Mapping Cerrado physiognomies using Landsat time series based phenological profiles

Marcel Schwieder [1]
Pedro Jorge Paixão Leitão [1]
Andreas Rabe [1]
Mercedes Maria da Cunha Bustamante [2]
Laerte Guimarães Ferreira[3]
Patrick Hostert [1]

[1] Humboldt-Universität zu Berlin
Unter den Linden 6 - 10099 - Berlin - Berlin, Germany
marcel.schwieder@geo.hu-berlin.de
p.leitao@geo.hu-berlin.de
andreas.rabe@geo.hu-berlin.de
patrick.hostert@geo.hu-berlin.de

[2] Universidade de Brasília- UNB
Universitário Darcy Ribeiro, Asa Norte - 70910-900 - Brasília- DF, Brasil
mercedesmcb@gmail.com

[3] Universidade Federal de Goiás - UFG
Prédio da Reitoria - 74001-970 - Goiânia - GO, Brasil
lapig.ufg@gmail.com

Abstract. Monitoring of the Brazilian Cerrado biome is crucial in order to fully understand its ecosystem functions, its response to changes, and to keep track of ongoing change processes in one of the world's biodiversity hotspots. The huge extent, heterogeneity and complex diversity of the Cerrado makes monitoring very challenging, but optical remote sensing based approaches have been shown to be valuable to tackle this task. We explored the potential of Landsat data to derive seasonal phenological profiles for the main Cerrado physiognomies, in order to use this information for classification purposes. Therefore we gathered Landsat TM, ETM+ and OLI data from the open Landsat data archive. To overcome data gaps in the time series, we applied a Gaussian kernel based convolution filter on Tasseled Cap transformed Landsat data. Thus, it was possible to derive comprehensible phenological profiles for distinct Cerrado physiognomies at Landsat spatial resolution. We used these derived seasonal profiles to train a Support Vector Machine classification model in order to map the main Cerrado physiognomies around Brasília, DF. Although classification errors were observed between similar classes in terms of their vegetation structure and density, we were able produce an accurate map that captures the spatial patterns of the vegetation physiognomies.

Keywords: support vector classification, remote sensing, tasseled cap, Cerrado monitoring.

1. Introduction

The Brazilian Cerrado is a highly heterogeneous biome that is rich in biodiversity with approximately 160,000 species in flora and fauna (Furley 1999). Following the Amazon it is the second largest biome in Brazil with an extent of around 2 million km² (Ratter et al. 1997). It is characterized by gradients in vegetation heights and tree densities, which are mainly influenced by soil traits and fertility, climatic seasonality and long term fluctuations as well as the frequency of fire events (Oliveira-Filho and Ratter 2002). The resulting vegetative formations range from open grasslands and shrublands to dense forests and can be comprehended to structurally coherent vegetation physiognomies. Large scale conversions from natural to agricultural land already affected around 40% of the Cerrado and land use models even suggest an increase in future deforestation rates (Ferreira et al. 2012). This rapid land cover change is one of the reasons why the Brazilian Cerrado was defined as one of the global biodiversity hotspot that needs to be under special focus in terms of assessment and conservation planning (Myers et al. 2000).

To keep track of the changes in the Cerrado, to support decision makers and to better understand ongoing processes, accurate mapping and monitoring is crucial (Sano et al. 2010). Even though field observations are inevitable for detailed local understanding of processes or for calibration and validation purposes (de Miranda et al. 2014), remote sensing data have been shown to be adequate for large scale assessments (Ferreira et al. 2004, Asner et al. 2005). Various Cerrado physiognomy mapping approaches using optical, radar or a combination of both types of remote sensing data have already been shown to be successful (Sano et al. 2005). However, these approaches are often time and labor intensive especially for wide biome assessments (Sano et al. 2010). Other studies used the temporal domain of frequently acquired remote sensing data like AVHRR or MODIS to make use of remote sensing derived phenologies (Ratana et al. 2005). Nevertheless, this approaches often lack spatial detail which is crucial for a detailed mapping of the complex vegetation heterogeneity in the Brazilian Cerrado. This is especially necessary for change detection approaches in order to assess environmental policy decisions and anthropogenic impacts (Jepson 2005). The opening of the Landsat archive enables to explore multi-temporal data from the past with a sufficient spatial resolution, while still covering large extents. Furthermore, the launch of the Landsat 8 and Sentinel-2 satellites assures continuous acquisitions of comparable data in the future. Even though Landsat data have a sufficient spatial resolution, its temporal resolution of 16 days (8 days when two sensors are combined) is not appropriate to derive continuous phenological profiles, especially when considering the low clear data availability due to cloud contamination (Sano et al. 2007).

To overcome these limiting factors, we created a phenological time series with Landsat spatial resolution, by approximating original Tasseled Cap transformed Landsat data comparable to the function fitting approach in TIMESAT (Jonsson and Eklundh 2004). Instead of applying a Savitzky-Golay or a double logistic function, we approximated our data with a weighted ensemble of Gaussian convolution filters. We tested the potential of the derived phenological profiles by using them to train a Support Vector Machine classification (Vapnik 1998), as this machine learning algorithm has been shown to be capable of solving complex non-linear classification problems (van der Linden et al. 2007). Thus our objectives are to analyze i) if it is possible to derive comprehensible seasonal phenological profiles from Landsat imagery using a Gaussian kernel convolution filter and ii) if these derived profiles allow us to distinguish the main Cerrado physiognomies in selected study areas around Brasília, DF.

2. Methodologies

The focus of this research is on a study area in central Brazil, covering Cerrado vegetation within the Distrito Federal (Figure 1). The selected sites include all the main Cerrado physiognomies and are located in protected areas.

2.1 Cerrado Physiognomies

As reference data we used two maps of protected areas around Brasília, DF that contain the main Cerrado physiognomies, classified by means of their vertical and horizontal vegetation density, as described by Ribeiro and Walter (Ribeiro and Walter 2008). The first map covers the whole extent of Parque Nacional de Brasília (PNB) with an area of approximately 30 000 ha. The map is based on a vegetation map and was refined in 2001 through visual interpretation of two Landsat scenes from 2001 (July and September), two high-resolution (4m spatial resolution) Ikonos scenes aerial photographs and field survey (Ferreira et al. 2007). It was manually up-dated in 2012 through visual interpretation of aerial photography from 2009[1]. It includes 13 classes (including physiognomy subclasses) which are representative for the Cerrado biome (Ferreira et al. 2007). The second map covers an area of approximately 4400 ha in the Reserva Ecológica do IBGE. It contains 20 classes, including subclasses of physiognomies. By using high resolution imagery and field data from the nearby Jardim Botânico de Brasília we added an area of 56 ha of Cerradão to the map mosaic, as this major Cerrado physiognomy was not covered in the other maps. Our final reference product ended up with 7 classes ranging from open grasslands to dense forests: Campo Limpo, Campo Sujo, Campo Cerrado, Cerrado Sensu Stricto, Cerrado Denso, Cerradao and Mata de Galeria (Figure 1).

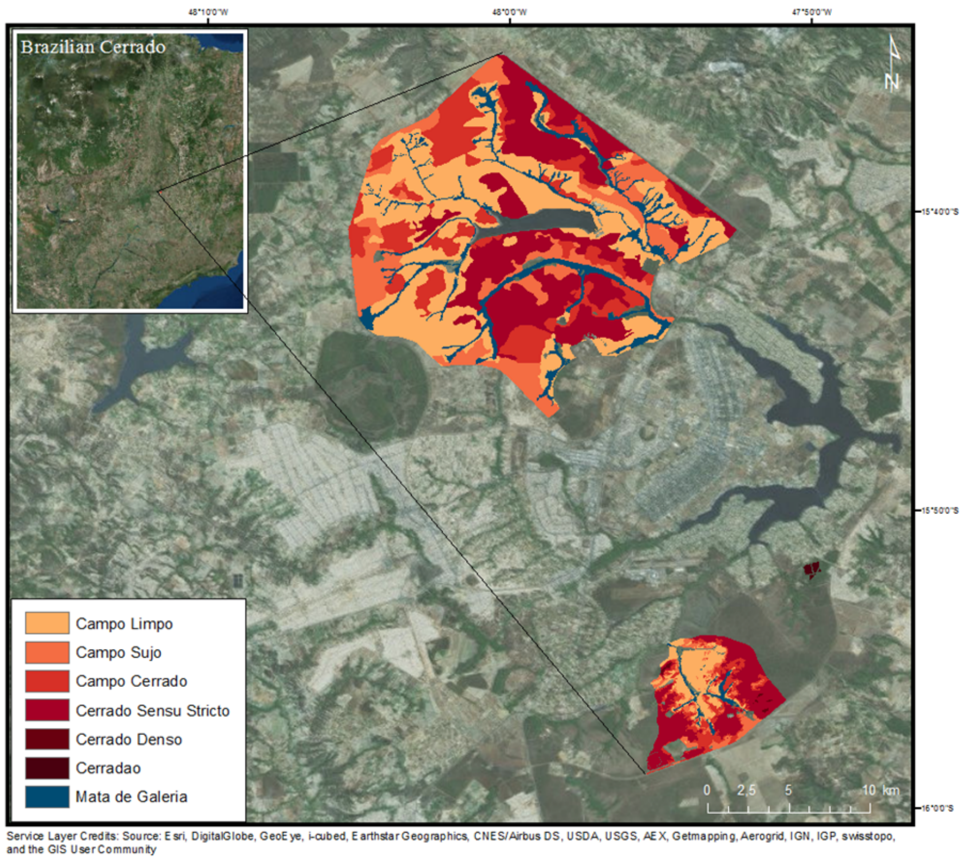

Figure 1. Study area covering Brazilian Cerrado physiognomies around Brasília, DF.

2.2 Phenological Landsat Time Series

In order to derive phenological profiles in the selected study sites we gathered all useful Landsat TM and ETM+ Level 1T corrected data (path: 221, row: 071) that have been recorded between January 2000 (DOY 007) and September 2014 (DOY 221), resulting in a total of 431 overpasses. The data were converted to top of atmosphere reflectance values and cloud masks were derived using the FMASK algorithm (Zhu and Woodcock 2012). The converted data were then transformed using the tasseled cap (TC) transformation coefficients for top of atmosphere reflectance, which are recommended for applications where

[1] Data from the Image Processing and GIS Laboratory (LAPIG), Federal University of Goiás (UFG). www.lapig.iesa.ufg.br

atmospheric correction approaches are not feasible (Huang et al. 2002). Furthermore, has the tasseled cap transformation been shown to be valuable for phenology based classification approaches (e.g. Dymond et al. 2002). For further analysis we chose the transformed TC greenness (TC_{green}) and wetness (TC_{wet}) values, as they are known to correlate with vegetation structure and vegetation/soil water content (Crist and Kauth 1986). As various factors like cloud contamination or sensor errors make it difficult to derive consistent phenological profiles from Landsat data, we used a function fitting approach to fill the data gaps in our timeseries. We used a Gaussian kernel based convolution filter approach to approximate the given clear observations into a dense 8-days-sampled timeseries. Gaussian kernel based convolution filtering is comparable to a weighted moving window average filter, where the weights are given by a Gaussian function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (1)$$

in which σ defines the smoothness or sharpness of the fitted function. For outlier (noise) detection in our observations we used a relatively smooth Gaussian kernel with σ = 5*8 days. The kernel includes all coefficients of the +/- 3* σ interval, i.e. an overall kernel size of 31 observations (15 to the left and 15 to the right), resulting in a kernel size of 31x8 = 248 in terms of days. The weights for each observation are much higher around the observed pixel and decrease towards the end of the kernel. Given this approximation, we basically follow the TIMESAT (Jonsson and Eklundh 2004) approach and eliminate all real observations that are substantially different from the approximation. The threshold for the difference from the approximation is measured in units of standard deviation of the timeseries. For our case study we chose one standard deviation as outlier detection threshold. Following the outlier elimination, we used an ensemble of multiple Gaussian kernel based convolution approximations to derive the final phenological profiles. We applied kernels with sigma=3x8, 5x8, 7x8 days. The final approximation is then a weighted average of the three individual approximations, where the weights depended on the original data density within the given kernels. This has the advantage that in sparse situations, the approximation of a smaller (i.e. sharper) kernel with lots of missing observations, can be corrected by a larger (i.e. smoother) kernel seeing potentially more observations. In that sense, our approach is adaptive to data availability.

2.4 Cerrado mapping

The derived TC_{green} and TC_{wet} profiles were truncated to represent one phenological season (July 2009 – June 2010), which corresponds to the date of the reference map actualization. We used these profiles to train a Support Vector Machine (Vapnik 1998) classification (SVC). Therefore 300 trainings pixel were randomly selected from each of the large classes in the physiognomy map (Campo Limpo, Campo Sujo, Campo Cerrado and Cerrado Sensu Stricto), whereat in the smaller classes (Cerrado Denso, Cerradão and Mata de Galeria) 25, 25 and 100 samples were randomly chosen. This sampling strategy resulted in a total of 1350 trainings pixel. A SVC model was trained using the sampled training pixel and its regularization parameters were optimized via a cross-validated grid search. Afterwards the model was applied to classify the complete study area of PNB and IBGE. All steps were performed with the EnMAP-Box Version 2.03 (Rabe et al. 2014) using the imageSVM implementation (Rabe et al. 2014) which is based on LIBSVM (Chang and Lin 2011). The accuracy of the result was validated by visual interpretation of 400 randomly stratified drawn pixel from the classification map and high resolution imagery in Google Earth. The results were comprehended in a confusion matrix and the accuracies adjusted with respect to the actual area they cover in the classification result (Congalton and Green 2008).

3. Results and Discussion

The application of the Gaussian convolution filter led to an approximation of phenological profiles (season 2009-2010) for each pixel. Based on the reference vegetation map we averaged the profiles for each physiognomie (Figure 2). The distinct profiles underline our assumption that the phenologies, as derived from remote sensing data, are physiognomy specific and differ from each other. The differences can be observed in the TC_{green} as well as in the TC_{wet} based phenological profiles. The profiles are comprehensible as they follow the typical seasonal climate trend, with high precipitations from October to April and very dry months from May to September (Ferreira et al. 2004). This goes along with findings in similar studies (Ratana et al. 2005). The trends show a green-up starting in October and a green peak between December and February, which differs in the distinct physiognomies. However, the TC_{wet} trend is much smoother and stretched over the whole rain season with a peak around April. It can be observed that the shape of the TC_{green} seasonal trend is quite similar in all physiognomies, but shifted to the beginning of the rain season in denser vegetation along with higher values. The TC_{wet} seasonal trend, which can be interpreted as proxy for water content, flattens out with higher vegetation density. Very dense vegetation structures like in the Cerradão or Mata de Galeria show highest TC_{wet} values with the smallest range in seasonality compared to the sparser physiognomies. These differences in the characteristics of both indices stress the added value of using them in combination as classification input.

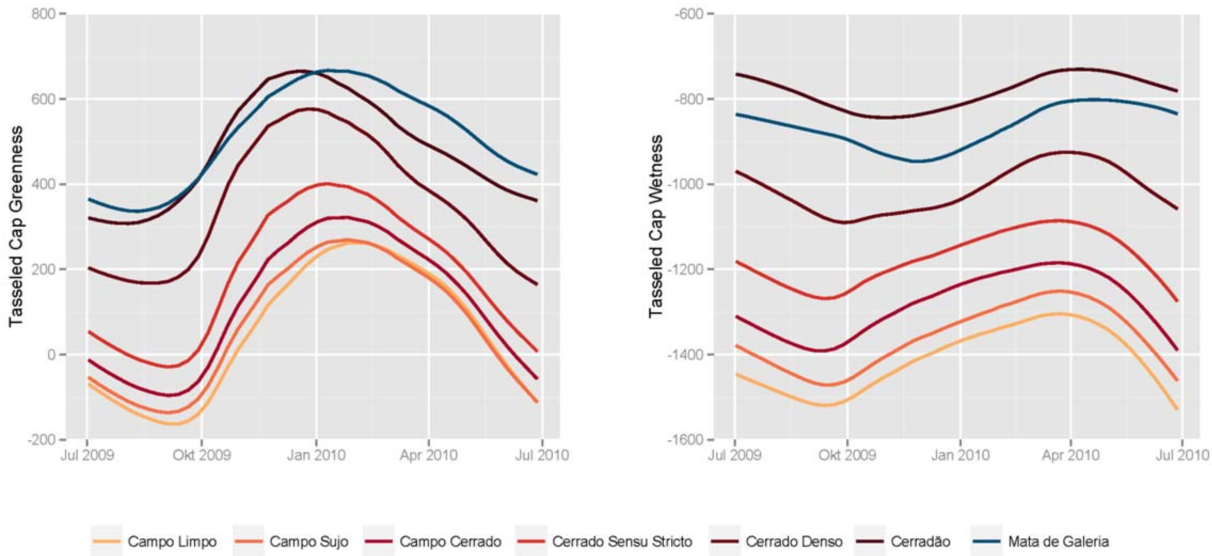

Figure 2. TC_{green} and TC_{wet} based phenological profiles, averaged for each physiognomie in the reference map.

The output of the SVM classification model is shown in Figure 3. It reveals the potential of utilizing the phenological profiles on Landsat spatial resolution for physiognomies distinction. All large scale spatial physiognomie patterns within the study areas could be clearly derived by their phenological profiles. Although the Gaussian kernel convolution filter was enabled to overcome data gaps, some artifacts of sensor errors, such as SLC-off regions of Landsat ETM+, remain in the final product. This might be due to very sparse data availability. Applying a convolution filter that also considers adjacent pixel values could help to solve this issue, but it would also generalize the final result and should therefore be decided on with respect to the application of the final map.

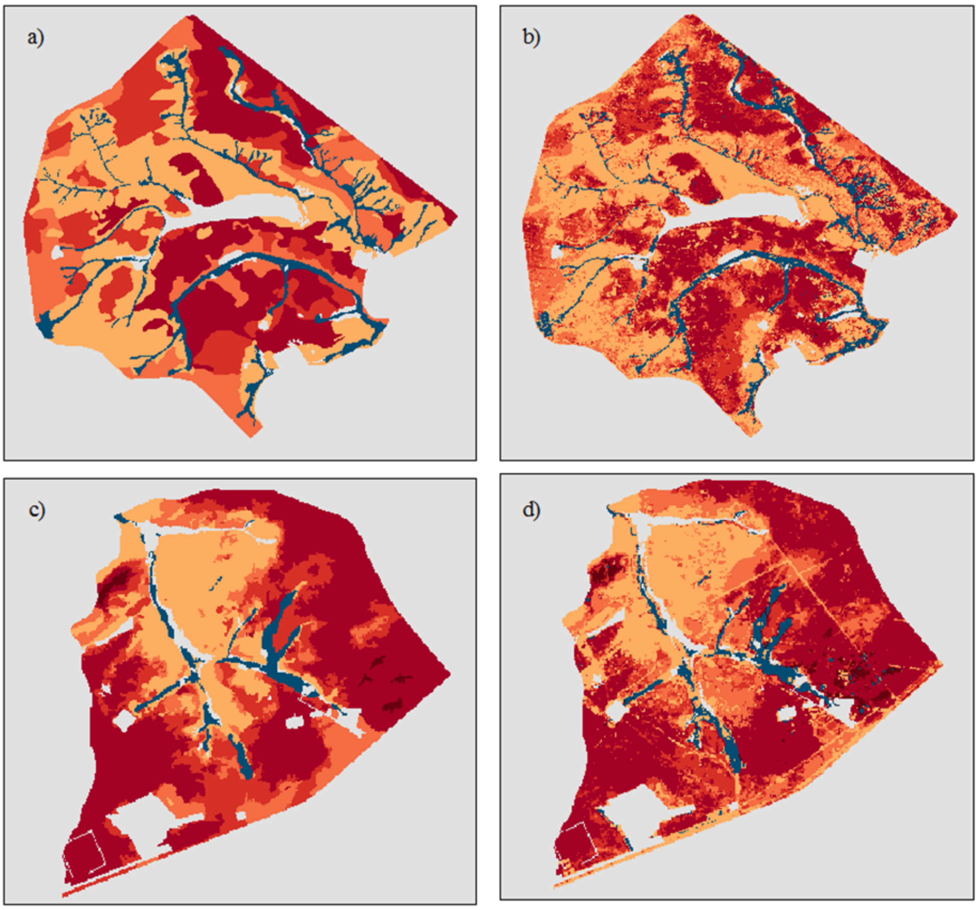

Figure 3. Reference and classified physiognomy maps of the study area: a) Physiognomy reference map PNB, b) Classification PNB, c) Physiognomy reference map IBGE, d) Classification IBGE. Grey areas are masked out because of a lack of reference data or classes that do not fit in the described physiognomie scheme.

A hard validation of the classification showed an overall accuracy of 61%, although most of the class confusions occurred between neighboring classes. However, as the defined classes mostly follow a vegetation density gradient (from Campo Limpo to Cerradão) it poses a challenge for describing it with hard classes. By applying a more tolerant validation approach (i.e. allowing pixels to be assigned to the adjacent classes within the vegetation gradient) we achieved an overall accuracy of 88%. Furthermore, the scale mismatch between the vector based reference map and a Landsat pixel based classification approach might induce errors in the model training. For example a Campo Limpo polygon (used to derive reference pixels for model training) could contain smaller patches of denser vegetation, which by definition belong to other physiognomies. Using pixel based reference data would further improve the proposed method.

4. Conclusion

We analyzed the potential use of Landsat derived phenological profiles to map the gradient of vegetation densities in the Brazilian Cerrado around Brasília, DF. It was possible to approximate typical phenological profiles for each Cerrado physiognomy by applying a Gaussian kernel based convolution filter on Tasseled Cap transformed Landsat data. Thus we were able to overcome most of the data gaps, which are due to sensor defects and cloud contaminated observations. The derived pixelwise profiles enabled us to train a Support Vector Classification model, which could be used to produce a map that captured the spatial patterns of the vegetation physiognomies in the observed study area. Highest errors were identified between structurally more similar classes. Although we achieved modest results these were mostly due to generalisation in the reference data. We could, however, highlight the potential of Landsat's spatial resolution in terms of mapping the complexity of the Cerrado vegetation density gradient, especially when taking the temporal domain into account. Future research should focus on the applicability of the proposed method in terms of transferability in the temporal domain, for example by deriving seasonal phenological metrics. Furthermore, should the spatial transferability of the approach be evaluated, as it might be usefull for frequent large scale Cerrado assessments. The open Landsat data archive, the new Landsat 8 and Sentinel sensors and the increasing trend in big data handling and processing performance facilitate these steps.

Acknowledgements

This research is part of the research activities of the EnMAP Science Advisory Group (EnSAG), and was funded by the German Aerospace Centre (DLR) - Project Management Agency, granted by the Ministry of Economics and Technology (BMWi; grant nr.50EE1309).

Special thanks to Manuel Eduardo Ferreira for providing the Parque Nacional de Brasília map and to Betânia Goes for the Reserva Ecológica do IBGE map.

References

Asner, G. P.; Knapp, D. E.; Cooper, A. N.; Bustamante, M. M. C.; Olander, L. P. Ecosystem structure throughout the Brazilian amazon from Landsat observations and automated spectral unmixing. **Earth Interactions,** v. 9, n., 2005.

Chang, C.-C.; Lin, C.-J. LIBSVM. **ACM Transactions on Intelligent Systems and Technology,** v. 2, n. 3, pp. 1–27, 2011.

Congalton, R. G.; Green, K. **Assessing the Accuracy of Remotely Sensed Data: Principles and Practices** Boca Raton, FL, 2008.

Crist, E. P.; Kauth, R. J. The Tasseled Cap demystified. **Photogrammetric Engineering and Remote Sensing,** v. 52, n. 1, pp. 81-86, 1986.

de Miranda, S. d. C.; Bustamante, M.; Palace, M.; Hagen, S.; Keller, M.; Ferreira, L. G. Regional Variations in Biomass Distribution in Brazilian Savanna Woodland. **Biotropica,** v. 46, n. 2, pp. 125-138, 2014.

Dymond, C. C.; Mladenoff, D. J.; Radeloff, V. C. Phenological differences in Tasseled Cap indices improve deciduous forest classification. **Remote Sensing of Environment,** v. 80, n. 3, pp. 460-472, 2002.

Ferreira, L. G.; Yoshioka, H.; Huete, A.; Sano, E. E. Optical characterization of the Brazilian Savanna physiognomies for improved land cover monitoring of the cerrado biome: preliminary assessments from an airborne campaign over an LBA core site. **Journal of Arid Environments,** v. 56, n. 3, pp. 425-447, 2004.

Ferreira, M. E.; Ferreira, L. G.; Miziara, F.; Soares-Filho, B. S. Modeling landscape dynamics in the central Brazilian savanna biome: future scenarios and perspectives for conservation. **Journal of Land Use Science,** v. 8, n. 4, pp. 403-421, 2012.

Ferreira, M. E.; Ferreira, L. G.; Sano, E. E.; Shimabukuro, Y. E. Spectral linear mixture modelling approaches for land cover mapping of tropical savanna areas in Brazil. **International Journal of Remote Sensing,** v. 28, n. 1-2, pp. 413-429, 2007.

Furley, P. A. The nature and diversity of neotropical savanna vegetation with particular reference to the Brazilian cerrados. **Global Ecology and Biogeography,** v. 8, n. 3-4, pp. 223-241, 1999.

Huang, C.; Wylie, B.; Yang, L.; Homer, C.; Zylstra, G. Derivation of a tasselled cap transformation based on Landsat 7 at-satellite reflectance. **International Journal of Remote Sensing,** v. 23, n. 8, pp. 1741-1748, 2002.

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

Jonsson, P.; Eklundh, L. TIMESAT - a program for analyzing time-series of satellite sensor data. **Computers & Geosciences,** v. 30, n. 8, pp. 833-845, 2004.

Myers, N.; Mittermeier, R. A.; Mittermeier, C. G.; da Fonseca, G. A. B.; Kent, J. Biodiversity hotspots for conservation priorities. **Nature,** v. 403, n. 6772, pp. 853-858, 2000.

Oliveira-Filho, A. T.; Ratter, J. A. Vegetation Physiognomies and Woody Flora of the Cerrado Biome. In: P. E. M. Oliveira and R. J. Marquis **The Cerrados of Brazil - Ecology and Natural History of a Neotropical Savanna**. New York: Columbia University Press, 2002.

Rabe, A.; Jakimow, B.; Held, M.; Van der Linden, S.; Hostert, P. **EnMAP-Box Version 2.03**, software available at www.enmap.org. v., 2014.

Rabe, A.; van der Linden, S.; Hostert, P. **imageSVM, Version 3.0**, software availbale at http://imagesvm.net/. v., 2014.

Ratana, P.; Huete, A. R.; Ferreira, L. Analysis of cerrado physiognomies and conversion in the MODIS seasonal-temporal domain. **Earth Interactions,** v. 9, n., 2005.

Ratter, J. A.; Ribeiro, J. F.; Bridgewater, S. The Brazilian Cerrado Vegetation and Threats to its Biodiversity. **Annals of Botany,** v. 80, n. 3, pp. 223-230, 1997.

Ribeiro, J. F.; Walter, B. M. T. As principais fitofisionomias do bioma Cerrado. In: S. M. Sano, S. P. Almeida and J. F. Ribeiro **Cerrado: Ecologia e flora**. Brasilia, DF: Embrapa Cerrados, 2008. 153-212.

Sano, E. E.; Ferreira, L. G.; Asner, G. P.; Steinke, E. T. Spatial and temporal probabilities of obtaining cloud-free Landsat images over the Brazilian tropical savanna. **International Journal of Remote Sensing,** v. 28, n. 12, pp. 2739-2752, 2007.

Sano, E. E.; Ferreira, L. G.; Huete, A. R. Synthetic aperture radar (L band) and optical vegetation indices for discriminating the Brazilian savanna physiognomies: A comparative analysis. **Earth Interactions,** v. 9, n. p. 15, 2005.

Sano, E. E.; Rosa, R.; Brito, J. L.; Ferreira, L. G. Land cover mapping of the tropical savanna region in Brazil. **Environ Monit Assess,** v. 166, n. 1-4, pp. 113-124, 2010.

van der Linden, S.; Janz, A.; Waske, B.; Eiden, M.; Hostert, P. Classifying segmented hyperspectral data from a heterogeneous urban environment using support vector machines. **Journal of Applied Remote Sensing,** v. 1, n., 2007.

Vapnik, V. **Statistical Learning Theory**. New York, 1998.

Zhu, Z.; Woodcock, C. E. Object-based cloud and cloud shadow detection in Landsat imagery. **Remote Sensing of Environment,** v. 118, n. pp. 83-94, 2012.