Satellite derived phenology of southern Africa for 1985-2000 and functional classification of vegetation based on phenometrics

K Steenkamp [a] , K Wessels [a], S. Archibald [b], G von Maltitz [b]

[a] Meraka Institute, CSIR, Meiring Naude Ave, Pretoria, 0001, South Africa – (ksteenkamp, kwessels)@csir.co.za
[b] Natural Resources and the Environment, CSIR, Meiring Naude Ave, Pretoria, 0001, South Africa – (sarchibald, gvmalt)@csir.co.za

Abstract – Remotely-sensed phenological metrics (or phenometrics) were derived from AVHRR vegetation-index time-series data and to describes seasonal growth in terms of start, end, length of season and estimates of net primary production (NPP). This study analyzed vegetation phenometrics across South Africa (SA) in order to characterize phenological patterns and their inter-annual variability. A second objective is to distinguish biomes and sub-biome "bioregions" based on functional patterns. The long term phenometrics gave ecologically-meaningful results which reflect our current understanding of the spatial patterns of production and seasonality of vegetation growth. The results suggest that phenometrics capture sufficient functional diversity to classify and map vegetation based on function.

Keywords: vegetation phenology, time-series, AVHRR, inter-annual variability,biomes, bioregions, South Africa

1. INTRODUCTION

The dynamic phenology of terrestrial ecosystems reflects response of the biosphere to proximal climatic factors (e.g. temperature and rainfall) and these climatic drivers, as well as fire, are largely responsible for the geographic distribution of different vegetation zones, e.g. biomes. Satellite-derived phenology provides the opportunity for defining and mapping vegetation zones (e.g. biomes) based on vegetation function and dynamics.

Mucina & Rutherford (2006) defines a biome as a broad ecological unit representing major life zones of large natural areas, defined mainly by vegetation structure, climate and major large scale disturbance factors (such as fires). A bioregion is viewed as a composite spatial terrestrial unit defined on the basis of similar biotic and physical features and processes at the regional scale. It is the intermediate level of vegetation organisation between that of vegetation type and biome (Mucina & Rutherford, 2006). Figure 1 shows the biomes of South Africa with the Savanna biome divided into bioregions.

The objective was to investigate the long-term spatial patterns and inter-annual variability in satellite-derived vegetation phenology in relation to the recently revised biome map of South Africa, as well as the savanna bioregions. Furthermore, to identify the phenological attributes that differentiate different biomes and bioregions.

2. METHODOLOGY

2.1 Extracting phenometrics from AVHRR data

The 1km^2 AVHRR data were previously processed and calibrated for sensor degradation. For details see (Wessels *et al.*, 2006). Daily NDVI data were composited into 10-day maximum value composites. This study was limited to the period 1985-2000 in order to avoid changes in spectral response function of NOAA-16, post 2000. A data gap exists for 1994 due to the failure of NOAA-13.

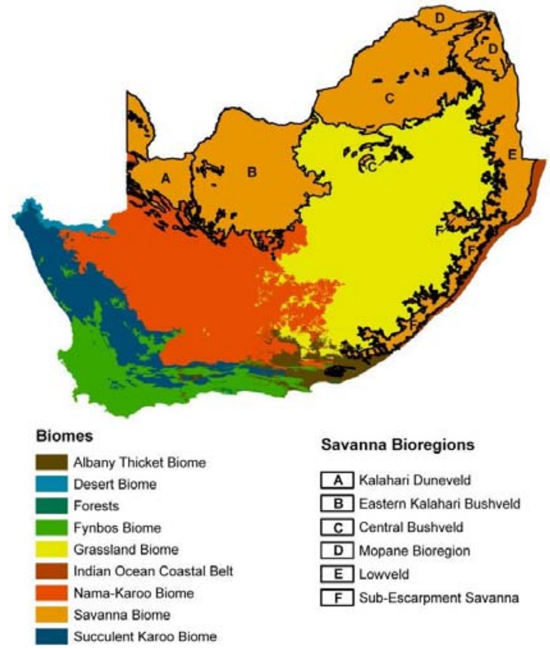

Figure 1. Biome map of South Africa with the Savanna biome divided into bioregions indicated by letters A…F (after Mucina & Rutherford 2006).

The long-term, 1km^2 NDVI data were analyzed using the TIMESAT program developed for the exploration and extraction of seasonality parameters from time-series data (Jönsson & Eklundh, 2004). An adaptive Savitsky-Golay filter, proved to be the most successful at producing a smoothed curve while capturing rapid phenological changes. Seasonality parameters (hereafter referred to as phenometrics) such as start date, end date and length of growing season were identified throughout the data set. A user-defined threshold of 20% of the seasonal amplitude (as measured from the left minima of a seasonal curve) is set as the start of growing season (SGS) date. Similarly the end of growing season (EGS) is defined as the date at which the right edge has declined to 20% as measured from the right minima.

Phenometrics for each of the growing season were extracted (Figure 2) with a distinction between "date-related phenometrics (e.g. start or end of season) and "NPP-related phenometrics" (e.g. large integral) shown in Table 1. Long-term means, standard deviations (SD) or coefficients of variation (CV) were calculated and mapped for all phenometrics (Table 1) across the periods 1985-1993 and 1995-2000 (data gap 1994).

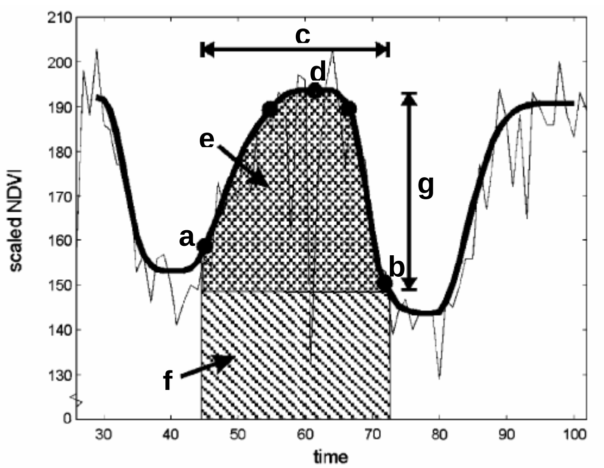

Figure 2. Phenometrics extracted from the seasonal NDVI curve, as defined in TIMESAT. See Table1 (after (Jönsson & Eklundh, 2004).

Table 1. Date-related and productivity (NPP) – related phenometrics indicated in Figure 2.

Date-related metrics	NPP-related metrics
a. Start of growing season (SGS)	d. Maximum NDVI value (MAX)
b. End of growing season (EGS)	e. Small Integral (SI)
c. Length of growth season	f. Large Integral (LI)
d. Mid position of growth season	g. Amplitude

Transformed areas such as cultivated land, plantations and built-up areas mapped in National Land Cover 2000 were excluded from further analyses which were only concerned with natural vegetation. A buffer of 1km around the transformed areas was also excluded to avoid adjacency effects. 400 pixels per biome were randomly selected from the remaining untransformed areas. The same points were used for the savanna bioregion analysis.

2.2 Phenology-based regression tree analyses

A random forest regression tree (Breiman, 2001; Prasad et al., 2006) was run using a range of phenometrics as the input variables and the biomes and savanna bioregions (Mucina & Rutherford, 2006) respectively as dependent variables. No prior probabilities were used. To see how much information was contributed by date-related versus NPP-related phenometrics (Table 1), two regression trees were applied: (i) using all the phenometrics, (ii) using date-related metrics only. The importance of the different phenometrics was assessed by comparing the results of the original model, with models run on random data for each input variable (Gini index) (Breiman, 2001). For each tree the model was re-run while randomly permuting a single input variable. The resulting random forest model was used to run a prediction which mapped the biomes and savanna bioregions based on the phenometric data which was then assessed in terms of users and producers accuracy.

3. RESULTS AND DISCUSSION

3.1 Mean and Inter-annual variability of phenometrics

The long-term mean and inter-annual variability of phenometrics were mapped and summarised per biome (Table 2).

Table 2. Summary of phenometrics per biome.

Biome	Start date	Start Date SD (decades)	Middle of season	Middle of Season SD (decades)	End Date	End Date SD (decades	Large Integral	Large Integral CV (%)
Grassland	11-20 Oct	2	11-20 Feb	2.5	1-10 Jun	2	8	15
Savanna	21-30 Sep	2.5	1-10 Feb	3	21-30 Jun	2.5	6	20
Indian Ocean Coastal Belt	1-10 Oct	2	11-20 Feb	2.5	21-30 Jun	2.5	12	20
Forest	21-31 Aug	4	1-10 Jan	2	21-31 May	4	15.5	20
Albany Thicket	11-20 Sep	5.5	21-31 Jan	5	11-20 Jun	4.5	9	35
Nama Karoo	21-31 Oct	9	21-28 Feb	5.5	1-10 Jul	8.5	4.5	35
Succulent Karoo	11-20 May	3.5	11-20 Aug	3	1-10 Jan	6	4.5	25
Fynbos	11-20 May	4	21-31 Aug	3	1-10 Feb	5	5	30
Desert	1-10 May	8	21-31 Aug	5	11-20 Feb	6	2	40

The winter rainfall area in the south western part of South Africa can clearly be distinguished by having mean start dates in May. In contrast, the growing season in the summer rainfall region starts in late September and October (Figure 3).

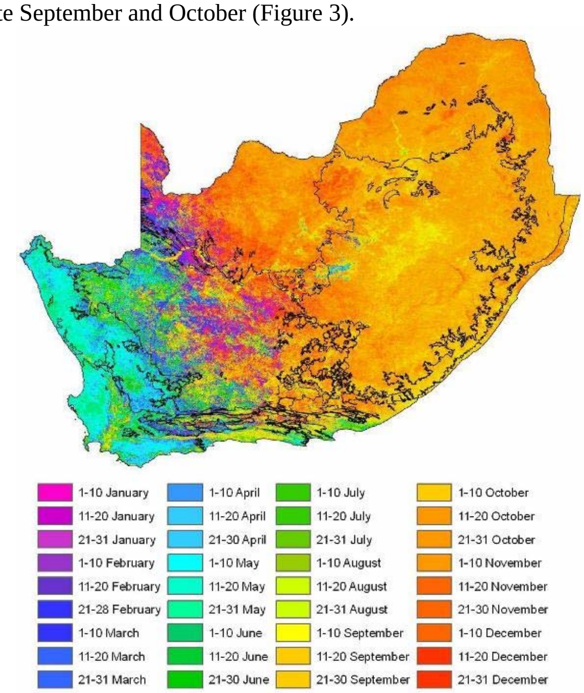

Figure 3. Mean start of the growing season (SGS).

The arid Nama Karoo and Desert biomes have the highest variability in SD SGS (Figure 4) due to highly variable rainfall. In contrast the Grassland, Savanna and Indian Ocean Coastal Belt biomes have the lowest SD for all date-related phenometrics. An area of exceptional low SD SGS can be seen in the Western Cape close to the Cape Peninsula (Figure 4). Although this is part of the Fynbos biome, this area is characterised by wheat farming with consistent planting and harvesting dates. This is in contrast with dryland agriculture in the Free State (Grassland biome) with approximately 80 days variability in SGS.

Mean LI was the lowest for the Succulent Karoo, Nama Karoo and Desert biomes ranging from 0.5-5.0 (Figure 5). The Nama Karoo and Desert biomes had LI CV values of 35% and 40% respectively, while the LI CV for Succulent Karoo was much lower at 25%. The Forest and Indian Coastal Belt biomes showed the highest mean LI values reaching 15.5 and 12 respectively (Figure 5) indicating the highest level of seasonal growth of all the biomes.

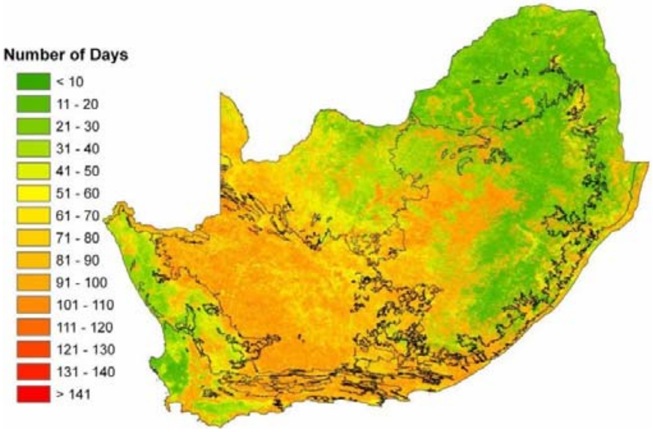

Figure 4. Standard deviation in start date (SD SGS) expressed in number of days.

Their LI CV is only 20% and show low inter-annual variability. Although Albany Thicket showed high productivity with mean LI values at 9, its inter-annual variability is high at 35%, similar to Nama Karoo. Fynbos exhibited lower mean LI values (5) and LI CV (30%) than Albany Thicket. The Grassland and Savanna biomes had mean LI values of 8 and 6 respectively and their LI CV was low at 15% and 20%.

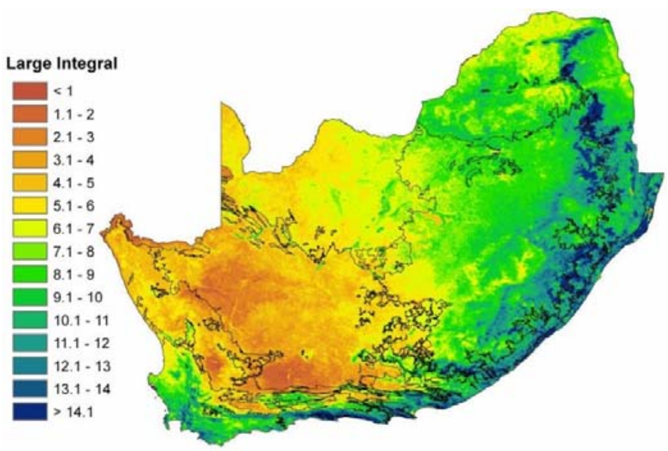

Figure 5. Mean large integral (LI) with biome outlines in black.

3.2 Regression tree analyses of phenometrics of the biomes

The random forest method produced reliable predictions from the input sample data (Table 3). Using all the phenometrics the overall prediction had an R^2 of 0.75, while R^2 values for individual biomes ranged from 0.66 to 0.90. Using date-related metrics only (Table 3), reduced the overall explanatory power by 10%.

Table 3. Accuracy of the random forest regression tree model developed on 3400 sample points from nine biomes. Values represent proportion of the sample points which were correctly classified by the predictive model.

Biome	NPP-related phenometrics	Date-specific phenometrics	All phenometrics
Desert	0.83	0.72	0.89
Succulent Karoo	0.50	0.58	0.69
Nama Karoo	0.39	0.49	0.67
Fynbos	0.43	0.43	0.66
Albany Thicket	0.47	0.56	0.70
Grassland	0.71	0.67	0.76
Savanna	0.15	0.62	0.71
Forests	0.80	0.68	0.79
Indian Ocean Coastal Belt	0.65	0.77	0.90
Total	0.54	0.61	0.75

The importance of different phenometrics in predicting biome class was analysed by calculating the Gini index. This indicated that the mean LI, mean SI, mean MAX and SGS were the most important phenometrics for distinguishing between biomes (Figure 6).

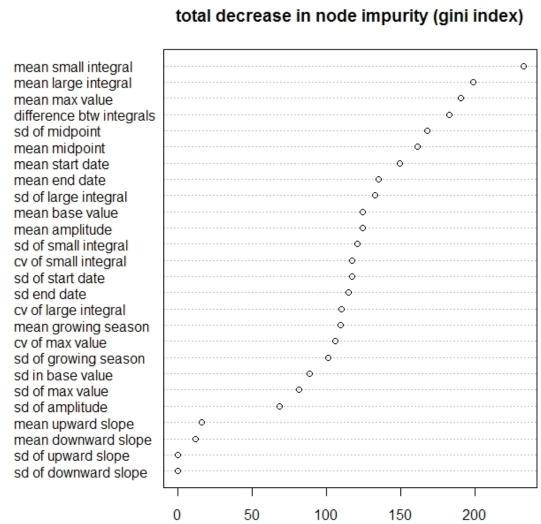

Figure 6. The importance of different phenological variables in predicting biome class, measured as the decrease in node impurity that occurs when the variable is randomly permuted.

The predicted biomes calculated from phenometrics are shown in Figure 7. It is clear that areas of confusion are associated with boundaries between biomes e.g. Desert and Nama Karoo and the Grassland and Savanna biomes, where the transition is gradual rather than abrupt. The extent of Desert was overestimated to include very dry areas of the Nama and Succulent Karoo.

3.3 Regression tree analysis of phenometrics of the Savanna bioregions

Using all the phenometrics the overall prediction had an R^2 of 0.87, while R^2 values for individual biomes ranged from 0.56 (Mopane) to 0.93 (Eastern Kalahari Bushveld)(Table 4). Using date-metrics only (Table 4) in random forests reduced the overall explanatory power by 8%.

The mean SI was the most important factor in firstly separating the arid Kalahari Bushveld and Duneveld from the wetter eastern savanna bioregions. The NPP-related mean SI, mean AMP and mean MAX were the most important phenometrics for distinguishing between bioregions (Figure 8).

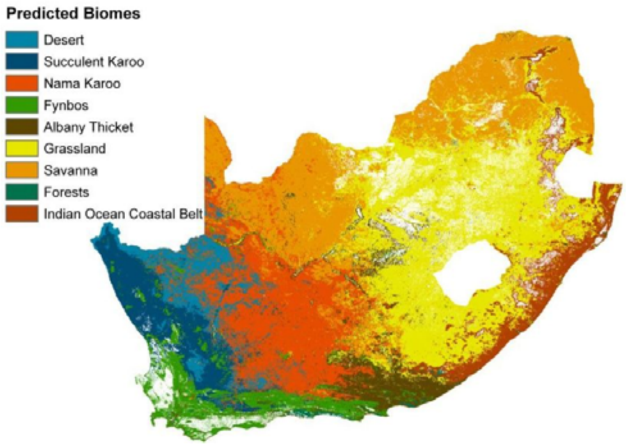

Figure 7. Biomes of South Africa predicted by the phenology-based regression tree. Transformed areas shown in white. Compare to Fig. 1.

Table 4. Accuracy of the random forest regression tree model developed on 399 sample points from six bioregions of the Savanna biome. Values represent proportion of the sample points which were correctly classified by the predictive model.

Savanna bioregion	NPP-related phenometrics	Date-specific phenometrics	All phenometrics
Kalahari Duneveld	0.87	0.89	0.93
Eastern Kalahari Bushveld	0.87	0.88	0.92
Central Bushveld	0.89	0.84	0.92
Mopane	0.59	0.38	0.56
Lowveld	0.83	0.56	0.78
Sub-escarpment savanna	0.37	0.47	0.67
Total	0.81	0.79	0.87

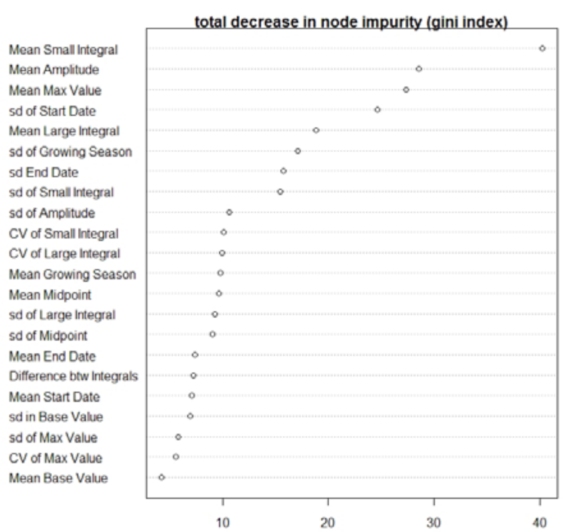

Figure 8. The importance of different phenological variables in predicting biome class.

The most important date-related phenometrics for distinguishing the bioregions were SD of SGS which split the arid Kalahari bioregions from the rest, followed by mean SGS and SD of LGS. The mean EGS was used to differentiate the deciduous Mopane bioregions from the Bushveld and Lowveld bioregions in both the date-related and NPP-related trees.

4. CONCLUSION

The phenometrics derived from remote sensing data gave ecologically-meaningful results which reflect the current understanding of the spatial patterns of production and seasonality of vegetation growth. The phenometrics captured functional processes that were not readily predictable from the combination of floristic data and climate variables alone.

Despite the fact that the biomes are internally very diverse in function, the phenology-based decision tree analysis was just as successful at distinguishing the biomes as the climate-based regression tree (Rutherford *et al.*, 2006). At sub-biome level, the phenology-based regression tree performed even better and was able to distinguish the savanna bioregions with a reliability of 87%. The split conditions derived from the phenometrics matched our understanding of the differences in functional dynamics of the biomes and bioregions. This ultimately suggests a convergence of composition, structure and function at a biome and sub-biome level.

REFERENCES

Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32.

Jönsson, P. & Eklundh, L. (2004) TIMESAT—a program for analyzing time-series of satellite sensor data. *Computers and Geosciences*, **30**, 833–845.

Mucina, L. & Rutherford, M. C. (2006) The vegetation of South Africa, Lesotho and Swaziland. *Strelitzia 19*. South African National Biodiversity Institute, Pretoria.

Prasad, A. M., Iverson, L. R. & Liaw, A. (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, **9**, 181-199.

Rutherford, M. C., Muncina, L. & Powrie, L. W. (2006) Biomes and bioregions of Southern Africa. *The vegetation of South Africa, Lesotho and Swaziland.* (ed. by L. Muncina & M.C. Rutherford), pp 32-50. Strelitzia, Cape Town.

Wessels, K. J., Prince, S. D., Zambatis, N., Macfadyen, S., Frost, P. E. & Vanzyl, D. (2006) Relationship between herbaceous biomass and 1-km^2 Advanced Very High Resolution Radiometer (AVHRR) NDVI in Kruger National Park, South Africa. *International Journal of Remote Sensing*, **27**, 951-973.