Guo, S., Zhu, Z., Dilley, R.D., Beckmann, T., ... Laue, B. (2017). Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sensing of Environment*, 194, 379–390. <http://doi.org/10.1016/j.rse.2017.03.026>
- Forkel, M., Carvalhais, N., Verbesselt, J., Mahecha, M. D., Neigh, C. S. R., & Reichstein, M. (2013). Trend Change detection in NDVI time series: Effects of inter-annual variability and methodology. *Remote Sensing*, 5(5), 2113–2144. <https://doi.org/10.3390/rs5052113>
- Forkel, M., & Wutzler, T. (2015). Greenbrown - land surface phenology and trend analysis. A package for the R software. Version 2.2, 2015-04-15, <http://greenbrown.r-forge.r-project.org/>
- Friedl, M., Sulla-Menashe, D. (2015). *MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006* [Data set]. NASA EOSDIS Land Processes DAAC. doi: 10.5067/MODIS/MCD12Q1.006
- Geerken, R. A. (2009). An algorithm to classify and monitor seasonal variations in vegetation phenologies and their inter-annual change. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, 42–431. <https://doi.org/10.1016/j.isprsjprs.2009.03.001>
- The Geology, Oil and Gas Map showing geology published by the U.S. Geological Survey (2002). Download at [http://www.uni-koeln.de/sfb389/e/e1/e1\\_download\\_e.htm#geology](http://www.uni-koeln.de/sfb389/e/e1/e1_download_e.htm#geology)



- Giglio, L., Randerson, J. T., & Van Der Werf, G. R. (2013). Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). *Journal of Geophysical Research: Biogeosciences*, 118(1), 317–328. <https://doi.org/10.1002/jgrg.20042>.
- Githumbi, E. N., Courtney Mustaphi, C. J., Yun, K. J., Muiruri, V., Rucina, S. M., & Marchant, R. (2018). Late Holocene wetland transgression and 500 years of vegetation and fire variability in the semi-arid Amboseli landscape, southern Kenya. *Ambio*, 47(6), 682–696. <https://doi.org/10.1007/s13280-018-1014-2>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*.
- Grafius, D. R., Corstanje, R., & Harris, J. A. (2018). Linking ecosystem services, urban form and green space configuration using multivariate landscape metric analysis. *Landscape Ecology*, 33, 557–573. <https://doi.org/10.1007/s10980-018-0618-z>.
- Gu, J., Li, X., Huang, C.L., Okin, G.S. (2009). A simplified data assimilation method for reconstructing time-series MODIS NDVI data. *Adv. Space Res.* 44, 501–509.
- Guan, K., Wood, E. F., Medvigy, D., Kimball, J., Pan, M., Caylor, K. K., ... Jones, M. O. (2014). Terrestrial hydrological controls on land surface phenology of African savannas and woodlands. *Journal of Geophysical Research: Biogeosciences*, 119, 1652–1669. <https://doi.org/10.1002/2013JG002572>
- Hennig, C. (2018). fpc: Flexible Procedures for Clustering. R package version 2.1-11. <https://CRAN.R-project.org/package=fpc>
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15), 1965–1978. <https://doi.org/10.1002/joc.1276>
- Hijmans, R. J. (2017). raster: Geographic Data Analysis and Modeling. R package version 2.6-7. <https://CRAN.R-project.org/package=raster>
- Hoagland, S. J., Beier, P., & Lee, D. (2018). Using MODIS NDVI phenoclasses and phenoclusters to characterize wildlife habitat: Mexican spotted owl as a case study. *Forest Ecology and Management*, 412, 80–93. <https://doi.org/10.1016/J.FORECO.2017.12.017>
- Hoffman, F. M., Hargrove, W. W., Mills, R. T., Mahajan, S., Erickson, D. J., & Oglesby, R. J. (2008). Multivariate Spatio-Temporal Clustering (MSTC) as a Data Mining Tool for Environmental Applications. *IEMSs 2008: International Congress on Environmental Modelling and Software: Integrating Sciences and Information Technology for Environmental Assessment and Decision Making*, 1774–1781.
- Hurley, C. (2012). gclus: Clustering Graphics. R package version 1.3.1. <https://CRAN.R-project.org/package=gclus>
- Jakubauskas, Mark E., Legates, David R., Kastens, Jude H. (2003). Harmonic analysis of time-series AVHRR NDVI Data. *Photogrammetric Engineering & Remote Sensing*, 67(4), 461-470.

- Jolliffe, I. T. (2002). Principal Component Analysis, Second Edition. *Encyclopedia of Statistics in Behavioral Science*, 30(3), 487. <https://doi.org/10.2307/1270093>
- Jönsson, P., & Eklundh, L. (2002). Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing*, 40(8), 1824–1832. <https://doi.org/10.1109/TGRS.2002.802519>
- Jönsson, P., & Eklundh, L. (2004). TIMESAT - A program for analyzing time-series of satellite sensor data. *Computers and Geosciences*, 30(8), 833–845. <https://doi.org/10.1016/j.cageo.2004.05.006>
- Kangalawe, R., Mwakalila, S., & Masolwa, P. (2011). Climate change impacts, local knowledge and coping strategies in the Great Ruaha River Catchment Area, Tanzania. *Natural Resources*, 2(4), 212–223. <https://doi.org/10.4236/nr.2011.24027>
- Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., ... Vertenstein, M. (2015). The community earth system model (CESM) large ensemble project : A community resource for studying climate change in the presence of internal climate variability. *American Meteorological Society*, 96(8), 1333–1349. <https://doi.org/10.1175/BAMS-D-13-00255.1>
- Keenan, T. F., Gray, J., Friedl, M. A., Toomey, M., Bohrer, G., Hollinger, D. Y., ... Richardson, A. D. (2014). Net carbon uptake has increased through warming-induced changes in temperate forest phenology. *Nature Climate Change*, 4(June), 598–604. <https://doi.org/10.1038/NCLIMATE2253>
- Kröber, W., Böhnke, M., Welk, E., Wirth, C., & Bruelheide, H. (2012). Leaf trait-environment relationships in a subtropical broadleaved forest in South-East China. *PLoS ONE*, 7(4). <https://doi.org/10.1371/journal.pone.0035742>
- Kumar, J., Mills, R. T., Hoffman, F. M., & Hargrove, W. W. (2011). Parallel k-means clustering for quantitative ecoregion delineation using large data sets. In *Procedia Computer Science*. <https://doi.org/10.1016/j.procs.2011.04.173>
- Liu, R., Shang, R., Liu, Y., & Lu, X. (2017). Global evaluation of gap-filling approaches for seasonal NDVI with considering vegetation growth trajectory, protection of key point, noise resistance and curve stability. *Remote Sensing of Environment*, 189, 164–179. <https://doi.org/10.1016/j.rse.2016.11.023>
- Ma, X., Huete, A., Yu, Q., Coupe, N. R., Davies, K., Broich, M., ... Eamus, D. (2013). Spatial patterns and temporal dynamics in savanna vegetation phenology across the North Australian Tropical Transect. *Remote Sensing of Environment*, 139, 97–115. <https://doi.org/10.1016/j.rse.2013.07.030>
- MacGarigal, K., Cushman, S., & Susan Stafford. (2000). *Multivariate Statistics for Wildlife and Ecology Research*. <https://doi.org/10.1007/978-1-4612-1288-1>
- Marthews, T. R., Dadson, S. J., Lehner, B., Abele, S., & Gedney, N. (2015). High-resolution global topographic index values for use in large-scale hydrological modelling. *Hydrology and Earth System Sciences*, 19(1), 91–104. <https://doi.org/10.5194/hess-19-91-2015>

- Maxwell, R. M. (2010). Infiltration in arid environments: Spatial patterns between subsurface heterogeneity and water-energy balances. *Vadose Zone Journal*, 9(4), 970–983. <https://doi.org/10.2136/vzj2010.0014>
- Mbungu, W. B., & Kashaigili, J. J. (2017). Assessing the hydrology of a data-scarce tropical watershed using the soil and water assessment tool: Case of the Little Ruaha River Watershed in Iringa, Tanzania. *Open Journal of Modern Hydrology*, 07(02), 65–89. <https://doi.org/10.4236/ojmh.2017.72004>
- Mills, R. T., Hoffman, F. M., Kumar, J., & Hargrove, W. W. (2011). Cluster analysis-based approaches for geospatiotemporal data mining of massive data sets for identification of forest threats. *Procedia Computer Science*, 4, 1612–1621. <https://doi.org/10.1016/j.procs.2011.04.174>
- Morrison, T. A., Holdo, R. M., & Anderson, T. M. (2016). Elephant damage, not fire or rainfall, explains mortality of overstorey trees in Serengeti. *Journal of Ecology*, 104(2), 409–418. <https://doi.org/10.1111/1365-2745.12517>
- Mosleh, Z., Salehi, M. H., Jafari, A., Borujeni, I. E., & Mehnatkesh, A. (2016). The effectiveness of digital soil mapping to predict soil properties over low-relief areas. *Environmental Monitoring and Assessment*. <https://doi.org/10.1007/s10661-016-5204-8>
- Muvengwi, J., Davies, A. B., Parrini, F., & Witkowski, E. T. F. (2018). Geology drives the spatial patterning and structure of termite mounds in an African savanna. *Ecosphere*, 9(3). <https://doi.org/10.1002/ecs2.2148>
- Notter, B., MacMillan, L., Viviroli, D., Weingartner, R., & Liniger, H. P. (2007). Impacts of environmental change on water resources in the Mt. Kenya region. *Journal of Hydrology*, 343(3–4), 266–278. <https://doi.org/10.1016/j.jhydrol.2007.06.022>
- Ogutu, J. O., Piepho, H. P., Dublin, H. T., Bhola, N., & Reid, R. S. (2008). Rainfall influences on ungulate population abundance in the Mara-Serengeti ecosystem. *Journal of Animal Ecology*, 77(4), 814–829. <https://doi.org/10.1111/j.1365-2656.2008.01392.x>
- Parmesan, C. (2007). Influences of species, latitudes and methodologies on estimates of phenological response to global warming. *Global Change Biology*, 13(9), 1860–1872. <https://doi.org/10.1111/j.1365-2486.2007.01404.x>
- Parveen, R., Kulkarni, S., & Mytri, V. D. (2016). Vegetation extraction from multi sensor IRS-1C LISS III image by Principle Component Analysis and comparative assessment. In *Proceeding of IEEE - 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics, IEEE - AEEICB 2016*. <https://doi.org/10.1109/AEEICB.2016.7538310>
- Pascucci, S., Francesca Carfora, M., Palombo, A., Pignatti, S., Casa, R., Pepe, M., & Castaldi, F. (2018). A Comparison between standard and functional clustering methodologies: Application to agricultural fields for yield pattern assessment. *Remote Sensing*, 10(585), 1–21. <https://doi.org/10.3390/rs10040585>

- Pebesma, E. J., Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News* 5 (2), <https://cran.r-project.org/doc/Rnews/>
- Penuelas, J., Filella, I., Zhang, X., Llorens, L., Ogaya, R., Lloret, F., ... Terradas, J. (2004). Complex spatiotemporal phenological shifts as a response to rainfall changes. *New Phytologist*, 161(3), 837–846. <https://doi.org/10.1111/j.1469-8137.2003.01003.x>
- Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., ... Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, 344(6187). <https://doi.org/10.1126/science.1246752>
- Potter, S. F., Dawson, E. J., & Frierson, D. M. W. (2017). Southern African orography impacts on low clouds and the Atlantic ITCZ in a coupled model. *Geophysical Research Letters*, 44(7), 3283–3289. <https://doi.org/10.1002/2017GL073098>
- Poulter, B., MacBean, N., Hartley, A., Khlystova, I., Arino, O., Betts, R., ... Peylin, P. (2015). Plant functional type classification for earth system models: Results from the European Space Agency's Land Cover Climate Change Initiative. *Geoscientific Model Development*, 8(7), 2315–2328. <https://doi.org/10.5194/gmd-8-2315-2015>
- QGIS Development Team (2017). QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>
- Reason, C. J. C., Hachigonta, S., & Phaladi, R. F. (2005). Interannual variability in rainy season characteristics over the Limpopo region of southern Africa. *International Journal of Climatology*, 25(14), 1835–1853. <https://doi.org/10.1002/joc.1228>
- Recha, J.W., Lehmann, J., Walter, M.T., Pell, A., Verchot, L. and Johnson, M. (2012). Stream Discharge in Tropical Headwater Catchments as a Result of Forest Clearing and Soil Degradation. *Earth Interactions*, 16, 1-18. <https://doi.org/10.1175/2012EI000439.1>
- Reid, R. S., Nkedianye, D., Said, M. Y., Kaelo, D., Neselle, M., Makui, O., ... Clark, W. C. (2016). Evolution of models to support community and policy action with science: Balancing pastoral livelihoods and wildlife conservation in savannas of East Africa. *Proceedings of the National Academy of Sciences*, 113(17), 4579–4584. <https://doi.org/10.1073/pnas.0900313106>
- Rial, M., Martínez Cortizas, A., & Rodríguez-Lado, L. (2017). Understanding the spatial distribution of factors controlling topsoil organic carbon content in European soils. *Science of The Total Environment*, 609, 1411–1422. <https://doi.org/10.1016/J.SCITOTENV.2017.08.012>
- Rodríguez-Moreno, V. M., Kretzschmar, T. G., & Padilla-Ramírez, J. S. (2015). The geospatial relationship of geologic strata, geological fractures, and land use attained by a time-series aridity index in a

semiarid region. *Environmental Monitoring and Assessment*, 187(457), 1–14.  
<https://doi.org/10.1007/s10661-015-4676-2>

Roger S. Bivand, Edzer Pebesma, Virgilio Gomez-Rubio, (2013). Applied spatial data analysis with R, Second edition. Springer, NY. <http://www.asdar-book.org/>

Scharsich, V., Mtata, K., Hauhs, M., Lange, H., & Bogner, C. (2017). Analysing land cover and land use change in the Matobo National Park and surroundings in Zimbabwe. *Remote Sensing of Environment*, 194, 278–286. <https://doi.org/10.1016/J.RSE.2017.03.037>

Scrivani, R., Gonclaves, R. R. V, Zullo Jr., J., & Romani, L. A. S. (2015). Agricultural monitoring assessment based on low spatial and high temporal resolution satellite: A comparison of AVHRR and MODIS. *Geoscience and Remote Sensing (IGARSS), IEEE International Symposium*, 3405–3408.

Senthilnath, J., Kandukuri, M., Dokania, A., & Ramesh, K. N. (2017). Application of UAV imaging platform for vegetation analysis based on spectral-spatial methods. *Computers and Electronics in Agriculture*, 140, 8–24. <https://doi.org/10.1016/j.compag.2017.05.027>

Stalmans, M., Gertenbach, W. P. D., & Carvalho-Serfontein, F. (2004). Plant communities and landscapes of the Parque Nacional de Zinave, Mozambique. *Koedoe*, 47(2), 61–81.  
<https://doi.org/10.4102/koedoe.v52i1.703>

Subbalakshmi, C., Rama Krishna, G., Krishna Mohan Rao, S., & Venketeswa Rao, P. (2015). A method to find optimum number of clusters based on fuzzy silhouette on dynamic data set. *Procedia Computer Science*, 46, 346–353. <https://doi.org/10.1016/j.procs.2015.02.030>

Tucker, C. J., Pinzon, J. E., Brown, M. E., Slayback, D. A., Pak, E. W., Mahoney, R., ... El Saleous, N. (2005). An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing*, 26(20), 4485–4498.  
<https://doi.org/10.1080/01431160500168686>

Tzortzis, G., & Likas, A. (2014). The MinMax k-Means clustering algorithm. *Pattern Recognition*, 47(7), 2505–2516. <https://doi.org/10.1016/J.PATCOG.2014.01.015>

van Vegten, J. A., & Straatweg, K. (1983). Thornbush invasion in a savanna ecosystem in eastern Botswana. *Vegetatio*, 56, 3–7.

Venables, W. N., & Ripley, B. D. (2002). Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Verbesselt, J., Hyndman, R., Newnham, G., & Culvenor, D. (2010). Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment*, 114(1), 106–115.  
<https://doi.org/10.1016/j.rse.2009.08.014>

Verbesselt, J., Hyndman, R., Zeileis, A., & Culvenor, D. (2010). Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. *Remote Sensing of Environment*, 114(12), 2970–2980. <https://doi.org/10.1016/j.rse.2010.08.003>

- Walker, J., & Gillison, A. N. (1982). Australian savannas. In B. J. Huntley, & B. H. Walker (Eds.), *Ecological studies* (pp. 5–24). Springer Berlin Heidelberg.
- Wang, Y., Tang, X., Yu, L., Hou, X., & Munger, J. W. (2016). Comparison of net ecosystem carbon exchange estimation in a mixed temperate forest using field eddy covariance and MODIS data. *SpringerPlus*, 5(491), 1–11. <https://doi.org/10.1186/s40064-016-2134-4>
- Wang, Q., Wang, Y., Niu, R., & Peng, L. (2017). Integration of Information Theory, K-Means Cluster Analysis and the Logistic Regression Model for landslide susceptibility mapping in the Three Gorges Area, China. *Remote Sensing*, 9(9), 1–28. <https://doi.org/10.3390/rs9090938>.
- Wang, S., Dong, P., & Tian, Y. (2017). A novel method of statistical line loss estimation for distribution feeders based on feeder cluster and modified XGBoost. *Energies*, 10(12), 1–17. <https://doi.org/10.3390/en10122067>
- Weihs, C., Ligges, U., Luebke, K., Raabe, N. (2005). klaR Analyzing German Business Cycles. In Baier, D., Decker, R. and Schmidt-Thieme, L. (eds.). *Data Analysis and Decision Support*, 335-343, Springer-Verlag, Berlin.
- White, M.A., Thornton, P.E. & Running, S.W. (1997). A continental phenology model for monitoring vegetation responses to interannual climatic variability. *Global Biogeochem. Cycles* 11: 217-234.
- White, M. A., de Beurs, K. M., Didan, K., Inouye, D. W., Richardson, A. D., Jensen, O. P., ... Lauenroth, W. K. (2009). Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982-2006. *Global Change Biology*, 15(10), 2335–2359. <https://doi.org/10.1111/j.1365-2486.2009.01910.x>
- White, M. A., Hoffman, F., Hargrove, W. W., & Nemani, R. R. (2005). A global framework for monitoring phenological responses to climate change. *Geophysical Research Letters*, 32(4), 1–4. <https://doi.org/10.1029/2004GL021961>
- Wickham, H., Francois, R., Henry, L., Müller, K. (2017). dplyr: A Grammar of Data Manipulation. R package version 0.7.4. <https://CRAN.R-project.org/package=dplyr>
- Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., & Woodcock, C. E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment*, 122, 2–10. <https://doi.org/10.1016/j.rse.2012.01.010>
- Xie, Y. (2014). knitr: A Comprehensive Tool for Reproducible Research in R. In Victoria Stodden, Friedrich Leisch and Roger D. Peng, editors, *Implementing Reproducible Computational Research*. Chapman and Hall/CRC. ISBN 978-1466561595
- Xie, Y. (2015). *Dynamic Documents with R and knitr*. 2nd edition. Chapman and Hall/CRC. ISBN 978-1498716963
- Xie, Y. (2018). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.20.

- Yang, Y. C. E., & Wi, S. (2018). Informing regional water-energy-food nexus with system analysis and interactive visualization – A case study in the Great Ruaha River of Tanzania. *Agricultural Water Management*, 196, 75–86. <https://doi.org/10.1016/j.agwat.2017.10.022>
- Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., ... Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, 84(3), 471–475. [https://doi.org/10.1016/S0034-4257\(02\)00135-9](https://doi.org/10.1016/S0034-4257(02)00135-9)
- Zhou, J., Jia, L., & Menenti, M. (2015). Reconstruction of global MODIS NDVI time series: Performance of Harmonic ANalysis of Time Series (HANTS). *Remote Sensing of Environment*, 163, 217–228. <https://doi.org/10.1016/j.rse.2015.03.018>
- Ziegler, M., Simon, M. H., Hall, I. R., Barker, S., Stringer, C., & Zahn, R. (2013). Development of Middle Stone Age innovation linked to rapid climate change. *Nature Communications*, 4, 1905–1909. <https://doi.org/10.1038/ncomms2897>