Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape

Hannes Müller[1]
Philippe Rufin[1]
Patrick Griffiths[1]
Auberto José Barros Siqueira[2]
Patrick Hostert[1,3]

[1]Geography Department, Humboldt Universität zu Berlin, Unter den Linden 6, D-10099
Berlin, Germany
Hannes.Mueller@geo.hu-berlin.de
Philippe.Rufin@geo.hu-berlin.de
Patrick.Griffiths@geo.hu-berlin.de

[2]Department of Sanitary and Environmental Engineering, Universidade Federal de Mato
Grosso, Av. Fernando Corrêa da Costa, B-78060-900 Cuiabá, Brazil
Ajbsiqueira@gmail.com

[3]Integrative Research Institute on Transformations of Human–Environment Systems,
Humboldt Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany
Patrick.Hostert@geo.hu-berlin.de

Abstract: Better remote sensing based information on the global distribution of croplands and pastures is urgently needed. Without reliable cropland-pasture separation it will be impossible to retrieve high-quality information on agricultural expansion or land use intensification, and on related ecosystem service provision. In this context, the savanna biome is critically important but information on land use and land cover (LULC) is notoriously inaccurate in these areas. This is due to pronounced spatial-temporal dynamics of agricultural land use and spectral similarities between cropland, pasture, and natural savanna vegetation. In this study, we investigated the potential to reliably separate cropland, pasture, natural savanna vegetation, and other relevant land cover classes employing Landsat-derived spectral-temporal variability metrics for a savanna landscape in the Brazilian Cerrado. Spectral-temporal variability metrics were derived from 344 Landsat images across four footprints between 2009 and 2012. Our results showed a reliable separation between cropland, pasture, and natural savanna vegetation achieving an overall accuracy of 93%. There is great potential for expanding our approach towards large parts of the Cerrado biome and to other savanna systems which still suffer from inaccurate LULC information.

Keywords: Compositing; Spectral-temporal metrics; Landsat time series; Land use and land cover classification; Brazilian Cerrado; Savannas

1. Introduction

Food security is of global concern, and land-based production has to meet the demands of an estimated 9 billion people in 2050 (Godfray et al., 2010). To define an optimum trade-off between agricultural expansion and intensification on the one hand, and sustainable ecosystem service provision on the other hand, reliable information about the global distribution of croplands and pasture areas is required (Fritz et al., 2013; Garnett et al., 2013). In this context, the savanna biome is critically important because it suffers from enormous land conversions and spatially explicit mapping of land use and land cover (LULC) is notoriously challenging and inaccurate in this area, which hampers monitoring of ecosystem changes (Herold et al., 2008).

So far, the majority of remote sensing studies in Brazil focused on the Amazon biome, including governmental programs like the annual deforestation monitoring (PRODES), the real time system for detection of deforestation (DETER), and the program for land use classification in deforested areas (TerraClass) (INPE, 2008). First attempts to build similar

knowledge for the Cerrado region are conducted by the remote sensing department of the Federal University of Goiás, who created a Systematic Monitoring of Deforestation in the Cerrado Biome (SIAD) since the year 2006 (LAPIG, 2014). However, no program has been established for monitoring general land use in the Cerrado biome and remote sensing faces numerous challenges there.

Besides the strong seasonality of natural vegetation and the diversity of crop types in space and time (Sano et al., 2007), spectral similarities between cropland, pasture, and natural savanna vegetation complicate differentiating LULC in the Brazilian Cerrado (Sano et al., 2010; Grecchi et al., 2013). This confusion results from spectral similarity between - as well as a high spectral heterogeneity within each land cover type (Brannstrom et al., 2008). Due to these spectral ambiguities, most classification approaches predominantly rely on dense temporal information, such as time series data obtained from the Moderate Resolution Imaging Spectrometer (MODIS) for separating natural savanna vegetation, cropland, and pasture areas (Adami et al., 2011; Arvor et al., 2011a). However, MODIS-based analyses cannot monitor LULC before 2000 (when satellite became operational) and often do not capture fine-scale patterns in heterogeneous savanna ecosystems such as the Brazilian Cerrado or South-African savannas (Munyati and Mboweni, 2013). Recent approaches therefore include Landsat imagery (30m resolution) to overcome the limitation of the MODIS spatial scale (250m resolution) and temporal extent in heterogeneous savanna regions (Grecchi et al., 2013).

So far, most Landsat-based time series have been created on an annual basis to monitor forest cover dynamics (Kennedy et al., 2012; Griffiths et al., 2012). We therefore hypothesize that identifying complex LULC classes within agricultural systems will profit from intra-annual information to capture phenological characteristics as well. However, due to Landsat's 16-day repeat cycle, cloud contaminations, and long term acquisition plans, regular spacing of intra-annual acquisitions is difficult and direct quantification of phenological metrics is challenging. To overcome these limitations, interpolation and curve fitting methods have been successfully employed for regions with high data availability, notably in North America (Zhong et al., 2014). However, it is questionable whether single dates of phenological characteristics in regions of low observation density should be extracted, given the degree of generalization involved in bridging large temporal data gaps. Less specific but more robust phenological indicators can be derived by simply calculating mean, range, and standard deviation of available reflectance observations for one or for multiple years (Griffiths et al., 2013a; Hansen et al., 2013). These spectral-temporal variability metrics (from now on "spectral-temporal metrics") capture important phenological information and can be analyzed on a seasonal, annual or multi-annual basis, allowing for a wide range of applications for characterizing land use systems in space and time.

In this study, we employ Landsat-derived spectral-temporal metrics to reliably separate cropland, pasture, natural savanna vegetation, and forest in a heterogeneous savanna landscape. In order to profit from temporal information we used a dense Landsat time series of 344 Landsat scenes across four footprints between 2009 and 2012.

2. Data and methods
2.1 Study area and description of LULC classes
To evaluate the potential of the Landsat-derived spectral-temporal information for distinguishing LULC classes in the Brazilian Cerrado, we conducted a case study in the Rio das Mortes watershed, a tributary of the Araguaia River that drains into the Tocantins River The watershed boundaries were derived from a digital elevation model with 90m spatial resolution (SRTM, 2008) and delineate an area of about 18,000 km² (Fig. 1). The regional

climate exhibits a distinct dry season from May to September and corresponds to a tropical savanna (Aw), according to the Köppen climate classification (Moreno et al., 2005).

The watershed has undergone significant LULC change between the 1970 and 2005 becoming one of the major agricultural production centers of Mato Grosso. After 2005, cropland expansion stopped, which is important in context of our classification approach because our approach requires a high stability of land use between 2009 and 2012. This assumption is supported by field interviews, existing literature and agrarian statistics (IBGE, 2010; Grecchi et al., 2013). Within our study period, the main sowing season for soy starts with the wet season between September and October. Sowing and harvesting practices differ for individual rotation systems ("single-" or "double-cropping") and crop type (e.g. cotton, corn), but the production cycle usually ends with the last harvests at the beginning of the dry season in July (Arvor et al., 2011b).
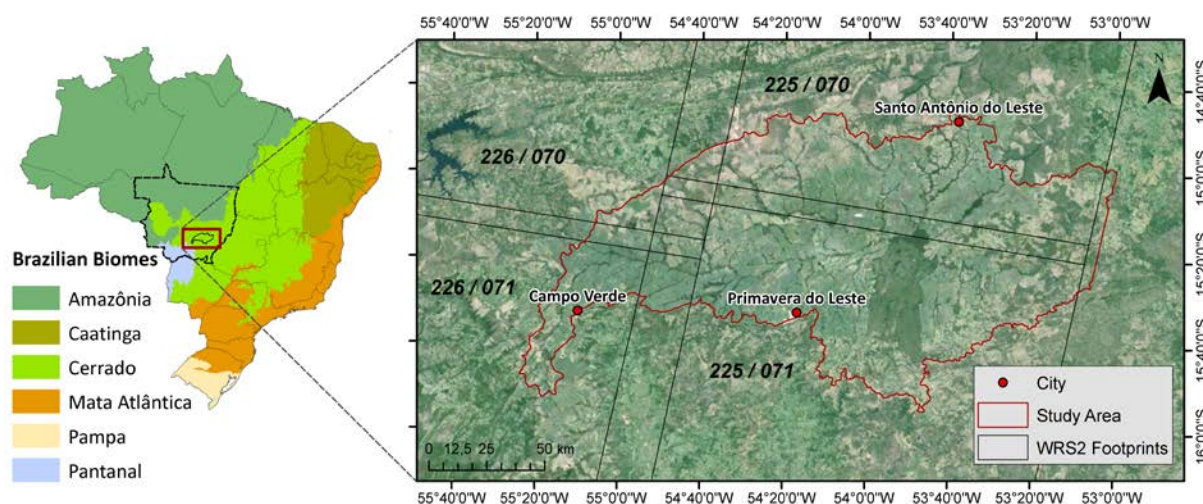

Figure 1: Left: Distribution of the predominant biomes in Brazil and Mato Grosso state. Right: Close-up of the study area based on Google Earth imagery.

To characterize the predominant LULC classes in the study region, we differentiate forest, savanna, cropland, pasture and water bodies and non-vegetated land. The "non-vegetated land" class includes the built-up environment, open soils and rock outcrops. The savanna class comprises woody grassland (Campo Cerrado), woodland Cerrado (Cerrado sensu stricto) and wooded Cerrado (Cerradão). Croplands include both "single" and "double-cropping" systems for soy, corn and cotton. During our field visits it was apparent that these systems can also include fallow cycles with grassland and temporary cattle ranching. Pasture areas consist of planted grasslands used exclusively for cattle ranching. These areas are often characterized by scattered residuals of woody savanna vegetation, indicating the absence of large-scale mechanized agriculture.

2.2 Satellite imagery, atmospheric correction and cloud masking

Satellite imagery was obtained from all available Landsat acquisitions within a three-year period, from January 2nd 2009 to February 28th 2012. We used TM and ETM+ Landsat data exclusively in precision terrain corrected (L1T) format resulting into 344 images across four footprints. To ensure comparability of input data across different Landsat sensors, footprints, and acquisition dates, all imagery was converted to surface reflectance using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) atmospheric correction algorithm (Masek et al., 2006). Cloud and shadow masks were calculated for all imagery using FMask (Zhu and Woodcock, 2012). The FMask algorithm uses textural and spectral

information to detect cloud and cloud shadows on a probabilistic basis. To capture as many clouds and cloud shadows as possible, the cloud probability threshold was set to a highly conservative value (1%). All scenes from different UTM zones were reprojected to WGS 84 UTM 21 South.

2.3 Extraction of spectral-temporal variability metrics

To capture most phenological characteristics of the investigated LULC classes, five statistical metrics (mean, median, standard deviation, 75% quartile, and interquartile range) were computed from all cloud/cloud shadow-free observations between 2009 and 2012, using the composing approach described by Griffiths et al. (2013b). To include all spectral information, the five statistical metrics were calculated for each spectral band individually (Table 1). We also calculated a SWIR-Index as the sum of SWIR/NIR reflectance normalized over the number of clear observations. The SWIR-Index serves as an integrated measure which potentially supports the separation of pasture and cropland areas.

Table 1: Overview of the 31 spectral-temporal variability metrics.

No. of variables	Metric	Spectral bands
6	Mean	1-5, 7
6	Median	1-5, 7
6	Standard deviation	1-5, 7
6	75% Quantile	1-5, 7
6	Interquartile range (25%-75%)	1-5, 7
1	SWIR-Index	4, 5, 7

2.4 Land cover mapping and validation

LULC mapping was performed using a random forest (RF) classifier (Breiman, 2001). Model calibration and prediction was carried out using the statistical software CRAN R (R Development Core Team, 2011) and the "randomForest" package (version 4.6-7 by Liaw and Wiener, 2002). We employed the random forest model for all further analyses, using a standard parameterization for the number of trees (500), the number of training data for each tree (63% of sample size per class), and the number of predictor variables tried at each split (5, approximated by the square root of 31 predictor variables).

For training of the RF model, we randomly selected 626 training points with a minimum distance of 350m. The final LULC map was validated using a stratified random sampling of 470 validation points with a minimum distance of 350m. To classify the training and validation points, we utilized very high-resolution imagery of Google Earth from 2009 to 2012 (Quickbird, GeoEye). If recent Google Earth imagery was not available, Landsat images and calculated variability metrics were used as reference data. To account for the stratified random sampling design, overall accuracies and area estimates were adjusted after Olofsson et al. (2013). In addition, confidence intervals for the adjusted overall accuracy were calculated to report uncertainty of the accuracy estimates introduced by the number of validation points (Congalton and Green, 2009).

3. Results

Cropland was identified as the dominant land use class, covering approximately 51% (9,096 km²) of the study area (Figure 2). Savanna areas were ranked as the second largest land cover type (24%, 4,298 km²) and can be found in the indigenous reserve, along gallery forests and within the extensively managed eastern and northern parts of the watershed. The savanna

areas form a mosaic with spots of cattle farming which cover 15% (2,682 km²) of the study site. Forests are dominantly located along river streams and account for 9% (1,737 km²) of the total land cover. Non-vegetated land was identified in the cities of Campo Verde, Primavera do Leste and Santo Antônio do Leste. In addition, this class was detected in close proximity to cropland and pasture areas representing 0.8% of the study area.
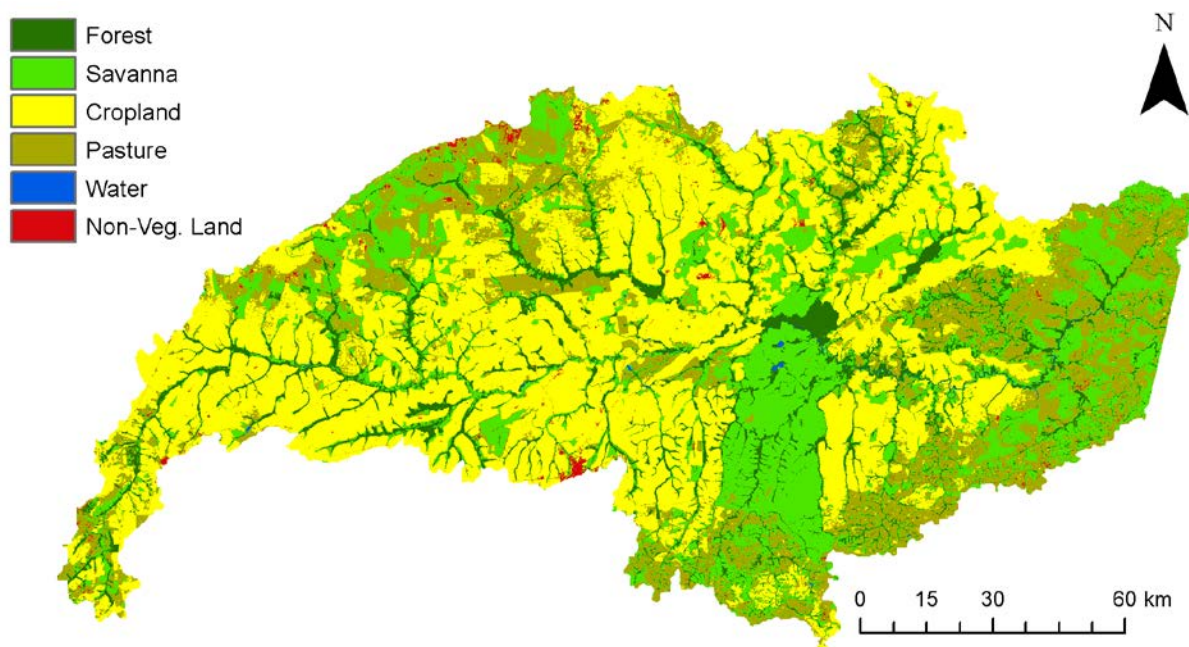

Figure 2: Classification results of the Rio das Mortes watershed.

The LULC classification achieved an adjusted overall accuracy of 93% with a 95% confidence interval margin of ±2%. Highest reliability was observed for the forest, savanna and water classes that revealed user`s and producer`s accuracies greater than 90% (Table 3). No commission error was observed for the cropland class.

Table 3: User`s and producer`s accuracy of validation results, normalized between 1 (100%) and 0 (0%). User`s and producer`s accuracy translate into omission and commission error as follows: omission error = 1-producers accuracy, commission error = 1-user`s accuracy.

	Forest	Savanna	Cropland	Pasture	Water	Non-veg. land
User`s accuracy	0.97	0.98	1	0.70	1	0.84
Producer`s accuracy (error adjusted)	0.97 (0.91)	0.93 (0.97)	0.80 (0.89)	0.94 (0.99)	0.99 (0.92)	0.98 (0.67)

Main class confusion arose from the overestimation of pasture at the expense of cropland. This effect resulted in a commission error of 30% for the pasture class and an omission error of 11% for the cropland class after error adjustment. A minor omission error of 1% for pasture originated from the overestimation of non-vegetated land on pasture areas, leading to a commission error of 16% for the non-vegetated land.

4. Discussion and conclusions

The LULC classification achieved an overall accuracy of 93% and class-wise uncertainty was generally low, including the target classes of cropland and pasture. We therefore conclude that our spectral-temporal classification approach provided a reliable separation between cropland, pasture, and natural savanna vegetation. Remaining errors mainly relate to the slight overestimation of the pasture class and can be explained by two factors: Firstly, we labeled our training data in favor of a conservative estimation of the cropland class, as it is the most dominant and heterogeneous land use class in the study area. Therefore, the commission error of the cropland class is generally low at the expense of a slightly overestimated pasture class. Secondly, there are extensively managed cropland areas that are used as pastures during one or two rotational cycles. These mixed systems appear as croplands in the validation data but are most likely classified as pastures. Land conversions during our observation period could also hamper the explanatory power of our spectral-temporal variability metrics. However, such land conversions are unlikely, given our analysis period compared to knowledge on recent land use dynamics (section 2.1).

To our best knowledge, this is the first study employing Landsat-derived, spectral-temporal information for classifying LULC in the heterogeneous Brazilian Cerrado. In a regionally comparable Landsat-based study, Grecchi et al. (2013) reported an overall accuracy of 85% when including MODIS data and pre-existing land cover maps, and not separating forest and savanna areas. This supports our finding that multitemporal information is important in such a setting to separate spectrally similar classes reliably. Other studies that exclusively employed Landsat data reached overall accuracies of above 80% only when fusing cropland and pasture into an agro-pastoral class (Jepson, 2005; Brannstrom et al., 2008). Such aggregation, however, limits linking remote sensing based results with ecological or management processes of great importance, e.g. how pastoralism versus intensive cropping alters savanna ecosystems. Our results therefore emphasize the potential of employing high-resolution spectral-temporal variability metrics for identifying LULC in heterogeneous savanna regions.

So far, it is a great challenge to operationally observe LULC in savanna landscapes on a large spatial scale. Sano et al. (2010) used a monotemporal wall to wall mosaic for manual interpretation of 170 Landsat images for the entire Cerrado Biome and reported high class confusion between croplands and pastures (OA= 71%). These results suggest great potential for improvement by using multitemporal Landsat data. A preliminary screening of the Landsat archive showed that 80% of the entire Cerrado biome has a similar Landsat coverage compared to our study area (+-20% scene availability in 2009-2012), rendering our multitemporal classification approach highly applicable over large spatial extents.

7. Literature

Adami, M., Rudorff, B.F.T., Freitas, R.M., Aguiar, D.A., & Mello, M.P. (2011). Remote Sensing Time Series to Evaluate Direct Land Use Change of Recent Expanded Sugarcane Crop in Brazil. In *1st World Sustain. Forum*.

Arvor, D., Jonathan Milton, Meirelles, M.S.P., Dubreuil, V., & Durieux, L. (2011a). Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. *International Journal of Remote Sensing*, *32*, 7847-7871.

Arvor, D., Meirelles, M.S.P., Dubreuil, V., Bégué, A., & Shimabukuro, Y.E. (2011b). Analyzing the agricultural transition in Mato Grosso, Brazil, using satellite-derived indices. *Applied Geography*, *32*, 702-713.

Brannstrom, C., Jepson, W., Filippi, A., Redo, D., Xu, Z., & Ganesh, S. (2008). Land change in the Brazilian savanna (Cerrado), 1986-2002: Comparative analysis and implications for land-use policy. *Land Use Policy*, *25*, 579-595.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.

Congalton, R.G., & Green, K. (2009). *Assessing the accuracy of remotely sensed data. Principles and practices*. (2nd ed.). Boca Raton: CRC Press/Taylor & Francis.

Fritz, S., See, L., You, L., Justice, C., Becker-Reshef, I., Bydekerke, L., Cumani, R., Defourny, P., Erb, K.H., Foley, J., & others (2013). The Need for Improved Maps of Global Cropland. *Eos, Transactions American Geophysical Union*, *94*, 31-32.

Garnett, T., Appleby, M.C., Balmford, A., Bateman, I.J., Benton, T.G., Bloomer, P., Burlingame, B., Dawkins, M., Dolan, L., Fraser, D., & others (2013). Sustainable intensification in agriculture: premises and policies. *Science*, *341*, 33-34.

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., & Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. *Science*, *327*, 812-818.

Grecchi, R.C., Gwyn, Q.H.J., Bénié, G.B., & Formaggio, A.R. (2013). Assessing the spatio-temporal rates and patterns of land-use and land-cover changes in the Cerrados of southeastern Mato Grosso, Brazil. *International Journal of Remote Sensing*, *34*, 5369-5392.

Griffiths, P., Kuemmerle, T., Kennedy, R., Abrudan, I., Knorn, J., & Hostert, P. (2012). Using annual time-series of Landsat images to assess the effects of forest restitution in post-socialist Romania. *Remote Sensing of Environment*, *118*, 199-214.

Griffiths, P., Müller, D., Kuemmerle, T., & Hostert, P. (2013a). Agricultural land change in the Carpathian ecoregion after the breakdown of socialism and expansion of the European Union. *Environmental Research Letters*, *8*, 45024.

Griffiths, P., van der Linden, S., Kuemmerle, T., & Hostert, P. (2013b). A Pixel-Based Landsat Compositing Algorithm for Large Area Land Cover Mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1–14.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., & Townshend, J.R.G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, *342*, 850–853.

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.

IBGE (2010). Estatística Agropecuária do Mato Grosso 1988-2010. Insituto Brasileiro de Geographia e Estatística. http://www.ipeadata.gov.br/. Last access : 28.1.2014

INPE (2008). Monitoramento Da Cobertura Florestal Da Amazônia Por Satélites. Sistemas Prodes, Deter, Degrad E Queimadas 2007-2008. Coordenação Geral de Observação da Terra São José dos Campos. http://www.obt.inpe.br/prodes/Relatorio_Prodes2008.pdf. Last access : 28.1.2014

Jepson, W. (2005). A disappearing biome? Reconsidering land-cover change in the Brazilian savanna. *The Geographical Journal*, *171*, 99-111.

Kennedy, R., Yang, Z., Cohen, W., Pfaff, E., Braaten, J., & Nelson, P. (2012). Spatial and temporal patterns of forest disturbance and regrowth within the area of the Northwest Forest Plan. *Remote Sensing of Environment*, 117–133.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R news*, *2*, 18-22.

LAPIG (2014). Monitoramento Sistemático dos Desmatamentos no Bioma Cerrado (SIAD-Cerrado). Universidade Federal de Goiás. http://www.lapig.iesa.ufg.br/lapig/index.php?option=com_content&view=article&id=38&Itemid=52. Last access : 28.7.2014

Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., & Lim, T.-K. (2006). A Landsat surface reflectance dataset for North America, 1990-2000. *Geoscience and Remote Sensing Letters, IEEE*, *3*, 68-72.

Moreno, G., Souza-Higa, T.C.C. de, Maitelli, G.T., & Oliveira, A.U. de (2005). *Geografia de Mato Grosso: território, sociedade, ambiente:* Entrelinhas.

Munyati, C., & Mboweni, G. (2013). Variation in NDVI values with change in spatial resolution for semi-arid savanna vegetation: a case study in northwestern South Africa. *International Journal of Remote Sensing*, *34*, 2253-2267.

Olofsson, P., Foody, G.M., Stehman, S.V., & Woodcock, C.E. (2013). Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, *129*, 122-131.

R Development Core Team (2011). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria*.

Sano, E.E., Rosa, R., Brito, J.L.S., & Ferreira, L.G. (2007). Mapeamento de cobertura vegetal do bioma Cerrado. *Planaltina: Embrapa Cerrados*.

Sano, E.E., Rosa, R., Brito, J.L.S., & Ferreira, L.G. (2010). Land cover mapping of the tropical savanna region in Brazil. Environmental monitoring and assessment, 166, 113–124.

SRTM (2008). Shuttle Radar Topographic Mission. Dataset available at: http://srtm.csi.cgiar.org.

Zhong, L., Gong, P., & Biging, G.S. (2014). Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. *Remote Sensing of Environment*, *140*, 1-13.

Zhu, Z., & Woodcock, C.E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, *118*, 83-94.