



Estimating the fractional cover of growth forms and bare surface in savannas. A multi-resolution approach based on regression tree ensembles

Ursula Gessner^{a,b,*}, Miriam Machwitz^{a,c}, Christopher Conrad^a, Stefan Dech^{a,b}

^a Department of Remote Sensing, University of Wuerzburg, Am Hubland, 97074 Wuerzburg, Germany

^b German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Muenchner Strasse 20, 82234 Oberpfaffenhofen, Germany

^c Centre de Recherche Public – Gabriel Lippmann, Department of Environment and Agro-Biotechnologies (EVA), 41, rue du Brill, 4422 Belvaux, Luxembourg

ARTICLE INFO

Article history:

Received 13 February 2012

Received in revised form 25 September 2012

Accepted 12 October 2012

Available online xxxx

Keywords:

Sub-pixel fractional cover

Savanna

Grassland and shrubland biome

Africa

Vegetation structure

Land cover

Multi-resolution analysis

ABSTRACT

Evaluations of existing land cover maps have revealed that high landscape heterogeneity and small patch sizes are a major reason for misclassification. These problems globally occur in biomes of mixed vegetation structure and are particularly relevant for African savannas. This paper presents a multi-resolution approach to derive fractional cover of vegetation growth forms at sub-pixel level, aiming at an improved mapping of land cover in the African grassland, savanna and shrubland biome. Fractional cover is delineated for woody growth forms (trees and shrubs), herbaceous growth forms, and bare surface. The approach incorporates very high resolution (QuickBird/IKONOS, 0.6–1 m), high resolution (Landsat TM/ETM+, 30 m), and medium resolution data (MODIS, 250 m). While QuickBird/IKONOS data are classified into discrete classes, at Landsat and MODIS resolutions, sub-pixel cover is delineated using non-parametric ensemble regression trees from the random forest family. The propagation of errors in the hierarchical multi-resolution approach is assessed with Monte Carlo simulations.

The multi-resolution approach allows the adequate description of the heterogeneous vegetation structure in selected study regions of Southern Africa. The RMSE of the delineated fractional cover values range between 3.1% and 8.2% when compared with higher resolution data and between 4.4% and 9.9% when compared with field surveys. Errors at the Landsat resolution show minor influence on the accuracy of the MODIS results. Regarding the inter-resolution error propagation, for 90% of the Monte Carlo simulations, errors at the Landsat resolution resulted in RMSEs for MODIS increased by less than 4% (woody vegetation), 3.5% (herbaceous vegetation) and 2% (bare surface).

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Information on land cover and land cover change is fundamental in many fields of environmental research. Land cover datasets are an important basis for monitoring processes in the context of global change such as deforestation, degradation, and urban expansion. Moreover, they are required in fields such as biodiversity, water management, carbon storage, and ecosystem functions. The Global Terrestrial Observing System (GTOS) selected land cover as one of the Terrestrial Essential Climate Variables, i.e. as a terrestrial property which controls

physical, biological and chemical processes with relevance to climate (GTOS, 2008).

Accuracy assessments and comparisons between land cover maps have proven their general value. But likewise, studies have revealed that best accuracies are found for homogeneous land cover classes while heterogeneous classes with a mixed vegetation structure such as savannas are not as well represented. Small patch sizes and high heterogeneity were identified to be major reasons for misclassifications (Latifovic & Olthof, 2004; Smith et al., 2002, 2003). The validation of the global datasets GlobCover and MODIS Land Cover Type (MCD12Q1) revealed elevated confusions for mixed and spatially heterogeneous classes (Bicheron et al., 2008; Bontemps et al., 2011; Friedl et al., 2010). In a comparative study of two global land cover maps, Giri et al. (2005) found major differences in class assignment for heterogeneous classes. Likewise, Herold et al. (2008) identified a correlation between the homogeneity of a landscape and the agreement among four global land cover datasets. They identified low accuracies in all datasets particularly for classes with mixed vegetation such as ‘mixed trees’, ‘shrublands’ and ‘herbaceous vegetation’. Herold et al. (2008) see

* Corresponding author at: DLR-DFD, Muenchner Strasse 20, 82234 Oberpfaffenhofen, Germany. Tel.: +49 8153 28 1250; fax: +49 8153 28 1445.

E-mail addresses: ursula.gessner@dlr.de (U. Gessner),

miriam.machwitz@lippmann.lu (M. Machwitz),

christopher.conrad@uni-wuerzburg.de (C. Conrad), stefan.dech@dlr.de (S. Dech).

¹ Ursula Gessner did this work at the Department of Remote Sensing at the University of Wuerzburg. Her present affiliation and contact address is at DLR-DFD. Miriam Machwitz also did this work at the Department of Remote Sensing at the University of Wuerzburg. Her present affiliation and contact address is at Centre de Recherche Public – Gabriel Lippmann.

an outstanding challenge for improved global land cover derivation in enhancing the mapping of heterogeneous landscapes. For the African continent, these mixed classes are mainly found in the tropical and subtropical grassland, savanna and shrubland biome (biome definition according to Olson et al., 2001).

The mentioned inconsistencies may arise from different causes. Evidently, an important fact is that current land cover mapping activities rely on the distinction of discrete land cover classes. In biomes such as the tropical and subtropical savanna, grassland and shrubland biome, land cover is not discrete but rather characterized by a mixture of growth forms, primarily grasses, shrubs and trees. Small patches and gradual transitions between open and closed vegetation cover are typical features of these landscapes while clear boundaries between distinct land cover classes are rare.

A better representation of heterogeneous landscapes could be achieved by mapping the fractional cover of relevant land cover elements on a continuous scale, a concept that has been suggested by DeFries et al. (1995), Hansen et al. (2003), Heiskanen (2008) and Herold et al. (2008). Datasets following this concept of continuous categories have already been derived from medium resolution remote sensing data. The first maps of this kind focused exclusively on the derivation of tree cover. Examples are the global map of percentage tree cover, produced by DeFries et al. (2000) from AVHRR data of 1992/93 at a spatial resolution of 1 km, and a regional mapping approach for tropical woodlands and parklands in Zambia based on MODIS data (Hansen et al., 2002b, 2005). Another regional approach for forest biomes in southeast Asia was presented by Tottrup et al. (2007), considering even different forest types. A dataset that differentiates fractional cover of more than just one growth form is the MODIS Vegetation Continuous Fields product (VCF, MOD44B; Hansen et al., 2002a, 2003). Two editions of this product are currently available. The collection 3 dataset gives information on the sub-pixel proportion of trees, herbaceous vegetation and bare area for the year 2001 at a resolution of 500 m. The MODIS-VCF collection 5 contains proportional estimates only for tree cover and for the years 2000 to 2010 at a resolution of 250 m. The MODIS-VCF dataset has proven valuable especially in areas of high tree cover for mapping percent tree cover in Zambia (Hansen et al., 2005) and for depicting general tendencies of forest change in southeast Asia (Tottrup et al., 2007). However, regional assessments showed that pixel-level MODIS-VCF information has clear limitations for the monitoring of tree cover (White et al., 2005), particularly in landscapes of low tree density (Hansen et al., 2005; Heiskanen, 2008; Montesano et al., 2009). Hansen et al. (2005) state that the global land cover map that was used as a training dataset for MODIS-VCF exhibits particularly low accuracies for wooded grassland and woodland classes (i.e. savanna classes) which are due to high within-class heterogeneity. Thus, fractional vegetation cover mapping needs enhancement, especially in regions of mixed vegetation structure, where discrete land cover maps also exhibit inconsistencies (e.g. Bicheron et al., 2008; Friedl et al., 2010; Herold et al., 2008).

Apart from qualitative considerations, definitional problems arise when using tree cover or MODIS-VCF datasets in the African grassland, savanna and shrubland biome. These datasets account only for trees and partly (in MODIS-VCF, collection 3) delineate non-tree vegetation as a whole, i.e. herbaceous and shrub vegetation cover as one layer. However, in this biome, the differentiation between woody (trees and shrubs) and herbaceous growth forms seems more adequate as the ecosystem services (ESS) of shrubs and trees are generally similar, while grasses show rather different characteristics. Woody and herbaceous components exhibit for example clear differences in magnitude and seasonality of carbon sequestration, in provision of fodder and food, in the suitability of ecosystems as habitats, in their influence on the water cycle, water quality, and soil nutrients (e.g. Eldridge et al., 2011; Lloyd et al., 2008; Meik et al., 2002; Scholes & Archer, 1997).

With the aim to improve mapping of land cover in heterogeneous landscapes, this paper applies a multi-resolution approach for deriving fractional vegetation cover at regional scale. The procedure is tailored

to the African grassland, savanna and shrubland biome. Thus, fractional cover is derived for the three most prevalent – and with respect to ESS most meaningful – elements of land cover in this biome: woody growth forms (trees and shrubs), herbaceous growth forms and bare surface. The approach operates on three spatial resolutions incorporating very high resolution imagery (QuickBird/IKONOS, 0.6–1 m), multi-temporal high resolution data (Landsat TM/ETM+, 30 m) and medium resolution, annual time series (MODIS, 250 m). Sub-pixel proportions of growth forms are considered at Landsat and MODIS resolution. The fractional cover derivation is based on non-parametric ensemble regression trees from the random forest family. The approach is applied for two Southern African savanna ecosystems, the Kalahari Woodland Savanna and the Central and Eastern Namibian Savanna. The propagation of errors during the hierarchical multi-resolution approach is considered using Monte Carlo simulations.

2. Study regions

The first study region, the Kalahari Woodland Savanna, is located in northeastern Namibia (Fig. 1). The area is covered by Kalahari sands with longitudinal dunes and only minor variations in elevation (1050–1250 m above sea level). Typical for the region are deep, sandy soils with low nutrient contents (ferralic Arenosols) and associated petric Calcisols (FAO, 1998; Strohbach & Petersen, 2007). The mean annual precipitation ranges between 450 mm in the southwest and 750 mm in the northeast (Mendelsohn et al., 2002). In winter, temperatures can fall below 8 °C while summer temperatures rise up to 32–36 °C (Mendelsohn et al., 2002). The vegetation is characterized by a heterogeneous mixture of open woodland savanna (*Burkeo-Pterocarpetea*) and closed to open shrublands (*Acaciatea*; Strohbach & Petersen, 2007). The natural savanna vegetation is considerably influenced by human activities as the rural population strongly depends on natural resources for their livelihoods. In the Kalahari Woodland Savanna, grazing and cropping are typical land use activities, and large hardwood species are logged for timber. Frequent and mostly land-use-related savanna fires are typical and lead to a decrease in shrub density (Strohbach & Petersen, 2007).

The second study region, the Central and Eastern Namibian Savanna (Fig. 1), comprises the vegetation types ‘Camelthorn Savanna’ and ‘Thornbush Shrubland’ (Giess, 1998). The elevation above sea level varies between 1200 m and 1700 m. The most important soils in this region are ferralic Arenosols on Kalahari sands with associated petric Calcisols as well as chromic Cambisols and Regosols on sandstones and schists (Mendelsohn et al., 2002). Annual rainfall ranges from 250 mm in the southeast to 500 mm in the northeast. In winter, temperatures can drop to 0 °C while maximum summer temperatures range between 30 and 36 °C (Mendelsohn et al., 2002). The vegetation of the study area is dominated by fine-leaved and broad-leaved shrub- and grassland savannas. Fine-leaved *Acaciateae* can be found on relatively nutrient-rich soils while broad-leaved *Combretaceae* grow on poorer substrates (Rothauge, 2006). Towards the north of the region, *Terminalia-Combretum* associations are common (Strohbach et al., 2004). Land use in the Central and Eastern Namibian Savanna is mainly characterized by livestock farming on large tracts. Farmers suppress fires as far as possible and thus the savanna vegetation burns only exceptionally in this region.

Both study areas show a mixed composition of vegetation growth forms with gradual transitions between more open and more closed areas, which is a typical feature of semi-natural savanna ecosystems. Fig. 2 illustrates this characteristic savanna vegetation structure in the Kalahari Woodland Savanna. The scatter plot is based on a QuickBird classification and displays the sub-pixel proportions of growth forms within 30 m × 30 m grid cells. It is obvious that already at this spatial resolution, the savanna landscape is dominated by a heterogeneous mixture of herbaceous and woody vegetation with

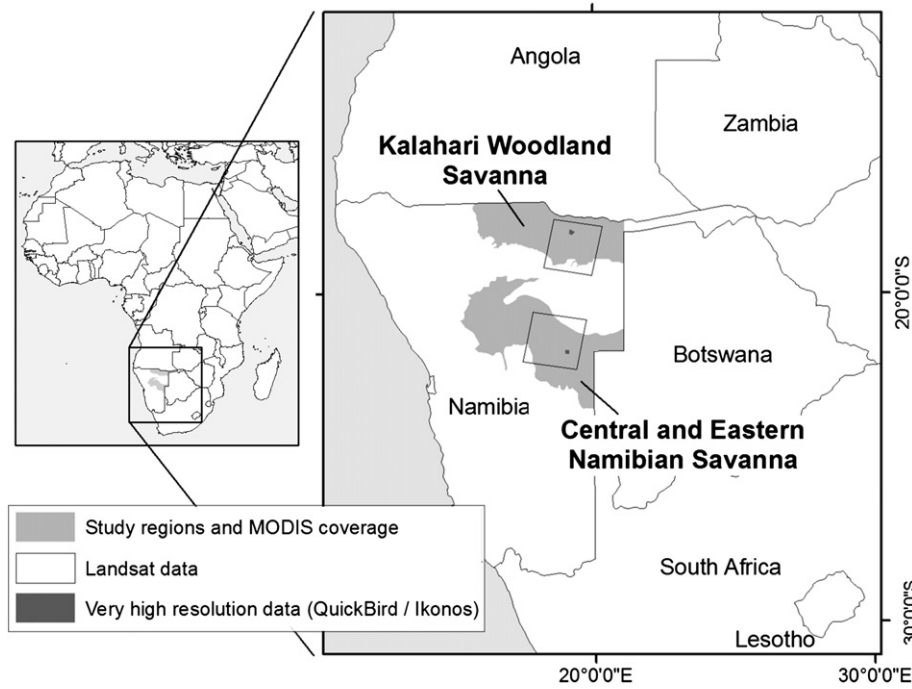


Fig. 1. Location of the study regions Kalahari Woodland Savanna and Central and Eastern Namibian Savanna and spatial coverage of applied satellite data.

gradual transitions. It is apparently impossible to subdivide the point cloud of the scatter plot by distinct thresholds into discrete land cover classes without a considerable loss of thematic content. Accordingly, at $30\text{ m} \times 30\text{ m}$ spatial resolution it is not possible to divide this landscape into discrete land cover classes.

3. Data

Data of one year was analyzed for each of the study regions. The years of analysis start in October and end in September in order to capture a complete vegetation phase (approximately October to April). According to the availability of very high resolution data, the years

2006/2007 and 2007/2008 were selected for the Kalahari Woodland Savanna and the Central and Eastern Namibian Savanna, respectively.

3.1. QuickBird and IKONOS data

One very high resolution image was available for each study region. A QuickBird dataset is located in the central part of the Kalahari Woodland Savanna (center coordinate: $18^{\circ}21'N/19^{\circ}16'E$, Fig. 1). It was acquired on April 15 2007 and covers an area of approximately 220 km^2 . The IKONOS image of the Central and Eastern Namibian Savanna (center coordinates: $22^{\circ}0'N/19^{\circ}9'E$, Fig. 1) with an area of approximately 120 km^2 was recorded on March 24 2008. For each image, the panchromatic band with a spatial resolution of 0.6 m (QuickBird) and 1 m (IKONOS) was fused with the 2.4 m (QuickBird) and 4 m (IKONOS) multispectral bands (blue, green, red and near-infrared). This resulted in four pan-sharpened bands of 0.6 m and 1 m spatial resolution for QuickBird and IKONOS respectively.

3.2. Landsat TM and ETM+ data

Landsat TM and ETM+ data were provided by the Landsat archive of the United States Geological Survey (USGS) and by the archive of the South African Council for Scientific and Industrial Research (CSIR). For the Kalahari Woodland Savanna, four Landsat TM acquisitions (path 177/row 73) of 2007 were available (Table 1, Fig. 1). For the Central and Eastern Namibian Savanna, one TM and two ETM+ datasets of 2008 were selected (path 177/row 75, Table 1, Fig. 1). The ETM+ images exhibit data gaps due to the failure of the Scan Line Corrector (SLC-off) on Landsat 7. The Landsat data provided by USGS showed high geometric accuracy when compared with GPS measurements from field campaigns and with IKONOS data and thus did not need any further geometric correction. The TM images provided by CSIR were co-registered to reference scenes of the USGS archive. For all TM acquisitions, the co-registration resulted in a root mean squared error (RMSE) of less than 15 m . Atmospheric correction was performed with the ATCOR-2 software implemented in IDL (Richter, 1996, 2007). Given the gaps in the ETM+ SLC-off acquisitions, all Landsat data selected for the Central and Eastern Namibian Savanna were spatially

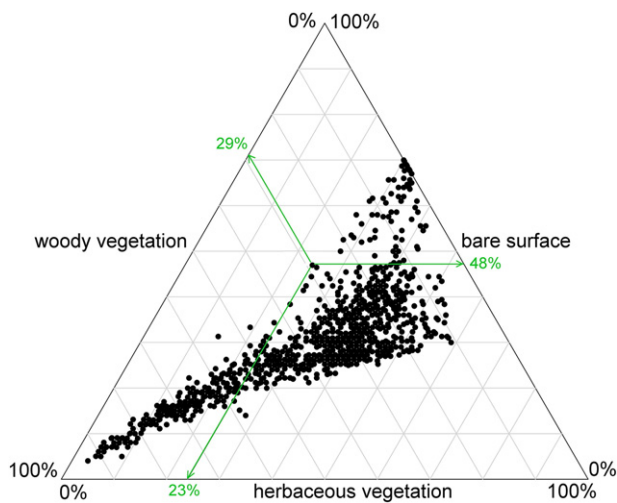


Fig. 2. Vegetation composition of the central Kalahari Woodland Savanna in northern Namibia with regard to fractional cover of herbaceous growth forms, woody growth forms and bare surface. The plot displays the two growth forms and bare fractions according to a classification of a QuickBird image (Fig. 1) summarized in grid cells of $30\text{ m} \times 30\text{ m}$. The green lines indicate how data values are to be read on the plot axes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

Landsat TM and ETM+ data analyzed at the intermediate resolution of the multi-resolution approach.

Study area	WRS-2 path/row	Date of acquisition	Sensor	Archive
Kalahari Woodland Savanna	177/73	15/02/2007	TM	CSIR-SAC
Kalahari Woodland Savanna	177/73	06/05/2007	TM	CSIR-SAC
Kalahari Woodland Savanna	177/73	23/06/2007	TM	CSIR-SAC
Kalahari Woodland Savanna	177/73	25/07/2007	TM	CSIR-SAC
Central and Eastern Namibian Savanna	177/75	09/01/2008	ETM+	USGS
Central and Eastern Namibian Savanna	177/75	22/04/2008	TM	CSIR-SAC
Central and Eastern Namibian Savanna	177/75	16/05/2008	ETM+	USGS

restricted to those areas where none of the three images had missing data. Consequently, the fractional cover results at Landsat resolution show data gaps for this study region.

Based on the Landsat datasets, a feature space was created that consists of reflectance values, multispectral and multitemporal indices. In addition to the reflectance values of the Landsat bands ranging from visible to short-wave infrared wavelengths (bands 1 to 5 and band 7), the Normalized Difference Vegetation Index (NDVI), Simple Ratio (SR, $\rho_{\text{NIR}}/\rho_{\text{RED}}$), Soil Adjusted Vegetation Index (SAVI; Huete, 1988) and the Tasseled Cap Indices brightness, greenness and wetness (Crist et al., 1986; Jensen, 1996) were derived from all TM and ETM+ reflectances in the relevant bands. Furthermore, multitemporal indices were created based on these monotemporal features. This was done by calculating the temporal mean, median, minimum, maximum and amplitude for each of the above mentioned indices and reflectances. An example of such a multitemporal index is the median NDVI of the four Landsat TM datasets of the Kalahari Woodland Savanna (Table 1).

3.3. Annual MODIS time series

The MODIS dataset comprised 16-day composites of Enhanced Vegetation Index (EVI), NDVI and the MODIS band 1 (red), band 2 (near-infrared, NIR), band 3 (blue) and band 7 (mid-infrared, MIR) reflectances. All datasets were extracted from the 250 m MODIS standard product MOD13Q1, collection 5, distributed by the NASA Land Processes Distributed Active Archive Center (LP DAAC). The 16-day composites were assembled as time series for the years 2006/2007 (Kalahari Woodland Savanna) and 2007/2008 (Central and Eastern Namibian Savanna). Each time series starts in October and ends in September, according to the vegetation seasonality in Southern African savannas. To avoid measurements of low quality, the MODIS Quality Assessment Science Datasets (QA-SDS) were inspected and all composites with the layer 'VI usefulness' being lower than the fourth best quality level were rejected. The data gaps were filled by linear temporal interpolation. The analysis of QA-SDS and the subsequent interpolation was performed using the tool TiSeG (Colditz et al., 2008).

From the annual MODIS time series, 90 metrics were generated as input features at the coarsest level of the multi-resolution approach. The six time series (i.e. EVI, NDVI, red, NIR, blue, MIR) were subdivided into dry season and vegetation period. The duration of vegetation period and dry season was determined for each pixel individually in order to account for spatial variability in the onset of rainfall. They were defined by calculating the inflection points of the EVI time series which had been smoothed beforehand by discrete Fourier transformation (Jakubauskas et al., 2001). From each of the reflectances and indices we calculated mean, median, minimum, maximum and amplitude of the dry season, of the rainy season, and of the complete year.

3.4. Acquisition dates and temporal data coverage

Fig. 3 shows the acquisition dates of the employed very high and high resolution data embedded in the phenological cycle of one year (MODIS-NDVI for October to September). At the highest spatial resolution, data from a single acquisition date was analyzed. The QuickBird and IKONOS imagery were scheduled such that they approximately capture the peak of annual vegetation development. In this way they can provide a high quality reference dataset in accordance with the UN Land Cover Classification System (LCCS) requirement to assess land cover in terms of the fullest seasonal plant development (Di Gregorio, 2005). At the intermediate resolution, the three and four Landsat acquisitions represent snapshots of different phenological phases like the approximate state of maximum vegetation development but also the green-up phase, senescence and dry season, as far as data were available. In terms of MODIS, continuous annual time series describe the full vegetation cycle of one year in time-steps of 16 days.

4. Methods

4.1. Multi-resolution approach for deriving the fractional cover of vegetation growth forms

The procedure for deriving the fractional cover of vegetation growth forms and bare surface is presented in Fig. 4. The workflow is repeated for each cover component separately. Very high resolution information on vegetation cover is delineated by classification of QuickBird or IKONOS imagery. This information is spatially aggregated to the 30 m resolution of Landsat and used as learning and validation data for this intermediate level. Random forest regression trees are then employed on Landsat TM and ETM+ data for delineating sub-pixel fractional cover of woody vegetation, herbaceous vegetation and bare surface respectively. The resulting 30 m continuous cover values are aggregated to the 250 m resolution of MODIS. Finally, at the coarsest resolution, continuous cover values are delineated from MODIS data, again using random forest regression trees. This time, the learning samples for the regression trees are a subset of the degraded 250 m results of the Landsat level. The following sections describe the individual steps of the multi-resolution approach and its validation in detail.

4.2. Classification of QuickBird and IKONOS imagery

QuickBird and IKONOS data were classified into discrete classes using a supervised object- and pixel-based maximum likelihood approach. The classification was based on training areas extracted from 15 to 20 rectangular subsets of the very high resolution data. These subsets, with a size of 600 m × 600 m, were equally spaced over the images on a 2.5 km grid. The subsets were segmented using an optimized segmentation approach after Esch et al. (2008). From the resulting segments, training areas for four classes were selected for the Kalahari Woodland Savanna. In the Central and Eastern Namibian Savanna – due to higher variations in the density of grassland patches – six classes were distinguished. For each class, the surface cover composition was specified (Table 2) with the help of field data that had been collected close to the acquisition dates of the imagery (± 1 week). These field data consisted of 60 m × 60 m plots, where woody vegetation cover, herbaceous vegetation cover and bare surface were estimated visually on the ground. For trees, shrubs, and grasses, the average height was additionally recorded. Furthermore, from the center of the field plots, pictures were taken in a northward, eastward, southward and westward direction.

Even at a spatial resolution of 0.6 m to 1 m, some classes consist of a mixture of bare surface, herbaceous growth forms and woody growth forms (e.g. dwarf shrubs). For the class 'cast shadow' that is caused by trees, an assignment of growth form and bare surface

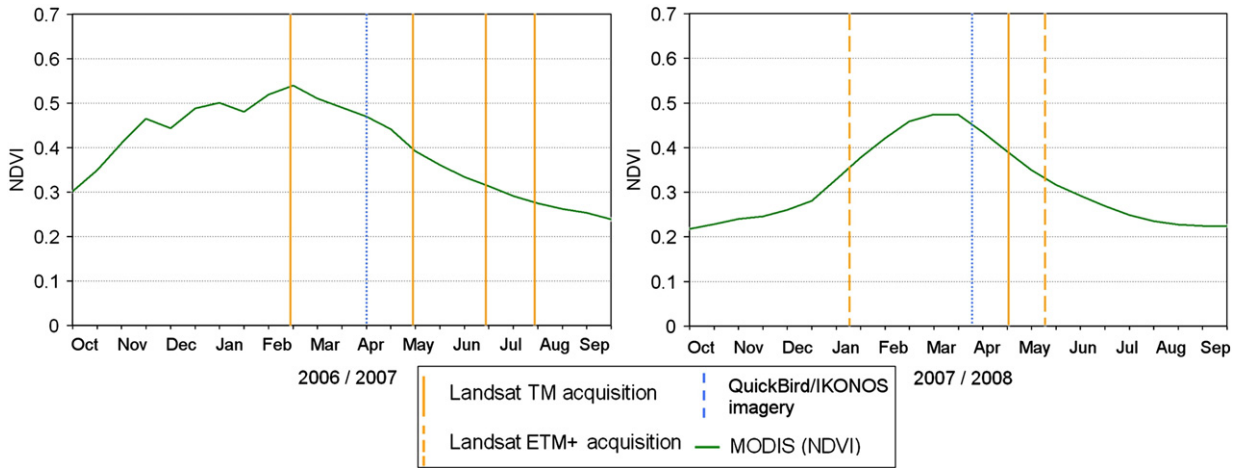


Fig. 3. Acquisition dates and temporal coverage of the remote sensing datasets for the Kalahari Woodland Savanna (left) and the Central and Eastern Namibian Savanna (right).

fractions is not straightforward. It was assumed that the growth forms in the study regions are, by approximation, equally distributed in space. Thus, for estimating the growth form composition of surfaces obscured by cast shadows of trees, the overall proportion of woody vegetation, herbaceous vegetation and bare surface in each image was identified. Cast shadows were then assumed to approximately obscure surfaces with a growth form composition equal to these overall proportions.

4.3. Spatial aggregation and sampling for training and validation

The IKONOS and QuickBird classifications were aggregated to the Landsat TM pixel size of 30 m by averaging the cover values over the coarser grid cells. The results were layers of continuous sub-pixel percentages of woody cover, herbaceous cover and bare surface. These three layers served as a data set for the learning procedure of the regression trees and for their validation at Landsat resolution. From each layer, approximately 4000 learning and 2000 validation samples (no overlap) were extracted. The samples were equally distributed over the extent of

the image. The share of samples with a certain cover value (0–100%) in the number of all extracted samples was defined according to the percentage to which this cover value occurred in the reference data set (stratified weighted sampling). Later in the multi-resolution approach, this scaling and sampling procedure was repeated for the 30 m fractional cover results of the Landsat resolution which again resulted in approximately 4000 learning and 2000 validation samples for the 250 m MODIS resolution.

4.4. Random forest regression tree ensembles

At the Landsat and MODIS resolutions, sub-pixel cover of woody vegetation, herbaceous vegetation and bare surface were delineated using non-parametric random forest regression tree ensembles, which belong to the Classification and Regression Tree methods (CART; Breiman et al., 1984; Breiman, 2001). Studies have shown that single regression trees deliver relatively low accuracies (e.g. Garzon et al., 2006; Liu & Wu, 2005) but combining several trees to regression tree ensembles can significantly improve the results (e.g. Breiman, 1996,

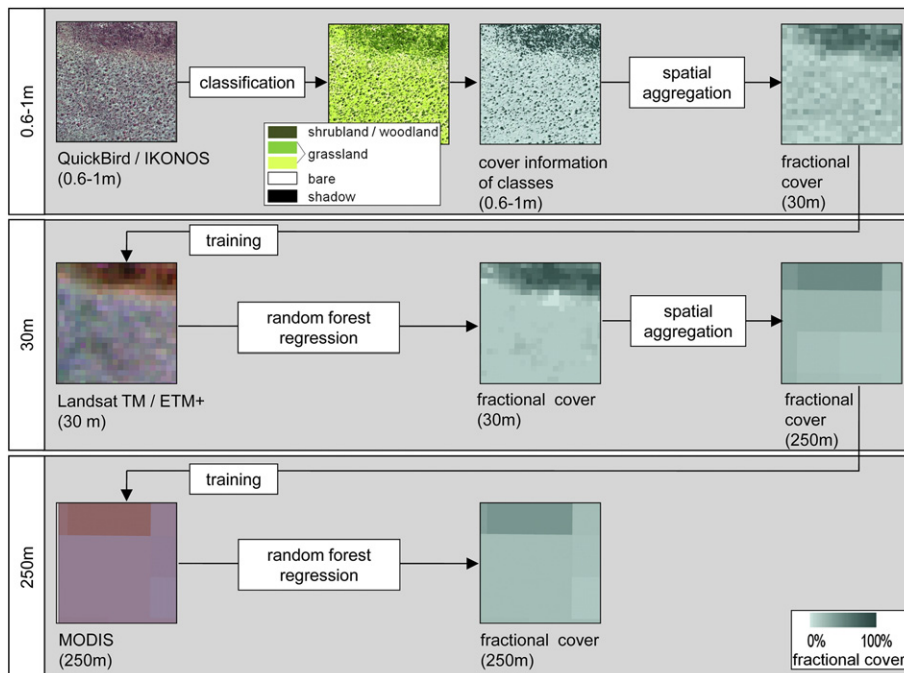


Fig. 4. Schematic workflow of the multi-resolution approach for the derivation of fractional cover of woody vegetation, herbaceous vegetation and bare surface.

Table 2

Classes of the QuickBird- and IKONOS-based classification and the class-specific growth form composition.

Class	Woody cover [%]	Herbaceous cover [%]	Bare [%]
<i>QuickBird image of Kalahari Woodland Savanna</i>			
Shrubland/woodland	100	0	0
Grassland	10	60	30
Bare	0	30	70
Cast shadow	40	30	30
<i>IKONOS image of Central and Eastern Namibian Savanna</i>			
Shrubland/woodland	90	5	5
Open grassland	0	30	70
Medium dense grassland	5	55	40
Closed grassland	0	100	0
Bare	0	0	100
Cast shadow	70	15	15

2001; Garzon et al., 2006; Xu et al., 2005). Therefore, in this study, ensemble random forest regression trees were generated using bagging techniques (Breiman, 2001). First, the individual regression trees were constructed. During this construction process, binary recursive partitioning is applied with the aim to estimate a dependent variable (here: the fractional cover) by means of multiple independent variables (here: the remote sensing metrics, see Sections 3.2 and 3.3). The regression trees are constructed on the basis of a set of so called learning samples which are comparable to the training data of traditional supervised classification approaches. Starting at the root of a regression tree (root node) and in a hierarchical sequence of binary splits, the dataset is divided into subspaces (nodes) which are as homogeneous as possible. As splitting criterion, a threshold value s is chosen such that Eq. (1) is minimized.

$$\sum_{y_i | X_{i,j} \leq s} (y_i - \bar{y})^2 + \sum_{y_i | X_{i,j} > s} (y_i - \bar{y})^2 \tag{1}$$

Here, Y denotes the dependent and X_j the independent variables. \bar{y} represents the mean of all elements y_i , and the threshold value s can take any value of the independent variables X_j . The final nodes of a tree are called leaves. The mean value of the dependent variable of all leaf elements is assigned to each of these leaves. As the constructed trees are partly fitted to noise in the learning samples, they show reduced generalization capability to make correct predictions for data that have not been included in the learning process. To avoid this overfitting, pruning methods can be applied which cut off less important branches. Here, the relevance of branches and splits

is identified by means of training data, that have not been included in the tree growing process (pruning data). After a number of test runs with varying types of pruning, we decided to perform pruning based on a cost complexity sequence of sub-trees (Breiman et al., 1984) where the sub-tree with the 50 best predicting leaves was selected.

With the aim to generate ensembles of several single regression trees, the random forest technique of bagging (Breiman, 1996) was applied. Test runs showed that the accuracy increased with increasing ensemble size but stabilized as soon as 12–15 trees were included. Thus, we constructed 15 regression trees on subsets of the learning samples by randomly drawing 50% of the samples each time. The results of the 15 single trees were combined by weighted averaging. The weight for each leaf was defined according to the leaf-specific RMSE calculated from the learning samples.

A measure that indicates the importance of each of the independent variables for the fractional cover estimation is the increase in mean squared error (%IncMSE). This measure is calculated by constructing each tree of an ensemble with and without the specific variable. For all trees, the differences in error of these two variants are recorded, averaged and normalized by their standard deviation. For all regression tree analyses, the software R (R Development Core Team, 2011) with the packages tree (Ripley, 2012) and randomForest (Liaw & Wiener, 2002) were used.

4.5. Assessment of accuracy and error propagation

Both at the Landsat and MODIS resolutions and for each cover type, validation sample sets of approximately 2000 pixels (Section 4.3) were analyzed. The results were compared to the validation samples of QuickBird/IKONOS (degraded to 30 m) and Landsat (degraded to 250 m) respectively, and RMSEs were calculated.

In addition, the Landsat-based results of the Central and Eastern Namibian Savanna were validated with field data collected in March and April 2008. Cover fractions were surveyed on the ground with the point intercept method (Bonham, 1989) on 14 plots of 60 m × 60 m (Fig. 5, left). Depending on the heterogeneity of the landscape, 3–5 transects with a length of 60 m were placed within each of these plots. Along the transects, the dominant cover type (woody vegetation, herbaceous vegetation and bare soil) was recorded in intervals of 1 m. At each interval, a circular area with a diameter of 10 cm was assessed (Fig. 5, right). For comparison with Landsat data, average values of woody vegetation cover, herbaceous vegetation cover and bare surface cover were calculated for each field plot from the 180 to 300 individual

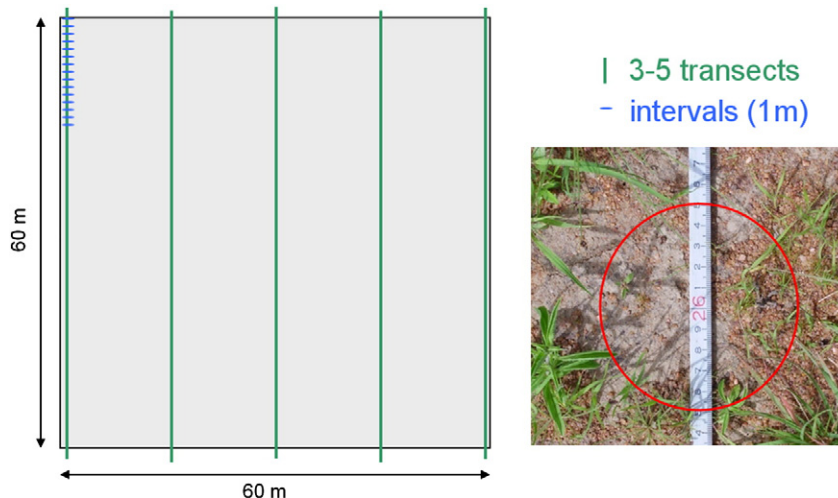


Fig. 5. Layout of a point intercept field plot for fractional vegetation cover surveys.

Table 3
Root mean squared errors (RMSE) of fractional cover datasets delineated from Landsat and MODIS when compared to the validation datasets based on QuickBird/IKONOS and Landsat data, respectively and RMSE delineated from field surveys.

Study region	Sensor		RMSE higher resolution	RMSE field surveys
Kalahari Woodland Savanna	Landsat TM (path 177, row 73)	Woody vegetation	8.2	
		Herbaceous vegetation	5.3	
		Bare surface	4.1	
	MODIS	Woody vegetation	8.1	
		Herbaceous vegetation	4.6	
		Bare surface	3.6	
Central and Eastern Namibian Savanna	Landsat TM and ETM+ (path 177, row 75)	Woody vegetation	5.9	6.5
		Herbaceous vegetation	5.2	9.9
		Bare surface	6.0	4.4
	MODIS	Woody vegetation	3.1	
		Herbaceous vegetation	3.1	
		Bare surface	3.2	

records. These average values were compared with the mean fractional cover values of the four nearest Landsat pixels.

As a consequence of the hierarchical, multi-resolution structure of the fractional cover delineation, the learning samples at MODIS resolution are contaminated by an error which has been committed at Landsat resolution. Therefore, inter-resolution error propagation was evaluated for the Kalahari Woodland Savanna using Monte Carlo simulations (e.g. Heuvelink, 1998). The basic idea of this approach is to repeatedly construct regression tree ensembles at MODIS resolution after having contaminated the fractional cover values of the learning samples with an estimate of their error. For each Monte Carlo simulation, this error was randomly sampled from the estimated error distribution of the Landsat-based learning samples at 250 m resolution. To estimate this error distribution, the validation samples of the QuickBird level were aggregated to 250 m and subtracted from the regression tree results at Landsat resolution, also aggregated to 250 m. The resulting differences were tested for normality and the corresponding normal distribution was used as an estimate for the distribution of errors of the Landsat-derived learning samples. Based on this error distribution, 1000

Monte Carlo simulations were calculated for each cover type. The results of the Monte Carlo simulations indicate the typical variation of accuracy associated with the error of the input data.

5. Results and discussion

5.1. Accuracy and error propagation

The validation results for woody vegetation, herbaceous vegetation and bare surface with the higher resolution data showed RMSE values ranging between 3.1% and 8.2%. Table 3 gives an overview of the RMSE for the datasets at 30 m and 250 m spatial resolutions and for the three cover types. The respective scatter plots of the validation samples versus the multi-resolution results are shown in Figs. 6 and 7. The validation with field plots resulted in similar RMSE values ranging from 4.4% to 9.9% (Table 3).

Field validation showed highest errors for herbaceous vegetation with an overestimation of herbaceous cover in the Landsat product. A reason for this overestimation could be a misinterpretation of dwarf

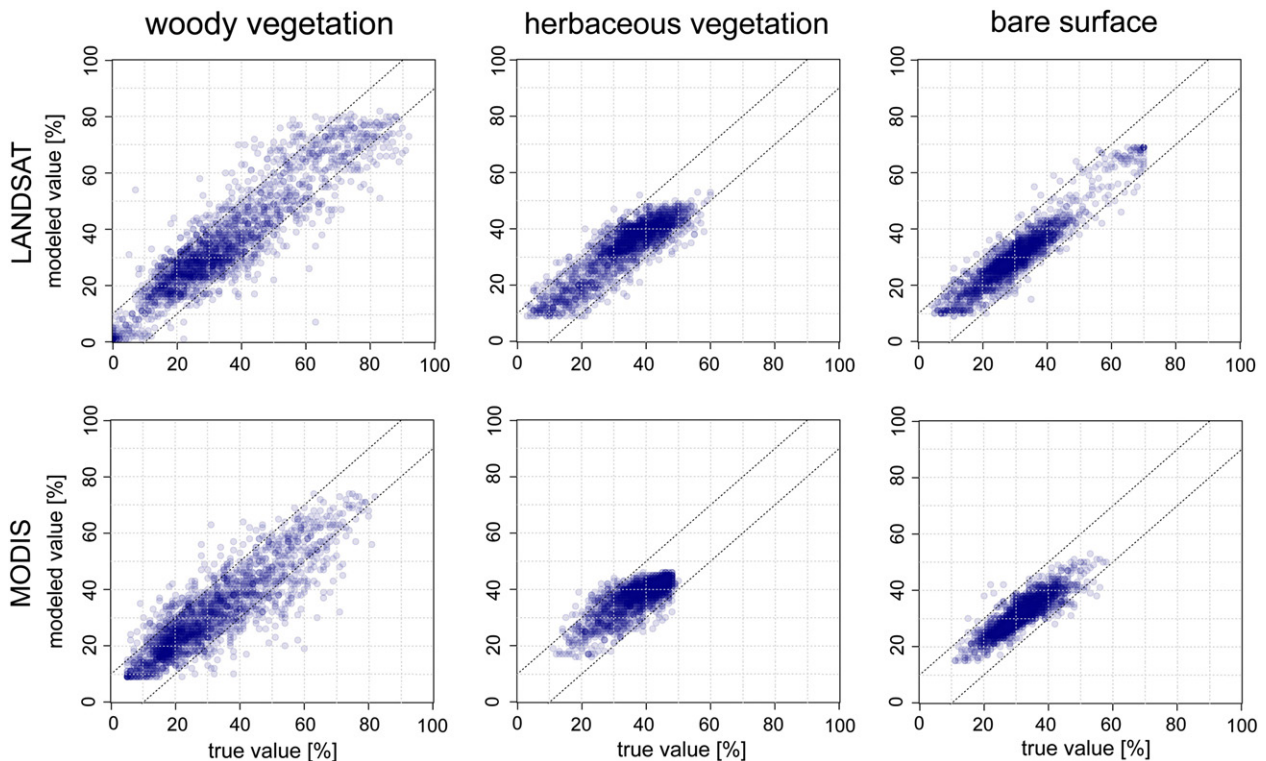


Fig. 6. Scatterplots of regression tree results and validation data for the Kalahari Woodland Savanna. The diagonal lines indicate a $\pm 10\%$ deviation.

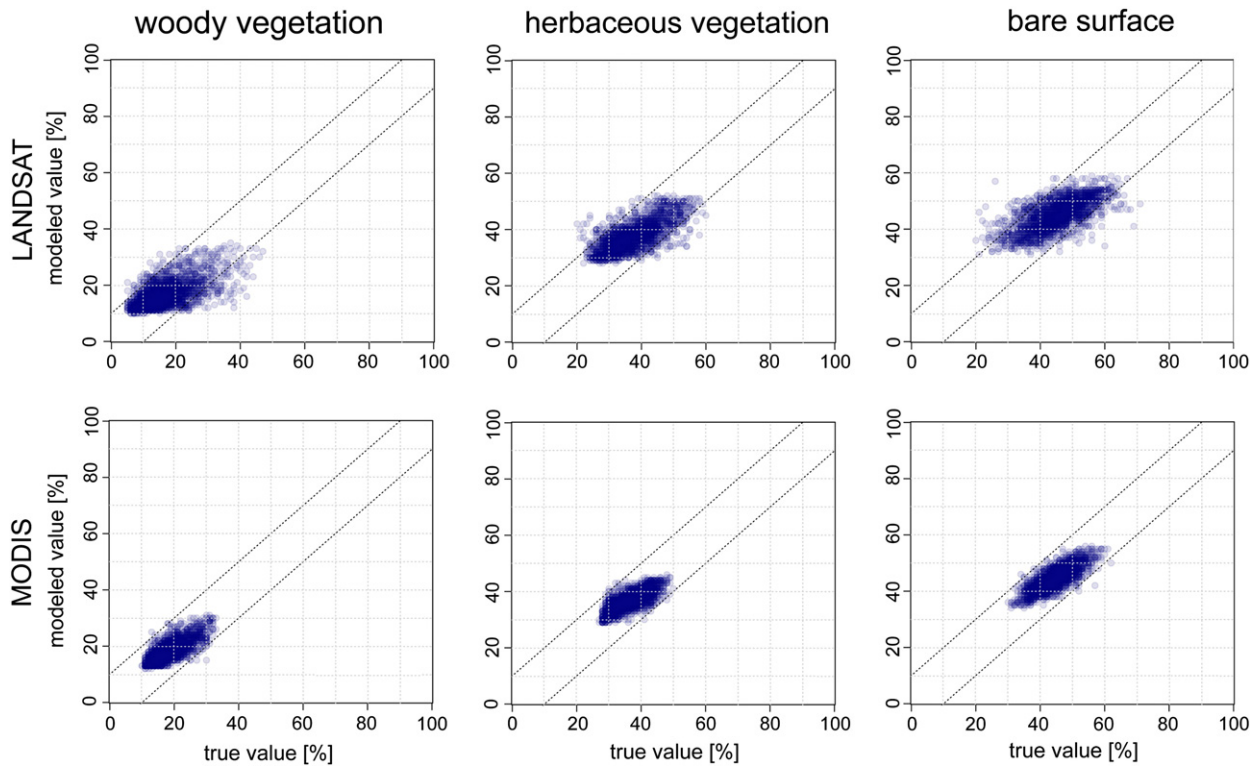


Fig. 7. Scatterplots of regression tree results and validation data for the Central and Eastern Namibian Savanna. The diagonal lines indicate a $\pm 10\%$ deviation.

shrubs as grasses in the IKONOS classification and consequently also in the Landsat data.

In the Kalahari Woodland Savanna, best accuracies were achieved for bare surface (RMSE: 4.1% and 3.6%). In contrast, woody vegetation turned out to be the most difficult cover type in this study area (RMSE: 8.2% and 8.1%). When comparing the spectral characteristics of woody vegetation and soil in the Kalahari Woodland Savanna, soils show a higher homogeneity, due to the rather uniform Kalahari sands. The woody vegetation component in contrast is composed of a high variety of shrub and tree species even including a smaller proportion of evergreen trees (Strohbach & Petersen, 2007). As a consequence, slight variations exist in the spectral characteristics of leaves and canopies of different species that are related to phenology, chemical composition, leaf structure and leaf angle inclination. These variations affect the Bidirectional Reflectance Distribution Function BRDF and shadow fractions in the canopy that might complicate the delineation of woody cover in this region. A different situation exists in the Central and Eastern Namibian Savanna where, when compared to higher resolution data, woody vegetation and bare surface are delineated with similar accuracies (Table 3). Here, woody species are exclusively deciduous. The geology however is relatively heterogeneous and includes sands, sandstones and schists, resulting in a higher variety of the soil’s spectral characteristics. Thus for both bare surface and woody cover, the results in both study regions suggest a relationship between the homogeneity of the spectral and phenological characteristics of the target cover type and the accuracy of delineated cover values.

In both study regions and for all cover types, RMSEs for MODIS are lower than those for Landsat based results. Partly, this difference in accuracy may be attributed to resolution effects, which typically have an influence on the accuracy of remote sensing products. In many cases, accuracy improves with increasing spatial resolution. However, there is not a general rule in terms of direction and magnitude of resolution effects, and they seem to depend on the spatial, temporal and spectral characteristics of the regarded land cover and landscape (e.g. Hansen et al., 2002b; Irons et al., 1985; Marceau et al., 1994;

Moody & Woodcock, 1995; Yanchen, 2008). Fig. 8 can help to better interpret the influence of resolution on the error of the presented results at Landsat and MODIS levels. The plot shows the mean absolute errors of the Landsat-based results – compared to the QuickBird reference – when Landsat results are continuously degraded to coarser grid cells of up to 1 km. For all three cover types, errors decline with increasing grid cell size. The decline is markedly steeper between 30 m and 250 m than between 250 m and 1 km, which is similar to the scaling behavior found by Hansen et al. (2002b) for woodlands in Zambia.

The inter-resolution propagation of errors from Landsat to MODIS level is not accounted for in Table 3. This error propagation was assessed with Monte Carlo simulations for the Kalahari Woodland Savanna.

The error of the Landsat-based learning samples was identified by comparing the aggregated 250 m Landsat results with the aggregated 250 m QuickBird classification. Errors were found to be by approximation normally distributed with mean errors of 5.6%, 3.4% and 2.7% for woody, herbaceous and bare cover respectively. Standard deviations were 5.5%, 3.3% and 2.7% respectively. The frequency distributions of the RMSE of 1000 Monte Carlo simulations for woody cover, herbaceous

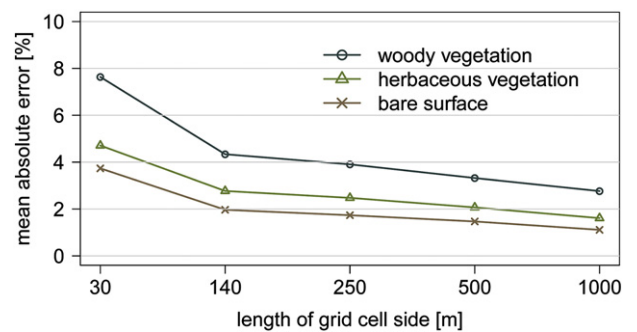


Fig. 8. Mean absolute error of fractional cover derived from Landsat data in the Kalahari Woodland Savanna. The errors are based on a comparison between the QuickBird classification and the Landsat-derived results at grid cell sizes between 30 m \times 30 m and 1 km \times 1 km.

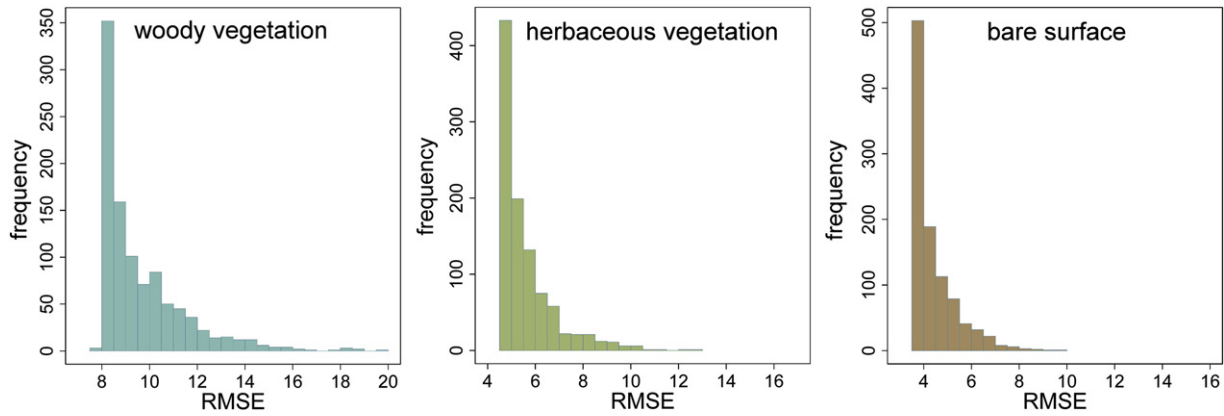


Fig. 9. Frequency of RMSE resulting from 1000 Monte Carlo simulations for woody growth forms, herbaceous growth forms and bare surface.

cover and bare surface (Fig. 9) indicate that extreme RMSEs of up to 20% (Fig. 9, woody vegetation) can occur occasionally with woody vegetation, but for all cover types, the frequency of RMSE decreases nearly exponentially with increasing error. 90% of the Monte Carlo simulations resulted in an RMSE of less than approximately 12% for woody vegetation, 8% for herbaceous vegetation and 5.5% for bare surface. Thus, the pure effect of inter-resolution error propagation was – for 90% of the simulations – a decrease in RMSE of less than 4% (woody vegetation), 3.5% (herbaceous vegetation) and 2% (bare surface).

In a woodland to savanna region in Zambia, Hansen et al. (2002b) derived tree cover in a similar approach to the one presented here. Major differences between the procedure of Hansen et al. (2002b) and the approach of this study are the use of single regression trees versus random forest regression ensembles and the derivation of only tree cover versus the derivation of woody vegetation cover, herbaceous vegetation cover and bare surface fractions. Hansen et al. (2002b) created a 250 m MODIS tree cover map with an overall RMSE of 6.76%, but as they examined validation samples which had also been used for training, an independently calculated RMSE

would be higher. For the same region in Zambia, Hansen et al. (2005) derived tree cover using the MODIS-VCF algorithm and single date images, composites and metrics of 500 m MODIS data. They achieved standard errors of 7.1% to 7.8% when using the training samples for validation, and a standard error of 11.5% when excluding the training data from validation. They noted problems in particular for the Kalahari woodlands at the low end of tree cover and assigned these difficulties to poor training data and high soil reflectances of the bright Kalahari sands. A quantitative comparison of our results with the MODIS-VCF product would also be interesting but is limited for two reasons. First, the MODIS-VCF layer bare cover and non-tree vegetation cover are only available for the year 2001. As in semi-arid areas, these cover types are highly variable at interannual time-scales, a comparison with our results of 2006/07 and 2007/08 is not meaningful. Second, concerning the temporally more stable woody growth forms, the MODIS-VCF layer tree cover (available for 2000–2010) is not quantitatively comparable with the layer woody cover, as the latter includes not only trees but also shrubs.

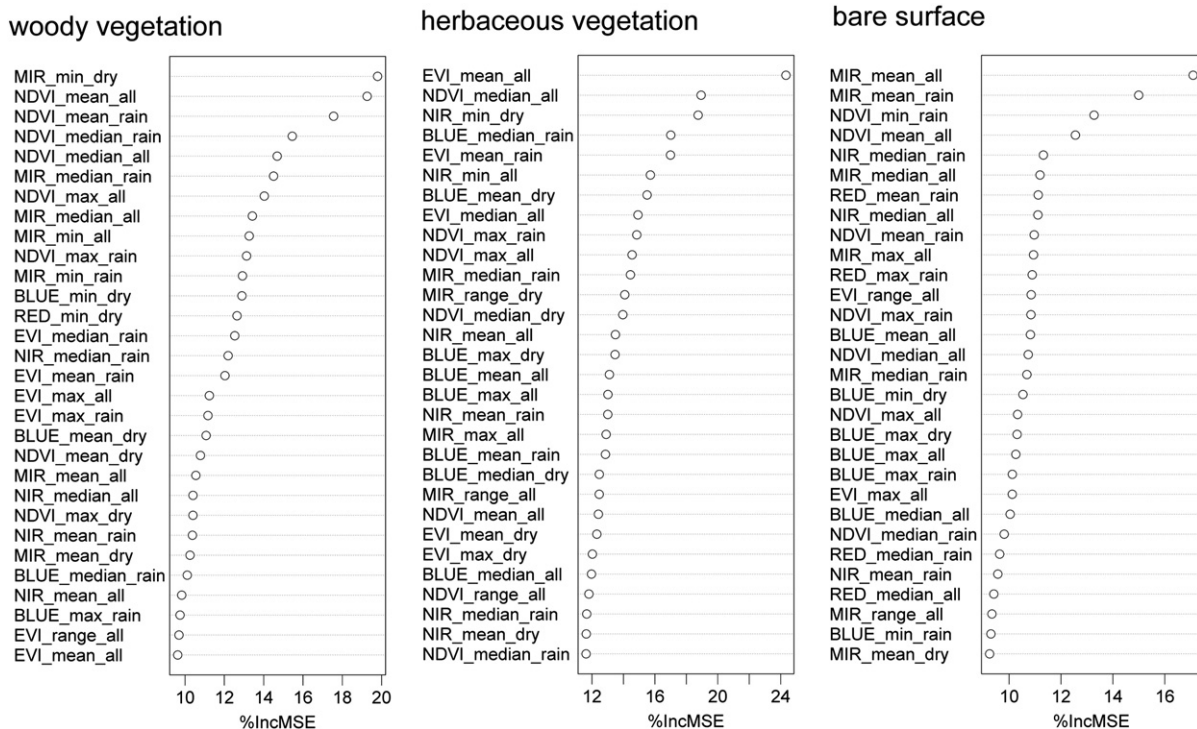


Fig. 10. Importance of MODIS variables for the delineation of woody vegetation cover, herbaceous vegetation cover and bare surface in the Kalahari Woodland Savanna.

5.2. Variable importance

The relevance of the MODIS variables for the delineation of fractional cover in the Kalahari Woodland Savanna is illustrated in Fig. 10. The plots list the 30 most important variables for the estimation of woody vegetation cover, herbaceous vegetation cover and bare surface. For the identification of woody vegetation, variables delineated from MIR and NDVI time series are most important, while metrics of EVI, NIR, NDVI and blue are of highest relevance for herbaceous vegetation. For both vegetation growth forms, metrics of the rainy season, of the dry season and of the full year are found among the 30 most important variables. For the delineation of bare surface, highest relevance was identified for variables based on MIR and NDVI, but dry season metrics are of minor importance.

5.3. Mapping results

Fig. 11 shows the fractional cover of woody growth forms, herba- ceous growth forms and bare surface in the Kalahari Woodland Savanna as delineated from Landsat data. The maps reproduce the vegetation structure of the region as a heterogeneous mixture of open woodland savanna and closed to open shrublands with hardly any abrupt borders between distinct land cover classes. The extensive patches of

comparatively low woody vegetation cover and high bare fractions southeast of the road B8 are mainly due to the deeper Kalahari sands and to recurring savanna fires in this area. In the western third of the Landsat scene, the typical vegetation structure of longitudinal Kalahari dune systems is clearly depicted by the maps of woody and herbaceous vegetation cover. On dune crests we mainly find open woodlands with higher herbaceous proportions while in interdunal depressions, woody vegetation cover is typically higher and grass cover is sparser (Strohbach & Petersen, 2007).

Also at MODIS resolution, the heterogeneous vegetation structure and gradual transitions of the Kalahari Woodland Savanna are well depicted (Fig. 12). In the vicinity to the urban areas of Rundu, the extent of cleared savanna is observable. The areas of agricultural land use along the Okavango River and in the west of the study region are characterized by low woody, high bare and high herbaceous fractions. In the central and western study region, the longitudinal dunes with their distinct alterations of vegetation structure are delineated also at MODIS resolution.

The delineated fractional cover datasets capture and characterize different typical land cover elements of the savanna landscapes at the three regarded spatial resolutions. Fig. 13 focuses on a subregion in the center of the Kalahari Woodland Savanna and compares the cover of woody vegetation delineated from QuickBird (left), Landsat

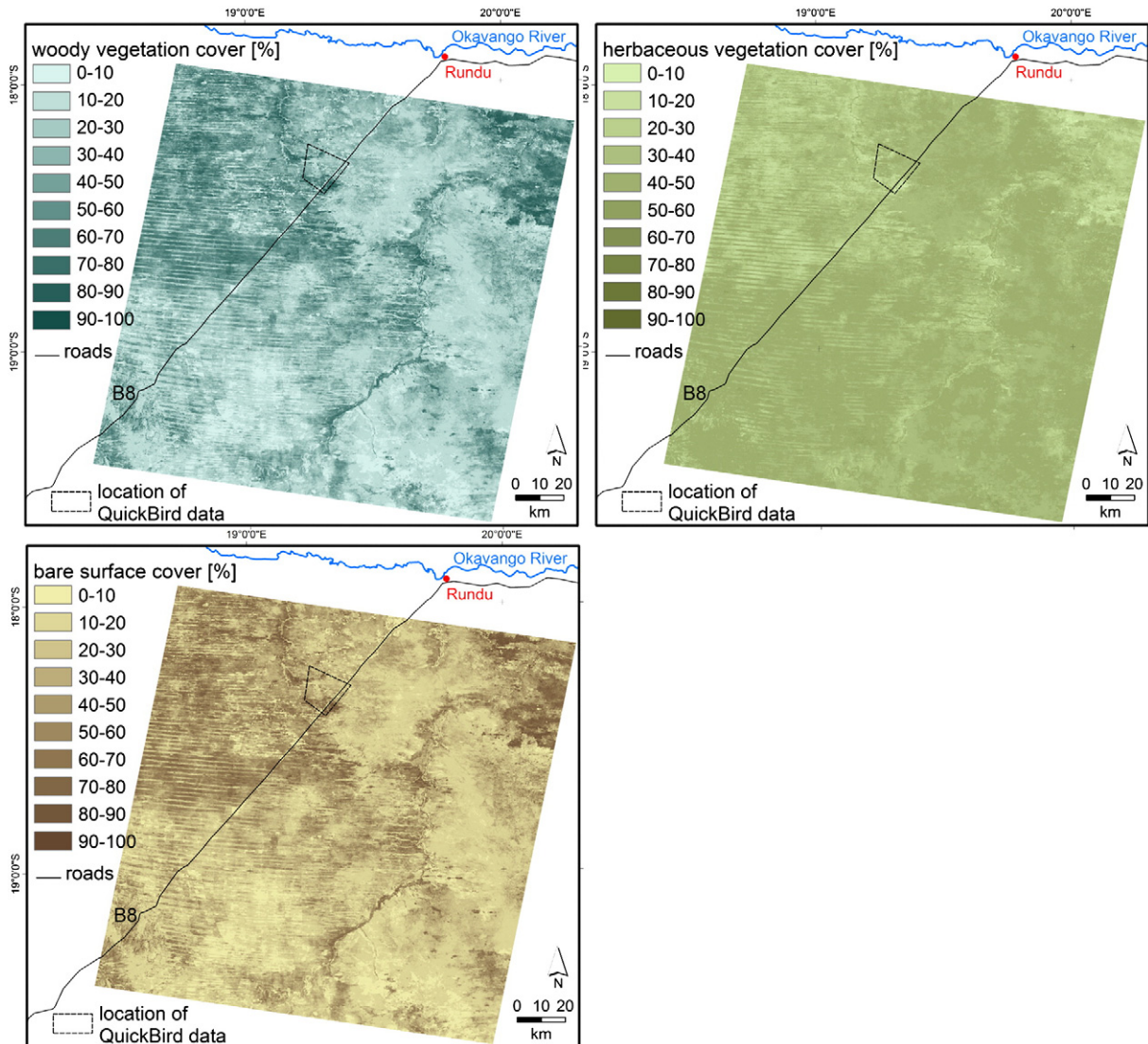


Fig. 11. Fractional cover of woody growth forms, herbaceous growth forms and bare surface delineated from Landsat TM data for the Kalahari Woodland Savanna.

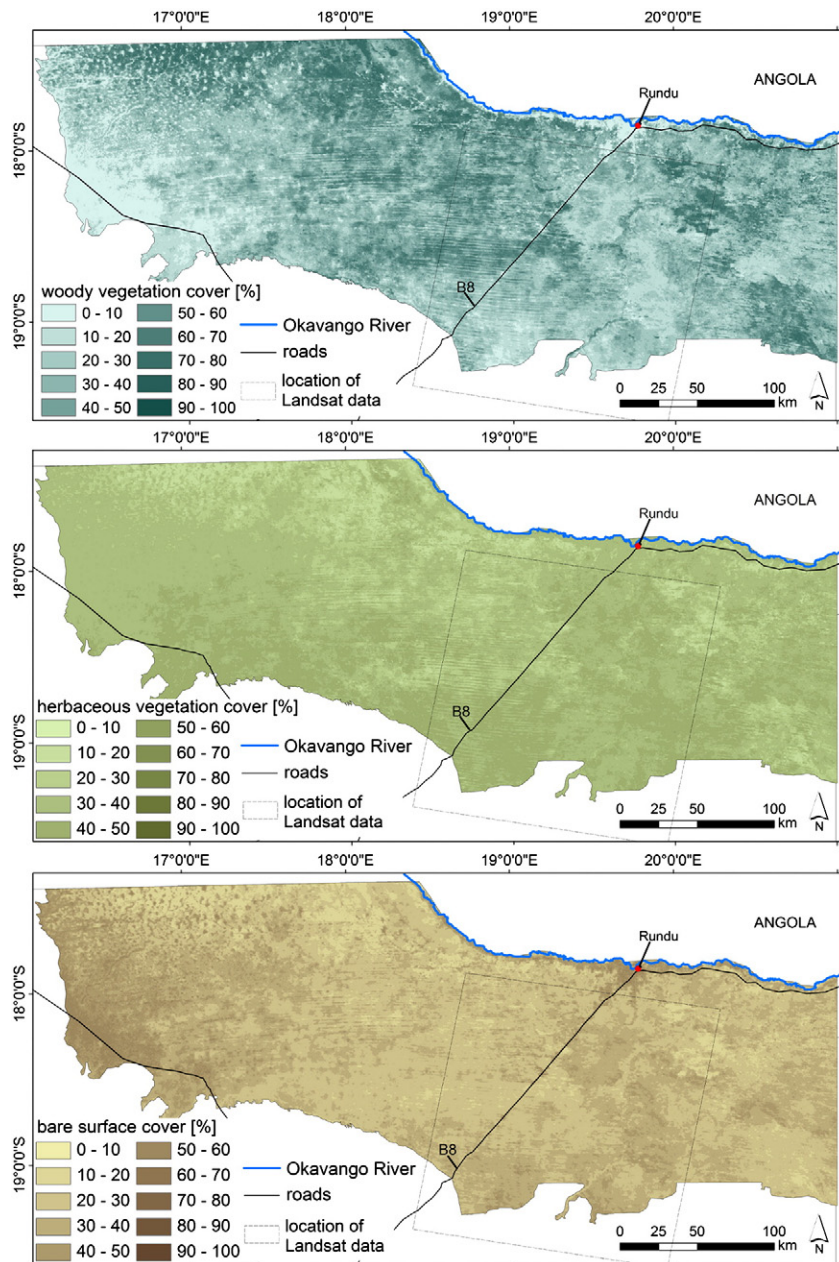


Fig. 12. Fractional cover of woody growth forms, herbaceous growth forms and bare surface delineated from MODIS data for the Kalahari Woodland Savanna.

(center) and MODIS (right). A north–south oriented fire-proof fence subdivides the region into a western, fire protected area and an eastern area where fires occur every second year on average. At the highest resolution, individual shrubs and trees could be identified in the classification. Agricultural fields and the oval thicket structures with gradual transitions to neighboring open savanna vegetation are well depicted at the highest resolution. The slight differences in woody cover inside and outside the fire-proof fencing are observable with rather dense and open shrub understory respectively. At the resolution of Landsat, shrubs and trees are captured as cover fractions at sub-pixel level (Fig. 13, center). This allows for an appropriate characterization of the savanna vegetation structure with gradual transitions between areas of open savanna and closed thickets at the 30 m spatial resolution. Based on discrete land cover classes it would not be possible to capture these spatial patterns with their gradual characteristics. Furthermore, in the Landsat-based results, agricultural fields can clearly be identified as well as the slight differences in vegetation structure between areas of low and high

fire frequency west and east of the fence. At MODIS resolution, smaller landscape elements such as small agricultural fields and thickets are blurred (Fig. 13, right). However, the general characteristics with gradual transitions between open and closed vegetation structure are still well reproduced in the form of sub-pixel fractional cover. Also at this coarsest level, the slight differences in vegetation structure due to differences in fire frequency are clearly depicted.

6. Conclusions

In this study, sub-pixel fractional cover of woody growth forms (trees and shrubs), herbaceous growth forms, and bare surface were delineated for two savanna landscapes in Southern Africa. The considered cover types constitute the main components of land cover in savannas and represent meaningful units with respect to ESS. The fractional cover values were derived with a multi-resolution approach based on random forest regression techniques and optical remote sensing data of three spatial resolutions (QuickBird/IKONOS, Landsat, MODIS). The

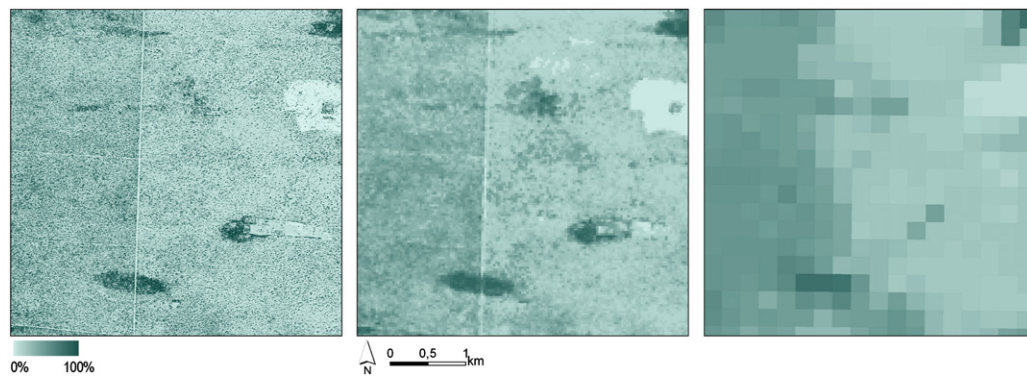


Fig. 13. Woody cover delineated from QuickBird (left), Landsat TM (center) and MODIS (right) data for a subregion in the central Kalahari Woodland Savanna. A fire-proof fence crosses the region from north to south. It subdivides the area into savanna vegetation with low (west) and high (east) fire frequency.

results showed that the approach allows the description of the characteristic features of vegetation structure in savanna landscapes. Root mean squared errors of the fractional cover datasets range between 3.1% and 8.2% when compared to higher resolution data, and between 4.4% and 9.9% when compared to field surveys. In general, spectrally homogeneous cover types could be estimated with higher accuracies than heterogeneous cover types. The influence of error propagation from the Landsat to the MODIS level was estimated by Monte Carlo simulations. These analyses showed minor inter-resolution propagation of errors with 90% of the Monte Carlo simulations resulting in an increase of RMSE of less than 4% (woody vegetation), 3.5% (herbaceous vegetation) and 2% (bare surface) at MODIS level.

In the presented procedure, the vegetation structure is delineated as sub-pixel cover already at a spatial resolution of 30 m. This allows for a better characterization of the grassland, savanna and shrubland biome which has a highly mixed vegetation structure at this intermediate spatial resolution. Accordingly, the mapping results of the intermediate resolution are less generalized and consequently build a sound basis for the estimation of fractional cover at MODIS resolution.

The presented approach is generally suitable for grasslands, savannas and shrublands, also outside Africa. However, high resolution data (~1 m) for the region is a prerequisite, and Landsat and MODIS feature spaces need to be adapted to the prevalent phenological characteristics. For regional land cover mapping activities in such landscapes, the presented fractional cover delineation has a high potential to improve traditional discrete classification approaches. If required for certain applications or for reason of comparison with other datasets, the presented fractional cover information can be aggregated to discrete classes in accordance with the UN LCCS (Di Gregorio, 2005) where class definitions are generally based on the coverage of growth forms. This kind of concept could lead to an improved representation of the particularly problematic heterogeneous classes in global land cover maps not only for the African grasslands, savannas and shrublands but presumably also for other heterogeneous biomes and ecotones such as the tundra, the taiga-tundra ecotone and the steppe regions.

Acknowledgments

This study was funded by the German Federal Ministry of Education and Research (BMBF) as part of the project BIOTA South (Biodiversity Transect Analysis). We appreciate that MODIS data could be obtained free of charge through the online Data Pool at the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (http://lpdaac.usgs.gov/get_data) and Landsat data could be obtained free of charge through the Landsat archive of the United States Geological Survey (USGS) and through the archive of the South African Council for Scientific and Industrial Research (CSIR).

References

- Bicheron, P., Defourny, P., Brockmann, C., Schouten, L., Vancutsem, C., Huc, M., et al. (2008). *GLOBCOVER*. Products Description and Validation Report (<http://ionia1.esrin.esa.int/>), last date accessed: December 11 2011)
- Bonham, C. (1989). *Measurements for terrestrial vegetation*. New York: John Wiley & Sons.
- Bontemps, S., Defourny, P., Bogaert, E. V., Arino, O., Kalogirou, V., & Perez, J. R. (2011). *GLOBCOVER 2009*. Products Description and Validation Report (http://ionia1.esrin.esa.int/docs/GLOBCOVER2009_Validation_Report_2.2.pdf), last date accessed: December 11 2011)
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123–140.
- Breiman, R. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Boca Raton: Chapman & Hall/CRC.
- Colditz, R. R., Conrad, C., Wehrmann, T., Schmidt, M., & Dech, S. (2008). TiSeG: A flexible software tool for time-series generation of MODIS data utilizing the quality assessment science data set. *IEEE Transactions on Geoscience and Remote Sensing*, 46, 3296–3308.
- Crist, E. P., Laurin, R., & Cicone, R. C. (1986). Vegetation and soils information contained in transformed Thematic Mapper Data. *Proceedings of IGARSS' 86 Symposium* (pp. 1465–1470).
- DeFries, R. S., Field, C. B., Fung, I., Justice, C. O., Los, S., Matson, P. A., et al. (1995). Mapping the land surface for global atmosphere-biosphere models: Towards continuous distributions of vegetation's functional properties. *Journal of Geophysical Research*, 100, 20867–20882.
- DeFries, R. S., Hansen, M. C., Townshend, J. R. G., Janetos, A. C., & Loveland, T. R. (2000). A new global 1-km dataset of percentage tree cover derived from remote sensing. *Global Change Biology*, 6, 247–254.
- Di Gregorio, A. (2005). *Land cover classification system. Classification concepts and user manual. Software version 2*. Based on the original software version 1 Prepared by Antonio Di Gregorio and Louisa J.M. Jansen (<http://www.fao.org/docrep/008/y7220e/y7220e00.htm>), last date accessed: December 11 2011)
- Eldridge, D. J., Bowker, M. A., Maestre, F. T., Roger, E., Reynolds, J. F., & Whitford, W. G. (2011). Impacts of shrub encroachment on ecosystem structure and functioning: towards a global synthesis. *Ecology Letters*, 14, 709–722.
- Esch, T., Thiel, M., Bock, M., Roth, A., & Dech, S. (2008). Improvement of image segmentation accuracy based on multi-scale optimization procedure. *IEEE Geoscience and Remote Sensing Letters*, 5, 463–467.
- FAO (1998). *World reference base for soil resources. World Soil Resources Report*, 84, Rome: FAO (ISBN: 92-5-104141-5).
- Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., et al. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 114, 168–182.
- Garzon, M. B., Blazek, R., Neteler, M., de Dios, R. S., Ollero, H. S., & Furlanello, C. (2006). Predicting habitat suitability with machine learning models: The potential area of *Pinus sylvestris* L. in the Iberian Peninsula. *Ecological Modelling*, 197, 383–393.
- Giess, W. (1998). *A preliminary vegetation map of Namibia*. (3rd revised edition). *Dinteria*, 4, (pp. 5–15).
- Giri, C., Zhu, Z., & Reed, B. (2005). A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets. *Remote Sensing of Environment*, 94, 123–132.
- GTOS (2008). *Terrestrial essential climate variables for climate change assessment, mitigation and adaptation*. Biennial report supplement. (Rome: FAO (<http://www.fao.org/gtos/doc/pub52.pdf>), last date accessed: December 11 2011).
- Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Sohlberg, R., Dimiceli, C., & Carroll, M. (2002a). Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. *Remote Sensing of Environment*, 83, 303–319.
- Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Marufu, L., & Sohlberg, R. (2002b). Development of a MODIS tree cover validation data set for Western Province, Zambia. *Remote Sensing of Environment*, 83, 320–335.
- Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C., & Sohlberg, R. A. (2003). Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS Vegetation Continuous Fields algorithm. *Earth Interactions*, 7, 1–15.

- Hansen, M. C., Townshend, J. R. G., DeFries, R. S., & Carroll, M. (2005). Estimation of tree cover using MODIS data at global, continental and regional/local scales. *International Journal of Remote Sensing*, 26, 4359–4380.
- Heiskanen, J. (2008). Evaluation of global land cover data sets over the tundra–taiga transition zone in northernmost Finland. *International Journal of Remote Sensing*, 29, 3727–3751.
- Herold, M., Mayaux, P., Woodcock, C. E., Baccini, A., & Schmullius, C. (2008). Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment*, 112, 2538–2556.
- Heuvelink, G. B. M. (1998). *Error propagation in environmental modeling with GIS*. London/Bristol: Taylor & Francis.
- Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25, 295–309.
- Irons, J. R., Markham, B. L., Nelson, R. F., Toll, D. L., Williams, D. L., Latty, R. S., et al. (1985). The effects of spatial resolution on the classification of Thematic Mapper data. *International Journal of Remote Sensing*, 6, 1385–1403.
- Jakubauskas, M. E., Legates, D. R., & Kastens, J. H. (2001). Harmonic analysis of time-series AVHRR NDVI data. *Photogrammetric Engineering and Remote Sensing*, 67, 461–470.
- Jensen, J. R. (1996). *Introductory digital image processing: A remote sensing perspective*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Latifovic, R., & Olthof, I. (2004). Accuracy assessment using sub-pixel fractional error matrices of global land cover products derived from satellite data. *Remote Sensing of Environment*, 90, 153–165.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.
- Liu, W., & Wu, E. Y. (2005). Comparison of non-linear mixture models: Sub-pixel classification. *Remote Sensing of Environment*, 94, 145–154.
- Lloyd, J., Bird, M. I., Vellen, L., Miranda, A. C., Veenendaal, E. M., Djagbletey, G., et al. (2008). Contributions of woody and herbaceous vegetation to tropical savanna ecosystem productivity: A quasi-global estimate. *Tree Physiology*, 28, 451–468.
- Marceau, D. J., Howarth, P. J., & Gratton, D. J. (1994). Remote sensing and measurement of geographical entities in a forested environment. 1. *The scale and spatial aggregation problem*. *Remote Sensing of Environment*, 49, 93–104.
- Meik, J. M., Jeo, R. M., Mendelsohn, J. R., & Jenks, K. E. (2002). Effects of bush encroachment on an assemblage of diurnal lizard species in central Namibia. *Biological Conservation*, 106, 29–36.
- Mendelsohn, J., Jarvis, A., Roberts, C., & Robertson, T. (2002). *Atlas of Namibia. A portrait of the land and its people*. Cape Town: David Philip Publishers.
- Montesano, P. M., Nelson, R., Sun, G., Margolis, H., Kerber, A., & Ranson, K. J. (2009). MODIS tree cover validation for the circumpolar taiga–tundra transition zone. *Remote Sensing of Environment*, 113, 2130–2141.
- Moody, A., & Woodcock, C. E. (1995). The influence of scale and spatial characteristics of landscapes on land-cover mapping using remote sensing. *Landscape Ecology*, 10, 363–379.
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., et al. (2001). Terrestrial ecoregions of the world: A new map of life on Earth. *BioScience*, 51, 933–938.
- R Development Core Team (2011). *R: A Language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing (<http://www.R-project.org>, last date accessed: December 11 2011).
- Richter, R. (1996). A spatially adaptive fast atmospheric correction algorithm. *International Journal of Remote Sensing*, 17, 1201–1214.
- Richter, R. (2007). *Atmospheric/topographic correction for satellite imagery*. ATCOR-2/3 User Guide, Version 6.3 (DLR-IB 565-01/07).
- Ripley, B. (2012). *tree: Classification and regression trees*. R package version 1.0-31 (<http://CRAN.R-project.org/package=tree>).
- Rothauge, A. (2006). *The effect of frame size and stocking rate on diet selection of cattle and range condition in the Camelthorn savanna of east-central Namibia*. PhD thesis. Windhoek: University of Namibia.
- Scholes, R. J., & Archer, S. R. (1997). Tree-grass interactions in savannas. *Annual Review of Ecology and Systematics*, 28, 517–544.
- Smith, J. H., Wickham, J. D., Stehman, S. V., & Yang, L. (2002). Impacts of patch size and land-cover heterogeneity on thematic image classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 68, 65–70.
- Smith, J. H., Stehman, S. V., Wickham, J. D., & Yang, L. (2003). Effects of landscape characteristics on land-cover class accuracy. *Remote Sensing of Environment*, 84, 342–349.
- Strohbach, B. J., & Petersen, A. (2007). Vegetation of the central Kavango woodlands in Namibia: An example from the Mile 46 Livestock Development Centre. *South African Journal of Botany*, 73, 391–401.
- Strohbach, B. J., Strohbach, M., Kutuahuripa, J. T., & Mouton, H. D. (2004). *A reconnaissance survey of the landscapes, soils and vegetation of the Eastern Communal Areas (Otjozondjupa and Omaheke Regions), Namibia*. Windhoek: National Botanical Research Institute and Agro-Ecological Survey Programme, Directorate Agriculture Research and Training, Ministry of Agriculture, Water and Rural Development.
- Tottrup, C., Rasmussen, M. S., Eklundh, L., & Jönsson, P. (2007). Mapping fractional forest cover across the highlands of mainland Southeast Asia using MODIS data and regression tree modelling. *International Journal of Remote Sensing*, 28, 23–46.
- White, M. A., Shaw, J. D., & Ramsey, R. D. (2005). Accuracy assessment of the vegetation continuous field tree cover product using 3954 ground plots in the south-western USA. *International Journal of Remote Sensing*, 26, 2699–2704.
- Xu, M., Watanachaturaporn, P., Varshney, P. K., & Arora, M. K. (2005). Decision tree regression for soft classification of remote sensing data. *Remote Sensing of Environment*, 97, 322–336.
- Yanchen, B. (2008). A method to improve the accuracy of remote sensing data classification by exploiting the multi-scale properties in the scene. *Proceedings of the 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences Shanghai, P. R. China, June 25–27, 2008* (pp. 183–188).