Title: Phenology and diversity in Zambia

Authors: Godlee, J. L.[1],

[1]: School of GeoSciences, University of Edinburgh, Edinburgh, United Kingdom

Corresponding author:

John L. Godlee

johngodlee@gmail.com

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

Acknowledgements

Author constribution statement

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 acquisative 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

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

Variation in tree species composition, as well as species richness, is also expected to have an effect on savanna phenology in southern Africa. Savannas of a number of different types (species composition and structure) are found across southern Africa, but these are often poorly differentiated in regional-scale phenological studies (), resulting in a dearth of information on the phenological behaviour of different woodlands. As our ability to remotely sense tree species composition improves, it allows us to create 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.

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

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

In this study we contend that, across Zambian deciduous savannas, tree species diversity and composition influence three key measurable aspects of the tree phenological cycle: (1) the rates of greening and senescence at the start and end of the seasonal growth phase, (2) the overall length of the growth period, and (3) the lag time between green-up/senescence and the start/end of the rainy season. It is hypothesised that: (H_1) due to variation among species in minimum viable water availability for growth, plots with greater tree species richness will exhibit slower rates of greening and senescence as different species green-up and senesce at different times. We expect that: (H_2) in plots with greater species richness the start of the growing season will occur earlier in respect to the onset of rain due to an increased likelihood of containing a species which can green-up early, facilitating other species to initiate the growing season. We hypothesise that: (H_3) plots with greater species richness will exhibit a longer growth period and greater cumulative green-ness over the course of the growth period, due to a higher resilience to variation in water availability, acting as a buffer to ecosystem-level productivity. Finally, we hypothesise that: (H_4) irrespective of species diversity, variation in tree species composition and vegetation type will

cause variation in the phenological metrics outlined above.

2 Materials and methods

2.1 Data collection

We used plot-level data on tree species diversity across 709 sites from the Zambian Integrated Land Use Assessment Phase II (ILUA-II), conducted in 2014 (Mukosha and Siampale, 2009; Pelletier et al., 2018). Each site consisted of four 20x50 m (0.2 ha) plots positioned in a square around a central point, with a distance of 500 m between each plot (Figure 2). The original census contained 993 sites, which was filtered in order to define study bounds and to ensure data quality. Only sites with $\geq$50 stems ha^{-1} $\geq$10 cm DBH (Diameter at Breast Height) were included in the analysis, to ensure all sites represented woody savanna rather than 'grassy savanna', which is considered a separate biome with very different species composition and ecosystem processes governing phenology (Parr et al., 2014). Sites in Mopane woodland were removed by filtering sites with greater than 50% of individuals belonging to *Colophospermum mopane*, preserving only plots with Zambesian tree savanna / woodland. To eliminate compositional outliers, plots with fewer than five species with more than one individual were excluded. Plots dominated by non-native tree species ($\geq$50% of individuals), e.g. *Pinus* spp. and *Eucalyptus* spp. were also excluded, as these species may exhibit non-seasonal patterns of leaf production ().

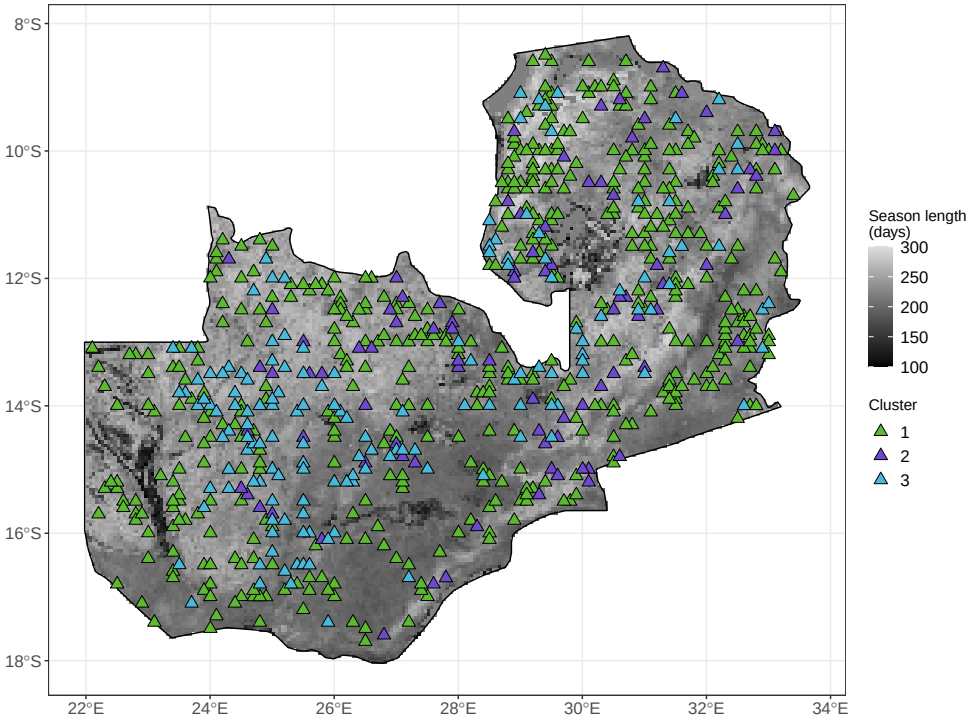

Figure 1: Distribution of study sites within Zambia as triangles, each consisting of four plots. Sites are oloured according to vegetation compositional cluster as identifed by Ward's clustering algorithm on euclidean distance of plots in the first two axes of NSCA ordination space. Zambia is shaded according to growing season length as estimated by the MODIS VIPPHEN-EVI2 product, at 0.05 degrees spatial resolution (Didan and Barreto, 2016).

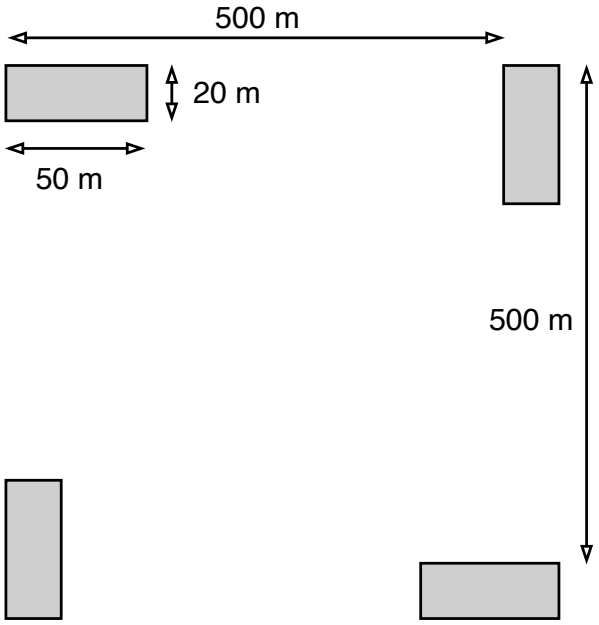

Figure 2: Schematic diagram of plot layout within a site. Each 20x50 m (0.2 ha) plot is shaded grey. The site centre is denoted by a circle. Note that the plot dimensions are not to scale.

Within each plot, the species of all trees with at least one stem ≥10 cm DBH were recorded. Plot data was aggregated to the site level for analyses to avoid pseudo-replication caused by the more spatially coarse phenology data. Tree species composition varied little among the four plots within a site, and were treated as representative of the woodland in the local area. Using the Bray-Curtis dissimilarity index of species abundance data, we calculated that the mean pairwise compositional distance between plots within a site was lower than the mean compositional distance across all pairs of plots in 93.9% of cases.

To quantify phenology at each site, we used the MODIS MOD13Q1 satellite data product at 250 m resolution (Didan, 2015). The MOD13Q1 product provides an Enhanced Vegetation Index (EVI) time series at 16 day intervals. EVI is widely used as a measure of vegetation growth, as an improvement to NDVI (Normalised Differential Vegetation Index), which tends to saturate at higher values. EVI is well-correlated with gross primary productivity and so can act as a suitable proxy (). We used all scenes from January 2015 to August 2020 with less than 20% cloud cover covering the study area. All sites were determined to have a single annual growth season according to the MODIS VIPPHEN product (), which assigns pixels (0.05°, 5.55 km at equator) up to three growth seasons per year. We stacked yearly data between 2015 and 2020 and fit a General Additive Model (GAM) to produce an average EVI curve. We estimated the start and end of the growing season using the first derivative of the GAM. We identified the start of the growing season as the day where the slope of the curve first exceeds a slope of 0.05, which is maintained or exceeded for 10 or more days and the end of the growing season as the last day where the slope of the curve falls below -0.05, which has been maintained for 10 or more days. We estimated the length of the growing season as the number of days between the start and end of the growing season defined as above. We estimated the green-up rate as the slope of a linear model across EVI values between the start of the growing season and the point at which the slope of increase fell below 0.05. Similarly the senescence rate was estimated as the slope of a linear model between the point where the slope of decrease fell below -0.05 and the end of the growing season Figure 3.

Precipitation data was gathered using the "GPM IMERG Final Precipitation L3 1 day V06" dataset, which has a pixel size of 0.1°(11.1 km at the equator) (**GPM**), between 2015 and 2020. Daily total precipitation was separated into two periods: precipitation during the growing season (growing season precipitation), and precipitation in the 90 day period before the onset of the growing season (dry season precipitation). Similar to estimation of the growing season, the rainy season was defined using the first derivative of a GAM to create a curve for each site using stacked yearly precipitation data. The slope coefficient used to identify the start and end of the rainy season was 0.06. Mean diurnal temperature range (Diurnal δT) was calculated as the mean of monthly temperature range from the WorldClim database, using the BioClim variables, with a pixel size of 30 arc seconds (926 m at the equator) (Fick and Hijmans, 2017). averaged across all years of available data (1970-2000).

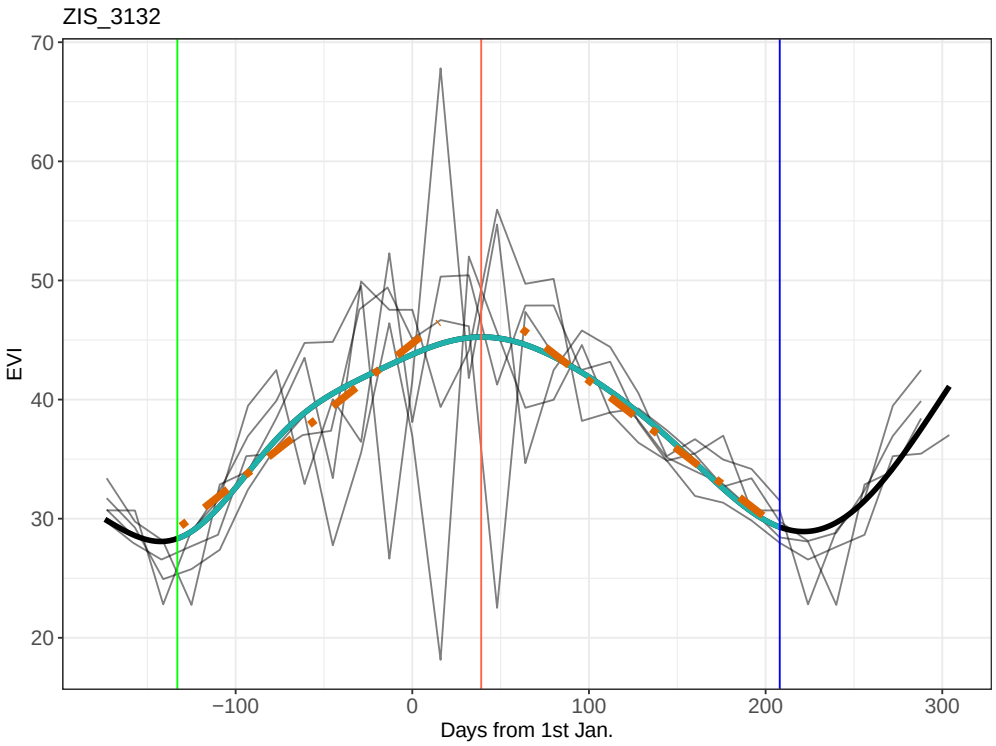

Figure 3: Example EVI time series, demonstrating the metrics derived from it. Thin black lines show the raw EVI time series, with one line for each annual growth season. The thick black line shows the GAM fit. The thin blue lines show the minima which bound the growing season. The red line shows the 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 green-up 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 measure variation in tree species composition we a used combination of Non-symmetric Correspondence Analysis (NSCA) and agglomerative hirerarchical clustering on species abundance data (**Kreft2010**; **Fayolle2014**), and was performed using the `ade4` R package (Dray and Dufour, 2007). To guard against sensitivity to rare individuals, which can preclude meaningful cluster delineation across such a large species compositional range, we restricted the NSCA to species with greater than five records, and to sites with fewer than five species (). We used Ward's algorithm to define clusters (**Murtagh2014**), based on the euclidean distance of sites in NSCA ordination space. We determined the optimal number of clusters by maximising the mean silhouette width among clusters (**Rousseeuw1987**) **??**. Vegetation type clusters were used later as interaction terms in linear models. We described the vegetation types represented by each of the clusters using a Dufrene-Legendre indicator species analysis (**Dufrene1997**).

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 as separate independent variables. 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 quality of 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. Where two similar models were within 2 AIC points of each other, the model with fewer terms was chosen as the best model, to maximise model parsimony. 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 to allow comparison of slope coefficients within a given model. All statistical analyses were conducted in R version 4.0.2 (R Core Team, 2020).

3 Results

Model selection showed that richness and evenness are important determinants of each of the chosen phenological metrics, across vegetation types. Species richness featured in all models, though was not significant in cumulative EVI or senescence rate, while evenness was included in models for cumulative EVI and season length only Figure 4.

NSCA used 4 axes, which accounted for 12.8% of the variance according to eigenvalue decay. 3 were identified during clustering. Pairwise Wilcoxon Signed Rank Tests showed that clusters were significantly different in species composition ($p<0.05$).

In models for green-up lag and senescence lag, species richness had consistent effects effects on the phenological response to rainy season onset and decline. Species richness caused green-up to occur increasingly earlier with respect to the rainy season onset, while richness caused senescence onset to occur later after the end of the rainy season.

Against expectations, tree species evenness had a negative effect on cumulative EVI, and richness had a negligible, non-significant effect. Wet season precipitation had a positive effect and diurnal temperature range had a negative effect, as expected. Despite this, species richness had a significant positive effect on season length. It is striking that richness and evenness have contrasting effects on season length.

All models were of better quality than models which included only climatic variables Table 2. The phenological metrics best predicted were green-up lag and season length, where models explained 25% and 24% of the variance in these variables, respectively. Senescence rate and senescence lag were the least well predicted phenological metrics, with the best model explaining 5% and 10% of their variance, respectively.

The slope of the relationship between species richness and phenological metrics varied among vegetation types, but largely maintained the same direction Figure 5. Models predicting green-up and

senescence lag had the tightest confidence intervals among vegetation type marginal effects.

The hierarchical clustering analysis demonstrated that there was some degree of spatial structure to the vegetation types. Sites classified as within cluster 3 were found predominantly in the south and southwest of the country, while sites in cluster 2 were found predominantly in the north. Cluster 1 sites were restricted mostly to the centre and northwest of the country Figure 1.

Cluster	Species	Indicator value
1	*Pterocarpus angolensis*	0.262
	Diplorhynchus condylocarpon	0.254
	Brachystegia longifolia	0.242
2	*Brachystegia boehmii*	0.783
	Psuedolachnostylis maprouneifolia	0.244
	Brachystegia spiciformis	0.215
3	*Julbernardia paniculata*	0.701
	Diplorhynchus condylocarpon	0.274
	Psuedolachnostylis maprouneifolia	0.236

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

Response	δAIC	δBIC	R^2_{adj}	δlogLik
Cumulative EVI	4.8	-13.5	0.11	-6.38
Season length	11.4	-2.3	0.23	-8.68
Green-up rate	3.6	-10.1	0.10	-4.78
Senescence rate	10.2	-3.5	0.06	-8.08
Green-up lag	75.1	56.8	0.25	-41.53
Senescence lag	34.0	20.3	0.10	-19.98

Table 2: Model fit statistics for each phenological metric.

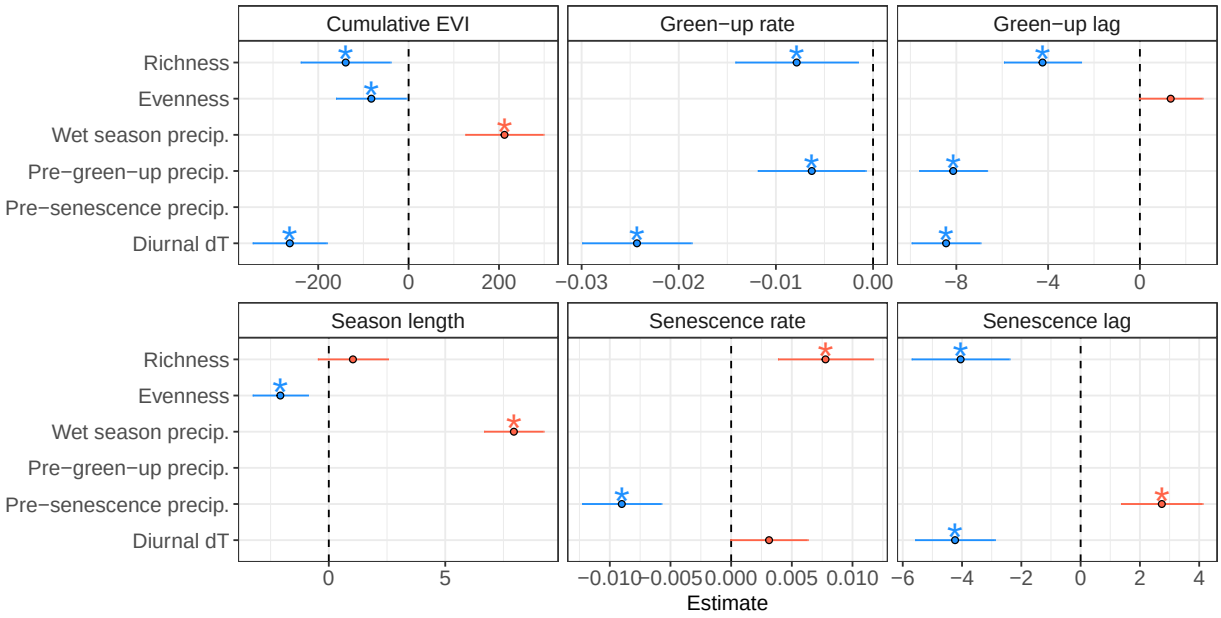

Figure 4: Standardized slope coefficients for each best model of a phenological metric. Slope estimates are ±1 standard error. Slope estimates where the interval (standard error) does not overlap zero are considered to be significant effects.

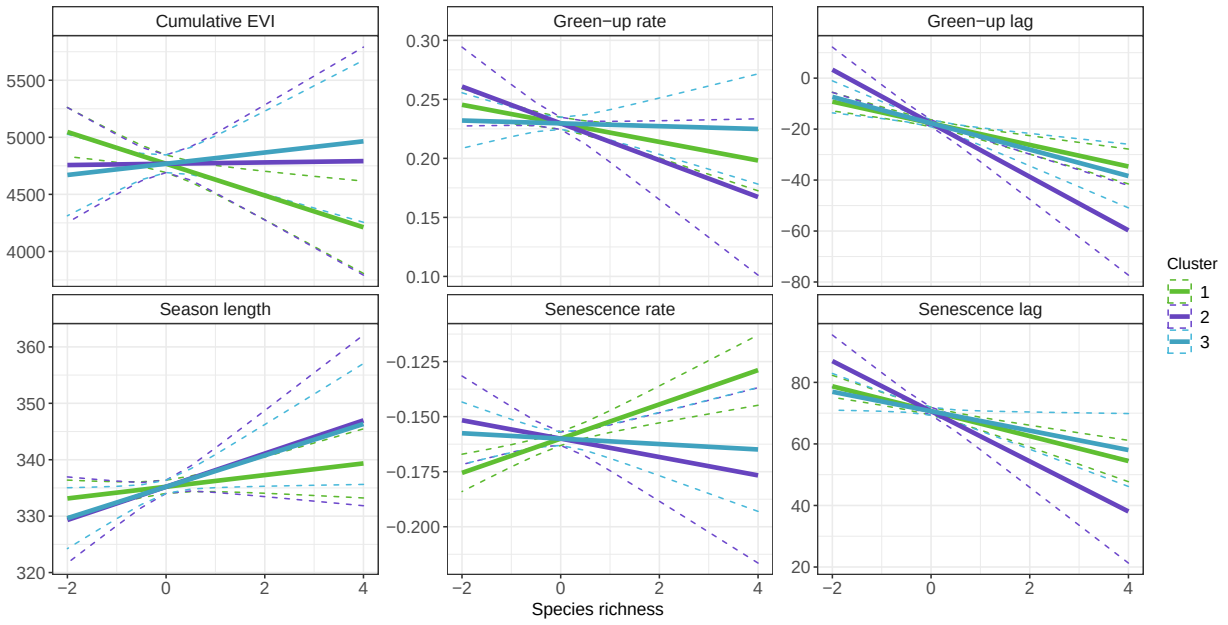

Figure 5: Marginal effects of tree species richness on each of the phenological metrics, for each vegetation type, using the best model for each phenological metric.

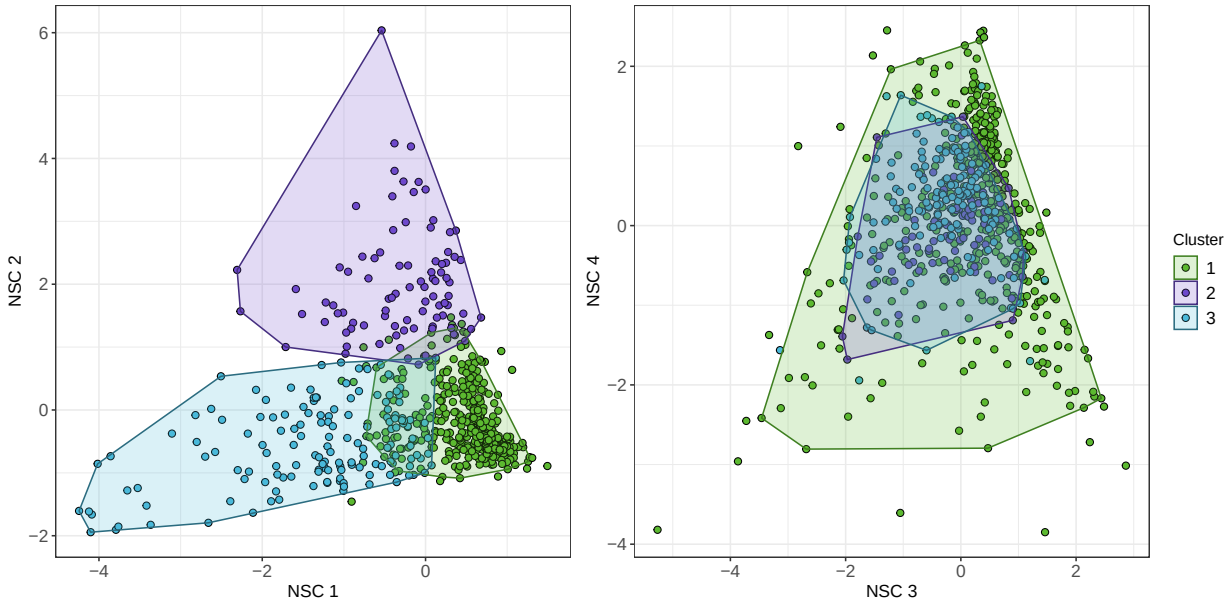

Figure 6: Plot scores of the (A) first and second, and (B) third and fourth axes of the Non-Symmetric Correspondence Analysis of tree species composition. Points are coloured according to clusters defined by Ward's algorithm on euclidean distances of the NSCA ordination axes, along with a convex hull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

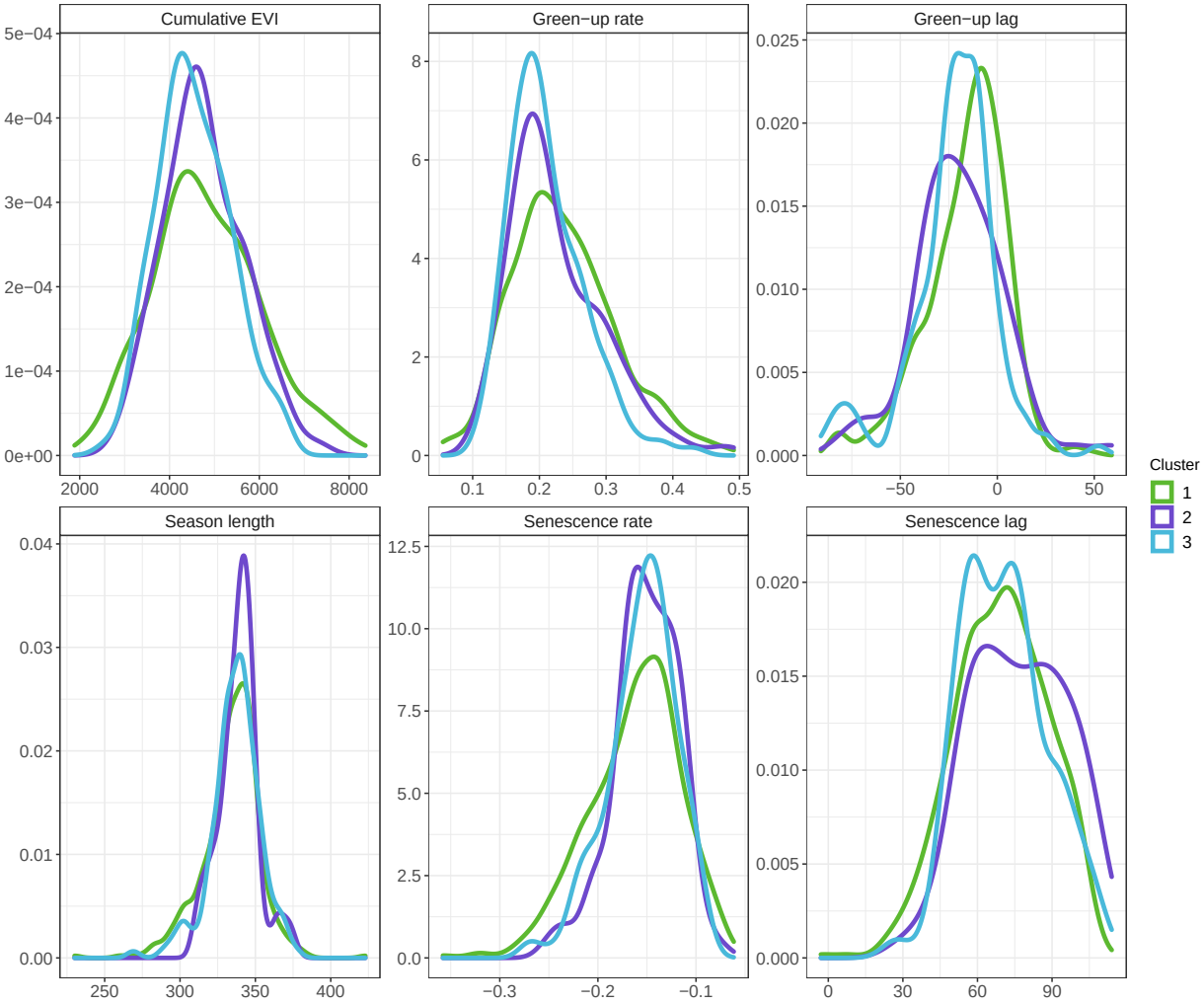

Figure 7

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.

It's possible that the reason we couldn't predict season length well is that there is very little variation between sites. Notably, only cluster three deviates from the density distribution of the other two clusters, with more sites at the lower end of the distribution. Cluster 3 also differed in significantly in senescence rate from the other two clusters, with many more plots showing a very steep senescence rate. Interestingly, Cluster three had wider variation in cumulative EVI than the other clusters and this is reflected in the marginal effects of species richness on these phenological metrics. Species richness had a greater effect on cumulative EVI than the other two clusters, and led to a steeper senescence rate.. **WHYTHO?**

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.

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

Didan, L. and A. Barreto (2016). *NASA MEaSUREs Vegetation Index and Phenology (VIP) Phenology EVI2 Yearly Global 0.05Deg CMG [Data set]*. NASA EOSDIS Land Processes DAAC. DOI: 10.5067/MEaSUREs/VIP/VIPPHEN_EVI2.004. (Visited on 08/05/2020).

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

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

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

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

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

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

6 Supplementary material

Rank	Precipitation	Diurnal dT	Evenness	Richness	logLik	AIC	ΔIC	W_i
1	✓	✓	✓	✓	-5933	11885	0.00	0.333
2	✓	✓	✓	✓	-5935	11885	0.21	0.299
3	✓	✓		✓	-5937	11887	2.08	0.118
4	✓	✓	✓		-5937	11887	2.14	0.115
5	✓	✓		✓	-5935	11887	2.31	0.105
6	✓	✓			-5940	11890	4.77	0.031
7		✓	✓		-5947	11904	18.38	0.000
8		✓			-5948	11905	20.15	0.000
9		✓	✓	✓	-5946	11905	20.26	0.000
10		✓	✓	✓	-5945	11906	20.99	0.000

Table 3: Cumulative EVI model selection candidate models, with fit statistics.

Rank	Precipitation	Diurnal dT	Evenness	Richness	logLik	AIC	ΔIC	W_i
1	✓	✓		✓	-3041	6094	0.00	0.346
2	✓	✓		✓	-3039	6094	0.05	0.338
3	✓	✓	✓	✓	-3039	6096	1.53	0.161
4	✓	✓	✓	✓	-3041	6096	1.63	0.153
5		✓		✓	-3046	6107	13.27	0.000
6		✓		✓	-3048	6107	13.28	0.000
7		✓	✓	✓	-3046	6109	14.93	0.000
8		✓	✓	✓	-3048	6109	15.04	0.000
9	✓	✓			-3059	6128	34.02	0.000
10	✓	✓	✓		-3058	6129	34.77	0.000

Table 4: Senescence lag model selection candidate models, with fit statistics.

Rank	Precipitation	Diurnal dT	Evenness	Richness	logLik	AIC	ΔIC	W_i
1	✓	✓		✓	881	-1751	0.00	0.369
2	✓	✓	✓	✓	881	-1749	1.35	0.188
3	✓	✓		✓	882	-1749	1.91	0.142
4		✓		✓	879	-1748	2.88	0.087
5	✓	✓	✓	✓	883	-1748	3.05	0.080
6		✓	✓	✓	879	-1746	4.45	0.040
7		✓		✓	880	-1746	4.89	0.032
8	✓	✓			877	-1745	5.46	0.024
9	✓	✓	✓		878	-1744	6.28	0.016
10		✓	✓	✓	880	-1744	6.29	0.016

Table 5: Green-up rate model selection candidate models, with fit statistics.

Rank	Precipitation	Diurnal dT	Evenness	Richness	logLik	AIC	ΔIC	W_i
1	✓		✓	✓	-2966	5945	0.00	0.448
2	✓	✓	✓	✓	-2966	5946	1.43	0.220
3	✓		✓	✓	-2965	5947	2.08	0.158
4	✓	✓	✓	✓	-2965	5948	3.53	0.077
5	✓		✓		-2969	5949	3.96	0.062
6	✓	✓	✓		-2969	5950	5.69	0.026
7	✓			✓	-2972	5954	9.49	0.004
8	✓	✓		✓	-2971	5955	10.67	0.002
9	✓			✓	-2971	5957	11.90	0.001
10	✓				-2974	5957	11.92	0.001

Table 6: Season length model selection candidate models, with fit statistics.

Rank	Precipitation	Diurnal dT	Evenness	Richness	logLik	AIC	ΔIC	W_i
1	✓	✓		✓	1245	-2474	0.00	0.449
2	✓			✓	1243	-2472	1.70	0.192
3	✓	✓	✓	✓	1245	-2472	1.80	0.182
4	✓		✓	✓	1243	-2470	3.56	0.076
5	✓	✓		✓	1240	-2469	4.63	0.044
6	✓			✓	1239	-2468	5.98	0.023
7	✓	✓	✓	✓	1240	-2467	6.51	0.017
8	✓		✓	✓	1239	-2466	7.92	0.009
9	✓				1236	-2464	10.04	0.003
10	✓	✓			1237	-2464	10.16	0.003

Table 7: Senescence rate model selection candidate models, with fit statistics.

Rank	Precipitation	Diurnal dT	Evenness	Richness	logLik	AIC	ΔIC	W_i
1	✓	✓	✓	✓	-3074	6167	0.00	0.563
2	✓	✓		✓	-3076	6168	1.63	0.249
3	✓	✓	✓	✓	-3077	6169	2.85	0.135
4	✓	✓		✓	-3079	6171	4.72	0.053
5	✓	✓			-3102	6215	48.86	0.000
6	✓	✓	✓		-3102	6216	49.48	0.000
7		✓	✓	✓	-3130	6273	106.35	0.000
8		✓	✓	✓	-3128	6273	106.77	0.000
9		✓		✓	-3133	6277	110.46	0.000
10		✓		✓	-3131	6277	110.63	0.000

Table 8: Green-up lag model selection candidate models, with fit statistics.