



EDUARDA MARTINIANO DE OLIVEIRA SILVEIRA

**THE GEOSTATISTICAL CONTEXT
EMPLOYED IN REMOTE SENSING
APPLICATIONS: IMAGE CLASSIFICATION,
CHANGE DETECTION AND FOREST
INVENTORY**

LAVRAS – MG

2018

EDUARDA MARTINIANO DE OLIVEIRA SILVEIRA

**THE GEOSTATISTICAL CONTEXT EMPLOYED IN REMOTE
SENSING APPLICATIONS: IMAGE CLASSIFICATION, CHANGE
DETECTION AND FOREST INVENTORY**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Ciências Florestais, para a obtenção do título de Doutor

Prof. Dr. José Márcio de Mello

Orientador

Prof. Dr. Fausto Weimar Acerbi Júnior

Coorientador

LAVRAS – MG

2018

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Silveira, Eduarda Martiniano de Oliveira.

The geostatistical context employed in remote sensing applications: image classification, change detection and forest inventory / Eduarda Martiniano de Oliveira Silveira. - 2018.

270 p. : il.

Orientador: José Márcio de Mello.

Coorientador: Fausto Weimar Acerbi-Júnior.

Tese (doutorado) - Universidade Federal de Lavras, 2018.

Bibliografia.

1. Geostatistical context. 2. Semivariogram. 3. Remote sensing. I. Mello, José Márcio de. II. Acerbi-Júnior, Fausto Weimar. III. Título.

EDUARDA MARTINIANO DE OLIVEIRA SILVEIRA

**THE GEOSTATISTICAL CONTEXT EMPLOYED IN REMOTE
SENSING APPLICATIONS: IMAGE CLASSIFICATION, CHANGE
DETECTION AND FOREST INVENTORY**

**O CONTEXTO GEOESTATÍSTICO EMPREGADO NAS APLICAÇÕES
DE SENSORIAMENTO REMOTO: CLASSIFICAÇÃO DE IMAGENS,
DETECÇÃO DE MUDANÇAS E INVENTÁRIO FLORESTAL**

Tese apresentada à Universidade Federal
de Lavras, como parte das exigências do
Programa de Pós-Graduação em
Engenharia Florestal, área de
concentração em Ciências Florestais, para
a obtenção do título de Doutor

APROVADA em 08 de março de 2018.

Dr. Alan de Brito	FUNDEP/INPE
Prof. Dr. Sérgio Henrique Godinho	UFLA
Dr. Marcela Castro Nunes Santos Terra	UFLA
Prof. Dr. Luis Marcelo Tavares de Carvalho	UFLA

Prof. Dr. José Márcio de Mello
Orientador

Prof. Dr. Fausto Weimar Acerbi Júnior
Coorientador

LAVRAS – MG

2018

*To my love Faustinho, my family, and
my true friends...*

ACKNOWLEDGEMENTS

I would like to thank the following people for their contributions of academic and technical support on this Thesis:

José Márcio de Mello

Fausto Weimar Acerbi Júnior

Fernando Del Bon Espírito-Santo

Michael A. Wulder

Kieran Daniel Withey

Luis Angel Ruiz

Michele Duarte Menezes

Luis Marcelo Tavares de Carvalho

Carlos Rogério de Mello

Mônica Canaan Carvalho

Inácio Tomaz Bueno

Marcela Castro Nunes Santos Terra

Aliny Aparecida dos Reis

I would also like to thank the Universidade Federal de Lavras (UFLA), the Programa de Pós-Graduação em Engenharia Florestal and the Laboratório de Manejo Florestal (LEMAF) for providing the research opportunities. The Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fapemig for funding my studies, including the time abroad at at Lancaster Environmental Centre (LEC), Lancaster University (UK).

RESUMO GERAL

O contexto espacial foi empregado nas aplicações de sensoriamento remoto: classificação de imagens, detecção de mudanças e inventário florestal. Primeiramente, os parâmetros do semivariograma derivados de imagens de satélite com diferentes resoluções espaciais foram avaliados para caracterizar a heterogeneidade espacial da cobertura de solo. No estudo referente a classificação de imagens, o objetivo foi avaliar o potencial de atributos geoestatísticos em uma classificação orientada a objetos para melhorar o resultado da classificação de imagens. No estudo de detecção de mudanças foi avaliado a performance dos índices do semivariograma derivados de imagens NDVI Landsat, utilizando análise orientada a objetos, para detectar com acurácia mudanças na cobertura do solo, desconsiderando as mudanças associadas aos efeitos fenológicos da vegetação. No artigo aplicado ao inventário florestal, foi investigado o potencial de dados extraídos de imagens Landsat, MODIS e variáveis espaço-temporais para mapear a distribuição espacial da biomassa aérea no Estado de Minas Gerais, utilizando o algoritmo Random forest e krigagem regressiva através da amostragem estratificado. Os resultados demonstraram que: (1) a resolução espacial das imagens influencia os parâmetros alcance e patamar do semivariograma, que podem ser utilizados como simples indicadores da heterogeneidade da cobertura do solo; (2) os semivariogramas foram eficientes para caracterizar a heterogeneidade da cobertura do solo, aumentando significativamente a acurácia da classificação de imagens combinando atributos geoestatísticos com dados espectrais; (3) os atributos geoestatísticos apresentam potencial para distinguir classes homogêneas e heterogêneas, não são afetados pela sazonalidade da vegetação, e podem produzir séries temporais que minimizam os efeitos da fenologia da vegetação, e (4) a estratificação em fitofisionomias não permite apenas a estimativa da biomassa aérea da vegetação com alta precisão, mas também permite que o algoritmo Random forest selecione o menor número de variáveis que resulte em um modelo com menor erro. O contexto espacial apresentado nesta tese é novo e aplicado para a classificação de imagens que contenha fitofisionomias espectralmente semelhantes; para a detecção de mudanças em áreas onde a vegetação apresenta alta sazonalidade, e o mapa detalhado de biomassa aérea da vegetação e o entendimento que como as características das variáveis estão associadas com a modelagem de cada fitofisionomia, permite pesquisadores melhorarem as estimativas grosseiras de emissão de gases de efeito estufa, auxiliando a seleção das variáveis mais apropriadas para modelar a biomassa em áreas de transição entre o Cerrado e Florestas.

Palavras-chave: Contexto geoestatístico. Semivariograma. Sensoriamento remoto.

GENERAL ABSTRACT

We used the spatial context, specifically the geostatistical techniques to improve remote sensing applications: image classification, change detection and forest inventory. We first evaluate the potential of semivariogram parameters, derived from satellite images with different spatial resolutions to characterize landscape spatial heterogeneity. In the image classification study, the goal was to assess the potential of geostatistical features at the object level to improve the image classification of contrasted landscape vegetation cover. In the change detection approaches we explored and evaluated the performance of semivariogram indices in an object-based approach to detecting land-cover changes using the NDVI derived from Landsat images using the support vector machines and random forest algorithms. We assessed the potential of geostatistical features to accurately detect land-cover changes, disregarding those associated with phenological differences. In the forest inventory manuscript, we investigated the potential of data extracted from Landsat TM, MODIS products and spatial-environmental variables to map the spatial distribution of aboveground biomass in Minas Gerais State, using random forest regression algorithm and regression kriging technique using a stratified design. The applications results indicate that: (1) image spatial resolution does in fact influence the sill and range parameters, that can be used as a simple indicator of landscape heterogeneity; (2) semivariogram curves were efficient for characterizing spatial heterogeneity, significantly improving the image classification accuracy when combining geostatistical features with spectral data; (3) geostatistical features have the potential to discriminate between homogeneous and heterogeneous classes within objects, are not affected by vegetation seasonality, and can produce times series that accurately differentiate forest changes from seasonal changes, resulting in fewer classification errors and (4) the stratification of data into vegetation types not only improved the accuracy of aboveground biomass estimative, but also allowed random forest regression to select the lowest number of variables that offer the best predictive model performance to AGB mapping. The spatial context approach we presented in this thesis is a novel and useful remote sensing method for the image classification of spectrally similar land-cover types, detection of forest change events in areas where forests exhibit strong seasonality, and therefining aboveground biomass map and the understanding of how the variables properties are associated with the biomass enable researches to improve the roughly estimates of greenhouse gas emission and also helps the selection of appropriate variables that best model the aboveground biomass in savanna-forest transition areas.

Keywords: Geostatistical context. Semivariogram. Remote sensing.

SUMÁRIO

FIRST PART	9
1 INTRODUCTION	9
2 LITERATURE REVIEW	15
2.1 Image classification using the spatial context	19
2.2 Change detection using the spatial context	22
2.3 Aboveground biomass modelling using remote sensing	24
3 FINAL CONSIDERATIONS	29
REFERENCES	31
SECOND PART – ARTICLES	31
ARTICLE 1 - CHARACTERIZING LANDSCAPE SPATIAL HETEROGENEITY USING SEMIVARIOGRAM PARAMETERS DERIVED FROM NDVI IMAGES	37
ARTICLE 2 - ASSESSMENT OF GEOSTATISTICAL FEATURES FOR OBJECT-BASED IMAGE CLASSIFICATION OF CONTRASTED LANDSCAPE VEGETATION COVER	63
ARTICLE 3 - OBJECT-BASED CHANGE DETECTION USING SEMIVARIOGRAM INDICES DERIVED FROM NDVI IMAGES: THE ENVIRONMENTAL DISASTER IN MARIANA, BRAZIL	93
ARTICLE 4 - OBJECT-BASED LAND USE/COVER CHANGE DETECTION APPLIED TO BRAZILIAN SEASONAL SAVANNAS USING GEOSTATISTICAL FEATURES	117
ARTICLE 5 - DISENTANGLING THE EFFECTS OF FOREST PHENOLOGY AND LAND-COVER CHANGES IN BRAZILIAN SEASONAL BIOMES COMBINING SPATIAL AND SPECTRAL REMOTE SENSING FEATURES	159
ARTICLE 6 - USING SPATIAL FEATURES TO REDUCE IMPACT OF SEASONALITY FOR DETECTING FOREST CHANGE FROM LANDSAT TIME SERIES	191
ARTICLE 7 - STRATIFICATION IMPROVEMENTS IN ABOVEGROUND BIOMASS MODELLING OF SAVANNA-FOREST TRANSITION DRIVEN BY SATELLITE IMAGES AND SPATIO-ENVIRONMENTAL DATA	225

FIRST PART

1 INTRODUCTION

The Earth's surface and remotely sensed images of that surface have distinct spatial properties. Once quantified, these properties can be used for many tasks in remote sensing, including image classification, change detection and forest inventory estimates. Many methods have potential for the quantification of these spatial properties (CURRAN, 1988). Spatial feature extraction algorithms are classified into four categories: (1) statistical (grey-level co-occurrence matrix – GLCM and semivariogram analysis), (2) geometrical (primitive textures), (3) model-based (fractal dimension), and (4) signal processing (Fourier transformation and wavelet transformation). A semivariogram, that is the primary tool of geostatistics, provides a concise and unbiased description of the scale and pattern of spatial variability and can be used to investigate and quantify spatial variability. This spatial variability is the key to improve remote sensing applications.

The integration between remote sensing and geostatistical theory was consolidated in the late 1980s by various groups that used semivariogram analysis to quantify image structure (CURRAN, 1988; WOODCOCK; STRAHLER; JUPP, 1988a, 1988b). Many studies have demonstrated that the semivariogram has strong potential for heterogeneity analysis (BALAGUER-BESER et al., 2013; GARRIGUES et al., 2006, 2008; SILVEIRA et al., 2017a), image classification of remotely sensed data (ATKINSON; LEWIS, 2000; BALAGUER et al., 2010; CHICA-OLMO; ABARCA-HERNÁNDEZ, 2000; SILVEIRA et al., 2017b; WU et al., 2015; YUE et al., 2013), change detection (ACERBI JUNIOR et al., 2015; GIL-YEPES et al., 2016; SILVEIRA et al., 2017c, 2018) and to improve forest inventory estimates (SCOLFORO et al.,

2015). Therefore, the main objective of this thesis was to indicate how the geostatistical context could be employed in remote sensing applications. The hypotheses were:

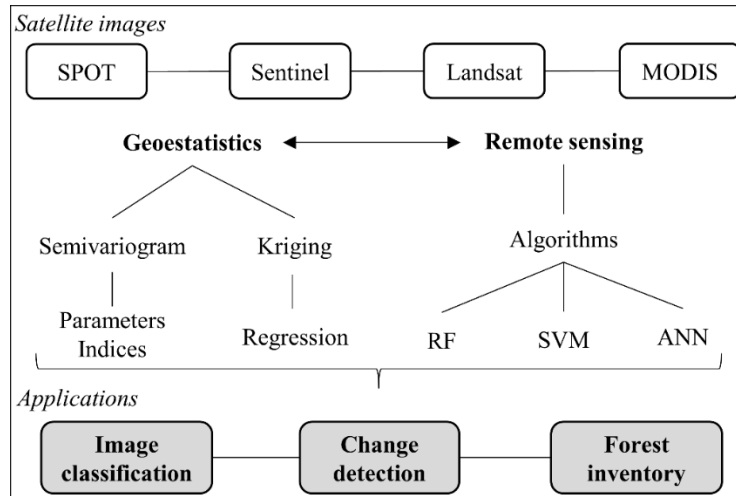
- 1) There is a relationship between landscape heterogeneity and measures of spatial dependence from remotely sensed data;
- 2) The NDVI (normalized difference vegetation index) and NIR (near-infrared channel) spatial variability provided by semivariogram features from different landscape vegetation cover could be used as input data to train a RF (random forest) algorithm, thereby improving object-based image classification (indicate image classification application).
- 3) Land-use/cover changes could be accurately detected using geostatistical features, such as semivariogram indices and parameters calculated from NDVI images using an object-based approach (change detection application).
- 4) The geostatistical features obtained at the object level are not affected by vegetation seasonality as the NDVI variability captured by these indices remains constant in the presence of seasonal changes and substantially increases in the presence of forest changes (change detection application).
- 5) Stratification of inventory plots into forest types could improve the precision of aboveground biomass (AGB) estimates using regression kriging technique and remote-sensing/spatial-environmental variables.

Thus, we organized the thesis in seven articles (TABLE 1). In the **first** we evaluated the potential of semivariogram parameters, derived from satellite

images with different spatial resolutions, to characterize landscape spatial heterogeneity of forested and human modified areas. In the **second** we studied the potential of geostatistical features, derived from medium spatial resolution satellite imagery, to characterize contrasted landscape vegetation cover and improve object-based image classification. In the **third** we explored and evaluated the performance of semivariogram indices in an object-based approach to detect land-cover changes caused by the 2015 dam-collapse disaster in Brazil. In the **fourth** and **fifth** articles we utilized NDVI derived from bitemporal Landsat images obtained during the wet and dry seasons to assess the potential of individual geostatistical features and their combination with spectral features to accurately detect forest changes, respectively. In the **sixth** we proposed a new method to reduce impact of seasonality for detecting forest change from Landsat time series using spatial indices, and in the **seventh** we investigated the potential of data extracted from Landsat TM, MODIS products and spatio-environmental variables to map the spatial distribution of aboveground biomass (ABG) of six heterogeneous vegetation types in Atlantic Forest, Savanna, and Semi-arid woodland Biomes in Minas Gerais State, Brazil.

The integration between geostatistics and remote sensing (FIGURE 1) is presented in the Literature Review Section and the techniques used are presented in the articles. The main research questions are presented in Table 2.

Figure 1 - Methodology workflow. RF-random forest; SVM-support vector machines; ANN-artificial neural network.



Source: Author (2018).

Table 1 – Articles comprised in this Thesis.

N°	Title	Application
1	Characterizing landscape spatial heterogeneity using semivariogram parameters derived from NDVI images	Landscape heterogeneity
2	Assessment of geostatistical features for object-based image classification of contrasted landscape vegetation cover	Image classification
3	Object-based change detection using semivariogram indices derived from NDVI images: The environmental disaster in Mariana, Brazil	Change detection
4	Object-based land cover change detection applied to Brazilian seasonal Savannas using geostatistical features	Change detection
5	Disentangling the Effects of Forest Phenology and Land-Use Changes in Brazilian Seasonal Biomes Combining Spatial and Spectral Remote Sensing Features	Change detection
6	Using spatial features to reduce impact of seasonality for detecting forest change from Landsat time series	Change detection
7	Stratification improves aboveground biomass modeling of Savanna-Forest transition driven by satellite images and spatio-environmental data	Forest Inventory

Source: Author (2018).

Table 2 – Main research questions. AGB-aboveground biomass; NDVI - Normalized Difference Vegetation Index

Questions	Article
How is image spatial resolution linked with the spatial variability of NDVI values? Which semivariogram parameter and image spatial resolution is the most appropriate to produce a landscape heterogeneity map?	1
Combining geostatistical features with spectral data improves the object-based image classification procedure?	2
Are the semivariogram indices derived from Landsat NDVI images able to discriminate between homogeneous and heterogeneous pixels within objects?	3
Do changes caused by vegetation seasonality affect the spatial variability of NDVI values? Are geostatistical features derived from semivariograms able to accurately detect land-cover changes?	4
How well can we differentiate between seasonal and land-cover changes combining spatial and spectral features derived from NDVI Landsat images?	5
Are the geostatistical features efficient to eliminate seasonal variations in satellite images time series?	6
How does the stratification into vegetation types improves the predictive quality of AGB random forest model? How do the remote sensing and spatio environmental variables are associated with the vegetation types?	7

Source: Author (2018).

2 LITERATURE REVIEW

This section reviews geostatistical integration with remote sensing applications: image classification, change detection and forest inventory estimate (specifically aboveground biomass-AGB). Since the applications described depend on the semivariogram (the geostatistical tool), it is introduced here briefly.

The theoretical conception of geostatistics is the Theory of Regionalized Variables, proposed by Matheron in 1965, inspired by Wijs (1951) and Krige (1951). Matheron (1963) defined Regionalized Variable as a numerical spatial function, which varies from one place to another, with an apparent continuity and whose variation cannot be represented by a simple mathematical function, but by a variogram. According to this theory, the difference in the values of a given variable taken at two points in the field depends on the distance between them. Thus, the difference between values of an attribute taken at two points closest to the space must be smaller than the difference between values taken at two distant points. Therefore, each value carries with it a strong interference of the values of its neighbourhood, illustrating a spatial continuity (ISAACS; SRIVASTAVA, 1988).

Some hypotheses of stationarity must be assumed for the characterization of a random variable along the space. A variable is stationary if the development of the process in time or space occurs homogeneously, with continuous random oscillations around an average value, in which both the mean amplitude and the oscillations change sharply in time or space. The most common are the second-order stationarity and Intrinsic Hypothesis. If the mathematical expectation of a random variable is constant, regardless of the origin adopted in space or time, it can be said that the variable is stationary of

first order and therefore the average will be the same for the whole process according to the intrinsic hypothesis. In the intrinsic hypothesis we have that:

$$E[Z(x)] = m$$

$$\text{Var}[Z(x+h)-Z(x)] = E[Z(x+h)-Z(x)]^2 = 2\gamma(h)$$

The key to the theory of regionalized variables is the semivariogram. This function relates semivariance to spatial separation and provides a concise and unbiased description of the scale and pattern of spatial variability. For continuous variables, such as reflectance in a given spectral band or vegetation indices, the experimental semivariogram is defined as half of the average squared difference between values separated by a given lag, where this lag is a vector in both distance and direction (ATKINSON; LEWIS, 2000). The experimental semivariogram is:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x) - Z(x+h)]^2$$

where $z(x)$ represents the value of the variable at the location x , h the separation between elements in a given direction, and $N(h)$ the number of data pairs occurring at locations x and $x + h$.

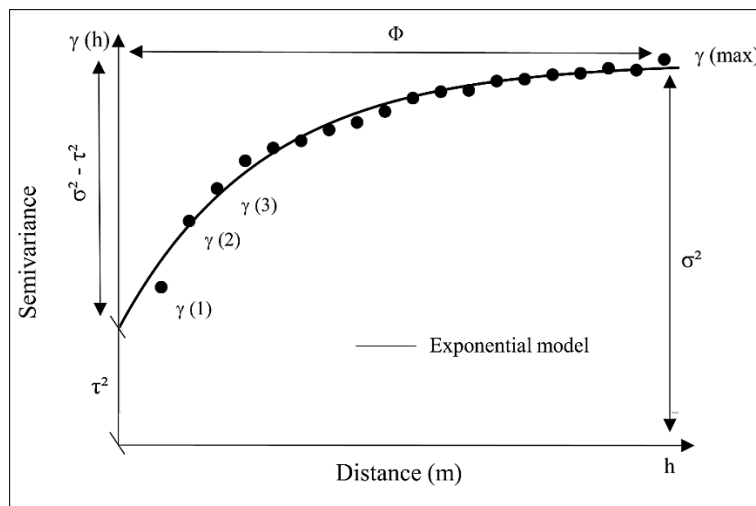
The semivariance functions are characterized by three parameters: sill (σ^2), range (Φ), and nugget effect (τ^2). The sill parameter is the plateau reached by semivariance values and shows the quantity of variation explained by the spatial structure of the data. The range parameter is the distance where the semivariogram reaches the sill, showing the distance until the data are correlated. The nugget effect is the combination of sampling errors and

variations that happen in scales smaller than the distance between the sampled points (CURRAN, 1988).

The spatial features provided by the semivariogram can be achieved by (FIGURE 2):

- Modelling the semivariogram by fitting a model to extract semivariogram parameters sill (σ^2), range (Φ), and nugget effect (τ^2);
- Using the semivariance at various lags $\gamma(1)$, $\gamma(2)$, $\gamma(3)$ $\gamma(\max)$. It is free of the problems caused by modeling and becomes more popular for describing spatial properties of remote sensing image;
- Using indices that synthesize the most relevant information about the shape of the semivariogram and enhance the spatial information, i.e. RVF (Ratio between the values of the total variance and the semivariance at first lag) index, that is given by $\frac{\gamma(\max)}{\gamma(1)}$.

Figure 2 - Example of a classic semivariogram.



Source: Author (2018).

Table 3 indicates the terms used in the description of the semivariogram. To obtain both the semivariogram parameters, raw semivariance values and semivariogram indices from the images, we can use both pixel-based and object-based approaches. When using the pixel-based approach it is necessary to define a moving window, that is easy to implement, however, it is computational expensive (ZHU, 2017), biased along their diagonals and will likely straddle the boundary between two landscape features, especially when a large window size is used (LALIBERTE; RANGO, 2009). When using the object-based approach, these measures are obtained within each individual object, and it is necessary to define the lag distance and the number of lags to generate the semivariogram. According to Woodcock, Strahler and Jupp (1988a), the lag distance needs to be larger than the range of influence in order to characterize the initial part of the semivariogram, and large enough to reveal the presence of periodicity.

Table 3- Terms and symbols of the semivariogram curve.

Term	Symbol	Definition
Lag	h	Separation distance between sample pairs
Sill	σ^2	Maximum level of $\gamma(h)$
Range	ϕ	Maximum distance of spatial dependency
Nugget effect	τ^2	Independent variance
Partial sill	$\sigma^2 - \tau^2$	Sill minus Nugget effect

Source: Curran (1988).

Ruiz et al. (2011) presented and described a software application for automatic extract descriptive feature from image-objects, FETEX 2.0. The input data include a multispectral digital image or single band, such as NDVI and a vector file in shapefile format containing the objects. The output file is a table that can be produced in four alternative formats, containing a vector of features for every object processed. This table of numeric values describing the objects from different points of view can be externally used as input data for any

classification software. Additionally, several types of graphs and images describing the feature extraction procedure are produced, useful for interpretation and understanding the process. A test of the processing times is included, as well as an application of the program in a real parcel-based classification problem, providing some results and analysing the applicability, the future improvement of the methodologies, and the use of additional types of data sets. This software is intended to be a dynamic tool, integrating further data and feature extraction algorithms for the progressive improvement of the integration between geostatistics and remote sensing applications.

2.1 Image classification using the spatial context

In general, classification accuracies are improved by the use of the spatial context. Most classical mathematical algorithms for image classification do not consider the spectral dependence existing between a pixel and its neighbours, i.e the spatial autocorrelation. In this way, results obtained by using both spectral and spatial information can improve image classification results. This improvement arises the hypothesis that a pixel is not independent of its neighbours and furthermore, this dependence can be quantified and incorporated into the classifiers. Figure 4 indicates three different landscape vegetation cover (cerrado sensu stricto, deciduous forest and palm swamps) that present different spatial variability provided by sill parameter (σ^2) computed by NIR (near infrared) band. In this example, the σ^2 semivariogram parameter can be used as spatial feature to minimize the inter-class confusion and improve the discrimination among these classes (note the sill values of each class).

The use of semivariogram in image classification was made popular by Miranda, Fonseca and Carr (1998) and Miranda, Macdonald and Carr (1992) by using semivariogram textural classifiers (STC) using a pixel-based approach to

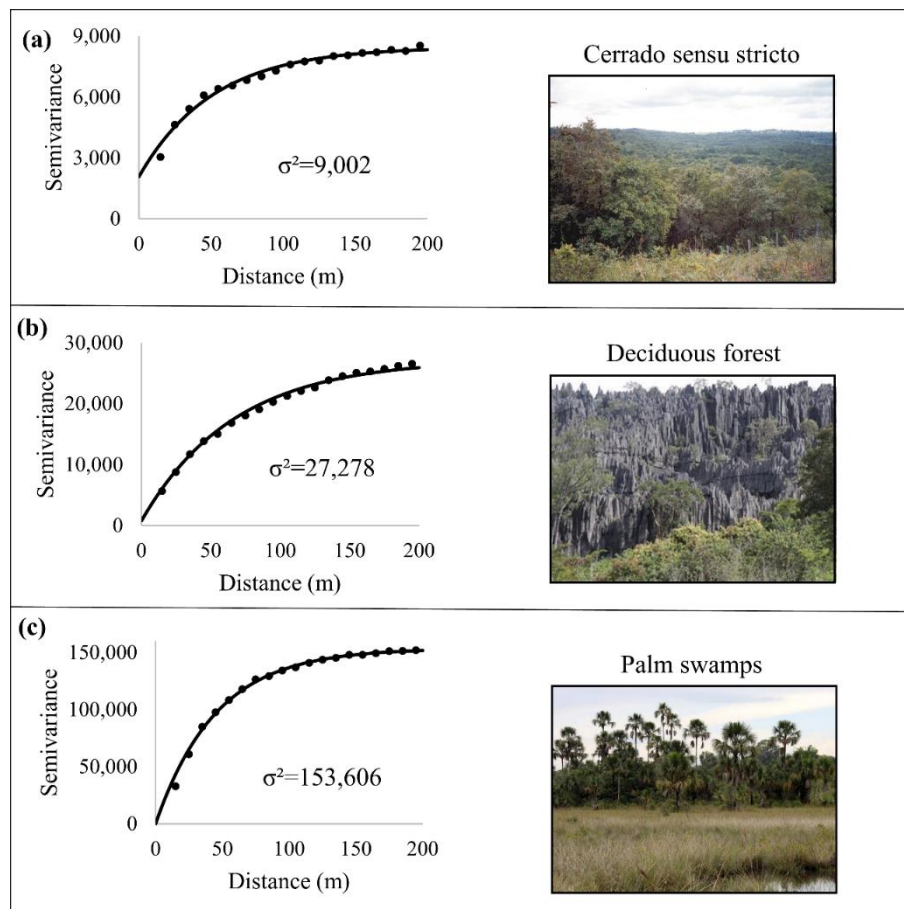
improve the discrimination between land cover types in Borneo and the Amazon Forest, Brazil. Chica-Olmo and Abarca-Hernandez (2000), used a set of textural measures within a moving window in a Landsat TM image to resolve inter-class confusion of indistinct landcover categories. Their results provided significant increase in accuracy compared to standard methodologies.

Balaguer et al. (2010) proposed a comprehensive set of texture features extracted from the experimental semivariogram of specific image objects to improve image classification of high resolution images. Fourteen features were defined and categorized into three different groups, according to the location of their respective parameters in the semivariogram curve. The suitability of the proposed features for object-based image classification has been evaluated using digital aerial images from an agricultural area on the Mediterranean coast of Spain. The performance of the selected semivariogram features has been compared with two different sets of texture features: those derived from the grey level co-occurrence matrix (GLCM), and the values of raw semivariance directly extracted from the semivariogram at different positions. As a result of the tests, the classification accuracies obtained using the proposed semivariogram features are, in general, higher than those obtained using the other two sets of standard texture features.

Wu et al. (2015) studied an efficient use of semivariogram features for object-based high-resolution image classification. First, an input image is divided into segments, for each of which a semivariogram is then calculated. Second, candidate features are extracted as a number of key locations of the semivariogram functions. Then they used an improved Relief algorithm and principal component analysis to select independent and significant features. Then the selected prominent semivariogram features and the conventional spectral features are combined to constitute a feature vector for a support vector machine classifier. The effect of such selected semivariogram features is

compared with those of the GLCM features and window-based semivariogram texture features. Tests with aerial and satellite images show that such selected semivariogram features are of a more beneficial supplement to spectral features.

Figure 3 - Discrimination among (a) cerrado sensu stricto; (b) deciduous forest and (c) palm swamps vegetation cover by using the sill parameter derived from NIR band obtained by Sentinel 2 images.



Source: Author (2018).

Silveira et al. (2017b) tested the potential of geostatistical features, the semivariogram indices proposed by Balaguer et al. (2010), derived from medium

spatial resolution satellite imagery to characterize contrasted landscape vegetation cover and improve object-based image classification. The study site in Brazil included cerrado sensu stricto, deciduous forest, and palm swamp vegetation cover. Sentinel 2 and Landsat 8 images were used to test the potential of the indices derived from near-infrared (NIR) and normalized difference vegetation index (NDVI). The results showed that semivariogram curves were efficient to characterize the spatial heterogeneity of the classes analysed, with similar results using NIR and NDVI from Sentinel 2 and Landsat 8. Accuracy was significantly greater when combining geostatistical features with spectral data, suggesting that this method can improve image classification results.

2.2 Change detection using the spatial context

Several authors have used the spatial context to extract forest damage and detect land-use/cover (LULC) changes. Sertel, Kaya and Curran (2007) investigated the relationship between semivariogram metrics and degree of earthquake damage. They concluded that in severely damaged areas, the earthquake caused large spatial variations that were quantified by the semivariogram variables of range, nugget, and sill. The range captured coarse-scale spatial variability, the nugget captured fine-scale spatial variability, and the sill captured overall variability in the landscape.

Costantini et al. (2012) tested the hypothesis of existence of a relationship between the vegetation spatial heterogeneity and the occurrence/spread of disturbance events, integrating the change detection and spatial analyses of the NDVI in forested areas. They combined techniques of remote sensing, geostatistics and landscape ecology in mixed forests to propose the range of semivariogram as a simple indirect estimator of this susceptibility of forests.

Acerbi Junior et al. (2015) analysed the potential of semivariograms generated from the NDVI values, derived from Landsat TM images, to detect changes in an area covered by Brazilian savanna's vegetation. This study had demonstrated the usefulness of semivariogram shape and parameters to detect deforestation. In deforested areas, the landscape change has caused spatial variations that were quantified by the semivariogram parameters of range, sill and shape. The range and sill were the two most important and complementary parameters. Both increased their values after deforestation and remain similar if the land cover had not been changed.

Gil-Yepes et al. (2016) proposed and evaluated a set of new temporal features based on geostatistics functions that relate the spatial distribution of the elements within agricultural parcels at two different dates. They concluded that the new set of features extracted from cross-semivariogram are suitable change indicators when changes affect the heterogeneity of the elements within objects, having in this cases a high discrimination power between change and no change class. Furthermore, by combining spectral features with cross-semivariogram, higher accuracies were achieved and confusion between classes was reduced.

Hamunyela, Verbesselt and Herold (2016) published the manuscript "Using spatial context to improve early detection of deforestation from Landsat time series". They proposed a new approach that reduces seasonality in satellite image time series using spatial context, with a spatial moving window to calculate a median value of Landsat NDVI time series, and a spatial moving window (defined spatial neighbourhood) to calculate a median value that is divided by the value of the focal pixel resulting into a spatially normalised time series. The assumption was: if no disturbance has occurred, the forest at the focal pixel would exhibit temporal dynamics similar to that of forest pixels within its neighbourhood. Dividing the value of the focal pixel with median value of its neighbours would yield a value closer to one if no disturbance has

occurred or if the disturbance is regional. When the focal pixel becomes disturbed and the disturbance is local, dividing the value of the focal pixel by the median value calculated from pixels in the neighbourhood, is likely to amplify the impact of the disturbance on the time series.

The authors demonstrated that the spatial context is a useful approach for timely detection of deforestation events in areas where forests exhibit strong seasonality. However, as they used a pixel-based approach that requires a moving window, the selection of the appropriate window size is necessary, and also, it is biased along their diagonals and will likely straddle the boundary between two landscape features, especially when a large window size is used (LALIBERTE; RANGO, 2009), implying in limitations.

Recently Silveira et al. (2018) proposed a method to minimize the effects of forest phenology on forest change detection, exploiting the spatial context, represented by geostatistical features, using bitemporal Landsat NDVI images and an object-based approach. The study findings indicated that using the geostatistical context it is possible to use bitemporal NDVI images to address the challenge of accurately detecting forest changes, eliminating the effects of forest phenology, without the need to use a dense time series of remote sensing images.

2.3 Aboveground biomass modelling using remote sensing

In recent years remote sensing techniques have become prevalent in estimating aboveground biomass (AGB). The advantages of remotely sensed data, such as in repetitivity of data collection, a synoptic view, a digital format that allows fast processing of large quantities of data, and the high correlations between spectral bands and vegetation parameters, make it the primary source for large area AGB estimation, especially in areas of difficult access (LU, 2006).

Although remotely sensed observations do not directly measure biomass, the radiometry is sensitive to vegetation structure (crown size and tree density), texture and shadow, which are correlated with AGB, particularly in the short wave infrared bands (BACCINI et al., 2008).

This has been mainly achieved by the use of vegetation indices such as the normalized difference vegetation index (NDVI). The main problem associated with indices computed from multispectral sensors is that they reach a saturation level on high density biomass estimation (CHEN et al., 2009). NDVI calculated from broad band sensors, asymptotically approach a saturation level after a certain AGB of about 0.3 g cm^{-1} (HURCOM; HARRISON, 1998) or vegetation age of about 15 years in tropical forests (STEININGER, 2000). Therefore, NDVI yields poor estimates during peak growing seasons and in more densely vegetated areas (THENKABAIL et al., 2004).

High-spatial resolution data from both airborne and satellite platforms can provide accurate aboveground biomass estimates at local scales; however, for regional scales, a large volume of data is required, which is not only expensive, but also difficult to process; this limits its application for larger areas. Landsat TM data have been found more effective for AGB estimation at a regional scale, however, mixed pixels and data saturation problems have been reported with these data in biomass estimation for complex environments. At the national and global scales, coarse-spatial resolution data, such as AVHRR or MODIS, have been found useful in AGB estimation, however, the data have not been used much because of the difficulty in linking coarse-spatial resolution data and field measurement (KUMAR et al., 2015).

Baccini et al. (2008) described methods to map aboveground biomass over tropical Africa using multi-year MODIS satellite observations and a wide range of field measurements. The results indicated that the MODIS data sets

captured the amount and spatial distribution of aboveground biomass across tropical Africa.

Mutanga, Adam and Cho (2012) evaluated the performance of data extracted from WorldView-2 imagery to estimate biomass in the South Africa. The major conclusion was that the NDVI extracted from WorldView-2 can be used to model and predict wetland biomass in a high density and vegetated wetland. The NDVI involving the additional spectral bands 2, such as the red-edge and near infrared regions of the electromagnetic spectrum, improved the prediction accuracy compared with the traditional NDVIs.

Basuki et al. (2013) proposed an approach to quantification of AGB by integrating phased array L-band synthetic aperture radar (PALSAR) with Landsat-7 ETM+. The proposed model showed the potential use of the integration of radar and optical imagery increasing the estimation accuracy of AGB in mixed forests.

Latifi et al. (2015) explored the question of stratification the sampling units into major forest types to improve the predictive quality of forest biomass modeling using LiDAR (light detection and ranging) and hyperspectral data. The results indicated that the sensor type (hyperspectral/LiDAR) showed to be the essential source of impact on the yielded predictive performance of the models and the based on forest strata did show slight improvements over the use of unstratified models.

Fayad et al. (2016) proposed a calibrated regression model to map the AGB by using LiDAR, SAR, optical and environmental data. The goals were to decrease the bias and increase the precision of the AGB estimates (which can be linked to the saturation of the radar and optical data) using GLAS data that are freely accessible and globally available but with sparse, and inhomogeneous coverage. They used the regression kriging technique with the random forest algorithm, that allowed to achieve the best precision for the AGB estimates.

A review of various remote sensing applications (optical, hyperspectral, and active remote sensing) in forest aboveground biomass inventorying and monitoring was conducted by Timothy et al. (2016). The review concluded that the use of remote sensing in large-scale forest aboveground biomass quantification provides plausible alternatives, when compared to the use of conventional approaches, which are laborious costly, and time-intensive and sometimes inapplicable due to poor accessibility. It was noted that although remote sensing provides reasonably accurate forest aboveground biomass estimates, active sensors, such as LiDAR and radar are not fully operational as yet due to complex pre-processing and high cost of data acquisition. They also concluded that more research is needed on the application of remote sensing for estimating the AGB.

3 FINAL CONSIDERATIONS

We recommend the integration between geostatistical techniques applied to image classification, change detection and biomass estimate. The general conclusions were:

- a) There is a relationship between landscape heterogeneity and measures of spatial dependence from remotely sensed data; Image spatial resolution does in fact influence the sill and range parameters and, medium spatial resolutions are the most appropriate to derive indicators of landscape heterogeneity;
- b) The semivariogram curves obtained by NDVI and NIR generated from Sentinel 2 and Landsat 8 were efficient for characterizing spatial heterogeneity within contrasting landscape vegetation cover, such as cerrado sensu stricto, deciduous forests and palm swamps; The image classification accuracy was significantly improved by combining geostatistical features with spectral data, thereby supporting the use of the semivariogram indices to improve object-based image classification procedures;
- c) Land-use/cover changes can be accurately detected using only semivariogram indices obtained by semivariograms derived from Landsat NDVI images, due to the NDVI spatial variability have the potential to discriminate between homogeneous and heterogeneous pixels within objects;
- d) The spatial variability of NDVI values, represented by geostatistical features (most of the semivariogram indices and sill parameter) are not affected by vegetation seasonality, and therefore, are able to accurately detect land-use/cover changes in an object level, disregarding those

associated with forest phenology, resulting in fewer classification errors, without the need to use a dense time series of remote sensing images;

- e) The combination of spectral features and semivariogram indices derived from Landsat NDVI times series provides the best results do disentangling the effects of forest phenology and land-use/cover changes in seasonal biomes, underscoring the use of spatial features to reduce the need of multi-temporal satellite images to accurately extract land use/cover changes such as deforestation, logging or fire while disregarding the ones caused by phenological differences;
- f) The NDVI spatial variability, captured by AFM semivariogram index is not affected by vegetation seasonality, and therefore, can produce times series that accurately differentiate land-use/cover changes from seasonal changes, resulting in fewer classification errors;
- g) The spatial context approach is useful for the detection of deforestation and fires events using NDVI Landsat data in areas where forests exhibit strong seasonality, addressing the challenge of accurately detecting land-use/cover changes, eliminating the effects of forest phenology, without the need of using de-seasoning models;
- h) The stratification of data into vegetation types not only improved the accuracy of aboveground biomass estimative, but also allowed random forest regression to select the lowest number of variables that offer the best predictive model performance to AGB mapping. The improvement in AGB estimates is driven by the spatial distribution and seasonality effects of each vegetation type, and it is achieved by stratifying the models which minimize the Savanna-Forest transition heterogeneity.

REFERENCES

ACERBI JUNIOR, F. W. et al. Change detection in Brazilian savannas using semivariograms derived from NDVI images. **Ciencia e Agrotecnologia**, Lavras, v. 39, n. 2, p. 103–109, mar./abr. 2015.

ATKINSON, P. M.; LEWIS, P. Geostatistical classification for remote sensing: an introduction. **Computers & Geosciences**, New York, v. 26, n. 4, p. 361–371, May 2000.

BACCINI, A. et al. A first map of tropical Africa's above-ground biomass derived from satellite imagery. **Environmental Research Letters**, Bristol, v. 3, n. 4, p. 1-9, Dec. 2008.

BALAGUER, A. et al. Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification. **Computers & Geosciences**, New York, v. 36, n. 2, p. 231–240, Feb. 2010.

BALAGUER-BESER, A. et al. Using semivariogram indices to analyse heterogeneity in spatial patterns in remotely sensed images. **Computers & Geosciences**, New York, v. 50, p. 115–127, Jan. 2013.

BASUKI, T. M. et al. Estimating tropical forest biomass more accurately by integrating ALOS PALSAR and Landsat-7 ETM+ data. **International Journal of Remote Sensing**, London, v. 34, n. 13, p. 4871–4888, Apr. 2013.

CHEN, J. et al. Estimating aboveground biomass of grassland having a high canopy cover: an exploratory analysis of in situ hyperspectral data. **International Journal of Remote Sensing**, London, v. 1161, n. 24, p. 37–41, Dec. 2009.

CHICA-OLMO, M.; ABARCA-HERNÁNDEZ, F. Computing geostatistical image texture for remotely sensed data classification. **Computers & Geosciences**, New York, v. 26, n. 4, p. 373–383, May 2000.

COSTANTINI, M. L. et al. NDVI spatial pattern and the potential fragility of mixed forested areas in volcanic lake watersheds. **Forest Ecology and Management**, Amsterdam, v. 285, p. 133–141, Dec. 2012.

CURRAN, P. J. The semivariogram in remote sensing: an introduction. **Remote Sensing of Environment**, New York, v. 24, n. 3, p. 493–507, Apr. 1988.

FAYAD, I. et al. Aboveground biomass mapping in French Guiana by combining remote sensing, forest inventories and environmental data. **International Journal of Applied Earth Observation and Geoinformation**, Enschede, v. 52, p. 502–514, Oct. 2016.

GARRIGUES, S. et al. Multivariate quantification of landscape spatial heterogeneity using variogram models. **Remote Sensing of Environment**, New York, v. 112, n. 1, p. 216–230, Jan. 2008.

_____. Quantifying spatial heterogeneity at the landscape scale using variogram models. **Remote Sensing of Environment**, New York, v. 103, n. 1, p. 81–96, July 2006.

GIL-YEPES, J. L. et al. Description and validation of a new set of object-based temporal geostatistical features for land-use/land-cover change detection. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 121, p. 77–91, Nov. 2016.

HAMUNYELA, E.; VERBESSELT, J.; HEROLD, M. Using spatial context to improve early detection of deforestation from Landsat time series. **Remote Sensing of Environment**, New York, v. 172, p. 126–138, Jan. 2016.

HURCOM, S. J.; HARRISON, A. R. The NDVI and spectral decomposition for semi-arid vegetation abundance estimation. **International Journal of Remote Sensing**, London, v. 19, n. 16, p. 3109–3125, 1998.

ISAAKS, E. H.; SRIVASTAVA, R. M. Spatial continuity measures for probabilistic and deterministic geostatistics. **Mathematical Geology**, New York, v. 20, n. 4, p. 313–341, May 1988.

KRIGE, D. G. A statistical approach to some basic mine valuation problems on the Witwatersrand. **Journal of the Chemistry, Metallurgy and Mining Society of South Africa**, Elmsford, v. 52, n. 9, p. 119–139, Dec. 1951.

KUMAR, L. et al. Review of the use of remote sensing for biomass estimation to support renewable energy generation. **Journal of Applied Remote Sensing**, Bellingham, v. 9, n. 1, p. 97696, June 2015.

LALIBERTE, A. S.; RANGO, A. Texture and scale in object-based analysis of subdecimeter resolution Unmanned Aerial Vehicle (UAV) Imagery. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 47, n. 3, p. 761–770, Mar. 2009.

LATIFI, H. et al. Stratified aboveground forest biomass estimation by remote sensing data. **International Journal of Applied Earth Observation and Geoinformation**, Enschede, v. 38, p. 229–241, June 2015.

LU, D. The potential and challenge of remote sensing-based biomass estimation. **International Journal of Remote Sensing**, London, v. 27, n. 7, p. 1297–1328, 2006.

MATHERON, G. Principles of geostatistics. **Economic Geology**, New York, v. 58, n. 8, p. 1246–1266, 1963.

MIRANDA, F. P.; FONSECA, L. E. N.; CARR, J. R. Semivariogram textural classification of JERS-1 (Fuyo-1) SAR data obtained over a flooded area of the Amazon rainforest. **International Journal of Remote Sensing**, London, v. 19, n. 3, p. 549–556, 1998.

MIRANDA, F. P.; MACDONALD, J. A.; CARR, J. R. Application of the semivariogram textural classifier (STC) for vegetation discrimination using SIR-B data of borneo. **International Journal of Remote Sensing**, London, v. 13, n. 12, p. 2349–2354, Jan. 1992.

MUTANGA, O.; ADAM, E.; CHO, M. A. High density biomass estimation for wetland vegetation using worldview-2 imagery and random forest regression algorithm. **International Journal of Applied Earth Observation and Geoinformation**, Enschede, v. 18, p. 399–406, Aug. 2012.

RUIZ, L. A. et al. A feature extraction software tool for agricultural object-based image analysis. **Computers and Electronics in Agriculture**, Amsterdam, v. 76, n. 2, p. 284–296, May 2011.

SCOLFORO, H. F. et al. Spatial distribution of aboveground carbon stock of the arboreal vegetation in Brazilian Biomes of Savanna, Atlantic Forest and Semi-arid woodland. **PLoS One**, San Francisco, v. 10, n. 6, p. 1–20, June 2015.

SERTEL, E.; KAYA, S.; CURRAN, P. J. Use of semivariograms to identify earthquake damage in an Urban Area. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 45, n. 6, p. 1590–1594, June 2007.

SILVEIRA, E. M. O. et al. Assessment of geostatistical features for object-based image classification of contrasted landscape vegetation cover. **Journal of Applied Remote Sensing**, Bellingham, v. 11, n. 3, p. 36004, July 2017b.

_____. Characterizing landscape spatial heterogeneity using semivariogram parameters derived from NDVI images. **Cerne**, Lavras, v. 23, n. 4, p. 413-422, out./dez. 2017a.

_____. Object-based change detection using semivariogram indices derived from NDVI images: the environmental disaster in Mariana, Brazil. **Ciencia e Agrotecnologia**, Lavras, v. 41, n. 5, p. 554–564, set./out. 2017c.

_____. Object-based land-cover change detection applied to Brazilian seasonal savannahs using geostatistical features. **International Journal of Remote Sensing**, London, v. 39, n. 8, p. 2597-2619, Jan. 2018.

STEININGER, M. K. Satellite estimation of tropical secondary forest above-ground biomass: Data from Brazil and Bolivi. **International Journal of Remote Sensing**, London, v. 21, n. 6/7, p. 1139–1157, 2000.

THENKABAIL, P. S. et al. Biomass estimations and carbon stock calculations in the oil palm plantations of African derived savannas using IKONOS data. **International Journal of Remote Sensing**, London, v. 25, n. 23, p. 5447–5472, 2004.

TIMOTHY, D. et al. Remote sensing of aboveground forest biomass: a review. **Tropical Ecology**, Varanasi, v. 57, n. 2, p. 125–132, 2016.

WIJS, H. J. de. Statistical ore distribution. **Geologie en Mijnbouw**, Dordrecht, v. 30, p. 365-375, Nov. 1951.

WOODCOCK, C. E.; STRAHLER, A. H.; JUPP, D. L. B. The use of variograms in remote sensing: II. Real digital images. **Remote Sensing of Environment**, New York, v. 25, n. 3, p. 349–379, Aug. 1988b.

_____. The use of variograms in remote sensing: I. Scene models and simulated images. **Remote Sensing of Environment**, New York, v. 25, n. 3, p. 323–348, Aug. 1988a.

WU, X. et al. Evaluation of semivariogram features for object-based image classification. **Geo-spatial Information Science**, Wuhan, v. 18, n. 4, p. 159–170, Sept. 2015.

YUE, A. et al. Texture extraction for object-oriented classification of high spatial resolution remotely sensed images using a semivariogram. **International Journal of Remote Sensing**, London, v. 34, n. 11, p. 3736–3759, Feb. 2013.

ZHU, Z. Change detection using landsat time series: a review of frequencies, preprocessing , algorithms, and applications. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 130, p. 370–384, Aug. 2017.

SECOND PART – ARTICLES**ARTICLE 1 - CHARACTERIZING LANDSCAPE SPATIAL
HETEROGENEITY USING SEMIVARIOGRAM PARAMETERS
DERIVED FROM NDVI IMAGES**

Eduarda Martiniano de Oliveira Silveira¹, José Márcio de Mello¹, Fausto Weimar Acerbi Júnior¹, Aliny Aparecida dos Reis¹, Kieran Daniel Withey², Luis Angel Ruiz³

¹ Federal University of Lavras - Lavras, Minas Gerais, Brazil

² Lancaster University - Lancaster, Lancashire, United Kingdom of Great Britain and Northern Ireland

³ Universitat Politècnica de València - Valencia, Valencia, Spain

Publication status: Published in the **Cerne**

10.1590/01047760201723042370

Abstract. Assuming a relationship between landscape heterogeneity and measures of spatial dependence by using remotely sensed data, the aim of this work was to evaluate the potential of semivariogram parameters, derived from satellite images with different spatial resolutions, to characterize landscape spatial heterogeneity of forested and human modified areas. The NDVI (Normalized Difference Vegetation Index) was generated in an area of Brazilian amazon tropical forest (1,000 km²). We selected samples (1 x 1 km) from forested and human modified areas distributed throughout the study area, to generate the semivariogram and extract the sill (σ^2 -overall spatial variability of the surface property) and range (ϕ -the length scale of the spatial structures of objects) parameters. The analysis revealed that image resolution influenced the sill and range parameters. The average sill and range values increase from forested to human modified areas and the greatest between-class variation was found for LANDSAT 8 imagery, indicating that this image resolution is the most appropriate for deriving sill and range parameters with the intention of describing landscape spatial heterogeneity. By combining remote sensing and geostatistical techniques, we have shown that the sill and range parameters of semivariograms derived from NDVI images are a simple indicator of landscape heterogeneity and can be used to provide landscape heterogeneity maps to enable researchers to design appropriate sampling regimes. In the future, more applications combining remote sensing and geostatistical features should be further investigated and developed, such as change detection and image classification using object-based image analysis (OBIA) approaches.

Key words: remote sensing, geostatistics, forested areas, human-modified landscapes.

Introduction

Recent years have seen a dramatic increase in the attention being given to the condition of tropical forests. In recognition of the considerable impact human activities are having on tropical forest systems, a range of initiatives have been launched to mitigate the adverse effects of tropical forest loss (DEVRIES et al., 2016). Remote sensing based approaches play a key role in forest monitoring, as they are of low cost and provide an opportunity for mapping forest change over large areas (DEVRIES et al., 2015).

Understanding the negative and positive effects of agricultural land use for the conservation of biodiversity, and its relation to ecosystem services, needs a landscape perspective (TSCHRNTKE et al., 2005). Landscapes exhibit various degrees of spatial heterogeneity due to the interactions of natural and anthropogenic processes (BIE et al., 2012).

Remote sensing using satellite imagery has emerged as a key geospatial tool to meet the growing information needs of landscape and forest managers (COSTANTINI et al., 2012). Different methods to quantify changes in landscape complexity have been developed in the last few decades (WU, 2013). Most of these involve the use of remotely sensed images and geospatial techniques (MONMANY et al., 2015; BERBEROGLU et al., 2000; BERBEROGLU; AKIN, 2009; GARCIA-PEDRERO et al., 2015). Several works have investigated environmental changes, using spatial heterogeneity derived from various types of remote sensing data (WU et al., 2000; CHEN; HENEGBRY, 2009; FENG et al., 2010). A reasonably number of studies have focused on the variation of spatial heterogeneity (WEN et al., 2012) using multi-scale remote sensing data (CHEN; HENEGBRY, 2009).

In some studies, the Normalized Difference Vegetation Index (NDVI) has been used to detect and analyse spatial heterogeneity using semivariograms (GARRIGUES et al., 2006; GARRIGUES et al., 2008; BALAGUER-BESER et

al., 2013). NDVI, which represents an especially informative vector for landscape structure and temporal change analyses (GRIFFITH et al., 2002), has been used in numerous studies of vegetation dynamics because of its simplicity and close relationship to variables of ecological interest such as land cover change and disturbance propagation at multiple scales (ZURLINI et al., 2006; ZACCARELLI et al., 2008).

Geostatistical semivariograms are used as measures of texture (CURRAN, 1988; WOODCOCK; et al., 1988) and have been widely used for heterogeneity analysis (GARRIGUES et al., 2006, 2008; CADENASSO et al., 2007; HUANG et al., 2013; LAUSCH et al., 2013; BALAGUER-BESER et al., 2013; WU et al., 2000; QIU et al., 2013), improved image classification (BALAGUER et al., 2010; BALAGUER-BESER et al., 2011; WU et al., 2015; YUE et al., 2013; POWERS et al., 2015) and change detection (SERTEL et al., 2007; COSTANTINI et al., 2012; ACERBI JUNIOR et al., 2015; GIL-YEPES et al., 2016). In short, the spatial heterogeneity of surface reflectance values is dependent on the spatial resolution of the image, spectral bands and the size of the image or sample analysed.

Sampling heterogeneity in a statistically robust manner has proven to be a challenge, with researchers often stratifying their sampling regimes along subjectively chosen landscape features (WHITE et al., 2004). Sampling regimes can be improved by an initial estimate of the spatial heterogeneity, provided by an accurate map of landscape heterogeneity. The mosaics of land use and land cover appear with boundaries and edges between them and this variability can be captured and extracted by using the spatial variability of remote sensing images using geostatistical approaches (GUARRIGUES et al., 2006).

Semivariograms are reportedly an efficient method to characterize the structure of spatial continuity (GUEDES et al. 2015), due to their potential to describe the spatial variability of data. Therefore, assuming a relationship

between landscape heterogeneity and measures of spatial dependence of remotely sensed data, the aim of this work was to evaluate the potential of semivariogram parameters, derived from satellite images with different spatial resolutions, to characterize landscape spatial heterogeneity of forested and human modified areas. The questions that motivated this study were: (i) How is image spatial resolution linked with the spatial variability of NDVI values? (ii) Which semivariogram parameter and image spatial resolution is the most appropriate to produce a landscape heterogeneity map?

This manuscript makes a significant contribution to the understanding of how the manipulation of satellite imagery parameters can be used to detect variation in the landscape, indicating the most appropriate image spatial resolution to improve the analysis and also the most appropriate semivariogram parameter to generate a map of landscape heterogeneity. This enables researchers to design appropriate sampling regimes to capture landscape heterogeneity.

Material and Methods

We performed three main steps: (1) Acquisition of SPOT 6 (Satellite Pour l'Observation de la Terre), LANDSAT 8 (Land Remote Sensing Satellite) and MODIS TERRA (Moderate Resolution Imaging Spectroradiometer) imagery and generation of NDVI images; (2) Stratified sampling of forested and human modified areas distributed throughout the study area; (3) Analysis of semivariograms generation using the NDVI pixels inside the samples. The sill (σ^2 -overall spatial variability of the surface property) and range (ϕ -the length scale of the spatial structures of objects or patches) parameters were generated by fitting mathematical models (exponential, spherical and gaussian) to the experimental semivariograms using the weighted least squares method. Figure 1 presents a diagrammatical overview of the methods.

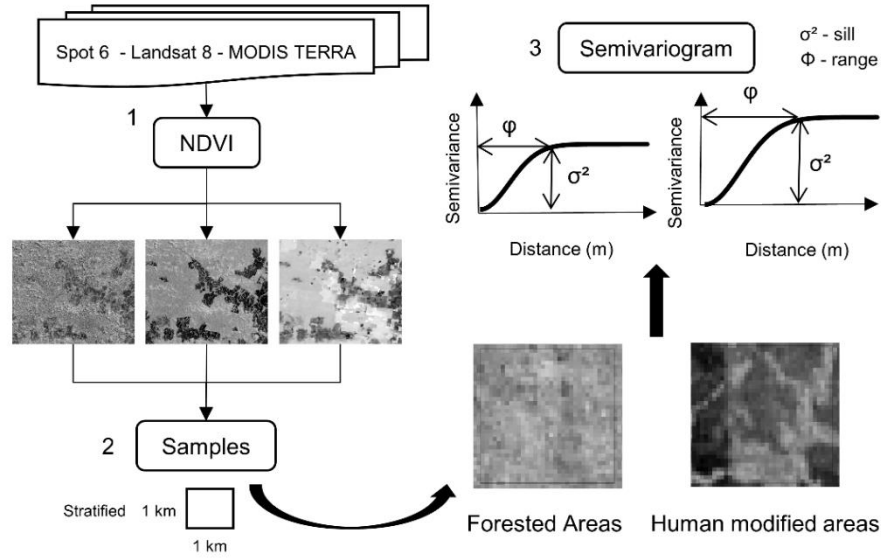


Figure 1. Diagrammatical overview of methods divided into three main steps.

Study area

The study area is a total of 1,000 km² set within the Cotriguaçu municipality, located in the northeast of the state of Mato Grosso (MT), Brazil, central coordinates 09° 48' 9.14" south latitude and 58° 47' 40.25" west longitude (Figure 2).

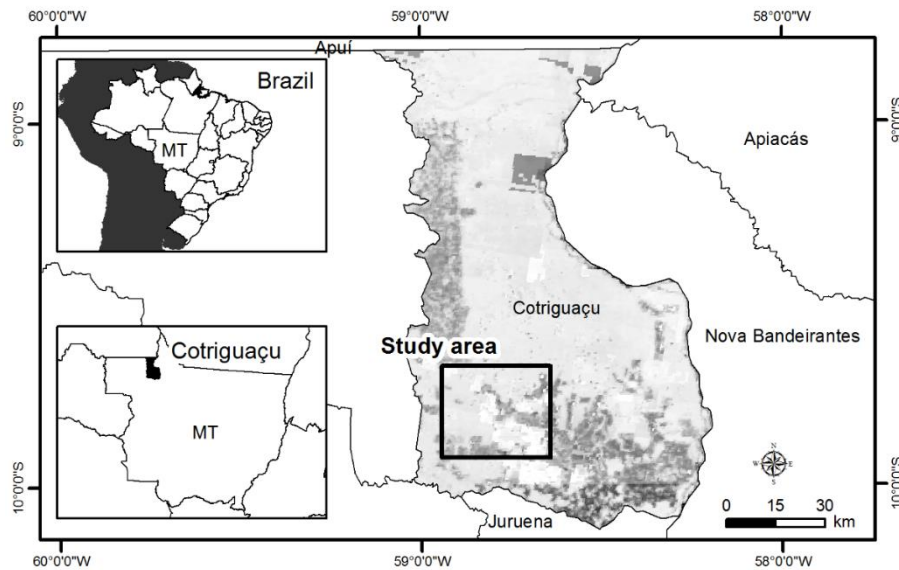


Figure 2. Location of the study area in the municipality of Cotriguaçu, Mato Grosso (MT) State, Brazil.

The municipality of Cotriguaçu, which has a total area of 9,149 km², forms part of the Amazon Watershed and is drained by the Juruena River, the largest volume of water in Mato Grosso State. A total of 25% of its landmass has a flat relief, 60% has an irregular relief and 15% has a mountainous relief. The municipality has a mean altitude of 240 meters. This region has a warm and humid equatorial climate with two dry months (June and July). The mean annual rainfall is 2,750 mm, with low rainfall from May to September (period in which the majority of logging occurs) and high precipitation from January to March. Average annual temperatures ranges from 22°C to 35°C (FINGER, 2005).

Timber is the main commodity and the forested areas are characterized by a wide diversity of valuable tropical species. By 2004 Cotriguaçu had lost 13% of its forest cover. Extensive livestock farming is the main reason for conversion of forest areas in this region. The wood poles located in the Amazon

are established from an economy based exclusively on forestry, on which the local residents depend for socio-economic development. These municipalities are the most susceptible to the effects of predatory exploitation of natural resources. Cotriguaçu is undergoing expansion with excessive pressure on natural resources (FERREIRA et al., 2005).

Image acquisition

We used images with three different spatial resolutions to assess the effect of spatial resolution on the semivariogram parameters generated from NDVI images (Table 1). The SPOT 6 image was provided by ONF (Office National des Forêts) Brazil, a Brazilian Forestry Company, subsidiary of ONF International, a company of the ONF group, leading a reforestation pilot project for carbon sequestration "Peugeot-ONF carbon sink" in the northwest of Mato Grosso.

The LANDSAT 8 and MODIS TERRA images were acquired from the United States Geological Survey for Earth Observation and Science (USGS\EROS). The LANDSAT 8 image was acquired in the CDR processing level (Landsat Surface Reflectance Climate Data Record), that is, with the necessary geometric corrections and reflectance values at ground level. The MODIS product used was MOD13Q1 - Vegetation Index.

Table 1. Acquisition dates and spatial resolution of SPOT 6, LANDSAT 8 and MODIS TERRA images.

Images	Acquisition date	Spatial resolution (m)
SPOT 6	09/July/2014	6
LANDSAT 8	09/August/2014	30
MODIS TERRA	10/August/2014	250

NDVI was used to describe the spatial heterogeneity of vegetation cover. This index is based on quotients and uses red spectral and near infrared bands to enhance vegetation and, at the same time, minimize the effects of shadows caused by topography (VOROVENCII, 2014). Although NDVI is sensitive to soil and atmospheric effects, it is a good indicator of the total amount of vegetation (HENEGBRY, 1993) and is considered important for the analysis of land cover structure and temporal changes (GRIFFITH et al., 2002; SADER, 2003).

Sampling design

According to a previous map provided by ONF Brazil in 2014, the study area comprises six land use/land cover (LULC) classes: croplands, degraded forests, forests, plantation, settlements and water. We merged these LULC into two classes: (1) Forested areas and (2) Human modified areas (Figure 3).

Forested areas are densely tropical rainforest, that have not undergone major changes due to human activities and present only on patch covering the total grid area.

Human modified areas are croplands, settlements, degraded forests and plantations, with more than one patch in the grid. The cropland class is composed of small farm plots around villages and in forest edges (agroforest mosaics), as well as more intensive subsistence agricultural use (large permanent plots). Towns, villages and other human infrastructures were classified as 'settlements'. The water class is composed of rivers and open water surfaces. Degraded forest and saplings were designed to identify areas with significant tree density (typically tree cover in the order of 25 to 35%). The plantation class is composed of tree plantations.

We selected 20 square samples (1 x 1 km) using a stratified sampling approach (Figure 3), comprising 10 samples of human modified areas and 10

samples of forested areas in each image, totalizing 60 samples. We then extracted all the NDVI values from the SPOT 6, LANDSAT 8 and MODIS TERRA pixels inside the sampled areas.

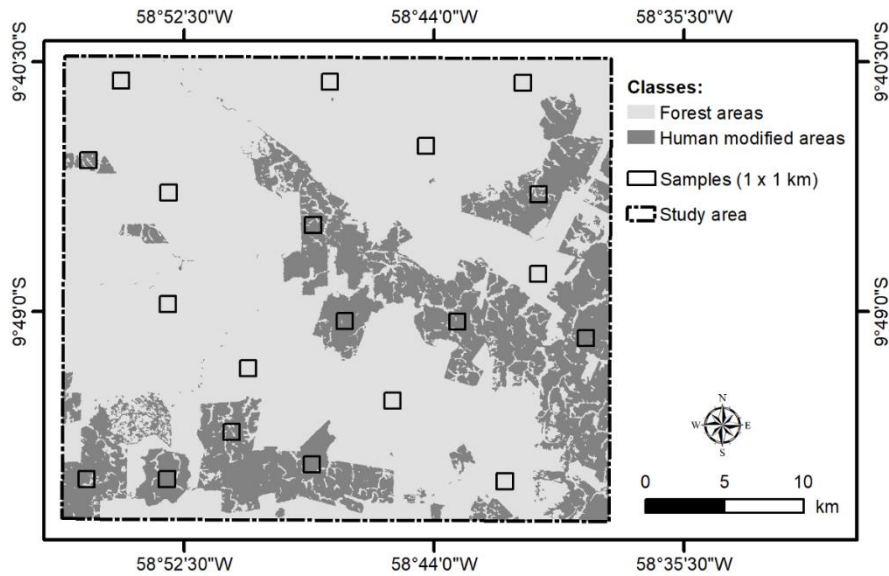


Figure 3. Sampling design in forest and human modified areas.

Semivariogram parameters

The semivariogram is a graphical representation of the spatial variability in a given set of data (COHEN et al., 1990). The relationship between a pair of pixels can be calculated with the variogram function (Equation 1), called $2\gamma(h)$, which corresponds to the mathematical expectation of the squared difference between pairs of points separated by a distance h , where $Z(x)$ is the value of the regionalized variable at point x , $Z(x+h)$ is the value at $x+h$. The semivariogram function depends on the location x , and the distance between samples h . For the variogram to be based solely on the distance between the sampling units, it is necessary to adopt the intrinsic hypothesis (stationarity), which implies that the variance of the differences between two sample points

depends only on the distance (h).

$$2\gamma(h)=E\{[Z(x)-Z(x+h)]^2\} \quad (1)$$

For continuous variables, such as the NDVI, the experimental semivariogram is defined as the half of the average squared difference between values separated by a given lag, where this lag is a vector in both distance and direction (ATKINSON; LEWIS, 2000). It was estimated using equation 2:

$$\gamma(h)=\frac{1}{2N(h)}\sum_{i=1}^{N(h)}[Z(x)-Z(x+h)]^2 \quad (2)$$

Where $\gamma(h)$ is the estimator of the semivariance for each distance h, $N(h)$ is the number of pairs of points separated by the distance h, $Z(x)$ is the value of the regionalized variable at point x and $Z(x+h)$ is the value of the point x+h.

Spatial variance versus distance h, is the graphical representation of the semivariogram, which allows obtaining an estimate of the variance value for different combinations of pairs of points. The semivariogram (Figure 4) is characterized by three parameters: sill (σ^2), range (ϕ) and nugget effect (τ^2). The sill parameter is the plateau reached by the semivariance values and shows the quantity of variation explained by the spatial structure of the data. The range parameter is the distance at which the semivariogram reaches the sill, showing the distance at which the data cease to be correlated. The nugget effect is the combination of sampling errors and variations that happen at scales smaller than the distance between the sampled points (CURRAN, 1988).

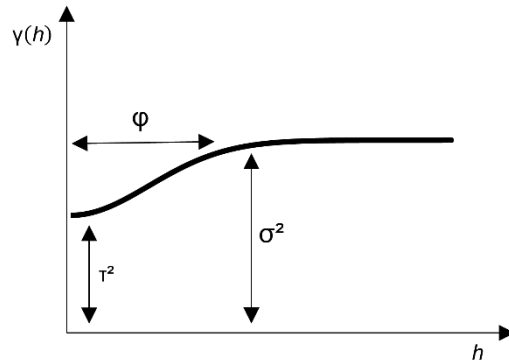


Figure 4. Illustration of a classical semivariogram: σ^2 - sill; (ϕ) range and (τ^2) nugget effect.

We generate the semivariograms using the NDVI values extracted from the square samples in the SPOT 6, LANDSAT 8 and MODIS images. The size of the samples needs to be larger than the range of influence to characterize the initial part of the semivariogram and large enough to reveal the presence of periodicity (WOODCOCK et al., 1988).

Geostatistical methods are optimal when data are normally distributed and stationary (mean and variance do not vary significantly in space). Significant deviations from normality and stationarity can cause problems, so it is always best to begin by looking at a histogram to check for normality and a plotting of the data values in space to check for significant trends.

Thus, the first step was an exploratory data analysis in order to understand the overall behaviour of NDVI values, i.e. to study the tendency, shape and distribution of the data. The position measures (average, minimum and maximum values), dispersion (standard deviation and coefficient of variation) and normality assessment were determined through the frequency histogram. In addition, to verify the presence of outliers, a boxplot graph was

generated. After the exploratory analysis, the semivariograms were generated to analyse the spatial dependence of the data.

Theoretical semivariograms were estimated by fitting mathematical models to the experimental semivariogram using the weighted least squares method. Exponential, spherical and gaussian models were tested (Table 2). The fitted models were cross-validated, analysing the reduced mean error (ER) and the standard deviation of reduced errors (SRE). We used R (R Core Team 2016) and ArcGis version 10.1 (Esri 2010) to perform this analysis.

Table 2. Semivariogram models.

Model	Formula
Exponential	$\gamma(h)=\sigma^2 \left[1-e^{-3\left(\frac{h}{\phi}\right)} \right]$
Spherical	$\gamma(h)=\left\{ \sigma^2 \left[\frac{3}{2} \left(\frac{h}{\phi}\right) - \frac{1}{2} \left(\frac{h}{\phi}\right)^3 \right] \right\}$
Gaussian	$\gamma(h)=\sigma^2 \left[1-e^{-3\left(\frac{h}{\phi}\right)^2} \right]$

$\gamma(h)$ =semivariance; h =distance; σ^2 =sill and ϕ =range.

Results and discussion

Exploratory analysis

The distribution of NDVI data was found not to deviate from normality. The distribution of latitude and longitude values showed that the data do not have any spatial tendencies, and the spatial dependence can be explained only by the distance among the samples for all images, assuming the intrinsic hypothesis of stationarity.

The descriptive analysis of NDVI values inside samples (Table 3) showed that the average NDVI values for the forested areas were 0.48 for the SPOT 6 image, 0.43 for the LANDSAT 8 image and 0.86 for the MODIS

TERRA image. For the human modified samples, the average NDVI values were 0.39, 0.29 and 0.59, respectively. Thus, as the spatial resolution decreases (from 6 to 250 m), the average NDVI values between the classes increases, showing that the NDVI values between the land-use classes are more similar, using high resolution images (i.e. SPOT 6).

For forested and human modified areas, the minimum NDVI values were obtained from samples of the SPOT 6 image; this is probably due to the presence of shadows that are captured by high resolution images, causing a reduction in NDVI values. The maximum values for these classes, were provided by MODIS TERRA, reaching 0.90.

Table 3. Descriptive statistics.

Image	Samples	Forested areas				Human modified areas			
		Min	Mean	Max	Std	Min	Mean	Max	Std
SPOT 6	1	0.12	0.51	0.69	0.06	-0.30	0.35	0.64	0.07
	2	0.09	0.48	0.65	0.07	-0.12	0.44	0.67	0.07
	3	0.00	0.49	0.67	0.06	-0.45	0.41	0.60	0.10
	4	0.11	0.47	0.65	0.06	-0.25	0.37	0.64	0.07
	5	0.12	0.47	0.67	0.06	-0.25	0.38	0.62	0.08
	6	0.11	0.49	0.68	0.06	-0.44	0.41	0.61	0.08
	7	0.10	0.47	0.65	0.06	0.12	0.45	0.63	0.08
	8	0.04	0.47	0.64	0.06	-0.47	0.38	0.63	0.09
	9	0.06	0.48	0.66	0.07	-0.12	0.38	0.64	0.07
	10	0.07	0.48	0.66	0.07	0.14	0.38	0.66	0.08
	Average	0.08	0.48	0.66	0.06	-0.21	0.39	0.63	0.08
LANDSAT 8	1	0.37	0.44	0.53	0.02	0.16	0.25	0.46	0.06
	2	0.33	0.42	0.49	0.02	0.19	0.31	0.48	0.08
	3	0.36	0.43	0.50	0.02	0.01	0.28	0.47	0.07
	4	0.35	0.42	0.48	0.02	0.16	0.27	0.48	0.07
	5	0.36	0.42	0.49	0.02	0.11	0.27	0.46	0.07
	6	0.34	0.43	0.50	0.02	0.04	0.31	0.47	0.08

	7	0.36	0.43	0.50	0.02	0.23	0.37	0.50	0.06
	8	0.33	0.42	0.50	0.02	0.00	0.25	0.47	0.06
	9	0.36	0.43	0.50	0.02	0.18	0.29	0.48	0.07
	10	0.35	0.44	0.50	0.02	0.19	0.31	0.52	0.07
	Average	0.35	0.43	0.50	0.02	0.13	0.29	0.48	0.07
MODIS TERRA	1	0.88	0.89	0.90	0.01	0.41	0.55	0.80	0.08
	2	0.84	0.85	0.89	0.02	0.43	0.56	0.70	0.08
	3	0.86	0.87	0.88	0.00	0.64	0.69	0.71	0.02
	4	0.72	0.85	0.88	0.05	0.49	0.58	0.68	0.05
	5	0.67	0.84	0.88	0.05	0.38	0.48	0.62	0.08
	6	0.82	0.86	0.86	0.01	0.46	0.63	0.79	0.09
	7	0.85	0.86	0.87	0.01	0.59	0.67	0.76	0.06
	8	0.75	0.83	0.89	0.04	0.34	0.55	0.75	0.17
	9	0.86	0.87	0.88	0.01	0.49	0.61	0.73	0.08
	10	0.84	0.87	0.88	0.01	0.55	0.60	0.69	0.04
	Average	0.81	0.86	0.88	0.02	0.48	0.59	0.72	0.08

Min=minimum; Max=maximum; Std=standard deviation.

Semivariogram analysis

The semivariograms reached the sill within the calculated distance (Figure 5), indicating that their spatial extents were sufficiently large to encompass the entire spatial variability. The Gaussian model provided the best fit for the data. The cross-validated models, showed values close to 0 (~ 0.0005) for ER and close to 1 (~ 1.11) for SRE.

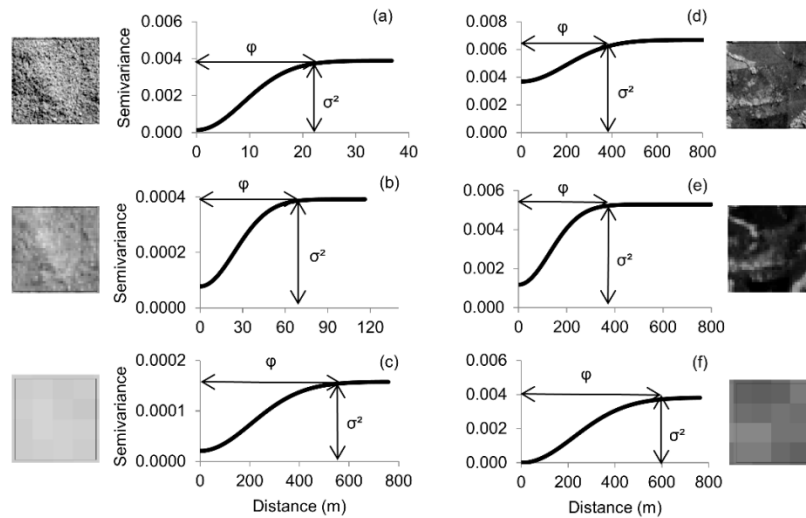


Figure 5. Example semivariogram curves for: (a) forested NDVI values of SPOT 6; (b) forested NDVI values of LANDSAT 8; (c) forested NDVI values of MODIS TERRA; (d) human modified NDVI values of SPOT 6; (e) human modified NDVI values of LANDSAT 8; (f) human modified NDVI values of MODIS TERRA. Dashed lines illustrate the σ^2 (sill) and ϕ (range) parameters.

We analysed the semivariogram parameters obtained from SPOT 6, LANDSAT 8 and MODIS TERRA in both classes (Table 4). The spatial variability of each image, represented by the sill (σ^2) parameter increases considerably from forested to human modified areas. The sill (σ^2) parameter indicates the amount of variance explained by the spatial structure. GARRIGUES et al. (2006) found that forested areas have a low sill values because the important development of vegetation and the presence of green understory limit the variability of the landscape vegetation cover. The high variability of human modified areas is explained by the mosaic of vegetation with high NDVI values and bare soil and pastures with low NDVI values. The highest sill value was found in the human modified areas using SPOT 6 (0.0064) and the lowest was found in forested areas using MODIS TERRA (0.0001).

The average range (ϕ) also increases from forested to human modified areas. The range parameter obtained from satellite images is the ratio of the area covered by the dominant objects (TREITZ, 2000). In forested areas, the range (ϕ) represents the size of the largest elements, which are the tree crowns of the forest canopy. On the other hand, in human modified areas, the dominant objects will vary according to the land cover classes present in the samples. Using SPOT 6, the sill values of forested areas reach the overall image variance at a short range (22 m), indicating that this class is poorly structured at the observational scale. The nugget effect is high, representing variability at distances smaller than the sample spacing (1 pixel). The greatest range was found in human modified areas using MODIS TERRA (600 m). The semivariogram range can be used to quantify coarse spatial variability since it increases as a result of changes in land-use throughout the landscape (SERTEL et al. 2007).

ACERBI JUNIOR et al. (2015) analysed semivariogram parameters to detect changes in a Brazilian savanna biome using NDVI LANDSAT TM images. They showed that sill and range semivariogram parameters were different when deforestation occurred and were similar when the area had not been changed.

Table 4. Average semivariogram parameters obtained from 20 samples.

Image	Semivariogram parameters	Forested	Human Modified	Variation
SPOT 6	σ^2	0.0039	0.0064	0.0025
	Φ	22	338	316
LANDSAT 8	σ^2	0.0004	0.0053	0.0049
	Φ	70	400	330
MODIS TERRA	σ^2	0.0001	0.0043	0.0042
	Φ	585	600	15

σ^2 =sill; ϕ =range.

The sill and range semivariogram parameters extracted from NDVI images are affected by image spatial resolution. Images with high spatial resolution have lower internal intra-pixel variability and high inter-pixel variability. This inter-pixel variability is linked with the sill and range parameters and is affected by the LULC classes. As image spatial resolution decreases, inter-pixel variability also decreases (Figure 6). Low inter-pixel variability provides low sill values and high range values, meaning that NDVI SPOT 6 should provide higher sill and lower range values than LANDSAT 8 and MODIS TERRA. For forested areas, this pattern is more pronounced than in human modified areas.

To select the most appropriate semivariogram parameter and image spatial resolution we analysed which one provided the greatest variation between the LULC classes. The variation of sill and range values between the classes increases between SPOT 6 (0.0025-316) and LANDSAT 8 (0.0049-330) and decreases from LANDSAT 8 to MODIS TERRA (0.0042-15).

In the SPOT 6 image, the variation in sill values between the LULC classes is low and the variation in range values is high. SPOT 6 images have high spatial resolution (6 meters), and the largest elements in forested areas are the tree crown shadows. The high NDVI values of trees and low values of shadows increase the overall variability captured by the sill semivariogram parameter, making the sill values of forested areas as high as the sill values of human modified areas. The range parameter is low in forested areas due to the size of the elements also being low. In human modified areas, the overall variability is influenced by the type of landscape (bare soil, plantation, pasture and cropland), presenting high range values, thus increasing the variation in range values between the classes.

In the LANDSAT 8 images, the sill and range present a high variation between the LULC classes. MODIS TERRA presents a high variation in sill values and low variation in range values.

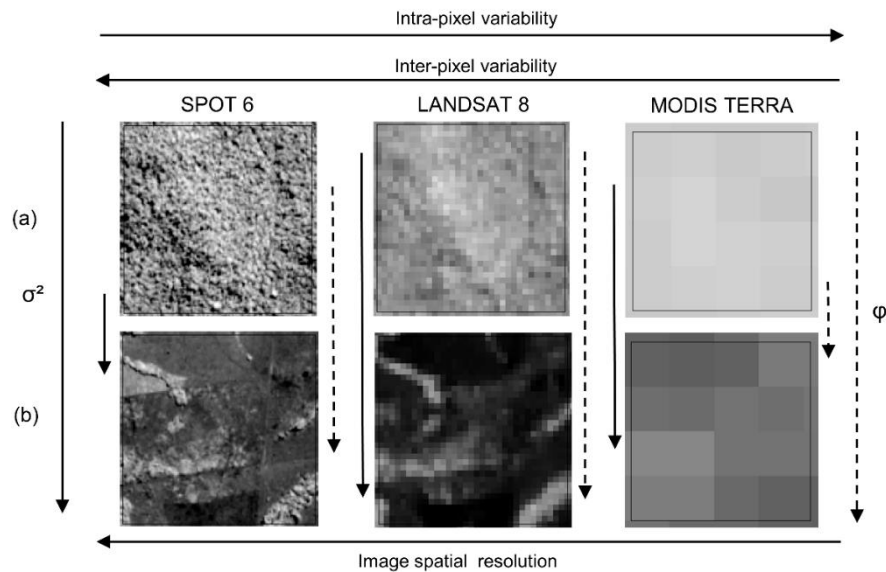


Figure 6. Effect of image resolution on sill (σ^2) and range (ϕ) parameters for the characterization of landscape spatial heterogeneity: (a) Forest areas and (b) Human modified areas.

The spatial heterogeneity between the LULC classes follows the same pattern using the semivariogram parameters obtained from three different image spatial resolutions. The sill (σ^2) and range (ϕ) parameters increase from forested to human modified areas. This increase is more pronounced using medium spatial resolution images (Landsat – 30 m).

The semivariogram parameters were efficient at describing landscape spatial heterogeneity, being able to distinguish the analysed classes. LANDSAT 8 showed the greatest differentiation of land cover classes with the sill and range parameters. We advise using both sill and range parameters derived from NDVI

LANDSAT images to identify forested and human modified areas, to capture the landscape heterogeneity.

Conclusions

We have evaluated the potential of semivariogram parameters derived from NDVI images to describe landscape spatial heterogeneity of forested and human modified areas.

Our analyses have revealed that image resolution does in fact influence the sill and range parameters. Average sill and range values increased from forested to human modified areas at the three image spatial resolutions analysed, however the greatest between-class variation was provided by LANDSAT 8 imagery, indicating that medium spatial resolution is the most appropriate for deriving the sill and range parameters with the intention of capture and map the landscape spatial heterogeneity.

By combining remote sensing and geostatistical techniques, we have shown that the sill and range parameters of semivariograms derived from NDVI images are a simple indicator of landscape heterogeneity and can be used to provide landscape heterogeneity maps to enable researchers to design appropriate sampling regimes. In the future, more applications combining remote sensing and geostatistical features should be further investigated and developed, such as change detection and image classification using object-based image analysis (OBIA) approaches.

Acknowledgments

The authors are grateful to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Department of Forest Science of the Federal University of Lavras (UFLA) and the ONF Brazil group for supporting this work.

References

ACERBI JUNIOR, F. W. et al. Change detection in Brazilian savannas using semivariograms derived from NDVI images. **Ciencia e Agrotecnologia**, v. 39, n. 2, p. 103–109, 2015.

ATKINSON, P. M.; LEWIS, P. Geostatistical classification for remote sensing: an introduction. **Computers & Geosciences**, v. 26, n. 4, p. 361–371, maio 2000.

BALAGUER-BESER, A. et al. Semivariogram calculation optimization for object-oriented image classification. **Modelling in Science Education and Learning**, v. 4, n. 7, p. 91–104, 2011.

BALAGUER-BESER, A. et al. Using semivariogram indices to analyse heterogeneity in spatial patterns in remotely sensed images. **Computers and Geosciences**, v. 50, p. 115–127, 2013.

BALAGUER, A. et al. Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification. **Computers and Geosciences**, v. 36, n. 2, p. 231–240, 2010.

BERBEROGLU, S. et al. The integration of spectral and textural information using neural networks for land cover mapping in the Mediterranean. **Computers and Geosciences**, v. 26, n. 4, p. 385–396, 2000.

BERBEROGLU, S.; AKIN, A. Assessing different remote sensing techniques to detect land use/cover changes in the eastern Mediterranean. **International Journal of Applied Earth Observation and Geoinformation**, v. 11, n. 1, p. 46–53, 2009.

BIE, C.A.J.M.et al. LaHMa: a landscape heterogeneity mapping method using hyper-temporal datasets. **International Journal of Geographical Information Science**, v.26, n.11, p. 2177-2192. 2012.

CADENASSO, M. L.; PICKETT, S. T. A.; SCHWARZ, K. Spatial heterogeneity in urban ecosystems: Reconceptualizing land cover and a

framework for classification. **Frontiers in Ecology and the Environment**, v. 5, n. 2, p. 80–88, 2007.

CHEN, W.; HENEGBRY, G. M. Change of spatial information under rescaling: A case study using multi-resolution image series. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 64, n. 6, p. 592–597, 2009.

COHEN, W. B.; SPIES, T. A.; BRADSHAW, G. A. Semivariograms of digital imagery for analysis of conifer canopy structure. **Remote Sensing of Environment**, v. 34, n. 3, p. 167–178, 1990.

COSTANTINI, M. L. et al. NDVI spatial pattern and the potential fragility of mixed forested areas in volcanic lake watersheds. **Forest Ecology and Management**, v. 285, p. 133–141, 2012.

CURRAN, P. J. The semivariogram in remote sensing: An introduction. **Remote Sensing of Environment**, v. 24, n. 3, p. 493–507, 1988.

DEVRIES, B. et al. Robust monitoring of small-scale forest disturbances in a tropical montane forest using Landsat time series. **Remote Sensing of Environment**, v. 161, p. 107–121, 2015.

DEVRIES, B. et al. Characterizing forest change using community-based monitoring data and landsat time series. **PLoS ONE**, v. 11, n. 3, p. 1–25, 2016.

FENG, X. et al. Remote sensing of ecosystem services: An opportunity for spatially explicit assessment. **Chinese Geographical Science**, v. 20, n. 6, p. 522–535, 2010.

FERREIRA, L. V.; VENTICINQUE, E.; ALMEIDA, S. O desmatamento na Amazônia e a importância das áreas protegidas. **Estudos avançados**, v. 19, n. 53, p. 157–166, 2005.

FINGER, F. A. Diagnóstico do setor florestal no município de Cotriguaçu, Mato Grosso: perspectivas e desafios na percepção dos dirigentes das empresas florestais. 2005.

GARCIA-PEDRERO, A. et al. A GEOBIA methodology for fragmented agricultural landscapes. **Remote Sensing**, v. 7, n. 1, p. 767–787, 2015.

GARRIGUES, S. et al. Quantifying spatial heterogeneity at the landscape scale using variogram models. **Remote Sensing of Environment**, v. 103, n. 1, p. 81–96, 2006.

GARRIGUES, S. et al. Multivariate quantification of landscape spatial heterogeneity using variogram models. **Remote Sensing of Environment**, v. 112, n. 1, p. 216–230, 2008.

GARRIGUES, S.; ALLARD, D.; BARET, F. Modeling temporal changes in surface spatial heterogeneity over an agricultural site. **Remote Sensing of Environment**, v. 112, n. 2, p. 588–602, 2008.

GIL-YEPES, J. L. et al. Description and validation of a new set of object-based temporal geostatistical features for land-use/land-cover change detection. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 121, p. 77–91, 2016.

GRIFFITH, J. A. et al. Preliminary comparison of landscape pattern-normalized difference vegetation index (NDVI) relationships to Central Plains stream conditions. **Journal of environmental quality**, v. 31, n. 3, p. 846–859, 2002.

GUEDES, I. C. DE L. et al. Continuidade espacial de características dendrométricas em povoamentos clonais de Eucalyptus sp. avaliada ao longo do tempo. **Cerne**, v. 21, n. 4, p. 527–534, 2015.

HENEGBRY, G. M. Detecting change in grasslands using measures of spatial dependence with landsat TM data. **Remote Sensing of Environment**, v. 46, n. 2, p. 223–234, 1993.

HUANG, Y. et al. Spatio-temporal variation of landscape heterogeneity under influence of human activities in Xiamen City of China in recent decade. **Chinese Geographical Science**, v. 23, n. 2, p. 227–236, 2013.

LAUSCH, A. et al. Monitoring and assessing of landscape heterogeneity at different scales. **Environmental Monitoring and Assessment**, v. 185, n. 11, p. 9419–9434, 2013.

POWERS, R. P. et al. Remote sensing and object-based techniques for mapping fine-scale industrial disturbances. **International Journal of Applied Earth Observation and Geoinformation**, v. 34, n. 1, p. 51–57, 2015.

QIU, B. et al. Characterizing landscape spatial heterogeneity in multisensor images with variogram models. **Chinese Geographical Science**, v. 24, n. 3, p. 1–11, 2013.

SADER, S. A. Forest Harvest Patterns on an Industrial. n. October 2016, 2003.

SERTEL, E.; KAYA, S.; CURRAN, P. J. Use of semivariograms to identify earthquake damage in an Urban Area. **IEEE Transactions on Geoscience and Remote Sensing**, v. 45, n. 6, p. 1590–1594, 2007.

TSCHARMTKE, T. et al. Landscape perspectives on agricultural intensification and biodiversity - Ecosystem service management. **Ecology Letters**, v.8, n.8, p. 857-874. 2005.

TREITZ, P. High Spatial Resolution Remote Sensing Data for Forest Ecosystem Classification An Examination of Spatial Scale. **Remote Sensing of Environment**, v. 72, n. 3, p. 268–289, 2000.

VOROVENCII, I. Assessment of some remote sensing techniques used to detect land use/land cover changes in South-East Transilvania, Romania. **Environmental Monitoring and Assessment**, v. 186, n. 5, p. 2685–2699, 2014.

WEN, Z. et al. Effects of normalized difference vegetation index and related wavebands' characteristics on detecting spatial heterogeneity using variogram-based analysis. **Chinese Geographical Science**, v. 22, n. 2, p. 188–195, 2012.

WOODCOCK, C. E.; STRAHLER, A. H.; JUPP, D. L. B. The use of variograms in remote sensing: I. Scene models and simulated images. **Remote Sensing of Environment**, v. 25, n. 3, p. 323–348, 1988.

WHITE, D. A., HOOD, C.S., HARCOMBE, P. Vegetation patterns and environmental gradients in tropical dry forests of the northern Yucatan Peninsula. **Journal of Vegetation Science**, v. 15, n.2, p. 151-161. 2004.

WU, J. et al. Multiscale Analysis of Landscape Heterogeneity: Scale Variance and Pattern Metrics. **Annals of GIS**, v. 6, n. 1, p. 6–19, 2000.

WU, J. Key concepts and research topics in landscape ecology revisited : 30 years after the Allerton Park workshop. p. 1–11, 2013.

WU, X. et al. Evaluation of semivariogram features for object-based image classification. **Geo-spatial Information Science**, v. 18, n. 4, p. 159–170, 2015.

YUE, A. et al. Texture extraction for object-oriented classification of high spatial resolution remotely sensed images using a semivariogram. **International Journal of Remote Sensing**, v. 34, n. 11, p. 3736–3759, 2013.

ZACCARELLI, N. et al. Indicating disturbance content and context for preserved areas. **Ecological Indicators**, v. 8, n. 6, p. 841–853, 2008.

ZURLINI, G. et al. Disturbance patterns in a socio-ecological system at multiple scales. **Ecological Complexity**, v. 3, n. 2, p. 119–128, 2006.

**ARTICLE 2 - ASSESSMENT OF GEOSTATISTICAL FEATURES FOR
OBJECT-BASED IMAGE CLASSIFICATION OF CONTRASTED
LANDSCAPE VEGETATION COVER**

Eduarda Martiniano de Oliveira Silveira,^{a*} Michele Duarte Menezes,^b Fausto Weimar Acerbi Júnior,^a Marcela Castro Nunes Santos Terra^c, José Márcio de Mello^a

^a Federal University of Lavras, Forest Science Department, Campus UFLA, Lavras, Brazil, 3037

^b Federal University of Lavras, Soil Science Department, Campus UFLA, Lavras, Brazil, 3037

^c Federal University of Lavras, Engineering Department, Campus UFLA, Lavras, Brazil, 3037

Publication status: Published in the **Journal of Applied Remote Sensing**

10.1117/1.JRS.11.036004

Abstract. Accurate mapping and monitoring of savanna and semi-arid woodland biomes is needed to support the selection of new areas of conservation; to provide sustainable land use; and to improve the understanding of vegetation. This paper studies the potential of geostatistical features, derived from medium spatial resolution satellite imagery, to characterize contrasted landscape vegetation cover and improve object-based image classification. The study site in Brazil includes cerrado *sensu stricto*, deciduous forest, and palm swamp vegetation cover. Sentinel 2 and Landsat 8 images were acquired and divided into objects, for each of which a semivariogram was calculated using near-infrared (NIR) and normalized difference vegetation index (NDVI) to extract the set of geostatistical features. Features selected by principal component analysis (PCA) were used as input data to train a random forest (RF) algorithm. Tests were conducted, combining spectral and geostatistical features. Change detection evaluation was performed using a confusion matrix and its accuracies. The semivariogram curves were efficient to characterize spatial heterogeneity, with similar results using NIR and NDVI from Sentinel 2 and Landsat 8. Accuracy was significantly greater when combining geostatistical features with spectral data, suggesting that this method can improve image classification results.

Keywords: semivariogram, remote sensing, mapping, cerrado, deciduous forest, palm swamps

1.Introduction

Minas Gerais is the fourth-largest state in Brazil, covering an area (586,528 km²) equivalent to that of France or Spain. Savanna biome is predominant in the west/northwest, with warmer and wetter climate during summer, and a pronounced dry period. Semi-arid woodland biome is predominant in the northern area, characterized by semi-arid and sub-humid climates, with higher temperatures and lower rainfall throughout the year.¹

Accurate mapping and monitoring of savanna and semi-arid woodland biomes are needed to support the selection of new areas of conservation, to provide sustainable land use, and to improve the understanding vegetation.² Identification of land cover types provides basic information for generating other thematic maps, and establishes a baseline for monitoring activities.³ The production of accurate land cover maps is a difficult task, mainly due to large areas, great seasonality, land use pressure highly dynamic, and presence of clouds.⁴

Remote sensing data have successfully been used to map this biome across large areas and in inaccessible terrain.⁵ However, there are large areas of spectrally similar but compositionally different vegetation cover, or continuums of cover of varying densities of tree cover, that might be difficult to differentiate, potentially increasing classification uncertainty. Furthermore, most algorithms for image classification do not consider the spatial dependence between pixels and their neighbors. For discriminating land cover classes, it would be advisable to complement spectral bands with sensor information on the textural features of an image.⁶

Recently, many studies have used geostatistics, by means of semivariograms, to describe the textural and spatial features of remote sensing images.⁷⁻¹⁴ Semivariograms describe data variability patterns and can be used to investigate and quantify spatial variability.¹⁵

Object-based classifications have been made using semivariogram features for image classification. Ref. 16 used the semivariogram to extract texture information in agricultural parcel limits. Ref. 10 achieved high classification performance with semivariogram features, using various texture feature sets in rural parcels. Ref. 17 used semivariogram features with other spectral, textural, and shape features extracted from high-resolution imagery for object classification in urban areas. Ref.13 used semivariogram features to improve the accuracy of object-based classification using QuickBird satellite images. Ref. 18 used semivariogram features for object-based classification to map industrial disturbances in forest areas. For object-based classification of generic land cover types, Ref. 10 and Ref. 19 demonstrated better performance by combining spectral information and features derived from semivariograms.

Although semivariograms have been applied for image classification, few studies have focused on object-based classification using medium spatial resolution (10–100 m)³ remotely sensed images, such as Sentinel and Landsat, to map landscape vegetation areas.

Remote sensing has proved its value in many fields but the success of any image classification depends on various factors, including the choice of a suitable classification procedure.²⁰ Over the last two decades the use of the random forest (RF) classifier²¹ has received increasing attention due to the excellent classification results obtained and the speed of processing.²²⁻²⁶ The RF classifier is less sensitive than other streamline machine learning classifiers to the quality of training samples and to overfitting, due to the large number of decision trees produced by randomly selecting a subset of training samples and a subset of variables for splitting at each tree node.²⁴

Thus, the hypothesis here is: the NDVI (normalized difference vegetation index) and NIR (near-infrared) spatial variability provided by semivariogram features from different landscape vegetation cover could be used

as input data to train a RF classifier, thereby improving object-based classification of images. In this sense, the present study aims to: (1) characterize the NDVI and NIR spatial variability of contrasted landscape vegetation cover using sill parameter and semivariogram shape; (2) analyze the effects of image resolution from different satellite sources; and (3) evaluate a set of geostatistical features for object-based image classification.

2.Methodology

In this section, we describe the data, study area, proposed set of geostatistical features (semivariogram indices), and the steps involved in spatial variability characterization and image classification. The methodology consists of the following main steps, also presented in Fig. 1:

- (1) Image acquisition;
- (2) NDVI transformation and NIR data;
- (3) Collection of training samples of land cover vegetation types;
- (4) Image segmentation by a multiresolution algorithm to generate the image objects;
- (5) Feature extraction from NDVI and NIR computed within objects for semivariogram modeling;
- (6) Feature extraction from NDVI and NIR computed within objects for semivariogram and spectral features generation;
- (7) Principal component analysis (PCA) to group and interpret the redundancies in the information.
- (8) Image classification by Random Forest (RF) algorithm;
- (9) Classification evaluation by the confusion matrix and its accuracy measures.

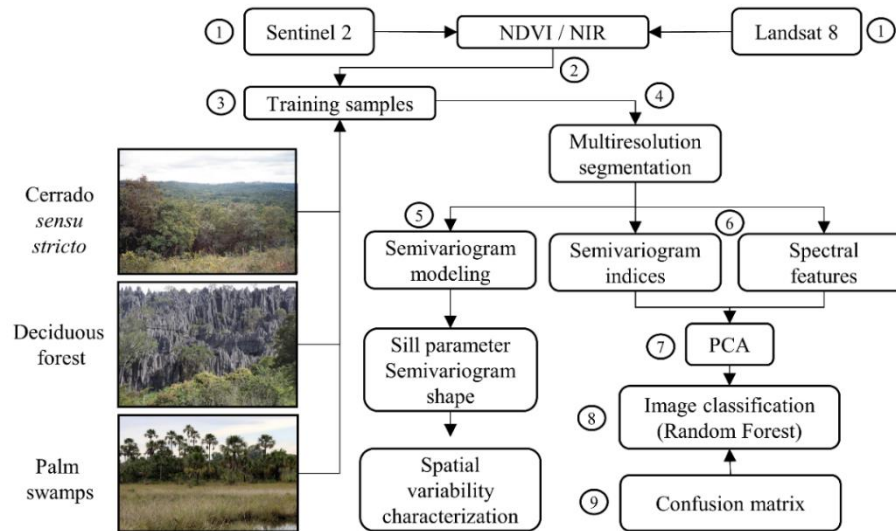


Fig. 1 Methodology workflow.

2.1 Study Area and Data Set

The study site is located in the north of Minas Gerais state, Brazil, with transition of savanna and semi-arid woodland biomes (Fig. 2). A site of 745 km² with contrasting landscape vegetation cover was chosen for semivariogram modeling.

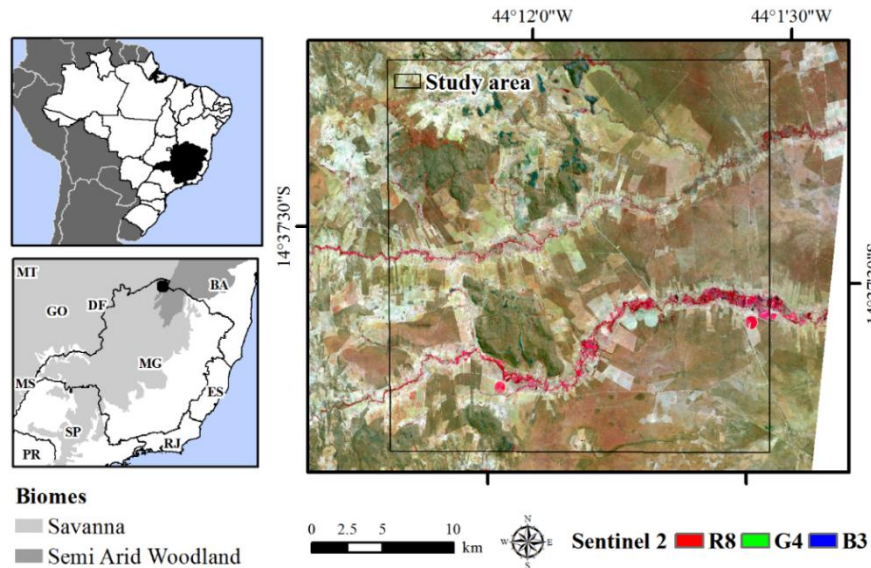


Fig. 2 Study area location in Minas Gerais state, Brazil.

Savanna biome includes vegetation cover ranging from dry grassland to densely wooded savanna.²⁷ In Brazil, savanna is the second-largest biome, after the Amazonian biome. Savanna is also a world biodiversity hotspot, including many endemic and rare species.²⁸ Semi-arid woodland is represented by deciduous seasonal forest, and is an ecosystem occupied by tropical dry forest and shrub vegetation.²⁷

The predominant land cover vegetation types in the study area are: cerrado sensu stricto, deciduous forest, and palm swamp (veredas). One of the main forms of landscape vegetation cover in savanna biome is cerrado sensu stricto²⁹, which demonstrates well-defined arboreal and shrub-herbaceous layers³⁰ with arboreal coverage varying from 10% to 60%. Deciduous forest is marked by a dry season and a well-defined rainy season. The main difference of this forest type in relation to the semi-deciduous seasonal forest is the predominance of deciduous individuals whose loss of foliage reaches more than

50% of these in the unfavorable season.³⁰ Palm swamps are included in the savanna biome. These areas are preferred for crops due to higher soil fertility and water availability.³¹ However, palm swamps are legally protected as Permanent Preservation Areas, according to the Brazilian Forest Code, and so any use is restricted.

Sentinel 2 and Landsat 8 satellite images from September 2016 were used. The data were acquired from the United States Geological Survey for Earth Observation and Science (USGS/EROS) at the processing level of Landsat Surface Reflectance, with the appropriate geometrical corrections and reflectance values to the soil level.

The Sentinel-2 mission, launched in 2015, is a land monitoring constellation of two satellites (Sentinel 2a and Sentinel 2b) providing global optical imagery with 13 spectral bands using the MSI (Multispectral Imager) instrument, and has spatial resolution of 10 m to 60 m. The Landsat 8 satellite, launched in 2013, has 11 bands and spatial resolution of 30 m to 100 m. From each corrected image, the NDVI and NIR bands were used.

The selection of sample classes was based on the map produced by the Forest Inventory of Minas Gerais project, which generated abundant qualitative and quantitative information about the forest remnants within the state.³²

2.2 Image Segmentation

In the object-based image analysis (OBIA), pixels are not individually classified but are combined into homogenous groups (objects) and classified together.³³⁻³⁵ The object becomes the basic unit of analysis, and is characterized by a large number of descriptive features derived from the images.

Objects have additional spectral information compared to single pixels (e.g. mean values, minimum and maximum values, standard deviation etc.),³⁶ but of even greater advantage than the diversification of spectral value

descriptions of objects is the additional spatial information for objects^{37,38}, including features derived from geostatistical functions.^{10,12,13,16-19} It has been frequently claimed that this spatial dimension is crucial to OBIA methods, and that this is a major reason for the marked increase in the usage of segmentation-based methods in recent times.³⁶

The first step in OBIA is image-object extraction, which is achieved by segmentation or stratification of the images that may be applied using external information to like-parcel boundaries.³⁹ Image segmentation is the division of the satellite image into spatially continuous and homogeneous objects.³⁵ Image segmentation is the core of OBIA, and various segmentation techniques have been developed around it.³⁶ In this study, the Sentinel 2 and Landsat 8 images were segmented according to their spectral and spatial attributes, using the multiresolution segmentation algorithm.⁴⁰

Multiresolution segmentation is a basic procedure in the software eCognition for object-based image analysis, and was used here to produce image objects as a first step before carrying out further feature extraction and image classification analysis. The multiresolution segmentation produces highly homogeneous image objects at an arbitrary resolution on different types of data.⁴⁰ Three key segmentation parameters (shape, compactness and scale) control the size, shape, and spectral variation of segmented image objects.⁴¹ The shape parameter was set to 0.1 and compactness was set to 0.5. The most critical step is the selection of the scale parameter, which controls the size of the image objects. The scale parameter sets a homogeneity threshold that determines the number of neighbouring pixels that can be merged together to form an image object.⁴² The scale of segmentation directly influences the size of the objects connected to the semivariogram predefined criteria (lag distance) and the minimum number of pixels inside each object necessary to generate the semivariogram. We used a “trial and error” approach⁴³ to find the appropriate

scale parameter⁴⁴ to guarantee a minimum number of samples (25 pixels) inside the objects.

2.3 Semivariogram Modeling

Spatial variability of landscape vegetation cover was characterized from semivariogram modeling of medium spatial resolution with NIR and NDVI data. We used the sill parameter and semivariogram shape. Representative forest remains of cerrado sensu stricto, deciduous forest, and palm swamp were selected, and corresponding semivariograms were generated using NIR and NDVI from both images (Fig. 3).

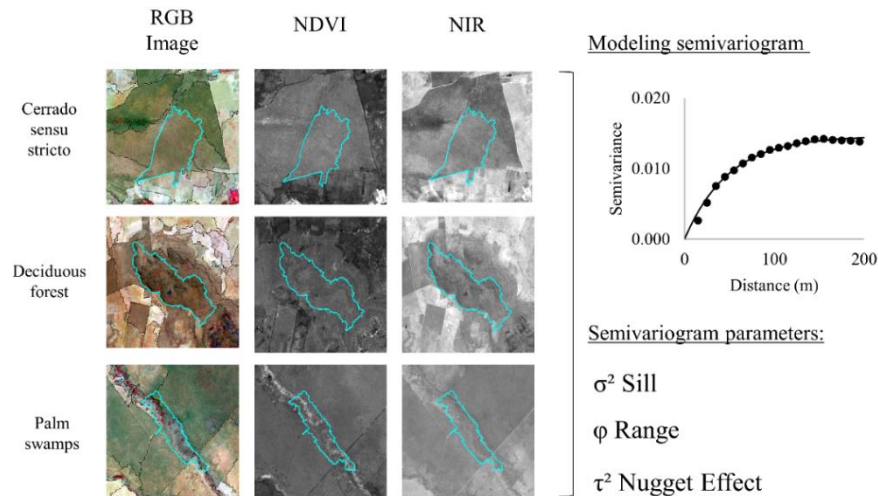


Fig. 3 Example of semivariogram modeling using NIR and NDVI data for three types of landscape vegetation cover.

For continuous variables such as NIR and NDVI, the experimental semivariogram is defined as half of the average squared difference between values separated by a given lag, where this lag is a vector in both distance and direction.¹⁵

The semivariance is defined from the spatial variance of measures performed in

samples from a determined distance (h) being the sum of the squares' difference between the sampled values separated by a distance (h), divided by two times the number of possible pairs on each distance, as estimated by Equation 1.

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x) - Z(x+h)]^2 \quad (1)$$

Here, $\gamma(h)$ is the estimator of the semivariance for each distance h, $N(h)$ is the number of pairs of points separated by the distance h, $Z(x)$ is the value of the regionalized variable at point x, and $Z(x+h)$ is the value of the point (x+h).

The semivariogram provides graphical representation of spatial variance versus distance h, which enables estimation of the variance for combinations of different pairs of points. The semivariance functions are characterized by three parameters: sill, range, and nugget effect. Sill is the plateau reached by the semivariance values, and shows the quantity of variation explained by the spatial structure of the data. Range is the distance at which the semivariogram reaches the sill, showing the distance until the point where the data are correlated. Nugget effect is the combination of sampling errors and variations in slight scales that occurs at scales smaller than the distance between the sampled points. The sill was used to characterize the landscape spatial variability of vegetation cover.

When the semivariogram is calculated for each individual object, an important factor to be considered is the lag distance. It should not be larger than the spatial extent of the object; on the other hand, an exceedingly small distance fails to provide a complete description of textural features. We attempted to find an optimal lag distance to ensure that sill values would provide a concise description of data variability. The interval used between two lags (distance between pairs of points of semivariogram calculation) was one pixel, in order to obtain maximum detail on the spatial heterogeneity of landscape vegetation cover. Thus, the lag size was equivalent to the size of the pixel (10 m for

Sentinel 2; 30 m for Landsat 8). The number of lags was fixed at 20 pixels to ensure that sill values provide a concise description of data variability, resulting in a lag distance of 200 m using Sentinel 2 and 600m using Landsat 8. The size of the samples needs to be larger than the range of influence in order to characterize the initial part of the semivariogram curves, and large enough to reveal the presence of periodicity.⁴⁵

The theoretical semivariograms were estimated by fitting mathematical models to the experimental semivariogram using weighted least squares. Exponential, spherical, and Gaussian models were tested.

2.4 Geostatistical Features

In addition to the semivariogram parameters commonly used (mentioned above), after computing the experimental semivariogram, indices proposed by Ref.10 were generated to evaluate geostatistical features as input data to image classification. These were included in the object-based descriptive feature, and the values were extracted by FETEX 2.0 software.⁴⁶

Such indices enable a more detailed description of experimental semivariogram shapes and spatial patterns, providing textural information that may be used for change detection analysis (Table 1). They are categorized as: (1) near the origin, and (2) up to the first maximum. The group of features is defined to provide such information as the change ratio, slope, concavity, and convexity level of the image at short distance.

Table 1 Parameters described in Ref.10 based on the points defining the experimental semivariogram.

Indices	Description	Formula
Near the Origin		
RVF	Ratio between the values of the total variance and the semivariance at first lag	$RVF = \frac{\text{Variance}}{\gamma_1}$
RSF	Ratio between semivariance values at second and first lag	$RSF = \frac{\gamma_2}{\gamma_1}$
FDO	First derivative near the origin	$FDO = \frac{\gamma_2 - \gamma_1}{h}$
SDT	Second Derivative at third lag	$SDT = \frac{\gamma_4 - 2\gamma_3 + \gamma_2}{h^2}$
Up to the First Maximum		
FML	The lag value where the curve reaches the first local maximum	$FML = h_{\max_1}$
MFM	Mean of the semivariogram values up to the first maximum	$MFM = \frac{1}{\text{Max_1}} \sum \gamma_i$
VFM	Variance of the semivariogram values up to the first maximum	$VFM = \frac{1}{\text{Max_1}} \sum (\gamma_i - \gamma)^2$
DMF	Difference between MFM and the semivariance at first lag	$DMF = MFM - \gamma_1$
RMM	Ratio between the semivariance at first local maximum and MFM	$RMM = \frac{\gamma_{\max_1}}{\gamma_{\max_1}^{\text{mean}}}$
SDF	Second-order difference between first lag and first maximum.	$SDF = \gamma_{\max_1} - 2\gamma_{\frac{\max_1}{2}} + \gamma_2$
AFM	Semivariance curvature	$AFM = \frac{h}{2} \left(\gamma_1 + 2 \left(\sum_{i=2}^{\max_1-1} \gamma_i \right) \right) \cdot (\gamma_1 (h_{\max_1} - h_1))$

According to Ref.10, an advantage of the proposed set of features, as opposed to the raw semivariance values, is that they synthesize the most relevant information about the shape of the semivariogram in a few features. They identify the singular points and enhance the information contained on the first lags, where spatial correlation at short distances is higher. In addition, fitting a

model is unnecessary, which improves the processing time and reduces the errors generated by choosing an incorrect model.

For comparative purpose, spectral features such as minimum (MIN), mean (MEAN), maximum (MAX) and standard deviation (DEVST) were included to assess the contribution of geostatistical features (semivariogram indices). And also, the combined performance of semivariogram indices and spectral features was tested. Since a large number of spectral and spatial features were generated, principal component analysis (PCA) was performed to group and interpret the redundancies in the information. Thus, the number of variables is reduced and multicollinearity is avoided.

2.5 Classification and Evaluation

Within the extent of the objects, the semivariogram indices and spectral data were extracted as inputs for training the classifier. A group of samples was selected manually by visual interpretation, and was divided into two parts for classifier training and classification assessment respectively. A data set of 300 objects, with 50 samples per class, was employed, and the samples were randomly allocated (50:50) to training or evaluation.

A RF classifier was used to model the landscape vegetation cover. The RF algorithm, initially proposed by Ref. 21, is an ensemble method which generates a set of individually trained decision trees and combines their results. As described by Ref. 47, such classification is based on a machine learning algorithm, is a robust non-parametric classifier, and has the ability to accommodate many predictor variables. The advantages of RFs include excellent accuracy, efficient implementation on large datasets, and a structure that enables the future use of pre-generated trees ²¹.

The RF algorithm fitted in this study is implemented in the open source software WEKA 3.8. Two parameters need to be set in order to produce the

forest trees: the number of decision trees to be generated (Ntree) and the number of variables to be selected and tested for the best split when growing the trees (Mtry).²⁴

Five hundred trees were grown for each classification.⁴⁸ The Ntree required to maintain a certain level of accuracy has been assessed by several authors, and the minimum number of trees for optimal classification appears to be somewhat fewer than 100⁴⁹ to 300 trees⁵⁰ and the majority of the studies set the Ntree value to 500 because the errors stabilize before this number of classification trees is achieved.⁴⁹ Therefore, using 500 may not be necessary, but does not harm the model.²¹ Since RF classifier is computationally efficient and does not overfit, Ntree can be as large as possible.⁵¹

Since theoretical and empirical research has highlighted that classification accuracy is less sensitive to Ntree than to the Mtry parameter⁵², the number of features was left at their default values (log of the number of features + 1)⁴⁸. The best results were identified using a cross-validation method.

Tests were done combining spectral and geostatistical features, resulting in three groups of classification assessment: (1) the spectral features, comprising the MIN, MEAN, MAX and DEVST chosen by PCA Analysis; (2) the geostatistical features, comprising the semivariogram indices chosen by PCA analysis and (3) the combination of spectral and geostatistical features.

Image classification evaluation was performed using a confusion matrix⁵³ and its accuracies: (1) the overall accuracy, which is computed by dividing the total number of correct results by the total number of samples in the error matrix; (2) the producer's accuracy, which indicates the probability of a reference pixel being correctly classified and is actually a measure of omission error; (3) the user's accuracy, which is indicative of the probability that a sample classified on the image actually represents that category on the ground and (4) the kappa coefficient, which is a statistic that measures the inter-rater agreement

for qualitative items. It is generally thought to be a more robust measure than simple percent agreement calculation, since it takes into account the possibility of the agreement occurring by chance.

3 Results and Discussion

3.1 Semivariogram Analysis

The semivariograms reached the sill within the calculated distance, indicating that their spatial extent of 20 pixels was large enough to encompass the entire spatial variability. Table 2 presents the sill semivariogram parameters obtained from each landscape vegetation cover using NIR and NDVI data from Sentinel 2 and Landsat 8 images.

Table 2 Sill semivariogram parameters.

Satellite image	Landscape vegetation cover	σ^2 NIR	σ^2 NDVI
Sentinel 2	Cerrado <i>sensu stricto</i>	9,002	0.00012
	Deciduous forest	27,278	0.00043
	Palm swamp	153,606	0.0146
Landsat 8	Cerrado <i>sensu stricto</i>	137,751	0.00007
	Deciduous forest	489,861	0.00012
	Palm swamp	2,069,336	0.0035

σ^2 - sill.

The Sentinel 2 spatial variability obtained for NIR and NDVI (represented by σ^2) increased from cerrado *sensu stricto* to deciduous forest and palm swamp. The two first vegetation covers have lower sill values, because of the presence of green understory, which decreases the variation explained by the spatial structure. The higher variability of palm swamps is explained by the influence of water, since it is riparian vegetation. Such landscape has a mixture of vegetation and water on NIR and NDVI values, increasing the internal variability within objects, contrasting high values for palm, and low

values for water.

The low variability of cerrado sensu stricto compared to deciduous forest is due to its dense vegetation with high density of trees: approximately 1,074 per hectare versus 973 trees/ha in deciduous forest.²⁷ It homogenizes the distribution of NIR and NDVI values. Furthermore, the presence of rock outcrops in deciduous forest increases the internal variability.

Analyzing the semivariogram parameters generated using NIR and NDVI from Landsat 8, the trends were the same as those for Sentinel 2 (Table 2). The sill (σ^2) semivariogram parameter increased considerably from cerrado sensu stricto and deciduous forest to palm swamp. According to Ref.7, semivariogram shape provides an understanding of the data spatial structures. The semivariogram curves of vegetation cover obtained from NIR and NDVI are shown for Sentinel 2 (Fig. 4) and Landsat 8 (Fig. 5).

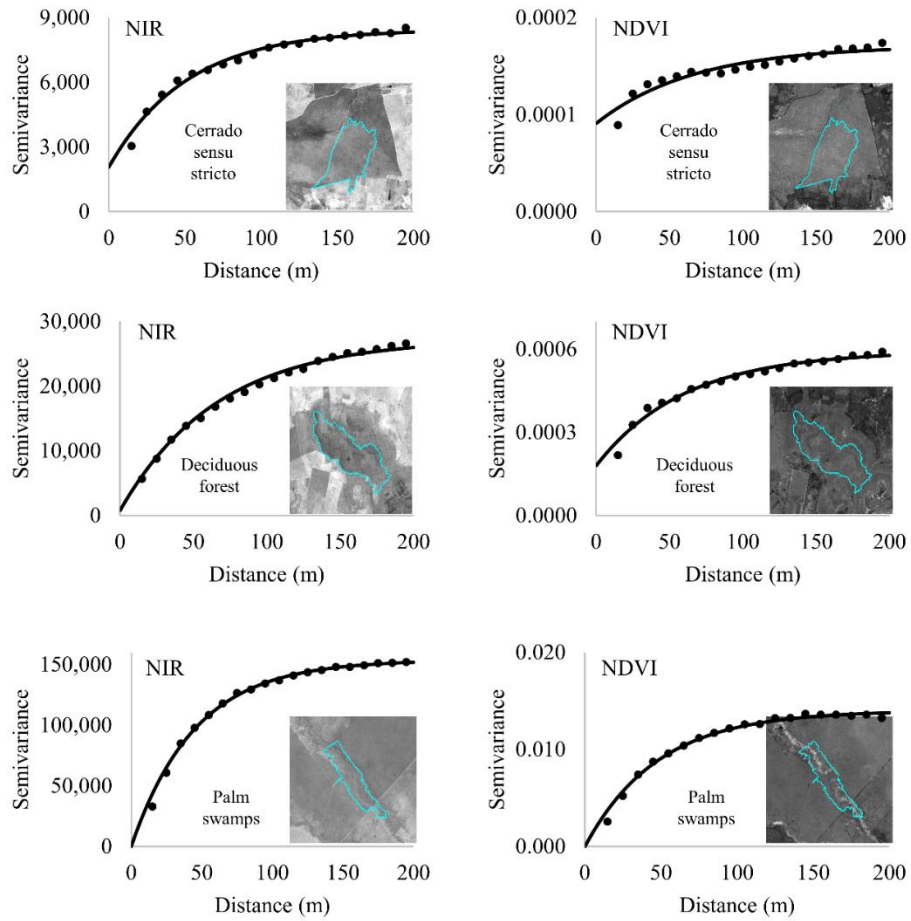


Fig. 4 Semivariograms generated using Sentinel 2.

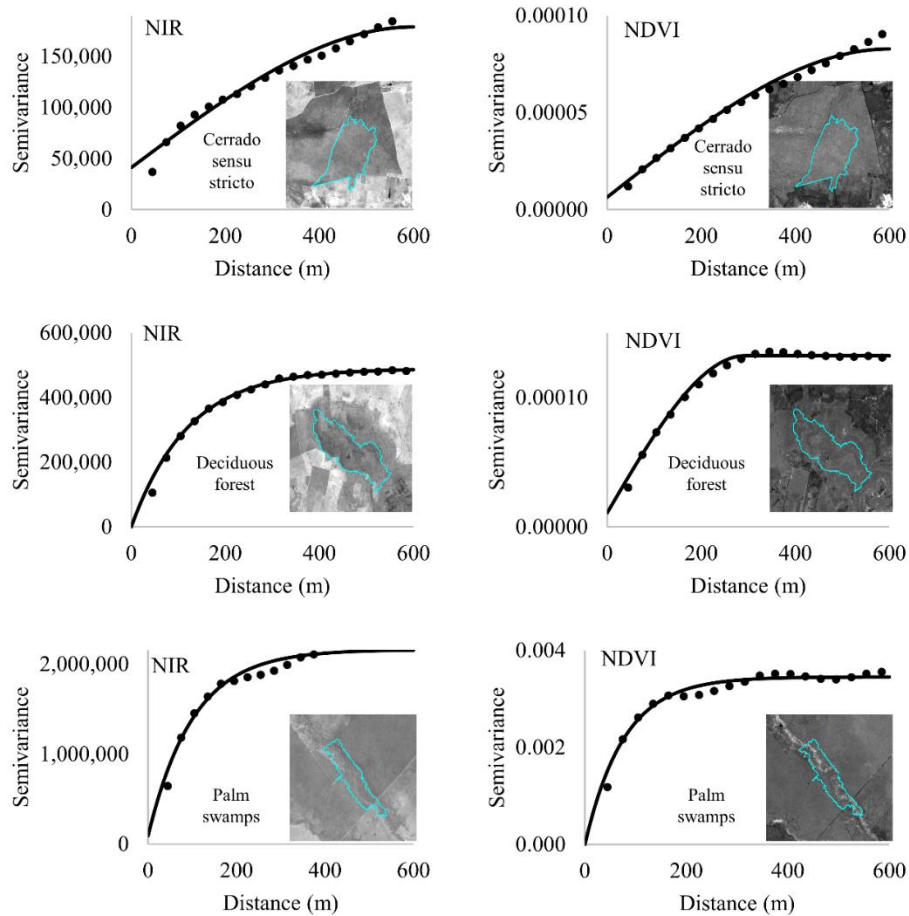


Fig. 5 Semivariograms generated using Landsat 8.

The semivariogram curves using NIR and NDVI from Sentinel 2 and Landsat 8 presented similar trends for the three contrasted landscape vegetation covers. The range among palm swamps and the other classes was higher with NDVI; however, it was smaller between cerrado sensu stricto and deciduous forest. The semivariogram curves described efficiently the spatial heterogeneity of landscape vegetation cover and also presented similar trends for the three contrasted vegetation types. These results demonstrate that the image spatial

resolution tested (10 and 30 m) did not affect semivariogram characterization.

3.2 Feature Selection

PCA computed across the complete set of features shows that most of the data variability is concentrated within the first two principal components (Table 3). Therefore, the features of each group (Spectral, Near the Origin and Up to First Maxima) with higher correlations to the first two PCA components were selected. Visualization of the ordinations graphics enables better understanding of redundancies (Fig. 6).

Table 3 Selected features according to the PCA.

	Spectral	Near the Origin	Up to First Maxima	PC1 / PC2
Sentinel NIR	DEVST MEAN	FDO RVF	MFM SDF FML	60.56% of total variation
Sentinel NDVI	DEVST MAX	FDO SDT	MFM SDF RMM	62.90% of total variation
Landsat NIR	DEVST MEAN	FDO RVF	MFM SDF FML	55.68% of total variation
Landsat NDVI	DEVST MIN	FDO RVF	MFM RMM	53.28% of total variation

In the group of spectral features, near the origin, and up to first maxima, the representative ones, were the DEVST, FDO and MFM, respectively. Thus, we recommend these group of features as general input data to train a classifier. Among the spectral features, the DEVST is the unique that represent spatial information, and probably this characteristic made this feature the most important. About the group of features that provides information near the origin, FDO represents the slope of the semivariogram at the first two lags,

approximating the first derivative near the origin. It shows the variability changes of the data at short distances. MFM is the mean of the semivariogram values up to the first maximum. It is an indicator of the average of the semivariogram values between the first lag and the first maximum. It provides information about the changes in the variability of the data, and is related to the concavity or convexity of the semivariogram in that interval.

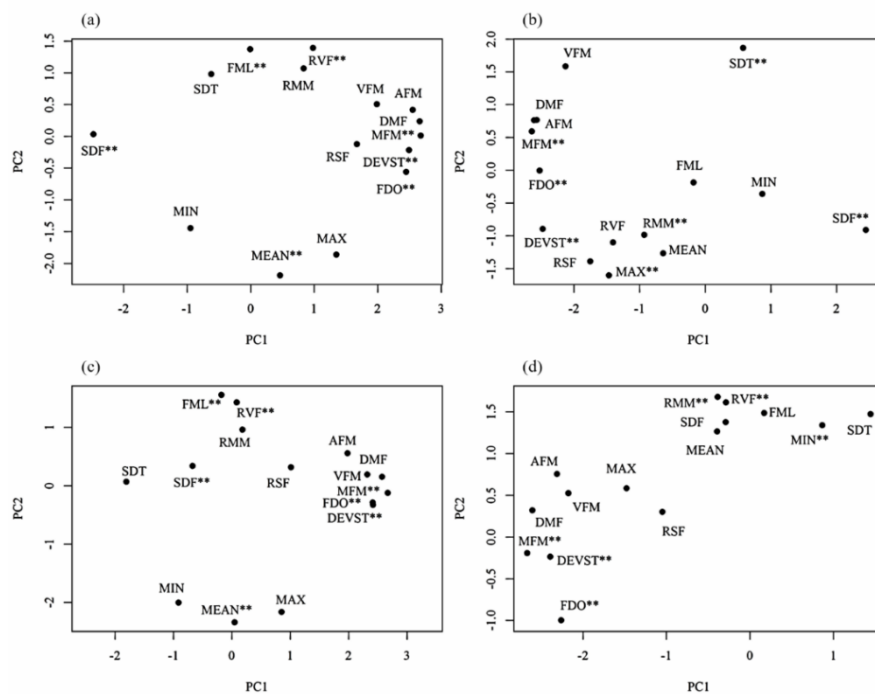


Fig. 6. Projection of features in principal component planes: (a) Sentinel NIR; (b) Sentinel NDVI; (c) Landsat NIR; (d) Landsat NDVI.

3.3 Classification Assessment

Table 4 compares the producer, user, and overall accuracies obtained using Sentinel 2 imagery with NIR and NDVI data. The features tested are effective for classifying landscape vegetation cover, presenting overall

accuracies greater than 80%.

The highest overall accuracy (93.63%) was obtained by combining NDVI and the indices provided by the semivariogram, followed by the combination of NIR spectral data and spatial indices (91.81%). The lowest overall, user, and producer accuracies were obtained using only the indices. Analyzing the classes' accuracies, the Cerrado *sensu stricto* and Deciduous forest classes also presented the highest producer's and user's accuracies using NDVI and the semivariogram indices with all values higher than 86%. Most important to notice is that the Deciduous forest, a difficult class to map, presented only 1.92% of commission errors, without any omission errors, using this mapping strategy. The exception was the Palm swamp class that presented the lowest commission error (user accuracy equal to 84 %) using NIR and spatial indices.

Table 4 Classification results using Sentinel 2.

Classes	NIR		Indices		NIR + Indices	
	PA	UA	PA	UA	PA	UA
Cerrado <i>sensu stricto</i>	85.71	90.91	77.14	81.82	85.29	87.88
Deciduous forest	90.91	96.15	85.45	90.38	92.73	98.08
Palm swamp	100.00	80.00	100.00	80.00	100.00	84.00
OA	90.90		85.45		91.81	
Classes	NDVI		Indices		NDVI + Indices	
	PA	UA	PA	UA	PA	UA
Cerrado <i>sensu stricto</i>	80.56	87.88	84.38	81.82	86.49	96.97
Deciduous forest	96.15	96.15	82.14	88.46	100.00	98.08
Palm swamp	90.91	80.00	86.36	76.00	90.91	80.00
OA	90.91		83.63		93.63	

OA – overall accuracy; PA – producer accuracy; UA – user accuracy.

Using Landsat 8 image, the highest overall accuracy (95.41%) was obtained by combining NDVI spectral data and the indices provided by the semivariogram, followed by the combination of NIR and spatial indices

(91.74%). The lowest overall accuracy was obtained using only the geostatistical indices (69.72% and 88.08%) for NIR and NDVI, respectively (Table 5). Analyzing the classes' accuracies, combining NDVI and the indices was the best strategy for all classes, with values higher than 87%. The exception was the Palm swamp class that presented the lowest commission error (user accuracy equal to 91.67 %) using only the indices.

Table 5 Classification results using Landsat 8.

Classes	NIR		Indices		NIR + Indices	
	PA	UA	PA	UA	PA	UA
Cerrado <i>sensu stricto</i>	85.29	87.88	78.13	75.76	91.18	93.94
Deciduous forest	90.57	92.31	67.21	78.85	94.23	94.23
Palm swamp	81.82	75.00	62.50	41.67	86.96	83.33
OA	87.15		69.72		91.74	
Classes	NDVI		Indices		NDVI + Indices	
	PA	UA	PA	UA	PA	UA
Cerrado <i>sensu stricto</i>	87.88	87.88	84.85	84.85	94.12	96.97
Deciduous forest	92.59	96.15	88.46	88.46	96.23	98.08
Palm swamp	90.91	83.33	91.67	91.67	95.45	87.50
OA	90.82		88.07		95.41	

OA – overall accuracy; PA – producer accuracy; UA – user accuracy.

The results show significantly increased accuracy by combining geostatistical features with spectral data for all types of vegetation cover, considering also NIR and NDVI, and Sentinel 2 and Landsat 8 images.

4 Conclusions

This study examined whether spatial variability described by semivariograms for different types of landscape vegetation cover, derived from NIR and NDVI data from Sentinel 2 and Landsat 8, has potential use as input data to train a RF classifier and improve object-based image classification.

We demonstrated that: (1) semivariogram curves were efficient for

characterizing spatial heterogeneity within contrasting landscape vegetation cover; (2) the results from Sentinel 2 and Landsat 8 satellite images were similar using both NIR and NDVI data; and (3) accuracy was significantly improved by combining geostatistical features with spectral data, thereby supporting the use of the selected indices to improve image classification procedures.

The two major limitations of the approach used in this study are the predefined criteria required to generate the semivariogram and the object size of the OBIA approach. An image object should be adequately sized to sufficiently represent its textural pattern, with the minimum number of necessary pixels to generate the semivariogram. The main limitation is the presence of long and narrow objects.

The relationship between the scale of segmentation and the lag distance need be further explored, and also we recommend further studies to improve this methodology and its evaluation in other areas/vegetation types for comparison purposes.

5 References

1. H. F. Scolforo et al., "Spatial interpolators for improving the mapping of carbon stock of the arboreal vegetation in Brazilian biomes of Atlantic forest and savanna," *For. Ecol. Manage.* **376**, 24–35 (2016).
2. E. E. Sano et al., "Land cover mapping of the tropical savanna region in Brazil," *Env. Monit Assess.* **166**, 113–124 (2010).
3. C. Gomez, J. C. White, and M. A. Wulder, "Optical remotely sensed time series data for land cover classification: A review," *ISPRS J. Photogramm. Remote Sens.* **116**, 55–72 (2016).
4. G. B. S. Silva et al., "Discriminação da cobertura vegetal do Cerrado matogrossense por meio de imagens MODIS," *Pesqui. Agropecu. Bras.*, **45**(2), 186–194 (2010).

5. M. Schwieder et al., "Mapping Brazilian savanna vegetation gradients with Landsat time series," *Int. J. Appl. Earth Obs. Geoinf.* **52**, 361–370 (2016).
6. M. Chica-Olmo and F. Abarca-Hernández, "Computing geostatistical image texture for remotely sensed data classification," *Comput. Geosci.* **26**(4), 373–383 (2000).
7. S. Garrigues, D. Allard, F. Baret, and M. Weiss, "Quantifying spatial heterogeneity at the landscape scale using variogram models," *Remote Sens. Environ.* **103**(1), 81–96 (2006).
8. S. Garrigues et al., "Multivariate quantification of landscape spatial heterogeneity using variogram models," *Remote Sens. Environ.* **112**(1), 216–230 (2008).
9. S. Berberoglu and A. Akin, "Assessing different remote sensing techniques to detect land use/cover changes in the eastern Mediterranean," *Int. J. Appl. Earth Obs. Geoinf.* **11**(1), 46–53 (2009).
10. A. Balaguer et al., "Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification," *Comput. Geosci.* **36**(2), 231–240 (2010).
11. F. Van der Meer, "Remote-sensing image analysis and geostatistics," *Int. J. Remote Sens.* **33**, 5644–5676 (2012)
12. A. Balaguer-Beser et al., "Using semivariogram indices to analyse heterogeneity in spatial patterns in remotely sensed images," *Comput. Geosci.* **50**, 115–127 (2013).
13. A. Yue et al., "Texture extraction for object-oriented classification of high spatial resolution remotely sensed images using a semivariogram," *Int. J. Remote Sens.* **34**(11), 3736–3759 (2013).
14. J. L. Gil-Yepes et al., "Description and validation of a new set of object-based temporal geostatistical features for land-use/land-cover change detection," *ISPRS J. Photogramm. Remote Sens.* **121**, 77–91 (2016).

15. P. J. Curran, "The semivariogram in remote sensing: An introduction," *Remote Sens. Environ.* **24**(3), 493–507 (1988).
16. S. Berberoglu et al., "The integration of spectral and textural information using neural networks for land cover mapping in the Mediterranean," *Comput. Geosci.* **264**, 385–396 (2000).
17. T. Hermosilla et al., "Change detection in peri-urban areas based on contextual classification," *Photogramm. Fernerkund. Geoinf.* **4**, 355–366 (2012).
18. R. P. Powers et al., "Remote sensing and object-based techniques for mapping fine-scale industrial disturbances," *Int. J. Appl. Earth Obs. Geoinf.* **34** (1), 51–57 (2015).
19. X. Wu et al., "Evaluation of semivariogram features for object-based image classification," *Geo-spatial Inf. Sci.* **18**(4), 159–170 (2015).
20. D. Lu and Q. A. Weng, "A survey of image classification methods and techniques for improving classification performance," *Int. J. Remote Sens.* **28**, 823–870 (2007).
21. L. Breiman, "Random forests," *Mach. Learn.* **45**(1), 5–32 (2001).
22. B. Ghimire et al., "Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic," *Remote Sens. Lett.* **1**, 45–54 (2010).
23. V. F. Rodriguez-Galiano et al., "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS J. Photogramm. Remote Sens.* **67**, 93–104 (2012).
24. M. Belgiu and L. Drăgu, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.* **114**, 24–31 (2016).
25. K. J. Wessels et al., "Rapid land cover map updates using change detection and robust random forest classifiers," *Remote Sens.* **8**, 1–24 (2016).

26. M. F. A. Vogels et al., “Agricultural cropland mapping using black-and-white aerial photography, Object-Based Image Analysis and Random Forests,” *Int. J. Appl. Earth Obs. Geoinf.* **54**, 114–123 (2017).
27. H. F. Scolforo et al., “Spatial distribution of aboveground carbon stock of the arboreal vegetation in Brazilian biomes of savanna, Atlantic Forest and Semi-arid woodland,” *PLoS One.* **10**(6), 1–20 (2015).
28. N. Myers et al., “Biodiversity hotspots for conservation priorities,” *Nature.* **403**(6772), 853–858 (2000).
29. T. R. R. Santos et al., “The tree-shrub vegetation in rocky outcrop cerrado areas in Goiás State, Brazil,” *Brazilian J. Bot.* **35**(3), 281–294 (2012).
30. J. F. Ribeiro and B.M.T. Walter, “As principais fitofisionomias do Bioma Cerrado”, in *Cerrado: Ecologia e Flora*, S.M. Sano, S.P. Almeida and J.F. Ribeiro, Eds., pp. 151–212, Embrapa Cerrados, Planaltina, Brazil (2008).
31. J. F. Ribeiro and B.M.T. Walter, “Fitofisionomias do bioma Cerrado”, in S.M. Sano and S.P. Almeida, Eds., pp. 89–166, Embrapa-CPAC, Planaltina, Brazil (1998).
32. L. M. T. Carvalho and J. R. S. Scolforo, “Inventário florestal de Minas Gerais: Monitoramento da Flora Nativa 2005-2007”, UFLA, Lavras (2008).
33. C. Chen et al., “Object-based change detection,” *Int. J. Remote Sens.* **33**(14), 4434–4457 (2012).
34. B. Desclée, P. Bogaert, and P. Defourny, “Forest change detection by statistical object-based method,” *Remote Sens. Environ.* **102**(1–2), 1–11 (2006).
35. M. Hussain et al., “Change detection from remotely sensed images: From pixel-based to object-based approaches,” *ISPRS J. Photogramm. Remote Sens.* **80**, 91–106 (2013).
36. T. Blaschke, “Object based image analysis for remote sensing,” *ISPRS J. Photogramm. Remote Sens.* **65**, 2–16 (2010).

37. R. M. Haralick et al., "Textural Features for Image Classification," *Syst. Man Cybern. IEEE Trans.* **6**, 610–621 (1973).
38. Q. Chen and P. Gong, "Automatic Variogram Parameter Extraction for Textural Classification of the Panchromatic IKONOS Imagery," *IEEE Trans.* **42**, 1106–1115 (2004).
39. E. A. Addink et al., "Introduction to the GEOBIA 2010 special issue: From pixels to geographic objects in remote sensing image analysis," *Int. J. Appl. Earth Obs. Geoinf.* **15**, 1–6 (2012).
40. M. Baatz and A. Schäpe, "Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation," *J. Photogramm. Remote Sens.* **58**(3–4), 12–23 (2000).
41. A. Mui et al., "An object-based approach to delineate wetlands across landscapes of varied disturbance with high spatial resolution satellite imagery," *ISPRS J. Photogramm. Remote Sens.* **109**, 30–46 (2015).
42. U. C. Benz et al., "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information," *ISPRS J. Photogramm. Remote Sens.* **58**, 239–258 (2004).
43. D. C. Duro et al., "A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery," *Remote Sens. Environ.* **118**, 259–272 (2012).
44. J. Yang et al., "An Automated Method to Parameterize Segmentation Scale by Enhancing Intra-segment Homogeneity and Inter-segment Heterogeneity," *IEEE Geosci. Remote Sens. Lett.* **6**(1), 29–38 (2015).
45. C. E. Woodcock, A. H. Strahler, and D. L. B. Jupp, "The use of variograms in remote sensing: I. Scene models and simulated images," *Remote Sens. Environ.*, **25**(3), 323–348 (1988).

46. L. A. Ruiz et al., "A feature extraction software tool for agricultural object-based image analysis," *Comput. Electron. Agric.* **76**(2), 284–296 (2011).
47. B. Devries et al., "Characterizing forest change using community-based monitoring data and Landsat time series," *PLoS One*. **11**(3), 1–25 (2016).
48. K. Millard and M. Richardson, "On the importance of training data sample selection in Random Forest image classification: A case study in peatland ecosystem mapping," *Remote Sens.* **7**, 8489–8515 (2015).
49. R. L. Lawrence, "Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest)," *Remote Sens.* **100**, 356–362 (2006).
50. O. Akar and O. Gungor, "Integrating multiple texture methods and NDVI to the Random Forest classification algorithm to detect tea and hazelnut plantation areas in northeast Turkey," *Int. J. Remote Sens.* **36**, 442–464 (2015).
51. H. Guan et al., "Integration of orthoimagery and lidar data for object-based urban thematic mapping using random forests," *Int. J. Remote Sens.* **34**, 5166–5186 (2013).
52. A. Ghosh and P. K. A., "H. Guan et al., "comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery," *Int. J. Appl. Earth Obs. Geoinf.* **26**, 298–311 (2014).3
53. R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sens. Environ.* **37**(1), 35–46 (1991).

**ARTICLE 3 - OBJECT-BASED CHANGE DETECTION USING
SEMIVARIOGRAM INDICES DERIVED FROM NDVI IMAGES: THE
ENVIRONMENTAL DISASTER IN MARIANA, BRAZIL**

**Detecção de mudanças baseada em objetos utilizando índices do
semivariograma derivados de imagens NDVI: O desastre ambiental em
Mariana, Brasil**

Eduarda Martiniano de Oliveira Silveira^{1*}, Fausto Weimar Acerbi Júnior¹, José
Márcio de Mello¹, Inácio Thomaz Bueno¹

¹Universidade Federal de Lavras/UFLA, Departamento de Ciências
Florestais/DCF, Lavras, MG, Brasil

Publication status: Published in **Ciência e Agrotecnologia**

doi: 10.1590/1413-70542017415009817

ABSTRACT: Object-based change detection is a powerful analysis tool for remote sensing data, but few studies consider the potential of temporal semivariogram indices for mapping land-cover changes using object-based approaches. In this study, we explored and evaluated the performance of semivariogram indices calculated from remote sensing imagery, using the Normalized Differential Vegetation Index (NDVI) to detect changes in spatial features related to land cover caused by a disastrous 2015 dam failure in Brazil's Mariana district. We calculated the NDVI from Landsat 8 images acquired before and after the disaster, then created objects by multiresolution segmentation analysis based on post-disaster images. Experimental semivariograms were computed within the image objects and semivariogram indices were calculated and selected by principal component analysis. We used the selected indices as input data to a support vector machine algorithm for classifying change and no-change classes. The selected semivariogram indices showed their effectiveness as input data for object-based change detection analysis, producing highly accurate maps of areas affected by post-dam-failure flooding in the region. This approach can be used in many other contexts for rapid and accurate assessment of such land-cover changes.

Index terms: Remote sensing; geostatistics; feature extraction

RESUMO: Recentemente, variáveis geostatísticas derivadas de imagens de sensoriamento remoto ganharam espaço dentre os procedimentos de detecção de mudanças, porém, o potencial temporal destas variáveis para o mapeamento das mudanças baseado na análise por objetos ainda é pouco estudado. Neste estudo, o desempenho de um conjunto de índices calculados de semivariogramas derivados de imagens NDVI bitemporais para detectar mudanças na cobertura do solo foi analisado e avaliado. O município de Mariana foi selecionado para teste e validação da metodologia devido ao grande impacto ocasionado pelo desastre. O processo iniciou-se com a aquisição de imagens Landsat 8 antes e após o desastre e o cálculo do NDVI. Os objetos foram criados através da segmentação em multiresolução baseada na imagem pós-desastre. Os semivariogramas experimentais foram gerados dentro de cada objeto e os índices foram extraídos e selecionados através da análise de componentes principais. Os índices selecionados foram utilizados como dados de entrada para o algoritmo support vector machines para a classificação de áreas de mudança e não mudança. Os índices selecionados se mostraram efetivos para a detecção de mudanças, indicando a possibilidade de utilização para a detecção de mudanças baseada em objetos, resultando em um mapa precisos das áreas inundadas afetadas pelo desastre. Esta abordagem pode ser usada em muitos outros contextos para uma avaliação rápida e precisa de tais mudanças na cobertura do solo.

Termos para indexação: Sensoriamento remoto; geostatística; extração de atributos.

INTRODUCTION

The collapse of a mining dam in the Brazilian state of Minas Gerais on November 5th 2015, considered one of the biggest environmental disasters in the country's history, resulted in the destruction of whole communities by a river of mud and mining waste. This calamity affected the Gualaxo River, a tributary to the Carmo River and ultimately the Doce River, waterways that supply water to a significant number of municipalities. The flood affected 600 kilometers of riverbed and destroyed human and animal lives as well as several land-cover classes (such as grasslands, urban areas, and native vegetation), including in permanent preservation areas. The full extent of the environmental impacts is yet unknown, and the changes within the affected area have yet to be fully quantified.

Remote sensing techniques are effective in capturing the structure, rates, and changes of land cover. They can supply essential information concerning the ecological status of a region, including changes that modify plant phenological standards and deforestation (Munroe; Southworth; Tucker 2002; Tucker et al., 2005; Yue et al., 2003). The Normalized Difference Vegetation Index (NDVI) is an important approach to the analysis of land-cover structure analysis and its temporal modifications (Griffith et al., 2007). According to Costantini et al. (2012) and Garrigues et al. (2006), NDVI images are the most robust variable used to describe the spatial and temporal heterogeneity of a landscape's biosphere. In addition, these data can be treated as regionalized variables once the information contained in a pixel is highly correlated with the information contained in neighboring pixels (Acerbi Junior et al., 2015; Curran, 1988).

Studies of environmental disasters have emphasized the importance of damage determination to assist environmental management programs and stressed the use of remote sensing images and geostatistical techniques as central

tools for this kind of analytical approach (Sertel; Kaya; Curran, 2007). Combining remote sensing information with GIS techniques and geospatial databases can increase the accuracy and reduce the processing time of change detection and classification procedures (Berberoglu et al., 2000; Berberoglu; Akin 2009; Garcia-Pedrero et al., 2015).

For example, semivariograms are an analytical technique used to assess the relationship and variance between points based on distance and a given variable. These have been used as measures of texture (Curran 1988; Woodcock; Strahler; Jupp et al. 1988), for improved image classification (Balaguer et al., 2010; Balaguer-Beser et al., 2011; Wu et al., 2015 Yue et al., 2013; Powers et al., 2015), and more recently, in change detection studies (Costantini et al., 2012; Sertel et al., 2007; Gil-Yepes et al., 2016). Acerbi Junior et al. (2015) demonstrated the potential of semivariogram parameters (derived from bitemporal NDVI images) to detect changes in Brazilian savanna vegetation, showing that these parameters increased on deforested areas and remained constant in regions where the land cover had not changed.

In recent years, semivariograms have also contributed to object-based image analysis (OBIA) (Meer, 2012). Powers et al. (2015) used semivariogram features and OBIA for classification of industrial disturbances in forest areas. Balaguer et al. (2010) achieved high-accuracy measurements by combining semivariogram features and spectral information in land cover mapping. Gil-Yepes et al. (2016) proposed and evaluated a set of new temporal geostatistical features for object-based change detection (OBCD) analysis within agricultural plots at two different dates, showing that the new set of cross-semivariogram and codispersion features provided high global accuracy measures when compared to the use of only spectral information.

Textural features have proven to be more effective than spectral bands alone for change detection (Chen et al., 2012; Wu et al., 2000). However, few

studies have explored the potential of temporal semivariogram features for mapping land cover changes using the OBCD approach. We hypothesized that landscape changes could be accurately detected using only semivariograms calculated from NDVI images and so we explored and evaluated the performance of semivariogram indices in an object-based approach to detecting land-cover changes caused by the 2015 dam-collapse disaster in Brazil.

MATERIAL AND METHODS

We derived the NDVI from Landsat 8 images for use in an object-based change detection approach to analyzing land-cover changes in the afflicted area, using the following methodology (graphically summarized in Figure 1):

- (1) Image acquisition and NDVI transformation
- (2) Object delimitation by multiresolution algorithm based on the post-disaster image
- (3) Experimental semivariogram computed within the objects
- (4) Generation of semivariogram indices, as proposed by Balaguer et al. (2010)
- (5) Selection of the most important semivariogram indices by PCA analysis
- (6) Change detection using the Support Vector Machine (SVM) algorithm
- (7) Evaluation by the confusion matrix and its accuracy measures

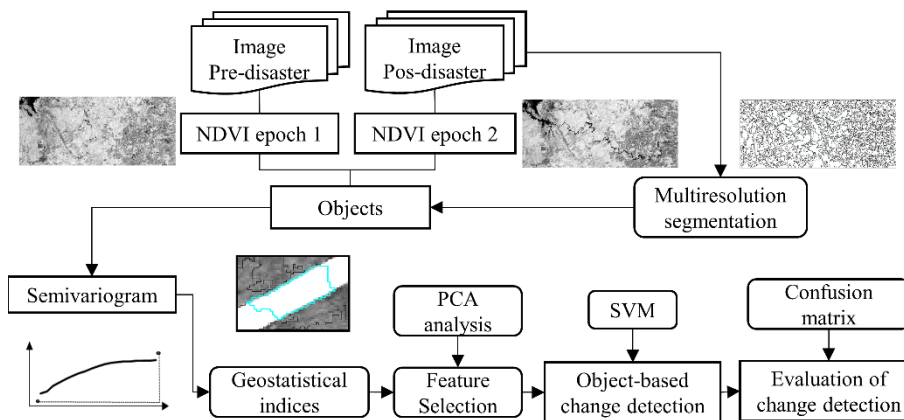


Figure 1: Methodology workflow.

Study area and data

The district of Mariana is located in the central region of Minas Gerais state, Brazil, between the 43° 05' 00" and 43° 30' 00" meridians and the 20° 08' 00" and 20° 35' 00" parallels (Figure 2). The district includes the upper portion of the Doce River basin and is characterized by hilly relief and abundant tablelands. The climatic conditions are typical of humid tropical highlands, with hot and rainy summers. The vegetation is predominantly composed of Atlantic Forest and Savanna biomes.

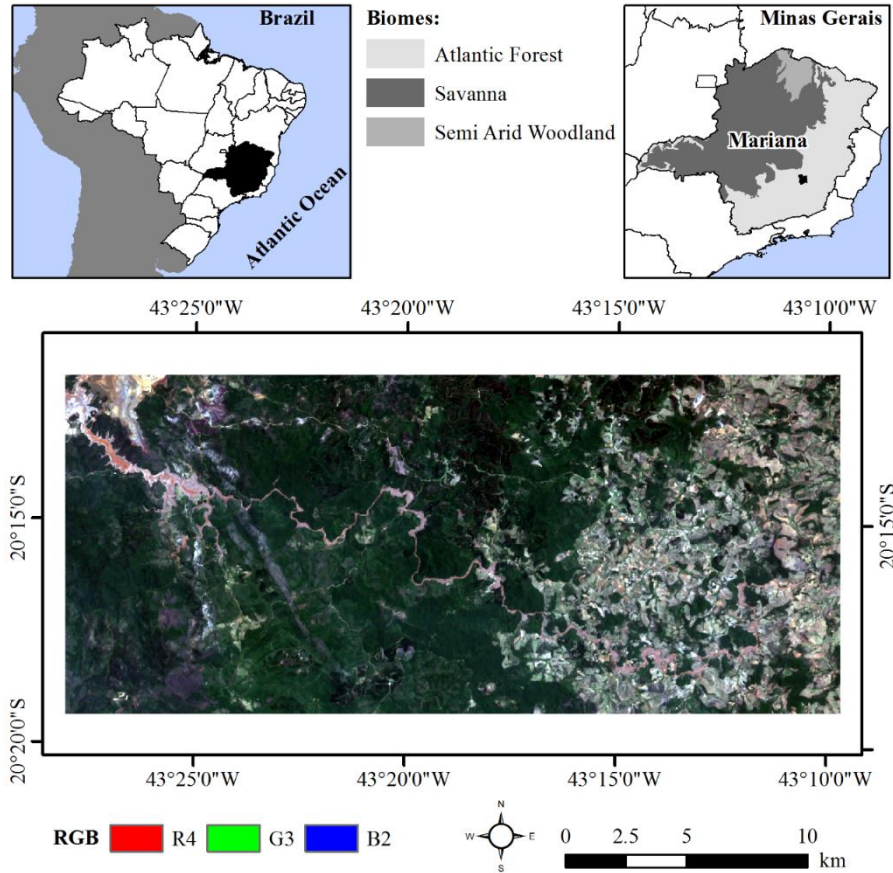


Figure 2: Study area location within Minas Gerais state, Brazil.

We acquired Landsat 8 satellite images from the United States Geological Survey for Earth Observation and Science (USGS/EROS) from October 2015 (predisaster) and November 2015 (post-disaster), at the processing level of Landsat Surface Reflectance, with the appropriate geometrical corrections and reflectance values to the soil level. We then generated the NDVI, which is based on quotients and uses the spectral bands from the red and near-infrared bands to enhance vegetative characteristics and minimize the effects of shadows caused by the terrain's topography (Berra et al., 2012; Vorovencii,

2014). The values of this index vary from -1 to 1, calculated as:

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}} \quad (1)$$

where ρ_{NIR} and ρ_{RED} are the reflectance values for the near-infrared and red wavelengths, respectively.

Image segmentation

In the object-based change detection method, pixels are not individually classified but rather combined into homogenous groups (objects) and classified together (Chen et al., 2012; Desclée; Bogaert; Defourny 2006; Hussain et al., 2013). The object is characterized using a large number of descriptive features derived from the images and becomes the basic unit of analysis. In comparison with pixel-based methods, additional spatial and contextual information can be obtained from the objects (Blaschke 2010; Hussain et al., 2013; Ruiz et al., 2011; Wu et al., 2015).

Object-based semivariogram analysis is based on the delimitation of homogeneous groups, in which the objects' boundaries are pre-defined and the semivariogram features are extracted from each object. Multiresolution segmentation is a basic procedure in the eCognition software employed in this study; we used a multiresolution segmentation algorithm (Baatz; Schäpe, 2000) to generate objects based on the post-disaster NDVI image. The size, shape, and spectral variation of each object are controlled by three key segmentation parameters: shape, compactness, and scale. The shape parameter was set to 0.1 and the compactness to 0.5. The most critical step is the selection of the scale parameter, which controls the size of the image objects. This sets a threshold of homogeneity determining how many neighboring pixels can be merged together to form an image object (Mui et al., 2015). We tested values from 80 to 200 for this parameter and obtained the best segmentation result using the value 150. Figure 3 shows the image segmentation procedure.

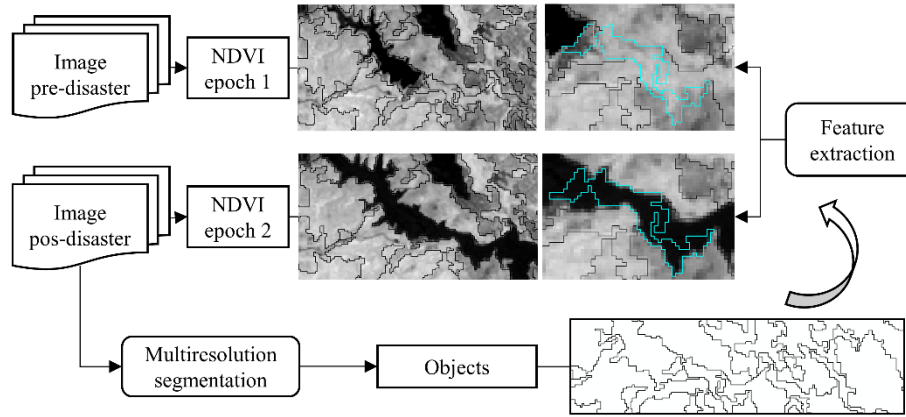


Figure 3: Image segmentation procedure for feature extraction.

Experimental semivariogram

For continuous variables, such as the NDVI, the experimental semivariogram is defined as half of the average squared difference between values separated by a given lag, where this lag is a vector in both distance and direction (Atkinson; Lewis, 2000). The semivariance is defined from the spatial variance of measures performed in samples from a determined distance “h”, being the sum of the squares’ difference between the sampled values separated by a distance “h”, divided by two times the number of possible pairs on each distance. This was estimated using Equation 2:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x) - Z(x+h)]^2 \quad (2)$$

where $N(h)$ is the number of pairs of points separated by the distance h , $Z(x)$ is the value of the regionalized variable in the point x , and $Z(x+h)$ is the value of the point $(x+h)$.

The semivariogram is the graphic representation of the spatial variance

versus distance h , which allows an estimate of the variance value for different combinations of pairs of points. The semivariance functions are characterized by three parameters: sill (σ^2), range (ϕ), and nugget effect (τ^2). The sill parameter is the plateau reached by semivariance values and shows the quantity of variation explained by the spatial structure of the data. The range parameter is the distance where the semivariogram reaches the sill, showing the distance until the data are correlated. The nugget effect is the combination of sampling errors and variations that happen in scales smaller than the distance between the sampled points (Curran, 1988).

Since we wanted to characterize the NDVI spatial variability to obtain maximum detail, we used a onepixel interval between two lags (the distance between pairs of points in the semivariogram calculation), so the lag size was equivalent to the pixel size (30 m). After some experimentation to find an appropriate optimal lag distance, we fixed the number of lags at 20 pixels (resulting in a lag distance of 600 m) to ensure that sill values would provide a concise description of data variability. According to Woodcock, Strahler and Jupp (1988), the size of the samples needs to be larger than the range of influence to characterize the initial part of the semivariogram and large enough to reveal the presence of periodicity.

Set of semivariogram indices

The set of semivariogram indices we used was described by Balaguer et al. (2010) based on the points defining the experimental semivariogram. These indices describe the shape of the experimental semivariograms and therefore the properties that characterize the spatial patterns of the image object (Table 1); they have been categorized according to the position of the lags used in their definition (near the origin and up to the first maximum). The devised feature groups provide information such as the change ratio, slope, concavity, and

convexity (curvature) level of the images and data variability.

Table 1: Semivariogram indices described by Balaguer et al. (2010).

Group	Description	Formula
a	1. Ratio between the values of the total variance and the semivariance at first lag	$RVF = \frac{\text{Variance}}{\gamma_1}$
	2. Ratio between semivariance values at second and first lag	$RSF = \frac{\gamma_2}{\gamma_1}$
	3. First derivative near the origin	$FDO = \frac{\gamma_2 - \gamma_1}{h}$
	4. Second Derivative at third lag	$SDT = \frac{\gamma_4 - 2\gamma_3 + \gamma_2}{h^2}$
b	5. First maximum lag value	$FML = h_{\max_1}$
	6. Mean of the semivariogram values up to the first maximum	$MFM = \frac{1}{\text{Max_1}} \sum \gamma_i$
	7. Variance of the semivariogram values up to the first maximum	$VFM = \frac{1}{\text{Max_1}} \sum (\gamma_i - \bar{\gamma})^2$
	8. Ratio between the semivariance at first local maximum and the mean semivariogram values up to this maximum	$RMM = \frac{\gamma_{\text{Max_1}}}{\bar{\gamma}_{\text{Max_1}}}$
	9. Difference between the mean of the semivariogram values up to the first maximum (MFM) and the semivariance at first lag	$DMF = MFM - \gamma_1$
	10. Second-order difference between first lag and first maximum	$SDF = \gamma_{\text{Max_1}} - 2\frac{\gamma_{\text{Max_1}}}{2} + \gamma_2$
	11. Semivariance curvature	$AFM = \frac{h}{2} \left(\gamma_1 + 2 \left(\sum_{i=2}^{\text{max_1}-1} \gamma_i \right) + \gamma_{\text{max_1}} \right) - \left(\gamma_1 (h_{\text{max_1}} - h_1) \right)$

a=Indices that provide semivariogram information near the origin; b=Indices that provide semivariogram information at first maxima.

The semivariogram texture description is traditionally achieved by fitting a mathematical function (i.e. exponential model, gaussian model and spherical model) whose parameters (such as sill and range) are adopted as

texture measures (Chen; Gong, 2004; Woodcock; Strahler; Jupp, 1988). This method often suffers from the selection of a proper function because simple functions are not sufficiently distinguishable and complex ones may be subject to overfitting (Chica-Olmo; Abarca-Hernández, 2000). The semivariogram indices are free of the problems caused by modeling the experimental semivariogram and thus have become more popular for describing the spatial properties of remote sensing images (Wu et al., 2015).

Feature extraction

We focused on two classes in this study: (1) no change objects consisting of areas with the same cover in both images and (2) change objects consisting of areas affected by flooding from the dam failure. A data set of 200 objects (with 100 objects per class) was sampled with 50% of the samples randomly chosen as training samples and the rest used as evaluation samples. Within the objects, the semivariogram indices were extracted in both images using FETEX 2.0 software (Ruiz et al., 2011), a feature extraction tool for object-based image analysis.

Due to the high number of indices, some of the information they provide may overlap with others, and so are probably redundant in terms of efficiently describing the objects. Thus we employed principal component analysis (PCA) in order to group and interpret the redundancies in the information provided by the analyzed semivariogram indices. By choosing the variables with higher impact on the first two principal components, we were able to reduce the number of variables, avoid redundant variables (multicollinearity), and make further analyses more efficient.

Change detection and evaluation

In order to detect changes in the images, we chose to use a support

vector machine (SVM) algorithm. Consisting of a group of theoretically superior machine learning algorithms, this approach is especially advantageous in the presence of heterogeneous classes for which only a few training samples are available (Wu et al., 2015).

SVMs operate by assuming that each set of inputs will have a unique relation to the response variable, and that the grouping and relation of these predictors to one another is sufficient to identify rules that can be used to predict the response variable from new input sets. To do this, SVMs project the input space data into a feature space with a much larger dimension, enabling linearly nonseparable data to become separable in the feature space. For example, this method has been successfully used in forestry classification problems (García-Gutiérrez et al., 2015; Wu et al., 2015). We used the Gaussian or radial basis function (RBF) as the Kernel function and performed change detection evaluation using a confusion matrix (Congalton 1991) and its accuracy measures, validating the results with a manually-produced map.

RESULTS AND DISCUSSION

Semivariogram indices selection

By computing the PCA over the complete set of semivariogram features, we concentrated most of the data's variability in the first components; the resulting visualization of the data allows for a better understanding of redundancies (Figure 4). The proportion of variability explained by PC1 and PC2 (the first two principal components) was 53.15%.

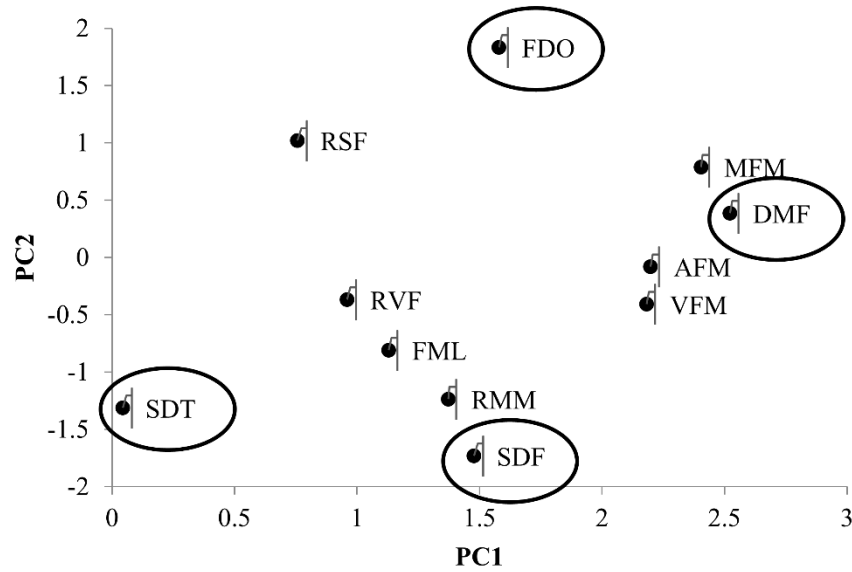


Figure 4: Projection of proposed indices in principal component planes (PC1-PC2).

As a result of PCA analysis for the group of indices that provide information near the origin, we removed RVF and RSF and included FDO and SDT as input data for the change detection analysis. After analyzing the indices that provided information up to the first maxima, we also removed AFM, VFM, FML and RMM and included DMF and SDF as further input data. We selected the variables that presented higher values in module in the first two components (Table 2).

Table 2: PCA eigenvalues.

Group	Indices	PC1	PC2	PC3	PC4	PC5	PC6
a	RVF	1.454	0.138	0.901	-0.234	-0.351	-1.001
	RSF	1.098	0.238	1.362	1.461	-0.613	-0.158
	FDO	1.830	0.627	0.007	1.165	0.468	0.131
	SDT	0.352	-0.969	-1.340	0.921	0.526	-1.536
b	FML	1.101	-0.908	0.473	-1.401	-0.373	-0.692
	MFM	2.410	0.359	-0.220	0.105	0.428	0.260
	VFM	2.107	-0.252	-0.824	-0.113	0.169	0.653
	DMF	2.454	0.230	-0.280	-0.209	0.108	0.174
	RMM	0.916	-1.817	1.014	0.320	-0.365	0.328
	SDF	-0.122	-2.230	-0.357	0.399	0.071	0.633
	AFM	2.253	0.075	-0.280	-0.887	-0.142	0.010

a=Indices that provide semivariogram information near the origin; b=Indices that provide semivariogram information at first maxima.

Exploring the semivariogram indices

We analyzed the semivariogram curves considering both the change (Figure 5a) and no-change (Figure 5b) classes. In the former, the image's spatial variability changed considerably from native vegetation (pre disaster image) to flooded areas (post-disaster image). The flooded areas had a low overall variability due to the homogeneity of NDVI pixels with low internal variation. The high relative variability of native vegetation is explained by the presence of high and low NDVI values in the same object. In contrast, the semivariogram curves for the no-change objects presented similar values.

The pre-selected semivariogram indices decreased (FDO and DMF) or increased (SDT and SDF) considerably in the presence of changes (Figure 6a) and remained almost constant in the absence of changes (Figure 6b).

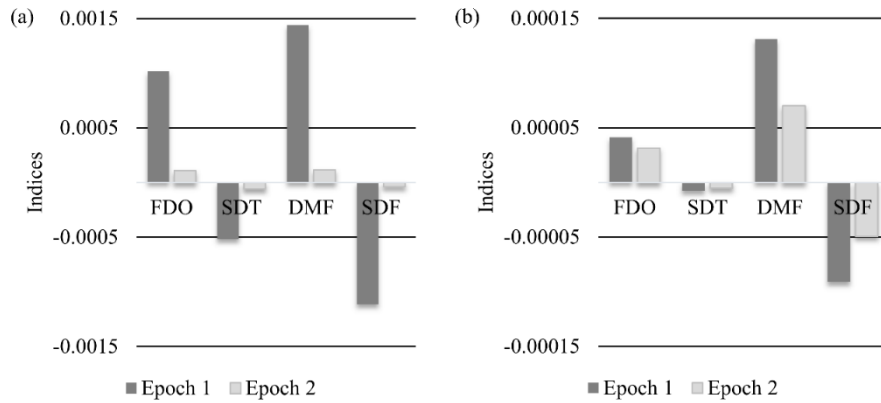


Figure 6: Values of pre-selected semivariogram indices from image epoch 1 and image epoch 2 for: (a) change objects; (b) no-change objects

FDO is the first derivative near the origin and represents the slope of the semivariogram at the first two lags; it shows the variability changes of the data at short distances. FDO presented high values for heterogeneous objects (Figure 7a) and low values for homogeneous objects (Figure 7b). SDT is the second derivative at the third lag. This index approximates the value of the second derivative of the semivariogram at the third lag. It quantifies the concavity or convexity level of the semivariogram at short distances, corresponding with the heterogeneity of the objects in the image. Negative values indicate that the semivariogram is convex and thus that the image is heterogeneous at short distances. SDT presented high negative values for change objects (Figure 7a) and low negative values for no-change objects (Figure 7b).

DMF is the difference between the mean of the semivariogram values up to the first maximum (MFM) and the semivariance at the first lag (difference mean of semivariogram and first lag semivariance). This index shows the decreasing rate of the spatial correlation in the image up to the lags where the semivariogram theoretically tends to be stabilized. The results showed a high

variation of DMF values for change objects and a relatively low variation of DMF values for no-change objects. SDF is the second-order difference between the first lag and first maximum. This parameter provides information about the semivariogram curvature in that interval, also representing the low frequency values in the image. SDF values presented a high variation for change objects and low variation for no-change objects.

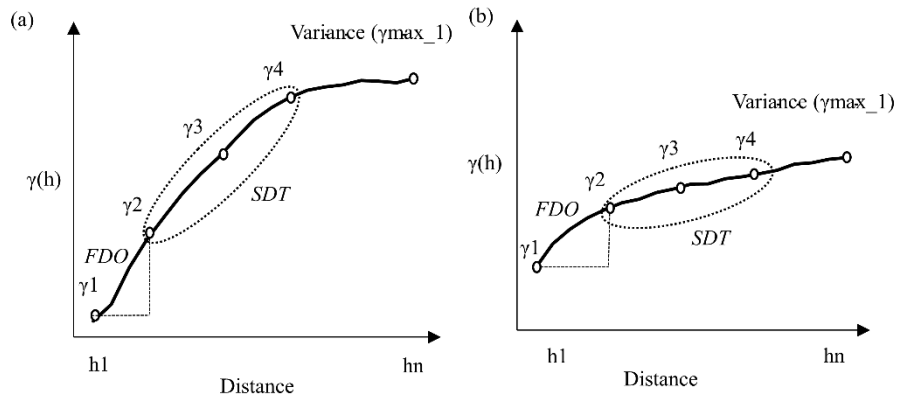


Figure 7: Semivariogram representation of the total data variance for the FDO and SDT indices: (a) heterogeneous objects, and (b) homogeneous objects.

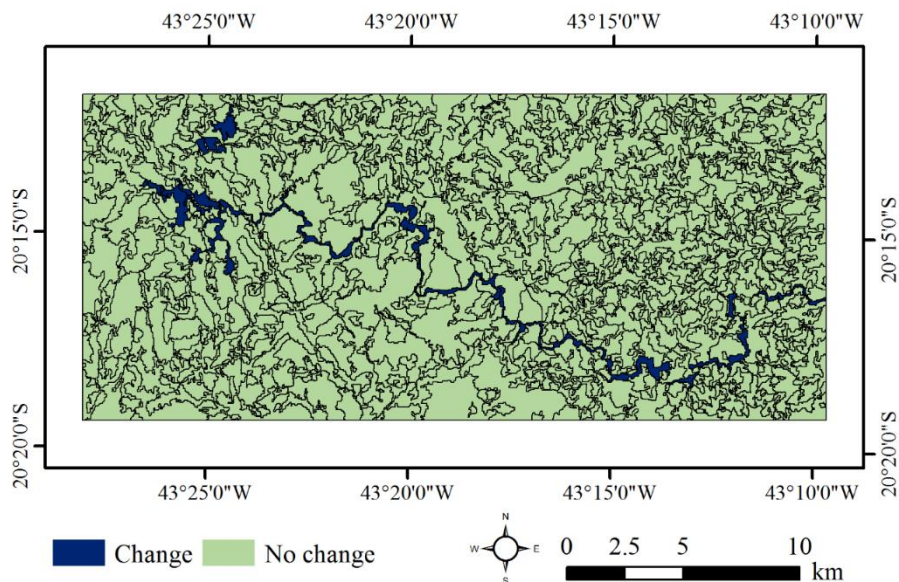
Change detection assessment

The classification accuracy measures, using the selected semivariogram indices as input for the SVM algorithm, are shown in Table 3. The semivariogram indices showed their effectiveness in the classification of change and no-change classes, presenting an overall accuracy of 95.12% and producer's and user's accuracies higher than 85%.

Table 3: Confusion matrix of the support vector machines classification.

Confusion Matrix	Producer's accuracy (%)	User's accuracy (%)
Change	100.00	85.71
No Change	93.10	100.00
Overall accuracy (%)	95.12	
Kappa	0.88	

Figure 8 shows the change detection map (producer's accuracy = 100%); all objects classified as no-change in the map are correct (user's accuracy = 100%). However, according to the validation data set, there are still some misclassification problems with 14.29% of the objects classified erroneously as change (user's accuracy = 85.71%) and the omission of 6.9% of change-class objects in the map.

**Figure 8:** Change detection results.

In summary, the semivariogram indices synthesized the most relevant information about the shape of the semivariogram (slope) in a few features.

They identified the singular points (maxima) and enhanced the information contained in the first lags, where spatial correlation at short distances is higher. These indices also have a specific meaning, allowing them to be easily interpreted.

CONCLUSIONS

In this study, we used spatial context to detect land cover changes resulting from a Brazilian dam failure using an object-based approach. We explored and investigated the potential of semivariogram indices as inputs for training the support vector machines algorithm for change detection. Our results indicate that landscape changes can be accurately detected using only textural features calculated from semivariograms derived from NDVI images.

The semivariogram indices selected by PCA analysis showed their effectiveness in the classification results, presenting high accuracy values. Using semivariograms as the main geostatistical tool to describe spatial variability standards in data means that indices derived from NDVI variability have the potential to discriminate between homogeneous and heterogeneous classes within objects. This approach can be used in many other contexts for rapid and accurate assessment of such land-cover changes.

Further research should explore the use of geostatistical features to characterize the degree of changes as well as the impact of the initial land cover class and the image segmentation epoch on the analysis results. Other studies could analyze the influence of seasonality on change detection in vegetated areas.

ACKNOWLEDGEMENTS

The authors are grateful to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), and the Departamento de Ciências Florestais da Universidade Federal de Lavras (UFLA) for supporting this work.

REFERENCES

- ACERBI JUNIOR, F. W. et al. Change detection in Brazilian savannas using semivariograms derived from NDVI images. **Ciencia e Agrotecnologia**, 39(2):103-109, 2015.
- ATKINSON, P. M.; LEWIS, P. Geostatistical classification for remote sensing: An introduction. **Computers & Geosciences**, 26(4):361-371, 2000.
- BAATZ, M.; SCHÄPE, A. Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation. **ISPRS Journal of Photogrammetry and Remote Sensing**, 58(3-4):12-23, 2000.
- BALAGUER, A. et al. Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification. **Computers and Geosciences**, 36(2):231-240, 2010.
- BALAGUER-BESER, A. et al. Semivariogram calculation optimization for object-oriented image classification. **Modelling in Science Education and Learning**, 4(7):91-104, 2011.
- BERBEROGLU, S. et al. The integration of spectral and textural information using neural networks for land cover mapping in the Mediterranean. **Computers and Geosciences**, 26(4):385-396, 2000.
- BERBEROGLU, S.; AKIN, A. Assessing different remote sensing techniques to detect land use/cover changes in the eastern Mediterranean.

International Journal of Applied Earth Observation and Geoinformation, 11(1):46-53, 2009.

BERRA, E. F. et al. Estimativa do volume total de madeira em espécies de eucalipto a partir de imagens de satélite landsat. **Ciencia Florestal**, 22(4):853-864, 2012.

BLASCHKE, T. Object based image analysis for remote sensing. **ISPRS Journal of Photogrammetry and Remote Sensing**, 65(1):2-16, 2010.

CHEN, G. et al. Object-based Change Detection. **International Journal of Remote Sensing**, 33(14):4434-4457, 2012.

CHEN, Q.; GONG, P. Automatic variogram parameter extraction for textural classification of the panchromatic IKONOS Imagery. **IEEE Transactions on Geoscience and Remote Sensing**, 42(5):1106-1115, 2004.

CHICA-OLMO, M.; ABARCA-HERNÁNDEZ, F. Computing geostatistical image texture for remotely sensed data classification. **Computers & Geosciences**, 26(4):373-383, 2000.

CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, 37(1):35-46, 1991. COSTANTINI, M. L. et al. NDVI spatial pattern and the potential fragility of mixed forested areas in volcanic lake watersheds. **Forest Ecology and Management**, 285:133-141, 2012.

CURRAN, P. J. The semivariogram in remote sensing: An introduction. **Remote Sensing of Environment**, 24(3):493-507, 1988.

DESCLÉE, B.; BOGAERT, P.; DEFOURNY, P. Forest change detection by statistical object-based method. **Remote Sensing of Environment**, 102(1-2):1-11, 2006.

GARCÍA-GUTIÉRREZ, J. et al. An evolutionary-weighted majority voting and support vector machines applied to contextual classification of LiDAR and imagery data fusion. **Neurocomputing**, 163:17-24, 2015.

GARCIA-PEDRERO, A. et al. A GEOBIA methodology for fragmented agricultural landscapes. **Remote Sensing**, 7(1):767-787, 2015.

GARRIGUES, S. et al. Quantifying spatial heterogeneity at the landscape scale using variogram models. **Remote Sensing of Environment**, 103(1):81-96, 2006.

GIL-YEPES, J. L. et al. Description and validation of a new set of object-based temporal geostatistical features for land-use/land-cover change detection. **ISPRS Journal of Photogrammetry and Remote Sensing**, 121:77-91, 2016.

GRIFFITH, J. A. et al. Preliminary comparison of landscape pattern-normalized difference vegetation index (NDVI) relationships to Central Plains stream conditions. **Journal of Environmental Quality**, 31(3):846-859, 2002.

GRIFFITH, J. A. Interrelationships among landscapes, NDVI, and stream water quality in the U. S. Central Plains. **Ecological Applications**, 12(6):1702-1718, 2007.

HUSSAIN, M. et al. Change detection from remotely sensed images: From pixel-based to object-based approaches. **ISPRS Journal of Photogrammetry and Remote Sensing**, 80:91-106, 2013.

MEER, F. VAN DER. Remote-sensing image analysis and geostatistics. **International Journal of Remote Sensing**, 33(18):5644-5676, 2012.

MUI, A.; HE, Y.; WENG, Q. An object-based approach to delineate wetlands across landscapes of varied disturbance with high spatial resolution satellite imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, 109:30-46, 2015.

MUNROE, D. K.; SOUTHWORTH, J.; TUCKER, C. M. The dynamics of land-cover change in western Honduras: Exploring spatial and temporal complexity. **Agricultural Economics**, 27(3):355-369, 2002.

POWERS, R. P. et al. Remote sensing and object-based techniques for

mapping fine-scale industrial disturbances. **International Journal of Applied Earth Observation and Geoinformation**, 34(1):51-57, 2015.

RUIZ, L. A. et al. A feature extraction software tool for agricultural object-based image analysis. **Computers and Electronics in Agriculture**, 76(2):284-296, 2011.

SERTEL, E.; KAYA, S.; CURRAN, P. J. Use of semivariograms to identify earthquake damage in an urban area. **IEEE Transactions on Geoscience and Remote Sensing**, 45(6):1590-1594, 2007.

TUCKER, C. M. et al. Comparative spatial analyses of forest conservation and change in Honduras and Guatemala. **Conservation and Society**, 3(1):174-200, 2005.

VOROVENCII, I. Assessment of some remote sensing techniques used to detect land use/land cover changes in South-East Transylvania, Romania. **Environmental Monitoring and Assessment**, 186(5):2685-2699, 2014.

WOODCOCK, C. E.; STRAHLER, A. H.; JUPP, D. L. B. The use of variograms in remote sensing: I. Scene models and simulated images. **Remote Sensing of Environment**, 25(3):323-348, 1988.

WU, X. et al. Evaluation of semivariogram features for objectbased image classification. **Geo-spatial Information Science**, 18(4):159-170, 2015.

YUE, T. X. et al. Landscape change detection of the newly created wetland in Yellow River Delta. **Ecological Modelling**, 164(1):21-31, 2003.

**ARTICLE 4 - OBJECT-BASED LAND USE/COVER CHANGE
DETECTION APPLIED TO BRAZILIAN SEASONAL SAVANNAS
USING GEOSTATISTICAL FEATURES**

Eduarda Martiniano de Oliveira Silveira^{1*}, José Márcio de Mello², Fausto
Weimar Acerbi Júnior³ and Luis Marcelo Tavares de Carvalho⁴

¹*Forest Science Department, Federal University of Lavras, Lavras, Brazil, PO
Box 3037, +55 12 996446259, dudalavras@hotmail.com*

²*Forest Science Department, Federal University of Lavras, Lavras, Brazil, PO
Box 3037, +55 35 991422599, josemarcio@dcf.ufla.br*

³*Forest Science Department, Federal University of Lavras, Lavras, Brazil, PO
Box 3037, +55 35 999790898, fausto@dcf.ufla.br*

⁴*Forest Science Department, Federal University of Lavras, Lavras, Brazil, PO
Box 3037, +55 35 998256208, passarinho@dcf.ufla.br*

Publication status: Published in the **Internacional Journal of Remote Sensing**

doi: 10.1080/01431161.2018.1430397

Abstract

A new method for remote-sensing land-use/land-cover (LULC) change detection is proposed to eliminate the effects of forest phenology on classification results. This method is insensitive to spectral changes caused by vegetation seasonality and uses an object-based approach to extract geostatistical features from bitemporal Landsat TM (Thematic Mapper) images. We first create image objects by multiresolution segmentation to extract geostatistical features (semivariogram parameters and indices) and spectral information (average values) from NDVI (normalized difference vegetation index), acquired in the wet and dry seasons, as input data to train a Support Vector Machine algorithm. We also used the image difference traditional change-detection method to validate the effectiveness of the proposed method. We used two classes: (1) LULC change class and (2) seasonal change class. Using the most geostatistical features, the change detection results are considerably improved compared with the spectral features and image differencing technique. The highest accuracy was achieved by the sill (σ^2 overall variability) semivariogram parameter (95%) and the AFM (area first lag–first maximum) semivariogram index (88.33%), which were not affected by vegetation seasonality. The results indicate that the geostatistical context makes possible the use of bitemporal NDVI images to address the challenge of accurately detecting LULC changes in Brazilian seasonal savannahs, disregarding changes caused by phenological differences, without using a dense time series of remote-sensing images. The challenge of extracting accurate semivariogram curves from objects of long and narrow shapes requires further study, along with the relationship between the scale of segmentation and image spatial resolution, including the type of change and the initial land-cover class.

Keywords: geostatistical; semivariogram; remote sensing; change detection; feature extraction; vegetation seasonality

1. Introduction

Change detection is the process of identifying changes from analysing multitemporal images (Radke et al. 2005). Such changes may be due to natural causes (natural forest phenology) or land-use/land-cover (LULC) alterations (e.g. urban growth and deforestation; Ghosh, Roy, and Ghosh 2014). A significant challenge in remote-sensing change detection is accurately extracting LULC changes while disregarding those associated with phenological difference (Chen et al. 2013, 2014; Jin et al. 2013). When images from different seasons are acquired, changes caused by phenological differences are inevitable and pose a significant challenge to LULC change detection (Lu et al. 2016).

Utilizing change detection based on time series of normalized difference vegetation index (NDVI) could avoid this problem, since it can accurately track seasonal characteristics and capture information on vegetation phenology (Chen et al. 2015). Verbesselt et al. (2010) proposed a generic change detection approach for time series by detecting and characterizing breaks for additive seasonal and trend (BFAST) to separate disturbances such as deforestation, urbanization, floods, and fire from phenological changes. Jin et al. (2013) proposed an object-based spatial and temporal vegetation index unmixing model to solve problems related to phenological differences, using spectral data from Moderate Resolution Imaging Spectroradiometer NDVI time series in Landsat objects. Hamunyela, Verbesselt, and Herold (2016) proposed a new approach that reduces seasonality in satellite image time series using spatial context in a pixel-based approach. However, implementation of these methods is dependent on the availability of an appropriate time series, without cloud contamination and appropriate image temporal resolution.

Blended data with high spatio-temporal resolution could also be used to monitor LULC changes and eliminate those changes associated with phenological differences (Lu et al. 2016). However, these methods are generally

pixel based and are sensitive to registration errors. In object-based image analysis (OBIA) approaches, pixels are not individually classified, but combined into homogeneous groups (objects) and classified together (Mui, He, and Weng 2015), optimizing the delineation of individual features, resulting in more accurate change detection (Johansen et al. 2010) by reducing small spurious changes (Hussain et al. 2013).

In comparison with pixel-based approaches, additional spatial information can be obtained from objects in OBIA (Blaschke 2010; Ruiz et al. 2011; Hussain et al. 2013; Wu et al. 2015) to improve change detection analysis (Li and Leung 2002; Wu et al. 2010; Chen et al. 2012). The semivariograms of geostatistics are used as spatial measures (Curran 1988; Woodcock, Strahler, and Jupp 1988) and have widely been used in heterogeneity analyses (Wu et al. 2000; Garrigues et al. 2006; Garrigues et al. 2008; Cadenasso, Pickett, and Schwarz 2007; Balaguer-Beser et al. 2013; Huang et al. 2013; Lausch et al. 2013; Qiu et al. 2013), image classification (Balaguer et al. 2010; Balaguer-Beser et al. 2011; Yue et al. 2013; Powers et al. 2015; Wu et al. 2015; Silveira et al. 2017), and change detection studies (Sertel, Kaya, and Curran 2007; Costantini et al. 2012; Acerbi Júnior et al. 2015; Gil-Yepes et al. 2016). Most of the studies did not use the spatial context to help with change detection to climate seasonal variations (Zhu 2017), except Hamunyela, Verbesselt, and Herold (2016), which used NDVI spatially normalized to reduce phenological difference in time series.

Here, instead of using dense time series to minimize the effects of forest phenology on LULC change detection, we exploit the spatial context, represented by geostatistical features, using bitemporal NDVI images and an object-based approach. The main objective of this study was to develop a method for remote-sensing LULC change detection that is insensitive to changes caused by vegetation seasonality, without the need to use a dense time series.

We focused on the following research questions: (1) Do changes caused by vegetation seasonality affect the spatial variability of NDVI values? (2) Are geostatistical features derived from semivariograms able to accurately detect LULC changes?

We utilize NDVI derived from bitemporal Landsat TM (Thematic Mapper) images obtained during the wet and dry seasons to assess the potential of individual geostatistical features to accurately detect LULC changes, disregarding those associated with phenological differences, in a Brazilian seasonal savannah. Brazilian savannah is a highly heterogeneous biome in terms of biodiversity and vegetation types, and it shows strong seasonality, where phenological differences affect the results of LULC change detection.

2. Materials and methods

The study exploits the potential of individual geostatistical features extracted from bitemporal NDVI images to accurately detect LULC changes, disregarding those associated with phenological differences, using an object-based approach. We tested (1) spectral and (2) geostatistical features as input data to train Support Vector Machine (SVM) algorithm and also we used the (3) NDVI image difference as a traditional change detection method to compare the proposed method.

We defined two classes: (1) LULC change objects and (2) seasonal change (SC) objects. The methodology consists of the following main steps, also presented in Figure 1.

- (1) Image acquisition and NDVI transformation
- (2) Image segmentation by a multiresolution algorithm using the first epoch wavebands of Landsat TM to generate the image objects
- (3) Feature extraction from NDVI of wet and dry seasons computed within objects

(4) Change detection

(5) Evaluation by confusion matrix and accuracy measurement

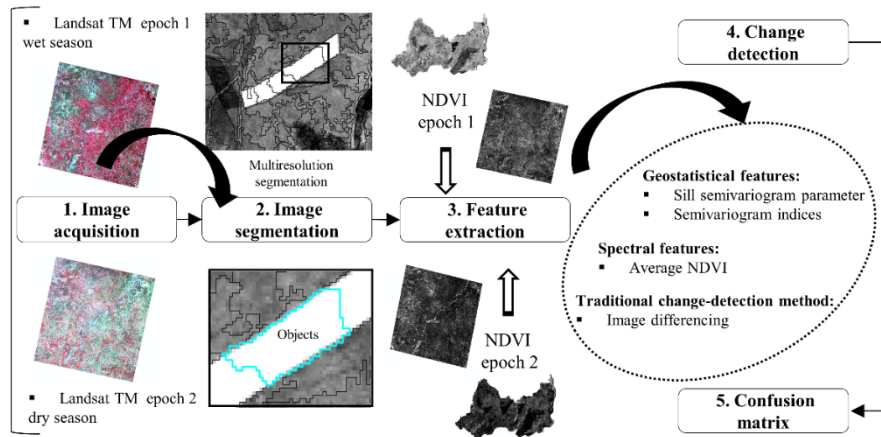


Figure 1. Methodology workflow.

2.1. Data and study area

The study area, the municipality of São Romão, is located in northern Minas Gerais (MG), Brazil (Figure 2). The main vegetation type is typical of a tropical savannah, with physiognomies ranging from grasslands to densely vegetated areas dominated by shrubs and trees. The study area occupies 2,200 km², of which approximately 84% exhibits native vegetation. The gradient of vegetation density in the area is noteworthy, containing areas of grassland with no shrubs or trees, sparse and dense shrublands, as well as woodlands dominated by palm trees along watercourses.

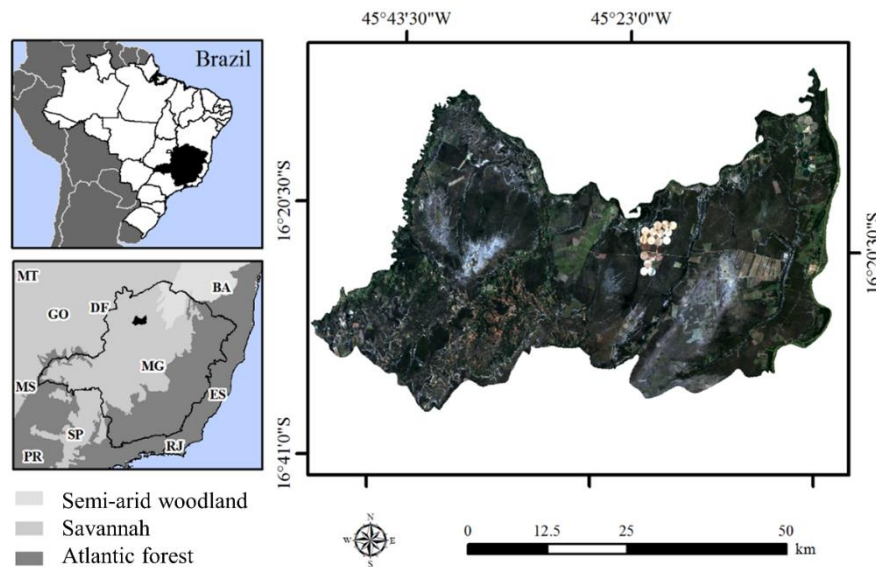


Figure 2. Location of Minas Gerais (MG) in Brazil, MG Biomes, São Romão, and its political boundaries.

Landsat TM images from May 2010 and September 2011 were obtained from the United States geological survey for earth observation and science (USGS\EROS). The images were acquired at the CDR processing level (Landsat surface reflectance climate data record), with the necessary geometric corrections and reflectance values at the ground level. Image selection was based on the following criteria: presence of vegetation seasonality, occurrence of deforestation, and absence of clouds.

The NDVI was calculated for both Landsat TM images. This index is based on quotients, and uses the red and near-infrared spectral bands to enhance vegetation cover and minimize the effects of shadows caused by the terrain's topography (Vorovencii 2014).

According to Peel, Finlayson and McMahon (2006), the climate in the region is typical of a tropical savannah, where rains are concentrated from

October to May, characterizing a pronounced seasonality in the region. The Brazilian savannah vegetation is influenced by strong climatic seasonality with a dry season between May and September (Schwieder et al. 2016). For this reason, we used the images acquired on May 2010 as representatives of the end of the rainy season (high NDVI values), due to the rainfall in the previous months. On the other hand, images acquired on September 2011 were used as representatives of the end of the dry season (low NDVI values), due to the lack of rainfall in the previous months. Figure 3 shows the NDVI average values from savannahs vegetation object in both seasons.

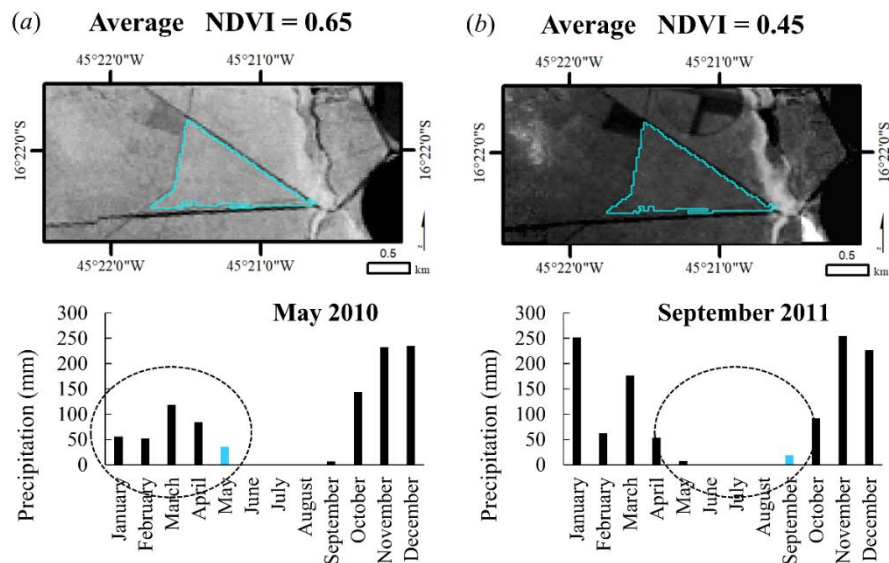


Figure 3. Monthly precipitation pattern: Example object's average NDVI values in (a) May 2010 and (b) September 2011.

2.2. Class definition

The study focused on two classes: (1) LULC change areas comprising deforestation and burned areas (mainly savannahs that were converted to bare soil and pastures, and (2) seasonal change (SC) areas comprising the same cover

in both epochs, albeit with changes caused by vegetation seasonality (Figure 4). These classes were set up according to prior visual interpretation of vegetation maps (Carvalho and Scolforo 2008), which described qualitative and quantitative information.

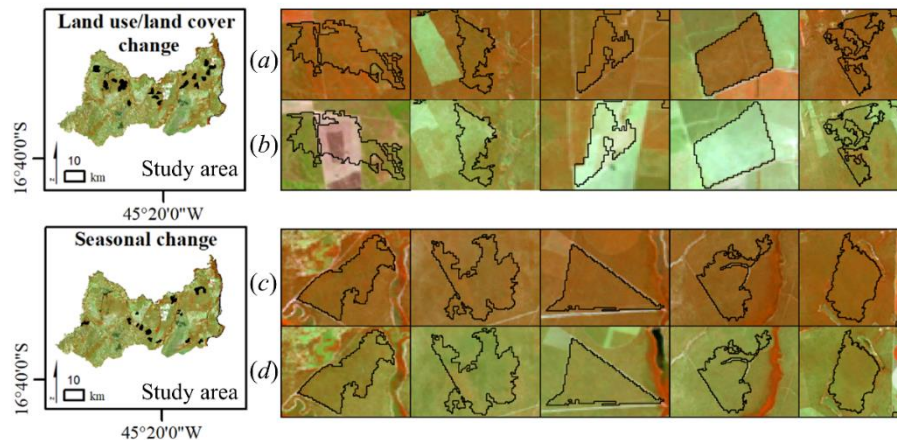


Figure 4. Examples of class definitions using Landsat TM image R4 G5 B3. Land use/land cover (LULC) change class: (a) savannah's first epoch during the wet season; (b) savannah deforestation in the second epoch during the dry season. Seasonal change (SC) class: (c) savannah's first epoch during the wet season; (d) savannah's second epoch during the dry season.

A data set of 200 objects well-distributed over the study area, with 100 samples per class, was assigned based on visual interpretation. Fifty per cent of the samples were randomly chosen as training samples, while the rest were used as test samples. The objects used to train (50) and evaluate (50) the LULC change class were collected considering both expressive and slight changes, comprising both small and big changes within the objects (Figure 5). And also, the objects used to train (50) and evaluate (50) the SC class were collected from preserved savannah's vegetation areas, affected by a pronounced seasonality from the wet to the dry season.

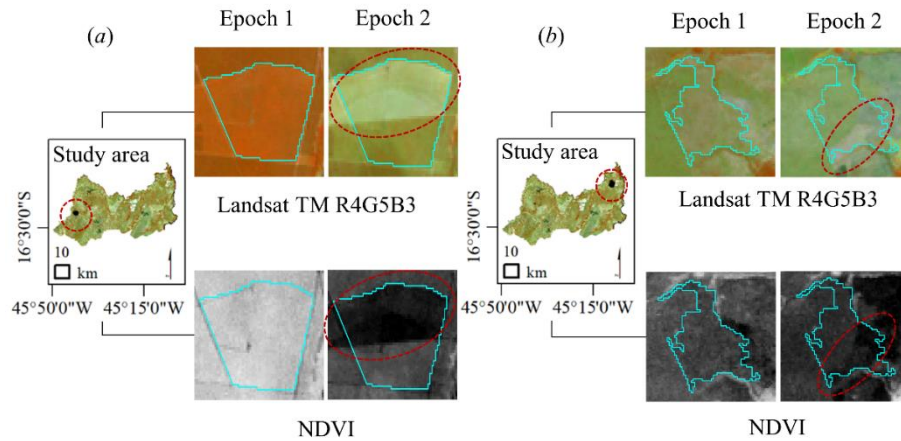


Figure 5. Examples of samples comprising (a) expressive and (b) slight land use/land cover (LULC) changes.

2.3. Multiresolution segmentation

The first step in OBIA and object-based change detection (OBCD) is image-object extraction, which is achieved by segmentation or stratification of the images (Addink, Van Coillie, and de Jong 2012) into spatially continuous and homogeneous objects (Hussain et al. 2013). Image segmentation is the core of OBIA, and various segmentation techniques have been developed around it (Blaschke 2010). Multiresolution segmentation (Baatz and Schäpe 2000) is a basic procedure available in eCognition software for object-based image analysis, and was used here to produce image objects as a first step prior to further feature extraction and change detection analysis. According to Tewkesbury et al. (2015) three different choices might be considered when defining the analysis unit (image objects) for multitemporal object-based image analysis:

(i) image-object overlay: image-objects are generated by segmenting one of the images in the time series. A comparison against other images is then made by simple overlay;

(ii) image-object comparison: image-objects are generated by segmenting each image in the time series independently;

(iii) multitemporal image-object: image-objects are generated by segmenting the entire time series together.

Our study focused on the image-object overlay approach. We chose this technique because it provides a more meaningful framework for spatial measures (i.e. geostatistical features) (Tewkesbury et al. 2015), capable of capturing the intra-object heterogeneity (Listner and Niemeyer 2011). Image segmentation used wavebands of the Landsat TM image acquired in May 2010 (Epoch 1). Three key segmentation parameters (shape, compactness, and scale) control the size, shape, and spectral variation of segmented image objects. The shape parameter was set to 0.1 and compactness was set to 0.5. The most critical step is the selection of the scale parameter, which controls the size of the image objects. The scale parameter sets a homogeneity threshold that determines the number of neighbouring pixels that can be merged to form an image object (Mui, He, and Weng 2015).

The scale of segmentation directly influences the size of the objects connected to the semivariogram predefined criteria (lag distance) and the minimum number of pixels inside each object necessary to generate the semivariogram. The scale parameter was set to 200 to guarantee a minimum number of samples (25 pixels). To find this appropriate scale parameter we used a “trial and error” approach (Duro, Franklin, Dubé 2012; Chen et al. 2015). The segmentation results were evaluated based on a visual assessment (Zhang, Fritts, and Goldman 2008; Whiteside, Boggs, and Maier 2011).

2.4. Feature extraction

In this section, we describe the geostatistical (the semivariogram parameter and the set of semivariogram indices), and the spectral features

(NDVI average values) extracted within the extent of the objects used to train the SVM algorithm.

We used the geostatistical features achieved by the following two ways: (1) modelling the semivariogram by fitting a model to extract semivariogram parameters (Acerbi Júnior et al. 2015) and (2) by using indices that synthesize the most relevant information about the shape of the semivariogram and enhance the spatial information (Balaguer et al. 2010). As spectral feature we used the average NDVI values inside the objects.

2.4.1 Geostatistical features: semivariogram parameter

We used as semivariogram parameter, the overall variability (sill - σ^2) of NDVI values inside the objects, captured by the semivariogram. The semivariogram, was originally developed for the mining industry and has since been adapted for application to textural information of remotely sensed images because it is a relatively simple function for investigating spatial correlations (Miranda et al. 1996).

The semivariogram is a graphical representation of the spatial variability in a given dataset (Cohen, Spies and Bradshaw 1990). The relationship between a pair of pixels can be calculated by the variogram function (Eq. 1), called $2\gamma(h)$, which corresponds to the mathematical expectation of the squared difference between pairs of points separated by a distance h . The semivariogram function depends on the location x and the distance between samples h . For the variogram to be based solely on the distance between the sampling units, it is necessary to adopt the intrinsic hypothesis (stationarity), which implies that the variance of the differences between two sample points depends only on the distance h .

$$2\gamma(h) = E\left\{\left(Z(x)-Z(x+h)\right)^2\right\} \quad (1)$$

For continuous variables such as the NDVI, the experimental

semivariogram is defined as half of the average squared difference between the values separated by a given lag, where the lag is a vector having both distance and direction (Atkinson and Lewis 2000). It was estimated using Eq. 2:

$$\gamma(h) = \left(\frac{1}{2N(h)} \right) \sum_{i=1}^{N(h)} (Z(x) - Z(x+h))^2 \quad (2)$$

where $\gamma(h)$ is the estimator of the semivariance for each distance h , $N(h)$ is the number of pairs of points separated by the distance h , $Z(x)$ is the value of the regionalized variable at point x , and $Z(x+h)$ is the value at point $(x+h)$.

The graph of spatial variance versus distance h represents the semivariogram (Figure 6), from which the estimated variance for different pairs of point combinations can be obtained. The semivariance functions are characterized by three parameters: sill (σ^2), range (ϕ), and nugget effect (τ^2). The σ^2 parameter is the plateau reached by semivariance values, and shows the quantity of variation explained by the spatial structure of the data. The ϕ parameter is the distance at which the semivariogram reaches the sill; it shows the distance to which the data are correlated. The τ^2 is the combination of sampling errors and variations that occur at scales smaller than the distance between the sampled points (Curran 1988).

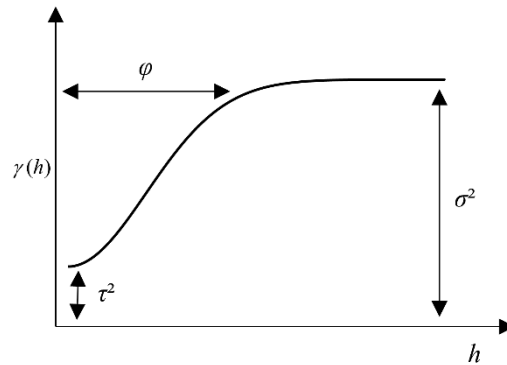


Figure 6. Classical semivariogram: $\gamma(h)$ - estimator of the semivariance for each distance h ; (σ^2) - sill; (ϕ) - range; (τ^2) - nugget effect.

When the semivariogram is calculated for each individual object, an important factor to be considered is the lag distance. It should not be larger than the spatial extent of the object; on the other hand, an exceedingly small distance fails to provide a complete description of textural features. We attempted to find an optimal lag distance to ensure that σ^2 values would provide a concise description of data variability. We fixed the number of lags as 20 pixels and the lag size equivalent to image spatial resolution (30 m), resulting in a lag distance of 600 m. According to Woodcock, Strahler, and Jupp (1988), the lag distance needs to be larger than the range of influence in order to characterize the initial part of the semivariogram, and large enough to reveal the presence of periodicity.

The theoretical semivariograms were estimated by fitting mathematical models to the experimental semivariogram using weighted least squares. Exponential, spherical, and Gaussian models were tested. The fitted models were validated through cross-validation, by analysing the reduced mean error (ER) and the standard deviation of reduced errors (SRE).

2.4.2. Geostatistical features: set of semivariogram indices

We chose the set of semivariogram indices included in the object-based descriptive feature extraction software FETEX 2.0 (Ruiz et al. 2011). These indices have been categorized according to the position of the lags used in their definition (Table 1). The groups of features are devised to provide information such as the change ratio, slope, levels of image concavity and convexity (curvature), and data variability (Balaguer et al. 2010).

Table 1. Indices described in Balaguer et al. (2010).

Index	Description	Formula
Near the origin		
RVF (Ratio variance – first lag)	Ratio between total variance and semivariance at first lag	$RVF = \left(\frac{\text{Variance}}{\gamma_1} \right)$
RSF (Ratio second – first lag)	Ratio between semivariance values at second and first lags	$RSF = \left(\frac{\gamma_2}{\gamma_1} \right)$
Up to the first maximum		
MFM (Mean first maximum)	Mean of semivariogram values up to the first maximum	$MFM = \left(\frac{1}{\max_1} \right) \sum \gamma_i$
DMF (Difference mean first lag)	Difference between mean semivariogram value up to the first maximum (MFM), and semivariance at first lag	$DMF = (MFM) - \gamma_i$
RMM (Ratio maximum mean)	Ratio between semivariance at first local maximum, and mean semivariogram values up to this maximum	$RMM = \left(\frac{\gamma_{\max_1}}{\gamma_{\max_1}^{\text{mean}}} \right)$
AFM (Area first lag – first maximum)	Semivariance curvature	$AFM = \frac{h}{2} \left(\gamma_1 + 2 \left(\sum_{i=2}^{\max_1-1} \gamma_i \right) + \right. \\ \left. \gamma_{\max_1} \right) - \left(\gamma_1 (h_{\max_1} - h_1) \right)$

2.4.3. Spectral features

As a complement to geostatistical features, the average NDVI values of the pixels inside each object were included to serve as a spectral feature. This allowed comparison of the performance of spatial versus spectral features.

2.5. Change detection

2.5.1. Features analysis and classification

We analysed the potential of the individual geostatistical and spectral features to accurately detect LULC changes, by the following two ways: (i) semivariogram and graphic analysis and (ii) change detection classification. The precept was: If from the wet to the dry season, the individual features values remain constant and also increase or decrease in the presence of LULC changes, the feature is able to discriminate between the classes, due to it is not affected by vegetation seasonality.

We chose the SVM algorithm to perform the change detection classification. As a group of theoretically superior machine learning algorithms, SVM algorithms appear to be especially advantageous in the presence of heterogeneous classes for which only a few training samples are available (Wu et al. 2015).

As described by Kulkarni and Lowe (2016), SVMs are a supervised nonparametric statistical learning method. In their simplest form, SVMs are binary classifiers that assign the given test sample to one of the two possible classes. SVM algorithms are extended to nonlinear cases by mapping samples within a multidimensional feature space using a kernel function. SVMs are particularly appealing in remote sensing image processing because of their ability to successfully handle small training datasets, often yielding higher classification accuracy than traditional methods (Mantero, Moser and Serpico 2005).

The radial basis function (RBF) kernel was used for its effectiveness and accuracy (Zuo, John and Carranza 2011; Shao and Lunetta 2012). The algorithm used is implemented in WEKA 3.8 software under the sequential minimal

optimization (SMO) function. Values of parameters C and σ (bandwidth or influence range of each training point in the RBF) were tested within the interval 10^i , $i = -3, -2, -1, 0, 1, 2, 3$, where the least squared mean error configuration was chosen for application. The best combination was identified using a cross-validation method.

The SVM classifier consists of a selection of image objects used as samples to train the SVM module, after which the classifier is provided to all objects to derive the final classification. To perform a classification accuracy assessment, it is necessary to compare two sources of information: the classification map produced and reference test samples (Jensen 2004). The group of samples selected manually by visual interpretation was divided into two parts, one for training the algorithm and the other for classification assessment (see Section Class definition).

2.5.2. Traditional change detection method: image differencing technique

To validate the effectiveness of the proposed method we applied an image differencing technique (Lu and Brondízio 2004). This technique has long been used to highlight areas of image change quickly with minimal supervision and is still in use today, typically applied to image-objects (Desclée, Bogaert and Defourny 2006; Tewkesbury et al. 2015). Similar to pixel-based change detection, OBCD can be performed by directly comparing image-objects defined by a threshold (Chen et al. 2012; Constantini et al. 2012). Selecting threshold NDVI difference values to indicate actual land cover change is a decision-making process that can be modified based on field verification and interactive image interpretation.

Since the threshold value is often intuitively defined by researchers, a bias may be introduced. Thresholding NDVI differences based upon SD (standard deviation) values is one approach that has been commonly utilized

(Coppin and Bauer 1996; Morisette and Khorram 2000). This approach was implemented as a consistent way and arbitrary NDVI threshold values at plus and minus one SD from zero difference were used to discriminate the binary classes of change from no-change (Coulter et al. 2011).

Change objects were identified as those with NDVI difference values less than negative one SD from zero difference or greater than one SD from zero difference. Objects classified as no-change were those with values within or equal to one SD from zero difference. We did not test threshold values, due the intention of applying this technique is for comparative purpose. The choice of one SD reduces the possibility of analysing either the pattern of background noise that could be obtained with much higher thresholds (Zurlini et al 2006) and also, was appropriate given our knowledge of changes within the study area.

2.6. Evaluation of change detection

We assessed the quality of the change detection classification generating a confusion matrix (Congalton 1991) and comparing its accuracies (1) the overall accuracy, which is computed by dividing the total number of correct results by the total number of samples in the error matrix; (2) the absolute error and confidence interval; (3) the producer's accuracy, which indicates the probability of a reference object being correctly classified and is actually a measure of omission errors; (4) the user's accuracy, which is indicative of the probability that a sample classified on the image actually represents that category on the ground; and (5) the kappa coefficient (k), which is a statistic that measures the inter-rater agreement for qualitative items. It is generally thought to be a more robust measure than simple percent agreement calculation, since it takes into account the possibility of the agreement occurring by chance.

We compared the individual performance of SVM algorithm trained by geostatistical and spectral features and the image differencing using one SD as

threshold (Figure 7).

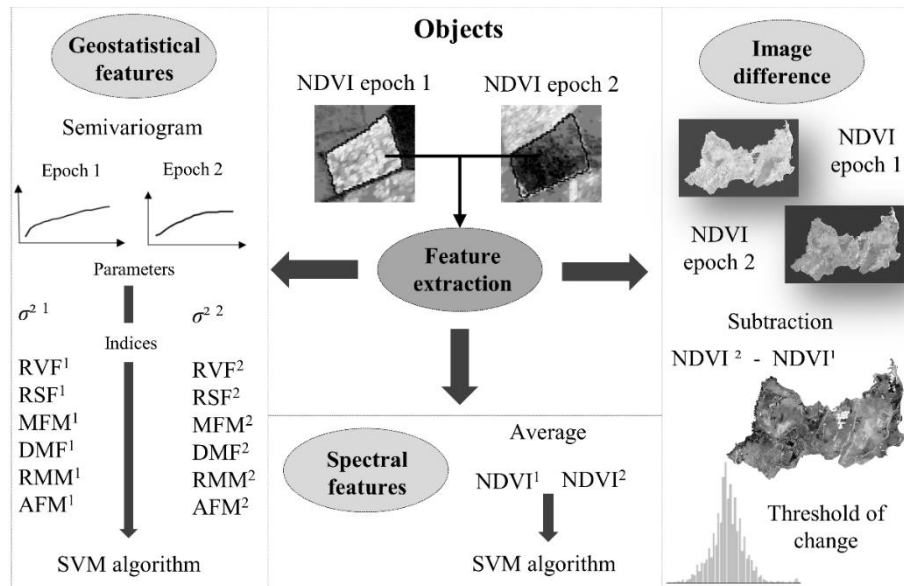


Figure 7. Features used to train the SVM algorithm and the NDVI image differencing change detection method.

3. Results and discussion

3.1. Evaluation of geostatistical and spectral features

We first analysed the potential of the individual features to discriminate between LULC change class and SC class. The semivariograms reached the σ^2 within the calculated distance, indicating that the lag distance of 600 m was large enough to encompass the entire spatial variability of the NDVI values within the objects. The objects were identified as homogeneous or heterogeneous: homogeneous objects have internally stable structures without distinct textural variation. Their semivariograms are similar to Figures 8a, c, and d; heterogeneous objects (Figure 8b) usually contain abundant distinct textural information and show a more complicated spatial structure than the

homogeneous objects.

The objects covered by savannah's vegetation in the wet season (Figures 8a and c) are composed of homogeneous pixels (low internal NDVI variability) with low σ^2 values between 0.0004 and 0.0006. After LULC change, i.e. deforestation (Figure 8b), there was an increase in internal variability of NDVI values due to the presence of bare soil and/or pasture, evidenced by a higher σ^2 value (0.0040). When the area presented SC (Figure 8d), the semivariogram shape and σ^2 remained constant, with low σ^2 values, indicating that seasonal change did not affect the spatial variability of NDVI data.

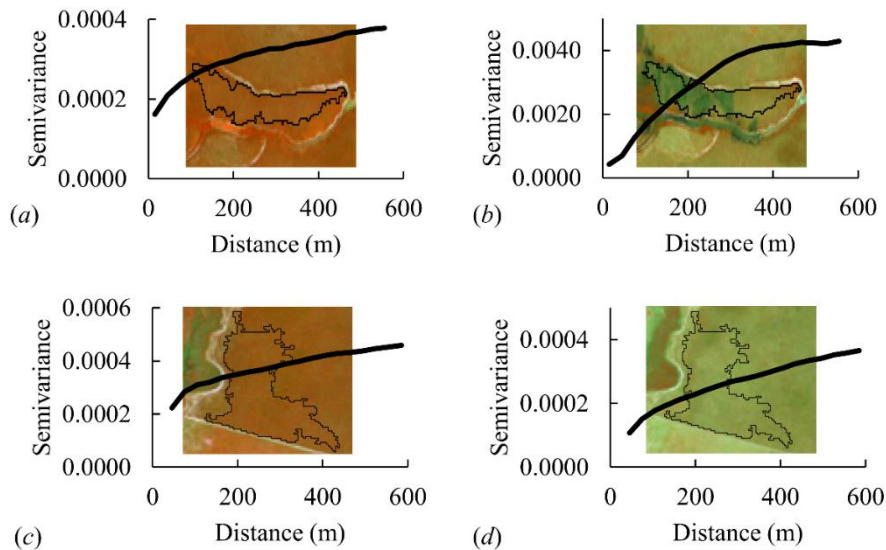


Figure 8. Semivariogram examples: (a) preserved savannah; (b) savannah undergoing a deforestation; (c) preserved savannah; (d) savannah undergoing the effects of forest phenology. Landsat TM image R4 G5 B3.

The analysis of the individuals features, indicates that the average NDVI values (spectral feature) within the objects decreases from the initial to second epoch both in the presence of LULC changes and SC (Figure 9a). In contrast, the

σ^2 parameter increased from the initial epoch (0.0004) to the second epoch (0.0046) in the presence of LULC changes and remained almost constant (approximately 0.0002, 0.0004) in the presence of SC (Figure 9b).

Analysis of the semivariogram indices that provide information near the origin, the RVF index (ratio variance first lag - indicates the relationship between spatial correlations at large and short distances) presented low values in both initial and second epochs (1.83 and 3.39) for SC objects. In the LULC change objects, it presented low values in the initial epoch (1.77) and relatively higher values in the second epoch (11.67), indicating that, the initial part of the semivariogram is strongly affected by LULC changes, and is not affect by SC (Figure 9c).

The RSF index (ratio second first lag - ratio between semivariance values at the second and first lags), also provide information near the origin. This index represents the proportion of the semivariogram value at the second lag to that at the first lag. It provides information about changes in data variability at short distances. The average RSF values for LULC change and SC objects were almost constant, (Figure 9d), indicating that this index is not affected by vegetation seasonality, however, this index is not useful in detecting LULC changes. RSF index is affected by the lag distance of the semivariogram. The lag size of 30 m (pixel size) was not sufficient to provide variations between the second and first lags and capture the landscape changes using this index.

Among the indices that provide information up to the first maxima, the MFM (mean first maximum), DMF (difference mean first lag), and AFM (area first lag – first maximum) presented the same trend (Figure 9e, f, g), with similar values for SC objects from the initial to the second epoch and increased average values in the presence of LULC changes. The MFM is an indicator of the average values between the first lag and the first maximum. The DMF index is the difference between the MFM and the semivariance at the first lag. This

parameter complements the information provided by the RVF index and showed similar results. The AFM index is the area between the semivariogram value at the first lag and the semivariogram function until the first maximum. It provides information about the curvature of the semivariogram and is also related to the variability of the data. For SC objects, the AFM is almost constant (0.0014 and 0.0034). In the presence of LULC changes, the semivariogram area presented a significant increase from 0.0015 to 0.0396.

The semivariogram index RMM (ratio maximum mean) presented similar values from the first and second epochs in both classes (Figure 9h). This index is the ratio between the semivariance at the first local maxima and the mean semivariogram values up to this maximum. It could not efficiently discriminate changes from LULC changes, classifying SC as LULC changes.

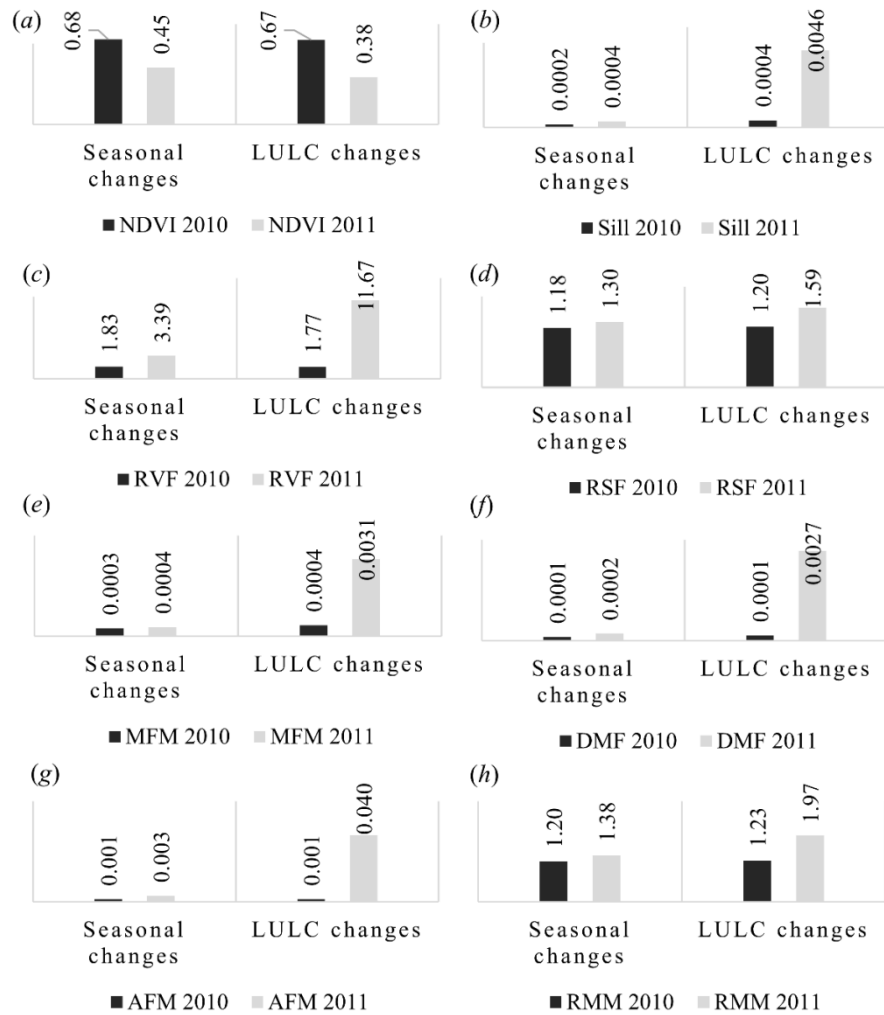


Figure 9. First and second epochs average object values for LULC change and SC classes obtained by: (a) NDVI spectral data; (b) σ^2 parameter from semivariogram; (c) RVF semivariogram index; (d) RSF semivariogram index; (e) MFM semivariogram index; (f) DMF semivariogram index; (g) AFM semivariogram index; and (h) RMM index.

As a result, the average NDVI (spectral feature) and RSF/RMM semivariogram indices were not able to capture the differences between LULC and SC classes. These features are sensitive to phenological variations, posing a challenge to OBCD using bitemporal images (Lu et al. 2016), probably resulting in false detection of LULC changes, without disregard the changes caused by vegetation phenology (SC).

The RVF, MFM, DMF, and AFM semivariogram indices have the potential to distinguish between homogeneous and heterogeneous objects. These indices are directly influenced by the σ^2 semivariogram parameter (overall variability), which is sensitive to LULC changes and is not affected by changes caused by vegetation seasonality (SC).

3.2. Evaluation of change detection

Table 2 summarizes the comparative results in terms of the producer's and user's accuracies, overall accuracy, and κ . The most of the geostatistical features exhibit their effectiveness in change detection analysis, with the highest overall accuracy value (95%), κ (90%). We obtained the highest producer's and user's accuracies using the σ^2 semivariogram parameter. With regard to the SC class, all the objects in the map were correctly classified (user's accuracy = 100%) with an omission error of 9.09%. The most important point to note is that all LULC changed objects were correctly mapped (producer's accuracy = 100%) with an inclusion error of 10%. The SC class also present the best producer's accuracy, reaching 90.91%.

The AFM index also provided good results, with an overall accuracy of 88.33%. Among the geostatistical features, RMM and RSF semivariogram indices provided the lower accuracies, with overall accuracy of 76.67% and 81.67%, respectively. This was expected, due to the previous graphic analysis indicates these indices were not able to accurately discriminate between LULC

change classes and SC.

Table 2. Change detection results based on confusion matrix.

Feature	Accuracy	Seasonal change	LULC change
Δ NDVI	Overall accuracy (%)		66.67
	Kappa coefficient (κ)		0.33
	Producer's accuracy (%)	60	100
	User's accuracy (%)	100	33
	Error (%)		6.55
	Confidence interval	60.12 < 66.67 < 73.22	
NDVI	Overall accuracy (%)		78.33
	Kappa coefficient (κ)		0.57
	Producer's accuracy (%)	77.42	79.31
	User's accuracy (%)	80	76.67
	Error (%)		5.72
	Confidence interval	72.6 < 78.33 < 84.05	
Sill	Overall accuracy (%)		95
	Kappa coefficient (κ)		0.9
	Producer's accuracy (%)	90.91	100
	User's accuracy (%)	100	90
	Error (%)		3.03
	Confidence interval	91.97 < 95 < 98.02	
RVF	Overall accuracy (%)		85
	Kappa coefficient (κ)		0.7
	Producer's accuracy (%)	78.38	95.65
	User's accuracy (%)	96.67	73.33
	Error (%)		4.96
	Confidence interval	80.03 < 85 < 89.96	
RSF	Overall accuracy (%)		81.67
	Kappa coefficient (κ)		0.63
	Producer's accuracy (%)	74.36	95.24
	User's accuracy (%)	96.67	66.67
	Error (%)		5.38

	Confidence interval	76.29 < 81.67 < 87.04	
	Overall accuracy (%)	85	
	Kappa coefficient (κ)	0.7	
MFM	Producer's accuracy (%)	76.92	100
	User's accuracy (%)	100	70
	Error (%)	4.96	
	Confidence interval	80.03 < 85 < 89.96	
	Overall accuracy (%)	85	
	Kappa coefficient (κ)	0.7	
DMF	Producer's accuracy (%)	76.92	100
	User's accuracy (%)	100	70
	Error (%)	4.96	
	Confidence interval	80.03 < 85 < 89.96	
	Overall accuracy (%)	76.67	
	Kappa coefficient (κ)	0.53	
RMM	Producer's accuracy (%)	73.53	80.77
	User's accuracy (%)	83.33	70
	Error (%)	5.88	
	Confidence interval	70.79 < 76.67 < 82.54	
	Overall accuracy (%)	88.33	
	Kappa coefficient (κ)	0.76	
AFM	Producer's accuracy (%)	81.08	100
	User's accuracy (%)	100	76.67
	Error (%)	4.46	
	Confidence interval	83.86 < 88.33 < 92.79	
	Overall accuracy (%)	88.33	

The σ^2 semivariogram parameter quantifies the variation explained by the spatial structure. An increase in this value over time indicates that the initial image is more homogeneous than the final image. A σ^2 semivariogram parameter that remains constant with time indicates that the variability between the two images remained constant. Similar results have been reported by other scholars.

Sertel, Kaya, and Curran (2007) investigated the relationship between semivariogram metrics and the degree of earthquake damage using satellite pour l'observation de la terre – SPOT HRVIR (high-resolution visible & infrared) imagery. They found that semivariogram shapes for pre- and post-earthquake samples differed if there was damage but were similar if the area was not damaged severely. Acerbi Júnior et al. (2015) used the semivariogram derived from NDVI images to detect changes in the Brazilian savannah and found that the values of the range and σ^2 metrics increased after deforestation, but remained similar if the land cover had not been changed.

Classification employing average NDVI values produced relatively low accuracies (overall accuracy of 78.33%), and presented values superior only to the RMM semivariogram index (overall accuracy of 76.67%) and the image differencing traditional change detection (overall accuracy of 66.67%). The variations in the NDVI values for LULC change and SC classes during the analysed period were very similar, and these small differences contributed to the poor results reported here. The image differencing change detection provided the lowest overall accuracy, however, the user's accuracy of the SC class of 100%, as well the producer's accuracy of the land use/land cover change.

Most geostatistical features provided better results when compared with spectral NDVI data and the image differencing traditional change detection. Among the geostatistical features tested, six measures of spatial dependence reached an overall accuracy of more than 80% (Figure 10).

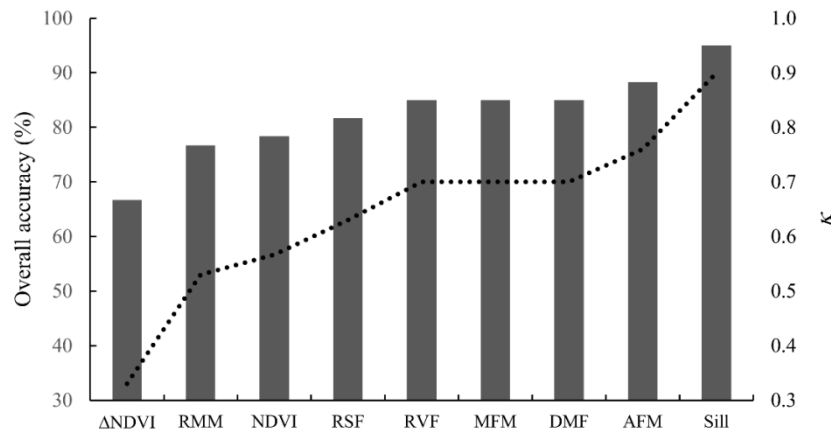


Figure 10. Features sorted by overall accuracy (bars) and kappa coefficient- κ (lines).

Geostatistical features have high accuracies in SVM classifications, indicating that these features are able to discriminate between (1) land use/land cover (LULC) change class, comprising deforestation and burned areas, and (2) seasonal changes (SC) class comprising the same cover in both epochs, but in different seasonal conditions. They provide information regarding the variability of the internal structure of objects that are not influenced by vegetation seasonality, thereby reducing confusion errors between the classes.

The change detection classes analysed were distinguished by the overall variability provided by the parameter of the semivariogram. The accuracies of the σ^2 parameter and AFM semivariogram indices were higher than those of the other indices, as they provide information that represented the overall variability of the NDVI, and the other indices provide information related to specific parts of the semivariogram.

The advantage of using geostatistical features (RVF, MFM, DMF, AFM and σ^2), instead of using spectral features and/or traditional change detection methods, is that these features are insensitive to changes caused by vegetation

seasonality; therefore, they allow for the generation of accurate maps of deforestation/burned areas, without the need of use a dense time series to eliminate the effects of forest phenology. These indices not only are able to detect LULC changes, but also, do not detect SC as LULC changes. The change detection map produced using the geostatistical, spectral feature and image differencing is presented in Figure 11.

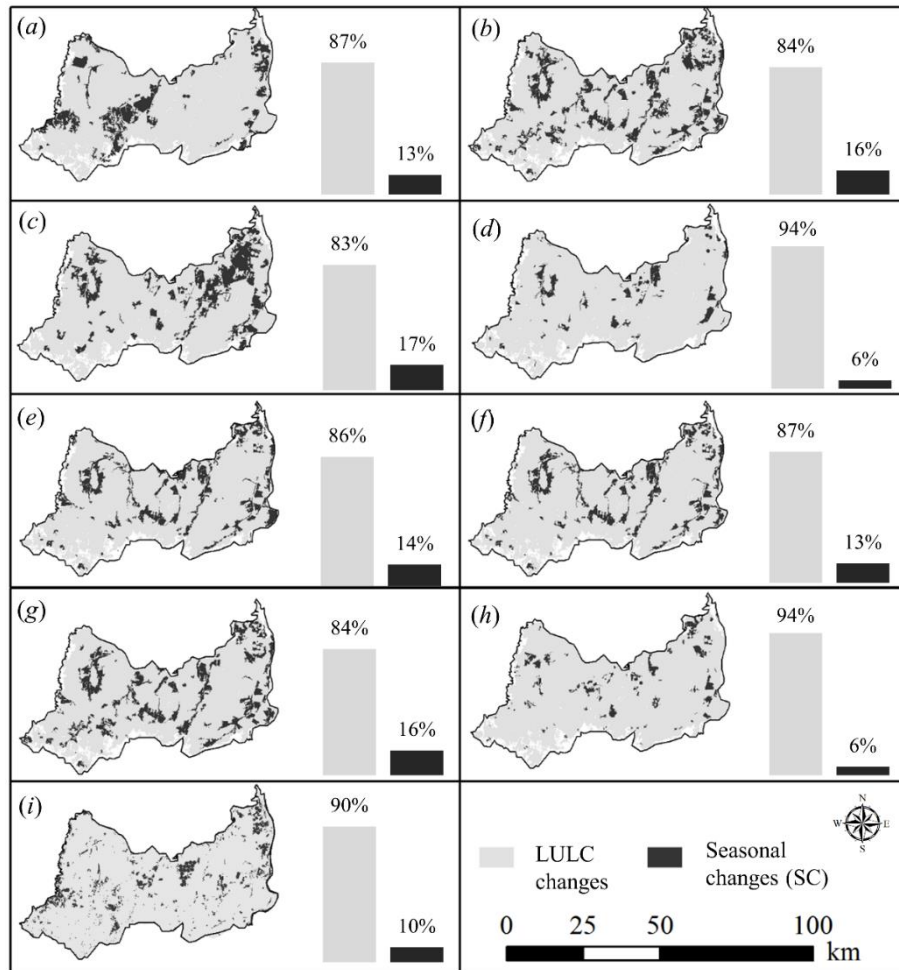


Figure 11. Deforestation map and the percentage area of LULC change and seasonal changes (SC) obtained by (a) average NDVI; (b) σ^2 semivariogram parameter; (c) RVF semivariogram index; (d) RSF semivariogram index; (e) MFM semivariogram index; (f) DMF semivariogram index; (g) AFM semivariogram index; (h) RMM semivariogram index and (i) Image differencing traditional change detection method.

3.3. Limitations to the method and further studies

In this section, we present the two major limitations of the method used in this study: (i) the predefined criteria required to generate the semivariogram and (ii) the object size of the OBIA approach. We also (iii) recommend further studies to improve this methodology and its evaluation in other areas/vegetation types for comparison purposes.

(i) Predefined criteria: In OBIA, an important input to the segmentation algorithms is the scale parameter, which controls the output object size that directly influences the semivariogram lag distance. The lag distance should not be larger than the spatial extent of the object; a small distance fails to provide a complete description of textural features, and if the lag distance is too large then short-range autocorrelation may be masked. It is necessary to find an optimal lag distance that works well for both landscape objects and image spatial resolution. To obtain the lag distance, it is necessary to set the lag size and the number of lags. When samples are located on a sampling grid, the grid spacing is usually a good indicator of the lag size. As satellite images are continuous, the lag size is usually equivalent to the size of the image spatial resolution. Tests are required to estimate the appropriate number of lags.

(ii) Object size: An image object should be adequately sized to sufficiently represent its textural pattern, with the minimum number of necessary pixels to generate the semivariogram. The main limitation in using

semivariogram indices or semivariogram parameters in OBCD is the presence of long and narrow objects. The question remains, of how to deal with semivariogram extraction for some curiously shaped objects (Wu et al. 2015).

(iii) Further studies: The following issues are of interest for future research: (1) establishing the relationship between image segmentation parameters and image spatial resolution; (2) resolving the limitations associated with narrow and long objects in detecting changes using geostatistical features in the OBIA approaches; (3) identifying and assessing the types of changes (areas converted to bare soil, pasture, or burned areas) including the initial savannah physiognomies (i.e., the grassland, shrub, wooded, and woodland forms of savannah).

4. Conclusions

We developed a method for remote sensing land use/land cover (LULC) change detection that uses the spatial context to eliminate the effects of forest phenology. The method applies an object-based approach using geostatistical features obtained from bitemporal NDVI images.

We conclude that the spatial variability of NDVI values, here represented by geostatistical features, are not affected by vegetation seasonality, and therefore, are able to accurately detect LULC changes, disregarding those associated with forest phenology, resulting in fewer classification errors. Using the most geostatistical features, the change detection results are considerably improved compared with the spectral features and image differencing technique. The highest accuracy was achieved by the sill (σ^2) parameter and the AFM (area first lag – first maximum) index, which were not affected by vegetation seasonality because they represent the information provided by the entire semivariogram rather than specific parts of it.

The study findings indicate that using the geostatistical context it is

possible to use bitemporal NDVI images to address the challenge of accurately detecting LULC changes, eliminating the effects of forest phenology, without the need to use a dense time series of remote sensing images.

The method of extracting accurate semivariogram curves from objects of long and narrow shapes needs to be further studied, along with the relationship between the scale of segmentation and image spatial resolution, including the type of change and the initial land cover class.

Acknowledgments

The authors are grateful to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Department of Forest Science of the Federal University of Lavras (UFLA) for supporting this work.

References

- Acerbi Júnior F. W., E. M. O. Silveira, J. M. Mello, C. R. Mello, and J. R. S. Scolforo. 2015. "Change Detection in Brazilian Savannas Using Semivariograms Derived from NDVI Images." *Ciencia e Agrotecnologia* 39 (2): 103–109. doi:10.1590/S1413-70542015000200001.
- Addink, E. A., F. M. B. Van Coillie, and S. M. de Jong. 2012. "Introduction to the GEOBIA 2010 Special Issue: From Pixels to Geographic Objects in Remote Sensing Image Analysis." *International Journal of Applied Earth Observation and Geoinformation* 15 (1): 1–6. doi:10.1016/j.jag.2011.12.001.
- Atkinson, P. M., and P. Lewis. 2000. "Geostatistical Classification for Remote Sensing: An Introduction." *Computers and Geosciences* 26 (4): 361–371. doi:http://dx.doi.org/10.1016/S0098-3004(99)00117-X.
- Baatz, M., and A. Schäpe. 2000. "Multiresolution Segmentation: An Optimization Approach for High Quality Multi-Scale Image

- Segmentation.” *Journal of Photogrammetry and Remote Sensing* 58 (3–4): 12–23.
- Balaguer, A., L. A. Ruiz, T. Hermosilla, and J. A. Recio. 2010. “Definition of a Comprehensive Set of Texture Semivariogram Features and Their Evaluation for Object-Oriented Image Classification.” *Computers and Geosciences* 36 (2). Elsevier: 231–240. doi:10.1016/j.cageo.2009.05.003.
- Balaguer-Beser, A., T. Hermosilla, J. A. Recio, and L. A. Ruiz. 2011. “Semivariogram Calculation Optimization for Object-Oriented Image Classification.” *Modelling in Science Education and Learning* 4 (7): 91–104. doi:10.4995/msel.2011.3057.
- Balaguer-Beser, A., L. A. Ruiz, T. Hermosilla, and J. A. Recio. 2013. “Using Semivariogram Indices to Analyse Heterogeneity in Spatial Patterns in Remotely Sensed Images.” *Computers and Geosciences* 50: 115–127. doi:10.1016/j.cageo.2012.08.001.
- Blaschke, T. 2010. “Object Based Image Analysis for Remote Sensing.” *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (1). Elsevier B.V.: 2–16. doi:10.1016/j.isprsjprs.2009.06.004.
- Cadenasso, M. L., S. T. A. Pickett, and K. Schwarz. 2007. “Spatial Heterogeneity in Urban Ecosystems: Reconceptualizing Land Cover and a Framework for Classification.” *Frontiers in Ecology and the Environment* 5 (2): 80–88. doi:10.1890/1540-9295(2007)5[80:SHIUER]2.0.CO;2.
- Carvalho L.M.T and J.R.S. Scolforo. 2008. *Inventário Florestal de Minas Gerais: Monitoramento da Flora Nativa 2005-2007*. UFLA, Lavras, p. 11.
- Chen, G., G. J. Hay, L. M. T. Carvalho, and A. A. Wulder. 2012. “Object-Based Change Detection.” *International Journal of Remote Sensing* 33 (14): 4434–4457. doi:10.1127/1432-8364/2011/0085.

- Chen, J., M. Lu, X. Chen, J. Chen, and L. Chen. 2013. "A Spectral Gradient Difference Based Approach for Land Cover Change Detection." *ISPRS Journal of Photogrammetry and Remote Sensing* 85: 1–12. doi:10.1016/j.isprsjprs.2013.07.009.
- Chen, J., J. Chen, A. Liao, X. Cao, L. Chen, X. Chen, C. He, et al. 2014. "Global Land Cover Mapping at 30 M Resolution: A POK-Based Operational Approach." *ISPRS Journal of Photogrammetry and Remote Sensing* 103: 7–27. doi:10.1016/j.isprsjprs.2014.09.002.
- Chen, X., D. Yang, J. Chen, and X. Cao. 2015. "An Improved Automated Land Cover Updating Approach by Integrating with Downscaled NDVI Time Series Data." *Remote Sensing Letters* 6 (1): 29–38. doi:10.1080/2150704X.2014.998793.
- Cohen, W. B., T. A. Spies, and G. A. Bradshaw. 1990. "Semivariograms of Digital Imagery for Analysis of Conifer Canopy Structure." *Remote Sensing of Environment* 34 (3): 167–178. doi:10.1016/0034-4257(90)90066-U.
- Congalton, R. G. 1991. "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data." *Remote Sensing of Environment* 37 (1): 35–46. doi:10.1016/0034-4257(91)90048-B.
- Coppin, P., and Bauer, M.E. 1996. "Change detection in forest ecosystems with remote sensing digital imagery." *Remote Sensing Reviews* 13: 207–234.
- Costantini, M. L., N. Zaccarelli, S. Mandrone, D. Rossi, E. Calizza, and L. Rossi. 2012. "NDVI Spatial Pattern and the Potential Fragility of Mixed Forested Areas in Volcanic Lake Watersheds." *Forest Ecology and Management* 285: 133–141. doi:10.1016/j.foreco.2012.08.029.
- Coulter, L., A.S. Hope, D. A. Stow, C. D Lippitt and S. J. Lahthrop. 2011.

- “Time–space radiometric normalization of TM/ ETM+ images for land cover change detection.” *International Journal of Remote Sensing* 32 (20): 3556–7539. doi: 10.1080/01431161.2010.524676.
- Curran, P. J. 1988. “The Semivariogram in Remote Sensing: An Introduction.” *Remote Sensing of Environment* 24 (3): 493–507. doi:10.1016/0034-4257(88)90021-1.
- Desclée, B., P. Bogaert, and P. Defourny. 2006. “Forest Change Detection by Statistical Object-Based Method.” *Remote Sensing of Environment* 102 (1–2): 1–11. doi:10.1016/j.rse.2006.01.013.
- Duro, D.C., S.E. Franklin, and M.G. Dubé. 2012. “A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery.” *Remote Sensing of Environment* 118: 259–272. doi:10.1016/j.rse.2011.11.020.
- Garrigues, S., D. Allard, F. Baret, and M. Weiss. 2006. “Quantifying Spatial Heterogeneity at the Landscape Scale Using Variogram Models.” *Remote Sensing of Environment* 103 (1): 81–96. doi:10.1016/j.rse.2006.03.013.
- Garrigues, S., D. Allard, F. Baret, and J. Morisette. 2008. “Multivariate Quantification of Landscape Spatial Heterogeneity Using Variogram Models.” *Remote Sensing of Environment* 112 (1): 216–230. doi:10.1016/j.rse.2007.04.017.
- Gil-Yepes, J. L., L. A. Ruiz, J. A. Recio, A. Balaguer-Beser, and T. Hermosilla. 2016. “Description and Validation of a New Set of Object-Based Temporal Geostatistical Features for Land-Use/land-Cover Change Detection.” *ISPRS Journal of Photogrammetry and Remote Sensing* 121: 77–91. doi:10.1016/j.isprsjprs.2016.08.010.

- Ghosh, S., M. Roy, and A. Ghosh. 2014. "Semi-Supervised Change Detection Using Modified Self-Organizing Feature Map Neural Network." *Applied Soft Computing Journal* 15: 1–20. doi:10.1016/j.asoc.2013.09.010.
- Hamunyela, E., J. Verbesselt, and M. Herold. 2016. "Using spatial context to improve early detection of deforestation from Landsat time series." *Remote Sensing of Environment* 172: 126 – 138. doi: 10.1016/j.rse.2015.11.006.
- Huang, Y., X. Yin, G. Ye, J. Lin, R. Huang, N. Wang, L. Wang, and Y. Sun. 2013. "Spatio-Temporal Variation of Landscape Heterogeneity under Influence of Human Activities in Xiamen City of China in Recent Decade." *Chinese Geographical Science* 23 (2): 227–236. doi:10.1007/s11769-012-0577-2.
- Hussain, M., D. Chen, A. Cheng, H. Wei, and D. Stanley. 2013. "Change Detection from Remotely Sensed Images: From Pixel-Based to Object-Based Approaches." *ISPRS Journal of Photogrammetry and Remote Sensing* 80: 91–106. doi:10.1016/j.isprsjprs.2013.03.006.
- Jensen J. R. 2004. *Introductory Digital Image Processing: A Remote Sensing Perspective. Third edition.* Upper Saddle River, NJ, p. 526.
- Jin, S., L. Yang, P. Danielson, C. Homer, J. Fry, and G. Xian. 2013. "A Comprehensive Change Detection Method for Updating the National Land Cover Database to circa 2011." *Remote Sensing of Environment* 132: 159–175. doi:10.1016/j.rse.2013.01.012.
- Johansen, K., L. A. Arroyo, S. Phinn, and C. Witte. 2010. "Comparison of Geo-Object Based and Pixel-Based Change Detection of Riparian Environments Using High Spatial Resolution Multi-Spectral Imagery." *Photogrammetric Engineering and Remote Sensing* 76 (2): 123–136. doi:10.14358/PERS.76.2.123.

- Kulkarni, A. D., and B. Lowe. 2016. "Random Forest for Land Cover Classification." *International Journal on Recent and Innovation Trends in Computing and Communication* 4 (3): 58–63.
- Lausch, A., M. Pause, D. Doktor, S. Preidl, and K. Schulz. 2013. "Monitoring and Assessing of Landscape Heterogeneity at Different Scales." *Environmental Monitoring and Assessment* 185 (11): 9419–9434. doi:10.1007/s10661-013-3262-8.
- Li, L., and M. K. H. Leung. 2002. "Integrating Intensity and Texture Differences for Robust Change Detection." *IEEE Transactions on Image Processing* 11 (2): 105–112. doi:10.1109/83.982818.
- Listner, C., and I. Niemeyer. 2011. "Object-based change detection." *Photogrammetrie - Fernerkundung - Geoinformation* 4: 233–245. doi:10.1127/1432-8364/2011/0085.
- Lu, D., and E. S. Brondizio. 2004. "Change Detection Techniques." *International Journal of Remote Sensing* 25 (12): 2365 – 2407. doi:10.1080/0143116031000139863.
- Lu, M., J. Chen, H. Tang, Y. Rao, P. Yang, and W. Wu. 2016. "Land Cover Change Detection by Integrating Object-Based Data Blending Model of Landsat and MODIS." *Remote Sensing of Environment* 184: 374–386. doi:10.1016/j.rse.2016.07.028.
- Mantero, P., G. Moser, and S. B. Serpico. 2005. "Partially Supervised Classification of Remote Sensing Images through SVM-Based Probability Density Estimation." *IEEE Transactions on Geoscience and Remote Sensing* 43 (3): 559–570. doi:10.1109/TGRS.2004.842022.
- Miranda, F. P., L. E. N. Fonseca, J. R. Carr, and J. V. Taranik. 1996. "Analysis of JERS-1 (Fuyo-1) SAR Data for Vegetation Discrimination in

- Northwestern Brazil Using the Semivariogram Textural Classifier (STC).” *International Journal of Remote Sensing* 17 (17): 3523–3529. doi:10.1080/01431169608949167.
- Morisette, J. T., and S. Khorram. 2000. “Accuracy Assessment Curves for Satellite-Based Change Detection,” *Photogrammetric Engineering & Remote Sensing* 66 (7): 875–80.
- Mui, A., Y. He, and O. Weng. 2015. “An Object-Based Approach to Delineate Wetlands across Landscapes of Varied Disturbance with High Spatial Resolution Satellite Imagery.” *ISPRS Journal of Photogrammetry and Remote Sensing* 109: 30–46. doi:10.1016/j.isprsjprs.2015.08.005.
- Peel, M. C., B. L. Finlayson, and T. A. McMahon. 2006. “Updated World Map of the Köppen -Geiger Climate Classification.” *Meteorologische Zeitschrift* 15: 259–263. doi:10.1127/0941-2948/2006/0130.
- Powers, R. P., T. Hermosilla, N. C. Coops, and G. Chen. 2015. “Remote Sensing and Object-Based Techniques for Mapping Fine-Scale Industrial Disturbances.” *International Journal of Applied Earth Observation and Geoinformation* 34 (1): 51–57. doi:10.1016/j.jag.2014.06.015.
- Qiu, B., C. Zeng, C. Cheng, Z. Tang, J. Gao, and Y. Sui. 2013. “Characterizing Landscape Spatial Heterogeneity in Multisensor Images with Variogram Models.” *Chinese Geographical Science* 24 (3): 1–11. doi:10.1007/s11769-013-0649-y.
- Radke, R.J., S. Andra, O. Al-Kofahi, and B. Roysam. 2005. “Image Change Detection Algorithms: A Systematic Survey.” *IEEE Transactions on Image Processing* 14 (3): 294–307. doi:10.1109/TIP.2004.838698.
- Ruiz, L. A., J. A. Recio, A. Fernández-Sarría, and T. Hermosilla. 2011. “A Feature Extraction Software Tool for Agricultural Object-Based Image

- Analysis.” *Computers and Electronics in Agriculture* 76 (2): 284–296. doi:10.1016/j.compag.2011.02.007.
- Schwieder, M., P.J. Leitão, M.M.C. Bustamante, L.F Guimarães, A. Rabe, and P. Hostert. 2016. “Mapping Brazilian savanna vegetation gradients with Landsat time series.” *International Journal of Applied Earth Observations and Geoinformation* 52, 361–370.
- Sertel, E., S. Kaya, and P. J. Curran. 2007. “Use of Semivariograms to Identify Earthquake Damage in an Urban Area.” *IEEE Transactions on Geoscience and Remote Sensing* 45 (6): 1590–1594. doi:10.1109/TGRS.2007.894019.
- Shao, Y., and R. S. Lunetta. 2012. “Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points.” *ISPRS Journal of Photogrammetry and Remote Sensing* 70: 78–87. doi:10.1016/j.isprsjprs.2012.04.001.
- Silveira, E. M.O., M. D. Menezes, F. W. Acerbi Júnior, M. C. N. Santos, and J. M. Mello. 2017. “Assessment of Geostatistical Features for Object-Based Image Classification of Contrasted Landscape Vegetation Cover.” *Journal of Applied Remote Sensing* 11 (3): 1-15. doi:10.1117/1.JRS.11.036004.
- Tewkesbury, A. P., A. J. Comber, N. J. Tate, A. Lamb, and P. F. Fisher. 2015. “A Critical Synthesis of Remotely Sensed Optical Image Change Detection Techniques.” *Remote Sensing of Environment* 160: 1–14. doi:10.1016/j.rse.2015.01.006.
- Verbesselt, J., R. Hyndman, G. Newnham, and D. Culvenor. 2010. “Detecting trend and seasonal changes in satellite image time series.” *Remote Sensing of Environment* 114 (1): 106-115. doi: 10.1016/j.rse.2009.08.014
- Vorovencii, I. 2014. “Assessment of Some Remote Sensing Techniques Used to Detect Land Use/land Cover Changes in South-East Transilvania,

- Romania.” *Environmental Monitoring and Assessment* 186 (5): 2685–2699. doi:10.1007/s10661-013-3571-y.
- Whiteside, T. G., G. S. Boggs, and S. W. Maier. 2011. “Comparing Object-Based and Pixel-Based Classifications for Mapping Savannas.” *International Journal of Applied Earth Observation and Geoinformation* 13 (6): 884–93. doi:10.1016/j.jag.2011.06.008.
- Woodcock, C. E., A. H. Strahler, and D. L. B. Jupp. 1988. “The Use of Variograms in Remote Sensing: I. Scene Models and Simulated Images.” *Remote Sensing of Environment* 25 (3): 323–348. doi:10.1016/0034-4257(88)90108-3.
- Wu, J., D. E. Jelinski, M. Luck, and P. T. Tueller. 2000. “Multiscale Analysis of Landscape Heterogeneity: Scale Variance and Pattern Metrics.” *Annals of GIS* 6 (1): 6–19. doi:10.1080/10824000009480529.
- Wu, X., D. E. Jelinski, M. Luck, and P. T. Tueller. 2010. “Land Cover Change Detection Using Texture Analysis.” *Journal of Computer Science* 6 (1): 92-100.
- Wu, X., J. Peng, J. Shan, and W. Cui. 2015. “Evaluation of Semivariogram Features for Object-Based Image Classification.” *Geo-Spatial Information Science* 18 (4): 159–170. doi:10.1080/10095020.2015.1116206.
- Yue, A., C. Zhang, J. Yang, W. Su, W. Yun, and D. Zhu. 2013. “Texture Extraction for Object-Oriented Classification of High Spatial Resolution Remotely Sensed Images Using a Semivariogram.”

- International Journal of Remote Sensing* 34 (11): 3736–3759.
doi:10.1080/01431161.2012.759298.
- Zhang, H., J. E. Fritts, and S. A. Goldman. 2008. “Image Segmentation Evaluation : A Survey of Unsupervised Methods” *Computer Vision and Image Understanding* 110: 260–80.
doi:10.1016/j.cviu.2007.08.003.
- Zhu, Z. 2017. “Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications.” *ISPRS Journal of Photogrammetry and Remote Sensing* 130: 370-384.
doi:10.1016/j.isprsjprs.2017.06.013
- Zuo, R., E. John, M. Carranza. 2011. “Support vector machine : A tool for mapping mineral prospectivity.” *Computers and Geosciences*: 37: 1967–1975. doi:10.1016/j.cageo.2010.09.014.
- Zurlini, G., K. Riitters, N. Zaccarelli, I. Petrosillo, K.B. Jones, and L. Rossi.2006. “Disturbance patterns in a socio-ecological system at multiple scales.” *Ecological Complexity* 3(2):119-128. doi: 10.1016/j.ecocom.2005.11.002.

**ARTICLE 5 - DISENTANGLING THE EFFECTS OF FOREST
PHENOLOGY AND LAND-COVER CHANGES IN BRAZILIAN
SEASONAL BIOMES COMBINING SPATIAL AND SPECTRAL
REMOTE SENSING FEATURES**

Eduarda Martiniano de Oliveira Silveira^{1*}, Fernando Del Bon Espírito-Santo²,
Fausto Weimar Acerbi Júnior¹, Kieran Daniel Withey², José Márcio de Mello¹,
Peter Atkinson²

¹ Forest Science Department (DCF), Federal University of Lavras (UFLA),
Lavras, Brazil;

² Lancaster Environment Center (LEC), Lancaster University, UK, LA1 4YQ;

* Correspondence: dudalavras@hotmail.com; Tel.: +55-012-99644-6259;

Publication status: In prep. for submission to **Remote Sensing**

Abstract: The Brazilian Savannas and Semi-arid woodland biomes have been under increasing anthropic pressure for many years, and land-cover (LC) changes in these seasonal ecosystems have been largely neglected. Remote sensing provides useful tools to detect forest changes, but previous studies have not attempted to separate the effects of forest phenology and land-cover changes, as deforestation, logging and fires to improve the accuracy of change detection using bitemporal Landsat data. Here we developed an object-based remote sensing method that is able to disentangle the effects of forest phenology and land-cover changes by combining spectral and the spatial context using traditional spectral features (SF) and semivariogram indices (SI), exploring the full capability of NDVI (normalized difference vegetation index) Landsat images. We used random forest (RF), support vector machines (SVM) and artificial neural network (ANN) algorithms to train the input features. We found that the temporal spatial variability of NDVI values is not affected by vegetation seasonality and, therefore, the combination of spectral features and semivariogram indices provided the best global accuracy (~92.27%) to separate the seasonal changes and land-cover changes. This study underscores that the use of spatial features reduces the need of multi-temporal satellite images to accurately extract land-cover changes such as deforestation, logging or fire while disregarding the ones caused by phenological differences.

Keywords: Geostatistics, Semivariogram, Change detection, Savannas, Semi-arid woodland

1. Introduction

The Brazilian Savannas and Semi-arid woodland biomes, which together cover approximately 35% of the Brazilian territory, are placed among the most endangered ecoregions on Earth due to high rates of conversion and few protected areas (Hoekstra et al., 2005). The Brazilian Savannas (also known as Cerrado) have suffered heavy losses of natural vegetation due to agricultural expansion (Silva et al., 2006) and the Semi-arid woodland (Caatinga) has been indicated as one of the least known and most neglected of Brazilian biomes (Santos et al., 2011). Most of the efforts for estimating forest cover changes have been focused on the tropical rain forests, with far less attention dedicated to the less humid seasonal regions (Beuchle et al., 2015). Although projects and studies assessed land-cover (LC) changes in these biomes based on remote sensing images, the accuracy of these results is under question due to their large areas, highly dynamic and the influence of vegetation seasonality on change detection that poses a significant challenge in remote sensing approaches (Lu et al., 2016).

Estimating land-cover changes by remote sensing is not a trivial task, since satellite images contain a combination of natural forest phenology (driven by annual temperature and rainfall interactions impacting plant phenology), anthropogenic disturbances (i.e. deforestation, urbanization, floods, and fire), cloud cover, atmospheric scattering and geometric errors (Roy et al., 2002). The vast majority of change detection algorithms and techniques involve the comparison of images from just two dates (bitemporal) (Healey et al., 2005; Masek et al., 2008), interpreting natural phenological changes as land-cover changes (Verbesselt et al., 2010; Zhu et al., 2012). Thus, an important issue in remote sensing change detection is how to accurately extract land-cover changes such as deforestation, logging and fires while disregarding the ones caused by phenological differences (Chen et al., 2013; Jin et al., 2013).

Verbesselt et al., (2010) developed a powerful change detection approach for time series by detecting and characterizing Breaks For Additive Seasonal and Trend (BFAST). Changes occurring in the trend component often indicate disturbances, while changes occurring in the seasonal component indicate phenological changes. The use of a seasonal model is based on an assumption that there is an identifiable seasonal pattern in the time series that can be described mathematically. However, this assumption may not always hold if the time series is for satellite images which are not acquired at regular interval, or have wide temporal gaps due to persistent cloud cover (Asner, 2001). Instead of using a seasonal model to account for seasonality in image time series, Hamunyela et al., (2016) used the spatial context, in a pixel-based approach. They concluded that the spatial context approach is useful for timely detection of deforestation events in areas where forests exhibit strong seasonality. However, the gain in the overall spatial accuracy was only marginally higher when using spatial context than when using a seasonal model.

Recently, Silveira et al., (2018) developed a method for remote-sensing change detection that is insensitive to changes caused by vegetation seasonality, using the spatial context and bitemporal images instead of time series. The study findings indicate that using the geostatistical context, it is possible to use bitemporal normalized difference vegetation index (NDVI) images to address the challenge of accurately detecting land-cover changes, eliminating the effects of forest phenology, resulting in fewer classification errors.

Motivated by the critical need of development of methods that can address the issue of seasonality using bitemporal images, without the need of use dense time series and seasonal models, we evaluated a method to disentangle the effects of forest phenology and land-cover changes using an object-based approach that combines spatial and spectral features using bitemporal NDVI images. We compared the performance of three machine learning algorithms

that represent new directions in change detection: artificial neural networks (ANN), random forests (RF), and support vector machines (SVM). Specifically, we address the following research question: How well can we differentiate between seasonal and land-cover changes combining spatial and spectral features derived from NDVI Landsat images?

To address these research questions, we used a set of semivariogram indices (SI) calculated from semivariograms obtained from NDVI Landsat images as spatial features. The semivariogram indices were combined with spectral features (SF) to improve the accuracy of change detection. We performed the analysis at the object level, that reduces the small spurious changes (Hussain et al., 2013), and also allow the incorporation of spatial information (Blaschke, 2010). We used an area of seasonal Savanna and Semi-arid woodland Brazilian biome that is undergoing rapid land-cover changes in the past years to test the efficacy of this method.

2. Materials and Methods

We propose a new method based on object-based change detection (OBCD) that combines semivariogram indices and spectral features obtained from bitemporal NDVI Landsat images as input data to train machine learning algorithms, focusing on separate the effects of forest phenology from land-cover changes. The method is divided into six steps (Figure 1).

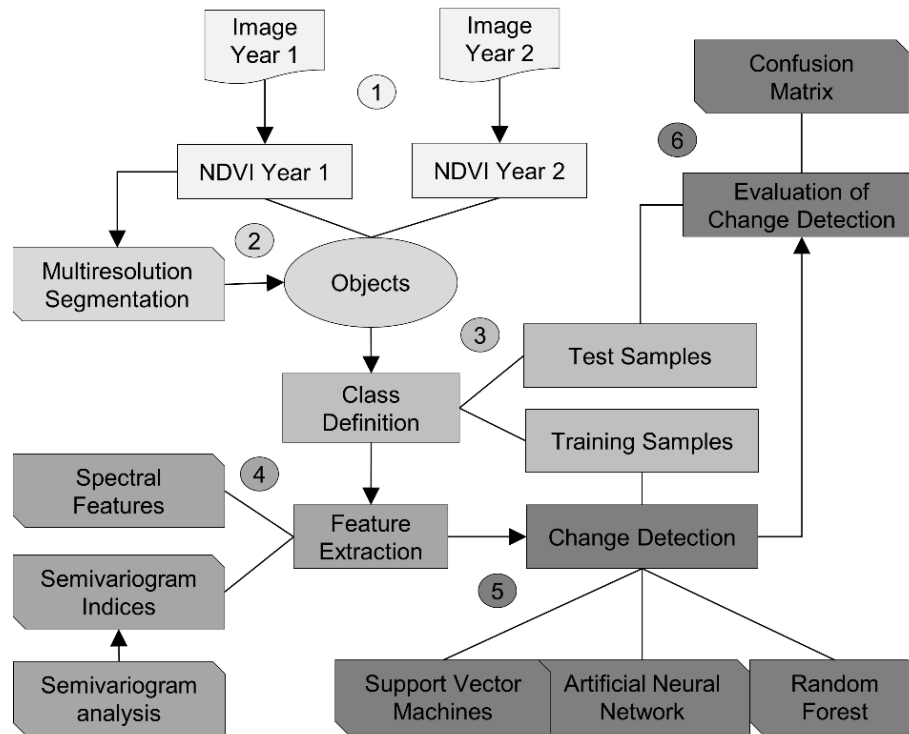


Figure 1. Workflow of the proposed method.

2.1. Study Area and Data

Our study comprises Landsat path/row 219/71 in the north of the state of Minas Gerais (MG), Brazil (Figure 2). This region, which totals approximately 31,742 km², exhibits a transition between Savanna and Semi-arid woodland biomes. The Savanna and Semi-arid woodland biomes are seasonal ecosystems characterized by distinct dry and wet seasons. They are located in the centre and the Northeast of Brazil, covering nearly 35% of the Brazilian territory.

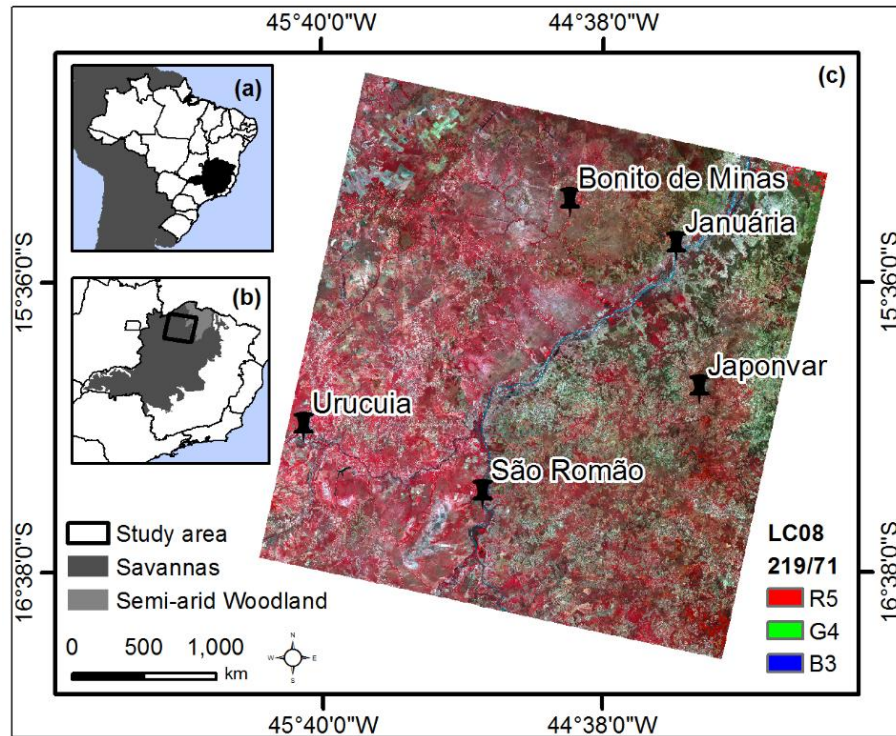


Figure 2. (a) Minas Gerais state located in the southeast of Brazil; (b) Study area, covering Brazilian Savanna and Semi-arid woodland biomes; (c) Landsat 8 OLI image of the study area.

The diversity of the vegetation types of these biomes (Figure 3) is well documented ranging from closed or open canopy deciduous and semi-deciduous forest, grassland to woodland cerrado (Ferreira et al., 2004; Arantes et al., 2016) and deciduous forests ranging from deciduous low shrub to small patches of tall dry forests (Santos et al., 2012). The climate is tropical with rain concentrated in October–May, while the dry season has close to zero rainfall in some months, with air humidity less than 20% in August and September, which characterizes the high seasonality in the region (Peel et al., 2006). The rainfall in Caatinga biome is extremely irregular, in both its temporal and geographical distribution;

usually more than 75% of the total annual rainfall occurs within three months. The annual variations are large; droughts can last for years (Leal et al., 2005).

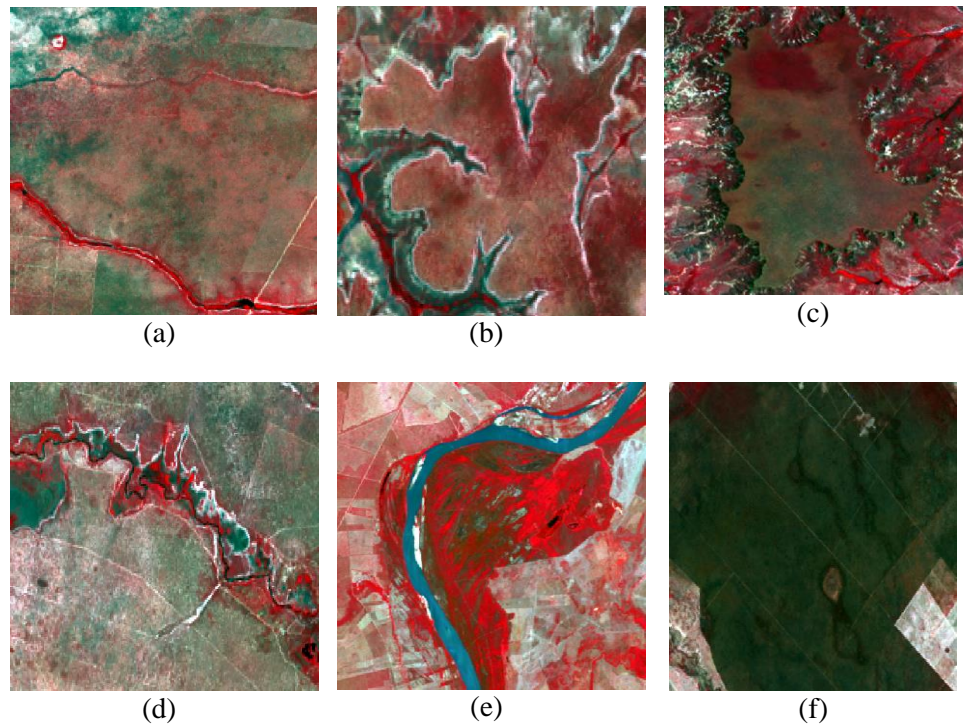


Figure 3. Landsat 8 OLI images (RGB colour composite, 543) of the vegetation types of the study area: (a) Cerrado grassland (open grassland); (b) Shrub cerrado (open grassland with sparse shrubs); (c). Cerrado woodland (mixed grassland, shrublands, and trees up to seven meters in height); (d) Palm swamps (riparian vegetation); (e) Semideciduous forest (semideciduous canopy foliage); (f) Deciduous forest (predominance of deciduous individuals whose loss of foliage reaches more than 50%).

We used two Landsat images (8 OLI – Operational Land Imager from 2015/06/19 and 2016/10/27) from the United States Geological Survey (USGS).

The acquired images were geometrically and atmospherically (surface reflectance) corrected by USGS (high level product). We selected the images not only avoiding presence of clouds, but also selecting images around dry and wet months to maximize the effects of vegetation seasonality. We used the normalized difference vegetation index (NDVI) (Tucker, 1979) of each period. Although NDVI is a simple remote sensing index, it minimizes the effect of shadows caused by the terrain's topography (Vorovencii, 2014).

2.2. Image Segmentation

We used object-based image analysis (OBIA) because, in general, it provides better results of image classifications than traditional pixel by pixel methods (Tewkesbury et al., 2015). OBIA makes use not only of the spectral information, but also the spatial features of remote sensing objects (Balaguer et al., 2010; Balaguer-Beser et al., 2013; Yue et al., 2013; Powers et al., 2015; Wu et al., 2015; Gil-Yepes et al., 2016; Silveira et al., 2017).

For each NDVI Landsat image we applied an automatic segmentation (Addink et al., 2012) using the eCognition software for object-based image analysis. The multiresolution segmentation algorithm (Baatz and Shaped, 2000) was set with three parameters: shape, compactness, and scale. These parameters control the shape, size, and spectral variation of segmented image objects. The shape parameter was set to 0.1 and compactness was set to 0.5. The most critical step is the selection of the scale parameter, which controls the size of the image objects. The scale parameter sets a homogeneity threshold that determines the number of neighbouring pixels that can be merged together to form an image object (Benz et al., 2004) and directly influences the size of the objects connected to the semivariogram predefined criteria (lag distance) and also the minimum number of pixels inside each object necessary to generate the semivariogram. We used a trial and error approach (Duro et al., 2012) to find the

appropriate scale parameter (Chen et al., 2015) to guarantee a minimum number of samples (25 pixels) inside the objects.

The main limitation of using semivariogram features in OBCD is the presence of long and narrow objects as well as small objects without the minimum number of pixels necessary to generate the semivariogram. To minimize this issue, the multiresolution segmentation was conducted based on the initial Landsat 8 OLI image (Year 1). We extracted the NDVI values within the objects to obtain the SI and SF used here as input data for the change detection analysis.

2.3. Class Definition

This study focused in two classes: (i) no-change: landscape vegetation cover that had the same land-cover in both time, encompassing seasonal variations due vegetation phenology (Figure 4ab); (ii) land-cover changes: landscape vegetation cover under deforestation, logging or fires. The land-cover changes samples comprised small changes reaching up to 50% of the objects (Figure 4cd) and big changes ranging from 50 to 100% of the objects (Figure 4ef). The classes were defined according to a prior visual interpretation and sampled using a stratified random design.

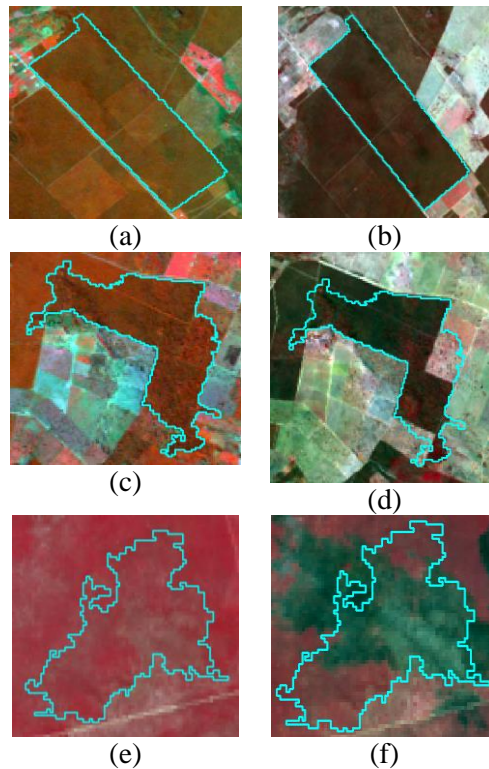


Figure 4. Landsat images 8 OLI (RGB 543 composite) illustrating the LULC classes: (a) Wet deciduous forest in 2015; (b) Dry deciduous forest in 2016; (c) Wet cerrado woodland in 2015; (d) Dry cerrado woodland with 25% of loss caused by deforestation in 2016; (e) Grassland cerrado in 2015 and (f) Grassland cerrado in 2016 with 75% of loss caused by fire 2016.

2.4. Spatial Features

We used as spatial features the information provided by semivariogram derived from NDVI images (Atkinson and Lewis, 2000). The experimental semivariogram (1) is given by:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x) - Z(x+h)]^2 \quad (1)$$

Where $\gamma(h)$ is the estimator of the semivariance for each distance h , $N(h)$ is the number of pairs of points separated by distance h , $Z(x)$ is the value of the regionalized variable at point x , and $Z(x+h)$ is the value at point $(x+h)$.

The semivariogram is given by plotting the spatial variance against the distance h , and enables the variance among different combinations of pairs of points to be estimated. The semivariance functions are characterized by three parameters: sill (σ^2), range (ϕ), and nugget effect (τ^2). The sill is the plateau reached by the semivariance values, and measures the quantity of variation explained by the spatial structure of the data. The range is the distance until the semivariogram reaches the sill, reflecting the distance at which the data become correlated. The nugget effect is the combination of sampling errors and variations in scale that occurs over scales of less than the distance between the sampled points (Curran, 1988).

To verify how the temporal spatial variability of NDVI values within objects are affected by vegetation seasonality, we compared the semivariance sill (σ^2) of each individual object in both years. We attempted to find an optimal lag distance to ensure that sill values would provide a concise description of data variability. We fixed the number of lags as 30 pixels and the lag size equivalent to image spatial resolution (30 m), resulting in a lag distance of 900 m. We then analysed the shape and overall variability of the semivariograms.

The information generated by semivariogram can be achieved by: (i) Modelling the semivariogram by fitting a model to extract the parameters (Acerbi Junior et al., 2015; Silveira et al., 2018); (ii) By using raw semivariograms at various lag and window sizes (Berberoğlu et al., 2010) and (iii) By using indices calculated from the semivariograms (Balaguer et al., 2010). Fitting a model to extract semivariogram parameters require the selection of a proper function (Wu et al., 2015), raw semivariograms values and windows sizes makes necessary to find an appropriate window size and to considerer the

border effects. The semivariogram indices synthesize the most relevant information about the shape of the semivariogram and enhance the information contained on the first lags, where spatial correlation at short distances is higher. In addition, fitting a model is unnecessary, which improves the processing time and reduces the errors generated by choosing an incorrect model.

Thus, to verify how well can we differentiate between seasonal and land-use changes combining geostatistical and spectral features derived from NDVI Landsat images, we used eight semivariogram indices (Balaguer et al., 2010) available in the software FETEX 2.0 (Ruiz et al., 2011) (Table 1). These indices describe the shape of the experimental semivariograms, and therefore the properties that characterize the spatial patterns of the image objects, thus providing textural information that may be used for change detection analysis (Table 1).

Table 1. Semivariogram indices used in this research.

Index	Description	Formula
RVF*	Ratio between the values of the total variance and the semivariance at first lag	$RVF = \frac{\text{Variance}}{\gamma_1}$
RSF*	Ratio between semivariance values at second and first lag	$RSF = \frac{\gamma_2}{\gamma_1}$
FDO*	First derivative near the origin	$FDO = \frac{\gamma_2 - \gamma_1}{h}$
SDT*	Second derivative at third lag	$SDT = \frac{\gamma_4 - 2\gamma_3 + \gamma_2}{h^2}$
MFM**	Mean of the semivariogram values up to the first maximum	$MFM = \frac{1}{\text{Max}_1} \sum \gamma_i$
DMF**	Difference between the mean of the semivariogram values up to the first maximum (MFM) and the semivariance at first lag	$DMF = MFM - \gamma_1$
RMM**	Ratio between the semivariance at first local maximum and the mean semivariogram values up to this maximum	$RMM = \frac{\gamma_{\text{max}_1}}{\gamma_{\text{max}_1}^{\text{mean}}}$
AFM**	Semivariance curvature	$AFM = \frac{h}{2} \left(\gamma_1 + 2 \left(\sum_{i=2}^{\text{max}_1-1} \gamma_i \right) + \gamma_{\text{max}_1} \right) - (\gamma_1 (h_{\text{max}_1} - h_1))$

* Near the origin; **Up to the first maximum.

2.5. Spectral Features

As spectral features, we considered the minimum (MIN), mean (MEAN), maximum (MAX), and standard deviation (STDEV) of NDVI values inside each object. This allows the performance of spatial and spectral features to be compared and combined.

2.6. Change Detection Algorithms

We chose three different machine learning algorithms to classify the changes. We compared the performance of an artificial neural network (ANN), support vector machine (SVM), and a random forest (RF). The samples were divided into two parts, one for classifier training and the other for classification assessment. Half of the samples were chosen at random for training, with the remaining 50% used as evaluation samples.

2.6.1. Artificial Neural Network (ANN)

We used an ANN algorithm as they are nonparametric and make no assumptions about data distribution and independency. They adaptively estimate continuous functions from data without specifying mathematically how outputs depend on inputs (Im and Jensen, 2005). ANN algorithms learn from the training dataset and build relationships (networks) between input (features) and output nodes (changes) (Hussain et al., 2013). The trained network then is applied to the test dataset (Gopal and Woodcock, 1996). There are many different types of ANN, and one that is widely used in remote sensing applications is the multilayer perceptron (MLP) (Benediktsson et al., 1990). This type of model consists of three or more interconnected layers. The first layer, termed the input layer, serves as a distribution structure for the data being presented to the network. The final processing layer is called the output layer, and the layers between the input and output layers are termed hidden layers (Berberoglu et al., 2000).

We used the ANN obtained by running the Multilayer Perceptron function (of the multilayer perceptron type) provided by WEKA 3.8 software. The training of neural networks occurred through the back-propagation algorithm, which fit the weights of all the layers of the network from the back-propagation of the error, obtained in the output layer. The updating of weights

was carried out according to the error, learning rate, and momentum terms (Delta rule).

2.6.1. Support Vector Machine (SVM)

The SVM algorithm is extended to nonlinearly separable classes by mapping the samples to a higher-dimensional feature space using a kernel function (Kulkarni and Lowe, 2016). We used the radial basis function (RBF) kernel, as this is known to be effective and accurate. To train the SVM classifier, an error parameter C and a kernel parameter γ were set. SVMs are particularly useful for remote sensing because they have the ability to handle small training datasets, often producing higher classification accuracy than traditional methods (Mantero et al., 2005). The algorithm was implemented in the WEKA 3.8 software under the Sequential Minimal Optimization function. The values of C and σ (the bandwidth or influence range of each training point in the RBF) were tested within intervals of 10^i ($i = -3, \dots, 3$), and the least-squares mean error configuration was chosen for the application.

2.6.2. Random Forest (RF)

The RF algorithm, initially proposed by (Breiman, 2001) is an ensemble method which generates a set of individually trained decision trees and combines their results. RF is a robust non-parametric classifier and has the ability to accommodate many predictor variables (Devries et al., 2016). The advantages of RFs include excellent accuracy, efficient implementation on large datasets, and a structure that enables the future use of pre-generated trees (Breiman, 2001).

We used the open-source software WEKA 3.8 to fit the RF. Two parameters need to be set in order to produce the forest trees: the number of decision trees to be generated (Ntree) and the number of variables to be selected and tested for the best split when growing the trees (Mtry) (Belgiu and Drăgu, 2016). Five hundred trees were grown for each classification (Millard and

Richardson, 2015). The Ntree required to maintain a certain level of accuracy has been assessed by several authors, and the minimum number of trees for optimal classification appears to be between 100 (Lawrence et al., 2006) and 300 (Akar and Güngör, 2015), with the majority of studies setting the Ntree value to 500 because the errors stabilize before this number of classification trees is achieved (Lawrence et al., 2006). Since the RF classifier is computationally efficient and does not over fit, Ntree can be arbitrarily large (Guan et al., 2013). Since theoretical and empirical research has highlighted that classification accuracy is less sensitive to Ntree than to the Mtry parameter (Guan et al., 2013), the number of features was left at its default value (log of the number of features + 1) (Millard and Richardson, 2015).

2.7. Change Detection Evaluation

We tested (i) semivariogram indices; (ii) spectral features; and (iii) their combination. Change detection performance was evaluated using a confusion matrix (Congalton, 1991). From this, we measured the overall accuracy, producer's and user's accuracy and the kappa coefficient.

3. Results

3.1. Semivariogram Analysis

We analysed the temporal spatial variability of NDVI data comparing the semivariogram shape and parameters. The semivariograms reached the sill within the calculated distance (900m), indicating that their spatial extents were sufficiently large to encompass the entire spatial variability. From several remote sensing object of no-change class (landscape vegetation cover that had the same land-cover classes in both time, encompassing seasonal variations due to vegetation phenology), the semivariogram generated using NDVI images of 2015 had similar behaviour to the semivariogram of NDVI 2016 (Figure 5a and

5b). On the other hand, the semivariogram representing objects of land-cover change class (i.e deforestation, logging or fires) increased considerably from 2015 to 2016 (Figure 5c and 5d).

The low variability of NDVI values of the landscape types covering the study area (physiognomies of Savanna and Semi-arid woodland biomes) in 2015 is explained by the high and homogeneous values of this index inside the objects. In 2016 the land-cover did not change, however the NDVI values decreased due to the effect of vegetation seasonality. Nevertheless, the NDVI variability did not change (overall variability ~ 0.0005) because the phenology affected the whole object. The values decreased from 2015 to 2016, but its variability did not.

In the presence of land-cover changes, the high variability of the objects is explained by the mix of bare soil/burnt areas and vegetation. The increase in overall variability (~ 0.0030) after change is explained by the combination of high NDVI values for the remaining vegetation inside the objects and low NDVI values for bare soil/burnt areas.

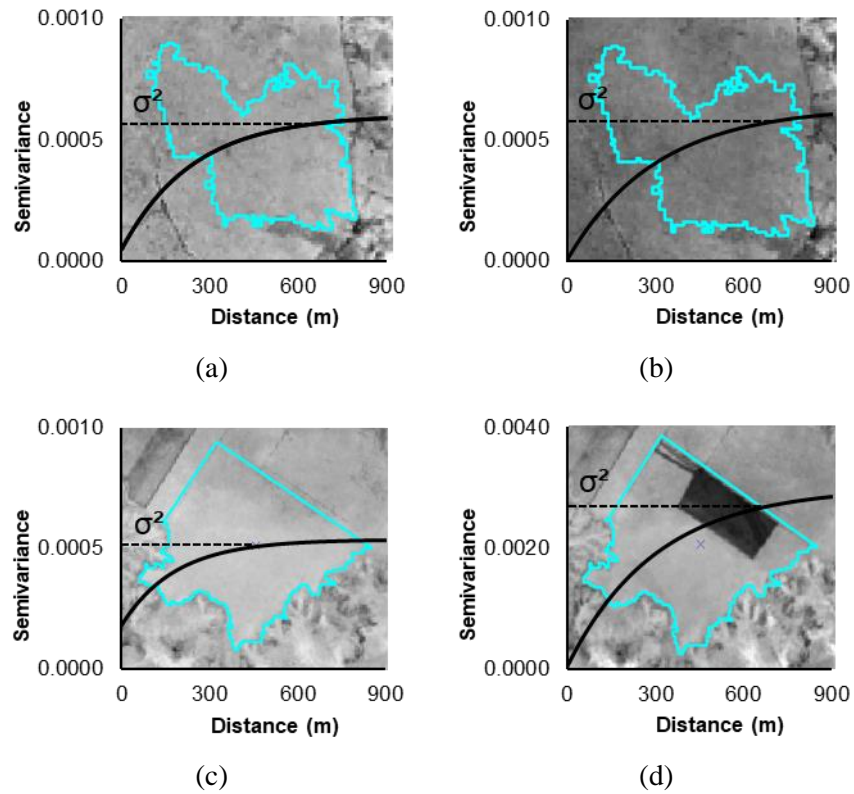


Figure 5. Semivariogram examples of NDVI values extracted from objects covering by: (a) Wet physiognomy in 2015; (b) Dry physiognomy in 2016; (c) Wet physiognomy in 2015; (d) Dry physiognomy in 2016 affected by deforestation.

These results indicate a very clear trend where by the overall variability and shape, the semivariogram between 2015 and 2016 were different when land-cover changes (i.e. deforestation) occurred, and were similar when the area had not undergone any land-cover change. These results demonstrate that the shape of semivariograms derived from NDVI images, is not affected by phenological changes, while also being capable of detecting land-cover changes.

3.2. Change Detection Evaluation

The comparative results in terms of overall, producer's and user's accuracies, along with the kappa index for each group of features and machine learning algorithms indicated that features are effective in differentiating no-change and land-cover changes, with overall accuracies greater than 80% (Table 2). We obtained the highest overall accuracy (92.27%) and kappa index (0.84) through the combination of semivariogram indices and spectral features using the support vector machine algorithm. The combination of semivariogram and spectral features reduced the overall confusion between classes and produced the best accuracy results with the MPL, SVM and RF algorithms, with accuracy greater than 90.82%, 92.27% and 91.30%, respectively.

The highest accuracies considering all the features groups and machine learning algorithms were obtained for the no-change user's accuracy. This means that this class presented a low inclusion error, reaching 95.33% using the combination of the features and SVM. This is very significant, because it means that objects with phenological changes are not being classified as land-cover changes.

Table 2. Change detection results using the group of features and machine learning algorithms.

ML	Class	SF		SI		SF + SI	
		PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)
	Change	88.24	75.00	87.21	75.00	93.55	87.00
	No-change	79.51	90.65	79.34	89.72	88.60	94.39
MPL	Global accuracy	83.09		82.6		90.82	
	Kappa index	0.65		0.65		0.81	
	Change	88.76	79.00	94.05	79.00	94.68	90.00
	No-change	82.20	90.65	82.93	95.33	90.27	95.33
SVM	Global accuracy	85.02		87.43		92.27	
	Kappa index	0.69		0.74		0.84	
	Change	84.54	82.00	87.37	83.00	91.84	90.00
	No-change	83.64	85.98	84.82	88.79	90.83	92.52
RF	Global accuracy	84.05		85.99		91.3	
	Kappa index	0.68		0.71		0.82	

* PA=producer accuracy; UA=user accuracy; SF=spectral features; SI=semivariogram indices; MPL=multilayer perceptron; SVM=support vector machine; RF=random forest.

Using spatial features, here represented by semivariogram indices, the change detection analysis was improved. These indices synthesized the most relevant information from semivariogram enabling us to extract the spatial information contained in the remote sensing objects, thus enhancing the NDVI spatial variability.

In summary, the semivariogram indices combined with the spectral

features provided improved change detection results, disentangling seasonal changes and land-cover changes, which are easily confused using only spectral features. The SVM algorithm was the most effective method to classify changes in areas of Savanna and Semi-arid woodland biomes under strong degradation processes.

4. Discussion

We propose a new method of object-based change detection to disentangle the effects of forest phenology and land-cover changes in Brazilian seasonal biomes combining spatial and spectral remote sensing features. We achieved the spatial information by extracting the spatial variability of NDVI provided by semivariogram indices.

We found that areas with land-cover changes such as deforestation, logging or fires provide a singular semivariogram, with higher values of sill parameter than the ones generated from natural physiognomies of Savanna and Semi-arid woodland. Similar results were found by Acerbi Junior et al., (2015) and Silveira et al., (2018) studying change detection in Brazilian Savanas using semivariograms derived from NDVI images. Their results showed a very clear trend where the shape of semivariograms, and the overall variability were different when deforestation occurred and were similar when the area had not been changed. Sertel et al., (2007) used semivariograms to identify earthquake damage and found that semivariance values of post-earthquake data were higher than semivariance values of pre-earthquake data if the area had been affected severely.

These differences of NDVI spatial variability between landscapes are mainly explained by the type of landscape. Objects that comprise more than one type of landscape are the most heterogeneous (Garrigues et al., 2006). Their spatial variability is explained by the differences of NDVI values, due the

mosaic of types of landscape with low NDVI and high NDVI increases the spatial variability. Natural vegetation and forest sites are homogeneous; the vegetation cover, with NDVI around 0.7 (Garrigues et al., 2006) which includes the green understory, the high density of trees, or the presence of broadleaves, homogenizes the distribution of NDVI values. In the presence of forest phenology effects, the NDVI values decrease, reaching about 0.3. In this condition of seasonality changes, the spatial distribution of NDVI values remains almost constant, indicating that the temporal spatial variability of NDVI values is not affected by vegetation seasonality. On the contrary, the spectral information of NDVI data are sensitivity to phenological effects, decreasing their values from around 0.7 to 0.3.

Thus, combining spatial and spectral remote sensing features increases the accuracy of change detection algorithms, mainly the support vector machines. As a group of theoretically superior machine learning algorithms, it has been frequently cited in image classification and achieved empirical successes (Foody and Mathur, 2004). They appear to be especially advantageous in the presence of heterogeneous classes for which only a few training samples are available (Huang et al., 2002; Wu et al., 2015).

How to accurately extract land-cover changes while disregarding the ones caused by phenological differences in Brazilian seasonal biomes undergoing rapid land-changes can be achieved by adding semivariogram indices as input data to train machine learning algorithms.

5. Conclusions

We have developed a method to separate the effects of forest phenology from land-cover changes using an object-based approach and Landsat images. Our method was based on traditional spectral NDVI data and spatial features provided by semivariogram as input data to train machine learning algorithms.

This study demonstrated that the temporal spatial variability of NDVI values is not affected by vegetation seasonality inducing the addition of semivariogram indices to improve the change detection results. The combination of spectral features and semivariogram indices provided the best results to disentangle the effects of forest phenology and land-cover changes in Brazilian seasonal biomes of Savana and Semi-arid woodland.

Moreover, our study underscores that the use of spatial features reduces the need of multi-temporal satellite images to accurately extract land-cover changes such as deforestation, logging or fire while disregarding the ones caused by seasonality changes.

6. References

Acerbi Junior, F.W., Silveira, E.M. de O., de Mello, J.M., de Mello, C.R., Scolforo, J.R.S., 2015. Change detection in Brazilian savannas using semivariograms derived from NDVI images. *Cienc. e Agrotecnologia* 39, 103–109. doi:10.1590/S1413-70542015000200001

Addink, E.A., Van Coillie, F.M.B., de Jong, S.M., 2012. Introduction to the GEOBIA 2010 special issue: From pixels to geographic objects in remote sensing image analysis. *Int. J. Appl. Earth Obs. Geoinf.* 15, 1–6. doi:10.1016/j.jag.2011.12.001

Akar, Ö., Güngör, O., 2015. Integrating multiple texture methods and NDVI to the Random Forest classification algorithm to detect tea and hazelnut plantation areas in northeast Turkey. *Int. J. Remote Sens.* 36, 442–464. doi:10.1080/01431161.2014.995276

Arantes, A.E., Ferreira, L.G., Coe, M.T., 2016. The seasonal carbon and water balances of the Cerrado environment of Brazil: Past, present, and future influences of land cover and land use. *ISPRS J. Photogramm. Remote Sens.* 117, 66–78. doi:10.1016/j.isprsjprs.2016.02.008

Asner, G. P. 2001. Cloud cover in Landsat observations of the Brazilian Amazon. *International Journal of Remote Sensing*. 22, 3855–3862. doi.org/10.1080/01431160010006926.

Atkinson, P.M., Lewis, P., 2000. Geostatistical classification for remote sensing: an introduction. *Comput. Geosci.* 26, 361–371. doi.org/10.1016/S0098-3004(99)00117-X

Baatz, M., Schäpe, A., 2000. Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. *J. Photogramm. Remote Sens.* 58, 12–23. doi:Export Date 6 May 2013

Balaguer-Beser, A., Ruiz, L.A., Hermosilla, T., Recio, J.A., 2013. Using semivariogram indices to analyse heterogeneity in spatial patterns in remotely sensed images. *Comput. Geosci.* 50, 115–127. doi:10.1016/j.cageo.2012.08.001

Balaguer, A., Ruiz, L.A., Hermosilla, T., Recio, J.A., 2010. Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification. *Comput. Geosci.* 36, 231–240. doi:10.1016/j.cageo.2009.05.003

Belgiu, M., Drăgu, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31. doi:10.1016/j.isprsjprs.2016.01.011

Benediktsson, J.A., Swain, P.H., Ersoy, O.K., 1990. Neural Network Approaches Versus Statistical Methods in Classification of Multisource Remote Sensing Data. 12th Can. Symp. Remote Sens. Geosci. Remote Sens. Symp. 28, 540–552. doi:10.1109/IGARSS.1989.578748

Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS J. Photogramm. Remote Sens.* 58, 239–258. doi:10.1016/j.isprsjprs.2003.10.002

Berberoğlu, S., Akin, a., Atkinson, P.M., Curran, P.J., 2010. Utilizing

image texture to detect land-cover change in Mediterranean coastal wetlands. *Int. J. Remote Sens.* 31, 2793–2815. doi:10.1080/01431160903111077

Beuchle, R., Grecchi, R.C., Shimabukuro, Y.E., Seliger, R., Eva, H.D., Sano, E., Achard, F., 2015. Land cover changes in the Brazilian Cerrado and Caatinga biomes from 1990 to 2010 based on a systematic remote sensing sampling approach. *Appl. Geogr.* 58, 116–127. doi:10.1016/j.apgeog.2015.01.017

Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* 65, 2–16. doi:10.1016/j.isprsjprs.2009.06.004

Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. doi:10.1023/A:1010933404324

Chen, J., Lu, M., Chen, X., Chen, J., Chen, L., 2013. A spectral gradient difference based approach for land cover change detection. *ISPRS J. Photogramm. Remote Sens.* 85, 1–12. doi:10.1016/j.isprsjprs.2013.07.009

Chen, X., Yang, D., Chen, J., Cao, X., 2015. An improved automated land cover updating approach by integrating with downscaled NDVI time series data. *Remote Sens. Lett.* 6, 29–38. doi:10.1080/2150704X.2014.998793

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* 37, 35–46. doi:10.1016/0034-4257(91)90048-B

Curran, P.J., 1988. The semivariogram in remote sensing: An introduction. *Remote Sens. Environ.* 24, 493–507. doi:10.1016/0034-4257(88)90021-1

Devries, B., Pratihast, A.K., Verbesselt, J., Kooistra, L., Herold, M., 2016. Characterizing forest change using community-based monitoring data and landsat time series. *PLoS One* 11, 1–25. doi:10.1371/journal.pone.0147121

DeVries, B., Verbesselt, J., Kooistra, L., Herold, M., 2015. Robust

monitoring of small-scale forest disturbances in a tropical montane forest using Landsat time series. *Remote Sens. Environ.* 161, 107–121. doi:10.1016/j.rse.2015.02.012

Duro, D.C., Franklin, S.E., Dub??, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sens. Environ.* 118, 259–272. doi:10.1016/j.rse.2011.11.020

Ferreira, L.G., Yoshioka, H., Huete, A., Sano, E.E., 2004. 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. *J. Arid Environ.* 56, 425–447. doi:10.1016/S0140-1963(03)00068-5

Foody, G.M., Mathur, A., 2004. A relative evaluation of multiclass image classification by support vector machines. *IEEE Trans. Geosci. Remote Sens.* 42, 1335–1343. doi:10.1109/TGRS.2004.827257

Garrigues, S., Allard, D., Baret, F., Weiss, M., 2006. Quantifying spatial heterogeneity at the landscape scale using variogram models. *Remote Sens. Environ.* 103, 81–96. doi:10.1016/j.rse.2006.03.013

Gil-Yepes, J.L., Ruiz, L.A., Recio, J.A., Balaguer-Beser, ??ngel, Hermosilla, T., 2016. Description and validation of a new set of object-based temporal geostatistical features for land-use/land-cover change detection. *ISPRS J. Photogramm. Remote Sens.* 121, 77–91. doi:10.1016/j.isprsjprs.2016.08.010

Gopal, S., Woodcock, C., 1996. Remote sensing of forest change using artificial neural networks. *IEEE Trans. Geosci. Remote Sens.* 34, 398–404. doi:10.1109/36.485117

Guan, H., Li, J., Chapman, M., Deng, F., Ji, Z., Yang, X., 2013. Integration of orthoimagery and lidar data for object-based urban thematic mapping using random forests. *Int. J. Remote Sens.* 34, 5166–5186.

doi:10.1080/01431161.2013.788261

Hamunyela, E., Verbesselt, J., Herold, M., 2016. Using spatial context to improve early detection of deforestation from Landsat time series. *Remote Sens. Environ.* 172, 126–138. doi:10.1016/j.rse.2015.11.006

Healey, S.P., Cohen, W.B., Zhiqiang, Y., Krankina, O.N., 2005. Comparison of Tasseled Cap-based Landsat data structures for use in forest disturbance detection 97, 301–310. doi:10.1016/j.rse.2005.05.009

Hoekstra, J.M., Boucher, T.M., Ricketts, T.H., Roberts, C., 2005. Confronting a biome crisis: Global disparities of habitat loss and protection. *Ecol. Lett.* 8, 23–29. doi:10.1111/j.1461-0248.2004.00686.x

Huang, C., Davis, L.S., Townshend, J.R.G., 2002. An assessment of support vector machines for land cover classification. *Int. J. Remote Sens.* 23, 725–749. doi:10.1080/01431160110040323

Hussain, M., Chen, D., Cheng, A., Wei, H., Stanley, D., 2013. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS J. Photogramm. Remote Sens.* 80, 91–106. doi:10.1016/j.isprsjprs.2013.03.006

Im, J., Jensen, J.R., 2005. A change detection model based on neighborhood correlation image analysis and decision tree classification 99, 326–340. doi:10.1016/j.rse.2005.09.008

Jin, S., Yang, L., Danielson, P., Homer, C., Fry, J., Xian, G., 2013. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. *Remote Sens. Environ.* 132, 159–175. doi:10.1016/j.rse.2013.01.012

Kulkarni, A.D., Lowe, B., 2016. Random Forest for Land Cover Classification. *Int. J. Recent Innov. Trends Comput. Commun.* 4, 58–63.

Lawrence, R.L., Wood, S.D., Sheley, R.L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications

(randomForest). *Remote Sens. Environ.* 100, 356–362. doi:10.1016/j.rse.2005.10.014

Leal, I.R., Da Silva, J.M.C., Tabarelli, M., Lacher, T.E., 2005. Changing the Course of Biodiversity Conservation in the Caatinga of Northeastern Brazil\Cambiando el Curso de la Conservación de Biodiversidad en la Caatinga del Noreste de Brasil. *Conserv. Biol.* 19, 701–706. doi:10.1111/j.1523-1739.2005.00703.x

Lu, M., Chen, J., Tang, H., Rao, Y., Yang, P., Wu, W., 2016. Land cover change detection by integrating object-based data blending model of Landsat and MODIS. *Remote Sens. Environ.* 184, 374–386. doi:10.1016/j.rse.2016.07.028

Mantero, P., Moser, G., Serpico, S.B., 2005. Partially supervised classification of remote sensing images through SVM-based probability density estimation. *IEEE Trans. Geosci. Remote Sens.* 43, 559–570. doi:10.1109/TGRS.2004.842022

Masek, J.G., Huang, C., Wolfe, R., Cohen, W., Hall, F., Kutler, J., Nelson, P., 2008. Remote Sensing of Environment North American forest disturbance mapped from a decadal Landsat record 112, 2914–2926. doi:10.1016/j.rse.2008.02.010

Millard, K., Richardson, M., 2015. On the importance of training data sample selection in Random Forest image classification: A case study in peatland ecosystem mapping. *Remote Sens.* 7, 8489–8515. doi:10.3390/rs70708489

Powers, R.P., Hermosilla, T., Coops, N.C., Chen, G., 2015. Remote sensing and object-based techniques for mapping fine-scale industrial disturbances. *Int. J. Appl. Earth Obs. Geoinf.* 34, 51–57. doi:10.1016/j.jag.2014.06.015

Roy, D.P., Borak, J.S., Devadiga, S., Wolfe, R.E., Zheng, M.,

Descloitres, J., 2002. The MODIS Land product quality assessment approach. *Remote Sens. Environ.* 83, 62–76. doi:10.1016/S0034-4257(02)00087-1

Ruiz, L.A., Recio, J.A., Fernández-Sarría, A., Hermosilla, T., 2011. A feature extraction software tool for agricultural object-based image analysis. *Comput. Electron. Agric.* 76, 284–296. doi:10.1016/j.compag.2011.02.007

Santos, J.J.C., Leal, I.R., Almeida-Cortez, J.S., Fernandes, G.W., Tabarelli, M., 2011. Caatinga: the scientific negligence experienced by a dry tropical forest. *Trop. Conserv. Sci.* 4, 276–286. doi:10.1177/194008291100400306

Santos, R.M., Oliveira-Filho, A.T., Eisenlohr, P. V., Queiroz, L.P., Cardoso, D.B.O.S., Rodal, M.J.N., 2012. Identity and relationships of the Arboreal Caatinga among other floristic units of seasonally dry tropical forests (SDTFs) of north-eastern and Central Brazil. *Ecol. Evol.* 2, 409–428. doi:10.1002/ece3.91

Sertel, E., Kaya, S., Curran, P.J., 2007. Use of semivariograms to identify earthquake damage in an Urban Area. *IEEE Trans. Geosci. Remote Sens.* 45, 1590–1594. doi:10.1109/TGRS.2007.894019

Silva, J.F., Fariñas, M.R., Felfili, J.M., Klink, C.A., 2006. Spatial heterogeneity, land use and conservation in the cerrado region of Brazil. *J. Biogeogr.* 33, 536–548. doi:10.1111/j.1365-2699.2005.01422.x

Silveira, E.M.O., Menezes, M.D., Acerbi Júnior, F.W., Terra, M.C.N.S, Mello, J.M., 2017. Assessment of geostatistical features for object-based image classification of contrasted landscape vegetation cover. *J. Appl. Remote Sens.* 11, 36004. doi:10.1117/1.JRS.11.036004

Silveira, E. M. O., Mello, J. M., Acerbi Júnior, F. W., Carvalho, L. M. T. 2018. Object-based land-cover change detection applied to Brazilian seasonal savannahs using geostatistical features. *Int. J. Remote Sens.* 39, 2597-2619. Doi.10.1080/01431161.2018.1430397

Tewkesbury, A.P., Comber, A.J., Tate, N.J., Lamb, A., Fisher, P.F., 2015. A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sens. Environ.* 160, 1–14. doi:10.1016/j.rse.2015.01.006

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8, 127–150. doi:10.1016/0034-4257(79)90013-0

Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D., 2010. Detecting trend and seasonal changes in satellite image time series. *Remote Sens. Environ.* 114, 106–115. doi:10.1016/j.rse.2009.08.014

Vorovencii, I., 2014. Assessment of some remote sensing techniques used to detect land use/land cover changes in South-East Transilvania, Romania. *Environ. Monit. Assess.* 186, 2685–2699. doi:10.1007/s10661-013-3571-y

Wu, X., Peng, J., Shan, J., Cui, W., 2015. Evaluation of semivariogram features for object-based image classification. *Geo-spatial Inf. Sci.* 18, 159–170. doi:10.1080/10095020.2015.1116206

Yue, A., Zhang, C., Yang, J., Su, W., Yun, W., Zhu, D., 2013. Texture extraction for object-oriented classification of high spatial resolution remotely sensed images using a semivariogram. *Int. J. Remote Sens.* 34, 3736–3759. doi:10.1080/01431161.2012.759298

Zhu, Z., Woodcock, C.E., Olofsson, P., 2012. Continuous monitoring of forest disturbance using all available Landsat imagery. *Remote Sens. Environ.* 122, 75–91. doi:10.1016/j.rse.2011.10.030

**ARTICLE 6 - USING SPATIAL FEATURES TO REDUCE IMPACT OF
SEASONALITY FOR DETECTING FOREST CHANGE FROM
LANDSAT TIME SERIES**

**Eduarda M.O. Silveira ^{1,*}, Inácio T. Bueno ¹, Fausto W. Acerbi-Junior ¹,
José M. Mello ¹, José R. S. Scolforo ¹ and Michael A. Wulder ²**

¹ Forest Science Department, Federal University of Lavras, Brazil;
dudalavras@hotmail.com (E.M.O.S.); inaciotbueno@gmail.com (I.T.B.);
fausto@dcf.ufla.br (F.W.A.J.); josemarcio@dcf.ufla.br (J.M.M.)

² Canadian Forest Service (Pacific Forestry Centre), Natural Resources
Canada, 506 West Burnside Road, Victoria, BC V8Z 1M5, Canada;
mike.wulder@canada.ca

* Correspondence: dudalavras@hotmail.com; Tel.: +55-35-3822-1700

Publication status: In prep. for submission to **Remote Sensing**

Abstract: In forest areas that experience strong seasonality and are undergoing rapid land-cover conversion (i.e., Brazilian savannas), the accuracy of remote sensing change detection is affected by seasonal changes that are erroneously classified as having changed. To improve the quality and consistency of regionally important forest change maps, we aim to separate process related to spectral variability due to phenology from changes related to deforestation or fire. A seasonal model is typically used to account for seasonality, but fitting a model is difficult when there are not enough data in the time series. In this research, we combine remote sensing and the spatial context at the object level to evaluate the performance of geostatistical features to reduce the impact of seasonality from NDVI (Normalized Difference Vegetation Index) Landsat time series, using both spectral and spatial information. The study area is the São Romão municipality, totalizing 2,440 km², part of a Brazilian savannas biome. We first create image objects by multiresolution segmentation, basing the objects on the characteristics found in the first image of the time series. We overlapped the objects with the remaining NDVI images in order to extract semivariogram indices, the RVF (Ratio Variance - First lag) and AFM (Area First lag - First Maximum) and spectral information (average and standard deviation of NDVI values) to generate time series from these features and derive spatio-temporal metrics (change and trend) to train a Random Forest (RF) algorithm. The NDVI spatial variability, captured by AFM semivariogram index time series produced the best result, reaching 96.53% of overall accuracy (OA) to separate no-change from forest change, while the greatest inter-class confusion occurred using the average of NDVI values time series (OA=63.72%). The spatial context approach we presented is a novel and useful object-based remote sensing method for the detection of forest change events in seasonal areas, eliminating the effects of forest phenology, without the need of use de-seasoning models.

Keywords: remote sensing; geostatistical; semivariogram; change detection; NDVI; forest phenology; savannas

1. Introduction

Savannas are the dominant biome in South Hemisphere, covering approximately 45% of the area of South America [1]. In Brazil, this biome consists of a mosaic of land cover types, undergoing a strong seasonality in climate, accompanied with a widespread occurrence of fires that imposes environmental pressures with the most rapid land conversion in Brazil, exceeding that of the tropical forests [2]. The high rates of deforestation and few formally protected areas, the savanna biome is among the most endangered ecoregions on Earth [3]. Accurate mapping and monitoring of areas undergoing conversion are needed to indicate the priority areas of conservation and to inform sustainable land-use management, as well as to improve the understanding of the dynamics and related impacts on carbon balance, nutrient cycling, and water resources [1]. At present, much of the effort estimating forest changes in Brazil have been focused on the tropical rain forests with far less attention dedicated to the less humid seasonal regions [4].

Remotely sensed data have been widely recognized as essential data sources for comprehensive mapping and quantification of land-cover changes [5-11]. Studies have used the NDVI (Normalized Difference Vegetation Index) time series acquired by moderate resolution sensors (i.e. MODIS) for land-cover change detection because of their frequent revisit time [12-14]. However, images from medium spatial resolution sources such as Landsat offer more detailed spatial information, providing insights at smaller scales over a long period of time. Initiated in 1972, Landsat is the longest running cross-calibrated globally consistent record of the Earth's surface at medium resolution [10]. The ability to utilize the dense time series of imagery from Landsat have been demonstrated for mapping land-cover changes, such as deforestation, forest degradation, and impacts of fire [15-19].

However, estimating forest change by remote sensing is not a trivial task. Satellite images can contain a combination of natural forest phenology, cloud cover, atmospheric scattering, and geometric errors [20], impairing the precision of change estimates, especially in areas that exhibit strong seasonality in photosynthetic activity [9]. When images from different seasons are acquired, changes caused by phenological differences are present, and can result in mapping of a change in condition, rather than a change in land cover or land use [21]. To reduce the seasonal variations captured by remotely sensed images, techniques for de-seasoning time series have been used [22] on the assumption that seasonal patterns can be identified and removed [12,14,21,23,24]. However, this assumption may not always hold if the time series is for satellite images which are not acquired at regular interval or have gaps due to presence of clouds [9]. To minimize the impacts of seasonality, methods combining different satellites [25,26] and others that explore the spatial context [9] have been developed.

Previous studies have shown that pixel-based change detection can suffer from not including additional information regarding spatial context [27]. Inclusion of spatial information can reduce the small or spurious changes [28] and also allow for the incorporation of supplemental spatial information, such as geostatistical features [29-31]. Although pixel-based approaches also generate spatial information, the neighbourhood used to retrieve the information is often defined by a square window, that are easy to implement, however, they are computational expensive [22], biased along their diagonals and can straddle the boundary between two landscape features, especially when a large window size is used [32].

Here, we propose a new approach to reduce the impacts of the seasonal changes introduced by forest phenology on forest change detection using the spatial variability of Landsat NDVI time series at the object level. We

hypothesize that semivariogram indices are not affected by vegetation seasonality as the NDVI variability captured by these indices remains constant in the presence of seasonal changes and substantially increases in the presence of land-cover changes. Thus, we investigated if the geostatistical features can be used to eliminate seasonal variations in satellite image time series and improve the change detection results. We used an object-based approach associated to a set of spatio-temporal metrics by computing geostatistical features, specifically the semivariogram indices developed by [29]. We also compared the change detection results to those provided by traditional spectral features (average and standard deviation of NDVI values). Implementation of this approach enables us to enrich object-based remote sensing methods, using the spatial information provided by semivariograms, that capture the spatial variability of images which are directly related to land-cover changes, improving the change detection classification performance in highly seasonal and ecologically important environments.

2. Study Area

The study area is the São Romão municipality, state of Minas Gerais (MG), Brazil, (Figure 1). Totalizing 2,440 km², approximately 84% of this area is covered by land cover classes of Brazilian savannas biome. The vegetation diversity of this biome is well documented (Table 1), ranging from grassland, shrublands to woodlands and palm swamps [33,34]. The climate is typically tropical (Aw) with rains concentrated in the October–May period, which characterizes the high seasonality in the region [35].

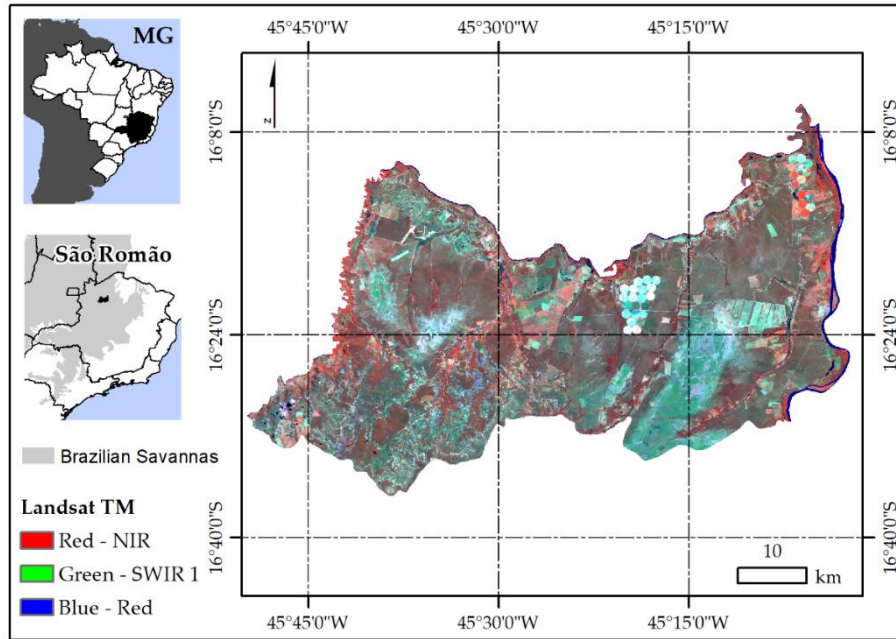

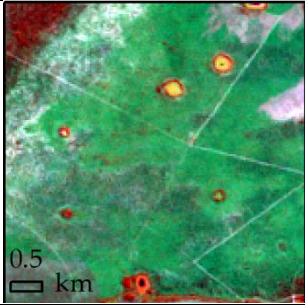

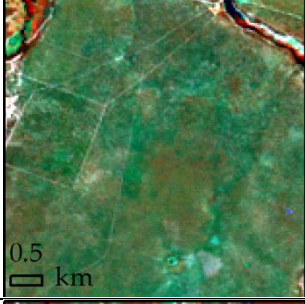

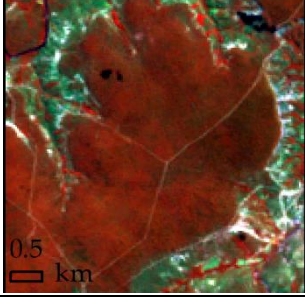


Figure 1. Location of MG state in Brazil, and São Romão inserted in Brazilian savannas biome. Landsat TM (Thematic Mapper) composite: Red-NIR (Near Infrared), Green-SWIR 1 (Shortwave Infrared) and Blue-Red.

Table 1. Examples of the savannas land-cover classes found in the study area. Landsat TM images Red-NIR, Green-SWIR 1 and Blue-Red composite.

Land-cover classes	Description	Field view	Landsat TM images
Grasslands	Open grassland		
Shrublands	Open grassland with sparse shrubs		
Woodlands	Mixed grassland, shrubland, and trees up to seven meters		



3. Methodology

We introduce a method for object-based change detection using geostatistical features to derive spatio-temporal metrics from time series, eliminating the effect of forest phenology. The method is divided into three main steps (Figure 2).

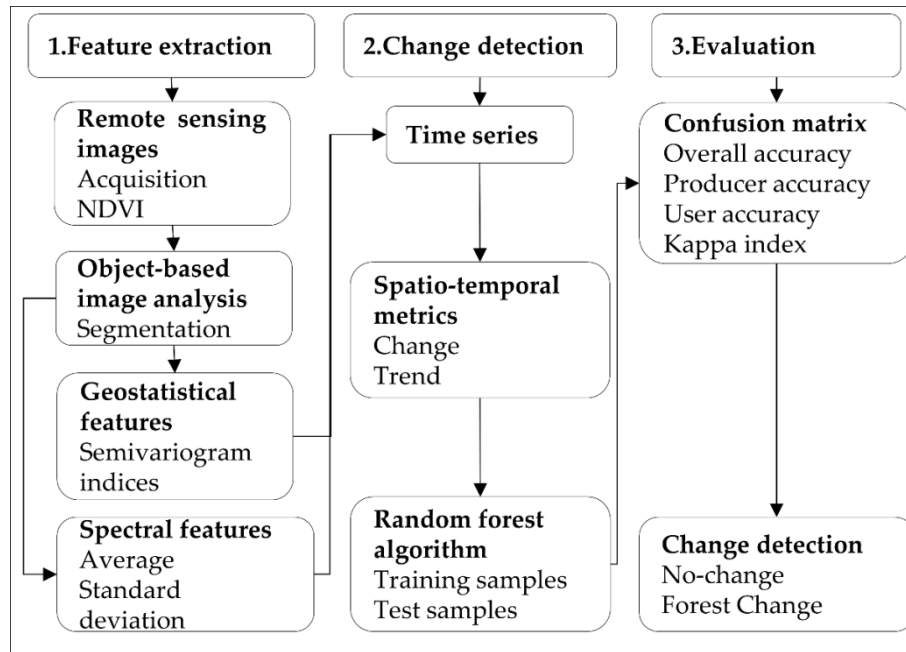


Figure 2. Methodology flowchart detailing the tree main steps to detect land-cover changes over the savanna biome: 1. Feature extraction from time series using NDVI trends at the object level; 2. Change detection using random forest to model change type as a function of spatio-temporal metrics and; 3. Evaluation of change detection using accuracy measures.

We choose as change target the Brazilian savannas land-cover classes, and the change agents (Figure 3):

- No-change: seasonal changes due phenological effects (natural events)
- Forest change: land-cover changes such as deforestation (human induced) and fires (both natural events and human induces)

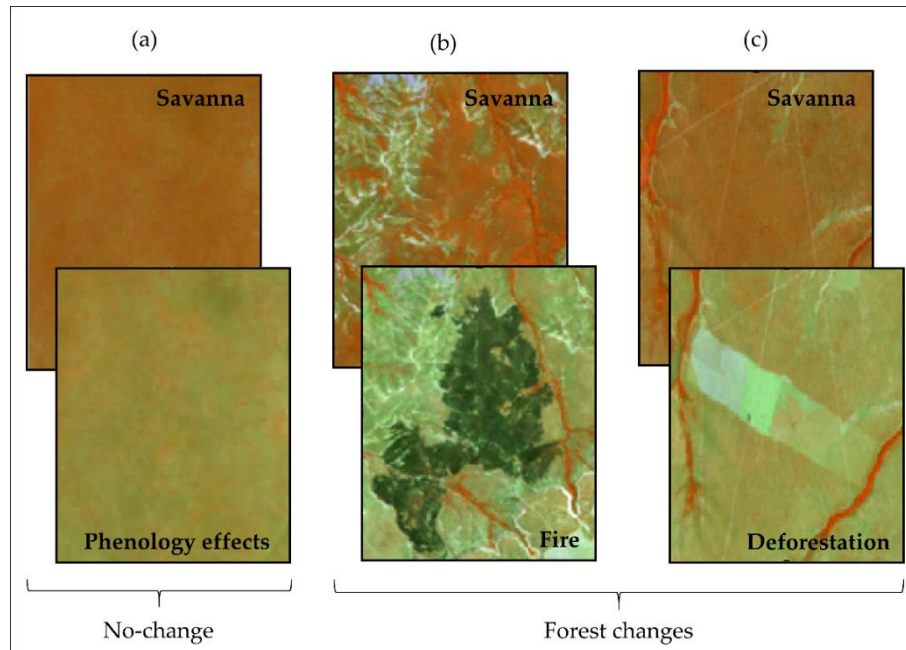


Figure 3. Landsat TM Red-NIR, Green-SWIR 1 and Blue-Red composite. Examples of no-change class: (a) savannas affected by phenological effects; and forest change classes: (b) savannas undergoing a change by fire and (c) savannas undergoing a change by deforestation.

3.1. Satellite Data and Processing

We downloaded all available cloud free Landsat TM and OLI images between 2003 to 2016 (total of 66 images) from the free and open access USGS Earth Resources Observation and Science (EROS) Centre archive. The data as downloaded is already processed to surface reflectance using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) and L8SR algorithm, respectively. Clouds and cloud shadows have been masked with the CFmask function [36].

We used the NDVI vegetation index [37] for further analysis. The advantages of using indices rather than the original spectral bands observations include: minimizing soil and other background effects, reducing data dimensionality, providing a degree of standardization for comparison, and enhancing the vegetation signal [38].

3.2. Object-based Image Analysis

Image segmentation is the division of the satellite image into group of pixels spatially continuous and spectrally homogeneous, minimizing the within-object variability compared to the between-object variability [39]. We used the multiresolution segmentation algorithm [40] implemented in the eCognition software, that produces highly homogeneous image objects based on spatial and spectral homogeneity criteria. Three different choices are typically considered when defining the analysis unit (image objects) for multitemporal object-based image analysis [41]: (i) image-object overlay: image-objects are generated by segmenting one of the images in the time series. A comparison against other images is then made by simple overlay; (ii) image-object comparison: image-objects are generated by segmenting each image in the time series independently; and (iii) multitemporal image-object: image-objects are generated by segmenting the entire time series together. Our study focused on the image-object overlay approach, as expected to capture the intra-object heterogeneity well.

Three key segmentation parameters (shape, compactness, and scale) control the size, shape, and spectral variation of segmented image objects [42]. We set the shape and compactness parameter to 0.1 and 0.5, respectively. The scale parameter controls the size of the image objects and sets a homogeneity threshold that determines the number of neighbouring pixels that can be merged together to form an image object [43]. The scale of segmentation directly

influences the size of the objects connected to the semivariogram predefined criteria (lag distance) and the minimum number of pixels inside each object necessary to generate the semivariogram, which makes this segmentation parameter crucial at this point. Based upon starting values found in the literature [39], we then used a “trial-and-error” approach [44] to find the appropriate scale parameter [45] that was set to 150 [31]. The image segmentation outputs were assessed once the objects produced to confirm a visual correspondence to meaningful objects (from a forest ecosystem perspective). The segmentation parameters were also defined to generate sufficiently small objects to not be spatially or spectrally confused, yet large enough to allow for calculation of the semivariogram and related geostatistical features.

3.3. Geostatistical and Spectral Features

We used as geostatistical features, two semivariogram indices (Table 2) developed by [29]. Inside each object, we extracted the NDVI values to generate the experimental semivariogram and calculate the semivariogram indices. The experimental semivariogram is defined as half of the average squared difference between values separated by a given lag, where this lag is a vector in both, distance and directions [46]. It was estimated using Equation 1:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x) - Z(x+h)]^2, \quad (1)$$

Where $\gamma(h)$ is the estimator of the semivariance for each distance h , $N(h)$ is the number of pairs of points separated by the distance h , $Z(x)$ is the value of the regionalized variable in the point x and $Z(x+h)$ is the value of the point $(x+h)$.

The graphic for spatial variance versus distance h represents the semivariogram, which allows obtaining the estimative of the variance value for different combinations of pairs of points. The semivariance functions are characterized by three parameters: sill (σ^2), range (ϕ) and nugget effect (τ^2). The

sill parameter is the plateau reached by semivariance values and shows the quantity of variation explained by the spatial structure of the data. The range parameter is the distance where the semivariogram reaches the sill, showing the distance until where the data are correlated. The nugget effect is the combination of sampling errors and variations that happen in scales smaller than the distance between the sampled points [47].

When the semivariogram is calculated at object level, an important factor to be considered is the lag distance. It should not be larger than the spatial extent of the object; on the other hand, an exceedingly small distance fails to provide a complete description of textural features. We attempted to find an optimal lag distance to ensure that sill values would provide a concise description of data variability. We fixed the number of lags as 20 pixels and the lag size equivalent to image spatial resolution (30 m), resulting in a lag distance of 600 m [31,48].

Using FETEX 2.0 [49] we calculated the semivariogram indices, that describe the shape of the experimental semivariograms, and therefore the properties that characterize the spatial patterns of the image object. In addition, fitting a model is unnecessary, which shortens processing time and reduces the errors generated by choosing a sub-optimal model.

Table 2. Geostatistical features: semivariogram indices described in [29].

Indices	Description	Formula
RVF	Ratio variance – first lag	$RVF = \frac{\gamma_{\max_1}}{\gamma_1}$
AFM	Area first lag-first maximum	$AFM = \frac{h}{2} \left(\gamma_1 + 2 \left(\sum_{i=2}^{\max_1-1} \gamma_i \right) + \gamma_{\max_1} \right) - (\gamma_1 (h_{\max_1} - h_1))$

RVF is the ratio between the values of the total variance γ_{\max_1} and the semivariance at first lag (γ_1). Since the semivariogram tends to reach the sill near the total variance, this parameter is an indicator of the relationship between the spatial correlation at long and short distances. Its value increases when high variability at long distances and low variability at short distances occurs. This feature also presents high values for images with large primitives or periodic patterns [29]. AFM is the area between the semivariogram value in the first lag and the semivariogram. This parameter provides information about the semivariogram curvature and is also related to the variability of the data (Figure 4).

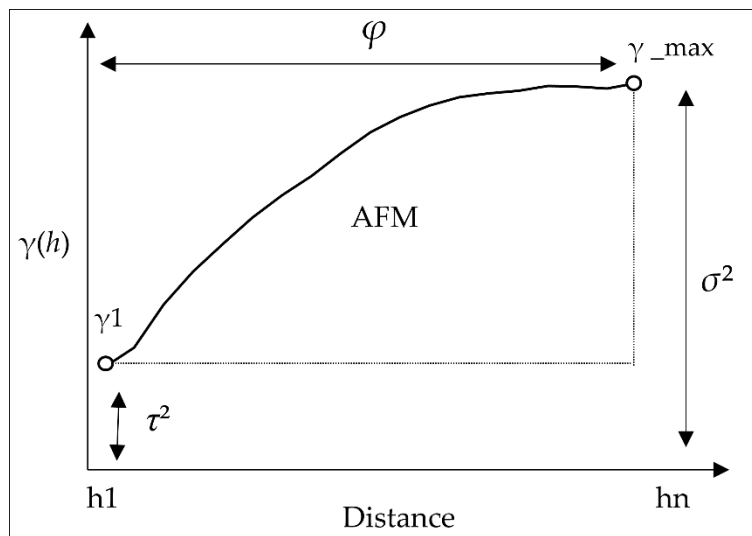


Figure 4. Graphic representation of AFM index in the classical semivariogram: (σ^2) – sill, (ϕ) – range, (τ^2) – nugget effect, $\gamma(h)$ - semivariance, γ_{\max_1} - total variance and γ_1 - semivariance at first lag.

To verify the potential of the semivariogram indices, we also used spectral features for comparative purpose. We extracted the average and standard deviation of NDVI values inside the objects.

3.4. Time Series

To generate the time series, we first masked the objects cover by savanna's land cover classes in the first image and eliminated classes as water, pasture, bare soil, crops and urban areas. We overlapped the objects with the remaining NDVI images in order to extract the geostatistical (RVF and AFM indices) and spectral features (average and standard deviation), and then generate the time series (Figure 5).

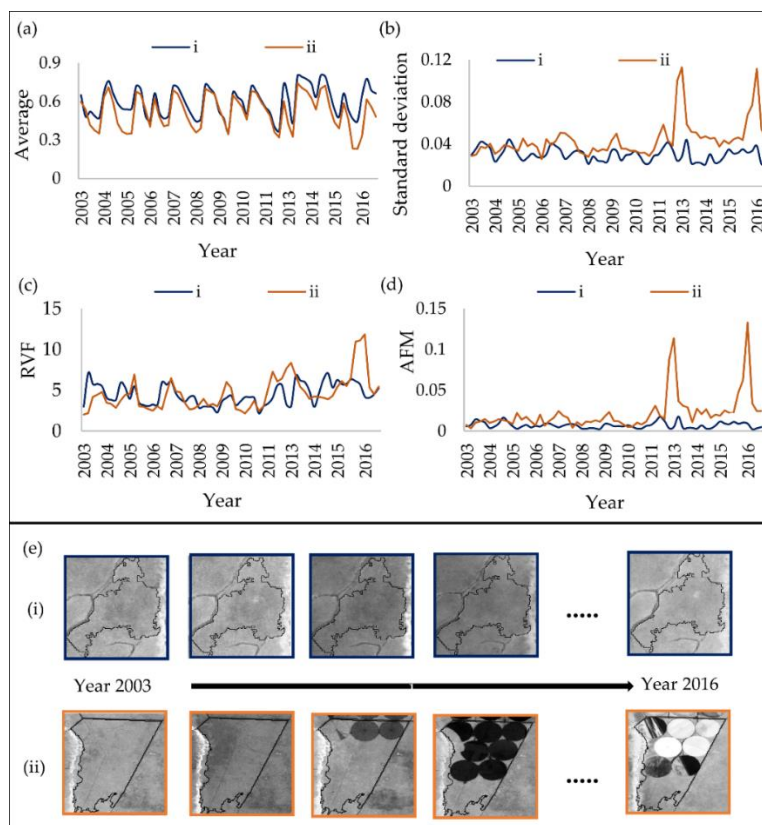


Figure 5. Example of Landsat time series from 2003 to 2016 of (a) average NDVI values; (b) standard deviation NDVI values; (c) RVF index and (d) AFM index, (e) comprising objects containing (i) seasonal changes and (ii) land-cover changes.

3.5. Spatio-Temporal Metrics

We transformed the geostatistical and spectral features time series into two summary variables: Change and Trend, adapted from [14]. As described by [41], change is the maximum interannual absolute difference in the time series data within the observation period (Equation 2). Trend is the slope on linear regression applied to the full time series (Equation 3).

$$\text{Change} = \text{Max} (\text{Feature}_t - \text{Feature}_{t+1}), \quad (2)$$

$$\text{Trend} = \frac{n(\sum_{t=1}^{t=n} t * \text{Feature}_t) - (\sum_{t=1}^{t=n} \text{Feature}_t) * (\sum_{t=1}^{t=n} t)}{n(\sum_{t=1}^{t=n} t^2) - (\sum_{t=1}^{t=n} t)^2}, \quad (3)$$

The Change variable contains information on the greatest positive or negative change between consecutive data. A high value indicates that a high magnitude change occurred or there is an influence of seasonal variation of the feature response. The Trend variable indicates the overall (increasing, decreasing, or none) trend in the amount of green vegetation across the entire period [14].

3.6. Random Forest Algorithm

We used the spatio-temporal metrics derived from geostatistical and spectral features to train the random forest (RF) algorithm [50]. RF is an ensemble method which generates a set of individually trained decision trees and combines their results. RF is a robust non-parametric classifier and can accommodate many predictor variables [17]. The advantages of RF include effectiveness in prediction, efficient implementation on large datasets [50], and insensitivity to noise in the training data [51].

Two parameters need to be set in order to produce the random forest trees: the number of decision trees to be generated (Ntree) and the number of variables

to be selected and tested for the best split when growing the trees (Mtry) [52]. The Gini criterion is used to select the split with the lowest impurity (highest homogeneity) at each node. For each tree, the predicted class for each observation is obtained and the class with the maximum number of votes among the trees is the predicted class of an observation [53].

The Ntree required to maintain a given level of accuracy has been assessed by several authors, and the minimum number of trees for optimal classification appears to be between 100 [54] and 300 [55], with the majority of studies setting the Ntree value to 500 because the errors stabilize before this number of classification trees is achieved [54]. Five hundred trees were grown for each classification and the number of features (Mtry) was left at its default values.

3.6. Change Detection Evaluation

Evaluation of change detection outcomes is communicated using a confusion matrix [56] and its accuracy measures: overall accuracy (OA), producer accuracy (PA), user accuracy (UA) and the kappa coefficient (k). A data set of 1,000 objects well-distributed over the study area, with 500 samples per class (Table 3), was assigned based on visual interpretation. Seventy per cent of the samples were randomly chosen as training samples, while the rest were used as validation samples. We used the class stratified random sampling design, that provides the option to increase the sample size in classes that occupy a small proportion of area and is one of the easier designs to implement [57].

Table 3. Classification scheme.

Classes	Description	Training samples	Validation samples
No-change	Areas comprising the same cover in both epochs, albeit with changes caused by vegetation seasonality	350	150
Forest change	Land-cover areas comprising deforestation and fires (savannas that were converted to bare soil and pastures)	350	150

The evaluated performances were compared among the four input features (average, standard deviation, RVF and AFM) through the confusion matrix measures. We also generated a validation map to identify the savannas land-cover changes (deforestation and fires) by visual interpretation to spatialize the areas classified as correct and incorrect. We used as reference data, the map provided by [58], which described qualitative and quantitative information of savannas vegetation type in the state of Minas Gerais, Brazil.

4. Results

4.1. Change Detection

We used as input data to train the random forest (RF) algorithm the Change and Trend spatio-temporal metrics derived from geostatistical and spectral features time series. The classification results (Figure 6) achieved overall accuracies from 63.72% to 96.53%, with user's accuracies ranging from 61.53 to 98.23% and producer's accuracies 58.94 to 97.67% (Table 4). The AFM index provided the best results, reaching 96.53% of overall accuracy and 0.92 of kappa index. The standard deviation also produced good results, with a kappa index reaching 0.76. The greatest inter-class confusion occurred using the average of NDVI values and RVF index, revealing low values of overall, producer and user accuracies and kappa index.

Table 4. Classification results using groups of features tested. PA: Producer's accuracy in %; UA: User's accuracy in %.

Classes	Average		Standard deviation		RVF		AFM	
	PA	UA	PA	UA	PA	UA	PA	UA
No-change	58.94	61.53	88.57	86.91	69.79	74.44	97.67	94.38
Forest change	67.88	65.48	88.13	89.66	78.50	74.33	95.68	98.23
Overall accuracy	63.72		88.34		74.38		96.53	
Kappa coefficient	0.26		0.76		0.48		0.92	

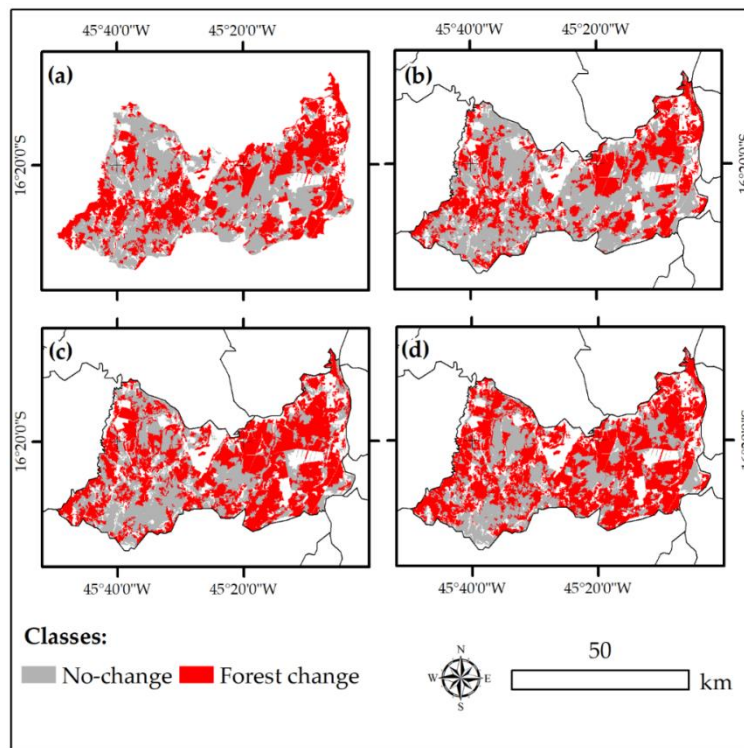


Figure 6. Change maps generated by the features: (a) average; (b) Standard deviation; (c) RVF index and (d) AFM index.

As a result, the feature average and the semivariogram index RVF, were not able to accurately capture the differences between forest change and seasonal

changes. These features are sensitive to phenological variations, resulting in false detection of LULC changes, overestimating the areas undergoing a fire or deforestation (Figure 7a, and Figure 7c). The standard deviation spectral feature presented good accuracy measures (overall accuracy reaching 88.34%), however, the map generated was not the more accurate (Figure 7b). AFM semivariogram index provided the best result. This index is directly influenced by the sill (σ^2) semivariogram parameter (overall variability), which is sensitive to land-cover changes and is not affected by changes caused by vegetation seasonality, providing good classification results (Figure 7d).

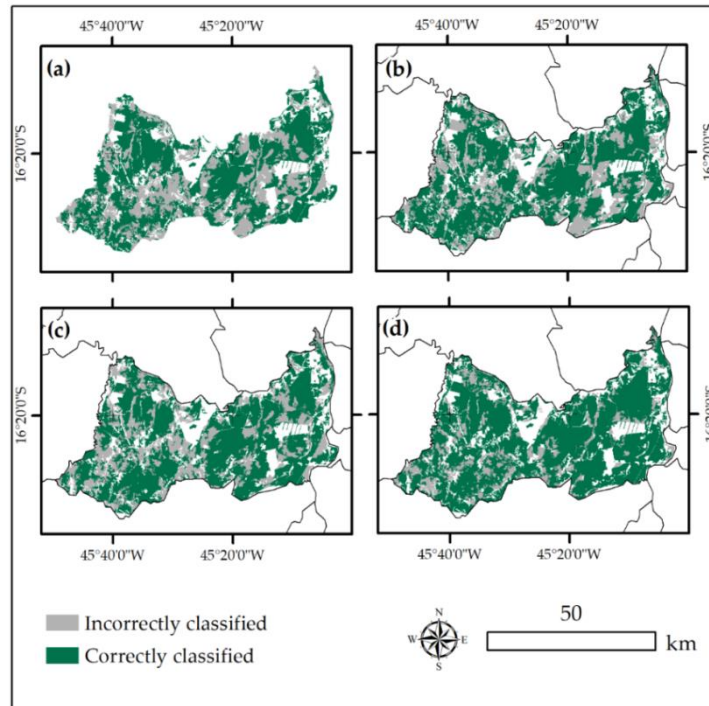


Figure 7. Comparative analysis among the validation map and the estimates using (a) average spectral feature; (b) standard deviation spectral feature; (c) RVF semivariogram index and (d) AFM semivariogram index.

5. Discussion

5.1. NDVI Spatial Variability

To capture the NDVI variability inside the objects we used the semivariogram, that is a graphical representation of the spatial variability in a given dataset [59], provided by the sill (σ^2), range (ϕ) and nugget effect (τ^2) parameters. [60] showed the ability of the semivariogram to depict landscape spatial heterogeneity. The differences of NDVI spatial variability between landscapes are mainly explained by the type of landscape, where the image spatial variability (sill- σ^2), increases considerably from homogeneous to heterogeneous areas.

[61] investigated the relationship between semivariogram metrics and degree of earthquake-induced changes. In severely changed areas, the earthquake caused large spatial variations that were quantified by the semivariogram variables of range, nugget, and sill. The range captured coarse-scale spatial variability, the nugget captured fine-scale spatial variability, and the sill captured overall variability in the landscape. The results indicated that the dominant earthquake-induced changes were textural rather than tonal and so lend themselves to quantification using semivariograms.

[62] demonstrated the usefulness of semivariogram shape and metrics, generated from NDVI values, derived from Landsat TM images, to detect deforestation in an area covered by savanna vegetation. In deforested areas, the landscape change caused spatial variations that were quantified by the semivariogram metrics of range, sill and shape. The range and sill were the two most important and complementary metrics. Both metrics increased their values after deforestation and remain similar if the land-cover had not been changed.

Here, instead of using the semivariogram parameters to quantify the NDVI spatial variability we used the RVF and AFM semivariogram indices, as they

synthesize the most relevant information about the shape of the semivariogram. In addition, fitting a model is unnecessary, which shortens processing time and reduces the errors generated by choosing a sub-optimal model [29]. From the wet to dry season the NDVI spatial variability is almost constant and in the presence of land-cover changes, such as deforestation and fires, the NDVI variability highly increase (Figure 8). Preserved savannas land-cover classes (both in wet and dry seasons) have low NDVI variability compared to disturbed savannas land-cover classes. This high is explained by the presence of high and low NDVI values in the same object, as land cover changes exceed seasonal variability.

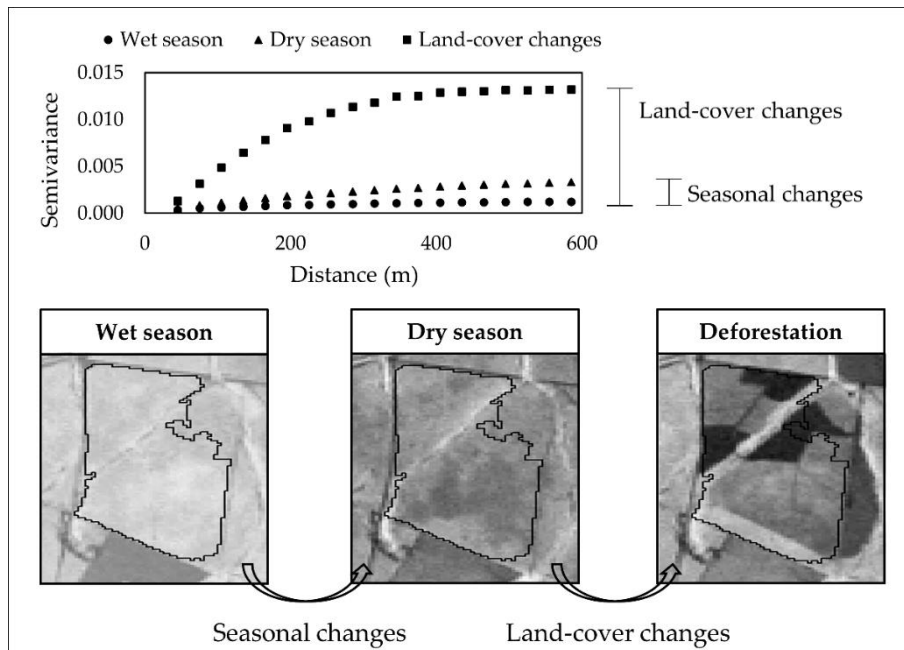


Figure 8. Example of NDVI spatial variability of objects under a seasonal change and deforestation land-cover change.

The total variance of NDVI values increases considerably from homogeneous to heterogeneous landscapes. In homogeneous objects as preserved savannas the NDVI semivariance among the pixels is low, thus the RVF index is also low because the y_{max} (~ 0.0013) is almost similar with the semivariance in the first lag (~ 0.0005). The AFM is also small (0.011) (Figure 9a). In heterogeneous objects, such as savannas undergoing a deforestation or fires, the RVF and AFM indices considerably increase, ranging from 3.33 to 8.38 and 0.011 to 0.13, respectively (Figure 9b).

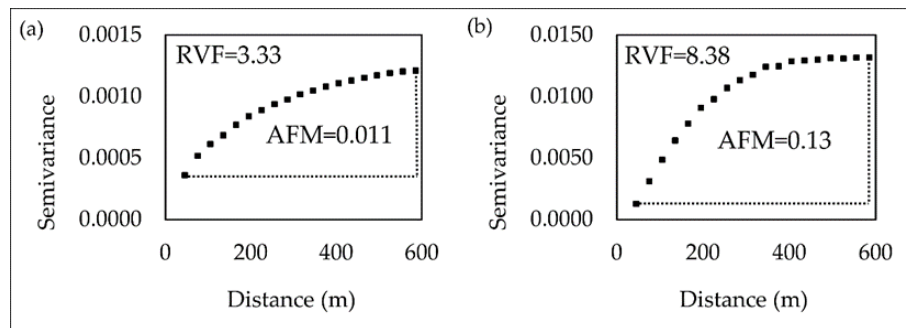


Figure 9. Example of NDVI spatial variability of objects under (a) a seasonal change and (b) deforestation land-cover change.

[31] explored and evaluated the performance of semivariogram indices derived from NDVI images in an object-based approach to detect land-cover changes. The image's spatial variability changed considerably from native vegetation (pre-disaster image) to flooded areas (post-disaster image). The flooded areas had a low overall variability compared to native vegetation, making possible the detection of change using the semivariogram.

5.2. Seasonal Effects on Change Detection

We hypothesized that the semivariogram indices (RVF and AFM) are not affected by vegetation seasonality, due to the NDVI variability remaining

constant in the presence of seasonal changes and substantially increases in the presence of land-cover changes, such as deforestation and fires (both natural events and human induced).

Remote sensing of vegetation has an implied link to vegetation phenology. Phenology is an environmental process that operates independent of land cover or land use change. The ability to use remote sensing images to detect change in seasonal areas relies on the capacity to remove the effects of seasonal variations, while still identifying changes caused by deforestation, urbanization, floods, and fires. To avoid the classification of seasonal effects as land-cover change, time series have been used to derive the seasonal patterns when mapping deforestation. For instance, [12] developed a powerful change detection approach for time series, using 16-day MODIS NDVI images, involving the detection and characterization of Breaks For Additive Seasonal and Trend (BFAST), that integrates the iterative decomposition of time series into trend, seasonal and noise components with methods for detecting changes, without the need to select a reference period, set a threshold, or define a change trajectory.

Rather than using the BFAST seasonal model to account for seasonality in image time series, [9] used the spatial context in a pixel-based approach using Landsat data assessing how the size of spatial window influences the deforestation detection and related computational time. They demonstrated that, in dry tropical forest, deforestation events are detected much earlier when the spatial context approach is used than when a seasonal model is used. Smaller spatial windows achieved higher overall spatial accuracy than larger windows and increasing the size of the spatial window resulted in a linear increase in the computational time per pixel.

The window size is an important aspect for the spatial context method, since it influences the resultant spatial information due to the amount of variance included [63,64]. The spatial moving window needs to be sufficiently larger than

the size of the deforestation event to avoid smoothing out the disturbance impact from the time series, and also for it to capture local variations [9].

Here, we used the spatial context (semivariogram indices) in an object-based approach, without the need of use of the spatial window, basing the objects on the characteristics found in the first image of the time series. The time series of AFM semivariogram index, provided the best separability (global accuracy = 96.53%) between objects with seasonal changes and objects undergoing a forest change. Similar results were found by [65] assessing the potential of individual geostatistical features derived from bitemporal NDVI images to accurately detect land-cover changes. They demonstrated that the accuracies of the sill parameter and AFM semivariogram indices were higher than those of the other indices, which were not affected by vegetation seasonality because they represent the information provided by the entire semivariogram rather than specific parts of it.

6. Conclusions

We have developed a method to mitigate the presence of phenological effects from time series using NDVI trends to detect changes over a savannah forest ecosystem. The approach combining spatial and spectral information allowed for the detection of changes without being impacted by seasonal variations. We demonstrated that NDVI spatial variability, captured by AFM semivariogram index is not affected by vegetation seasonality, and therefore, can produce times series that accurately differentiate forest changes from seasonal changes, resulting in fewer classification errors. Our proposed method is simple and accurate, efficiently detecting changes while eliminating the effects of phenology.

The spatial context approach we presented in this paper is a novel and useful object-based remote sensing method for the detection of forest change

events using NDVI Landsat data in areas where forests exhibit strong seasonality (i.e., in Brazilian seasonal biomes), addressing the challenge of accurately detecting deforestation and fires, eliminating the effects of forest phenology, without the need of use de-seasoning models.

References

1. Sano, E.E; Rosa, R.; Brito, J.L.S; Ferreira, L.G. Land cover mapping of the tropical savanna region in Brazil. *Env. Monit. Assess.* **2010**, *166*, 113-124, DOI: 10.1007/s10661-009-0988-4.
2. Ferreira, L.G.; Yoshioka, H.; Huete, A.; Sano, E.E. Seasonal landscape and spectral vegetation index dynamics in the Brazilian Cerrado: An analysis within the Large-Scale Biosphere-Atmosphere Experiment in Amazônia (LBA). *Remote Sens. Environ.* **2003**, *87*, 534–50. DOI: 10.1016/j.rse.2002.09.003.
3. Hoekstra, J.M.; Boucher, T.M.; Ricketts, T.H.; Roberts, C. Confronting a biome crisis: Global disparities of habitat loss and protection. *Ecol. Lett.* **2005**, *8*, 23–29, DOI: 10.1111/j.1461-0248.2004.00686.x.
4. Beuchle, R.; Grecchi, R.C.; Shimabukuro, Y.E.; Seliger, R.; Eva, H.D.; Sano, E.; Achard, F. Land cover changes in the Brazilian Cerrado and Caatinga biomes from 1990 to 2010 based on a systematic remote sensing sampling approach. *Appl. Geogr.* **2015**, *58*, 116–27. DOI: 10.1016/j.apgeog.2015.01.017.
5. Yuan, F.; Sawaya, K.E.; Loeffelholz, B.C.; Bauer, M.E. Land cover classification and change analysis of the Twin Cities (Minnesota) metropolitan area by multitemporal Landsat remote sensing. *Remote Sens. Environ.* **2005**, *98*, 317–28. DOI: 10.1016/j.rse.2005.08.006.
6. De Sy, V.; Herold, M.; Achard, F. Asner, G.P.; Held, A.; Kellndorfer J.; Verbesselt, J. Synergies of multiple remote sensing data sources for

- REDD+ monitoring. *Current Opinion in Environmental Sustainability*. **2012**, *4*, 696–706. DOI: 10.1016/j.cosust.2012.09.013.
7. Grecchi RC, Gwyn, Q.H.J.; Bénié, G.B.; Formaggio, A.R.; Fahl, F.C. Land use and land cover changes in the Brazilian Cerrado: A multidisciplinary approach to assess the impacts of agricultural expansion. *Appl. Geogr.* **2014**, *55*: 300–312. DOI: 10.1016/j.apgeog.2014.09.014.
 8. DeVries, B.; Verbesselt, J.; Kooistra, L.; Herold, M. Robust monitoring of small-scale forest disturbances in a tropical montane forest using Landsat time series. *Remote Sens. Environ.* **2015**, *161*, 107–121. DOI: 10.1016/j.rse.2015.02.012.
 9. Hamunyela, E.; Verbesselt, J.; Herold, M. Using spatial context to improve early detection of deforestation from Landsat time series. *Remote Sens. Environ.* **2016**, *172*, 126–138. DOI: 10.1016/j.rse.2015.11.006.
 10. Schultz M.; Clevers, J.G.P.W.; Carter, S.; Verbesselt, J.; Avitabile, V.; Quang, H.V.; Martin, H. Performance of vegetation indices from Landsat time series in deforestation monitoring. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *52*, 318–27. DOI: 10.1016/j.jag.2016.06.020.
 11. Hamunyela, E.; Reiche, J.; Verbesselt, J.; Herold, M. Using space-time features to improve detection of forest disturbances from Landsat time series. *Remote Sens.* **2017**, *9*, 1–17. DOI: 10.3390/rs9060515.
 12. Verbesselt, J.; Hyndman, R.; Newnham, G.; Culvenor, D. Detecting trend and seasonal changes in satellite image time series. *Remote Sens. Environ.* **2010**, *114*, 106–15. DOI: 10.1016/j.rse.2009.08.014.
 13. Zhao, P.; Lu, D.; Wang, G.; Wu, C.; Huang, Y.; Yu, S. Examining spectral reflectance saturation in landsat imagery and corresponding solutions to improve forest aboveground biomass estimation. *Remote Sens.* **2016**, *8*, 1–26. DOI: 10.3390/rs8060469.

14. Hird, J.N.; Castilla, G.; McDermid, G.J.; Bueno, I.T. A Simple Transformation for Visualizing Non-seasonal Landscape Change from Dense Time Series of Satellite Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3372–3383. DOI: 10.1109/JSTARS.2015.2419594.
15. Wulder, M. A.; Masek, J.G.; Cohen, W.B.; Loveland, T.R.; Woodcock, C.E. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sens. Environ.* **2012**, *122*, 2–10. DOI: 10.1016/j.rse.2012.01.010.
16. Zhu, Z.; Woodcock, C.E.; Olofsson, P. Continuous monitoring of forest disturbance using all available Landsat imagery. *Remote Sens. Environ.* **2012**, *122*, 75–91. DOI: 10.1016/j.rse.2011.10.030.
17. DeVries, B.; Pratihast, A.K.; Verbesselt, J.; Kooistra, L.; Herold, M. Characterizing forest change using community-based monitoring data and landsat time series. *PLoS One.* **2016**, *11*, 1–25. DOI: 10.1371/journal.pone.0147121.
18. Bartels, S.F.; Chen, H.Y.H.; Wulder, M.A.; White, J.C. Trends in post-disturbance recovery rates of Canada's forests following wildfire and harvest. *For. Ecol. Manage.* **2016**, *361*, 194–207. DOI: 10.1016/j.foreco.2015.11.015.
19. White, J.C.; Wulder, M.A.; Hermosilla, T.; Coops, N.C.; Hobart, G.W. A nationwide annual characterization of 25 years of forest disturbance and recovery for Canada using Landsat time series. *Remote Sens. Environ.* **2017**, *194*, 303–21. DOI: 10.1016/j.rse.2017.03.035.
20. Roy, D.P.; Borak, J.S.; Devadiga, S.; Wolfe, R.E.; Zheng, M.; Descloitres, J. The MODIS Land product quality assessment approach. *Remote Sens. Environ.* **2002**, *83*, 62–76. DOI: 10.1016/S0034-4257(02)00087-1.
21. Lu, M.; Chen, J.; Tang, H.; Rao, Y.; Yang, P.; Wu, W. Land cover change detection by integrating object-based data blending model of Landsat and

- MODIS. *Remote Sens Environ.* **2016**, *184*, 374–86. DOI: 10.1016/j.rse.2016.07.028.
22. Zhu Z. ISPRS Journal of Photogrammetry and Remote Sensing Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. *ISPRS J. Photogramm Remote Sens.* **2017**; *130*, 370–384. DI: 10.1016/j.isprsjprs.2017.06.013.
 23. Verbesselt, J.; Hyndman, R.; Zeileis, A.; Culvenor, D. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. *Remote Sens. Environ.* **2010**, *114*, 2970–2980. DOI: doi.org/10.1016/j.rse.2010.08.003.
 24. Verbesselt, J.; Zeileis, A.; Herold, M. Near real-time disturbance detection using satellite image time series. *Remote Sens. Environ.* **2012**, *123*, 98–108. DOI: 10.1016/j.rse.2012.02.022.
 25. Dutrieux, L. P.; Verbesselt, J.; Kooistra, L., Herold, M. Monitoring forest cover loss using multiple data streams, a case study of a tropical dry forest in Bolivia. *ISPRS J. Photogramm. Remote Sens.* **2015**, *107*, 112–125. DOI: 10.1016/S0034-4257(02)00087-1.
 26. Zhang, J. Multi-source remote sensing data fusion: Status and trends. *International J. of Image and Data Fusion.* **2010**, *1*, 5–24. DOI: 10.1080/19479830903561035.
 27. Chen, G.; Hay, G.J.; Carvalho, L.M.T.; Wulder, M.A. Object-based change detection. *International Journal of Remote Sensing.* **2012**, *33*, 4434–4457. DOI: 10.1080/01431161.2011.648285,
 28. Hussain, M.; Chen, D.; Cheng, A.; Wei, H.; Stanley, D. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS J. Photogramm Remote Sens.* **2013**, *80*, 91–106. DOI: 10.1016/j.isprsjprs.2013.03.006.

29. Balaguer, A.; Ruiz, L.A.; Hermosilla, T.; Recio, J.A. Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification. *Comput. Geosci.* **2010**, *36*, 231–40. DOI: 10.1016/j.cageo.2009.05.003.
30. Gil-Yepes, J.L.; Ruiz, L.A.; Recio, J.A.; Balaguer-Beser, A.; Hermosilla, T. Description and validation of a new set of object-based temporal geostatistical features for land-use/land-cover change detection. *ISPRS J. Photogramm. Remote Sens.* **2016**, *121*, 77–91. DOI: 10.1016/j.isprsjprs.2016.08.010.
31. Silveira, E.M.O.; Acerbi Júnior, F.W.; Mello, J.M.; Bueno, I.T. Object-based change detection using semivariogram indices derived from NDVI images: The environmental disaster in Mariana, Brazil. *Cienc e Agrotecnologia.* **2017**, *41*, 554–564. DOI: 10.1590/1413-70542017415009817.
32. Laliberte, A.S.; Rango, A. Texture and Scale in Object-Based Analysis of Subdecimeter Resolution Unmanned Aerial Vehicle (UAV) Imagery. *Geosci. Remote Sensing, IEEE Trans.* **2009**, *47*, 761–70. DOI: 10.1109/TGRS.2008.2009355.
33. 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. *J. Arid Environ.* **2004**, *56*, 425–47. DOI: 10.1016/S0140-1963(03)00068-5.
34. Arantes, A.E.; Ferreira, L.G.; Coe, M.T. The seasonal carbon and water balances of the Cerrado environment of Brazil: Past, present, and future influences of land cover and land use. *ISPRS J Photogramm. Remote Sen.s.* **2016**, *117*, 66–78. DOI: 10.1016/j.isprsjprs.2016.02.008.

35. Peel, M.C.; Finlayson, B.L., McMahon, T. A. Updated world map of the Koppen-Geiger climate classification. *Meteorol Zeitschrift*. **2006**, *15*, 259–263. DOI: 10.1127/0941-2948/2006/0130.
36. Zhu, Z.; Woodcock, C.E. Remote Sensing of Environment Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* **2012**, *118*, 83–94. DOI: 10.1016/j.rse.2011.10.028.
37. Rouse, J.W.; Smith, M.O.; Adams, J.B. Monitoring vegetation systems in the great plains with ERTS. *Third ERTS Symposium*. **1973**, *351*, 39-317.
38. Goward, S.N. Satellite bioclimatology. *Journal of Climate*. **1989**, *2*, 710–20. DOI: 10.1175/1520-0442(1989)002<0710:SB>2.0.CO;2.
39. Desclée, B.; Bogaert, P.; Defourny, P. Forest change detection by statistical object-based method. *Remote Sens Environ.* **2006**, *102*, 1–11. DOI: 10.1016/j.rse.2006.01.013.
40. Baatz, M. Schäpe, A. Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. *J. Photogramm. Remote Sens.* **2000**, *58*, 12–23.
41. Tewkesbury, A.P.; Comber, A.J.; Tate, N.J.; Lamb, A.; Fisher, P.F. A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sens. Environ.* **2015**, *160*, 1–14. DOI: 10.1016/j.rse.2015.01.006.
42. Mui, A.; He, Y.; Weng, Q. An object-based approach to delineate wetlands across landscapes of varied disturbance with high spatial resolution satellite imagery. *ISPRS J. Photogramm. Remote Sens.* **2015**, *109*, 30–46. DOI: 10.1016/j.isprsjprs.2015.08.005.
43. Benz, U.C.; Hofmann, P.; Willhauck, G.; Lingenfelder, I.; Heynen, M. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS J. Photogramm. Remote Sens.* **2004**, *58*, 239–258. DOI: 10.1016/j.isprsjprs.2003.10.002.

44. Duro, D.C.; Franklin, S.E.; Dubé, M.G. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sens. Environ.* **2012**, *118*, 259–72. DOI: 10.1016/j.rse.2011.11.020.
45. Yang, J.; He, Y.; Weng, Q. An Automated Method to Parameterize Segmentation Scale by Enhancing Intra-segment Homogeneity and Inter-segment Heterogeneity. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1282–1286. DOI: 10.1109/LGRS.2015.2393255.
46. Atkinson, P.M.; Lewis, P. Geostatistical classification for remote sensing: an introduction. *Comput. Geosci.* **2000**, *26*, 361–71. Available from: DOI: 10.1016/S0098-3004(99)00117-X.
47. Curran, P.J. The semivariogram in remote sensing: An introduction. *Remote Sens. Environ.* **1988**, *24*, 493–507. DOI: 10.1016/0034-4257(88)90021-1.
48. Silveira, E.M.O.; Menezes, M.D.; Acerbi Júnior, F.W.; Castro, N. S. T. M.; Mello, J.M. Assessment of geostatistical features for object-based image classification of contrasted landscape vegetation cover. *J. Appl. Remote Sens.* **2017**, *11*, 1–15. DOI: 10.1117/1.JRS.11.036004.
49. Ruiz, L.A.; Recio, J.A.; Fernández-Sarría, A.; Hermosilla, T. A feature extraction software tool for agricultural object-based image analysis. *Comput. Electron. Agric.* **2011**; *76*, 284–96. DOI: 10.1016/j.compag.2011.02.007.
50. Breiman L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. DOI: 10.1023/A:1010933404324.
51. Baccini, A.; Laporte, N.; Goetz, S.J.; Sun, M.; Dong, H. A first map of tropical Africa's above-ground biomass derived from satellite imagery. *Environ. Res. Lett.* **2008**, *3*, 1–9. DOI: 10.1088/1748-9326/3/4/045011.

52. Belgiu, M.; Drăgu, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. DOI: 10.1016/j.isprsjprs.2016.01.011.
53. Archer, K.J.; Kimes, R. V. Empirical characterization of random forest variable importance measures. *Comput. Stat. Data Anal.* **2008**, *52*, 2249–2260. DOI: 10.1016/j.csda.2007.08.015.
54. Lawrence, R.L.; Wood, S.D.; Sheley, R.L. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). *Remote Sens. Environ.* **2006**, *100*, 356–362. DOI: 10.1016/j.rse.2005.10.014.
55. Akar, Ö.; Güngör, O. Integrating multiple texture methods and NDVI to the Random Forest classification algorithm to detect tea and hazelnut plantation areas in northeast Turkey. *Int. J. Remote Sens.* **2015**, *36*, 442–64. DOI: 10.1080/01431161.2014.995276.
56. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens Environ.* **1991**, *37*, 35–46. DOI: 10.1201/9781420055139.
57. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S. V.; Woodcock, C.E.; Wulder, M. A. Good Practices for Assessing Accuracy and Estimating Area of Land Change Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2017**, *148*, 42–57. DOI: [10.1016/j.rse.2014.02.015](https://doi.org/10.1016/j.rse.2014.02.015).
58. Carvalho, L. M. T.; Scolforo, J. R. S. *Inventário Florestal de Minas Gerais: Monitoramento da Flora Nativa 2005–2007*. Lavras: UFLA, 2008.
59. Cohen, W.B.; Spies, T.A.; Bradshaw, G.A.; Semivariograms of digital imagery for analysis of conifer canopy structure. *Remote Sens. Environ.* **1990**, *34*, 167–78. DOI: 10.1016/0034-4257(90)90066-U.

60. Garrigues, S.; Allard, D.; Baret, F.; Weiss, M. Quantifying spatial heterogeneity at the landscape scale using variogram models. *Remote Sens Environ.* **2006**, *103*, 81–96. DOI: 10.1016/j.rse.2006.03.013.
61. Sertel, E.; Kaya, S.; Curran, P.J. Use of semivariograms to identify earthquake damage in an Urban Area. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1590–1594. DOI: 10.1109/TGRS.2007.894019.
62. Acerbi Junior, F.W.; Silveira, E.M.O.; Mello, J.M.; Mello, C.R.; Scolforo, J.R.S. Change detection in Brazilian savannas using semivariograms derived from NDVI images. *Cienc. e Agrotecnologia.* **2015**, *39*, 103–9. DOI: 10.1590/S1413-70542015000200001.
63. Wulder, M.A.; LeDrew, E.F.; Franklin, S.E.; Lavigne, M.B. Aerial image texture information in the estimation of northern deciduous and mixed wood forest leaf area index (LAI). *Remote Sens. Environ.* **1998**, *64*, 64–76. DOI: 10.1016/S0034-4257(97)00169-7.
64. Chen, Q.; Gong, P. Automatic variogram parameter extraction for textural classification of the panchromatic IKONOS imagery. *IEEE Trans Geosci. Remote Sens.* **2004**, *42*, 1106–15. DOI: 10.1109/TGRS.2004.825591.
65. Silveira, E. M. O.; Mello, J. M.; Acerbi Júnior, F. W.; Carvalho, L. M. T. Object-based land-cover change detection applied to Brazilian seasonal savannas using geostatistical features. *Int. J. Remote Sens.* **2018**, *39*, 2597-2619. DOI: 10.1080/01431161.2018.1430397.

**ARTICLE 7 - STRATIFICATION IMPROVEMENTS IN
ABOVEGROUND BIOMASS MODELLING OF SAVANNA-FOREST
TRANSITION DRIVEN BY SATELLITE IMAGES AND SPATIO-
ENVIRONMENTAL DATA**

Eduarda Martiniano de Oliveira Silveira^{1*}, Fernando Delbon Espírito Santo²,
José Márcio de Mello¹, Carlos Rogério de Mello³, Fausto Weimar Acerbi
Júnior¹, Mônica Canaan Carvalho¹, José Roberto Soares Scolforo¹, Jos Barlow²

¹ Forest Science Department (DCF), Federal University of Lavras (UFLA),
Lavras, Brazil; dudalavras@hotmail.com; josemarcio@dcf.ufla.br;
fausto@dcf.ufla.br; monicacanaan@gmail.com; jscolforo@dcf.ufla.br

² Lancaster Environment Center (LEC), Lancaster University, UK, LA1 4YQ;
f.espirito-santo@lancaster.ac.uk; jos.barlow@lancaster.ac.uk

³ Soil and Water Engineering Department, Federal University of Lavras
(UFLA), Lavras, Brazil; crmello@deg.ufla.br

*Corresponding author: dudalavras@hotmail.com

Publication status: In prep. for submission to **Carbon Balance and
Management**

Background: Improved maps of aboveground biomass (AGB) of large and heterogeneous areas are required, mainly in threatened Brazilian biomes such as Savannas, Atlantic Forest and Semi-arid woodland. The quantification of biomass and carbon in these biomes are roughly estimated due to the uncertainty of global maps that are not focused on the stratification of inventory plots into forest types. Given this gap, we investigated the potential of data extracted from remote sensing (Landsat TM and MODIS products) and spatio-environmental variables for mapping the spatial distribution AGB of six heterogeneous vegetation types in a Savanna-Forest transition in Minas Gerais state, southeast Brazil. We selected the best variables to predict the AGB and compared the performance of random forest (RF) regression using non-stratified and stratified models based on vegetation types. We used regression-kriging technique to correct the maps developed by RF regression.

Results: The results suggest that the vegetation types stratification decreases the root mean square error (RMSE) as well as the mean absolute error (MAE), mainly in grassland cerrado (RMSE reduction of 53.48% and MAE reduction of 44.62%), woodland cerrado (RMSE reduction of 43.71% and MAE reduction of 34.73%) and cerrado *sensu stricto* (RMSE reduction of 20.68% and MAE reduction of 17.46%). Residuals analysis of non-stratified model indicated slight trends that were eliminated using the stratified models. The variables selected are associated with each vegetation type through their spatial distribution and vegetation phenology. The number of variables is driven by the spatial distribution and it is a characteristic of the seasonality effects. The total area of AGB in Minas Gerais state accounts 839,375,640 tones, with mean values ranging from 13.32 t/ha (grassland cerrado) to 124.03 t/ha (wetland forests).

Conclusions: The stratification into vegetation types not only improved the accuracy of AGB estimative, but also allowed RF regression to select the

smallest number of variables that offer the best predictive model performance. The improvement in AGB estimates is given by spatial distribution and seasonality effects of each vegetation type, which is achieved by stratifying the models to minimize the Savanna-Forest transition heterogeneity. The refining map and the understanding of how the variables characteristics are associated with the AGB enable researches to improve the roughly estimates of greenhouse gas emission and also helps on the selection of appropriate variables that best modelling the aboveground biomass in Savanna-Forest transition areas.

Keywords: Remote Sensing; Random Forest; Regression Kriging; Savanna; Atlantic Forest, Semi-arid woodland; Biomass.

Background

The high levels of terrestrial carbon stocks found in Brazil's forest biomes associated with the land-cover changes such as deforestation and fires, make Brazil one of the five biggest carbon dioxide emitting nations globally (Baccini et al. 2012; Matthews et al. 2014). Brazil has a total area of about 8,514,877 km², from which 7% occurs in Minas Gerais State, southeast region (586,528 km²). This large area encompasses landscape variations ranging from Savanna, Atlantic Forest, and Semi-arid woodland, representing 57%, 41%, and 2% of the native vegetation, respectively (Scolforo et al. 2015). Atlantic Forest and Savanna are the Brazilian biomes that have most suffered anthropogenic impacts (land-cover changes and deforestation are responsible for 54% of the total greenhouse gas (GHG) emissions and have been classified among the most threatened biomes of the world (Myers et al. 2000). The high rates of deforestation caused GHG, roughly estimated as there are only a very limited number of studies that deal with the quantification of biomass and carbon in these biomes with a comprehensive manner (Ribeiro et al. 2011). Refining these estimates requires improved knowledge of the density and spatial distribution of forest biomass (Baccini et al. 2008), disaggregating large scale estimates (Wilson, Woodall, and Griffith 2013).

There are numerous approaches for estimating aboveground biomass (AGB) from remotely sensed data in large areas (Lu 2006). In general, these data are empirically linked to AGB measurements of field plots (Baccini et al. 2004; Baccini et al. 2008; Zheng et al. 2004; Yin et al. 2015; Zhu and Liu 2015; Rodríguez-veiga et al. 2016; Aslan et al. 2016; Gizachew et al. 2016; Hu et al. 2016), ranging from simple linear regression to complex machine learning algorithms (MLA) (Powell et al. 2010; Gleason and Im 2012; Zhou et al. 2016) and regression kriging technique (Viana et al. 2012; Su et al. 2016; Castillo-Santiago et al. 2013; Galeana-Pizaña et al. 2014; Scolforo et al. 2016).

Landsat images are the most medium-resolution data commonly used due to the longest data record along with a spatial resolution of 30 meters (Powell et al. 2010; Lu et al. 2012; Main-Knorn et al. 2013; Zhu and Liu 2015). Seasonal time-series data at coarse resolution (e.g. MODIS) have been explored for estimating AGB at large scales (Yin et al. 2015; Arantes, Ferreira, and Coe 2016; Rodríguez-veiga et al. 2016), in order to improve the accuracy of the monitoring forest attributes (Dymond, Mladenoff, and Radeloff 2002).

Identifying suitable variables for developing a model is often critical (Lu 2006), mainly in large and high heterogeneity areas comprising vegetation types ranging from Savannas to Forests, that have their own phenology, local climate and geographic location that makes AGB mapping more difficult in comparison with homogeneous regions (Asner et al. 2012). Recent studies suggest that the stratification of inventory plots into forest types can improve the precision of AGB estimates (Heurich and Thoma 2008; Ribeiro et al. 2011; Dahlke et al. 2013; Latifi et al. 2015; Fayad et al. 2016) making easier the appropriate selection of variables that best explain the data variability behaviour (Zhao et al. 2016).

Based on this gap, the questions that motivate this study were: (i) how does the stratification into vegetation types improves the predictive quality of AGB random forest model? (ii) How do the remote-sensing and spatio-environmental variables are associated with the vegetation types?

To respond these questions, we investigated the potential of data extracted from Landsat TM, MODIS products and spatio-environmental variables to map the spatial distribution of aboveground biomass (ABG) of six heterogeneous vegetation types in the Atlantic Forest, Savanna, and Semi-arid woodland Biomes in Minas Gerais State, Brazil. The approach leverages a combination of extensive field data that provide accurate information at the plot level, and variables that are continuous in space. The random forest (RF)

regression algorithm was adopted due to its capability to select and rank important variables for AGB prediction. We used regression-kriging technique, by kriging the residuals generated by the models, to improve the map produced by RF. Specifically, the objectives of this study were:

- (i) to compare the performance of random forest regression model for AGB using non-stratified and stratified models based on vegetation types; and
- (ii) to select and drive the best variables to predict the AGB of each vegetation type.

Methods

We linked the aboveground biomass field inventory data with the following variables: spatio-environmental, comprising geographic, topographic and climate data; Landsat TM spectral bands and vegetation indices and monthly temporal series of MODIS products. As image processing, we resampled and mosaic the images to accurately extract the data values as inputs to training random forest regression algorithm. We classified the plots into six vegetation types in Savanna, Atlantic forest and Semi-arid woodland biomes to perform the stratified random forest regression model. We identified the smallest number of variables that offer the best predictive model and evaluated the models performance: non-stratified AGB map versus stratified AGB map. In the map that provided the best result, we applied regression kriging technique to produce the improved AGB map (Figure 1).

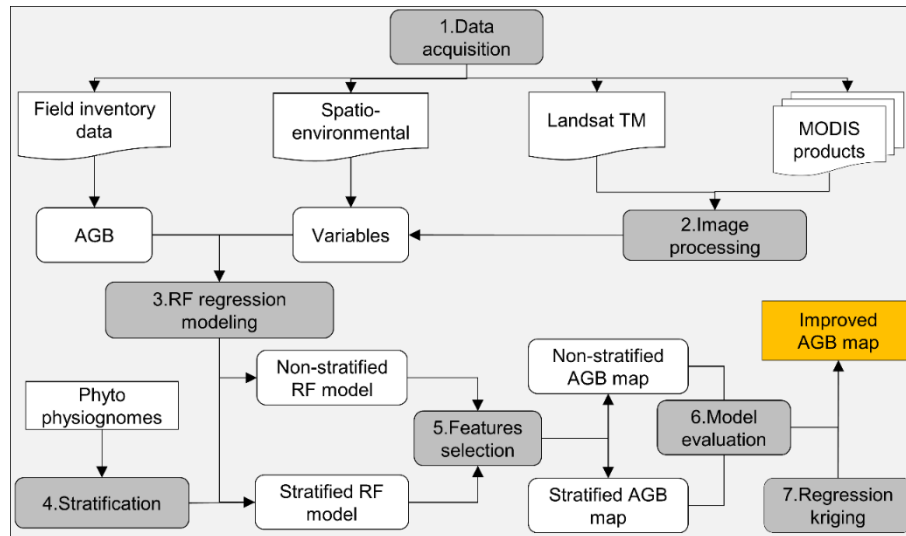


Figure 1. The workflow of the aboveground biomass (AGB) modeling approach.

Study area and field inventory data

Minas Gerais (MG) state is located in the southeast Brazil (Figure 2a), encompassing the Savanna, Atlantic Forest, and Semi-arid woodland biomes (Figure 2b). The Brazilian Savanna is a heterogeneous biome (Miranda et al. 1997), comprising vegetation types of grassland cerrado (shrub type of savanna, encompassing both herbaceous vegetation and scattered small trees), cerrado sensu stricto (savanna formation with trees and shrubs up to 8–10m high and with a grass understory) and woodland cerrado (forest formation with trees up to a height of 20 m) (Ribeiro et al. 2011). Atlantic forest occurs in the south-central and east of MG state and is composed of semideciduous and rain forests (Ombrophylous Forest) vegetation types. Semi-arid woodland is an ecosystem occupied by tropical dry forest and represents the vegetation type of deciduous forest. They are the unique biome for being in areas with the highest temperatures and the lowest amount of rain, being characterized as the region

with the greatest meteorological limitations of the MG state (Scolforo et al. 2015).

The climate variability of MG state indicates a negative precipitation and a positive temperature gradient from south to north (Figure 2c and Figure 2d). This variability helps to explain the predominance of these biomes. The elevation ranges from 30 to 2,824 meters and the greatest altitude variation is found in the eastern region (Figure 2e).

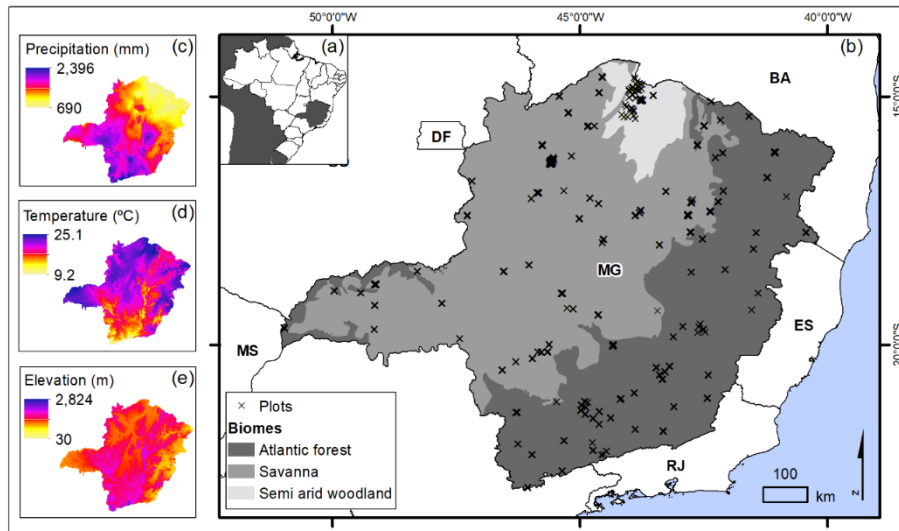


Figure 2. (a) Geographic location of MG in Brazil; (b) Biomes and plots spatially distributed in MG State; (c) Average annual precipitation in MG state; (d) Average annual temperature in MG state; and (e) Elevation of MG state.

A total of 3,284 field plots were located within the study area (see Figure 2a). The square sample plots (10 x 100m) were established in 2006 during the Project “Forest Inventory of Minas Gerais”, conducted by the Federal University of Lavras, MG, Brazil (Scolforo et al. 2008). The data used in the analysis was the aboveground biomass containing in the plots. The methodology

to determine the AGB is described in Scolforo et al. (2008). There are plots in different degrees of anthropization, different sites, different successional stages, and trees with different diameters and heights, leading to a high variability. The descriptive statistics for each vegetation types (Table 1) highlight the structural variability among them. High biomass and standard deviation was observed in semideciduous, wetland and deciduous forest. The lowest values in grasslands occur because these vegetation type is characterized by herbaceous vegetation with scattered small trees.

Table 1. Descriptive statistics of the measured above ground biomass (t/ha).

Vegetation types	Min	Mean	Max	Standard Deviation
Grassland cerrado	1.08	13.54	49.67	10.32
Woodland cerrado	10.81	56.30	177.97	29.98
Cerrado sensu stricto	10.02	34.57	170.23	22.38
Deciduous forest	10.61	74.78	295.45	55.05
Semideciduous forest	10.34	102.52	398.16	62.31
Rain forest	10.07	133.92	345.25	69.38

Spatio-environmental data

We used as spatio-environmental data, 19 climatic variables (1 km² of spatial resolution) acquired from WorldClim (Hijmans et al. 2005), a digital elevation model (30 meters of spatial resolution) developed from the “Shuttle Radar Topography Mission” (SRTM) and latitude and longitude geographical coordinates (Table 2).

Table 2. Spatio-environmental variables used to model AGB.

Variables	Description
Bio 1	Annual Mean Temperature (°C)
Bio 2	Mean Diurnal Range (Mean of monthly (°C)
Bio 3	Isothermality (BIO2/BIO7) (* 100) (°C)
Bio 4	Temperature Seasonality (standard deviation *100) (°C)
Bio 5	Max Temperature of Warmest Month (°C)
Bio 6	Min Temperature of Coldest Month (°C)
Bio 7	Temperature Annual Range (BIO5-BIO6) (°C)
Bio 8	Mean Temperature of Wettest Quarter (°C)
Bio 9	Mean Temperature of Driest Quarter (°C)
Bio 10	Mean Temperature of Warmest Quarter (°C)
Bio 11	Mean Temperature of Coldest Quarter (°C)
Bio 12	Annual Precipitation (mm)
Bio 13	Precipitation of Wettest Month (mm)
Bio 14	Precipitation of Driest Month (mm)
Bio 15	Precipitation Seasonality (Coefficient of Variation) (mm)
Bio 16	Precipitation of Wettest Quarter (mm)
Bio 17	Precipitation of Driest Quarter (mm)
Bio 18	Precipitation of Warmest Quarter (mm)
Bio 19	Precipitation of Coldest Quarter (mm)
DEM	Elevation (m)
Latitude	Latitude

Longitude

Longitude

Remote sensing data

We used Landsat TM (Thematic Mapper) and six MODIS (Moderate Resolution Imaging Spectroradiometer) products (Table 3) as remote sensing data. We acquired 35 scenes of Landsat TM to recover the study area, downloaded from the United States Geological Survey for Earth Observation and Science (USGS\EROS), with the necessary geometric corrections and reflectance values at the ground level (Young et al. 2017). For the year of 2006 (the same as inventory data collection), we selected one image date in which the scene was entirely without clouds.

The MODIS sensors assembled on the Terra and Aqua satellites have got a total of 36 spectral bands, seven of which are designed specifically for land applications with spatial resolutions that range from 250 m to 1,000 m and temporal resolution of one day (Mitchard et al. 2014). We used the Terra MODIS products computed from atmospherically corrected, bidirectional surface reflectance masked for clouds, heavy aerosols, and cloud shadows. We selected one scene per month (during the year 2006) to explore the temporal resolution of these products. Four MODIS tiles, namely, h13v10, h13v11, h14v10, and h14v11, were required to cover Minas Gerais state as a whole.

The images were mosaicked and resampled to Albers Equal Area Conic projection due to area and shape preserving characteristics of this projection, since to use a standard Universal Transverse Mercator (UTM) projection would have spanned multiple zones, introducing potential projection related errors in the final map output (Duro, Franklin, and Dube 2012).

Table 3. Remote sensing variables used to model AGB.

Remote sensing data	Variables	Spatial resolution (m)
Landsat TM	TM B1 (blue) TM B2 (green) TM B3 (red) TM B4 (NIR) TM B5 (SWIR 1) TM B7 (SWIR 2) NDVI EVI SAVI	30
MOD9Q1 MODIS Surface Reflectance	Surface Reflectance Band 1 Surface Reflectance Band 2	250
MOD13Q1 vegetation index (VI)	Red reflectance Band 1 NIR reflectance Band 2 Blue reflectance Band 3 MIR reflectance Band 7 NDVI EVI	250
MOD44B Vegetation Continuous Cover/Fields	Percent Tree Cover	250
MOD17A2H Gross primary productivity	GPP (Gross Primary Production) PSN (Net Photosynthesis)	500
MOD15A2H LAI/FPAR	LAI (Leaf area index) FPAR (Fraction of photosynthetically active radiation absorbed by vegetation)	500
M11A2 Land Surface Temperature and Emissivity	LSTd (Dailyday time Land-surface Temperature) LSTn (Daily nighttime Land-surface Temperature) Band 31 emissivity Band 32 emissivity	1000

Random forest regression algorithm

In random forest (Breiman 2001), a large number of trees are grown with the root node containing a different bootstrap sample of the data with the same number of cases as the original data. At each node, splitting is performed using a randomly selected subset of the predictor variables. Random forest is less sensitive to noise in the training data and tends to result in more accurate models (Baccini et al. 2008). We used random forest package (Liaw & Wiener 2002) available in the R software (R Development Core Team, 2014).

About 70% of the data were randomly selected and used as training data to fit the model. The remaining 30% was used to test the model's performance. We computed the coefficient of determination (R^2 , in %) which indicates the part of the observed variability that is explained by the model; the root-mean-square error (RMSE, in t/ha) which measures the average difference between values predicted by the model and observations; and, the mean absolute error (MAE, in %) which indicates an average over or underestimation bias by the model (Vieilledent et al. 2016).

To optimize the model, we examined the effect of the number of randomly selected variables (mtry) on the prediction error (Hamza and Larocque 2005). We optimized the mtry value by creating random forest ensembles for all possible mtry values and then selecting the optimal mtry value based on the lowest RMSE. The number of decision trees (Ntree) was set to 1,000 (Millard and Richardson 2015).

Variables selection

In order to improve the AGB estimative achievement of a good predictive performance, we identified the smallest number of variables that are able to fit the best predictive power and help in the interpretation of the final random forest (RF) regression model, employing a backward feature elimination

(Díaz-Uriarte and Alvarez de Andrés 2006). We first took into account the measures of importance to obtain an initial variable ranking and then proceeded with an iterative backward elimination of the least important variables. In each iteration, we eliminated the 20% least important variables and a new random forest was built by training with the remaining features for the assessment of RMSE. The backward stepwise selection allows us to put all variables into the random forest and eliminate unnecessary or partially correlated features (Guan et al. 2013). Besides obtaining the best overall predictive accuracy, variables selection allowed us to simplify the modelling process and identify the minimum number of features that offer the best predictive accuracy (Ismail and Mutanga 2010).

Stratification into inventory plots

We stratified the plots into six vegetation types described in Scolforo et al. (2008) (Figure 3). In each stratum, we modelled the AGB using the random forest algorithm method described above. The backward elimination procedure was also used to identify the smallest number of features that provides the best predictive results.

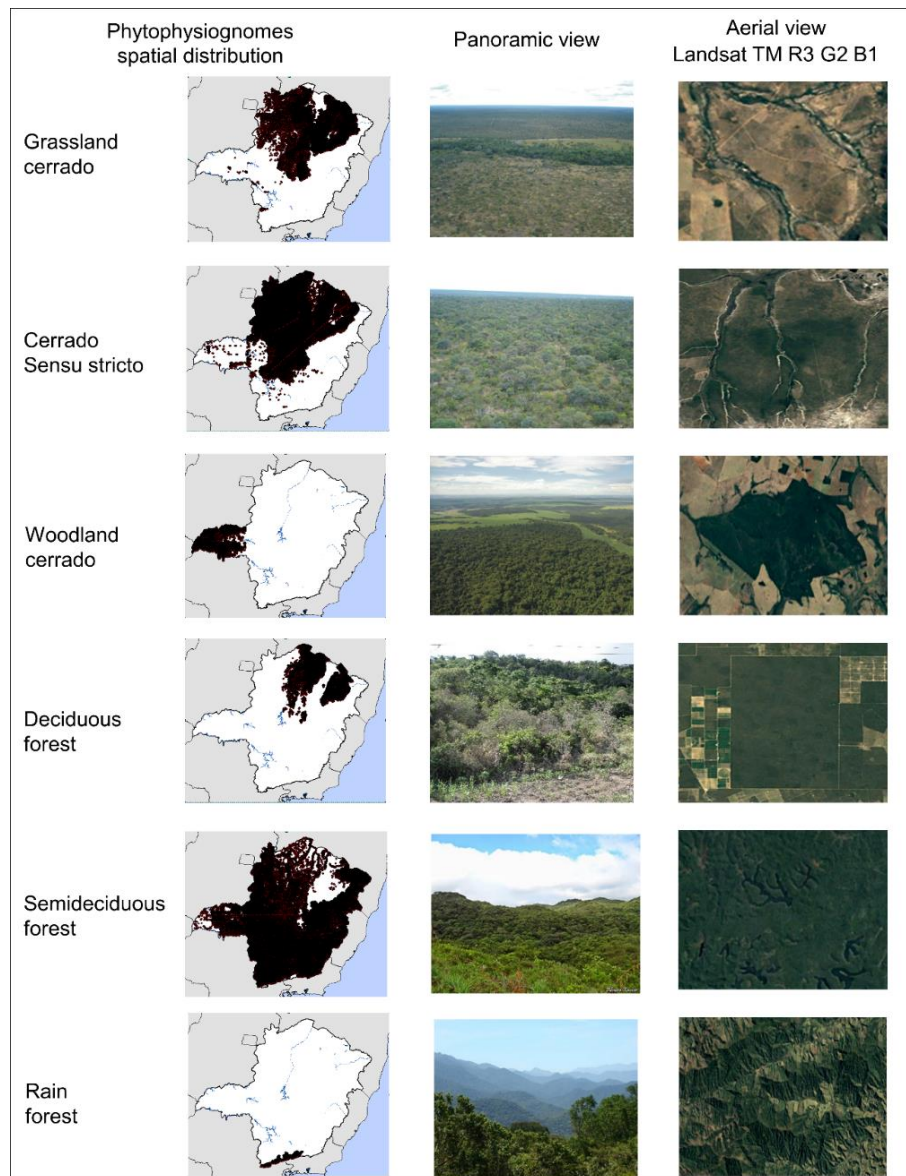


Figure 3. Vegetation types spatial distribution used to generate the stratified AGB random forest model.

Deriving the aboveground biomass maps

To derive the AGB maps, we created continuous cells with dimensions of 1 ha (100 x 100 m) throughout the length of the vegetation in Minas Gerais State (Carvalho et al. 2008). In each cell containing the values of the selected variables, we applied the random forest regression model to predict the AGB. We generated the following AGB maps: (i) non-stratified; and (ii) stratified by vegetation types (6 maps).

We compared the predictions of the maps capability computing the statistical precision mean absolute error (MAE, in %), root mean squared error (RMSE, in t/ha) and by the visual analysis of residuals graphs.

Regression-kriging

We applied regression-kriging technique to improve the map generated by random forest regression (Guo et al. 2015) by kriging the residuals produced by the model. The residuals are obtained by subtracting the predicted from the observed values, and then interpolated them using ordinary kriging. The residuals map is created and merged into AGB map previously developed. This procedure allows for removing the trends on the estimates by the geographical model, promoting a final unbiased map. This procedure was applied only in the map that provided the best results (non-stratified versus stratified).

The residuals were spatially modelled using ordinary kriging by fitting theoretical semivariogram models (specifically the Gaussian, the Spherical, and the Exponential) using the weighted least squares method. The selection and validation of the best semivariogram model was based on reduced mean error (ER) and standard deviation of reduced mean error (SDE), which were calculated on the basis of the cross validation process (Cressie 1991).

Results

Non-stratified random forest model performance

The random forest backward method selected seven variables that offered the best predictive result: Longitude, SRTM, MOD11_17LSTd (product MODIS 11, julian day 17 and dailyday time land-surface temperature variable), MOD11_113LSTd (product MODIS 11, julian day 113 and dailyday time land-surface temperature variable), TM B3, MOD11_145LSTn (product MODIS 11, julian day 17 and daily night time land-surface temperature variable) and TM B5 (Landsat SWIR 1). These variables, except SRTM, have a negative Pearson's correlation with AGB, reaching 0.25 with TM B5 (Figure 4).

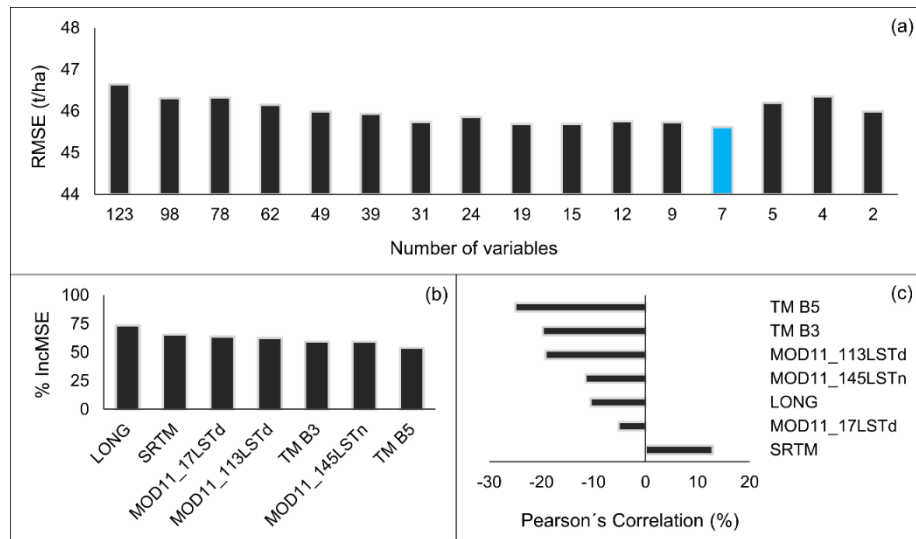


Figure 4. Variables of non-stratified model: (a) Number of features versus RMSE; (b) Selected features versus % IncMSE and (c) Person's Correlation among the selected features and AGB (t/ha).

The non-stratified model presented an R^2 value equal to 0.45. This value is explained by the wide variation found in the data. The RMSE and MAE were,

respectively, 44.68 t/ha and 38.51%. In addition, the model presented randomly distributed residuals with a slight trend, indicating underestimates of AGB (Figure 5).

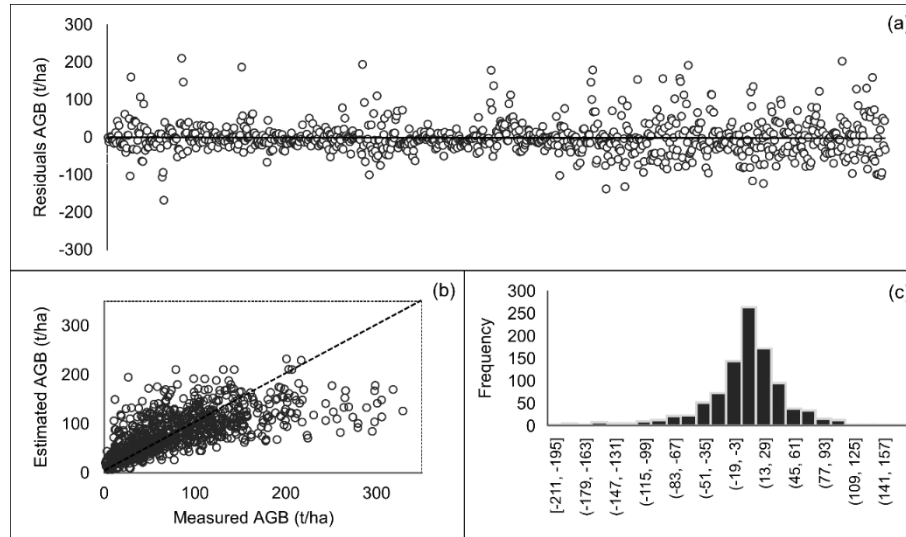


Figure 5. Residuals analysis of RF non-stratified model: (a) Distribution of the residuals; (b) Scatter plots of measured value versus estimated value; and (c) Histogram of AGB residuals.

We also analysed the predictive performance of RF regression non-stratified model by vegetation types (Table 4). Grassland cerrado, deciduous forest and cerrado sensu stricto presented the highest MAE%. The highest RMSE was found in wetland forests, semideciduous and deciduous forests. The residual distribution of each vegetation types indicates that the model present slight trends, mainly towards underestimates in rainforest, deciduous forest, and woodland cerrado, and overestimates in grassland cerrado (Figure 6).

Table 4. Predictive performance of the random forest regression non-stratified model by vegetation types.

Vegetation types	RMSE (t/ha)	MAE%
Rain forests	62.08	36.52
Deciduous Forest	45.31	35.36
Grassland Cerrado	16.60	82.69
Woodland cerrado	29.26	32.02
Cerrado stricto sensu	18.32	41.50
Semideciduous forest	52.50	39.29

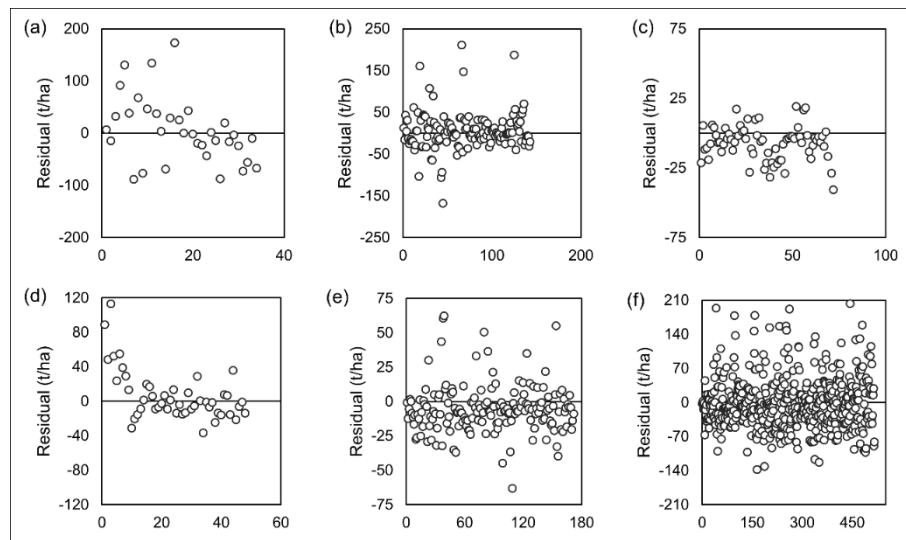


Figure 6. Residuals analysis by non-stratified RF model by vegetation type: (a): Rain forests; (b) Deciduous forest; (c) Grassland cerrado; (d) Woodland cerrado; (e) Cerrado stricto sensu; (f) Semideciduous forest.

From the AGB map generated by non-stratified model (Figure 7) using the seven variables selected, some discussion about the AGB distribution in the state of MG can be highlighted. The total area of AGB estimated by non-stratified model is 968,646,297 tons (Table 5) ranged from 5.53 to 278.76 t/ha.

The central, northwest and north regions presented low values of AGB. Cerrado sensu stricto, grassland cerrado and deciduous forest are predominant in these regions, which helps to explain the low amount AGB. In addition, the low AGB in these regions results from the combination of high temperatures and low amount rain. The west region presents intermediate values, due to the presence of woodland cerrado, and the south, central and east of MG presented high values of AGB, due to the predominance of semideciduous forest and rain forests, with lower temperatures and higher amount rain over the year.

Table 5. Total aboveground biomass by vegetation type estimated by non-stratified random forest model.

Vegetation types	AGB (T)	Area (ha)
Rain forests	25,809,743	224,108
Deciduous forest	128,060,012	2,029,065
Grassland Cerrado	82,490,430	1,484,230
Woodland cerrado	22,923,814	353,531
Cerrado stricto sensu	300,647,793	5,460,099
Semideciduous forest	408,714,504	5,171,549
Total	968,646,297	14,722,582

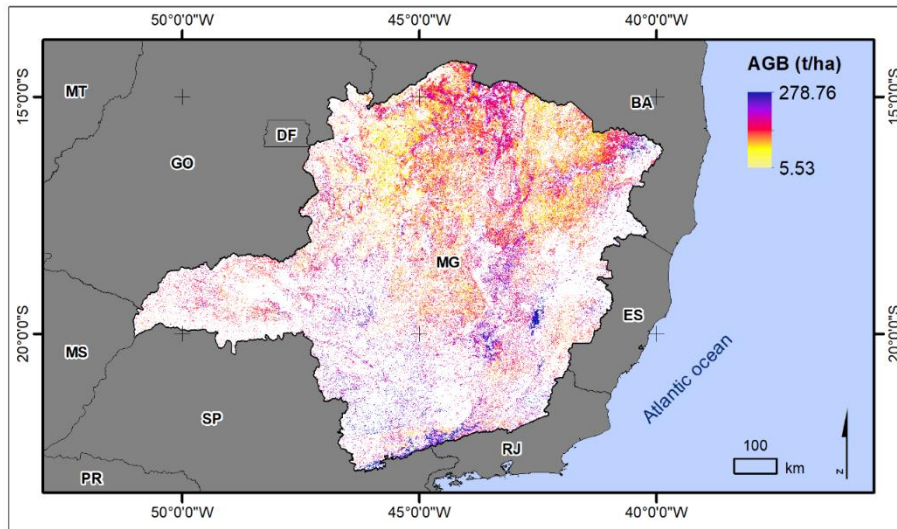


Figure 7. AGB map of Minas Gerais state estimated by non-stratified RF model.

Stratified random forest model performance

The random forest backward procedure method selected singular variables for each vegetation type. For those widely distributed in Minas Gerais state (semideciduous forest, cerrado sensu stricto, and grassland cerrado) the number of variables was greater, ranging from 9 to 24. For those that are concentrated in specific regions of Minas Gerais state (woodland cerrado, deciduous forest, and rain forest) the number of variables selected was lower, ranging from 2 to 4 (Figure 8).

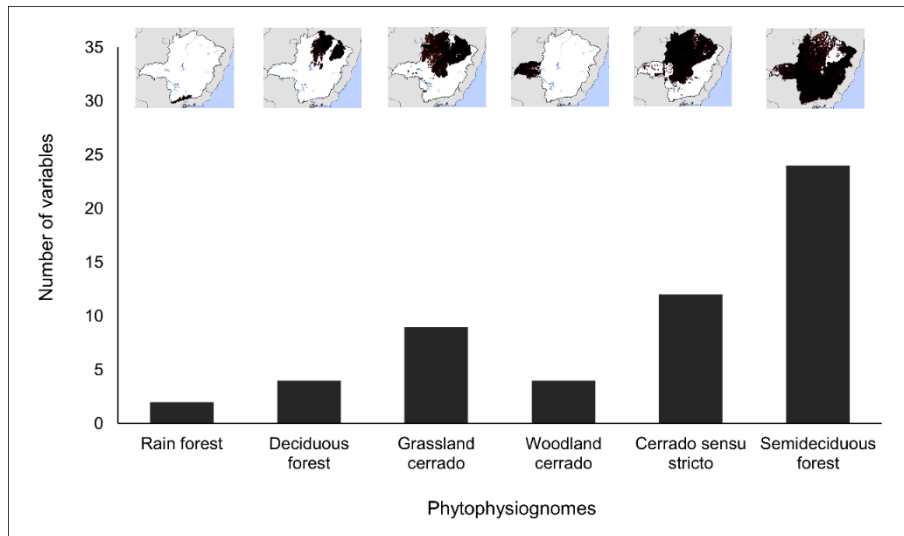


Figure 8. Number of variables selected by random forest stratified model for each vegetation type.

The stratified models presented an R^2 value ranging from 0.15 to 0.52. The rain forest presented the lowest R^2 and the highest root mean square error (RMSE). Woodland cerrado presented the greatest R^2 and the lowest mean absolute error (MAE%), indicating this prediction as the most accurate (Table 6). The residuals distribution of each vegetation types presented a balanced spatial distribution with non-characterized trend, indicating an accurate prediction of AGB (Figure 9).

Table 6. Predictive performance of the random forest regression stratified model by vegetation types.

Vegetation types	R ²	RMSE	MAE%
Rain forest	15.39	57.36	34.56
Deciduous forest	37.05	42.76	34.74
Grassland cerrado	35.51	7.72	54.73
Woodland cerrado	52.26	16.47	23.62
Cerrado stricto sensu	51.97	14.53	31.51
Semideciduous forest	20.19	51.25	37.84

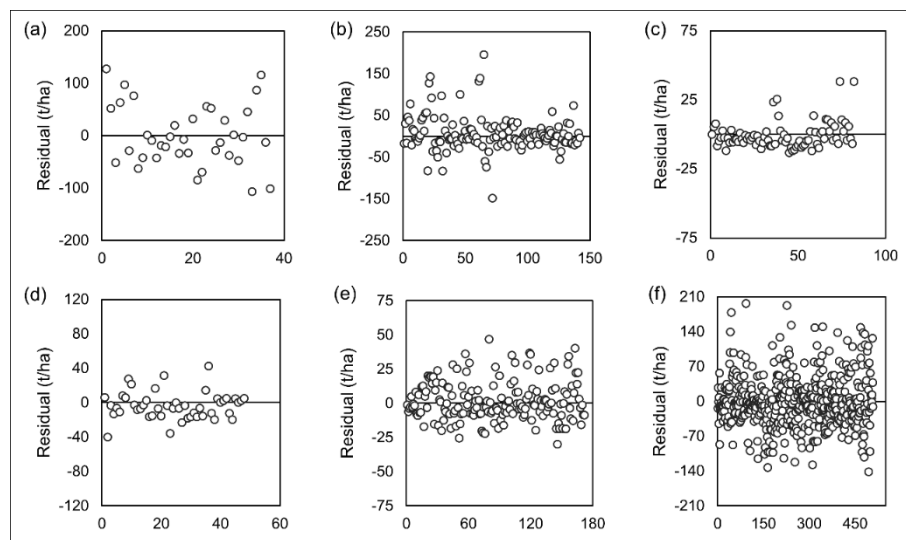


Figure 9. Residuals analysis by stratified RF model by vegetation type: (a): Rain forests; (b) Deciduous forest; (c) Grassland cerrado; (d) Woodland cerrado; (e) Cerrado stricto sensu; (f) Semideciduous forest.

We modelled the rain forest AGB using Landsat TM B3 (red waveband) and EVI (enhanced vegetation index). The R² was low (0.15), however the MAE (%) presented good results (36.52%). The highest values of rain forest aboveground biomass are concentrated in the west strip of this vegetation type and

the lowest ones are distributed throughout the area. The AGB values ranged from 39.35 to 261.98 t/ha (Figure 10a). These high values explain the high RMSE obtained (57.36%).

The deciduous forest AGB ranged from 15.97 to 166.34 t/ha (Figure 10b). The random forest regression selected MOD9_17SR2 (product MODIS 9, julian day 9 and surface reflectance band 2 variable), MOD11_17LSTd (product MODIS 11, julian day 17 and dailyday time Land-surface temperature variable), MOD15_121FPAR (product MODIS 15, julian day 121 and fraction of photosynthetically active radiation absorbed by vegetation variable) and MOD17_241GPP variables (product MODIS 17, julian day 241 and gross primary production variable).

The grassland cerrado map (Figure 10c) presented AGB ranging from 3.66 to 34.01 with higher values in the west part of its occurrence. Random forest regression selected nine variables: three from Landsat TM (SAVI, EVI and B4-NIR waveband), four from MODIS 13 product (MOD13_113NIR, MOD13_145NIR, MOD13_177red and MOD13_209NIR), MOD9_145RS2 and MOD11_113LSTd).

About woodland cerrado, random forest regression selected four variables: EVI (enhanced vegetation index derived from Landsat TM), MOD15_241FPAR (product MODIS 15, julian day 241 and fraction of photosynthetically active radiation absorbed by vegetation variable), Latitude and TM B3 (red waveband). The AGB ranged from 29.23 to 134.72 t/ha with the high values concentrated in the north of its occurrence (Figure 10d).

The random forest regression selected 12 variables to model cerrado sensu stricto and 24 variables to model semideciduous forest (Table 7). The AGB ranged from 14.25 to 113.60 t/ha and 38.29 to 245.95 t/ha for cerrado sensu stricto (Figure 10e) and deciduous forest (Figure 10f), respectively.

Table 7. Selected variables by random forest regression for cerrado sensu stricto and semideciduos forest.

Cerrado sensu stricto	Semideciduous forest
TM B5	TM B1
SAVI	TM B2
MOD9_145SR1	TM B3
MOD9_241SR2	TM B5
MOD13_17red	TM B7
MOD13_17blue	NDVI
MOD13_17NDVI	Bio 2
MOD13_241blue	Bio 4
MOD17_241NET	Bio 6
MOD17_289NET	Bio 8
MOD17_289GPP	Bio 9
Longitude	Bio 12
	Bio 18
	Bio 19
	MOD11_113LSTd
	MOD11_145LSTd
	MOD11_145LSTn
	MOD11_17LSTd
	MOD13_177NDVI
	MOD44_065
	MOD15_209FPAR
	Longitude
	Latitude
	SRTM

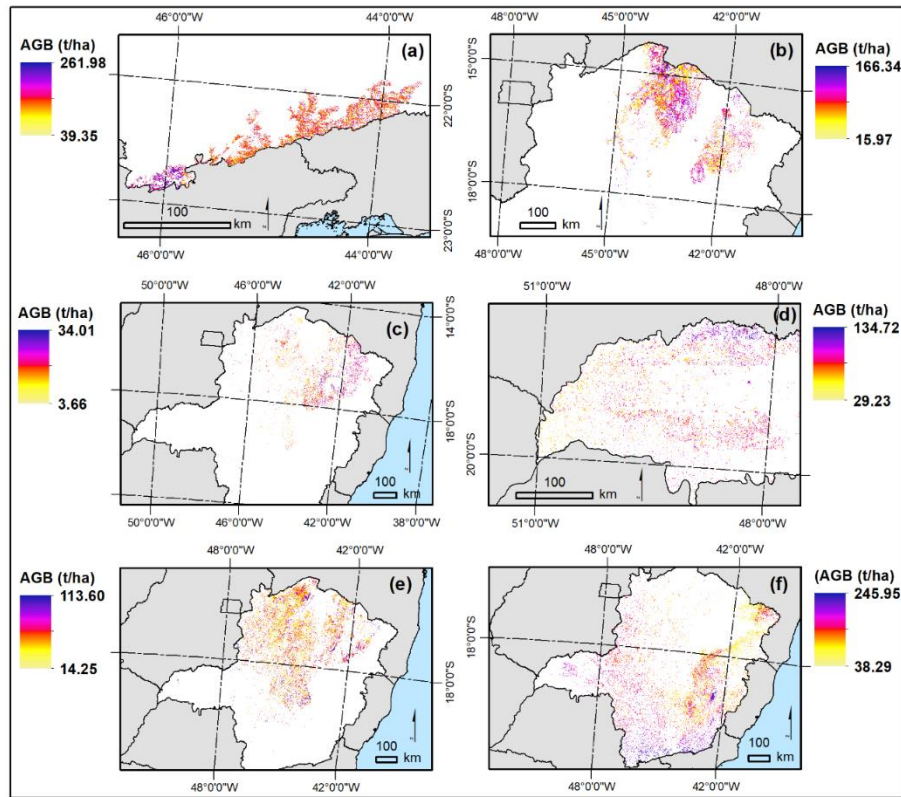


Figure 10. AGB map of Minas Gerais state estimated by stratified RF model:

(a): Rain forest; (b) Deciduous forest; (c) Grassland cerrado; (d) Wooded cerrado; (e) Cerrado stricto sensu; (f) Semideciduous forest.

The total area of AGB estimated by stratified model is 871,939,468 tons (Table 8), ranging from 3.66 to 261.98 t/ha.

Table 8. Total aboveground biomass by vegetation type estimated by non-stratified random forest model.

Vegetation types	AGB (T)	Area (ha)
Rain forests	29,787,278	224,108
Deciduous Seasonal Forest	138,614,744	2,029,065
Grassland Cerrado	21,606,086	1,484,230
Woodland cerrado	27,616,639	353,531
Cerrado stricto sensu	160,311,870	5,460,099
Semideciduous forest	494,002,853	5,171,549
Total	871,939,468	14,722,582

Comparative analysis of random forest models performance

The performance of non-stratified versus stratified AGB models was assessed based on the mean absolute error (MAE) and root mean squared error (RMSE) statistics (Table 9). And also, we analysed the scatter plots of measured values versus estimated values of each vegetation type.

The stratified models significantly improved the AGB prediction by reducing the MAE and RMSE for all vegetation types, mainly woodland cerrado (RMSE decreased from 29.26 to 16.47 t/ha), grassland cerrado (MSE reduced from 82.69 to 54.73%) and cerrado sensu stricto (RMSE decreased from 18.32 to 14.53 t/ha and MSE reduced from 41.50 to 31.51%).

Table 9. Root mean squared error (RMSE) and mean absolute error (MAE (%)) of random forest models.

Vegetation types	Non-stratified model		Stratified model	
	RMSE	MAE%	RMSE	MAE%
Rain forests	62.08	36.52	57.36	34.56
Deciduous forest	45.31	35.36	42.76	34.74
Grassland cerrado	16.60	82.69	7.72	54.73
Woodland cerrado	29.26	32.02	16.47	23.62
Cerrado stricto sensu	18.32	41.50	14.53	31.51
Semideciduous forest	52.50	39.29	51.25	37.84

The increasing on the performance of random forest stratified models is also highlighted in the scatter plot graphics. The measured AGB versus estimated by non-stratified model (Figure 11) indicates a slight trend of underestimates (rain forest, deciduous forest, and woodland cerrado) and overestimates (grassland cerrado). These trends did not occur in the stratified random forest models (Figure 12), highlighting the improvements of using these models.

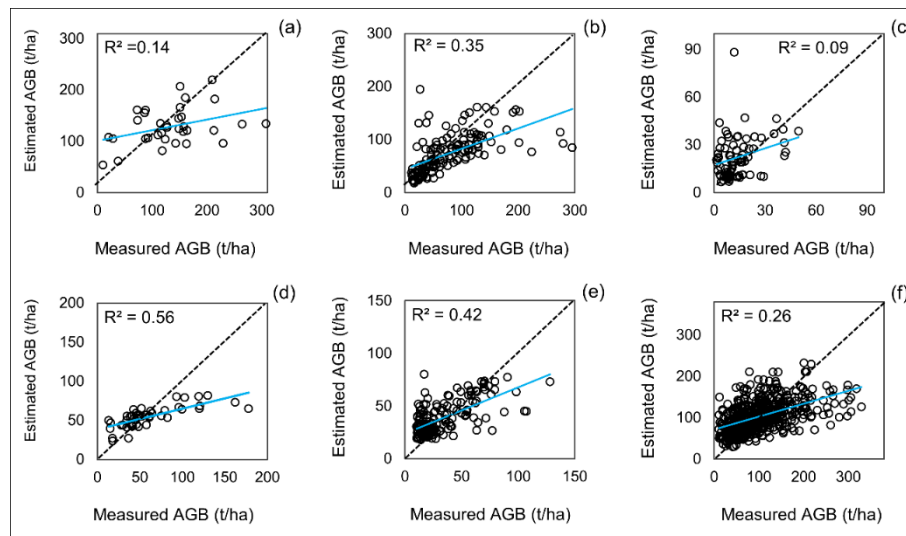


Figure 11. Scatter plots of measured value versus estimated value by non-stratified RF model by vegetation type: (a): Rain forest; (b) Deciduous forest; (c) Grassland cerrado; (d) Woodland cerrado; (e) Cerrado stricto sensu; (f) Semideciduous forest. A 1:1 line (black, dashed) is provided for reference.

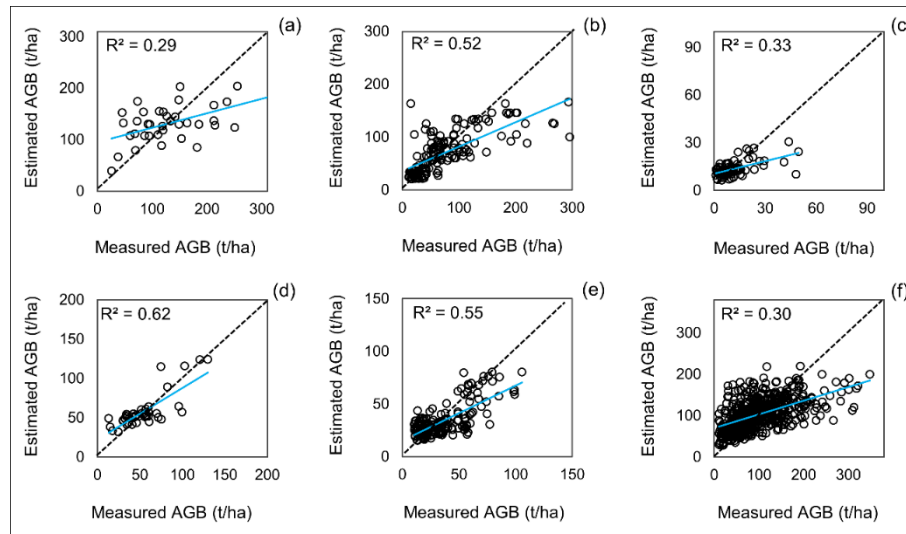


Figure 12. Scatter plots of measured value versus estimated value by stratified RF model by vegetation type: (a): Rain forests; (b) Deciduous forest; (c) Grassland cerrado; (d) Woodland cerrado; (e) Cerrado stricto sensu; (f) Semideciduous forest. A 1:1 line (black, dashed) is provided for reference.

Regression-kriging

We modeled the residuals by the stratified random forest regression. The experimental and modeled semivariograms for the residuals derived from each vegetation type for aboveground biomass (Figure 13) presented reduction in both mean error (RE) and standard deviation of reduced errors (SRE)(Table 10).

Table 10. Semivariogram statistics for residuals from the regression models.

Vegetation types	RE	SRE
Rains forest	0.023	0.85
Deciduous forest	-0.004	1.18
Grassland cerrado	0.024	1.24
Woodland cerrado	-0.012	0.89
Cerrado stricto sensu	0.019	1.00
Semideciduous forest	0.010	1.10

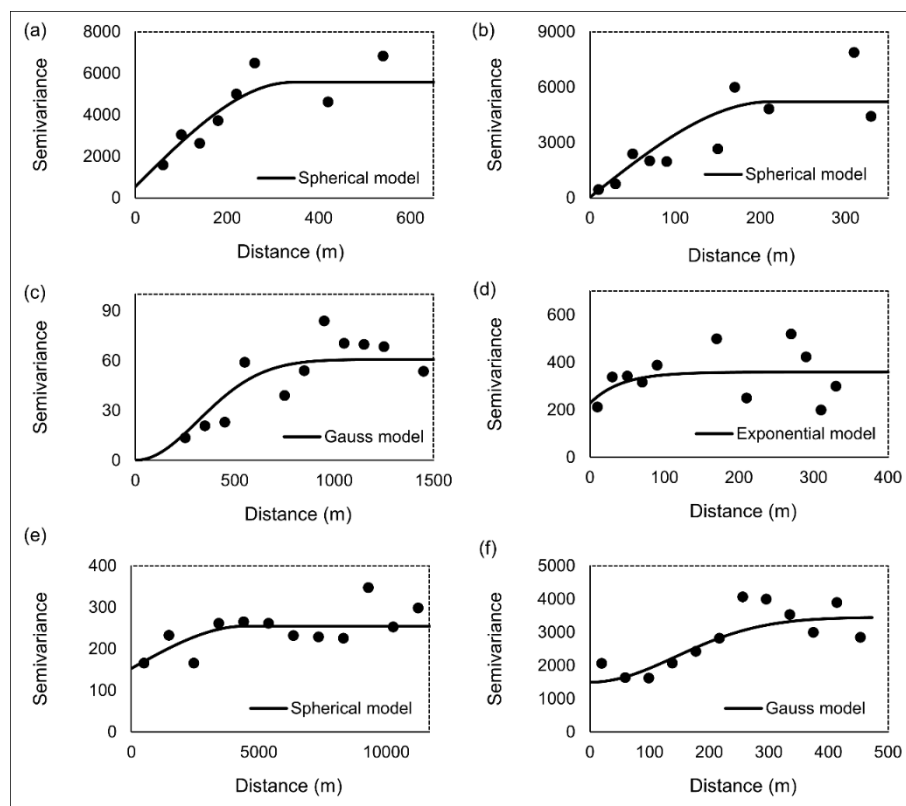


Figure 13. Theoretical and experimental semivariogram for the residual aboveground biomass (t/ha) by vegetation type.

Improved aboveground biomass map

We applied the regression-kriging technique into the six maps generated by the stratified models. We corrected the individual vegetation type AGB maps, by adding the residuals map into the predicted values map by the models. The total area of AGB for Minas Gerais State is about 839,375,640 tones, with mean values ranging from 13.32 t/ha (grassland cerrado) to 124.03 t/ha (rain forests) (Table 11). The individual's maps area presented in Figure 14.

Table 11. Total and mean aboveground biomass (t/ha) of Minas Gerais state by each vegetation type.

Vegetation types	AGB (T)	Mean AGB (t/ha)	Area (ha)
Rain forests	27,795,005	124.03	224,108
Deciduous Seasonal Forest	139,216,378	68.61	2,029,065
Grassland Cerrado	19,762,640	13.32	1,484,230
Woodland cerrado	25,893,228	73.24	353,531
Cerrado stricto sensu	163,977,289	30.03	5,460,099
Semideciduous forest	462,731,101	89.48	5,171,549
Total	839,375,640		14,722,582

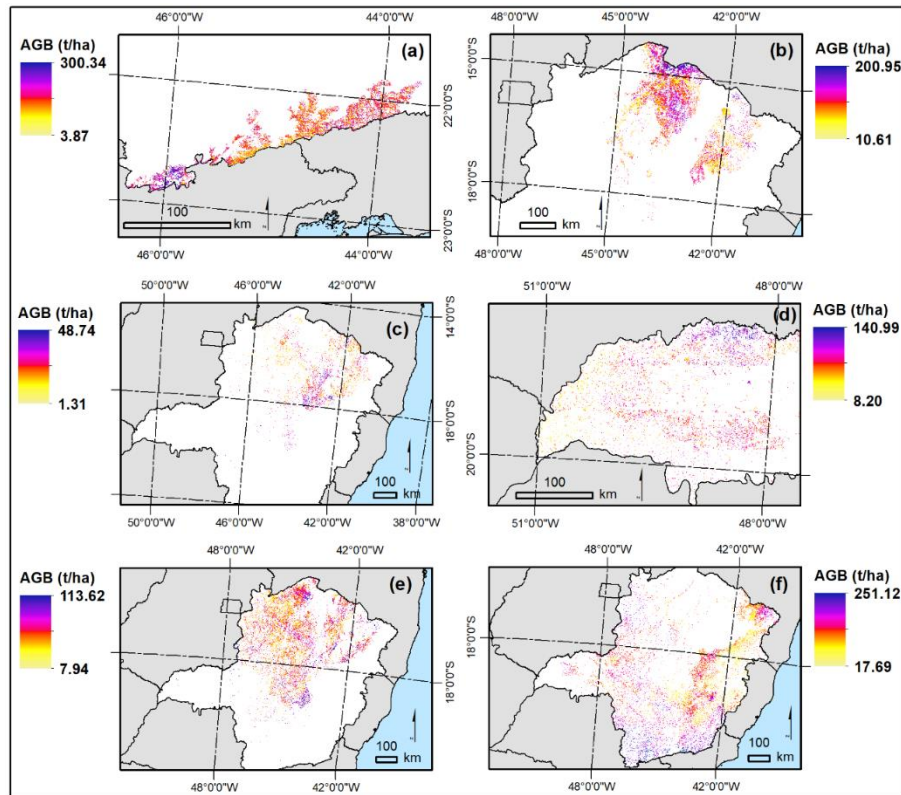


Figure 14. Improved AGB maps by vegetation types: (a) Rain forest; (b) Deciduous forest; (c) Grassland cerrado; (d) Woodland cerrado; (e) Cerrado stricto sensu; (f) Semideciduous forest.

Discussion

We proposed a stratification approach combined with regression-kriging technique to improve aboveground biomass models in large and heterogeneous Savanna-Forest transition areas in Minas Gerais state, Brazil. The effect of stratification on the AGB predictions in combination with the appropriate selection of variables was investigated. First, we compared the predictive performance of non-stratified map with the stratified one. Second, we applied the regression kriging, by kriging the model's residuals, and then merged it into the

stratified map to generate the improved aboveground biomass of Minas Gerais state, Brazil (Figure 15).

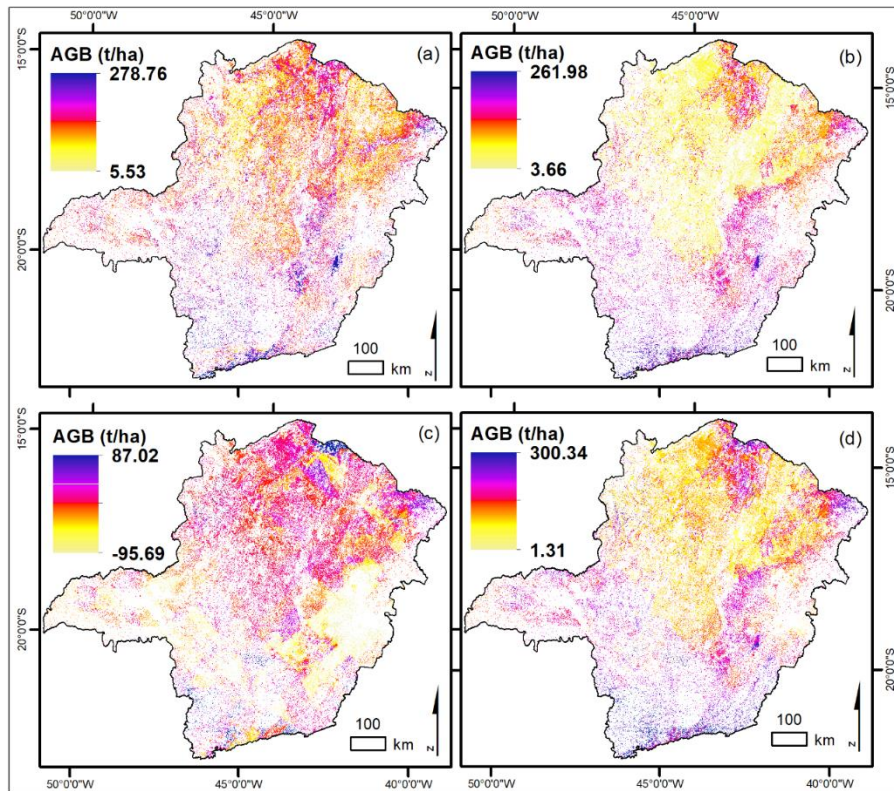


Figure 15. (a) aboveground biomass map generated using non-stratified model; (b) aboveground biomass map generated using stratified model; (c) residuals kriging of stratified models and (d) final aboveground map of Minas Gerais state.

The results indicate that the stratification into vegetation types decreases the root mean square error and mean absolute error, mainly in grassland cerrado (RMSE reduction of 53.48% and MAE reduction of 44.62%), woodland cerrado (RMSE reduction of 43.71% and MAE reduction of 34.73%) and cerrado sensu

stricto (RMSE reduction of 20.68% and MAR reduction of 17.46%). The stratified rain forest, deciduous and semideciduous forests model were slightly better than the non-stratified one (see Table 9). Latifi et al. (2015) analyzed the impact of stratifying forest data into three classes (broadleaved, coniferous and mixed forest) to estimate aboveground forest biomass. The results revealed marginal advantages for the strata prediction models over the non-stratified ones. A previous study clearly illustrated that stratification of the sample plots can help to improve the accuracy of the estimation of forestry stand parameters. However, this holds true to a different degree for each one of the individual stand parameters, whereas the standing volume and density were significantly enhanced (Heurich and Thoma 2008).

In the non-stratified model, that mixed the Savanna-Forest areas, seven variables were selected (see Figure 6). Among them, two are geographical (longitude and elevation), and the remaining five were taken from remote sensing. From the remote sensing, three are climate variables, displaying temperature and two are from medium spatial resolution, displaying spectral information that are negative correlated with aboveground biomass. Scolforo et al. (2015) showed correlations between forest carbon and geographical variables (latitude and altitude), which is expected due to their relation with the climate variables. In summary, the variables that provide the best performance of non-stratified model are related with regional scale, providing spatio-environmental information.

In the stratified models, the variables selection of each vegetation type occurred as a function of their spatial distribution and vegetation phenology. The number of variables was driven by the spatial distribution and it is a characteristic of the seasonality effects (Figure 16). For vegetation types that are widely distributed throughout Minas Gerais (semideciduous forest, cerrado sensu stricto and grassland cerrado) the random forest regression selected a

higher number of variables to best fit the model than those vegetation types that are local distributed in the study area (woodland cerrado, deciduous forest and rain forest). Also, as the vegetation phenology exists, the temporal variables were selected instead of the spectral and spatial ones.

Random forest modelled rainforest using a small number of variables, comprising only remote sensing data (TM B3 and EVI). This vegetation type is local distribute (southern Minas Gerais), not requiring the use of spatio-environmental variables to explain and predict the AGB distribution. And also, the rainforests are evergreen and do not need the use of seasonal variables to improve its capability.

Deciduous forest are seasonally dry tropical forest, submitted to a seasonal rainfall regime (Santos et al. 2011). Thus, random forest regression selected only temporal remote sensing variables derived from MODIS products to model the AGB of this vegetation type (MOD9_17SR2, MOD11_17LSTd, MOD15_121FPAR and MOD17_241GPP). These variables cover both the wet and dry seasons, capturing the phenological effects of deciduous forests. The low number of variables is explained by the local distribution of this vegetation type in the state of Minas Gerais.

Random forest modelled grassland cerrado using 9 remote sensing variables, mixing 33% of Landsat TM and 67% of MODIS products. Grassland cerrado are widespread in MG state and also are influenced by seasonality effects (Veldman et al. 2015).

Woodland cerrado comprising dense woodland with high trees (Schwieder et al. 2016) are located in the western Minas Gerais state. Because of its local vegetation type, random forest regression selected four variables comprising 25% of environmental and 75% of remote sensing data (EVI, TM B3 and MOD15_241FPAR).

Random forest modelled semideciduous forest using 24 variables, mixing spatio-environmental (46%) and remote sensing data (54%). From remote sensing data, both spectral (Landsat TM band and indices) and temporal data (MODIS products) were selected. The high number of variables selected is explained by the large spatial distribution of this vegetation type (Scolforo et al. 2015).

The AGB estimates from Savannas-Forest transitions using stratified models are more accurate than the non-stratified one, since the relationship among AGB and the variables are peculiar to each vegetation type. Seasonal vegetation types (i.e. deciduous forest) are associated with the temporal characteristics of remote sensing images, requiring high temporal image resolution (~monthly) instead of medium spatial image resolution (~30 m). On the contrary, non-seasonal vegetation types (i.e. rain forest) require medium spatial resolution instead of high temporal image resolution. Those variables that are widely spatial distributed require a higher number of variables and the combination not only of high temporal and medium spatial resolutions, but also the inclusion of spatio-environmental variables.

Conclusions

Considering that Savanna, Atlantic forest and Semi-arid woodland biome are large and heterogeneous in Minas Geraus state, Brazil, our study contributes to the understanding of the relationship between spatio-environmental/remote sensing data and aboveground biomass (AGB) distribution. The stratification of data into vegetation types not only improved the accuracy of AGB estimative, but also allowed random forest regression to select the lowest number of variables that offer the best predictive model performance to AGB mapping. The improvement in AGB estimates is driven by the spatial distribution and seasonality effects of each vegetation type, and it is

achieved by stratifying the models which minimize the Savanna-Forest transition heterogeneity.

The refining map and the understanding of how the variables properties are associated with the AGB enable researches to improve the roughly estimates of greenhouse gas emission and also helps the selection of appropriate variables that best model the aboveground biomass in Savanna-Forest transition areas.

References

Arantes, Arielle Elias, Laerte G. Ferreira, and Michael T. Coe. 2016. "The Seasonal Carbon and Water Balances of the Cerrado Environment of Brazil: Past, Present, and Future Influences of Land Cover and Land Use." *ISPRS Journal of Photogrammetry and Remote Sensing* 117. International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS): 66–78. doi:10.1016/j.isprsjprs.2016.02.008.

Aslan, Aslan, Abdullah F. Rahman, Matthew W. Warren, and Scott M. Robeson. 2016. "Mapping Spatial Distribution and Biomass of Coastal Wetland Vegetation in Indonesian Papua by Combining Active and Passive Remotely Sensed Data." *Remote Sensing of Environment* 183. Elsevier Inc.: 65–81. doi:10.1016/j.rse.2016.04.026.

Asner, Gregory P, Veronique De Sy, Martin Herold, Alex Held, Josef Kellndorfer, and Jan Verbesselt. 2012. "Synergies of Multiple Remote Sensing Data Sources for REDD + Monitoring ' de," 696–706. doi:10.1016/j.cosust.2012.09.013.

Baccini, A., M. A. Friedl, C. E. Woodcock, and R. Warbington. 2004. "Forest Biomass Estimation over Regional Scales Using Multisource Data." *Geophysical Research Letters* 31 (10): 2–5. doi:10.1029/2004GL019782.

Baccini, A., S. J. Goetz, W. S. Walker, N. T. Laporte, M. Sun, D. Sulla-Menashe, J. Hackler, et al. 2012. "Estimated Carbon Dioxide Emissions from

Tropical Deforestation Improved by Carbon-Density Maps.” *Nature Climate Change* 2 (3). Nature Publishing Group: 182–85. doi:10.1038/nclimate1354.

Baccini, A, N Laporte, S J Goetz, M Sun, and H Dong. 2008. “A First Map of Tropical Africa ’ S above-Ground Biomass Derived from” 45011. doi:10.1088/1748-9326/3/4/045011.

Breiman, Leo. 2001. “Random Forests.” *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.

Castillo-Santiago, Miguel Ángel, Adrián Ghilardi, Ken Oyama, José Luis Hernández-Stefanoni, Ignacio Torres, Alejandro Flamenco-Sandoval, Ana Fernández, and Jean François Mas. 2013. “Estimating the Spatial Distribution of Woody Biomass Suitable for Charcoal Making from Remote Sensing and Geostatistics in Central Mexico.” *Energy for Sustainable Development* 17 (2). International Energy Initiative: 177–88. doi:10.1016/j.esd.2012.10.007.

Cressie, N. 1991. “Statistics for Spatial Data: Wiley Series in Probability and Statistics,” 900. <http://textbooks-to-succeed.info/wp-content/uploads/pdfs/Statistics for Spatial Data Wiley Series in Probability and Statistics by Noel A C Cressie - Thorough Account Of Spatial Statistical Methods Including Resampling Approaches.pdf>.

Dahlke, Mark, F Jay Breidt, Jean D Opsomer, and Ingrid Van Keilegom. 2013. “POST-STRATIFICATION ESTIMATION” 23: 189–211.

Díaz-Uriarte, Ramón, and Sara Alvarez de Andrés. 2006. “Gene Selection and Classification of Microarray Data Using Random Forest.” *BMC Bioinformatics* 7: 3. doi:10.1186/1471-2105-7-3.

Duro, Dennis C., Steven E. Franklin, and Monique G. Dub?? 2012. “A Comparison of Pixel-Based and Object-Based Image Analysis with Selected Machine Learning Algorithms for the Classification of Agricultural Landscapes Using SPOT-5 HRG Imagery.” *Remote Sensing of Environment* 118. Elsevier Inc.: 259–72. doi:10.1016/j.rse.2011.11.020.

Dymond, Caren C., David J. Mladenoff, and Volker C. Radeloff. 2002. "Phenological Differences in Tasseled Cap Indices Improve Deciduous Forest Classification." *Remote Sensing of Environment* 80 (3): 460–72. doi:10.1016/S0034-4257(01)00324-8.

Fayad, Ibrahim, Nicolas Baghdadi, Stéphane Guitet, Jean Stéphane Bailly, Bruno Hérault, Valéry Gond, Mahmoud El Hajj, and Dinh Ho Tong Minh. 2016. "Aboveground Biomass Mapping in French Guiana by Combining Remote Sensing, Forest Inventories and Environmental Data." *International Journal of Applied Earth Observation and Geoinformation* 52. Elsevier B.V.: 502–14. doi:10.1016/j.jag.2016.07.015.

Galeana-Pizaña, J. Mauricio, Alejandra López-Caloca, Penélope López-Quiroz, José Luis Silván-Cárdenas, and Stéphane Couturier. 2014. "Modeling the Spatial Distribution of above-Ground Carbon in Mexican Coniferous Forests Using Remote Sensing and a Geostatistical Approach." *International Journal of Applied Earth Observation and Geoinformation* 30 (1): 179–89. doi:10.1016/j.jag.2014.02.005.

Gizachew, Belachew, Svein Solberg, Erik Næsset, Terje Gobakken, Ole Martin Bollandsås, Johannes Breidenbach, Eliakimu Zahabu, and Ernest William Mauya. 2016. "Mapping and Estimating the Total Living Biomass and Carbon in Low-Biomass Woodlands Using Landsat 8 CDR Data." *Carbon Balance and Management* 11 (1). Springer International Publishing: 1–14. doi:10.1186/s13021-016-0055-8.

Gleason, Colin J., and Jungho Im. 2012. "Forest Biomass Estimation from Airborne LiDAR Data Using Machine Learning Approaches." *Remote Sensing of Environment* 125. Elsevier Inc.: 80–91. doi:10.1016/j.rse.2012.07.006.

Guan, Haiyan, Jonathan Li, Michael Chapman, Fei Deng, Zheng Ji, and Xu Yang. 2013. "Integration of Orthoimagery and Lidar Data for Object-Based Urban Thematic Mapping Using Random Forests." *International Journal of*

Remote Sensing 34 (14): 5166–86. doi:10.1080/01431161.2013.788261.

Guo, Peng Tao, Mao Fen Li, Wei Luo, Qun Feng Tang, Zhi Wei Liu, and Zhao Mu Lin. 2015. “Digital Mapping of Soil Organic Matter for Rubber Plantation at Regional Scale: An Application of Random Forest plus Residuals Kriging Approach.” *Geoderma* 237–238. Elsevier B.V.: 49–59. doi:10.1016/j.geoderma.2014.08.009.

Hamza, Mounir, and Denis Larocque. 2005. “An Empirical Comparison of Ensemble Methods Based on Classification Trees.” *Journal of Statistical Computation and Simulation* 75 (8): 629–43. doi:10.1080/00949650410001729472.

Heurich, Marco, and Franz Thoma. 2008. “Estimation of Forestry Stand Parameters Using Laser Scanning Data in Temperate, Structurally Rich Natural European Beech (*Fagus Sylvatica*) and Norway Spruce (*Picea Abies*) Forests.” *Forestry* 81 (5): 645–61. doi:10.1093/forestry/cpn038.

Hijmans, Robert J., Susan E. Cameron, Juan L. Parra, Peter G. Jones, and Andy Jarvis. 2005. “Very High Resolution Interpolated Climate Surfaces for Global Land Areas.” *International Journal of Climatology* 25 (15): 1965–78. doi:10.1002/joc.1276.

Hu, Tianyu, Yanjun Su, Baolin Xue, Jin Liu, Xiaoqian Zhao, Jingyun Fang, and Qinghua Guo. 2016. “Mapping Global Forest Aboveground Biomass with Spaceborne LiDAR, Optical Imagery, and Forest Inventory Data.” *Remote Sensing* 8 (7). doi:10.3390/rs8070565.

Ismail, R., and O. Mutanga. 2010. “A Comparison of Regression Tree Ensembles: Predicting Sirex Noctilio Induced Water Stress in *Pinus Patula* Forests of KwaZulu-Natal, South Africa.” *International Journal of Applied Earth Observation and Geoinformation* 12 (SUPPL. 1): 45–51. doi:10.1016/j.jag.2009.09.004.

Latifi, Hooman, Fabian E. Fassnacht, Florian Hartig, Christian Berger,

Jaime Hernández, Patricio Corvalán, and Barbara Koch. 2015. “Stratified Aboveground Forest Biomass Estimation by Remote Sensing Data.” *International Journal of Applied Earth Observation and Geoinformation* 38. Elsevier B.V.: 229–41. doi:10.1016/j.jag.2015.01.016.

Lu, Dengsheng. 2006. “The Potential and Challenge of Remote Sensing-based Biomass Estimation.” *International Journal of Remote Sensing* 27 (7): 1297–1328. doi:10.1080/01431160500486732.

Lu, Dengsheng, Qi Chen, Guangxing Wang, Emilio Moran, Mateus Batistella, Maozhen Zhang, Gaia Vaglio Laurin, and David Saah. 2012. “Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates.” *International Journal of Forestry Research* 2012 (1): 1–16. doi:10.1155/2012/436537.

Main-Knorn, Magdalena, Warren B. Cohen, Robert E. Kennedy, Wojciech Grodzki, Dirk Pflugmacher, Patrick Griffiths, and Patrick Hostert. 2013. “Monitoring Coniferous Forest Biomass Change Using a Landsat Trajectory-Based Approach.” *Remote Sensing of Environment* 139: 277–90. doi:10.1016/j.rse.2013.08.010.

Matthews, H Damon, Tanya L Graham, Serge Keverian, Cassandra Lamontagne, Donny Seto, and Trevor J Smith. 2014. “National Contributions to Observed Global Warming.” *Environmental Research Letters* 9 (1): 14010. doi:10.1088/1748-9326/9/1/014010.

Millard, Koreen, and Murray Richardson. 2015. “On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping.” *Remote Sensing* 7 (7): 8489–8515. doi:10.3390/rs70708489.

Miranda, a C, H S Mirnda, J LLoyd, J Grace, R J Francey, J a McIntyre, P Meir, P Riggan, R Lockwood, and J Brass. 1997. “Fluxes of Carbon, Water and Energy over Brazilian Cerrado: An Analysis Using Eddy Covariance and Stable

Isotopes.” *Plant, Cell and Environment* 20: 315–28. doi:10.1046/j.1365-3040.1997.d01-80.x.

Mitchard, Edward T A, Ted R. Feldpausch, Roel J W Brienen, Gabriela Lopez-Gonzalez, Abel Monteagudo, Timothy R. Baker, Simon L. Lewis, et al. 2014. “Markedly Divergent Estimates of Amazon Forest Carbon Density from Ground Plots and Satellites.” *Global Ecology and Biogeography* 23 (8): 935–46. doi:10.1111/geb.12168.

Ministry of Agriculture, Livestock and Food Supply (MALFS). 2010. Brazilian agribusiness in Brazilian government.

Myers, Norman, Russell A. Mittermeier, Cristina G Mittermeier, Gustavo A. B. da Fonseca, and Jennifer Kent. 2000. “Biodiversity Hotspots for Conservation Priorities.” *Nature* 403 (6772): 853–58. doi:10.1038/35002501.

Oliveira Filho, A.T., Scolforo, J.R.S., 2008. Inventário Florestal De Minas Gerais: Espécies Arbóreas Da Flora Nativa, first ed. Editora UFLA, Lavras

Powell, Scott L, Warren B Cohen, Sean P Healey, Robert E Kennedy, Gretchen G Moisen, Kenneth B Pierce, and Janet L Ohmann. 2010. “Remote Sensing of Environment Quanti Fi Cation of Live Aboveground Forest Biomass Dynamics with Landsat Time-Series and Fi Eld Inventory Data : A Comparison of Empirical Modeling Approaches.” *Remote Sensing of Environment* 114 (5). Elsevier Inc.: 1053–68. doi:10.1016/j.rse.2009.12.018.

Ribeiro, Sabina Cerruto, Lutz Fehrmann, Carlos Pedro Boechat Soares, Laércio Antônio Gonçalves Jacovine, Christoph Kleinn, and Ricardo de Oliveira Gaspar. 2011. “Above- and Belowground Biomass in a Brazilian Cerrado.” *Forest Ecology and Management* 262 (3). Elsevier B.V.: 491–99. doi:10.1016/j.foreco.2011.04.017.

Rodríguez-veiga, Pedro, Sassan Saatchi, Kevin Tansey, and Heiko Balzter. 2016. “Remote Sensing of Environment Magnitude , Spatial Distribution and Uncertainty of Forest Biomass Stocks in Mexico.” *Remote*

Sensing of Environment 183. The Authors: 265–81. doi:10.1016/j.rse.2016.06.004.

Santos, Jc Jean Carlos, INARA R Leal, Jarcilene Silva Almeida-Cortez, G Wilson Fernandes, and Marcelo Tabarelli. 2011. “Caatinga: The Scientific Negligence Experienced by a Dry Tropical Forest.” *Tropical Conservation Science* 4 (3): 276–86. doi:10.1177/194008291100400306.

Schwieder, Marcel, Pedro J. Leitão, Mercedes Maria da Cunha Bustamante, Laerte Ferreira Guimarães, Andreas Rabe, and Patrick Hostert. 2016. “Mapping Brazilian Savanna Vegetation Gradients with Landsat Time Series.” *International Journal of Applied Earth Observations and Geoinformation* 52. Elsevier B.V.: 361–70. doi:10.1016/j.jag.2016.06.019.

Scolforo, Henrique Ferraco, Jose Roberto Soares Scolforo, Jose Marcio de Mello, Carlos Rogerio de Mello, and Vinicius Augusto Morais. 2016. “Spatial Interpolators for Improving the Mapping of Carbon Stock of the Arboreal Vegetation in Brazilian Biomes of Atlantic Forest and Savanna.” *Forest Ecology and Management* 376 (November): 24–35. doi:10.1016/j.foreco.2016.05.047.

Scolforo, Henrique Ferraco, Jose Roberto Soares Scolforo, Carlos Rogerio Mello, Jose Marcio Mello, and Antonio Carlos Ferraz Filho. 2015. “Spatial Distribution of Aboveground Carbon Stock of the Arboreal Vegetation in Brazilian Biomes of Savanna, Atlantic Forest and Semi-Arid Woodland.” *PLoS ONE* 10 (6): 1–20. doi:10.1371/journal.pone.0128781.

SCOLFORO, J. R., J. M. MELLO, A. D. OLIVEIRA, R.M. PEREIRA, F.N. SOUZA, AND I. C. L. GUEDES. 2008. Volumetria, peso de materia seca e carbono. In J. R. Scolforo, J. M. Mello, and A. D. Oliveira (Eds.). *Inventario Florestal de Minas Gerais: Cerrado - Floristica, Estrutura, Diversidade, Similaridade, Distribuicao Diametrica e de Altura, Volumetria, Tendencias de Crescimento e Areas Aptas para Manejo Florestal*, pp. 361–438. Lavras, UFLA

Su, Yanjun, Qinghua Guo, Baolin Xue, Tianyu Hu, Otto Alvarez, Shengli

Tao, and Jingyun Fang. 2016. “Spatial Distribution of Forest Aboveground Biomass in China: Estimation through Combination of Spaceborne Lidar, Optical Imagery, and Forest Inventory Data.” *Remote Sensing of Environment* 173. Elsevier Inc.: 187–99. doi:10.1016/j.rse.2015.12.002.

Veldman, Joseph W., Elise Buisson, Giselda Durigan, G. Wilson Fernandes, Soizig Le Stradic, Gregory Mahy, Daniel Negreiros, et al. 2015. “Toward an Old-Growth Concept for Grasslands, Savannas, and Woodlands.” *Frontiers in Ecology and the Environment* 13 (3): 154–62. doi:10.1890/140270.

Viana, H., J. Aranha, D. Lopes, and Warren B. Cohen. 2012. “Estimation of Crown Biomass of Pinus Pinaster Stands and Shrubland above-Ground Biomass Using Forest Inventory Data, Remotely Sensed Imagery and Spatial Prediction Models.” *Ecological Modelling* 226. Elsevier B.V.: 22–35. doi:10.1016/j.ecolmodel.2011.11.027.

Vieilledent, Ghislain, Oliver Gardi, Clovis Grinand, Christian Burren, Christian Camara, Charlie J Gardner, and Leah Glass. 2016. “Bioclimatic Envelope Models Predict a Decrease in Tropical Forest Carbon Stocks with Climate Change in Madagascar,” 703–15. doi:10.1111/1365-2745.12548.

Wilson, Barry Tyler, Christopher W Woodall, and Douglas M Griffith. 2013. “Imputing Forest Carbon Stock Estimates from Inventory Plots to a Nationally Continuous Coverage.” *Carbon Balance and Management* 8 (1): 1. doi:10.1186/1750-0680-8-1.

Yin, Guodong, Yuan Zhang, Yan Sun, Tao Wang, Zhenzhong Zeng, and Shilong Piao. 2015. “MODIS Based Estimation of Forest Aboveground Biomass in China.” *PLoS ONE* 10 (6): 1–13. doi:10.1371/journal.pone.0130143.

Young, Nicholas E., Ryan S. Anderson, Stephen M. Chignell, Anthony G. Vorster, Rick Lawrence, and Paul H. Evangelista. 2017. “A Survival Guide to Landsat Preprocessing.” *Ecology* 98 (4): 920–32. doi:10.1002/ecy.1730.

Zhao, Panpan, Dengsheng Lu, Guangxing Wang, Chuping Wu, Yujie

Huang, and Shuquan Yu. 2016. "Examining Spectral Reflectance Saturation in Landsat Imagery and Corresponding Solutions to Improve Forest Aboveground Biomass Estimation." *Remote Sensing* 8 (6). doi:10.3390/rs8060469.

Zheng, Daolan, John Rademacher, Jiquan Chen, Thomas Crow, Mary Bresee, James Le Moine, and Soung Ryoul Ryu. 2004. "Estimating Aboveground Biomass Using Landsat 7 ETM+ Data across a Managed Landscape in Northern Wisconsin, USA." *Remote Sensing of Environment* 93 (3): 402–11. doi:10.1016/j.rse.2004.08.008.

Zhou, Xudong, Xinkai Zhu, Zhaodi Dong, and Wenshan Guo. 2016. "ScienceDirect Estimation of Biomass in Wheat Using Random Forest Regression Algorithm and Remote Sensing Data." *CJ. Elsevier B.V.*, 1–8. doi:10.1016/j.cj.2016.01.008.

Zhu, Xiaolin, and Desheng Liu. 2015. "Improving Forest Aboveground Biomass Estimation Using Seasonal Landsat NDVI Time-Series." *ISPRS Journal of Photogrammetry and Remote Sensing* 102. International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS): 222–31. doi:10.1016/j.isprsjprs.2014.08.014.