

1 Title: Phenology and diversity in Zambia

2 Authors: Godlee, J. L.<sup>1</sup>,

3 <sup>1</sup>: School of GeoSciences, University of Edinburgh, Edinburgh, United Kingdom

4 Corresponding author:

5 John L. Godlee

6 johngodlee@gmail.com

7 School of GeoSciences, University of Edinburgh, Edinburgh, United Kingdom

8 **Acknowledgements**

9 **Author contribution statement**

10 **Data accessibility statement**

# Abstract

## 1 Introduction

The seasonal timing of tree leaf production in dry deciduous savannas directly influences ecosystem processes and structure (). Leaf Area Index (LAI), leaf area per unit ground area, is tightly coupled with photosynthetic activity and therefore Gross Primary Productivity (GPP) (). Directional shifts in GPP influence the accumulation rate of woody biomass, and affect the delicate balance between tree and grass co-occurrence in these ecosystems (), with potential consequences for transition between closed-canopy forest and open savanna. From a conservation perspective, deciduous savannas with a longer growth period support a greater diversity and abundance of wildlife, particularly bird species but also browsing mammals (). Extreme weather patterns as a result of climate change are leading to shorter but more intense leaf production cycles in these ecosystems which exist at the precipice of their climatic envelope, with severe negative consequences for biodiversity (). Understanding the determinants of seasonal patterns of tree leaf production (land-surface phenology) in dry deciduous savannas can provide valuable information on spatial variation in vulnerability to climate change, and help to model their contribution to land surface models under climate change.

Previous studies have shown that diurnal temperature variation and precipitation are the primary determinants of tree phenological activity in water-limited savannas (). At regional spatial scales, savanna phenological activity can be predicted well using only climatic factors and light environment (Adole, Dash, and Atkinson, 2018), but local variation exists in leaf production cycles which cannot be attributed solely to abiotic environment (). It has been repeatedly suggested that information on biotic environment play a larger role in predicting land-surface phenology (), but implementation is most often limited to coarse ecoregions or functional vegetation types (), which lack the fine-scale resolution which can now be paired with state-of-the-art earth observation data.

Tree species vary in their life history strategy regarding the timing of leaf production (). More conservative species (i.e. slower growing, robust leaves, denser wood) tend to initiate leaf production (green-up) before rainfall has commenced, and persist after the rainy season has finished, despite having lower overall GPP, while more resource acquisitive species and juvenile individuals tend to green-up during the rainy season, and create a dense leaf-flush during the mid-season peak of growth (). It has been suggested that this variation in leaf phenological activity between species is one aspect by which increased tree species richness causes an increase in ecosystem-level productivity in deciduous savannas (). Building on research linking biodiversity and ecosystem function, one might expect that an ecosystem with a greater diversity of tree species might be better able to maintain consistent leaf coverage for a longer period over the year, as species vary in their optimal growing conditions due to niche complementarity, whereby coexisting species vary in their occupation of niche space due to competitive exclusion ().

In the water-limited savannas such as those found in large areas of southern Africa (), the ability of conservative tree species to maintain consistent leaf coverage in the upper canopy strata over the growing season, but particularly at the start and end of the growing season, may provide facilitative effects to other tree species and juveniles occupying lower canopy strata that are less

51 well-adapted to moisture-limiting conditions, but are more productive, by providing shade and  
52 influencing below ground water availability through hydraulic lift ().

53 Variation in tree species composition, as well as species richness, is also expected to have an effect  
54 on savanna phenology in southern Africa. Savannas of a number of different types (species compo-  
55 sition and structure) are found across southern Africa, but these are often poorly differentiated in  
56 regional-scale phenological studies (), resulting in a dearth of information on the phenological be-  
57 haviour of different woodlands. As our ability to remotely sense tree species composition improves,  
58 it allows us to create more tailored models of the carbon cycle which incorporate not only climatic  
59 factors, but also biotic factors which govern productivity. We therefore need to understand how  
60 species composition and biodiversity metrics affect land-surface phenology.

61 In the deciduous woodlands of Zambia, a highly pronounced single wet-dry season annual oscilla-  
62 tion is observed across the majority of land area, with local exceptions in some mountainous areas  
63 (). Variation in leaf phenological activity across the country has a large influence on annual gross  
64 primary productivity. Using Zambia as a case study, we can expect similar response from decid-  
65 uous woodlands across southern Africa, with important consequences for the global carbon cycle  
66 ().

67 While cumulative leaf production across the growing season may be the most important aspect of  
68 leaf phenology for GPP, other phenological metrics may be more important for ecosystem function  
69 and habitat provision for wildlife. Periods of green-up and senescence which bookend the growing  
70 season are key times for invertebrate reproduction (), soil biotic activity () and herbivore brows-  
71 ing activity (). Pre-rainy season green-up in water-limited savannas provides a valuable source of  
72 moisture and nutrients before the rainy season, and can moderate the understorey microclimate,  
73 increasing humidity, reducing UV exposure, and moderating diurnal oscillations in temperature,  
74 reducing ecophysiological stress which can lead to mortality during the dry season. An increase in  
75 the time between leading tree growth and the onset of seasonal rains provides a buffer to stress-  
76 ful dry season climatic conditions and wildlife activity. A slower rate of green-up caused by tree  
77 species greening at different times provides an extended period of bud-burst, thus maintaining the  
78 important food source of nutrient rich young leaves for longer ().

79 In this study we contend that, across Zambian deciduous savannas, tree species diversity and com-  
80 position influence three key measurable aspects of the tree phenological cycle: (1) the rates of  
81 greening and senescence at the start and end of the seasonal growth phase, (2) the overall length  
82 of the growth period, and (3) the lag time between green-up/senescence and the start/end of the  
83 rainy season. It is hypothesised that: (H<sub>1</sub>) due to variation among species in minimum viable  
84 water availability for growth, plots with greater tree species richness will exhibit slower rates of  
85 greening and senescence as different species green-up and senesce at different times. We expect  
86 that: (H<sub>2</sub>) in plots with greater species richness the start of the growing season will occur ear-  
87 lier in respect to the onset of rain due to an increased likelihood of containing a species which  
88 can green-up early, facilitating other species to initiate the growing season. We hypothesise that:  
89 (H<sub>3</sub>) plots with greater species richness will exhibit a longer growth period and greater cumula-  
90 tive green-ness over the course of the growth period, due to a higher resilience to variation in wa-  
91 ter availability, acting as a buffer to ecosystem-level productivity. Finally, we hypothesise that:  
92 (H<sub>4</sub>) irrespective of species diversity, variation in tree species composition and vegetation type will

93 cause variation in the phenological metrics outlined above.

## 94 2 Materials and methods

### 95 2.1 Data collection

96 We used plot-level data on tree species diversity across 710 sites from the Zambian Integrated  
97 Land Use Assessment Phase II (ILUA-II), conducted in 2014 (Mukosha and Siampale, 2009; Pel-  
98 letier et al., 2018). Each site consisted of four 20x50 m (0.2 ha) plots positioned in a square around  
99 a central point, with a distance of 500 m between each plot (Figure 2). The original census con-  
100 tained 993 sites, which was filtered in order to define study bounds and to ensure data quality.  
101 Only sites with  $\geq 50$  stems  $\text{ha}^{-1}$   $\geq 10$  cm DBH (Diameter at Breast Height) were included in the  
102 analysis, to ensure all sites represented woody savanna rather than ‘grassy savanna’, which is con-  
103 sidered a separate biome with very different species composition and ecosystem processes govern-  
104 ing phenology (Parr et al., 2014). Sites in Mopane woodland were removed by filtering sites with  
105 greater than 50% of individuals belonging to *Colophospermum mopane*, preserving only plots with  
106 Zambesian tree savanna / woodland. To eliminate compositional outliers, plots with fewer than  
107 five species with more than one individual were excluded. Plots dominated by non-native tree  
108 species ( $\geq 50\%$  of individuals), e.g. *Pinus* spp. and *Eucalyptus* spp. were also excluded, as these  
109 species may exhibit non-seasonal patterns of leaf production ().

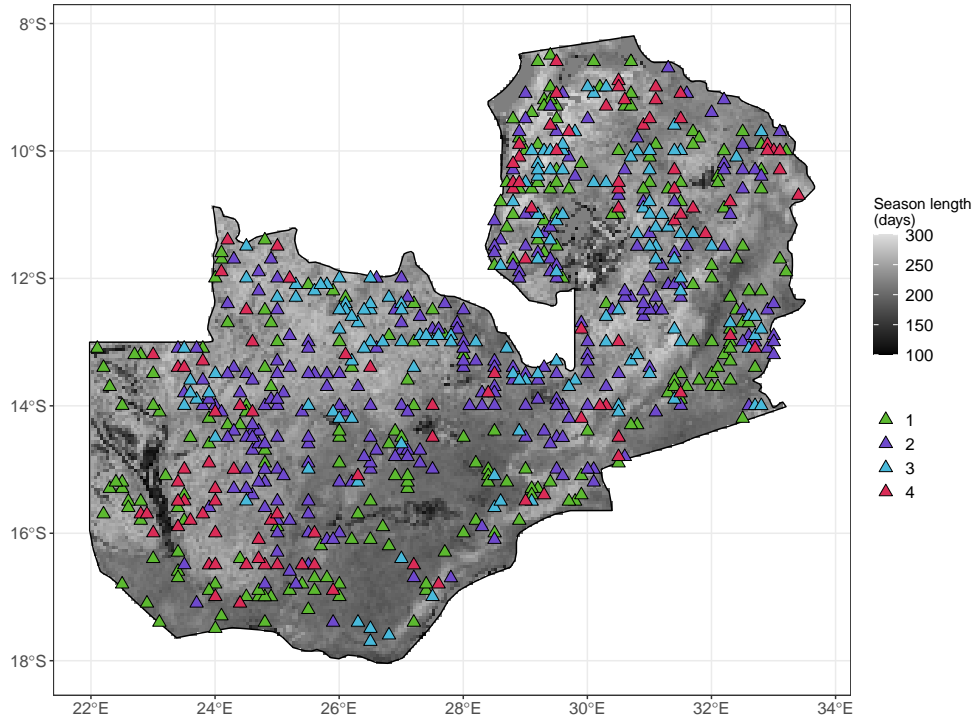


Figure 1: Distribution of study sites within Zambia as triangles, each consisting of four plots. Sites are coloured according to vegetation compositional cluster as identified by the PAM clustering algorithm on NSCA ordination axes of species abundance data. Zambia is shaded according to growing season length as estimated by the MODIS VIPPHEN-EVI2 product, at 0.05 degrees spatial resolution (Didan and Barreto, 2016).

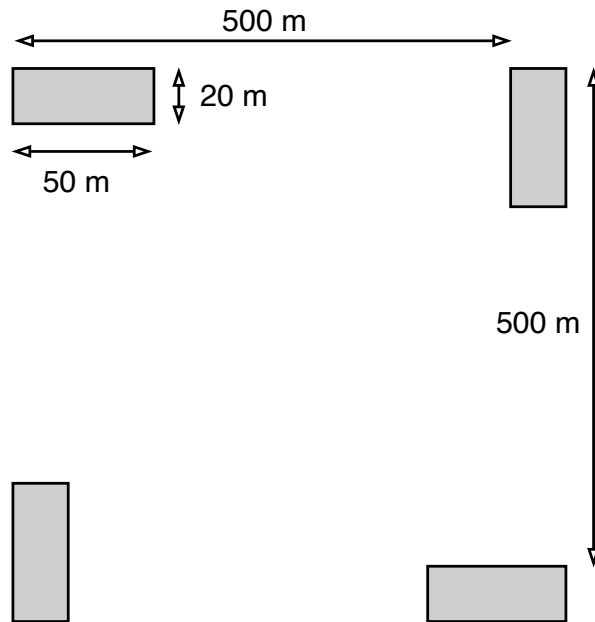


Figure 2: Schematic diagram of plot layout within a site. Each 20x50 m (0.2 ha) plot is shaded grey. The site centre is denoted by a circle. Note that the plot dimensions are not to scale.

110 Within each plot, the species of all trees with at least one stem  $\geq 10$  cm DBH were recorded. Plot  
111 data was aggregated to the site level for analyses to avoid pseudo-replication caused by the more  
112 spatially coarse phenology data. Tree species composition varied little among the four plots within  
113 a site, and were treated as representative of the woodland in the local area. Using the Bray-Curtis  
114 dissimilarity index of species abundance data, we calculated that the mean pairwise compositional  
115 distance between plots within a site was lower than the mean compositional distance across all  
116 pairs of plots in 93.4% of cases.

117 To quantify phenology at each site, we used the MODIS MOD13Q1 satellite data product at 250  
118 m resolution (Didan, 2015). The MOD13Q1 product provides an Enhanced Vegetation Index (EVI)  
119 time series at 16 day intervals. EVI is widely used as a measure of vegetation growth, as an im-  
120 provement to NDVI (Normalised Differential Vegetation Index), which tends to saturate at higher  
121 values. EVI is well-correlated with gross primary productivity and so can act as a suitable proxy  
122 (). We used all scenes from January 2015 to August 2020 with less than 20% cloud cover cover-  
123 ing the study area. All sites were determined to have a single annual growth season according to  
124 the MODIS VIPPHEN product (), which assigns pixels ( $0.05^\circ$ , 5.55 km at equator) up to three  
125 growth seasons per year. We stacked yearly data between 2015 and 2020 and fit a General Addi-  
126 tive Model (GAM) to produce an average EVI curve. We estimated the start and end of the grow-  
127 ing season using the first derivative of the GAM. We identified the start of the growing season as  
128 the day where the slope of the curve first exceeds a slope of 0.05, which is maintained or exceeded  
129 for 10 or more days and the end of the growing season as the last day where the slope of the curve  
130 falls below -0.05, which has been maintained for 10 or more days. We estimated the length of the  
131 growing season as the number of days between the start and end of the growing season defined as  
132 above. We estimated the greening rate as the slope of a linear model across EVI values between  
133 the start of the growing season and the point at which the slope of increase fell below 0.05. Simi-  
134 larly the senescence rate was estimated as the slope of a linear model between the point where the  
135 slope of decrease fell below -0.05 and the end of the growing season Figure 3.

136 Precipitation data was gathered using the “GPM IMERG Final Precipitation L3 1 day V06” dataset,  
137 which has a pixel size of  $0.1^\circ$  (11.1 km at the equator) (**GPM**), between 2015 and 2020. Daily to-  
138 tal precipitation was separated into two periods: precipitation during the growing season (growing  
139 season precipitation), and precipitation in the 90 day period before the onset of the growing sea-  
140 son (dry season precipitation). Similar to estimation of the growing season, the rainy season was  
141 defined using the first derivative of a GAM to create a curve for each site using stacked yearly pre-  
142 cipitation data. The slope coefficient used to identify the start and end of the rainy season was  
143 0.06. Mean diurnal temperature range (Diurnal  $\delta T$ ) was calculated as the mean of monthly tem-  
144 perature range from the WorldClim database, using the BioClim variables, with a pixel size of 30  
145 arc seconds (926 m at the equator) (Fick and Hijmans, 2017). averaged across all years of avail-  
146 able data (1970-2000).

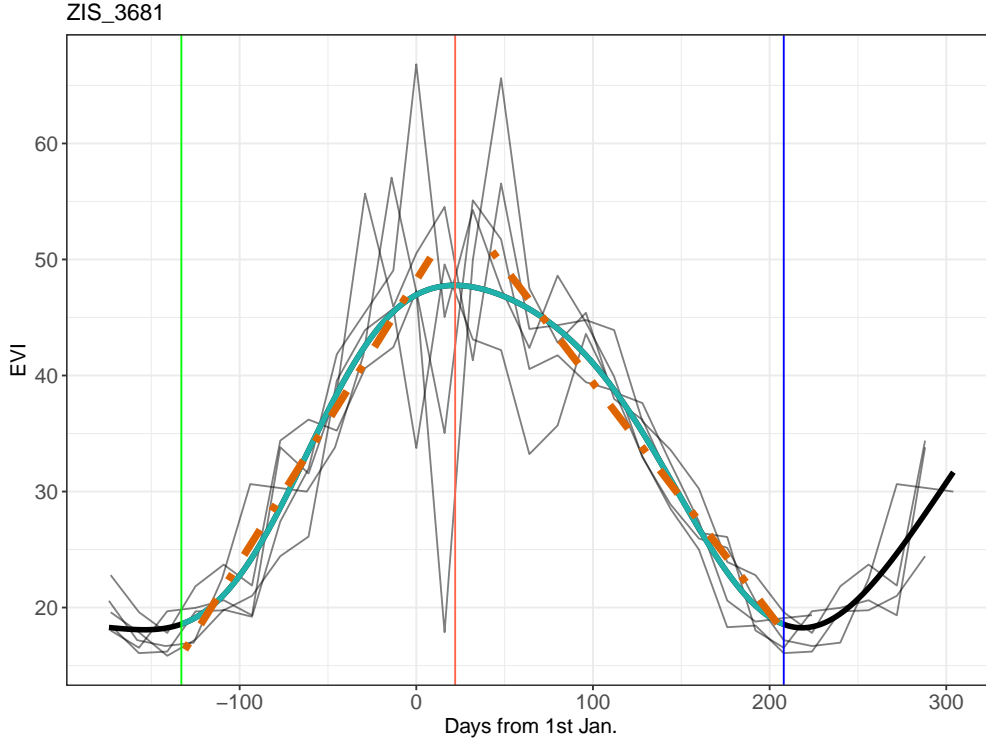


Figure 3: Example EVI time series, demonstrating the metrics derived from it. Thin black lines show the raw EVI time series, with one line for each annual growth season. The thick black line shows the GAM fit. The thin blue lines show the minima which bound the growing season. The red line shows the maximum EVI value reached within the growing season. The shaded cyan area of the GAM fit shows the growing season, as defined by the first derivative of the GAM curve. The two orange dashed lines are linear regressions predicting the greening rate and senescence rate at the start and end of the growing season, respectively. Note that while the raw EVI time series fluctuate greatly around the middle of the growing season, mostly due to cloud cover, the GAM fit effectively smooths this variation to estimate the average EVI during the mid-season period.

## 2.2 Data analysis

To quantify variation in tree species composition we used a Non-Symmetric Correspondence Analysis (NSCA), based on tree species abundance per site, with four axes (Inertia = 1.81), using the `ade4` R package (Dray and Dufour, 2007). The first 4 axes of the NSCA explained `nscaper` of the variation in species composition among sites according to eigenvalue analysis. We performed clustering on the NSCA axes, using the PAM (Partitioning Around Medoids) algorithm available in the `cluster` R package (Maechler et al., 2019). We identified four compositional groups based on this analysis which were used as random effects in statistical models. **More technical detail on NSCA procedure**

We specified multivariate linear models to assess the role of tree species diversity on each of the chosen phenological metrics. We defined tree species diversity using both species richness and abundance evenness. Abundance evenness was calculated as the Shannon Equitability index ( $E_{H'}$ ) (Smith1996) was calculated as the ratio of the Shannon diversity index to the natural log of species richness. We defined a maximal model structure including tree species richness, abundance

evenness, the interaction of species richness and vegetation type, and climatic variables shown by previous studies to strongly influence phenology. The maximal model was compared to models with different subsets of independent variables using the model log likelihood, AIC (Akaike Information Criteria), BIC (Bayesian Information Criteria), and adjusted  $R^2$  values for each model. For each phenological metric, the best model according to the model quality statistics is reported in the results. All models were fitted using Maximum Likelihood to allow comparison of models (). Independent variables in each model were transformed to achieve normality where necessary and standardised to Z-scores prior to modelling. All statistical analyses were conducted in R version 4.0.2 (R Core Team, 2020).

### 3 Results

All the best models for each phenological metric included richness and evenness in their fixed effects Figure 4. Models predicting rainy season - growing season lags included only species and richness.

Against expectations, tree species richness and evenness had negative effects on cumulative EVI, while wet season precipitation had a positive effect and diurnal temperature range had a negative effect, as expected. Despite this, species richness had a positive effect on season length.

Model estimates for both greening rate were poorly constrained, with wide confidence intervals on model coefficients which overlapped zero, indicating that other unmeasured drivers had important influence over these variables. The model for greening rate explained  $<0.001\%$  of variance in the response variable. As expected, greening rate was negatively affected by species richness and abundance evenness. A higher species richness led to a slower rate of greening.

Interestingly, the best models for start of season and end of season lag time included only species richness and evenness as fixed effects. An increase in species richness led to an increase in the negative lag between the start of the growing season and the start of the rainy season, while the opposite was true for the lag between the end of the growing season and the end of the rainy season. For all phenological metrics, models including species richness and evenness were of better quality than those containing only climatic variables.

The slope of the relationship between species richness and phenological metrics varied among vegetation types, but largely maintained the same direction. Only in greening rate, senescence rate, and season length did vegetation types have differing slope estimates.



Table 1: Legendre indicator species analysis for the four vegetation type clusters identified by the PAM algorithm

Cluster	Species	Indicator value
1	<i>Diplorhynchus condylocarpon</i>	0.256
	<i>Combretum molle</i>	0.243
	<i>Combretum zeyheri</i>	0.241
2	<i>Julbernardia paniculata</i>	0.559
	<i>Brachystegia boehmii</i>	0.455
	<i>Psuedolachnostylis maprouneifolia</i>	0.189
3	<i>Brachystegia longifolia</i>	0.722
	<i>Uapaca kirkiana</i>	0.299
	<i>Isoberlinia angolensis</i>	0.239
4	<i>Brachystegia spiciformis</i>	0.718
	<i>Pterocarpus angolensis</i>	0.242
	<i>Diospyros batocana</i>	0.229

Response	$\delta AIC$	$\delta BIC$	$R^2_{adj}$	$\delta \log Lik$
Cumulative EVI	-22.1	-31.2	0.11	9.06
Season length	-2.9	-12.0	0.24	-0.53
Greening rate	-60.2	-69.3	0.00	28.12
Senescence rate	-69.5	-78.6	0.02	32.77
Start lag	14.2	5.1	0.06	-9.08
End lag	13.1	4.0	0.03	-8.57

Table 2: Model fit statistics for each phenological metric.

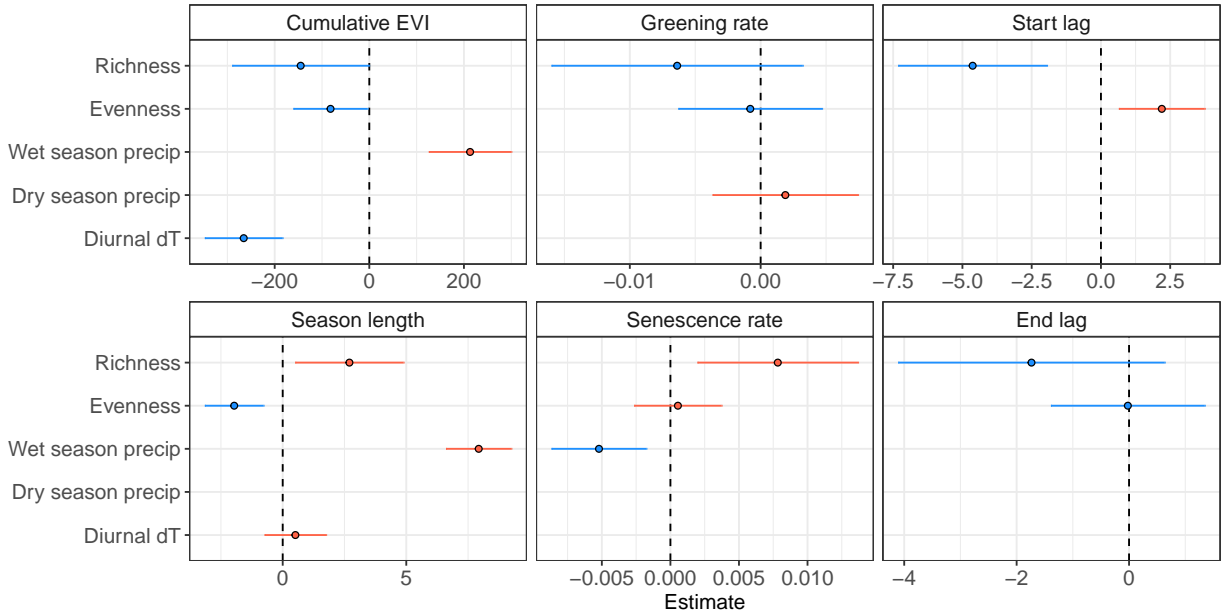


Figure 4: Predictor slope estimates for each maximal model of a phenological metric. Slope estimates are  $\pm 1$  standard error. Slope estimates where the interval (standard error) does not overlap zero are considered to be significant effects.

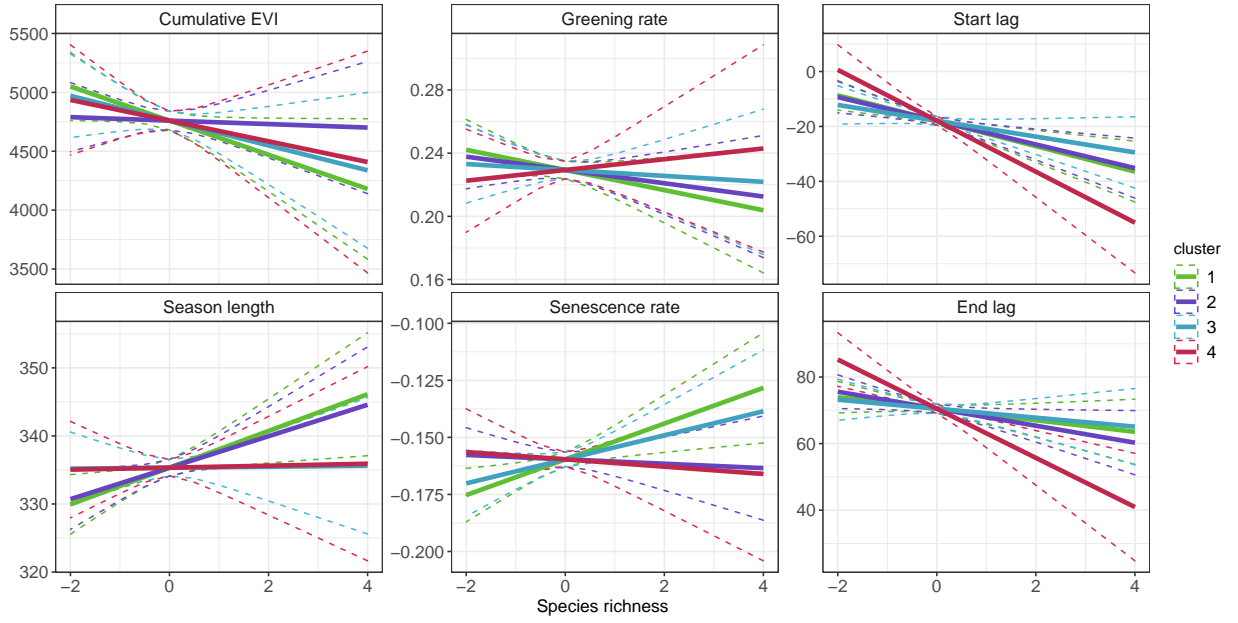


Figure 5: Marginal effects of tree species richness on each of the phenological metrics, for each vegetation type, using the best model for each phenological metric.

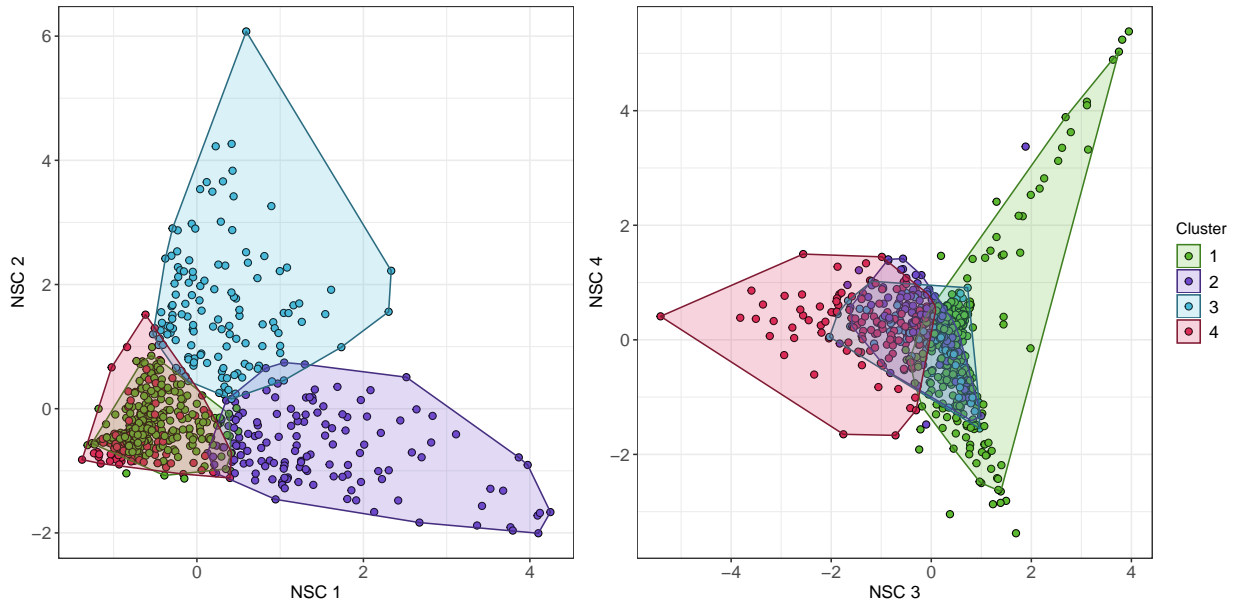


Figure 6: Plot scores of the (A) first and second, and (B) third and fourth axes of the Non-Symmetric Correspondence Analysis of tree species composition. Points are coloured according to clusters defined by the PAM algorithm on the NSCA ordination axes, along with a convex hull encompassing 95% of the points in each cluster.

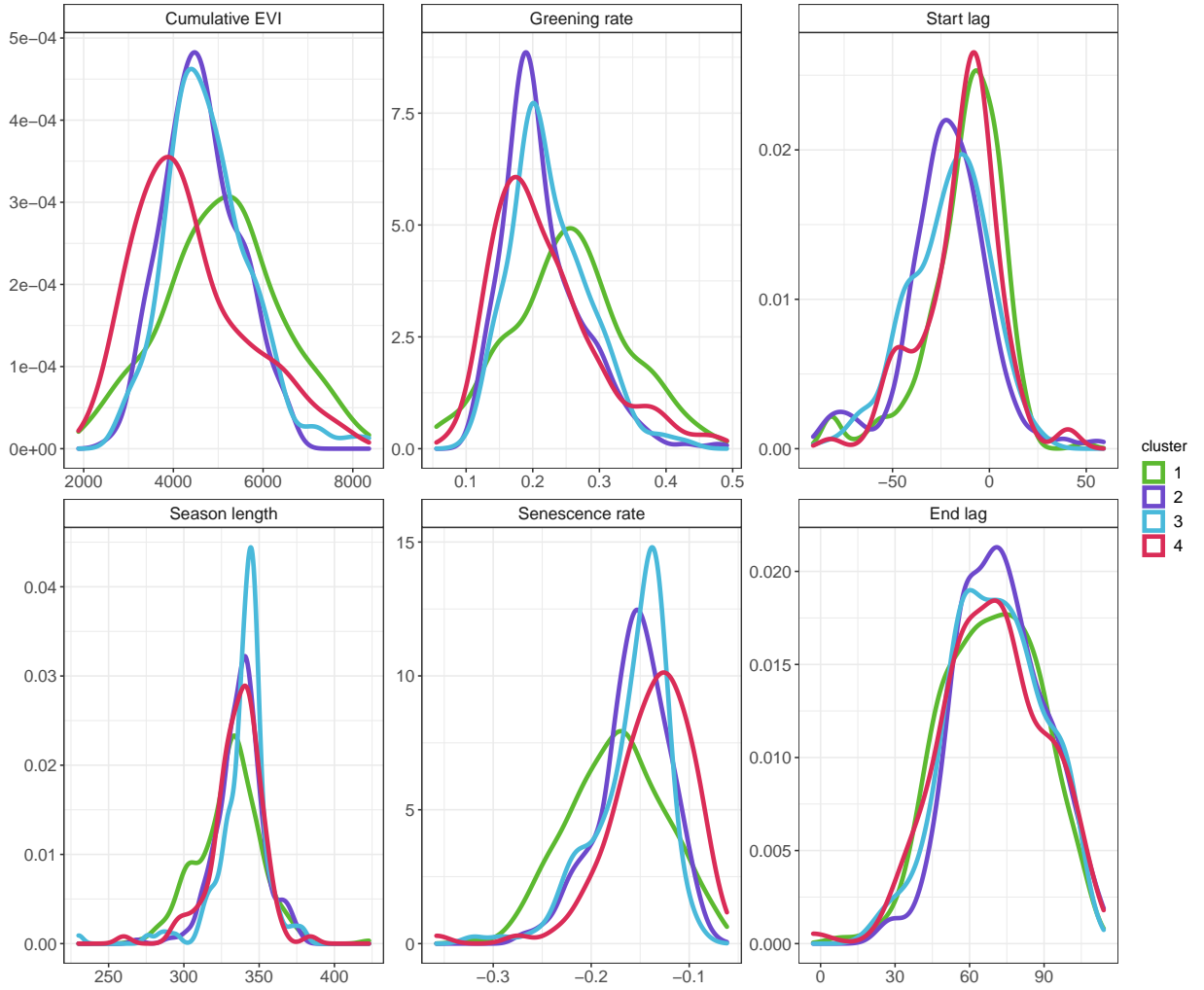


Figure 7

## 4 Discussion

The ability for us now nearing to be able to remotely sense tree species diversity allows us to make more tailored models of the carbon cycle which incorporate not only climatic factors, but also biotic factors which govern productivity. We therefore need to understand how species composition and biodiversity metrics affect land-surface phenology.

## 5 Conclusion

## References

Adole, Tracy, Jadunandan Dash, and Peter M. Atkinson (2018). “Large-scale prerain vegetation green-up across Africa”. In: *Global Change Biology* 24.9, pp. 4054–4068. DOI: 10.1111/gcb.14310.

201 Didan, L. (2015). *MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid*  
202 *V006 [Data set]*. NASA EOSDIS Land Processes DAAC. DOI: 10.5067/MODIS/MOD13Q1.006.  
203 (Visited on 08/05/2020).

204 Didan, L. and A. Barreto (2016). *NASA MEaSUREs Vegetation Index and Phenology (VIP) Phenology*  
205 *EVI2 Yearly Global 0.05Deg CMG [Data set]*. NASA EOSDIS Land Processes DAAC.  
206 DOI: 10.5067/MEaSUREs/VIP/VIPPHEN\_EVI2.004. (Visited on 08/05/2020).

207 Dray, Stéphane and Anne-Béatrice Dufour (2007). “The ade4 Package: Implementing the Duality  
208 Diagram for Ecologists”. In: *Journal of Statistical Software* 22.4, pp. 1–20. DOI: 10.18637/jss.  
209 v022.i04.

210 Fick, S. E. and R. J. Hijmans (2017). “WorldClim 2: New 1-km spatial resolution climate surfaces  
211 for global land areas”. In: *International Journal of Climatology* 37.12, pp. 4302–4315. DOI: <http://dx.doi.org/10.1002/joc.5086>.

212 Maechler, Martin et al. (2019). *cluster: Cluster Analysis Basics and Extensions*. R package version  
213 2.1.0 — For new features, see the ‘Changelog’ file (in the package source).

214 Mukosha, J and A Siampale (2009). *Integrated land use assessment Zambia 2005–2008*. Lusaka,  
215 Zambia: Ministry of Tourism, Environment et al.

216 Parr, C. L. et al. (2014). “Tropical grassy biomes: misunderstood, neglected, and under threat”. In:  
217 *Trends in Ecology and Evolution* 29, pp. 205–213. DOI: 10.1016/j.tree.2014.02.004.

218 Pelletier, J. et al. (2018). “Carbon sink despite large deforestation in African tropical dry forests  
219 (miombo woodlands)”. In: *Environmental Research Letters* 13, p. 094017. DOI: 10.1088/1748-  
220 9326/aadc9a.

221 R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation  
222 for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.