SPATIAL ESTIMATION OF HERBACEOUS BIOMASS USING REMOTE SENSING IN SOUTHERN AFRICAN SAVANNAS

Patrick Christopher Dwyer

A thesis submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in partial fulfillment of the requirements for the Degree of Master of Science.

Johannesburg, 2011

DECLARATION

I declare that this thesis is my own, unaided work. It is being submitted for the Degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

(Signature of candidate)

23rd day of May 2011

ABSTRACT

The Savanna biome covers around 60% of sub-Saharan Africa. The goods and services it provides are utilised and often depended upon by rural communities, commercial farmers and managers of conservation areas existing within it. The benefits derivable by these parties depend largely on vegetation structure and species composition which can show great variation within savannas. Fire has long been used as an effective means of manipulating savanna vegetation to maximise the provision of specific benefits, usually the provision of new herbaceous growth, and to a lesser extent to control woody cover. Information on the abundance and distribution of herbaceous biomass, which is the primary fuel source for savanna fires, has emerged as one of the most important inputs for savanna management planning. Although the most popular and reliable means of obtaining this information remains field-based sampling, estimation using remote sensing data is increasingly being incorporated into the process. Its increased popularity stems from the fact that it can greatly expand the extent of the areas for which herbaceous biomass estimations can be provided.

Although there have been studies conducted on the performance of individual remote sensing based herbaceous biomass estimation methods, few have focused on the relative performance of available methods. Information on the accuracy of methods when applied in relatively densely wooded savannas, or those where a large amount of herbaceous material is retained between seasons is also limited. This presents a problem for savanna managers in South Africa where these conditions prevail. It was the aim of this study to compare the accuracy and precision of two different remote sensing based herbaceous biomass estimation techniques (the use of a regression model and cokriging) when applied under such conditions.

To achieve this aim a large amount of herbaceous biomass data were required to form testing and training datasets. These were acquired from

the Kruger National Park's Veld Condition Assessment (VCA) datasets for the growth seasons between 2000 and 2006, which contains herbaceous biomass estimates based on disk pasture meter readings. It was suspected early on in the study that the VCA field data was not ideal for use as remote sensing (ground truthing) field data because of the limited size of the field plots relative to the pixels of the remotely sensed imagery used. It was decided to include an additional section of analysis to determine the possible contribution of this issue to the estimation error of the methods assessed. This involved measuring and comparing mean herbaceous biomass in co-located trial 60x60m VCA sites and trial 250x250m, The Moderate Resolution Imaging Spectroradiometer (MODIS) pixels.

The main section of analysis involved (i) gathering and deriving the required variables for use in the two estimation methods assessed, (ii) producing the estimates and (iii) comparing their accuracy and precision. The first method assessed was the use of a linear regression model. Seven regression models were created in total, one for each year of the growth seasons occurring between 2000 and 2006, plus another using all of the data combined. The models included variables to account for vegetation production (based on MODIS EVI), tree cover and fire history. These variables were derived using data supplied by the CSIR and Kruger National Park Scientific Services. The second method assessed was cokriging performed with the VCA herbaceous biomass field estimates as the primary variable and the MODIS EVI data as a secondary variable.

The regression models were unable to account for more than 46% of the variation in herbaceous biomass, usually accounting for between just 20 and 30% (R^2 of between 0.2 and 0.3). Three potential methods were identified that could improve the model fits obtained in the future, namely:

1. Increasing the dimensions of the field sample plots
2. Improving the calibration of the disk pasture meter used to collect the field data

3. Using EVI from previous seasons in conjunction with fire scar data to account for the presence of dry material from previous seasons.

Cokriging produced estimates that were on average 119 kg/ha more accurate than those of the regression models. However, the performance of cokriging was poorer than expected given the results of previous studies in the area. A possible explanation for this discrepancy is that the ArcGIS geostatistical analysis extension used in this study is limited in its capabilities. Even with the poorer than expected performance recorded in this study, the cokriged maps remain the best option for fire managers as they are the most accurate to date and require the fewest resources to produce. Neither method produced estimates with less than 1000 kg/ha of error (RMSE), the upper limit initially considered useful in this study. However this error limit could be considered unrealistic given the well documented high level of heterogeneity typical of southern African savannas.

ACKNOWLEDGEMENTS

The following people deserve thanks for their assistance with completing this study:

The CSIR for funding my MSc. Special thanks to Dr. Konrad Wessels, Dr. Frans van den Berg, Karen Steenkamp, Philip Frost, Waldo Kleynhans, Brian Salmon, Willem, Seare Araya, Andre Hauptfleisch and Rui Vieira of the CSIR Meraka Institute Remote Sensing Research Unit

The National Research Foundation (NRF2069152) for the grant holder bursary (via Ed Witkowski)

The University of the Witwatersrand for supplying a merit bursary

My Supervisors: Barend Erasmus, Edward Witkowski and Konrad Wessels

Members of the lab in APES who provided advice and assistance: Marco Vieira and Christopher Barichevy

The Kruger National Park Scientific Services for allowing me access to GIS and VCA data and for allowing me to conduct fieldwork. Special thanks to Navashni Govender, Dr. Izak Smit , Nick Zambatis and Sandra MacFadyen

Stella, Clare and Tessa Carthy for allowing me to live with them in Johannesburg while I completed this MSc.

Special thanks go to my Mother, Father and to Tessa for their support.

Chapter 1

INTRODUCTION

1.1. Rationale

The Savanna biome is of great importance in sub Saharan Africa. It covers 60% of the region, contains an exceptionally high diversity of both plant and animal species and is a major provider of resources that sustain both rural livelihoods and commercial activity (Van Wilgen et al. 2003; Twine et al. 2003; Shackleton et al. 2007). Savannas provide grazing for commercial and domestic livestock, medicinal plants, timber for construction and fuel wood for cooking and heating (Shackleton et al. 2007). In the majority of cases the consumption and trade of savanna resources to secure livelihoods is not a choice but a necessity, with those depending on them having few if any alternatives. In South Africa alone there are 9.2 million rural people living in and deriving direct benefits from savannas through resource extraction (Twine et al. 2003). Benefits are also derived from non consumptive use of savanna resources such as wildlife tourism. The revenue generated through tourist spending in and around the National Parks, conservation areas, game farms and various related enterprises located in savannas generates a significant portion of income in many areas (Wells 1997; Shackleton et al. 2007).

Even though all savannas contain both tree and grass layers (Archibald and Scholes 2007), the density of the woody layer and the species richness, abundance and dominant growth form in either layer can vary through both space and time (Smit 2004). Throughout this study the term ‘herbaceous biomass’ is used to refer to the biomass of both grass and forbs which collectively make up the herbaceous layer while ‘woody biomass’ refers to trees and shrubs. Variation in the above mentioned factors causes different stocks and flows of goods and services to become available. The density of the woody layer determines the availability of fuel wood and construction timber, the quality of which depends on the species present and their growth form. From a conservation perspective it affects the type of habitat available and the fauna it will support. The quality of the grazing available will depend on the species composition of the herbaceous layer and the amount of dead accumulated material persisting from previous seasons.

Management of savannas to maximise their value, be it in terms of supporting livelihoods or conservation of biodiversity, is therefore focused on manipulating factors that alter vegetation properties. Fire's ability to do just that has long been recognised and harnessed by man (Sheuyange, Oba, and Weladji 2005). The efficiency with which it enables the manipulation of vegetation properties has led to it being recognised as one of the most important tools in contemporary savanna management.

Successful prediction of the effect of fire on savanna vegetation requires among other things information on fuel load because of its role in determining fire intensity and hence a fires affect on vegetation (Trollope, Trollope and Hartnett 2002). In savannas this is provided through information on herbaceous biomass because herbaceous biomass constitutes the primary source of fuel for wildfires (Trollope, Trollope, and Hartnett 2002). Knowing how much herbaceous biomass is present and how it is distributed enables better planning of fire suppression and controlled burning activities for the achievement of management objectives (de Ronde, Geldenhuys, and Trollope 2004; Flasse et al. 2004).

The most straightforward, and often the most accurate means of attaining herbaceous biomass information is through field based methods such as clipping and weighing biomass or the use of a Disk Pasture Meter (DPM). These methods, which are covered in more detail in the literature review and methods sections, are labour intensive and best suited to the detailed assessment of herbaceous biomass within limited areas.

There are however situations in which detailed information on the spatial distribution of herbaceous biomass is required over a large area. These requirements cannot be met using a purely field based approach (Flasse et al. 2004). Indeed one of the primary motivations for this study was the interest expressed by the Kruger National Park fire management team in some means of attaining annual, spatially explicit herbaceous biomass estimates at useful levels of accuracy for the entire park (Wessels et al. 2006). This is a task not achievable using field based sampling alone (see appendix 1 of this chapter).

Two approaches present themselves for making the transition from point data to continuous data that have been explored in the literature because they were deemed appropriate for the study area. The first is making use of a measurement strongly correlated to herbaceous biomass that can be taken at every point within the area of interest without requiring excessive resources. Once obtained, the relationship between the measurements and herbaceous biomass can be established through the use of a regression analysis, and a regression model created. A number of studies have been conducted investigating the relationship between Vegetation Indices (VI's) and herbaceous biomass (Al-Bakri and Taylor 2003; Moreau et al. 2003; Prince 1991; Cayrol et al. 2000; Verbesselt et al. 2006; Sannier, Taylor and Plessis 2002; Wessels et al. 2006; Mutanga and Rugege 2006). The strength of the relationship reported varies widely, most likely due to variation in the size of the field sample plots used, variation in the complexity of the vegetation layer in the different study sites and the fact that some studies use aggregated, instead of per pixel data. All of these studies, except Mutanga and Rugege (2006) stop short of actually producing spatially explicit, per pixel fuel load maps.

To understand why a correlation between end of season herbaceous biomass and VI values exists, and why the strength of the correlation reported varies so widely, one must be clear on what VI's measure. According to (Huete et al. 2006), "Vegetation Indices (VI) are optical measures of vegetation canopy 'greenness', a direct measure of photosynthetic potential resulting from the composite property of total leaf chlorophyll, leaf area, canopy cover, and structure". They provide this measure by combining information from the chlorophyll-absorbing red spectral region with the non-absorbing, leaf reflectance signal in the near-infrared (NIR). The extent to which photosynthetic potential is realised is determined by a range of climatic and biophysical factors including the amount of incoming Photosynthetically Active Radiation (PAR), ambient temperature and available soil moisture. In other words, VI's only provide information on the upper limit of the Fraction of PAR that can be absorbed (f_{PAR}) (Huete et al. 2006), they are not a direct measure of net primary production (NPP). They are however sufficiently well correlated to NPP to have resulted in them being widely used as proxies for NPP (Huete et al. 2006).

NPP for a given growth season is in turn correlated to the amount of herbaceous biomass present at the end of that growth season. There are however numerous additional sources of variation which affect the relationship (figure 1).

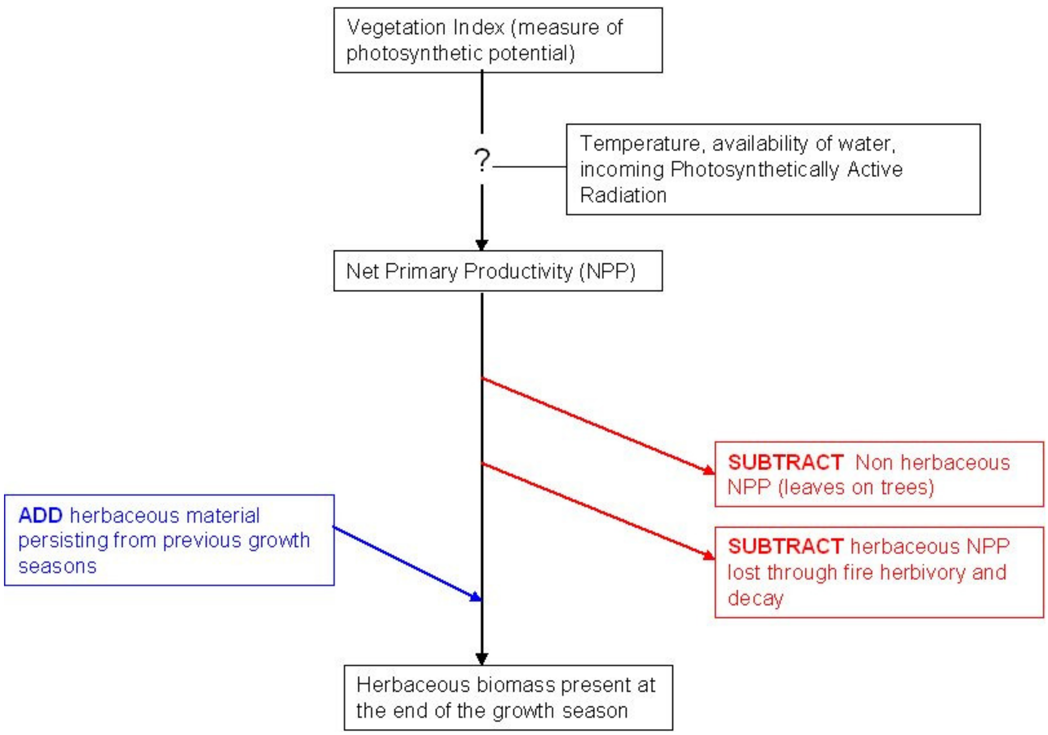


Figure 1: The relationship between Vegetation, Index values and herbaceous biomass.

The removal of herbaceous NPP through fire, herbivory and decay is constantly occurring. Some of the photosynthetic potential and resulting NPP will also be attributable to the tree layer where one is present such as in savannas. Production from previous growth seasons (also termed ‘carry-over’) which accumulates in the herbaceous layer also adds to the end of season herbaceous biomass but will not be related to the current seasons VI values. It should be clear then that the relationship between VI values and herbaceous biomass can be extremely complex, involving multiple potential sources of variation. The strength of the relationship varies considerably depending on the combination of perturbing factors existing within the location being observed.

The most widely referenced VI is the Normalised Difference Vegetation Index (NDVI). This is produced using the red and NIR (near infra-red) bands from optical sensors as follows:

$$NDVI = [\rho_{NIR} - \rho_{red}] / [\rho_{NIR} + \rho_{red}] \qquad \text{(Huete et al. 2006)}$$

There are however many variations on this formula designed to address various issues such as variation in background soil colour. One such variation, the Enhanced Vegetation Index (EVI) was developed to be implemented using the data from the Terra and Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) sensors (Huete, Justice and Van Leeuwen 1999). EVI differs from NDVI in that in addition to the red and near infrared bands, the blue band is used to overcome limitations identified in the NDVI, such as sensitivity to atmospheric interference and changes in background soil colour. It is calculated as follows:

$$EVI = 2.5 [\rho_{NIR} - \rho_{red}] / [L + \rho_{NIR} + C_1 \rho_{red} - C_2 \rho_{blue}]$$

where L is the canopy background adjustment factor, and C_1 and C_2 are the aerosol resistance weights. The coefficients of the EVI equation are L=1; C_1=6 and C_2 =7.5 (Huete et al. 2006).

The modelling approach pursued in this study is neither purely mechanistic nor is it purely statistical. Mechanistic modelling of vegetation properties is most often used at coarse continental scales, matching the resolution of the most readily available input variables such as incoming solar radiation and interpolated rainfall (see Higgins et al. (2010) for an example). Statistical modelling on the other hand is more common in the literature on localised modelling of vegetation properties such as herbaceous biomass (Mutanga and Rugege 2006; Verbesselt et al 2006; Wessels et al 2006), where the variables for mechanistic modelling are seldom available at the required resolution. At the outset of this study the intention was to pursue a basic statistical modelling approach. Preliminary results where however poor. Given the knowledge that the relationship underpinning the model varied to some extent in relation to variables already available as GIS layers, some effort was made to account for variation in production through direct adjustment of the surrogate measure of production, Vegetation Index values. The result was a statistical modelling approach with some elements of mechanistic modelling at various points in the study.

The second approach to making the transition from point data to continuous data is through the use of geostatistical interpolation. This produces estimates of herbaceous biomass at every point in the areas of interest using either the assumed or determined spatial trends in herbaceous biomass. One of the methods for determining the nature of the spatial trends in herbaceous biomass is known as Kriging (Clark and Harper 2000). The method can also be extended to make use of VI data (or any intensively sampled variable correlated to herbaceous biomass) to guide spatial estimates and increase estimation accuracy in a process called cokriging (Johnston, Sakala and Wrightsell 2001; Curran and Atkinson 1998; Mutanga and Rugege 2006).

Regardless of which of these methods is used to transform point data into continuous data, the results obtained will be affected by the characteristics of the VI data used. The MODIS sensor aboard the Terra and Aqua satellites offers' the best combination of spatial resolution, time span, temporal resolution, pixel quality information and VI products currently available for use in vegetation monitoring. Its potential and limitations therefore need to be tested and understood if remotely sensed estimations of herbaceous biomass are to be improved.

The accuracy and precision of the two methods mentioned above have only been investigated in three published studies in southern Africa (Mutanga and Rugege 2006; Verbesselt et al 2006; Wessels et al 2006). Only one of these studies, that by Mutanga and Rugege (2006) made use of MODIS data and assessed the relative accuracy of the two methods of herbaceous biomass estimation. It is also the only study to have addressed the per pixel accuracy of either method. Both Verbesselt et al (2006) and Wessels et al (2006) report correlations for aggregated data or smoothed data. Aggregated data is useful for illustrating the underlying relationships present, but is of no use in producing spatially explicit fuel load maps. Having only a single study addressing the production of spatially explicit fuel load maps makes it difficult for management agencies to make decisions confidently regarding the implementation of operational remote sensing based herbaceous biomass monitoring programs. Without more information on the relative performance of the two methods there is little information on which to base their decisions.

1.2. Aim and objectives

The aim of this study was to compare the relative accuracy and precision of cokriging and a linear regression model used to produce spatially explicit herbaceous biomass estimates from 250m MODIS VI data.

The objectives of the study were:

1. Quantify the accuracy and precision achieved when using a regression model, derived using the data currently available to the Kruger National Park, to produce herbaceous biomass estimates.

2. Quantify the accuracy and precision achieved when using cokriging, performed using the data currently available to the Kruger National Park, to produce herbaceous biomass estimates.

3. Provide a comparison of the two methods.

2. LITERATURE REVIEW

2.1. Fires in savannas

Climate, geology, fire and herbivory all interact to determine the tree grass balance in savannas. Of these, fire is the most easily manipulated and is thus a useful management tool. The changes it brings about depend largely on fire frequency and intensity and hence its effective use depends on ones ability to manipulate these components of the fire regime (Higgins, Bond and Trollope 2000).

Savanna trees are highly resilient to the effects of fire, especially fire of moderate intensity, often resisting top kill (death of the aerial biomass) because of thick cork like bark (Wilson and Witkowski 2003; de Ronde et al. 2004). They are also able to re-sprout from their base if top kill does occur (Hoffmann and Solbrig 2003; Higgins, Bond and Trollope 2000). Newly sprouted shoots and seedlings trapped within a savannas herbaceous layer are however extremely vulnerable to fire (Higgins, Bond, and Trollope 2000). They remain this way until they have grown to a sufficient height and produced sufficiently thick bark to survive frequent burns (Sankaran et al. 2005; Higgins, Bond and Trollope 2000). It is this vulnerable phase that allows a series of subsequent fires to result in significantly decreased density of the woody layer through accumulated mortalities and lowered recruitment rates by preventing trees from reaching reproductive size (Hoffmann and Solbrig 2003). In preventing trees from reaching reproductive size and killing of new seedlings, frequent fires also reduce the seed bank (Witkowski and Garner 2000) and lower future recruitment rates. From the above it is evident that both the frequency and the intensity of fires are therefore important in determining the impact of fire on the tree layer.

In contrast to the tree layer, fire intensity is less important, in terms of direct effects, than fire frequency in determining the properties of the herbaceous layer (Trollope 1996). Unpalatable grass species that might accumulate and shade out new growth or prevent recruitment of palatable species can be removed by occasional burning. Alternately their abundance can be increased by fire suppression (Van Wilgen et al. 2003; de Ronde et al. 2004). Perennial species that can reproduce by sending out sub surface runners may become more abundant than those that rely on seeds when there are frequent fires. This is because the seeds will be destroyed before becoming established (Garnier, Durand and Dajoz 2002). None of the above is especially sensitive to the intensity of the fires involved.

Although intensity has little bearing on the direct affects of fire on the herbaceous layer, indirectly it plays a significant role through affecting the density of the woody layer present. Herbaceous biomass production has been shown to be negatively related to the density of woody cover when assessed at the landscape scale (Wessels et al. 2006), although locally the reverse may be true. The relationship exists because increased woody cover reduces available light and water availability, which limits the production of herbaceous material (Savadogo et al. 2008). Because herbaceous material is the primary fuel for savanna wildfires, a decrease in herbaceous production reduces fire frequency and intensity. Frequency is decreased because fewer fires are successfully ignited and sustained given the lower fuel loads (Trollope 1996). Intensity is decreased because there is less fuel to burn. This creates a positive feedback loop. Decreased fire intensity and frequency caused by decreased herbaceous production leads to increased woody cover by allowing seedlings to escape the fire trap (Higgins, Bond, and Trollope 2000). As these seedlings grow and begin to intercept more light, they further reduce herbaceous production. In the absence of disturbance events such as the felling of trees by humans, or the damage and uprooting of trees by elephants, woody cover will increase to the limits set by climate and self shading (Smit 2005).

Even where trees are absent, herbaceous production can be reduced by shading. This occurs where dead herbaceous material (often termed moribund grass) is able to accumulate in sufficient amounts for it to shade out new herbaceous growth, decreasing production and accumulation rates. After 5 years, standing herbaceous biomass declines as dead material begins to decay faster than new material is produced (Govender, Trollope and Van Wilgen 2006). For production to resume, accumulated material needs to be removed. To summarise, prescribed burning can therefore be used to:

1. Remove dead herbaceous material and encourage new, palatable growth.
2. Encourage a reduction in the density of the woody layer through depleting the seed bank, stunting or damaging mature trees and killing seedlings.

Achieving either outcome, or their opposites, while maintaining control of prescribed burns requires information on prevailing climatic conditions, fuel moisture and fuel load as these all affect fire intensity. Broadly speaking, fuel loads of < 2000 kg/ha are insufficient for fire to spread, fuel loads of between 2000 and 4000 kg/ha produce cool to moderately intense fires of < 3000 kj/s/m and fuel loads of > 4000 kg/ha produce intense fires of >3000 kj/s/m (Trollope 1996). Fires of cool to moderate intensity will clear accumulated herbaceous material and encourage new palatable growth with little damage to mature trees. Intense fires on the other hand are likely to cause greater damage to mature trees and may reach the canopy layer, resulting in death of aerial biomass. The more detailed and accurate the information on herbaceous biomass that is available to savanna managers the greater their ability to plan, execute and achieve specific management objectives will be.

2.2. Herbaceous biomass estimation: regression

Regression models can serve two very useful purposes. Firstly, the process of creating a regression model and the model that results, provided variables are not just chosen at random, contributes to the understanding of the relationship being modelled. Secondly, once the relationship is represented as a mathematical equation it can be used to predict the value of the response variable if values for the predictor variable are available. Creating a regression model involves the following general steps:

1. Variable selection
2. Selection of model type and functional form
3. Data collection
4. Model fitting and evaluation

Each of these steps, and the corresponding information on how they have been addressed in past studies seeking to estimate herbaceous biomass using regression, are covered in more detail in the sections that follow.

2.2.1. Variable selection

Variable selection involves identifying all the variables affecting the relationship between the response and primary predictor variable as well as any important interactions between variables. Omission of variables or the interactions between variables affecting the relationship being modelled results in unexplained variation and error in predictions.

The simplest model possible for estimating herbaceous biomass in this study could contain just two variables, herbaceous biomass as the response variable and some form of VI variable as the predictor variable. Data to calculate VI's can be obtained from any optical sensor that records information from the red and near-infrared portions of the spectrum.

Although any optical imagery with the appropriate bands can be used to create VI's, the production of herbaceous biomass estimates for large areas is most easily accomplished using low or medium resolution imagery from a sensor and platform because of their high temporal resolution. This will provide regular and complete coverage of the area of interest required to monitor vegetation growth throughout a season. Historically the best source of such data has been the Advanced Very High Resolution Radiometer (AVHRR) sensor. This lead to its use in many herbaceous biomass and primary production estimation studies (Al-Bakri and Taylor 2003; Fensholt and

Sandholt 2005; Moreau et al. 2003; Prince and Tucker 1986; Tucker et al. 1985; Wessels et al. 2006). At the time of writing AVHRR's successor, the MODIS sensor aboard the Aqua and Terra satellites, has become the preferred source for such data as it offers much improved spatial and radiometric resolution (Anaya, Chuvieco and Palacios-Orueta 2009; Fensholt et al. 2006; Grigera, Oesterheld and Pacin 2007).

Both single VI images and summations of all the images within a growth season have been used in past studies. A single image can only provide information on the amount of photosynthetic potential at the time of acquisition (Funk and Budde 2009). This measure is only sensitive to the presence of *live* vegetation at a single point in time. True end of season biomass cannot be reliably inferred using a single season image because the end of the growth season only occurs once vegetation has dried out. Under these conditions the characteristics of vegetation VI's were designed to be sensitive to, primarily absorption in the red portion of the spectrum by chlorophyll, are absent or severely reduced in the herbaceous layer (Huete, Justice and Van Leeuwen 1999; Todd, Hoffer and Milchunas 1998).

It may be possible to work around this by using an image from earlier in the season when the vegetation is still green. If this is done, the problem of which point in the season the image should be acquired for then arises. Because a single image cannot account for vegetation which has dried out at any prior point in the season, it would be optimal to locate the image at the height of vegetation activity before the grass has begun to dry out. Not all regions in a study area will however experience maximum active vegetation levels simultaneously (Thein et al. 2008). The best possible solution, if using a single image, is to select the time period corresponding to mean peak in the presence of active vegetation for the study area for the year of interest. Error will still result from those areas when peak vegetation activity falls either side of the mean.

Summations of all the images within a growth season provide a measure of photosynthetic potential that existed during the growth season, rather than at a point in time. This approach is reported to maximise the herbaceous biomass – NDVI correlation within the study area (Verbesselt et al. 2006). Identification of appropriate images to include in a summation is complicated by the fact that the onset of rainfall events which trigger this activity is highly variable both spatially and temporally (Archibald and Scholes 2007). Summation of the VI over periods when photosynthetic potential is low has the potential to weaken its correlation to standing biomass through introducing noise to the VI signal. There is limited evidence that by basing the period of the summation on phenological cues derived from VI data a stronger correlation can be achieved between NDVI and crop biomass (Funk and Budde 2009) although to date this has not been tested for herbaceous biomass in savannas.

However, as outlined in figure 1, photosynthetic potential is not directly related to herbaceous biomass accumulation. Removal through herbivory is constantly occurring (Hely et al. 2003b). This removal is sufficient to have been identified as a possible source of error when using a VI summation to predict herbaceous biomass in the study area (Verbesselt et al. 2006). Incorporating the effects of herbivory into a model would require information on grazer distribution and abundance and herbaceous biomass consumption (Hely et al. 2003b). Accurate information on the distribution of large herbivores is difficult to obtain for the study area because of its size and the absence of internal divisions restricting animal movement. Acquiring such data would require extensive field work, the quality of which would ultimately be limited by cost and logistical constraints. No studies attempting to account for herbivory could be found to provide information on how best to do so or the improvements in estimation accuracy achievable.

Vegetation index values provide information on total photosynthetic potential, which includes the potential of both woody and herbaceous vegetation (Archibald and Scholes 2007). Only the portion of the signal relating to herbaceous production is of interest when predicting herbaceous biomass. One way to deal with this is the introduction of additional variables and interactions to account for the mixed signal. Alternately the signals can be unmixed and only the herbaceous component made use of. A number of studies have been conducted on ways in which the two signals can be unmixed (Lu et al. 2003; Scanlon et al. 2002; Archibald and Scholes 2007). These methods have however not been applied prior to the use of VI data in any of the attempts to estimate herbaceous biomass encountered in the literature. Successful implementation of such methods would do away with the need for additional variables and interactions. Fuller, Prince and Astle (1997) found that the issue of mixed signal can be ignored when few trees are present, and reflectance from the herbaceous layer dominates the VI signal during the growing season. Their study was however carried out in an area with limited woody cover. In areas where woody cover is in excess of 20% (Prince 1991b) and herbaceous production is limited, the herbaceous layer no longer dominates the signal and the woody cover needs to be accounted for. Given that 75% of the Kruger National Park has a woody crown cover of between 20% - 40% (Eckhardt, van Wilgen and Biggs 2000), it is possible that the contribution of the woody layer to VI values is significant. (Sannier, Taylor and Plessis 2002b) found that sample sites in areas with high wood cover constantly fell below the regression line fitted to their data. The effect was noticeable at 30% woody cover but became far more pronounced when it exceeded 60%. This indicates that for the same level of herbaceous biomass VI values will be significantly greater in heavily wooded areas. The relationship between VI data and herbaceous biomass therefore varies with changes in woody cover (Wessels et al 2006). This could be accounted for by adding an interaction term between the VI variable and a tree cover variable. Hely et al. (2003b) avoid the need for complex interaction terms by adjusting the VI data prior to analysis by penalising it based on canopy cover. Anaya, Chuvieco and Palacios-Orueta (2009) on the other hand adjust their herbaceous biomass estimates post production using a woody cover variable. At the time of writing there does not appear to be any consensus on which approach is best.

Four approaches can therefore be seen to exist: 1) ignore the issue if herbaceous cover dominates the signal, 2) un-mix the signal prior to analysis, 3) penalise the signal prior to analysis or 4) include a woody cover term in the regression model specifying an interaction between it and the VI variable.

There have not been any published studies to date comparing the effectiveness of the different approaches.

There also exists a negative relationship between herbaceous biomass and woody cover. Wessels et al. (2006), working in the KNP, investigated the affect of adding a percentage tree cover variable to the regression model without an interaction term which would account for this affect. This resulted in a 10% improvement in fit.

In southern African savannas the fuel load at the end of a season consists not only of that season's growth (less removal by herbivory and fire), but also of all dry, dead vegetation persisting from previous season's growth. This retained growth is also known as carry-over (Smit 2005). Studies conducted on Drakensberg highland sourveld indicate that after 3 years of being left un-burnt carry-over makes up between 60% and 90% of the herbaceous layer (Thompson and Everson 1993). Regression of NDVI against standing biomass under these conditions has produced an R^2 ranging from 0.003 to 0.28. In contrast carry-over in annually burnt veld accounts for between 0% to 10% of the herbaceous layer and regressions yielded an R^2 of between 0.55 and 0.79 (Thompson and Everson 1993). Similar differences in correlation between grazed and un-grazed sites were reported by Todd, Hoffer and Milchunas (1998) when working in the short-grass steppe of Eastern Colorado. They found that the R^2 for the regression predicting herbaceous biomass using NDVI for grazed sites was 0.66, whereas no significant relationship was found between NDVI and biomass on un-grazed sites. Clearly a variable to account for senesced material should be included in a model created for the study area as carry-over forms a significant percentage of the herbaceous layer and therefore potential fuel for wild fires.

2.2.2. Selection of model type and functional form

Having identified the variables to be included, a model type needs to be selected. There are a number of different model types that can be used with the most appropriate choice depending on the relationship being modelled and the characteristics of the available data. Past studies have found that a linear multiple regression model is appropriate for modelling the relationship between biomass and VI's (Wessels et al. 2006; Todd, Hoffer and Milchunas 1998; Prince 1991a). Given the past success with simple linear models for estimating herbaceous biomass there seems little reason to adopt the use of anything more complex.

2.2.3. Data Collection

The next step in model creation is data collection. There is a fair amount of variation in how data is collected and processed prior to analysis in the studies published in the literature. The data collection and pre processing methods most commonly encountered in the literature are described in the subsections that follow.

Ideally field data should be accurately measured and reflect the variation in herbaceous biomass within the pixels it will be assigned to. Measurement accuracy depends on the method used. The two most common methods are clipping and weighing of herbaceous biomass and the use of a Disc Pasture Meter (DPM). Clipping and weighing involves clipping and weighing all of the herbaceous material within numerous quadrates at each sample site (Hely et al. 2003). The quadrates are usually 0.25 m^2 to 1m^2 and the number used dependent on the variability of the herbaceous layer and size of the sample site. Because the method is labour intensive it is most suited to situations where data quality is more important than data quantity. A DPM comprises an aluminium disk, with a hole in the centre to which a section of aluminium pipe is fitted and a rod with graduations on it is threaded through the pipe to measure disc suspension height (Figure 2).

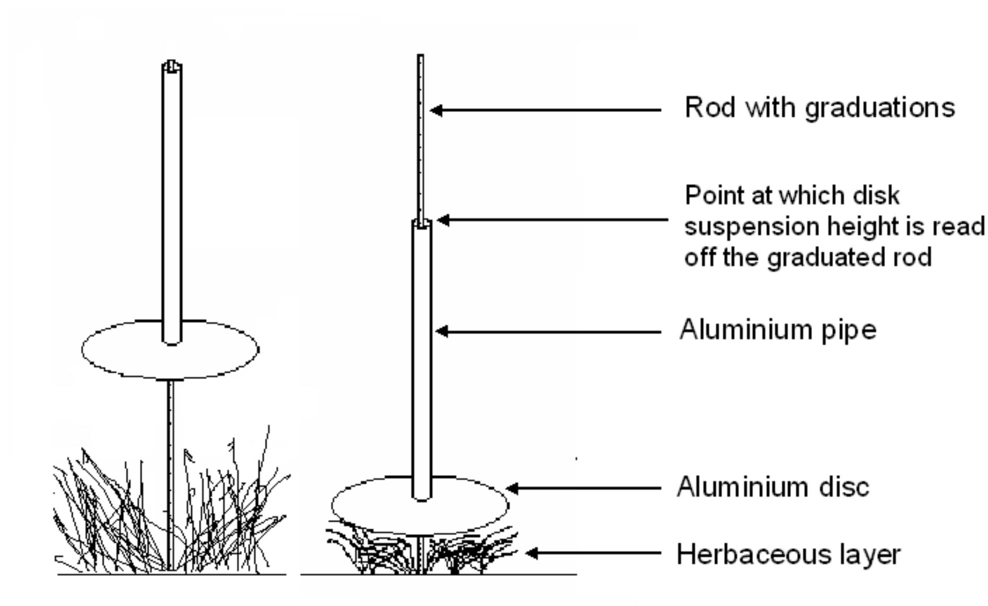


Figure 2: Operation of a pasture meter.

The pipe is slid up the rod so that their tops are level, the rod is then held upright with its end in contact with the ground and the pipe is released allowing the disk to fall. The height at which the disk is suspended is then read off the graduated rod and recorded.

Before the measurements from a pasture meter can be related to biomass the pasture meter must be calibrated (Sanderson et al. 2001). This involves gathering sets of co located pasture meter readings (height at which disc is suspended) and direct measurements of the herbaceous layer attained by clipping and weighing. A linear regression model is then created to enable herbaceous biomass to be estimated based on disk height. The biomass estimates produced in this way are often incorrectly treated and/or referred to as measurements. They are in fact estimates which have an error of more than 20% (Trollope and Potgieter 1986). The original calibration performed for the pasture meters used in the Kruger National Park, carried out by Trollop and Potgieter (1986), had a prediction error of $\pm$898 kg/ha. When the calibration was performed, mean herbaceous biomass for the study area was 3826 kg/ha, which means that the estimates produced had an error of $\pm$23%. This is comparable to the 25% error recorded by Sanderson *et al.* (2001)

when assessing the accuracy of a disk pasture meter calibrated for cultivated pasture in the United States. Accuracy is likely to decrease when estimating herbaceous biomass outside of the original calibration area. This is because the model lacks variables to account for the differences in the biomass / suspension height relationship caused by the differences in herbaceous species composition, vegetation condition and many other factors that vary between areas (Sanderson et al. 2001). Because the accuracy depends primarily on the model created during calibration, increasing accuracy requires an improved model. Variables could be added to account for vegetation type or separate calibrations performed for each. The presence of large amounts of dry material has also been shown to affect the relationship, which is more difficult to account for in the model created (Trollope and Potgieter 1986).

The accuracy of the measurement instrument is not the only factor which needs to be considered. It is also important to ensure that either the area being measured is the same for all the variables being used or is representative of that area. A slight mismatch in the measurement areas is less important when spatial variation in the property of interest is low and occurs at broad scales than when it is high and occurs over shorter distances. Variation in herbaceous biomass is ultimately controlled by the effect of topographic variation, disturbance and herbivore density on herbaceous production and accumulation (Augustine 2003). The more constant the mean and variance of herbaceous biomass within the area covered by a pixel the more limited the field sampling needs to be while still accurately reflecting the mean biomass within the pixel. The reverse is also true. The greater the variance in herbaceous biomass within a pixel the more extensive the sampling will need to be. Pixel 14 in Figure 3 illustrates just such a situation.

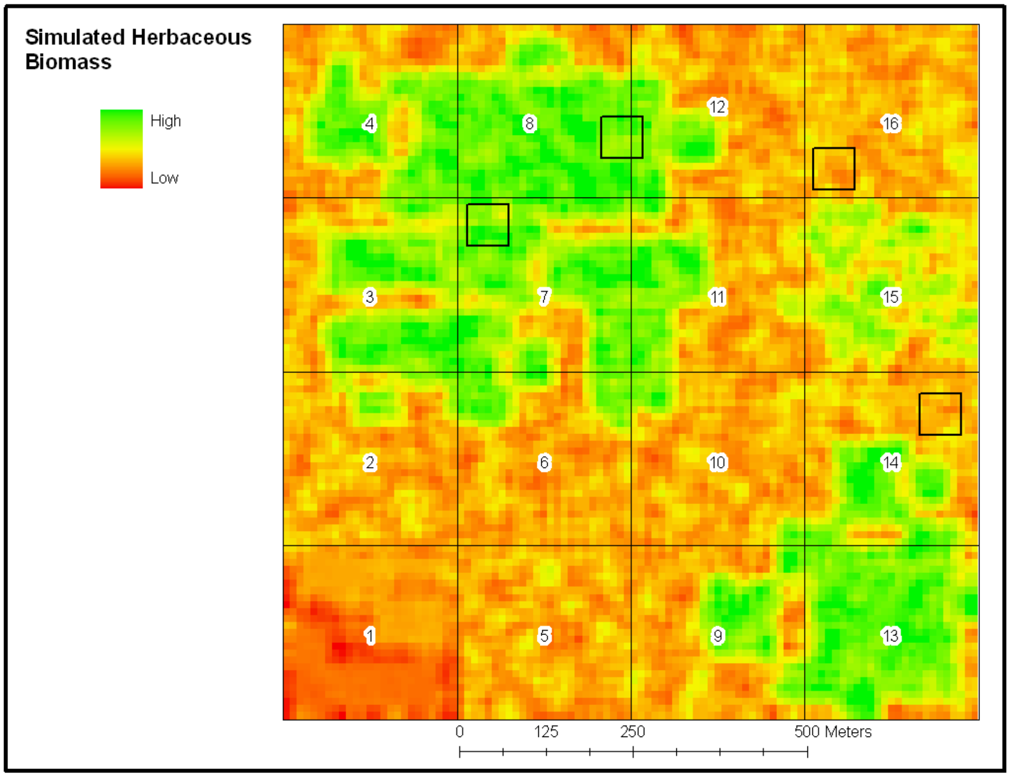


Figure 3: Simulated biomass raster overlaid with a 250m grid to reflect possible location of MODIS pixels. The smaller squares represent potential locations for 50x60m sample sites.

If the mean value from a 60x60m site were used it would differ greatly from the actual mean biomass within the area sampled by the MODIS pixel it would be matched to. Alternately if the sites and corresponding pixels occur as in pixels 7,8 and 16 in Figure 3, there would not be an issue.

To avoid encountering these issues, herbaceous biomass field measurements should be taken over areas equal to the size of the pixel they are to be matched to, if not larger (Sannier, Taylor and Plessis 2002). A common approach is to use a 1km transect located within a homogenous area. Multiple clipping or pasture meter readings are then taken either side of the length of this transect (Prince 1991; Sannier, Taylor and Plessis 2002; Moreau et al. 2003). Use of much smaller transects and sample sites is resorted to when only historical datasets, not designed for comparison with remotely sensed data, are the only ones available (Wessels et al. 2006; Sannier, Taylor and

Plessis 2002). As outlined above, measurements from smaller field sites can still be representative of the surrounding area, and the error introduced will be minimal if herbaceous biomass is fairly homogenous at scales larger than the pixels being used. Wessels et al. (2006) made use of herbaceous biomass measurement from the VCA (veld condition assessment) dataset maintained by the Kruger National Park. The dataset contains various vegetation measurements taken annually at over 500 sites across the park as well as herbaceous biomass estimates based on disk pasture meter measurements. Wessels et al. (2006) screened the VCA sites using the level of local heterogeneity in vegetation as measured by Landsat ETM NDVI in the areas surrounding VCA sites. This was done to exclude 60 x 60m sites located in areas too heterogeneous for use with 1km AVHRR pixels. The assumption made was that if there was a negative correlation between the level of variation in LANDSAT NDVI around a VCA site and the strength of the temporal relationship between biomass at the site and AVHRR NDVI, then the VCA sites were not representative of the surrounding area. They found that there was no correlation between the variation in LANDSAT NDVI within 700m of most VCA sites and the strength of the temporal relationship between biomass and NDVI within the study area. This suggests that patches of relatively homogenous biomass much larger than 700m in diameter exist resulting in most sites falling completely within such patches. This would result in 60 x 60m sample sites accurately characterising the mean herbaceous biomass sampled by a pixel. It should also be noted that the above approach detects variation in live material and not dead material because NDVI is only sensitive to green vegetation. Direct field based measurements would be required to provide a definitive answer to the variability question.

Up to this point only the quality of field based measurements has been discussed. The quality of the VI data used is also of great importance. End users have less control over this than the quality of other data sources. This is because cloud, atmospheric interference, and data acquisition gaps all reduce the useful information content of VI data, yet cannot be determined by the user (Kerr and Ostrovsky 2003). The only option available to a user wishing to avoid such issues is post acquisition processing. Cloud contaminated or atmospherically perturbed pixels can be excluded from the dataset or replaced with estimates based on temporal and / or spatial interpolation. Noise in VI signals caused by cloud contamination and other issues is most often negatively biased, causing dips in the time series profile (Thein et al. 2008). Fitting a curve that smoothes over these negative biases in VI time series data, for pixels where these issues are known to exist, and generating new values based on this curve, will minimise this noise (Thein et al. 2008). Although similar methods for pre-processing of VI data to account for cloud contamination and atmospheric interference is fairy common, no studies quantifying its effect on herbaceous biomass estimation were found.

2.2.4. Model fitting and evaluation

Ordinary least squares (OLS) regression is the standard method for fitting a regression line to data and is available in all statistical software packages. It is the method used in almost every simple statistical study published in the literature and therefore was adopted for use in this study.

There are a number of statistics that can be used for evaluating the performance of a model. The coefficient of determination, displayed as R^2 values, is one of the most commonly used statistics as it provides a relatively straightforward and easily interpretable measure of how well the model fits the data. It does so by representing the proportion of variance accounted for by the model.

$$R^2 = 1 - \frac{SS_{error}}{SS_{total}}$$

As such, it provides a simple means of evaluating how well a model fits the data. An R^2 of 0.2 for example can be interpreted as indication that the model to which it applies accounts for 20% of the variation in the data which it was created to describe. The major limitation associated with R^2 as the basis for comparing model fit is the fact that it increases as additional explanatory variables are added, regardless of whether there is a real correlation to the dependent variable. For this reason simple R^2 is a less reliable measure of fit for comparing models with differing numbers of explanatory variables. This is especially the case when the sample size is relatively small and the number of explanatory variables large,

Adjusted R^2 goes some way towards addressing the limitation of R^2 by penalising R^2 based on the number of explanatory variables used. The formula for calculating adjusted R^2 is:

$$adjR^2 = 1 - \frac{(n-i)\cdot SSE}{(n-k)\cdot SST}$$

Where n is the number of observations, *i* =1 if there is an intercept and *k* = the number of predictors + *i*. The difference between R^2 and Adjusted R^2 becomes far less pronounced as the *n* increases in very large samples.

Adjusted R^2 is a very basic metric on which to base model selection compared to more advanced metrics such as (AIC). It was used in this study despite its limitations for a number of reasons. The first was the absence of any mention of more advance methods in determining whether adding a variable to a model is acceptable in any of the literature consulted. The second was that extremely large samples and out of sample model verification were used in this study. Both increasing sample size relative to the number of predictor variables and using out of sample model verification improve the reliability of adjusted R^2 as a model selection metric. It was not used as the primary means by which the most promising model was selected. It was instead used as a means of rejecting variables which did not result in an increase in adjusted R^2, an indication that the addition of those variables was of no real value. Models containing variables that did result in an increase in Adjusted R^2 were compared based on their Root Mean Square Error (RMSE).

RMSE, literally the square root of the average value of the squared residuals (Willmott and Matsuura 2005), was selected because the figure it returns is in the same units as the models predictions, and is therefore more accessible than Adjusted R^2. It was also used as the statistic to facilitate the comparison of the two methods assessed in this study.

2.3. *Herbaceous Biomass Estimation: Cokriging*

The primary function of cokriging, as with any form of interpolation, is the prediction of values for the property of interest where no measurements have been taken (Krivoruchko 2009). All interpolation methods achieve this by assigning unknown locations values based on surrounding known values. The major difference between interpolation methods lies in how the relative contributions of the known points are determined (Clark and Harper 2000). Kriging, which forms the basis of cokriging, exploits the spatial autocorrelation inherent in the property to inform these weightings. Spatial autocorrelation simply refers to the tendency for things/objects spatially closer together to be more similar than things/objects further apart. Inverse distance weighting uses similar assumptions in that it assigns greater weight to those points closer to the point of estimation (Johnston et al. 2001). It is unlikely however that the nature of spatial autocorrelation in herbaceous biomass could be adequately characterised by a linear function derived from the inverse of the distance between points in a savanna.

Kriging allows for a better approximation of the nature of spatial autocorrelation by modelling the change in semivariance of a property through space. Semivariance refers to half of the squared difference between the value of a property measured at two points (Clark and Harper 2000). This is calculated for all possible point pairs and the values assigned to groups or ‘bins’ according to their separation distance e.g. the semivariance of points separated by between 0 and 20m, 20 and 40m, 40 and 60m, etc., the size of the bins is referred to as the ‘Lag’. The average semivariance is calculated for the point pairs in each bin and this value plotted against the distance corresponding to the centre point of that bin. The resulting plot is referred to as a semivariogram, although many authors simply refer to it as a variogram which has lead to considerable confusion (Clark and Harper 2000). The nature of the autocorrelation in a property is then approximated by fitting one of the standard autocorrelation models, such as the Spherical model (figure 4) to the semivariogram.

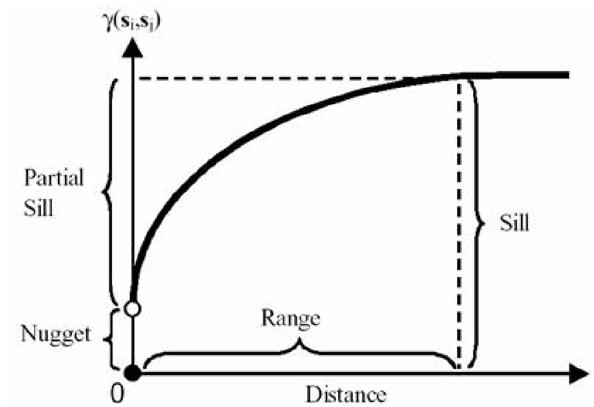


Figure 4: Spherical semi variogram and its associated parameters (http://planet.uwc.ac.za/nisl/GIS/spatial/chap_1_41.htm)

The Range value indicates the distance at which point measurements cease to exhibit spatial autocorrelation. The Sill value indicates what the variation in the sample population is beyond the range of autocorrelation. The Nugget value indicates both measurement error and the amount of variation occurring at scales finer than that of the field sample spacing (Clark and Harper 2000). The parameters values for the model can be arrived at through automated iterative fitting programmed to minimise the sum of the squared residuals by stepping through a range of values for each model parameters or through subjective fitting by the user. For more information on automated iterative fitting see the documentation for the 'sgeostat' package available at http://cran.r-project.org/web/packages/sgeostat/index.html.

Kriging can be extended to take advantage of the autocorrelation inherent in a second more intensively sample cross correlated variable in a process known as cokriging (Johnston, Sakala and Wrightsell 2001; Curran and Atkinson 1998). 'Cokriging accounts simultaneously for the autocorrelation in each variable, represented by the variograms and the crosscorrelation between the variables, represented by the crossvariograms' (Curran and Atkinson 1998). The stronger the cross correlation the greater the increase in prediction accuracy will be. Cross correlation in this context simply means the correlation between the primary and secondary variable in cokriging and crossvariograms means the variogram for the secondary variable.

Successful execution of cokriging involves the following steps:

1. Selection of an appropriate secondary variable

2. Data collection

3. Fitting of a standard model to the semi-variogram

4. Accuracy assessment of estimates produced

Only one study on the cokriging of herbaceous biomass was found for the whole of southern Africa. The study was conducted by Mutanga and Rugege (2006) in the Kruger National Park. The same study also provides the only available comparison of a regression based, kriging and co kriging approach to herbaceous biomass estimation for the region.

To identify the most appropriate secondary variable the authors regressed VI's, as well as individual MODIS bands used to calculate the indices, against herbaceous biomass estimates. They found MODIS band 2 (841–876nm, referred to as near infrared (NIR) to be the best correlated to biomass data, far better correlated than NDVI or the other VI's used. At first glance this is in conflict with most other studies published on relating remotely sensed data to plant biomass. The satellite data used in the study was however a single MODIS mod13 16 day composite corresponding to the beginning of July 2004, well into the dry season. Knowing that most vegetation activity in the region ceases during the dry season, especially in the herbaceous layer, the results make more sense as NDVI is insensitive to dry material (Thompson and Everson 1993).

Data used in the above study was taken from the KNP's Veld Condition Assessment (VCA) dataset which provided 463 samples spaced on average 1km apart (Mutanga and Rugege 2006). These samples are taken annually on between 450 and 500 50x60m sample plots using a disc pasture meter calibrated for the area. No information is available as to how optimal or sub optimal the VCA sample scheme is for use in kriging or how accuracy would be affected by an increase or decrease in sample intensity and site dimensions. It has however been noted by other authors that increasing the number of sample sites and decreasing the size of the sample plots increases the precision of kriging (Xiao et al. 2005).

The semi variogram models for kriging and cokriging in Mutanga and Rugege (2006) were arrived at by manual iterative alteration of the parameters (model form, total sill, range and nugget), obtained by an initial visual estimate of what would be optimal given the semi variogram plotted. The best model created using this approach was identified by comparing goodness of fit produced by all of the subjective model fittings. An alternative offered by some software is to obtain parameters through one subjective fitting and then allow a least squares iterative fitting algorithm to optimise those parameters (Rossiter 2007).

Accuracy assessment of kriged estimates can be performed either by validation or cross validation. Validation requires two sets of data one for creating the model and the other for assessing its accuracy. Mutanga and Rugege (2006) split the available VCA data assigning 75% to the training dataset and 25% to the testing dataset. Cross validation on the other hand does not require pre splitting of the data. Instead a single point is removed and used as validation data over a number of iterations or 'folds' and the average validation statistics calculated. This method is known to slightly inflate accuracy figures but is useful if insufficient data is available for conventional validation (Johnston et al. 2001).

Mutanga and Rugege (2006) found kriging, cokriging and regression based estimates to have RMSE's of 1008, 830 and 1374 kg/ha respectively when applied using the 2004 VCA herbaceous biomass field estimates and MODIS band 2 near infrared reflectance from a 16 day composite image corresponding to July 2004 as secondary data. This needs to be interpreted in light of the fact that herbaceous biomass at the end of the 2003 – 2004 growth season varied between 42 kg/ha and 9655 kg/ha, with an average of 3796 kg/ha and a standard deviation of 1628 kg/ha.

The herbaceous biomass – near infrared reflectance relationship produced an R^2 of 0.44. This was sufficient to provide the178 kg/ha improvement in cokriging accuracy over ordinary kriging recorded above. The spatial trends, as captured by the kriging model, produced estimates that were 366 kg/ha more accurate than the reflectance – herbaceous biomass relationship derived using regression modelling. By exploiting a combination of both the spatial patterns in herbaceous biomass and the correlation between reflectance and herbaceous biomass, cokriging was able to deliver a 544 kg/ha increase in estimation accuracy over a simple regression model. Although these results suggest that cokriging offers significant advantages over simple regression, the study used data from only a single growth season, providing no insight into whether similar results would arise given a different seasons data,

In this chapter the aim and objectives of this study have been laid out. A brief overview of the importance of information on herbaceous biomass and a brief introduction to remote sensing based herbaceous biomass estimation methods have also been provided for the reader. In the next chapter the methods and materials used in this study will be looked at in greater detail and their advantages and disadvantages discussed.

3. REFERENCES

Al-Bakri, J.T., and J. C. Taylor. 2003. Application of NOAA AVHRR for monitoring vegetation conditions and biomass in Jordan. *Journal of Arid Environments* 54: 579 - 593.

Anaya, A., Emilio Chuvieco, and Alicia Palacios-Orueta. 2009. Aboveground biomass assessment in Colombia: A remote sensing approach. *Forest Ecology and Management* 257, no. 4 (February 20): 1237-1246. doi:doi: DOI: 10.1016/j.foreco.2008.11.016.

Archibald, S., and R.J. Scholes. 2007. Leaf green-up in a semi-arid African savanna-separating tree and grass responses to environmental cues. *Journal of Vegetation Science* 18, no. 4: 583-594.

Augustine, D.J. 2003. Spatial heterogeneity in the herbaceous layer of a semi-arid savanna ecosystem. *Plant Ecology* 167, no. 2: 319-332. doi:10.1023/A:1023927512590.

Cayrol, P., A. Chehbouni, L. Kergoat, G. Dedieu, P. Mordelet, and Y. Nouvellon. 2000. Grassland modeling and monitoring with SPOT-4 VEGETATION instrument during the 1997-1999 SALSA experiment. *Agricultural and Forest Meteorology* 105, no. 1: 91-115.

Clark, I., and W. V. Harper. 2000. *Practical geostatistics 2000*. Ecosse North America.

Curran, P. J., and P. M. Atkinson. 1998. Geostatistics and remote sensing. *Progress in Physical Geography* 22, no. 1: 61.

Eckhardt, H. C., B. W. Wilgen, and H. C. Biggs. 2000. Trends in woody vegetation cover in the Kruger National Park, South Africa, between 1940 and 1998. *African Journal of Ecology* 38, no. 2: 108-115.

Fensholt, R., and I. Sandholt. 2005. Evaluation of MODIS and NOAA AVHRR vegetation indices with in situ measurements in a semi-arid environment. *International Journal of Remote Sensing* 26, no. 12: 2561-2594.

Fensholt, R., I. Sandholt, M. S. Rasmussen, S. Stisen, and A. Diouf. 2006. Evaluation of satellite based primary production modelling in the semi-arid Sahel. *Remote Sensing of Environment* 105, no. 3: 173-188.

Flasse, S.P., S.N. Trigg, P.N. Ceccato, A.H. Perrymann, A.T. Hudak, M.W. Tompson, B.H. Brockett, et al. 2004. Remote Sensing of Vegetation Fires and its Contrabution to a Fire Magement Information System. In *Wildland Fire Management Handbook for Sub-Saharan Africa*, ed. J.G. Goldammer and C. de Ronde, 158-211. Global Fire Monitoring Center.

Fuller, D. O., S. D. Prince, and W. L. Astle. 1997. The influence of canopy strata on remotely sensed observations of savanna-woodlands. *International Journal of Remote Sensing* 18, no. 14: 2985-3009.

Funk, C., and M. E. Budde. 2009. Phenologically-tuned MODIS NDVI-based production anomaly estimates for Zimbabwe. *Remote Sensing of*

Environment 113, no. 1: 115-125.
Garnier, L. K. M., J. Durand, and I. Dajoz. 2002. Limited seed dispersal and microspatial population structure of an agamospermous grass of West African savannahs, Hyparrhenia diplandra (Poaceae). *American journal of botany* 89, no. 11: 1785.
Govender, N., W. S. W. Trollope, and B. W. Van Wilgen. 2006. The effect of fire season, fire frequency, rainfall and management on fire intensity in savanna vegetation in South Africa. *Journal of Applied Ecology* 43, no. 4: 748-758.
Grigera, G., M. Oesterheld, and F. Pacin. 2007. Monitoring forage production for farmers' decision making. *Agricultural Systems*, no. 94: 637-648.
Hely, C., S. Alleaume, R. J. Swap, H. H. Shugart and C. O. Justice. 2003a. SAFARI-2000 characterization of fuels, fire behavior, combustion completeness, and emissions from experimental burns in infertile grass savannas in western Zambia. *Journal of Arid Environments* 54, no. 2 (June): 381-394. doi:doi: DOI: 10.1006/jare.2002.1097.
Hely, C., P. R. Dowty, S. Alleaume, K. K. Caylor, S. Korontzi, R. J. Swap, H. H. Shugart, and C. O. Justice. 2003b. Regional fuel load for two climatically contrasting years in southern Africa. *Journal of Geophysical Research* 108: 8475.
Higgins, S.I., W.J. Bond, and W.S.W. Trollope. 2000. Fire, resprouting and variability: a recipe for grass–tree coexistence in savanna. *Journal of Ecology* 88, no. 2 (April): 213-229.
Higgins, S.I., Scheiter, S. & Sankaran, M. (2010). The stability of African savannas: Insights from the indirect estimation of the parameters of a dynamic model. *Ecology*, 91, 1682-1692
Hoffmann, William A., and Otto T. Solbrig. 2003. The role of topkill in the differential response of savanna woody species to fire. *Forest Ecology and Management* 180, no. 1 (July 17): 273-286. doi:doi: DOI: 10.1016/S0378-1127(02)00566-2.
Huete, A., C. Justice, and W. Van Leeuwen. 1999. MODIS vegetation index (MOD13) algorithm theoretical basis document. *Greenbelt: NASA Goddard Space Flight Centre, http://modarch. gsfc. nasa. gov/MODIS/LAND/# vegetation-indices*.
Huete, A. R., K. F. Huemmrich, T. Miura, X. Xiao, K. Didan, W. van Leeuwen, F. Hall, and C. J. Tucker. 2006. Vegetation Index greenness global data set. *College Park, MD*.
Johnston, K., M. Sakala, and J. Wrightsell. 2001. *Using ArcGIS geostatistical analyst*. Environmental Systems Research Institute Redlands, CA.
Johnston, K., J. M. Ver Hoef, K. Krivoruchko, and N. Lucas. 2001. *Using ArcGIS geostatistical analyst*. Esri New York.
Kerr, J. T., and M. Ostrovsky. 2003. From space to species: ecological applications for remote sensing. *Trends in Ecology & Evolution* 18, no. 6: 299-305.
Krivoruchko, K. 2009. Introduction to Modeling Spatial Processes Using Geostatistical Analyst. ESRI. http://www.esri.com/library/whitepapers/pdfs/intro-modeling.pdf.
Lu, Hua, Michael R. Raupach, Tim R. McVicar, and Damian J. Barrett. 2003. Decomposition of vegetation cover into woody and herbaceous components using AVHRR NDVI time series. *Remote Sensing of*

Environment 86, no. 1 (June 30): 1-18. doi:doi: DOI: 10.1016/S0034-4257(03)00054-3.
Moreau, S., R. Bosseno, X. Fa Gu, and F. Baret. 2003. Assessing the biomass dynamics of Andean bofedal and totora high-protein wetland grasses from NOAA/AVHRR. *Remote Sensing of Environment*, no. 85: 516-592.
Mutanga, O., and D. Rugege. 2006. Integrating remote sensing and spatial statistics to model herbaceous biomass distribution in a tropical savanna. *International Journal of Remote Sensing* 27, no. 16: 3499-3514.
Prince, S. D., 1991. Satellite remote sensing of primary production: comparison of results for Sahelian grasslands 1981-1988. *International Journal of Remote Sensing* 12, no. 6: 1301-1311.
Prince, S. D., 1991. A model of regional primary production for use with coarse resolution satellite data. *International Journal of Remote Sensing* 12, no. 6: 1313-1330.
Prince, S. D., and C. J. Tucker. 1986. Satellite remote sensing of rangelands in Botswana II. NOAA AVHRR and herbaceous vegetation. *International Journal of Remote Sensing* 7, no. 11: 1555-1570.
de Ronde, C., C.J. Geldenhuys, and W. S. W. Trollope. 2004. Fire Behaviour. In *Wildland Fire Management Handbook for Sub-Saharan Africa*, ed. J.G. Goldammer and C. de Ronde, 27-59. Global Fire Monitoring Center.
de Ronde, C., C.J. Geldenhuys, W. S. W. Trollope, C.L. Parr, and B. Brockett. 2004. Fire Effect on Flora and Fauna. In *Wildland Fire Management Handbook for Sub-Saharan Africa*, ed. J.G. Goldammer and C. de Ronde, 60-87. Global Fire Monitoring Center.
Rossiter, D. G. 2007. *Technical Note: Co-kriging with the gstat package of the R environment for statistical computing*. Citeseer.
Sanderson, M.A., C.A. Rotz, S.W. Fultz, and E.B. Rayburn. 2001. Estimating Forage Mass with a Commercial Capacitance Meter, Rising Plate Meter and Pasture Ruler. *Agronomy Journal* 93: 1281-1286.
Sankaran, M., N.P. Hanan, R.J. Scholes, J. Ratnam, D.J. Augustine, B.S. Cade, J. Gignoux, et al. 2005. Determinants of woody cover in African savannas. *Nature* 438, no. 7069 (December 8): 846-849. doi:10.1038/nature04070.
Sannier, C. A. D., J. C. Taylor, and W. D. Plessis. 2002a. Real-time monitoring of vegetation biomass with NOAA-AVHRR in Etosha National Park, Namibia, for fire risk assessment. *International Journal of Remote Sensing* 23, no. 1: 71-89.
Savadogo, P., D. Tiveau, L. Swadogo, and M. Tigabu. 2008. Herbaceous species response to long-term effects of prescribed fire, grazing and selective tree cutting in the savanna-woodlands of West Africa. *Perspectives in Plant Ecology, Evolution and Systematics* 10: 179-195.
Scanlon, T. M., J. D. Albertson, K. K. Caylor, and C. A. Williams. 2002. Determining land surface fractional cover from NDVI and rainfall time series for a savanna ecosystem. *Remote Sensing of Environment* 82, no. 2: 376-388.
Shackleton, C.M., S.E. Shackleton, E. Buiten, and N. Bird. 2007. The importance of dry woodlands and forests in rural livelihoods and

poverty alleviation in South Africa. *Forest Policy and Economics* 9, no. 5 (January): 558-577. doi:doi: DOI: 10.1016/j.forpol.2006.03.004.
Sheuyange, Asser, Gufu Oba, and Robert B. Weladji. 2005. Effects of anthropogenic fire history on savanna vegetation in northeastern Namibia. *Journal of Environmental Management* 75, no. 3 (May): 189-198. doi:doi: DOI: 10.1016/j.jenvman.2004.11.004.
Smit, G. N. 2004. An approach to tree thinning to structure southern African savannas for long-term restoration from bush encroachment. *Journal of Environmental Management* 71, no. 2: 179-191. doi:doi: DOI: 10.1016/j.jenvman.2004.02.005.
———. 2005. Tree thinning as an option to increase herbaceous yield of an encroached semi-arid savanna in South Africa. *BMC ecology* 5, no. 1: 4.
Thein, T.R., F.G.R. Watson, S.S. Cornish, T.N. Anderson, W.B. Newman, and R.E. Lockwood. 2008. Chapter 7 Vegetation Dynamics of Yellowstone's Grazing System. *Terrestrial Ecology* 3: 113-133. doi:10.1016/S1936-7961(08)00207-8.
Thompson, M.W., and C.S. Everson. 1993. Development of spectral-biomass models for mapping and monitoring montane grassland resources. Division of Forest Science and Technology CSIR.
Todd, S. W., R. M. Hoffer, and D. G. Milchunas. 1998. Biomass estimation on grazed and ungrazed rangelands using spectral indices. *International Journal of Remote Sensing* 19, no. 3: 427-438.
Trollope, W, and A. L. F. Potgieter. 1986. Estimating grass fuel loads with a disc pasture meter in the Kruger National Park. *African Journal of Range and Forage Science* 3, no. 4.
Trollope, W. S. W. 1996. Biomass Burning in the savannas of Southern Africa with Particular Reference to the Kruger National Park in South Africa. In *Biomass Burning and Global Change: Remote sensing, modeling and inventory*, ed. J. S. Levine, 1:260 - 269. 1st ed. Massachusetts Institute of Technology.
Trollope, W.S.W., L.A. Trollope, and D.C. Hartnett. 2002. Fire behaviour a key factor in the fire ecology of African grasslands and savannas. *Forest fire Research & Wildland Fire Safety*.
Tucker, C. J., C.L. Vanpraet, M.J. Sharman, and G. Van Ittersum. 1985. Satelite Remote Sensing of Total Herbaceous Biomass Production in the Senegalese Sahel: 1980 - 1984. *Remote Sensing of Environment*, no. 17: 233-249.
Twine, W., D. Moshe, T. Netshiluvhi, and V. Siphugu. 2003. Consumption and direct-use values of savanna bio-resources used by rural households in Mametja, a semi-arid area of Limpopo province, South Africa.
Van Wilgen, B.W., W.S.W. Trollope, H.C. Biggs, A. L. F. Potgieter, and B.H. Brockett. 2003. Fire as a Driver of Ecosystem Variability. In *The Kruger Experience*, ed. J. du Toit, H. Biggs, and K.H. Rogers, 149-170. Island Press.
Verbesselt, J., B. Somers, J. van Aardt, I. Jonckheere, and P. Coppin. 2006. Monitoring herbaceous biomass and water content with SPOT VEGETATION time-series to improve fire risk assessment in savanna ecosystems. *Remote Sensing of Environment* 101, no. 3 (April 15): 399-414. doi:doi: DOI: 10.1016/j.rse.2006.01.005.

Wells, M. P. 1997. Economic perspectives on nature tourism, conservation and development. *Environment department paper* 55.
Wessels, K. J., S. D. Prince, N. Zambatis, S. MacFadyen, P. E. Frost, and D. Van Zyl. 2006. Relationship between herbaceous biomass and 1km 2 Advanced Very High Resolution Radiometer (AVHRR) NDVI in Kruger National Park, South Africa. *International Journal of Remote Sensing* 27, no. 5: 951-973.
Willmott, C. J., and K. Matsuura. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* 30, no. 1: 79.
Wilson, B.G. & Witkowski, E.T.F. (2003). Seed banks, bark thickness, and changes in age and size structure (1978 – 1999) of the African savanna tree, *Burkea africana. Plant Ecology* **167,** 151-162.
Witkowski, E.T.F. & Garner, R.D. (2000). Spatial distribution of soil seed banks of three African savanna woody species at contrasting sites. *Plant Ecology* **149**, 91-106.
Xiao, X., G. Gertner, G. Wang, and A. B. Anderson. 2005. Optimal sampling scheme for estimation landscape mapping of vegetation cover. *Landscape ecology* 20, no. 4: 375-387.

Appendix 1: Field based sampling

Resource requirements for providing detailed information on herbaceous biomass (spatial resolution of 250m) for the entire KNP using the 'clip and weigh' approach:

One would need to sample on a grid with nodes spaced at most 50m apart to gain a representative sample for each 250 m block. Assume that the fieldwork team:

1. could move between nodes at 5 km/hr
2. could clip and weigh herbaceous material at a node in 2 minutes (a very generous assumption)
3. would work for 9 hours a day with a 1 hour lunch break
4. would take 1 hour each way to travel to and from the field, a total of 2 hours each day (once again a very generous assumption)

If sampling at 50m intervals and using these assumptions then a team of fieldworkers (1 operator, 1 data recorder and 1 game guard) could cover a maximum of 8.08 km/day collecting 162 measurements. To move between and sample all the nodes needed to cover the KNP at this rate would take 2351 days or 6.4 years. Information on herbaceous biomass is however needed before the fire season starts each year so measurements need to be completed as soon as possible after the grass starts drying out. To complete all of the measurements within one month of the end of the growth season would require 84 teams containing in total 2100 people working 7 days a week.

Resource requirements for providing detailed information on herbaceous biomass (spatial resolution of 250m) for the entire KNP using a Disk Pasture Meter (DPM):

One would need to sample on a grid with nodes spaced at most 50m apart to gain a representative sample for each 250 m block. Assume that the fieldwork team:

1. could move between nodes at 5 km/hr
2. Take a DPM reading at a node in 10 seconds
3. would work for 9 hours a day with a 1 hour lunch break
4. would take 1 hour each way to travel to and from the field, a total of 2 hours each day (once again a very generous assumption)

If sampling at 50m intervals and using these assumptions then a team of fieldworkers (1 operator, 1 data recorder, and 1 game guard) could cover a maximum of 27.39 km a day collecting 548 measurements. This is 19.31 km further and 386 more measurements than possible when clipping and weighing material. To move between and sample all the nodes needed to cover the Kruger National Park at this rate would take 694 days or 1.9 years. To complete all of the measurements in one month would require 25 such teams containing in total 75 people working 7 days a week.

Chapter 2

METHODS AND MATERIALS

1. OVERVIEW

The materials and methods used in the completion of this study have been grouped into eight sections based largely on the order in which the work was completed. Figure 1 provides a summary of how these sections linked together to achieve the aim of this study.

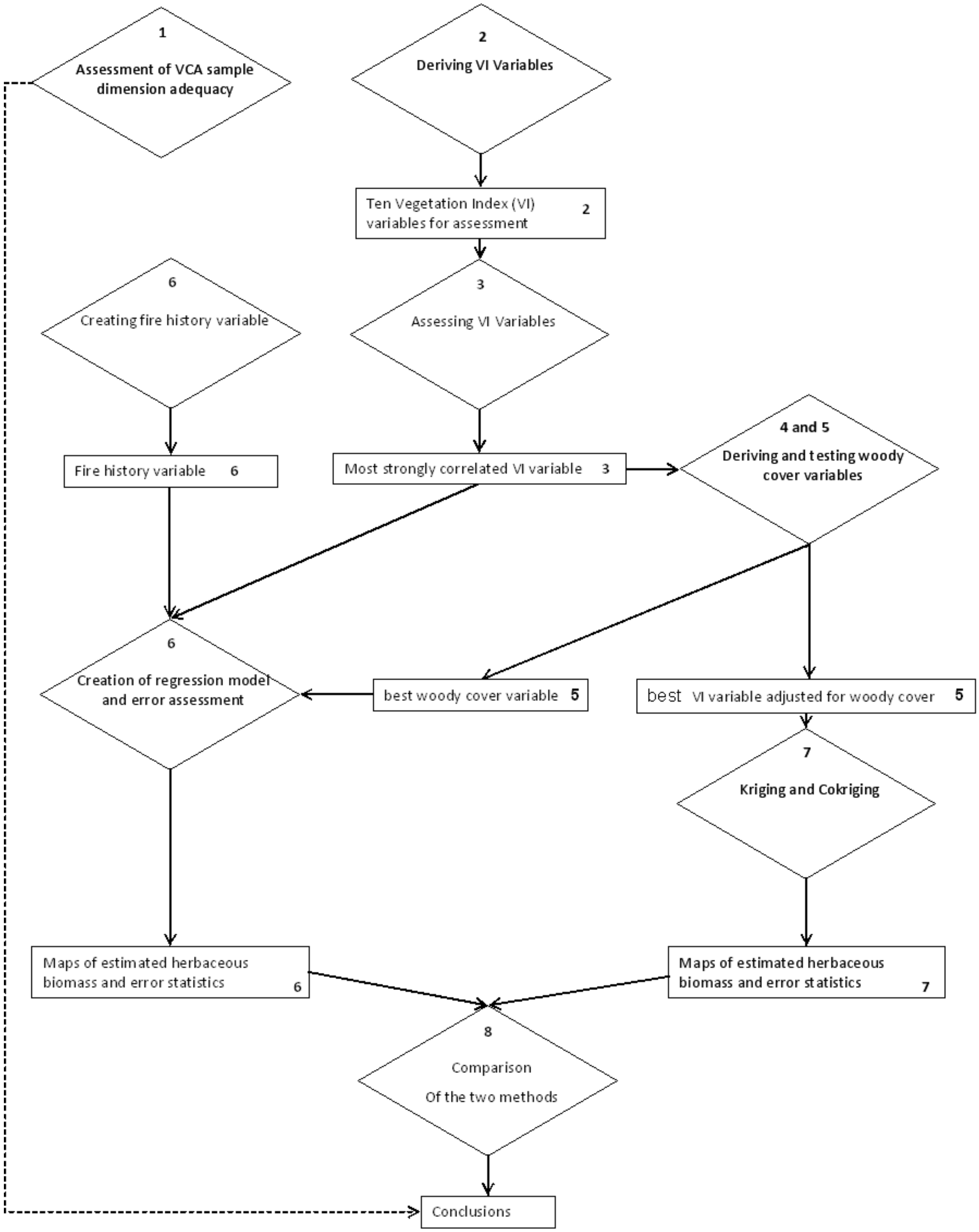


Figure 1: Flow chart summarising steps involved in achieving this studies aim.

Section 1 of the analysis was concerned with assessing the Veld Condition Assessment (VCA) field data available for use in the study. This step was included in light of the fact that the VCA sample plot dimensions were significantly smaller in area than those of the MODIS pixels the data was being paired with. The intention was to develop a context in which the final results of the study could be compared to other studies, and interpreted. Section 2 involved deriving 10 different herbaceous biomass production variables using both MODIS MOD13 EVI and NDVI. Section three involved identifying the herbaceous biomass production variable produced in section 2 that was best correlated to the herbaceous biomass field estimates from the VCA dataset. Sections 4 and 5 involved deriving woody cover variables and identifying the woody cover variable that accounted for the greatest improvement in estimation accuracy when:

a. included in a linear model to predict herbaceous biomass using the production variable identified in section 3, and
b. used to adjust the variable identified in section 3 to account for the contribution of trees to that variable's value.

Section 6 involved creating a fire history variable using fire scar and geology data and then creating a regression model using this variable, the production variable from section 3 and the woody cover variable form section 5 to predict herbaceous biomass. Error statistics and prediction maps for all seasons were also produced. Section 7 involved cokriging the herbaceous biomass map from the VCA field estimates with the production variable adjusted using the woody cover from section 5 as the secondary variable. Error statistics were also produced for the cokriged maps. Section 8 involved comparing the two methods based on the error statistics obtained from sections 6 and 7.

Detailed flowcharts outlining the activities undertaken for each section are provided later in this chapter. Each flowchart is preceded by information on the materials and methods used and followed by a detailed description of those methods. All flowcharts in this chapter use rectangles to represent both inputs and outputs while diamonds are used to represent actions or processes.

2. STUDY AREA

The Kruger National Park (KNP) was selected as the study area for a number of reasons. The primary reason was that a sufficiently large number of herbaceous biomass field estimates, those from the annual Veld Condition Assessment (VCA), already existed for the area. The VCA sample locations were also distributed over a large and diverse enough area, and the historical record long enough, to allow for meaningful statistical analysis to be conducted. No comparable data could be obtained for use from any other part of the country or region. The third reason was the interest in such a study expressed by the KNP remote sensing and fire management teams. The final reason was that numerous GIS layers detailing biophysical variables for the area were also available for use as additional predictor variables and to assess the conditions at each field sample location if required.

The KNP is located in the lowveld on the North Eastern border of South Africa adjacent to Zimbabwe and Mozambique (Figure 2). It falls entirely within the Savanna biome with mean annual rainfall varying from 350 mm/year in the north to 950 mm/year in the South West (Wessels et al. 2006). There is a rough West –East geological divide in the park with Granites in the West and Basalts in the East. This results in a similar divide in soil fertility with nutrient poor soils overlying the granites and relatively nutrient rich soils overlying the basalts (Venter, Scholes and Eckhardt 2003). Woody canopy cover varies from 5% to 60% with more than 75% consisting of 2-5 meter high trees with a canopy cover of between 20 and 40%. Canopy cover tends to be lower within the fertile soils overlying basalt where fires are more intense and higher within the infertile soils overlying granite where fires are less intense (Venter, Scholes and Eckhardt 2003).

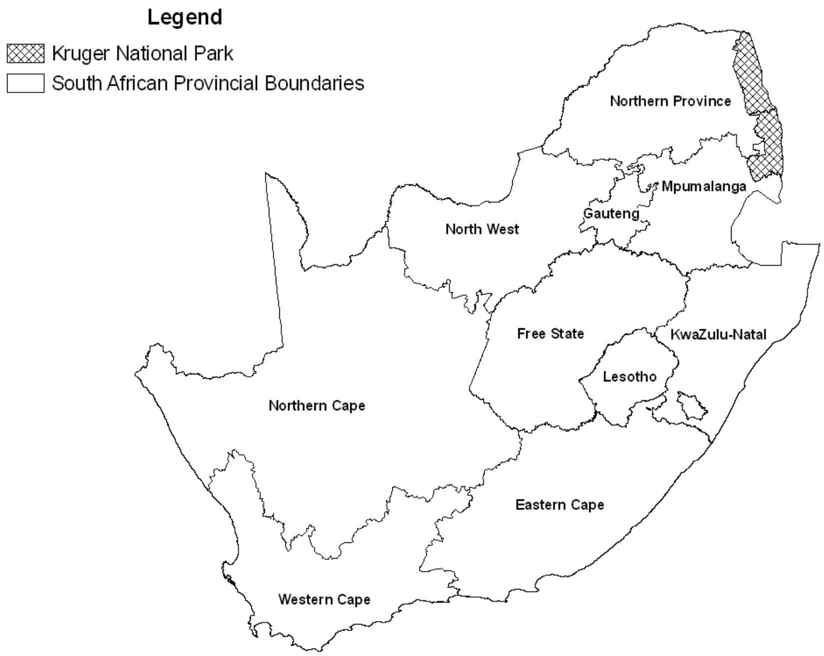


Figure 2: Location of the Kruger National Park within South Africa.

3. METHODS

3.1. *Assessment of the adequacy of the VCA sample site dimensions*

This section of work was not originally included in the project plan, however, initial exploratory analysis showed a discouragingly weak relationship through space between VI values and herbaceous biomass in the area (R^2 of 0.00 – 0.29). After revisiting the literature it became clear that this was not a fault with the initial analysis. This conclusion was reached based on the fact that other studies conducted in the area (Wessels et al. 2006; Mutanga and Rugege 2006), which made use of the VCA data, showed a weaker than expected relationship when compared to other published work (Al-Bakri and Taylor 2003; Moreau et al. 2003; Prince 1991). One of the major differences between the studies based in the KNP making use of the VCA sample sites and the other studies encountered was the size of the field sample plots used. The VCA field plots (50x60m = 3000 m^2, or 0.3 ha) dimensions differed significantly from those used in studies encountered in the literature. Sannier et al (2002) for example, used 1000x8m transects giving an area of 8000 m^2, or 0.8 ha, although the sampling intensity was the same as for the VCA sites, 100 DPM readings per transect. It was decided to further investigate the effects of the mismatch in the dimensions of the VCA field data and MODIS pixel data available for use in this study.

The VCA dataset consists of herbaceous vegetation species composition and biomass estimates recorded at approximately 533 fixed 50 x 60m sites across the KNP (Figure 3) between the end of March and the middle of April (i.e. at the end of the wet season) each year (Zambatis 2002).

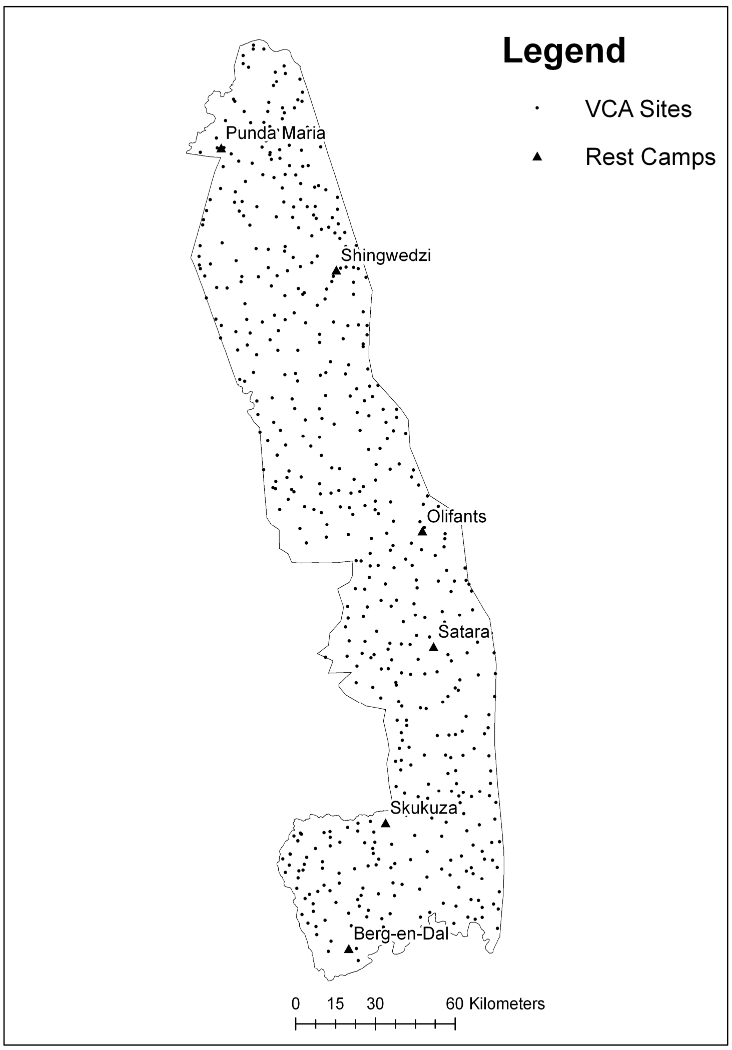


Figure 3: Location of VCA sample sites in the Kruger National Park.

Pasture meter readings are not taken at all the sites every year, resulting in slight variation in the number of estimates available for each season. Measurements have been recorded since 1986, providing 23 years of biomass estimates at the time of writing. There is no other comparable herbaceous biomass dataset available to researchers in southern Africa. This makes the VCA a valuable resource for those interested in the inter annual variation in anything associated with herbaceous biomass.

No information regarding the consistency of the teams who collected the data was evident in the excel spreadsheet in which the data was provided. Sampling commences when the grass begins to dry out, as determined by visual assessment of the herbaceous layer by those responsible for carrying out the survey (Zambatis 2002). Herbaceous biomass is measured using a disk pasture meter (DPM) calibrated for the area. 100 readings are taken on a grid paced out within a 50 x 60m area located at the sites co-ordinates and converted to biomass in kg/ha using the conversion equation detailed in Trollope and Potgieter (1986):

$$\text{Herbaceous biomass (kg/ha)} = -3019 + 2260\sqrt{\text{mean DPM height (cm)}}$$ [Equation 1]

The standard deviation of the residuals calculated during the creation of the equation using the field measurements as both testing and training data was reported to be 898 kg/ha (Trollope and Potgieter 1986). In other words, on average herbaceous biomass estimates produce using a pasture meter differ from the actual herbaceous biomass on the ground by 898 kg/ha.

The paper by Trollope and Potgieter (1986) on the calibration process states that measurements taken where the veld was moribund did not fit with the overall relationship (Figure 4) and thus were removed (Figure 5). Moribund veld in this context refers to veld in which live green herbaceous material is absent or scarce because new growth is shaded out by dead material from previous growth seasons.

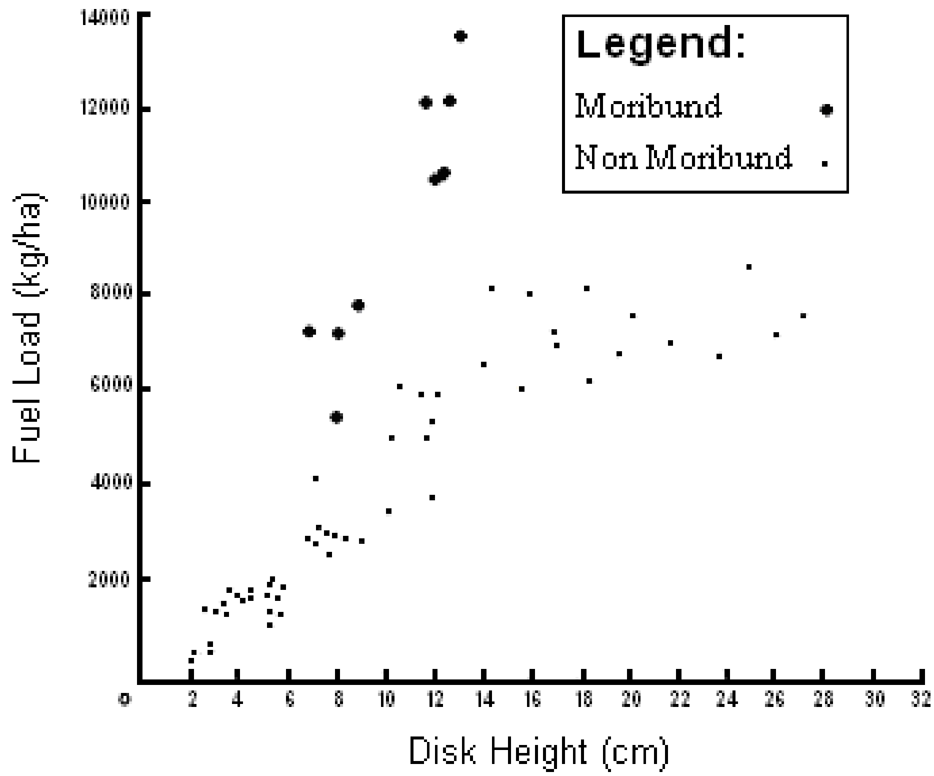


Figure 4: The relationship between clipped biomass measurements and disk pasture meter height readings in the Kruger National Park, adapted from Trollope and Potgieter (1986). The bold points indicate moribund site measurements removed from the final training data.

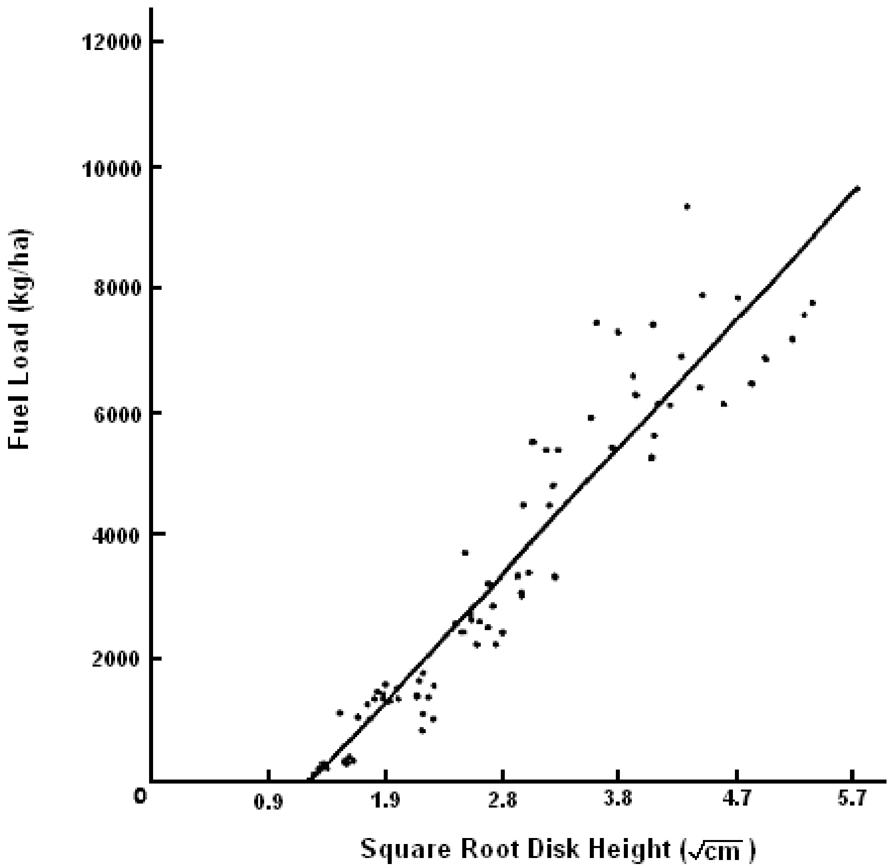


Figure 5: "The linear regression between the square root of disk height and fuel load for non moribund and erect grass swards in the major landscapes of the central and southern Kruger National Park" (Trollope and Potgieter 1986). Note that in contrast to Figure 3 the points where the veld was moribund have been removed. The square root of the disk heights was used because Trollope and Potgieter (1986) found they resulted in the best model fit.

To account for the points being removed a proviso was attached to the use of the conversion equation. The proviso was that the equation cannot be used to convert pasture meter readings taken in areas where the veld was dominated by dead material (Trollope and Potgieter 1986) because it would clearly *underestimate* the actual herbaceous biomass and thus fuel load in areas with significant moribund grass.

Two characteristics of the VCA dataset evident from this description are of relevance to a remote sensing study, these are:

1. The herbaceous biomass data are not measurements, they are estimates with a RMSE of $\pm$800kg/ha. This is a significant error considering the RMSE of the herbaceous estimates produced by (Mutanga and Rugege 2006) by combining this data with remotely sensed imagery ranged from 830 – 1374 kg/ha.
2. The field sites used are 50x60m in size (much smaller than MODIS or AVHRR pixels).

The affects of using inaccurate estimates as field data are fairly easy to predict. The issue here would be related to the distribution of areas with significant moribund grass. Sites that have been burnt recently would generally tend to have very little moribund material, but sites unburnt for several years are quite likely to be moribund. Similarly areas favoured by grazing herbivores would also have low levels of moribund grass. Measurement error in a variable will lead to unexplained variation in the model and increase estimation error. The effect of using small sample sites is more difficult to predict as it depends on both the nature of the spatial variation in the herbaceous layer and the size of the VI pixels. After consulting past studies it was apparent that in most cases where a strong relationship between herbaceous biomass and VI data was found, the area of the sample plots used to obtain the biomass field estimates were much larger than the 3000 m^2 used to obtain the VCA biomass field estimates (Diallo et al. 1991; Prince 1991; Prince and Tucker 1986; Sannier, Taylor, and Plessis 2002). This suggests that one of the factors involved in attaining a strong relationship between VI values and herbaceous biomass field estimates in savannas is the use of large field sample plots.

Wessels et al (2006), aware of the field data – pixel dimension mismatch, looked for a correlation between variability in NDVI generated from LANDSAT imagery in a 700m radius around each VCA site and the strength of AVHRR derived NDVI and the VCA biomass estimates through time. They found the sites with very high standard deviations in Landsat NDVI were generally closer than 600m to rivers and often contained riparian woodland vegetation along drainage channels with seasonal water or bare sand. These sites (n=37) were therefore excluded from further analysis. After this removal there was no relationship between the Landsat NDVI variation in the sites and their coefficient of determination between biomass and growth season sum AVHRR NDVI. All remaining sites (n=464) were therefore included in the subsequent analyses (Wessels et al. 2006). Although Wessels et al. (2006) provide a method for excluding highly heterogeneous sites they do not provide any information on the magnitude of the error that might arise by failing to do so. It was decided to pursue this via a field based assessment measuring the extent of the difference between herbaceous biomass measured on co located 50x60m and 250 x 250m sample plots. A summary of the assessment is provided in Figure 6.

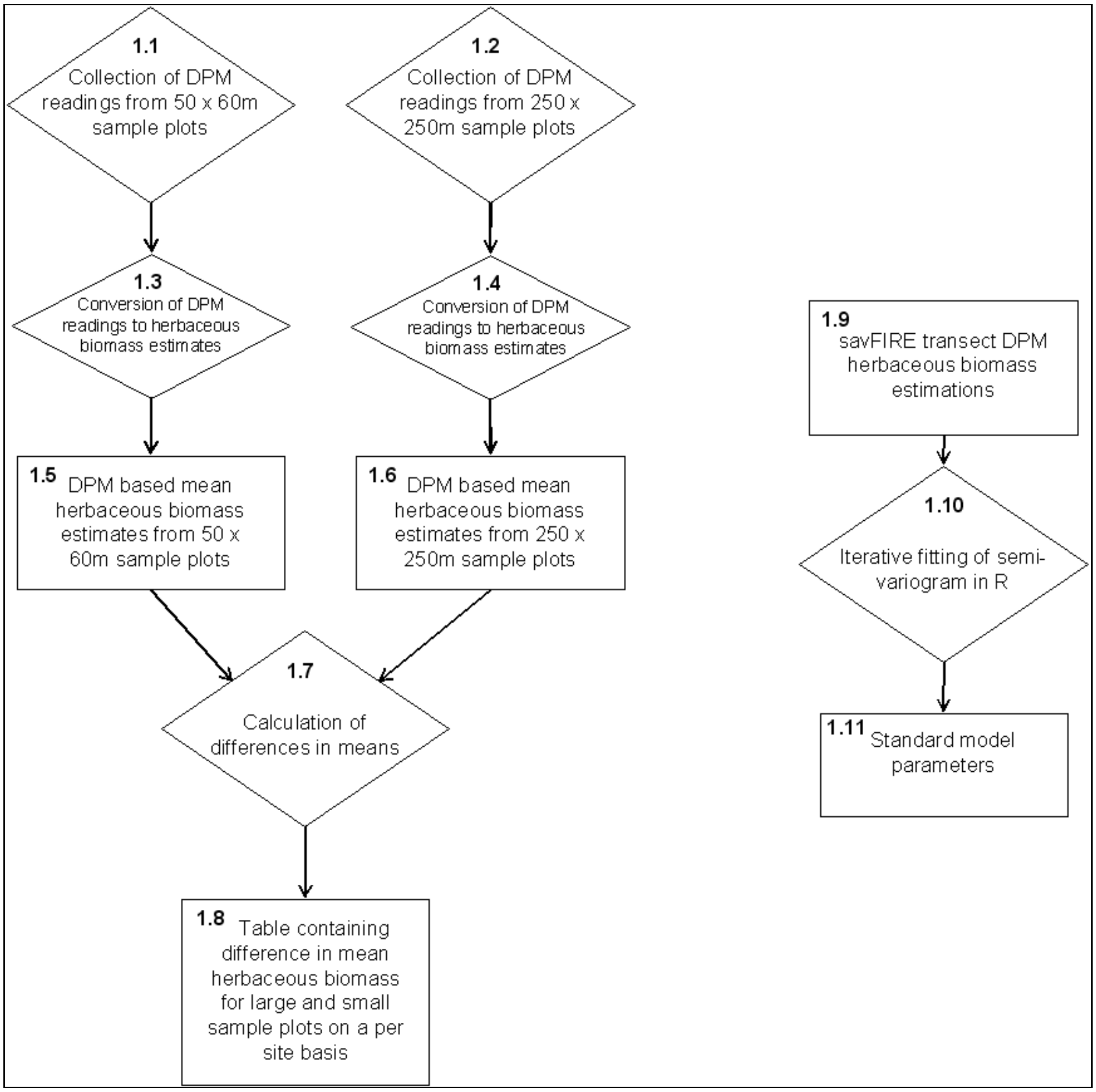


Figure 6: The steps involved in assessing the adequacy of the VCA sample site dimensions.

Collection of the DPM readings (Figure 6: 1.1 and 1.2) was conducted during the second week of April 2007 in the south of the KNP. Fieldwork was restricted to the south of the park within reasonable proximity of the research camp (adjacent to Skukuza rest camp) because of logistical constraints. Eight sample sites were selected based on them being within less than 2 hours drive from the research camp (Figure 7) and two sets of disk pasture meter readings taken at each site.

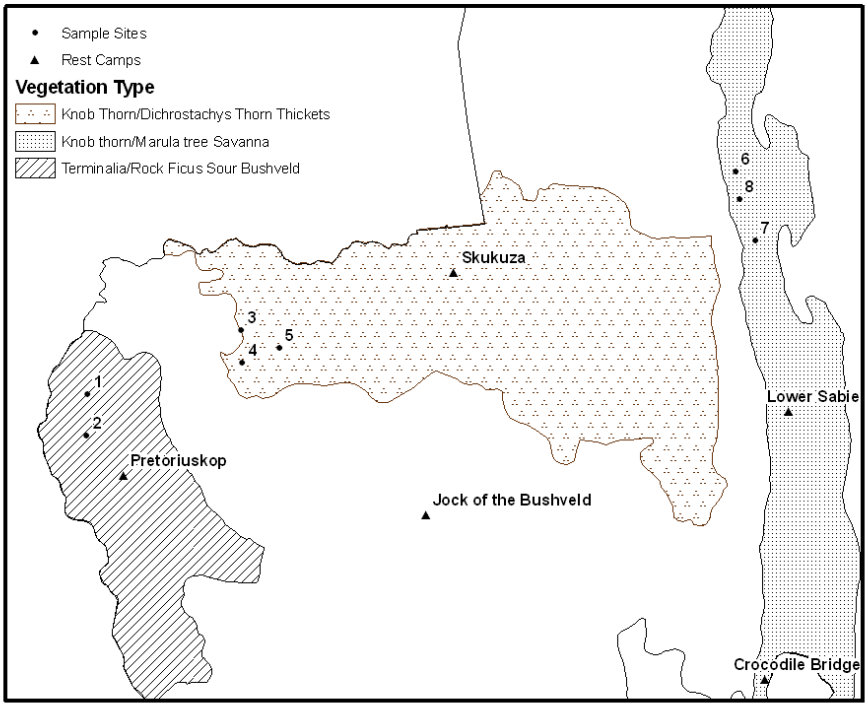


Figure 7: The location of the 50x60 m and 250x250 m sample sites within the Kruger National Park.

The first set comprised of 36 Disk Pasture Meter (DPM) readings taken over a 50x60 m area with the second set comprised of 60 readings taken over a 250x250 m area (Figure 8). The smaller set was always located within the area covered by the larger set.

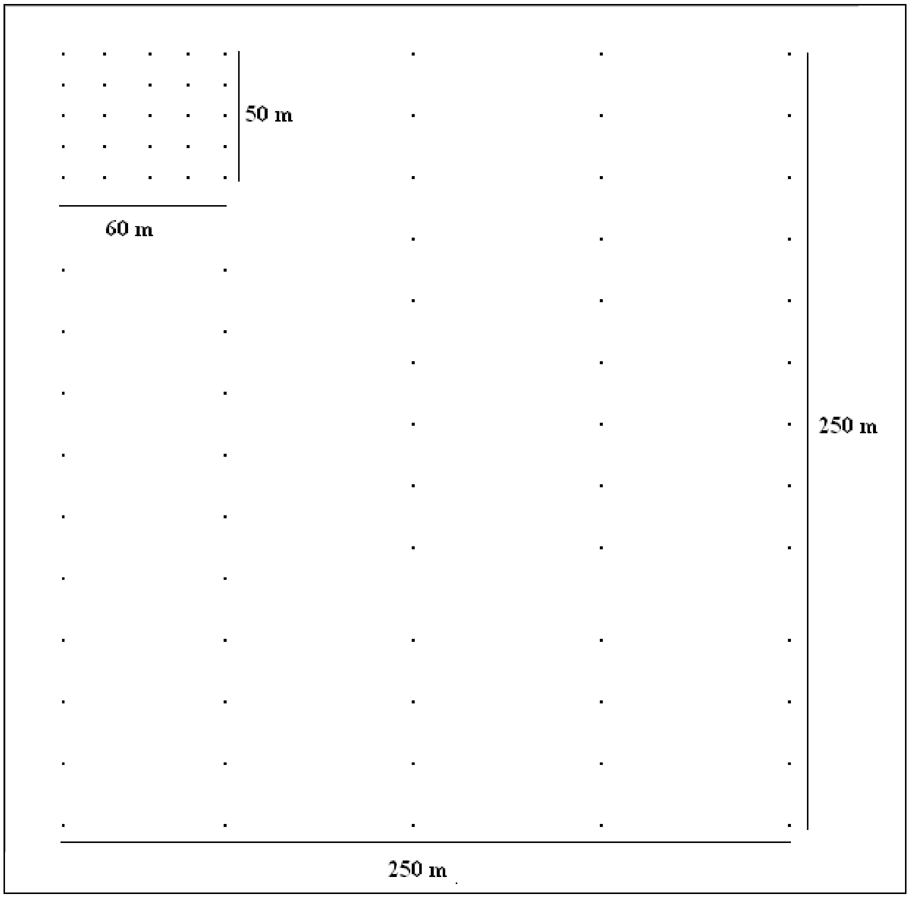


Figure 8: Layout of sample points to test the suitability of 50x60m sample plots as representative samples for 250x250m areas.

The pasture meter readings were converted to herbaceous biomass estimates (Figure 6: 1.3 and 1.4) using Equation 1, the equation specified by Trollope and Potgieter (1986). The mean biomass value for each site (Figure 6: 1.5 and 1.6) and the difference between the mean values for the corresponding large and small sites (Figure 6: 1.7) were calculated and the outputs entered into a table for comparison (Figure 6: 1.8).

3.2. Deriving VI variables

Having investigated concerns over the suitability of the VCA field data being used, attention was turned to the first step in the creation of the regression model, deriving a suitable primary explanatory variable based on Vegetation Index (VI) data.

A single VI image provides an indication of vegetation greenness and hence total photosynthetic potential at that specific point in time (Huete et al. 2006).

However, fire management planning in savannas requires estimates of herbaceous biomass at the end of the growing season once the herbaceous layer has dried out, but before prescribed burning has occurred.

To address this problem individual VI images can be integrated over the growing season to give a measure of total photosynthetic potential within a growth season (Huete et al. 2006; Wessels et al. 2006). Figure 9 summarises how the problem was dealt with, namely that individual VI images were integrated over the growth seasons assessed in this study.

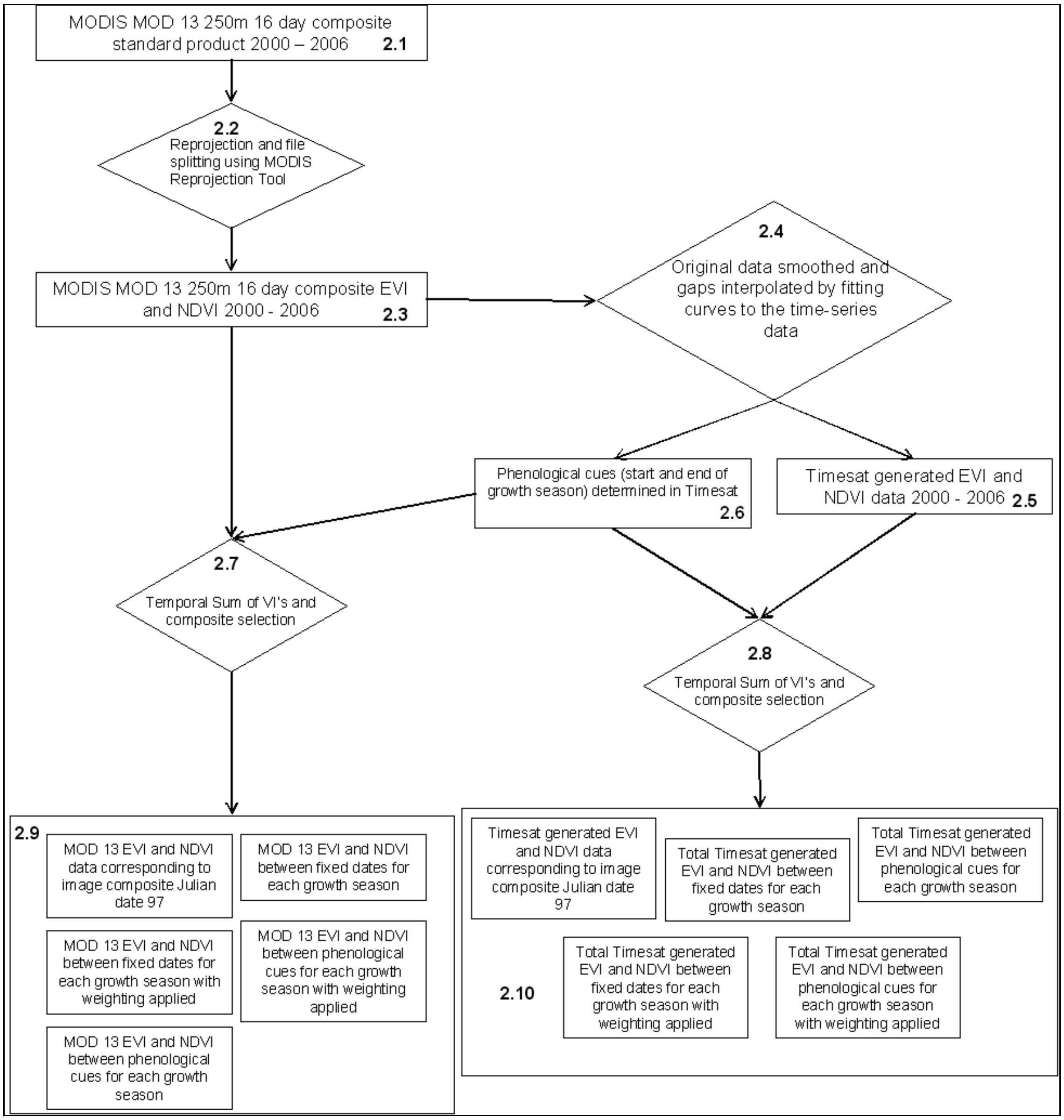


Figure 9: A summary of the process by which Vegetation Index variables were derived from separate MODIS MOD13 16 day composite Vegetation Index images.

MODIS data was selected for use in the study because it offers the best combination of radiometric, spatial and temporal resolutions currently available for monitoring of vegetation activity over large areas (Figure 9: 2.1). Before the specific MODIS product can be discussed some general background on the MODIS sensor is required. The MODIS sensors aboard the Aqua and Terra satellites record data in 36 spectral bands for the entire earth's surface every 1-2 days. The Red and Near

Infrared bands (bands 1 and 2) are acquired at 250m resolution while the blue band used to correct for atmospheric interference and background soil colour (band 3) in some VI's is acquired at 500m resolution (Huete, Justice and Van Leeuwen 1999a).

Data products derived from both satellites are available at a number of different spatial and temporal resolutions. Most are gridded products, meaning that the original observations from the sensor are resampled to fit into a predefined grid. Nearest neighbour resampling is used to assign observations to grid pixels (Tan et al. 2006). This causes 'pixel shift' because the pixel on the predefined grid will no longer correspond to the exact location at which the reflectance values which they assigned were measured (Tan et al. 2006). The difficulty this causes for those trying to match field data to a pixel is compounded by the fact that MODIS is a 'whisk broom' scanner. These scanners capture data from a scene by scanning it one row at a time perpendicular to the axis of travel or 'track' as it passes overhead. This produces pixels that vary in size depending on the angel at which the scanner was tilted to capture them, known as the viewing angle (Figure 10).

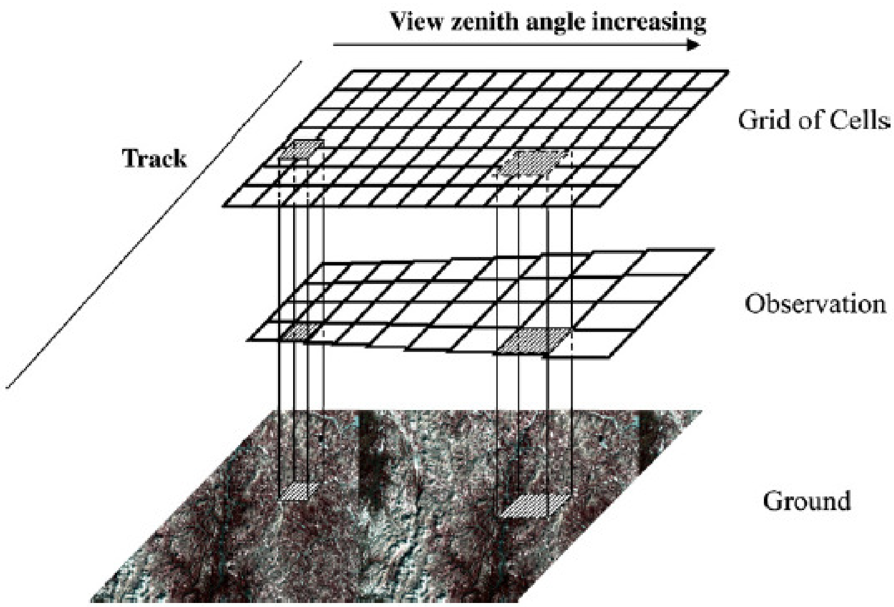


Figure 10: Illustration of how pixel size increases with increased view angle leading to a mismatch of pixel dimensions between observed and grid pixels (Tan et al. 2006).

Increased viewing angle leads to increased pixel size. This means that in addition to a mismatch in the location of the sample between grid position and the actual area sampled, it may also be larger than the pixel to which the reflectance value is assigned (Tan et al. 2006). This will result in an upward bias in reflectance values because reflection from a larger area is being recorded. Even in cases where the observed area and a grid cell overlap perfectly, only 75% of the value recorded is attributable to the area covered by the grid cell (Figure 11) (Tan et al. 2006; Huang et al. 2002). To limit these effects the MODIS VI is composited using a “view-angle” constraint (Huete, Justice and Van Leeuwen 1999a)

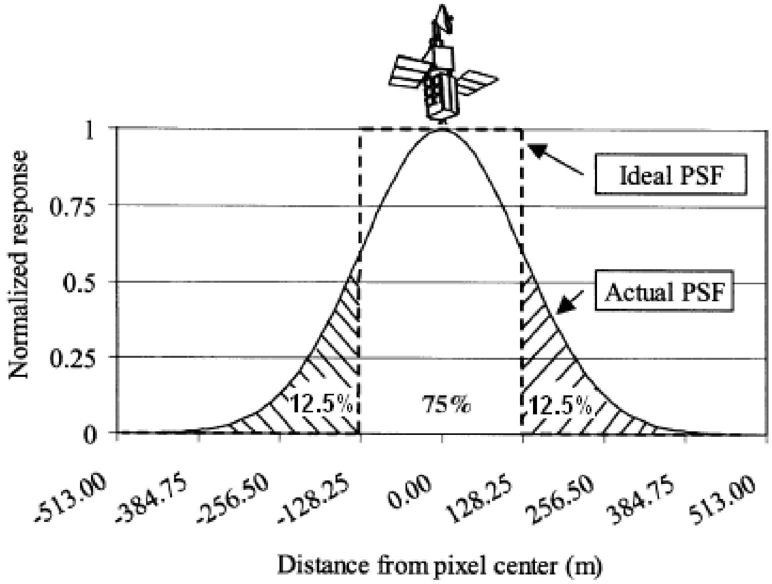


Figure 11: The 'triangular' or bell shaped point spread function (PSF) of the MODIS sensor results in only 75% of the reflectance in an observation at nadir originating from within the area observed. Modified from (Huang et al. 2002b).

This is because of the sensor's bell shaped point spread function (PSF) (Tan et al. 2006). A technical understanding of PSF and why it occurs falls outside of most natural scientist's sphere of knowledge. For the purpose of this study it is sufficient to state that PSF is dependent on sensor design and so cannot be controlled by end users as is the case with gridding error. When taking both into account it turns out that on average less than 30% of the signal received originates from the area covered by the pixel in the gridded product to which it is assigned (Tan et al. 2006).

As the two error sources cannot be removed by end users their effect must either be accepted at a given resolution or minimised through aggregating pixels together. It has been shown that by degrading the data's resolution by a factor of 8 through grouping pixels together, 80% of the signal from the group originates from the area covered by that group (Tan et al. 2006). This equates to moving from 250 m resolution to 2km resolution or from 500 m to 4 km resolution. This was not attempted in this study because the field data available would be grossly unrepresentative of a 2km pixel because of the dimensions of the field plots it was gathered from (50x60 m).

The specific MODIS data product of interest in this study is the MOD 13 Vegetation Index product which was created to provide a spatially and temporally consistent measure of vegetation conditions (Huete, Justice and Van Leeuwen 1999b). MOD 13 data is available at spatial resolutions of 250m and 500m as 16 day composites. It is also available at spatial resolutions of 1 km and 5.6 km as either 16 day or monthly composites. These composites are created using a special compositing algorithm designed to maximise VI data quality through a combination of methods discussed below.

The first step in creating a 16 day image composite is the collection of 16 days of data and filtering it to exclude cloud contaminated and extreme of-nadir pixels. This typically yields less than 10 pixels per 16 pixel stack deemed to be of adequate quality (Wolfe, Roy, and Vermote 1998). Pixels with an off-nadir viewing angle of <45 degrees and no cloud contamination are considered 'good', while those with >45 degree viewing angle and/or cloud contamination are deemed unacceptable and discarded (Wolfe, Roy, and Vermote 1998). Depending on the number and quality of the pixels retained, one of three compositing methods is applied to produce a single pixel representative of the average conditions over the 16 day period. The methods are the MVC: maximum value composite, CV-MVC: constraint-view angle - maximum value composite and the BRDF-C: bidirectional reflectance distribution function composite (Wolfe, Roy, and Vermote 1998). The maximum value composite involves calculating the VI for each of the pixels and selecting the pixel with the highest value. This is the most basic method and is used as a backup when no acceptable values are present. The CV-MVC involves taking the 3 highest VI values and selecting the one with the viewing angle closest to nadir. The method is used when there are less than 5 acceptable values. The BRDF method is the most complex, it involves first making use of all bands from all acceptable pixels to calculate the expected at nadir reflectance in each band. The interpolated reflectance values of the resulting pixel are then used to calculate the VI. This method is used if 5 or more good quality values exist (Wolfe, Roy, and Vermote 1998). The success of the compositing method at improving pixel quality depends on the prevalence of cloud cover. The more prevalent cloud cover during acquisition, the more cloud contaminated and extreme viewing angle pixels contained in the 16 day composites.

The Native projection of MODIS data is Sinusoidal and the native file format HDF5, both of which make it difficult to open and interact with in many standard GIS and remote sensing software packages. To address this issue the MODIS science team created the MODIS Reprojection Tool (MRT), which enables users to reproject the data into more user friendly projections and file formats. MRT was used to reproject the data used in this study into WGS 1984 UTM 36 south. This projection was selected to match the projection of all of the available GIS layers provided as support for the study by the Kruger National Park (Figure 9: 2.2). The resulting NDVI and EVI raster layers were saved as16 bit flat binary files to await further processing and summation (Figure 9: 2.3).

As part of the compositing process discussed earlier, information on the quality of the pixels produced by the compositing algorithm is recorded on a per pixel basis. A number of levels of detail are available ranging from an overview of quality to a complete breakdown of all contributing factors to pixel quality and values for each. This allows users to screen the data and identify poor quality pixels. Cloud contamination of pixels is one of the major issues in optical remote sensing. As such it was expected that this might be the case in this study. Analysis of the quality flags however revealed cloud contamination to be of minor concern (table 1). The percentage of pixels classed as “Marginal” quality was of greater concern as it, depending on the growth season, ranged from 17% to 41%, which is a significant portion of the time series.

Table 1: Pixel reliability information contained within the MOD13 version 5 quality flag layer for the study area.

Year	% of pixels classified as "Good" quality	% of pixels classified as "Marginal" quality	% cloud contaminated pixels
2001	56.5	41.1	2.3
2002	64.3	34.5	1.2
2003	75.6	23.6	0.8
2004	65.8	32.8	1.4
2005	82.6	17.0	0.4
2006	73.9	24.6	1.6

Marginal quality pixels are referred to in the MOD13 documentation as "useful", but have been produced using less than perfect data (USGS LP DAAC 2010). Less than perfect data can refer to pixels acquired at extreme viewing angles, pixels with clouds in adjacent pixels, pixels containing cloud shadow or pixels obtained under extreme aerosol interference. Determining the exact cause can be difficult because of the complexity of the detailed quality flag layer. The effects of the data quality issues on pixel values are likewise difficult to predict. If pixels obtained under extreme viewing angles are present they can result in increased VI values while cloud contamination can results in decreased VI values (Huete et al. 2002). These affects add increased variation to the temporal profile of a pixels VI values which is not related to actual changes in photosynthetic potential. Minimising this variation should therefore increase the correlation between VI values and herbaceous biomass.

Ideally one would like to tailor the adjustment of a pixel's VI value to the issue reflected in the pixel's quality flag. This would involve increasing the value of cloud contaminated pixels, decreasing those viewed at extreme angels and making no adjustment to pixels with no data quality issues. Doing so would be fairly complex, and time constraints meant that this could not be successfully pursued. Fortunately, a software tool called Timesat (Jonsson and Eklundh 2006), identified for use in deriving phonological cues used elsewhere in this study was found to offer a simple yet reportedly robust solution. The Timesat software fits a curve to the existing temporal profile of pixel values, smoothing it. New values can be generated from this curve in which the negative bias caused by clouds and atmospheric interference and the positive bias caused by extreme viewing angels should be significantly reduced. The software was used to run an adaptive Savitzky-Golay filtering algorithm over each pixel's NDVI and EVI temporal profiles to achieve this (Jonsson and Eklundh 2006). New values were then generated from the resulting curve (figure 9: 2.4, 2.5).

One of the questions identified in the literature review was 'for which dates should a growth season sum be performed'? Four options were included for evaluation in this study (figure 9: 2.7). The first was a summation over the same set of dates, September – April, each year this ensured that even early rains (September) and associated onset of herbaceous growth would be captured. April was used as the cut-off date to coincide with the time at which VCA field data is collected each year.

The second was a summation taking into account the variation in the onset of the growth season. This required the start and end dates for each growth season included in the study. Timesat identified the start and end of the growth season using the instant at which a curve fitted to the VI time series exceeded the user defined percentage of the pre-season VI minimum, in this case 20% (Jonsson and Eklundh 2006). The curve was fitted to the data using the Adaptive Savitzky-Golay filtering algorithm with 3 fitting steps of window size 2, 3 and 4 respectively. Points more than 2 standard deviations above the mean were also removed as they were assumed to be attributable to sensor errors (Figure 9: 2.6). The resulting start and end date raster were saved as flat binary files for use in the summation process. The end date raster was filtered to remove any end dates later than the 7th of April and replaced with that date to ensure that no summation included values from after the collection of the field data.

The third summation was designed to compensate for the removal of production by grazers. To do so a weighting was applied to the summation process based on the following reasoning:

1. The later in the season herbaceous growth occurs, the less likely it is to be removed by grazers before field measurements are taken.

2. The greater the removal through herbivory, the weaker the correlation between VI data from the early part of the growth season and the field estimates will be.
3. The VI data should therefore be summed in such a way as to increase the weight assigned to images as the season progresses. An arbitrary weighting of 30% for the first image in the time series and 100% for the last was chosen.

It was however acknowledged that these assumptions would be incorrect if:

1. There is significant vegetation growth early in the season
2. This activity is followed by extensive senescence of the herbaceous layer prior to the acquisition of the field data. The resulting error will be greatest in areas where herbaceous material remains palatable once senesced. This is more likely to occur on sweetveld occurring on basalts than sourveld occurring on granites.

The fourth and final summation was a combination of the variable dates based on actual growth season and the weighted summation described above.

All of the summations were conducted using scripts written in MATLAB. A single 16 day MODIS VI composite commencing on Julian date 97 (7^{th} April) was included to facilitate comparison between the summations and a single image. This date was chosen once again to correspond as close as possible to the time at which the field data was gathered. All 5 steps were then repeated using the values generated from the curve fitted to the EVI and NDVI time series in Timesat (figure 9: 2.8) to assess the effectiveness of the curve fitting in remedying the effects of cloud contamination and atmospheric interference. This resulted in 20 VI variables, 5 EVI variables and 5 NDVI variables from the raw data (figure 9: 2.9) and the same again from the fitted data (figure 9: 2.10).

3.3. Assessing the performance of the VI variables

Having created the VI variables, their correlation to the VCA herbaceous biomass field estimates needed to be assessed so that subsequent analysis would only involve the best correlated summation. A summary of how this was accomplished is contained in Figure 12.

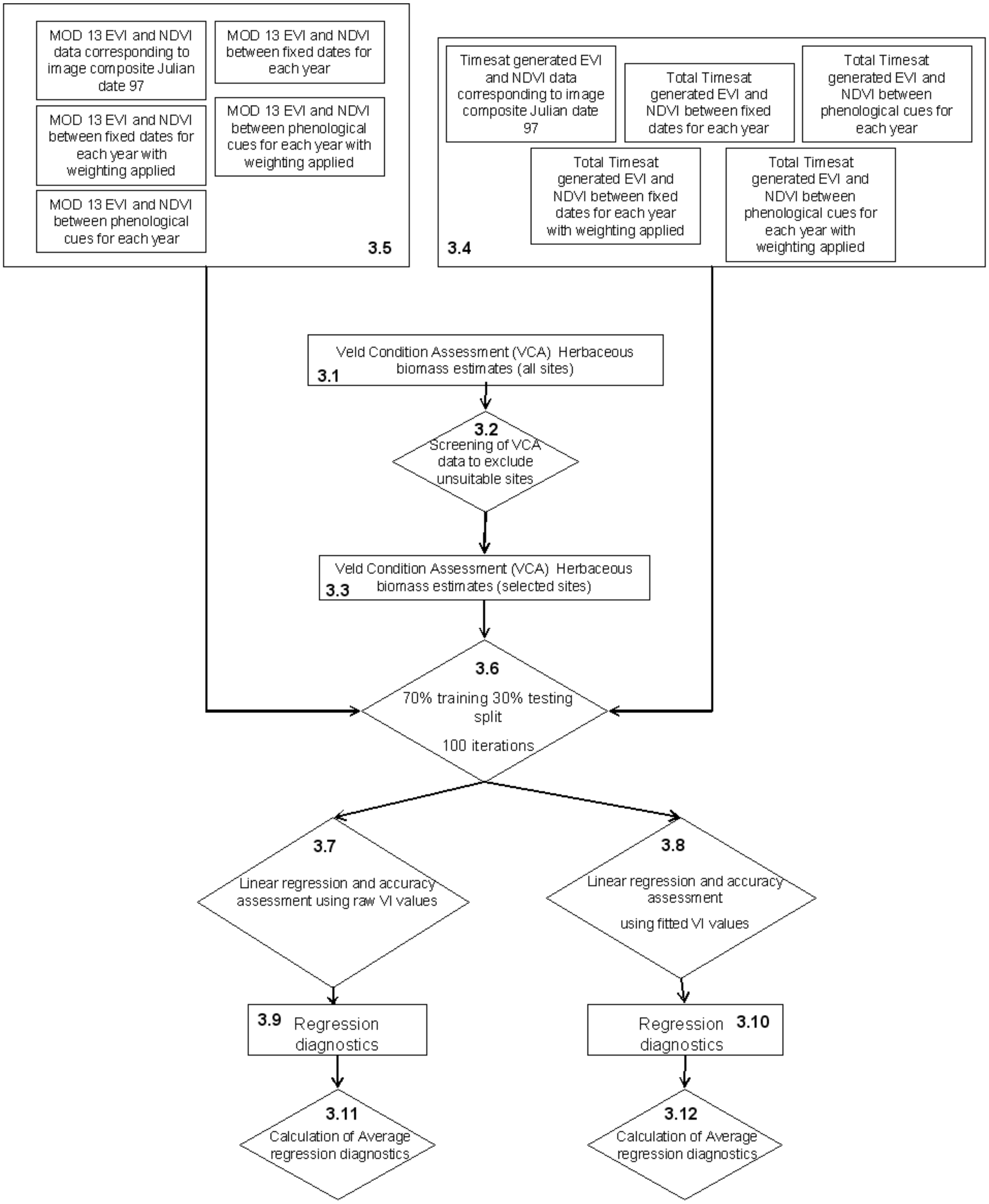


Figure 12: A summary of the process by which the best performing Vegetation Index variable derived for use in the study was selected for use in subsequent analysis steps.

The relative performance of the VI variables created was assessed by performing a simple linear regression between each variable and the herbaceous biomass field data for each growth season between 2000 and 2006. The field data used was once again drawn from the VCA dataset described previously (Figure 12: 3.1). The data was received as entries in an excel spreadsheet along with a shape file containing points indicating the location of each VCA site. The biomass estimates from the VCA data were assigned to their correct point location by using the unique site number identifier. Prior to inclusion in the regression the VCA data was screened to remove all sites within 500m of dams as well as 1st and 2nd order streams and rivers (Figure 12: 3.2). These sites were deemed to be too heterogeneous in terms of land cover to be included in the study. In all years, even after excluding highly heterogeneous sites, more than 400 acceptable sites were available (Figure 12: 3.3).

All of the VI variables (Figure 12: 3.4 and 3.5) were imported to ArcGIS and the pixel values underlying the VCA points extracted and added to the point files attribute table. The information in the attribute tables was then imported into the statistical program R. The following process was then run for each growth season:

1. A subset of 70% of the VCA sites was created (Figure 12: 3.6).
2. This data was used to train a regression model using each of the VI variables being assessed (Figure 12: 3.7, 3.8
3. The resulting models were used to predict the values of the remaining 30% of the data. Each of the models prediction accuracy and fit to the data was then determined by calculating RMSE and R^2 using the corresponding VCA biomass estimates as ground truth data.
4. This process was repeated 100 times and the average RMSE and R^2 for each growth season calculated. The best performing model based RMSE was then selected for use in further analysis steps (Figure 12: 3.11 and 3.12).

Monthly rainfall was obtained much later in the study in an attempt to aid understanding of the variability in the performance of the summations. This was plotted on bar charts and inserted alongside the above results.

3.4. Woody canopy cover derivation and accuracy assessment

Having selected the best VI variable for use as the herbaceous biomass production variable, work began on selecting an explanatory variable providing information on the presence of woody vegetation. The global MODIS MOD44 tree cover product was at the time of writing one of the few tree cover products available. The tree cover map for Kruger National Park derived by Gabriella Bucini's, regarded at the time of the final corrections to this dissertation as the most accurate tree cover map available, was not known to the author early enough in the study's project cycle for it to be made use of. Initial inspection of the MOD44 product raised doubts over its accuracy, especially in the relatively low tree cover of the Lowveld where the study area is located. Two approaches for deriving information on the presence of woody vegetation were therefore selected from the literature to be evaluated in this study in addition to the MODIS MOD 44 tree cover product. The first relied on the response of VI signal to rainfall relative to the temporal mean VI values (Scanlon et al. 2002). The second attempted to exploit the seasonal differences in woody and herbaceous vegetation activity (Archibald and Scholes 2007). The MODIS MOD 44 product estimates canopy cover based on reflectance in a number of different bands using a supervised regression tree algorithm (Hansen et al. 2002). Further details on the production and assessment of all three are presented below; Figure 13 provides a summary of the process.

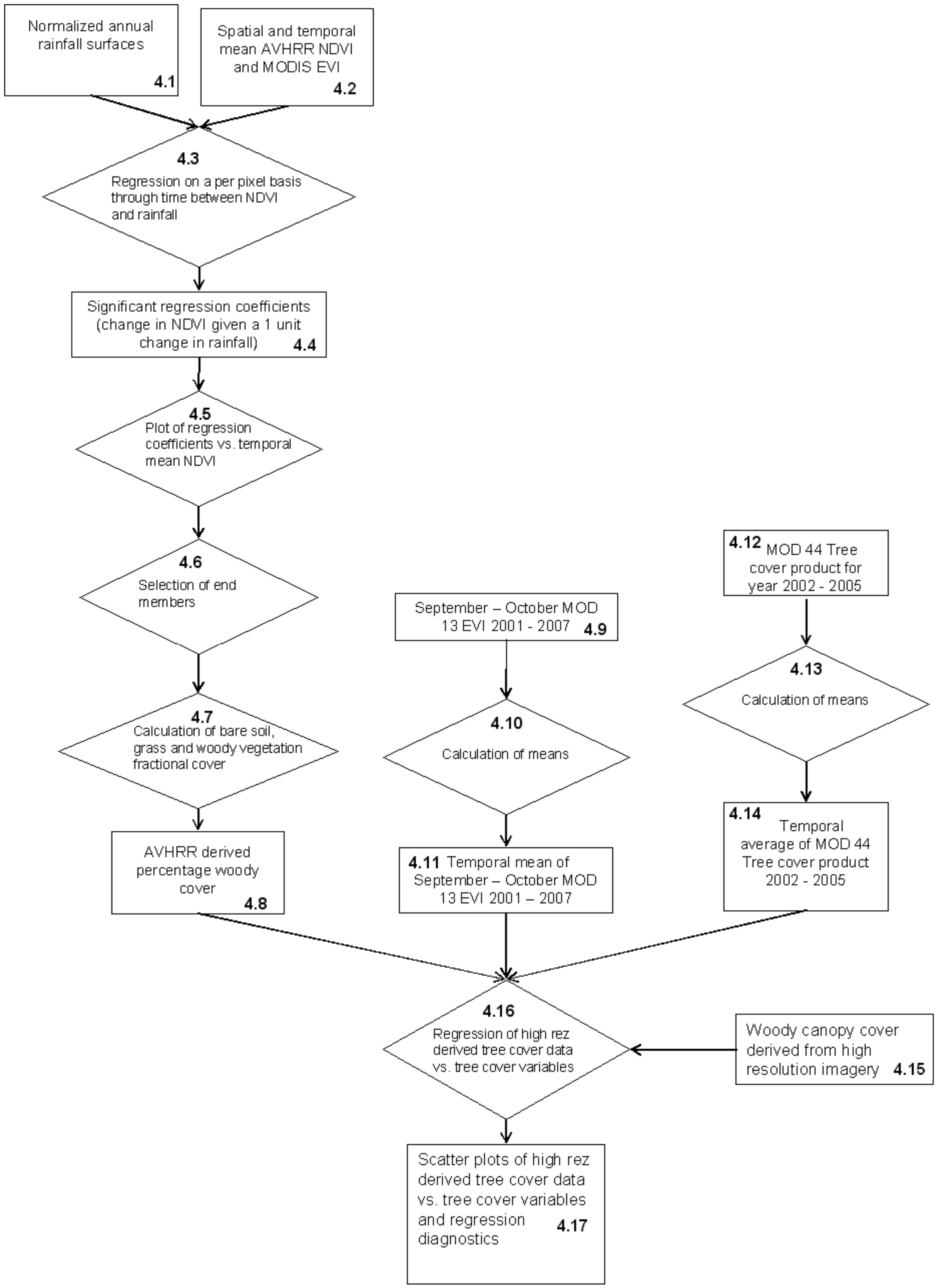


Figure 43: A summary of the process by which the variables used to account for the effect of woody vegetation on the VI – herbaceous biomass relationship were derived and evaluated. 'Selection of end members' in 4.6 refers to identifying pixels composed entirely of each of the cover types present. In the method adapted from Scanlon et al. (2002), these cover types were trees, grass and bare soil.

A method for deriving fractional woody vegetation cover adapted from a paper by Scanlon et al. (2002) was carried out using both AVHRR derived NDVI at 1km resolution for the years 1985 - 2003 and the MOD13 EVI at 250m resolution for the years 2001 – 2008 (Figure 13: 4.1 – 4.8). Unlike in the original method detailed in the paper, no spatial averaging was used as a spatially explicit fractional cover was needed. All calculations outlined below are therefore on a per pixel basis.

Temporal mean MODIS EVI and AVHRR NDVI images were calculated for each wet season in the time series. The wet season was taken to be between November and February for the study area. The long term temporal mean vegetation index values for all of the wet season mean images were also calculated (Figure 13: 4.2). Total rainfall for the wet season was normalised by subtracting the temporal mean rainfall and dividing by the temporal standard deviation in rainfall (Figure 13: 4.1).

OLS regression was performed using the time series of values from each pixel in the study area with normalized rainfall as the independent/predictor variable and the wet season vegetation index value as the dependent/response variable (Figure 13: 4.3). The beta coefficient for the pixel was recorded only if the regression has a P value of < 0.1. If the relationship was not significant, the pixel was excluded from further analysis (Figure 13: 4.4). It was found that less than 20% of the pixel time series regressions involving the MODIS data showed a significant relationship between VI and Rainfall. Less than 20% coverage of the study area was deemed insufficient coverage to be useful as tree cover could only be generated for those pixels with a significant relationship. Further assessment of the MODIS derived fractional cover product was therefore abandoned.

The significant coefficients were plotted against corresponding long term temporal mean vegetation index values (Figure 13: 4.5). The resulting plot was then used to visually select end member values for 100% tree cover, 100% grass cover and 100% bare soil (Figure 14 & Figure 13: 4.6).

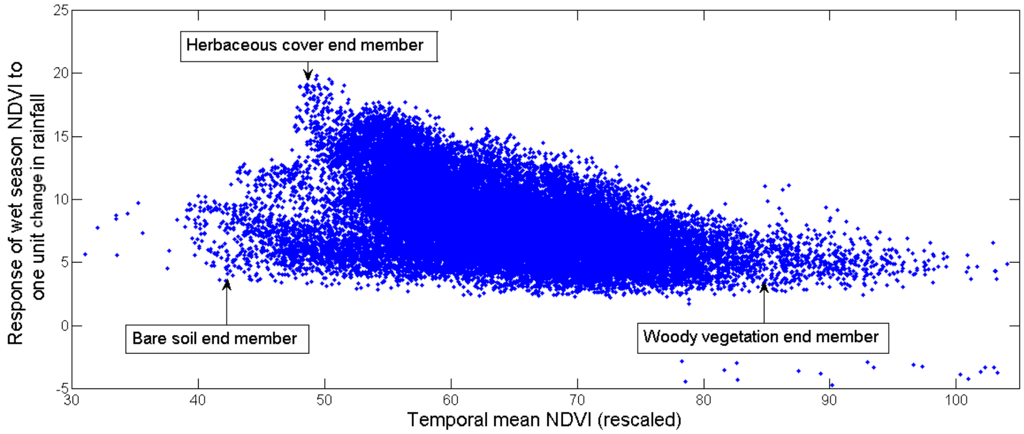


Figure14: End member selection for the AVHRR derived canopy cover method.

Fractional cover was then determined by solving equations 1, 2 and 3 simultaneously (Figure 14: 4.7).

1. $X_t(i) + X_b(i) + X_g(i) = 1$

2. $EVI_t(X_t(i)) + EVI_b(X_b(i)) + EVI_g(X_g(i)) = EVI(i)$

3. $\beta_t(X_t(i)) + B_b(X_b(i) + B_g(X_g(i)) = B(i)$

Where:

EVI(i) is the temporal mean vegetation index value of pixel (i)

EVI_t, EVI_b and EVI_g are the temporal mean EVI component of the tree, bare soil and grass end member values respectively

B(i) is the regression coefficient of pixel (i)

B_t, B_b and B_g are the regression coefficient component of tree, bare soil and grass end member values respectively

$X_t(i)$, $X_b(i)$ and $X_g(i)$ are the fraction of pixel (i) made up of tree, bare soil and grass.

The raster containing the values for the fraction of each pixel made up of trees was then imported to ArcGIS for use in later analysis (Figure 13: 4.8).

The second tree cover variable derived was the mean MODIS EVI for September and October of each year. The variable was an attempt to exploit the fact that certain trees green up before grasses in the study area and so dominate VI signal in September and October. The new leaves on the trees are the dominant source of green vegetation before the herbaceous layer begins growth again after inactivity in the dry season. The mean MODIS EVI for the September and October was calculated (Figure 13: 4.9 – 4.10) and the temporal average created to minimise the potential affect of large EVI decreases caused by fire and not changes in the tree-grass balance (Figure 13: 4.11). Because those species of trees that green up before grasses are not uniformly distributed across the study area the above method will be more effective in some areas than others.

The third tree cover variable, the MODIS v4 MOD 44 canopy cover product (Figure 13: 4.12), was obtained for the years 2002 - 2005 from the Global Land Cover Facility webpage (http://glcf.umiacs.umd.edu/data/vcf/). The product is derived using a supervised regression tree algorithm using 7 of the MODIS bands and 250 classified Landsat images as training data (Hansen et al. 2002). The product documentation defines trees as woody structures >5m. However, as the product is heavily reliant on the NDVI signal, it is not clear how other green vegetation is prevented from influencing the tree cover value reported by the product. For a more comprehensive account of the method, refer to the MOD44 users guide available online through the MODIS website (http://modis.gsfc.nasa.gov). Of the variation accounted for in the method, 70% arose from splits based on red reflectance levels. The presence of chlorophyll is the major determinant of red reflectance as chlorophyll absorbs in this region of the spectrum. Fire and fluctuations in herbaceous production related to rainfall are likely to introduce a large amount of variation into a product depending predominantly on the levels of chlorophyll present. This may not be a major issue where vegetation is dominated by evergreen trees. It is however a major issue when vegetation is dominated by highly variable herbaceous species as is the case in the study area. Initial inspections of the layers showed a larger inter-annual variation in canopy cover than seems likely (Figure 15).

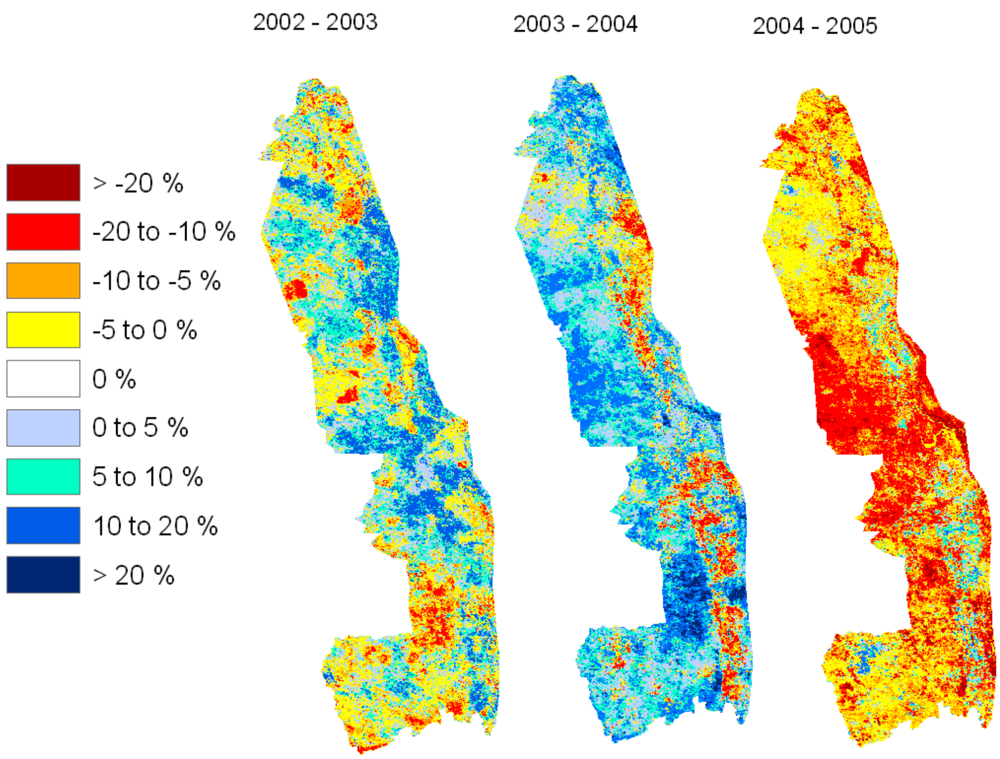


Figure15: Change in MOD44 canopy cover values for the periods indicated.

The validation results published for the MOD 44 product confirm the concerns over the product's accuracy. The limited agreement between measured and predicted values evident in Figure 16 suggests that the product is not suitable for use in quantitative studies. As such, expectations of the product's performance in this study were fairly low.

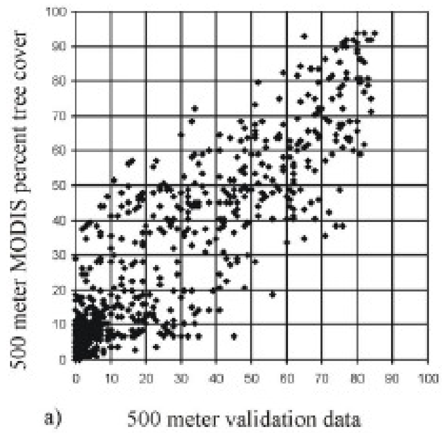


Figure 16 MOD 44 canopy cover validation results (Hansen et al. 2003). The R^2 value for the regression is given as 0.89.

In an attempt to compensate for what seems to be unrealistically large inter-annual variations in canopy cover, the product was averaged for the entire period (Figure 13: 4.13), producing a single temporally average MOD 44 tree cover layer (Figure 13: 4,14).

The accuracy of all three layers was assessed by regressing them against degraded 10m resolution canopy cover estimates derived from a combination of LiDAR; aerial photography and IKONOS Imagery (Figure 13: 4.16). The estimates were derived and provided courtesy of Dr. Sean Levick from the University of the Witwatersrand (now employed by the Carnegie Institute) and Russel Main and Melanie Vogel from the CSIR Ecosystems and Earth Observation department (Figure 13: 4.15). Scatter plots with R^2, P value and regression lines shown were then produced for comparison (Figure 13: 4.17).

3.5. Woody canopy cover re-evaluation

The regression of the woody canopy cover testing data against the Tree cover variables derived for this study did not reveal any statistically significant relationships. This meant that one or more of the following was true:

1. The tree cover variables investigated for use in this study contained little information on tree cover
2. The testing data was not an accurate measure of tree cover
3. Error was introduced when the testing data had its resolution degraded to match that of the tree cover variables being assessed.

Visual inspection of the tree cover variables by researchers with knowledge of the actual conditions on the ground suggested that the tree cover variables showed at least broad agreement with actual patterns of tree density in the study area. Inspection of the testing data alongside the high resolution imagery it was derived from suggested that the testing data reflected tree cover fairly accurately. This allowed possibilities 1 and 2 to be ruled out.

At the time the analysis was performed, possibility 3 did not register as a concern because I was not fully aware of the need to take into account PSF when degrading the spatial resolution of imagery to match coarse resolution imagery on a per pixel basis. In retrospect it seems the most likely candidate for the lack of correlation in this case is the September - October mean variable. It also seems likely that it would significantly affect the correlation with the other two tree cover variables as they were derived from similar coarse resolution imagery. In the absence of this knowledge, rather than attempting to correct for the PSF and re-running the accuracy assessment, a new approach was pursued. It was decided to re-assess the performance of the Tree cover variables based on their ability to account for additional variation in the regression between the best VI variable identified during previous analysis and the herbaceous biomass field estimates. The process for doing so is summarised in Figure 17 below.

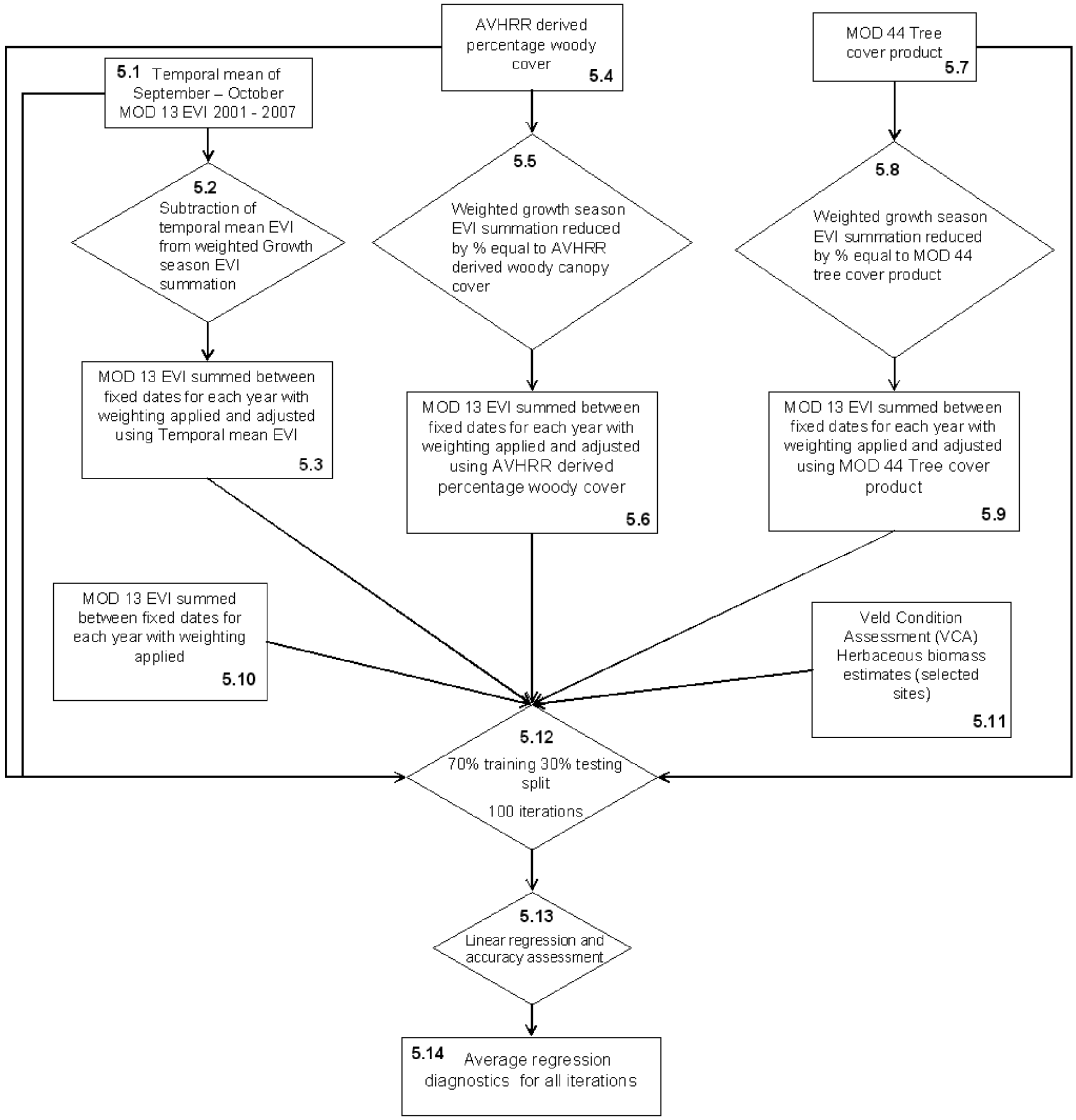


Figure17: A summary of the process by which the tree cover variables considered were re-assessed based on their ability to account for additional variation in the previously created herbaceous biomass estimation models.

Pre correction of the VI variables was conducted by reducing each pixels VI value by an amount proportional to the corresponding pixel value of one of the tree cover variables. In the case of the temporal mean September – October VI value tree cover variable (Figure 17: 5.1), this was achieved by subtracting it from each of the images used to create the original fixed date summations and then once again performing a weighted summation (Figure 17: 5.2). In the case of the two percentage measures (Figure 17: 5.4 and 5.7), the VI variable was simply reduced by the percentage tree cover indicated (Figure 17: 5.5 and 5.8), under the assumption that there was a 1:1 relationship between tree cover and VI contribution.

Performing the pre analysis correction provided four possible VI variables for each year, the original VI summation (Figure 17: 5.10) and three others adjusted using the tree cover variables (Figure 17: 5.3, 5.6 and 5.9). All of the VI variables (Figure 17: 5.3, 5.6, 5.9 and 5.10), as well as the tree cover variables (Figure 17: 5.1, 5.4 and 5.7), were imported to ArcGIS and the pixel values underlying the VCA points (Figure 17: 5.11) extracted and added to the point files attribute table. The information in the attribute tables was then imported into the statistical program R and the following process carried out for each growth season:

1. A subset of 70% of the VCA sites was created

2. This data was used to train a total of ten competing linear models. The first model had only the weighted summation of EVI (Figure 17: 5.10) as a predictor variable. Three models with the adjusted EVI variables (Figure 17: 5.3, 5.6 and 5.9) as the predictor variables were then trained. This was followed by three models with both the weighted summation of EVI (Figure 17: 5.10) and one of the tree cover variables (Figure 17: 5.1, 5.4 and 5.7) as predictor variables. Finally a last set of three models using the same set of variables as above but with interactions between the EVI summation and the tree cover variables were trained.

3. The resulting models were used to predict the values of the remaining 30% of the data. Each of the models prediction accuracy and fit to the data was then determined by calculating RMSE and adjusted R^2 using the corresponding VCA biomass estimates as ground truth data.

4. This process was repeated 100 times and the average RMSE and adjusted R^2 calculated for each growth season (Figure 17: 5.12).

The change in adjusted R^2 for each model relative to the base model was assessed and those not registering an increase rejected. The best performing of the remaining models, based RMSE, was then selected for use in further analysis steps.

3.6. Production of final regression model

Having identified the best performing tree cover variable and the most successful means of incorporating it into a regression model, a final refinement, the addition of a fire history variable, was investigated. The fire history variable was pursued in an attempt to account for the presence of dry material, a factor identified as having the potential to have a significant effect on the accuracy of a VI based prediction model (Thompson and Everson 1993). In addition to the usual regression diagnostics produced in previous modelling steps, the final model created for each year was used to produce herbaceous biomass estimates for the entire study area. Figure 18 provides a summary of the process.

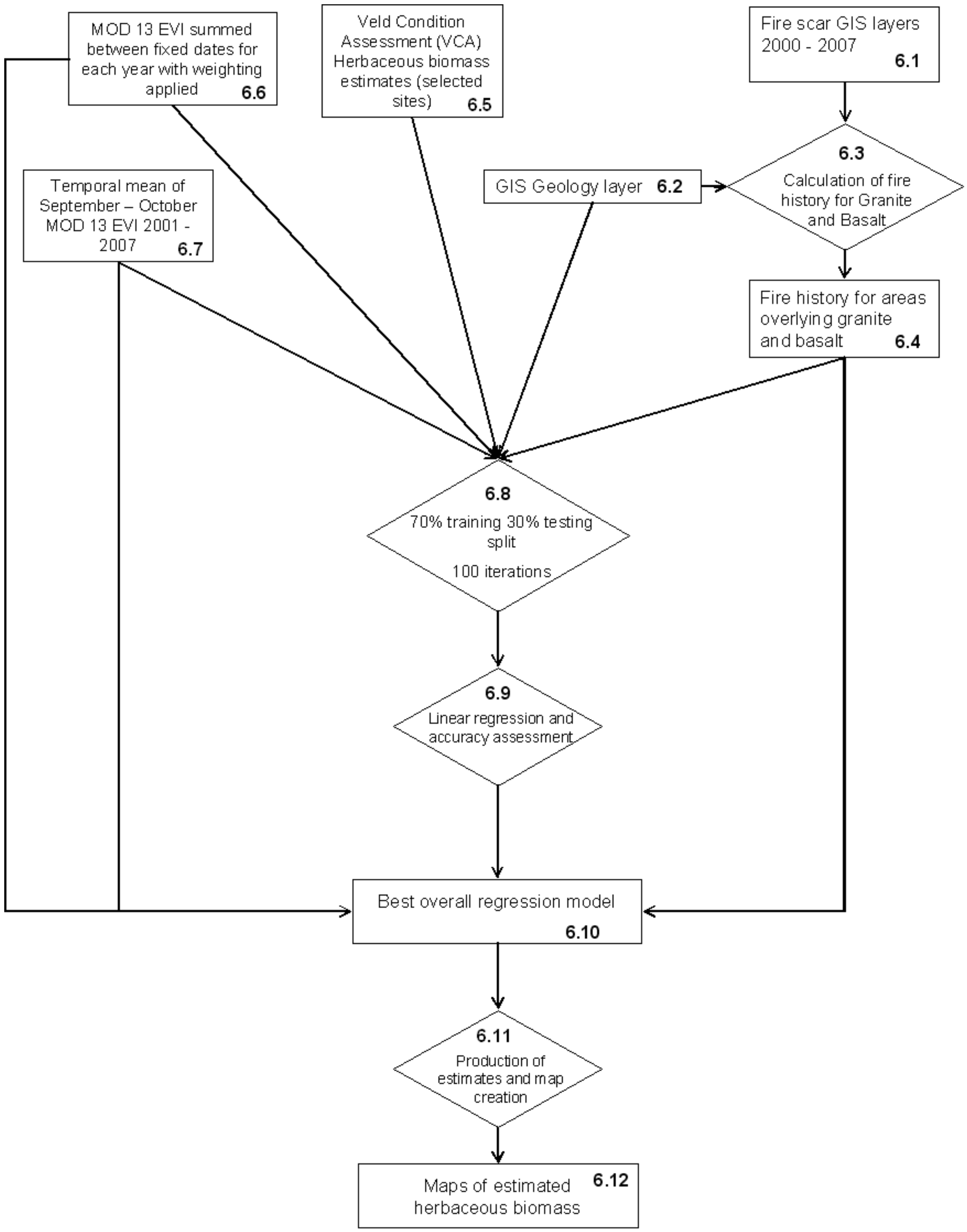


Figure 18: A summary of the process undertaken to create the final regression models used in this study. The arrows feeding into box 6.10 indicate variables that were used as inputs into the regression model identified as the most accurate predictor of herbaceous biomass.

Initially when searching for a suitable variable to account for dry material, VI summations from previous seasons were considered as they would provide an indication of past production. Average fire return period for the KNP is approximately 4.5 years (Van Wilgen et al. 2003). Production estimates for a maximum of four years prior to the season of interest, but fewer when a fire had occurred more recently, would therefore be sufficient to create a variable to account for carry-over. Attempting this approach would have necessitated the creation of a fairly complex model containing the following:

1. Separate VI variables for each of the season's prior to the season of interest until the first of the following is reached:
    - The season when the veld last burnt is reached
    - Four seasons without a burn event have lapsed

2. A three way interaction between each of these VI variables and variables accounting for:
    - Differences in herbivory. This would be required because the relationship between production and carry-over will be affected by the level of herbivory occurring in the dry season. Even a weighted summation of VI values could only account for herbivory occurring within the growth season.
    - The number of seasons between the season of interest and the season in which production occurred. This would be required because the greater the time lag, the more likely the material present is to be eaten.

There is no record of this approach being implemented in the literature, possibly because of the large amount of data required relative to the expected effectiveness of including such a variable. A simpler approach was therefore sought given the limited time and resources available for this project.

The approach decided on was to produce a categorical variable that was a combination of both fire history and geology as both affect accumulation of dry material. The effects of fire are obvious; burn events remove the majority of accumulated dry material resulting in areas most recently burnt having little dry material in the herbaceous layer. The effect of geology on the accumulation of dry material is more complex and is related to both production and herbivory which are controlled to some extent by underlying geology. The study area can be broadly divided into fertile regions located on basalts and infertile regions located on granites (Venter, Scholes and Eckhardt 2003). Infertile soils lead to lower herbaceous production and less palatable herbaceous material of little grazing value once dry. Fertile soils on the other hand enable greater herbaceous production and result in more palatable herbaceous material that remains palatable even once dry. It was therefore assumed that removal through herbivory would be higher where these soils occur.

By combining geology and fire history (Figure 18: 6.3), it was thought that the resulting categorical variable comprised of categories detailing geology and time since last burnt(Figure 18: 6.4) would improve estimation accuracy by allowing for variation in the intercept of the regression model. Information on time available for accumulation was provided by fire history determined using fire scar maps obtained from the KNP Scientific Services (Figure 18: 6.1). Information on differences in potential productivity and hence accumulation rate was provided by a geological layer indicating the location of granite and basalt parent material (Figure 18: 6.2). It is acknowledged that this fails to include the effect of rainfall on actual production.

The combined fire history and geological categorical variable, was imported to ArcGIS and the pixel values underlying the VCA points extracted and added to the point files attribute table already containing the best performing VI (Figure 18: 6.5) and tree cover variables (Figure 18: 6.6) identified in previous analysis steps. The information in the attribute tables was then once again imported into the statistical program R and the following process implemented for each growth season:

1. A subset of 70% of the VCA sites was created
2. This data was used to train a total of two competing linear models. The first model included the VI variable and the tree cover variable as predictors. The second contained the VI variable, the tree cover variable and the combined fire history – geology categorical variable.
3. The resulting models were used to predict the values of the remaining 30% of the data. Each of the models prediction accuracy and fit to the data was then determined by calculating RMSE and adjusted R^2 using the corresponding VCA biomass estimates as ground truth data.
4. This process was repeated 100 times and the average RMSE and adjusted R^2 calculated for each growth season (Figure 17: 5.12).

Average RMSE and adjusted R^2 values were calculated for each model and entered into tables for comparison. The change in adjusted R^2 for the models containing the geology-fire history variable relative to the simpler model was assessed. All of the more complex models registered an increase in adjusted R^2 and so were accepted for further scrutiny. The model providing the greatest average decrease in RMSE between seasons (Figure 18: 6.10), was re-trained using all the points for a particular year, and the resulting model used in the production of spatially explicit biomass estimates. The estimates were then imported back into Arc GIS and presented as maps (Figure 18: 6.12).

3.7. Cokriging

Having completed the regression component of the analysis for the study, attention was turned to producing estimates using cokriging. A summary of the process is provided below in figure 19.

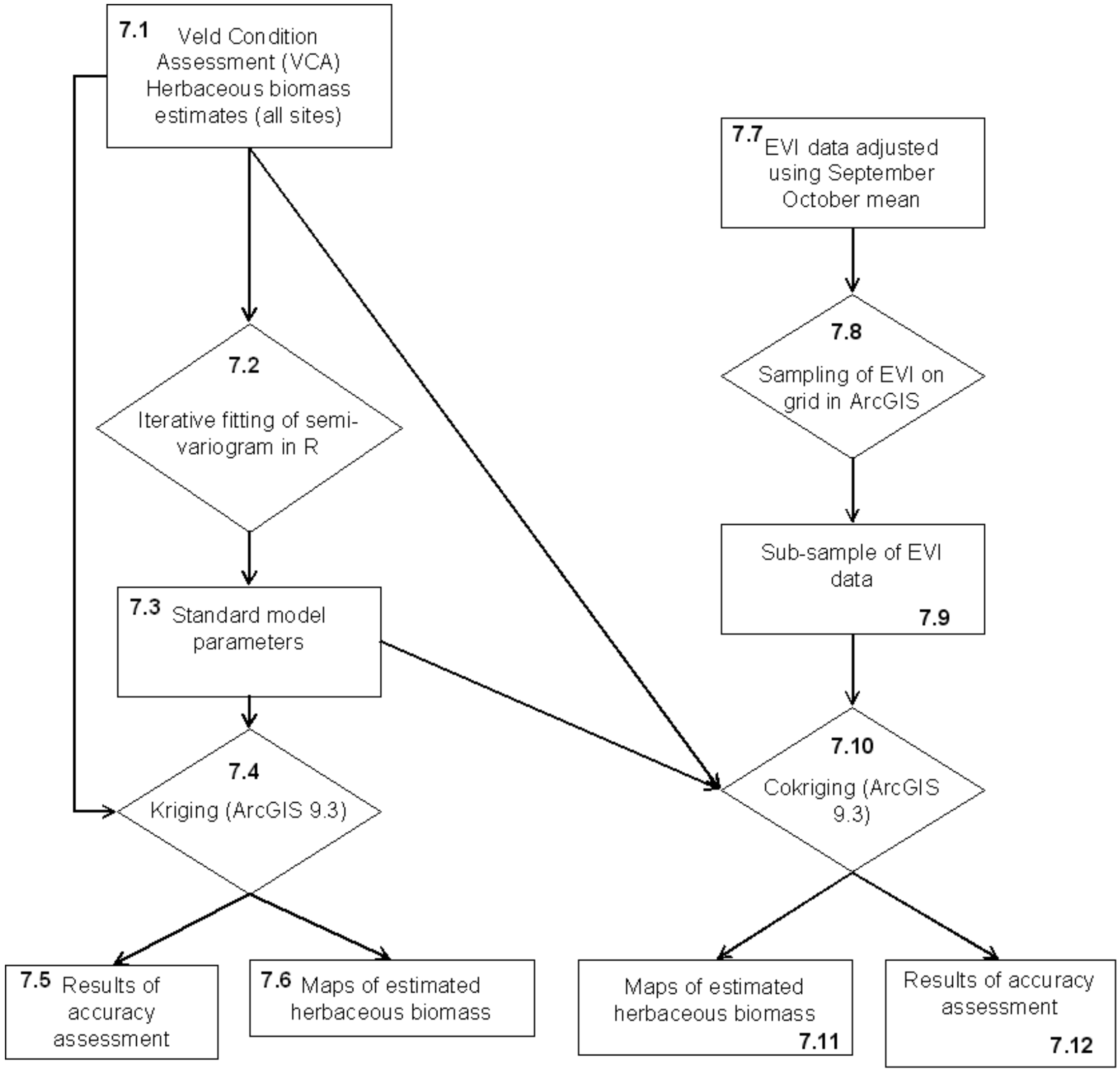


Figure 19: A summary of the process involved in producing herbaceous biomass estimates using geo-statistical interpolation in this study.

The VCA herbaceous biomass data for each year were imported into R (Figure 19: 7.1). Standard models were fitted using the 'sgeostat' package's iterative fitting command (Figure 19: 7.2), with the type of model selected based on visual inspection of the semi-variograms. The iterative fitting command cycles through a user defined number of iterations, altering the nugget and sill values, and selects the values that result in the best fit of the model type specified to the points on the semi-variogram. Because of difficulties experienced in successful use of the 'sgeostat' package, the actual kriging and cokriging interpolations were performed using the

ArcGIS 9.2 geostatistical wizard. The standard model parameter values determined through iterative fitting of the models to the semi-variogram data in R were entered into ArcGIS and the geostatistical wizard was allowed to optimise the number of lags, range, sill and nugget values. This provided a final set of parameters (Figure 19: 7.3) used by ArcGIS to perform the kriging (Figure 19: 7.4). The software provides two standard outputs when performing any type of kriging. The first is a set of cross-validation accuracy assessment figures (Figure 19: 7.5) and the second the actual kriged maps of herbaceous biomass (Figure 19: 7.6). As was the case in Mutanga and Rugege (2006) I used the default set of cross validation results to assess the accuracy of the kriged and cokrieged maps in this study. Cross-validation makes use of the entire set of training data to estimate the trend and autocorrelation models. Data points are then withheld from the dataset one at a time or in randomly selected subsets. A surface is interpolated using the remaining points and its accuracy assessed by comparing the actual value for the withheld validation points to the interpolated value. The validation points are then returned to the dataset and a new subset selected and the process repeated, until all points have been used in validation. Finally an average is calculated based on all of the accuracy assessment gathered (ESRI 2001).

Once again the VCA field estimates were used as the primary variable (Figure 19: 7.1 feeding into 7.10). The strongest correlated EVI variable identified in section three of this analysis that could be used as the secondary variable for cokriging was the weighted summation of growth season EVI adjusted for tree cover using the September – October temporal mean EVI (Figure 19: 7.7). Owing to the software repeatedly freezing when attempting to use all 304179 EVI points as secondary data, a 7000 point subset was randomly selected for each year in addition to the ±400 co-located with the VCA points (Figure 19: 7.9). The same standard model parameters as used for kriging were manually entered and the software was allowed to optimise the cokriging parameters (Figure 19: 7.10). Once again the software outputs the results for both the cross-validation accuracy assessment (Figure 19: 7.11) and the cokriged maps of herbaceous biomass (Figure 19: 7.12).

3.8. Comparison of methods

The regression model and cokriging approaches were compared by calculating the mean accuracy and standard deviation of accuracy for the study period. The difference between the herbaceous biomass prediction maps was calculated to identify if there were any patterns in how the predictions differed. Histograms of the residuals were calculated to determine if either method showed biased residuals. The residuals were plotted onto a map of the study area with symbols proportionate to the size of the residual to check for spatial trends in the magnitude of the residuals.

4. REFERENCES

Al-Bakri, J.T., and J. C. Taylor. 2003. Application of NOAA AVHRR for monitoring vegetation conditions and biomass in Jordan. *Journal of Arid Environments* 54: 579 - 593.

Archibald, S., and R.J. Scholes. 2007. Leaf green-up in a semi-arid African savanna-separating tree and grass responses to environmental cues. *Journal of Vegetation Science* 18, no. 4: 583-594.

Diallo, O., A. Diouf, N. P. Hanan, A. Ndiaye, and Y. Prevost. 1991. AVHRR monitoring of savanna primary production in Senegal, West Africa: 1987-1988. *International Journal of Remote Sensing* 12, no. 6: 1259-1279.

Didan, K., and A. Huete. 2006. MODIS Vegetation Index Product Series Collection 5 Change Summary. *TBRS Lab., University of Arizona*.

ESRI. 2001. Arc GIS Geostatistical Analysis: Statistical Tools for Data Exploration, Modelling, and Advance Surface Generation. http://www.esri.com/library/whitepapers/pdfs/geostat.pdf s

Hansen, M. C., R. S. DeFries, J. R. G. Townshend, R. Sohlberg, C. Dimiceli, and M. Carroll. 2002. Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. *Remote Sensing of Environment* 83, no. 1: 303-319.

Huang, C., J. R. G. Townshend, S. Liang, S. N. V. Kalluri, and R. S. DeFries. 2002a. Impact of sensor's point spread function on land cover characterization: assessment and deconvolution. *Remote Sensing of Environment* 80, no. 2: 203-212.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment* 83, no. 1: 195-213.

Huete, A., C. Justice, and W. Van Leeuwen. 1999a. MODIS vegetation index (MOD13) algorithm theoretical basis document. *Greenbelt: NASA Goddard Space Flight Centre, http://modarch. gsfc. nasa. gov/MODIS/LAND/# vegetation-indices*.

Huete, A. R., K. F. Huemmrich, T. Miura, X. Xiao, K. Didan, W. van Leeuwen, F. Hall, and C. J. Tucker. 2006. Vegetation Index greenness global data set. *College Park, MD*.

Jonsson, P, and L Eklundh. 2006. TIMESAT - a Program for Analyzing Time-Series of Satellite Sensor Data. Users Guide for TIMESAT 2.3. http://www.nateko.lu.se/personal/Lars.Eklundh/TIMESAT/timesat2_3_users_manual.pdf.

Moreau, S., R. Bosseno, X. Fa Gu, and F. Baret. 2003. Assessing the biomass dynamics of Andean bofedal and totora high-protein wetland grasses from NOAA/AVHRR. *Remote Sensing of Environment*, no. 85: 516-592.

Mutanga, O., and D. Rugege. 2006. Integrating remote sensing and spatial statistics to model herbaceous biomass distribution in a tropical savanna. *International Journal of Remote Sensing* 27, no. 16: 3499-3514.

Prince, S. D. 1991. Satellite remote sensing of primary production: comparison of results for Sahelian grasslands 1981-1988. *International Journal of Remote Sensing* 12, no. 6: 1301-1311.

Prince, S. D., and C. J. Tucker. 1986. Satellite remote sensing of rangelands in Botswana II. NOAA AVHRR and herbaceous vegetation. *International Journal of Remote Sensing* 7, no. 11: 1555-1570.

Sannier, C. A. D., J. C. Taylor, and W. D. Plessis. 2002. Real-time monitoring of vegetation biomass with NOAA-AVHRR in Etosha National Park, Namibia, for fire risk assessment. *International Journal of Remote Sensing* 23, no. 1: 71-89.

Scanlon, T. M., J. D. Albertson, K. K. Caylor, and C. A. Williams. 2002. Determining land surface fractional cover from NDVI and rainfall time series for a savanna ecosystem. *Remote Sensing of Environment* 82, no. 2: 376-388.

Tan, B., C. E. Woodcock, J. Hu, P. Zhang, M. Ozdogan, D. Huang, W. Yang, Y. Knyazikhin, and R. B. Myneni. 2006. The impact of gridding artifacts on the local spatial properties of MODIS data: Implications for validation, compositing, and band-to-band registration across resolutions. *Remote Sensing of Environment* 105, no. 2: 98-114.

Thompson, M.W., and C.S. Everson. 1993. Development of spectral-biomass models for mapping and monitoring montane grassland resources. Division of Forest Science and Technology CSIR.

Trollope, W, and A. L. F. Potgieter. 1986. Estimating grass fuel loads with a disc pasture meter in the Kruger National Park. *African Journal of Range and Forage Science* 3, no. 4.

Trollope, W. S. W. 1996. Biomass Burning in the savannas of Southern Africa with Particular Reference to the Kruger National Park in South Africa. In *Biomass Burning and Global Change: Remote sensing, modeling and inventory*, ed. J.

S. Levine, 1:260 - 269. 1st ed. Massachusetts Institute of Technology.

USGS LP DAAC. 2010. MOD13Q1 | 16-Day L3 Global 250m | Vegetation Indices | MODIS Products Table | Products | LP DAAC :: ASTER and MODIS Land Data Products and Services. https://lpdaac.usgs.gov/lpdaac/products/modis_products_table/vegetation_indices/16_day_l3_global_250m/mod13q1.

Van Wilgen, B.W., W.S.W. Trollope, H.C. Biggs, A. L. F. Potgieter, and B.H. Brockett. 2003. Fire as a Driver of Ecosystem Variability. In *The Kruger Experience*, ed. J. du Toit, H. Biggs, and K.H. Rogers, 149-170. Island Press.

Venter, F.J., R.J. Scholes, and H.C. Eckhardt. 2003. The Abiotic Template and Its Associated Vegetation Pattern. In *The Kruger Experience*, ed. J. du Toit, H. C. Biggs, and K.H. Rogers, 83-129. Island Press.

Wessels, K. J., S. D. Prince, N. Zambatis, S. MacFadyen, P. E. Frost, and D. Van Zyl. 2006. Relationship between herbaceous biomass and 1km 2 Advanced Very High Resolution Radiometer (AVHRR) NDVI in Kruger National Park, South Africa. *International Journal of Remote Sensing* 27, no. 5: 951-973.

Wolfe, R. E., D. P. Roy, and E. Vermote. 1998. MODIS land data storage, gridding, and compositing methodology: Level 2 grid. *IEEE Transactions on Geoscience and Remote Sensing* 36, no. 4: 1324-1338.

Zambatis, N. 2002. Revised procedures for veld condition assessment in the Kruger National Park. *Unpublished internal report, SANParks, Skukuza*.

Chapter 3

RESULTS

1. OVERVIEW

The results arising from the various analysis steps have been grouped as follows:

1. Assessment of the affects of the VCA sample site dimensions (section 1 on methods flowchart)
2. Creation of a regression model:
    - 2.1. Assessing the performance of the Vegetation Index variables (section 3 on methods flowchart)
    - 2.2. Assessment of a suitable woody canopy cover variable (sections 4 and 5 on methods flowchart)
    - 2.3. Assessment of a suitable variable to account for dry / moribund material (section 6 on methods flowchart)
    - 2.4. Assessment of the completed regression models.
3. Kriging and Cokriging (section 7 on methods flowchart).
4. Comparison of the accuracy and precision of the regression model and cokriging approaches (section 8 on methods flowchart).

2. ASSESSMENT OF THE ADEQUACY OF VCA SAMPLE PLOT DIMENSIONS

Table 1 contains the results from the analysis of the discrepancy in mean herbaceous biomass measured using a DPM on co-located 60 x 60m and 250 x 250m sample sites at the end of the 2007-2008 growth season. Biomass within 250 x 250m sites differed from that within co-located 60 x 60m sites between 42 kg/ha and 1308 kg/ha with a mean difference of 556 kg/ha and standard deviation of the differences of 410 kg/ha. On three occasions the biomass estimates for the 60 x 60 m plot were greater than biomass estimates for the 250 x 250 m, while it was the opposite for the other 5, and hence no directional bias was evident. No research was done in this study into what caused some sites to show a greater discrepancy in biomass estimates than others. The take home message here was that the discrepancy exists and researchers need to be aware of this source of error.

Table 1: The difference in mean biomass estimates based on pasture meter readings for co-located 60 x 60m and 250 x 250m sample sites in the Kruger National Park at the end of the 2007-2008 growth season. SD refers to DPM readings within each site. Refer back to the map in Figure 7 of chapter 2 for site locations.

Site	Mean and SD Biomass measurement 60*60m (kg/ha)	Mean and SD Biomass measurement 250*250m (kg/ha)	Difference(kg/ha)	Vegetation type
1	4163 (1478)	3647, (1276)	-516	Terminalia/Rock *Ficus* Sour Bushveld
2	4477 (1690)	4020, (1493)	-457	Terminalia/Rock *Ficus* Sour Bushveld
3	3087 (1139)	3045, (1268)	-42	Knob Thorn/Dichrostachys Thorn Thickets
4	2035 (632)	2831, (1302)	+796	Knob Thorn/Dichrostachys Thorn Thickets
5	2184 (1501)	3492, (1937)	+1308	Knob Thorn/Dichrostachys Thorn Thickets
6	5067 (1908)	5223, (1262)	+156	Knob thorn/Marula tree Savanna
7	5497 (1616)	5850, (1336)	+353	Knob thorn/Marula tree Savanna
8	5003 (1109)	5821, (1697)	+818	Knob thorn/Marula tree Savanna

3. CREATION OF THE REGRESSION MODELS

3.1. Assessing the performance of the vegetation index variables

Two sets of imagery were used for each growth season. The first set of imagery was used without any pre processing of the data apart from Reprojection. This set of images is referred to as the ‘raw’ data in this study. The second set of images was produced from the raw data by fitting a curve to each pixel’s time series and generating new pixel values from that curve. This was intended to minimise the effects of cloud contamination. The second set of data is referred to as the 'smoothed' data in this study. For each set of imagery and each growth season eight different growth season sum VI variables were created to predict herbaceous biomass.

The growth season sum variables were:

1. A summation over the same fixed set of dates, September – April, for each growth season (referred to as 'fixed' in the tables that follow).

2. A summation taking into account the variation in the onset of the growth season as determined by a 20% increased VI value on a pixel by pixel basis (referred to as 'variable' in the tables that follow).

3. A summation placing a 30% weighting on the first image in September and increasing the weighting linearly to 100% for the first image in April (the last image in the time series) (referred to as 'weighted' in the tables that follow).

4. A pixel by pixel summation placing a 30% weighting on the first value in the growth season time series (start of growing season) as determined by a 20% increased VI value from the previous season low and increasing the weighting linearly to 100% for the first image in April.

Each variable was created using both EVI and NDVI, yielding 8 growth season sum variables in total.

A single raw 16 day image composite was also selected from both the NDVI and EVI imagery for each growth season for use as VI variables. These 16 day composites corresponded to mid April in each growth season, the period in which the field measurement were reportedly taken (referred to as 'single composite' in the tables that follow).

The resulting data was randomly sampled to create 100 testing and training datasets using a 70% training 30% testing split. OLS regression was performed on each dataset using each VI variable in turn to predict herbaceous biomass.

Where appropriate, the best performing VI variable for each year based on either the average Root Mean Square Error (RMSE) or R^2 is highlighted in bold. R^2 has been used as a means of comparison in table 5 to facilitate comparison to other studies. In other tables RMSE has been used as it provides information on the accuracy of the estimates produced in kg/ha, which is of greater interest in this study than R^2, which is a measure of how well the models fit the data. The tables are followed by a figure (figure 2) showing the average rainfall within the study area for each month within the relevant growth season. This was added to aid in the interpretation of variations in the performance of the VI variables.

It is clear that on average EVI outperforms NDVI as a predictor of herbaceous biomass in the study area (table 2). The only instance in which this is not the case is in the 2002 – 2003 growth season, when the single NDVI composite image was the only variable with a significant correlation. The difference in performance is fairly minor in the 2004 – 2005 and 2001 – 2002 growth seasons, while it is most pronounced in the 2005 – 2006 growth season. It is also evident that vegetation indices were not related to herbaceous biomass in the 2002 – 2003 growth season, which was particularly dry. Having identified EVI as the best performing VI variable, it was selected as the VI with which the rest of the study was conducted. NDVI therefore does not appear in any of the subsequent results.

Table 2: The difference in the mean RMSE values of the linear models to predict herbaceous biomass using the NDVI and EVI to account for herbaceous production. Negative numbers indicate instances in which the NDVI based variable had a smaller RMSE value than its EVI equivalent. The missing values in 2002 -2003 indicate that neither the EVI nor NDVI variable were significantly correlated to herbaceous biomass. * indicates that one of the variables in the pair was not significantly correlated to herbaceous biomass.

VI Variable	2000 - 2001 (NDVI - EVI, kg/ha)	2001 - 2002 (NDVI - EVI, kg/ha)	2002 - 2003 (NDVI - EVI, kg/ha)	2003 - 2004 (NDVI - EVI, kg/ha)	2004 - 2005 (NDVI - EVI, kg/ha)	2005 - 2006 (NDVI - EVI, kg/ha)
Single Composite	54*	-2*	-11*	73	16	77
Fixed smoothed ∑	74	19	.	20	-3	68
Variable smoothed ∑	55	18	.	21	4*	100
Weighted smoothed ∑	86	18	.	47	3	78
Variable weighted smoothed ∑	72	17	.	58	6	106
Fixed Raw ∑	64	27	.	24	4	51
Variable Raw ∑	48	24	.	20	9	83
Weighted Raw ∑	74	25	.	55	9	67
Variable weighted Raw ∑	63	23	.	58	11	92
Average	**66**	**19**	**.**	**42**	**6**	**80**

The percentage of cloud contaminated pixels was extremely low for all growth seasons assessed in the study. The most contaminated growth season imagery was recorded in 2000 – 2001, when 2.3% of the pixels in the image stack were cloud contaminated according to the MODIS quality flags (table 3). The percentage of marginal pixels recorded within the study period was of greater concern, with between 17% and 41.1% of the pixels in the image stacks for the growth seasons being labelled 'marginal' in the quality flag layer (table 3).

Table 3: Pixel reliability information contained within the MOD13 quality flag layer for the study area. The quality flag layer is a raster image included with MODIS imagery with pixel values coded to provide information on the quality of the data in the acompanying layers on a per pixel basis.

Growth season	% of pixels classified as "Good" quality	% of pixels classified as "Marginal" quality	% cloud contaminated pixels
2000 - 2001	56.5	41.1	2.3
2001 - 2002	64.3	34.5	1.2
2002 - 2003	75.6	23.6	0.8
2003 - 2004	65.8	32.8	1.4
2004 - 2005	82.6	17.0	0.4
2005 - 2006	73.9	24.6	1.6

Smoothing the data to account for clouds and "marginal pixels" lead to greater prediction error, for all of the VI variables in the growth seasons between 2000 and 2003 (Table 4). The data from the latter half of the study period showed the opposite trend. In the growth seasons between 2003 and 2006, smoothing the data decreased prediction error, but only for some of the variables. The differences in RMSE between the raw and smoothed data were greatest in the 2000 – 2001 growth season, where they differed on average by 6 kg/ha (table 4). The performance of the raw and smoothed data also varied on average more widely for NDVI. All of the differences were extremely small (0 – 9 kg/ha) when compared to the size of the actual RMSEs' of the regression models created (1200 – 1700 kg/ha, appendix 1). In percentage terms these changes were in the region of 2%, similar to the % of cloud contaminated pixels present but far lower than the % of "marginal quality" pixels. It may be possible that:

1. Smoothing each pixels temporal profile helps remove the negative bias of cloud contaminated pixels, but that of marginal quality pixels obtained under extreme viewing angels remained a problem.

or

2. Cloud contamination and marginal quality pixels do not create significant perturbations to the VI time signal and therefore smoothing the data did little and these are not significant issues.

Table 4: The difference in the mean RMSE values between the smoothed data to account for the effect of cloud contamination and "marginal pixels", and the raw data for the EVI variables assessed in this study to account for herbaceous production. Negative values indicate instances in which using smoothed data resulted in improved estimation accuracy. Average values across the four summation types are provided for each growth season.

Summation	**2000 - 2001 (smoothed - raw, kg/ha)**	**2001 - 2002 (smoothed - raw, kg/ha)**	**2002 - 2003 (smoothed - raw, kg/ha)**	**2003 - 2004 (smoothed - raw, kg/ha)**	**2004 - 2005 (smoothed - raw, kg/ha)**	**2005 - 2006 (smoothed - raw, kg/ha)**
Fixed ∑ EVI	8	3	.	-6	-6	-1
Variable ∑ EVI	3	3	.	-2	-3	6
Weighted ∑ EVI	9	3	.	-6	-5	-4
Variable weighted ∑ EVI	5	3	.	-4	-3	3
Average	**6**	**3**	**.**	**-4**	**-4**	**1**

Based on the fact that the raw data on average performed marginally better than the smoothed data, the raw data was selected for use in all subsequent analysis.

The amount of variation accounted for in the herbaceous biomass data by the EVI variables assessed in the study exceeded 20% in only 2 of the 6 growth seasons, 2003 – 2004 and 2005 – 2006 (table 5). None of the variables could account for any variation in the data in the 2002 – 2003 growth season, or more than 6% in 2001 – 2002. 2005 – 2006 stands out as the only growth season in which all of the variables assessed accounted for more than 20% of the variation in the data (Table 5). All summations performed extremely badly in 2003 – 2004, while the single April EVI composite with an R^2 value of 0.37 performed relatively well. This leads on to the point that there is no single VI variable that consistently performed well or that could be considered the best. It is also interesting to note that there were numerous instances of multiple VI summations in a growth season showing no significant differences in the amount of variation which they accounted for or differences of only 1- 2%. These VI summations, although they vary in their preparation, must therefore contain much the same information.

Table 5: Mean R^2 values from the 100 iterations run for each of the models predicting herbaceous biomass using the EVI variables derived from the raw MODIS MOD13 data assessed in this study to account for herbaceous production. The best performing variable/s for the growth season are highlighted in bold. 0.00 indicates no significant correlation.

Summation	Mean R^2 2000 - 2001	Mean R^2 2001 - 2002	Mean R^2 2002 – 2003	Mean R^2 2003 - 2004	Mean R^2 2004 – 2005	Mean R^2 2005 - 2006
Single EVI Composite	0.06	0.00	0.00	**0.37**	0.05	**0.36**
Fixed Raw ∑ EVI	**0.16**	**0.06**	0.00	0.08	**0.18**	0.29
Variable Raw ∑ EVI	0.14	**0.06**	0.00	0.03	0.17	0.30
Weighted Raw ∑ EVI	**0.16**	**0.06**	0.00	0.15	**0.18**	0.34
Variable weighted Raw ∑ EVI	0.14	**0.06**	0.00	0.10	**0.18**	0.35

The similarity in the amount of variation accounted for by the fixed and variable date summations in table 5 is interesting to note, considering the fact that the images summed to create the two often varied considerably. The variation in the phenologically derived start dates for the growth seasons that led to this variation in images summed is apparent in figure 1. The similarity in variation accounted for suggests that although the start of the growth season as derived from changes in EVI value may vary considerably, the vegetation production occurring very early in the season contributes little to end of season herbaceous biomass. If this early season production was a significant contributor to herbaceous fuel load, the amount of variation accounted for by the fixed and variable summations could not be as similar as indicated in table 5. Accurate delineation of the growth appears to offer little benefit relative to the level of prediction error currently experienced.

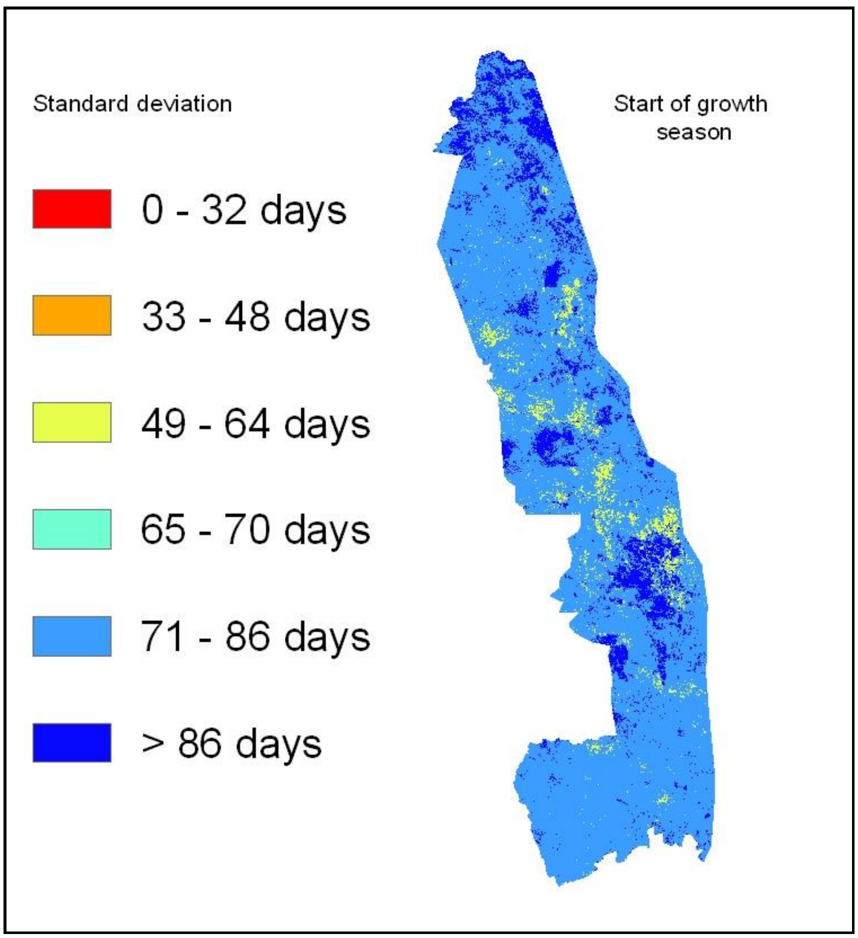


Figure 1: Standard deviation, measured in number of days, in the onset of vegetation growth for the growth seasons between 2000 and 2006. The onset of vegetation growth was established using a 20% increase in vegetation greenness from the previous seasons low, measured using EVI.

The onset of the growth season in savannas is dictated by the onset of the rains, while the amount of rainfall acts in conjunction with soil and other factors to determine the amount of production within a growth season. Both the distribution of rainfall through the year and the total amount of rain received by the study area as a whole shows a fair amount of variation between growth seasons (figure 2).

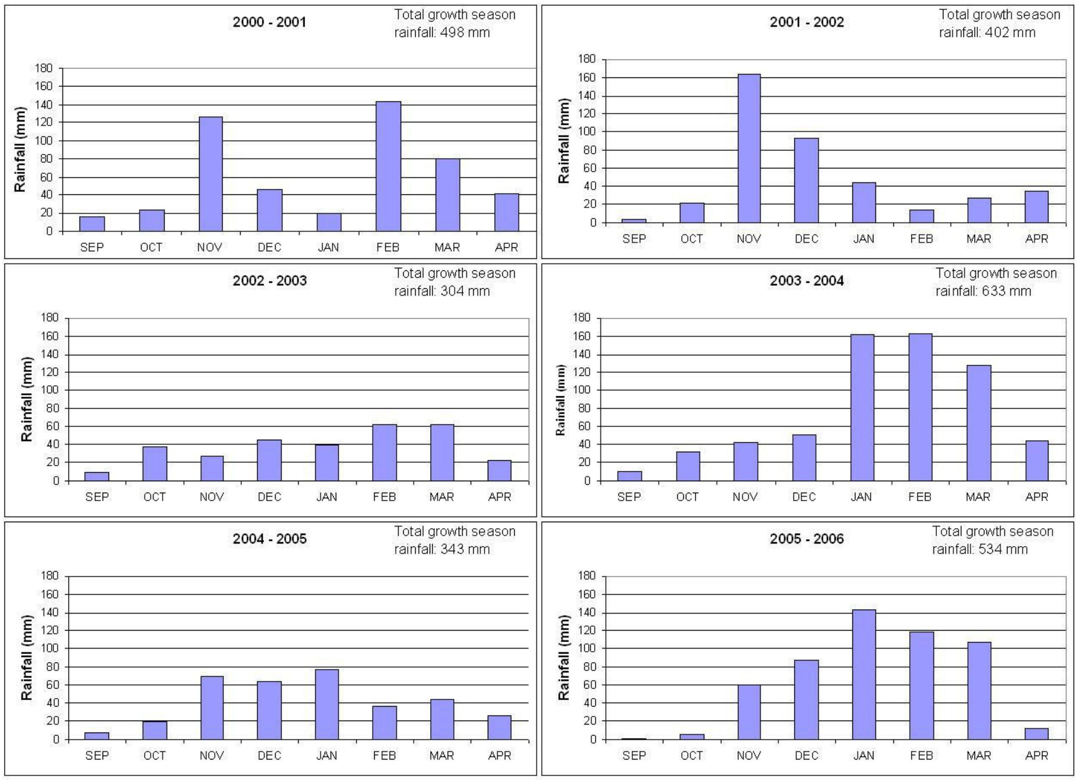


Figure 2: Monthly rainfall for each growth season within the study period.

The relevance of this variation to this study lies primarily in the relationship between production and biomass accumulation. Production must exceed removal for accumulation to occur. EVI has been used to measure production in this study. If there is insufficient rainfall to facilitate sufficient production to exceed the accumulation threshold, then EVI will not be correlated to end of season herbaceous biomass estimates. Timing of rainfall and hence production also plays a role in the correlation of EVI to end of season herbaceous biomass estimates. If the accumulation threshold is only exceeded very early or late in the season, an EVI summation for the entire season is unlikely to be strongly correlated to end of season herbaceous biomass estimates. This is apparent in the results contained in table 5.

The 2002 – 2003 growth seasons received fairly low but uniformly distributed rainfall. In comparison the 2000 – 2001 growth season received more rainfall but most of it was concentrated in two peaks, the first in November and the second in February. The 2001 – 2002 and 2003 – 2004 growth seasons were almost exact opposites in terms of the distribution of the rainfall they received. The former received the majority of its rainfall at the beginning of the season in November and December and the latter the majority in January, February and March, close to the end of the growth season. Late rainfall appears to increase the correlation between single date EVI and herbaceous biomass and weakens the correlation between growth season sum EVI (table 5).

3.2. Assessment of the woody canopy cover variables

Table 6 contains the results from the assessment of variables included in this study to account for tree cover. Assessment of the tree cover variables involved degrading the resolution of high resolution derived validation data (10m) to that of the three tree cover variables assessed (250m, 1km). Degrading of the resolution was achieved by assigning the average value of all of the 10m pixels contained within the corresponding cell of an overlaid 250m or 1km grid. The high resolution validation data was derived from a mixture of IKONOS and LiDAR data corresponding to three separate sites, two in the south and one in the centre of the study area. The three variables to be assessed were used as predictors of canopy cover using OLS regression.

All three variables can be seen to display negative relationships of varying strengths with the high resolution derived tree cover data (table 6), which is the opposite of what would logically be expected.

Table 6: Regression diagnostics for the relationships between the three potential tree cover variables originally assessed for use in this study and canopy cover measurements derived from high resolution imagery.

Tree cover variable	R	R Squared	P Value
MOD44	-0.04	0.00	0.10
September – October mean MODIS EVI	-0.13	0.01	P<0.01
AVHRR NDVI derived woody cover	-0.61	0.38	P<0.01

Figures 3, 4 and 5 show the fit of the above models to the data. No correlation was found for the MOD44 product. Figure 3 shows that the percentage tree cover indicated by the MOD44 product never exceeded 25% for the areas assessed. Based on experience in the field and the high resolution data available, this seems unlikely. It is likely that the inaccuracy of the MOD44 product arose because it was calibrated as a global product and therefore is not suited for use in studies carried out at a local scale. Although figures 4 and 5 correspond to models that are statistically significant, inspection of the plots, and the fact that the r values are negative, reveal the correlations found to be meaningless. There is no way that as tree cover increases, either of the indicators of tree cover could decrease.

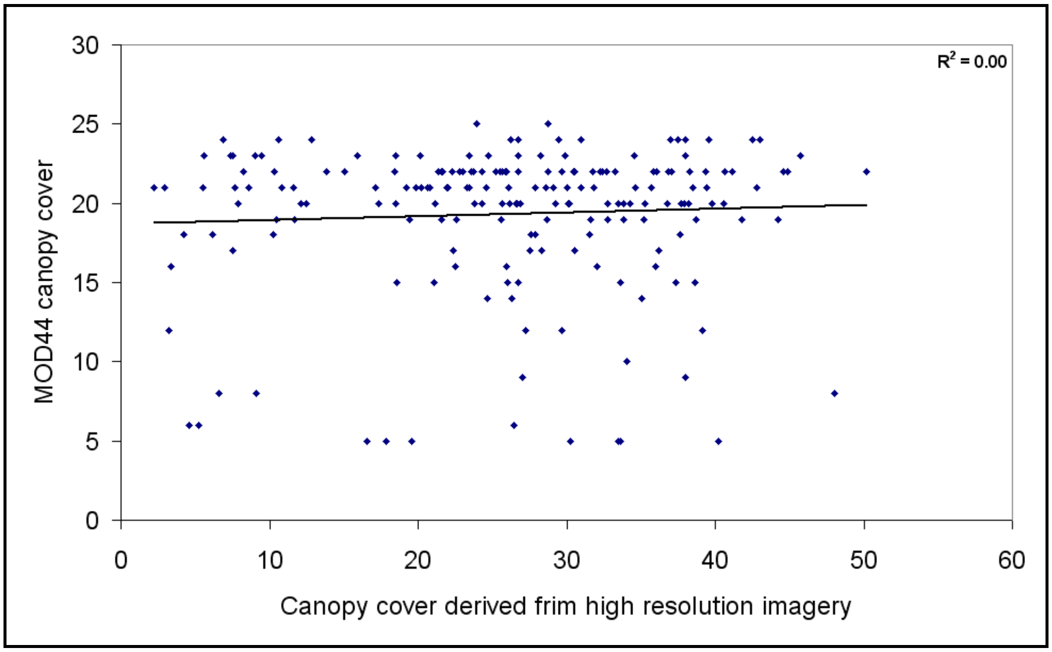


Figure 3: Relationship between MOD44 tree cover variable and high resolution tree cover data .

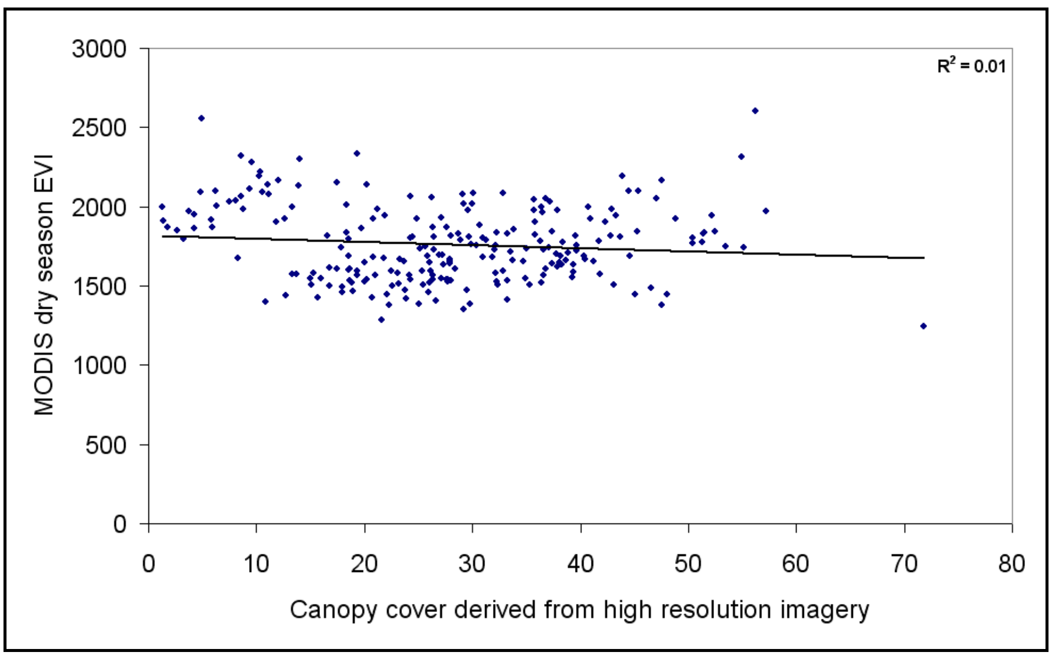


Figure 4: Relationship between the Modis September –October mean EVI variable and the high resolution tree cover data.

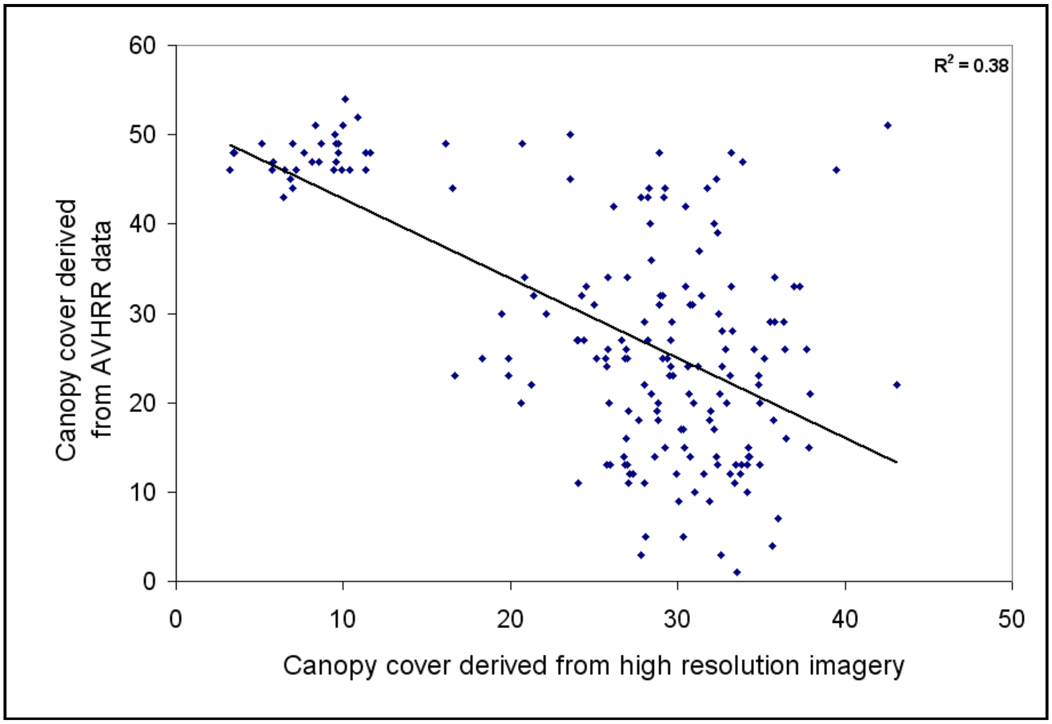


Figure 5: Relationship between the AVHRR derived tree cover variable and the high resolution tree cover data.

The above assessment, carried out using the high resolution derived tree cover data, was strongly suspected of being flawed. This was because it failed to show even a weak positive correlation as would be expected based on the literature consulted. A method for re-evaluating the above variables was therefore sought. It was decided that re-evaluation of the variables would be achieved by comparing the relative performance of regression models containing each of the different tree cover variables. As there was little guidance in the literature on how best to achieve this, they were included in three different ways. These were simple addition to the model, addition with an interaction term and finally, adjustment of the EVI variable prior to inclusion in the model as described in the methods section.

In four of the six years the September – October mean EVI variable resulted in the greatest decrease in RMSE of the three tree cover variables assessed (table 7). In two of the years this was achieved using an interaction term and in the other two it was achieved without one. The difference in performance between the inclusion and absence on an interaction term was never more than 10 kg/ha, which is less than 1% of the total error. Pre-correction, thereduction in the EVI value based on the tree cover variable, described in chapter 2, (section 3.5. Woody canopy cover re-evaluation), was the worst performing of the three inclusion methods, most noticeably in the very dry 2002 - 2003 growth season. In 2003 – 2004 and 2005 – 2006 the MODIS MOD44 product included with an interaction term resulted in the greatest increase in estimation accuracy, significantly outperforming the September –October mean EVI variable in the latter case. The AVHRR derived variable performed inconsistently and poorly in almost every instance regardless of inclusion method, with it being rejected entirely in numerous instances. Based on the fact that the September – October mean EVI woody cover variable resulted on average in the greatest improvement in accuracy, it was selected for use in subsequent analysis steps. Given the comparable performance of the two inclusion methods, the addition of the variable without an interaction term was selected because of its simplicity.

Table 7: The change in RMSE caused by adding a variable to account for the presence of woody vegetation to a model previously only containing a weighted summation of growth season EVI to account for herbaceous production. Posotive numbers highlighted in red indicate a decrease in performance (corresponding to an increase in RMSE). The greatest Decrease in RMSE for each year is highlighted in bold. Those instances in which the addition of the variable did not result in an increase in adjusted R^2 have been labelled as 'rejected'.

Explanatory variables to predict herbaceous biomass	Change in RMSE (kg/ha) 2000 - 2001	Change in RMSE (kg/ha) 2001 - 2002	Change in RMSE (kg/ha) 2002 - 2003	Change in RMSE (kg/ha) 2003 - 2004	Change in RMSE (kg/ha) 2004 - 2005	Change in RMSE (kg/ha) 2005 - 2006
EVI summation and September –October mean EVI tree cover variable	**-116 (6.9%)**	**-89 (5.9%)**	-87 (6.8%)	-97 (6.3%)	-62 (4.3%)	-39 (2.9%)
Interaction between EVI summation and September –October mean EVI tree	-112 (6.6%)	-85 (5.6%)	**-94 (7.4%)**	-89 (5.7%)	**-63 (4.4%)**	-41 (3.1%)
EVI summation adjusted prior to regression using September –October mean	-109 (6.5%)	-64 (4.2%)	-24 (1.9%)	-99 (6.4%)	-59 (4.1%)	-35 (2.6%)
EVI summation and AVHRR derived tree cover variable	0 (0%)	0 (0%)	-14 (1.1%)	-72 (4.6%)	+4 (-0.3%)	-48 (3.6%)
Interaction between EVI summation and AVHRR derived tree cover variable	+5 (0.3%)	rejected	-34 (2.7%)	-79 (5.1%)	+1 (0.1%)	-51 (3.8%)
EVI summation adjusted prior to regression using AVHRR derived tree cover	rejected	rejected	8 (0.6%)	rejected	rejected	rejected
EVI summation and mod44 tree cover variable	-35 (2.1%)	-4 (0.3%)	+-1 (0.1%)	-63 (4.1%)	-8 (0.6%)	-49 (3.7%)
Interaction between EVI summation and mod44 tree cover variable	-38 (2.3%)	-21 (1.4%)	+3 (0.2%)	**-116 (7.5%)**	-7 (0.5%)	**-82 (6.1%)**
EVI summation adjust prior to regression using MOD 44 tree cover variable	-42 (2.5%)	-14 (0.9%)	rejected	-81 (5.2%)	-14 (1.0%)	-62 (4.6%)

3.3. Assessment of a suitable variable to account for dry (moribund) material

Having investigated the best variable to account for tree cover and the best means of including it in the model, a combined fire history and geology categorical variable accounting for the presence of dry material was assessed. The fire history was arrived at by calculating the number of seasons since the veld was burned. This was achieved using digitised fire scar data provided by the KNP Scientific Services. The resulting fire history data was combined with a map of underlying geology to create a categorical variable with a value for every pixel of the MODIS imagery covering the study area. The resulting variable accounted for only slight improvements in both 2000 – 2001 and in the very dry growth season of 2002 – 2003 (table 8). In all other years it resulted in a minor yet moderately greater decrease in RMSE.

Table 8: The Reduction in RMSE caused by the addition of a variable to account for the presence of dry / moribund herbaceous material relative to a model including an EVI summation to account for herbaceous production and September –October mean EVI to account for the presence of woody cover. Adjusted R^2 increased in all cases, so RMSE has been reported for all models.

Explanatory variables used in addition to EVI to predict herbaceous biomass	Change in RMSE (kg/ha) 2000 – 2001	Change in RMSE (kg/ha) 2001 - 2002	Change in RMSE (kg/ha) 2002 - 2003	Change in RMSE (kg/ha) 2003 - 2004	Change in RMSE (kg/ha) 2004 - 2005	Change in RMSE (kg/ha) 2005 - 2006
September –October mean EVI tree cover variable and combined geology/fire history variable	-16 (1%)	-46 (3.2%)	-14 (1.2%)	-63 (4.3%)	-66 (4.8%)	-81 (6.2%)

3.4. Assessment of the completed regression models

Based on the above investigations into the performance of the available variables, the final model specification used for all of the growth seasons was:

> sqrt(biomass) = f(weighted growth season sum EVI, September and October mean EVI, number of seasons since last fire occurred and underlying geology)

Where:

a) weighted growth season sum EVI was used as a proxy for a measure of herbaceous biomass production

b) September and October mean EVI was used as a proxy for a measure of woody cover

c) The number of seasons since last fire occurred and underlying geology were used as an indicator of differing levels of accumulated dry material

Seven models were created in total. Six models were created using the variables derived for each growth season separately (one model using data from 2000 – 2001, another using only data from 2001 – 2002, etc.). A seventh model was created using all six of the growth season's data combined into one large dataset. The models created for the individual growth seasons had lower RMSE values than the seventh model, created using the combined dataset, for 5 out of the 6 growth seasons (Table 9). However, in 2001 – 2002, 2002 – 2003 and 2004 – 2005, there was very little difference in the performance of the models trained for individual season and the model trained using all the available data. The R^2 values in table 10 show how variable the performance of the individual growth season models were.

Table 9: RMSE (kg/ha) of the final models.

	2000 – 2001	2001 – 2002	2002 - 2003	2003 - 2004	2004 - 2005	2005 - 2006
One model for all growth seasons	1711	1374	1197	1417	1313	1410
Separate model for each growth season	1555	1383	1171	1390	1312	1221

Table 10: Adjusted R^2 Values and regression coefficients for the final models produced. In all cases the model specification was: sqrt(biomass estimate)=A(weighted growth season sum EVI)+B(September and October mean EVI)+C(combined geology and fire history dummy variable)+Intercept

	2000-2001	2001-2002	2002-2003	2003-2004	2004-2005	2005-2006	Combined
R^2	0.31	0.25	0.18	0.32	0.34	0.46	0.4
Intercept	34.930349	44.559838	62.586046	29.020591	47.192627	19.976594	39.94742
A	0.001976	0.001672	0.0009442	0.001778	0.001529	0.001659	0.0016
B	-0.022914	-0.024658	-0.025604	-0.015759	-0.022787	-0.007	-0.01905
C1		10.066876	6.9425826	7.011689	-23.75499	-5.303775	3.98792
C2	5.300167	5.439267	1.5010101	6.042806	10.760002	-4.675132	5.66033
C3	-2.081968	-0.591684	3.8396495	2.238204	0.526876	-4.961998	-0.90316
C4	-2.822972	-2.364269	-3.945986	-3.299088	-3.977131	-10.40284	-5.51498
C5	4.079999	2.449018	4.2885474	-1.392103	2.049056	-9.841138	-0.53004
C6	6.469278	6.48037	2.1914478	-1.011521	1.300199	-16.34216	-1.04851
C7	0.706152	0.604061	1.5305816	-3.84689	-3.227868	-13.48145	-4.19113

4. KRIGING AND COKRIGING

Table 11 contains the kriging parameters optimised by ArcGIS 9.2 Geostatistical Analyst for all years within the study period. The range of autocorrelation was shortest during the 2001 -2002 growth season when it was just 13 km. Autocorrelation ceased to exist beyond between 20 and 23 km in all other years.

Table 11: Optimised Kriging parameters provided by ArcGIS Geostatistical Analyst.

Growth Season	Lag size (m)	# lags	Model	Nugget	Partial sill	Range (m)
2000 - 2001	2000	12	Spherical	1230700	1887000	20207.8
2001 – 2002	2000	12	Spherical	590940	1453700	13076.3
2002 – 2003	2000	12	Spherical	725940	736230	23706.5
2003 – 2004	2000	12	Spherical	1149100	1433700	21754.3
2004 – 2005	2000	12	Spherical	1004000	809940	20689.6
2005 – 2006	2000	12	Spherical	821790	1272700	23230.1

Table 12 contains the Cokriging parameters optimised by ArcGIS 9.2 Geostatistical Analyst for all years within the study period. The Range value returned by the software as optimal (23.7 km) was identical in all 6 years. Experimentation within Arc showed that by using a lag of 3000m the ranges calculated by the software were no longer all equal, but that performance in terms of RMSE remained essentially unchanged.

Table 12: Optimised Cokriging parameters provided by ArcGIS Geostatistical Analyst. Weighted growth season sum of EVI values adjusted for the presence of woody vegetation, using September – October mean EVI values, were used as secondary variables for cokriging.

Growth Season	Lag size (m)	# lags	Model	Nugget	Partial sill	Range (m)	R squared of correlation between herbaceous biomass field estimates and the weighted growth season sum EVI variable corrected for tree cover
2000 - 2001	2000	12	Spherical	948050	2501400	23706.5	0.23
2001 - 2002	2000	12	Spherical	1088100	1147600	23706.5	0.18
2002 - 2003	2000	12	Spherical	685410	798950	23064.9	0.03
2003 - 2004	2000	12	Spherical	1133800	1457600	23706.5	0.23
2004 - 2005	2000	12	Spherical	172930	23706.5	23706.5	0.27
2005 - 2006	2000	12	Spherical	423090	1996700	23706.5	0.36

There was no consistency in which of the two kriging methods performed best in this study (table 13). What is of note is that in three of the growth season's cokriging performed worse than kriging, which was unexpected.

Table 13: The change in the prediction accuracy when using cokriging with EVI as a secondary variable rather than kriging to interpolate herbaceous biomass. Negative values indicate that kriging outperformed cokriging. RMSE values were arrived at using leave-one-out cross-validation as implemented by ArcGIS 9.2.

	2000 - 2001	2001 - 2002	2002 - 2003	2003 - 2004	2004 - 2005	2005 - 2006
Change In Cross Validation RMSE (kg/ha)	-22.62 (-1.5%)	-81.96 (-6.4%)	6.77 (0.7%)	76.22 (6.1%)	-18.45 (-1.5%)	76.37 (6.8%)

5. COMPARISON OF THE ACCURACY AND PRECISION OF THE REGRESSION MODEL AND COKRIGING APPROACHES

Cokriging using the weighted growth season sum EVI variable corrected for tree cover as the covariable was the most accurate means of herbaceous biomass prediction for every growth season in the study period (table 14). It performed on average 119 kg/ha better than the competing regression model approach for the growth seasons assessed (table 15). Although its precision is less than that of the regression model approach, it is only by 20 kg/ha, which is trivial compared to the average herbaceous biomass for the period.

Table 14: Prediction accuracy of the regression model and cokriging approaches for estimating end of season herbaceous biomass for each season in the study period. The lowest RMSE figures are highlighted in bold.

	2000 - 2001	2001 - 2002	2002 - 2003	2003 - 2004	2004 - 2005	2005 - 2006
Regression RMSE (kg/ha)	1555	1383	1171	1390	1312	1221
Cokriging RMSE (kg/ha)	**1464**	**1271**	**993**	**1249**	**1217**	**1124**
Difference	91	112	178	141	95	97

Table 15: Accuracy and precision figures for the regression model and cokriging approaches to estimating end of season herbaceous biomass. The lowest RMSE figures are highlighted in bold.

Method	Accuracy (mean of RMSE, kg/ha for growth seasons assessed)	Precision (STDEV of RMSE, kg/ha for growth seasons assessed)
Regression Model	1339	**137**
Cokriging	**1220**	157

The regression approach produced maps showing more abrupt changes in the level of herbaceous biomass, with drainage channels and river beds being identifiable (figure 6). Cokriging on the other hand produced maps of a more smoothed appearance, with gradual rather than abrupt transitions in the level of herbaceous biomass predicted. These differences are even more apparent in the prediction maps when the pixels are grouped into classes as has been done for the maps contained in appendix 2. Although the maps in appendix 2 represent the final product of the methods assessed in this study, they contain little additional information not obtainable from figure 6. They were included only as an illustration of what the final outputs might look like, and for the sake of completeness. It is for this reason that they receive no further attention in this study. The important findings are contained in the error statistics presented in the preceding tables and the differences in the prediction maps highlighted in the figures that follow.

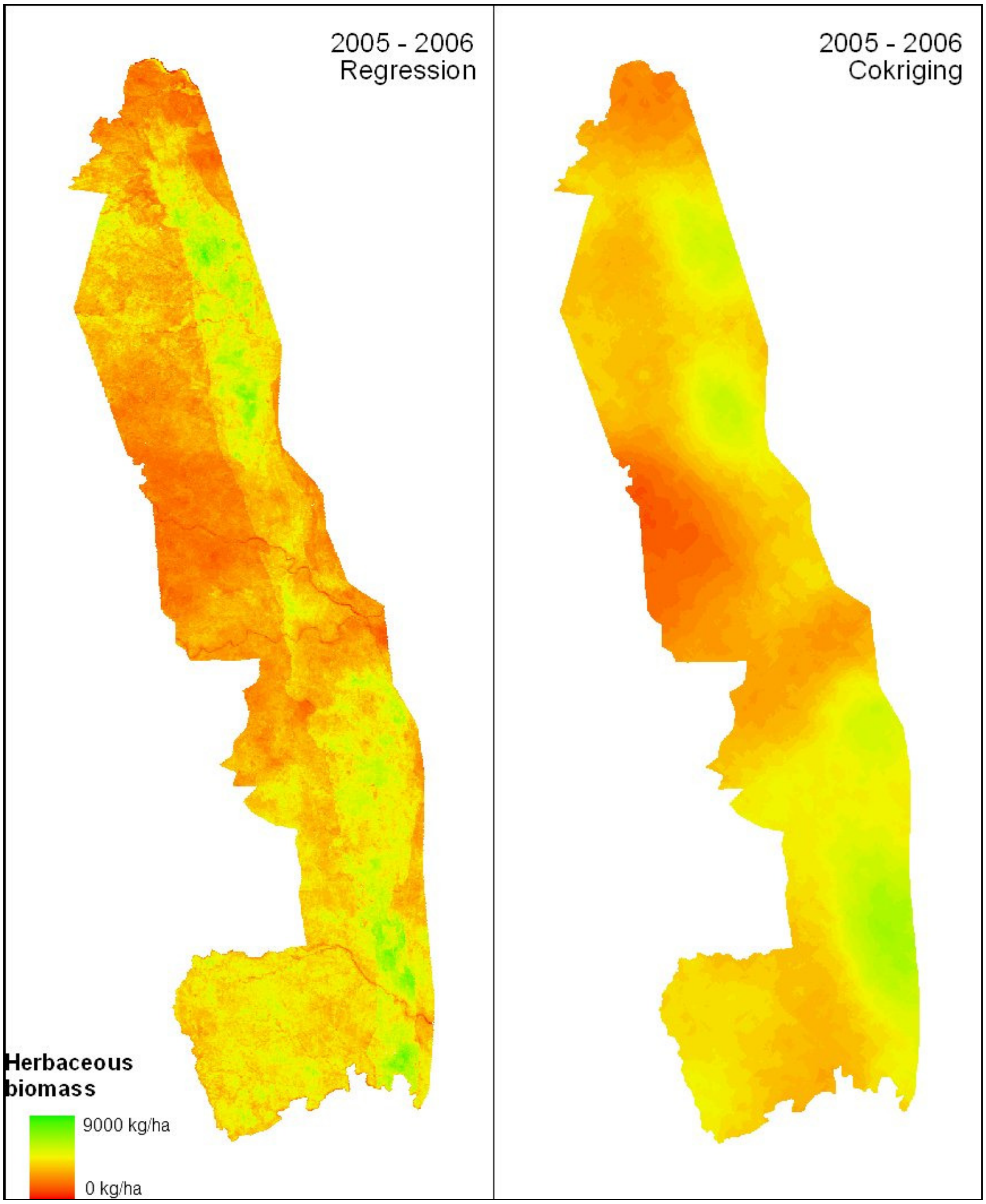


Figure 6: Maps of the herbaceous biomass estimates produced for the 2005 – 2006 growth season using Cokriging and a regression model.

Although the two methods were found to have accuracies that were within 120 kg/ha of one another, these figures hide the extent of their disagreement on the spatial distribution of herbaceous biomass within each growth season. Figure 7 illustrates the extent of these differences. Differences of >1000kg/ha for large areas of the park were found to be common.

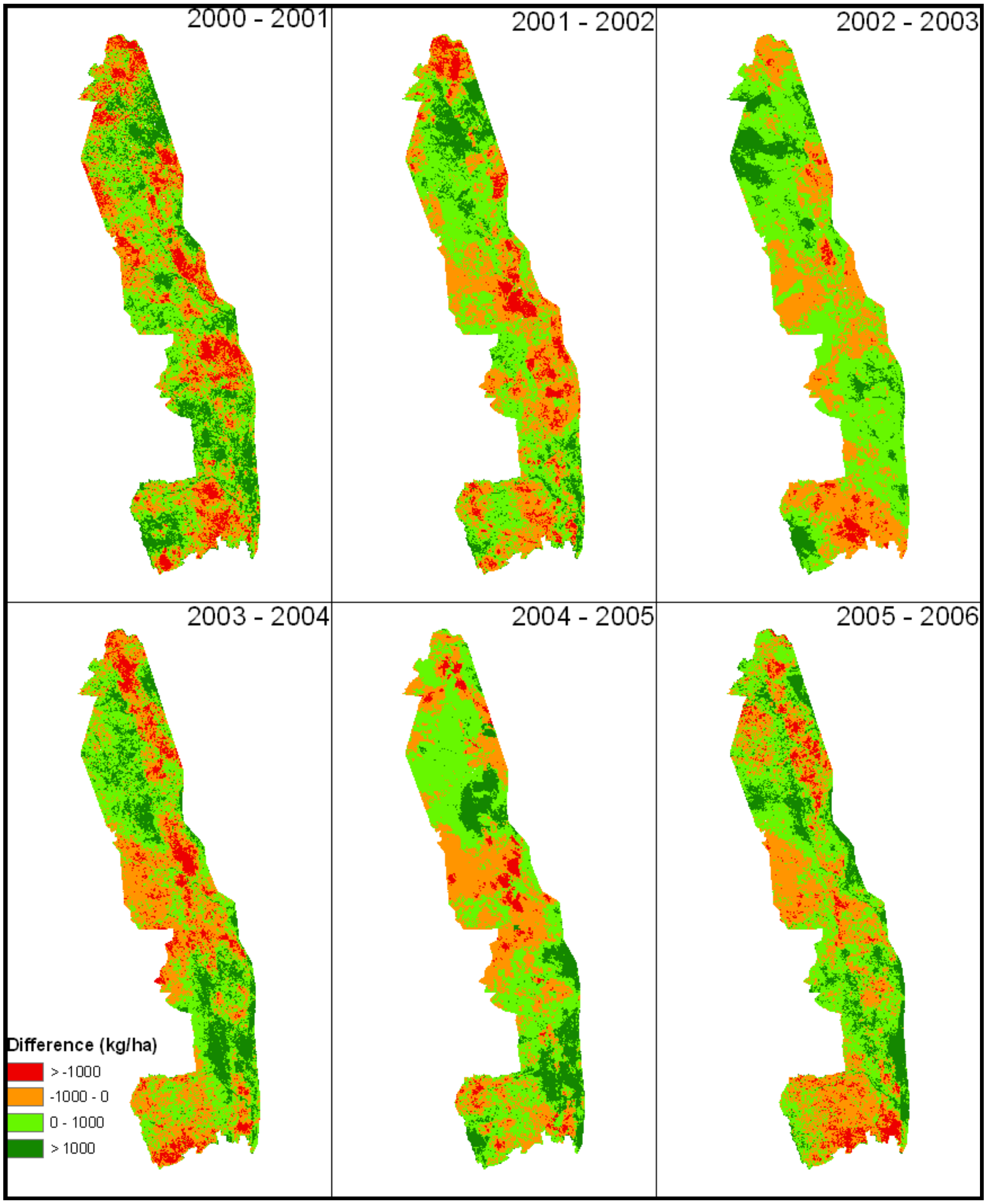


Figure 7: The difference between the herbaceous biomass estimates produced using the two methods for all growth seasons within the study period. Negative values (orange - red) indicate areas where cokriging estimates of herbaceous biomass exceeded those of the regression models. Positive values (light and dark green) indicate the opposite.

The distributions of the residuals for both methods in all seasons are centred on zero (figure 8), with slight tails to the left in some instances indicating a minor tendency for underestimates. Although there are slight differences identifiable between the error distributions of the two methods in each year, there are no clear trends in the differences.

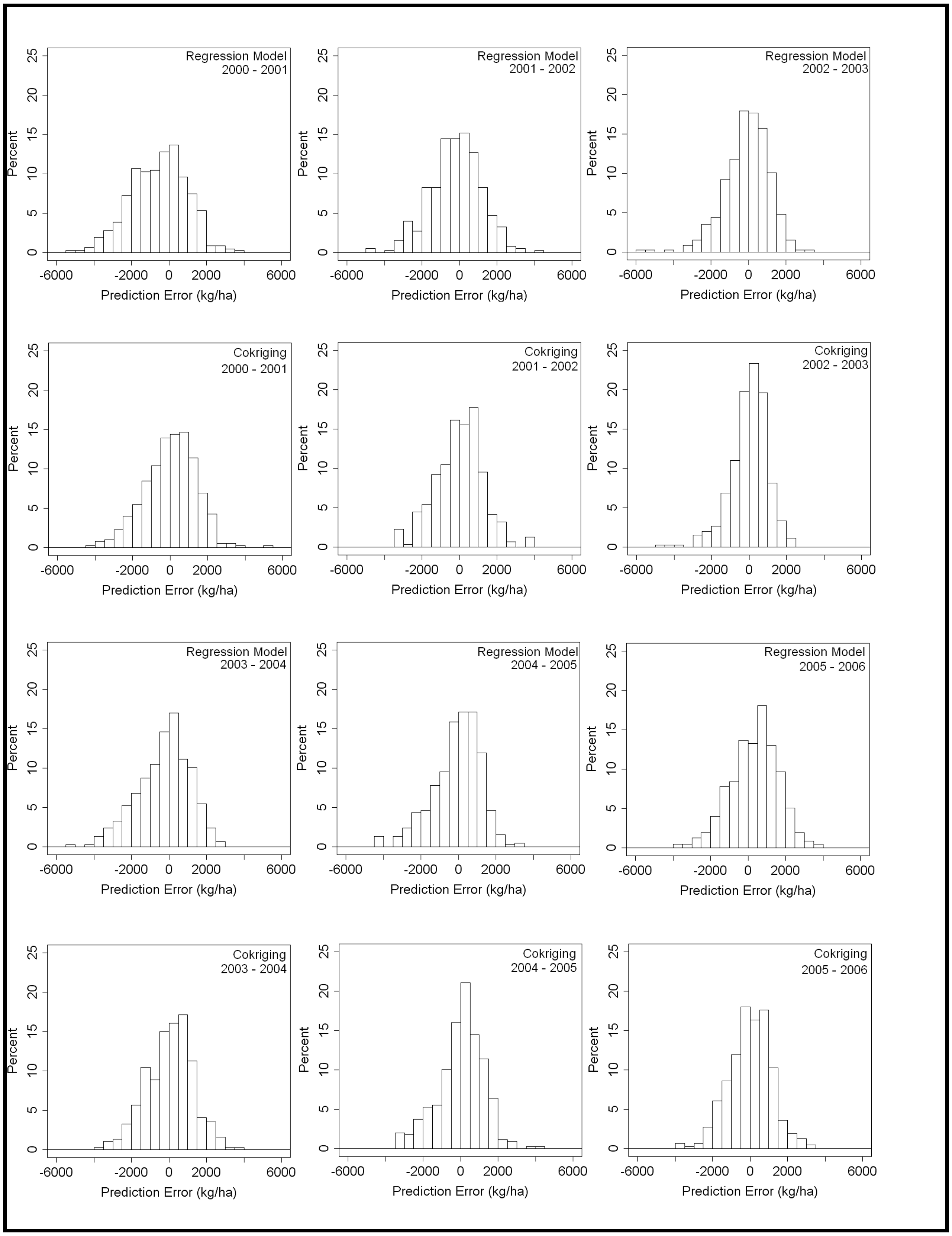


Figure 4: Distribution of the regression model and cokriging residuals for each growth season in the study period.

There are also no clear spatial patterns in the magnitude of the residuals for either the regression models or cokriging (figures 9 and 10). The magnitude of the residuals appears to depend more on the growth season than geographic location.

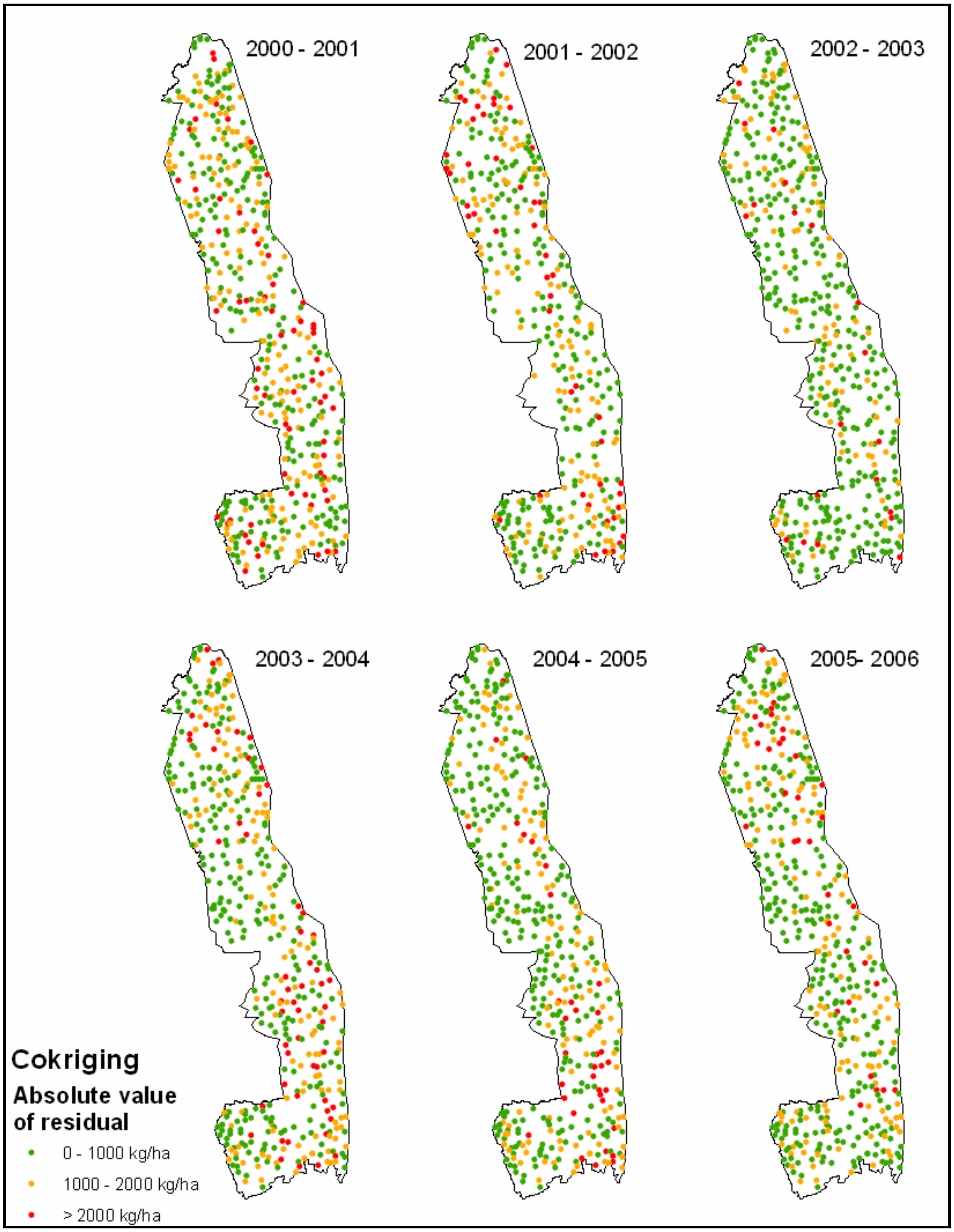


Figure 5: Location of the Veld Condition Assessment (VCA) sites (=field data) in the Kruger National Park and associated cokriging residuals.

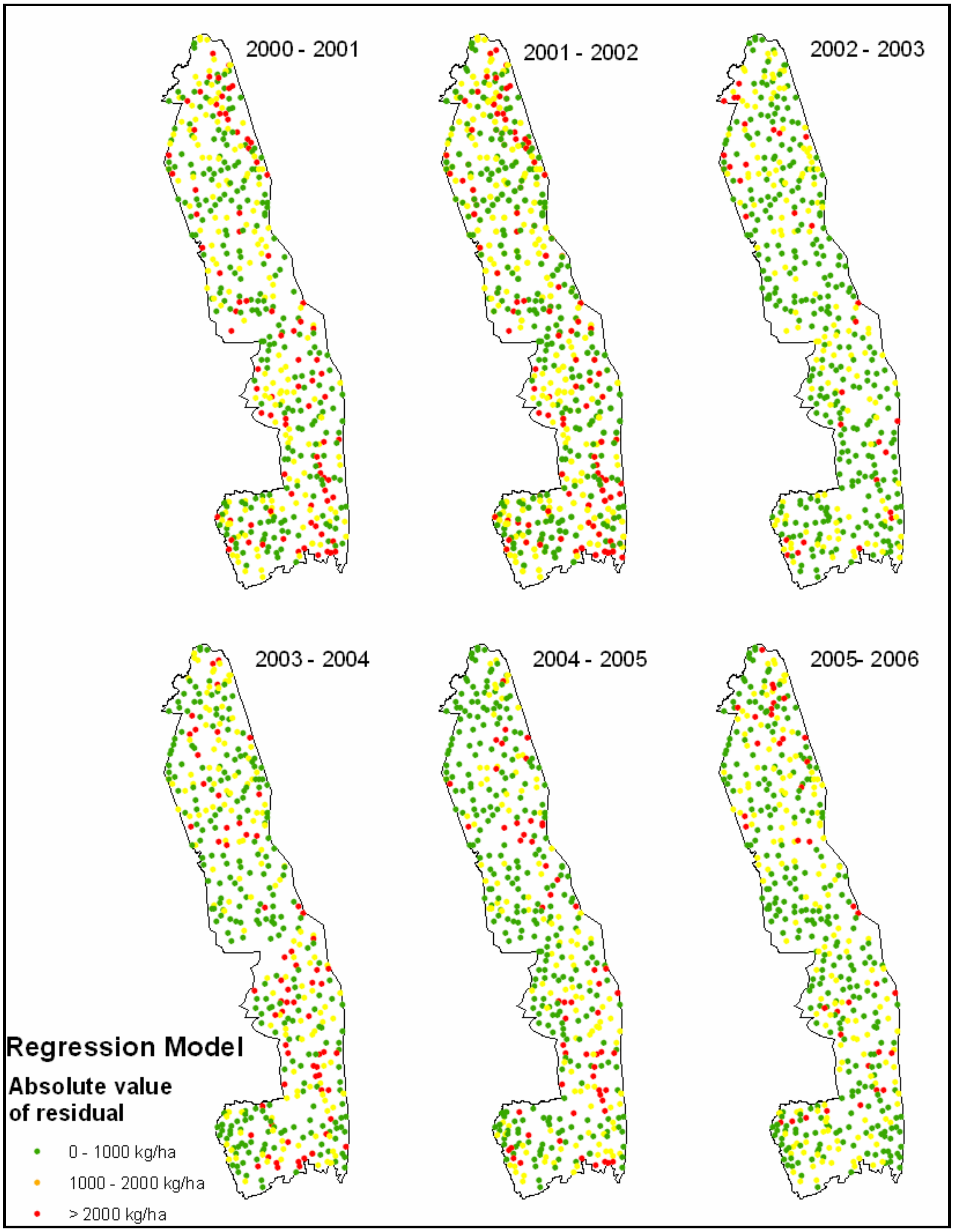


Figure 6: Location of the field data and associated regression model residuals.

Appendix 1: Regression statistics for the various EVI variables assessed in the study.

2001 Summation	Mean RMSE (kg/ha)	Mean R.Squared	Mean P.Value	Mean Absolute percentage Error
Single NDVI Composite	1835	0.01	0.182	51
Fixed Fitted ∑NDVI	1769	0.08	P<0.01	49
Variable Fitted ∑NDVI	1768	0.08	P<0.01	49
Weighted Fitted ∑NDVI	1782	0.06	P<0.01	49
Variable weighted Fitted ∑NDVI	1780	0.06	P<0.01	49
Fixed Raw ∑NDVI	1750	0.10	P<0.01	48
Variable Raw ∑NDVI	1758	0.09	P<0.01	49
Weighted Raw ∑NDVI	1760	0.09	P<0.01	49
Variable weighted Raw ∑NDVI	1765	0.08	P<0.01	49
Single EVI Composite	1781	0.06	P<0.01	48
Fixed Fitted ∑EVI	1695	0.16	P<0.01	45
Variable Fitted ∑ EVI	1713	0.14	P<0.01	46
Weighted Fitted ∑ EVI	1696	0.16	P<0.01	44
Variable weighted Fitted ∑ EVI	1708	0.14	P<0.01	45
Fixed Raw ∑ EVI	1687	0.16	P<0.01	45
Variable Raw ∑ EVI	1710	0.14	P<0.01	46
Weighted Raw ∑ EVI	1686	0.16	P<0.01	45
Variable weighted Raw ∑ EVI	1703	0.14	P<0.01	46

2002 Summation	Mean RMSE (kg/ha)	Mean R.Squared	Mean P.Value	Mean Absolute percentage Error
Single NDVI Composite	1556	0.01	0.040	57
Fixed Fitted ∑NDVI	1547	0.03	0.009	57
Variable Fitted ∑NDVI	1549	0.02	0.011	57
Weighted Fitted ∑NDVI	1546	0.03	0.008	57
Variable weighted Fitted ∑NDVI	1547	0.03	0.009	57
Fixed Raw ∑NDVI	1544	0.03	0.005	57
Variable Raw ∑NDVI	1546	0.03	0.006	57
Weighted Raw ∑NDVI	1543	0.03	0.005	57
Variable weighted Raw ∑NDVI	1544	0.03	0.005	57
Single EVI Composite	1558	0.01	0.082	58
Fixed Fitted ∑EVI	1528	0.05	0.001	57
Variable Fitted ∑ EVI	1531	0.05	0.001	57
Weighted Fitted ∑ EVI	1529	0.05	0.001	57
Variable weighted Fitted ∑ EVI	1530	0.05	0.001	57
Fixed Raw ∑ EVI	1517	0.06	P<0.01	56
Variable Raw ∑ EVI	1522	0.06	P<0.01	56
Weighted Raw ∑ EVI	1518	0.06	P<0.01	56
Variable weighted Raw ∑ EVI	1521	0.06	P<0.01	56

2003 Summation	Mean RMSE (kg/ha)	Mean R.Squared	Mean P.Value	Mean Absolute percentage Error
Single NDVI Composite	1256	0.02	0.023	88
Fixed Fitted ∑NDVI	1271	0.00	0.645	89
Variable Fitted ∑NDVI	1271	0.00	0.565	89
Weighted Fitted ∑NDVI	1271	0.00	0.592	89
Variable weighted Fitted ∑NDVI	1272	0.00	0.653	89
Fixed Raw ∑NDVI	1271	0.00	0.622	89
Variable Raw ∑NDVI	1271	0.00	0.584	89
Weighted Raw ∑NDVI	1270	0.00	0.538	89
Variable weighted Raw ∑NDVI	1272	0.00	0.667	89
Single EVI Composite	1267	0.00	0.216	89
Fixed Fitted ∑EVI	1272	0.00	0.625	89
Variable Fitted ∑ EVI	1271	0.00	0.508	89
Weighted Fitted ∑ EVI	1272	0.00	0.635	89
Variable weighted Fitted ∑ EVI	1272	0.00	0.622	89
Fixed Raw ∑ EVI	1272	0.00	0.594	89
Variable Raw ∑ EVI	1270	0.00	0.442	88
Weighted Raw ∑ EVI	1272	0.00	0.643	89
Variable weighted Raw ∑ EVI	1272	0.00	0.595	89

2004 Summation	Mean RMSE (kg/ha)	Mean R.Squared	Mean P.Value	Mean Absolute percentage Error
Single NDVI Composite	1384	0.32	P<0.01	40
Fixed Fitted ∑NDVI	1623	0.05	P<0.01	50
Variable Fitted ∑NDVI	1654	0.00	0.187	52
Weighted Fitted ∑NDVI	1599	0.08	P<0.01	49
Variable weighted Fitted ∑NDVI	1640	0.02	0.010	51
Fixed Raw ∑NDVI	1629	0.04	0.001	51
Variable Raw ∑NDVI	1656	0.00	0.268	52
Weighted Raw ∑NDVI	1605	0.08	P<0.01	49
Variable weighted Raw ∑NDVI	1644	0.02	0.016	51
Single EVI Composite	1311	0.37	P<0.01	38
Fixed Fitted ∑EVI	1602	0.08	P<0.01	49
Variable Fitted ∑ EVI	1633	0.04	0.002	51
Weighted Fitted ∑ EVI	1552	0.14	P<0.01	46
Variable weighted Fitted ∑ EVI	1582	0.10	P<0.01	48
Fixed Raw ∑ EVI	1605	0.08	P<0.01	49
Variable Raw ∑ EVI	1636	0.03	0.003	51
Weighted Raw ∑ EVI	1550	0.15	P<0.01	47
Variable weighted Raw ∑ EVI	1585	0.10	P<0.01	48

2005 Summation	Mean RMSE (kg/ha)	Mean R.Squared	Mean P.Value	Mean Absolute percentage Error
Single NDVI Composite	1578	0.04	0.002	134
Fixed Fitted ∑NDVI	1433	0.19	P<0.01	125
Variable Fitted ∑NDVI	1451	0.17	P<0.01	129
Weighted Fitted ∑NDVI	1444	0.18	P<0.01	125
Variable weighted Fitted ∑NDVI	1452	0.17	P<0.01	128
Fixed Raw ∑NDVI	1439	0.18	P<0.01	126
Variable Raw ∑NDVI	1455	0.17	P<0.01	130
Weighted Raw ∑NDVI	1450	0.18	P<0.01	126
Variable weighted Raw ∑NDVI	1456	0.17	P<0.01	128
Single EVI Composite	1563	0.05	P<0.01	132
Fixed Fitted ∑EVI	1437	0.18	P<0.01	130
Variable Fitted ∑ EVI	1447	0.17	P<0.01	132
Weighted Fitted ∑ EVI	1442	0.18	P<0.01	130
Variable weighted Fitted ∑ EVI	1447	0.17	P<0.01	131
Fixed Raw ∑ EVI	1436	0.18	P<0.01	131
Variable Raw ∑ EVI	1446	0.17	P<0.01	133
Weighted Raw ∑ EVI	1440	0.18	P<0.01	130
Variable weighted Raw ∑ EVI	1445	0.18	P<0.01	132

2006 Summation	Mean RMSE (kg/ha)	Mean R.Squared	Mean P.Value	Mean Absolute percentage Error
Single NDVI Composite	1395	0.30	P<0.01	40
Fixed Fitted ∑NDVI	1449	0.24	P<0.01	44
Variable Fitted ∑NDVI	1474	0.20	P<0.01	46
Weighted Fitted ∑NDVI	1405	0.29	P<0.01	42
Variable weighted Fitted ∑NDVI	1433	0.25	P<0.01	44
Fixed Raw ∑NDVI	1449	0.24	P<0.01	44
Variable Raw ∑NDVI	1468	0.21	P<0.01	45
Weighted Raw ∑NDVI	1409	0.29	P<0.01	42
Variable weighted Raw ∑NDVI	1430	0.26	P<0.01	43
Single EVI Composite	1318	0.36	P<0.01	39
Fixed Fitted ∑EVI	1381	0.31	P<0.01	42
Variable Fitted ∑ EVI	1374	0.32	P<0.01	41
Weighted Fitted ∑ EVI	1328	0.37	P<0.01	39
Variable weighted Fitted ∑ EVI	1328	0.37	P<0.01	39
Fixed Raw ∑ EVI	1398	0.29	P<0.01	43
Variable Raw ∑ EVI	1385	0.30	P<0.01	42
Weighted Raw ∑ EVI	1342	0.34	P<0.01	41
Variable weighted Raw ∑ EVI	1338	0.35	P<0.01	40

Appendix 2: Herbaceous biomass prediction maps produced using the regression models and cokriging.

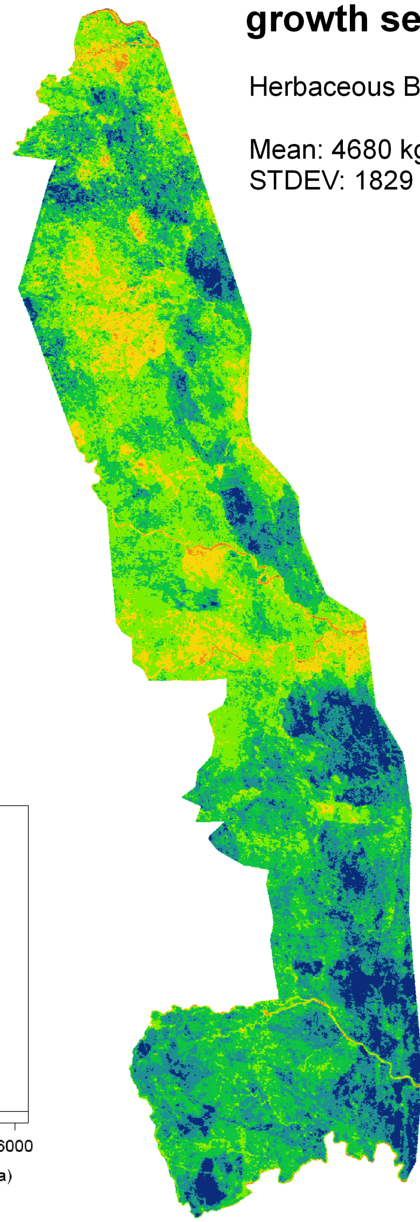
Herbaceous Biomass (kg/ha)

- 0 - 1000
- 1000 - 2000
- 2000 - 3000
- 3000 - 4000
- 4000 - 5000
- 5000 - 6000
- > 6000

End of 2000 - 2001 growth season

Herbaceous Biomass

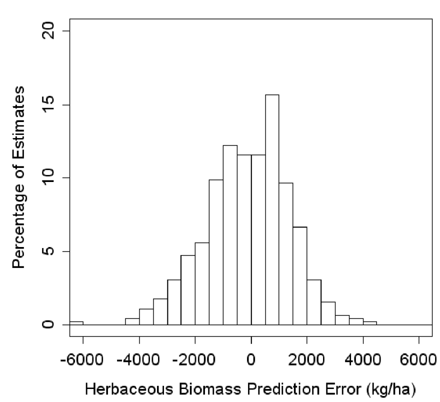
Mean: 4680 kg/ha
STDEV: 1829 kg/ha



Prediction statistics:

RMSE = 1531 kg/ha

Mean Absolute Prediction Error = 37%



Model: sqrt (biomass) ~ wer + sept + burn

Herbaceous Biomass (kg/ha)

0 - 1000

1000 - 2000

2000 - 3000

3000 - 4000

4000 - 5000

5000 - 6000

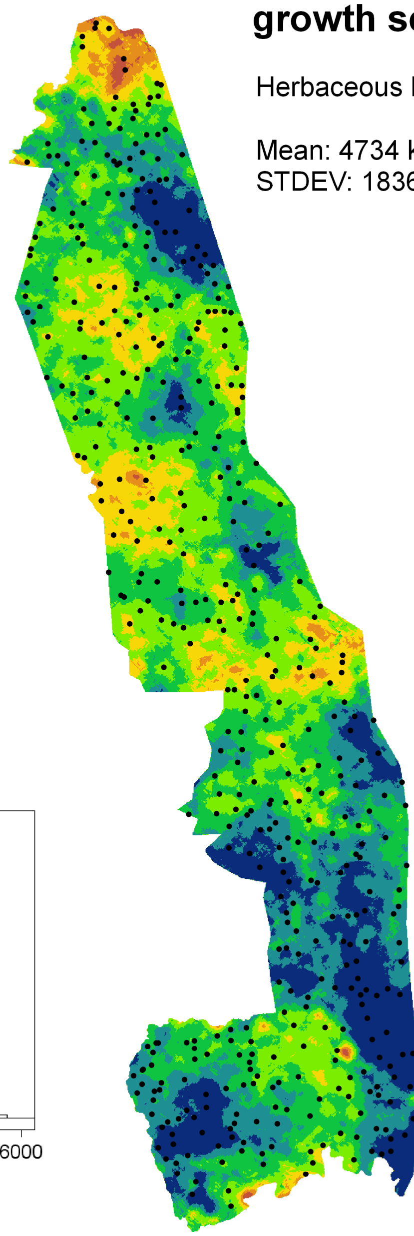
> 6000

· Pasture meter reading

End of 2000 - 2001 growth season

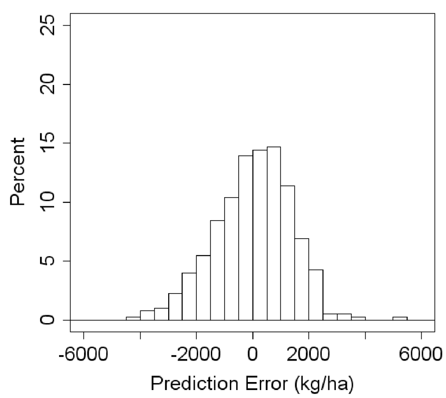
Herbaceous Biomass

Mean: 4734 kg/ha
STDEV: 1836 kg/ha



Prediction statistics:

MAE = 1126 kg/ha
RMSE = 1464 kg/ha



Ordinary Cokriging:
Secondary variable = weighted sum EVI penalised using mean dry season EVI

Herbaceous Biomass (kg/ha)

0 - 1000

1000 - 2000

2000 - 3000

3000 - 4000

4000 - 5000

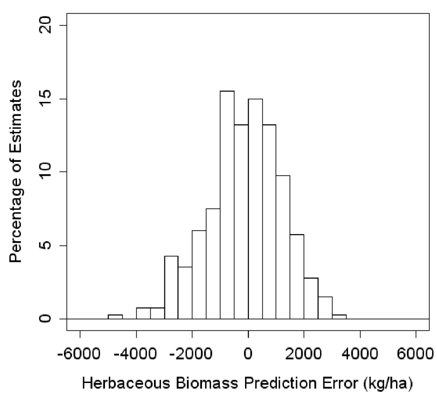
5000 - 6000

> 6000

Prediction statistics:

RMSE = 1339 kg/ha

Mean Absolute Prediction Error = 48%

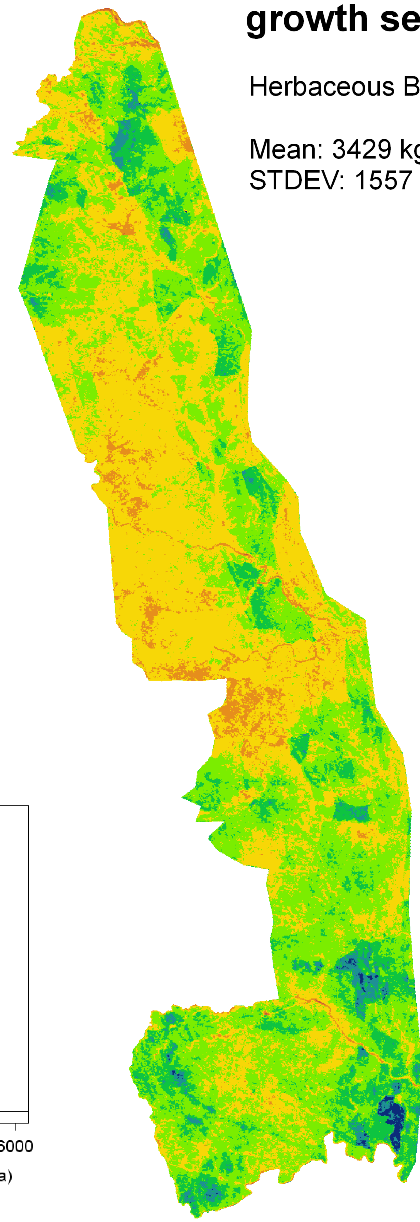


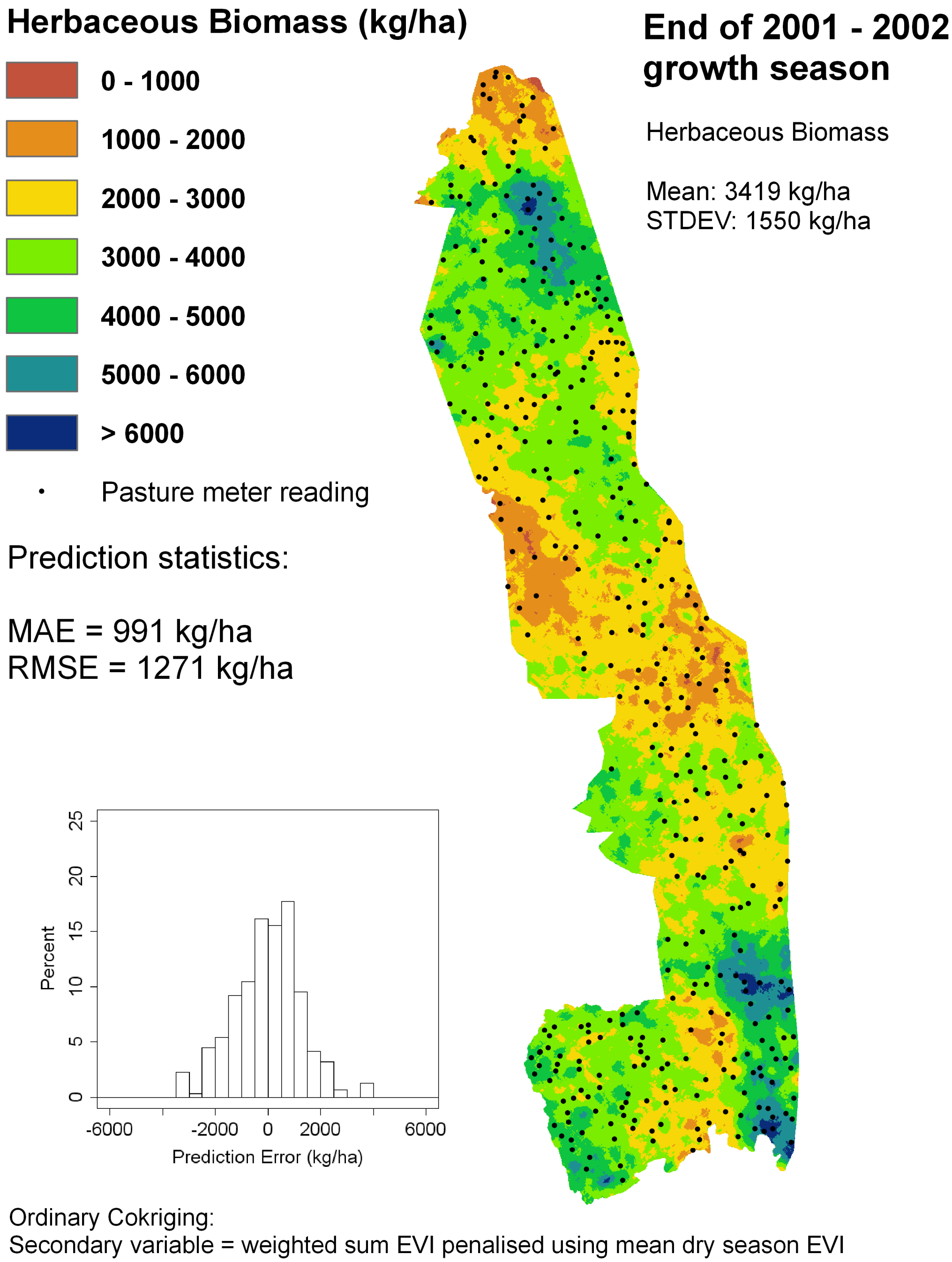
Model: sqrt (biomass) ~ wer + sept + burn

End of 2001 - 2002 growth season

Herbaceous Biomass

Mean: 3429 kg/ha
STDEV: 1557 kg/ha

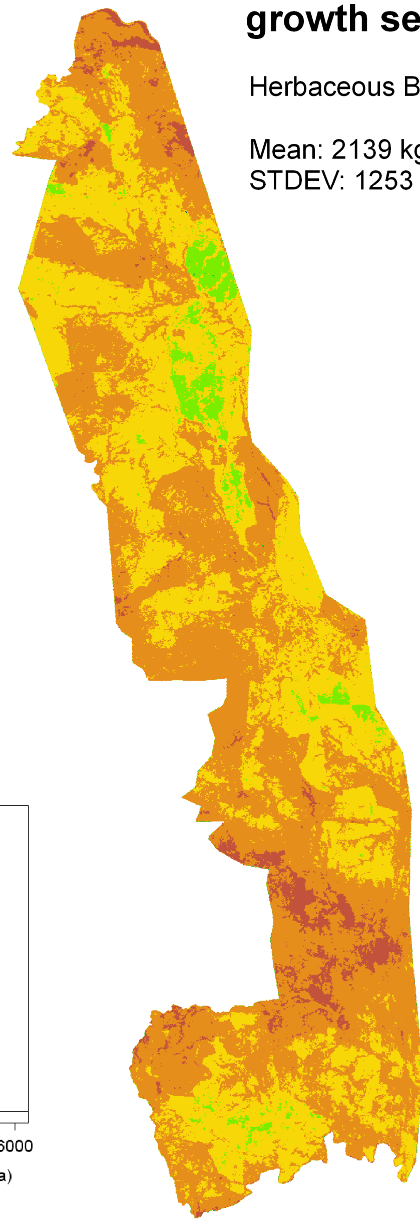



Herbaceous Biomass (kg/ha)
0 - 1000
1000 - 2000
2000 - 3000
3000 - 4000
4000 - 5000
5000 - 6000
> 6000
Pasture meter reading
Prediction statistics:
MAE = 991 kg/ha
RMSE = 1271 kg/ha
End of 2001 - 2002 growth season
Herbaceous Biomass
Mean: 3419 kg/ha
STDEV: 1550 kg/ha
Percent
0
5
10
15
20
25
-6000
-2000
0
2000
6000
Prediction Error (kg/ha)
Ordinary Cokriging:
Secondary variable = weighted sum EVI penalised using mean dry season EVI

Herbaceous Biomass (kg/ha)

- **0 - 1000**
- **1000 - 2000**
- **2000 - 3000**
- **3000 - 4000**
- **4000 - 5000**
- **5000 - 6000**
- **> 6000**

End of 2002 - 2003 growth season



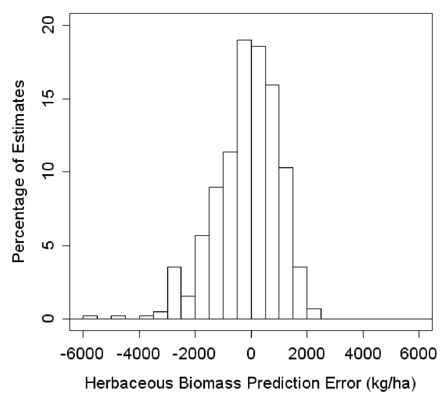
Herbaceous Biomass

Mean: 2139 kg/ha
STDEV: 1253 kg/ha

Prediction statistics:

RMSE = 1141 kg/ha

Mean Absolute Prediction Error = 75%



Model: sqrt (biomass) ~ wer + sept + burn

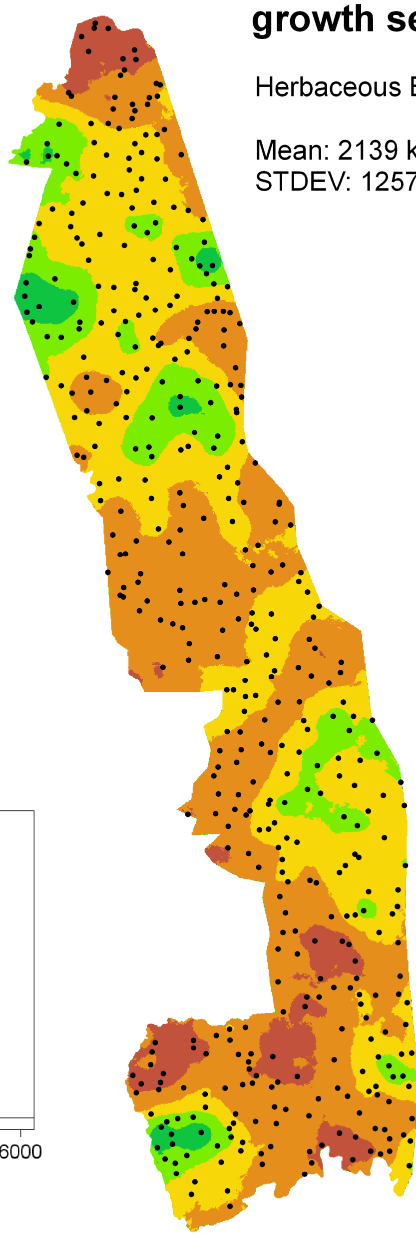
Herbaceous Biomass (kg/ha)

- 0 - 1000
- 1000 - 2000
- 2000 - 3000
- 3000 - 4000
- 4000 - 5000
- 5000 - 6000
- > 6000
- • Pasture meter reading

End of 2002 - 2003 growth season

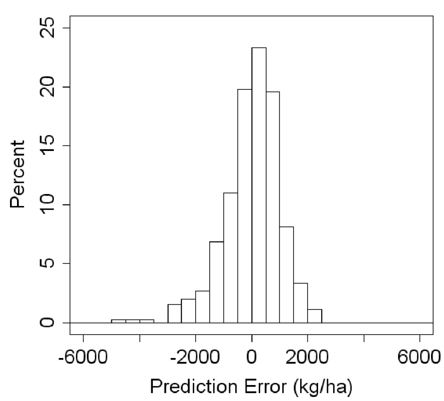
Herbaceous Biomass

Mean: 2139 kg/ha
STDEV: 1257 kg/ha



Prediction statistics:

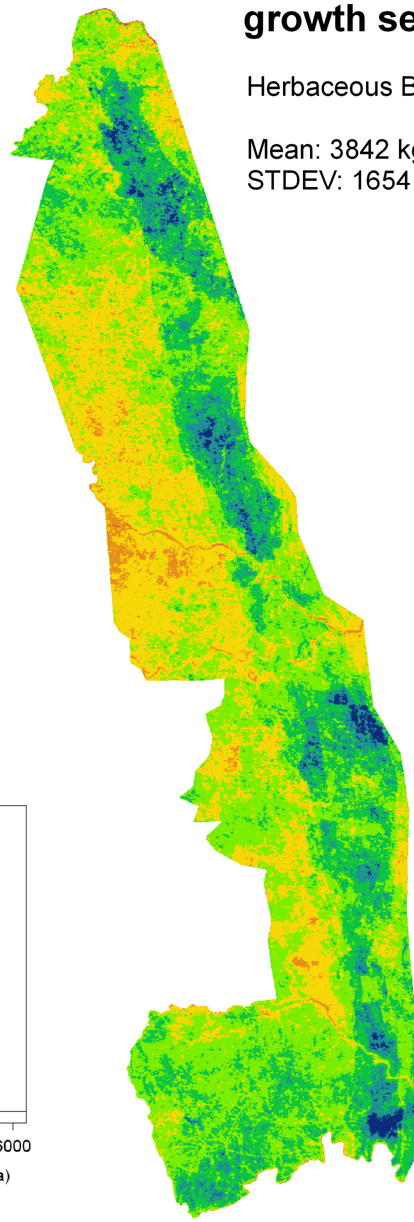
MAE = 749 kg/ha
RMSE = 993 kg/ha



Ordinary Cokriging:
Secondary variable = weighted sum EVI penalised using mean dry season EVI

Herbaceous Biomass (kg/ha)

- 0 - 1000
- 1000 - 2000
- 2000 - 3000
- 3000 - 4000
- 4000 - 5000
- 5000 - 6000
- > 6000



End of 2003 - 2004 growth season

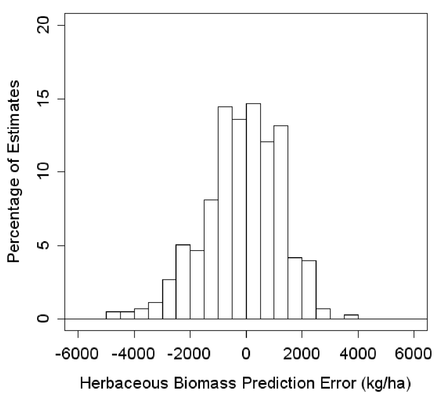
Herbaceous Biomass

Mean: 3842 kg/ha
STDEV: 1654 kg/ha

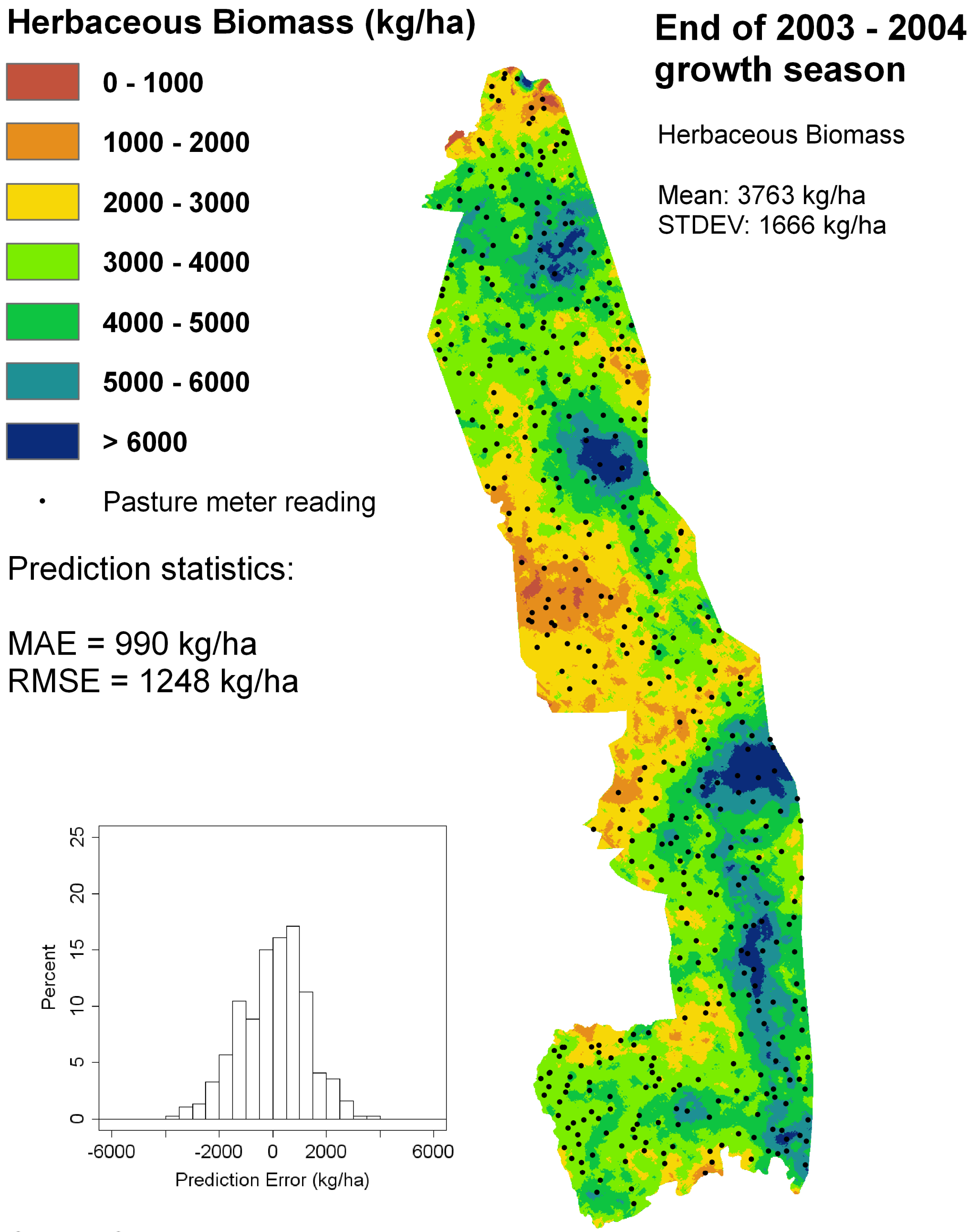
Prediction statistics:

RMSE = 1362 kg/ha

Mean Absolute Prediction Error = 39%



Model: sqrt (biomass) ~ wer + sept + burn



Ordinary Cokriging:
Secondary variable = weighted sum EVI penalised using mean dry season EVI

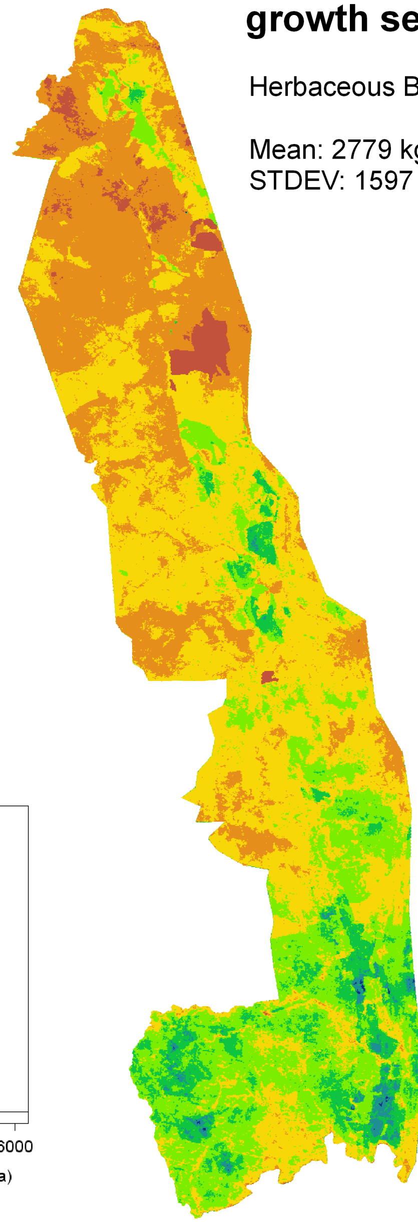
Herbaceous Biomass (kg/ha)

- **0 - 1000**
- **1000 - 2000**
- **2000 - 3000**
- **3000 - 4000**
- **4000 - 5000**
- **5000 - 6000**
- **> 6000**

End of 2004 - 2005 growth season

Herbaceous Biomass

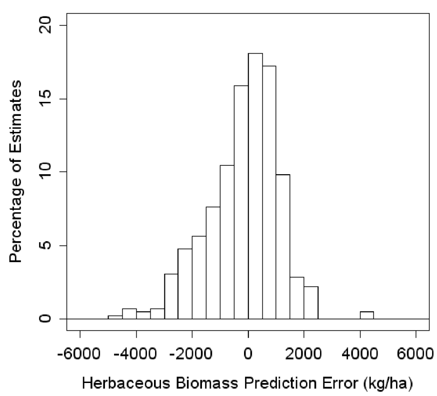
Mean: 2779 kg/ha
STDEV: 1597 kg/ha



Prediction statistics:

RMSE = 1289 kg/ha

Mean Absolute Prediction Error = 113%



Model: sqrt (biomass) ~ wer + sept + burn

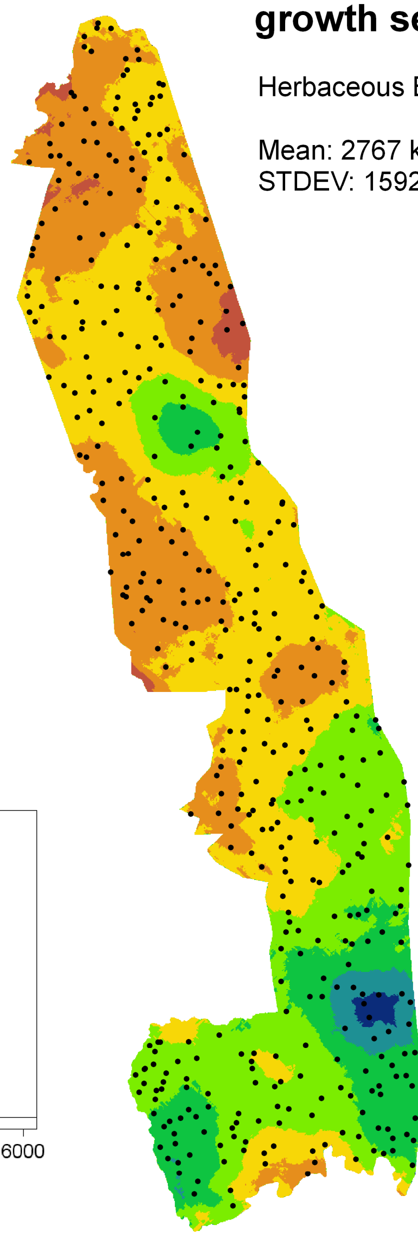
Herbaceous Biomass (kg/ha)

- **0 - 1000**
- **1000 - 2000**
- **2000 - 3000**
- **3000 - 4000**
- **4000 - 5000**
- **5000 - 6000**
- **> 6000**
- · Pasture meter reading

End of 2004 - 2005 growth season

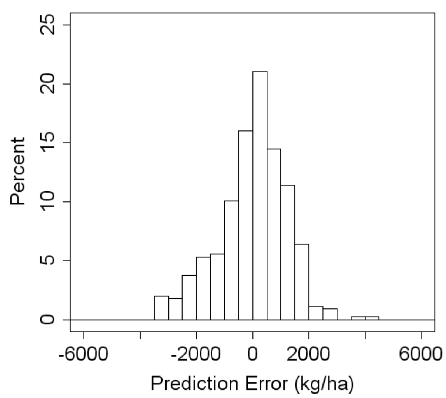
Herbaceous Biomass

Mean: 2767 kg/ha
STDEV: 1592 kg/ha



Prediction statistics:

MAE = 930 kg/ha
RMSE = 1216 kg/ha



Ordinary Cokriging:
Secondary variable = weighted sum EVI penalised using mean dry season EVI

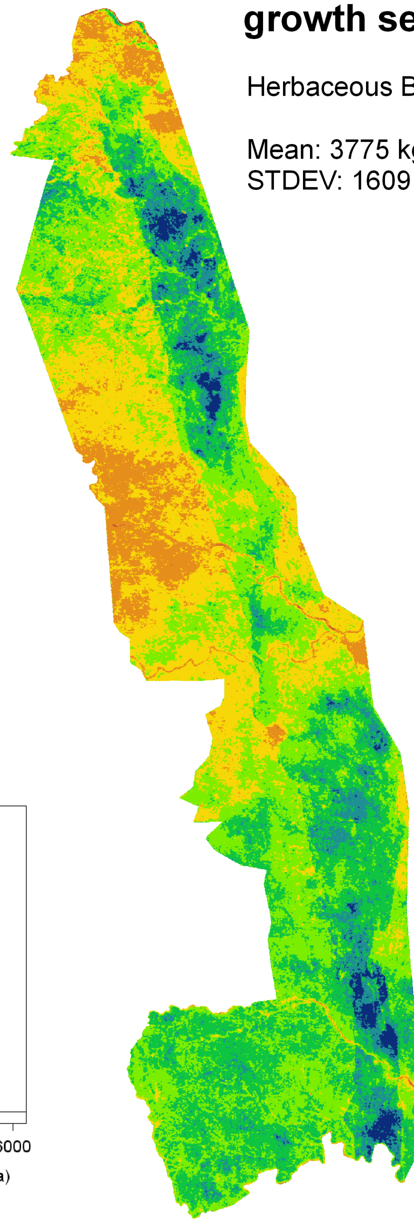
End of 2005 - 2006 growth season

Herbaceous Biomass

Mean: 3775 kg/ha
STDEV: 1609 kg/ha

Herbaceous Biomass (kg/ha)

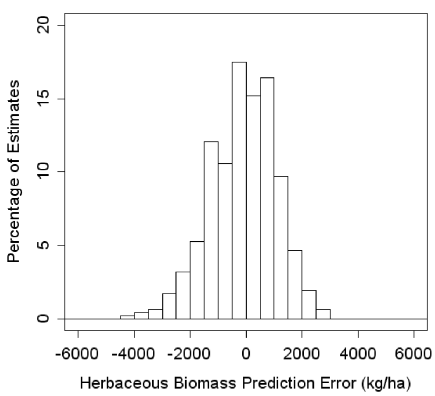
- **0 - 1000**
- **1000 - 2000**
- **2000 - 3000**
- **3000 - 4000**
- **4000 - 5000**
- **5000 - 6000**
- **> 6000**



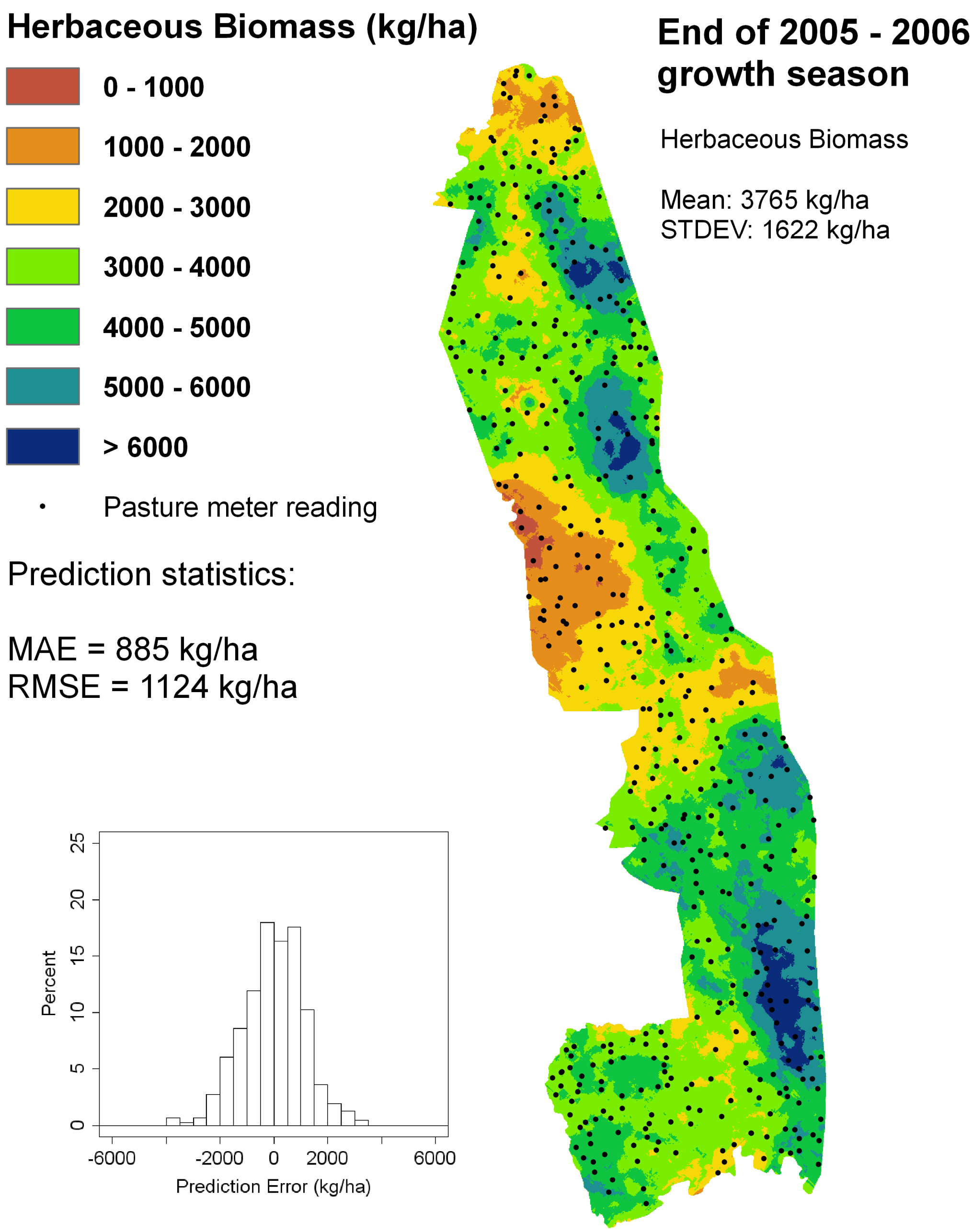
Prediction statistics:

RMSE = 1191kg/ha

Mean Absolute Prediction Error = 35%



Model: sqrt (biomass) ~ wer + sept + burn



Ordinary Cokriging:
Secondary variable = weighted sum EVI penalised using mean dry season EVI

Chapter 4

DISCUSSION AND CONCLUSIONS

1. FIELD DATA

1.1. Assessment of the affects of the VCA sample site dimensions

Field surveys conducted in this study revealed that the mean DPM based estimates for the 60x60m plots differed from the 250x250m plots by between 42 kg/ha and 1308 kg/ha with an average difference of 556 kg/ha. In other words, for the sites sampled in this study, the use of 60x60m plots to sample a 250x250m area result in measurement error of on average 556kg/ha.

The use of mismatched field sample plot and pixel dimensions is not an uncommon occurrence in remote sensing studies. Sannier, Taylor and Plessis (2002) note that because hyper temporal remote sensing imagery has relatively large pixels, that it is seldom feasible to collect field data on plots with comparable dimensions. The scale of the mismatch in this study, 50x60m field data matched to 250x250m MODIS pixels, is however far larger than in many other studies where transects of ±1km were matched to 1km AVHRR pixels (Al-Bakri and Taylor 2003; Moreau et al. 2003; Prince 1991). To minimize the error arising from a mismatch in dimensions it is common to select field sites located in areas showing the greatest level of homogeneity in herbaceous biomass and vegetation type (Wessels et al. 2006; Sannier, Taylor and Plessis 2002). Wessels et al. (2006) found heterogeneity, as measured by variation in LANDSAT NDVI, to increase sharply in the immediate proximity of dams and rivers. None of these sites were located within 250m of either feature. A similar amount of error is therefore likely to exist in the VCA sites that were retained in this study after those located within 250m of dams and rivers were excluded.

At the time of writing there were no published studies detailing the amount of herbaceous biomass prediction error attributable to the use of small field sample plots, highly heterogeneous or otherwise, for the study area. Although my results for this section of analysis are useful in that they place a tentative figure on the average magnitude of the error, the figure is unreliable as there were so few sites sampled. The analysis also falls short of answering one of the most interesting and useful questions we should be asking about the current VCA data. The question is: how much would the correlation between field data and satellite data increase, and the RMSE decrease, if instead of excluding heterogeneous sites we increased field sample plot dimensions to match those of the pixels used?

Answering this question would require a suitable number of the VCA sites to be selected and, during the next round of VCA sampling, DPM readings to be taken for co-located plots of varying sizes at these sites. At least 30-40 sites evenly distributed throughout the study area would be required to perform meaningful statistical analysis. The Inclusion of 60x60m and 250x250m plots at each site would be essential as these represent the current VCA and officially stated MODIS pixel dimensions. If possible 125x125m and 300x300m plots could be added to increase the information provided.

Comparing the R^2 and RMSE values of the regression models produced using the 60, 125, 250 and 300m data would provide a quantitative measure of how the correlation between field data and satellite data change as one approaches and then exceeds the stated pixel dimensions. This information would allow managers to make an informed decision on what trade-off between sample size and prediction accuracy would best suit their budget and management needs.

Based on my experience of similar fieldwork conducted in this study, a two person team could comfortably sample both 60x60m and 250x250m plots at three sites a day using a DPM. This may be reduced to two sites a day if 125x125m and 300x300m plots were included. Assuming one aimed to sample only two sites a day, obtaining 50 sets of site measurements would require 25 working days, or about a month of fieldwork. This is a significant amount of fieldwork to answer a single question but it provides a reliable way of assessing the affects of sample plot size on estimation accuracy. It is also important to note that regardless of the size of the field plots, the VCA field data has been and will continue to be used in remote sensing studies. This is because it is the longest running annually updated herbaceous biomass dataset in southern Africa. It is also, according to Wessels et al. (2006), the best field data available in South Africa for use in a study of this nature. For this reason I believe it is in the interest of the remote sensing and fire management communities that the error attributable to field sample dimensions is quantified. This will provide the data necessary to motivate for larger sites if it turns out that doing so would significantly improve the usefulness of the dataset.

1.2. Error associated with the use of a Disk Pasture Meter

The use of a DPM to collect the VCA herbaceous biomass field data makes it less than ideal for use in a quantitative remote sensing study. The herbaceous biomass data collected using a DPM, although often treated like measurements, are estimates of herbaceous biomass based on disk suspension height. These estimates, if produced using the equation derived during the calibration of the DPM for Kruger National Park by Trollope and Potgieter (1986), have a RMSE of 898 kg/ha. By using DPM data in this study to create models and interpolate surfaces we have used estimates to produce estimates. The model prediction errors presented in this study do not therefore reflect actual real world prediction error but rather how accurately DPM based estimates of herbaceous biomass can be estimated.

Dealing with this issue and arriving at actual estimation error was not included in the aim of this study of this study. Although the issue was not resolved in this study this study's findings have been interpreted by making the reasonable assumption that that actual error will be greater than the figures reported here.

2. REGRESSION

2.1. Assessing the performance of the Vegetation Index variables

It was realized at the outset of this study that there was no clear guidance as to the magnitude of the advantage EVI offered over NDVI as an indicator of production for use in estimating herbaceous biomass. Huete et al. (2002) provide evidence for how EVI avoids saturation over tropical rainforests, but it seems unlikely that biomass within savannas will ever reach that of tropical rainforests. There is also mention of EVI's superior resistance to atmospheric interference caused by water vapour in the MOD 13 algorithm theoretical basis document (Huete, Justice and Van Leeuwen 1999). There were however no published studies on the difference in the strength of the relationship between the two VI's and standing herbaceous biomass, and the resulting differences in estimation accuracy.

Depending on the summation and growth season, the differences in estimation accuracy recorded in this study ranged from 3 – 106 kg/ha, with EVI being better correlated to biomass than NDVI. Based on these findings, EVI is without a doubt the better of the two indexes for use in estimating herbaceous biomass.

In this study it was found that performance of the raw and smoothed data never varied by more than 22 kg/ha. This represents only a minor improvement considering that the total RMSE values are in the range of 1200 – 1700 kg/ha relative to average herbaceous biomass of 3000 kg/ha. Smoothing the data was intended to remove the variation not related to changes in photosynthetic potential introduced by cloud contamination and "marginal" quality pixels. The fact that such a minor change was recorded after soothing the data could therefore be because:

1. smoothing the data is not an efficient means of dealing with the variation in VI signal caused by cloud contamination and “marginal” quality pixels

or

2. the factors causing a pixel to be classified as ‘marginal’ quality do not significantly alter VI values.

If evidence could be found that cloud contamination, and other data quality issues, significantly decreased the herbaceous biomass – MODIS VI relationship for the study, then the second possibility could be dismissed. Cloud contamination is widely cited as a source of error in remote sensing studies but rarely quantified. Sannier, Taylor and Plessis (2002) exclude pixels that are cloud contaminated while Al-Bakri and Taylor (2003) interpolate new values for them. Neither paper assessed what the resulting error would have been had these steps not been taken. Likewise, although there is mention in the literature that viewing angel, and hence pixel quality, can affect VI values (Huete et al. 2002), no studies quantifying the effect that this has on the herbaceous biomass – VI correlation were found. Verbesselt et al. (2006) removed the possibility of extreme viewing angles in the SPOT vegetation data used in their study by excluding any such pixels from analysis. They did not however assess the effect on the correlation between *in situ* biomass estimates and the SPOT data when including pixels obtained at extreme viewing angles. Their results therefore do not provide any information on the extent to which the inclusion of pixels obtained under extreme viewing angles affect the herbaceous biomass VI correlation. No conclusion can therefore be reached as to the most likely explanation as to the minor change in estimation accuracy achieved by smoothing the data in this study.

To quantify the effect of pixel quality on the herbaceous biomass - VI correlation and answer the above question, one would need to extract three subsets from each growth season image stack. The first should consist purely of points overlaid by ‘good’ quality pixels. The second subset should consist of points overlying numerous ‘marginal’ quality pixels, but with none adjacent to one another in the temporal profile. The third should consist of points all overlying pixel stacks with one or more set of marginal quality pixels adjacent to one another in the temporal profile. Comparison of the estimation accuracy achieved when creating regression models with the three datasets would provide a fair indication of the effect of pixel quality on estimation accuracy.

Four different summations of growth season EVI and one 16 day composite were assessed as predictors of herbaceous biomass for each growth season. The performance of the variables was inconsistent between seasons, both in relation to themselves (the amount of variation they accounted for changed between seasons), and one another (the performance of the variables relative to one another changed between seasons). Variation in the amount of variation the variables accounted for between seasons is to be expected because the summations were the only explanatory variables included in the first round of analysis. Any change in an important explanatory variable not included in the regression would therefore result in this sort of variation. Variation in the performance of the variables relative to one another between growth seasons was of more interest. This is because ideally one would like to identify a single VI variable that performs consistently well across growth seasons.

Other studies have made use of either single images (Mutanga and Rugege 2006) or growth season sum VI values (Wessels et al. 2006), with no comparison of the two approaches being made. Verbesselt et al. (2006) compared the two approaches and concluded that the growth season sum approach provides the most strongly correlated variable. However, they combined their data into one large dataset and as a result no information was provided on the consistency of the performance of the temporal stability of either approach. There do not appear to be any studies directly investigating the causes of variation in the performance of the two approaches to which to compare the results obtained in this study.

To understand what caused the variation in the VI variables relative to one another between seasons identified in this study, one needs to look at how the summations differed from one another and how this could affect the correlation between the VI variables and herbaceous biomass. If this can be determined, the cause of the variation could be addressed, and a more consistently useful variable created. The difference in the summations, and hence the reason for the inconsistent performance, lies in the weightings assigned to the different images. All other variables within a season remain constant because the models created for each summation within a season were trained using the same data.

To recap, the summations included:

1. A summation over the same set of dates, September – April, for each growth season

2. A summation taking into account the variation in the onset and of the growth season as determined by a 20% increased VI value on a pixel by pixel basis

3. A summation placing a 30% weighting on the first image in September and increasing the weighting linearly to 100% for the first image in April

4. A pixel by pixel summation placing a 30% weighting on the first value in the growth season time series as determined by a 20% increased VI

value from the previous season low and increasing the weighting linearly to 100% for the first image in April.

The single 16 day image composite selected corresponded to mid April, the period in which the field measurement were reportedly taken (not a summation but included for comparison).

The weighted summations placed less weight on the contribution of images from early in the season than did the ‘fixed’ or ‘variable’ summations which included all growth season images at equal weighting. The single image composite on the other hand in effect places zero weight on all preceding images. In the 2003 – 2004 and 2005 – 2006 growth seasons, this difference led to the single image composite and the weighted growth season sum performing better than the fixed growth season sum. It can therefore be said that the additional production information the fixed and variable summations contained for these growth seasons was not well correlated to end of season herbaceous biomass. For this to be the case, the production they provided information on must not have resulted in the accumulation of herbaceous biomass.

For production to lead to accumulation of herbaceous biomass, production must exceed removal through fire, herbivory and decay. A plausible explanation for the above is that at certain times during these seasons this production threshold was not exceeded for parts of the study area. According to Trollope (2008) this often occurs in sweetveld areas due to the palatability of the grass, which persists even when it is dry. It is not impossible that the same could occur in sourveld regions when rainfall is sufficiently low. Assuming this was the case, the majority of the production as measured by the VI summation under these conditions would not be present in the herbaceous layer when the VCA pasture meter readings were taken and used to produce herbaceous biomass estimates. The fixed and variable summations would however have counted the contribution of the images from these periods as equal to the contribution from any other period.

This would have resulted in a weakened correlation of the VI sum to measured herbaceous biomass.

An alternate and equally plausible explanation is that some of the additional information on production in the fixed summation was attributable to woody vegetation canopy and not herbaceous production. During periods of limited water availability, the VI signal is dominated by the leaves of woody vegetation (Archibald and Scholes 2007). These remain green because the trees can access deeper water reserves than the shallow rooted grasses. This is most likely to occur at the beginning of the growth season because trees green up as day length increases regardless of whether the spring rains have begun, whereas grass does not (Archibald and Scholes 2007). In this case the variable summation could potentially contain even more information incorrectly attributed to herbaceous production. This is because the EVI threshold of a 10% increase from the previous season's lowest value, used to determine growth season start date, often resulted in a longer growth season than the fixed growth season. The 10% increase in EVI most likely came from the greening up of the woody canopy and not the herbaceous layer, thus making the summation even more prone to the error discussed above.

Distribution of the rainfall within these two growth seasons can be interpreted as providing support for both of the above explanations. If one looks at the rainfall in the 2003 – 2004 growth season (Figure 2, Chapter 3), it becomes apparent that the bulk of production would have occurred very late in the growth season. This is because herbaceous production is dependent on moisture availability and significant rainfall only occurred at the end of the season between January and March. This being the case, it is possible that the woody layer dominated the EVI signal in the early part of the season. It is also possible that production only marginally exceeded removal through herbivory in the early parts of the season and hence contributed little to end of season standing crop. This would have been exacerbated by the fact that there was very little rainfall and hence production in the 2002 – 2003 growth season. As a result very little production would have carried over and the limited amount of new growth would have constituted the bulk of available grazing, leading to a large percentage being removed. A similar rainfall distribution occurred in 2005 – 2006. The major difference seems to be that the onset of significant rainfall occurred slightly earlier in the season than in the 2003 – 2004 growth season. This would have resulted in the accumulation threshold being exceeded in more of the months captured by the weighted summation, improving its performance relative to the single image composite when compared to the 2003 – 2004 season.

Either of the explanations, or a combination of the two, would account for why the single image composite and the weighted summation variables, that place less weight on these periods, performed better than a fixed summation for the 2002 – 2003 growth season. Factoring in these possibilities and adjusting for them when creating summations in the future could lead to the creation of a single variable that performs consistently well across seasons.

Creating such a variable would require one to track the variation in accumulation determined by the interaction between rainfall, herbivory and production. It may be possible to achieve this by establishing a threshold for EVI values through further experimentation which must be exceeded before EVI values are added to the summation. The assumption behind this is that EVI values below the threshold are either reflecting production that is insufficient to result in accumulation or are attributable primarily to the woody layer.

Alternately, instead of determining thresholds through experimentation and creating a single VI variable through summation, all of the EVI images in a growth season could be entered as individual variables. Stepwise regression could then be used to empirically determine which EVI images should be included in each seasons regression model, Stepwise regression would also do away with the need to derive weightings for each image to account for removal through herbivory. In this study arbitrary weightings were applied because time prohibited anything more advance. By using stepwise regression optimal 'weightings' in the form of regression coefficients would be automatically calculated as part of least squares fitting.

If none of these options have been pursued it would be advisable to run exploratory regressions to identify the best performing of the EVI variable assessed in this study based on correlation to the available herbaceous biomass data.

2.2. Assessment of a suitable woody canopy cover variable

One of the challenges involved in estimating herbaceous biomass in the study area using satellite derived VI data is variation in woody canopy cover. Ideally one would like for there to be only one vegetation layer contributing to the VI signal, and for that layer to be the one of interest. This is not the case in the study area where in addition to the herbaceous layer there is often a significant woody layer.

The presence of a woody layer does not however mean that no information on the herbaceous layer can be extracted from a VI signal. Apart from riparian zones there are few areas in the study area where closed canopy forest exist. In the vast majority of cases there will be VI signal that is originating from un-obscured herbaceous layer somewhere in a MODIS Pixel. As canopy cover and density increase, a greater percentage of the herbaceous layer will be obscured from view by optical sensors, increasing the error associated with estimates based on remotely sensed VI data. The accuracy and therefore usefulness of optical remote sensing based herbaceous biomass estimation methods will therefore decrease with increased woody canopy cover.

The finding that there was an apparent lack of any significant relationship between the tree cover variables and high resolution derived tree cover validation data was unexpected. The tree cover variables were chosen because, based on the literature consulted (Scanlon et al. 2002; Archibald and Scholes 2007; Hansen et al. 2003), each was expected to show at least a weak positive correlation to tree cover. The high resolution derived data used to check them against seems unlikely to have been the problem as visual comparison of the tree cover layer with aerial photography shows a close match. The following possibilities remain:

1. The variables are in fact not correlated to tree cover
2. The accuracy assessment was performed incorrectly.

It is unlikely that option 1 could be the case for all three variables. The lack of a significant relationship across all three variables suggests that something else is more likely at fault. It is most likely therefore that there was a fault in the accuracy assessment performed.

The accuracy assessment was conducted before I had developed an understanding of the point spread function (PSF), or the occurrence of pixel shifts when gridded products are created. This led to a failure to account for the fact that the point spread function of MODIS data results in 25% of the pixel signal originating from outside the area of the pixel and that signal contribution increases from the pixel edge to the centre. Validation pixels exactly matched in size and location to the tree cover variables pixels were derived from the high resolution data by calculating the mean value of all of the high resolution pixels that fell within them. No weighting was applied to the individual high resolution pixels based on their location relative to the pixel centre to account for the PSF. Even if this alone does not fully account for the lack of any significant relationship in the validation step, it does go some way towards doing so.

The MOD44 variable performed poorly in four of the years and yet significantly outperformed the September – October mean EVI variable in 2003 -2004 and 2005 – 2006. The poor performance of the variable in those four years may be attributed to the difference in resolution between it and the EVI summation. Use of a variable not measuring the effect of tree cover over the exact area of the production variable would be expected to account for less variation than one that did. A variable designed to provide a global scale measure of canopy cover could also be expected to perform poorly at accurately identifying local variations in canopy cover. No logical explanation could be arrived at as to why the variable could suddenly perform better than all others in two growth seasons. It is possible that the improvement occurred by chance, but even if it was not, its inconsistent performance makes it difficult to recommend using in future studies of this nature.

The AVHRR NDVI derived variable, arrived at using the method outlined in Scanlon et al. (2002), performed extremely poorly. The method is complex and in retrospect may have been beyond my technical ability to correctly implement. The AVHRR data is also of coarser resolution than the EVI summation and resampling the data using nearest neighbour resampling to match the MODIS data may have introduced additional error. Even if executed correctly the method has its limitations. It assumes that there are end member pixels, made up of 100% forest, grass and bare soil in the training data. It is highly unlikely that any of these will occur within the study area, given a pixel resolution of 1km. This means that the end members selected will be compromises containing far more of the other cover types than intended. Furthermore the method requires the subjective selection of end members from the plot of response in NDVI to rainfall vs. long term mean NDVI. The tree cover variable produced will therefore be highly influenced by the person selecting the end members and the purity of the pixels that end up being used as end members in the study area. In the case of this study the above factors have produced a variable that is an extremely poor predictor of tree cover. The patches of homogenous cover required to produce good end members occur in the study area at scales finer than the spatial resolution of any imagery that provides the temporal resolution required for the method.

Overall the September – October mean EVI variable accounted for the greatest decrease in RMSE. Its superior performance when compared to the other two woody cover variables may be in part attributable to the resolution of this variable matching the EVI summation perfectly. This exact matching existed because both the EVI summation and the September – October mean EVI were derived using the same gridded MOD13 data product and so the pixels overlap perfectly. In all growth seasons adding the variable to the regression model resulted in appreciable decrees in RMSE. This is in agreement with Wessels et al. (2006) and Fuller, Prince and Astle (1997), who reported the existence of a negative relationship between the density of woody vegetation and herbaceous biomass in southern African savannas.

It is also apparent that there was little difference in the improvement in estimation accuracy when including or excluding an interaction term between the September – October mean EVI and the EVI summation. In other words, it did not matter very much how the September – October mean EVI tree cover variable was included, it still accounted for similar decreases in RMSE. Interaction terms are intended to account for the fact that the relationship between two variables is affected by a third. In this case it was assumed that the amount of woody vegetation present would affect the relationship between EVI and herbaceous biomass. One can imagine how this would manifest in a situation where the same amount of herbaceous biomass was present in two areas but that the density of trees was higher in the second area. The greater density of trees would lead to higher EVI values because the leaves of the trees cannot be distinguished from the grass by the satellite sensor. Failure to include an interaction term would result in this difference being introduced into the error term as variation unaccounted for in the relationship between EVI and herbaceous biomass. Very little variation in the RMSE, regardless of whether an interaction term was included, suggests that the signal is usually dominated by the herbaceous layer. Hence, an interaction term is not necessary. This is in line with the findings of Fuller, Prince and Astle (1997)

Adjustment of the EVI prior to analysis was found to be a less reliable means of accounting for the effect of trees on the herbaceous layer than including a variable in the regression equation. That such a simplistic and clearly flawed method could produce improvements only slightly less than including the variable in the regression in all growth seasons except for the very dry 2002 – 2003 was unexpected. I consider the method simplistic and flawed, and the results unexpected, because by subtracting the full amount of the September – October mean EVI from each EVI composite in the summation, the implicit assumption is that 100% of its value is attributable to the woody layer. Although it has been shown that trees green up earlier than the herbaceous layer for most years in the study area (Archibald and Scholes 2007), it is highly unlikely that there will consistently be no herbaceous activity in September or October. It was also unexpected because the failure of the interaction term to provide consistent additional reductions in RMSE also invalidated the primary assumption that the pre-correction was based on. This assumption was that the woody layer contributed enough to the EVI signal for the pre-correction of EVI to remove this contribution would provide significant improvements to RMSE. The improvements in RMSE yielded by the pre-adjustment method are therefore not because it corrects for the contribution of the woody layer to EVI. The only explanation I can offer for this is that the pre-adjustment is accounting for the inverse relationship between canopy cover and herbaceous biomass that exists when trees reach sufficient density to shade grasses enough to reduce their production.

Of all of the inclusion methods and variables assessed, the simple addition of the September - October temporal mean EVI performed best, reducing estimation error by on average 82 kg/ha and as such is recommended for future use in place of the other two methods. It is noted that the increased availability of RADAR and LiDAR data for the study area has already led to more accurate measure of tree cover being available than were used in this study. Boggs (2010) reported accuracies of > 85% when mapping tree cover for selected sites in the study area using Quickbird imagery. The greatest barrier to applying the method he describes to the entire Kruger National Park is the cost of the imagery. Assuming a cost per km^2 of $55 (www.eurimage.com, 2010), this would amount to $1,045,000, assuming 19000 km^2 of imagery would be sufficient to cover the entire park. Given an exchange rate of 7.8 rand to the US dollar, as it was at the time of writing, this equates to ZAR 8,151,000. With no knowledge of the KNP's budget I cannot say whether this would ever be a possibility, but I strongly suspect that this is more than management would be willing to spend on a single dataset. Unless RADAR based methods can deliver both affordable and sufficiently accurate estimates of woody cover, the use of the September - October temporal mean EVI may remain a viable option, regardless of the variable's limitations.

2.3. Assessment of a suitable variable to account for dry material

A large amount of dead herbaceous biomass is known to accumulate and persist between seasons in the study area (Govender, Trollope and Van Wilgen 2006). This material is not reflected in EVI values because of the absence of chlorophyll (Thompson and Everson 1993). For this reason a variable accounting for some of the variation in herbaceous biomass brought about by this dead material was sought. The variable arrived at was one that combined fire history and geology. Fire history provides information on how many seasons of accumulated growth could be present. Geology was intended to account for the fact that the amount accumulating during that time would vary between sites. Geology was, admittedly, a poor predictor of variation in production but was used because it does have some influence on it and was readily available.

Although the variable was simple to derive and clearly a compromise, it did lead to increases in estimation accuracy of over 40kg/ha in four of the growth seasons. Even so, this never represented more than a 6% increase in estimation accuracy. What's more, during the extremely dry 2002 – 2003 growth season, it accounted for just 14 kg/ha of additional variation. During this growth season the herbaceous layer must have consisted almost entirely of dry material because of limited rainfall. The majority of herbaceous biomass would have to have originated from previous seasons. One would expect a variable accounting for the presence of dry herbaceous biomass to have performed extremely well in these conditions, rather than performing this poorly.

It has been shown by Thompson and Everson (1993) that within grasslands, after three years of accumulation of dead herbaceous material, the correlation between field measurements and NDVI values can fall to zero. It is also known that within the study area, dead herbaceous material frequently constitutes a large percentage of the herbaceous layer (Trollope 2008).
Both of these factors suggest that there is a good chance that the presence of dry material is responsible for a large amount of the unexplained variation in the relationship between the DPM estimates and the EVI values. It would be beneficial to attempt to derive a variable that better accounts for its presence. A variable that could potentially do so is one derived by combining an EVI based production estimate from previous seasons with burn scar data to identify appropriate dates to sum between.

2.4. Assessment of the completed regression models

Even with two additional explanatory variables, one to account for the affect of canopy cover on the herbaceous layer and one to account for the presence of dry material, the models performed disappointingly. In 5 of the 6 growth seasons, the models created accounted for less than 35% of the variation in herbaceous biomass. This is in line with the findings of Wessels et al (2006), who regressed the VCA data against AVHRR NDVI, landscape groups and tree cover data, achieving R^2 values between 0.08 – 0.41.

In contrast, studies in which herbaceous biomass was estimated using AVHRR data in Senegal (Tucker et al. 1985) and Jordan (Al-Bakri and Taylor 2003), without any additional variables, reported R^2 values of > 0.6, which roughly equates to having accounted for >60% of the variation in herbaceous biomass. Unfortunately these studies can provide only limited insight into how estimation accuracies could be improved in this study. This is because they were not conducted in comparable vegetation types. Al-Bakri and Taylor (2003) conducted their study in Jordan. The study area was not reported as having any significant woody layer and experienced extremely low, 100–200mm mean annual rainfall, resulting in limited production and little or no carry-over. “The green flush lasts for a very short time and tends to be overgrazed shortly after it occurs” (Al-Bakri and Taylor 2003). Similar conditions are described for Senegal where Tucker et al. (1985) conducted their study, although tree cover ranging from 5% – 20% was reported and rainfall of up to 200-400 mm/annum. In both of these systems the relationship between photosynthetic potential and herbaceous biomass production is less complicated than in the savanna encountered in the Kruger National Park. There is little carry-over between seasons, the influence of trees would only be an issue in a handful of pixels and spatial heterogeneity is relatively low (Tucker et al 1985). The lower R^2 values obtained for this and other studies conducted in the study area, despite the presence of additional explanatory variables, is attributable to the numerous variables that must be accounted for when creating such models in southern African savannas. In retrospect, my disappointment at the modest correlations reported for this study arose because I failed to fully comprehend just how very different these study areas were compared to my study in the Kruger National Park.

The single model created using data from all of the years combined performed well relative to the individual models trained for the growth seasons between 2001 and 2005, averaging a decrease in accuracy of only 11 kg/ha. The prediction error from the single model for the first and last growth season on the other hand exceeded that of the individual growth season models by 156 and 198 kg/ha respectively. This indicates that there is a factor involved in determining end of season herbaceous biomass which is not well accounted for by the models assessed. Fluctuations in this factor cause the models created for individual growth seasons to outperform the general model for the entire period because the effect of the factor, held constant within each season, is captured in the individual season models. This is not a major issue in the Kruger National Park where field data is collected every year. New models can be created at the end of each growth season to ensure maximum estimation accuracy is achieved. It is however an important point to be aware of if similar models are to be created for other areas, located in savannas with similar vegetation structure and variation, where yearly biomass field estimates are not collected. If a single model is created and applied across years, users of the model should be aware that the accuracy of the estimates produced could potentially fluctuate by an amount comparable to that reported above. It would be up to the user of the herbaceous biomass estimates to decide if the resulting level of accuracy would be sufficient.

In this study a conscious decision was made to try and keep the model as general as possible. The complexity of the landscape within the study area does however make stratification an appealing option. Stratifying the landscape and creating individual models for each sub region would likely improve estimation accuracy and should be investigated further by those interested in producing herbaceous biomass estimation models to be applied only within the Kruger National Park. Geographically weighted regression is another alternative to stratification or the extensive use of dummy variables which may provide improved estimation accuracies.

3. KRIGING AND COKRIGING

The results obtained for kriging and cokriging are perplexing. Cokriging was expected to perform either on par with or better than kriging. Instead, in three of the six years, it performed worse. In those instances where it did result in an improvement, it resulted in less of an improvement than expected when compared to a similar study conducted in the area. Mutanga and Rugege (2006) working with the VCA data and MODIS band 2 reported an increase in estimation accuracy of 178 kg/ha over kriging and 554 kg/ha over a regression model. The R^2 of the relationship between the VCA herbaceous biomass estimates and MODIS band 2 in that study was 0.44. By comparison the use of an EVI summation from the 2005 – 2006 growth season in this study correlated to herbaceous biomass with an R^2 of 0.36, provided an improvement over kriging of only 76 kg/ha. The improvement over the regression model with an R^2 of 0.46 created specifically for 2005 - 2006 was only 97 kg/ha. All of this suggests that there was some major flaw in the cokriging implemented in this project.

The implementation of cokriging is dependent on both the decisions made by the operator and the algorithm embedded in the software. When implementing cokriging in this project, many of the default parameters related to trend removal and number of points to include suggested by the ArcGIS geo-statistical extension, were accepted. The semivariogram model for the primary variable was modeled for each growth season to the best of my ability through adjusting lag size and experimenting with different nugget and sill values before deciding on a compromise and specifying the lag value and accepting the optimized nugget sill and range values. No option was given for experimental modeling of the semivariogram model for the secondary variable or the cross variogram model.

Despite my attempts at ensuring the models specified fitted the semivariaograms, results comparable to those of Mutanga and Rugege (2006) could not be obtained. Inspection of the ILWIS software used in their study revealed that it does not provide an equivalent of the Geostatistical wizard available in ArcGIS and the associated option of automatically 'optimized' parameters. Instead it requires the user to work through each step of the process, separately fitting each model through experimentation and using the parameters obtained as inputs into the subsequent steps.

It seems most likely then that using the geostatistical wizard available in ArcGIS resulted in sub-optimal semi variogram models being fitted. This might explain the inconsistent performance of cokriging relative to kriging, and hence my inability to achieve accuracies comparable to those of (Mutanga and Rugege 2006). Due to time constraints I was unable to redo the analysis using the ILWIS software. Even so the results obtained produce the useful finding that kriging and cokriging implemented by the ArcGIS geostatistical wizard is unlikely to achieve the levels of accuracy possible when the user is forced to optimize all of the semivariaogram model parameters manually. This is an important finding as ArcGIS is one of the most widely used commercial GIS packages, and many ecologists might consider it their first option when needing to interpolate surfaces from point measurements or estimates such as rainfall and herbaceous biomass.

4. COMPARISON OF THE ACCURACY AND PRECISION OF THE REGRESSION MODEL AND COKRIGING APPROACHES

This section is critical in terms of achieving the aim of this study, which was the to compare the accuracy and precision achieved using cokriging and a linear regression model for producing spatially explicit herbaceous biomass estimates using 250m MODIS VI data. An attempt to go beyond a simple comparison of RMSE values and prediction map characteristics has also been made. The two methods are assessed in terms of the resources required for implementation, and the implications of the studies findings for savanna management are discussed.

In terms of accuracy, even though cokriging was not performing optimally in this study, with a mean RMSE of 1220 kg/ha, its estimates of herbaceous biomass were on average 119kg/ha better than those of the regression model. This is in line with the findings of Mutanga and Rugege (2006), although they achieved a much more significant 544 kg/ha improvement. That said, Mutanga and Rugege (2006) only used data from a single growth season in their study. Whether similar results could have been obtained in consistently in different seasons is unknown.

Histograms of the residuals reveal that both methods are unbiased estimators, producing estimates with errors centered on zero. The precision achieved using regression models, measured using Standard Deviation of estimation accuracy between seasons, was greater than that achieved with cokriging, but only by 20kg/ha. Considering that the accuracy of the cokriging estimates were consistently the best, a slightly lower precision figure would not, based on these figures alone, cause anyone to identify the use of a regression model as the preferable method.

Although the RMSE figures for the two methods differ on average by only 119kg/ha, the predictions as to the distribution of herbaceous biomass differ widely. Large areas on the maps produced show differences of more than 1000 kg/ha in the amount of herbaceous biomass predicted. The areas with the largest discrepancies differ from year to year indicating that the difference in performance it is not due to some geographically fixed underlying factor such as vegetation type. There are also no spatial trends apparent in the residuals of either method for any season. The difference in prediction accuracy cannot therefore be attributed to a specific factor in any one location. Whatever is causing the regression model to produce less accurate herbaceous biomass estimates than cokriging must therefore occur fairly evenly throughout the study area.

The maps produced using cokriging have a smoothed appearance when compared to those produced using the regression models. It is tempting to assume that because the regression maps *appear* to provide more detail, they will be of more use as decision aids than the cokrieged maps. However, as mentioned previously the seemingly less detailed cokriged maps provide, based on the ground truth data available, a more reliable indication of herbaceous biomass. This means that they should also be regarded as more reliable by decision makers even though they appear less detailed.

The fine spatial resolution of the remotely sensed data results in abrupt changes in vegetation properties being identifiable on the herbaceous biomass maps created. Good examples of this are the large sandy riverbeds which stand out clearly on the maps produced using regression models. The inclusion of categorical variables also adds to the sense of increased detail through the sharp boundaries created by the different intercepts associated with the different levels of the categorical variable. Rapid changes in EVI over a short distance or a transition form one type of geology or burn history to another may correspond to a real and significant change in herbaceous biomass. However, if the combination of measurements available does not explain more of the variation in herbaceous biomass than can be inferred from the location of a point relative to a series of suitably arranged field measurements, the extra detail is of no use. The maps produced may reflect in detail the variation in the properties measured, but they provide less information on the abundance of herbaceous biomass than the cokriged estimates.

Because resources are usually scarce, it is useful to compare the methods not just in terms of the accuracy and precision achievable, but also in terms of the resources required to implement them. The creation of a regression model for a single season was the most resource intensive method to implement in this study. It required both field measurements that were time-consuming and relatively expensive to gather and measurements derived from satellite imagery for the explanatory variables. Hence the method requiring the most resources to implement did not provide the most accurate estimates of herbaceous biomass, and only a marginal increase in precision. Cokriging required fewer resources, but only marginally so as it did not make use of the fire history variable included in the regression models created. Even though it required marginally fewer resources, and was marginally less precise, it produced the most accurate estimates in the study (apart from kriging in certain instances, although if cokriging had been optimally implemented it could not have produced estimates worse than, only equal to, kriged estimates).

Creation of a single regression model to predict herbaceous biomass over a number of seasons required the same amount of data as all of the individual models combined. However, if its resource costs were calculated on a per season basis over a greater number of seasons than were used to provide the training data, it would be found to require the fewest resources. All of the measurements of explanatory variables required for subsequent estimations can be produced using remotely sensed data. Burn scar maps can be digitized from high resolution imagery, while the EVI summation to approximate production and the September - October mean EVI to account for the effect of woody vegetation are both derived from freely available MODIS data. Provided the imagery used to derive the fire scar maps is not excessively expensive, the cost of obtaining these measurements should be modest. By avoiding the need for constantly acquiring herbaceous biomass field estimates, which are expensive and time consuming to obtain, the use of a single regression model saves resources. The estimates produced by such a model are however less accurate than either cokriging or the creation of a growth season specific model.

In the case of the Kruger National Park, where herbaceous biomass data is collected every year for the VCA dataset, the best choice of method based on accuracy measured using RMSE and resource requirements, is cokriging. It also seems likely that the accuracy of the predictions could be further improved by using software such as ILWIS, which allows for greater control over the variogram models fitted to the data. The regression models created in this study simply cannot compete in terms of accuracy or resources required and only offer a marginal increase in precision. They reflect the fine scale variation in vegetation greenness, fire history and geology within the study area far better than cokriging could by virtue of having measurements for every pixel present. Unfortunately, variation in these properties does not account for variation in herbaceous biomass as accurately as does the relative position in space of each pixel to the VCAestimate.

Given the current data available in the Kruger National Park for producing the estimates, cokrigings combination of greater accuracy, comparable precision and marginally lower resource requirements make it the easier of the two methods assessed to recommend for operational implementation. Given a lower sampling density, it is highly likely that the reverse would be true, although this was not tested in this study.

5. IMPLICATIONS OF THE STUDIES FINDINGS FOR SAVANNA MANAGEMENT

The accuracy of herbaceous biomass estimates desired by the KNP fire management team based on the figure reported in a paper by Wessels et al. (2006) is 500 kg/ha. Given that the DPM conversion equation described in Trollope and Potgieter (1986) has a residual error of 898 kg/ha, and that a DPM is used to collect the VCA data, it would be impossible to achieve estimates with an accuracy of the order of 500kg/ha using this data. This is because the measurement error in the DPM estimates sets the maximum level of estimation accuracy achievable. Improvements to the DPM calibration equation would need to be made to reduce this error to below 500 kg/ha, followed by improvements to the variables affecting the correlation between herbaceous biomass and the EVI growth season sum. However, given that disk suspension height, which is a direct field based measurement, has in the past delivered estimation accuracies of only 898 kg/ha, it seems unlikely that the relationship between the remotely sensed data and standing herbaceous biomass could deliver estimation accuracies less than 898 kg/ha. There are simply too many additional variables, present because of the distance between the herbaceous layer and the sensor, that weaken the relationship between remotely sensed data and end of season herbaceous biomass.

During an informal conversation with Prof. Winston Trollop, a fire ecologist who has worked extensively in the study site and across Southern and East Africa, it was ascertained that, although 500 kg/ha would be desirable, estimation error as high as 1000 kg/ha would be acceptable. Based on the assumption that measurement error in the dependent/response variable causes an increases in estimation error equal to its magnitude, any model created using DPM based estimates containing 900 kg/ha of error can add no more than 100 kg/ha of error if it is to achieve the required 1000 kg/ha accuracy. Given that the average error for the regression models created in this study was 1339 kg/ha, 1000kg/ha seems a far more realistic goal to aim for than 500gh/ha. Given the distance between the satellite sensor and the herbaceous layer, combined with the effects of herbivory and dead material from previous seasons, limiting additional error to 100kg/ha is not something that could be consistently achieved using a regression model, even with improvements to the variables used in this study. Cokriging on the other hand, using a combination of the information regarding spatial autocorrelation in the herbaceous layer and the herbaceous biomass – EVI sum relationship, offers greater hope. Cokriging produced average errors of 1220 kg/ha, just 220 kg/ha above the acceptable level of accuracy, It must however be remembered that this is not the actual accuracy as it does not reflect the almost 900kg/ha of measurement error introduced by using DPM based estimates of herbaceous biomass. If the improvements in estimation accuracy of the magnitude achieved by Mutanga and Rugege (2006) could be replicated by re-running the cokriging with appropriate software, and the measurement error from the DPM reduced, estimation errors of much closer to 1000 kg/ha might be achievable. As I have not explored the relationship between measurement error, model error and actual estimation error I cannot provide an informed opinion on how much closer to 1000 kg/ha might be achievable.

After the completion of this study I encountered an interpolation method known as regression kriging. Although I did not have time to explore this method I encountered a reference on it in which it is described by the JRC as currently the best statistical method available for interpolating surfaces (http://eusoils.jrc.ec.europa.eu/esdb_archive/eusoils_docs/other/eur22904en.pdf). Based on this reference I would advise any researcher interested in carrying this work further in familiarizing themselves with this method and considering it as an alternative to the methods explored in this study.

6. REFERENCES

Al-Bakri, J.T., and J. C. Taylor. 2003. Application of NOAA AVHRR for monitoring vegetation conditions and biomass in Jordan. *Journal of Arid Environments* 54: 579 - 593.

Archibald, S., and R.J. Scholes. 2007. Leaf green-up in a semi-arid African savanna-separating tree and grass responses to environmental cues. *Journal of Vegetation Science* 18, no. 4: 583-594.

Boggs, G. S. 2010. Assessment of SPOT 5 and QuickBird remotely sensed imagery for mapping tree cover in savannas. *International Journal of Applied Earth Observation and Geoinformation*.

Fuller, D. O., S. D. Prince, and W. L. Astle. 1997. The influence of canopy strata on remotely sensed observations of savanna-woodlands. *International Journal of Remote Sensing* 18, no. 14: 2985-3009.

Govender, N., W. S. W. Trollope, and B. W. Van Wilgen. 2006. The effect of fire season, fire frequency, rainfall and management on fire intensity in savanna vegetation in South Africa. *Journal of Applied Ecology* 43, no. 4: 748-758.

Hansen, M. C., R. S. DeFries, J. R. G. Townshend, M. Carroll, C. Dimiceli, and R. A. Sohlberg. 2003. Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm. *Earth Interactions* 7, no. 1234567891011131214: 1-15.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment* 83, no. 1: 195-213.

Huete, A., C. Justice, and W. Van Leeuwen. 1999. MODIS vegetation index (MOD13) algorithm theoretical basis document. *Greenbelt: NASA Goddard Space Flight Centre, http://modarch. gsfc. nasa. gov/MODIS/LAND/# vegetation-indices*.

Moreau, S., R. Bosseno, X. Fa Gu, and F. Baret. 2003. Assessing the biomass dynamics of Andean bofedal and totora high-protein wetland grasses from NOAA/AVHRR. *Remote Sensing of Environment*, no. 85: 516-592.

Mutanga, O., and D. Rugege. 2006. Integrating remote sensing and spatial statistics to model herbaceous biomass distribution in a tropical savanna. *International Journal of Remote Sensing* 27, no. 16: 3499-3514.

Prince, S. D. 1991. Satellite remote sensing of primary production: comparison of results for Sahelian grasslands 1981-1988. *International Journal of Remote Sensing* 12, no. 6: 1301-1311.

Sannier, C. A. D., J. C. Taylor, and W. D. Plessis. 2002. Real-time monitoring of vegetation biomass with NOAA-AVHRR in Etosha National Park, Namibia, for fire risk assessment. *International Journal of Remote Sensing* 23, no. 1: 71-89.

Scanlon, T. M., J. D. Albertson, K. K. Caylor, and C. A. Williams. 2002. Determining land surface fractional cover from NDVI and rainfall time series for a savanna ecosystem. *Remote Sensing of Environment* 82, no.

2: 376-388.
Thompson, M.W., and C.S. Everson. 1993. Development of spectral-biomass models for mapping and monitoring montane grassland resources. Division of Forest Science and Technology CSIR.
Trollope, W, and A. L. F. Potgieter. 1986. Estimating grass fuel loads with a disc pasture meter in the Kruger National Park. *African Journal of Range and Forage Science* 3, no. 4.
Trollope, W. S. W. 2008. Fire regime of the Kruger National Park for the period 1980-1992. *Koedoe-African Protected Area Conservation and Science* 36, no. 2: 45.
Tucker, C. J., C.L. Vanpraet, M.J. Sharman, and G. Van Ittersum. 1985. Satelite Remote Sensing of Total Herbaceous Biomass Production in the Senegalese Sahel: 1980 - 1984. *Remote Sensing of Environment*, no. 17: 233-249.
Verbesselt, J., B. Somers, J. van Aardt, I. Jonckheere, and P. Coppin. 2006. Monitoring herbaceous biomass and water content with SPOT VEGETATION time-series to improve fire risk assessment in savanna ecosystems. *Remote Sensing of Environment* 101, no. 3 (April 15): 399-414. doi:doi: DOI: 10.1016/j.rse.2006.01.005.
Wessels, K. J., S. D. Prince, N. Zambatis, S. MacFadyen, P. E. Frost, and D. Van Zyl. 2006. Relationship between herbaceous biomass and 1km 2 Advanced Very High Resolution Radiometer (AVHRR) NDVI in Kruger National Park, South Africa. *International Journal of Remote Sensing* 27, no. 5: 951-973.

Appendix 1. Threshold classification accuracy

In this study the accuracies of the prediction methods have been presented in terms of kg/ha. However, during the course of the study I became aware that land managers planning prescribed burns are more likely to make decisions based on fuel load thresholds than on exact values as the latter are seldom available. Neither method explored in this study provided categorical outputs corresponding to these thresholds. It is however relatively simple assign each prediction to a class. This makes it possible to assess the methods in terms of their ability to correctly predict which threshold the herbaceous fuel load for an area falls within.

After the completion of the study a basic classification error matrix was created showing the accuracy with which the two methods predicted which of three classes the herbaceous biomass for an area fell within. The three classes were "< 2000kg/ha", "2000 – 4000 kg/ha" and ">4000 kg/ha". These classes were chosen to match those laid out in (Trollope 1996). Table 16 provides an overview of "classification accuracy" for both methods when their estimates are assigned to one of three classes. The first class, < 2000 kg/ha, corresponds to the range of herbaceous fuel loads at which fire will not spread in savannas. The second class, 2000 – 4000 kg/ha refers to the range which produce fires of cool to moderate intensity (< 3000 kj/s/m). The third class, >4000 kg/ha, refers to the range which produce fires of high intensity (>3000 kj/s/m). Tables 17- 22 provide a more detail on the classification and misclassification of pixels for all growth seasons assessed.

It is immediately apparent from Table 1 that cokriging achieves higher classification accuracy for the "<2000 kg/ha" and ">4000 kg/ha" classes than the regression models. The regression models on the other hand achieve higher classification accuracy than cokriging in the "2000 – 4000 kg/ha" class. Numerous incorrect decisions regarding prescribed burning would result from the use of maps produced with either of these levels of classification errors. Of greater importance is which class the pixels are most often incorrectly classified to. In other words, the important question is: does the method tend to over or underestimate herbaceous biomass within the given classes?

Over estimation in the “<2000 kg/ha” class has the fewest consequences. A fire team may arrive on site incorrectly classified through an overestimate as falling into the “2000 – 4000 kg/ha class” to initiate a burn and find that the fire does not spread. Time and resources have been wasted but the problem is self limiting, the fire will simply die out. Underestimates are far more serious. If a prescribed burn is set in an area incorrectly classified through an underestimate of fuel load as belonging to the “2000 – 4000 kg/ha” class but in reality 10000kg/ha of herbaceous biomass is present, the fire will be far more intense than expected. This could result in unwanted damage to the woody layer and possibly injury to the fire crew.

In all growth seasons except for 2005 - 2006 (tables 2 - 7) cokriging produces fewer instances of underestimates for the “>4000 kg/ha” class. The percentage of pixels misclassified due to underestimates in this class for the regression method ranged from 22 – 81% and 13 – 77% for cokriging. Within the “2000 – 4000 kg/ha class” cokriging once again produced fewer underestimates than regression. The percentage of pixels misclassified because of underestimates was on the whole lower than for the “>4000 kg/ha” class. For the regression method it ranged from 0 - 38% while for cokriging it ranged from 1 - 23%. Consequences arising from the misclassifications produced by cokriging are less serious because they arise primarily out of overestimating the amount of herbaceous biomass. This will lead fire teams to err on the side of caution rather than being surprised by a fire of greater intensity than they were expecting. It is up to fire mangers to determine whether the levels of classification accuracy reported here are sufficient for the maps produced to be useful decision aids.

After the completion of the study I encountered a promising hybrid approach to interpolation of surfaces known as regression kriging. Although I did not have time to explore this method further it is possible that through combining the use of regression models and cokriging the resulting outputs may lack the biases evident in the two methods assessed here. The paper entitled "A Practical Guide to Geostatistical Mapping of Environmental Variables" provides a good overview of regression kriging can be accessed at the following URL:
http://eusoils.jrc.ec.europa.eu/esdb_archive/eusoils_docs/other/eur22904en.pdf

Table 1: Classification accuracy of the regression and cokriging methods, the highest classification accuracy for each class in each season is highlighted in bold

		Classification accuracy		
Growth Season	**Method**	<2000 kg/ha	2000 - 4000 kg/ha	> 4000 kg/ha
2000 - 2001	Regression	0%	**67%**	64%
2001 - 2001	Cokriging	**21%**	56%	**87%**
2001 - 2002	Regression	13%	**90%**	28%
2002 - 2002	Cokriging	**23%**	76%	**57%**
2002 - 2003	Regression	61%	62%	0%
2003 - 2003	Cokriging	**69%**	**73%**	**9%**
2003 - 2004	Regression	3%	**78%**	51%
2004 - 2004	Cokriging	**22%**	72%	**68%**
2004 - 2005	Regression	42%	**76%**	24%
2005 - 2005	Cokriging	**52%**	73%	**45%**
2005 - 2006	Regression	13%	52%	**78%**
2006 - 2006	Cokriging	**48%**	**72%**	67%
Average	*Regression*	*22%*	***71%***	*41%*
Average	*Cokriging*	***37%***	*65%*	***48%***

Table 2: Herbaceous biomass classification error matrix for the 2000 – 2001 growth season

		Regression Model		2000 - 2001		Cokriging		2000 - 2001
		Predicted (kg/ha)				**Predicted (kg/ha)**		
		<2000	2000 - 4000	> 4000		<2000	2000 - 4000	> 4000
Measured (kg/ha)	<2000	**0%**	88%	12%		**21%**	63%	16%
	2000 - 4000	0%	**67%**	33%		1%	**56%**	44%
	> 4000	0%	36%	**64%**		0%	13%	**87%**

Table 3: Herbaceous biomass classification error matrix for the 2001 – 2002 growth season

		Regression Model		2001 - 2002		Cokriging		2001 - 2002
		Predicted (kg/ha)				**Predicted (kg/ha)**		
		<2000	2000 - 4000	> 4000		<2000	2000 - 4000	> 4000
Measured (kg/ha)	<2000	**13%**	83%	5%		**23%**	70%	7%
	2000 - 4000	4%	**90%**	6%		6%	**76%**	18%
	> 4000	0%	72%	**28%**		1%	42%	**57%**

Table 4: Herbaceous biomass classification error matrix for the 2002 – 2003 growth season

		Regression Model		2002 - 2003		Cokriging		2002 - 2003
		Predicted (kg/ha)				**Predicted (kg/ha)**		
		<2000	2000 - 4000	> 4000		<2000	2000 - 4000	> 4000
Measured (kg/ha)	<2000	**61%**	39%	0%		**69%**	31%	0%
	2000 - 4000	38%	**62%**	0%		23%	**73%**	4%
	> 4000	19%	81%	**0%**		14%	77%	**9%**

Table 5: Herbaceous biomass classification error matrix for the 2003 – 2004 growth season

		Regression Model		2003 - 2004		Cokriging		2003 - 2004
		Predicted (kg/ha)				**Predicted (kg/ha)**		
		<2000	2000 - 4000	> 4000		<2000	2000 - 4000	> 4000
Measured (kg/ha)	<2000	**3%**	93%	3%		**22%**	72%	6%
	2000 - 4000	1%	**78%**	21%		4%	**72%**	25%
	> 4000	0%	49%	**51%**		0%	32%	**68%**

Table 6: Herbaceous biomass classification error matrix for the 2004 – 2005 growth season

		Regression Model		2004 - 2005		Cokriging		2004 - 2005
		Predicted (kg/ha)				**Predicted (kg/ha)**		
		<2000	2000 - 4000	> 4000		<2000	2000 - 4000	> 4000
Measured (kg/ha)	<2000	**42%**	57%	1%		**52%**	45%	3%
	2000 - 4000	19%	**76%**	5%		18%	**73%**	9%
	> 4000	4%	72%	**24%**		2%	53%	**45%**

Table 7: Herbaceous biomass classification error matrix for the 2005 – 2006 growth season

		Regression Model		2005 - 2006		Cokriging		2005 - 2006
		Predicted (kg/ha)				**Predicted (kg/ha)**		
		<2000	2000 - 4000	> 4000		<2000	2000 - 4000	> 4000
Measured (kg/ha)	<2000	**13%**	81%	6%		**48%**	49%	3%
	2000 - 4000	2%	**52%**	46%		2%	**72%**	26%
	> 4000	0%	22%	**78%**		0%	33%	**67%**